National Centers for Biomedical Computing
Mid-Course Program Review Report
July 13, 2007

Introduction:
In response to a request from the Roadmap Implementation Coordinating Committee (RICC), an external panel was convened and charged to assess the status and progress of the National Centers for Biomedical Computing Initiative and to provide guidance for the future course of the program. The panel was asked to address 7 questions in their review and to make recommendations for future investments by NIH as part of the ongoing NIH Roadmap Initiative.

For many years, scientists supported by NIH have advanced the frontiers of computing and its methodological infrastructure. This work has provided valuable biomedical computing support for a variety of biomedical research areas and applications to medicine, as well as the informatics infrastructure important to both. The 1999 BISTI report (Botstein, et al. 1999) recognized the critical impact that computational science and infrastructure could make on the advancement of discovery in biomedical science. The four overarching recommendations of that report were: 1) to establish five to 20 National Programs of Excellence in Biomedical Computing, 2) to develop principles and best practices for the storage, curation, analysis and retrieval of information, 3) to support the development and adoption of software tools for biomedical computing and 4) to foster a scalable national computer infrastructure. The investment by NIH in the establishment of 7 National Centers for Biomedical Computing directly addresses the first and third recommendations made in the BISTI report.

The planning process for a Roadmap for Medical Research in the 21st Century (http://nihroadmap.nih.gov) also recognized the importance of developing sustainable infrastructure that spans multiple NIH Institutes and Centers for advancing biomedical computing. The National Centers for Biomedical Computing are poised to address several of the Roadmap themes: “New Pathways for Discovery” as part of its focus on new tools and methods, “Research Teams of the Future”, developing sites where training of cross disciplinary researchers takes place and “Re-engineering the Clinical Research Enterprise”, where NCBC advances in informatics and biomedical computing provide critical support to that research arena as well as computational tools that facilitate the delivery of its findings to medical environments.

This focus on support for biomedical computing is not new at NIH. For over four decades, the NIH has supported research and development (R&D) on mathematical and computational methods and systems crucial to the advancement of biomedical research. The panel was concerned to learn that there have been previous extramural programs at NIH to support biomedical computing centers that were subsequently abandoned. Thus, in the past, the NIH has failed to develop stable administrative structures at NIH to support this critical research area. It is of paramount importance that the NIH recognizes the massive scale of computational needs anticipated in the future of biomedical research and that these NCBCs, though necessary, are not sufficient. The panel sees sustained investment in these seven NCBCs as only the beginning of the investment required for the creation of a stable computational platform to sustain biomedical research.

The breadth of the Project Team for the NCBC program is very encouraging, as are the large number of Institutes and Centers represented in the Bioinformatics and Computational Biology Implementation Group. This must continue as the NCBC program and future programs related to it evolve, – not only for the sake of stability, but also to ensure that the biomedical computing centers have a comprehensive view of challenging R&D opportunities available to them and that they and other computational scientists are welcomed at frontiers

being pursued by many of the ICs.

Current funding provided to the NCBCs does not appear to permit substantial investments of effort outside of their basic R&D missions, although a number of very worthwhile cross-center activities are in place. The panel recommends the consideration of increasing the budget for the entire program, to facilitate more interactions between the ICs and NCBCs, and also to increase the impact of education and outreach programs.

The panel enthusiastically endorses the investment NIH has made in these critical areas through the NCBC program. The panel believes that these 7 centers, although young, are meeting the challenge of developing a national network of research centers in biomedical computing. They are effectively engaging in multidisciplinary team-based research, developing an extensive collection of useful software tools, providing training opportunities and promoting biomedical computing as a discipline through education and outreach to the community.

Charge to the panel:
"In response to a request from the Roadmap Implementation Coordinating Committee (RICC), the NCBC initiative will undergo a mid-course program review on June 11, 2007. An external panel will be convened to assess the status and progress of the NCBC initiative and to provide guidance for the future course of the program. The members of the review panel have been selected for their expertise in the diverse scientific areas impacted by the NCBCs and for their ability to provide objective input and advice. The chair of the review panel will responsible for writing a report summarizing the views and recommendations of the panel. The report will be sent to the Office of Portfolio Analysis and Strategic Initiatives (OPASI) and forwarded to the RICC. The RICC is scheduled to review the NCBC initiative on August 14, 2007."

"*The review panel will be asked to consider the following questions:*

1) To what extent does the vision and direction of the NCBC initiative promote biomedical computing?
2) In what ways has the NCBC initiative advanced biomedical computing?
3) Are the NCBCs interfacing appropriately?
4) What new collaborations have been formed through the NCBC initiative?
5) What new training opportunities have the centers provided?
6) What changes could make the program more effective in the future?
7) What lessons have been learned from the NCBC initiative that can guide future NIH efforts in biomedical computing?"

Executive Summary:
The panel concurred that a long-term investment in biomedical computing by the NIH is critically important to addressing the health care needs of the country. The panel recommends the following actions to ensure the success of this important effort.

