Opportunities for Biomedical Research and the NIH through High Performance Computing and Data Management

Prepared by the Coalition for Advanced Scientific Computing (http://www.casc.org)[1].

10 January 2003

Executive Summary

The biomedical sciences are advancing at a tremendous rate. Some of the most notable recent accomplishments (such as the assembly of the human genome) have depended upon the use of high performance computing and data management (HPC). There are important areas of opportunity for the biomedical sciences to accelerate advances in knowledge and in practical medical treatments through the use of high performance computing.

As we enter into the "century of biology" there are critical challenges in the areas of data organization, management, and analysis; simulation; and translational biomedical research. These challenges can be met only through investment in training, tools, and infrastructure that will enable greater use of high performance computing in biomedical research.

The Coalition for Advanced Scientific Computing makes the following recommendations to the NIH, all of which are consistent with the 1999 Biomedical Information Science and Technology Initiative report:

- Invest significant resources in training professionals at the interface between computing, biology, and medicine via support and development centers specialized in HPC for biomedical research.
- Invest in the creation of new biomedical HPC tools — robust applications and new algorithms.
- Initiate funding for medium and large computational facilities and data grids that provide hardware and software computing infrastructures to accommodate the national need in biomedical computing.
- Create a better interface between the HPC and biomedical research communities.

Direct investment by the NIH will ensure that the infrastructure, training programs, and tools match the needs and priorities established by the NIH as it charts the nation's course in biomedical research. We invite the NIH to embark upon a partnership with the HPC community that will benefit biomedical science and improve human health throughout the U.S. and around the world.

[1] Lead authors: Craig A. Stewart, Indiana University; Ralph Z. Roskies, Pittsburgh Supercomputing Center; and Shankar Subramaniam, University of California San Diego. Stewart is a Ph.D. biologist active in High Performance Computing. Roskies has for many years been a leader in providing HPC infrastructure for biomedical scientists. Subramaniam is director of a program in bioinformatics. A full list of contributing authors is included at the end of this document.

1.Introduction

Biomedical science is advancing at a dramatic rate, in large part as a result of the rapid generation of biomedical data from high-throughput instrumentation. Some of the most dramatic biomedical research breakthroughs, such as sequencing the human genome, were made possible through the use of high performance computing and data management (referred to hereafter as HPC). Yet there are datasets today that cannot be adequately analyzed and theories and models that cannot be fully tested because the requisite HPC tools and infrastructure are either not used for the task, or not sufficient to address the task.

The Biomedical Information Science and Technology Initiative (BISTI) report [1] emphasizes the need for training, tools, and infrastructure to enable the biomedical research community to exploit the value of high performance computing in advancing biomedical research. Many problems in biomedical research can be solved only on the nation's largest, multi-TFLOPS[2] HPC systems. Demand for use of HPC systems to solve biomedical problems is very high. The largest non-classified HPC system in the U.S. is operated by the Pittsburgh Supercomputing Center [2]; more than half of the usage of this system for FY2001/2002 was for biomedical applications [3]. Yet many biomedical computing challenges cannot be solved even on the largest HPC systems in the US.

There are three critical areas of opportunity for the use of high performance computing in biomedical research:

- Organizing, managing, mining, and analyzing the large volume of biological data
- Simulation, particularly in multi-scale problems (modeling from the genome to the organism)
- Translational medicine – transforming research into clinical practice in the shortest time possible.

However, certain critical needs must be met if biomedical research is going to take the fullest advantage of high performance computing:

- Training for biomedical researchers to utilize high performance computing hardware and software effectively, together with ongoing support for biomedical researchers using this technology.
- The creation of robust software tools applicable to a wide variety of biomedical research problems, and the ongoing improvement and maintenance of these tools.
- The creation of a scalable, high performance computing infrastructure to advance the nation's biomedical research agenda, including very large facilities dedicated to biomedical research.

