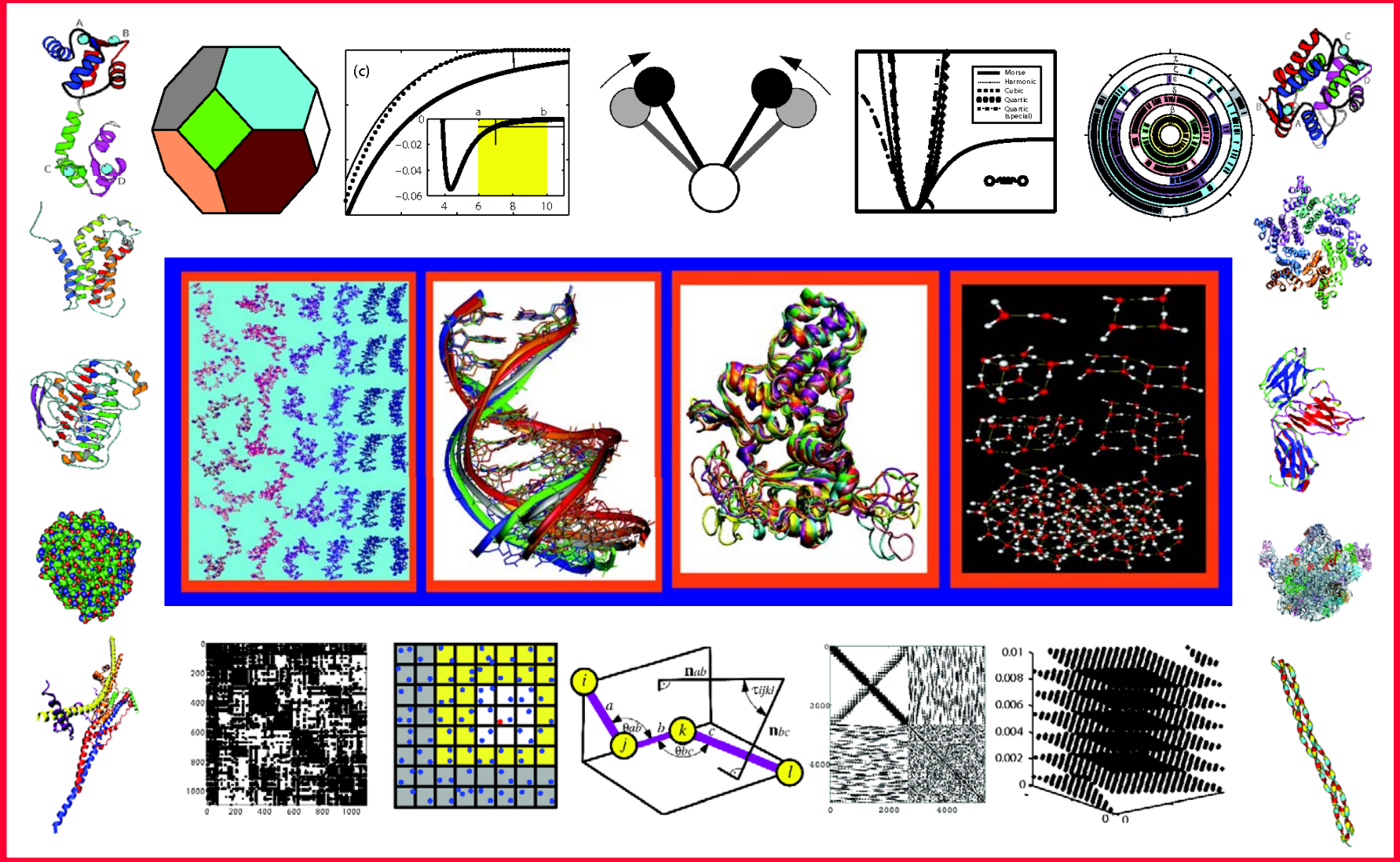
Biomolecular Modeling and Simulation: A Field Coming of Age

A Recent Assessment



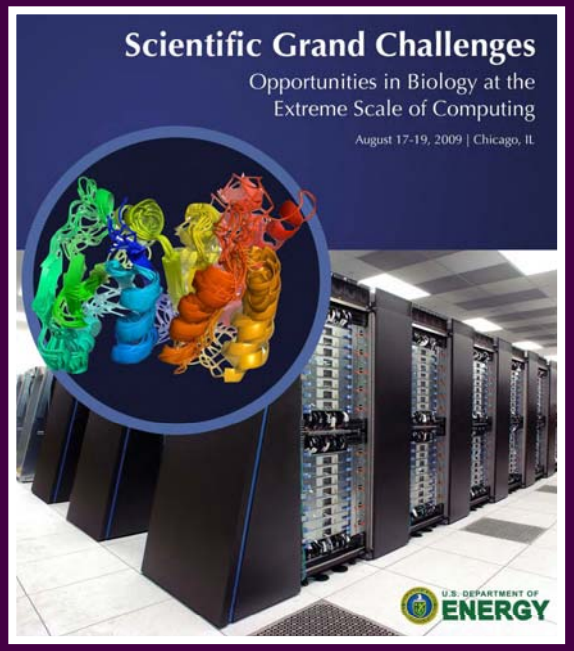
Why is a Perspective Timely?

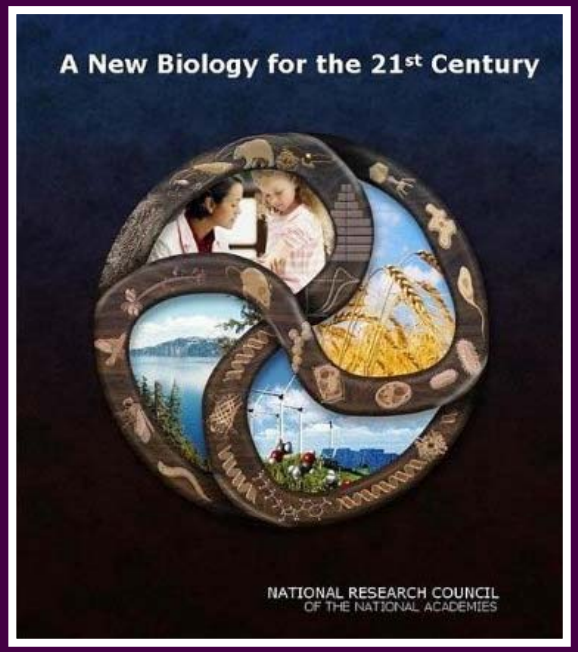
- Recent workshops have asked biologists to outline challenges that will best utilize next-generation computing

  - DOE, Aug. 2009, Opportunities in biology at the extreme scale of computing
  - NIH, Dec. 2009, IMAG Futures meeting: the impact of modeling on biomedical research

- Some have suggested that biologists have not sufficiently utilized available resources and pushed specific agendas

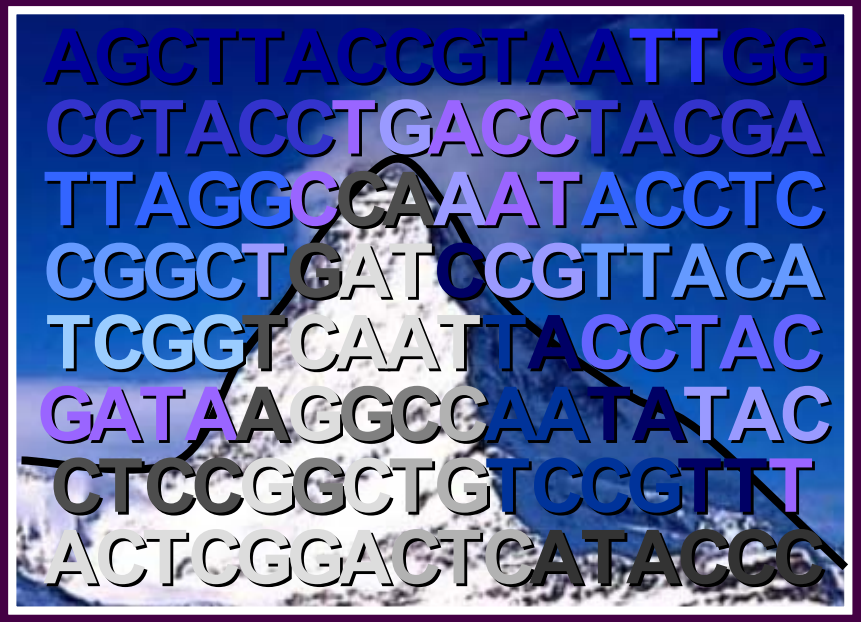
- NAS has urged for re-integration of biology's many subdisciplines and integration into biology of other fields (NAS Report, 2009)





Gaps and Complexity

- A growing gap between academe and industry has arisen
- A disparity between basic and applied research has also become apparent
- Increasing awareness of the complexity in biology has set in: this affects impact of modeling