1) Continue the support of biomedical computing as a key part of the NIH research portfolio over the long term. Computational biology, theoretical research and the development of robust software tools are critical to the understanding of biological processes and disease.

2) Begin developing a process to sustain and expand this effort now to anticipate support beyond the 10 year Roadmap funding horizon. The panel is concerned that the viability of this program and of biomedical computing in general depends on the relatively unstructured cooperative interactions of different NIH Institutes and Centers.

3) Focus research within and across the NCBC Centers on ambitious problems that other programs are unlikely to support, such as the development of cheaper, safer drugs, or new methods for multi-scale modeling

of biological processes. Consider partnership with industry, to address difficult problems of national importance, by taking advantage of longer more stable funding periods not possible within the biotech industry. Coordinate tool development with industry, since this is where the tools may have their biggest impact.

4) Continue to support the model of multidisciplinary, team based, collaborative research within the NCBCs. Extend the reach of the individual centers to collaborators outside the centers to increase the impact of the Centers on the community. Do not require interaction between NCBC centers where there is no obvious programmatic advantage. Continue the All Hands Meeting as an effective means for sharing best practices among the Centers and for fostering high impact Center-wide activities.

5) Develop an additional approach beyond the R01 and R21 collaborative grant program for developing and supporting collaborations with the Centers. The current peer-review system imposes delays in getting projects started and creates an additional administrative burden on the Centers to provide support to potential collaborators engaged in proposal submission. Support the NCBC Project Team in developing and implementing alternative approaches, such as streamlining the review process or providing funds for exploratory research, data collection or software design.

6) Develop a process to assess the impact of the software tools developed by the centers. Develop a simple assessment instrument to gauge how the software tools are advancing research and achieving widespread use within the community.

7) A focus on educating the next generation of computational scientists is critical to the success of biomedical computing as a discipline integrated within biomedical research. Continue to support the NCBCs and other programs in training multi-disciplinary researchers through collaborative research and outreach to the community. Leverage the efforts of the NCBCs and expand the educational programs designed to foster the education and training of computational biologists.

Answers to Questions:
The review panel was asked to address the following set of questions in their report. The panel's responses are based on their review of materials provided by program staff, additional material provided by the NCBC PIs, information provided by program staff and discussions within the panel.

1) To what extent does the vision and direction of the NCBC initiative promote biomedical computing?
The establishment of the seven NCBCs is an excellent start to what the panel hopes will be a long-term investment in biomedical computing. NIH has been less consistent than other federal agencies in recognizing the power of computing and promoting its multiple roles for advancing the biomedical sciences. This very visible program effectively reinforces the importance of biomedical computing to the research community. Moreover, by providing a longer planning horizon than is available in individual grants, or in the biotech industry, the NCBCs can propose projects that could not be accomplished otherwise. This program has also encouraged significant matching funds, and its visibility has helped the NCBCs to recruit and develop talent.

2) In what ways has the NCBC initiative advanced biomedical computing?
Despite the short time that these centers have been in place, many success stories are already evident. The NCBCs have developed widely available new software and web-based systems, created visibility for the discipline, and developed much-needed training programs. They have effectively paired experimentalists with computational biologists, and involved computer scientists and engineers in problems of biomedical interest. They have also generated a large number of relevant R01 and R21 collaborations. Besides their individual achievements, it is noteworthy that in such a short time, the NCBCs have collaborated as a whole on the categorization of biomedical ontologies by their degrees of acceptance in research communities, developing

software yellow pages and have begun to coordinate collaborative activities that center on Driving Biological Projects (DBPs).

3) Are the NCBCs interfacing appropriately?

The panel concurred that there is ample evidence of a considerable amount of appropriate trans-NCBC activity. For example, the July 2006 NIH Roadmap NCBC All Hands Meeting was particularly useful in this regard. These meetings should continue to be held and continue to include successful components such as the Building Bridges Compendium and Dissemination Events.

It is apparent that many productive interactions grew out of the All Hands Meeting, including the Software and Data Integration Working Group (SDIWG). The charter of the SDIWG is to promote software interoperability and data exchange, and to bring the collective knowledge and practices across the centers to wide publication. The SDIWG appears to have tapped a wellspring of endogenous enthusiasm in the Centers and has moved forward with leadership from within the centers to conduct regular electronic (and in some cases face-to-face) group conferencing to coordinate and direct the activities.

Three domains of SDIWG activity include: the Software Yellow Pages, Categorization of Scientific Ontologies, and Driving Biological Projects and Impact Working Group.

- The Yellow Pages project. Led by Ivo Dinov (CCB) and Daniel Rubin (NCBO) this project includes a NCBC iTools Prototype that supports a visualization interface for browsing and query of available NCBC tools.

- The Driving Biological Project Interactions, led by Andrea Califano (MAGNet) and Brian Athey (NCIBI), focuses on determining the research community needs for tools, data and methodologies for the analysis of cellular networks, with a focus on their use in complex trait and biological process analysis in current and future Driving Biological Projects. Currently, this activity is represented in a searchable graphical Interactome of potential DBP interactions among NCBCs.

