Next-Generation Simulations in Biology:
Investigating biomolecular structure, dynamics, and function through multiscale modeling

Presented by

Pratul K. Agarwal

Computer Science Research Group
Computer Science and Mathematics Division

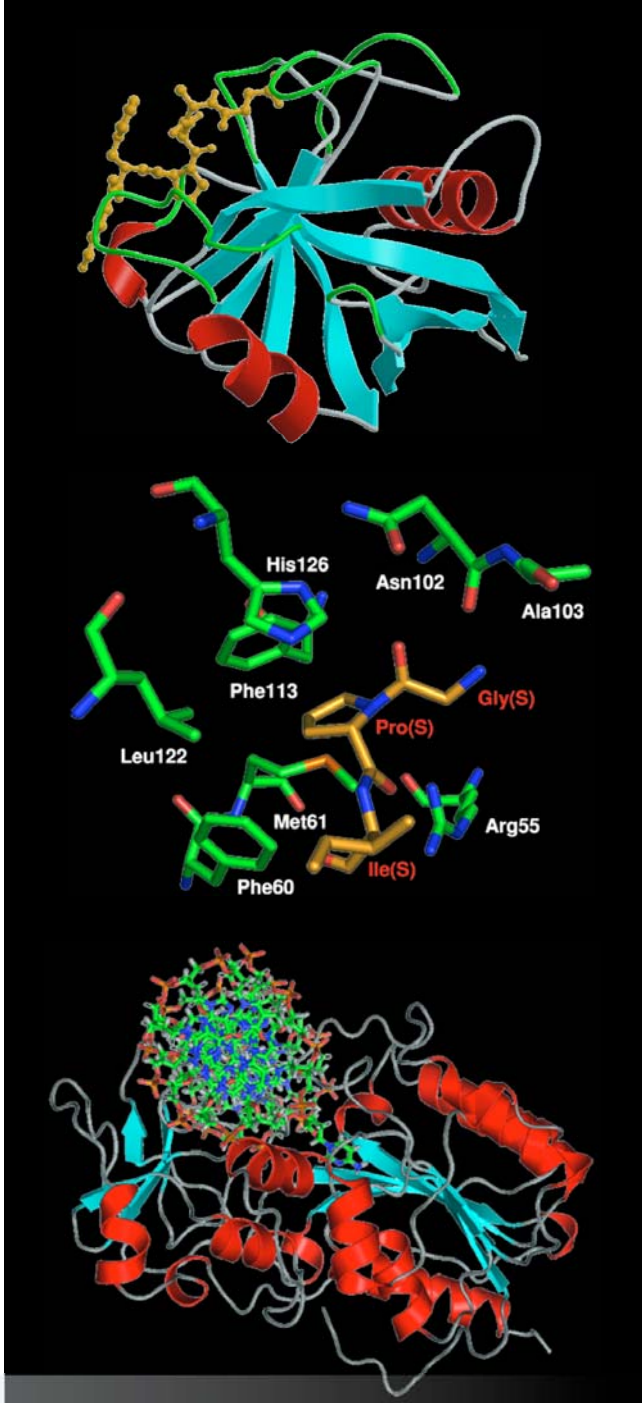
SC07

OAK RIDGE National Laboratory

Why study biomolecules?

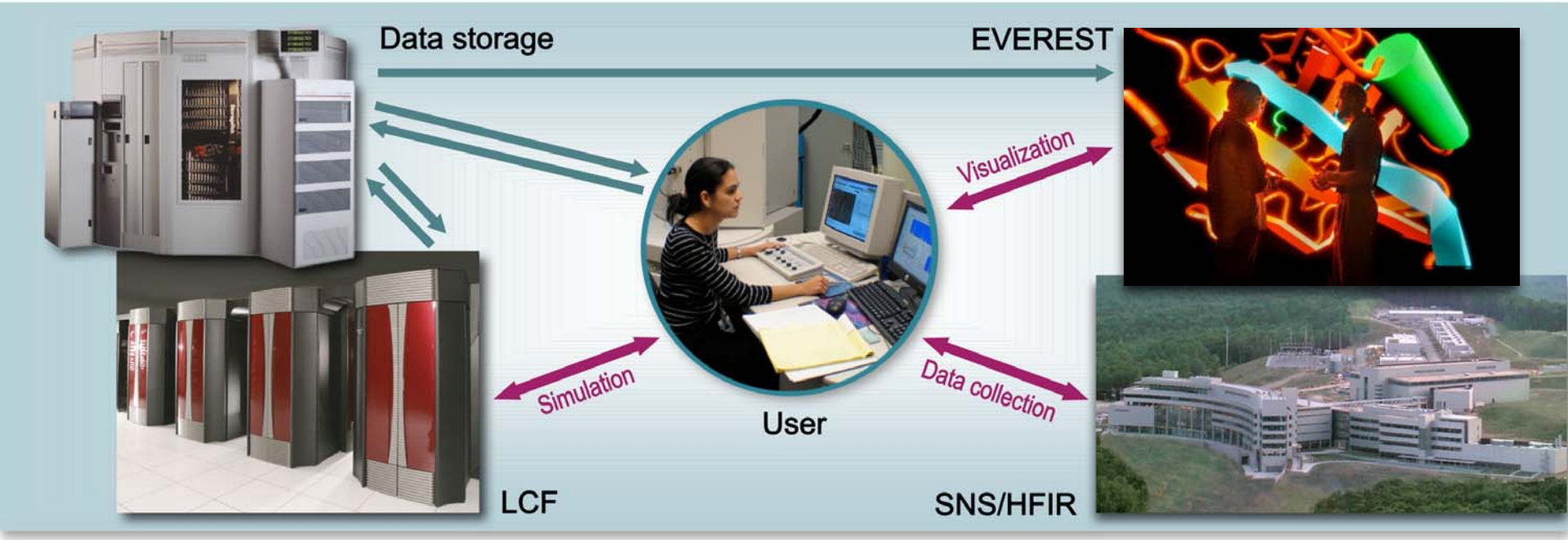
Biomolecules including proteins, enzymes and DNA are highly efficient molecular machines working in all cells

