CRA-NIH Computing Research Challenges in Biomedicine Workshop Recommendations

Dan Reed
Chair, Board of Directors
Computing Research Association (CRA)
Chancellor's Eminent Professor
University of North Carolina at Chapel Hill
Dan_Reed@unc.edu

Chris Johnson
Director,
Scientific Computing and Imaging Institute
Distinguished Professor
School of Computing
University of Utah
www.sci.utah.edu
crj@sci.utah.edu

Acknowledgments

The workshop co-chairs extend their thanks to the Computing Research Association (CRA) and the National Institutes of Health (NIH) for their generous sponsorship of this workshop. We thank the workshop attendees for their time and active participation. We extend special thanks to the workshop working group leaders: Jim Foley of the Georgia Institute of Technology, Jim Gray of Microsoft Research, Lee Hood of the Institute for Systems Biology, Gwen Jacobs from Montana State University, Ed Lazowska of the University of Washington and John Wooley of the University of California, San Diego.

Motivation and Goals

The status, goals, and impediments for 21st century biomedicine were well summarized in the 2004 NIH Roadmap. The Roadmap noted that computing has become absolutely essential to progress in biomedicine, stating:

The success of computational biology is shown by the fact that computation has become integral and critical to modern biomedical research.

However, the report also noted that both the substantial and substantive challenges biomedical researchers face in embracing and applying cutting-edge computing research, as well as those faced by computing researchers in understanding current and future biomedical computing needs, have inhibited biomedical research:

Because computation is integral to biomedical research, its deficiencies have become significant limiters on the rate of progress of biomedical research.

The productive synergies between these two fields can accelerate research in both, but only if we address these challenges through cooperative effort. The agencies and the communities, in other words, must work together to enhance frontier or cutting edge research at the interface. Hence, the Computing Research Association (CRA) and the National Institutes of Health (NIH) sponsored the CRA-NIH Computing Research Challenges in Biomedicine Workshop in Bethesda, MD on June 15-16, 2006. A number of leaders in computing and biomedicine participated, along with NIH Program Directors. There was an extended time for discussion and breakout groups. For a list of attendees and the text of the invited presentations, see the workshop web site at www.sci.utah.edu/ncrr/wiki/index.php/CIBC:Workshops:WorkshopCRC06.

The objective of the workshop was to develop a list of focused recommendations and action items that would guide the NIH and computing communities in addressing current impediments to fully realizing effective collaborations at the interface between computing and biomedical research. In this context, we use the term “computing” to include the entire range of contemporary approaches in computer science and engineering and information technology. Similar considerations apply to partnerships between computing and the biological domains funded by other agencies, most notably those of NSF, DOE and USDA, which includes basic plant, microbial and environmental biology and applications to bioenergy and agronomy.

The workshop’s focus was on identifying action items; there are already numerous reports on computational biology that describe the challenges and opportunities, but the limitations at this interface remain unaddressed. This document summarizes the outcomes of the workshop.

An Overview of the Workshop Discussions

Many recent reports [1-6] have discussed the potential, critical and necessary application of computing to biomedicine. Each has emphasized that computing is indispensable for future biomedical discovery and for translational medicine leading to advances in health care. Although scientists and administrators agree on the need to exploit the synergies from increased integration of computing and biomedicine, the NIH has been slow to embrace computing researchers fully. Similarly, computing researchers remain largely unaware of biomedical research challenges that could benefit from computing research and that could inform and stimulate innovative computing research.

Thus, this workshop articulated a prioritized list of recommendations or actionable items, which, if pursued, could realize many of the potential benefits detailed in [1-6]. We assigned working groups to focus on specific opportunities and current limitations (see the section below on Process and Resources). A consensus quickly emerged from the working groups (see the section below on **A Summary of Computing Opportunities Motivated by Biomedical Research and Healthcare Needs**). Although there was widespread agreement on many of the technical challenges and needs, we identified the following more difficult challenges:

1) The need to change legacy government and university **structures** by **significant incentives** for change and better coordination.
2) The need for more **interdisciplinary education, collaboration and communication**.
3) The need for **funding and reward mechanisms** that encourage, support and realize significant connections between computing and biomedical research.