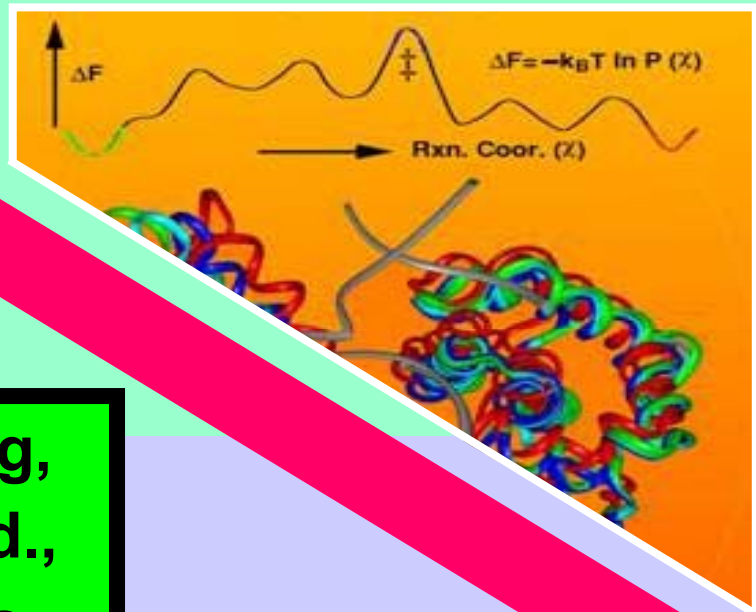


E. Check Hayden, *Nature*, 464:664 (2010)

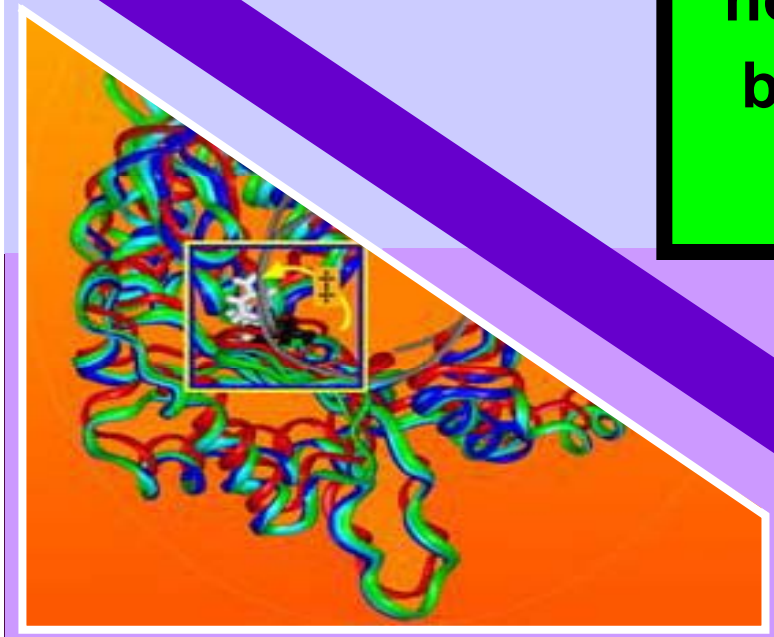
Modeling Tools in Academia vs. Industry

Industry top

Cheminformatics, ligand simulation, visualization



QSAR, docking, homology mod., bioinformatics, QM



Systems biology, de novo design, MD

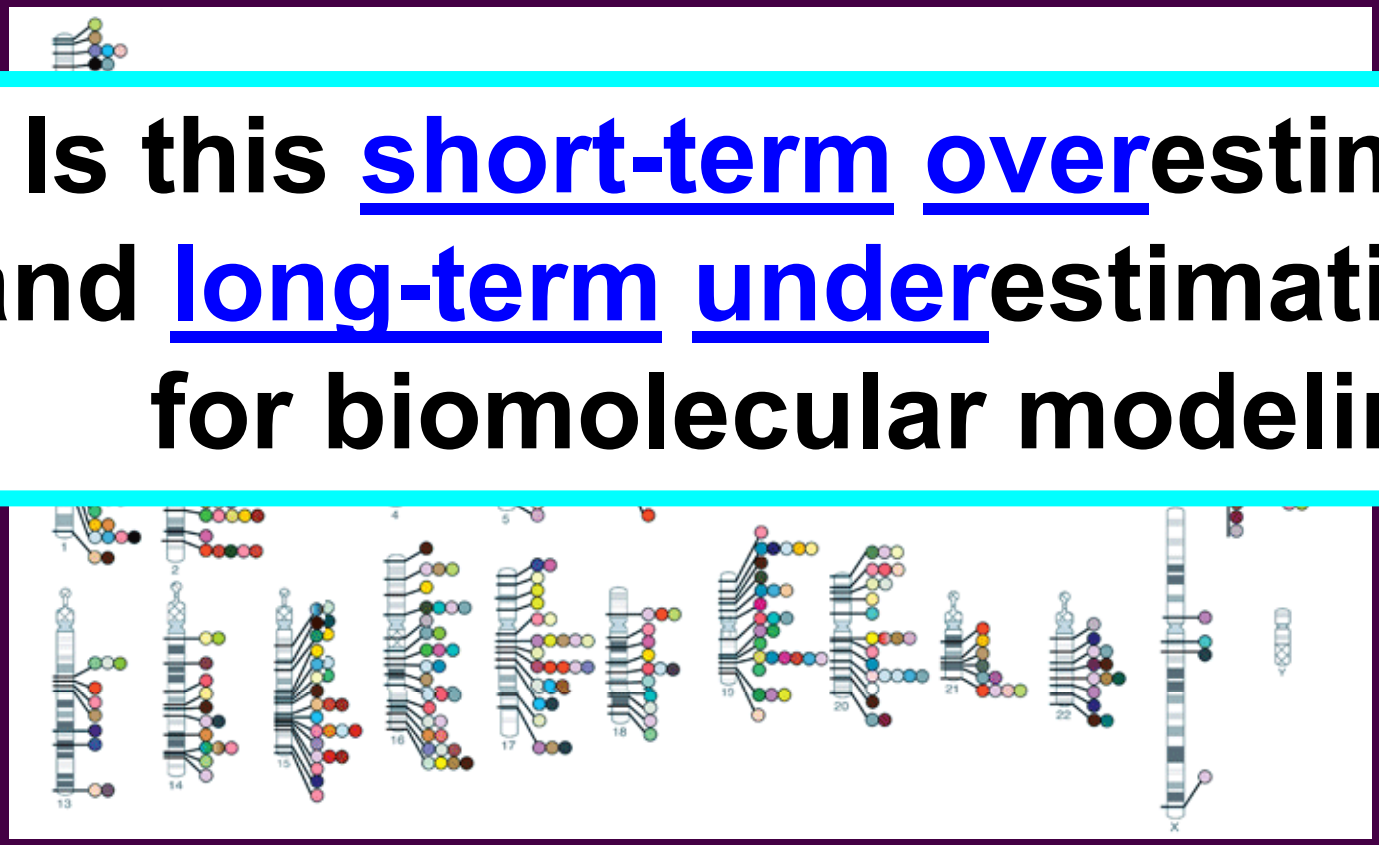
Academia top

Complexity and Innovation

"It's been said that a truly transformational technology will always have its immediate consequences overestimated and its long-term consequences underestimated"

Francis S. Collins, June 2010

Genetic variations and disease



Genetic

Is this short-term overestimation and long-term underestimation true for biomolecular modeling?

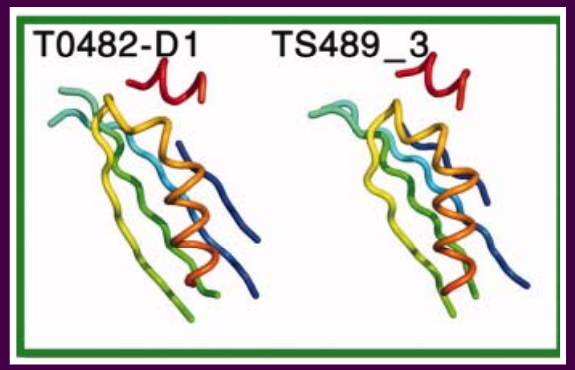
C&E News, June 21, p. 30 (2010)

Assessment Questions

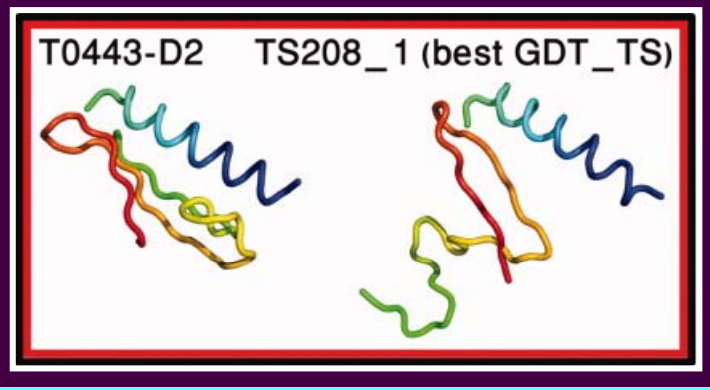
1. Is the field of biomolecular modeling and simulation in an existential crisis or just a mid-life crisis?

2. What are examples of success and failure?

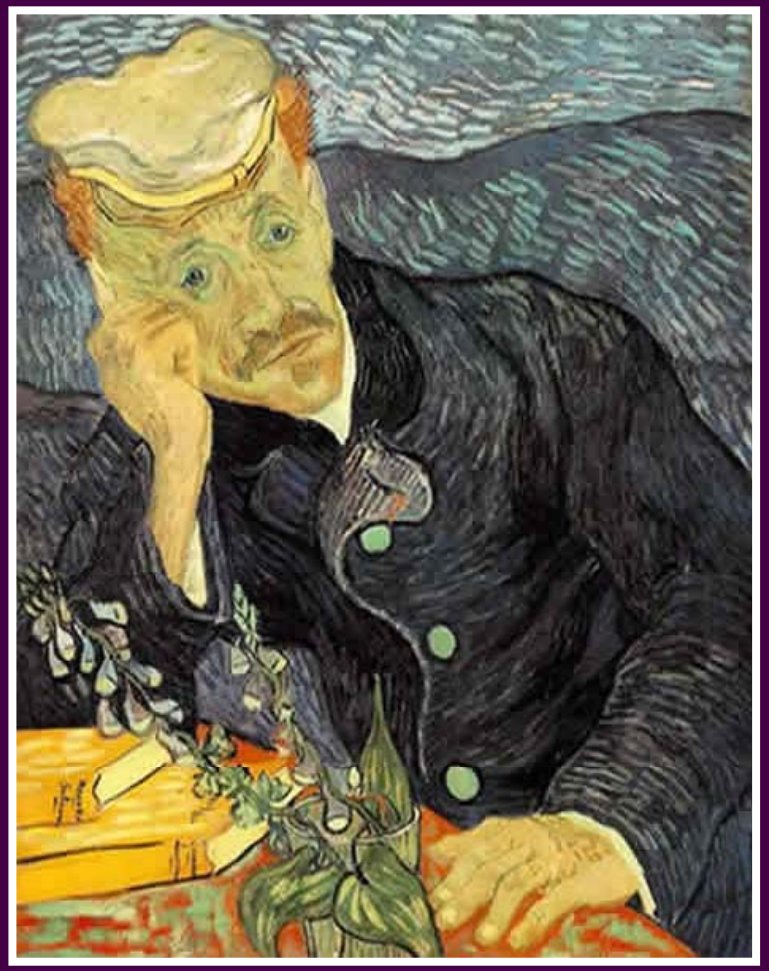
CASP8 Good prediction:



CASP8 Poor prediction:



M. Ben-David et al., *Proteins: Struc. Func. Gen., Suppl.* 9:50 (2009)



Vincent van Gogh: Portrait of Doctor Gachet, 1890

3. How do experimentalists and theoreticians view the field?

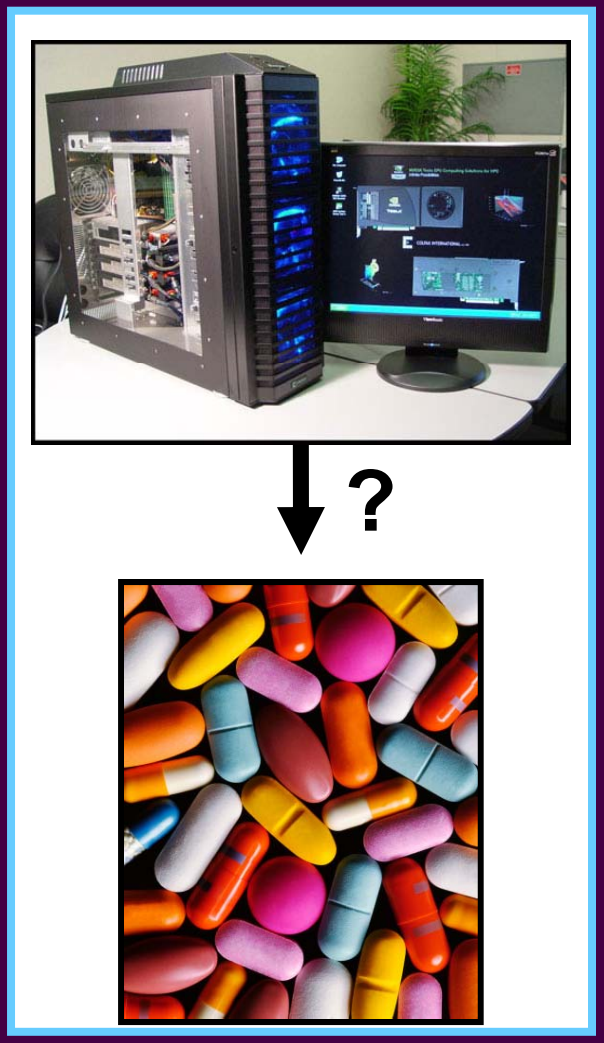
Assessment Questions (cont.)

4. How has computer technology impacted the field?
5. How important has the role of math / algorithms been in biomolecular modeling?
6. What are some of the scientific obstacles and how can they be overcome?
7. What are some of the cultural / infrastructural / educational obstacles and how can they be overcome?

Genetic Progr. Illustration, *Prof. Bull. US Army Logist.*, 37, 6 (2005)

Community Questionnaire

1. Describe your background
2. Has the field fulfilled initial expectations?
3. Is the field a complement to experiment or a field in its own right?
4. Describe success stories
5. Describe failure examples
6. What future applications are most promising?
7. What field advances are required?

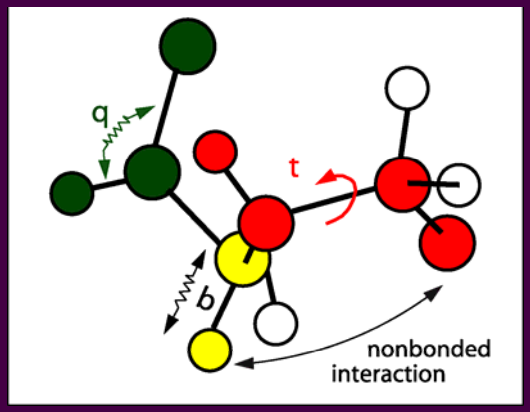


Molecular Modeling

Historical Perspective

1960s

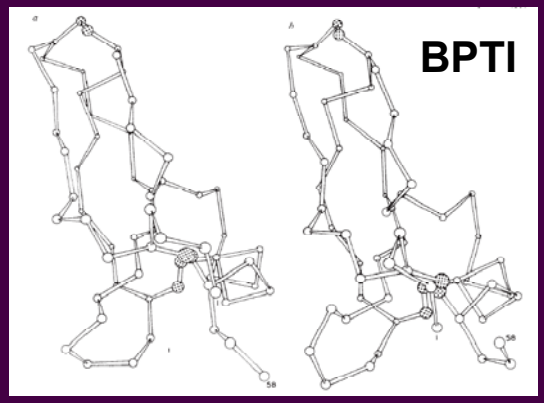
- Pioneers Lifson, Scheraga, Allinger, Levitt, Warshel, others began systematic force field (FF) development



T. Schlick., Molecular Modeling and Simulation, Springer-Verlag (2010)

1970s

- Rahman and Stillinger first report MD of liquid water; MD of BPTI followed
- Water FF developed (Berendsen, Jorgensen, etc.)
- Protein electrostatics concepts developed (Warshel, others)
- Pharma. applications by Kollman and others



McCammon et al., Nature, 267:585 (1977)

1980s

- MD simulations sprouted widely with advent of supercomputers

High Field Expectations

- 1986, Henry Schaefer (quantum chemistry pioneer):

"It is clear that theoretical chemistry has entered a new stage . . . with the goal of being no less than full partner with experiment." *Science*, 231:1100 (1986)

- 1998, Andrew Pollack (following advances in high-throughput technology):

"In a marriage of biotech and high tech, computers are beginning to transform the way drugs are developed, from the earliest stage of drug discovery to the late stage of testing the drugs in people." *The New York Times*, 10 November 1998

- 1998, The Economist editorial (following 1998 Chemistry Nobel Prize for quantum chemistry):

"In the real world, this could eventually mean that most chemical experiments are conducted inside the silicon of chips instead of in the glassware of laboratories. Turn off that Bunsen burner; it will not be wanted these ten years." *The Economist*, p. 57, 15 October 1998

