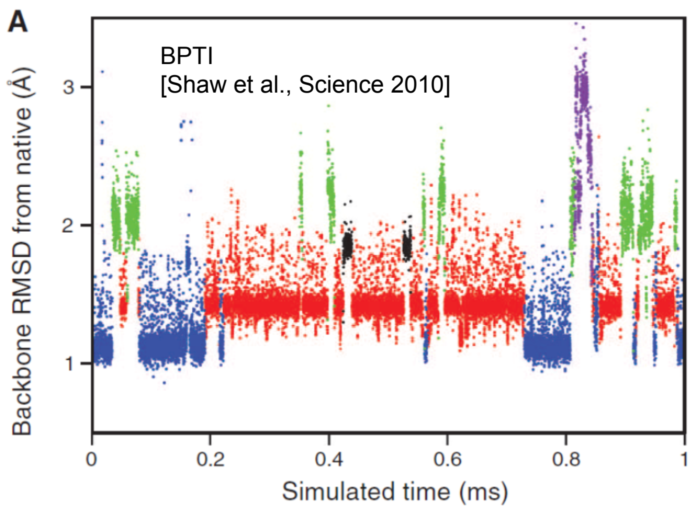
Physics of trajectories and the weighted ensemble method

Daniel M. Zuckerman

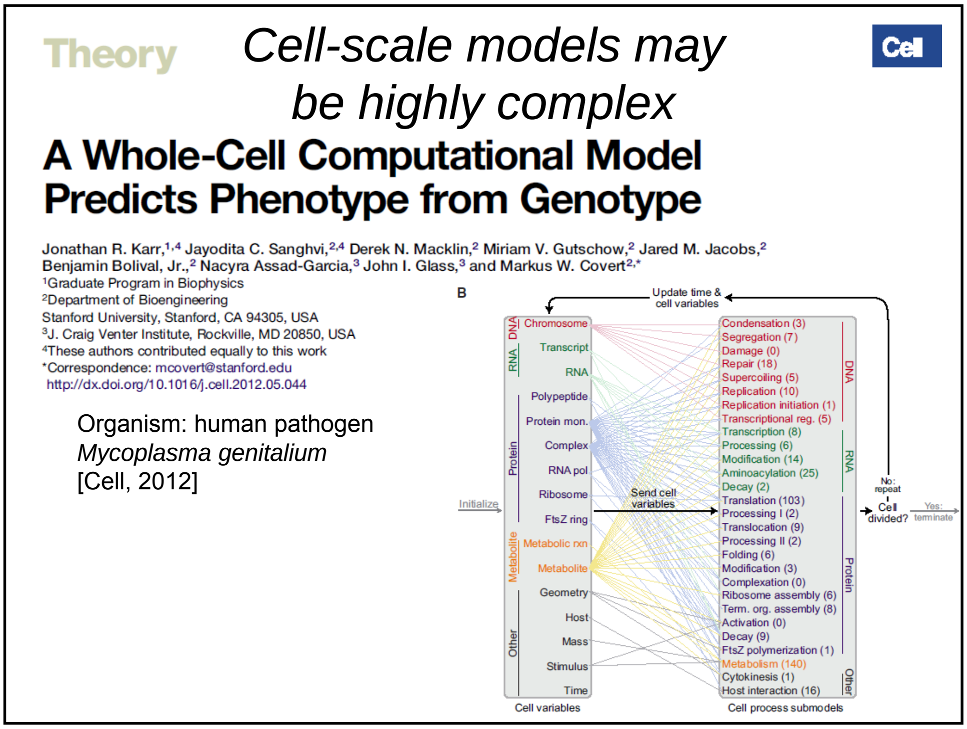
Department of Computational & Systems Biology

University of Pittsburgh School of Medicine

Molecular sampling is difficult!



BPTI
[Shaw et al., Science 2010]



Theory

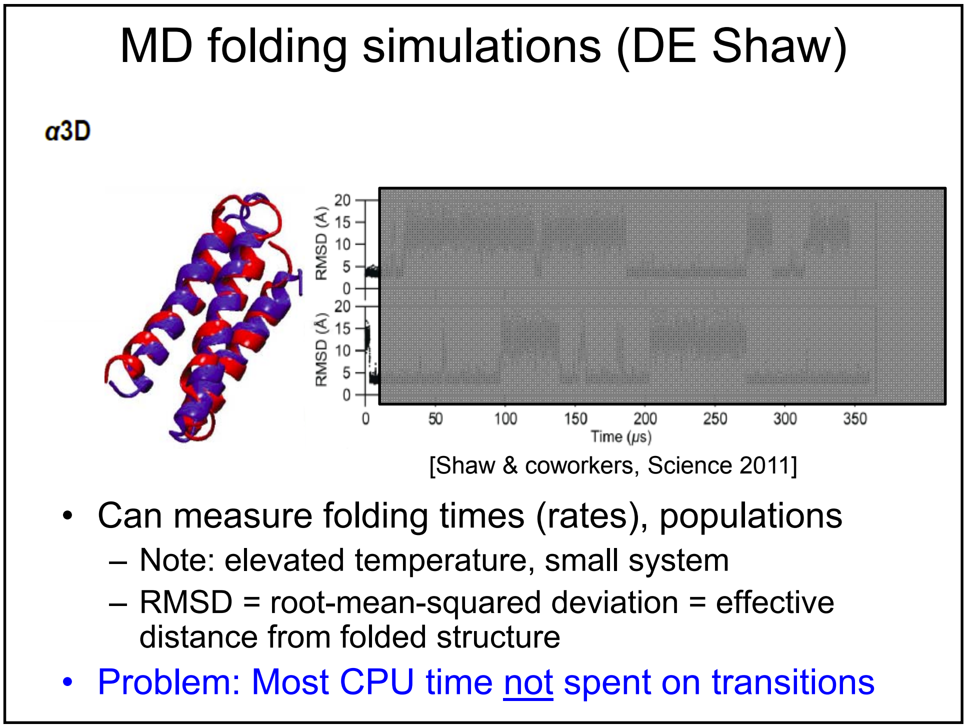
Cell-scale models may be highly complex

Cell

A Whole-Cell Computational Model Predicts Phenotype from Genotype

Jonathan R. Karr,[1,4] Jayodita C. Sanghvi,[2,4] Derek N. Macklin,[2] Miriam V. Gutschow,[2] Jared M. Jacobs,[2] Benjamin Bolival, Jr.,[2] Nacyra Assad-Garcia,[3] John I. Glass,[3] and Markus W. Covert[2,*]
[1]Graduate Program in Biophysics
[2]Department of Bioengineering
Stanford University, Stanford, CA 94305, USA
[3]J. Craig Venter Institute, Rockville, MD 20850, USA
[4]These authors contributed equally to this work
*Correspondence: mcovert@stanford.edu
http://dx.doi.org/10.1016/j.cell.2012.05.044

Organism: human pathogen *Mycoplasma genitalium* [Cell, 2012]



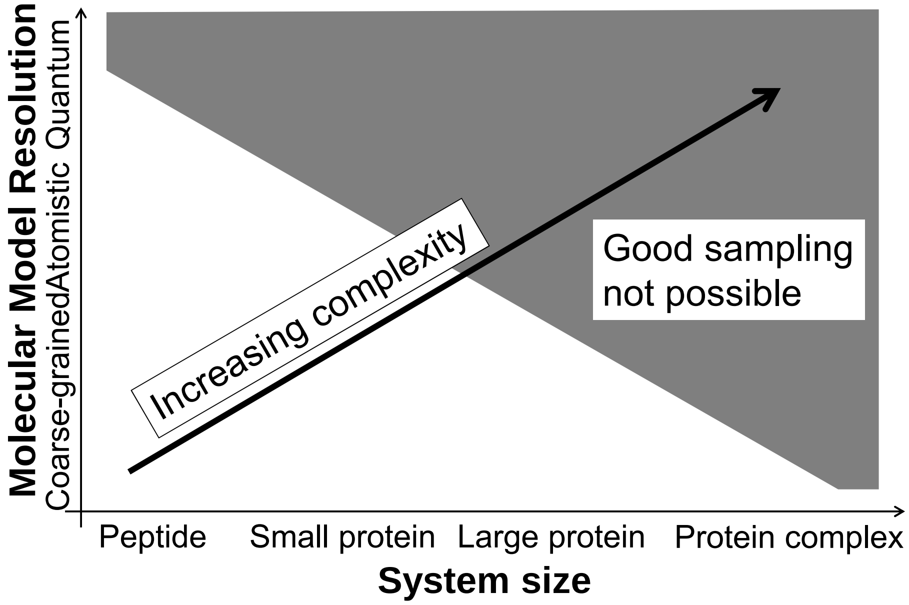
MD folding simulations (DE Shaw)

α3D

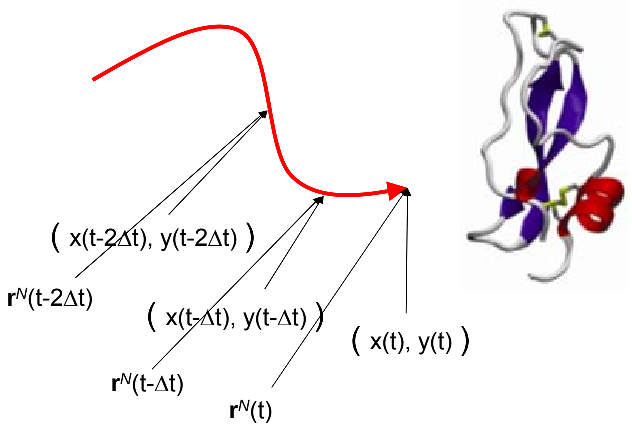
[Shaw & coworkers, Science 2011]

- Can measure folding times (rates), populations
  - Note: elevated temperature, small system
  - RMSD = root-mean-squared deviation = effective distance from folded structure
- Problem: Most CPU time not spent on transitions

MD sampling limitations

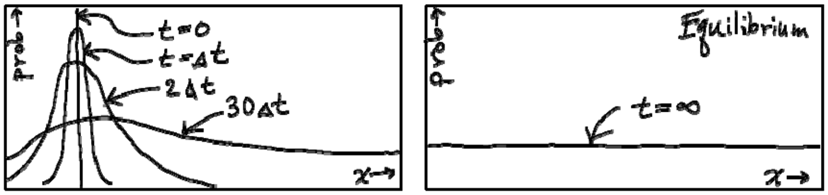


To improve on MD, we must understand trajectories

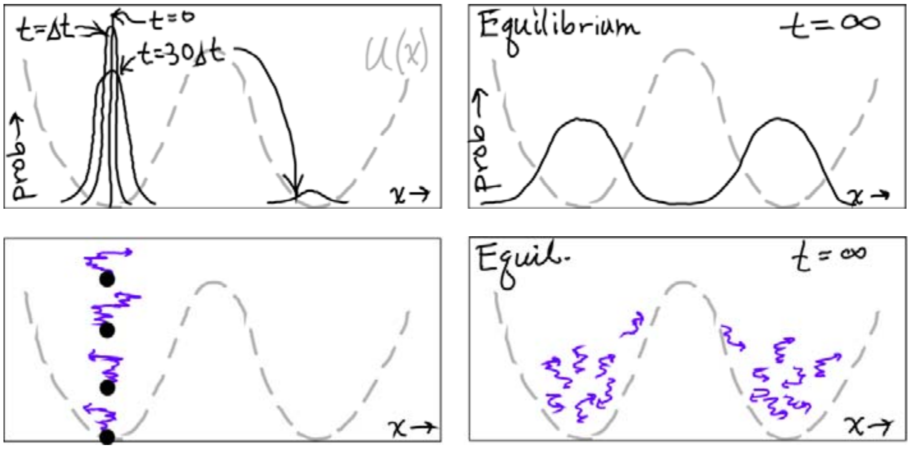


$\mathbf{r}^N = \{ (x_1, y_1, z_1), (x_2, y_2, z_2), \ldots (x_N, y_N, z_N) \}$

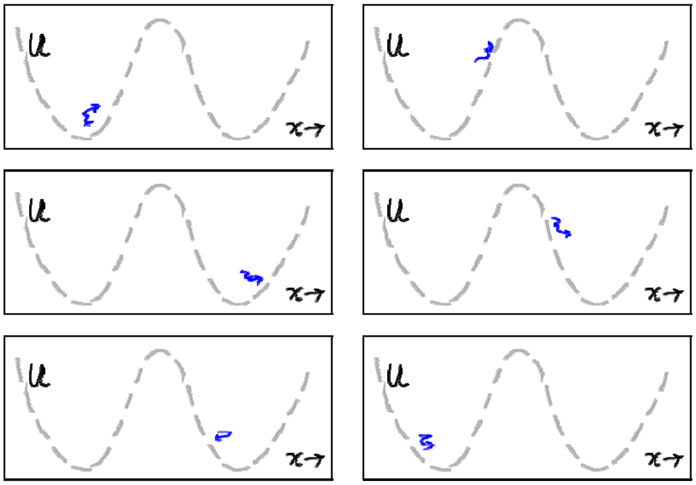
Representing diffusion with an ensemble of trajectories (Motion in real space)



Trajectory ensemble – activated process (Configuration space)



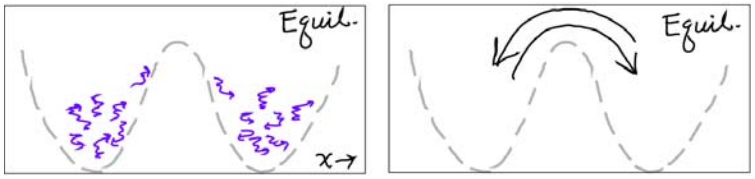
Trajectory ensembles of independent systems



