IBM Research Division

Blue Matter: Strong Scaling of Molecular Dynamics on Blue Gene/L

Robert S. Germain
Biomolecular Dynamics & Scalable Modeling
http://www.research.ibm.com/bluegene/

December 1999:

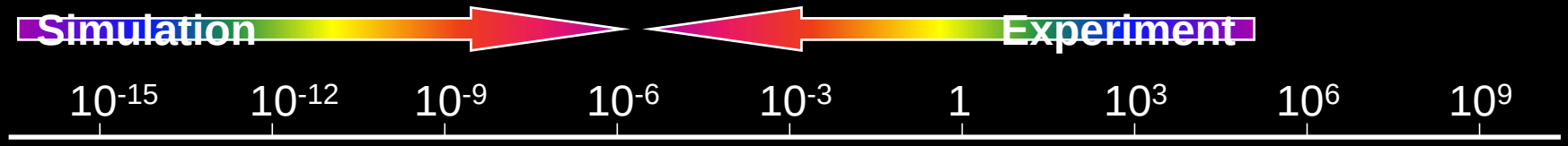
IBM Announces $100 Million Research Initiative to build World's Fastest Supercomputer

"Blue Gene" to Tackle Protein Folding Grand Challenge

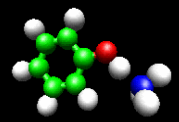
YORKTOWN HEIGHTS, NY, December 6, 1999 -- IBM today announced a new $100 million exploratory research initiative to build a supercomputer 500 times more powerful than the world's fastest computers today. The new computer -- nicknamed "Blue Gene" by IBM researchers -- will be capable of more than one quadrillion operations per second (one petaflop). This level of performance will make Blue Gene 1,000 times more powerful than the Deep Blue machine that beat world chess champion Garry Kasparov in 1997, and about 2 million times more powerful than today's top desktop PCs.

Blue Gene's massive computing power will initially be used to model the folding of human proteins, making this fundamental study of biology the company's first computing "grand challenge" since the Deep Blue experiment. Learning more about how proteins fold is expected to give medical researchers better understanding of diseases, as well as potential cures.

