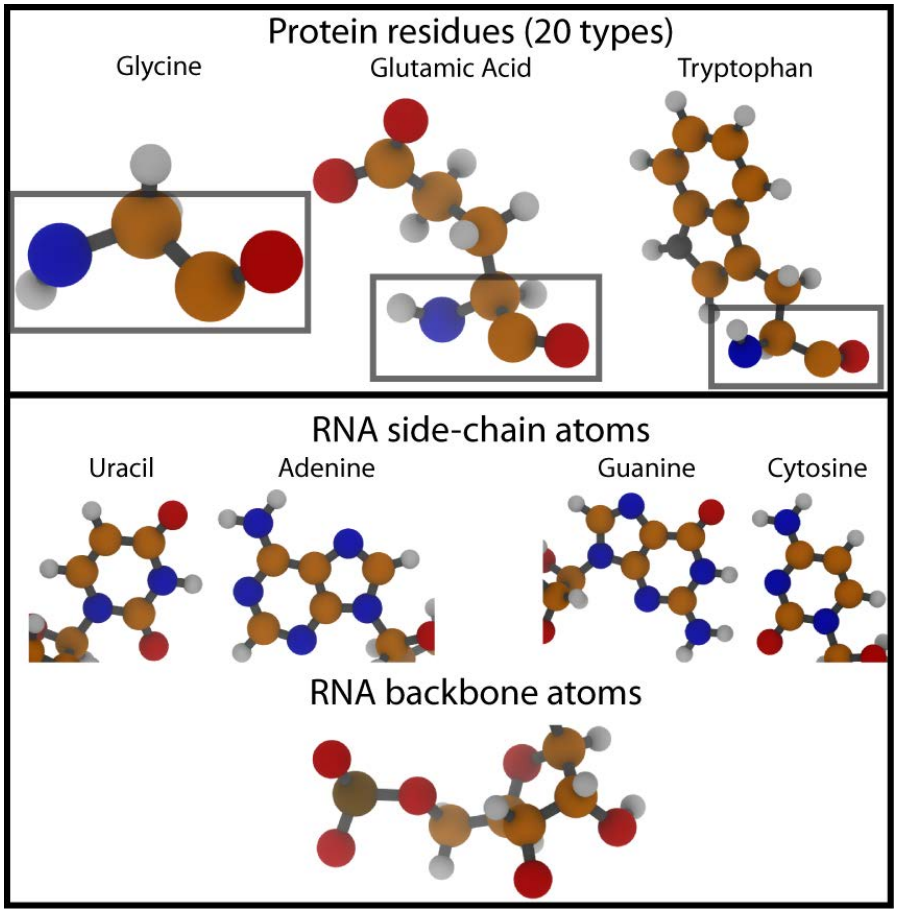
How HPC is transforming the study of biological machines

HPC Day
UMass Dartmouth
May 25th 2017

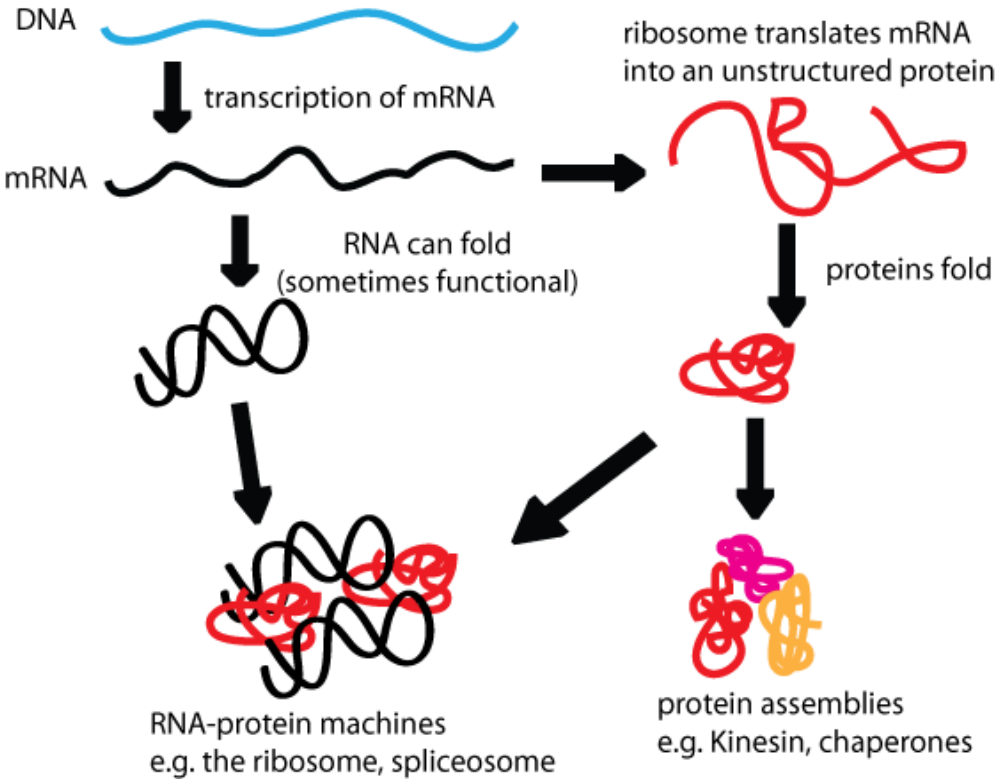
Paul Whitford
Department of Physics
Northeastern University

From Polymer Sequences to Molecular Assemblies

Residues form biopolymer chains



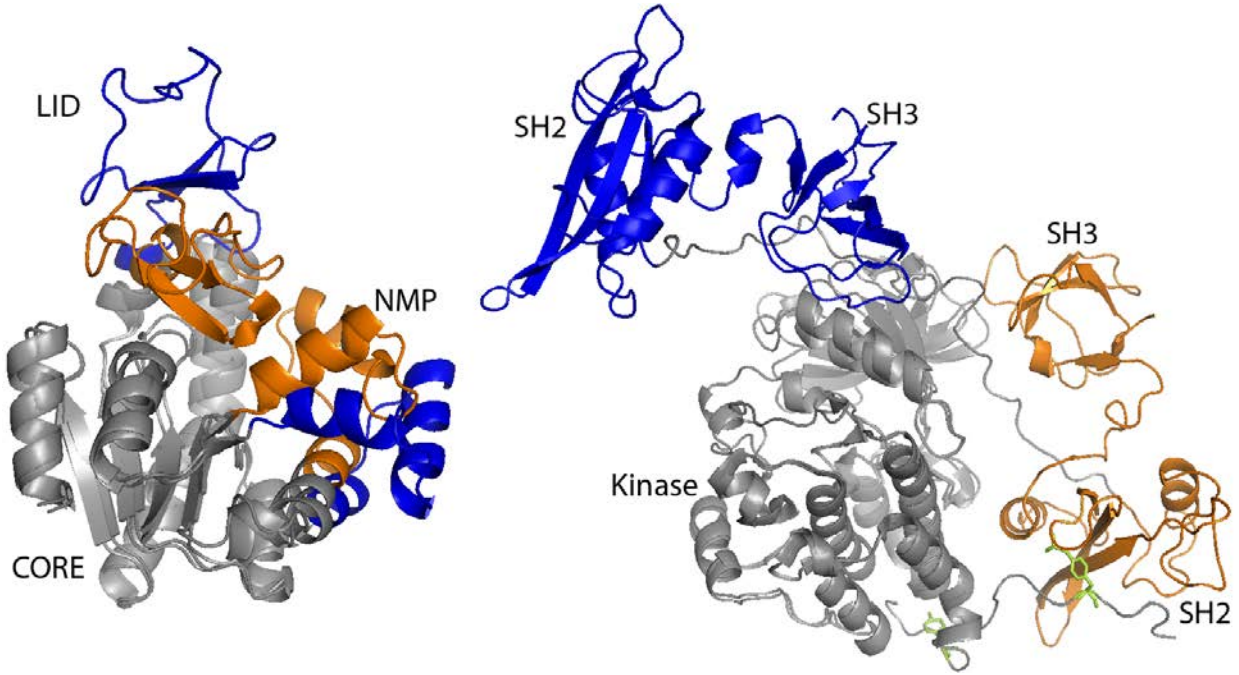
Biopolymers form functional assemblies



Functional aspects of molecules often require folding and assembly.
Biomolecular machines are extremely dynamic and flexible.

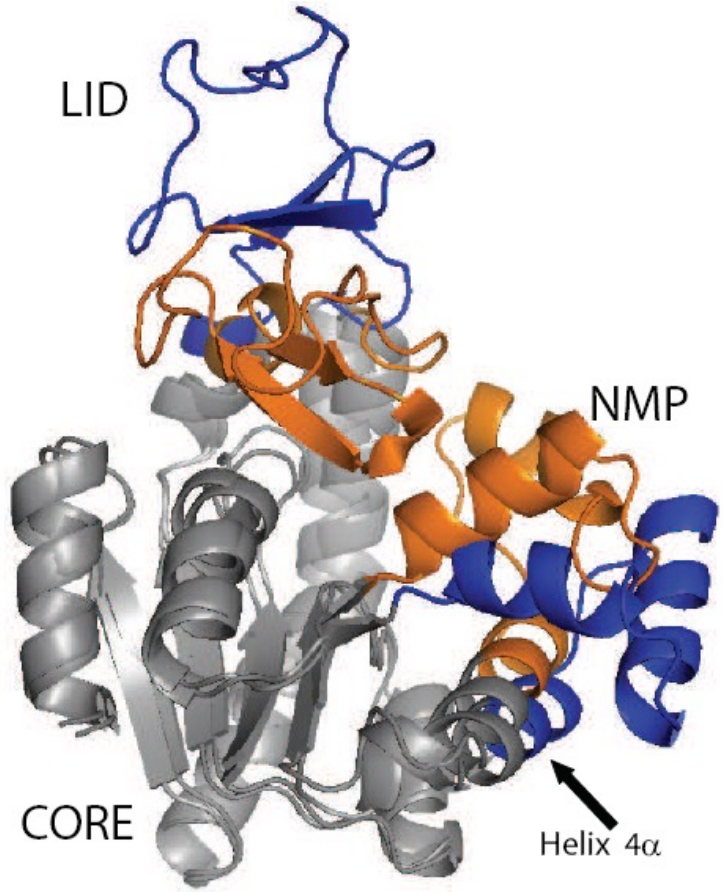
Conformational Changes in Proteins

- Believed to be important in many signaling pathways
- Can be rate limiting
- Static structures *may* suggest a “hinge-rod” description



Lieser, Shaffer & Adams, JBC, 2006
Wolf-Watz, Thai, Henzler-Wildman, Hadjipavlou, Eisenmesser & Kern, **NSB**, 2004

Adenylate Kinase



214 residues ~ 3000 atoms (excluding solvent)

LID and NMP domains undergo **5-10 Å** displacements during ligand binding

Classic system used for model/method development.

