

How to avoid embarrassment when quoting timescales (and temperatures) in coarse-grained/simplified models

Prof. Paul Whitford

CTBP Tech Talk

10/23/2024

Center for Theoretical Biological Physics



RICE



Northeastern

p.whitford@northeastern.edu

Disclaimer

- In order to make a point, some of the statements may be a bit hyperbolic.

Second Disclaimer

- The statements are not as hyperbolic as you may think...
- If you want to perform simulations, you should remember the basics of what you are quoting...

How the unwitting often get in trouble

You pick up your favorite software and run a simulation with a non-standard model (i.e. Team A wrote the software and Team B created the model).

You observe some large motion, such as a protein folding. The output indicates the temperature was 100K and it took 50 nanoseconds to occur.

You are excited and present your work.

Your supervisor sees you make an outrageous claim about nanosecond-scale folding at freezing temperatures. They lose all respect for your intellect. You start rethinking your career...

... a very slight exaggeration.



The bright side

- Units in coarse-grained/simplified models may appear mysterious, at first. However, getting them right isn't that complicated.
- If you make these mistakes, you are in good company. Papers that report incorrect times and temperatures are quite common... that doesn't make it acceptable.
- The CTBP community is available to help you navigate these topics. Just let us know.

Goals for today

- Prevent you from making outrageous (i.e. completely wrong) claims about simulated time and temperatures
- Discuss the meaning of units in physics (and therefore in models)
- Outline strategies for estimating time and energies/temperatures
 - Consider advantages and limitations of each approach

Back to basics: Units

- For any physical question, you can arbitrarily define a number of units.
- All other units are then derived from these, or quantities must be converted.
- For example, consider SI units:
 - There are 7 “base” units
 - Time, second
 - Length, meter
 - Mass, kilogram
 - Electric current, ampere
 - Thermodynamics temperature, kelvin
 - Amount, mole
 - Luminous intensity, candela
 - All other units are derived from base units, e.g.
 - Force = mass * acceleration = mass *length/time² = kg*m/s² = Newton
 - Energy = mass * length²/time² = kg*m²/s² = Joule

See wiki links on last slide

The precise definition of each unit is defined in terms of “defining constants”

- Defining constants are “the most fundamental feature of the definition of the system of units” – Wikipedia
- Each has an exactly defined value, for a specific set of units.
- What are they?
 - Boltzmann constant, k_B
 - Avagadro constant, N_A
 - Elementary charge, e
 - Planck constant, h
 - Hyperfine transition frequency of Cs
 - Speed of light, c
 - Luminous efficacy, K_{cd}

Base units and defining constants give us numeric values for the unit system

- For example, what is “a” Kelvin?
 - Boltzmann constant k_B has units of energy/K, and it is defined to be exactly $1.3806505 \times 10^{-23} \text{ J/K}$
 - Energy has units of Joules
 - Joules are derived from the base units kg, m, s
 - So, the numeric meaning of a Kelvin is only defined in terms of its exact relationship to the Boltzmann constant.
 - Logical. In statistical mechanics, T only enters as $k_B T$

What are other unit systems one could use?

- cgs – centimeter, gram, s
- British units: foot, ounce, s
 - Energy unit would be the ounce*foot²/s²
- Infinite number of possible systems – simply define the base units and everything else follows.
- This is true in physics, and therefore in simulations, too.
- We will consider “reduced units”
 - The units that “fits” the problem most naturally.
 - And set all constants (e.g. k_B) to 1
 - Examples of choices for reduced length units for different problems
 - Lennard-Jones fluid: the Lennard-Jones distance σ
 - Biomolecules: nm is a natural choice
 - galactic dynamics: astronomical units or light-years
 - It would be very inconvenient to use light-years to describe lengths in biomolecules.
 - “the atoms are initially separated by a distance of 6×10^{-26} light-years and then separate by an additional 9×10^{-26} light-years”
 - Or “the atoms are initially separated by 1 nm and they then separate by an additional 1.5nm”

Let's look at an example model

- All-atom structure-based “SMOG” models
 - Here, we assume you are using the default model in the SMOG 2 software (Noel et al. 2016)
- All non-hydrogen atoms are included
- Units in the model:
 - Each atomic interaction is given a strength of order 1 energy unit
 - Lengths are defined in terms of nm
 - The mass of each carbon/nitrogen/oxygen atom is given a value of 1
- Now, you run a simulation and the software reports time in “picoseconds” and temperature in “Kelvin”
 - How does Gromacs/NAMD/OpenMM/etc know your units?
 - Answer: It probably doesn’t!!!!!!! **The ps and K labels may be (likely are) meaningless.**
 - Tip: To avoid the fiery wrath of an expert (and keep your job), NEVER quote these labels without knowing where they came from.



Why is the unit in the time label meaningless?

- Consider how one defines units (in Gromacs, as an example)
 - Base units
 - Length, nm
 - Time, ps
 - Mass, u
 - Derived unit
 - Energy, kJ/mol
- You can arbitrarily define any three
 - But you must live with the fourth
- Typical software assume you are giving quantities in your model using nm, u and kJ/mol
 - Then, corresponding time unit is ps
- But, in our SMOG model, we are using different mass and energy units...

Table 2 Basic units used in GROMACS

Quantity	Symbol	Unit
length	r	nm = 10^{-9} m
mass	m	u (unified atomic mass unit) = $1.660\ 538\ 921 \times 10^{-27}$ kg
time	t	ps = 10^{-12} s
charge	q	e = elementary charge = $1.602\ 176\ 565 \times 10^{-19}$ C
temperature	T	K

Consistent with these units are a set of derived units, given in Table 3

Table 3 Derived units. Note that an additional conversion factor of 10^{28} a.m.u (≈ 16.6) is applied to get bar instead of internal MD units in the energy and log files

Quantity	Symbol	Unit
energy	E, V	kJ mol ⁻¹
Force	F	kJ mol ⁻¹ nm ⁻¹
pressure	p	bar
velocity	v	nm ps ⁻¹ = 1000 m s ⁻¹
dipole moment	μ	e nm
electric potential	Φ	kJ mol ⁻¹ e ⁻¹ = 0.010 364 269 19 Volt
electric field	E	kJ mol ⁻¹ nm ⁻¹ e ⁻¹ = $1.036\ 426\ 919 \times 10^7$ Vm ⁻¹

The SMOG model: Kelvin?

- Units in the model:
 - Each atomic interaction is given a strength of order 1 energy unit.
 - Length is defined in terms of nm.
 - The mass of a C atom is defined as 1.
- Derive the energy units: $E_{R.U.} = (\text{mass unit}) * (\text{length unit})^2 / (\text{time unit})^2$
 - The mass unit: 12 amu
 - The length unit: nm
 - We don't know the time unit....
- Since we don't know the time unit, we can not convert the Boltzmann constant from units of J/K to $E_{R.U.}/K$
 - If this is true, why does the software say "Kelvin" everywhere?
 - Because the software typically hard codes the value of the Boltzmann constant, assuming the model is using energy units of Joules or cal.
 - If your model is not using the hard-coded energy unit, then the label "Kelvin" does not mean your temperature is in Kelvin.

Refs on reduced units
Allen and Tildesley; Frenkel and Smit

The SMOG model: Kelvin?

How do we recover Kelvin? Or, how do we set the simulated temperature value to reflect a desired value in K (e.g. 300 K)?