Time Scales: Biopolymers and Membranes

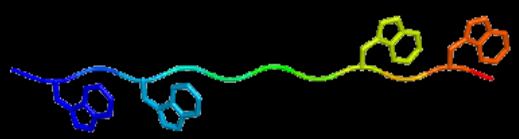


Bond Vibration



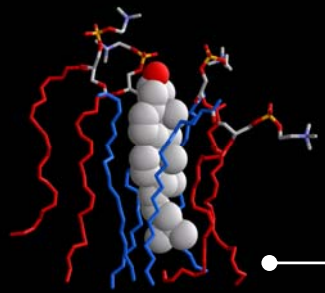
DNA Twisting

Hinge Motion



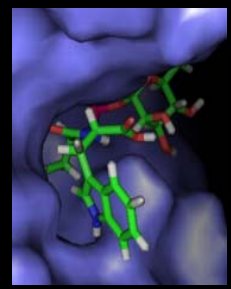
Helix-Coil Transition

Protein Folding



Lipid exchange via diffusion

Ligand-Protein Binding



Torsional correlation in lipid headgroups

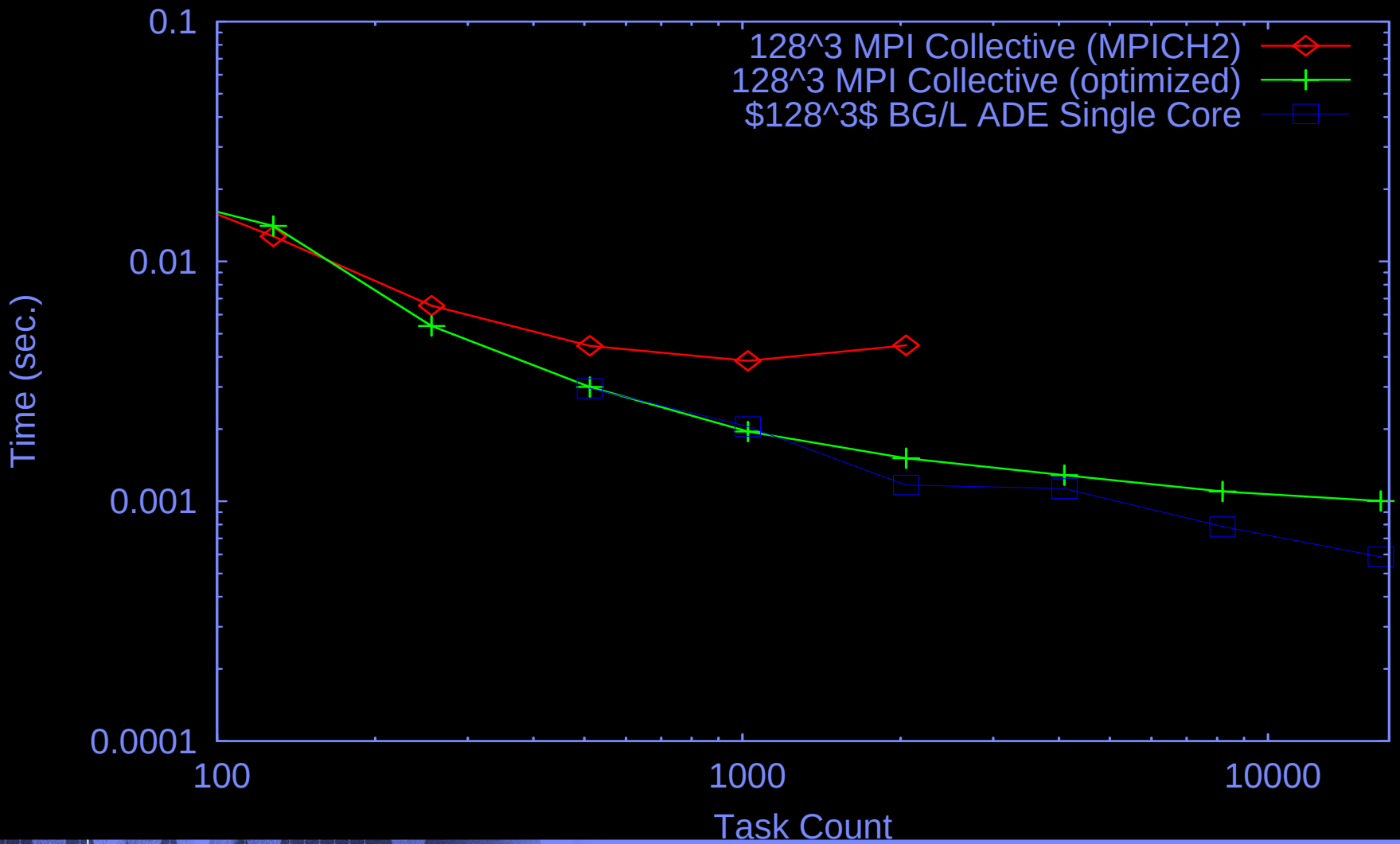
Electron Transfer

Adapted from “The Protein Folding Problem”, Chan and Dill, Physics Today, Feb. 1993

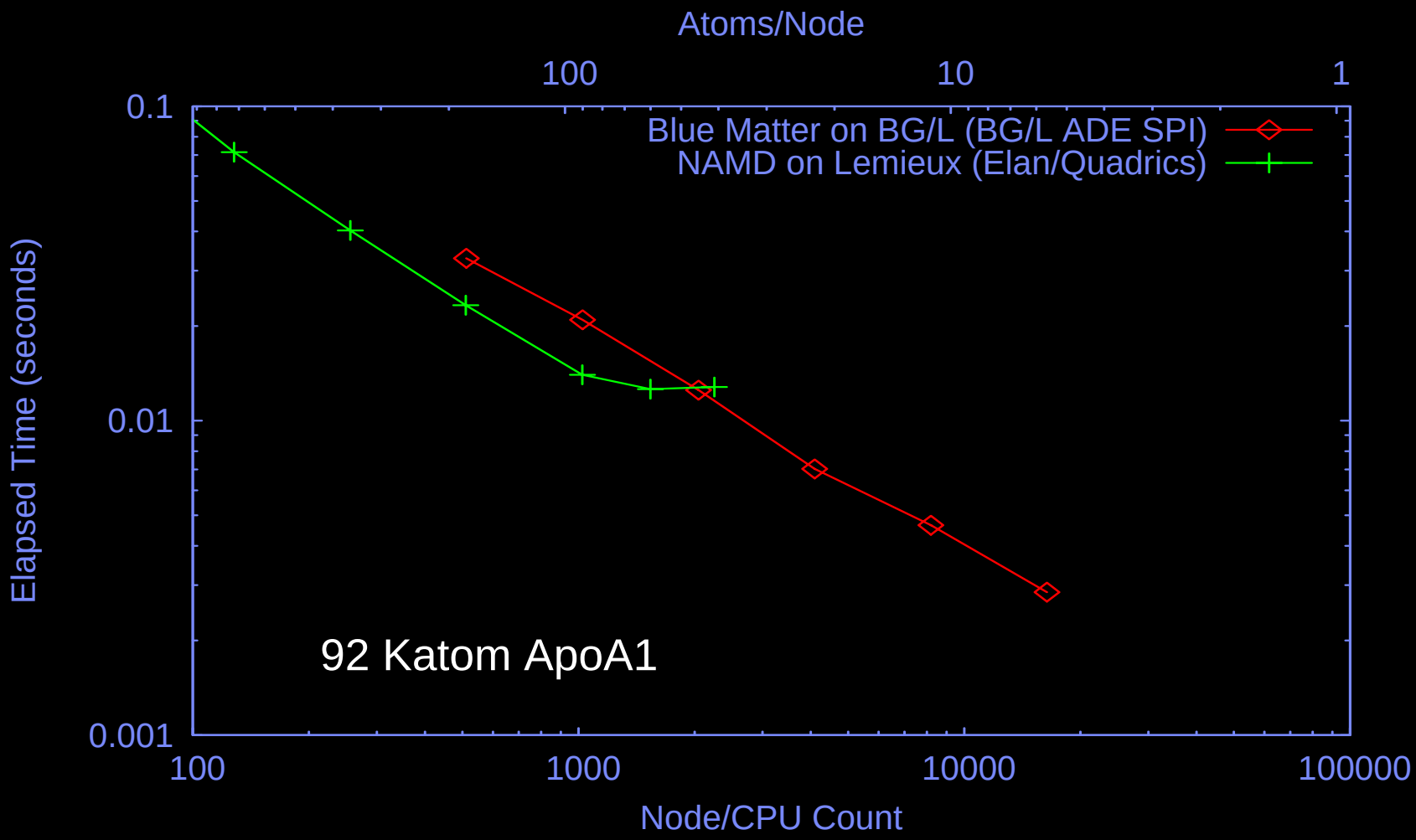
What Limits the Scalability of MD?