The "century of physics" has given way to the "century of biology." Understanding the factors that enabled the successes of the century of physics may help us understand ways to facilitate and accelerate biomedical research in the coming years. Key among the factors enabling success in the physical sciences was the close relationship between the physical science and HPC communities. This relationship was facilitated by the agencies that funded the basic scientific research (NSF and DOE) because they also funded the HPC infrastructure that supported this research. As a result, the creation, management, and use of the HPC infrastructure matched the priorities of the agencies funding the scientific research and the needs of the scientific communities being served. This successful model is one that the NIH, biomedical research community, and HPC community should follow in the coming years.

Existing NIH programs, such as the National Center for Research Resources [4], the National Library of Medicine's support of biomedical data repositories [5], and BISTI [6], have brought about significant successes in use of advanced information technology in biomedical research. These programs and the use of existing HPC centers by biomedical researchers have paved the way for widespread use of HPC techniques, tools, and infrastructure to biomedical science. In this paper we outline specific paths of action for HPC and the biomedical research community that will accelerate the process of achieving the goals set forth in the original BISTI report.

[2] TFLOPS - Trillion Floating Point (mathematical) Operations Per Second. A 1 TFLOPS computing system is about 400 times faster than a state-of the-art desktop PC.

2. Managing the Data Deluge

TeraByte-scale (TB[3]) data stores already exist in biomedical research [7], and they will become more common as digital storage of clinical data and high-throughput instrumentation create data at an accelerating pace. These data comprise the raw material for the process of data management and analysis that can lead to new insights. The dramatic growth in the size of biomedical data stores is well documented, dramatic, and surprisingly consistent among many subdisciplines of biomedical science.

Biomedical data is complex, highly context dependent, and heterogeneous. This makes the mapping of data into knowledge a challenging task. Much of the data organization to date has focused on genes and gene products. Given the defined alphabets of nucleic acids and proteins, this task has been addressed well in the community; many of the tools for managing large sequence data sets already exist in the form of massive data storehouses, distributed data grid software, and software for extracting and merging data from disparate data sources [e.g. 8-13].

Many innovative and successful projects are underway for managing large biomedical data stores; of these several have been funded by the NLM. At present much attention is being directed to the problems of large-scale integration of data across different levels of biological organization and multiple geographical locations. The Alliance for Cellular Signaling [14] is an example of a collaboration between biomedical researchers and bioinformaticians that aims to integrate information across multiple levels of biological organization. Data grids and data stores accessible via high-speed internet connections are similarly demonstrating success in linking geographically distributed data sources. Examples of such data grids and internet-accessible data stores include BIRN, a data grid [13]; the PDB, a large data storehouse [9]; and the Visible Human Project, a specialized data store [15].

The need for massive HPC systems to analyze these data remains a need, not yet met, facing the biomedical research community in their attempts to turn data stores into new biomedical understanding. For example, a full pairwise comparison of 30,000 proteins would take 14 years on a single microcomputer. This calculation, which must be re-done whenever new protein information becomes available, can now be completed in a matter of weeks using multi-TFLOPS HPC systems and parallel computing techniques (the application of several processor chips to a single problem). The results of these comparisons are then available to the biomedical research community at large [16]. As the size of these datasets becomes ever larger, it will be increasingly important to provide the most advanced HPC resources to analyze these data, along with new techniques to enable researchers to quickly gain insights from these data.

3. Simulation

Exciting developments in biomedical theory are now possible because scientists can work at new levels of detail to identify, define, and study the components of living organisms, down to the individual molecule. The use of simulations based on HPC tools and running on HPC systems will greatly facilitate such research [17]. Structural biology software programs are prodigious consumers of computing resources. Cell modeling programs are being developed by several research groups [e.g. 18-20]. Furthermore, simulation can now interact with experiments to create insights that could never be achieved through experiment alone. For example, it took thousands of hours of computer time on a nationally shared HPC system to provide fundamental new insights about the function of the aquaporin water channel [21]. This is an excellent exemplar of the interaction between simulation and experiment. The current understanding of how this water channel functions could not have been attained without the interplay of experimentation and advanced simulation.