Equilibrium ensemble

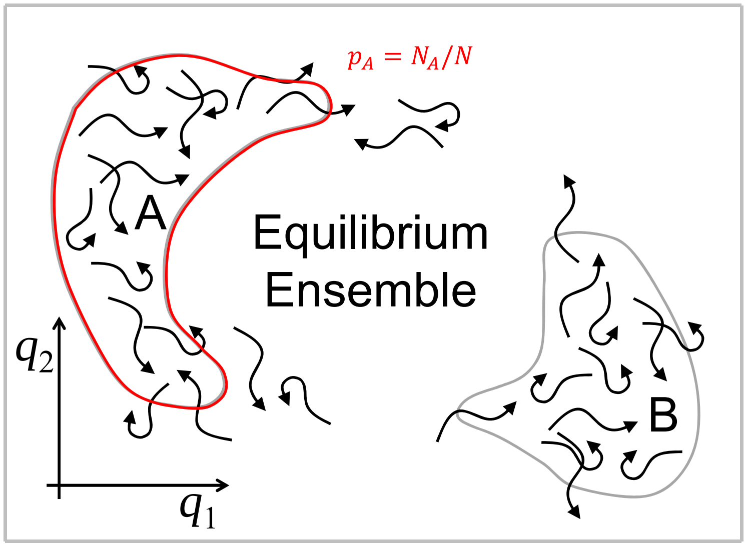
[See DM Zuckerman, *Statistical Physics of Biomolecules,* Ch 11]

Three key trajectory ensembles



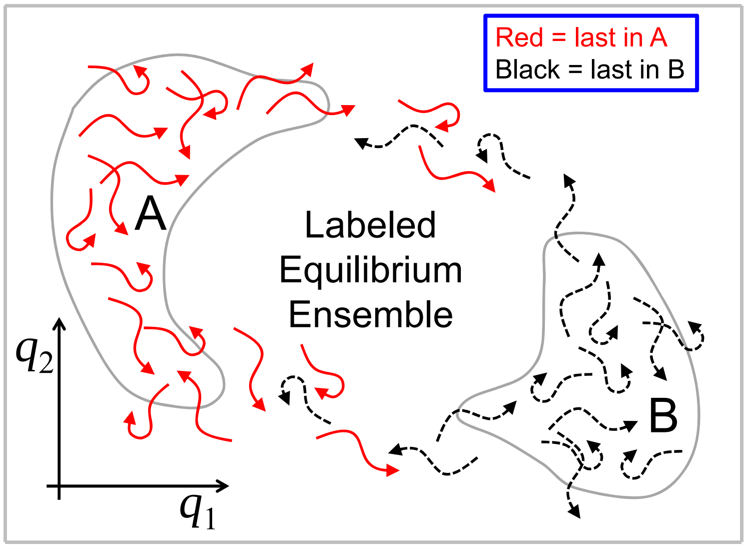
[See DM Zuckerman, *Statistical Physics of Biomolecules,* Ch 11]

Trajectory ensemble in two dimensions (independent systems)



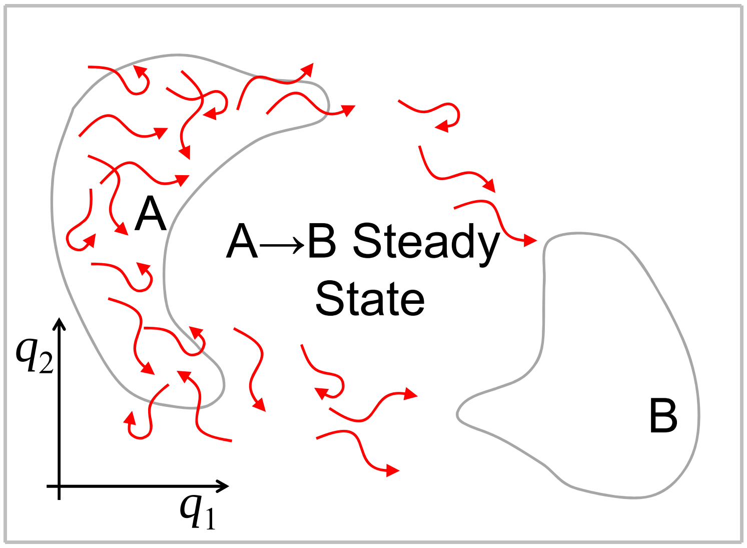
A large number of *independent* systems undergoing natural dynamics - constant conditions

The most important picture in non-equilibrium statistical mechanics?



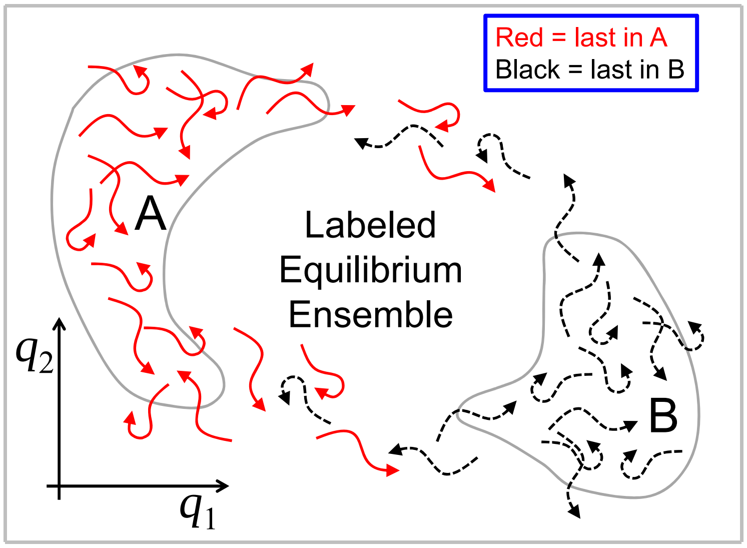
[Divide equil into SS: Bhatt & Zuckerman, *J Chem Theory Comp* 2011
cf. Vanden Eijnden & Venturoli, JCP 2009; Dinner & co, JCP 2009; Bolhuis & co, JCP 2003]

The path ensemble is the mechanism



A→B Steady State

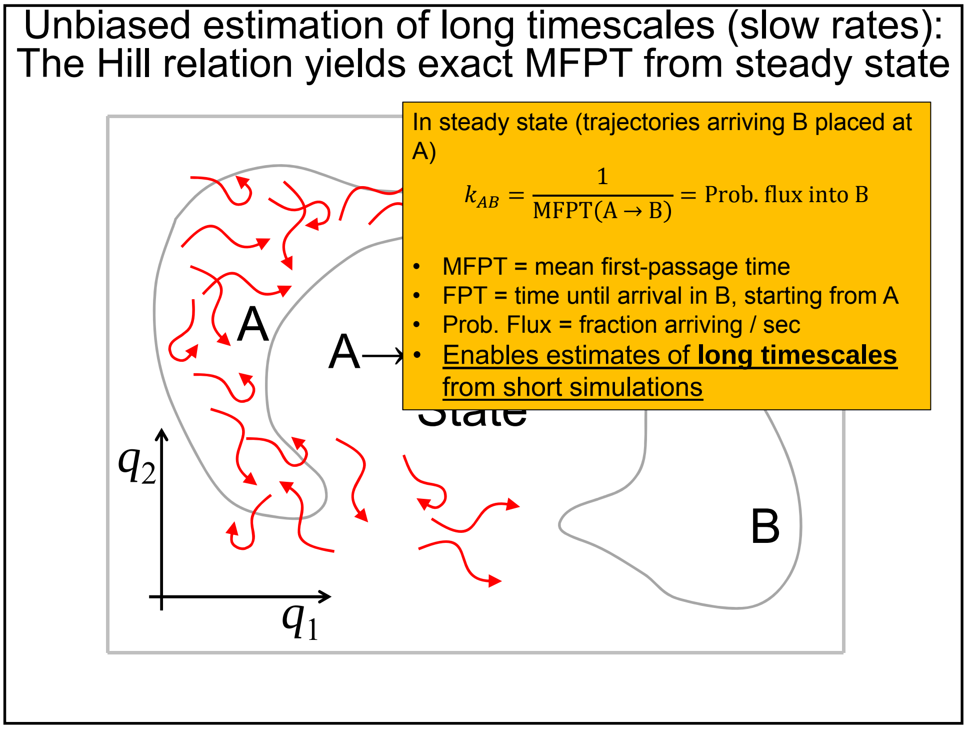
Macroscopic vs. microscopic reversibility



Red = last in A
Black = last in B

Labeled Equilibrium Ensemble

[Divide equil into SS: Bhatt & Zuckerman, *J Chem Theory Comp* 2011
cf. Vanden Eijnden & Venturoli, JCP 2009; Dinner & co, JCP 2009; Bolhuis & co, JCP 2003]

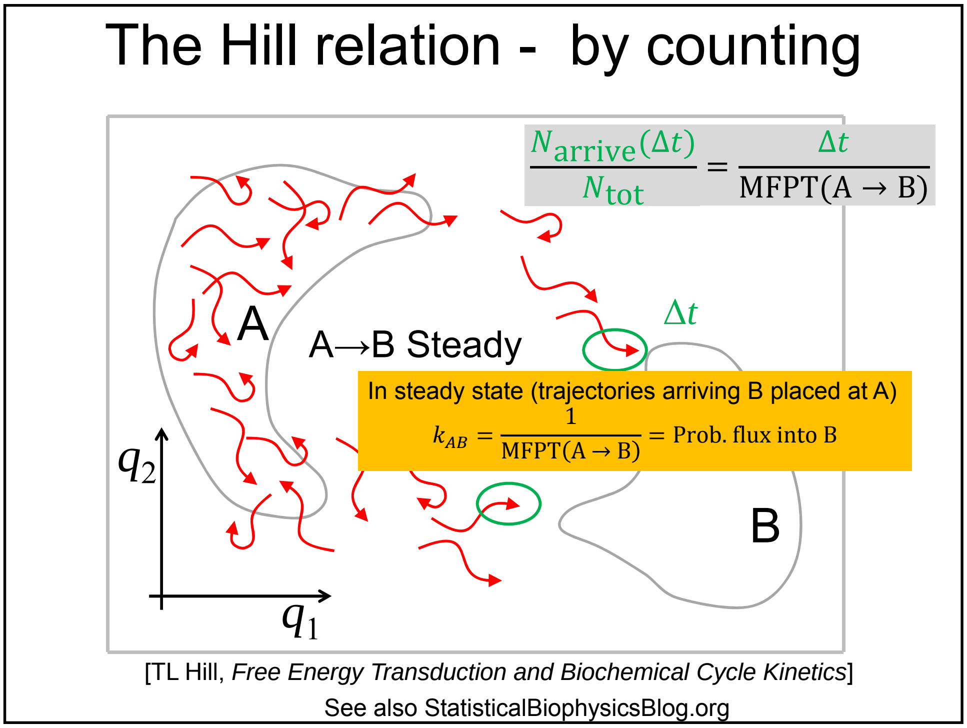


Unbiased estimation of long timescales (slow rates): The Hill relation yields exact MFPT from steady state

In steady state (trajectories arriving B placed at A)

$$k_{AB} = \frac{1}{\mathrm{MFPT(A \rightarrow B)}} = \text{Prob. flux into B}$$

- MFPT = mean first-passage time
- FPT = time until arrival in B, starting from A
- Prob. Flux = fraction arriving / sec
- <u>Enables estimates of **long timescales** from short simulations</u>



The Hill relation - by counting

$$\frac{N_{\mathrm{arrive}}(\Delta t)}{N_{\mathrm{tot}}} = \frac{\Delta t}{\mathrm{MFPT(A \rightarrow B)}}$$

In steady state (trajectories arriving B placed at A)

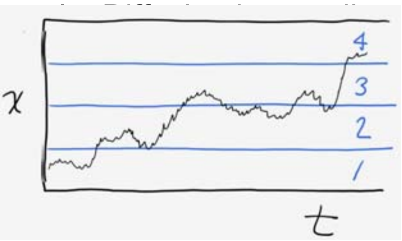
$$k_{AB} = \frac{1}{\mathrm{MFPT(A \rightarrow B)}} = \text{Prob. flux into B}$$

[TL Hill, *Free Energy Transduction and Biochemical Cycle Kinetics*]

See also StatisticalBiophysicsBlog.org

Markovian behavior – beware of assumptions!

- Markovian ⇔ Distribution of future outcomes depends only on present.
- *Everything is Markovian:* In molecular physics, current positions and velocities of all atoms fully determine future distribution of possible outcomes.
- *Nothing is Markovian:* Once a Markovian system is discretized (or projected onto a subset of coordinates), behavior in the reduced space is no longer Markovian
  - – Ex nsion



A discrete index is blind to whether the underlying continuous trajectory is near the upper or lower boundary.

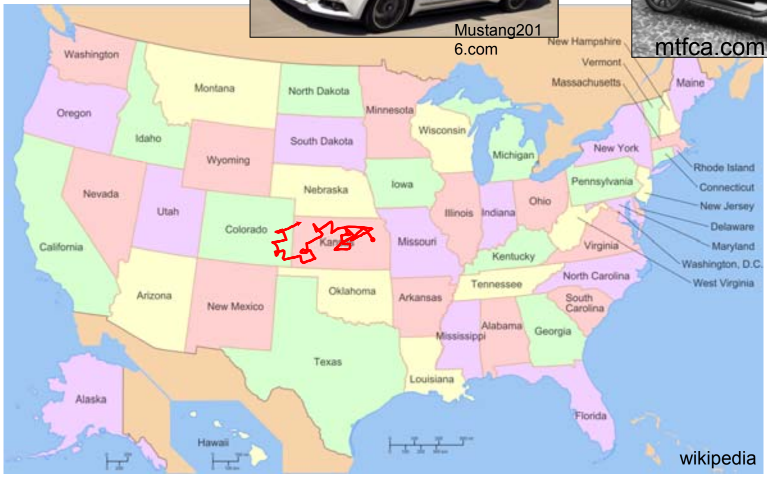
- Example: 3 → 4 transition, after coming from 2

StatisticalBiophysicsBlog.org

Random Driving: Non-Markovian in state-space

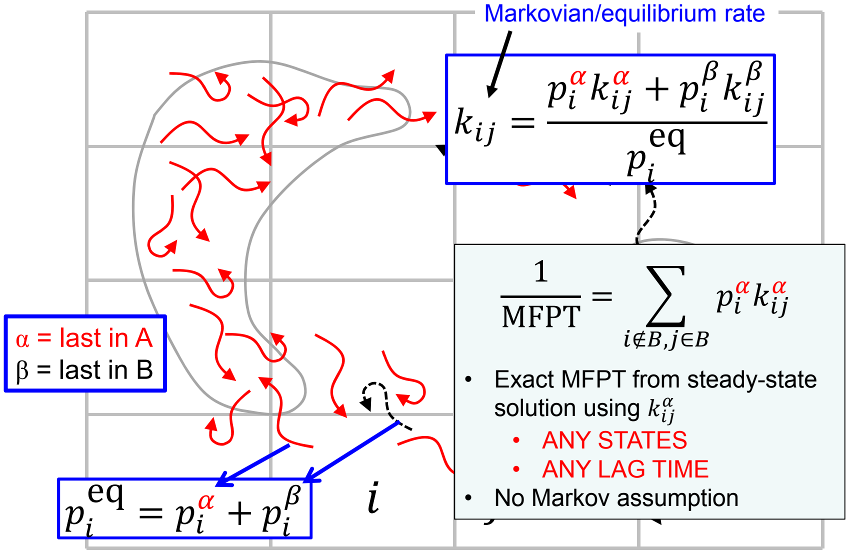
Mustang2016.com

mtfca.com



wikipedia

α/β labeling: Minimal history for kinetics



Markovian/equilibrium rate

$$k_{ij} = \frac{p_i^{\alpha} k_{ij}^{\alpha} + p_i^{\beta} k_{ij}^{\beta}}{p_i^{\text{eq}}}$$

α = last in A
β = last in B

$$p_i^{\text{eq}} = p_i^{\alpha} + p_i^{\beta}$$

$$\frac{1}{\text{MFPT}} = \sum_{i \notin B, j \in B} p_i^{\alpha} k_{ij}^{\alpha}$$

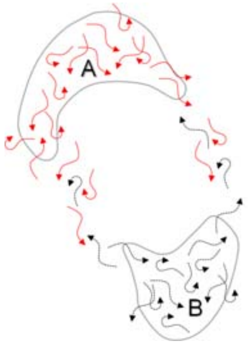
- Exact MFPT from steady-state solution using k_{ij}^{α}
  - ANY STATES
  - ANY LAG TIME
- No Markov assumption

[Suarez, Lettieri, … Zuckerman, JCTC, 2014]

Interim Summary

Essentials of trajectory physics

1. Equilibrium ensemble decomposed exactly into red (A→B) and black (B→A) steady states
2. Ensemble defines mechanism
3. MFPT calculated exactly from probability flux in steady state (Hill relation)
4. To analyze continuous trajectories in a reduced/discrete space, Markovian behavior cannot be assumed.