- **Inherent limitations on concurrency:**
  - Bonded force evaluation
    - *Represents only small fraction of computation, can be distributed moderately well.*
  - Real space non-bond force evaluation
    - *Large fraction of computation, but good distribution can be achieved using volume or interaction decompositions*.
  - Reciprocal space contribution to force evaluation for Ewald
    - *P3ME uses 3D FFT with global communication*
    - *Ewald with direct evaluation uses floating point reduction*
- **Load balancing**
- **System hardware/software overheads**

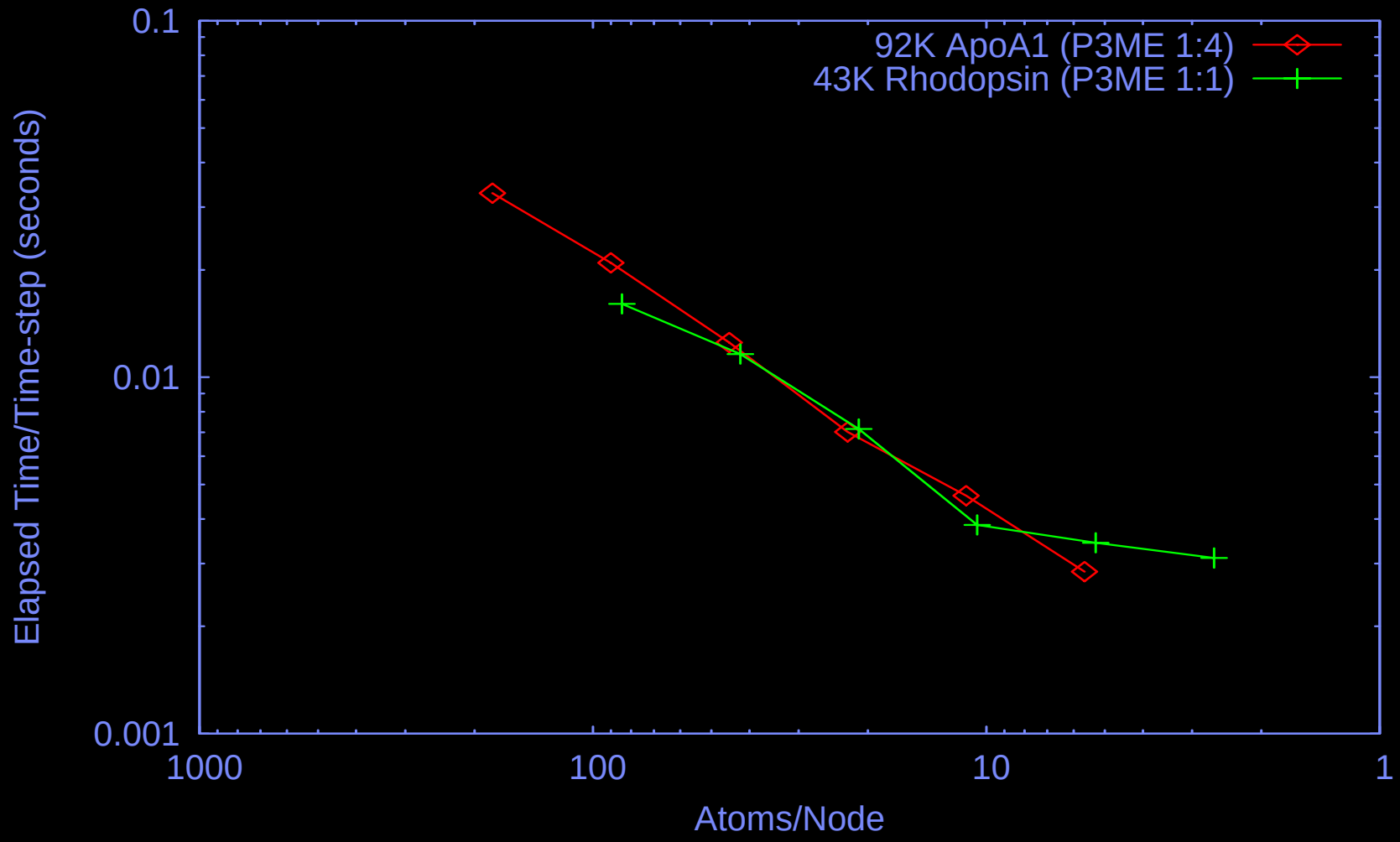
Distributed 3D-FFT Performance on Blue Gene/L