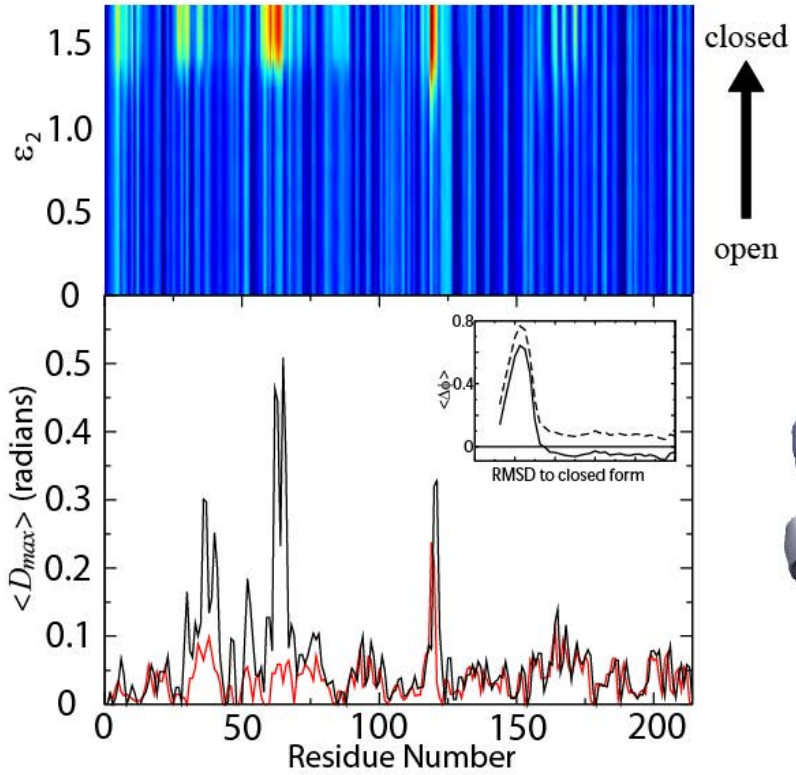
Computationally tractable with many models

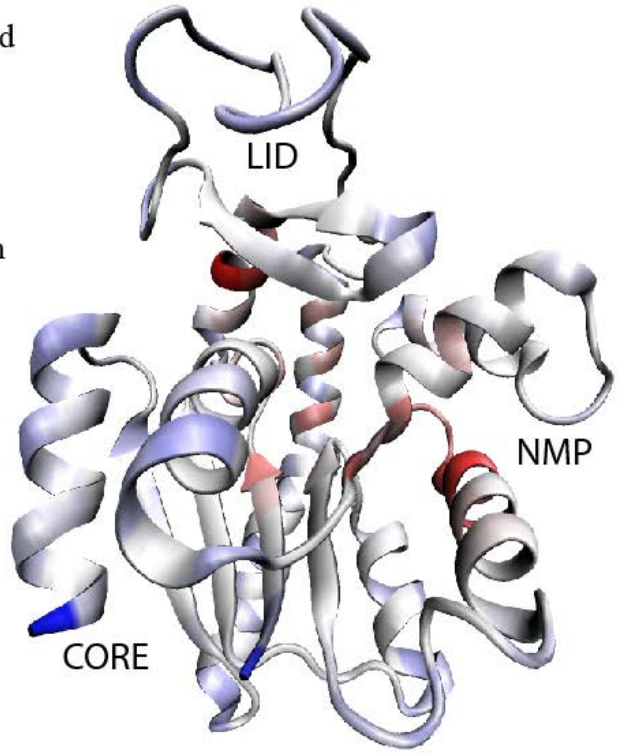
Extensively studied in the 2000s

Whitford, Miyashita, Levy & Onuchic, (2007) *JMB* **366**, 1661-1667

Predicted and Observed Cracking (Local unfolding/disorder)

