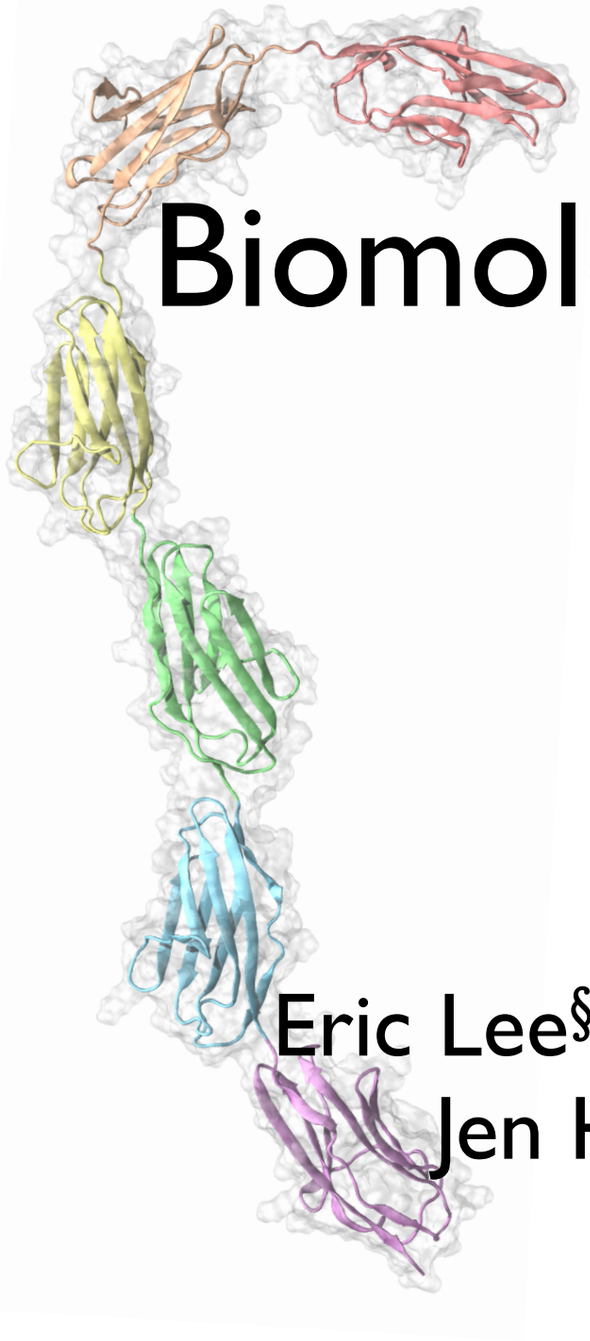
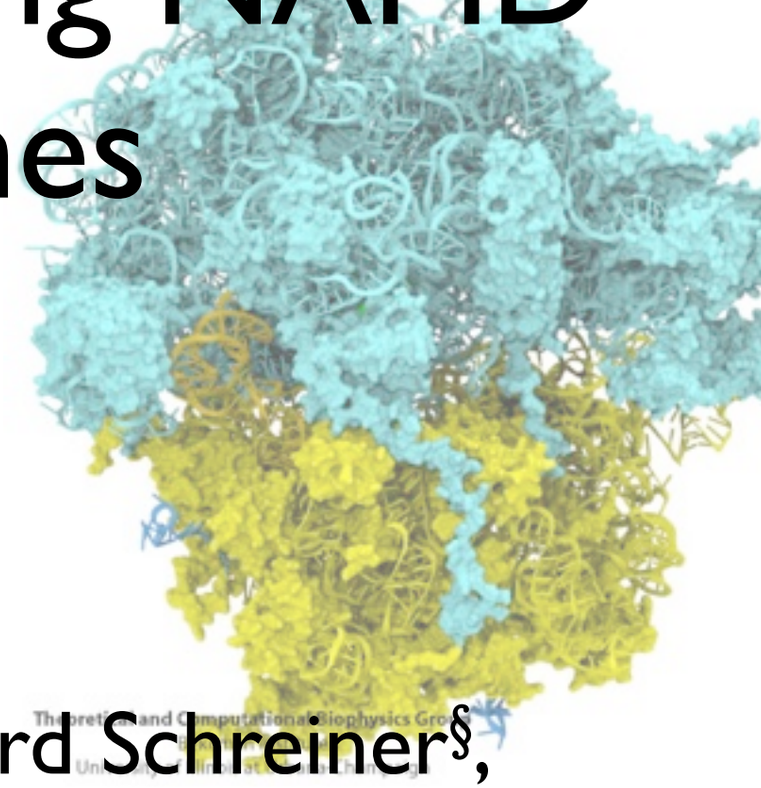


Biomolecular modeling using NAMD on TeraGrid machines



Abhinav Bhatele[†]

Eric Lee[§], Ly Le[§], Leonardo Trabuco[§], Eduard Schreiner[§], Jen Hsin[§], James C. Phillips[§], Laxmikant V. Kale[†], Klaus Schulten[§]

[†]Parallel Programming Laboratory (PPL) and [§]Theoretical and Computational Biophysics Group (TCBG)
University of Illinois at Urbana-Champaign

Scalable Molecular Dynamics

- NAMD: Parallel molecular dynamics software
- NIH funded and developed by the joint collaboration of Theoretical and Computational Biophysics Group and the Parallel Computing Laboratory at Illinois

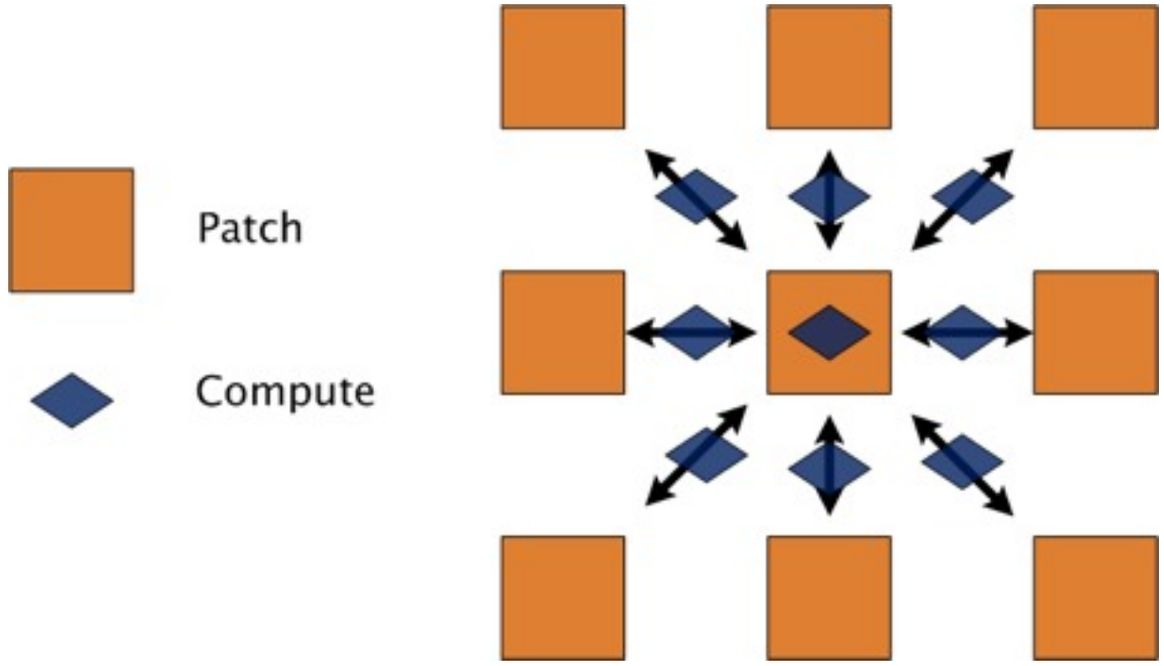
- Publicly available and installed at most supercomputing centers in US and elsewhere

Some history

- First publication on NAMD in 1996, NAMD 2.1 in 1999
  - Awarded the Gordon Bell Prize in 2002
- NAMD 2.6 was released in August 2006
  - In 2007, NAMD accounted for 20% and 15% of the compute cycles used at PSC and NCSA
- NAMD 2.7b3 was released in July 2010
  - CUDA version of NAMD for running on GPGPUs
- It has been downloaded by over 36,000 registered users
- 1776 citations as of March 2010

NAMD's Hybrid Method

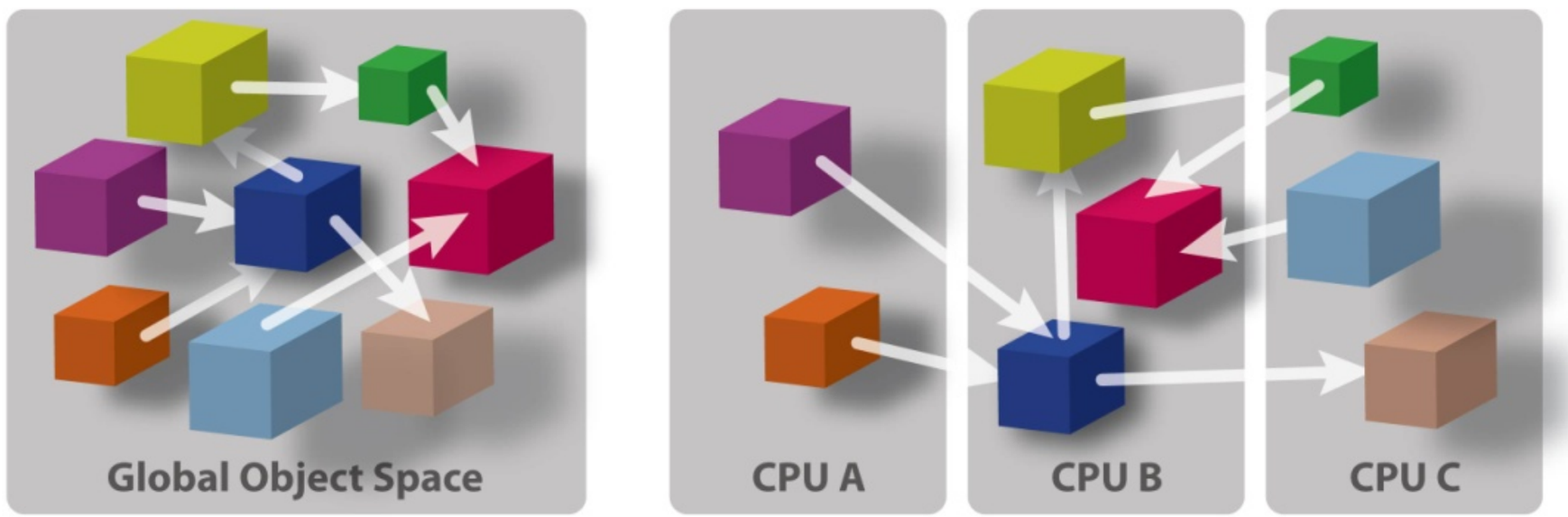
- Hybrid of spatial and force decomposition‡
- Similar methods (neutral territory, midpoint) later used in Blue Matter and Desmond



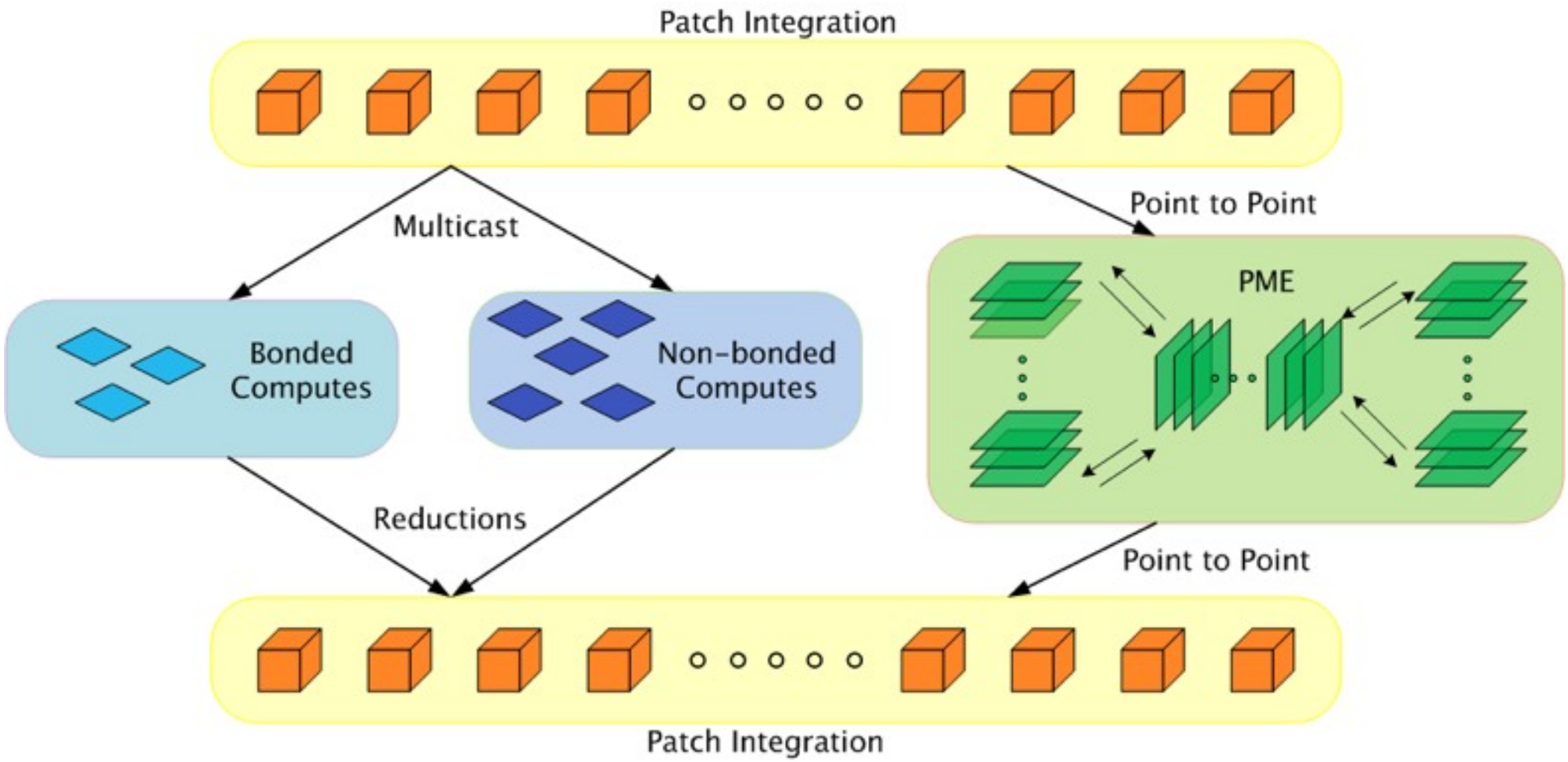
‡ L.V. Kalé , R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten. NAMD2: Greater scalability for parallel molecular dynamics. Journal of Computational Physics, 1998.