- In statistical mechanics, the (unnormalized) probability of a conformational state under equilibrium conditions follows the Boltzmann distribution: $P=\exp\left(-\frac{U}{k_B T}\right)$
- In your calculation, you can run a simulation at T “Kelvin” (i.e. the label says “Kelvin” and k_B has a non-unity value)
- Determine a value of your simulated temperature that should reflect the dynamics at 300 K
 - In Gromacs, when temperature is set to ~60 “Kelvin” in a default AA SMOG model, you can recover the scale of some observable in an explicit-solvent simulation (e.g. rmsf values) – strategy detailed in Jackson et al. 2015
 - Technical note: Temperature of 60 in Gromacs and OpenMM is the same as a temperature of 0.5 in OpenSMOG
- For 60 “Kelvin” in Gromacs/OpenMM you can obtain a numeric value for the thermal energy: $k_B T=0.0083144621*60=0.50 \text{ E}_{\text{R.U.}}$
 - This is the value of $k_B T$ at room temperature, in the energy units of your model (reduced energy units, $\text{E}_{\text{R.U.}}$).
- At room temperature, we know that $k_B T \sim 0.6 \text{ kcal/mol}$
- Based on this, the energy unit in the model represents $0.6/0.5 \text{ kcal/mol}=1.2 \text{ kcal/mol}$
- This is the energy scale that consistently defines 60 “Kelvin” (i.e. what the program thinks) to represent 300 K (the physical significance of T)
 - Tech note: this calibration can be system/process-specific. It is important to have a benchmark for verifying the estimated temperature.

The SMOG model: picoseconds?

- Units in the model:
 - Each atomic interaction is given a strength of order 1 energy unit
 - Length is defined in terms of nm
 - The mass of a C/N/O/S atom is defined as 1
- energy unit = (mass unit)*(length unit)²/(time unit)²
- Invert to get the time unit : $\tau_{R.U.} = (\text{length unit}) * ((\text{mass unit}) / (\text{energy unit}))^{1/2}$
- $\tau_{R.U.}$ is our reduced time unit
- In our model, the software will say “X ps”, but it really means “X $\tau_{R.U.}$.” *
 - * this assumes the software does not have any additional internal conversion factors. You can verify this by checking the number of timesteps simulated and the reported time.
- Can we estimate $\tau_{R.U.}$?
- The energy unit was estimated to be approximately 1 kcal/mol on the last slide
 - An estimate based on base stacking interactions for RNA in this model also suggested 1kcal/mol (Whitford et al. 2010a)
- In this model $\tau_{R.U.} = \text{nm} * ((\text{approx. mass of Carbon}) / (1 \text{ kcal/mol}))^{1/2} \sim 3 \text{ ps}$
- So if the software says 10 ps, it really means 10 physically means $10\tau_{R.U.}$, which is 30ps.
 - This is progress, but it is not perfect.

What are we still missing?

- In this model, there is no explicit water and the energetics are smooth.
- When using these models, Langevin dynamics protocols may be used, but typically with a small drag coefficient.
- In principle, one could use a larger drag coefficient that would mimic the presence of water, such that the diffusion coefficient of the molecule would match an experimental benchmark.
- This approach is not always desirable.
 - It would/could dramatically increase the computational requirements because it slows down the dynamics.
 - Drag would be uniformly applied, independent of the local environment (e.g. buried or exposed region).

How to get a better estimate of the significance of a time unit

- Various methods are available (ascending order of my preference)
 - Experimental comparison of mean first passage times (mfpt)
 - Calculating rates from free-energy barriers
 - Calculating rates using a diffusion coefficient and free-energy landscape
 - Directly interpreting the simulated time by estimating the time unit
- Each approach has pros and cons

Fundamentals relations in chemical kinetics

$$\tau = \frac{1}{R} = \frac{1}{C} \exp(\beta \Delta F) \quad \text{For a two-state transition}$$

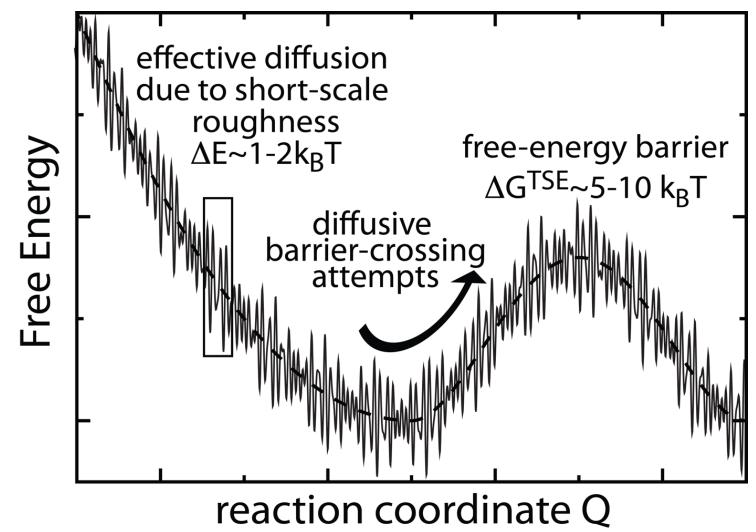
$$= \int_{Q_i}^{Q_f} \exp[F(Q')/kT] / D(Q') dQ' \int_{\infty}^{Q'} \exp[-F(Q'')/kT] dQ'' \quad \text{general}$$

Q A reaction coordinate

$F(Q)$ Effective, short-length-scale averaged free energy, as a function of Q

$D(Q)$ Effective diffusion coefficient, in Q -space, as a function of Q

τ Mean first passage time between Q_i and Q_f



Refs on kinetics

First 4 chapters of Zwanzig
Bryngelson and Wolynes 1989
Zwanzig 1988

Approach 1

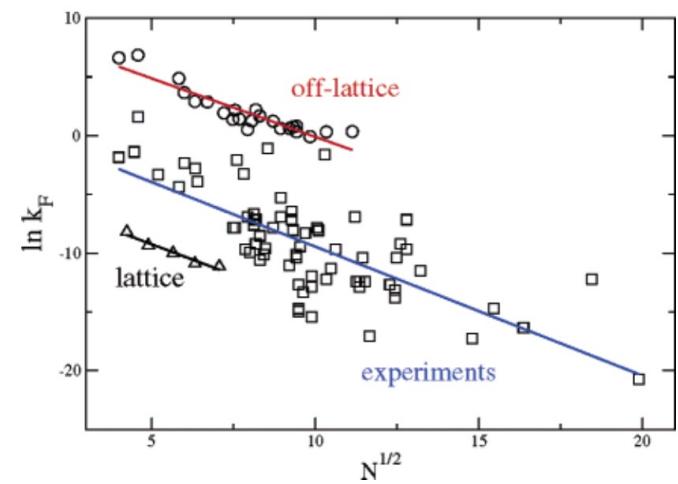
Experimental comparison of mean first passage times

- The idea

- For a specific model and class of phenomena, look for an empirical trend that can be corroborated experimentally.
- Then, assume this trend holds for other systems.

- Example from Kouza et al. 2006

- Simulated folding of many globular, two-state, proteins and calculated the folding time (in ps, after directly converting from reduced units)
- Noticed that the folding rate/time scales with the chain length in a manner consistent with experiments
- While the functional dependence was the same, the magnitude of the predicted rates was higher by $\sim e^8 \sim 3000$.
- This suggests that a single time unit numerically evaluates to ~ 3 ps, but the motions that occur in one time unit correspond to motions that occur on a timescale of approximately 9000 ps
- Note: The authors did *not* suggest this conversion scale should be generally applied to other systems.