- The Scientific Ontologies Group led by: Zak Kohane (i2b2), Suzi Lewis and Mark Musen (NCBO) aims to create a succinct categorization of available ontologies and terminologies. As a result, the particularly useful contribution of this effort has been the evaluation and categorization of existing biological ontologies into three groups, as (1) Fully Endorsed, (2) Promising and used with some reservations, or (3) Not quite ready for use, underdevelopment and for use under protest. An interactive table of these results is available at http://www.berkeleybop.org/sowg/table.cgi.

These activities show an appropriate focus on the tool development mission of the NCBCs. It is also encouraging that the ongoing interactions make programmatic sense in creating technical and biological synergies. *The panel recommends encouraging the NCBCs to continue to apply this metric and not develop forced interactions and collaborations that do not make programmatic sense. At the same time, the panel recommends encouraging the NCBC Project Team to broaden their search for additional potentially synergistic interactions outside the scope of the NCBCs where it makes programmatic sense, for example with P41s and with other agencies.*

4) What new collaborations have been formed through the NCBC initiative?

All seven of the NCBC centers have developed new collaborations as a result of the NCBC initiative. Those collaborations include interactions within the individual centers, among the different centers, and include a wide array of collaborations with new entities outside of the centers. (Specific examples are well detailed in the annual progress reports and summary documents provided by the Center Directors.) Within individual centers, collaborations have been forged across the individual core components, seminar series have been established and nascent links between biological, bio-informational, and computational components have been expanded and solidified, even beyond the activities proposed in the original applications.

- All NCBCs have cooperatively engaged in scientific and technical discussions of common interests under the auspices of the NIH Software and Data Integration Working Group (SDIWG) as described above. Other examples include the supplementary postdoctoral opportunity that helps bridge across centers, commonly attended conferences such as the DREAM (Dialogue on Reverse Engineering Assessment Methods Workshop: http://www.iscb.org/events/event_data.php?454), and examples of associations that have been developed with other NIH sponsored networks (e.g., caBIG, BIRN, and CTSA).

- Probably the most diverse set of new collaborations are those with entities from outside the NCBC initiative. Such collaborations have been spawned by the inclusion of DBPs (and the new round of DBPs that are being considered), the development of the R01 and R21 "Collaborating with NCBC" initiatives, and by the increasing visibility/capability of the individual centers. Examples include collaborations with industry, vendors, academia, hospitals, and foreign institutions, international, and healthcare organizations. The lists provided by the centers are truly impressive.

- The workshops, websites, and other dissemination efforts developed by the centers are serving to bring together diverse groups of people who might not otherwise interact. These efforts are catalyzing interactions and are likely to lead to new collaborations.

- The review committee believes that all centers have developed significant collaborative interactions. One useful tool used to build these interactions and to get tools out to the community is the Driving Biological Projects. However, the panel rose the question of what the optimal size/number of DBPs should be since they also place a burden on the other center components; affording additional flexibility may be warranted. The R01 and R21 initiatives help build collaborative activities with investigators "outside" the centers; but the peer review process (first at the NCBC and then the NIH) may unnecessarily delay the start of meritorious projects. *The panel recommends some form of flexible short term funding to jump start new collaborations, and/or funds to hire postdocs who bridge centers. The panel also recommends facilitating interactions between NCBC centers, (where it makes programmatic sense) with other P41s, caBIG, BIRN, Virtual Cell, etc - or other agencies.*

5) What new training opportunities have the centers provided?

Training is central to creating the next generation of multi-disciplinary scientists and to broadening the skills of existing scientists to pursue cross-disciplinary collaborations that are advancing the frontiers of biomedical sciences and their applications today. For recipients already committed to careers in this area, the training is of immediate value. For other students, exposure to major nationally supported centers where such multi-disciplinary research is thriving may become an important element in their choice of careers and in turn, a commitment to the very substantial preparation required to become leaders in this area.

The NCBC centers have provided a variety of training and educational opportunities to members of the centers, affiliated groups, and the broader biomedical scientific community. Most of these are at the post-doctoral level and involve special tutorials, workshops or meetings centered on topics of interest and research strength in a particular center. Training activities are not coordinated across the centers presently, and there is some debate as to whether coordination would be beneficial.

Some examples of training activities are listed briefly below:

- *Pre-college – graduate school training*: The CCB has hosted numerous visits by groups of pre-college students. It has provided both graduate and undergraduate courses for students at UCLA, offering research experience to both. I2b2 hosts a Summer Scholars program for undergraduates across the country, which includes both education and research projects. NCIBI participates in the training of 10 graduate students in the University of Michigan's Bioinformatics Graduate Program.