- Atomistic-level biomolecular modeling and simulations
  - Provides detailed biophysical characterization of molecular machines, protein structure prediction, and docking
  - Power of computational method: novel insights that are beyond the reach of experimental techniques
  - Enables multiscale approach: investigate a wide range of time and length scales

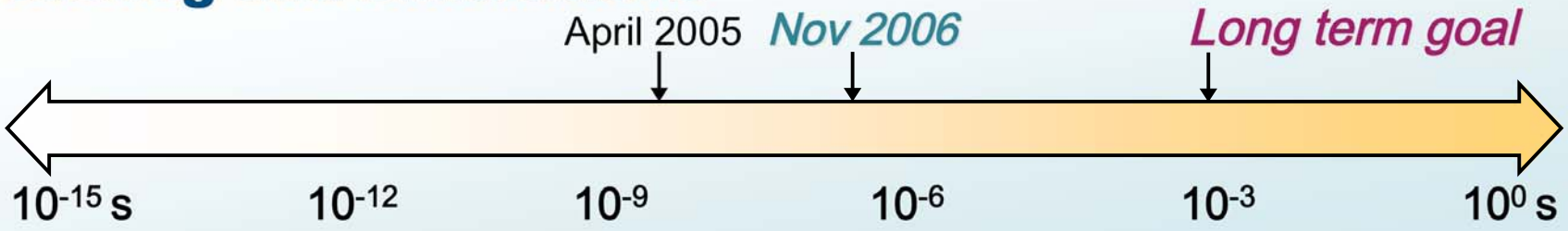


Multidisciplinary approach for multiscale problem

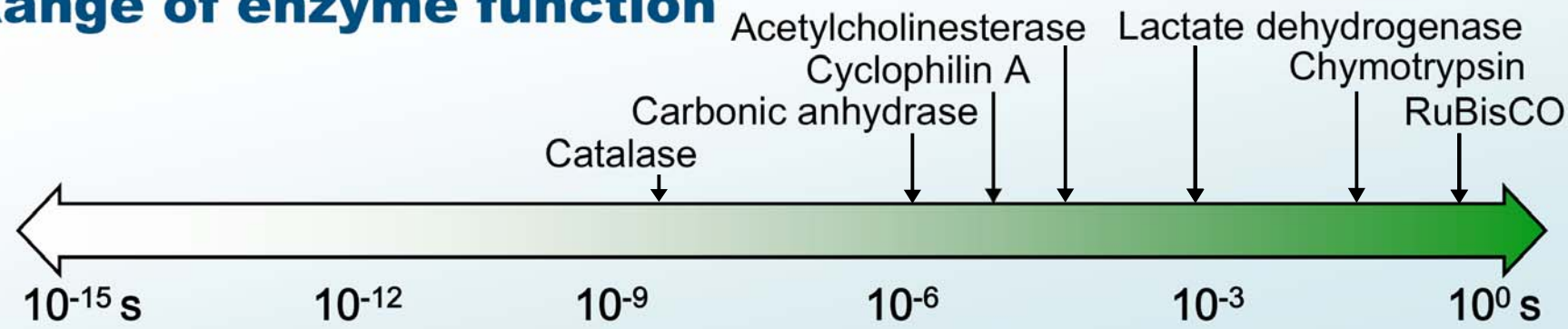
- Biomolecular structure spans multiple scales of length, and dynamics and function spans multiple scales of time.
- Multidisciplinary approach combining experimental and computational techniques is being developed at ORNL.
- Petascale computing at the LCF and neutron scattering at the SNS will provide novel insights into biomolecules.