The recent significant advances in systems biology provide other examples of the value of advanced simulation in biomedical research. A current aim is to model the human heart across multiple scales — from the genome up to the functioning of the entire organ [22]. One objective of this effort is to model a heart attack – an objective not yet achieved and not achievable with current HPC technology.

[3] TeraByte - TB - One TB is a trillion bytes of data, or approximately 1,490 CDs worth of data.

One of the critical challenges facing biomedical simulation is the problem of spanning multiple scales of biological organization. Given data ranging from genome sequences to digital X-rays, the question arises of how to manage the variety of levels of scale in a computer model. HPC applications for physics and astronomy have attacked such problems with the result that many useful techniques (including adaptive mesh refinement and coupled simulation models) now exist. Too little has been done to translate these techniques from the domains of physics and astronomy to biomedical research. We believe that more effort to translate existing HPC software tools from astronomy and physics would accelerate development of software tools for biomedical research.

The demand for the computational systems that support biological theory simulation encompasses a wide range of architectures and capabilities. Some simulations need only an individual PC. Cell modeling applications can run on systems ranging from Beowulf clusters [23] to TFLOPS-level supercomputers. Protein folding calculations may require PetaFLOPS (1,000 TFLOPS) computing systems; such systems do not yet exist.

4. Translational Medicine

Biomedical research is driven by the critical need to reach new insights, and then to translate these insights as quickly as possible into clinical practice and improvements in human health. The mapping of the human genome was a tremendous scientific breakthrough - but its true significance lies in the resultant improvements in human life that it may enable.

High performance computing can function as a "time machine" and accelerate the translation of biomedical breakthroughs into clinical practice. The speed and power of the central processing units of all computers are growing rapidly. Computations that today require an HPC system will someday be possible on a standard personal computer. One example of this is the development of functional MRI (fMRI) [24, 25]. In 1996, fMRI processing in real time required a then state-of-the-art HPC system — a Cray supercomputer. Today, such work is routine for a deskside workstation. The development of software and algorithms using a supercomputer greatly accelerated the availability of this technology on deskside systems, and enabled the medical community to plan for offering this technology as a routine service.

More recently, 60,000 CPU hours of supercomputer time were used to model a novel group of antimicrobial polypeptides that were subsequently successfully synthesized. As a result it should become possible to treat surfaces with these polypeptides and thus confer antimicrobial properties [26]. Another example of current translational research and high performance computing involves Gamma Knife radiation therapy for brain cancers [27]. Currently a generalized model of the human head is used for targeting the radiation. If a patient's head varies from this model, inaccuracies in targeting can result. Current work to adapt radiation transport codes from nuclear science will make it possible to use a model of each patient's head and thus improve the efficacy of Gamma Knife radiation therapy. Today this software requires a high-end supercomputer, but as with the fMRI example mentioned above, advances in computer power will make this technique routinely usable in clinical settings.

5. Challenges in achieving the promise of high performance computing as a tool for biomedical research.

a. Training and support

Biomedical researchers who use HPC systems consume computing and storage resources at prodigious rates. As mentioned previously, more than half of the usage at the largest non-classified HPC system in the U.S. (an NSF-funded system at the Pittsburgh Supercomputing Center) during FY 2001/2002 was for biomedical applications. Conversely, anecdotal evidence from the biomedical community and the HPC community consistently suggests that the fraction of the biomedical community that uses HPC techniques is, overall, rather low.

The learning curve for using HPC systems is quite steep, particularly given the relatively low percentage of biomedical researchers that use HPC techniques and consequent shortage of guidance and assistance available from

most researchers' local colleagues. The biomedical community would benefit significantly from a program of training for biomedical researchers interested in using HPC facilities, along with support for those who are adopting HPC techniques. Programs such as DOE's pre-doctoral computational science fellowship [28], NIH training grants [29], and American Mathematical Society mid-career fellowships [30] provide models for some aspects of such a program. However, such programs focus largely on pre-doctoral or immediate post-doctoral fellowships. A critical need at present is a focus on cross-training for established biomedical scientists so that they can more effectively exploit HPC technology.