As such, the workshop recommendations emphasize ways to overcome current roadblocks by modernizing outdated structures, providing guidelines and rewards for moving across traditional disciplinary boundaries in education and research, and engaging computing and biomedical researchers.

As succinctly stated in the 2005 PITAC Report on Computational Science [2]:

> *Organizational structures in academia have antecedents reaching back to the Renaissance…These structures evolve so slowly that creating a new department often requires years of negotiation and resource planning…The Federal R&D agencies have similar constraints on organizational change.…The result is an architecture of organizational structures trapped in time and constrained in rigid disciplinary silos whose mutually reinforcing boundaries limit adaptation to changing research needs and competitive pressures.*

> *Traditional disciplinary boundaries within academia and Federal R&D agencies severely inhibit the development of effective research and education in computational science. The paucity of incentives for longer term multidisciplinary, multiagency, or multisector efforts stifles structural innovation.*

Although significant changes to University or NIH organizational structures would be hugely beneficial, the enormity and formidability of such a task is beyond our scope (although we strongly recommend that the National Academies and NIH work together to start such a process as outlined in [2] and [7]). Instead, we provide a short list of recommendations that, if enacted, could significantly advance the level of collaboration between computing and biomedical research and both promote and enhance innovation and discovery.

As one of our working group leaders, Lee Hood from the Institute for Systems Biology, said: "*New ideas require new organizational structures.*" This is an apt summary of our six parallel and interconnected recommendations. The comprehensive implementation of these recommendations is needed for the full potential of computing to be applied to biomedical research and accelerate translational medicine.

Consensus Recommendations:

Recommendation 1: The NIH, NSF and DOE Office of Science should support biomedicine and computing research collaborations by three actions:

1) Initiate a program for small, interdisciplinary planning grants that requires conceptual proof-of-principle but minimal or no preliminary results and that involves both computing and biomedical researchers as full partners in the collaboration
2) Create a cross-disciplinary program or expand the scope of extant programs to fund computing and biomedicine research projects at the individual PI level (so-called “2 x 2” grants), as well as funding larger, collaborative projects with multiple computing and biomedical PIs. Many of the awards from the program would reflect the maturation of teams and their projects funded through action one.
3) Establish a cross-disciplinary, multiagency working group to identify, explore and recommend individual agency opportunities and define and coordinate joint agency programs. The working group would also track the delivery of actions one and two above and suggest improvements when needed. The multi-agency working group should also track the response to the other recommendations provided below.

These cross-disciplinary programs would be one of the positive outcomes from a balanced research funding portfolio arising from the proposed competitiveness initiatives (based on [7]). It would also provide proper resources and environments for addressing the central problems of biology and medicine. These problems should drive the biomedical computing agenda. *Only an increase in agency partnerships in calls for proposals can adequately address the deficiencies and limitations in the breadth and depth of research coverage and ensure full participation by the computing community.*

Timeline: Year 1: New programs created. Year 2: Call for proposal review. Year 3: Funded biomedicine and computing research programs start.

Estimated Cost: $25M/year.

Recommendation 2: The federal agencies, utilizing their respective mechanisms and available options, should enhance support for training at the interface. Training mechanisms would include summer schools for students, post-doctoral research associates, and professors; increased emphasis on extant undergraduate and graduate training programs; and funding to transform existing “silo” disciplinary education into new, multidisciplinary structures that support the integration of computing and biomedicine.

Timeline: Year 1: NIH, NSF, and CRA work together to create new educational programs. Year 2: Call for proposals and proposal review. Year 3: Funded trans-agency interdisciplinary educational programs start.