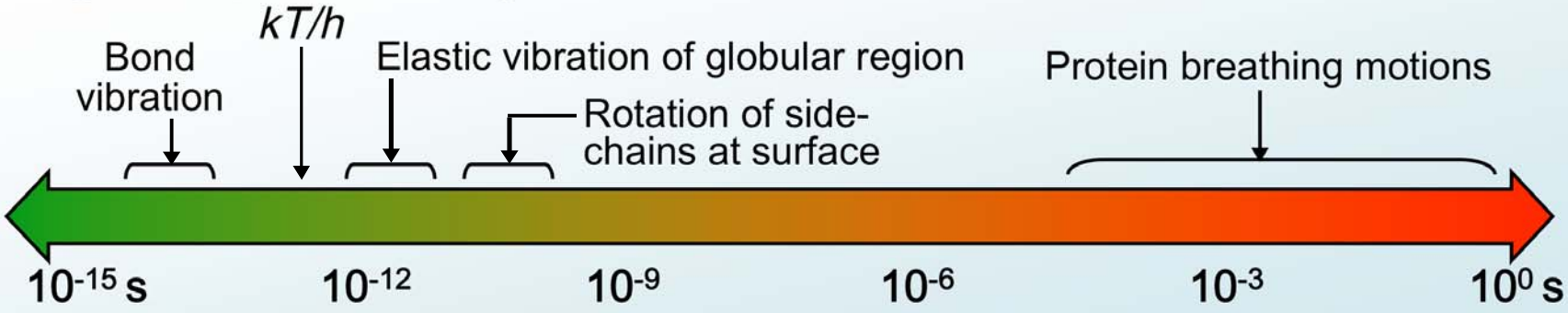
Modeling and simulations



Range of enzyme function



Range of protein dynamical events



Using supercomputers to study biomolecules

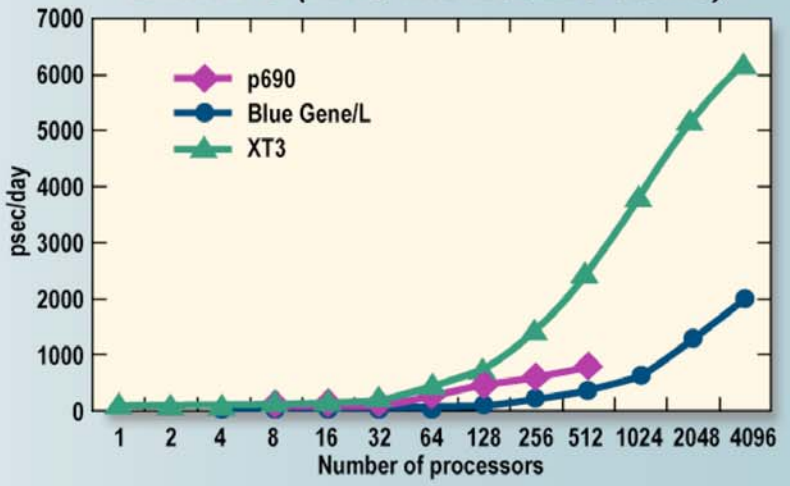
- Desktop computers can simulate only a fraction of scales.

Multiscale modeling – structure, dynamics, and function
Desired/current capability ratio: 10^4-10^6

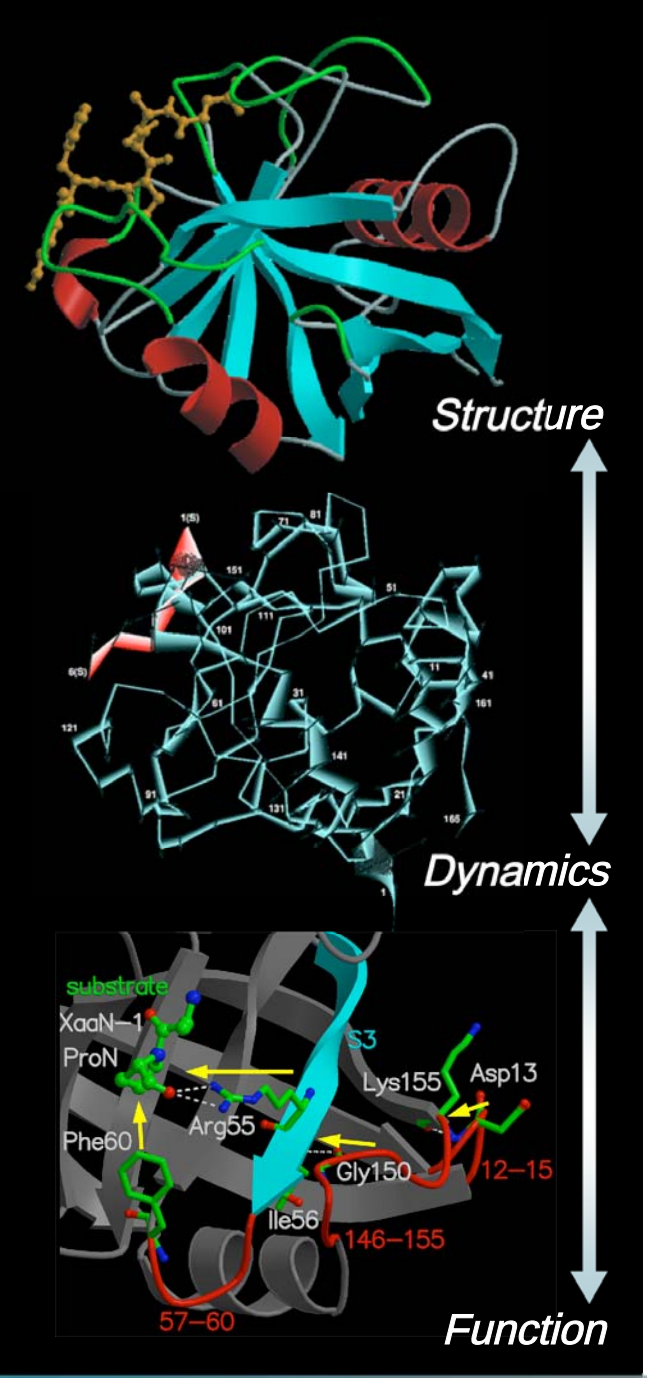
- Using LCF supercomputer Jaguar (Cray XT4), we can simulate scales relevant to the biomolecular processes.
- Using molecular dynamics (MD) and QM/MM codes for multiscale modeling.