- The most important picture in non-equilibrium statistical mechanics?
  - Powerful lessons from simple principles

Equilibrium ensemble → Path ensemble

$$\boxed{\rho(\mathbf{r}^N) \propto e^{-U(\mathbf{r}^N)/k_B T}}$$

$$\mathbf{r}^N = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$$

whole molecule

individual atoms

$$\mathbf{r}^{\text{traj}} = \left\{\mathbf{r}^N(t=0),\ \mathbf{r}^N(\Delta t),\ \mathbf{r}^N(2\Delta t), \ldots\right\}$$

$$\equiv \left\{\mathbf{r}_0^N,\ \mathbf{r}_1^N,\ \mathbf{r}_2^N, \ldots\right\}.$$

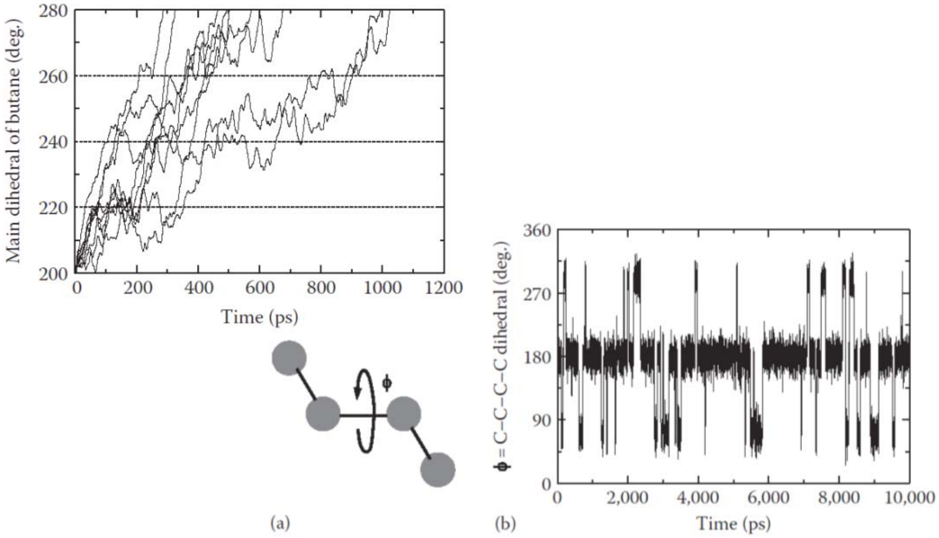
sequence of whole-system configurations

$$\boxed{\text{prob}\left(\mathbf{r}^{\text{traj}}\right) \propto \exp\left[-E^{\text{traj}}\left(\mathbf{r}^{\text{traj}}\right)/k_B T\right]}$$

[DM Zuckerman, *Statistical Physics of Biomolecules,* Ch 11]

One more: The transition-path ensemble

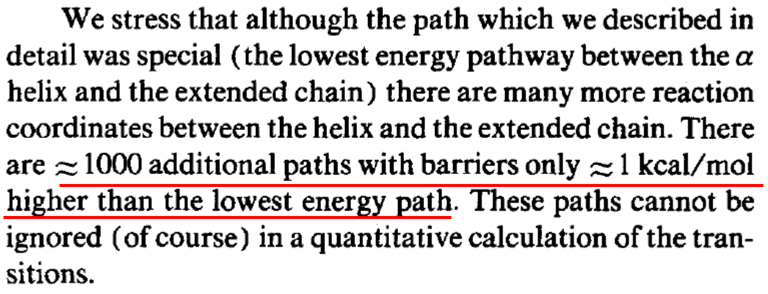
- Butane trans-to-gauche transitions



[DM Zuckerman, *Statistical Physics of Biomolecules,* Ch 11]

Tetra-alanine

- For a "tetra-alanine" (four peptide planes) …



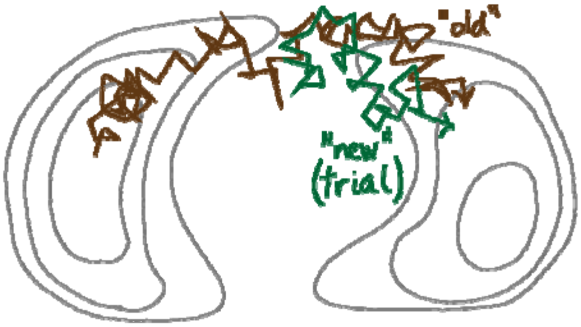
We stress that although the path which we described in detail was special (the lowest energy pathway between the α helix and the extended chain) there are many more reaction coordinates between the helix and the extended chain. There are ≈ 1000 additional paths with barriers only ≈ 1 kcal/mol higher than the lowest energy path. These paths cannot be ignored (of course) in a quantitative calculation of the transitions.

[Czerminksi & Elber, J Chem Phys, 1990]

Transition path ensemble: Intrinsic costs

- N_{ind} = number of independent paths desired
  - – Likely $10 \leq N_{\text{ind}} \leq 100$
- t_b = typical time for event
  - – Does not include dwell time in initial state
  - – Could include intermediate dwells, depending on context
- $N_{\text{ind}}\, t_b$ = minimum computation cost
  - – Minimum obtained when no correlated paths generated, and all paths are properly distributed (apparently impossible)

Transition Path Sampling: Monte Carlo in Path Space

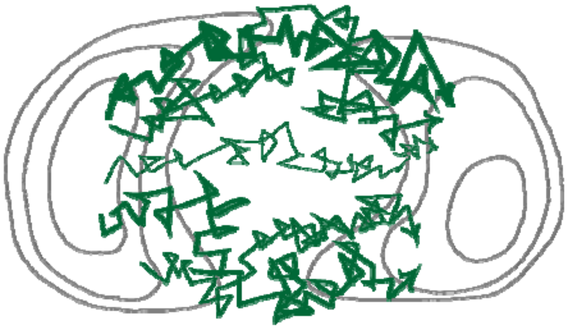


Developers:
- Pratt
- Chandler
- Bolhuis, van Erp

Basic Idea
- Trial trajectory generated from previous trajectory in ensemble
- Accept/reject via Metropolis criterion

Comments
- Connection to quantum path-integral methods
- Chance of trapping (like all Metropolis MC)
- Difficult to calculate rates – spurred improved variations
- Metastable intermediates lead to long trajectories – requires special treatment

Dynamic Importance Sampling: Reweighting in path space

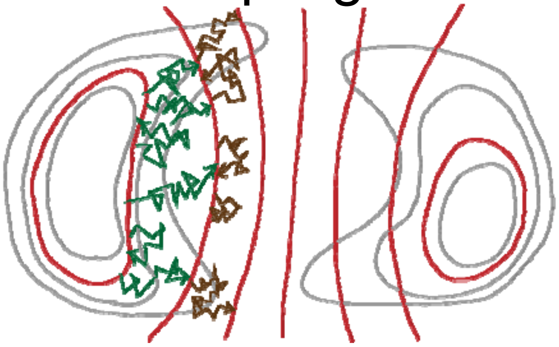


Developers:
- Woolf
- Zuckerman

Basic Idea
- Generate (biased) ensemble of trajectories
- Reweight using ratio of sampled to true probability

Comments
- Easily captures path diversity
- Difficult to have overlap between sampled and true ensemble (like all reweighting in high dimensions)

Milestoning, Forward-flux Sampling: Path sampling between interfaces

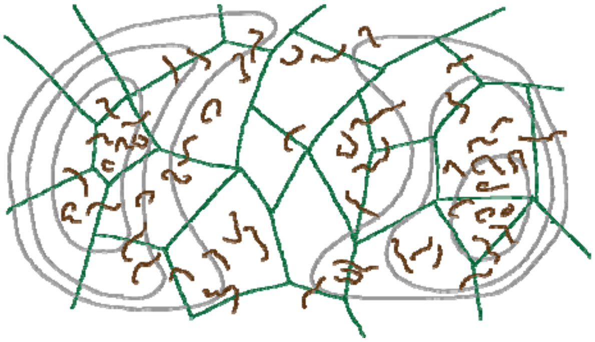


Developers:
- Elber
- ten Wolde, Allen

Basic Idea
- Set up interfaces interpolating between initial and final states
- Collect statistics on short trajectories initiated at interfaces

Comments
- Rigorous formulations possible; sometimes Markov assumption used
- Must "catch" trajectories as they cross boundaries
- Related to Transition Interface Sampling [Bolhuis, van Erp] and Non-equilibrium Umbrella Sampling [Dinner]

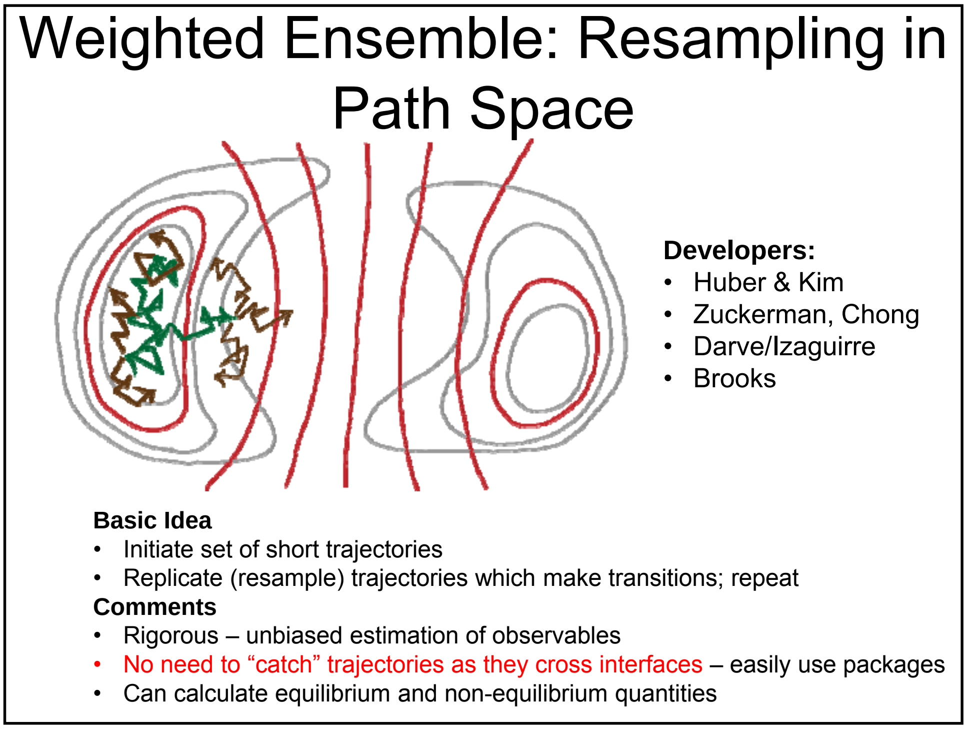
Markov State Modeling: A variation on interface methods



Developers:
- Bahar, Dill
- Pande
- Noe

Basic Idea
- Collect trajectories distributed in configuration space, possibly brief
- Decompose space and estimate transition probabilities, long timescales

Comments
- Literature reports use significant trajectory data for nearly Markovian behavior
- Non-trivial to generate optimal division of space



Weighted Ensemble: Resampling in Path Space

Developers:

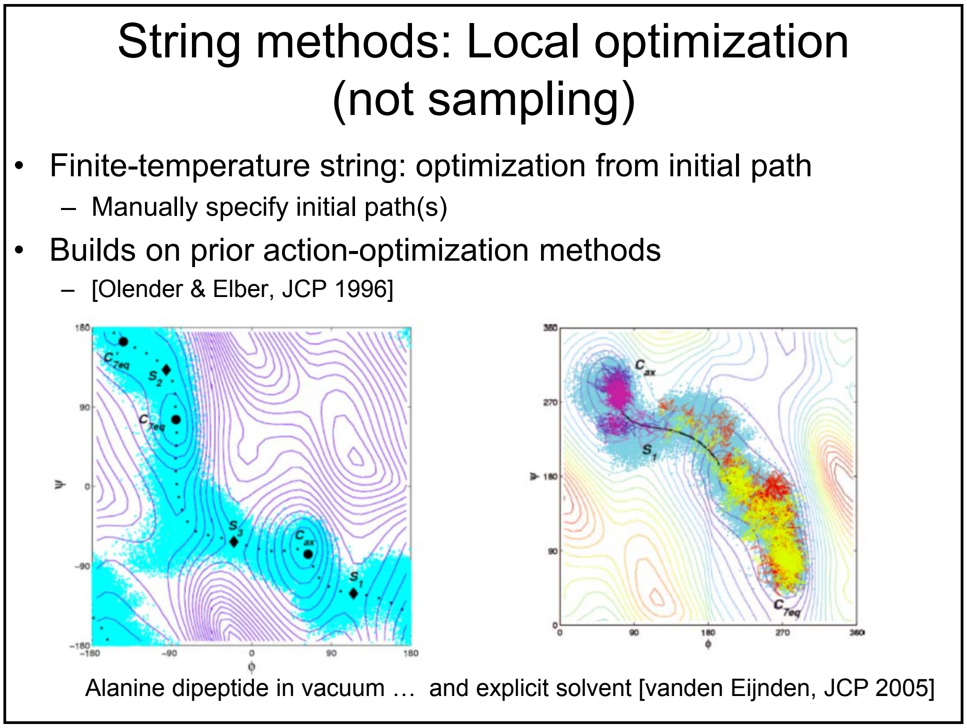
- Huber & Kim
- Zuckerman, Chong
- Darve/Izaguirre
- Brooks