Estimated Cost: $25M

Recommendation 3: The NIH should create a cross-institute software program to create and maintain high-quality, well-engineered biomedical computing software; to assess the quality and capabilities of existing biomedical computing software; and to create and support for biomedical computing software repositories that are deemed useful. This might involve setting up a portal or clearinghouse for the various repositories. DOE and NSF should each create a parallel program focused on the software for the research domains they fund.

Timeline: Year 1: Create cross Institute software program. Year 2: Call for proposals and proposal review. Year 3: Funded software program starts.

Estimated Cost: $20M/year

Recommendation 4: The NIH should fund a small number of large, distributed transformational centers to act as "expeditions to the future," as described in the 1999 PITAC report [11]. These centers would engage some of the best, most visionary researchers in both computing and biomedicine to attack a set of high impact (transformational) biomedical problems. We emphasize that these centers are distinct from and somewhat orthogonal to the successful NIH National Centers for Biomedical Computing (NCBC) program. While the NCBCs largely focus on creating important biomedical computing infrastructure, the expeditions to the future centers would focus on visionary, innovative, higher risk, transformational research in both computing and biomedicine.

To support the centers at sufficient funding levels, the NIH should mandate a specified level of non-federal support from private industry (much like the NSF Engineering Research Centers and Science and Technology Centers). We emphasize the criticality of avoiding micromanagement of these centers. Rather, they should be given the flexibility to invent the future unhindered by diverse federal bureaucratic obstructions.

Timeline: 5-10 years.

Estimated Cost: Each Center budget - $20M/year.

Recommendation 5: The NIH should invest in the range of computing research and existing technologies specified in detail in the following section; namely, the Summary of Computing Technology Motivated by Current and Future Biomedical Research and Healthcare Needs.

Timeline: Immediately, across all Institutes.

Estimated Cost: $40M/year

Recommendation 6: The NIH, NSF, DOE and the CRA should create a joint Interface Task Force (ITF) to recommend specific ways to accelerate and support advances at the interface between computing and biomedicine. Such a task force will seek to provide major outreach to the computing (CRA) and biomedicine (NIH) communities; this outreach activity would explain the specific needs and articulate the immediate possibilities for collaboration. The Interface Task Force, to guide the educational effort of action two, should establish a framework for broader multidisciplinary graduate and undergraduate training in both computing and biomedicine as recommended in *BIO2010: Transforming Undergraduate Education for Future Research Biologists* [8] and *Computing the Future: A Broader Agenda for Computer Science and Engineering* [9].

The nascent Computing Community Consortium (CCC) [10] is one possible vehicle for launching this activity within the computing community; the major professional societies, including FASEB, could pursue the parallel effort within the biology community. The agencies should identify their representatives, implement the ITF and provide operations support (which would be minimal).

Timeline: Create the Interface Task Force as soon as possible. The initial report and recommendations of the ITF will be due within six months, and the detailed report and recommendations within a year. As long as there remain significant limitations on the impact of computing on biomedicine, the ITF should continue to track actions one through three and to meet once a year with the multiagency working group.

Estimated Cost: $500K

A Summary of Computing Opportunities Motivated by Biomedical Research and Healthcare Needs

A diverse set of computing research and technologies can benefit biomedical researchers and healthcare professionals now and in the future, and the enhanced implementation would have early, extraordinary and even unanticipated impacts. Existing programs have clearly been inadequate [4]. To provide those benefits, we strongly and unanimously urge the NIH (and other agencies for their respective domains) to make significant investments in integrating computing research and existing technologies into the funded portfolio. Below, we summarizes these computing research domains and corresponding technologies.

Given the full agreement among the participants and the need to focus this report on policy, education, and collaboration issues, we have not tried to provide a detailed assessment, but rather we include an overview addressing larger categories. A number of recent reports that discuss these issues in more detail [1-6]. To organize and simplify the description of opportunities, we use conventional categories, which will each require attention by the NIH. That is, the relevant computing technologies can be broadly categorized into two main groups: 1) Data Analysis and Management Algorithms and Tools; and 2) Cyberinfrastructure and Software Infrastructure.