When can this be an acceptable strategy?

- You may “safely” use the correction factor to provide an estimate of the rate/time if you are :
 - Using the exact same model (their form of a CG Go model)
 - Using the same simulation protocols (e.g. Langevin parameters)
 - Simulating a similar system (single-domain protein)
 - Simulating a related process (two-state cooperative folding)
- If any of the above conditions are not true, the conversion factor may not be meaningful.
 - For example, if you use this to estimate timescale in an all-atom SMOG model, you should be very cautious about whether your timescale estimates are reliable

What is the significance of the correction factor?

- For simplicity, consider Kramers' relation for kinetics:

- $k = C \exp\left(-\frac{\Delta F}{k_B T}\right)$

- Why do we need a correction factor of 3000 for this model?

- Is our prefactor (which carries units of time in it) off by 3000?
- Are the free-energy barriers predicted by this model not accurate?

- That is, how should we think about accounting for the correction factor?

- $k = C/3000 \exp\left(-\frac{\Delta F}{k_B T}\right) ?$

- $k = C \exp\left(-\frac{\Delta F}{k_B T} - \ln(3000)\right) = C \exp\left(-\frac{\Delta F + 8k_B T}{k_B T}\right) ?$

- Why this distinction matters

- If we are off *only* because of the prefactor, then it is possible that the diffusion coefficient of all atoms is simply off by a common factor.
- If the free-energy barriers are underestimated for folding, then this correction may be completely meaningless if we want to convert timescales for a different process (e.g. conformational rearrangement)

- **Take home points**

- Because this approach is based on the comparison of a barrier-crossing process, we can not be certain whether diffusion coefficient or barriers need to be corrected. This limits the transferability of the estimates to other processes.
- This can be useful for comparing relative times of similar types of processes. However, since this may mask inaccuracies in the predicted free-energy barriers, it should not be considered a precise method for calibrating the timescale.

Why can this practice be risky?

We will consider a similar strategy that you are likely to find in the literature.

- Assume one knows the experimental timescale of a conformational motion in a protein: T_{exp}
- We use a multi-basin structure-based model
 - In this model, you can *arbitrarily* increase the stability of either/both endpoint
 - Increasing stability of the endpoints will almost always increase the height of the free-energy barrier
- Let's imagine two variants of the same model, each simulated for N timesteps:
 - Model one: transitions occur 100 times
 - Model two: transitions occur 2 times
- If we try to calibrate our time by comparing the time to transition between endpoints, then we come to the conclusion that:
 - In model one, N time steps represents $100T_{\text{exp}}$
 - In model two, N time steps represents $2T_{\text{exp}}$

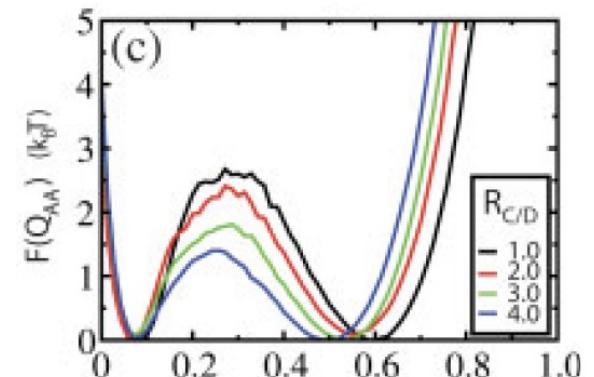


Image from Whitford et al. 2009.

Why can this practice be risky?

- Why is there a factor of 50 ambiguity? Is either estimate accurate?
 - Changing the model changed the barrier.
 - This approach absorbs any inaccuracies in the barrier.
 - Since the barrier is the result of our underlying energetic model, we are essentially ignoring anything that could be wrong with our energetic representation.
 - Both estimates may be completely wrong... Impossible to know without more information.
- **This strategy is (unfortunately) used frequently... Be extremely cautious about interpreting timescales that are derived from this approach.**

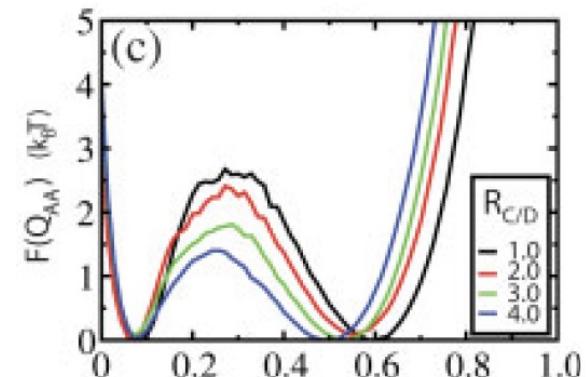


Image from Whitford et al. 2009.

Approach 2: estimate timescales from free-energy barriers

- If you can calculate a free-energy barrier, you may be tempted to immediately report a rate using $k = C \exp(-\frac{\Delta F}{k_B T})$
- In principle: This can/should work.
- In practice: Sometimes computationally difficult. Be careful not to make a trivial oversight.

The challenges of this approach

- $k = C \exp(-\frac{\Delta F}{k_B T})$ is nice and simple. How can you go wrong?
 - The prefactor C is system- and process-specific
 - The barrier must be calculated accurately
- Both factors can be addressed, you just need to be careful.
- Let's consider a hypothetical misuse of this relation
 - We calculate a free-energy barrier of $5 k_B T$ for the folding of a protein.
 - We then estimate the rate as: $k_B T/h^* \exp(-5) \sim 1/\text{ps}$.
 - This can't possibly be correct
 - What went wrong?

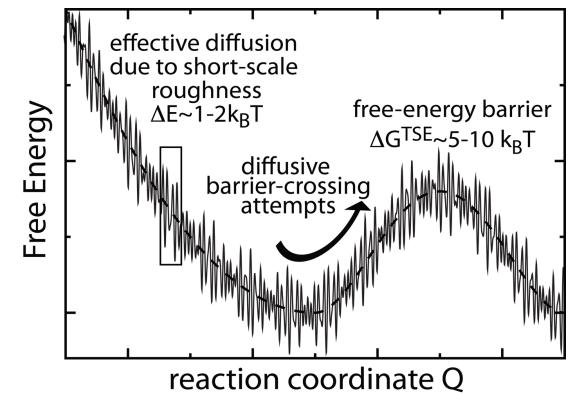
The exponential prefactor C

- How should we interpret the terms in $k = C \exp\left(-\frac{\Delta F}{k_B T}\right)$?
 - C : the frequency of attempting barrier-crossing events
 - $\exp\left(-\frac{\Delta F}{k_B T}\right)$: the probability of successfully crossing the barrier
- We used $C = k_B T/h$. Why?
 - Well, we shouldn't when looking at folding. This value is from transition state theory (TST) of chemical reactions, where the reaction depends on bond vibrations that occur in this range
 - Folding is not controlled by bond vibrations. Folding is a collective process involving many atoms diffusing through space. The "frequency" with which a protein will attempt to fold will be much much much... much slower.
- Related mistake you are certain to see in the literature: An experimental rate is measured for a large molecular motion (e.g. folding) and the barrier is estimated using the TST value of C. This artificially inflates the estimate of the barrier height to absurd values (e.g. 100 $k_B T$)

Estimating C for folding

- For folding of cooperative globular proteins, the timescale for the attempt frequency can be estimated based on correlation times or diffusion coefficients (Kubelka et al 2004).

