# Opportunities in Biology at the Extreme Scale of Computing

Mark Ellisman

Rick Stevens

# Workshop Charge

- Examine the scientific opportunities in advanced modeling and simulation at the exascale in the biological sciences
- Opportunity to shape the appropriate role for scientific computing at the exascale in the quest to advance the scientific frontiers in fundamental biological and ecology research
- Examine the role of extreme scale computing in applied biological research such as bioenergy, bioremediation and the understanding of the global carbon cycle.

# Workshop Charge

- A desired outcome of these meetings is to develop a short list of "global challenge" computational problems. Solving these problems should have the potential to transform our understanding of science and its impacts and to improve our ability to apply knowledge in applications important to science, engineering, industry, and society. We anticipate that a final workshop report will address these global challenge computational problems.

- The workshop will be a collaborative effort between BER and ASCR.

# The Big Questions

- What specific problem could be attacked and solved with the application of sustained multiple petaflops of computing power?
- What progress could be obtained on the problem at roughly the 10, 100, and 1000 petaflops levels of sustained performance?

# Focus on the Top 10

- Is the problem one of the "top 10" problems for the scientific discipline, independent of computing?
- Who would constitute the community of scientists and/or engineers that would enthusiastically address the problem?
- What would be the degree of international potential participation?

# Large-Scale Computing Required

- How is the use of petascale computational modeling and simulation irreplaceable in answering this question?
- Does it augment existing techniques or replace them?
- Is there history of large-scale computation being the preferred approach for this problem?

# Feasible to use other methods?

- Why are the other techniques (e.g., experiments/observation, more traditional theory) that could answer these questions not satisfactory? Is it even feasible to consider other techniques?

# Current State of the Tools

- What is the current status of the computing tools for the work being proposed: mathematical models, algorithms, software, and data analysis tools?
- What is the largest scale to date that codes have been run? (e.g. 1,000, 10,000, 100,000 cores)
- Are there existing code teams working on codes for this problem area, or is this a new area that would need seed investments?

# What Data do we need for Validation and Verification

- What experimental and observational data is there available to validate the codes? Is the validation method well established?

# What is Missing?

- What are the missing pieces in the areas of mathematical models, algorithms, software required to solve the problem?
- How would you rank them in terms of importance, cost, and risk?

# Breakout Themes

- *Tissues, Organs and Physiology*
- *Pathways, Cells and Organelles*
- *Macromolecules, Proteins and Protein Complexes*
- *Genomes and Populations to Ecosystems and Evolutionary Dynamics*
- *Advanced Imaging and Data Analysis*

# General Breakout Flow

- Introductions
- Quick Pass on Ideas From Participants
- Group Discussion
- Formulation of ~4 consensus “targets”
- Discussion of Requirements for Computing, Software Algorithms, Tools, etc.
- Generate Summary Descriptions
- Connect to DOE if possible