- *Postdoctoral training*: Most NCBCs are actively involved in post-doctoral training. Five of the seven NCBCs have close associations with NLM training programs at their universities. Simbios has created the "Simbios Distinguished Post-Doctoral Program." I2b2's post-doctoral program provides access to its leading investigators and DBP data resources. NCIBI would like to increase their post-doctoral fellowships to 3 years, to aid recruiting. NCBO, which supports two post-doctoral trainees and participates in training some from other programs, urges increasing budgets for training. CCB reports training over a dozen post-doctoral fellows and young investigators, and has a program for visiting scholars.

- *Cross training of established investigators*: Cross-training occurs as scientists collaborate at the Centers, in a variety of meetings, and via other resources. NCIBI produces a year-around weekly seminar series, "Tools and Technologies", in which NCBO, MAGNet, and others also participate. These also are broadcast live over the Internet via streaming video/audio, and are archived for later use. It also produces web-based interactive training and educational programs. I2b2's "Grand Rounds" seminars, to educate scientists about biomedical computing ingredients for discovery research in academic health-care centers, also are available via streaming video available from i2b2's site. CCB has organized three well-attended (> 100 each) international workshops. Simbios is providing a summer short-course and workshop on the use of some of their software. MAGNet's second retreat, this April, had about 150 attendees, and it has been recruited by a related network in Europe, ENFIN, to produce a joint conference next year.

6) What changes could make the program more effective in the future?

The panel was impressed with what the NCBCs have accomplished so far on a relatively small amount of funding. Limiting discussion to the constraint we were given of continuing the Centers for an additional five years at the current funding level, several suggestions emerged from the panel's discussions. These suggestions mirror the recommendations in the executive summary and provide additional detail and rationale for the recommendations.

- Centers should be given sufficient programmatic flexibility to jump-start new projects. For example, a joint post-doctoral program emerged from the All Hands Meeting brainstorm session. While this is a new positive idea *across* centers, there is little doubt that if each center's funds were less constrained and thereby they were given the opportunity to think broadly, many such ideas would emerge and be executed *within* each center.

- More could be done to encourage appropriate collaborations of other non-Center investigators with the Centers, while at the same time avoiding unnecessary collaborations. For example, simplifying the collaborative grant programs for the centers and streamlining the review process may be appropriate. The R01 and R21 programs, while getting strong responsiveness, require a conventional peer-review process that limits the risk-taking needed to quickly jump-start new ideas. It also appears to be particularly burdensome for the centers to "help" to write multiple R01 and R21 submissions most of which will never be funded. Perhaps a mechanism can be established whereby some of the IC funds that would otherwise go toward those R01 and R21 programs could be managed by the NIH NCBC Project Team for the ICs and be assigned to Centers for use in collaboration with others through a streamlined review process. Fundamentally, the development of the NIH NCBC Project Team is a terrific development that can be leveraged to assist the Centers in promoting the use of Center tools and collaborations. Although the Project Team has done a great job of stimulating new projects, they could be encouraged to take a more active role in supporting those projects. The shared ontology database is a good example where the NCBC Project Team stimulated new positive inter-Center work, but where additional support seems appropriate to fully execute and maintain it rather than pressuring the Center Directors to do it without new support.

- Encourage the NCBC Project Team to engage in developing an assessment program for software tools. Just as the Categorization of Ontologies performed by the NCBCs, stimulated by the Project Team, is seen generally as a very positive outcome, an objective method of assessment of the usefulness (effectiveness, ease of use, desirability, efficiency) of new biomedical software tools would be a great contribution. The panel believes the Project Team could develop and lead and/or subcontract a program to develop such an assessment instrument that would account for all the spectra that make a software tool a success (user base, computational efficiency, human-computer interaction, scientific papers/achievements using the software, specific niche it fills and why, etc.).

- Encourage the Project Team to work with the NCBCs to enhance the dissemination of their successful software development efforts. A variety of mechanisms could be used to help investigators adopt and use the software the NCBCs are creating. Websites that are deep enough to help investigators learn to use the software, rather than just download it, would be worth the effort. The NCBCs can lead this effort, since they are the experts in their software, however additional resources should be devoted to this effort would allow the NCBCs to focus on their strengths in building new tools.

- Consider how to leverage the NCBCs to enable training programs in computational biology. The NLM Medical Informatics Training Programs have certainly contributed to this domain and to the ability of the Centers to pursue their research agendas. Perhaps NIGMS, in collaboration with the NLM can establish additional training programs by leveraging the NLM model. These programs could be located outside the NCBCs, ideally linked with them expanding the new joint post-doctoral program.

- Adopt methods to improve the cross-disciplinary collaboration experience, among NCBC PIs, the Project Team, and others, perhaps through social networking approaches. This topic could be discussed at the next All-Hands Meeting.

7) What lessons have been learned from the NCBC initiative that can guide future NIH efforts in biomedical computing?

The panel concurred that the development of a successful NCBC is a very complex and challenging task. Each of the Centers has encountered different obstacles that can slow, or sometime prevent the achievement of the goals of the program. The panel identified several consistent themes that illuminate these challenges.