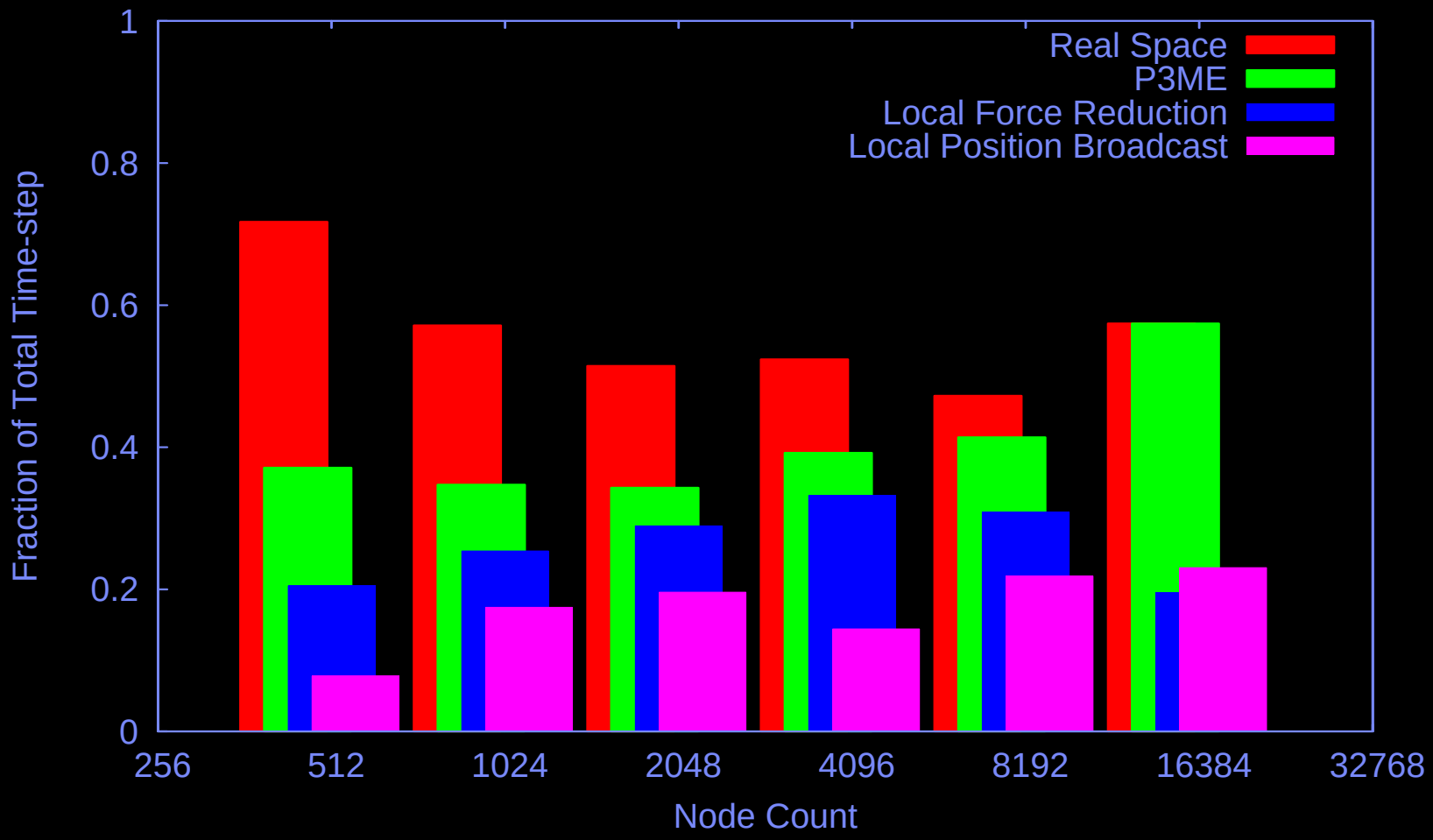
Blue Matter on BG/L vs. NAMD on PSC Lemieux



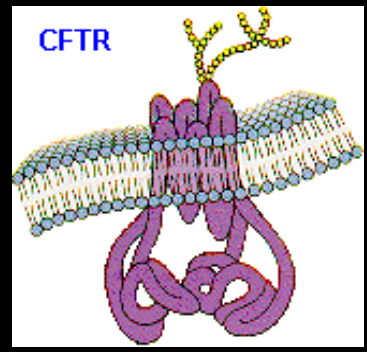
Strong Scaling of Blue Matter on Blue Gene/L



43 Katom Rhodopsin (Verlet, P3ME every step)

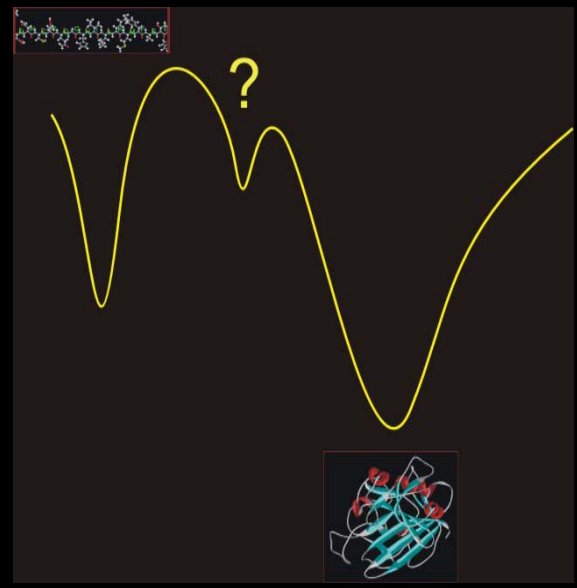


Why Protein Folding?



Folding-related diseases

Cystic Fibrosis, Alzheimer's, BSE



The Three Protein Folding Questions

Why does a protein fold?

Thermodynamics

How does it fold so quickly?

Kinetics

What structure does it fold to?