A training and support program for HPC in biomedical research must cover a broad spectrum of HPC tools and systems, ranging from small Beowulf clusters up to the nation's largest HPC systems. The assistance that such a program would give biomedical researchers in using a broad range of HPC tools and systems would greatly benefit the nation's biomedical research agenda.

b. Creating robust software tools and algorithms for biomedical research problems

Robust software tools and infrastructure, designed to meet the needs of biomedical researchers, are required in order for HPC technology to provide the greatest benefit in biomedical research. The learning curve for using HPC systems is steep, in part because the software tools available for biomedical research are still relatively primitive or limited in other ways. Further, many good first implementations of HPC tools fell by the wayside because the tools were not made sufficiently robust and were not supported long enough to attract a large community of users. The Alliance for Cell Signaling [14] and the Protein Data Bank [9] provide certain tools for biomedical scientists, but these initiatives are funded independently and on a limited timeframe. Examples of excellent new efforts to design interfaces to HPC infrastructure based on the needs of biomedical researchers include the Biology Workbench [10] and the GAMESS Web Portal [31].

HPC technology will be of greatest practical benefit to the biomedical research community once there is ongoing and reliable support for the creation and maintenance of robust software tools that are designed to meet the high-priority needs of those users. Such tools will encourage the adoption of HPC techniques and thereby accelerate progress in biomedical research.

c. National Biomedical Computing Infrastructure

The BISTI report called for the creation of a scalable high performance computing infrastructure to help advance the nation's biomedical research agenda – including very large facilities dedicated to biomedical research. This has yet to happen. The biomedical research community is enthusiastic about using HPC facilities. Its heavy use of the largest systems at the national Supercomputer Centers illustrates this, as does creation of many state-centric HPC facilities. New York, Pennsylvania, Utah, and Indiana are home to multi-million-dollar investments focused on HPC and biomedical research and funded by state or private agencies [32-35].

Still, the nation's high performance computing and data management facilities suffer from several significant problems. Advances in biomedical science are delayed by a lack of available computational resources. Further, state-centric facilities and Beowulf clusters in individual labs are not used in a coordinated fashion. Thus, some resources are idle while high demand for others frustrates researchers and impedes scientific progress.

A national infrastructure for high performance computing and data management for biomedical research would expand resources and integrate those that exist, creating the scalable infrastructure called for in the original BISTI report.