Basic Idea

- Initiate set of short trajectories
- Replicate (resample) trajectories which make transitions; repeat

Comments

- Rigorous – unbiased estimation of observables
- No need to “catch” trajectories as they cross interfaces – easily use packages
- Can calculate equilibrium and non-equilibrium quantities



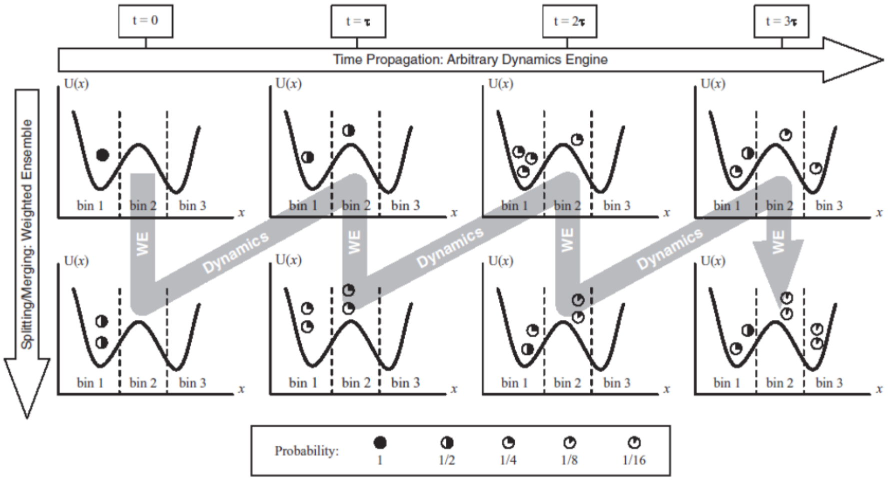
String methods: Local optimization (not sampling)

- Finite-temperature string: optimization from initial path
  - Manually specify initial path(s)
- Builds on prior action-optimization methods
  - [Olender & Elber, JCP 1996]

C'7eq
S2
C7eq
S3
Cax
S1
ψ
φ
180
90
0
-90
-180
Cax
S1
C7eq
ψ
φ
0
90
180
270
360

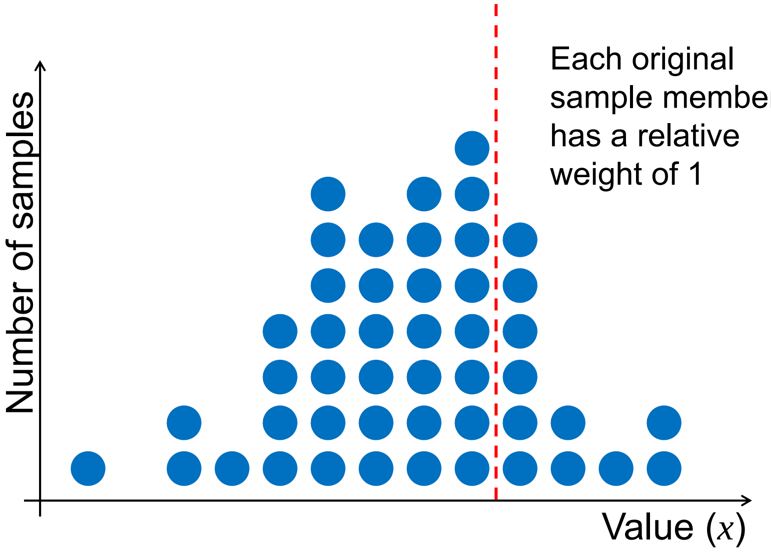
Alanine dipeptide in vacuum … and explicit solvent [vanden Eijnden, JCP 2005]

Weighted Ensemble – original algorithm


t = 0
t = τ
t = 2τ
t = 3τ
Time Propagation: Arbitrary Dynamics Engine
Splitting/Merging: Weighted Ensemble
U(x)
bin 1
bin 2
bin 3
x
WE
Dynamics
Probability:
1
1/2
1/4
1/8
1/16

[Original Weighted Ensemble: Huber & Kim, *Biophys J.* 1996;
Figure from Donvan et al., J Chem Phys 2013]

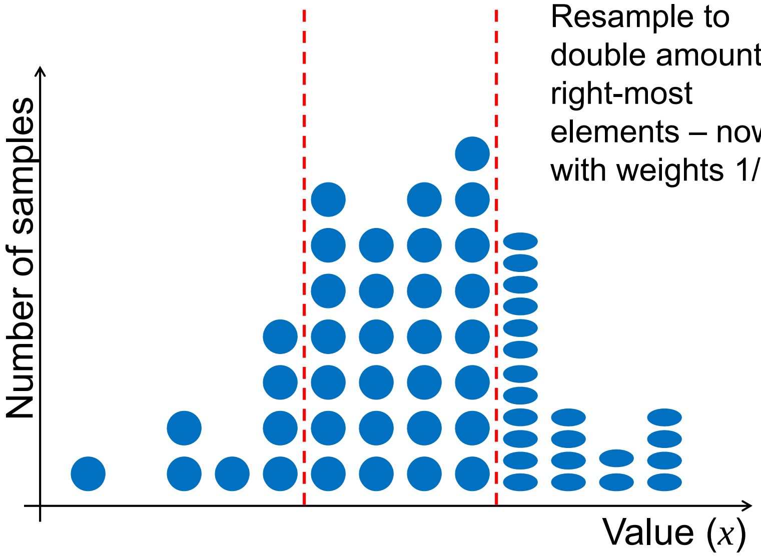
WE is based on resampling



Each original sample member has a relative weight of 1

[See Zhang, Jasnow, Zuckerman, J. Chem. Phys. 2010]

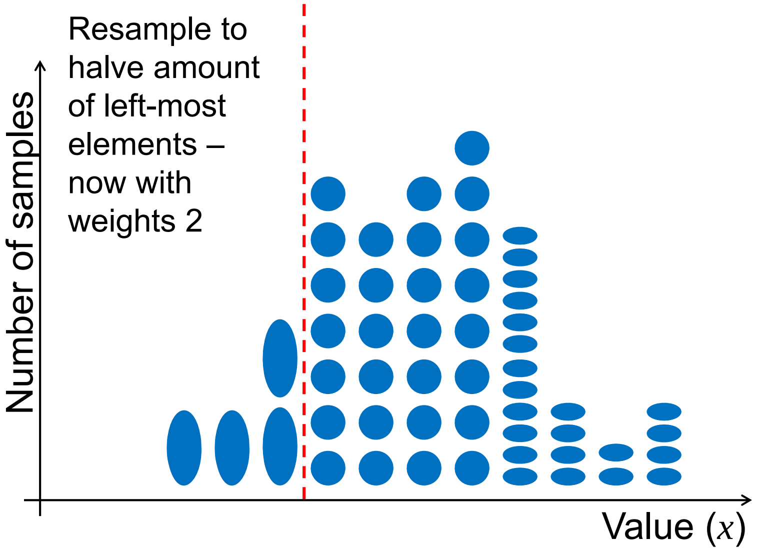
Resampling



Resample to double amount of right-most elements – now with weights 1/2

[See Zhang, Jasnow, Zuckerman, J. Chem. Phys. 2010]

Resampling



Resample to halve amount of left-most elements – now with weights 2

[See Zhang, Jasnow, Zuckerman, J. Chem. Phys. 2010]

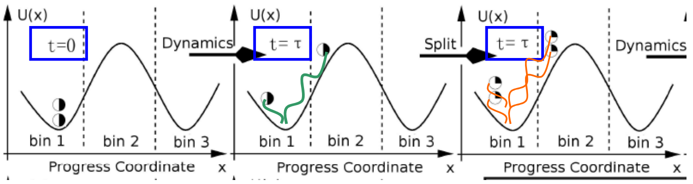
WE is resampling, in trajectory space

- Trajectories are objects in high dimensional space with well-defined distribution (see Zuckerman, *Statistical Physics of Biomolecules: An Introduction*)

$$\mathbf{x}^{\mathrm{traj}} = \{\mathbf{x}(t=0), \mathbf{x}(t=\delta t), \mathbf{x}(t=2\delta t), \ldots\}$$

  - WE starts from a correct path ensemble (multiple ordinary simulations)
  - All paths continuous & dynamical throughout
  - Occasional resampling in path space using splitting and combining
  - Probabilistic resampling: no assumption of equilibrium in bins
  - Correct for non-Markovian dynamics because history is included in resampling

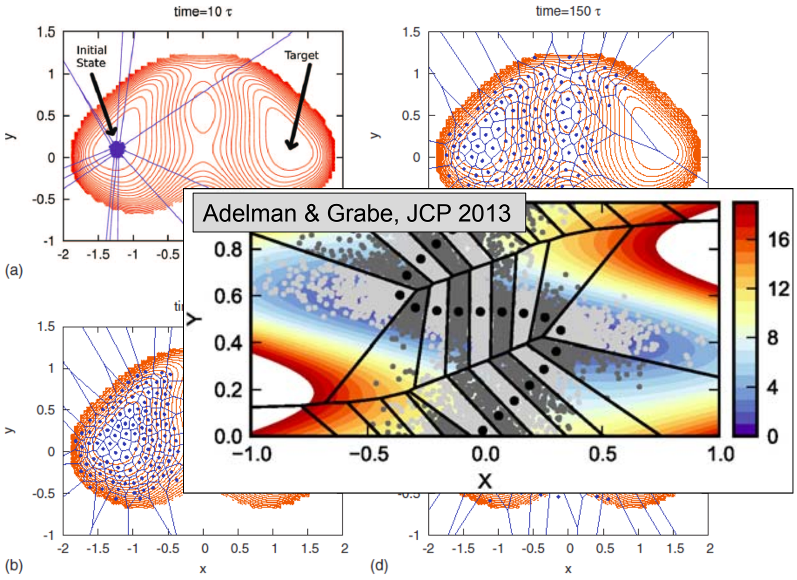
[Path integral formulation in Zhang, Jasnow, Zuckerman, *J. Chem. Phys.* 2010]



Limitations of WE

- Fundamental limitations:
  1. Orthogonal coordinates (which are uncorrelated with binned coordinates) must be sampled by “brute force” [Note: also true for other methods]
  2. Correlations result from splitting/merging [Note: other methods also yield path correlations]
  3. Not every observable can be sampled more efficiently – primarily slow coordinates improved
- Not required in WE:
  - Advance knowledge of slow coordinates
  - Static bins
  - Uniform bins
  - Bins themselves

Automated Voronoi binning

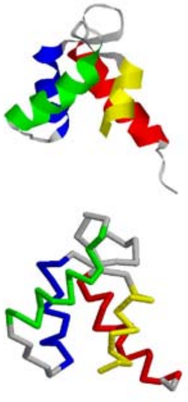


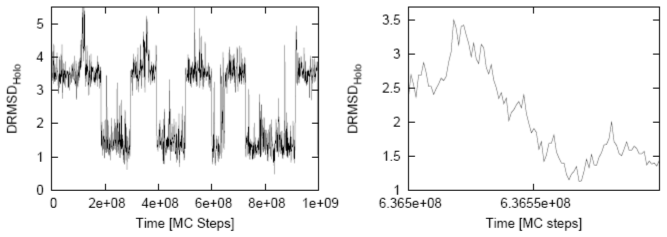
Adelman & Grabe, JCP 2013

[Zhang, Jasnow, Zuckerman, JCP 2010]

Validation of original WE

- Original WE algorithm [Huber and Kim]
- To check, we developed a verifiable system: dual-basin Gō model using alpha carbons



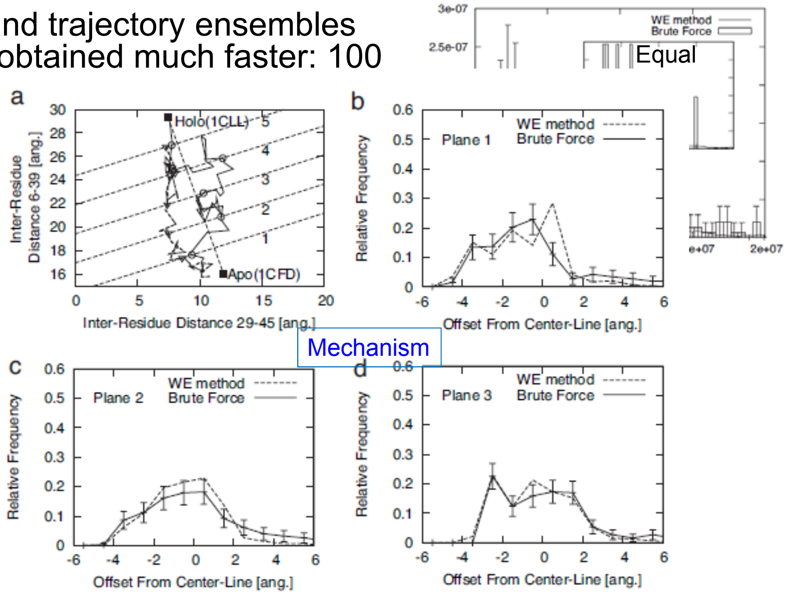


example brute-force trajectory, ~1 wk single CPU

[Zuckerman, *J. Phys. Chem. B*, 2004]

WE is correct & can be very efficient

- Rates and trajectory ensembles can be obtained much faster: 100 times
  - Cal
  - Exc forc
  - C-a [Zh *PN*