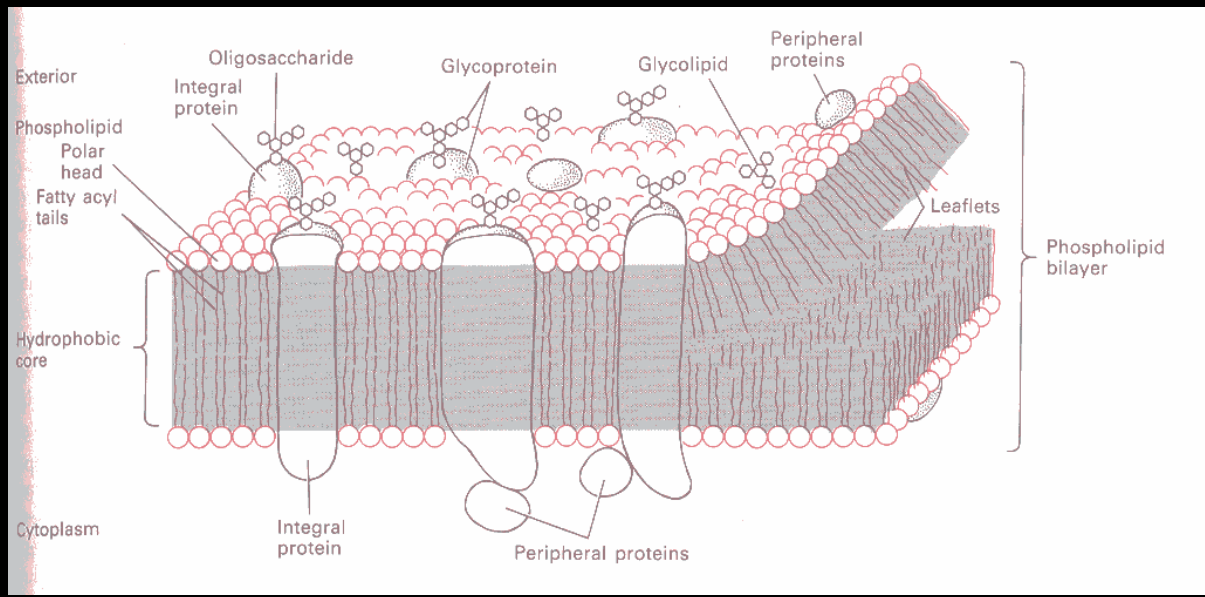
Structure Prediction

- Starting with simple systems, the Blue Gene science program is using high-quality thermodynamic and kinetic simulations to study the protein folding process.

Membrane Proteins

- **Membrane processes enable:**
  - cell signal detection, ion and nutrient transport
  - infection processes target specific membranes
  - Over 50% of drug discovery research targets are membrane proteins

Experiment and simulation play a concerted role in understanding membrane biophysics