Charm++ Programming Model

- Object-based message-driven asynchronous model
- Overdecomposition: number of objects >> number of processors
- Benefits from adaptive overlap of computation and communication and load balancing

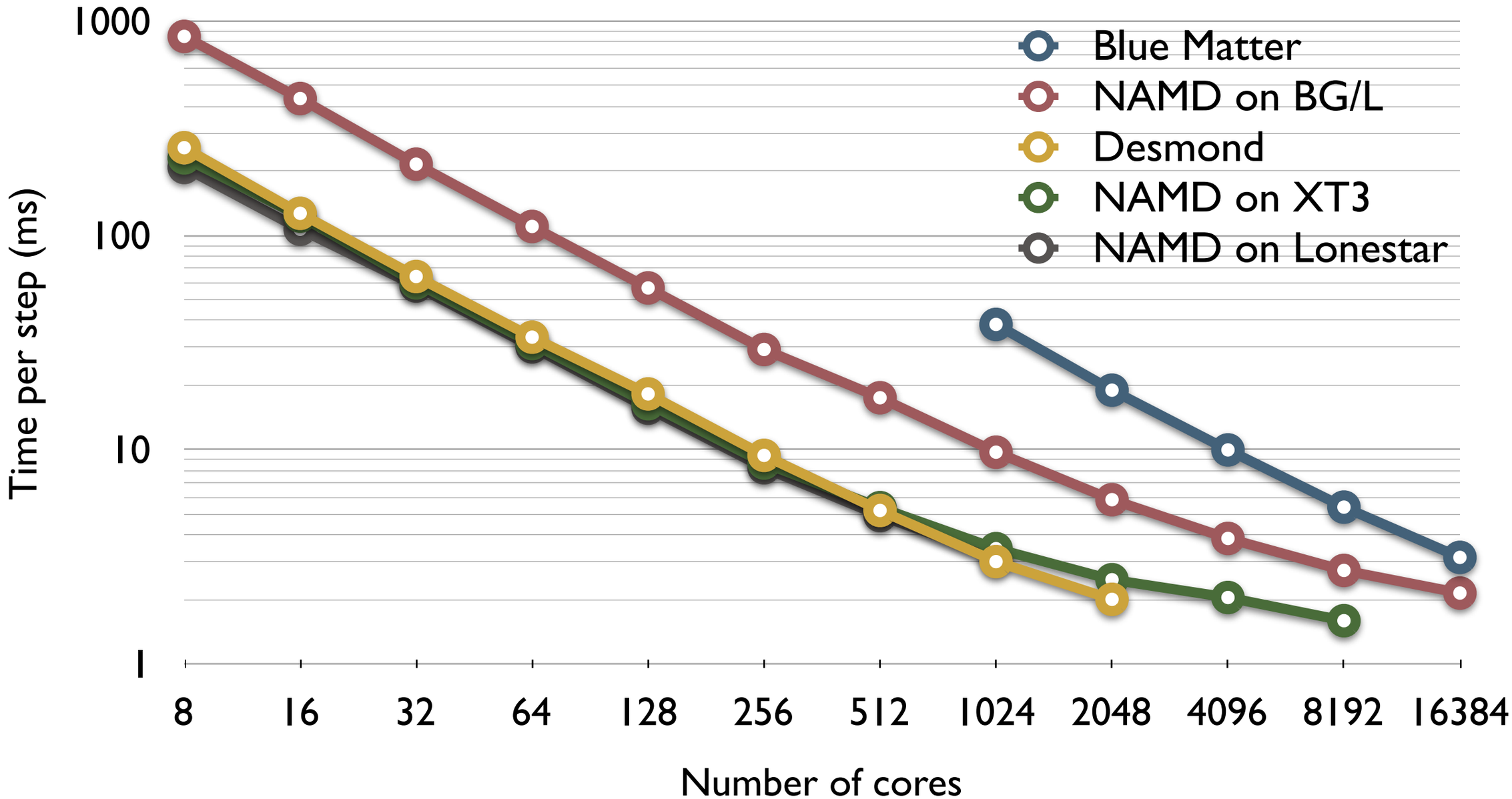


Parallelization using Charm++

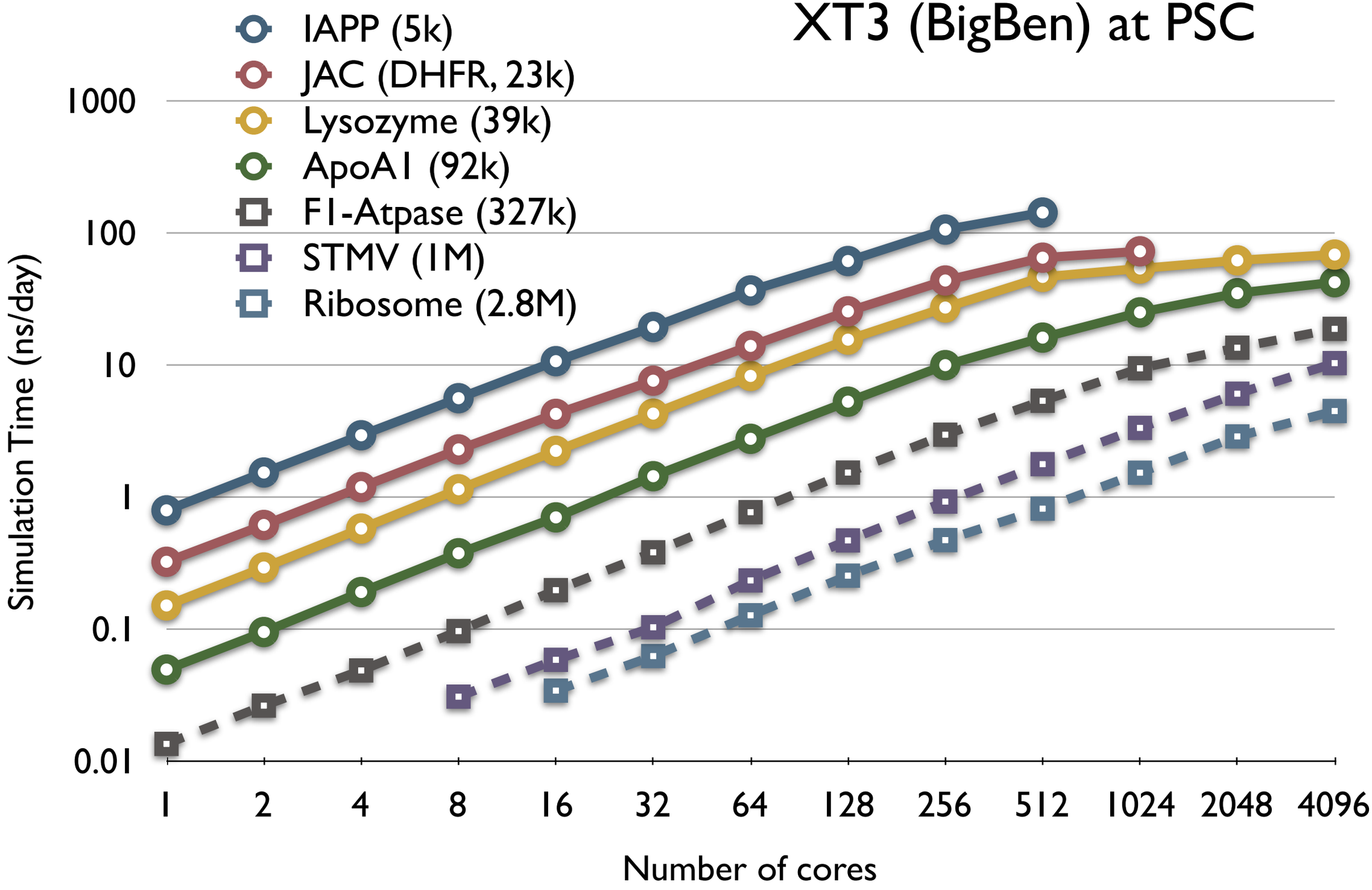


Bhatele, A., Kumar, S., Mei, C., Phillips, J. C., Zheng, G. & Kale, L.V., Overcoming Scaling Challenges in Biomolecular Simulations across Multiple Platforms. In Proceedings of IEEE International Parallel and Distributed Processing Symposium, Miami, FL, USA, April 2008.

Comparison with other MD codes

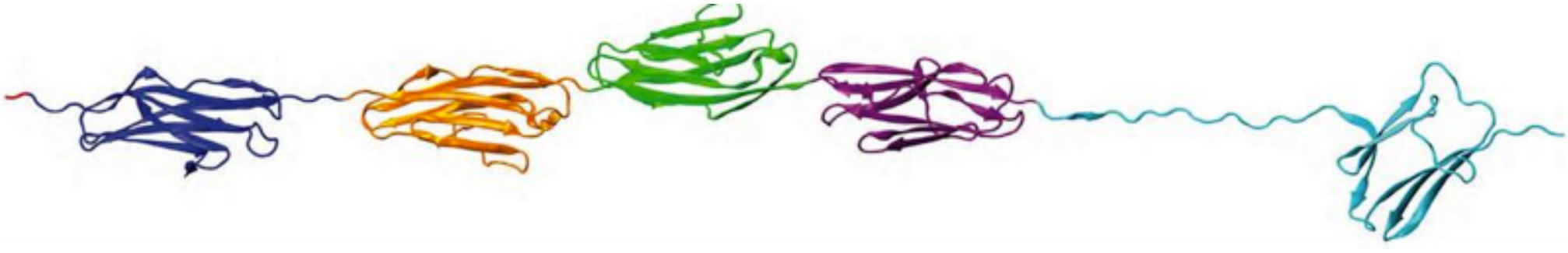


NAMD Running on Cray XT3 (BigBen) at PSC

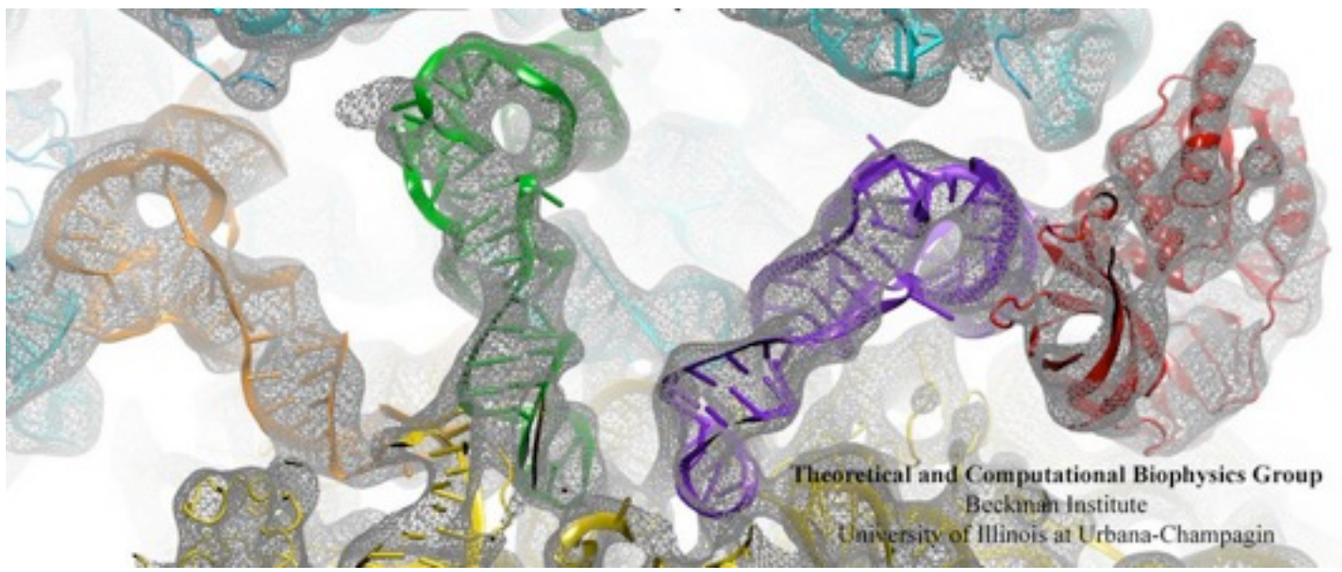


Two simulation studies

- Titin: the protein responsible for muscle elasticity



- Dynamics of ribosome: the protein factory

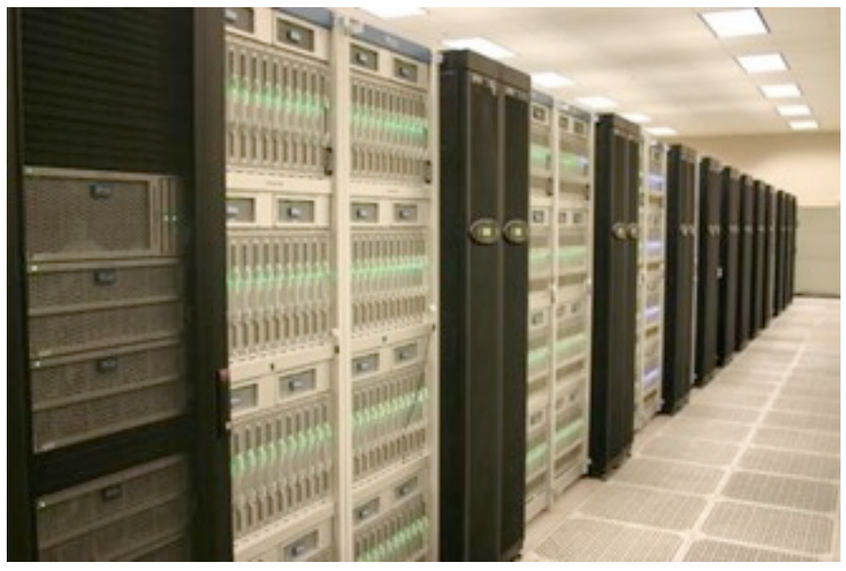


TeraGrid machines

- NCSA: Abe
  - 1200 8-core nodes
  - 2.33 GHz Clovertowns

http://www.ncsa.illinois.edu/UserInfo/Resources/Hardware/Intel64Cluster/

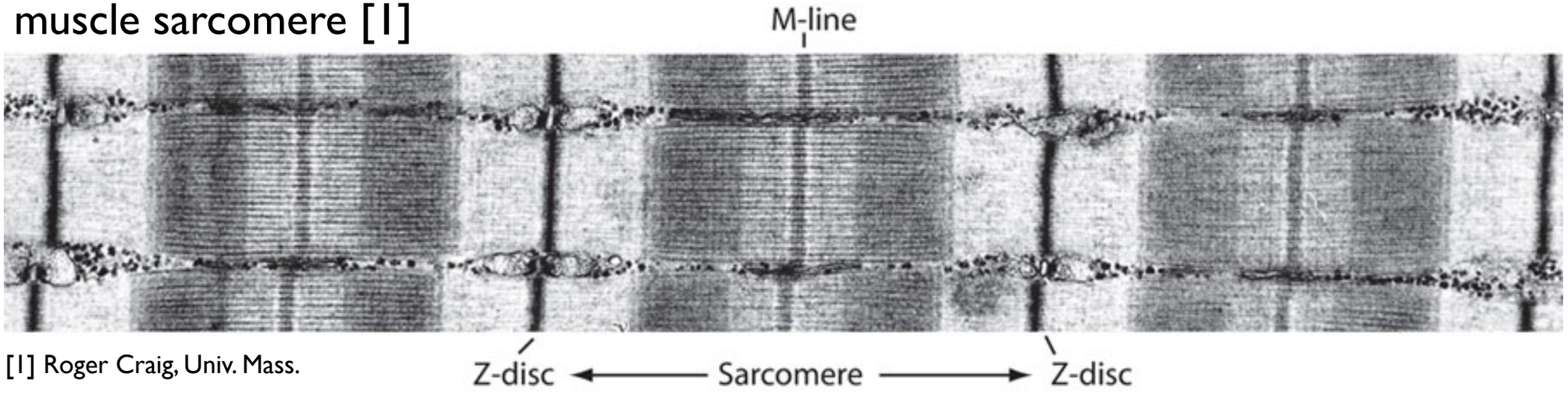
- TACC: Ranger
  - 3,936 16-core nodes
  - 2.3 GHz Opterons
- Infiniband fat-tree network



