REPORT OF THE PSI:BIOLOGY ADVISORY COMMITTEE – DECEMBER 13-14, 2012

The Protein Structure Initiative (PSI) has undertaken high-throughput determination of protein structures, resulting in the deposition in the Protein Data Bank of more than 5,000 macromolecular structures since 2000. During the current PSI:Biology phase, highly organized networks of investigators are successfully applying the high-throughput paradigm to study a broad range of significant biological and biomedical problems. PSI:Biology actively engages the scientific community, supporting community-nominated targets for structure determination and making resources widely available through the Resource Centers.

The third annual meeting has highlighted the breadth and excitement of the biology now being enabled by the PSI:Biology Network, which consists of the Centers for High-Throughput Structure Determination, Centers for Membrane Protein Structure Determination, Consortia for High-Throughput-Enabled Structural Biology Partnerships, and Resource Centers (the Structural Biology Knowledgebase and Materials Repository). As evidenced by presentations at the meeting and a 395-page progress report prepared by the Centers, the Advisory Committee feels that cohesion among the Centers is expanding and the scientific impact of the program is greatly enhanced through the diverse Biology Partnerships and the community targets. Now involving over 600 investigators, the penetration of the PSI:Biology science is impressive. As a whole, the PSI:Biology Network is working well towards the common goal of advancing biology through high-throughput structural biology.

Below we summarize key issues that were raised during the annual meeting and provide feedback intended to help ensure success of the program. In particular, as PSI:Biology reaches its mid-term, it is time to highlight the accomplishments and tell the stories of PSI:Biology.

A. CENTERS AND PARTNERSHIPS – THE NETWORK

1. *Centers for Membrane Protein Structure Determination:* The Membrane Protein Centers and their Partners continue to make excellent progress, with an admirable level of cooperation and synergy. It is a crown jewel of the program, representing a most impressive set of accomplishments. It goes a long way toward justifying and demonstrating the strength of PSI:Biology.
2. *Centers for High-Throughput Structure Determination:* The HTP Centers continue to make technological advances and maintain an impressive array of collaborative partnerships. The Centers have improved interactions with their Biology Partners with valuable lessons learned. However, the strong links between the four Centers that contributed greatly to the success of earlier PSI phases may have weakened somewhat. The Committee encourages the Centers to strengthen communication for optimal synergies, not only for technological advancements but also for lessons learned about how to optimize performance of partnerships.
3. *Biology Partners:* The progress of the Biological Partners is generally excellent with a broad range of projects relevant to all types of diseases and biological processes. The diversity allows the PSI:Biology to be relevant to the broader biology research community and is a strength of the program. Different Biological Partners seem to be doing a nice job of moving into different niches (protein design, bioinorganic chemistry, microbiome, etc.), which is a positive development.
4. *Optimizing the PSI: Biology Network:*
   a) Workshops and meeting sessions are useful for sharing technologies and recent developments and for discussion of issues and challenges. The open 1-day Technologies Meeting that preceded the annual meeting was a success. The forum displayed far-reaching new experimental and computational ap-

proaches from the Centers, the Partners and the public at large. Plans for similar workshops to be organized and sponsored where possible at national and international meetings are underway and are encouraged. The SBKB all-hands meeting (summer) and business meeting (Data Management Meeting at the PSI meeting) are good additions. The latter is particularly important to address the data management and computational challenges facing several Centers. A business meeting on metrics is also recommended for next year (see IMPACT section below).

b) Lessons learned from interacting with the diverse Biology Partners should be shared. A few observations from the meeting include: (i) In some cases, handling optimization of constructs and complexes, etc., is best done by the Biology Partner's lab, and in others, by the HTP Center with refinement and optimization of the pipelines to fit the project; (ii) Forming meaningful partnerships is challenging and may be most effective when a particular contact at the Center is responsible. Successful partnerships include technique dispersal to the Partner through hands-on training (workshops, etc.) and this is encouraged. Biologists are also finding themselves free to pursue more experiments when reagents are developed in the Centers; and (iii) Visits of investigators from the biology team to the Center, and vice versa, can facilitate the partnership.