Jaguar XT4

LAMMPS (RAQ with 290220 atoms)



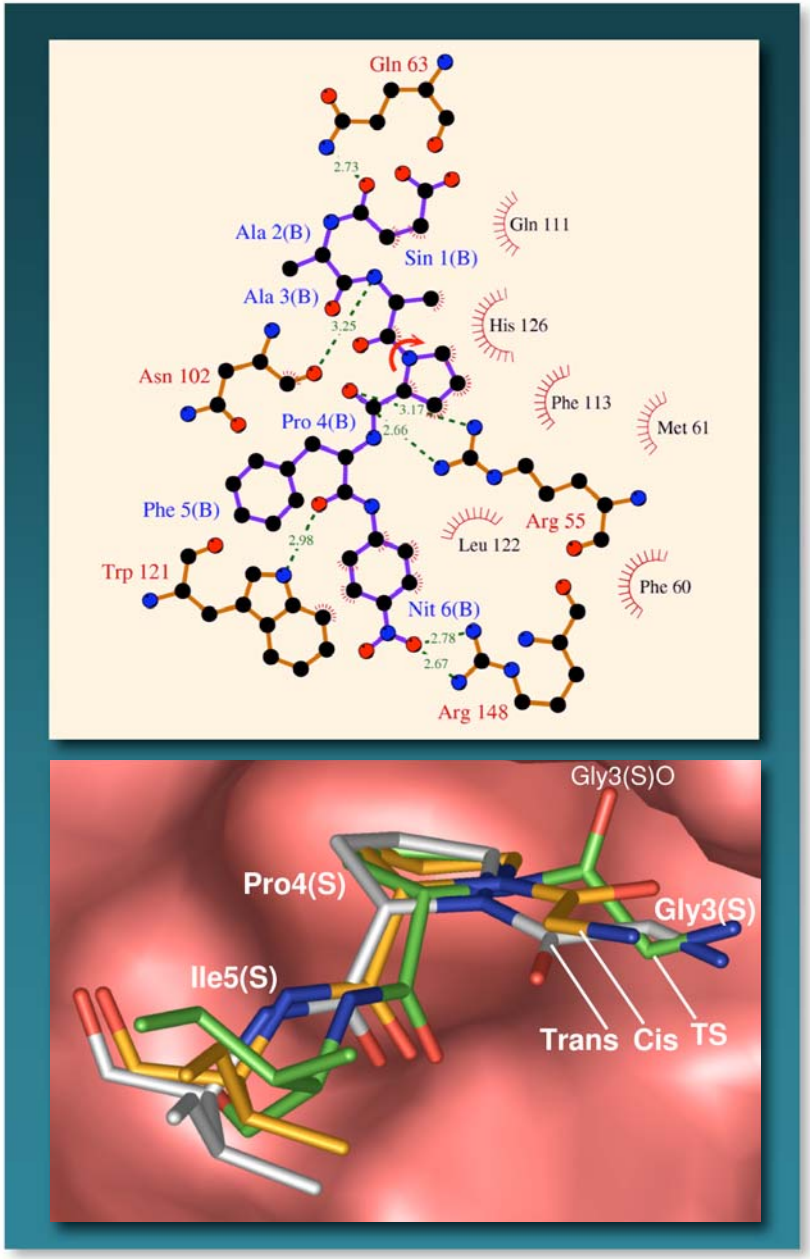
Enzymes: Nature's efficient molecular machines



- Enzymes are naturally occurring catalysts that participate in most biochemical processes in all living cells.
- Enhance reaction rates by many orders of magnitude.
- Studies suggest biomolecules, including enzymes, are dynamical assemblies.
- Increasing interest in the interconnection between structure, dynamics, and function.
- Hydration-shell and bulk solvent fluctuations impact internal motions.

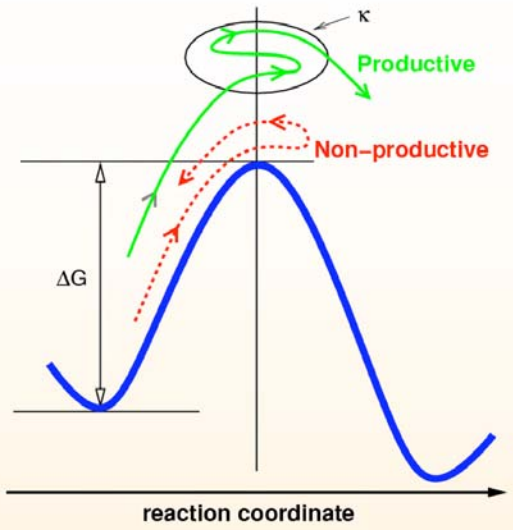
Modeling the enzymes at molecular level

- Investigating enzyme structure, dynamics, and catalytic step at multiple scales
- Detailed atomistic modeling using molecular mechanics
- Quantum effects investigated using electronic structure calculations (ab initio)
- Provides novel insights into the mechanism of the enzyme catalysis
- Small changes in the electronic environment lead to important events in the reaction



Novel computational insights into enzymes:
Protein vibrations promote catalysis

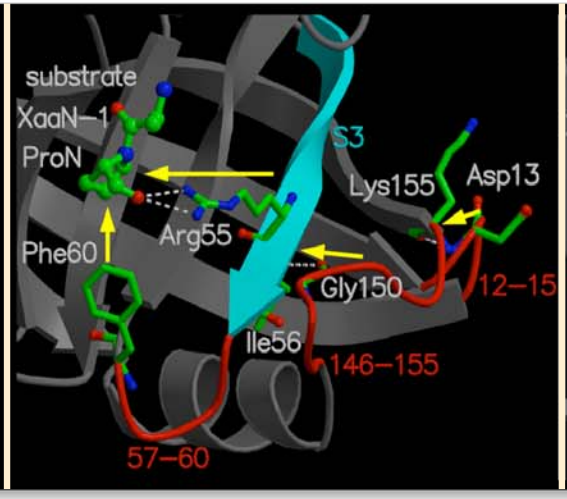
- Investigating the enzyme catalysis step using transition state theory framework
- Free energy profile generation using MD with umbrella sampling
- Discovered protein vibrational events that promote the reaction