1) Data Analysis and Management Algorithms and Tools. Given the increasing amount of high information content data with which biomedical researchers and clinicians have to deal, data analysis and management will continue to become ever increasingly important.

- Identify, Evaluate and Inventory Existing Models and Tools - There exists a large number of existing algorithms and software that are largely unknown in the biomedical community. There is a need to identify, evaluate, and inventory these tools.

- Data Representation and Management - While there has been some work in creating common data representations in imaging and genomic databases, there needs to be more agreement on data representations for biomedical data and corresponding tools to support such representations.

- Medical informatics – The information in patient records, including ever more sophisticated molecular analyses, needs to be organized for data mining toward such outcomes as improved diagnosis and prognosis evaluation. Traditional medical informatics has not incorporated the potential of the current, advanced information technology. Inclusive multiscale and multimodal knowledge repositories directly improving health care require significant and innovative involvement of computing with medical practice. These will extend advances in similar repositories for biomedicine and translational medicine, require that the challenges in data management are addressed, and will benefit from all of the other domains listed.

- Scientific and Information Visualization - As noted in [3], there are a number of visualization tools and techniques that could benefit biomedical scientists, engineers and clinicians. In addition, new visualization tools will need to be created by biomedical researchers collaborating directly with visualization researchers.

- Modeling and Data and Image Analysis - There is a need for both incorporating existing algorithms and software, as well as creating new algorithms and software for modeling and data and image analysis, including tools for image search, pattern detection, comparing data sets and images, dealing with large-scale, multi-scale, and (potentially high) multi-dimensional data and with the enormous signal to noise issues inherent in the validation and standardization of large data sets.

- Algorithm Classification, Analysis, and Improvement - Understanding of underlying computational complexity of various biomedical problems required

2) Cyberinfrastructure and Software Infrastructure - There is a significant need to support both the creation of new software tools and providing the necessary support for software infrastructure and software engineering. The biomedical software base is currently inadequate to keep pace with and support evolving applications needs. In addition, software development and support is significantly under funded and there is a need to rebalance our investments between infrastructure and research to maximize progress [2]. The infrastructure requires Open Science; namely, Open Software, Models and Data, and the establishment of National Data and Software Repositories.

- Collaboration Software - Information guided collaborative problem solving (between computer and user) - applied across broad variety of areas. Virtual organizations (VOs) / cyber frontier as the front-end to integrate teams that relies on infrastructure, grids, knowledge bases, facilitate international/large-scale studies

- Integrated Software Environments, Modular Components, Workflow capture, easier to use software, scalable software environments

- Computer Aided-Human Insight - Smart decision support, HCI, surgical planning, treatment

- Medical Informatics and current privacy laws – evaluate challenges in specific context of clinical med; Merging Personal Health record (digital) to

Clinical translational research data (Complete privacy & security solutions with appropriate attention to HIPAA concerns).

Process and Resources

The day-and-a-half workshop was held during June 15-16, 2006 in Bethesda, MD. After Chris Johnson from the University of Utah provided workshop attendees with an overview of workshop goals, the workshop began with four opening addresses, two on biomedical opportunities by Lee Hood (Institute for Systems Biology) and Jill Mesirov (MIT) and two from Dan Reed (UNC Chapel Hill) and Chris Johnson (Utah) on computing opportunities. The attendees then broke into three working groups, each led by a biomedical scientist and a computing scientist. The group leaders were:

Red Team: Gwen Jacobs (Montana State) and Ed Lazowska (Washington)

Blue Team: Lee Hood (ISB) and Jim Gray (Microsoft Research)

Green Team: John Wooley (UCSD) and Jim Foley (Georgia Tech)

The working groups were charged with exploring three questions:

Question 1: What are examples of computing research that could meet the present and future biomedical research needs?

- The problem is that there exist significant biomedical computing efforts that do not use state-of-the-art computing techniques.
- When identifying examples, give an idea of what kinds of improvements could be realized with new techniques and what the investment/cost would be.
- Computer scientists lead Biomedical scientists help with “reality check”.