http://services.tacc.utexas.edu/index.php/ranger-user-guide

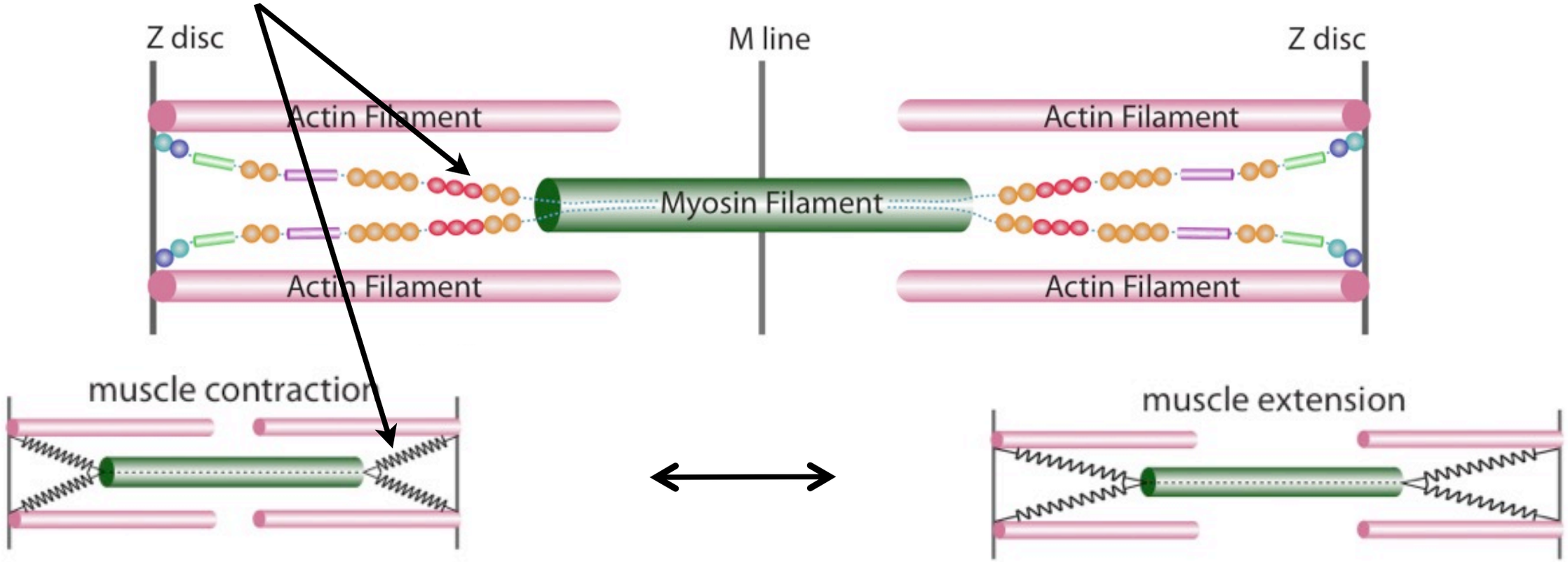
The Molecular Origin of Muscle Elasticity

muscle sarcomere [1]



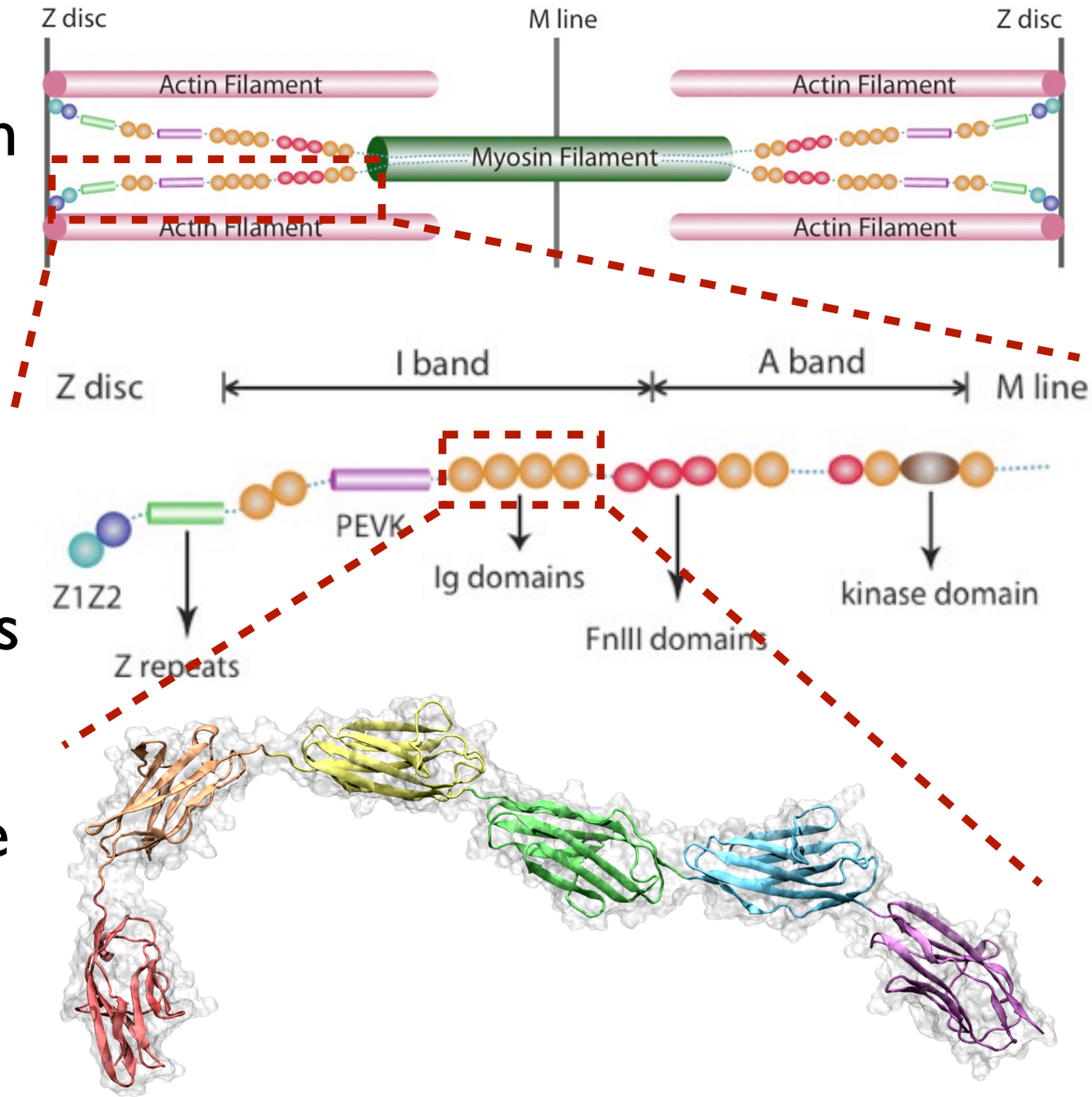
[1] Roger Craig, Univ. Mass.

Titin: the elastic component of sarcomere



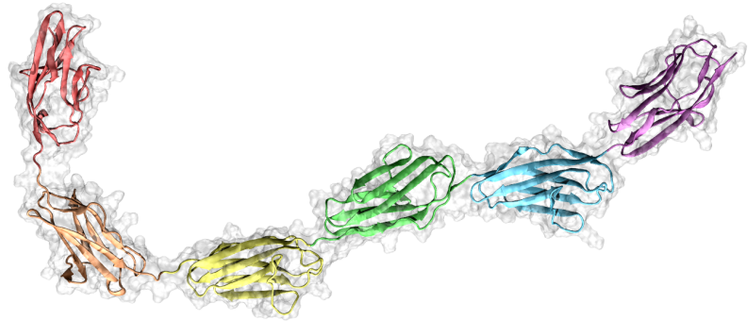
Titin: the molecular rubber band in muscle

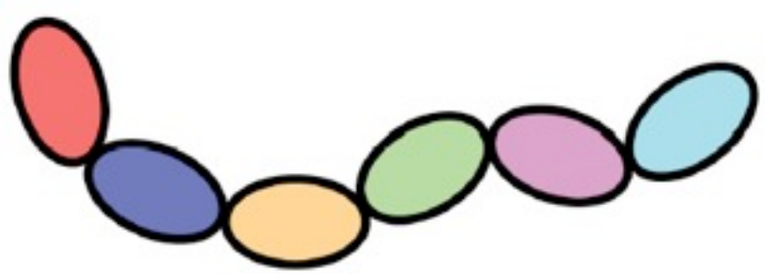
- Titin is a lengthy protein made of connected domains
- Titin is the largest protein known
- The Ig-chain contributes to the elasticity in titin
- Globular Ig domains are linked into a flexible chain

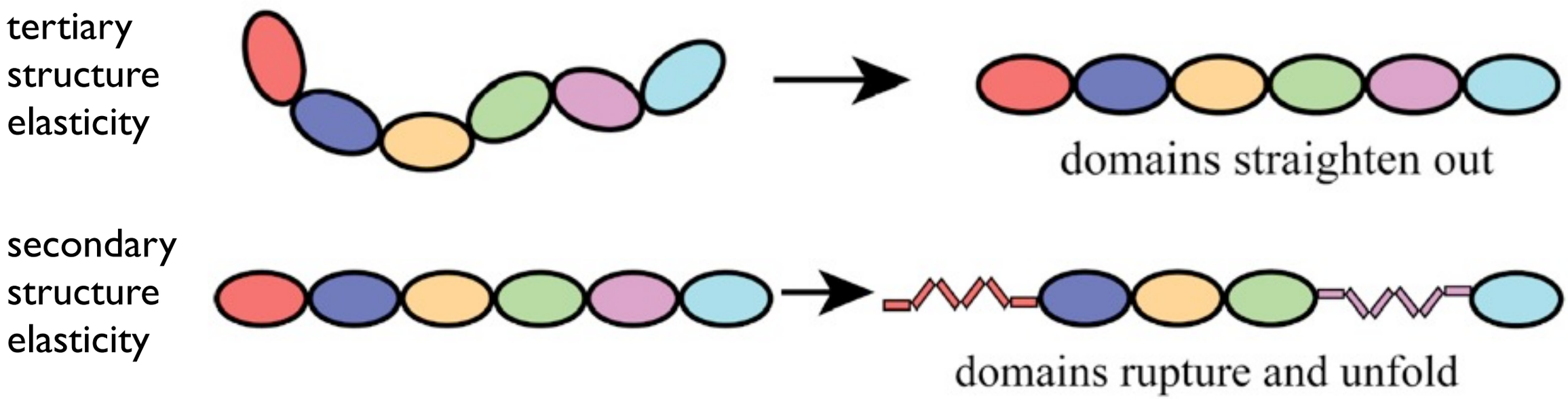


Two modes of force response in titin Ig chain

- At low stretching forces (few tens of pN), the Ig chain simply strengthens out - tertiary structure elasticity
- At physiologically extreme forces (> 100 pN), individual Ig domain can lose its secondary structure (unraveling) and prove further extension - secondary structure elasticity







Stretching Titin Ig chain softly with Molecular Dynamics

- Crystal structure of a titin Ig chain with six connected Ig domain was published in 2008 [1]
- Steered molecular dynamics, implemented in NAMD, was employed to stretch apart the crescent shaped Ig chain (without disrupting the structure of individual Ig domains) [2]
- From simulation, the relationship between force applied on the chain, and extension of the chain, is obtained (force-extension profile)

Extension of the protein (Å)

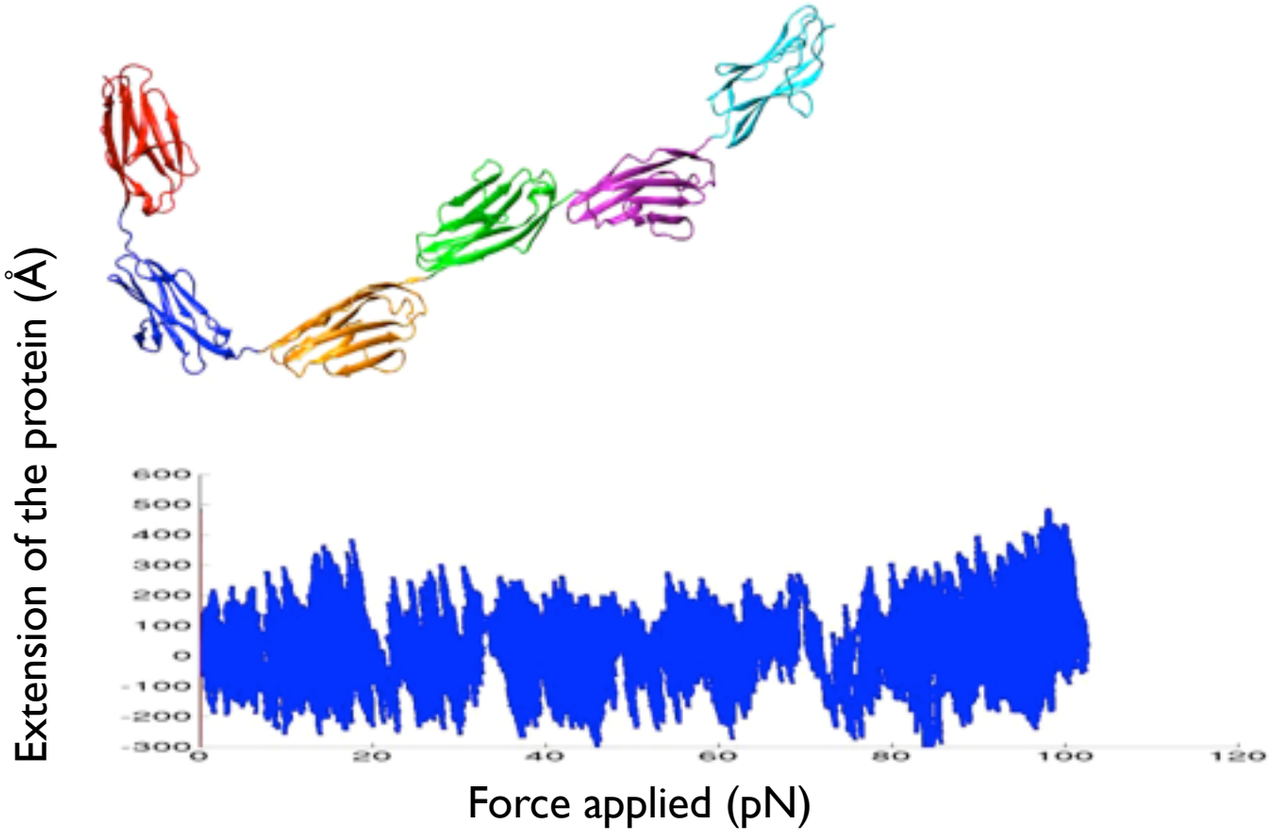
Force applied (pN)

- ~277,000 atoms, 10 ns
- Explicit water
- 512 cores, ~4 ns/day (on TACC's Ranger)

[1] Castelmur et al., PNAS, 2008.
[2] Lee, Hsin, von Castelmur, Mayans, Schulten. Biophysical Journal, 2010.