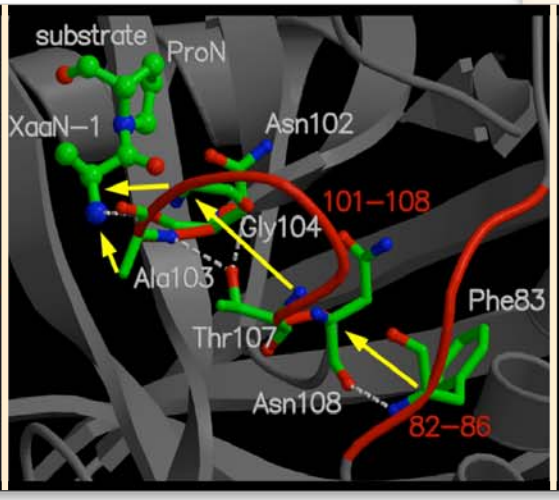


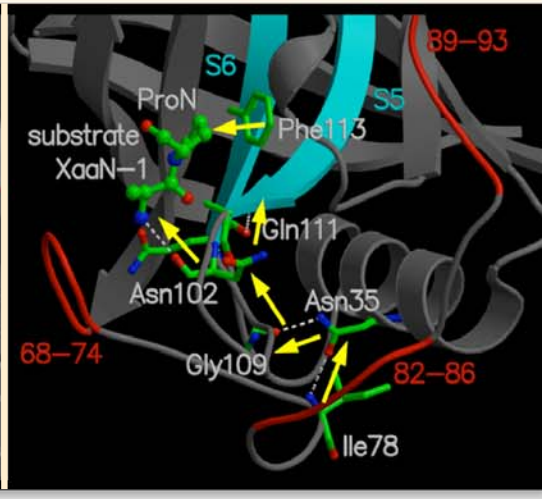
$$k_{TST} = \left(\frac{k_B T}{h}\right) \exp\left(\frac{-\Delta G^{\dagger}}{k_B T}\right)$$

$$k = \kappa \cdot k_{TST}$$

Network of protein vibrations promotes catalysis

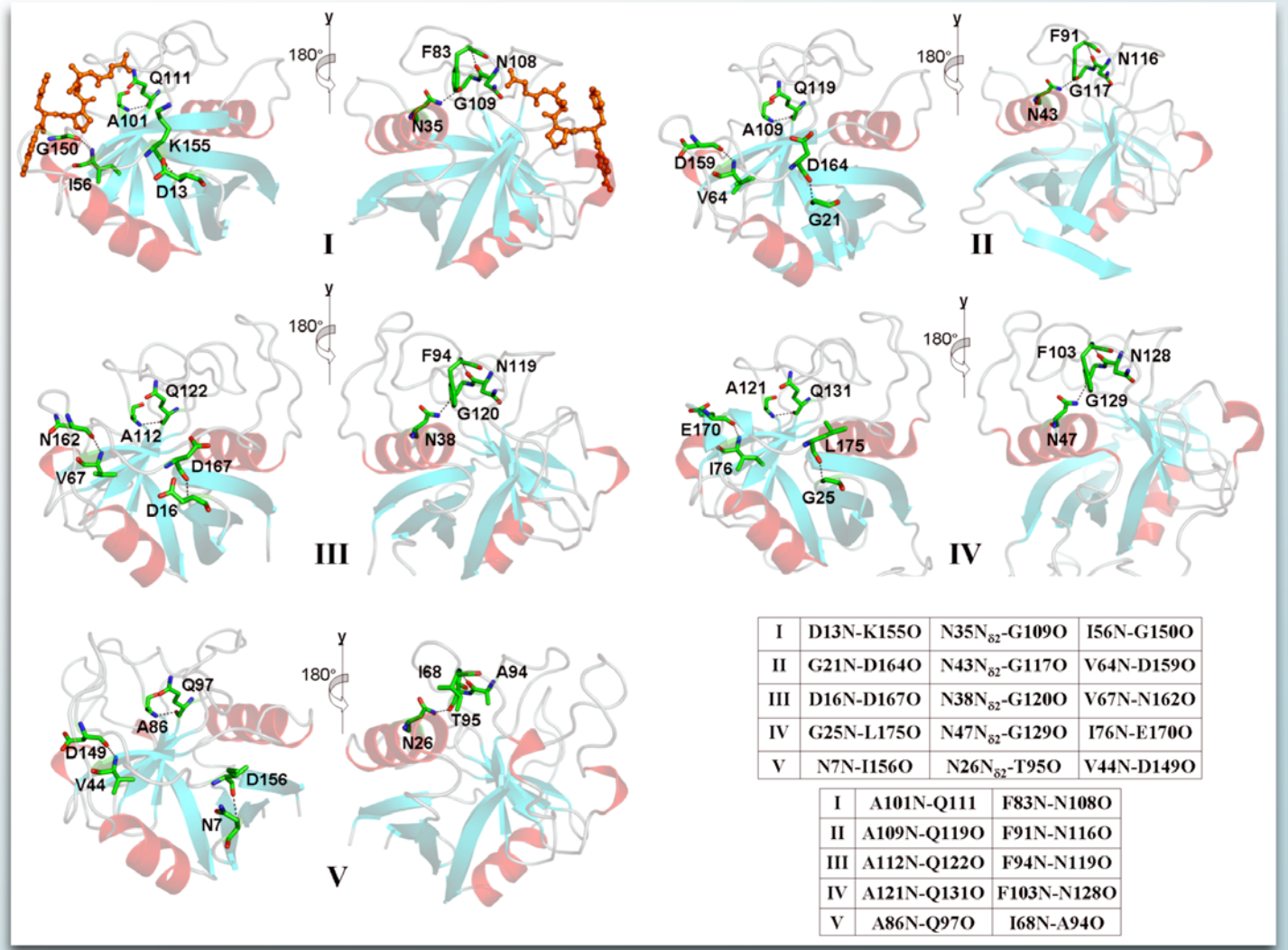






Enzymes have evolved to use dynamical effects

Genomic and structural analysis show conservation of dynamically active portions of the enzymes