Question 2: What are the high priority computing challenges motivated by biomedical research present and future?

- After listening to the biomedical keynote presentations, what areas of computing research would you consider most useful for biomedical applications?
- Biomedical scientists and computer scientists work together and perform mutual “reality checks”.

Question 3: What are the challenges for realizing the synergies between these Computing and Biomedicine that can drive both forward through cooperative effort?

- What are the most significant roadblocks?
- What are the best ideas for removing these roadblocks?
- What are ideas for having significant impact?

Day one ended with an overview of the recommendations from each working group followed by a workshop dinner. Day two focused on creating a list of prioritized recommendations and discussion further action items.

All of the information pertaining to the workshop, as well as links to a number of related reports can be found on the CRA-NIH Computing Research Challenges in Biomedicine Workshop Wiki:

www.sci.utah.edu/ncrr/wiki/index.php/CIBC:Workshops:WorkshopCRC06

Acknowledgements:
A special acknowledgment of our friend and colleague Jim Gray, whose loss will be felt throughout the computing community. Jim devoted his time, enthusiasm, and considerable intellect to bridging computer science and the natural sciences so that both might benefit. The authors would like to thank Peter M. Lyster of the NIH National Institute of General Medical Sciences Center for Bioinformatics and Computational Biology (CBCB) and Bret Peterson of the NIH National Center for Research Resources for helping to organize the CRA-NIH meeting on June 15-16, 2006.

References:
[1] J. Wooley and H. Lin. Catalyzing Inquiry at the Interface between Computing and Biology, National Academy Press, Washington, D.C., 2005

[2] Computational Science: Ensuring America's Competitiveness. President's Information Technology Advisory Committee (PITAC), Subcommittee on Computational Science, D. A. Reed, Chair, (http://www.nitrd.gov), June 2005.

[3] NIH-NSF Visualization Research Challenges, C.R. Johnson, R. Moorehead, T. Munzner, H.Pfister, R. Rheingans, and T. Yoo. (available from the IEEE Computer Society, http://tab.computer.org/vgtc/vrc/index.html, ISBN: 0-7695-2733-7), 2006.

[4] NIH Biomedical Information Science and Technology Initiative (BISTI), by the Biomedical Computing Advisory Committee to the Director: D. Botstein, L. Smarr, D. Agard, M. Levitt, D. Lippman, D. Herrington, C. R. Johnson, G Rose, G. Rubin, A. Levison, M. Spence, and H. Smith, C. Peskin, G. Jacobs. (http://www.bisti.nih.gov/June_1999_Report.cfm), June 1999.

[5] NIH BECON/BISTIC Symposium (BB2004): Symposium Report: "Biomedical Informatics for Clinical Decision Support: A Vision for the 21st Century", June 21-22, 2004, http://www.bisti.nih.gov/becon2004_final_report.doc

[6] NIH Knowledge Environments for Biomedical Research (KEBR) Conference, December 11-12, 2006, www.esi-bethesda.com/ncrrworkshops/kebr/index.aspx

[7] Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future. Committee on Prospering in the Global Economy of

the 21st Century: An Agenda for American Science and Technology, National Academy of Sciences, National Academy of Engineering, Institute of Medicine, 2005 (http://www.nap.edu/catalog/11463.html).

[8] BIO2010: Transforming Undergraduate Education for Future Research Biologists. Committee on Undergraduate Biology Education to Prepare Research Scientists for the 21st Century, National Research Council, 2003 (http://www.nap.edu/catalog/10497.html).

[9] Computing the Future: A Broader Agenda for Computer Science and Engineering. J. Hartmanis and H. Lin, Editors, Computer Science and Telecommunications Board, National Academies, 1992 (http://www7.nationalacademies.org/cstb/pub_computingthefuture.html).

[10] Computing Community Consortium (CCC), www.cra.org/ccc

[11] Information Technology Research: Investing in Our Future, President's Information Technology Advisory Committee (PITAC), February 1999, http://www.nitrd.gov/pitac/report/