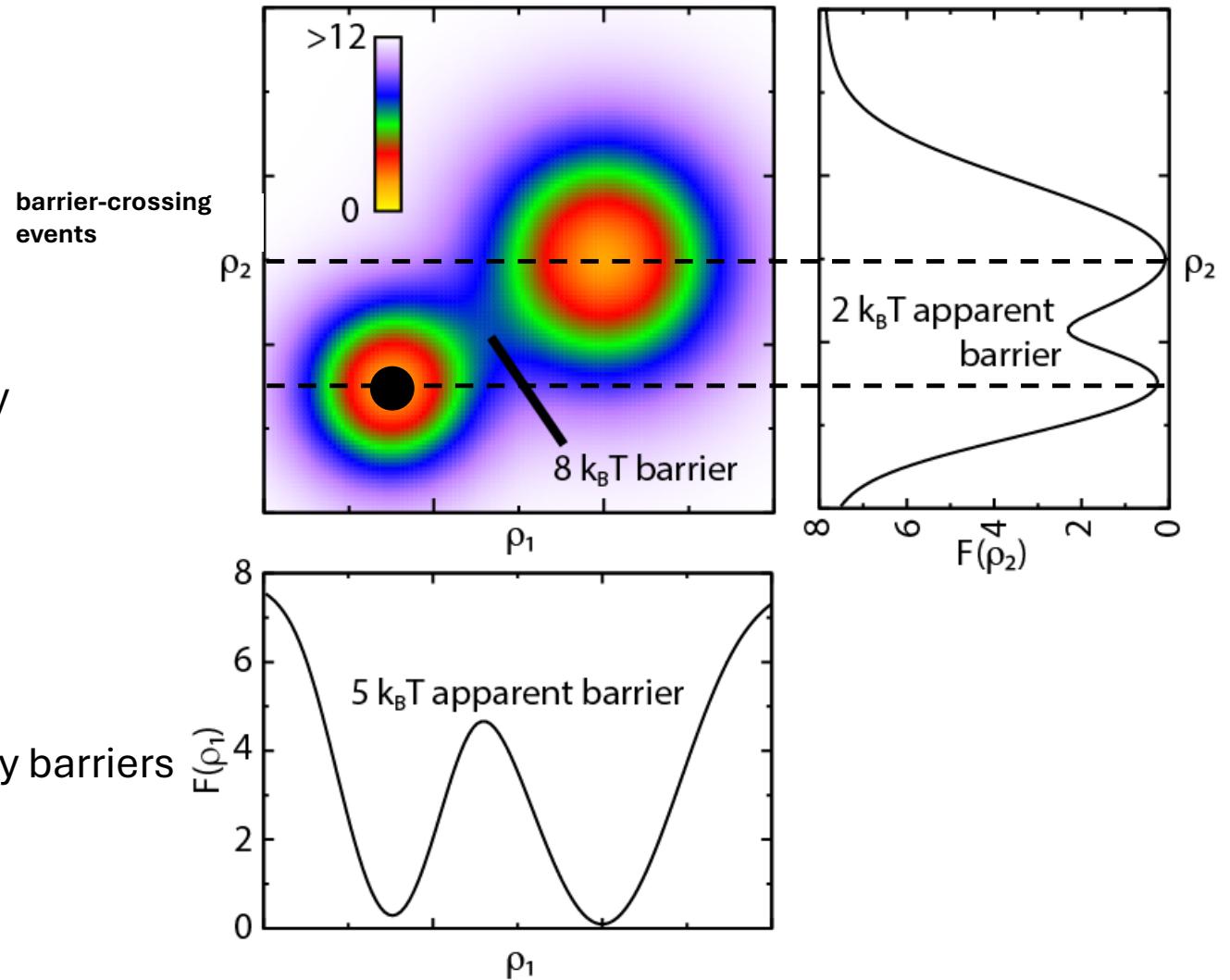
$$\tau_{\text{folding}} = \frac{2\pi k_B T}{\omega_{\min} \omega_{\max} D_{\max}} \exp\left(\frac{\Delta G^*}{k_B T}\right) \approx 2\pi \tau_{\text{corr}} \exp\left(\frac{\Delta G^*}{k_B T}\right)$$



- When this is done, $C \sim 1/\mu\text{s}$, not $0.1/\text{ps}$**
 - That's a difference of 5 orders of magnitude!!!!
- Important reminder: This value of C may not be suitable for anything other than cooperative folding of a single-domain protein.
- Note: Values of $\sim 1/\mu\text{s}$ have been found for other conformational motions, but it is still not a universal value (Whitford et al. 2010b; Whitford et al. 2013)

Another Challenge: accurately calculating the free-energy barrier

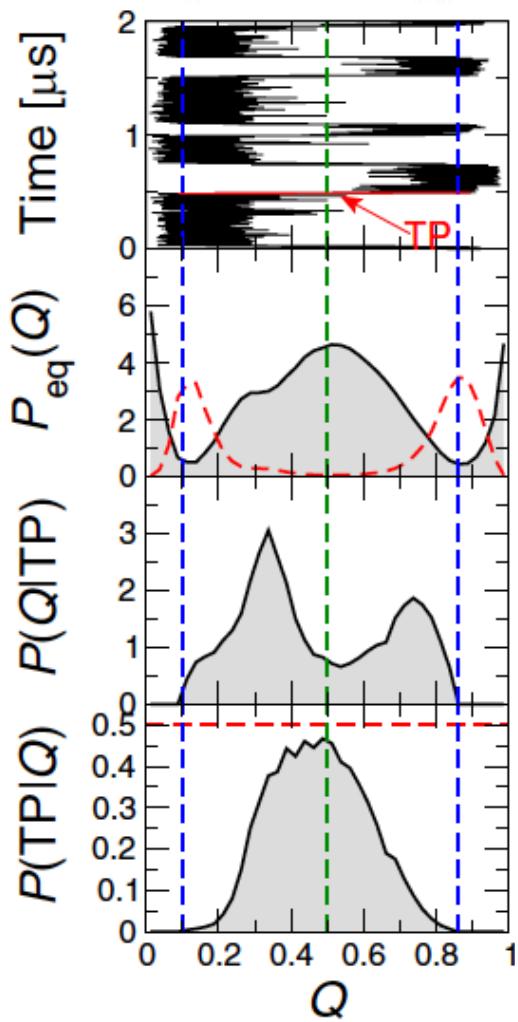
Endless number of possible metrics/probes for describing any process