Strain results in partial unfolding/refolding

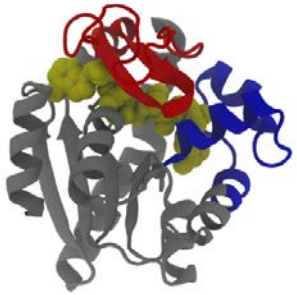




Whitford, Miyashita, Levy & Onuchic, (2007) *JMB* **366**, 1661-1667

Cracking is fundamentally distinct from dynamics in macroscopic systems

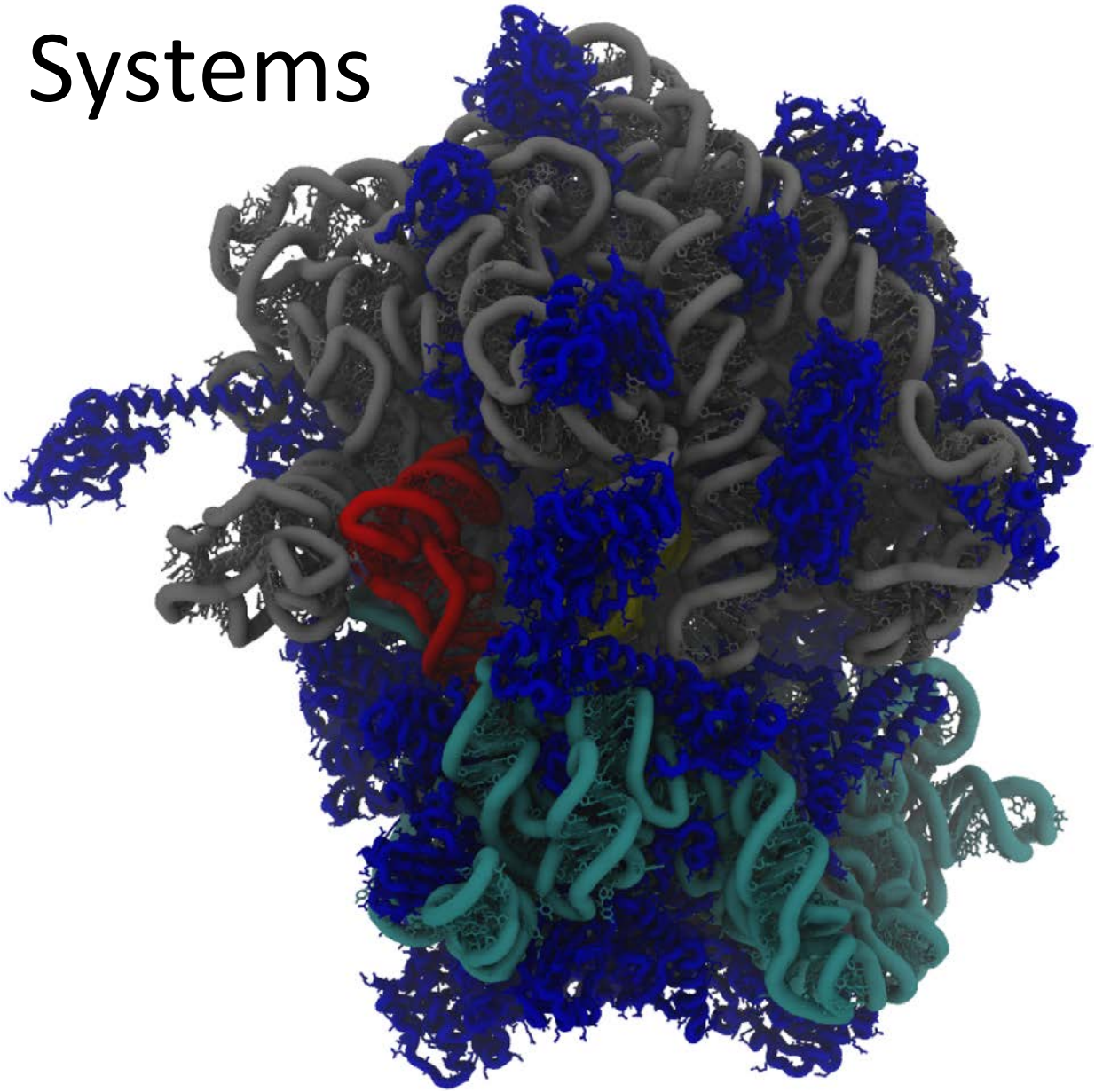
Larger-Scale Systems



Adenylate Kinase
214 protein residues
~ 20 kDa in mass

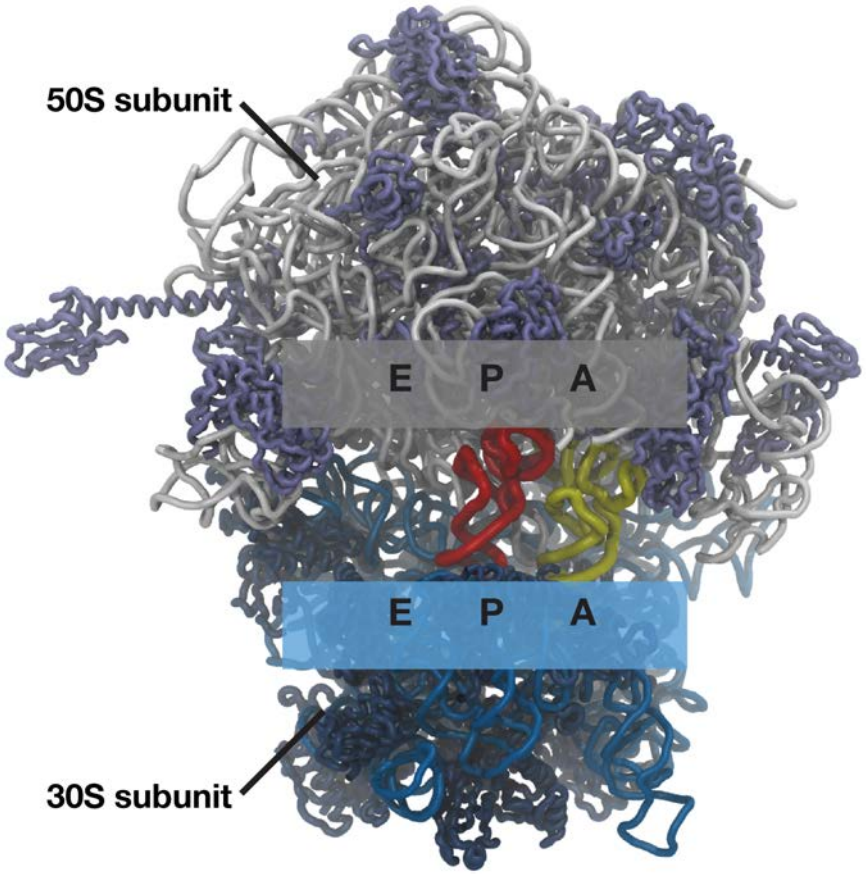
Paradigm system for exploring energetics of functional transitions in biopolymers.

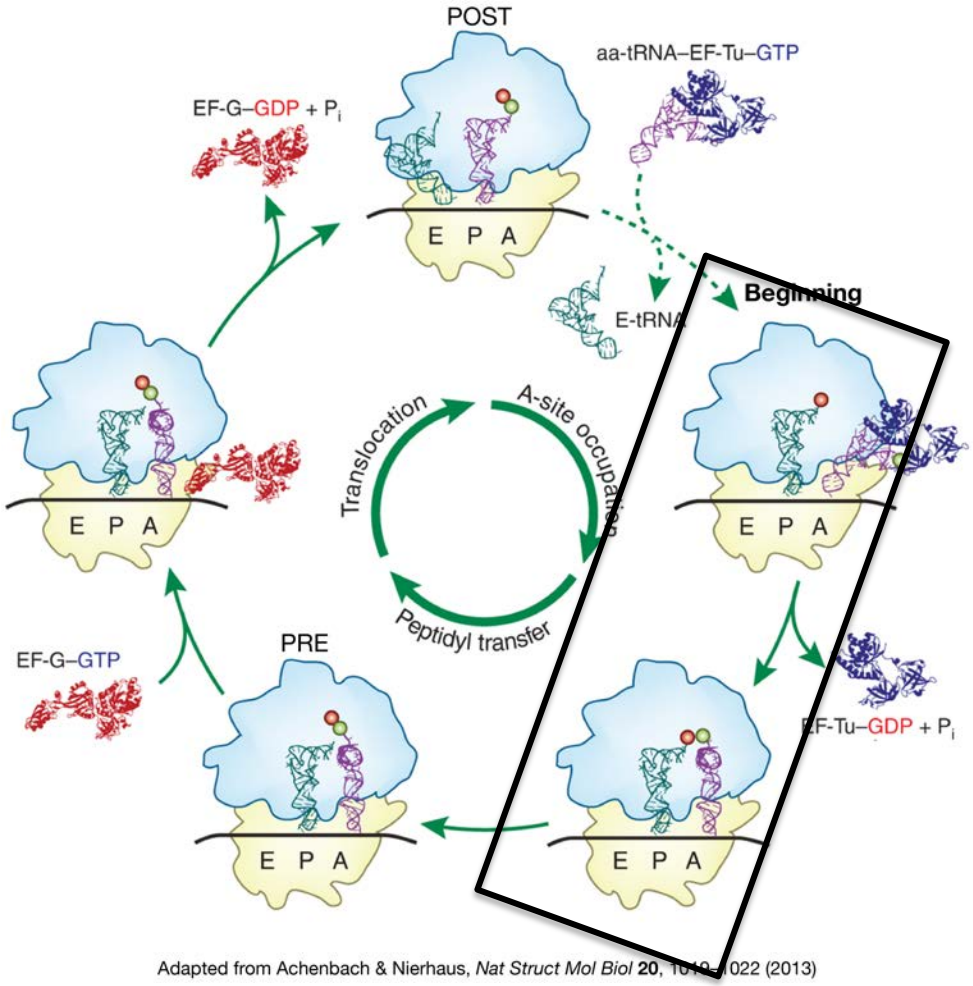
Many questions are computationally tractable



The 70S ribosome: ~4,000 RNA residues, >10,000 protein residues. > 2 MDa in mass.
Is there a prominent role of disorder in larger machines?

70S Ribosome





accommodation

Over 50 proteins
3 RNAs w/ ~4000 nt
~150,000 atoms

PDB ID: 4V6F
Jenner et al. (2010)

First simulation of a ribosome

Sanbonmatsu and colleagues (early 2000s)

Supplementary Movie 1:
Simulating movement of transfer RNA into the ribosome during decoding

Sanbonmatsu*, K.Y., Joseph, S. and C.S. Tung
Los Alamos National Laboratory

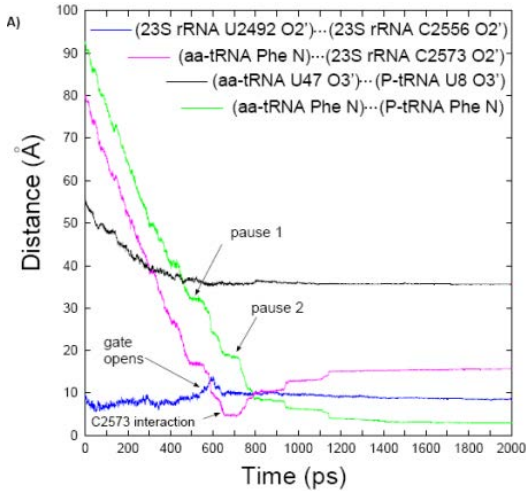
Explicit Solvent Targeted Molecular Dynamics

$N_{atoms} = 2.64 \times 10^6$

ASCI Q Machine (LANL)

***corresponding author: kys@lanl.gov**

www.t10.lanl.gov/kys



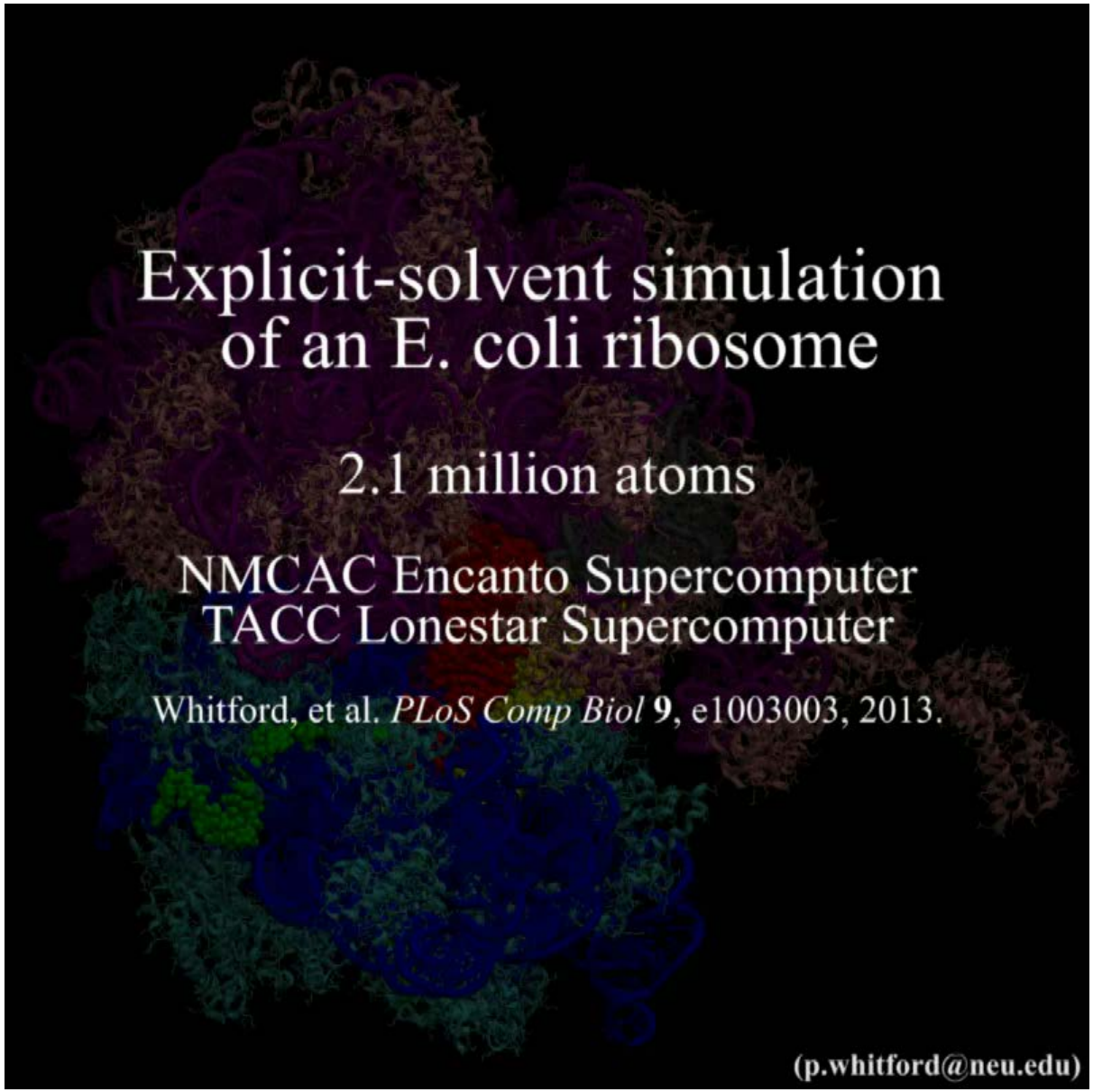
2 nanosecond simulation with targeting forces added
Required the ASCI Q Machine (#2/3 on top500 in 2003/2004)