- Carrying out true interdisciplinary work is hard enough when only two disciplines are involved; it is much more so when trying to achieve this with four or more disciplines (i.e. computer science, mathematics, engineering and biomedicine). NCBC programs will take time to fully develop.

- The concept of “Driving Biological Problems” is outstanding and critical for ensuring focus and utility. The best work is done when experimental biologists and computational researchers are equal partners and there is a tight feedback loop between experimental design and computational analysis.

- Co-location of various disciplines is a big advantage, especially during the initial period of program development. Geographical distribution within a single NCBC can inhibit cross-disciplinary exchange and increases administration complexity, but has the advantage of increasing the range of expertise available to apply to a specific problem. The importance of social networking has been under-appreciated and could help ameliorate these issues.

- Interactions among NCBCs, and with other NIH biomedical computing initiatives, may be very important for success but unnatural or artificial interfaces should not be forced just for the sake of interacting. A commitment to interoperability will enhance the success of interactions.

- While it may be sensible in some cases to attempt a broad and diversified portfolio of projects and applications, this approach may dilute efforts and should not be done at the expense of focus. Increased focus will increase impact.

- Because the NCBCs have such great potential, it is easy for unrealistic expectations and "scope creep" to develop, leading to a serious mismatch between objectives and the resources available to achieve them.

- Partnerships with industry (Pharma, other biosciences, translational medicine and IT) should be encouraged and supported. Open source software combined with the ability to license applications for commercial use may help leverage the work of the Centers and increase their impact.

- Sustainability plans should be formulated now and the NCBC initiative should evolve from a Roadmap activity to one that achieves ongoing and cross-ICD support (with competitive renewal). Partnerships with other Federal agencies (e.g. NSF) should be explored.

Vision and Grand Challenges for the Future of Biomedical Computing

The panel was encouraged to think broadly about a future vision for biomedical computing. The following sections summarize the panel's thoughts on this important issue.

Historically, some important advances in mathematics have been developed to meet challenges on the frontiers of research in the physical sciences. Today the rapidly advancing frontiers of biology and medicine invite similar opportunities in mathematics, statistics, and computational methodologies. There is a need to evolve productive environments and approaches for inspiring and facilitating cross-disciplinary collaborative research. There is also a need to increase our investments in training scientists who have substantial graduate-level training in both mathematical/computational disciplines and in a biological/medical discipline. Such scientists exist, but we need many more. The NCBCs can contribute in a variety of ways. They can explore and evaluate new approaches to improving cross-disciplinary techniques, research environments, and cultures. They can provide cross-disciplinary research opportunities for students from upper-division through post-doctoral levels.

The existing set of NCBCs is an extremely valuable start toward a maintained focus on biomedical computing at the NIH. While the centers cover a broad range of topics, at present they focus mostly on problems that can be handled with today's computers running algorithms that are not too far removed from current standard practice. Given the nature of the original NCBC proposal guidelines and the review process, this is not terribly surprising and is an appropriate beginning. Reviewers are naturally cautious at the outset of a new program, and pre-existing NIH funding was required for the candidate Driving Biological Projects. This set of constraints has produced centers that to some extent are plucking the low hanging fruit; that is, they are working mostly in areas that are already well established as part of computational biomedicine, and that have clearly discernible connections to the clinic in the here and now (imaging, genomics, high-throughput data handling, etc.).

From the standpoint of broad-ranging future needs, however, the program must expand to embrace more innovative cutting edge software and hardware development. Biology and medicine still lag far behind other scientific disciplines in the use of large-scale high performance computing (e.g., physics, chemistry, meteorology, climatology). In some circles the perception remains that biology and medicine just "aren't ready" for large-scale computing. This attitude is probably due to history, ignorance, and the admitted complexities of biomedical research. Even in those areas where large-scale computation is already necessary and established in biology (e.g., quantum mechanical and molecular mechanical simulations) the connection to classical chemistry (as opposed to biology) is strong and the overall scope remains narrow. But "waiting" for additional areas of biomedicine to be "ready" for large-scale computing will leave the US in a perpetual "catch-

up" position compared to other nations, and will dramatically delay breakthrough advances in predictive personalized health care.

To accelerate innovation, a new mechanism for collaborative interactions should be developed without an obligatory need for traditional, conservative, and lengthy R01/R21 review (as outlined in preceding sections). To make reasonable judgments about what projects are appropriate and productive, recommendations can be obtained from each Center's External Advisory Committee. Software development components of all interactions should strongly encourage innovation and maintain a long-term outlook on what needs to be achieved for the future, rather than remaining steadfastly centered on short-term applications. Current applications are important, of course, but the strategic outlook of the Centers should include increasing attention to innovative long-term goals beyond the obvious short-term gains.

Grand Challenges in Biomedical Computing

The following are examples of unsolved problems in the biomedical sciences that should be pursued as part of a sustained focus on biomedical computing. None of the current NCBCs are addressing these problems at present and while this does not represent an exhaustive list, these examples illustrate the range of problems that require the type of research infrastructure being developed in the NCBCs.