Stretching Titin Ig chain softly with Molecular Dynamics

- Crystal structure of a titin Ig chain with six connected Ig domain was published in 2008 [1]
- Steered molecular dynamics, implemented in NAMD, was employed to stretch apart the crescent shaped Ig chain (without disrupting the structure of individual Ig domains) [2]
- From simulation, the relationship between force applied on the chain, and extension of the chain, is obtained (force-extension profile)



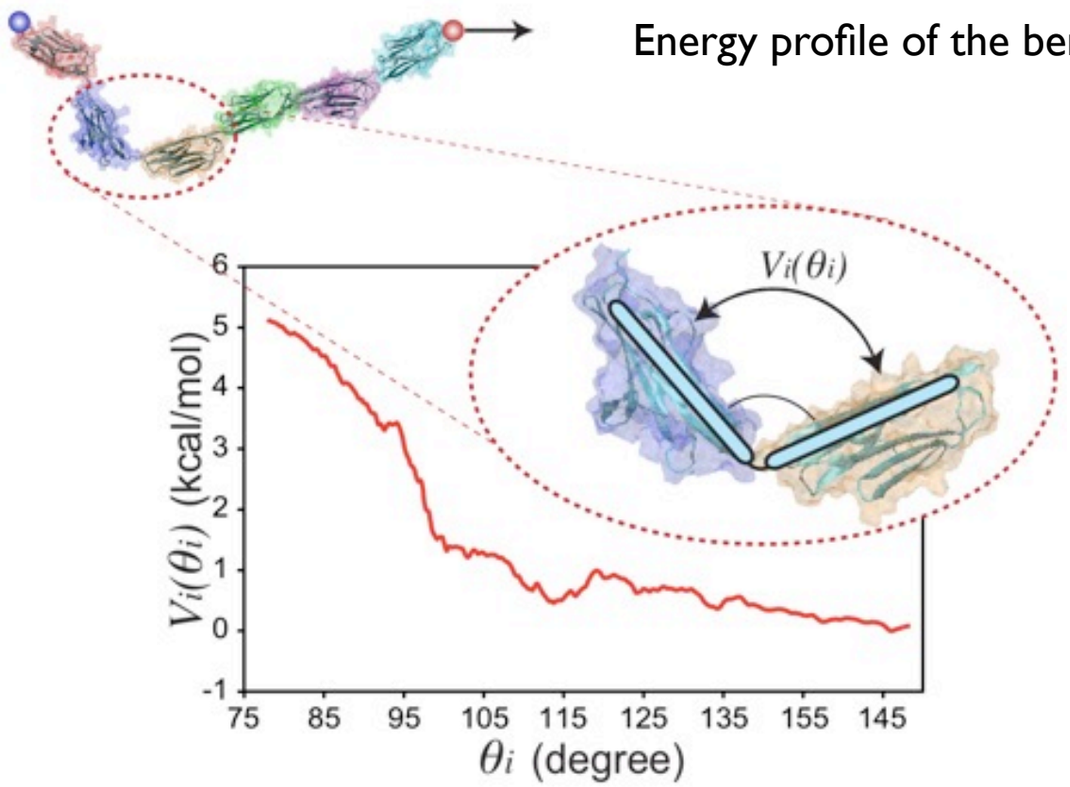
- ~277,000 atoms, 10 ns
- Explicit water
- 512 cores, ~4 ns/day (on TACC's Ranger)

[1] Castelmur et al., PNAS, 2008.
[2] Lee, Hsin, von Castelmur, Mayans, Schulten. Biophysical Journal, 2010.

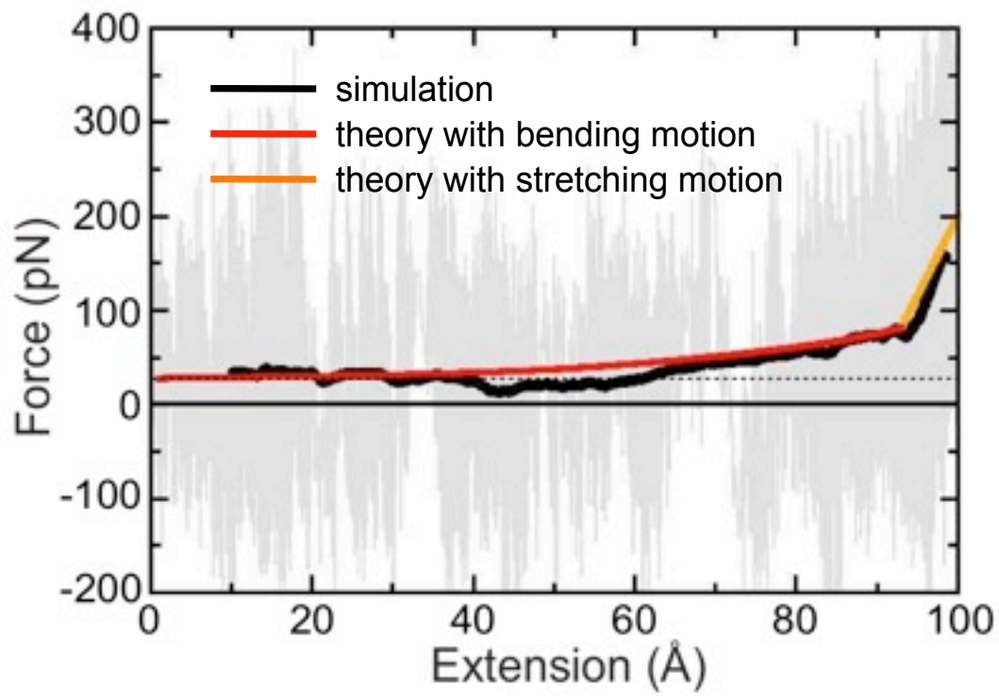
Understanding the tertiary elasticity of the Ig chain

- Characteristics of the force-extension profile can be understood completely by considering the domain-domain motion in the Ig chain
- Energetics of such domain-domain motion can be measured using adaptive biasing force method implemented in NAMD [3]
- Combining with statistical mechanical theory, the force-extension profile is described accurately [2]

[3] Henin and Chipot, J Chem Phys, 2004



Energy profile of the bending motion between an Ig domain pair



Secondary structure of Titin Ig chain

- Continued stretching on a linear Ig chain leads to unraveling of individual Ig domains
- This is the regime of secondary structure elasticity
- The Ig domains ruptures one-by-one, not concurrently, leading to force-extension profile with distinct peaks (known as the "sawtooth pattern") [2]

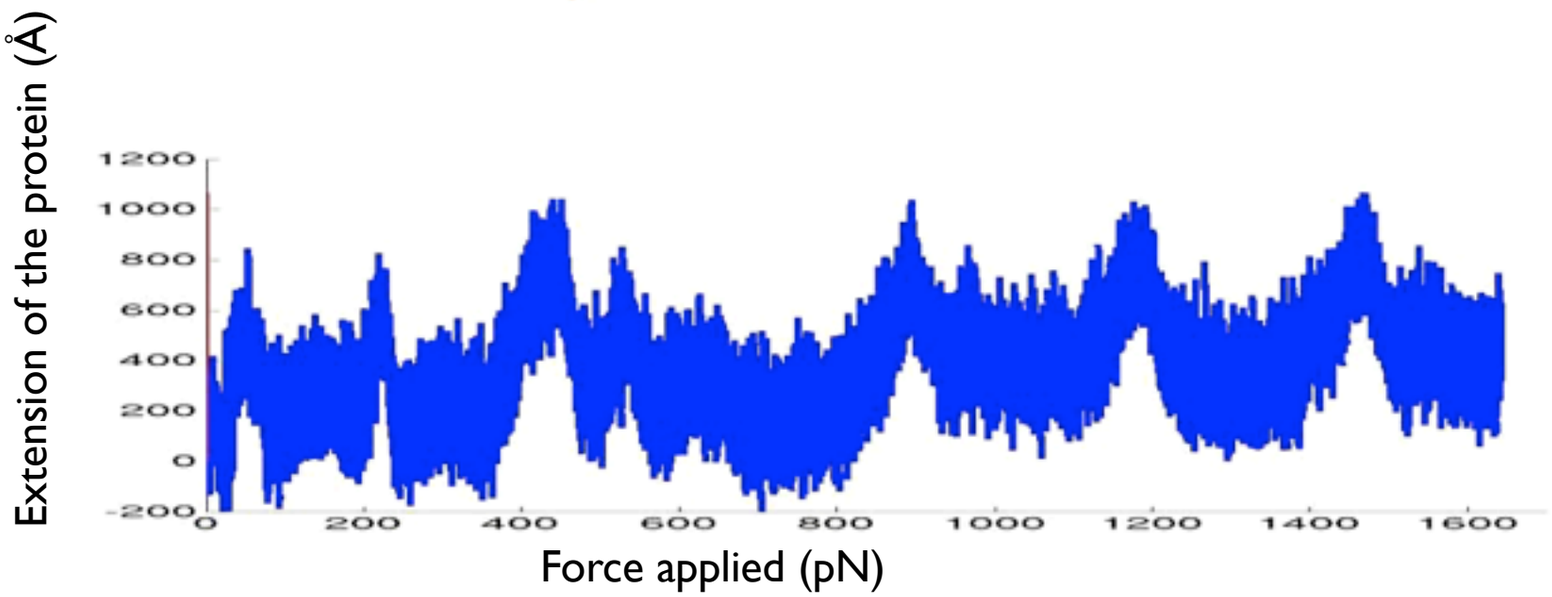
Extension of the protein (Å)

Force applied (pN)

- ~635,000 atoms, 66 ns
- Explicit water
- 1200 cores, ~2 ns/day (on TACC's Ranger)

Secondary structure of Titin Ig chain

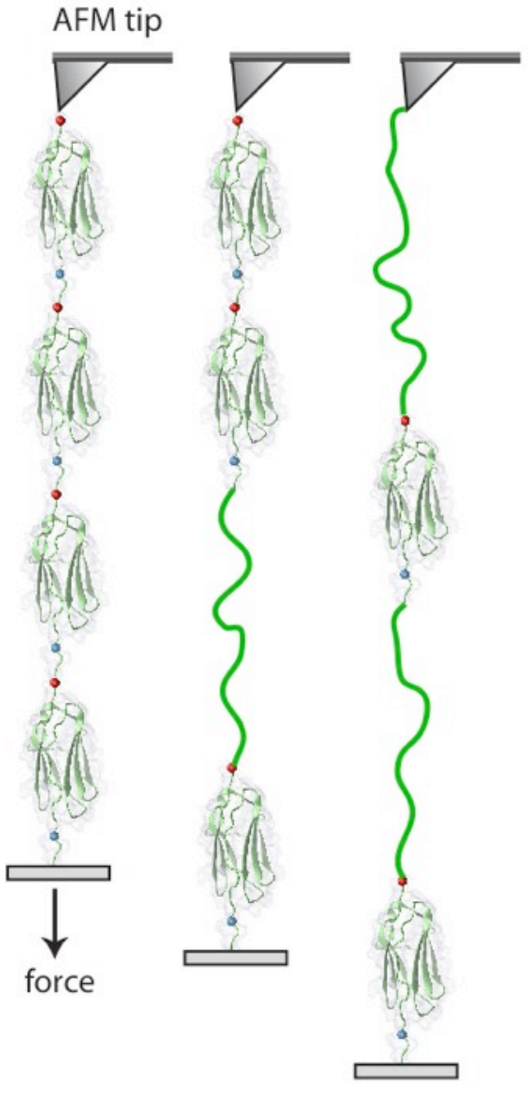
- Continued stretching on a linear Ig chain leads to unraveling of individual Ig domains
- This is the regime of secondary structure elasticity
- The Ig domains ruptures one-by-one, not concurrently, leading to force-extension profile with distinct peaks (known as the "sawtooth pattern") [2]

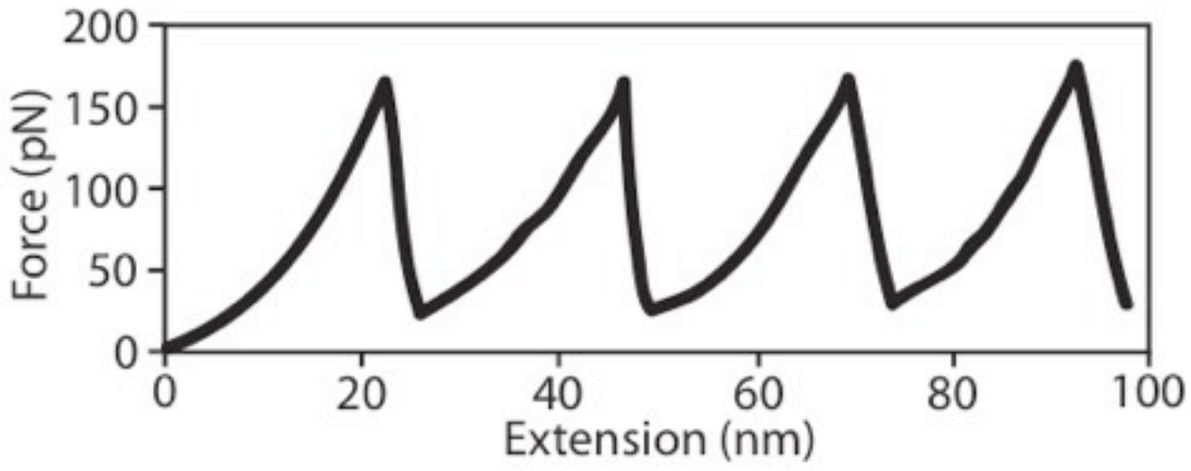


- ~635,000 atoms, 66 ns
- Explicit water
- 1200 cores, ~2 ns/day (on TACC's Ranger)

Secondary structure elasticity: Simulation vs. Experiment

- The "sawtooth" pattern in the force-extension profile of a multi-Ig construct is also seen in comparable experiment (e.g., using atomic force microscopy, AFM)
- But the force peaks seen in experiment is lower than in simulation