Extending the timescale of detailed models

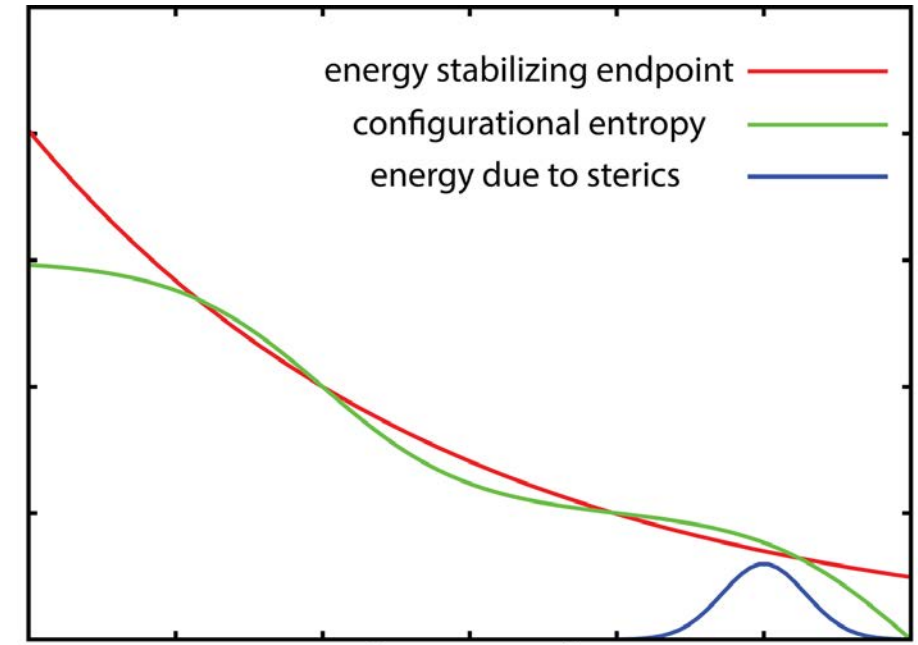
- Amber 99 forcefield
- Water, ions, hydrogen atoms explicitly represented
- General parameterization used for all simulations
- 70S Ribosome: [$MgCl_2$]=7mM, [KCl]=100m
- Simulations are typically nanosecond timescales (here, 200-1400 ns per run. 13.5 μs total) -- ~2000 cores for 5 months – Now Anton 2(Shaw group) can perform in ~8 hours

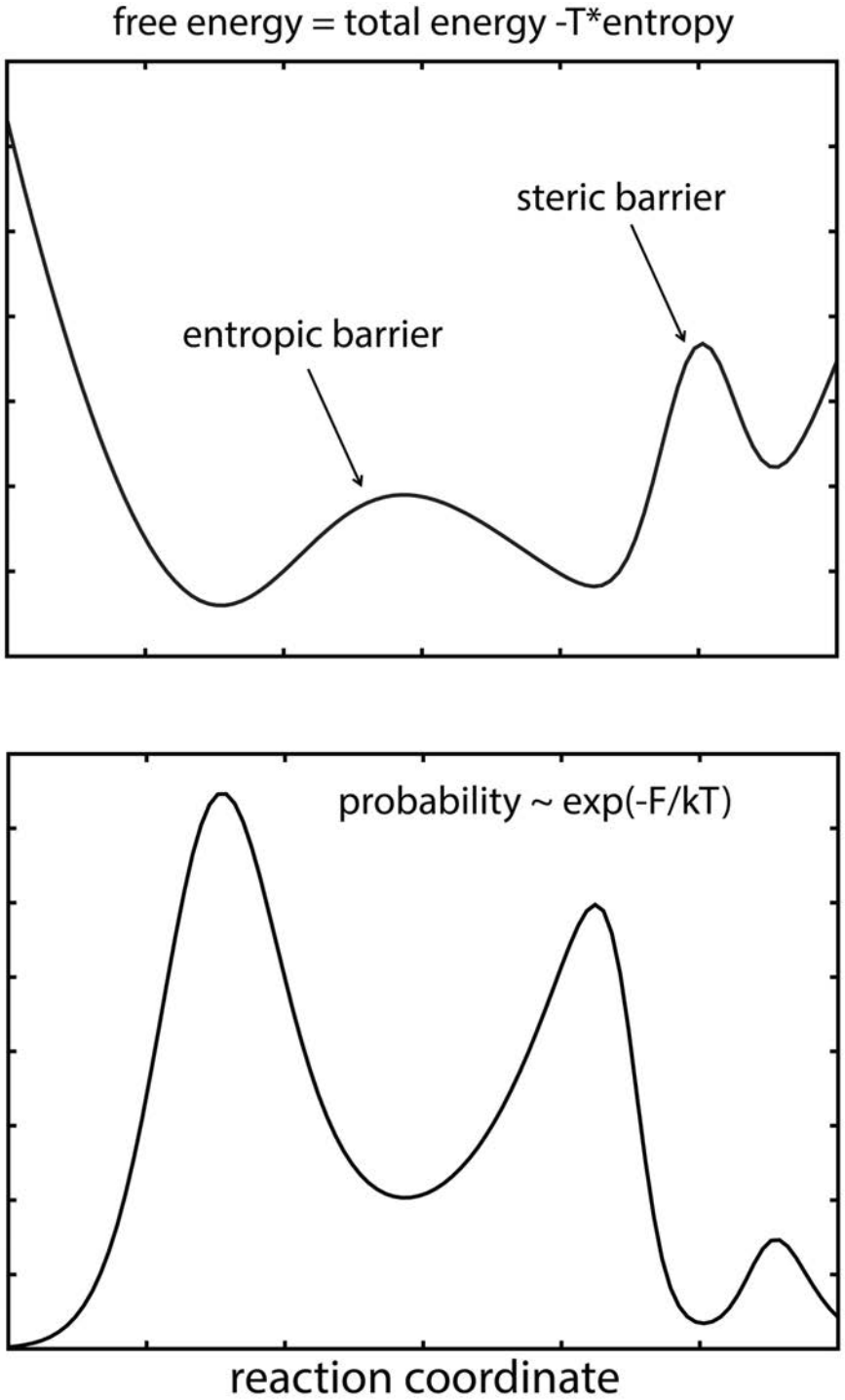
- Short-timescale (<100 ns) fluctuations can be observed
- **Biological timescales (many milliseconds) are not accessible**

PCW et al. *JACS* 2010
PLoS Comp Bio 2013



Applying simpler models (2008)



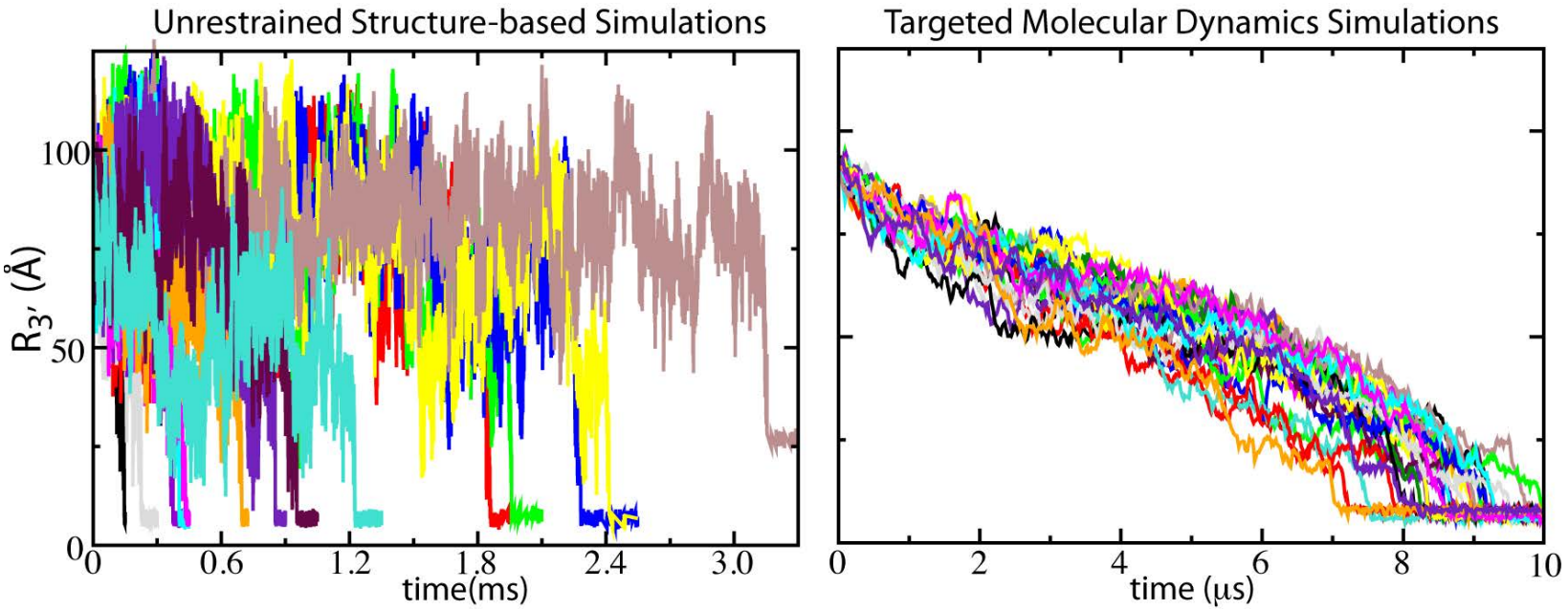


The Structure-based (SMOG) Model:

- Include all heavy (non-hydrogen atoms) atoms
- A/A state is the lowest effective-enthalpy state
- Consistent with the notion of tRNA as a 'molecular spring' (Frank et el. *FEBS* 2005), where tRNA has strain energy (Miyashita, Onuchic, Wolynes, *PNAS* 2003)

All-atom structure-based "SMOG" models (PCW et al. *Proteins* 2009)

Stochastic barrier-crossing events



Enabled the study of statistical properties of large-scale rearrangements

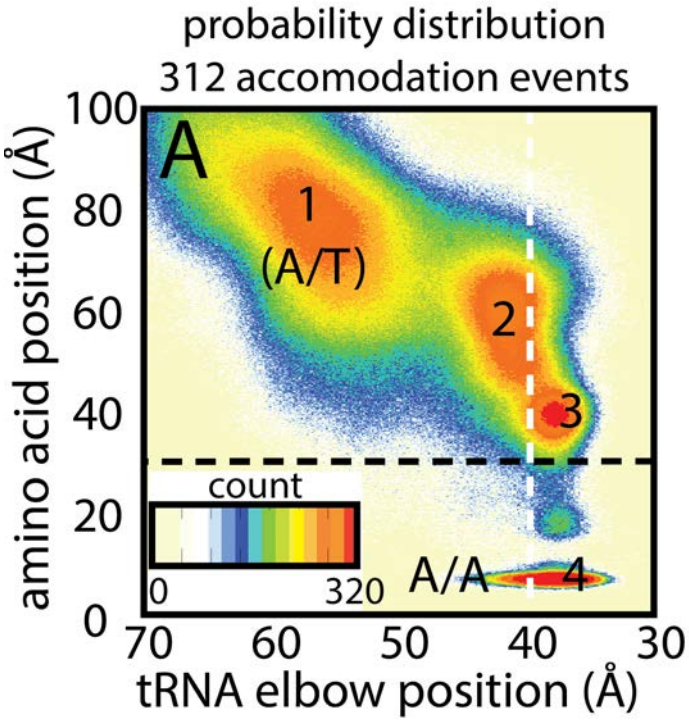
Even with simpler model, each simulation typically required weeks using 128 cores

Hundreds, or thousands, of simulated events are required

Performed using the Encanto Supercomputer (#8 on top500 in 2008)

PCW et al. *RNA* 2010

Improved statistics enables the study of distributions



tRNA accommodation with a SMOG model

Representative trajectory described in:
Whitford et al. *RNA* **16**, 1196-1402, 2010.

NMCAC Encanto Supercomputer

Provides signatures of the role of flexibility

PCW et al *RNA* 2010

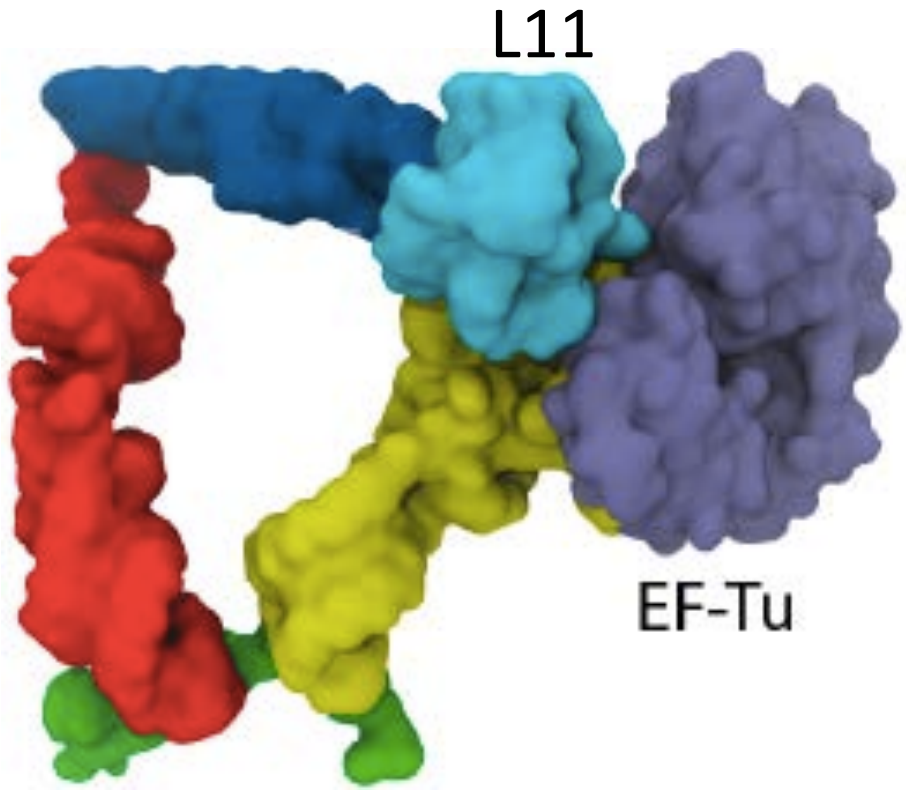
Dissecting the physical factors that govern kinetics: *Ongoing*

Observations:

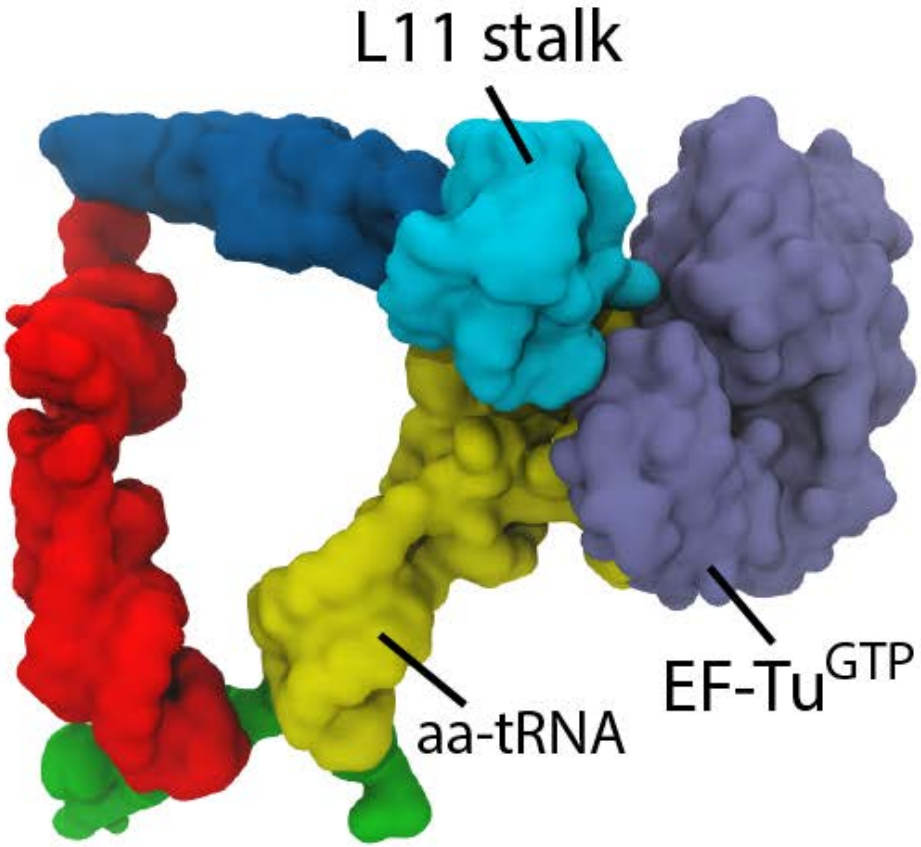
Experiments suggest EF-Tu dissociation can be slower, or faster than accommodation of aa-tRNA (Pape et al, EMBO J 1999) and that aa-tRNA accommodation rates can significantly depend on conditions (Johannson et al. *Mol Cell* 2008).

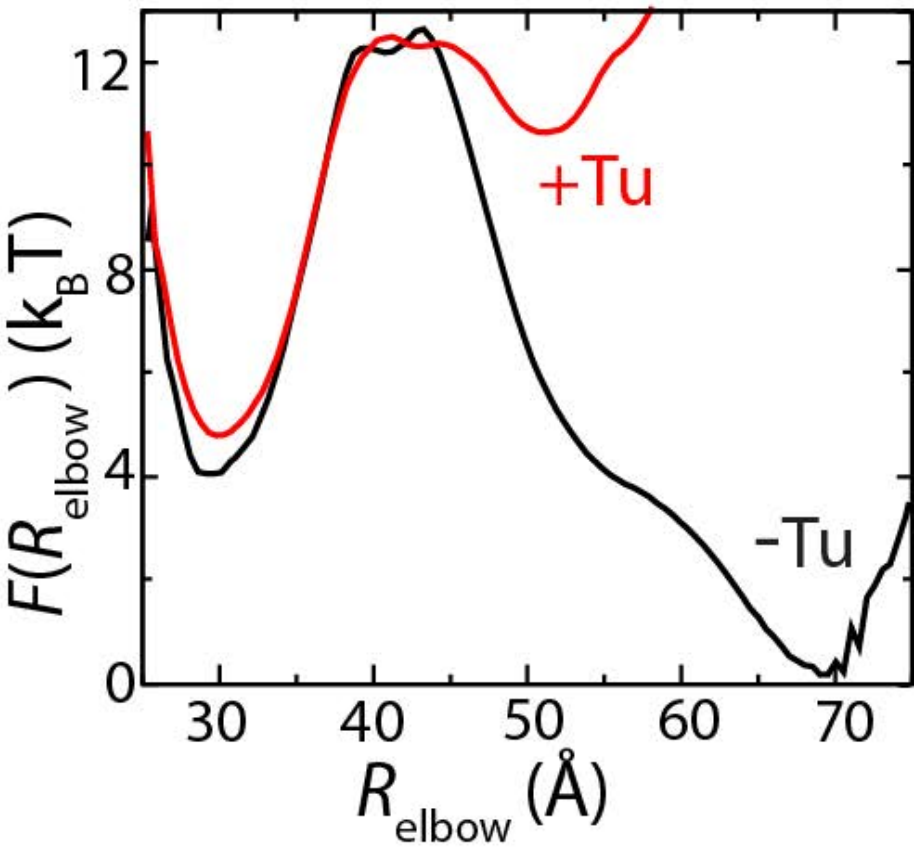
Models predict there are at least two conformational substeps: elbow accommodation and 3'-CCA accommodation

Question: Can the rate of EF-Tu dissociation from the ribosome impact accommodation rates?