Quantitative multiscale modeling is one example of an outstanding grand challenge in computational biomedicine. While many existing centers and investigators are already working on different aspects of the problem, no large-scale concerted effort has been created to date. Part of the problem is the need for new mathematical treatments of complex biological systems, and another is the need for new algorithms and approaches to "mesoscale" (cellular to tissue) problems in physiology and medicine. The eventual solution will require sophisticated and long-term software development that couples stochastic and continuous methods applied to problems of very large scale, and therefore will also require careful attention to hardware design and implementation. Despite a petascale computing initiative presently underway at the NSF, no federal agency has yet truly come to grips with the real software development requirements in any discipline. The NIH, in fact, remains conspicuously absent from such initiatives, and even within the NSF funding for petascale applications in biology, is conspicuously absent. One can view the present situation as a problem or an opportunity. To make it the latter the NIH will have to assume a leading role and push the frontier in collaboration with other federal funding agencies, and the NCBCs of the future are an obvious possible venue.

Predictive personalized medicine of the future will encompass routine knowledge of each individual's genome and proteome, high resolution real-time non-invasive imaging, and individualized modeling and simulation that will encompass molecular to cellular and tissue scales. For example, when a patient presents to the emergency room with an acute onset of heart attack or stroke, treatment decisions in the future will depend on massive amounts of individualized quantitative data coupled to quantitative predictive modeling of the evolving event. Thus the future of personalized interventional health care is similar to what is now taken for granted (but is still under active development) with storm forecasting, which is based on massive real time data collection and continually improving supercomputer models that likely will grow into first-case petascale applications. To accelerate this vision for the future of health care, it is incumbent on the NIH to break through traditional boundaries. New trans-agency funding mechanisms and long-term support must be developed to foster innovative breakthrough thinking in the NCBCs of the future, as well as new additional biomedical computing initiatives that have yet to be envisioned.

For patients with diseases, such as HIV and some kinds of cancer, which require long-term therapy during which there are changes in the patient's basic status (e.g. viability of immune, hematopoietic systems) and the treatment target (e.g. mutations conferring drug resistance), ***individualized optimal therapy*** will require continual modifications in the types and doses of therapies administered, and in their timing. The benefits to patients and the economy could be substantial. To achieve this, sophisticated techniques must be developed for the weighted and possibly model-based integration of the patient's information (e.g. functional, biochemical,

imaging). The resulting computational support should be physician-friendly and responsive to suggestions for improvement. NCBCs or other programs with the ability to provide stable support for the long-term cross-disciplinary R&D that is required here are essential. Recommendations made at the end of the preceding paragraph apply here as well.

Methods for answering complex queries over a continuously updatable semantic web is at the center of a new computational approach to integrating literature searches with methods of experimental science, and is stimulating a new paradigm of "active" computational scientific inquiry. The NCBCs have the potential to accelerate these developments by tying them to modeling, visualization and ontologically-based argumentation, helping researchers pose problems in entirely new ways, checking their conjectures "on the fly" against a synthesis of existing knowledge.

A critical missing component and challenge is how to combine the visual with the logical, simulation (model-based what-if), and statistical argumentation typical of such scientific reasoning – while also taking into account the ambiguity, risk and uncertainty inherent in many of the arguments. Developing effective tools which suggest new questions automatically from the results of such "computational models of inquiry" is a major challenge for the centers, which may increase the likelihood of "seeding' new ideas and scientific ventures in other bioscience laboratories. The development of such a "biosciences semantic web" will be essential to overcome the current "traditional discipline silos" which are the basis of most knowledge organization, both in the literature and even in the most flexible of current computational search systems.

An even more ambitious challenge is to see if specific experimental designs can be suggested automatically on the basis of a systematic search over current experimental conditions and results that have yielded conflicting interpretations under various model assumptions. This "grand challenge" involves cognitive science and learning in addition to the modeling, simulation, analysis and interpretation typical of current exploratory bioscience.

Working towards developing cheaper, safer drugs. The process of drug discovery, validation, development, testing and submission is long and complex and expensive. However, there are a number of points along the pipeline where better computation, the introduction of modeling, and more effective information retrieval and analysis could have a large impact in decreasing the time and the cost of preclinical and clinical development. Learning from a large body of prior experimental data about how to analyze interpretations of experimental results is a challenge that goes far beyond what current (largely static) ontologies and parameterized sets of models can handle. Overcoming this barrier, for even a restricted class of design problems, could generate automatic suggestions for designs within a "standard protocol" of experimental design and discovery, as is the current practice in drug design. To achieve this goal, the integrated analysis of meso or multi-scale (molecule-cell-tissue-organ, organism, population-environment system), and heterogeneous models will be required. However, these methods, once developed, could lead eventually to automatic search and discovery heuristics that could yield "cheap, effective drugs and procedures with a minimal number of side-effects."