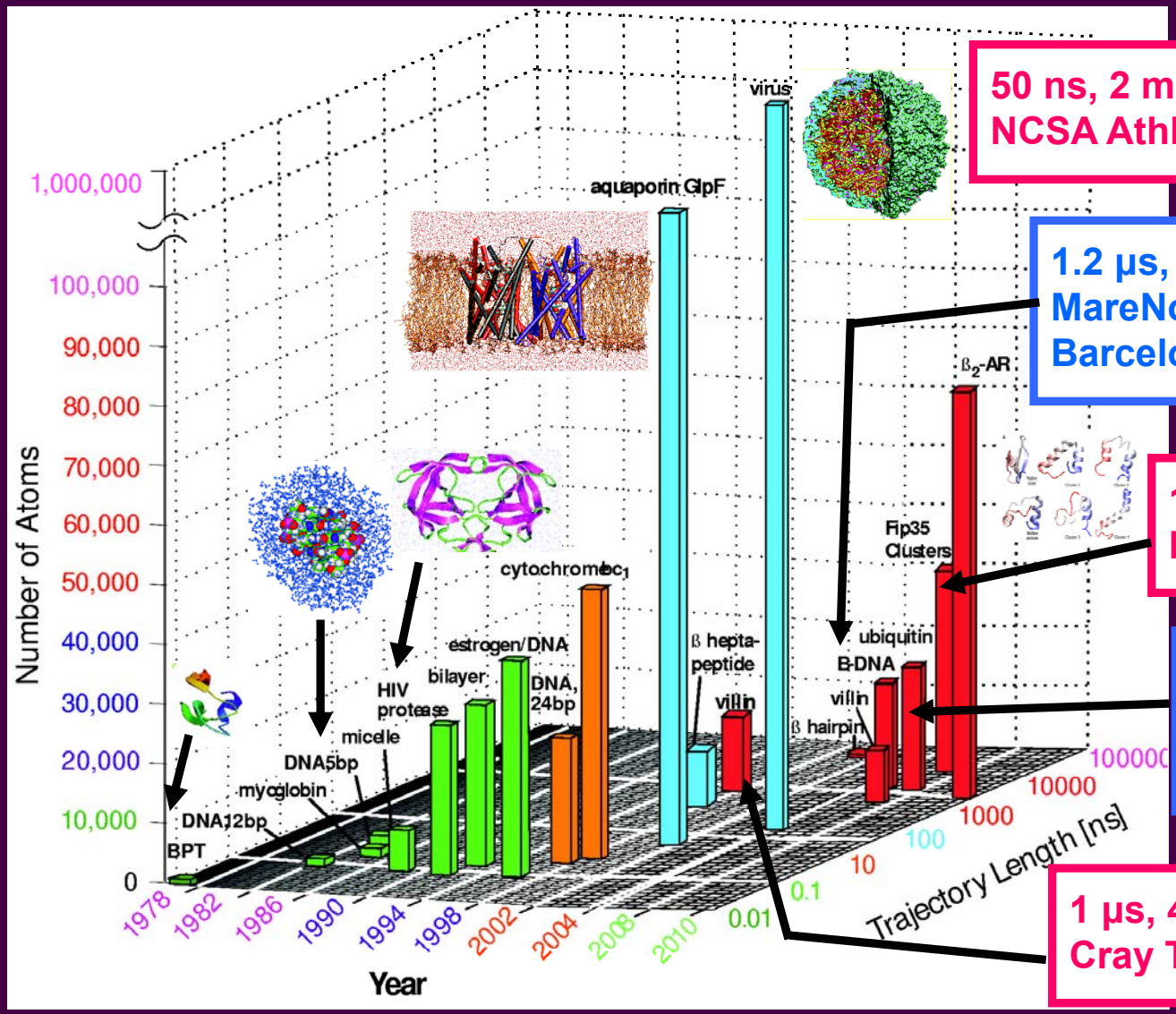
Molecular Modeling (cont.)

Historical Perspective

Today's simulations could almost be described by a

program design!

Evolution of MD (System Sizes and Simulation Lengths)



50 ns, 2 months, NCSA Athlon 2600+

1.2 µs, 20 wks, MareNostrum Supercomputer, Barcelona

10 µs, 14 wks, NAMD, NCSA Abe Cluster

1.2 µs, 2 wks, Operon Cluster Desmond program

1 µs, 4 months dedicated Cray T3D/E 256 processors

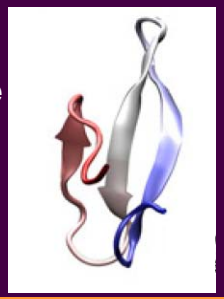
- MD simulations of proteins, nucleic acids, viruses and membranes help relate

  structure ⟷ flexibility ⟷ function
- Many approximations and limitations remain
- Adequate sampling of the vast configurational space remains a challenge, which led to numerous innovative ideas

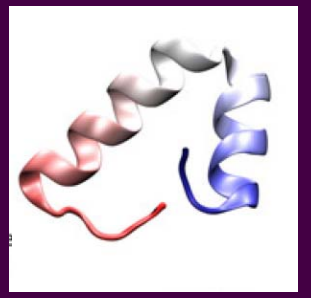
Long Simulations Raise More Questions Than Answers!

- Is computer power alone sufficient to observe folding?
- Are force fields sufficiently accurate?

Native structure



MD observed helical structure



P.L. Freddolino et al., *Biophys. J.* 94:L75 (2008)

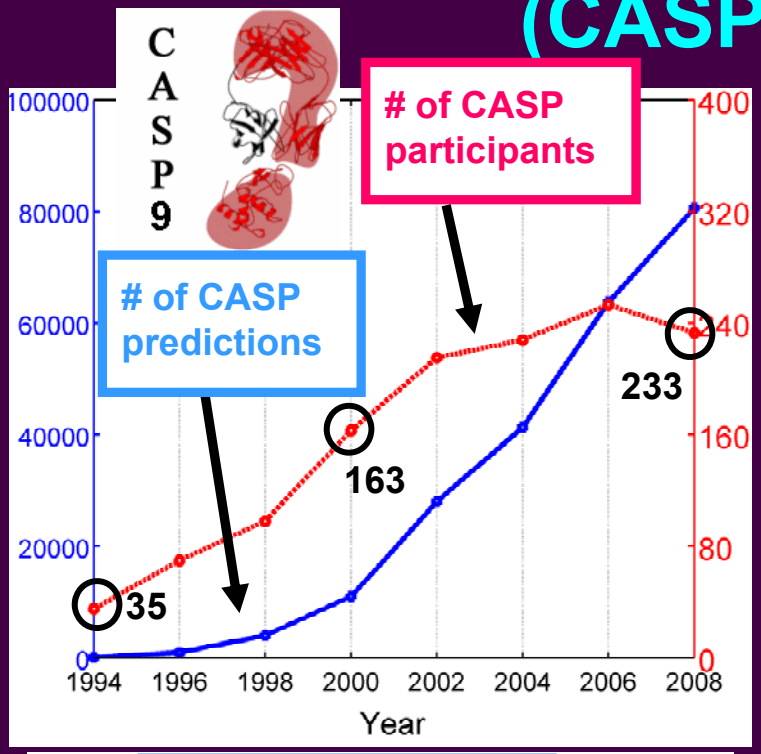
- Are algorithms sufficiently stable?

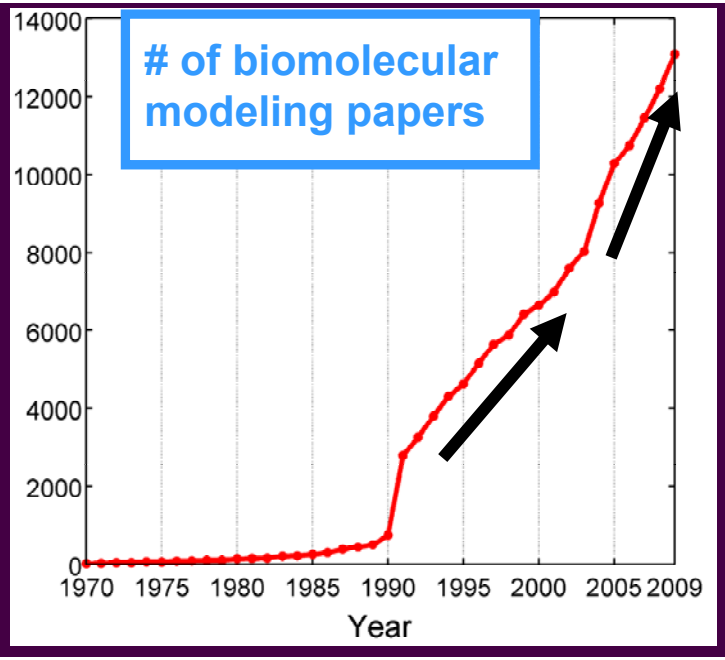
- Anton, a hard-wired supercomputer tailored to perform atomistic biomolecular simulations, was launched in Jan. 09 and will be available to the scientific community in late 2010 at PSC

"We are hopeful that Anton will ultimately allow scientists at D.E. Shaw Research and elsewhere to observe…structural changes…that have not previously been accessible to either computational or experimental study"

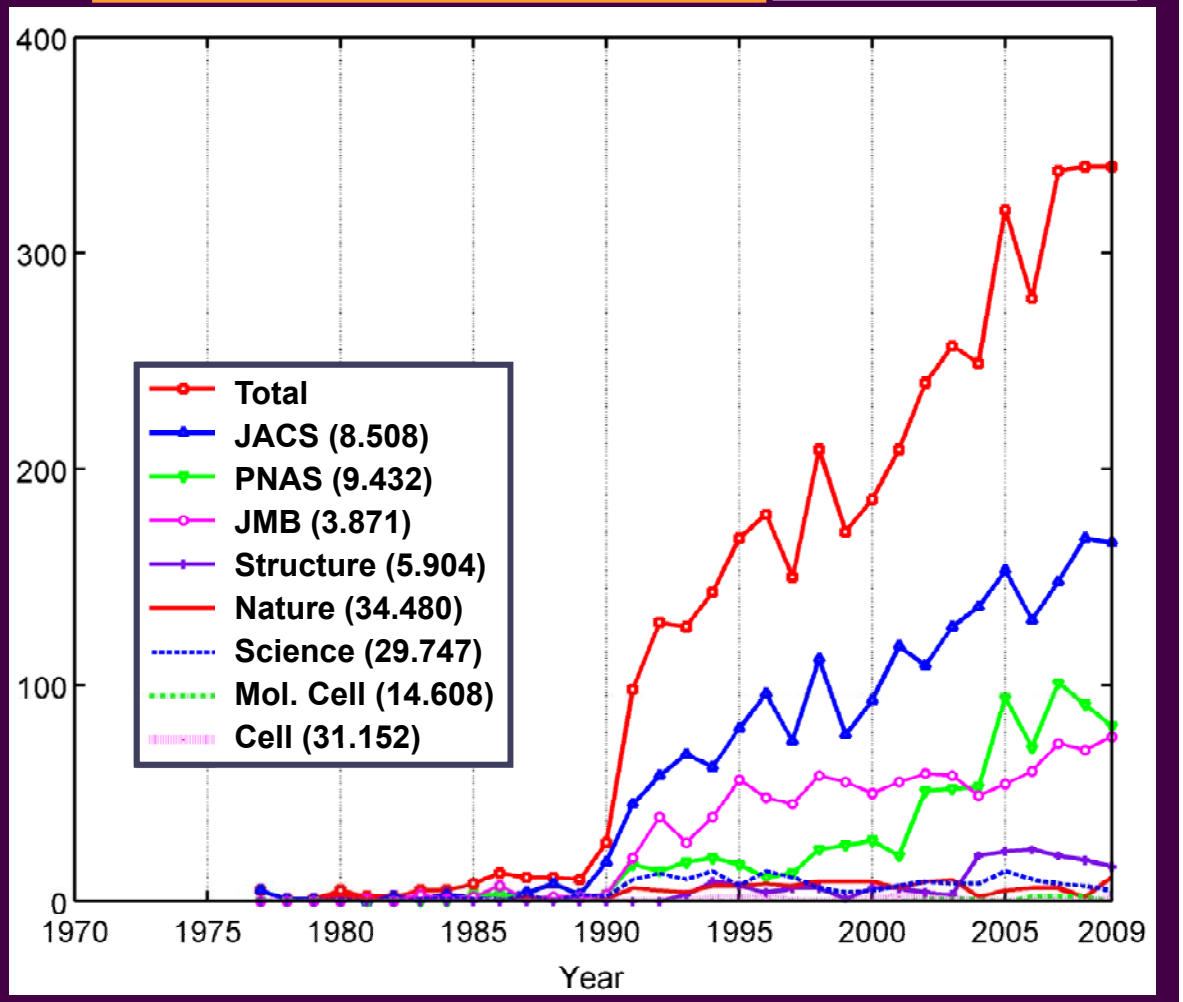
Metrics of the Field's Rise in Popularity

(CASP + Publications)