Controlling Kinetics





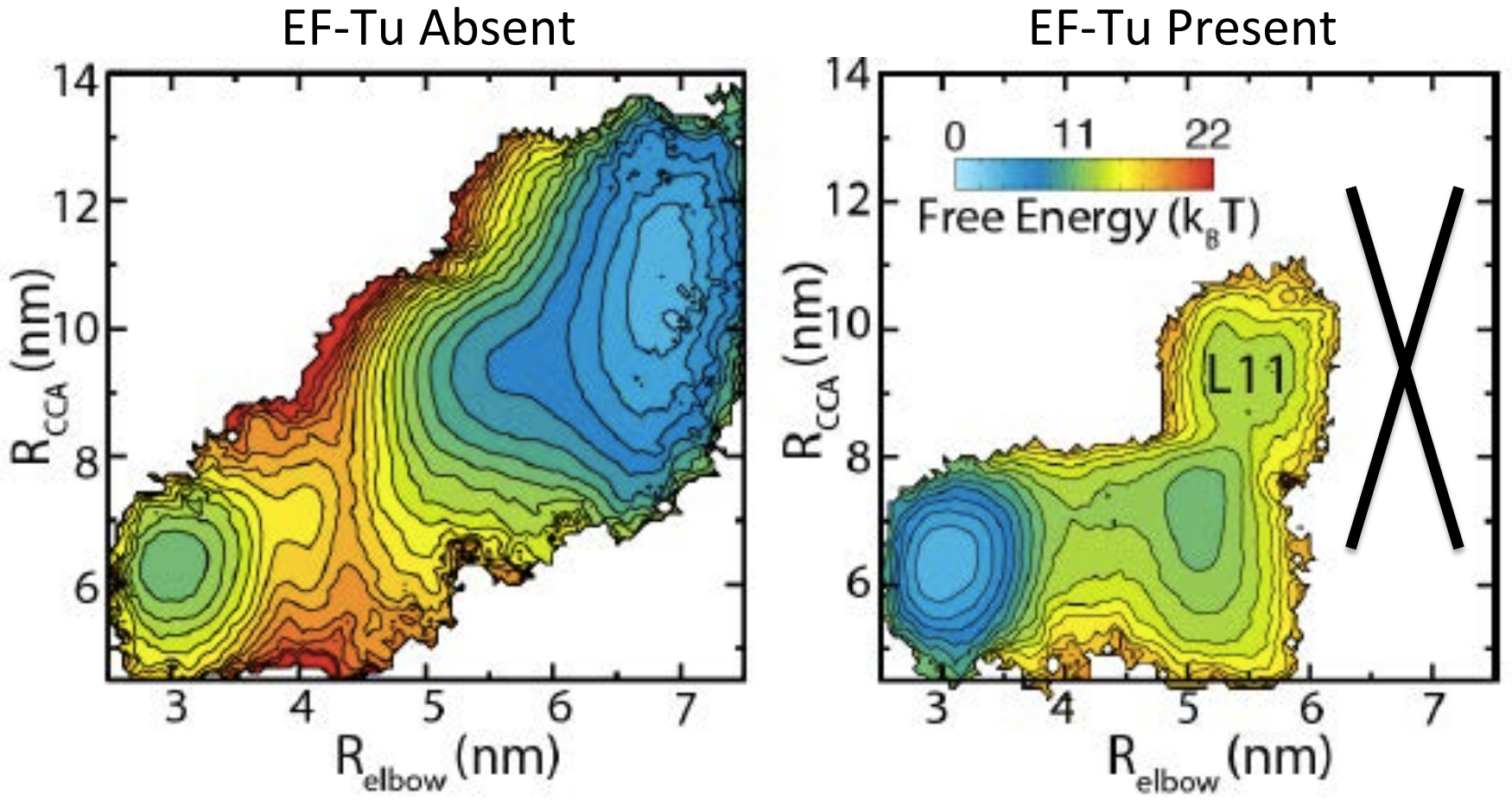
<u>Barrier much larger in the absence of EF-Tu (orders of magnitude slower kinetics)</u>

Each profile requires ~150 simulations, each for two days on 28 cores – aim to calculate profiles under a range of conditions

Red: EF-Tu bound to ribosome, tRNA released
Black: EF-Tu absent

Jeff Noel and PCW *Nature Comm.* (2016)

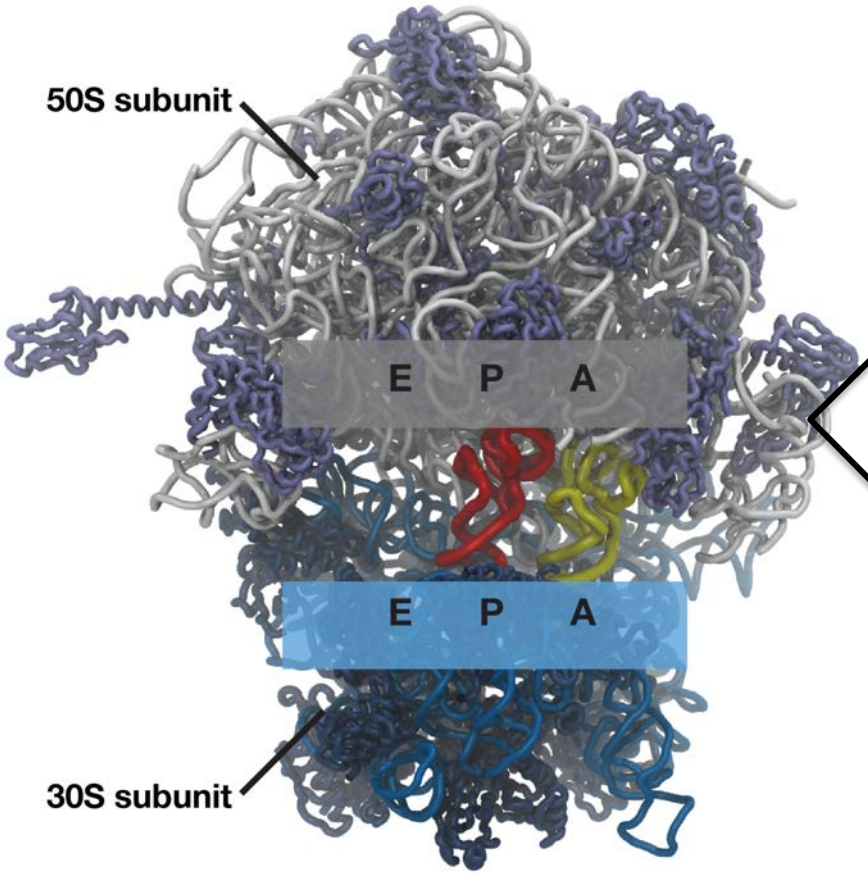
EF-Tu prevents sampling of a disordered ensemble



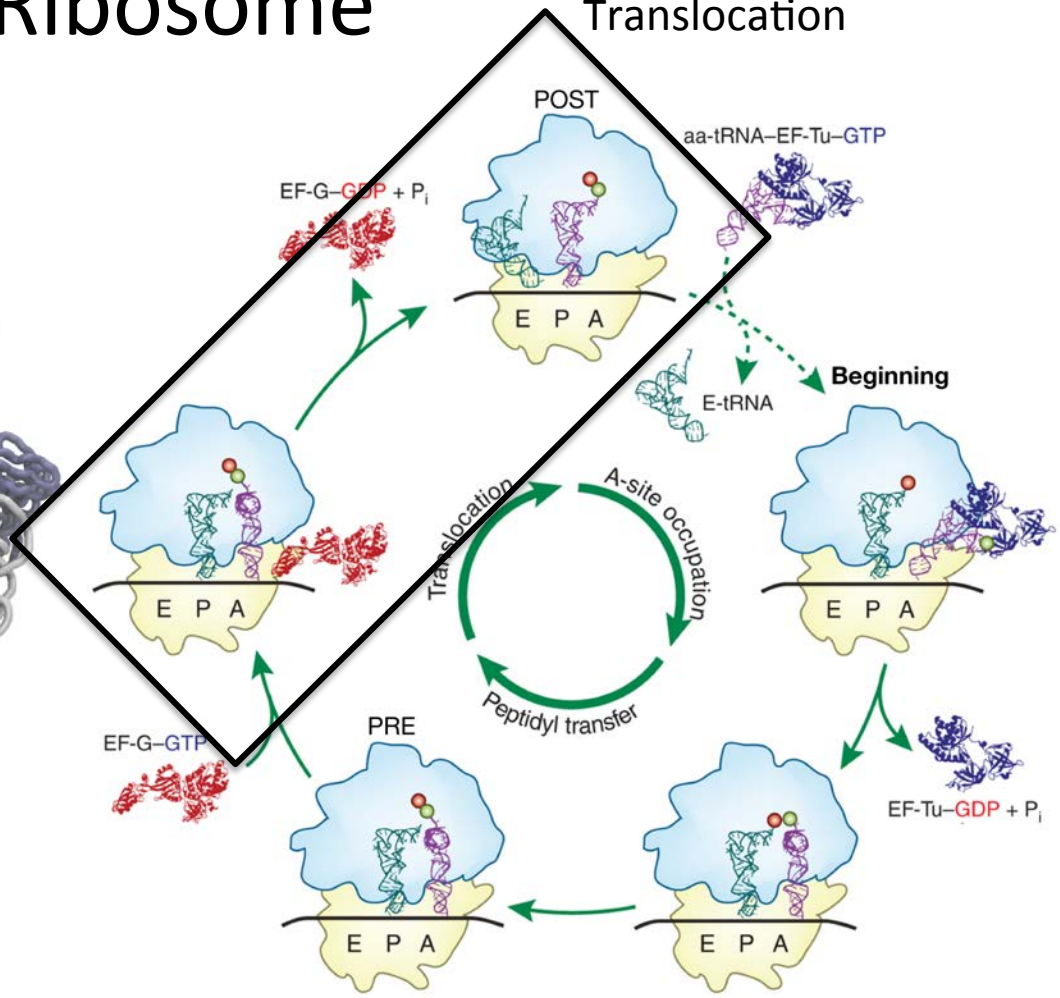
-- By limiting accessible phase space, forward barrier is lowered