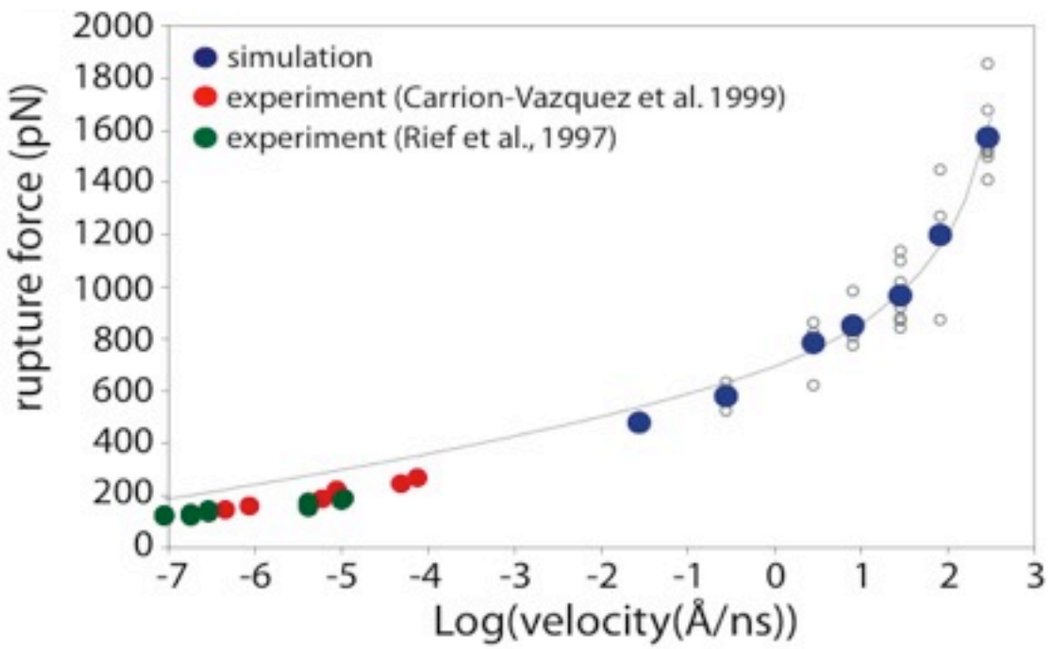


Typical force-extension profile seen in stretching experiment

The difference in force peak is due to the faster pulling velocity used in simulation, the faster velocity employed so that unraveling of the Ig domains can be completed within an reasonable amount of computer resources

Merging Simulation and Experiment Time scales

- A series of stretching simulations was performed for a single titin Ig domain to systematically test the relationship between pulling velocity and rupture peak force [4]
- The longest simulation being 1 μs long
- Results from the simulation and experiment can be described by the same theoretical model

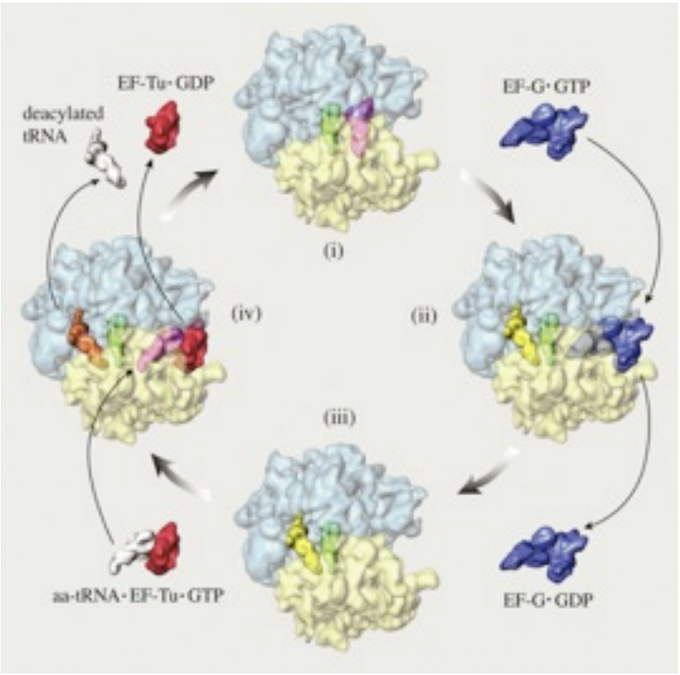


- ~52,000 atoms, 37 simulations totaling 2.4 μs
- Explicit water
- 546 cores, ~20 ns/day (on NCSA's Abe)

[4] Lee, Hsin, Sotomayor, Comellas, Schulten. Structure, 2009

The ribosome

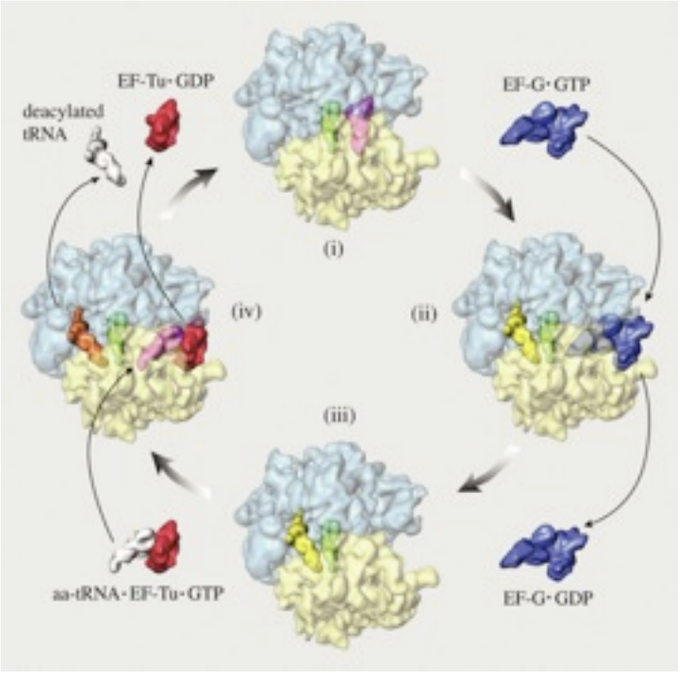
- large (~300,000 atoms) and dynamic molecular machine
- translates genetic code into proteins
- consists of a large and a small subunit
- interacts with many factors during its function
- can be controlled by the nascent chain
- can feed nascent chain into translocating pores
- many states are impossible to crystallize, but can be imaged by cryo-electron microscopy

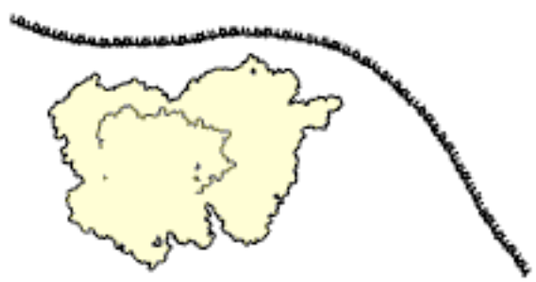


http://en.wikipedia.org/wiki/Ribosome

The ribosome

- large (~300,000 atoms) and dynamic molecular machine
- translates genetic code into proteins
- consists of a large and a small subunit
- interacts with many factors during its function
- can be controlled by the nascent chain
- can feed nascent chain into translocating pores
- many states are impossible to crystallize, but can be imaged by cryo-electron microscopy

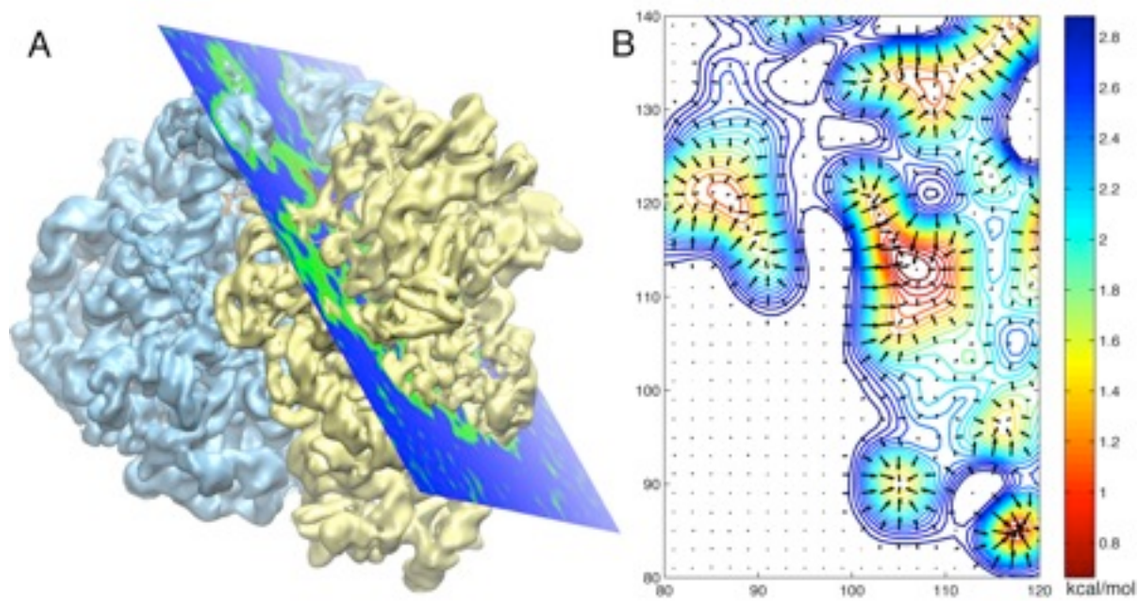




http://en.wikipedia.org/wiki/Ribosome

Molecular dynamics flexible fitting (MDFF)

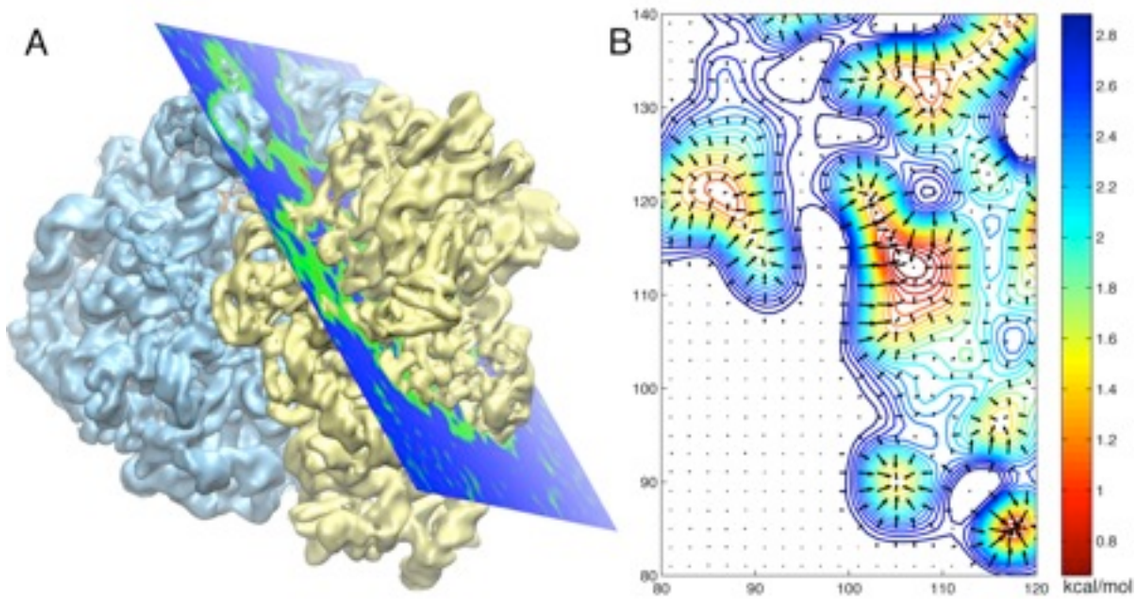
- Idea: add forces derived from the experimental cryo-EM map to the forces from the MD potential
- Cryo-EM map drives the structure into a specific conformation



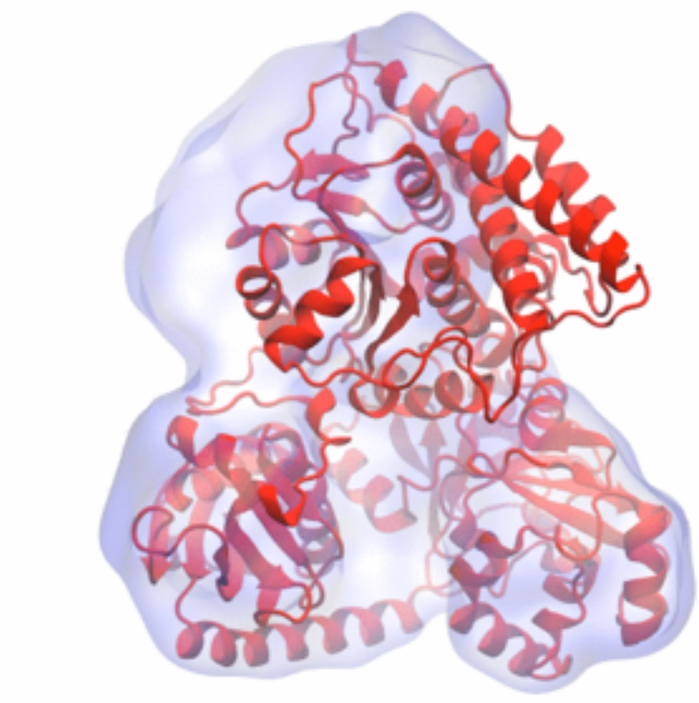
[1] Trabuco et al. Structure (2008) 16:673-683.
[2] Villa et al. PNAS (2009) 106:1063-1068.
[3] Sener et al. Chem Phys (2009) 357:188-197.
[4] Trabuco et al. Methods (2009) 49:174-180.
[5] Hsin et al. Biophys J (2009) 97:321-329.
[6] Gumbart et al. Structure (2009) 17:1453-1464.
[7] Seidelt et al. Science (2009) 326:1412-1415.
[8] Becker et al. Science (2009) 326:1369-1373.
[9] Trabuco et al. Structure (2010) 18:627-637.

Molecular dynamics flexible fitting (MDFF)

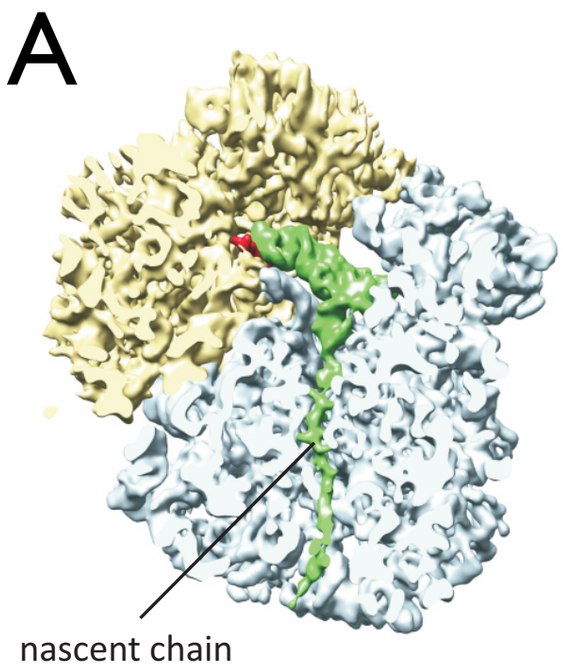
- Idea: add forces derived from the experimental cryo-EM map to the forces from the MD potential
- Cryo-EM map drives the structure into a specific conformation

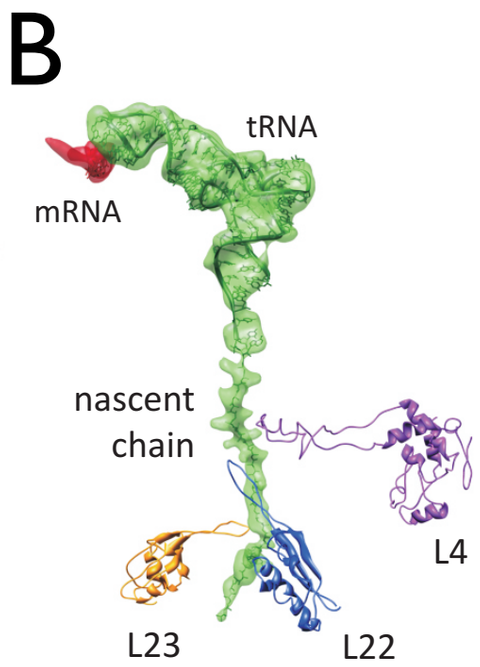


[1] Trabuco et al. Structure (2008) 16:673-683.
[2] Villa et al. PNAS (2009) 106:1063-1068.
[3] Sener et al. Chem Phys (2009) 357:188-197.
[4] Trabuco et al. Methods (2009) 49:174-180.
[5] Hsin et al. Biophys J (2009) 97:321-329.
[6] Gumbart et al. Structure (2009) 17:1453-1464.
[7] Seidelt et al. Science (2009) 326:1412-1415.
[8] Becker et al. Science (2009) 326:1369-1373.
[9] Trabuco et al. Structure (2010) 18:627-637.