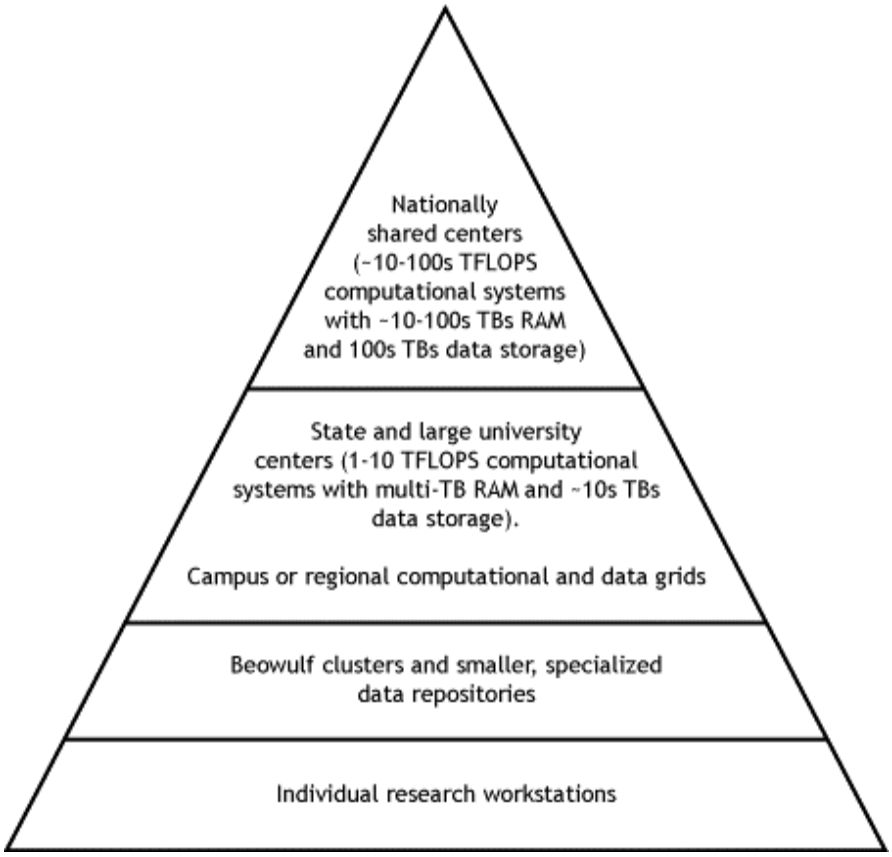

Figure 1. A vision for a national biomedical computing infrastructure viewed within the context of a Branscomb pyramid

The scalable computing infrastructure described in the BISTI report can be considered within the context of a concept called the "Branscomb Pyramid" [36]. This structure describes how the many and varied parts of a scalable national computing infrastructure might fit together. This pyramid, considerably modified and updated from its original, appears in Figure 1.

The width of the various levels of the Branscomb Pyramid indicates the numerical abundance of various types of systems.

- The lowest level represents the many individual microcomputers used by biomedical researchers.

- The second level includes Beowulf clusters used by individual workgroups, and smaller, shared data resources dedicated to specific topics. Such clusters play a critical role in the productivity of individual research labs. The spread of such facilities in the U.S. is also an important factor in increasing the use of HPC techniques – especially parallel processing – in biomedical sciences. Yet these clusters provide inadequate processing power or too little RAM for solving many kinds of problems.

- The third level — large, regional or state-centric centers and regional or local grids — provides resources for problems too large for computing systems in previous levels. Included here are the TFLOPS-level systems available at several universities and regional or local grid facilities. These play a significant role in biomedical research, but do not provide the capabilities available at the top level of the Branscomb Pyramid.

- At the apex are the most advanced computational systems and largest data repositories. This is the level at which “scalable” matters most. This level offers diversity in hardware architecture, including processor chips that incorporate matching facilities particularly useful in genomic and proteomic research. (By contrast, the lower levels are dominated by Intel-compatible processing chips.) Systems at this level also have very large complements of random access memory (RAM), required in many types of biomedical computing applications. No substitute exists elsewhere in the pyramid for these very large systems.

A coordinated, national, biomedical computing infrastructure providing scalable resources spanning the entire Branscomb pyramid would significantly accelerate research. It would ensure that biomedical computing applications are approached with the most advanced computing techniques available, providing the type of scalable computing infrastructure called for in the BISTI report. It would advance the nation's overall research agenda by ensuring that computing and data management systems at all levels of the Branscomb pyramid are used at optimal effectiveness.

6. The way forward

We believe that the NIH should enable and encourage advances in biomedical research by taking the following actions.

a. Invest significant resources in training professionals at the interface between computing, biology, and medicine via support and development centers specialized in HPC for biomedical research.

The importance of training cannot be overemphasized. Training in applying HPC in biomedical research can best be achieved through a strong partnership between the biomedical and HPC communities. A relatively modest investment by the NIH to create a training and support center focused on meeting the needs of the biomedical research community would pay great dividends in the short term and in the long run.

Such a training and support facility must include support staff trained in the use of HPC techniques for biomedical research, as well as biomedical researchers as pre-doctoral, postdoctoral, or mid-career Fellows. Because of modern networking and collaborative tools, such a facility could be located anywhere, and could provide support for a range of HPC resources ranging from Beowulf clusters up to nationally-shared HPC centers. A relatively modest investment in this by the NIH could have tremendous impact.

b. Invest in the creation of new biomedical HPC tools — robust applications and new algorithms.