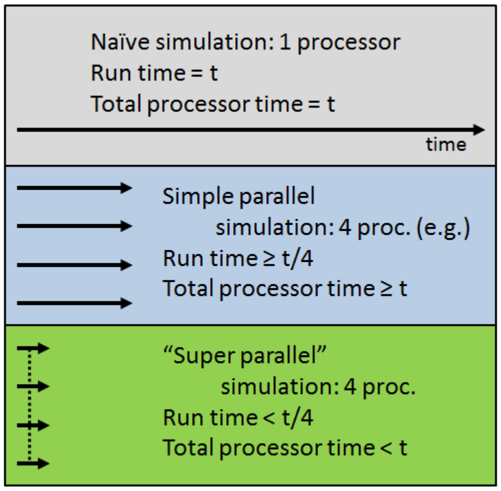


Equal

Mechanism

Super-Parallelism

Schematic of time for estimating observables to targeted precision

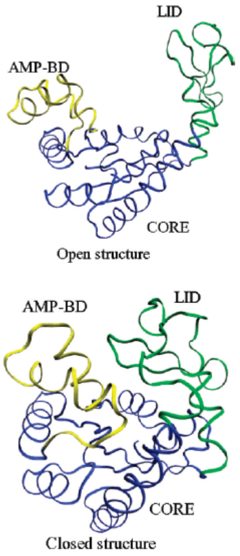


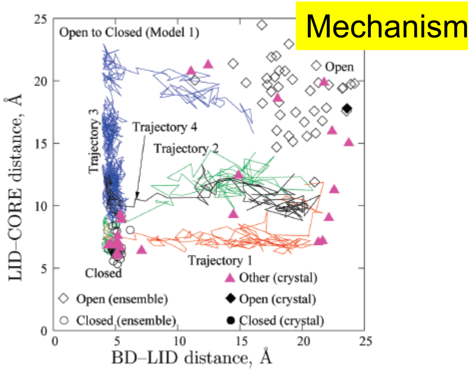
Naïve simulation: 1 processor
Run time = t
Total processor time = t

Simple parallel simulation: 4 proc. (e.g.)
Run time ≥ t/4
Total processor time ≥ t

"Super parallel" simulation: 4 proc.
Run time < t/4
Total processor time < t

Semi-atomistic model: Adenylate kinase



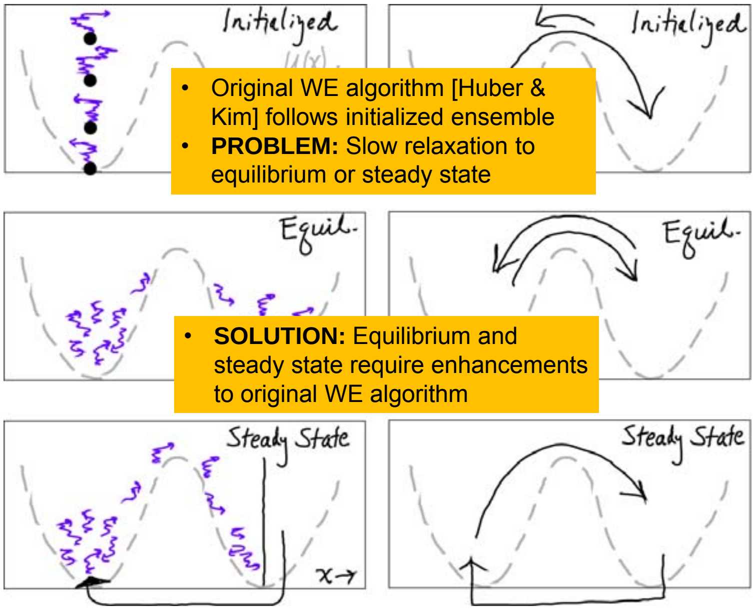


Adenylate kinase via semi-atomistic double-Go model & LMBC

- Brute force: 4 years for a single transition (one processor)
- Weighted ensemble: ~50 indep. transitions in 2 wks (one processor)

[Bhatt & Zuckerman, *J. Chem. Theory Comp.*, 2010]

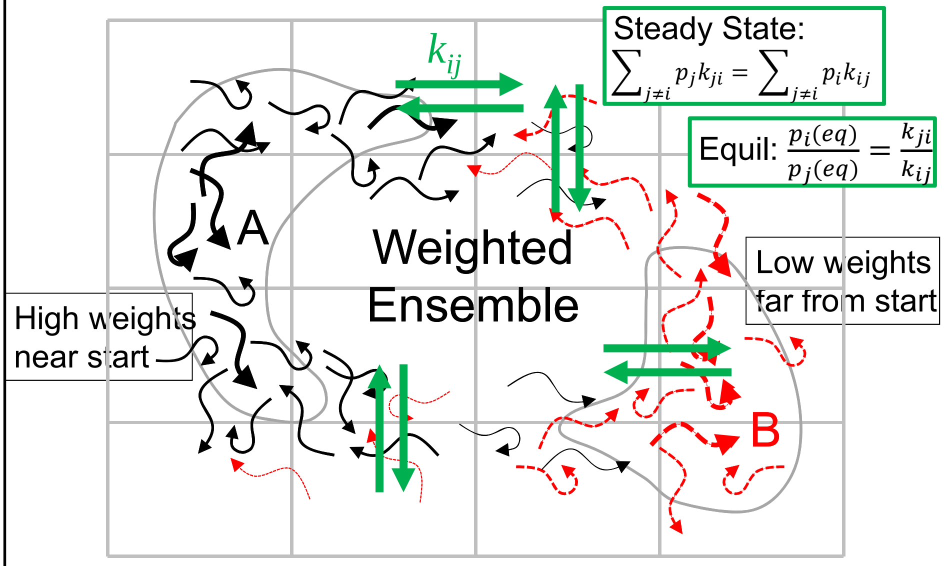
Three key trajectory ensembles



- Original WE algorithm [Huber & Kim] follows initialized ensemble
- **PROBLEM:** Slow relaxation to equilibrium or steady state

- **SOLUTION:** Equilibrium and steady state require enhancements to original WE algorithm

Extending WE to Equilibrium and Non-equilibrium Steady-State



Steady State: $\sum_{j\neq i} p_j k_{ji} = \sum_{j\neq i} p_i k_{ij}$

Equil: $\frac{p_i(eq)}{p_j(eq)} = \frac{k_{ji}}{k_{ij}}$

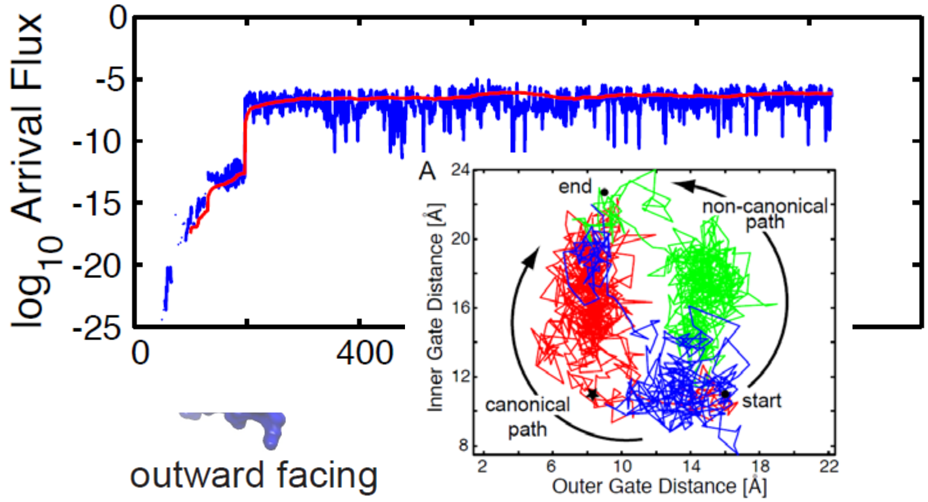
Weighted Ensemble

High weights near start

Low weights far from start

Steady-state/Equil WE: Bhatt, Zhang & Zuckerman, *J Chem Phys* 2010

Steady-state WE for a symporter

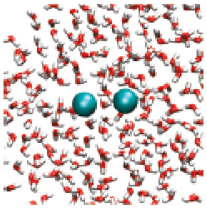


- Mhp1 sodium/benzylhydantoin co-transport – C-alpha model

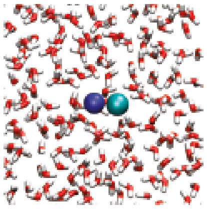
[Adelman et al, Biophys J, 2011]

Association in Explicit Solvent (Steady State)

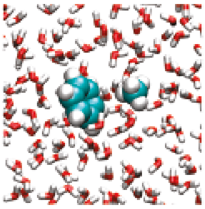
- Parallelized
- Efficient in terms of overall computer use



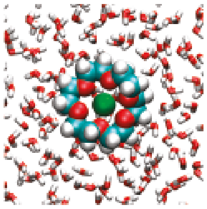
Methane/Methane
(7.0)



Na+/Cl-
(1.4)



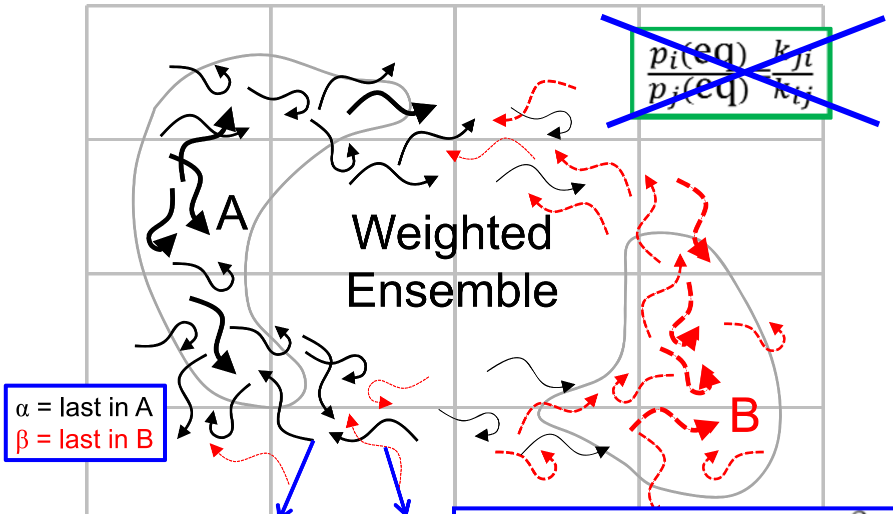
Benzene/Methane
(8.7)



K+ 18-crown-6 ether
(300)

[Zwier, Kaus, Chong, J Chem Theory Comp, 2011]

Non-Markov labeling for equilibrium and kinetics

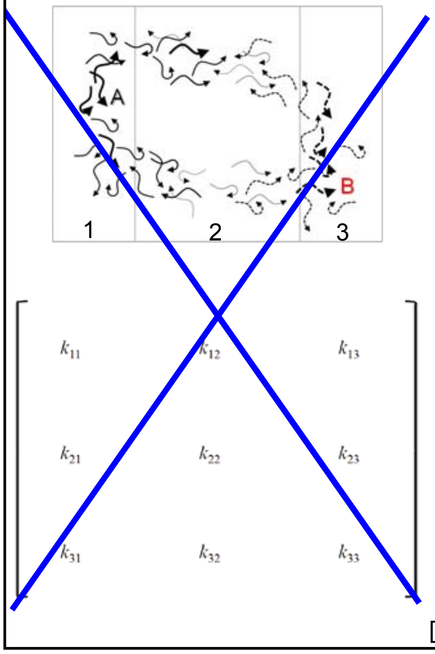


$$\cancel{\frac{p_i(\mathrm{eq})}{p_j(\mathrm{eq})} \; \frac{k_{ji}}{k_{ij}}}$$

$$\boxed{p_i = p_i^\alpha + p_i^\beta}$$

$$\boxed{k_{ij} = p_i^\alpha k_{ij}^\alpha + p_i^\beta k_{ij}^\beta}$$

Labeled/Non-Markovian matrix

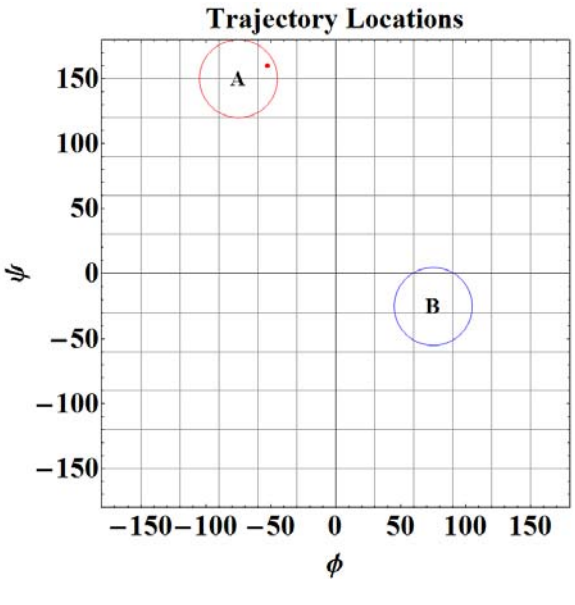


- Yields equilibrium and kinetic quantities without bias (Markov analysis yields biased kinetics)
- Matrix used to solve steady state: No Markov assumption
- Analysis performed after simulation: OK to change state definitions

$$\left[\begin{array}{cc:cc:cc}
k_{11}^{\alpha} & 0 & k_{12}^{\alpha} & 0 & 0 & k_{13}^{\alpha} \\
0 & 0 & 0 & 0 & 0 & 0 \\
\hdashline
k_{21}^{\alpha} & 0 & k_{22}^{\alpha} & 0 & 0 & k_{23}^{\alpha} \\
k_{21}^{\beta} & 0 & 0 & k_{22}^{\beta} & 0 & k_{23}^{\beta} \\
\hdashline
0 & 0 & 0 & 0 & 0 & 0 \\
k_{31}^{\beta} & 0 & 0 & k_{32}^{\beta} & 0 & k_{33}^{\beta}
\end{array}\right]$$

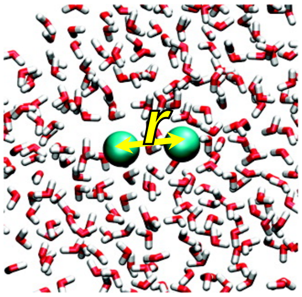
[Suarez, Lettieri, … Zuckerman, JCTC, 2014]