Poor measures will:
report false positives
underestimate free-energy barriers



Identifying better coordinates: Transition Path Analysis



Q – an arbitrary/putative reaction coordinate

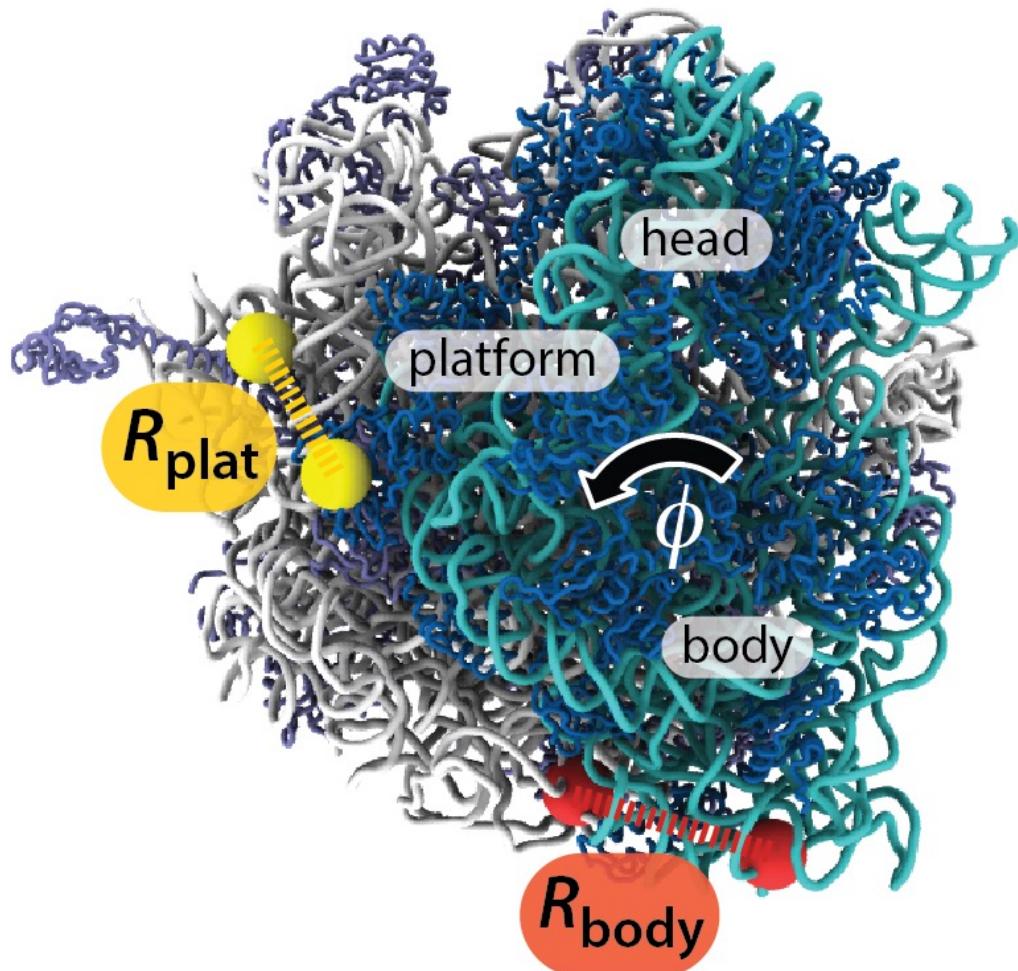
$P(TP|Q)$ - conditional probability of being on a transition path, given a value of the coordinate

For coordinates that capture the TSE: $P(TP|Q)$ should peak at 0.5

- i.e. from TSE, 0.5 probability of going either direction

This would be consistent with an accurate estimate of the probability of being at the TSE

Examples of poor and good coordinates

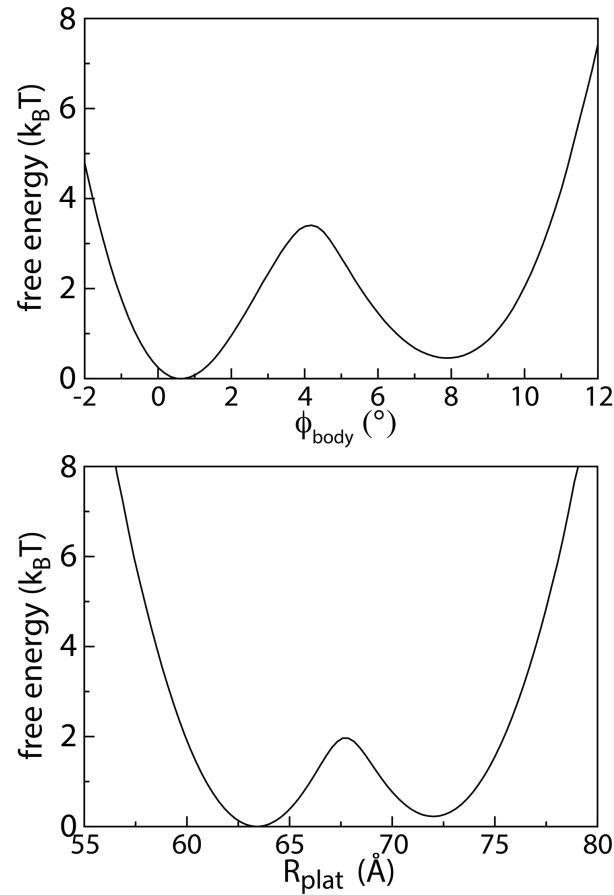


Subunits rotate in the ribosome.

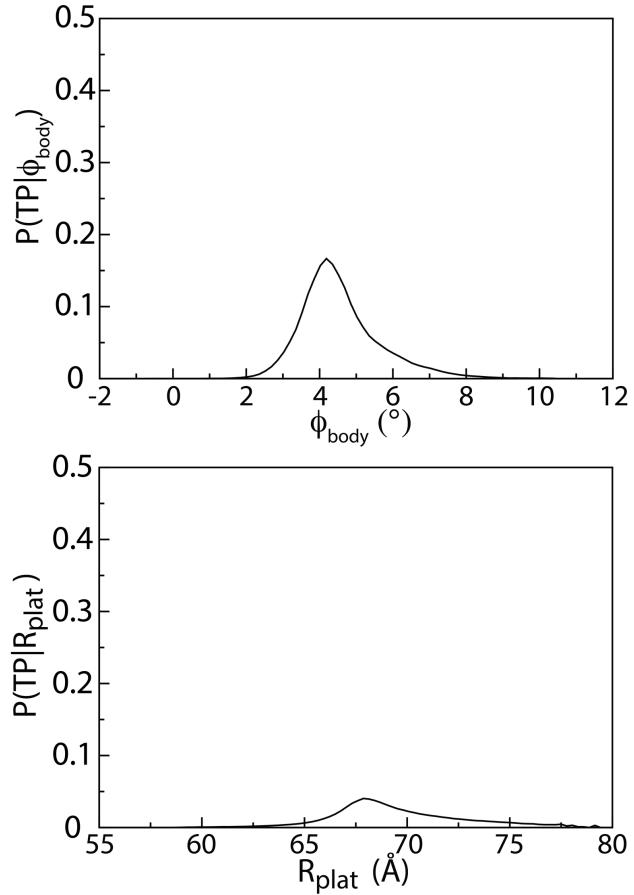
We can describe the dynamics using the angle ϕ , or an interatomic distance R_{plat}