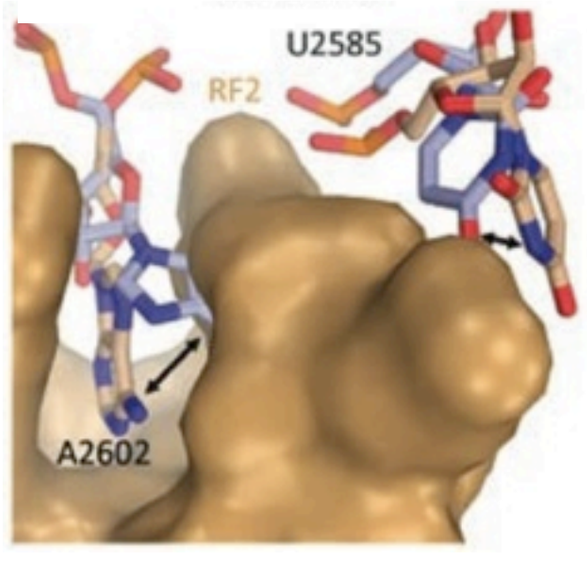


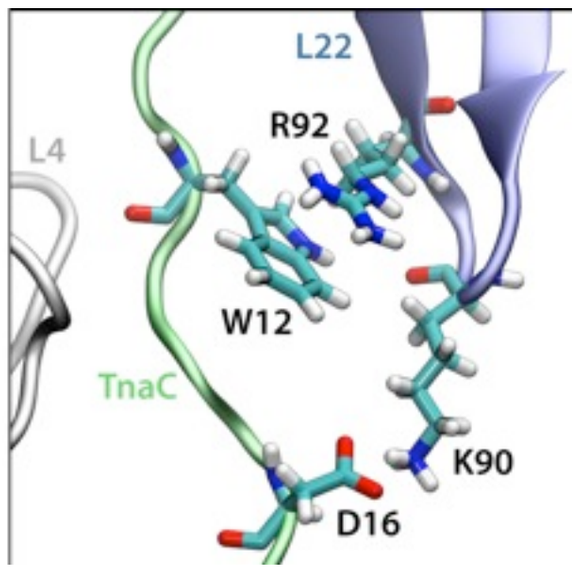
Regulatory chain TnaC





- regulatory nascent chain
- stops ribosome, but prevents termination
- system size: 300,000 atoms
- in vacuo, MDFF, 1024 cores, 48 h, 20 ns
- path of the backbone of TnaC through the ribosome identified
- active site residues A2602 and U2585 adopt conformations which prevent termination
- specific interactions were identified which promote stalling





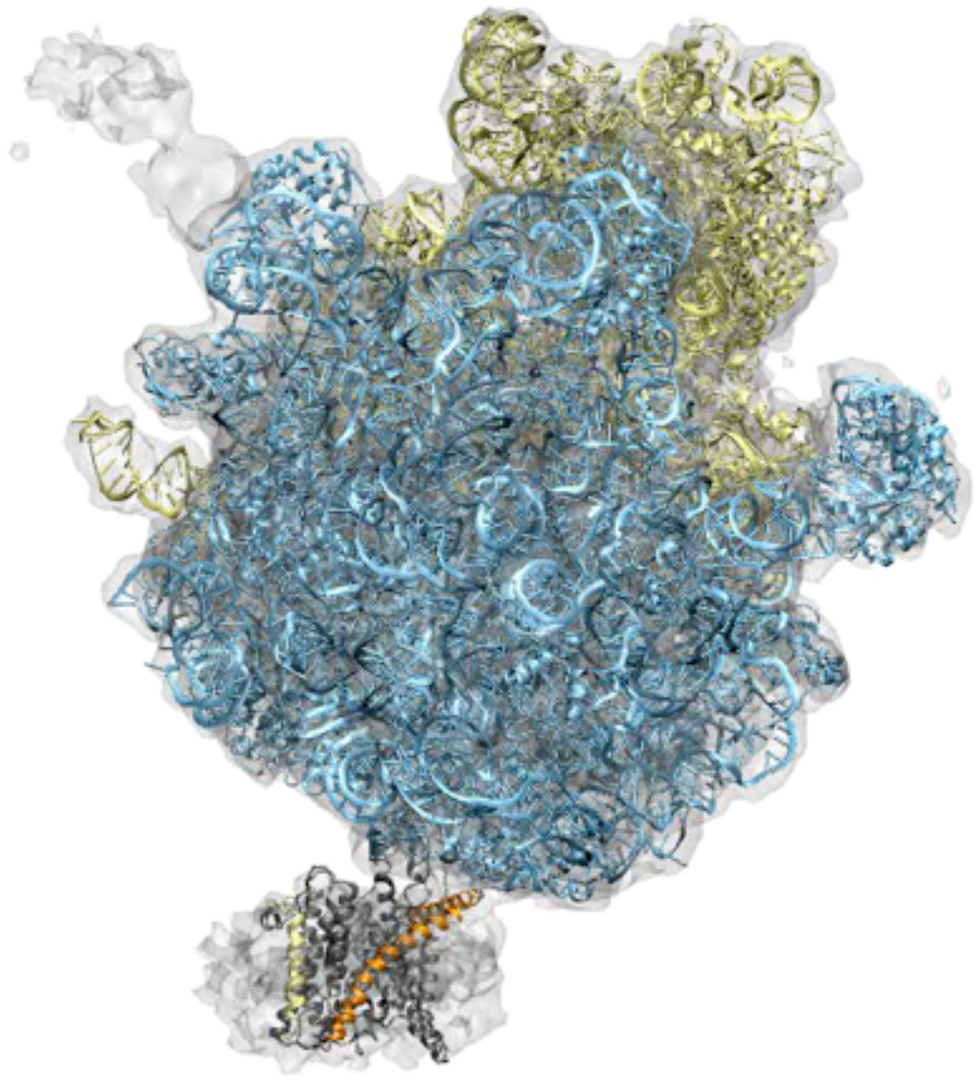
Seidelt et al. Science (2009) 326:1412-1415.
Trabuco et al. Structure (2010) 18:627-637.

Interactions between the ribosome and the protein conducting channel

- protein conducting channel in complex with ribosome
- system size: 2,700,000 atoms
- In water, MDFF
- 1024 cores, 2,000 h, 20 ns

Gumbart et al. Structure (2009) 17:1453-1464.

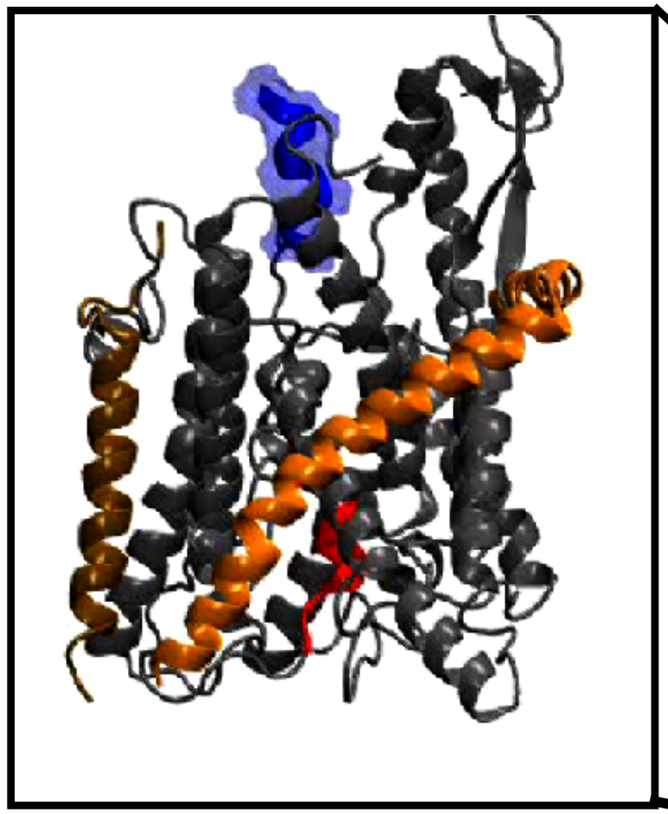
Interactions between the ribosome and the protein conducting channel



- protein conducting channel in complex with ribosome
- system size: 2,700,000 atoms
- In water, MDFF
- 1024 cores, 2,000 h, 20 ns

Gumbart et al. Structure (2009) 17:1453-1464.

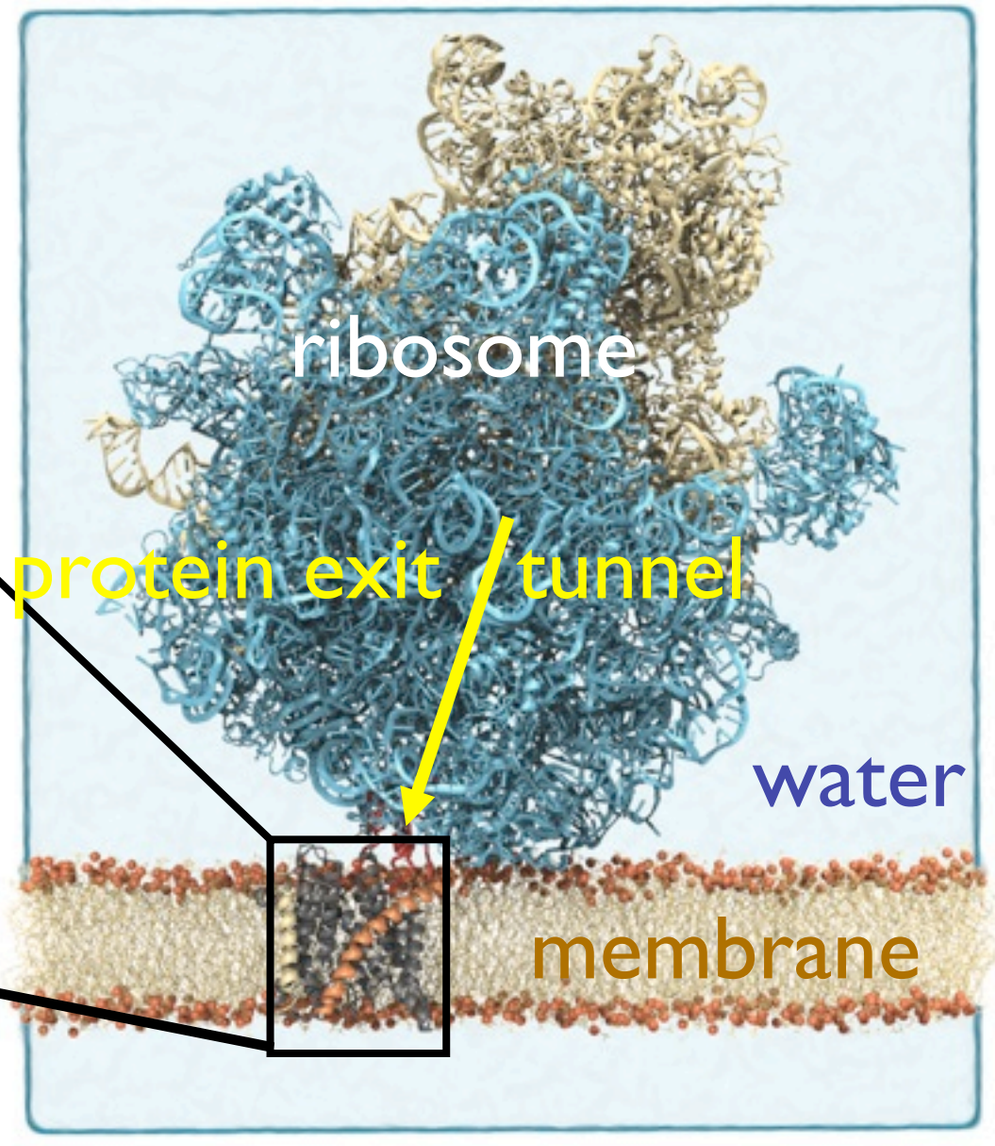
Regulation of the protein-conducting channel by a bound ribosome