Creating comprehensive, integrated computational modeling/statistical/information systems and related databases for major biomedical sectors. For example, the software platforms GENESIS, NEURON, and other computational modeling systems serving research in the neurosciences have evolved over the years into critical tools for this community. They provide information related to neuronal model elements and the software that facilitates linking these elements into a model. For one of these programs, an active users' group has made many contributions over the years, e.g. models and parameters for a variety of channels, compartmental models of various neurons.

Are there other important biomedical sectors that might benefit from, and do not as yet have, programs of this general type? – e.g. oncology, virology/HIV, genetics? When we examine the programs and consider overall

computational biology tools that might be applicable to research in those areas, can we identify methodological elements that perhaps should be added and integrated – e.g. the ability to embed stochastic models within a deterministic system?

Establishing a resource for advancing and maintaining software valuable for biomedical research and applications. Some non-commercial software may be of substantial use to various biomedical researchers long after the grant that funded its creation, or its creator, have ceased to exist. As time goes on, the operating systems and other elements for using the software change. If the community agrees that the software is well worth maintaining, and improved in specific ways, can NIH provide a resource for doing this?

Training the next generation of computational scientists: The panel recognized that the greatest grand challenge of all is recruiting and educating future computational scientists. The type of complex inter- and multi-disciplinary research, which is the hallmark of the NCBCs, will require long-term educational preparation by future members of such centers, including strong High School and Undergraduate studies in both the formal and analytical aspects of research methodology, as well as experimental science and/or engineering, and biological knowledge. This is a tall order for most people, but successful researchers in the future (and even now) often require the equivalent of two undergraduate degrees to prepare them for successful graduate studies and post-doctoral work in any of the computational biosciences. With this in mind, *the panel recommends that the NIH suggest to colleges and universities that they encourage such joint programs of study so future researchers can be as well prepared for deploying analytical and computational approaches to bioscience as they are for wet-lab experiments.*

The preparation needed for the discipline of computational biology is very different than that for bioinformatics. The former requires much more depth in mathematical modeling, algorithms, and simulation, and theoretical biology, whereas the latter tends to focus on large-scale data mining analysis and evaluation of experimental biological data. Both are needed in most NCBC projects, in addition to imaging and visualization methods, which tend to come from yet other disciplines: Computer Science, Cognitive Science, or Engineering. The appreciation of the complementary roles these different approaches play in experimental design and its implementation within the context of systems biology and the emerging semantic web of information is a central educational need for all the NCBCs. In addition, as we move increasingly into large-scale epidemiological studies, researchers will also have to appreciate the nuances of population modeling and study design, which presents yet another set of methodological challenges. In some sense the NCBCs could help develop a new “systems biology ontology” which has not yet emerged in most other sciences, which still tend to adhere to a physics-based paradigm of science.

Postdoctoral trainees within the NCBCs may be uniquely well positioned to make valuable contributions to research in the computational biosciences which require an exceptionally broad set of hybrid models and theories combining mathematical, logical (including temporal), and linguistic/semantic components to explain their heterogeneous data and knowledge over a wide range of scales, and levels of abstraction. Training experiences within the NCBCs may prepare these individuals to use ontologies in the representation of biological concepts and through “meta-experimentation” of an entirely new computational kind, develop methods to use ontologies in problem solving and biological experiment design, analysis, and interpretation.

The present “portfolio” of NCBC centers is likely to change over the years as new problems, methods, and technologies are developed. The only way for the NCBCs to yield productive long-term researchers who can themselves train a new generation of interdisciplinary investigators is to ensure that they disseminate and explain their results and produce educational materials about their novel computational and biological contributions. Connecting the NCBC centers to existing NIH training programs is one approach, though alternatively, once they are sufficiently mature, the NCBCs could develop their own stand-alone educational program(s).

On the practical, technology infrastructure side, the NCBCs are already developing systems and software for advanced modeling and analysis, but these frequently require highly sophisticated prior knowledge of the model assumptions, empirical data constraints and their application within a complex set of software tools. During the past 40 years there have been many misapplications of sophisticated statistical, graphics and image interpretation, mathematical modeling, simulation, data mining programs, and languages. In the future, the more heterogeneous and complex combinations of methods and software will make it yet more difficult for investigators to apply the results of NCBC software and evaluate their results. An advanced educational objective of the centers could involve educating others in the foundational assumptions behind the software, and the "rules of practical application" to specific examples with successful outcomes as well those problems that prove too difficult to solve with the current tools.

The NCBCs could develop an outstanding educational tool by providing a critical perspective on the application of their computational methods to specific biomedical problems. This perspective would be invaluable as a means to educate the next generation of investigators in how questions can be posed differently, problems reformulated, and different models chosen or sought as result of a careful documentation of NCBC "successes and failures". In contrast, training investigators to solve currently defined problems with current methods and technologies is merely an incremental component of center activities. A unique opportunity exists within the current and future NCBCs to create an educational environment for the next generation of computational scientists that can truly address the challenges of solving the most difficult problems in biomedical science now and in the future.

References:

NIH Biomedical Information Science and Technology Initiative, 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. (www.nih.gov/about/director/060399.htm), June 1999.