Advances in the power of computing systems have heightened the value of computation in biomedical research. Contributing just as much to reduced "time to solution" is the development of algorithms. A strong and continued investment in developing new algorithms and robust applications for biomedical computing will yield even greater gains. The development of robust software tools for HPC, created to meet the needs of biomedical researchers, is essential in enabling biomedical researchers to most effectively exploit HPC resources. Broadening the NPEBC program is one path toward this investment. Another is to add this tool to the kind of training and support center suggested above.

c. Initiate funding for medium and large computational facilities and data grids that provide hardware and software computing infrastructures to accommodate the national need in biomedical computing.

The scalable biomedical computing infrastructure called for in the BISTI report can be achieved by creating computational resources at the top layer of the Branscomb pyramid while integrating facilities at lower levels into a national biomedical computing infrastructure. High-end, nationally shared resources could be established by open competitions. The funding of at least one such center would create a proof of concept. Such nationally-shared centers would function as a national resource, accessible to any NIH-funded researcher, and possibly linked to the NCRR. Time allocations could be determined by a peer-review process run by experts within the biomedical community, so that the allocation of resources would follow and promote NIH research priorities.

Such national resources could serve as anchors for national computational and data grids, including efforts to harvest cycles from idle PCs (like fightAIDS@home [37]). National centers would serve a natural coordinating role in a national biomedical computing and data grid. Such centers would also serve as venues for knowledge transfer. Computational science and technologies applied to one area could immediately be applied to another area of biomedical research. By establishing at least one medium- to large-scale computing facility funded by and controlled by NIH, the NIH would have a relatively low-cost proof of concept of the key point in this paper: that HPC infrastructure funded by an agency with a science agenda will most effectively meet the needs and reflect the priorities of that science agenda.

d. Create a better interface between the HPC and biomedical research communities.

This process has already begun, notably through the BISTI report, the NCRR computational centers, services supported by the NLM, and via meetings between the HPC community and NIH leadership. A formal structure to encourage and enable dialog between the biomedical and HPC research communities would accelerate biomedical research and benefit both communities. A track within the BISTIC Symposia, expanded into an ongoing dialog for sharing expertise and setting standards, is one suggestion. Such a dialog must involve the Institutes that comprise the NIH and the many subareas of the HPC community.

7. Conclusions

A number of reports have recommended much greater investment in HPC for biomedical research, including the BISTI and PITAC reports [1, 38]. The recommendations presented here are largely in agreement with those regarding high performance computing in the report of the PITAC panel on Transforming Health Care [39]. The initiatives stemming from the BISTI report, including the BISTIC Symposia and the National Programs of Excellence in Biomedical Computing, are significant steps forward.

It is essential that the NIH take a strong stand in funding the initiatives outlined above. A process based on the NIH model of National Programs of Excellence in Biomedical Computing, targeted at creating nationally-shared training and support facilities, biomedical data centers, tool-building facilities, and one or more high-end HPC facilities for biomedicine, would seem particularly useful. Modest funding in a phase similar to the Pre-NPEBC [40] awards would serve as proof of concept and make progress in establishing the facilities and programs recommended above. Such programs would differ from the existing NPEBC program in that they would be geared specifically to supporting HPC facilities for national use by NIH-funded researchers, with allocations of resources at these centers directed by the NIH.

The NSF and DOE have historically been the providers of national infrastructure in HPC. This precedent was established during the "century of physics" when the physical sciences were the predominant users of shared national HPC facilities. High performance computing is becoming sufficiently mature to be of critical value to biomedical research, as the nation's research agenda is focused much more heavily on medicine and health. Direct investment by the NIH in high performance computing is required for this technology to deliver all it ought to the service of the biomedical research community. Direct and significant investment by the NIH will ensure that the infrastructure, training programs, and tools match the needs and priorities set by the NIH as it charts the nation's course in biomedical research. We invite and encourage the NIH to embark upon a partnership with the HPC community that will benefit biomedical science and improve human health throughout the U.S. and around the world.