Secβ

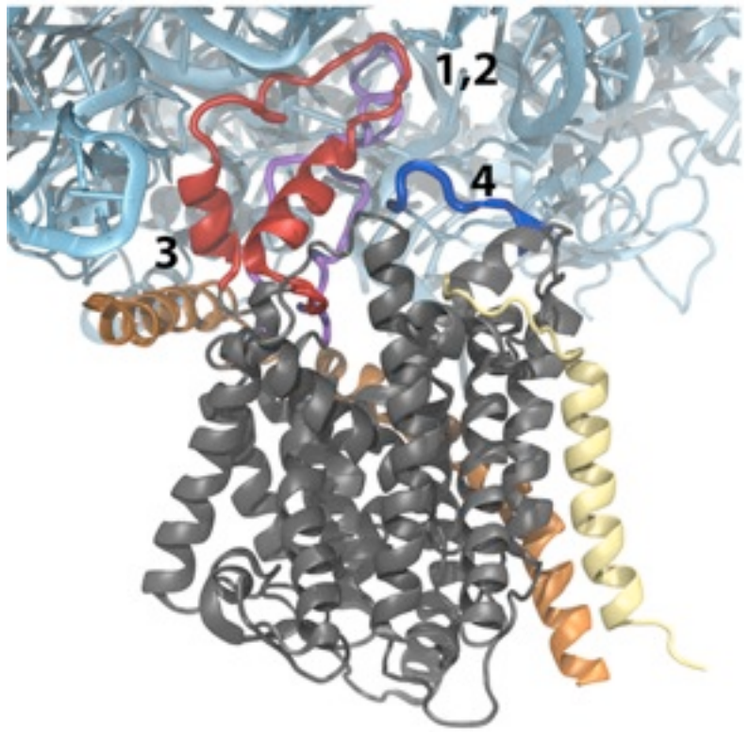
SecE

SecY

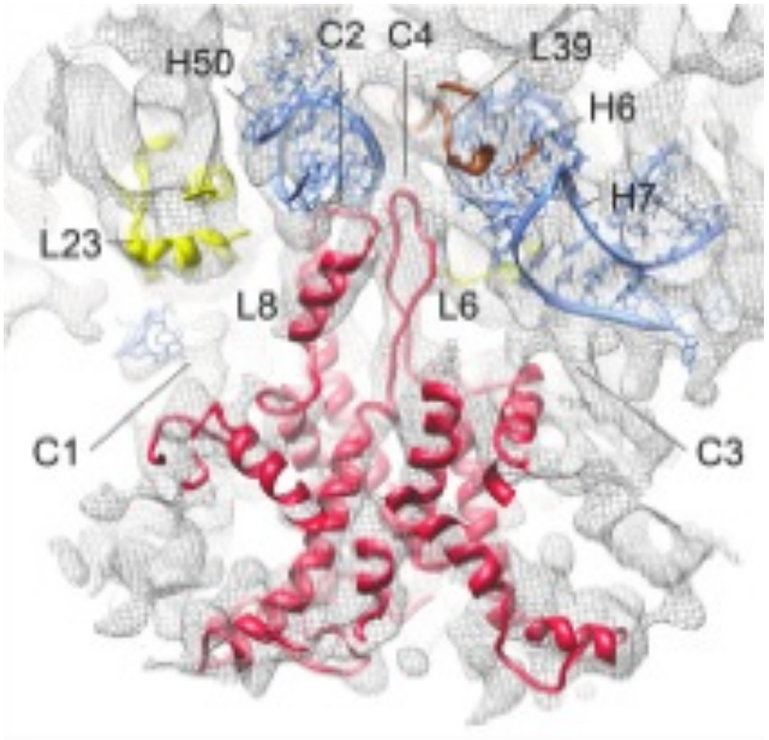


polypeptide translocation

Interactions between the ribosome and the protein conducting channel



Bacterial ribosome-SecYEβ complex



Mammalian ribosome-Sec61 complex

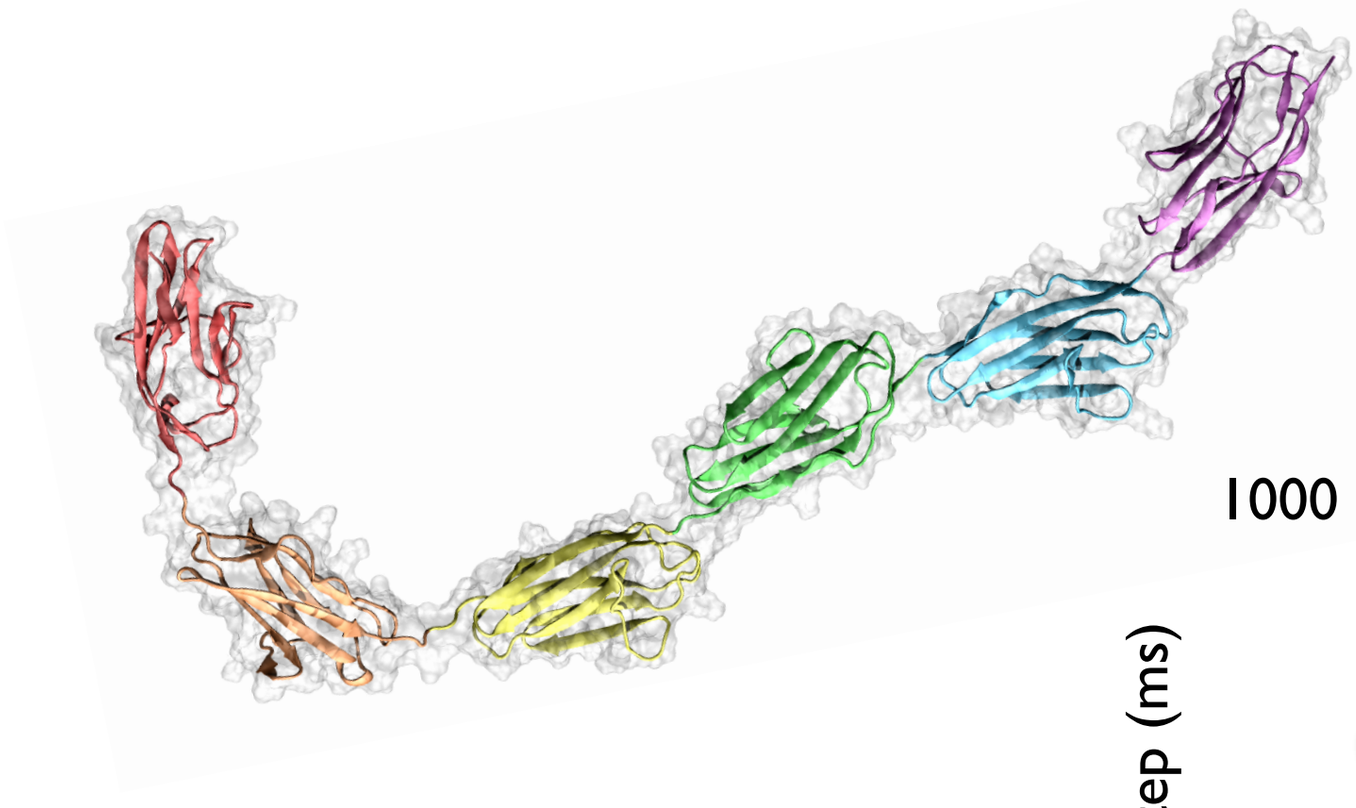
- Binding spots between channel and ribosome can be classified into four groups, all conserved
- Loops 6/7 and 8/9 insert into the exit tunnel and interact primarily with rRNA

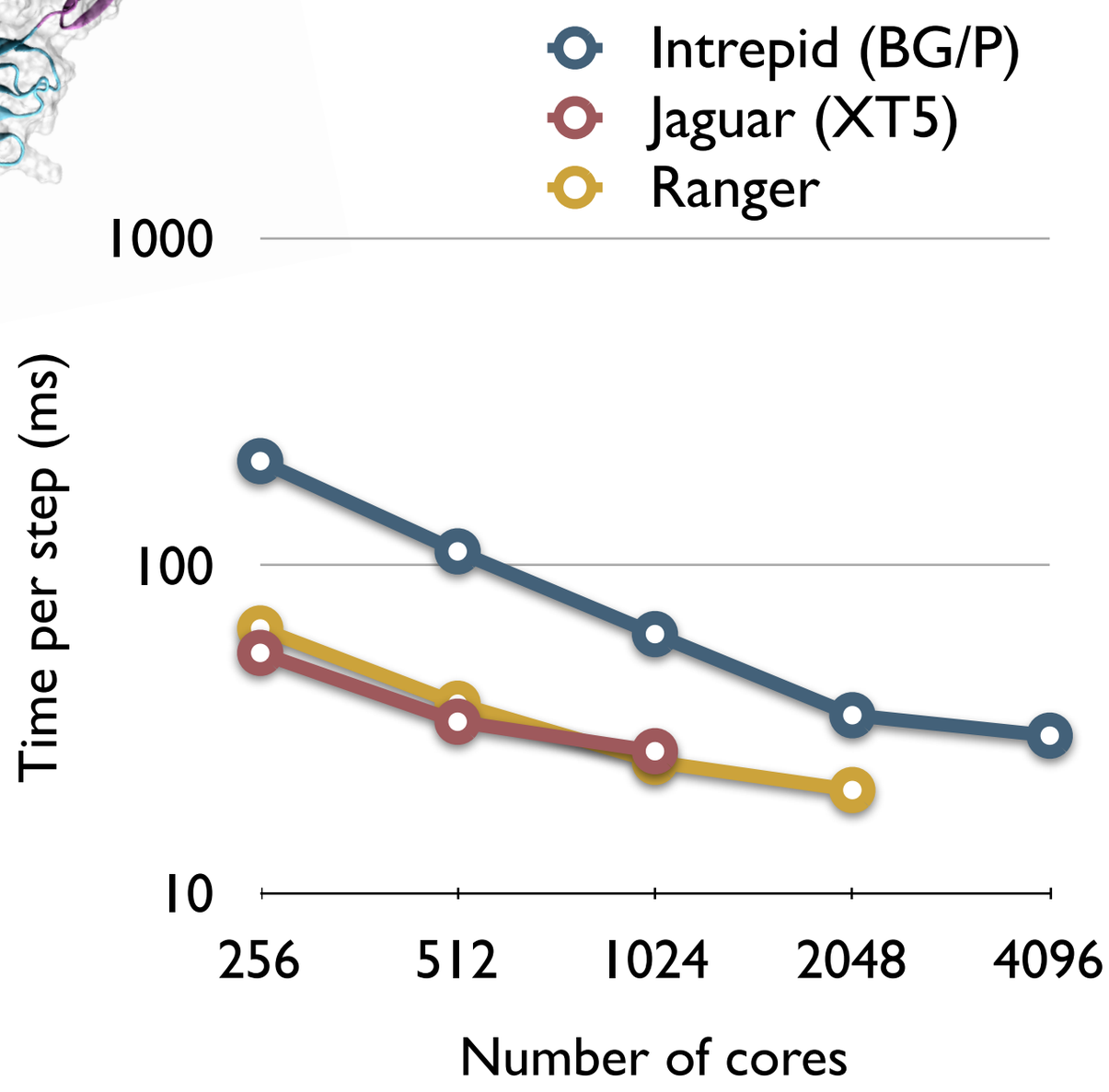
Gumbart et al. Structure (2009) 17:1453-1464.

- Connections are similar to bacterial complex; channel is still a monomer

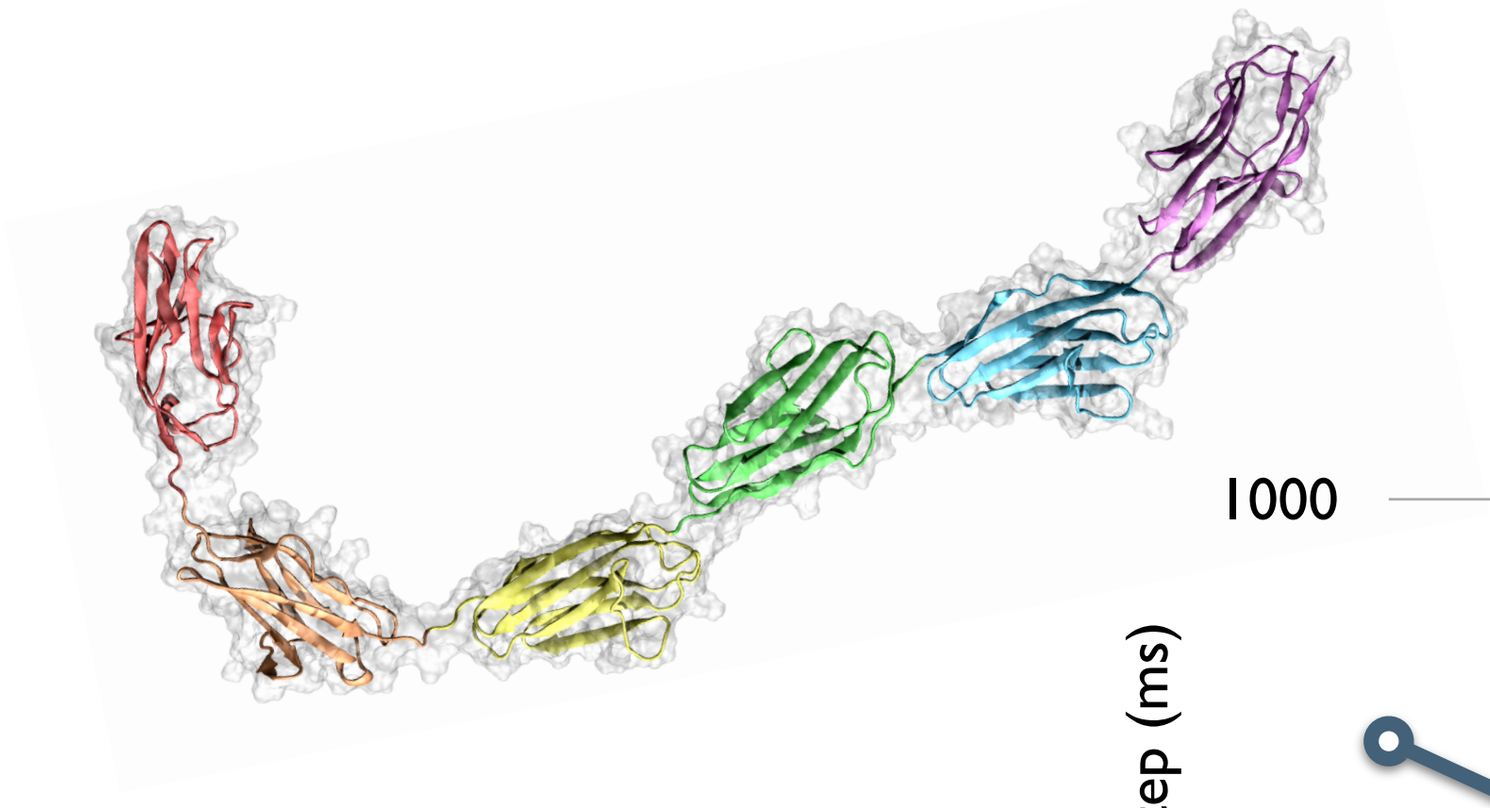
Becker et al. Science (2009) 326:1369-1373