Sub-optimal measures of rotation

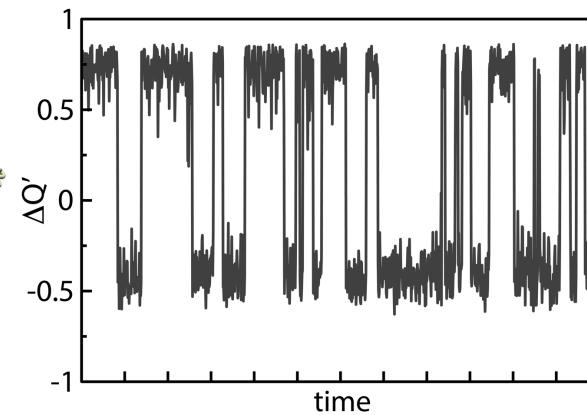
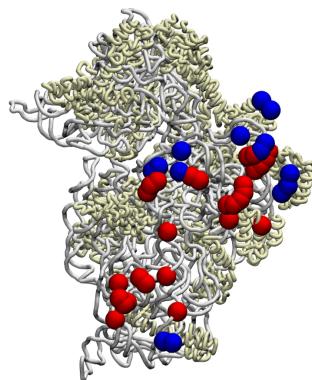
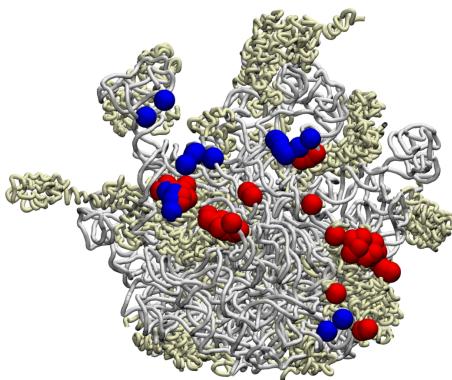
Barrier height, and therefore the implied kinetics, depends on the choice of coordinate



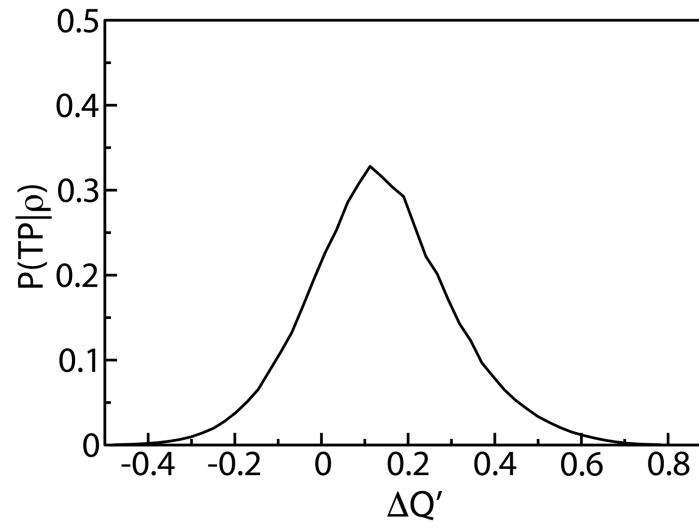
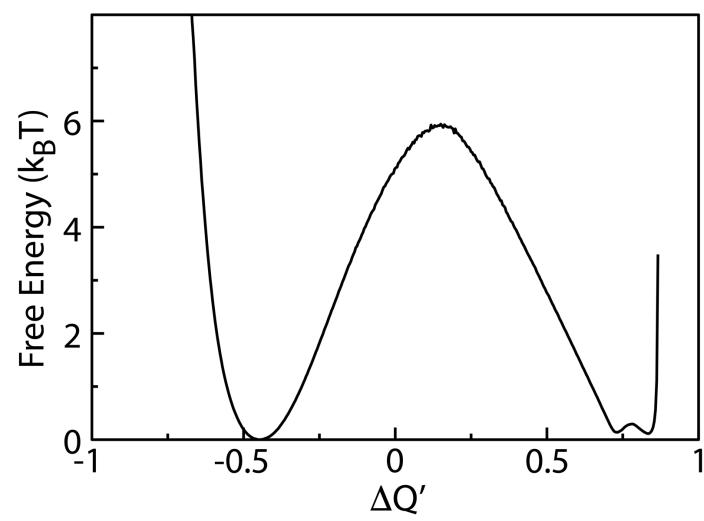
Neither coordinate captures the TSE.
 $P(\text{TP}|\text{Q}) < 0.2$ for all coordinates



Coordinate based on intersubunit contacts is superior



Monitoring formation of rotated/unrotated contacts can more clearly separate the endpoints



$$P(TP|Q) > 0.3$$

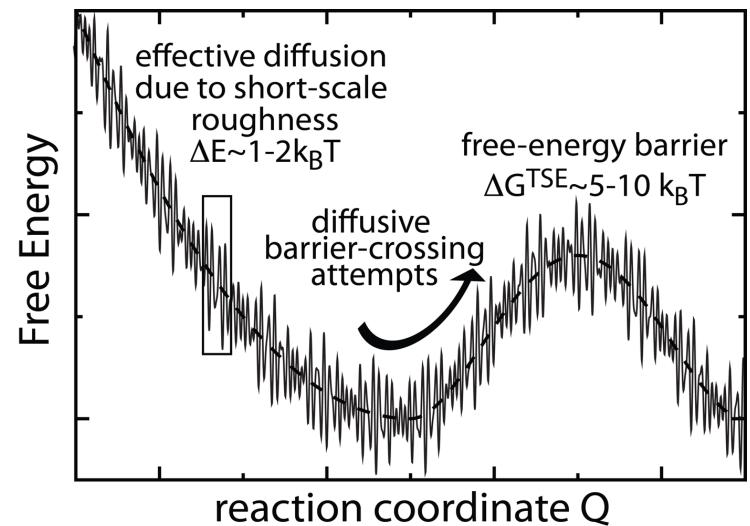
Measures of bridge interactions (red/blue) reveals a much larger free-energy barrier!!!!

Approach 3: calculating kinetics from diffusion coefficients and free-energy profiles

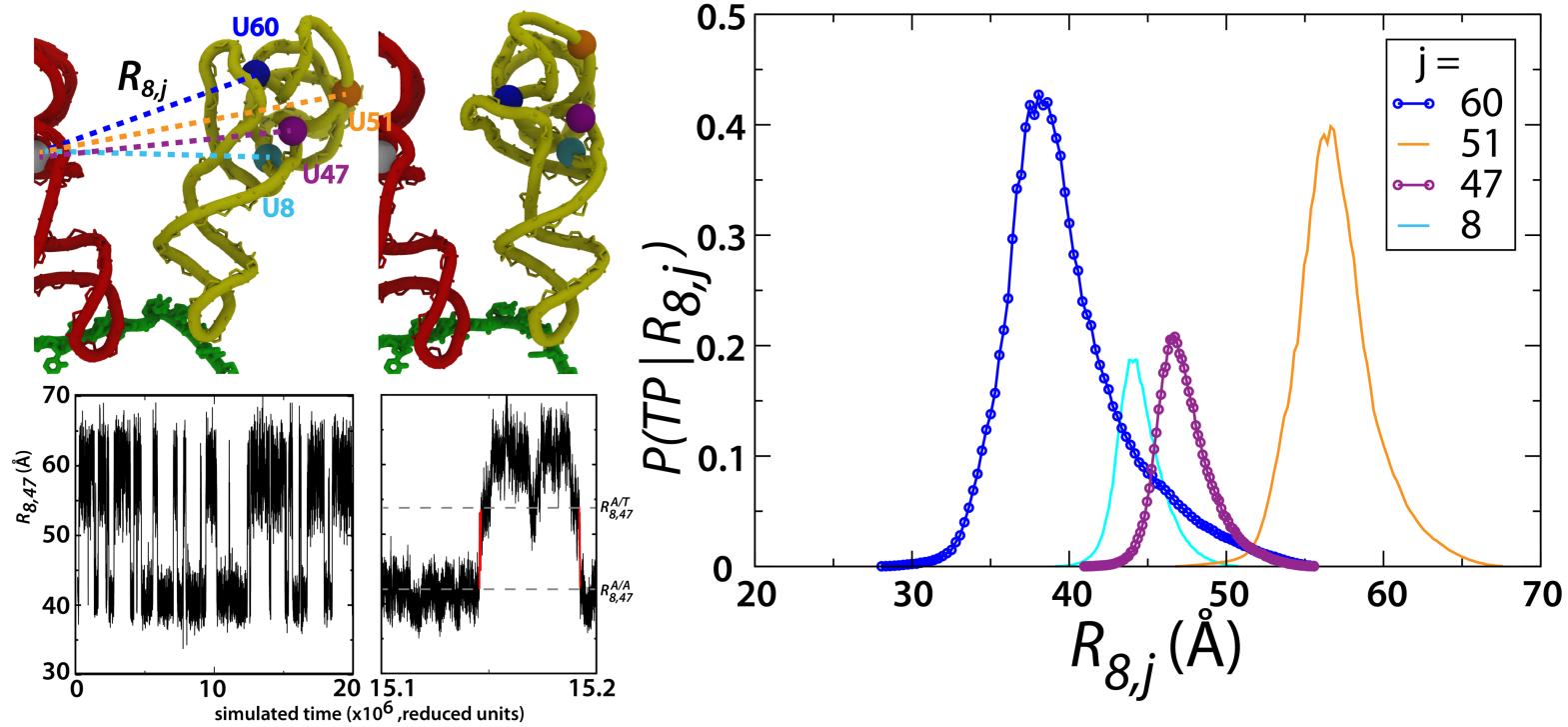
$$\begin{aligned}\tau &= \frac{1}{R} = \frac{1}{C} \exp(\beta\Delta F) \\ &= \int_{Q_i}^{Q_f} \exp[F(Q')/kT]/D(Q') dQ' \int_{\infty}^{Q'} \exp[-F(Q'')/kT] dQ''\end{aligned}$$