References

(all current as of 3 January 2003)

[1] The Biomedical Information Science and Technology Initiative. 1999. Working Group on Biomedical Computing Advisory Committee to the Director, National Institutes of Health. http://www.nih.gov/about/director/060399.htm

[2] Pittsburgh Supercomputing Center. http://www.psc.edu/

[3] Private Communication from Ralph Z. Roskies, Scientific Director, Pittsburgh Supercomputing Center, Nov. 13, 2002.

[4] National Center for Research Resources (NCRR). http://www.ncrr.nih.gov/

[5] National Library of Medicine (NLM). http://www.nlm.nih.gov/

[6] Bioinformatics at the NIH. http://grants1.nih.gov/grants/bistic/bistic.cfm

[7] Private Communication from Anurag Shankar, Manager, Distributed Storage Systems Group, Indiana University, January 10, 2003. As of the beginning of 2003 Indiana University stores approximately 1 TB of biomedical data in its massive data storage system.

[8] National Center for Biotechnology Information (NCBI). http://www.ncbi.nih.gov/

[9] Protein Data Bank. http://www.rcsb.org/pdb/

[10] Biology Workbench. http://workbench.sdsc.edu/

[11] The Globus Project. http://www.globus.org/

[12] North Carolina BioGrid. http://www.ncbiogrid.org/

[13] Biomedical Informatics Research Network (BIRN). http://birn.ncrr.nih.gov/birn/birn_getpage.php?fname=about_birn_bg.html

[14] Alliance for Cellular Signaling. http://www.afcs.org/

[15] The Visible Human Project. http://www.nlm.nih.gov/research/visible/visible_human.html

[16] Databases and Tools for 3-D Protein Structure Comparison and Alignment. http://cl.sdsc.edu/ce.html

[17] Csete, M.E., and J.C. Doyle. 2002. Reverse engineering of biological complexity. Science 295: 1664-1669

[18] E-CELL. http://www.e-cell.org/

[19] VCELL. http://www.nrcam.uchc.edu/vcell_development/vcell_dev.html

[20] Cyber Cell. http://biodynamics.indiana.edu/cyber_cell

[21] Tajkhorshid, E., P. Nollert, M.Ø. Jensen, L. J. W. Miercke, J. O'Connell, R. M. Stroud, and K. Schulten. 2002. Control of the selectivity of the aquaporin water channel family by global orientational tuning. Science 296:525-530.

[22] Noble, D. 2002. Modeling the heart - from genes to cells to the whole organ. Science 295:1678-1682.

[23] Beowulf project. http://www.beowulf.org/

[24] Goddard, N.H., Hood, G., Cohen, J.D., Eddy, W.F., Genovese, C.R., Noll, D.C., and Nystrom, L.E. Online Analysis of Functional MRI Datasets on Parallel Platforms. Journal of Supercomputing, 11(3): 295-318, 1997.

[25] Goddard, N.H., Hood, G., Cohen, J.D., Eddy, W.F., Genovese, C.R., Noll, D.C., and Nystrom, L.E. 2000. Functional magnetic resonance imaging dataset analysis. In: *Industrial Strength Parallel Computing*. Koniges, A. (Ed), Morgan Kaufman. pp. 431-452.

[26] Tew, G.N., D. Liu, B. Chen, R.J. Doerksen, J. Kaplan, P.J. Carroll, M.L. Klein, W.F. DeGrado. 2002. De novo design of biomimetic antimicrobial polymers. Proc. Natl. Acad. Sci. 99:5110-5114.