# of papers in top journal

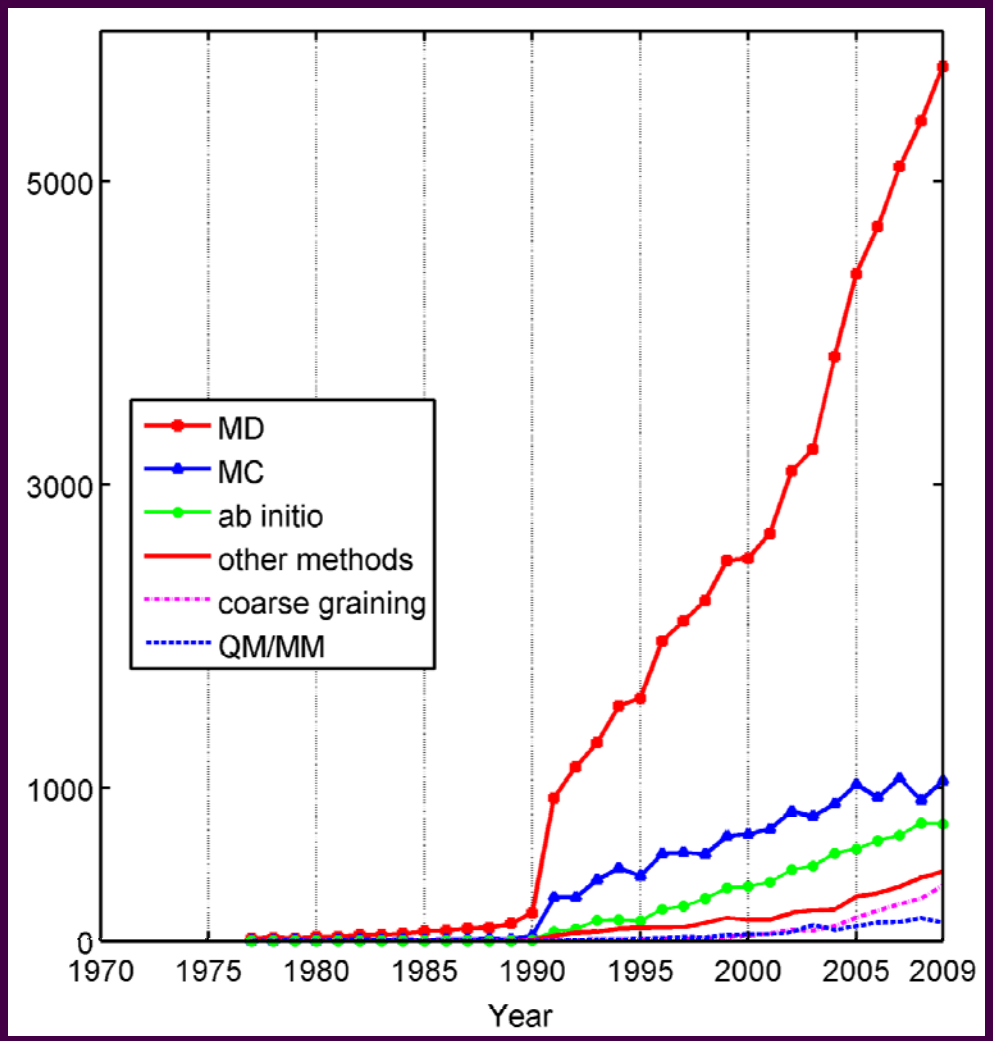


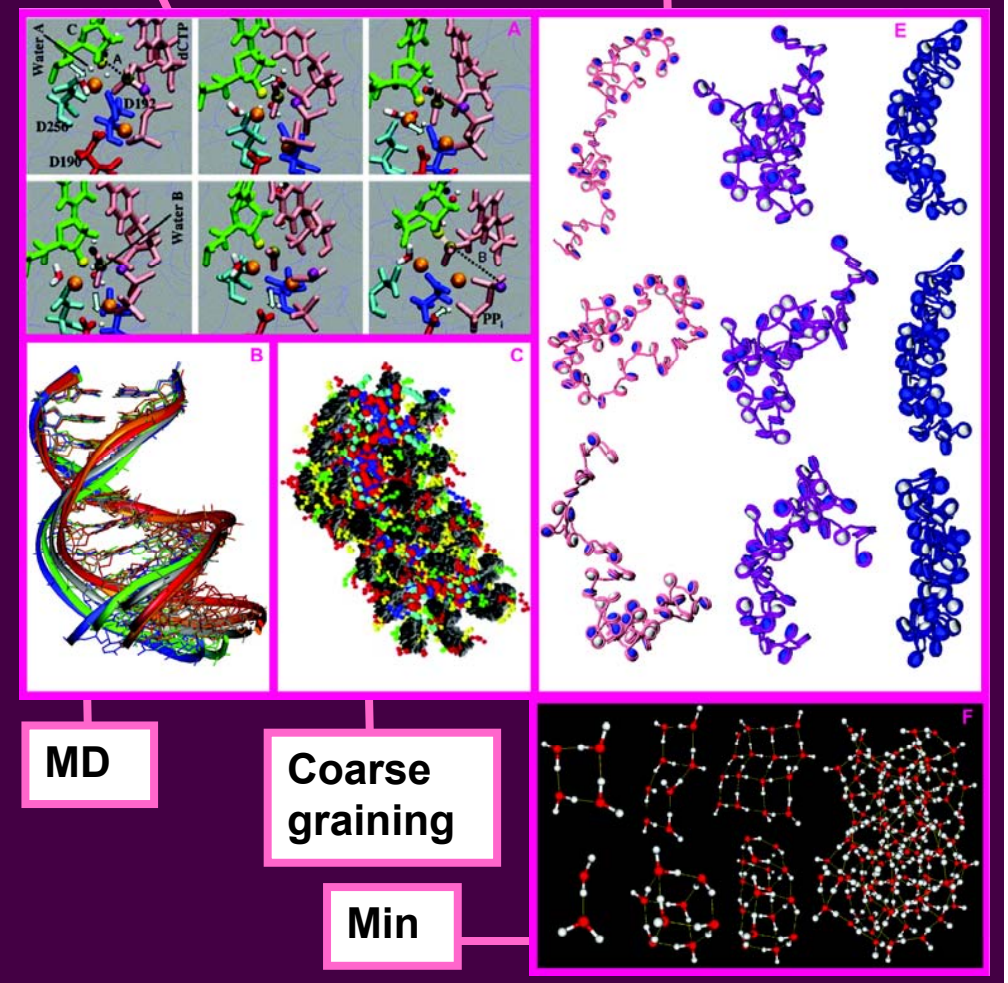
Schlick *et al.*, Biomolecular Modeling and Simulation: A Field Coming of Age (2010)

Metrics of the Field's Rise in Popularity

(MD Popularity)

# of publications per method

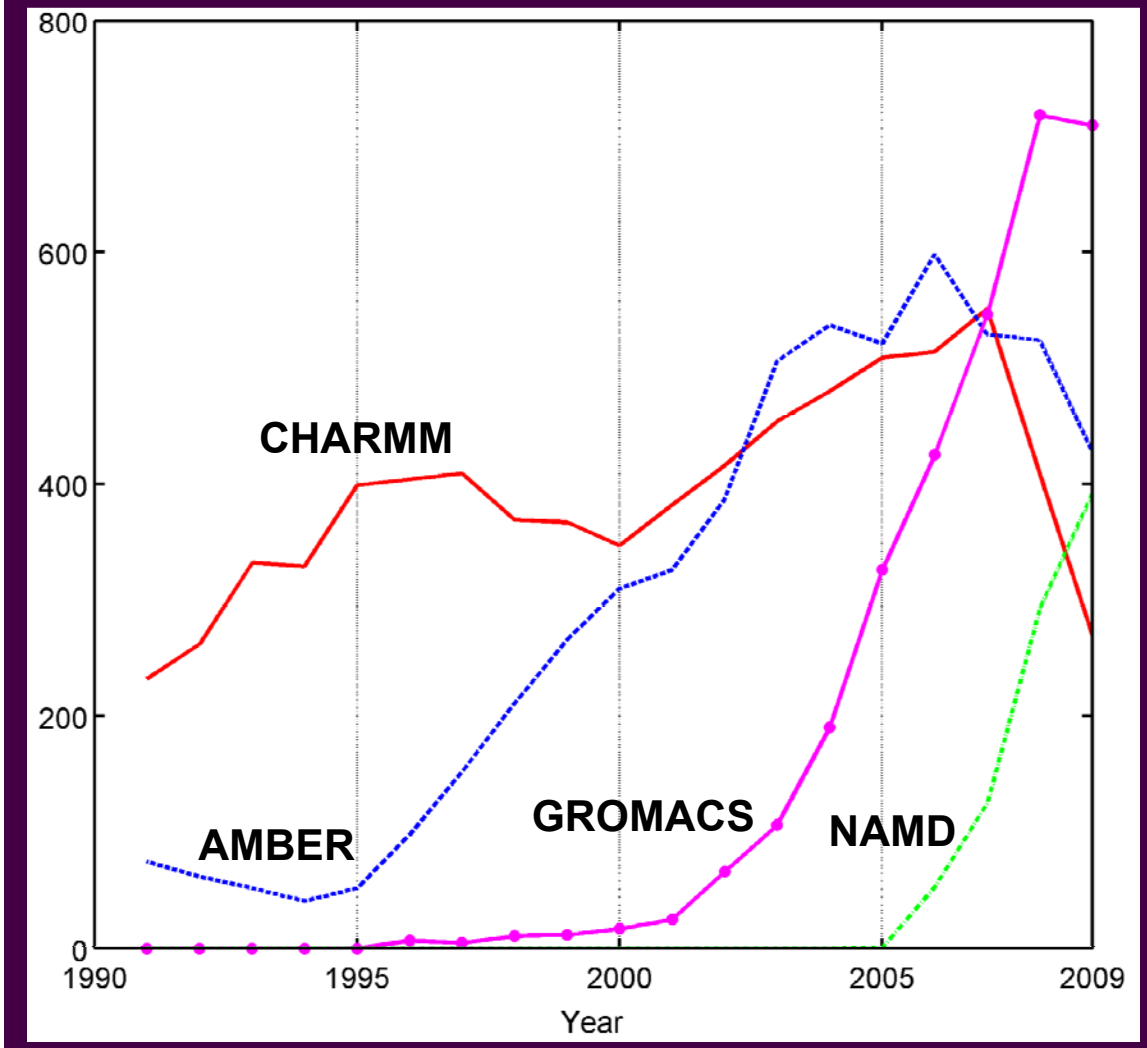




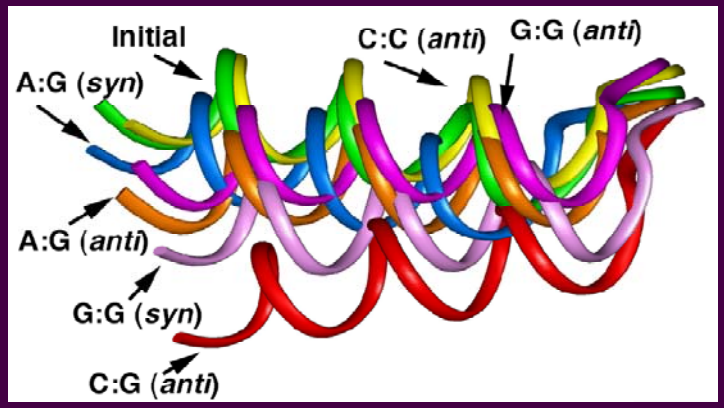
Schlick *et al.*, Biomolecular Modeling and Simulation: A Field Coming of Age (2010)

Metrics of the Field's Rise in Popularity
(Impact of Open Source Environment)

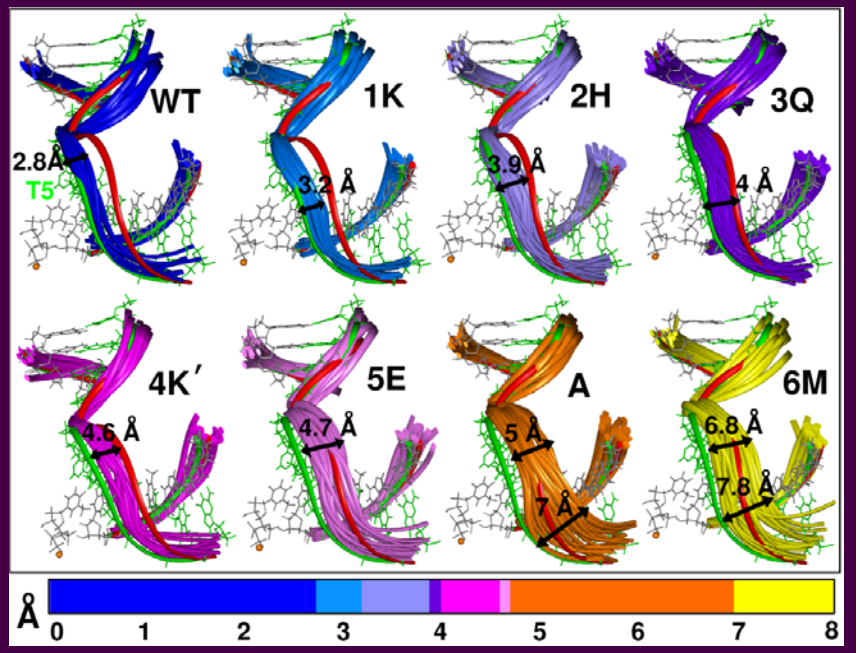
# of MD publications per package



Schlick *et al.*, Biomolecular Modeling and Simulation: A Field Coming of Age (2010)