Challenges:

- You need a good coordinate for calculating the free-energy profile
 - Same challenge as previous approach
- You need to have a decent estimate of the diffusion coefficient along the same coordinate
 - This diffusion coefficient needs to be the desired units (e.g. meters and seconds)



Another example comparing coordinates



This shows motion of a tRNA molecule and the $P(TP|R)$ values for different interatomic distances. One of the coordinates is pretty good, with $P(TP|R) \sim 0.45$

Noel et al., 2014

Calculating diffusion coefficients

Explicit-solvent (AMBER) simulations

>3000 simulations (15 ns, each)

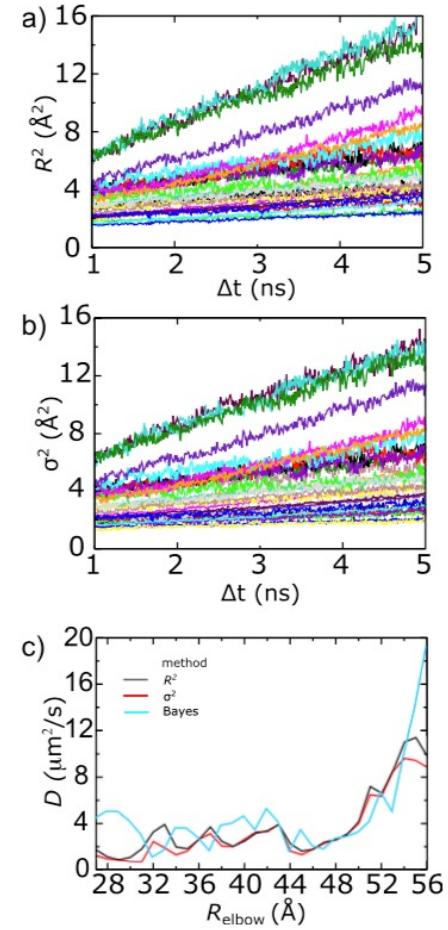
Calculate diffusion coefficients from the explicit solvent model and use it to determine the mfpt from the free-energybarrier.

$$R^2(\Delta t) = \langle (R_{\text{elbow}}(t + \Delta t) - R_{\text{elbow}}(t))^2 \rangle$$

$$D(R) = \lim_{\Delta t \rightarrow \infty} \frac{1}{2} \frac{\partial R^2(\Delta t)}{\partial \Delta t},$$

$$\tau = \frac{1}{R} = \frac{1}{C} \exp(\beta \Delta F)$$

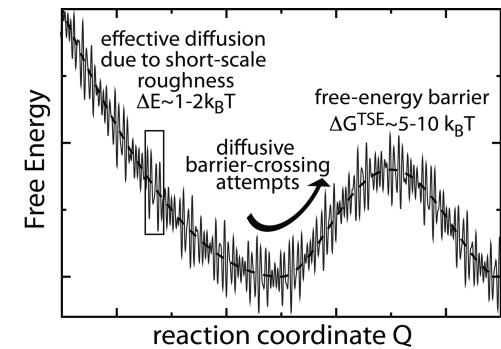
$$= \int_{Q_i}^{Q_f} \exp[F(Q')/kT] / D(Q') dQ' \int_{\infty}^{Q'} \exp[-F(Q'')/kT] dQ''$$



Yang, et al. 2019

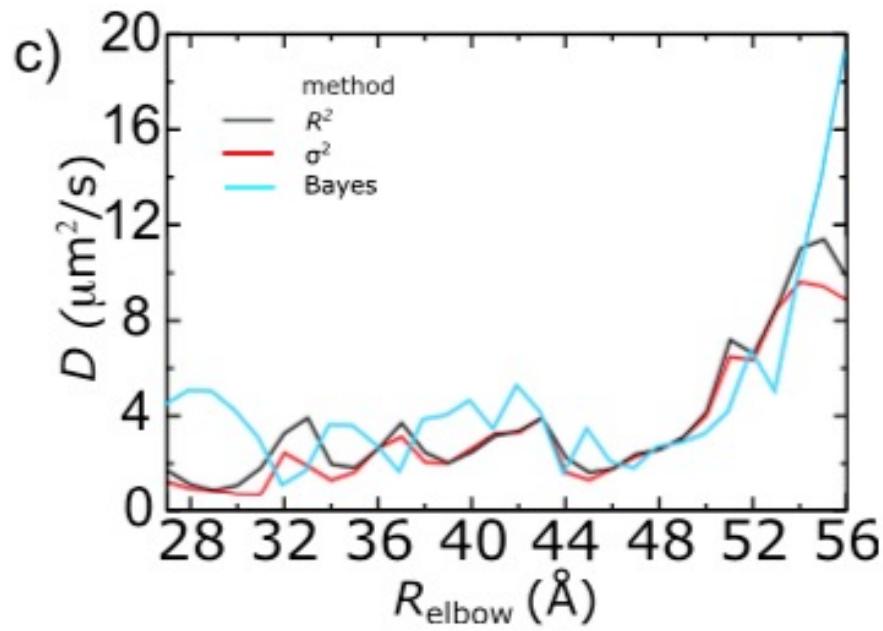
Approach 4: directly calibrate the time unit

- In many simplified models, the overall energetics is proposed, while the diffusion is not tuned/calibrated
- If you estimate the time unit from energy, mass and length, and then calculate the diffusion coefficient, you expect it to be too large
- Propose: Use the calculated and reference values of a diffusion coefficient to identify the time unit.
- Logic: The diffusion coefficient describes the characteristic time associated with a diffusive movement. If the calculated and reference diffusion coefficients describe the same underlying motion, then you calibrate your time unit by equating the time to enter the diffuse.
- Pros include:
 - No need to calculate a free-energy barrier
 - No need to identify an optimal coordinate
 - May directly use the estimated unit to report timescales
 - e.g. “N simulated steps equals M ps”
 - If done correctly, you will get the same values as approaches 2 and 3.