Jeff Noel and PCW *Nature Comm.* (2016)

70S Ribosome



Hybrid-state formation,
Translocation

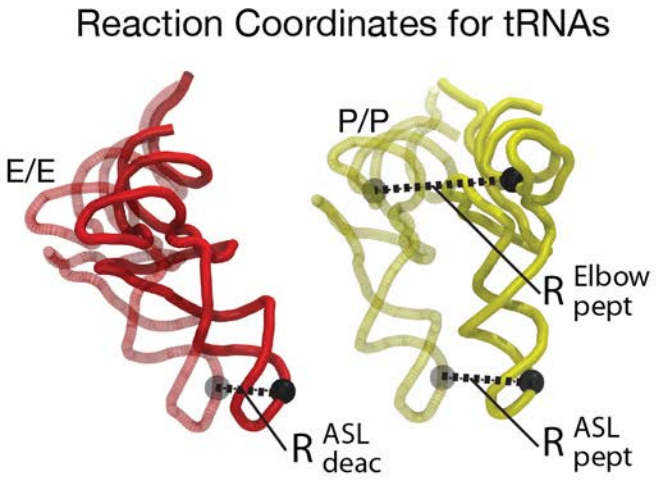


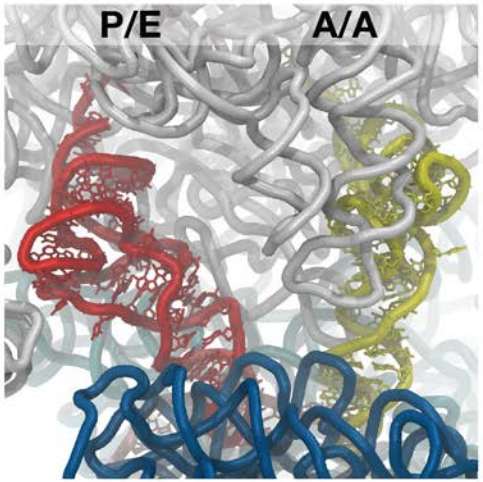
Adapted from Achenbach & Nierhaus, *Nat Struct Mol Biol* **20**, 1019–1022 (2013)

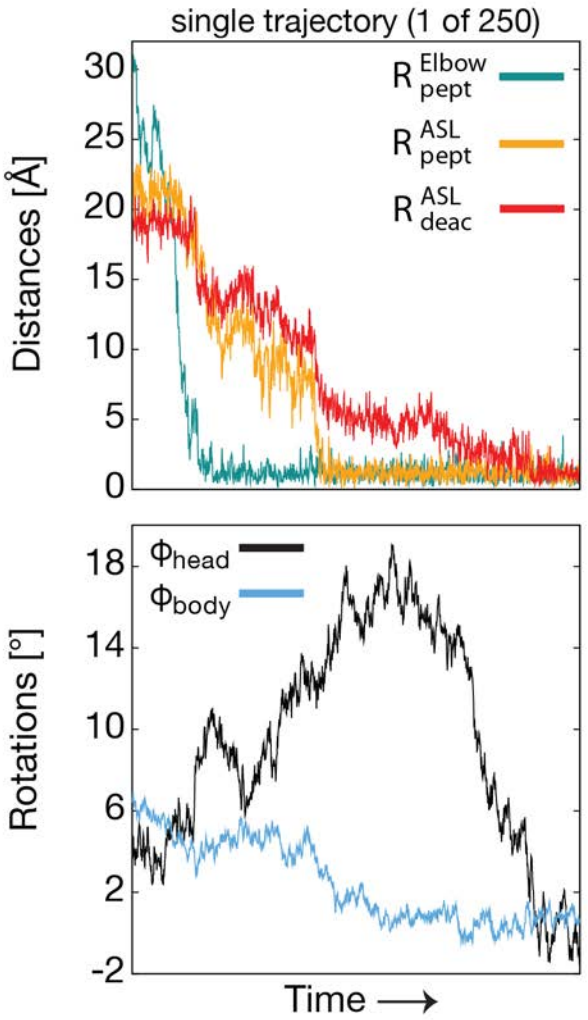
Over 50 proteins
3 RNAs w/ ~4000 nt
~150,000 atoms

PDB ID: 4V6F
Jenner et al. (2010)

Stochastic translocation events



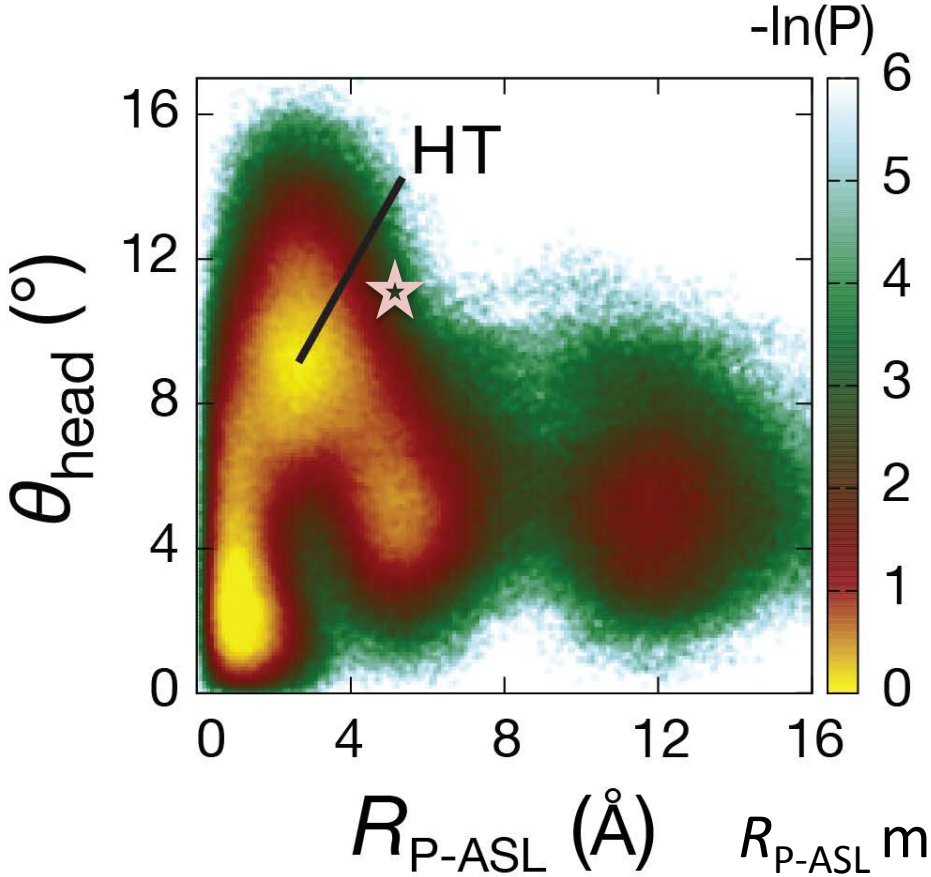


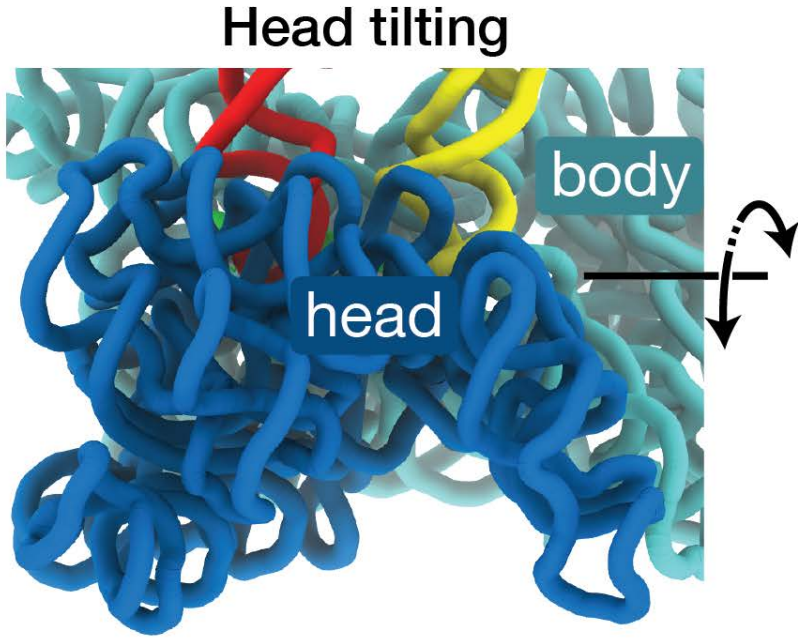


Initiate each simulation in a partially rotated A/A-P/E configuration
Spontaneously reaches the post configuration

Kien Nguyen and PCW *Nature Comm.* (2016)