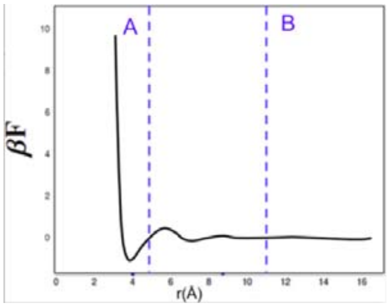
α/β Labeling



Methane-Methane Association

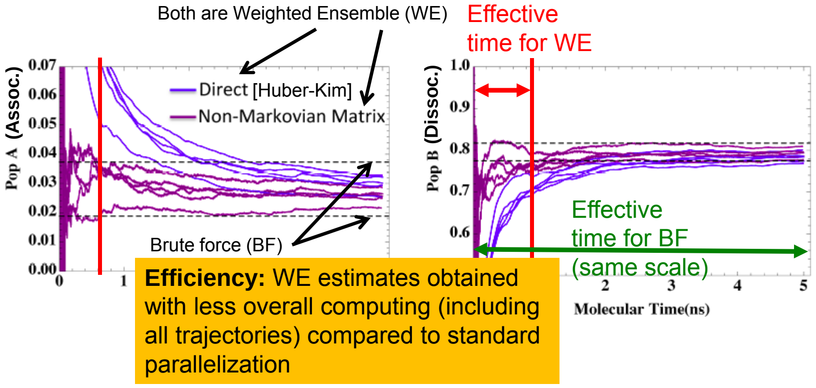
- Explicit solvent/united-atom methane
- Fast/easy system
- Good sampling by both
  - “brute force” (ordinary MD)
  - WE
- Single WE simulation (original Huber-Kim)
  - many different analyses
- Repeated runs to show variation





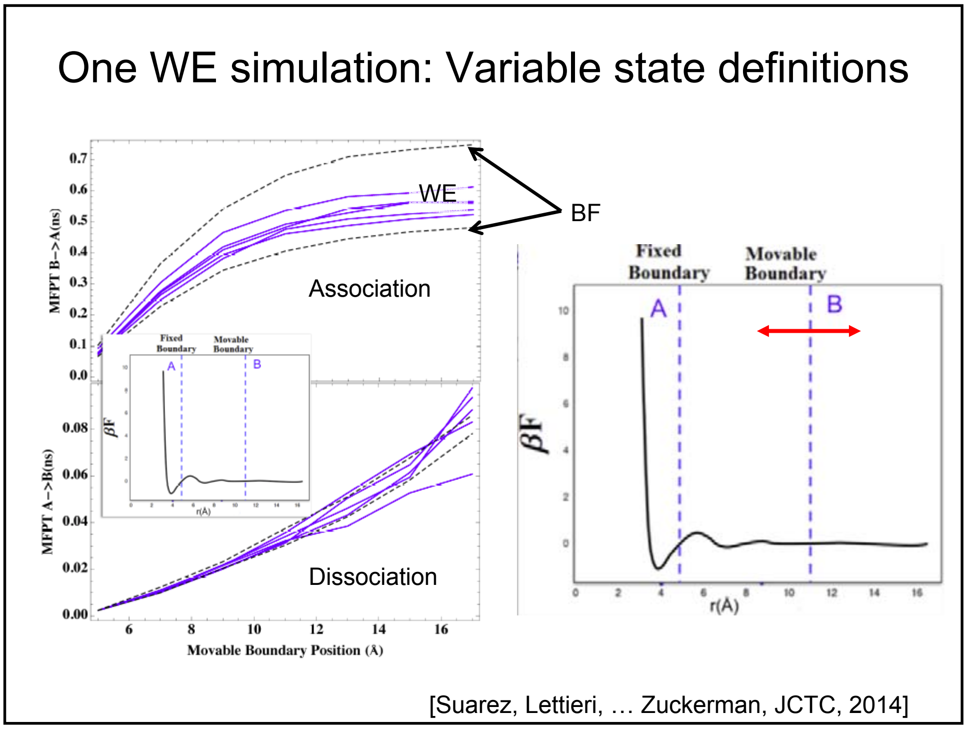
[Zwier, Kaus, Chong, JCTC 2011]

Equil/non-eq observables - Efficiency



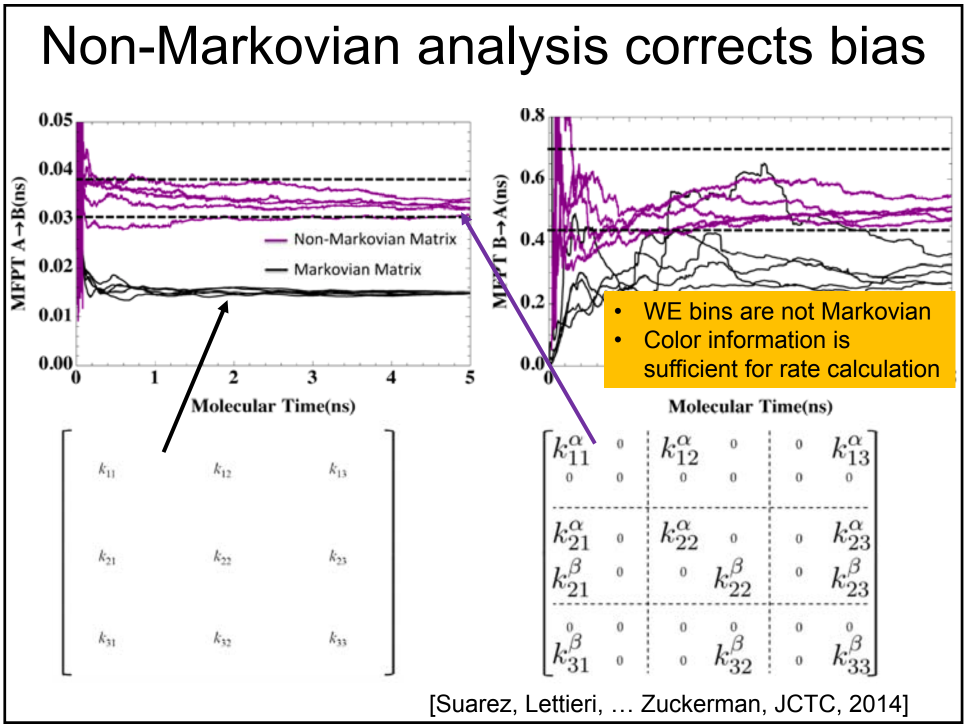
Efficiency: WE estimates obtained with less overall computing (including all trajectories) compared to standard parallelization

[Suarez, Lettieri, … Zuckerman, JCTC, 2014]



One WE simulation: Variable state definitions

[Suarez, Lettieri, … Zuckerman, JCTC, 2014]



Non-Markovian analysis corrects bias

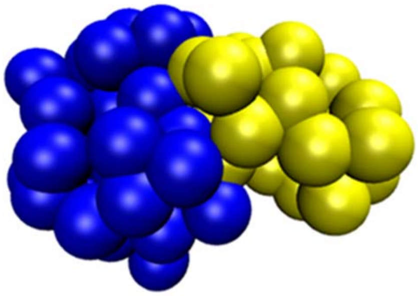
- WE bins are not Markovian
- Color information is sufficient for rate calculation

$$\begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix}$$

$$\left[\begin{array}{cc:cc:cc} k_{11}^{\alpha} & 0 & k_{12}^{\alpha} & 0 & 0 & k_{13}^{\alpha} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hdashline k_{21}^{\alpha} & 0 & k_{22}^{\alpha} & 0 & 0 & k_{23}^{\alpha} \\ k_{21}^{\beta} & 0 & 0 & k_{22}^{\beta} & 0 & k_{23}^{\beta} \\ \hdashline 0 & 0 & 0 & 0 & 0 & 0 \\ k_{31}^{\beta} & 0 & 0 & k_{32}^{\beta} & 0 & k_{33}^{\beta} \end{array}\right]$$

[Suarez, Lettieri, … Zuckerman, JCTC, 2014]

Flexible, coarse-grained models of barnase and barstar

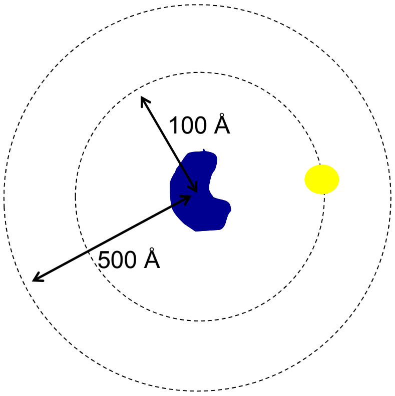


Retains molecular shapes, electrostatic potentials, and diffusion properties of all-atom models

- Approximately one pseudo-atom for every three residues with flexible harmonic bonds
- Electrostatic interactions calculated using Debye-Hückel equation
- Non-electrostatic interactions calculated using a *very* weak Gō-like potential

Frembgen-Kesner & Elcock, *Biophys. J.* 2010

Our simulation strategy

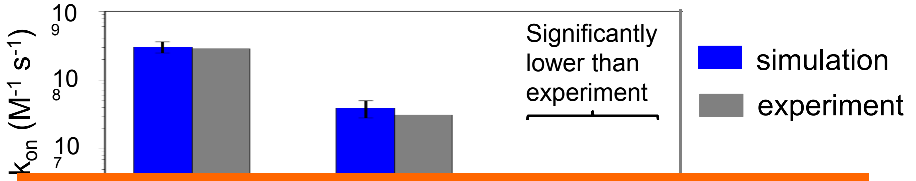


- Carried out five separate steady-state WE simulations
- Each simulation was initiated from 24 randomly oriented unbound states
- Applied the Northrup-Allison-McCammon (NAM) framework for recycling trajectories
- BD simulations with hydrodynamic interactions using UIOWA-BD software

WE parameters

- Progress coordinate divided into three zones:
  1) **Far zone:** Distance between the proteins
  2) **Intermediate zone:** RMSD of barstar after alignment of barnase from the native complex
  3) **Near zone:** a) RMSD of barstar after alignment of barnase from the native complex, and b) fraction of intermolecular native contacts
- 760 bins, 24 walkers/bin, τ = 2 ns, 1000τ (2 μs)

What is the computed ‘basal’ k_{on}?



Supports the use of simulations to directly obtain the basal k_{on} under regular salt conditions

- Suggests that electrostatic interactions are not completely eliminated in experiments at high salt concentrations

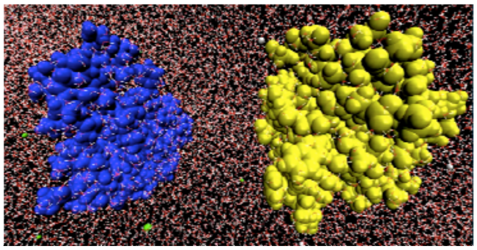
Saglam & Chong, *J. Phys. Chem. B* 2016

How efficient is WE sampling of these slow associations?

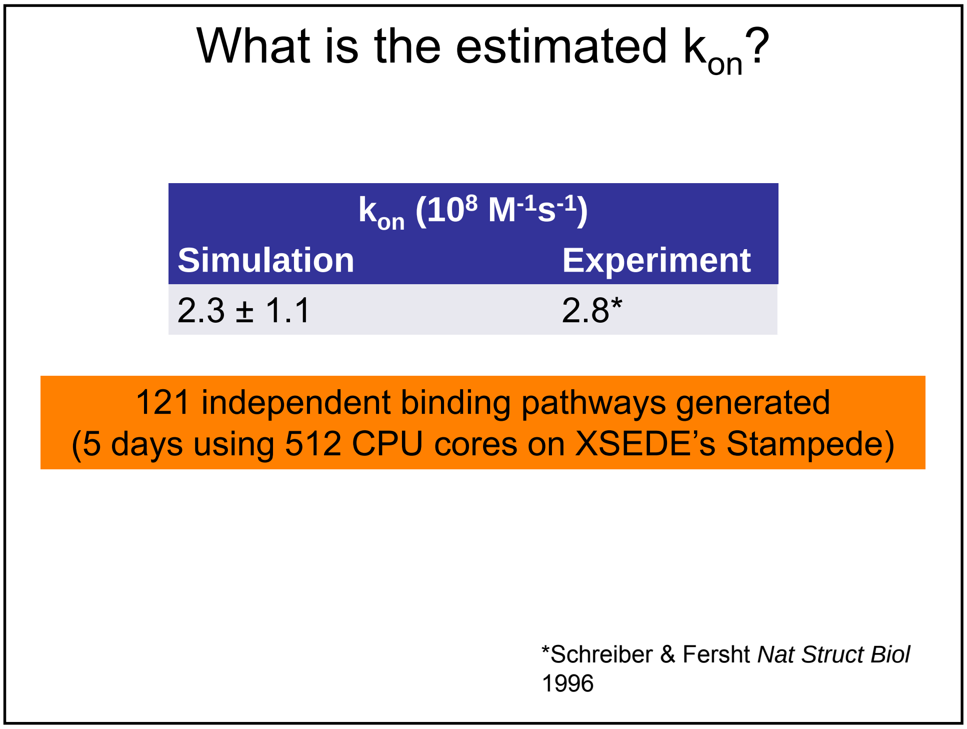
	WE	Brute force
Number of association events	>1000	
Number of CPU cores	512	
Wall-clock time	3 days	386 days (!)