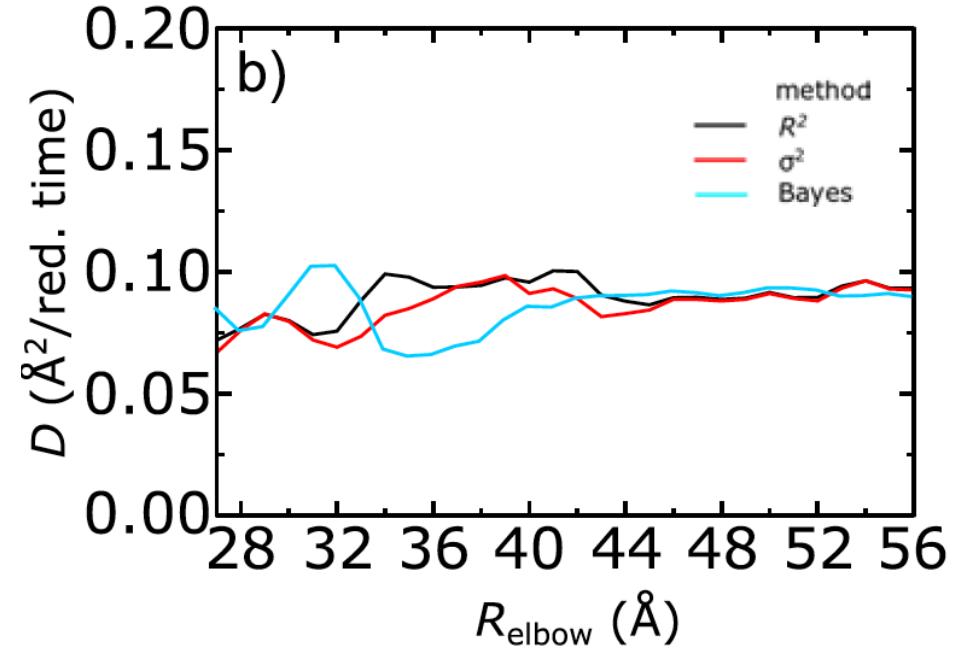


Example: same tRNA motion as before

Diffusion coefficients
from explicit-solvent sim.



Diffusion coefficients in an
all-atom SMOG model



Equate $2 \mu\text{m}^2/\text{s}$ and $0.1 \text{\AA}^2/\tau_{\text{R.U.}}$ which gives $\tau_{\text{R.U.}} = 0.1 \text{\AA}^2/(2 \mu\text{m}^2/\text{s}) = 500 \text{ps}$

Yang, et al. 2019

Imagine another application of this approach

- We will imagine a hypothetical model (inspired by MiChroM) where:
 - Each bead represents ~100k base pairs of DNA
 - Sequential beads are roughly 1 μm apart
 - Each bead is given mass of 1 (reduced mass unit)
 - Temperature of the simulation is ~ 1 (reduced unit, or $k_{\text{B}}T \sim 1$)
- Imagine a simulation of a single particle of this system
 - Expect a diffusion coefficient that is of order $1 \mu\text{m}^2/\tau_{\text{R.U.}}$.
 - Using Stokes-Einstein scaling (i.e. $\sim 1/\text{radius}$), we can estimate the diffusion of a particle of radius of $\sim 1/2$ micron
 - For tRNA (radius $\sim 3 \text{ nm}$) to have diffusion coefficients estimated at $1 \mu\text{m}^2/\text{s}$
 - For 0.5 micron radius, we expect a diffusion coefficient that is roughly 150-200 times smaller, or roughly $0.005 \mu\text{m}^2/\text{s}$
 - Equate diffusion coefficients: $1 \mu\text{m}^2/\tau_{\text{R.U.}} = 0.005 \mu\text{m}^2/\text{s}$ give us $\tau_{\text{R.U.}} = 200 \text{ seconds}$
- Reflect
 - $\tau_{\text{R.U.}}$ estimated as:
 - 50ps for tRNA in all-atom SMOG model
 - 200 seconds for the chromosome-like model
 - What might the software (e.g. OpenMM) label the time unit?
 - Yep, picoseconds for both!!!!

Final take home

- Use these approaches suitably and you should be ok.
- Just be careful.
- When in doubt, discuss with an expert.

References

- Definitions of quantities and Units
 - SI Units - https://en.wikipedia.org/wiki/International_System_of_Units
 - Kelvin - <https://en.wikipedia.org/wiki/Kelvin>
 - $k_B T$: [https://en.wikipedia.org/wiki/KT_\(energy\)](https://en.wikipedia.org/wiki/KT_(energy))
 - <https://manual.gromacs.org/documentation/2019/reference-manual/definitions.html>
- Allen and Tildesly. *Computer simulations of liquids*. Clarendon Press
- Best and Hummer 2005 : Reaction coordinates and rates from transition paths. *Proc. Nat. Acad. Sci. USA.* **102**, 6732-6737.
- Best et al. 2013 : Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Nat. Acad. Sci. USA.* **110**, 17874-17879.
- Bryngelson and Wolynes. 1989 : Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem.* **93**, 6901-6915.
- Frenkel and Smit. *Understanding Molecular Simulations: From algorithms to applications*. Academic Press
- Jackson et al. 2015 : Exploring the balance between folding and functional dynamics in proteins and RNA. *Intl. J. Mol. Sci.* **16**, 6868-6889.
- Kouza et al. 2006 : Effect of finite size on cooperativity and rates of protein folding. *J. Phys. Chem A.* **110**, 671-676.
- Kubelka et al. 2004 : The protein folding ‘speed limit’. *Curr Opin Struct Bio* **14**, 76-88.
- Levi and Whitford, 2019 : Dissecting the energetics of subunit rotation in the ribosome. *J. Phys. Chem. B.* **123**, 2812—2823.
- Noel et al. 2014 : Capturing transition paths and transition states for conformational rearrangements in the ribosome. *Biophys. J.* **107**, 2872-2881.
- Noel et al. 2016 :SMOG 2: A versatile software package for generating structure-based models. *PLoS Comp Biol.* **12**, e1004794..
- Whitford et al. 2009 : An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields. *Prot. Struct. Func. Bioinfo.* **75**, 430-441.
- Whitford et al. 2010a : Accommodation of aminoacyl-tRNA into the ribosome involved reversible excursions along multiple pathways. *RNA.* **16**, 1196-1204.
- Whitford et al. 2010b : Connecting energy landscapes with experimental rates for aminoacyl-tRNA accommodation. *J. Amer. Chem. Soc.,* **132**, 13170-13171.
- Whitford et al. 2013 : Connecting the kinetics and energy landscape of tRNA translocation on the ribosome. *PLoS Comp Biol.* **9**, e1003003.
- Yang et al. 2019 : Diffusion of tRNA inside the ribosome is position-dependent. *J. Chem. Phys.* **151**, 085102.
- Zwanzig, *Nonequilibrium Statistical Mechanics*, Oxford University Press
- Zwanzig 1988 : Diffusion in a rough potential. *Proc. Nat. Acad Sci. USA.* **85**, 2029-2030.