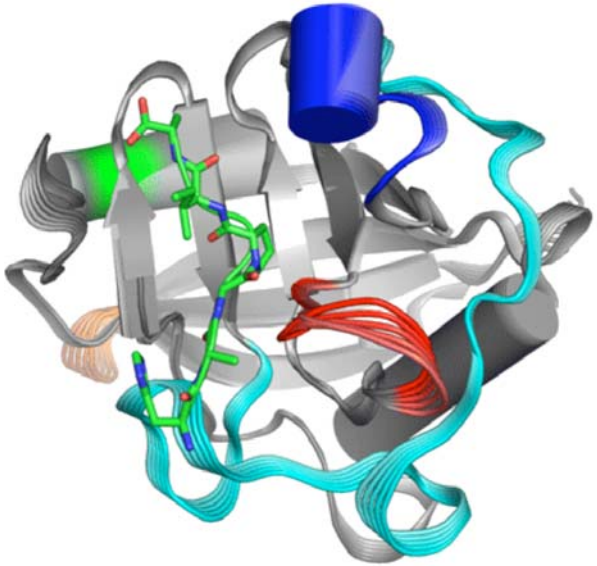


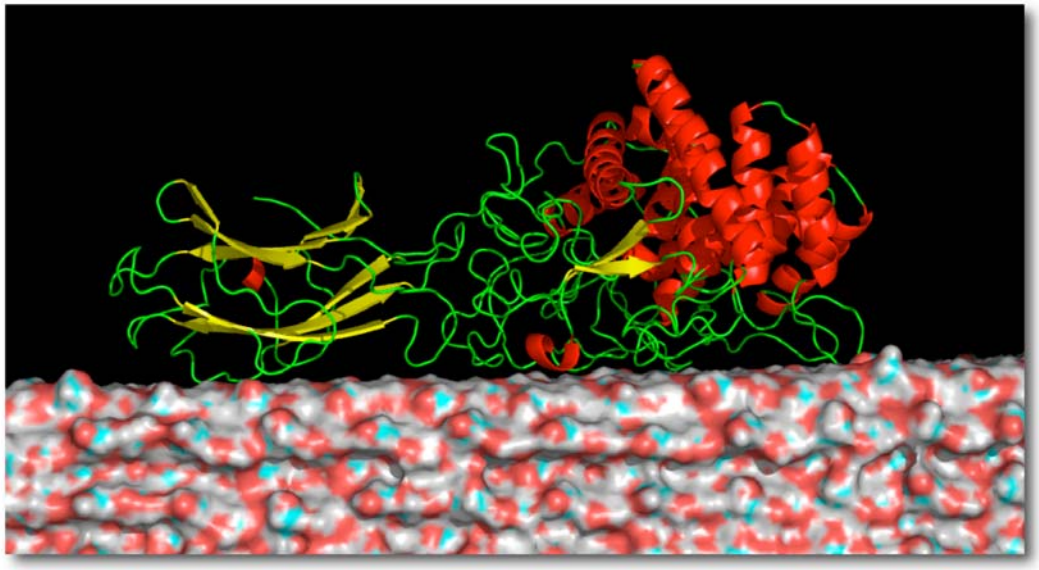
I	D13N-K155O	N35$N_{\delta 2}$-G109O	I56N-G150O
II	G21N-D164O	N43$N_{\delta 2}$-G117O	V64N-D159O
III	D16N-D167O	N38$N_{\delta 2}$-G120O	V67N-N162O
IV	G25N-L175O	N47$N_{\delta 2}$-G129O	I76N-E170O
V	N7N-I156O	N26$N_{\delta 2}$-T95O	V44N-D149O

I	A101N-Q111	F83N-N108O
II	A109N-Q119O	F91N-N116O
III	A112N-Q122O	F94N-N119O
IV	A121N-Q131O	F103N-N128O
V	A86N-Q97O	I68N-A94O