[27] The Gamma knife. http://www.radonc.iupui.edu/Gamma%20Knife%20pics_files/gamma.html

[28] DOE Computational Science Graduate Fellowship. http://www.Krellinst.org/csgf/

[29] NIH Research Training Opportunities. http://www.grants1.nih.gov/training/extramural.htm

[30] American Mathematical Society Centennial Fellowships for 2003-2004 Invitation for Applications. http://www.ams.org/employment/centflyer.html

[31] GAMESS Web Portal. gridport.npaci.edu/gamess

[32] Buffalo Center for Excellence in Bioinformatics. http://www.bioinformatics.buffalo.edu/

[33] Private Communication from David W. Deerfield, Assistant Director, Pittsburgh Supercomputing Center; Director, Biomedical Initiative, December 11, 2002

[34] Private Communication from Julio Facelli, Director, Center for High Performance Computing, University of Utah, January 10 2003.

[35] Indiana Genomics Initiative. http://www.ingen.iu.edu/

[36] Branscomb report: four challenges for NSF. http://www.anirudh.net/courses/ae597/branscomb/node1.html

[37] fightAIDS@home. http://www.fightaidsathome.org/

[38] President's Information Technology Advisory Committee Report to the President. 1999. http://www.ccic.gov/ac/report/

[39] President's Information Technology Advisory Committee Panel on Transforming Health Care. 2001. "Transforming Health Care Through Technology." http://www.itrd.gov/pubs/pitac/index.html

[40] Programs of Excellence in Biomedical Computing (PRE-NPEBC). http://grants2.nih.gov/grants/guide/pa-files/PAR-00-102.html

Contributing Authors

Craig A. Stewart, Director, Research and Academic Computing, and Director, IT Core, Indiana Genomics Initiative, Indiana University. stewart@iu.edu

Ralph Z. Roskies, Professor of Physics, University of Pittsburgh, and Scientific Director, Pittsburgh Supercomputing Center. roskies@psc.edu

Shankar Subramaniam, Professor of Bioengineering, Chemistry and Biochemistry and Director, Bioinformatics Program, University of California San Diego. shankar@sdsc.edu

Vijay Agarwala, Director, High Performance Computing and Visualization, Penn State University. vijay@psu.edu

Brian D. Athey, Director, Michigan Center for Biological Information, University of Michigan. bleu@umich.edu

Philip Bourne, Director of Integrative Biosciences, San Diego Supercomputer Center. bourne@sdsc.edu

Beverly Clayton, Executive Director, Pittsburgh Supercomputing Center. clayton@psc.edu

Thom Dunning, Director, UT-ORNL Joint Institute for Computational Sciences. dunning@jics.utk.edu.

Julio Facelli, Director, Center for High Performance Computing, University of Utah. facelli@chpc.utah.edu

James R. Kasdorf, Director of Special Programs, Pittsburgh Supercomputing Center. kasdorf@psc.edu

David P. Moffett, Associate Vice President, Purdue University. dpm@purdue.edu

Richard Moore, Executive Director, National Partnership for Advanced Computational Infrastructure, San Diego Supercomputer Center. rlm@sdsc.edu

Radha Nandkumar, Senior Research Scientist, National Computational Science Alliance. radha@ncsa.uiuc.edu

Jan E. Odegard, Executive Director, Computer and Information Technology Institute, Rice University. odegard@rice.edu

Bruce Pitman, Vice Provost for Educational Technology, State University of New York, Buffalo. pitman@buffalo.edu

Richard Repasky, Bioinformatics Specialist, Indiana University. rrepasky@indiana.edu

Quentin F. Stout, Professor, Electrical Engineering and Computer Science, University of Michigan. qstout@eecs.umich.edu

Joseph F. Thompson, Director, NSF/Engineering Research Center for Computational Field Simulation, Mississippi State University. joe@erc.msstate.edu

The editorial assistance of Jan Holloway (holloway@indiana.edu) and Malinda Lingwall (mlingwal@indiana.edu), of the Communications and Planning Office, Office of the Vice President for Information Technology, Indiana University, is gratefully acknowledged.