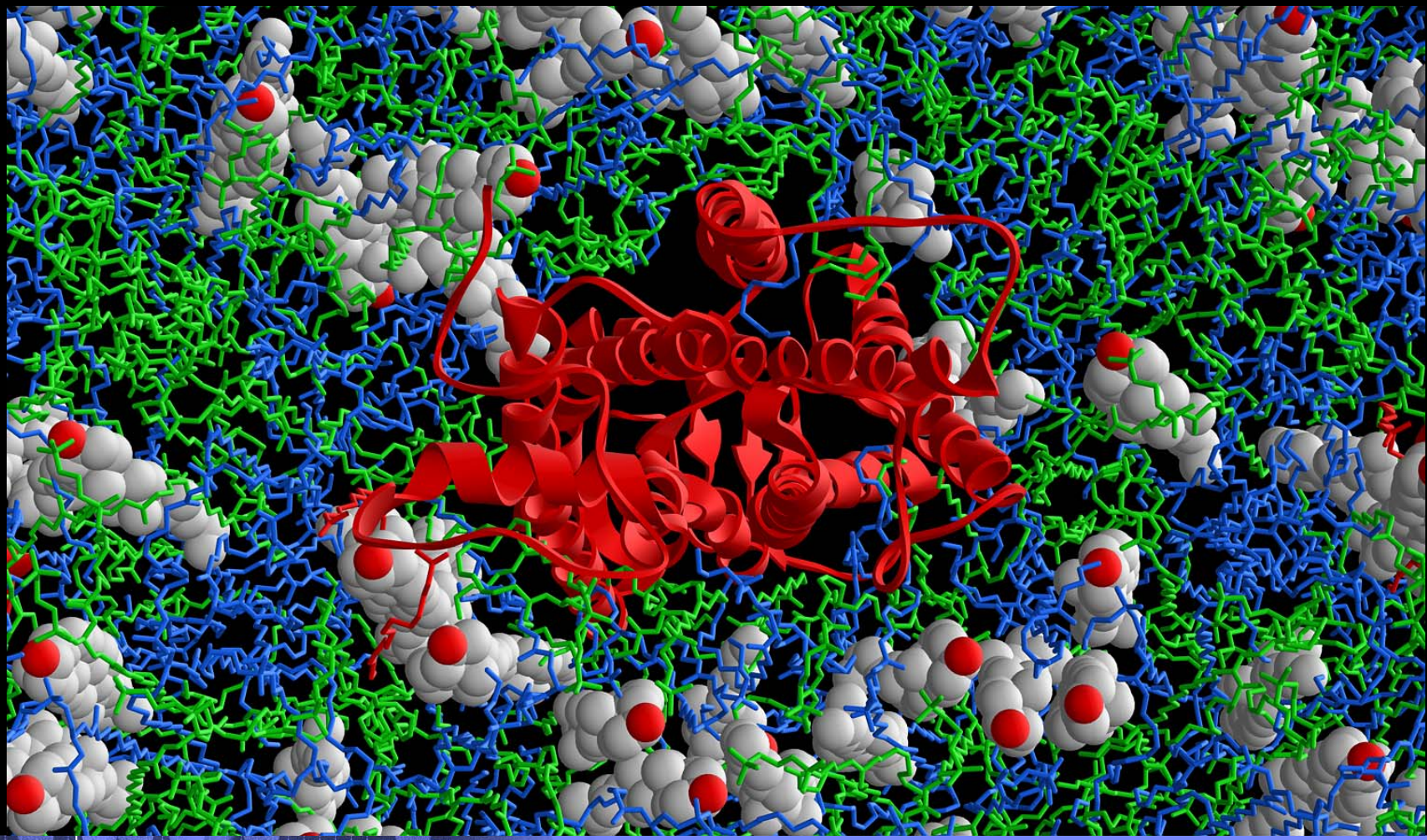


Simulation can be validated by experiment

Simulation can then help to interpret experiment

Current Simulations of Rhodopsin in Membrane

Rhodopsin in 2:2:1 SDPE/SDPC/Cholesterol after 120ns



Biomolecular Simulation on Blue Gene/L at Watson (BGW)

- **Rhodopsin in lipid bilayer with cholesterol (44,000 atoms)**
  - 118ns “dark-adapted”, 750+ ns “light-adapted”– used 512→1024→2048 →4096 nodes
- **Rhodopsin (“dark ensemble”):**
  - 26 simulations of rhodopsin, each > 100ns, used 1024→2048 nodes/trajectory
- **Lysozyme misfolding (40,000 atoms)**
  - 2 x 1 μsec trajectories, using 2048 → 4096 nodes/trajectory
- **Lambda Repressor (67,000 atoms)**
  - Replica exchange: 256 replicas on total of 8192 nodes
  - Multiple (2+) kinetic simulations (8 μsec): will use 4096 nodes/trajectory

Summary

- **Greatly improved times to solution for classical biomolecular simulations are possible on BG/L**
- **New scale of atomistic simulations is enabled**
- **Both capability and capacity applications of BG/L to biomolecular simulation are currently underway as part of the Blue Gene science effort**

Selected Publications

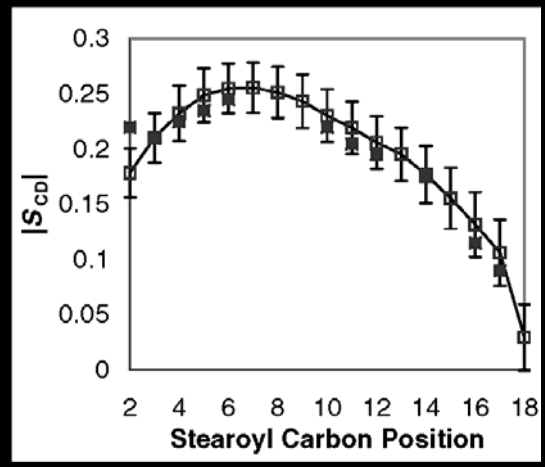
- Molecular dynamics investigation of dynamical properties of phosphatidylethanolamine lipid bilayers; The Journal of Chemical Physics 2005; 122(24) , 244715
- Molecular dynamics investigation of the structural properties of phosphatidylethanolamine lipid bilayers; The Journal of Chemical Physics 2005; 122(24) , 244714
- Role of Cholesterol and Polyunsaturated Chains in Lipid-Protein Interactions: Molecular Dynamics Simulation of Rhodopsin in a Realistic Membrane Environment J. Am. Chem. Soc.; 2005; *127*(13) pp 4576 - 4577
- Molecular-Level Organization of Saturated and Polyunsaturated Fatty Acids in a Phosphatidylcholine Bilayer Containing Cholesterol; Biochemistry 43(49); 2004; 15318-15328
- "Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew", J. Chem. Phys. v120 9665-9678, (2004)
- Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 1. Theory; The Journal of Physical Chemistry B; 2004; 108(21); 6571-6581
- Describing Protein Folding Kinetics by Molecular Dynamics Simulations. 2. Example Applications to Alanine Dipeptide and a beta-Hairpin Peptide; The Journal of Physical Chemistry B; 2004; 108(21); 6582-6594
- Understanding folding and design: Replica-exchange simulations of "Trp-cage" miniproteins, PNAS USA, Vol. 100, Issue 13, June 24, 2003, pp. 7587-7592
- Can a continuum solvent model reproduce the free energy landscape of a beta-hairpin folding in water?, Proc. Natl. Acad. Sci. USA, Vol. 99, Issue 20, October 1, 2002, pp. 12777-12782
- The free energy landscape for beta-hairpin folding in explicit water, Proc. Natl. Acad. Sci. USA, Vol. 98, Issue 26, December 18, 2001, pp. 14931-14936