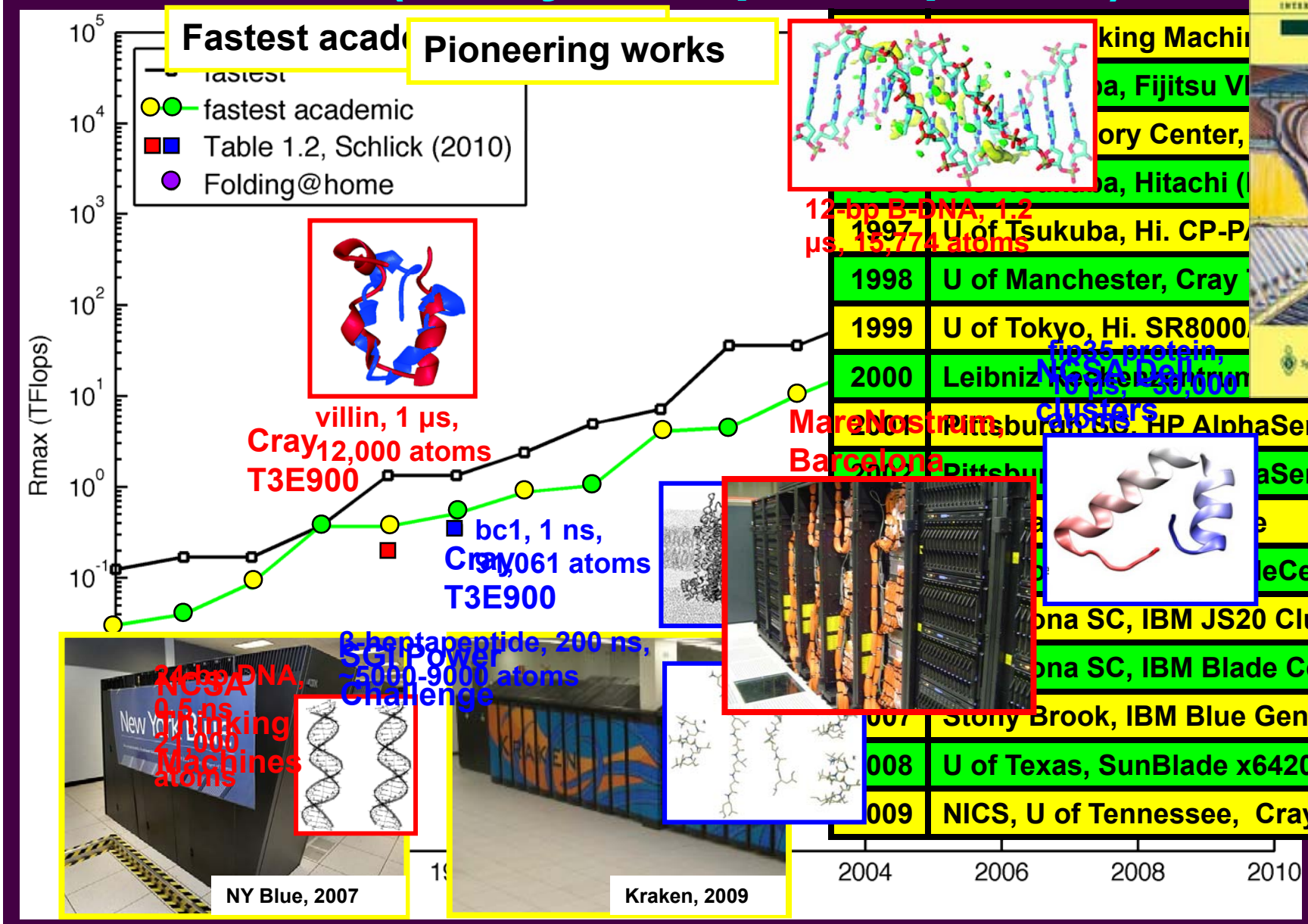


Sampoli Benitez et al., *Biophys. J.* 90: 42 (2006)

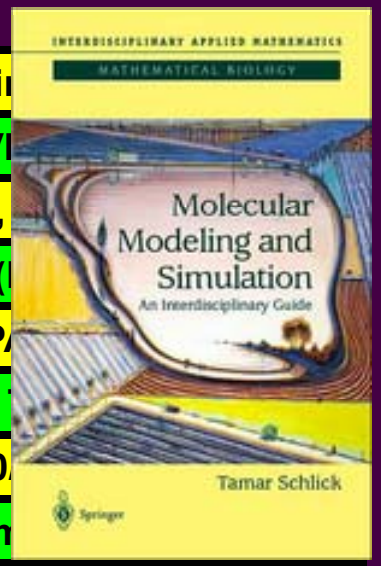


Foley & Schlick, *JACS* 130: 16500 (2008)

Metrics of the Field's Rise in Popularity
(Utility of Top Computers)



	king Machi
	a, Fijitsu V
	ory Center,
	a, Hitachi (
1997	U of Tsukuba, Hi. CP-P
1998	U of Manchester, Cray
1999	U of Tokyo, Hi. SR8000
2000	Leibniz Rechenzentrum
2001	Pittsburgh SC, HP AlphaServer SC45
	aServer SC45
	e
	leCenter JS20
	ona SC, IBM JS20 Cluster
	ona SC, IBM Blade Center JS21
007	Stony Brook, IBM Blue Gene
008	U of Texas, SunBlade x6420
009	NICS, U of Tennessee, Cray XT5-HE



Schlick *et al.*, Biomolecular Modeling and Simulation: A Field Coming of Age (2010)
Top 500 list: http://www.top500.org

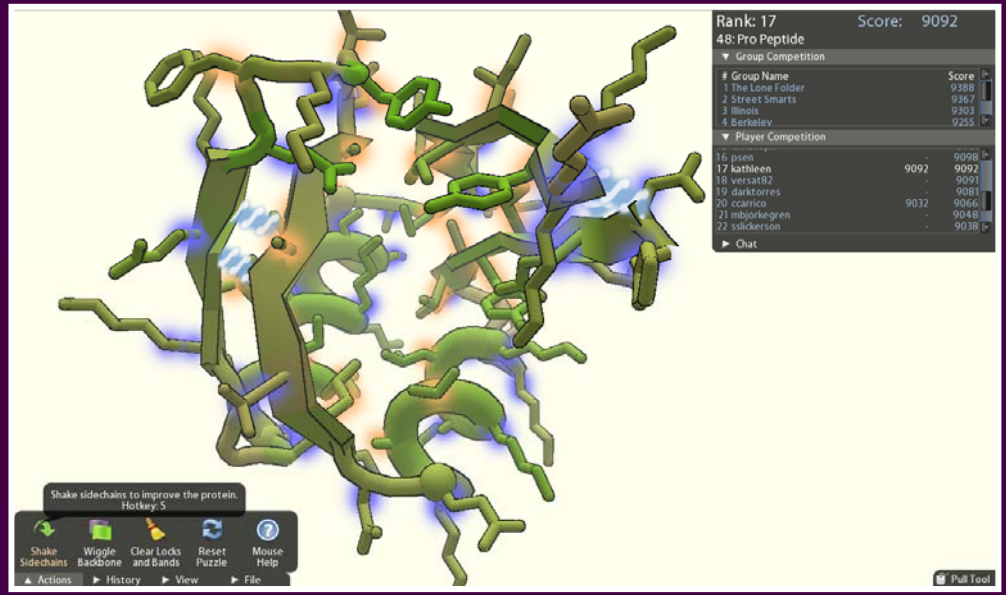
Computer Power Helped Popularize the Field

Karplus and Kuriyan in 2005, review on MD and protein function:

"The combination of increased computer power and improved potential functions has resulted in an ability to generate simulations that approach the point at which they can survive critical examination by the experimentalists who determine the structures of the proteins being simulated" Karplus & Kuriyan, *PNAS*, 102:6679 (2005)

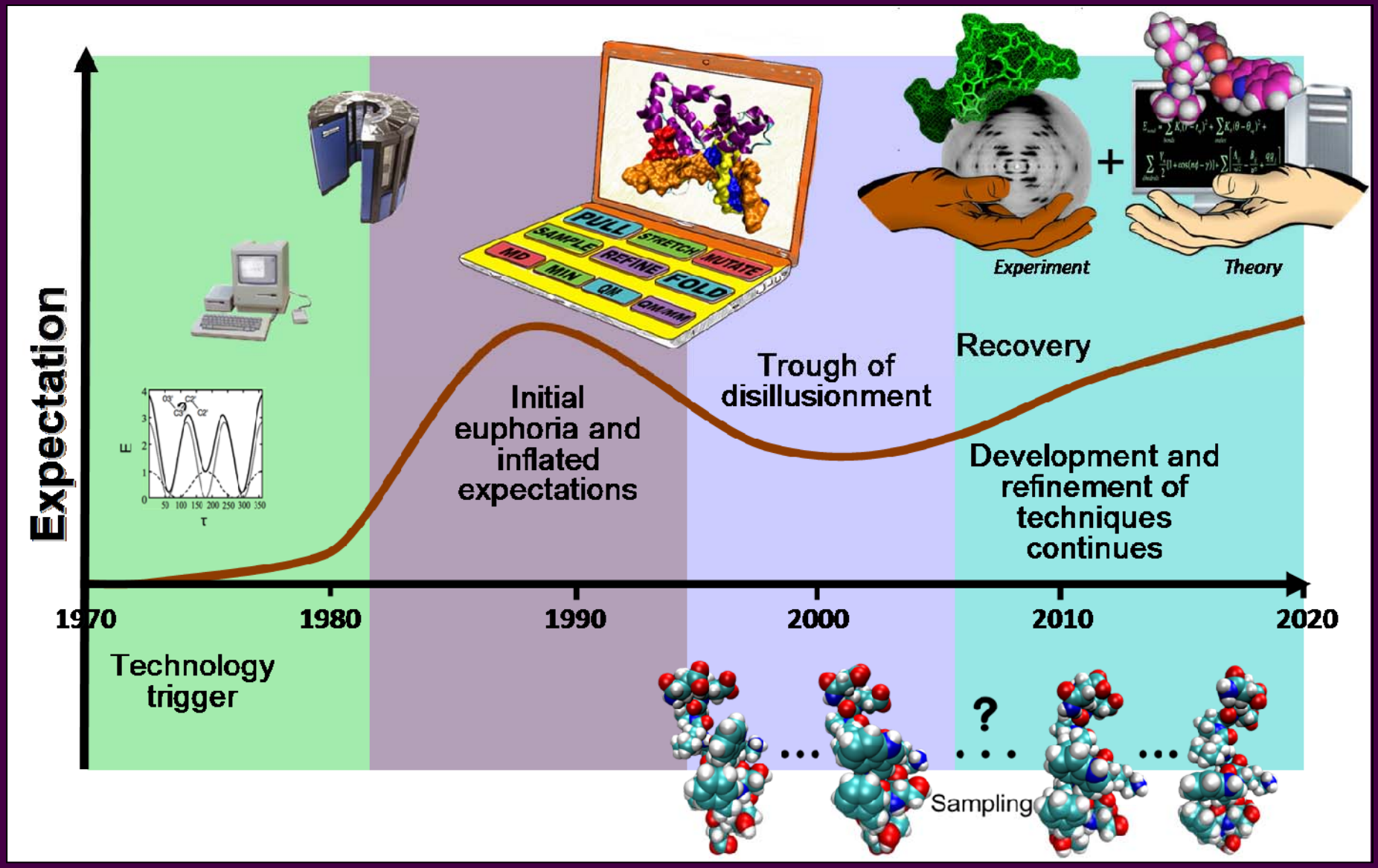
Protein folding is lucrative!