- WE is >100x more efficient than brute force simulation in generating association events

Moving on to atomistic simulations in explicit solvent…



- GROMACS software
- All-atom AMBER ff99SB-ILDN force field
- TIP3P explicit water molecules
- To match experiment: 25 ^{o}C, 1 atm, 50 mM NaCl

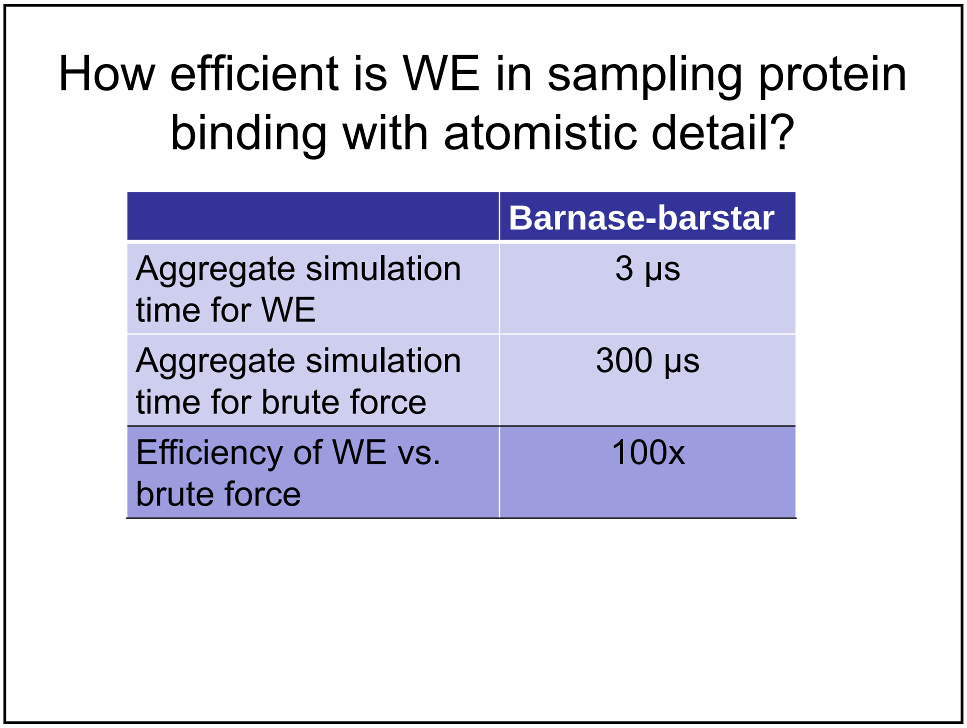


What is the estimated k_{on}?

k_{on} (10^8 $M^{-1}s^{-1}$)	
Simulation	**Experiment**
2.3 ± 1.1	2.8*

121 independent binding pathways generated
(5 days using 512 CPU cores on XSEDE's Stampede)

*Schreiber & Fersht *Nat Struct Biol* 1996



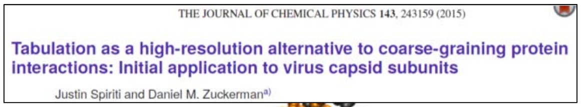
How efficient is WE in sampling protein binding with atomistic detail?

	Barnase-barstar
Aggregate simulation time for WE	3 μs
Aggregate simulation time for brute force	300 μs
Efficiency of WE vs. brute force	100x

WE is a "meta method"

- Key: WE checks trajectories at fixed time intervals
- No software-specific parallelization required
- Scripting-level: Requires only ability to start, stop, and re-start simulations
  - Competing methods require difficult modifications to source code
- Implemented with AMBER, GROMACS, NAMD
  - Easy to add new package
- Generality for other contexts
  - Example: Systems biology
- WESTPA software (LT Chong)
  - Scales to thousands of cores

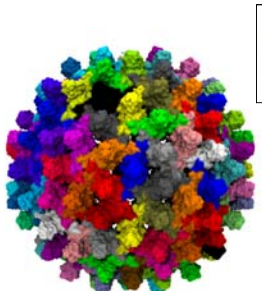
Virus Capsid Assembly

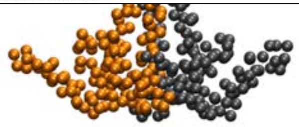


THE JOURNAL OF CHEMICAL PHYSICS 143, 243159 (2015)

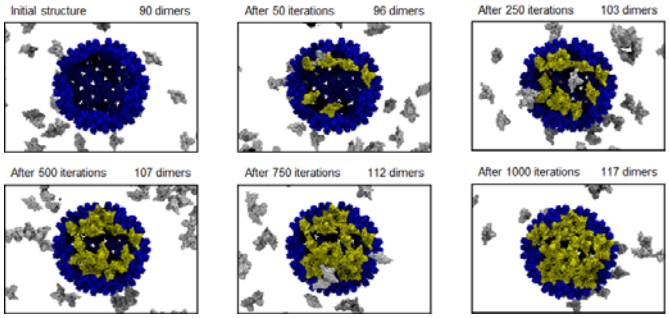
Tabulation as a high-resolution alternative to coarse-graining protein interactions: Initial application to virus capsid subunits

Justin Spiriti and Daniel M. Zuckerman[a)]



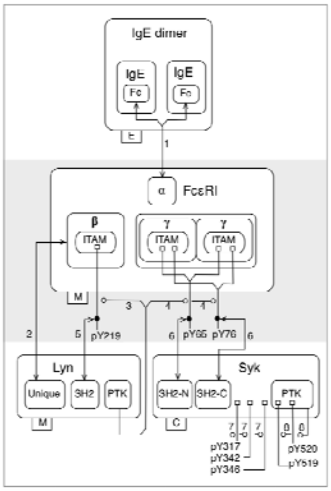


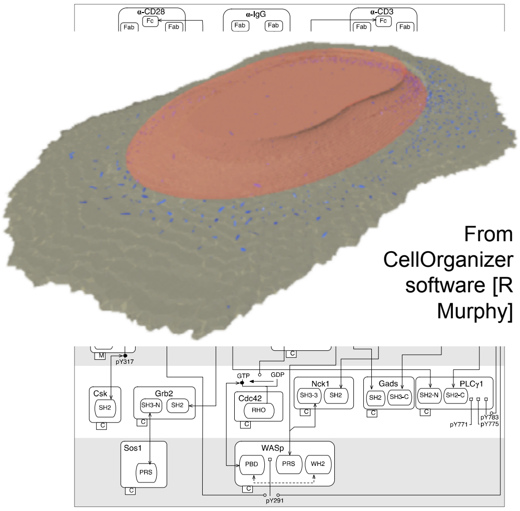
284 C-alpha model
(initial model)



Tabulation +
Weighted
Ensemble
simulation

Systems biology: Cell-scale networks, spatial models



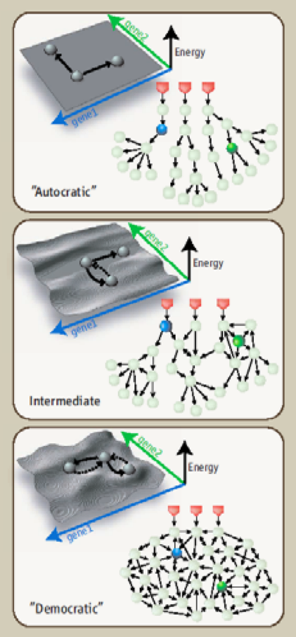


From CellOrganizer software [R Murphy]

[James Faeder, U. Pittsburgh]

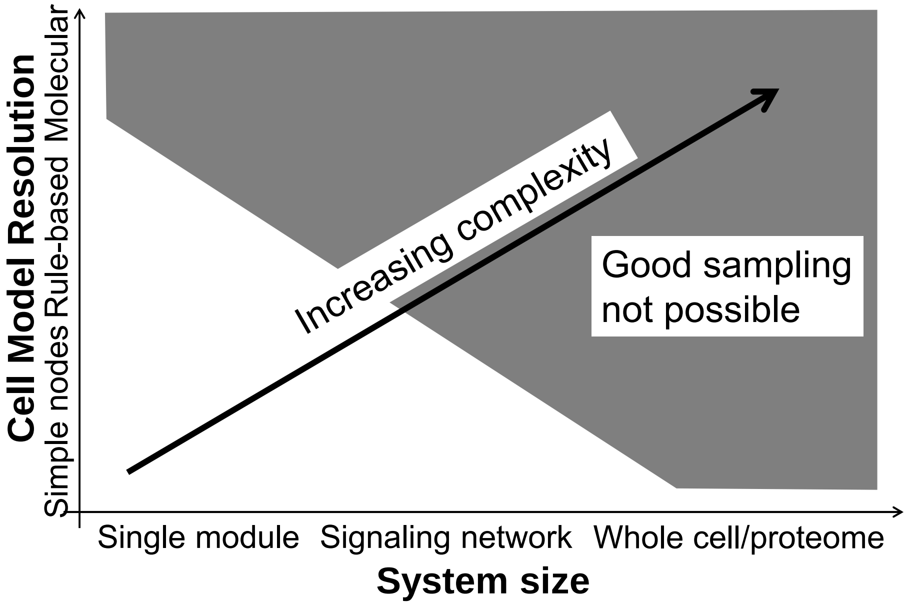
Energy landscapes in systems biology

- Properly constructed kinetic models (thermodynamically consistent) are equivalent to free energy landscapes



[Bar-Yam et al, Science, 2009]

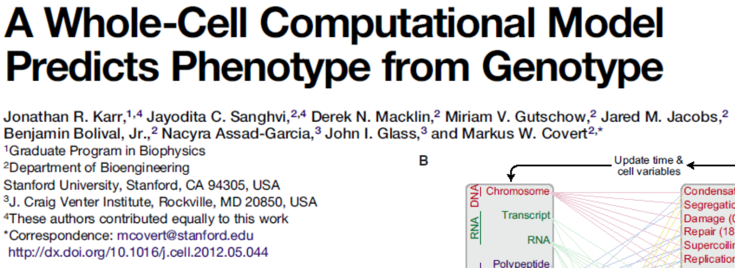
Similar challenges for biomolecules & systems



Complexity is here

Theory

Cell

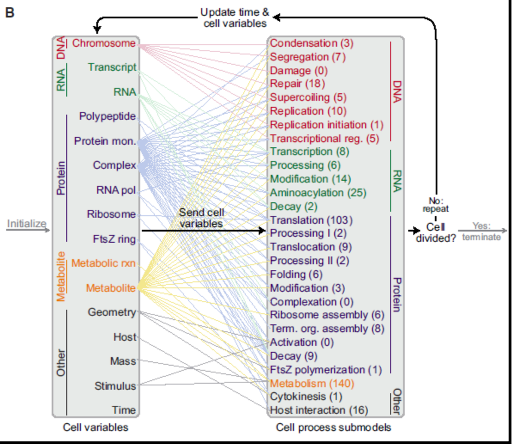


A Whole-Cell Computational Model Predicts Phenotype from Genotype

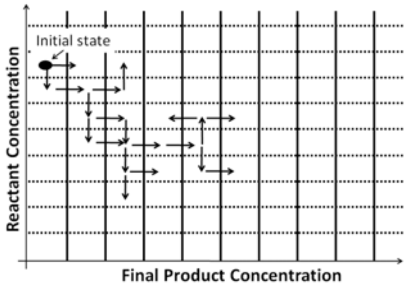
Jonathan R. Karr,[1,4] Jayodita C. Sanghvi,[2,4] Derek N. Macklin,[2] Miriam V. Gutschow,[2] Jared M. Jacobs,[2] Benjamin Bolival, Jr.,[2] Nacyra Assad-Garcia,[3] John I. Glass,[3] and Markus W. Covert[2,*]
[1]Graduate Program in Biophysics
[2]Department of Bioengineering
Stanford University, Stanford, CA 94305, USA
[3]J. Craig Venter Institute, Rockville, MD 20850, USA
[4]These authors contributed equally to this work
*Correspondence: mcovert@stanford.edu

http://dx.doi.org/10.1016/j.cell.2012.05.044

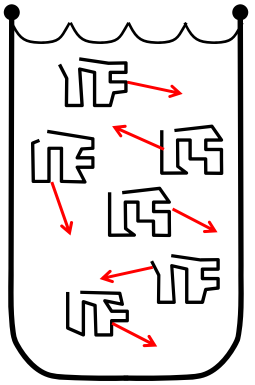
Organism: human pathogen
Mycoplasma genitalium
[Cell, 2012]



WE can work in other spaces: Species concentrations, Real space



Species-concentration space



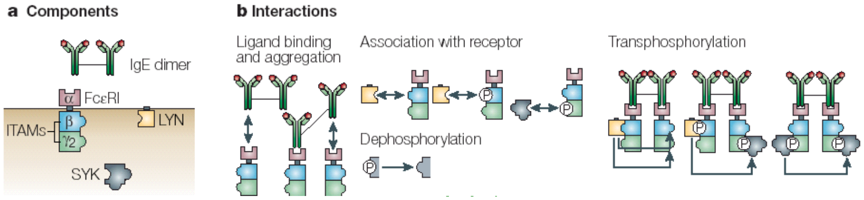
Real space

Immunological signaling via the high affinity receptor for IgE (FcεRI) – BioNetGen software

354 species, 3680 reactions

```
# Ligand-receptor binding
R1: Rec(a) + Lig(l,l) <-> Rec(a!1).Lig(l!1,l) kp1, km1
```

```
# Transphosphorylation of Syk by constitutive Lyn
R10: Lig(l!1,l!2).Lyn(U!3,SH2).Rec(a!2,b~Y!3).Rec(a!1,g~pY!4).Syk(tSH2!4,l~Y) -> \
```

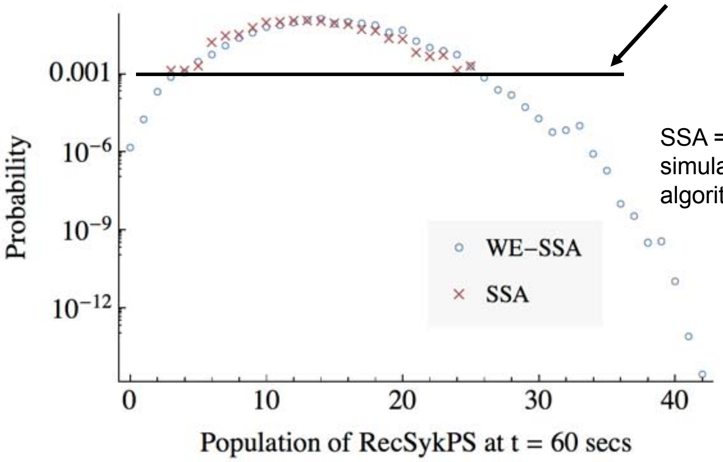


```
# Lyn-receptor binding through SH2 domain
R6: Rec(b~pY) + Lyn(U,SH2) <-> Rec(b~pY!1).Lyn(U,SH2!1) kpLs, kmLs

# Transphosphorylation of beta by SH2-bound Lyn
R7: Lig(l!1,l!2).Lyn(U,SH2!3).Rec(a!2,b~pY!3).Rec(a!1,b~Y) -> \
   Lig(l!1,l!2).Lyn(U,SH2!3).Rec(a!2,b~pY!3).Rec(a!1,b~pY) pLbs

# Transphosphorylation of gamma by SH2-bound Lyn
R8: Lig(l!1,l!2).Lyn(U,SH2!3).Rec(a!2,b~pY!3).Rec(a!1,g~Y) -> \
   Lig(l!1,l!2).Lyn(U,SH2!3).Rec(a!2,b~pY!3).Rec(a!1,g~pY) pLgs

# Syk-receptor binding through tSH2 domain
R9: Rec(g~pY) + Syk(tSH2) <-> Rec(g~pY!1).Syk(tSH2!1) kpS, kmS
```

```
R14: Rec(b~pY)-> Rec(b~Y) dm

# Dephosphorylation of Rec gamma
R15: Rec(g~pY)-> Rec(g~Y) dm

# Dephosphorylation of Syk at membrane
R16: Syk(tSH2!+,l~pY)-> Syk(tSH2!+,l~Y) dm
R17: Syk(tSH2!+,a~pY)-> Syk(tSH2!+,a~Y) dm

# Dephosphorylation of Syk in cytosol
R18: Syk(tSH2,l~pY)-> Syk(tSH2,l~Y) dc
R19: Syk(tSH2,a~pY)-> Syk(tSH2,a~Y) dc
```