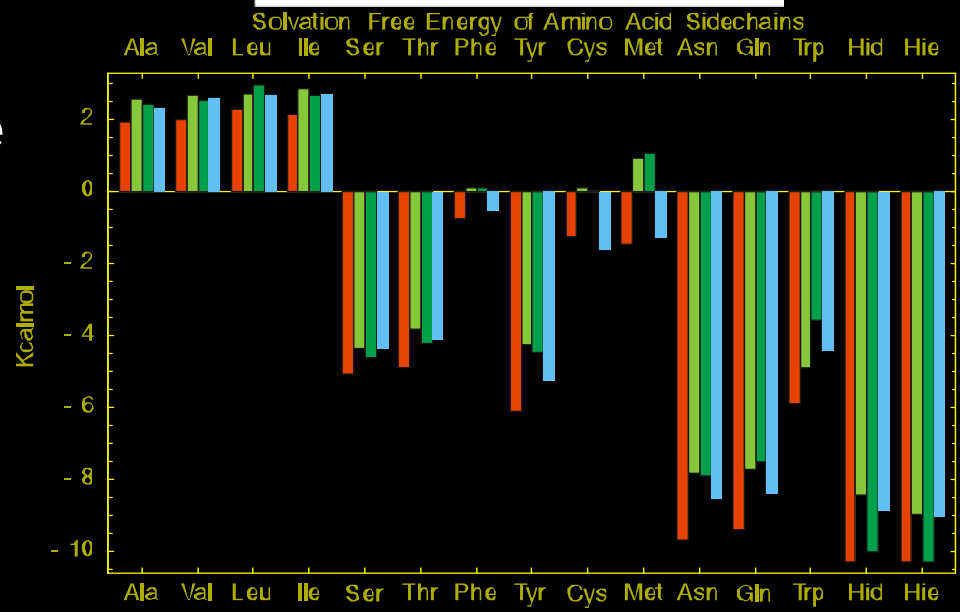
Acknowledgements

- Alex Balaeff
- Bruce Berne
- Maria Eleftheriou
- Scott Feller
- Blake Fitch
- Klaus Gawrisch
- Alan Grossfield
- Jed Pitera
- Blue Gene Hardware and System Software teams
- Mike Pitman
- Alex Rayshubskiy
- Yuk Sham
- Frank Suits
- Bill Swope
- Chris Ward
- Yuri Zhestkov
- Ruhong Zhou

How can we use large scale computational resources?

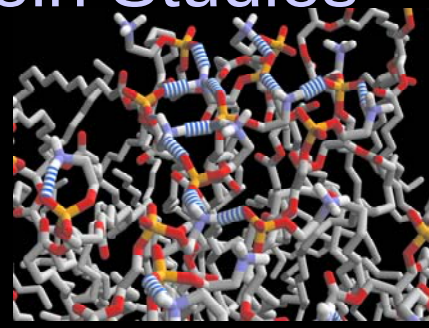
- **Capability**
  - Increase time scales probed (strong scaling)
  - Increase system size studied
- **Capacity**
  - Improve sampling to reduce statistical uncertainties
  - Run large ensembles of trajectories
- **Make contact with experiment**



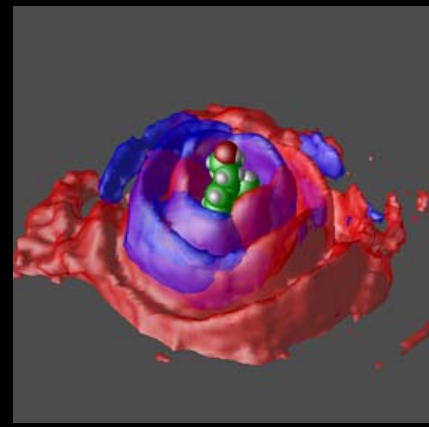


Overview of Blue Gene Membrane/Membrane Protein Studies

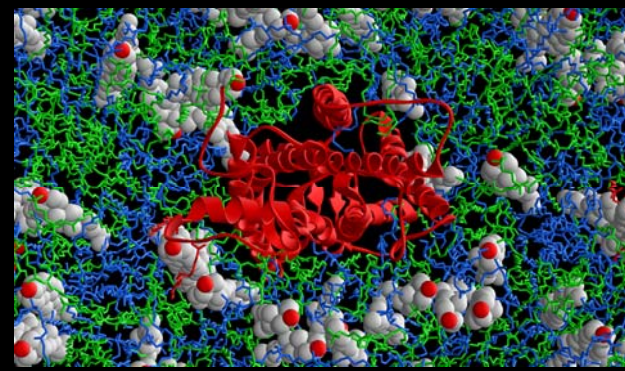
- **SOPE (JCP 2005)**
  - Extensive hydrogen bonding network with headgroups
  - Excellent agreement with experiment for both structural and dynamic properties



- **3:1 SDPC/Cholesterol (Biochemistry 2004)**
  - Cholesterol induces dramatic lateral organization
  - Cholesterol shows preference of STEA over DHA
  - Significant Angular anisotropy of Cholesterol Environment



- **GPCR in a membrane environment (JACS 2005, …)**
  - Rhodopsin with 2:2:1 SDPC/SDPE/CHOL
  - 100 ns *cis*-retinal - 650+ ns *trans*-retinal
  - Current production rate ~9.5ns/day on 2048 nodes BG/L