"Computer Game's High Score Could Earn The Nobel Prize in Medicine"
ScienceDaily, 9 May 2008



Field Expectation (or Hype) Curve

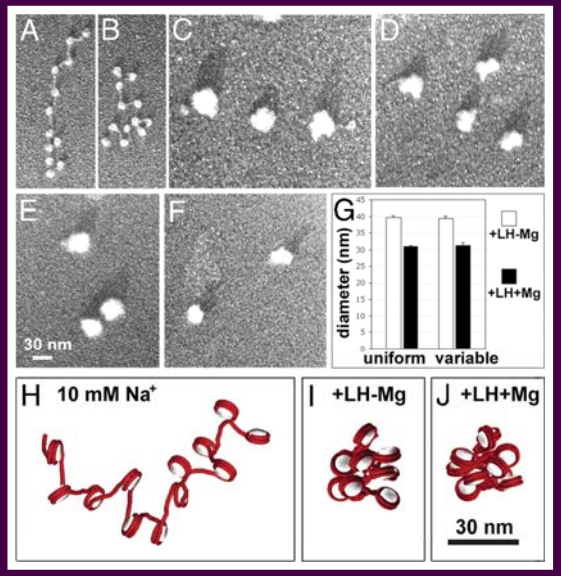
(idea of James Bezdek, founder of new journal on fuzzy mathematics in 1993)



Schlick *et al.*, Biomolecular Modeling and Simulation: A Field Coming of Age (2010)

Overall Community Views
(58 Responses)

- Mathematicians / computer scientists were more reserved / disappointed at the relatively modest impact of algorithmic advances
- All believed that modeling is a companion to experiment; half considered the field a discipline on its own right
- Most opined that field expectations were not met



Jacques Cohen (CS, Brandeis University):

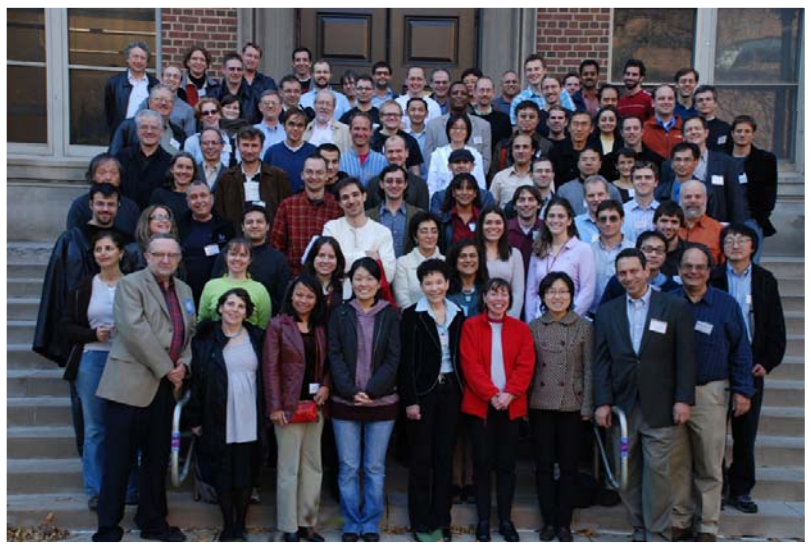
"These are very difficult modeling problems and nature has to be considered as the ultimate referee"

Stephen Neidle (Crystallography, U of London):

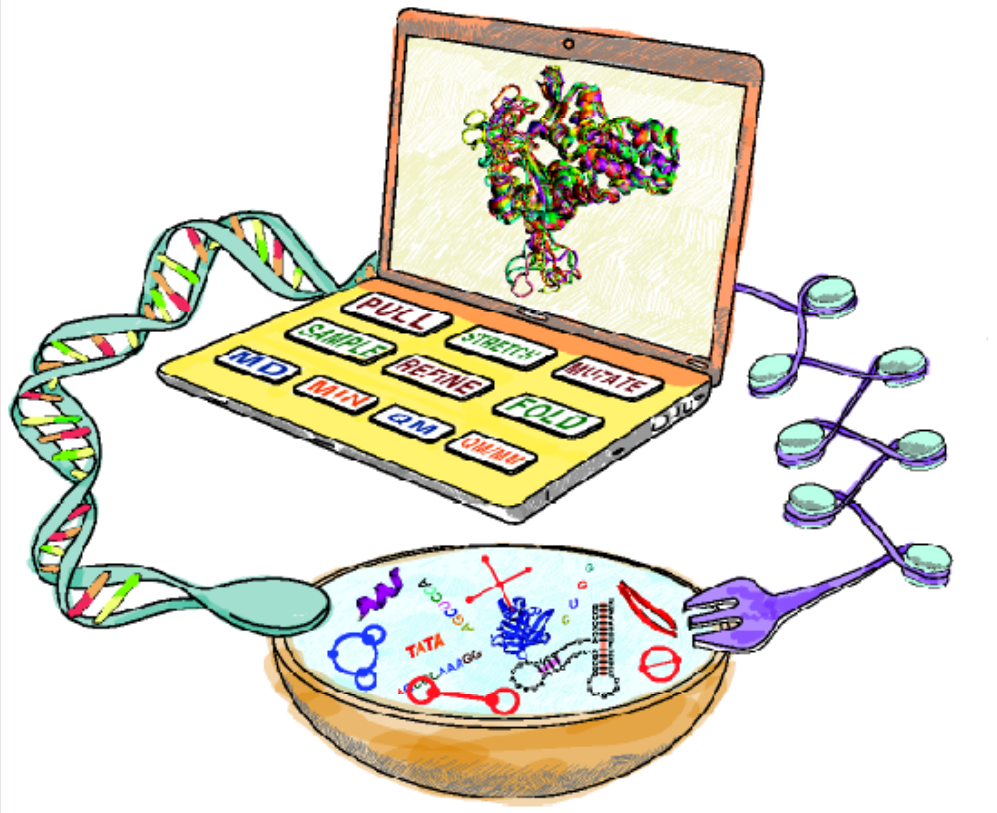
"My expectations have been conservative, and the reality has thus not disappointed me"

Overall Community Views (cont.)

- No single approach to drug design has succeeded
- MD has provided an invaluable tool to probe molecular structure and function (but no "BLACK BOX" exists!)
- Force Fields need most improvement
- More collaborations, rather than more funding, are needed to advance the field



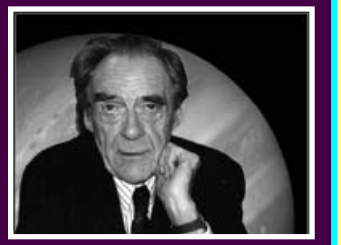
IMA RNA Bioinformatics Workshop (2008)



Modeling Successes!

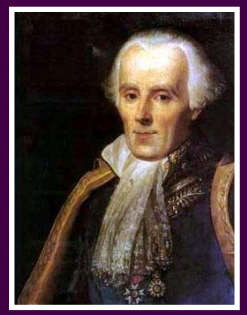
Prominent Example: MD
"Statistical Mechanics by Numbers"

- Sir John Maddox commented on MD of liquids and materials (1988):



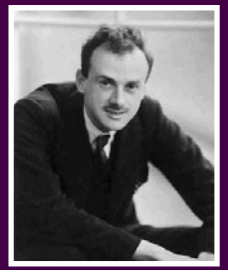
"Molecular dynamics is clearly well on the way to being a universal tool, as if it were the differential calculus. So much is plain from the range of problems now being tackled, and sometimes at least partially solved, by the application of this technique. Gone, it seems, are the days when people's ambitions in the field were restricted to the calculation of the properties of smallish molecules, with only distant hopes of being able to tackle the problems of, say, protein molecules. Now, molecular dynamics is being used to tackle problems which are quite general." Nature 334:561 (1988)

- MD is simply Pierre Simon de Laplace's vision on modern computers:



"An intelligence which could, at any moment, comprehend all the forces by which nature is animated and the respective position of the beings of which it is composed, and moreover, if this intelligence were far-reaching enough to subject these data to analysis, it would encompass in that formula both the movements of the largest bodies in the universe and those of the lightest atom: to it nothing would be uncertain, and the future, as well as the past, would be present to its eyes. The human mind offers us, in the perfection which it has given to astronomy, a faint sketch of this intelligence." Oevres Vol VII. Théorie Analytique de Probabilité, 1820

- And dazzling advances in computer technology have now made it possible:



"The fundamental laws necessary for the mathematical treatment of a large part of physics and the whole of chemistry are thus completely known, and the difficulty lies only in the fact that application of these laws leads to equations that are too complex to be solved." Paul Dirac (1929)

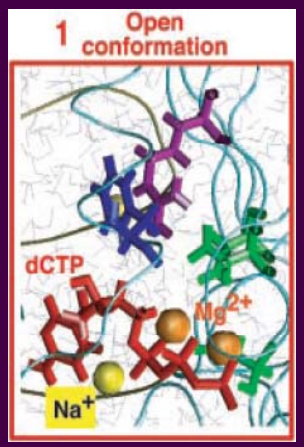
Modeling Successes (cont.)

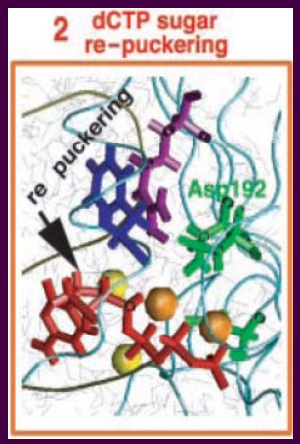
Molecular Dynamics - Current applications

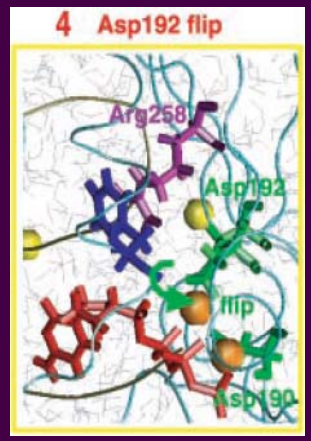
- Refine exp. structures
- Interpret exp. data (NMR relaxation, force / extension curves, improve structure-based function prediction)
- Link static structures to generate pathways
- Make structural prediction
- Resolve exp. ambiguities

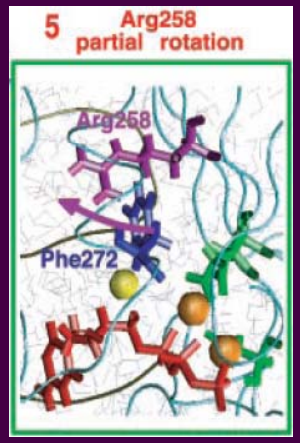
DNA polymerase β closing dynamics using deterministic path sampling approach

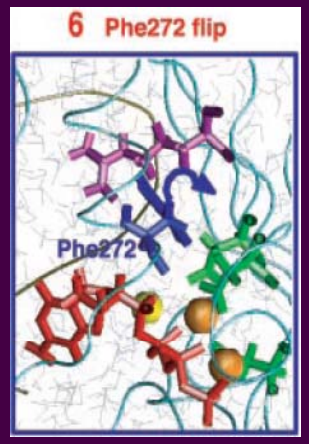
Ravi Radhakrishnan
Tamar Schlick

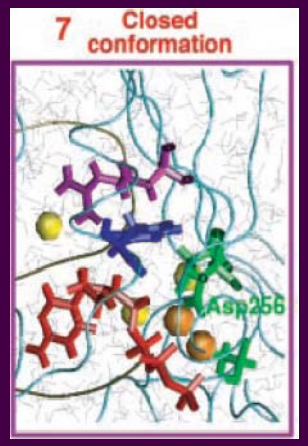












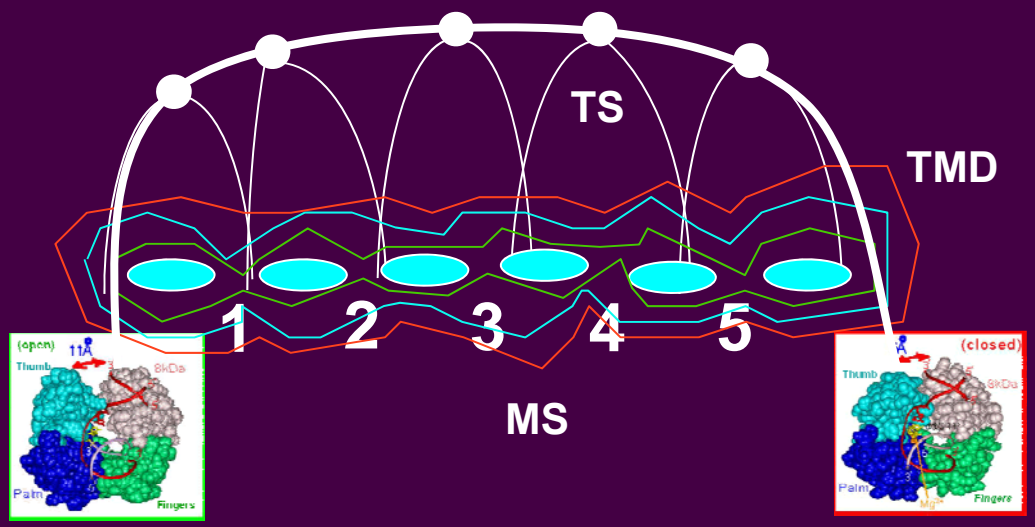
 F. Händel, Concerti Grossi, Opus 6, Nos. 11 & 12

Radhakrishnan & Schlick, *PNAS* 101: 5970 (2004)