5. *Community Nominated Targets (CNTs):* Community-based nominations provide good public outreach and are central to certain projects, as exemplified by the Baker project in the NESG and the outreach program of the Mitochondrial Protein Group. Since the role of community nominations varies from Center to Center, each needs to meet defined selection criteria and have good support from the nominating group to be successful. Each Center should be granted considerable discretion in finding the best balance for their Center. Overall, we recommend that:

   a) Each Center should have a nomination mechanism with clearly defined selection criteria. The rules of engagement may vary from project to project and change with time, but need to be clearly articulated up-front.

   b) The nominating groups should be required to provide scientific justification and add value beyond just listing potential targets. There should be enough information provided to insure the nominator can perform the backup experiments necessary to leverage the new structural information. The Centers may give email updates on the status or make clones and have relevant ligands on hand as some Centers are doing.

   c) To balance services provided to CNT's and Biology Partners, the distribution between CNT and Partner projects should be flexible and at the discretion of each Center, with endorsement from the NIGMS Staff. Center-solicited community nominated targets can be more successful since they are pre-filtered and pre-contacted. To avoid spreading the Centers too thin, we recommend against adding many more Partners to the Program. Centers should be reminded that they will ultimately be evaluated on the impact of their work, and then encouraged to maximize that as they see fit.

6. *Technological Advances:* The Centers continue to enhance their biochemical techniques and instrumentation, introducing a number of important innovations. The advances described by the Centers for mammalian protein expression were quite varied and impressive. Advances in discovering protein complexes through co-IP, cross-linking and bioinformatics were also varied and impressive. Efforts to uncover post-translational modifications were interesting but are currently more anecdotal than systematic. Other nice technical advances and contributions being made include the Honig Nature paper on the protein interactome, new lentiviral expression vectors, crystal engineering, nanobodies, and contributions to CASP-like public contests. The groups involved might benefit from a mechanism to share the techniques developed in various places.

7. *Annual Progress Reports:* The progress reports from the Centers and Partners have benefitted tremendously from the carefully structured format implemented by Program. The information requested for the annual reports was excellent, balancing progress with challenges underway. We suggest a few areas for improvement next year: (i) The copious duplication of the reports (especially between the reports from the HTP Centers and the Partner reports) should be avoided; (ii) The page limit should be strictly followed; (iii) The reports should be provided as one PDF file, but should also be made available on the website in such a way that all sections can easily be viewed per Center and per Partner; (iv) The reports should add a section on training impact, including the career paths of trainees.

B. PSI STRUCTURAL BIOLOGY KNOWLEDGEBASE AND MATERIALS REPOSITORY

1. *Structural Biology Knowledgebase:* The SBKB is impressive. The Web site has a wealth of information and has served effectively as the public face of the PSI:Biology, with significantly increased numbers of web hits. The team is doing a great job of raising visibility and utility, increasing participation and establishing liaisons with publishers such as Thompson Reuters. Further engagement with organizations such as the United Press to reach wider audience is encouraged. The SBKB has improved the usability of its Web site. It can still benefit from user studies to gather functional requirements and scientific use cases from the broader biology community to improve the utility and usability even further. The SBKB has done an excellent job in curating structural information and capturing data resulting from the collaborative efforts. We make the following suggestions for further improving the content of SBKB:
   a) Add a "Chemical Information" hub, including for instance: (i) Chemical information on bound compounds (e.g., substrates like ATP, etc.); (ii) Information on, or electron density of, unexpected bound ligands, a gift from the expression system or crystallization condition; (iii) Information on synthetic molecules bound; and (iv) (perhaps) information on pockets in the structures.
   b) Representation of biological complexes in PDB/SBKB: (i) Rename "Biological Unit" to "Biological Complex," the terminology used more often by experimentalists; (ii) Add text/citations to the "Biological Complex" description, so it doesn't only describe the oligomeric state, but also the experiments used to determine the oligomeric state; and (iii) Include required non-polymer ligands in the definition of the "Biological Complex." The biological complex of human hemoglobin, for example, should include not only two alpha and two beta protein chains, but also the four heme groups.
2. *Materials Repository:* The Materials Repository at DNASU is impressive and starting to reach greater audiences. The plasmid repository has become a useful resource for the scientific community within and outside PSI:Biology. Publicity could be increased by notifying every NIH award holder that the resource exists.
3. *Enhancing the impact of PSI:Biology via information collection and presentation at SBKB:*
   a) The Program should decide what information exactly to be captured by the SBKB, and enforce essentially 100% compliance with data reporting. In the interests of the Program, we recommend that the Centers and Partners provide copies of compliance reports to the Advisory Committee. This will allow us to evaluate what fraction of all of the data is ending up in the PDB, on the SBKB webpage, and at DNASU, and whether or not data reporting is an issue.
   b) Metrics information such as publications, structures solved, etc. is well presented per Center and Partner on SBKB. For progress assessment, a page should be added where output and metrics are presented to allow easy comparison of these data. Such a page could be supplemented with funding information (separating PSI funds per group and funds obtained from other sources by these groups) and be made available only for the Advisory Committee. The information should (also) be organized such that the

three major areas of activities of the HTP Centers can be followed per year – i.e. per HTP Center the progress in Partner projects, the community nominated projects, the sequence space exploration projects.