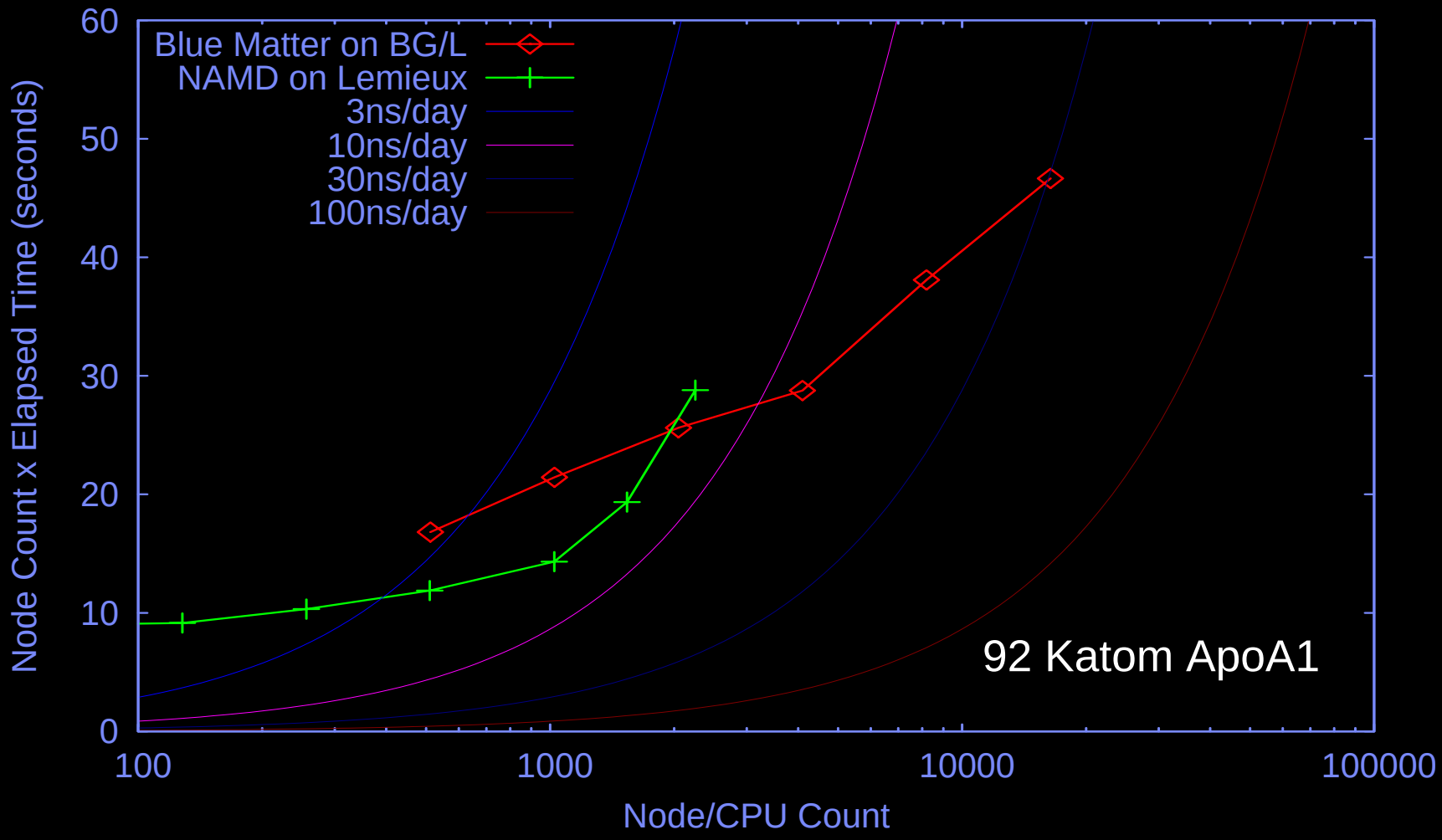


GPCR-based drugs among the 200 best-selling prescriptions, and their GPCR targets

GPCR target	Drug	Disease	Company	2000 sales(US $m)
Histamine receptors	Zantac	Ulcers	AstraZeneca	870
	Pepcid		Merck	850
	Claritin	Allergies	Schering-Plough	2,200
	Allegra		Aventia	1,100
5-HT receptors	Risperdal	Psychosis	Johnson & Johnson	1,600
	Imitrex	Migraine	GlaxoSmithKline	1,100
	BuSpar	Anxiety	Bristol-Myers Squibb	714
	Zyprexa	Schizophrenia	Eli Lilly	2,400
Angiotensin receptors	Cozaar	Hypertension	Merck	1,700
Adrenoceptors	Toprol-XL		AstraZeneca	580
	Coreg	Congestive heart failure	GlaxoSmithKline	250
	Serevent	Asthma	GlaxoSmithKline	940
Muscarinic acetylcholine receptors	Atrovent	COPD	Boehringer Ingelheim	600
GnRH receptors	Zoladex	Cancer	AstraZeneca	740
Dopamine receptors	Requip	Parkinson's diseases	AstraZeneca	90
Prostaglandin (PGE1) receptors	Cytotec	Ulcers	Pharmacia	100
ADP receptors	Plavix	Stroke	Bristol-Myers Squibb	900

http://www.predixpharm.com/market_table.htm

Blue Matter on BG/L vs. NAMD on PSC Lemieux



Blue Matter Performance on 43K Atom Rhodopsin

Nodes				Time/time-step (seconds)				Atoms/node
				MPI		BGL/ADE		
Total	Px	Py	Pz	Single	Dual(1)	Dual(1)	Dual(2)	
32	4	4	2	0.4471	0.3646			1351
128	8	4	4	0.1322	0.0911			338
512	8	8	8	0.0317	0.0253	0.0234	0.0161	84
1024	16	8	8	0.0206	0.0185	0.0162	0.0116	42
2048	16	16	8	0.0137	0.0102	0.0097	0.0072	21
4096	32	16	8	0.0156	0.0135		0.0067	11
4096	16	16	16	0.0104	0.009	0.0054	0.0039	11
8192	32	16	16				0.0034	5.3
16384	32	32	16				0.0031	2.6

For the 2 femtosecond time-step used in production, 3.1ms/ts → 55 ns/day

FLOP Rates for Classical MD (ApoA1 with MTS)