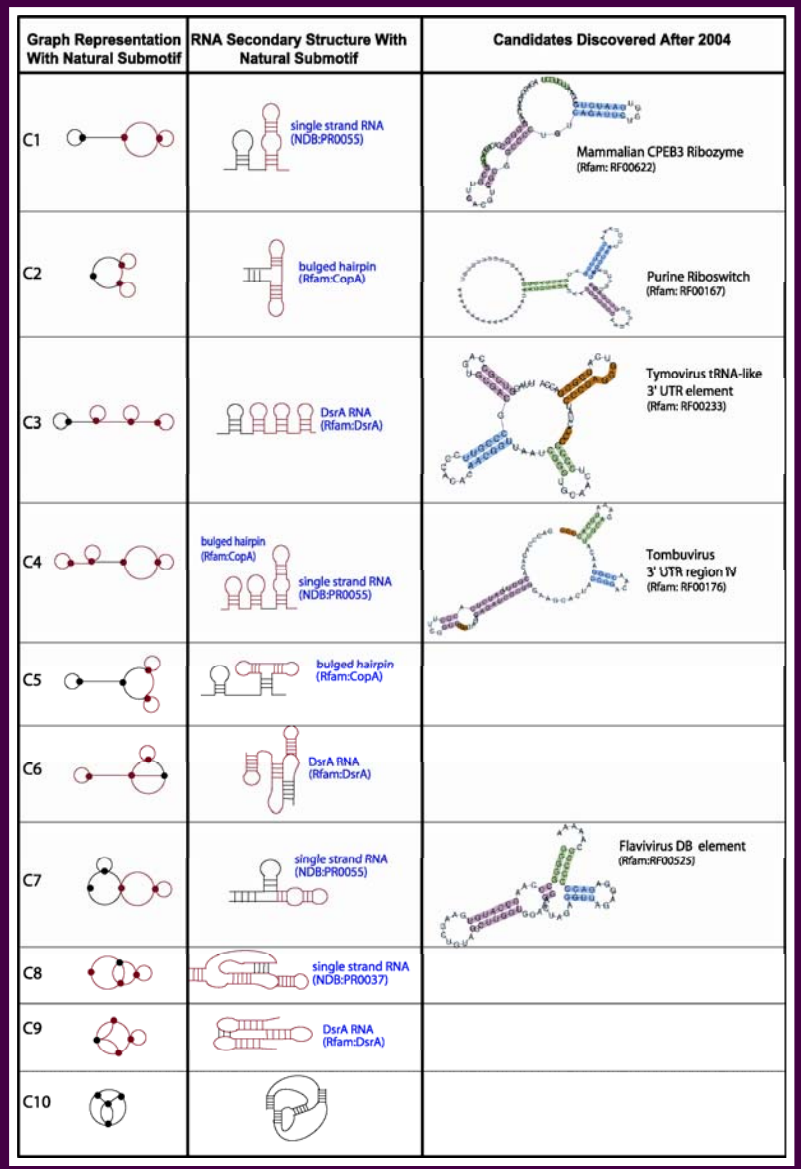
Other Modeling Successes

Establishment of Reliable Algorithms

- Symplectic integrators
- Resonance artifacts
- PME
- Enhanced sampling



- Graph theory for RNA for design applications

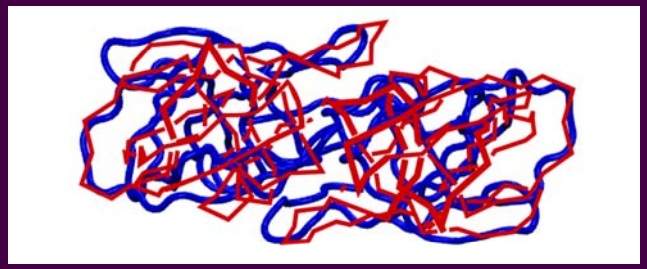


Kim et al., *J. Mol. Biol.*, 341:1129 (2004)

Modeling Successes (cont.)

Structure Prediction that Preceded Exp.

- HIV-1 Protease: Homodimer



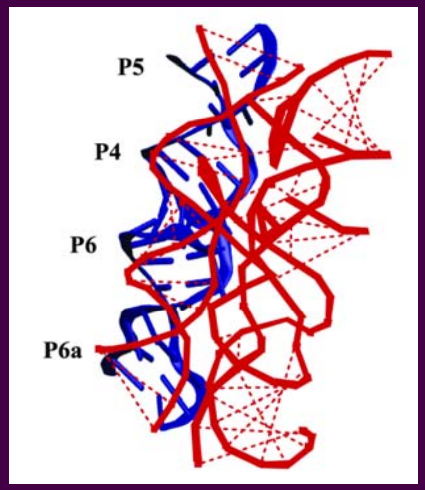
Crystal structure

Homology modeling

Mod.: L. H. Pearl and W. R. Taylor, *Nature*, 329:351 (1987)

Exp.: A. Wlodawer et al., *Science*, 245:616 (1989)

- Group-I intron: P4-P5 domain



Crystal structure

Comparative modeling

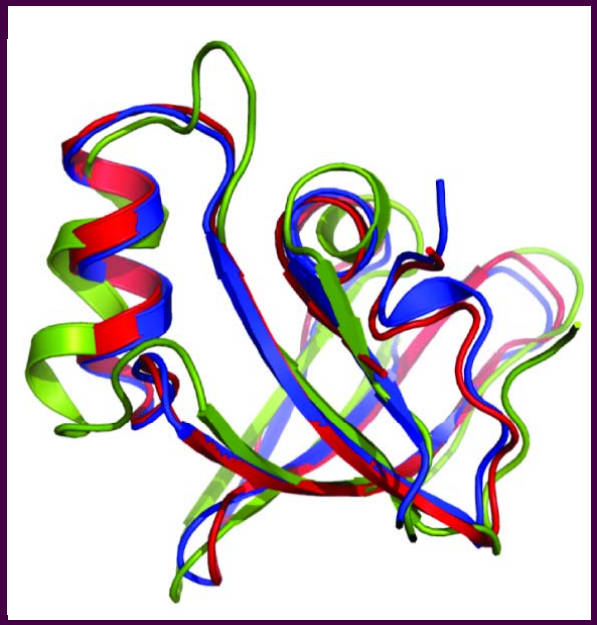
Mod.:F. Michel and E. Westhof, *J Mol. Bio.*, 216:585 (1990)

Exp.: J. H. Cate et al., *Science*, 273:1678 (1996)

- Peptide Structure Prediction (PEPstr, PePLook, Robetta)

- Protein Structure Prediction: CASP

Baker group prediction for target T0492 (all-β protein)



Crystal structure

Structure by comp. mod.

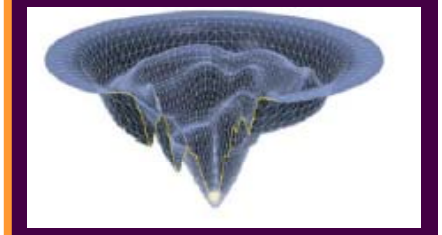
Prediction by Rosetta and comp. mod.

Mod.: S. Raman, *Proteins*, Suppl. 9:89 (2009)

Modeling Successes (cont.)

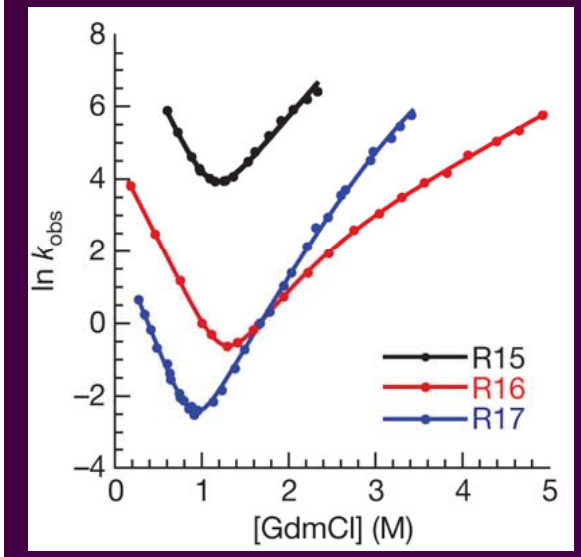
Protein Folding Theory

Kinetics theories of folding landscapes (Frauenfelder, Wolynes, Onuchic, Thirumalai, others) have provided framework for experimental analysis



P. G. Wolynes, *Phil. Trans. R. Soc. A*, 363:453 (2005)

Recent experiments confirmed slow kinetics ⇔ rugged landscapes for α-spectrin

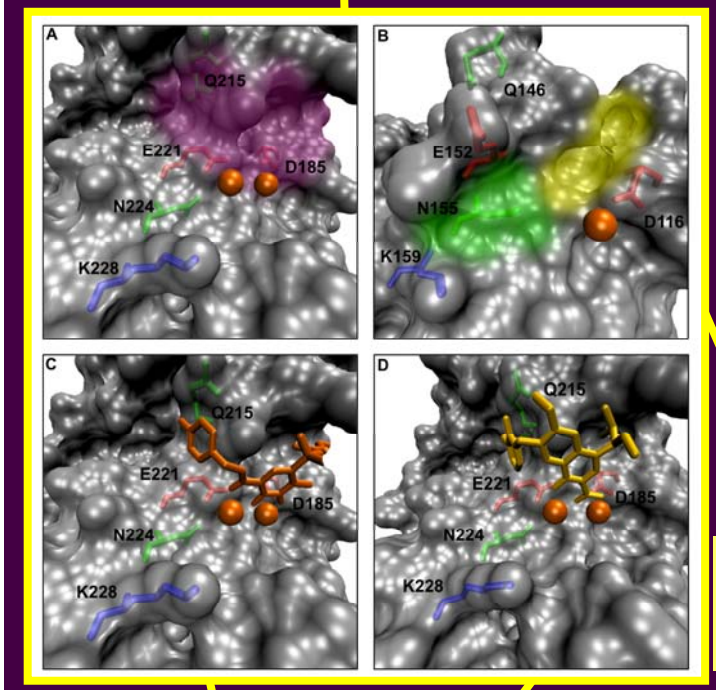


R16 and R17 fold / unfold much slower than R15

B.G. Wenseley et al., *Nature*, 463:685 (2010)

Mod.-aided Drug Design

Modeling helped outline drug binding modes and specificity



A. HIV integrase core domain with inhibitor (crystallography)

Y. Goldgur et al., *PNAS*, 96:13040 (1999)

B. Simulation predicted binding sites (preferred, flipped)

J. R. Schames et al., *J. Med. Chem.*, 47:1879 (2004)

C and D: Raltegravir and Elvitvavir drug complexes

S. Hare et al., *Nature*, 464:232 (2010)

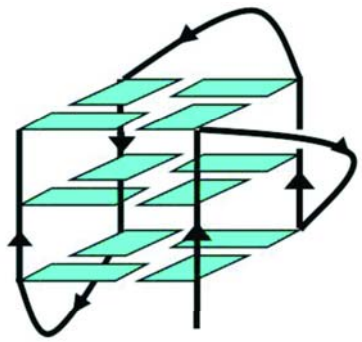
New experimental findings combined with modeling are expected to aid the development of antiretroviral drugs

Modeling Successes (cont.)

Interpretation of Experimental Data

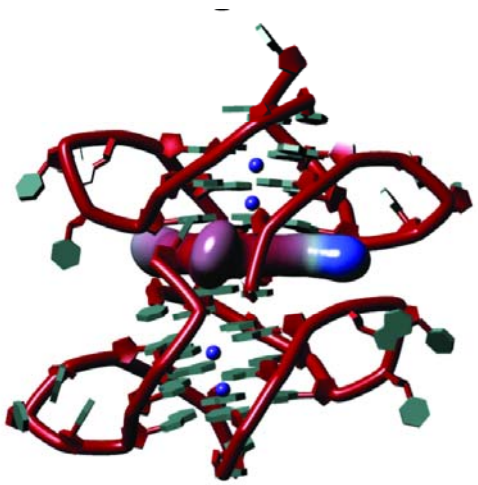
- G-Quadruplexes

Anti-par. (NMR)



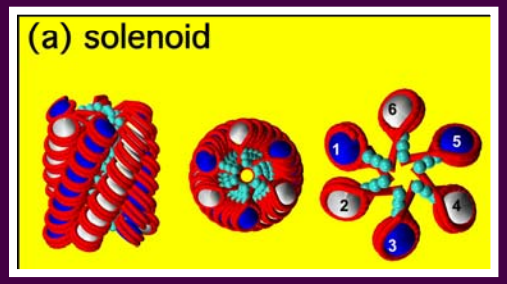
Parallel (cryst.)