c) The investigators should consider how best to implement technology transfer for instruments and techniques developed in the Centers so that they will become more broadly available to general scientific community. Accordingly, the Program should collect information on patents, licenses and spin-off companies and display this information on a special page of the SBKB.

d) It is critical to highlight the PSI:Biology accomplishments and strengthen dissemination of the technologies and structural results. This should be done both by individual investigators and also through the SBKB. The SBKB is leading the way with efforts to highlight breakthroughs and exploring a short publication format for high-throughput structures (see IMPACT section below). This is especially important given that in PSI2 there was an about 3 year lag between structure solution and publication and use of the structure in other studies.

C. OUTREACH AND IMPACT

1. *Highlighting the PSI:Biology accomplishments:* As PSI:Biology reaches its mid-term, it is time to tell some of the success stories. There has been progress in articulating how the PSI:Biology is succeeding where individual grants could not. However this will become increasingly important in the next two years. We wish to remind investigators that they must take every opportunity to articulate what accomplishments have been made possible by this program that could not have happened with traditional R01 funding. In particular, there needs to be material evidence that the whole is greater than the sum of the parts. The scientific community must hear what advances have been fostered by this program and appreciate the pioneering science. In addition, telling the stories may facilitate optimization of partnerships, as others may imitate mechanisms of partnering that have worked well. The Committee recommend that:

   a) It is time to “wrap things up” - to focus on a few key stories and get them to publication ("completion"). A goal of each group now should be to identify which of their developing efforts can be in the literature in 1-2 years, and testify strongly to the synergy between biology and structure determination; What are the cases wherein structure determination yielded an unanticipated explanation for biology, or suggested biological experiments that would not have been considered in the absence of the structure? This is particularly important for the Biology Partners, where there are concerns about the limited number of publications resulting from PSI support.

   b) Consider a special section in *Science* or *Nature* showcasing the accomplishments and the goals of the PSI:Biology. These can be presented as works in progress but they should make a case for molecular structure as an integral component of understanding biology.

   c) Publish newsy review of how the program is doing and what’s being learned, with vignettes highlighting top science and descriptions of the partnerships and CNT.

   d) The groups are strongly encouraged to: (i) submit short abstracts plus a picture on each significant structure, advance, etc. that they make; and (ii) enhance the mechanism for publishing these abstracts on the SBKB website (or perhaps even in another format like *PLoS ONE*). There seemed to be a number of nice surprises that should be highlighted (e.g., the calicheamicin double pentamer).

   e) Ultimately, the committee feels that each group will have to complete the statement “PSI:Biology is a success because….” in a meaningful manner.

2. *Metrics:* The Metrics Committee needs strong leadership and needs to work hard on developing more sophisticated metrics if the accomplishments of the group are to be rigorously evaluated and supported. Bet-

ter assessment indices of the return on PSI funding investment are required. A few recommendations from the Committee are:

a) In addition to "structure counting," it is necessary to have more "value-driven" evaluations that measure impact and the leveraging and enabling power of PSI:Biology. These may include: numbers of papers, measurements of the impact of those papers, measurements of who is using the structures, ways to "score" examples in which where cooperation or economies of scale enabled science that could otherwise not have been accomplished, measurements of overall scientific impact, and measures of training impact.

b) PSI:Biology has reached out to the computational community, and this has had an impact on ongoing structure prediction efforts. It is important that as many structures and complexes as possible continue to be submitted to CASP, CAPRI and similar "challenges." This "leveraging factor" resulting from computational prediction of structures could be included in metrics.

c) PSI:Biology investigators have continued to leverage other initiatives and funding sources. Such leveraging is strongly encouraged to harness the power of the Program. However, some way of distinguishing gains directly attributable to PSI:Biology is needed. A breakdown of percentage of the support from PSI and other sources in the research results and publications is needed to account for the contribution of PSI. The group should clearly list the published papers directly supported by PSI, with proper accreditation as it appears on the NIH eRA Commons site.

d) The group also needs to track how the relevant metrics change over time (so that accurate projections can be made).