Science breakthrough

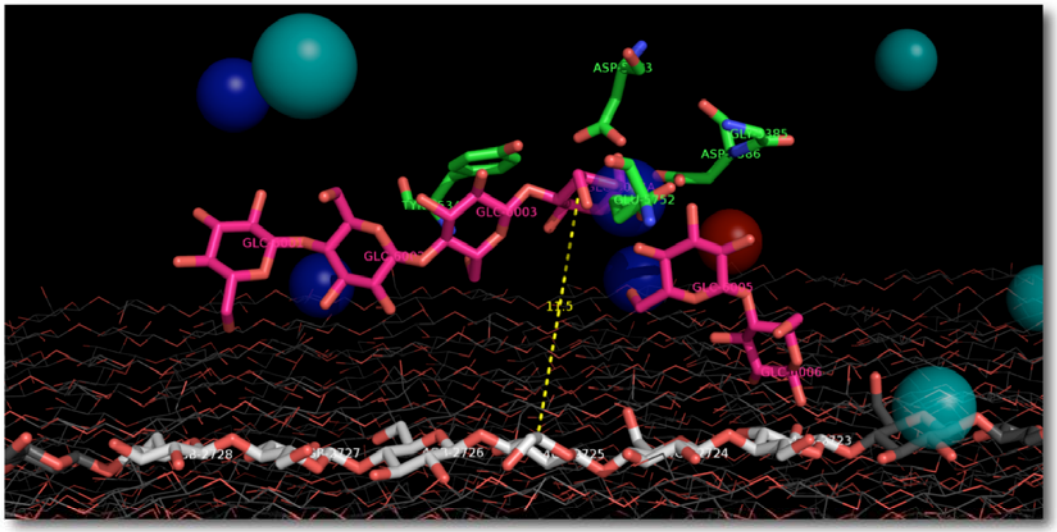
- Novel insights into linkage of enzyme structure, dynamics, and function
  - Evolutionary conservation of protein structure is based on the reaction-promoting vibrations.
- New insights into the mechanism of enzyme *cellulase*
  - Implications for low-cost bio-ethanol production.





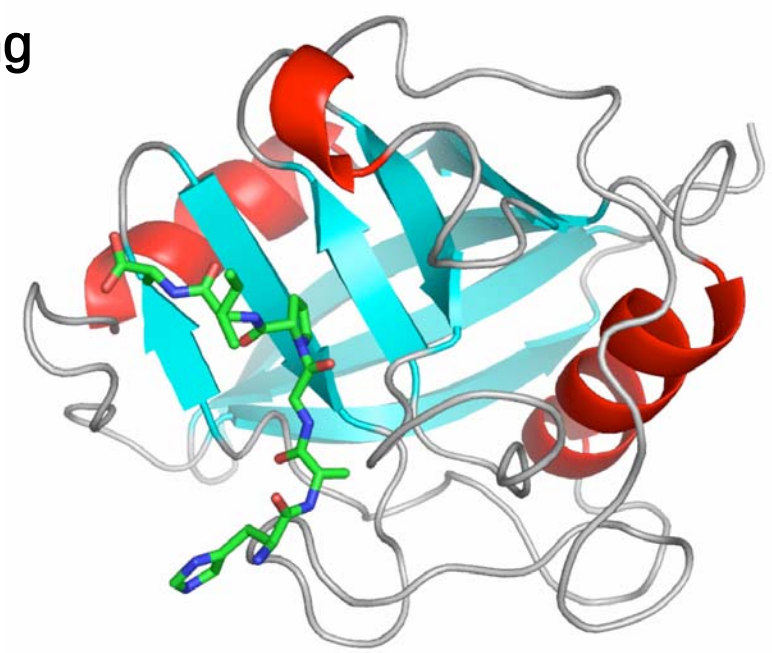
Vital insight: Dynamics are conserved through evolution

- Biomolecular modeling expanded to real systems
  - Enzymes: cyclophilin A, DHFR, and ribonuclease A
  - Human, bacteria, and other species
- Detailed analysis of the protein structure and dynamics
  - Multiscale dynamics: fast, intermediate, and slow modes
  - Active-site and distal to active-site residues explored



Impact of the recent results

- Computational modeling indicates that the “dynamic core” of proteins is conserved over evolution!
- New insights into the “dynamic personality” of enzymes
  - Insight into molecular basis of life
  - Better understanding of the biochemical processes
  - Drug design and protein engineering
- Future investigations exploring
  - Linking dynamics and function
  - Energetic coupling at long range