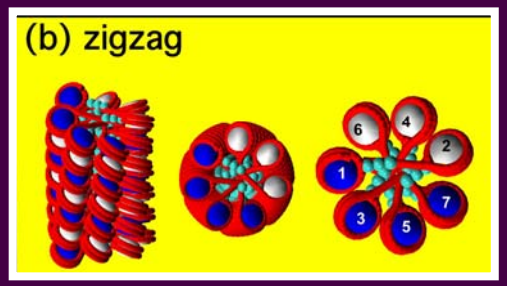
Drug candidate complexed with two human telomeric quadruplexes

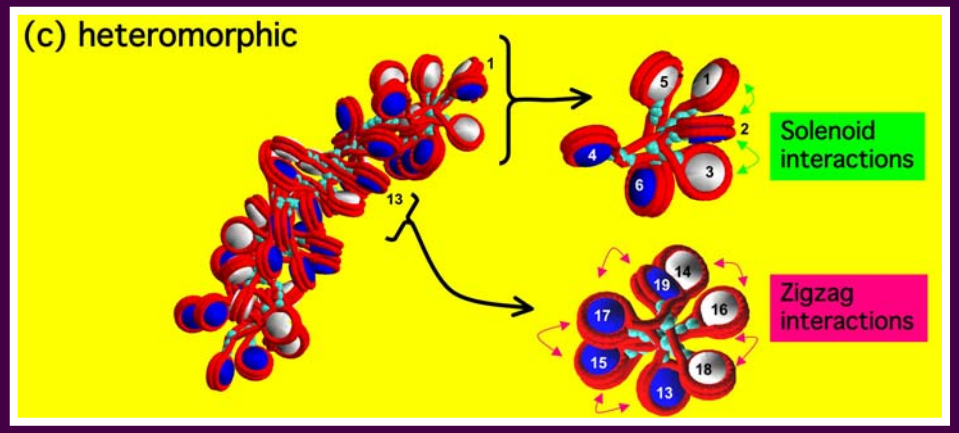


H. Campbell et al, JACS, 130:6722, (2008)

- Chromatin Organization







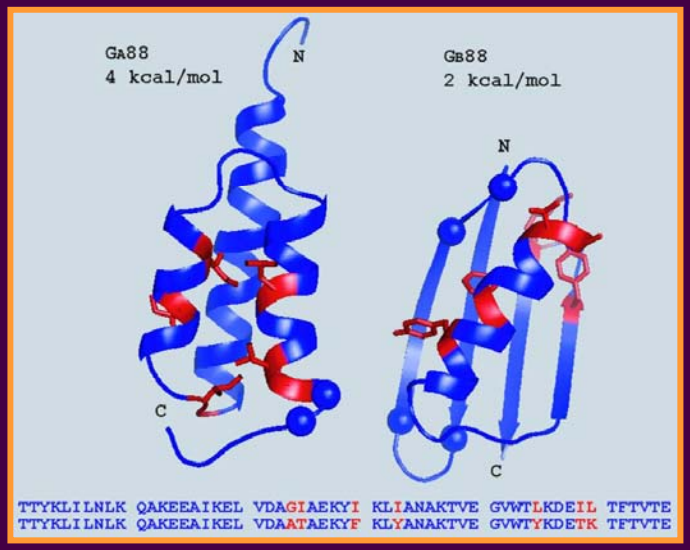
S. A. Grigoriev et al., *PNAS*, 106:13317, (2009)

Some Modeling Failures

- CASP failures in structure prediction (homology misleads, ab initio)

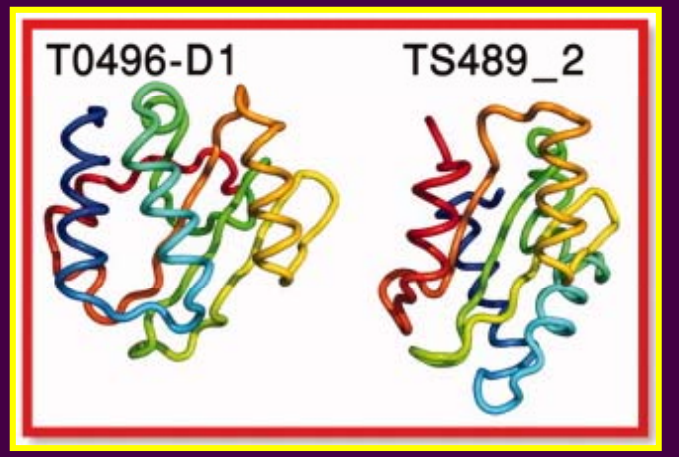
Failures for targets with ~90% sequence homology (10%: red)

P.A. Alexander et al., *PNAS*, 104:11963 (2007)

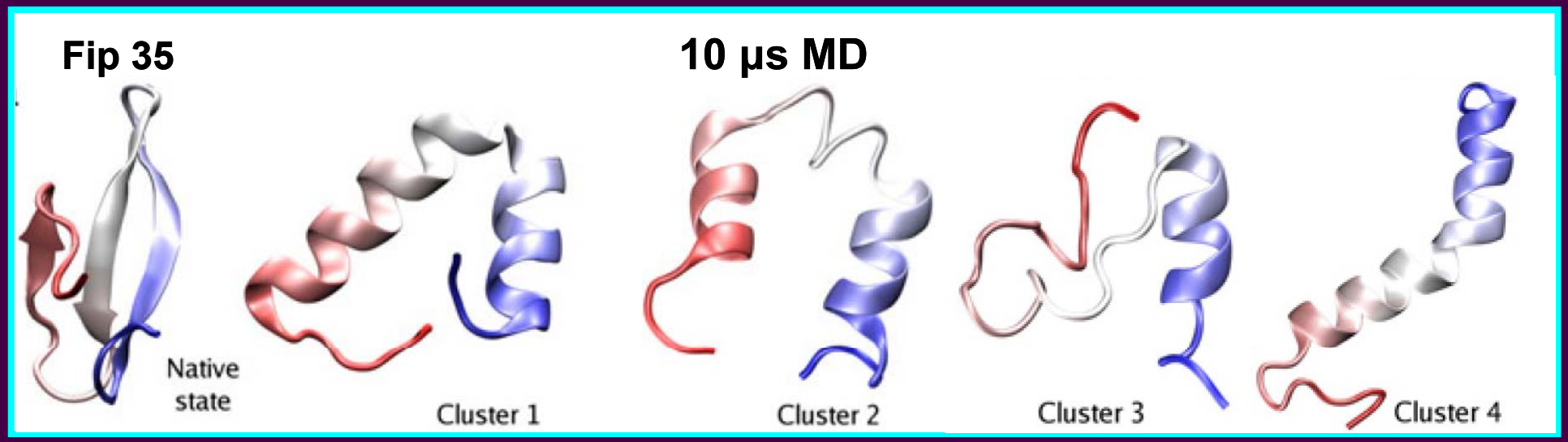


Failures for ab initio prediction

M. Ben-David et al., *Proteins: Struc. Func. Gen., Suppl.* 9:50 (2009)



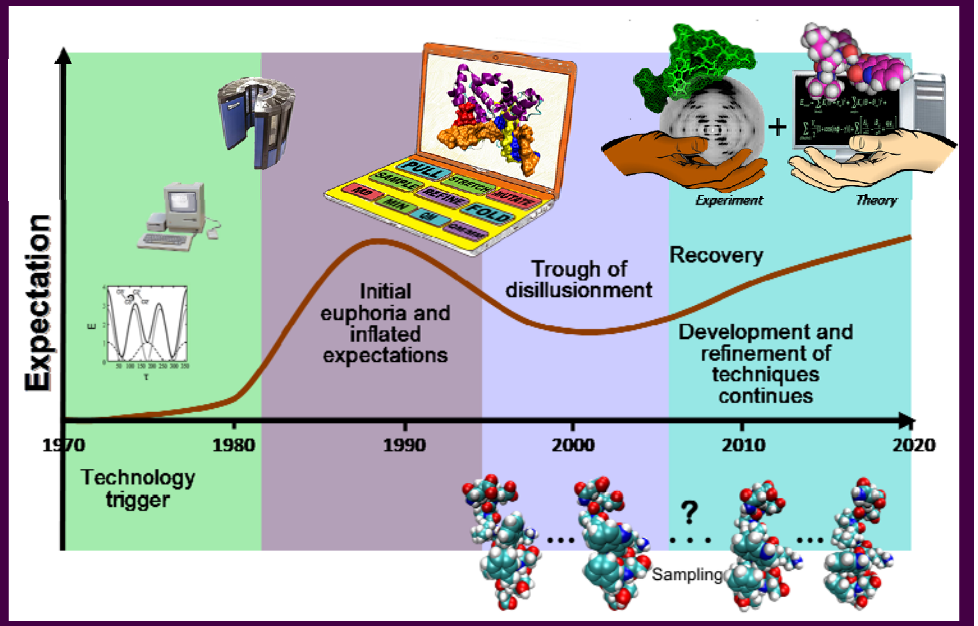
- Failures in configurational sampling
- Force field biases: overstabilization of α-helices



P.L. Freddolino et al., *Biophys. J.* 94:L75 (2008)

Conclusions

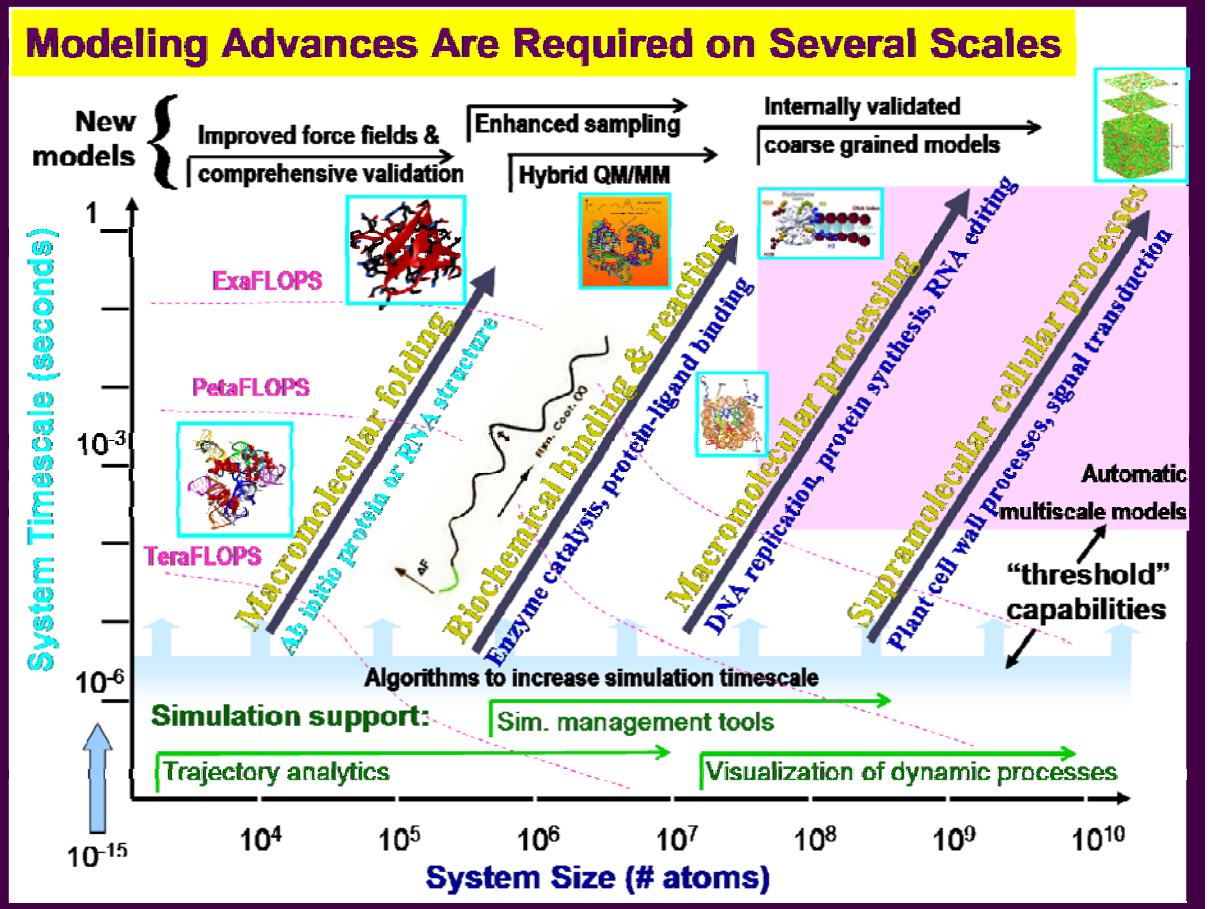
- Biomolecular modeling and simulation is a vibrant field with many proponents and numerous exciting areas of activity
- Recent doubts have been stirred by realization of:
  - General biological complexity
  - Limited ability of fast computers to bridge theory and experiment
  - Hype and expectations were unrealistically high; but a healthy productivity level is setting in



- These issues together with unprecedented opportunities in 21st century biology argue for focused activities

Recommendations

1. Bridging Scales
2. Methodology Advances
3. Force Fields
4. Data Sharing
5. Better Interdisciplinary Education
6. Shared Resources
7. Collaborations
8. Faster Processors
9. More Community Efforts
10. Better Communication between Academe and Industry
11. Reports of Failure



Opportunities in biology at the extreme scale of computing, DOE, Aug. 2009
http://www.er.doe.gov/ascr/ProgramDocuments/Docs/BiologyReport.pdf

Acknowledgments

- Karunesh Arora
- Gaurav Arya
- Rosana Collepardo
- Shereef Elmetwaly
- Meredith Foley
- Hin Hark Gan
- Leif Halvorsen
- Abdul Iqbal
- Joe Izzo
- Segun Jung
- Namhee Kim
- Christian Laing
- Yunlang Li
- Sean McGuffee
- Victoria Padow
- Ognjen Perišić
- Giulio Quarta
- Ravi Radhakrishnan
- Benedetta Sampoli Benitez
- Yanli Wang
- Yurong Xin
- Linjing Yang

- NIH
- NSF
- Philip-Morris
- NYU CTSI

- Bill Beard
- Sergei Grigoryev
- Tom Kunkel
- Sam Wilson
- Ada Yonath