3. *Publishing high-throughput structures:* Publishing the many high-throughput structures remains a challenge. It is necessary to get more than say 20% of structures into a real, substantive paper. The SBKB's effort in exploring short publications for these structures is strongly encouraged. Other suggestions are:

   a) Depositors should write an abstract for each submitted structure, the abstract should include a statement describing the functional role(s) of the molecule and the significance of the structure. This exercise would also be useful for target selection so that efforts are concentrated on targets of genuine interest and potential significance. This would be a big step towards giving submissions the status of mini-publications.

   b) A simple publication format would be useful, perhaps modeled after *Acta Crystallographica E* for small molecule structure reports and incorporating the strong validation activities of the PDB.

4. *Training:* The training impact of PSI:Biology must be assessed. The investigators should begin to track, and inform the SBKB, trainees who have been supported by the Program (from undergraduate and graduate students to postdoctoral and research scientists) and their career paths. Anecdotally, student impact appears quite good. Putting broad-based state-of-the-art technology at the disposal of students frees them to pursue more in-depth experiments and to focus on asking good questions. Exposure to the pipeline and breadth of training in PSI:Biology is a plus whereas in earlier incarnations of PSI the training was too narrow. Understanding the basic principles underlying the different methodologies is necessary in order for students to intelligently approach increasingly complex structure problems. Documentation manuals, workshop handouts and videos should be prepared for the investigators to enable them to use the pipelines more effectively.

5. *Outreach:* Some members of PSI:Biology are actively involved in outreach at the undergraduate or high-school level. This should be encouraged and facilitated, perhaps by making teaching and outreach resources

available to all members of the Program. NESG's NMR Wiki and Science Olympiad support by the RCSB PDB exemplify very valuable outreach to two other key audiences.

6. Diversity: Provide evidence that participation of under-represented groups at highest decision-making levels is on track to achieve parity with their participation at the lower implementation levels. We need to see more under-represented groups in more visible roles, and the science needs to be taking full advantage of the diverse experience and talent represented in these groups.

D. LOOKING FORWARD SCIENTIFICALLY

1. *Emerging Themes of Shared Interest*: PSI:Biology would benefit from a careful analysis of areas of shared interest, and find ways to promote the synergy of PSI:Biology in these areas. The synergy present among the Membrane Protein Centers should be a goal for other areas of common interests.
   a) An emerging trend among the PSI teams is towards biological complexes. This is highly encouraged, and a development period to create the biological and biochemical tools and methodologies necessary to realize this challenge is warranted.
   b) Another emerging theme for many Centers is the addition and application of biochemical and imaging techniques to understand the biological complexity of protein complexes under study. This too is well justified.
   c) The important area of protein complex preparation is advancing well. We encourage similar focused efforts for optimizing eukaryotic protein expression (particularly for human proteins or close homologs), and elucidating all aspects of post-translation modification in the process.
   d) More projects might benefit from establishing collaborative efforts in electron microcopy, in particular electron tomography, which would help to place the structures in a cellular context.
2. *Technological Innovation:* There are signs of scientific maturity as PSI:Biology is transitioning from a structural pipeline of "more" structures to a biological engine with "higher-value" structures. Continuing technological advances will be critically important to this transition. PSI:Biology should continue to make major contributions to free-electron laser crystallography (XFEL) and obtain as much experience as possible in this highly innovative, and potentially revolutionary technology area where PSI:Biology projects are already making a contribution (e.g., in providing nano-crystals).

In summary, the committee applauds all contributors to the PSI:Biology effort but exhorts them to move to publication on lines of work that are sufficiently mature, make explicit in all work the added-value of PSI's integrative funding, make the most of the partnerships in publishing work that synthesizes the insights of structure with the implications to biology and continue the fine work of advancing frontiers in technology, methodology and understanding of complex systems.

From the PSI:Biology Advisory Committee

Cathy H. Wu, Ph.D., Chair
University of Delaware &
Georgetown University Medical Center

Karen Allen, Ph.D.
Boston University

Steve Bryant, Ph.D.
National Center for Biotechnology Information/
National Library of Medicine/NIH

Wah Chiu, Ph.D.
Baylor College of Medicine

Lila Gierasch, Ph.D.
University of Massachusetts, Amherst

Wim Hol, Ph.D.
University of Washington

Karolin Luger, Ph.D.
HHMI and Colorado State University

Anne-Frances Miller, Ph.D.
University of Kentucky

William Montfort, Ph.D.
University of Arizona

Wes Sundquist, Ph.D.
University of Utah

Stephen H. White, Ph.D.
University of California at Irvine