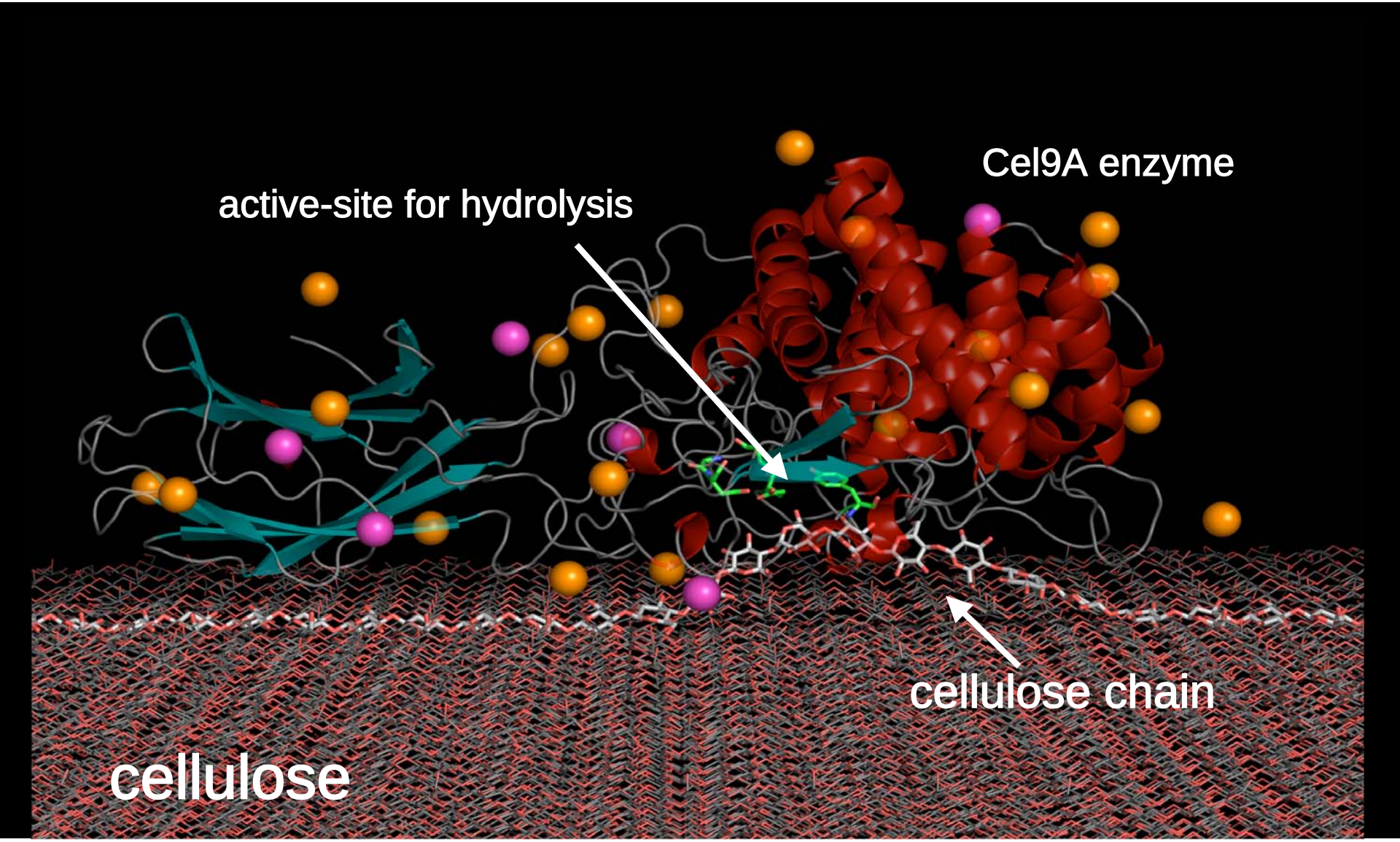


Modeling the cellulase enzyme

- Investigating enzyme structure, dynamics and catalytic step at multiple scales
- Cel9A from *Thermobifida fusca* based on public data (PDB 4TF4), cellulose crystal structure, GLYCAM force-field
- Mechanism of the processive activity of cellulase and the mechanism of glycosidic bond hydrolysis

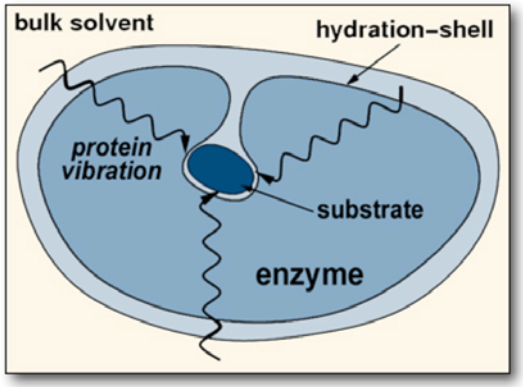
Model of Cel9A from *Thermobifida fusca* in complex with Iβ crystalline cellulose

T. fusca Cellulase enzyme in action



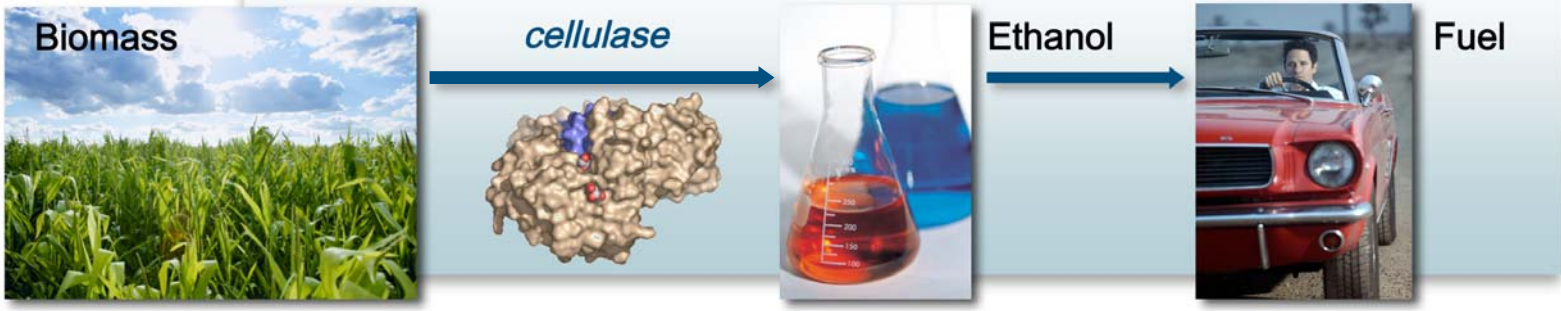
Benefits of the discovery: Improving enzymes

- Wide implications:
  - Protein engineering
  - More efficient enzymes
  - Novel enzymes



- Cleaner environment
  - Role in carbon sequestration
  - Enzyme acts on order of seconds
  - Very large enzyme
  - Multiscale modeling needed

- Ethanol production from biomass
  - Enzymatic degradation of cellulose (biomass) to simple sugars
  - Sugars converted to ethanol by fermentation
  - Naturally occurring cellulose degrading enzymes are very inefficient
  - Computational studies will help in engineering better enzymes and therefore lower costs for ethanol production



Contact

Pratul K. Agarwal

Computer Science Research Group
Computer Science and Mathematics Division
(865) 574-7184
agarwalpk@ornl.gov

Office of Science
U.S. DEPARTMENT OF ENERGY

SciDAC
Scientific Discovery through Advanced Computing

OAK RIDGE
National Laboratory