Benchmarking Titin

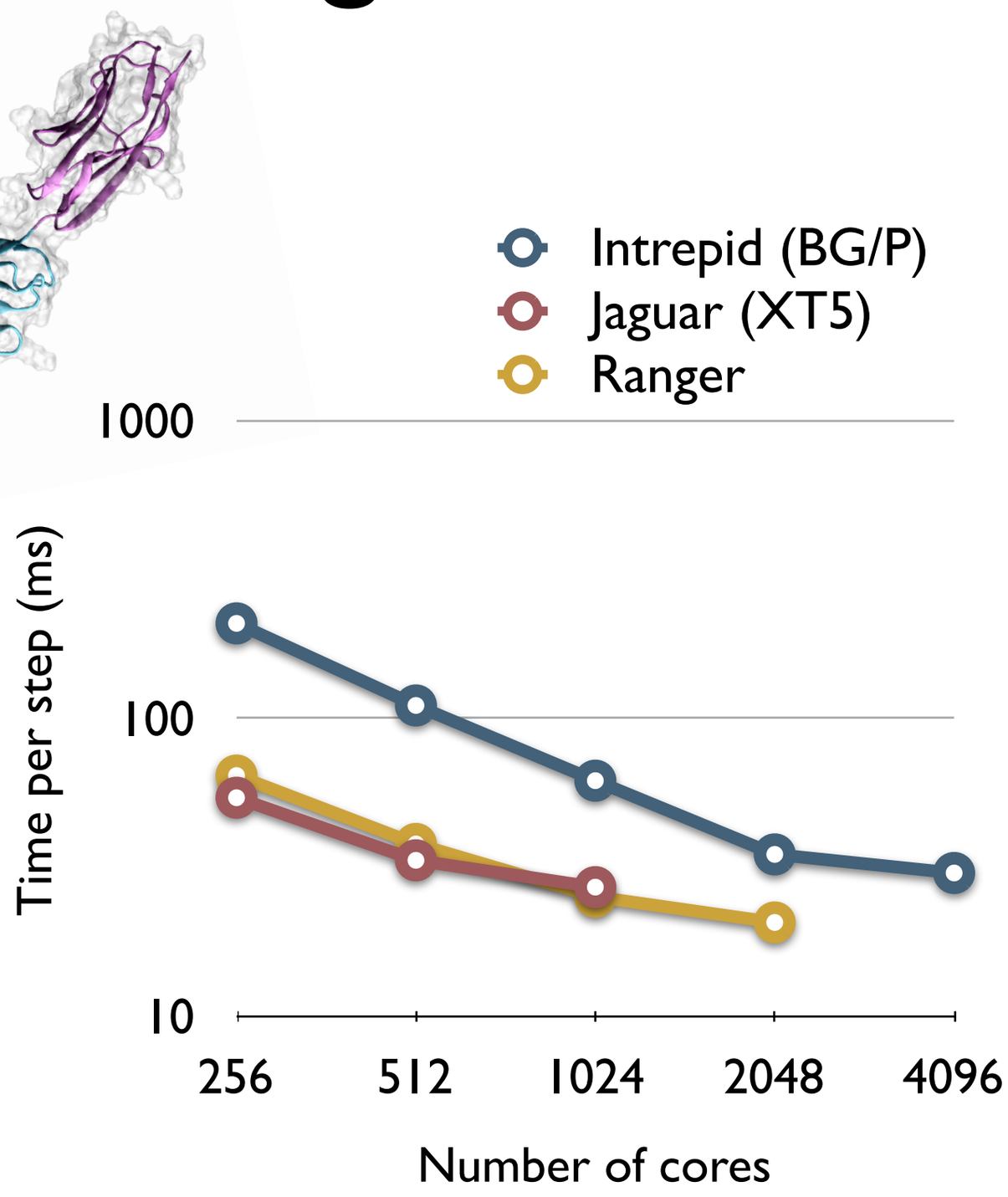




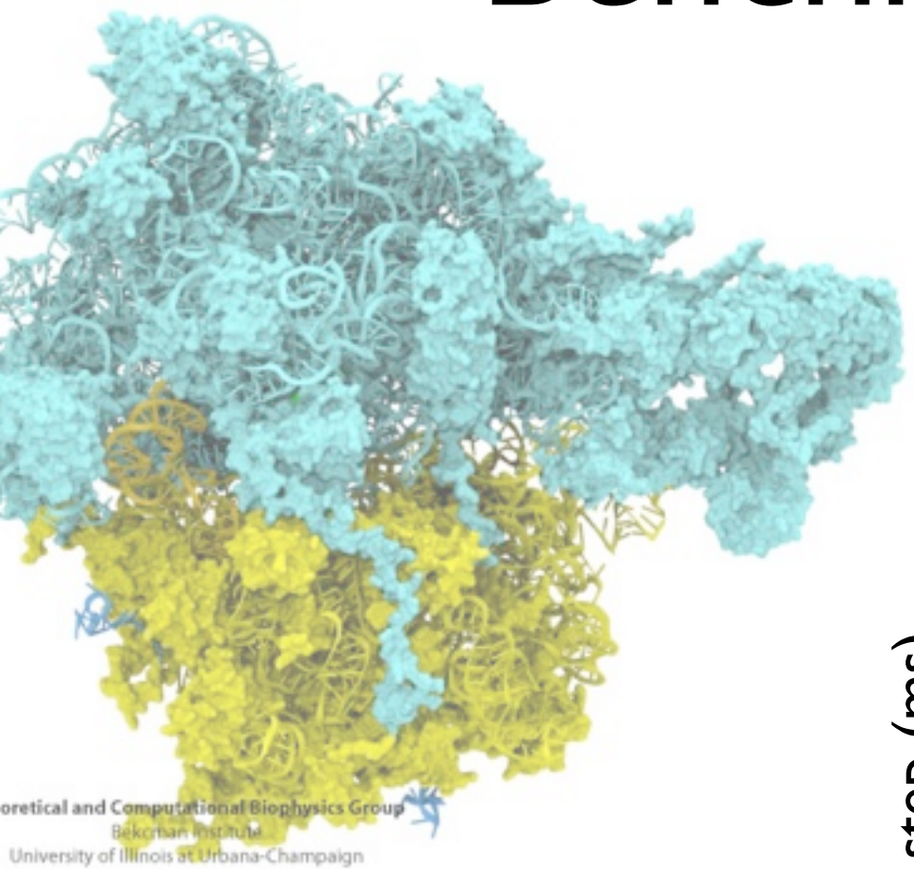
Benchmarking Titin

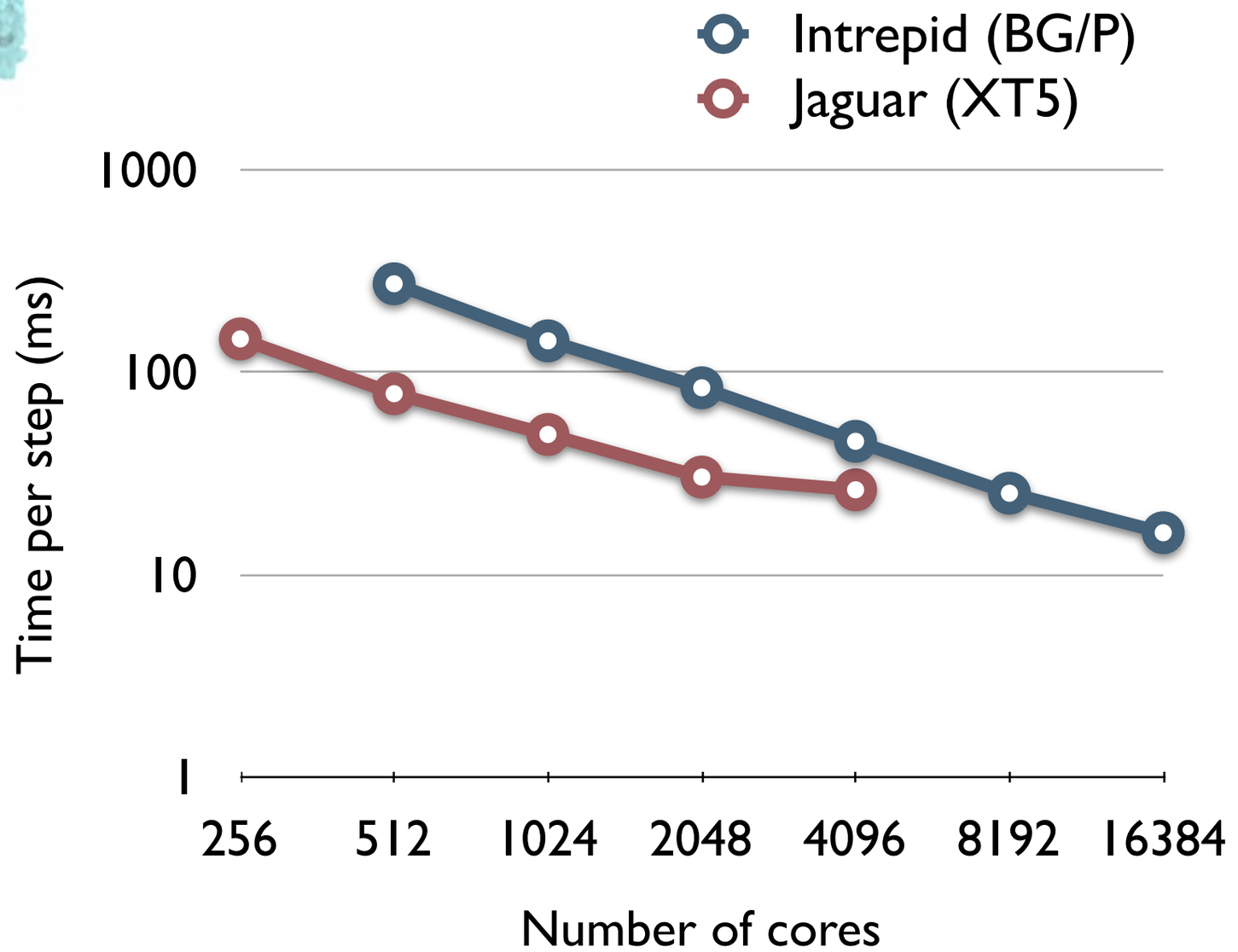


~4.2 ns/day on 2,048 cores of Ranger

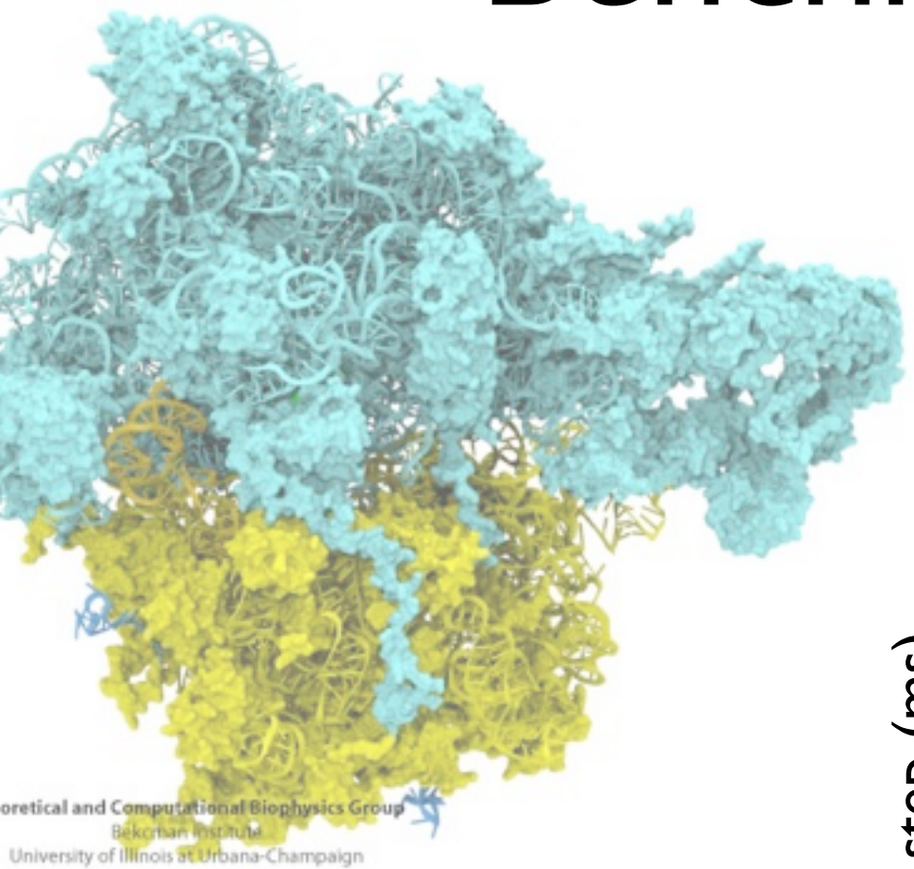


Benchmarking Ribosome

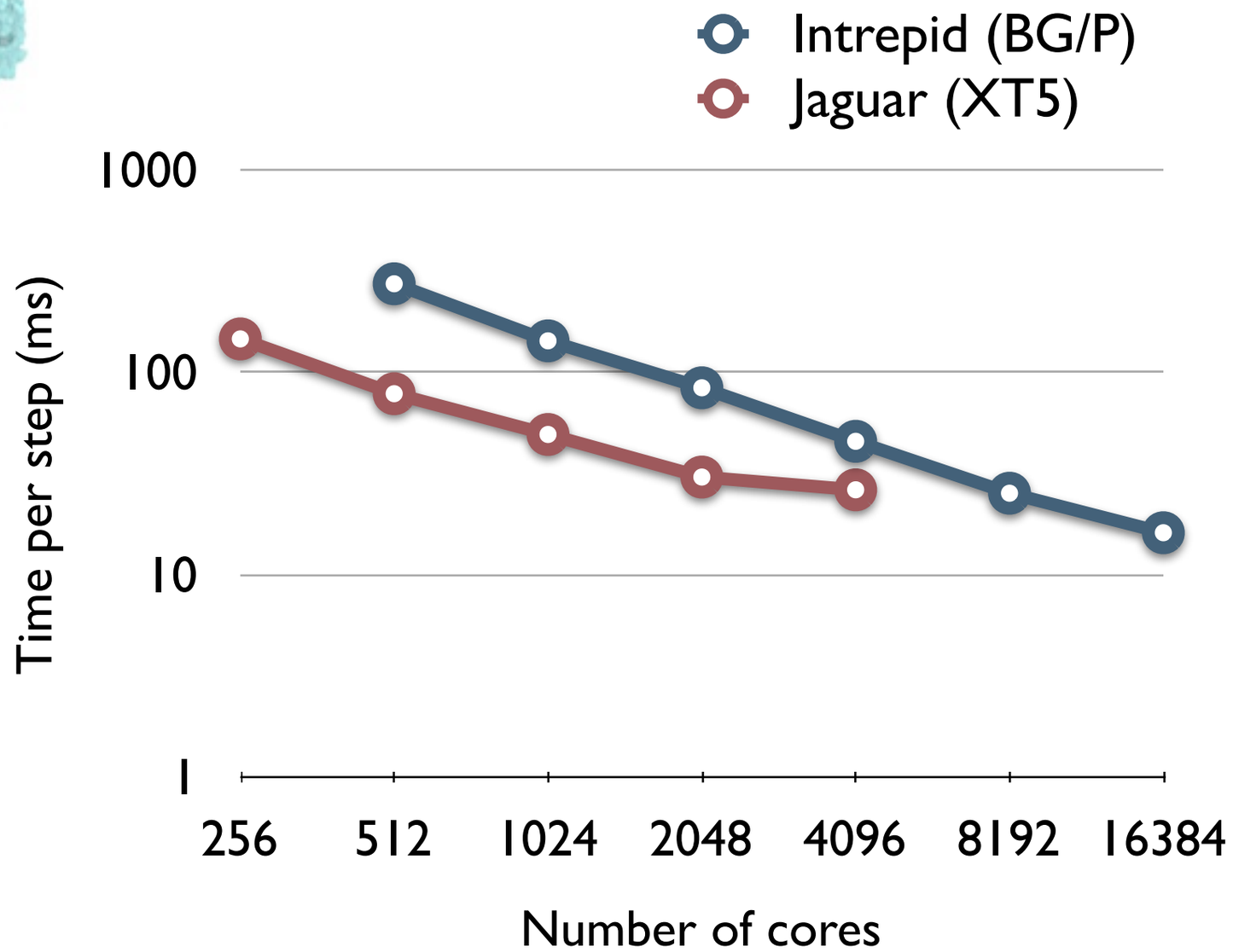




Benchmarking Ribosome



~5.4 ns/day on 16,384 cores of Blue Gene/P



Questions?

Acknowledgements

Allocations on TeraGrid: TG-MCA93S028, TG-ASC050039N
Allocation on Blue Gene/P: supported by DOE under contract DE-AC02-06CH11357
Allocation on Jaguar: supported by the DOE under contract DE-AC05-00OR22725
(under the PEAC project)