Statistical Signature of *Tilting* Intermediate





$R_{P\text{-}ASL}$ monitors tRNA movement between binding sites

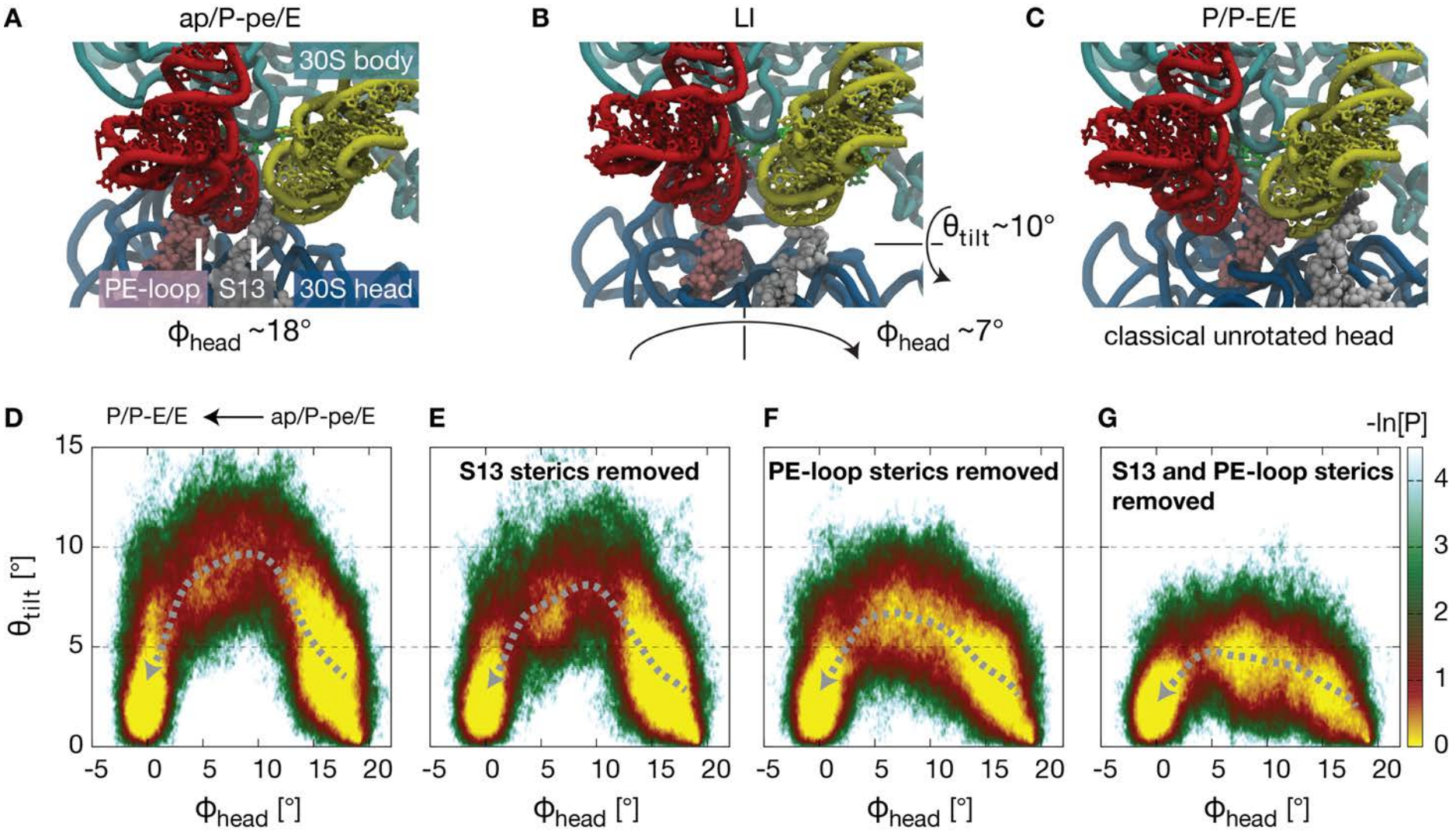
Cryo-em models (star)
Pink: Ramrath et al *Nature* 2013 (tmRNA)

250 simulated events

intermediates arise from steric effects

Kien Nguyen and PCW *Nature Comm.* (2016)

Dissecting Steric Factors of Tilting



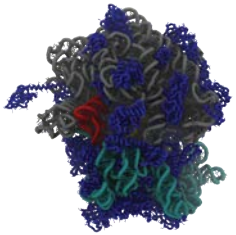
Kien Nguyen and PCW *Nature Comm.* (2016)

Simulation of mRNA-tRNA translocation

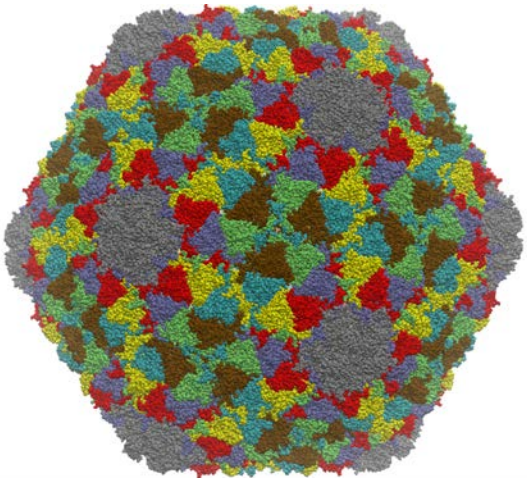
1 of 250 simulated events

Kien Nguyen
Paul Whitford
Northeastern University
Department of Physics

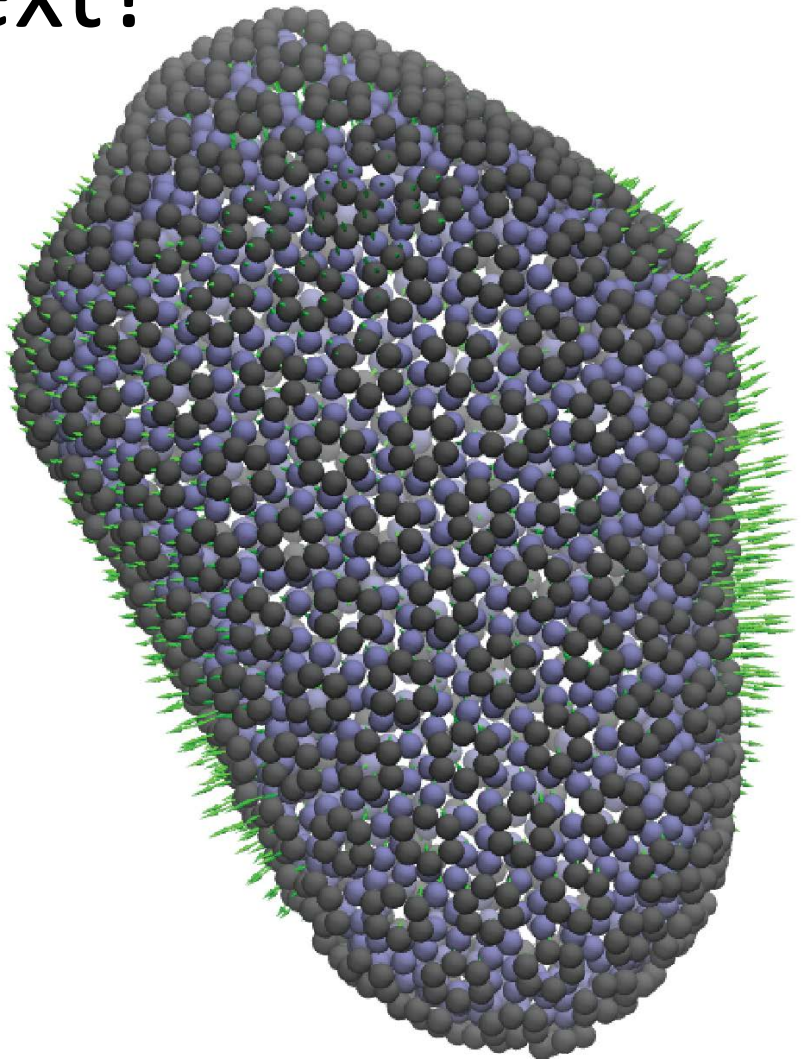
What's next?



Ribosomes
200-300k atoms
2M atoms with water
~100-1000 cores



Bacteriophages
~2 million atoms
~20M atoms with water
~500-10,000 cores



HIV-1 Capsid
~4 million atoms
~50M atoms with water
~2,000-100,000 cores

Endless list of biological assemblies exist

- New models and HPC resources have allowed us to move away from performing single simulations

- Can now dissect the physical factors that govern ***statistical*** properties

- In 2017, approaches for the ribosome are comparable to efforts to study small proteins around 2000

- Continued HPC development will allow systematic study of more complex, and larger assemblies

Thanks

- Reaction coordinates
  - **Jeff Noel (Max Delbrück Center, Berlin)**
  - Huan (Joyce) Yang (NEU)
  - Vitor Leite (Univ. Estadual São Paulo)
  - Jorge Chahine (Univ. Estadual São Paulo)
  - Mariana Levi (Northeastern)
  - Liah Dukaye (Northeastern)
- tRNA-mRNA translocation and hybrid motion
  - **Kien Nguyen (Northeastern)**
- Explicit-solvent simulations
  - Karissa Sanbonmatsu (LANL)

- SMOG team
  - Jeff Noel (Max Delbrück Center, Berlin)
  - Mariana Levi (Northeastern)
  - José Onuchic (Rice U)

Computing

- NEU Discovery Cluster
- C3DDB cluster
- TeraGrid/XSEDE
- Univ São Paulo/Rice Univ. Blue Gene/Q

NSF CAREER MCB/POLS

Downloadable SMOG software

http://smog-server.org

Noel, Levi, Raghunathan, Lammert, Hayes, Onuchic, Whitford (2016) PLoS Comput Biol 12(3): e1004794