Dynamics =
Ordinary stochastic
chemical kinetics

Probability density of a key species



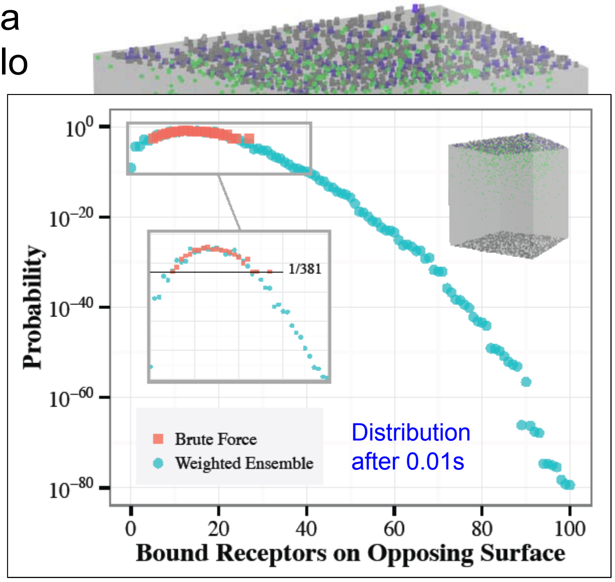
$\frac{1}{1,000}$ = Precision limit from 1,000 simulations

SSA = Stochastic simulation (Gillespie) algorithm

[Donovan, Sedgewick, Faeder, Zuckerman, J. Chem. Phys. 2013]

Spatial dynamics via kinetic Monte Carlo

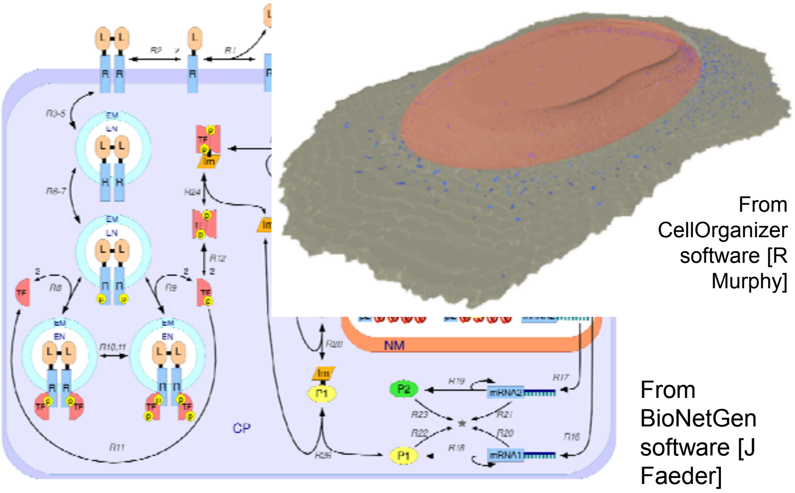
- Implementation via MCell (Monte Carlo Cell) simulator, controlled by WESTPA



Distribution after 0.01s

[Donovan et al., PLoS Comp Bio, 2016]

Biochemical reactions embedded in realistic cellular geometry

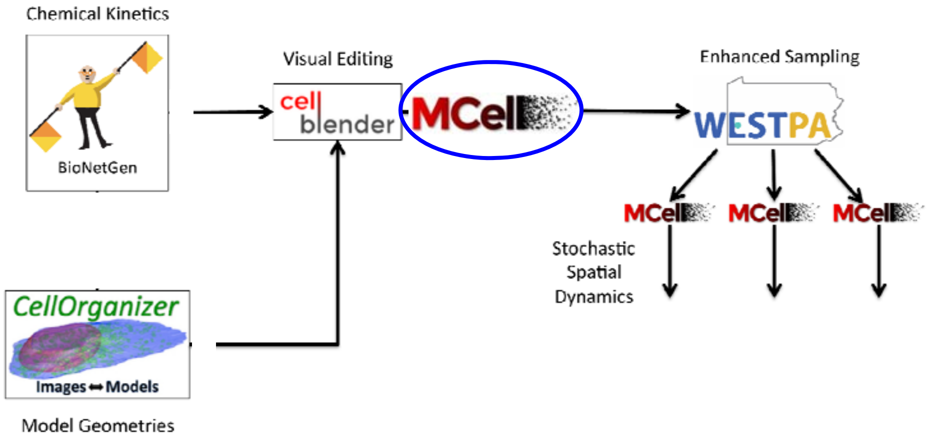


From CellOrganizer software [R Murphy]

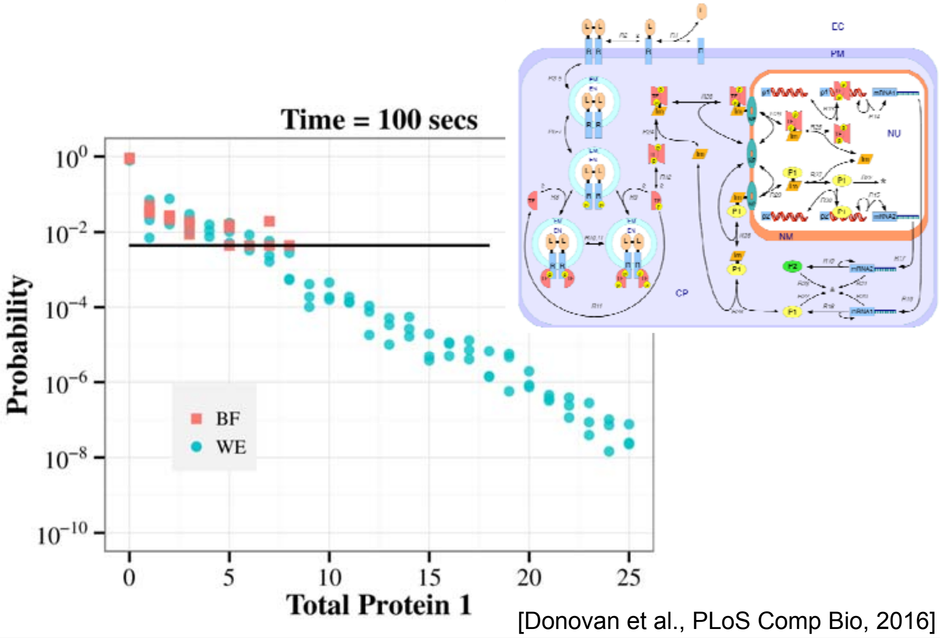
From BioNetGen software [J Faeder]

[Donovan et al., PLoS Comp Bio, 2016]

Reaction network & geometry → MCell



Cell/Compartment Signaling

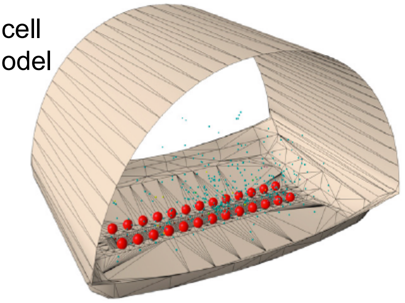


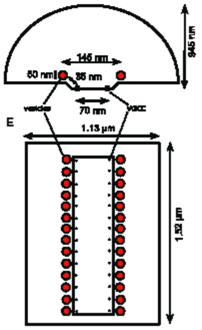
[Donovan et al., PLoS Comp Bio, 2016]

Spatial modeling: Frog NMJ

Neuromuscular Junction Model

Mcell model

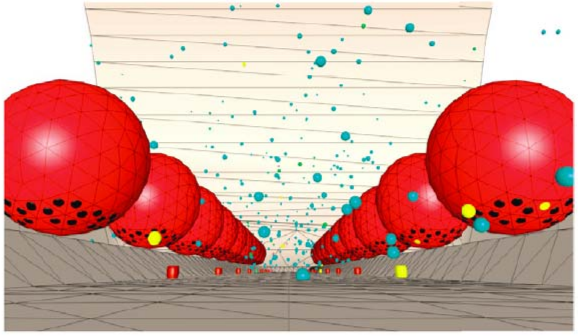




[Donovan et al., PLoS Comp Bio, 2016]

NMJ in MCell software

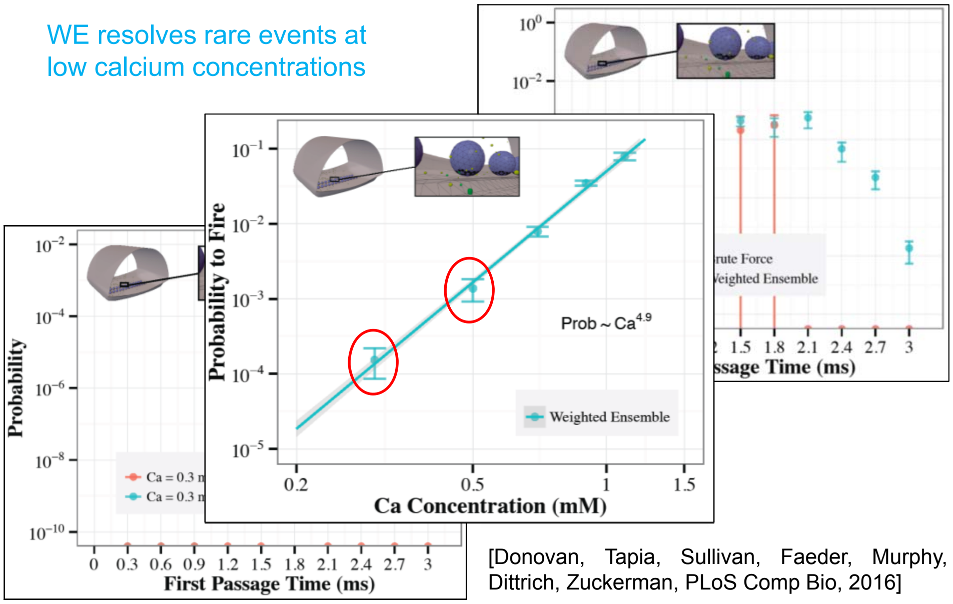
- MCell = Spatially resolved kinetic Monte Carlo
- NMJ model: Release of pre-synaptic vesicle triggered only when sufficient calcium ions bind in threshold configuration
  - [Dittrich et al., Biophysical J., 2013]



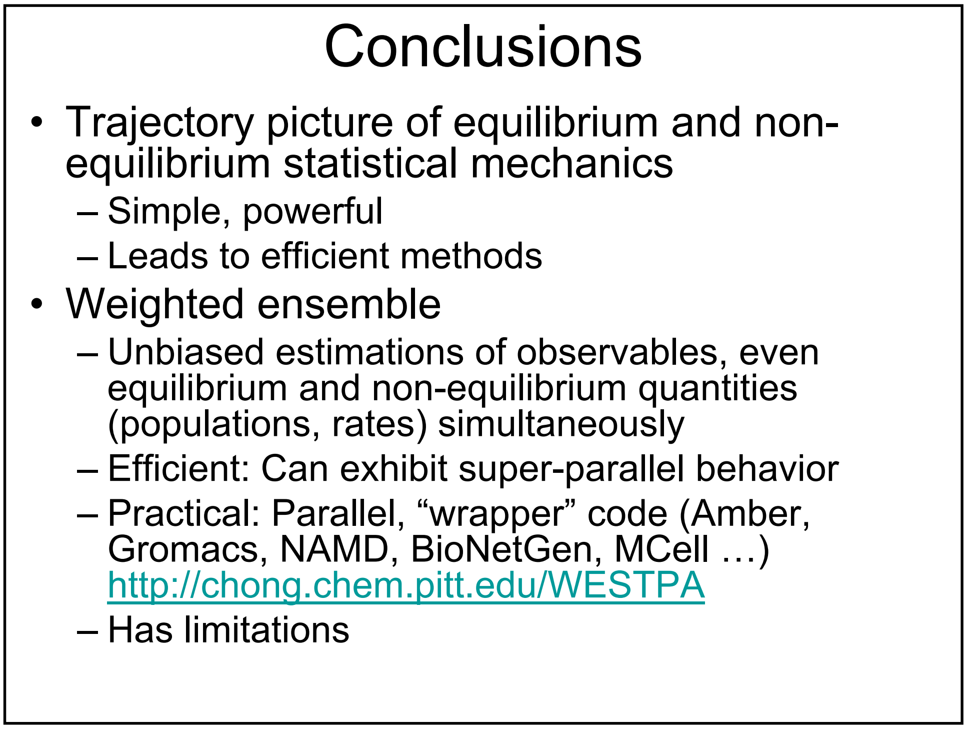
[Donovan et al., PLoS Comp Bio, 2016]

WE applied to Neuro-muscular junction model

WE resolves rare events at low calcium concentrations



[Donovan, Tapia, Sullivan, Faeder, Murphy, Dittrich, Zuckerman, PLoS Comp Bio, 2016]



Conclusions

- Trajectory picture of equilibrium and non-equilibrium statistical mechanics
  - Simple, powerful
  - Leads to efficient methods
- Weighted ensemble
  - Unbiased estimations of observables, even equilibrium and non-equilibrium quantities (populations, rates) simultaneously
  - Efficient: Can exhibit super-parallel behavior
  - Practical: Parallel, “wrapper” code (Amber, Gromacs, NAMD, BioNetGen, MCell …) http://chong.chem.pitt.edu/WESTPA
  - Has limitations