NIH Roadmap for Medical Research [Online] http://nihroadmap.nih.gov

Materials provided to the Panel:

BISTI Report – This report was written in June 1999 by the ad-hoc NIH Working Group on Biomedical Computing to the NIH Director.
RFA-RM-04-022 – The request for applications which established the NCBCs.
NCBC Descriptions – Brief descriptions with links to each center's website.
PAR-07-249 – The program announcement for collaborative R01 grants with the NCBCs.
PAR-07-250 – The program announcement for collaborative R21 grants with the NCBCs.
NCBC AHM Report – A report of the July 2006 All Hands Meeting
NCBC Management Plan – This is the organizational chart for the NIH management of the NCBC program.
The Software and Data Integration Working Group (SDIWG)
http://namic.org/Wiki/index.php/SDIWG:Software_and_Data_Integration_Working_Group

Biomedical Computation Review: http://biomedicalcomputationreview.org/

Review Panel:

Gwen Jacobs, Ph.D., Chair
Professor of Neuroscience
Asst. CIO & Director of Academic Computing
Cell Biology and Neuroscience
Montana State University
P.O. Box 173148
Bozeman, MT 59717
Phone: 406-994-5120
gwen@cns.montana.edu

Carol Newton, M.D., Ph.D.
Professor
Department of Biomathematics
School of Medicine
University of California, Los Angeles
AV-155 CHS 176620
Los Angeles, CA 90095
Phone: 310-825-5800
cnewton@biomath.medsch.ucla.edu

Mark Boguski, M.D., Ph.D.
Vice President and Global Head
Division of Genome and Proteome Sciences
Novartis Institute for Biomedical Research, Inc.
250 Massachusetts Avenue
Cambridge, MA 02139
Phone: 617-871-8000
mboguski@safe-mail.net

Ralph Roskies, Ph.D.
Co-Scientific Director
Pittsburgh Supercomputing Center
Professor of Physics
University of Pittsburgh
4400 Fifth Street
Pittsburgh, PA 15213
Phone: 412-268-4960
roskies@psc.edu

Craig Jordan, Ph.D.
Director
Division of Extramural Activities
National Institute on Deafness and Other Communication Disorders
6120 Executive Boulevard, MSC 7180
Rockville, MD 20892
Phone: 301-496-8693
jordanc@mail.nih.gov

James Schwaber, Ph.D.
Director, Daniel Baugh Institute for Functional Genomics/Computational Biology
Department of Pathology, Anatomy and Cell Biology
Thomas Jefferson University
1020 Locust Street
Room 381, Jefferson Alumni Hall
Philadelphia, PA 19107
Phone: 215-503-7823
james.schwaber@mail.dbi.tju.edu

Casimir Kulikowski, Ph.D.
Board of Governors Professor of Computer Science
Rutgers, The State University of New Jersey
110 Frelinghuysen Road
Room 323 Computer Science
Piscataway, NJ 08855
Phone: 732-445-2006
kulikows@cs.rutgers.edu

Jonathan Silverstein, M.D., M.S., F.A.C.S.
Assistant Professor of Surgery and Radiology
Director, The University of Chicago Hospitals' Center for Clinical Information
Scientific Director, Chicago Biomedical Consortium
Associate Director, Computation Institute
The University of Chicago
Research Institutes, Room 405
5640 South Ellis Avenue
Chicago, IL 60637
Phone: 773-834-8200
jcs@uchicago.edu

Joel Stiles, M.D., Ph.D.
Director
Center for Quantitative Biological Simulation
Pittsburgh Supercomputing Center
Associate Professor
Mellon College of Science
Carnegie Mellon University
300 South Craig, Room 406
Pittsburgh, PA 15213
Phone: 412-268-4786
stiles@psc.edu

NIH Roadmap Bioinformatics and Computational Biology Co-Chairs

Jeremy Berg, Ph.D.
Director
National Institute of General Medical Sciences
Phone: 301-594-2172
bergj@mail.nih.gov

Donald A.B. Lindberg, M.D.
Director
National Library of Medicine
Phone: 301-496-6221
lindbergd@mail.nih.gov

National Centers for Biomedical Computing Project Team

John Whitmarsh, *Leader*
NIGMS
Michael Ackerman
NLM
Carol Bean
NHLBI
Zohara Cohen
NIBIB
Milton Corn
NLM
Valentina Di Francesco
NIAID
Greg Farber
NCRR
Valerie Florance
NLM
Dan Gallahan
NCI
Peter Good
NHGRI
John Haller
NIBIB
Michael Huerta
NIMH
Don Jenkins
NLM
Jennie Larkin
NHLBI
Peter Lyster
NIGMS
Grace Peng
NIBIB
Salvatore Sechi
NIDDK
Karen Skinner
NIDA
Jen Villani
NIGMS

Office of Portfolio Analysis and Strategic Initiatives Staff

Rebecca Lipsitz
OD
Anne Menkens
OD
Krista Pietrangelo
OD