Report of the
National Institute of General Medical Sciences
Future of Structural Biology Committees

Recommendations for Continued Investment in Structural Biology Following the Sunsetting of the Protein Structure Initiative

December 2014

Introduction and Background

Structural biology is a major area of interest for NIGMS, with approximately $140 million invested annually. Over the past 15 years, a substantial portion of that investment has been through the Protein Structure Initiative (PSI). The PSI program has developed technologies and methods that have improved protein structure determination and should be useful for the community in the future. The final phase of the program, PSI: Biology ($70 million/year) will end in 2015. Although the PSI is drawing to a close, structural biology remains a high priority for NIGMS, including the development of new technologies and support of structural biology resources essential for the biomedical research community. As NIGMS support for structural biology changes, it is important to consider whether and how resources developed in the PSI program may continue to be important for the broader community going forward.

In September 2013, NIGMS Director Jon Lorsch commissioned two committees to develop recommendations for the future support of structural biology by NIGMS. The NIGMS Future of Structural Biology Committees (FSBC) included members with expertise in both the practice of structural biology as a field and its impact on other researchers. An external committee was composed of practitioners of structural biology and researchers who use structural biology data and resources in their work. That group was asked to focus primarily on articulating community needs and suggesting emerging challenges and opportunities in structural biology. An internal committee, composed of NIH staff, provided a complementary perspective in those areas and developed recommendations for implementation of the priorities identified by the external committee. The committees worked closely together throughout the process. Following a year of data gathering and discussion, the committees have developed the recommendations reported here.

Information Gathering and Deliberation

The committees had access to all previous evaluations of the PSI, including the mid-course review of PSI:Biology completed in September 2013. The committees consulted written information provided by the PSI investigators, conducted interviews with PSI investigators and other relevant experts, issued a Request for Information (RFI), and in some instances, assigned individual committee members to discuss specific questions with outside experts.

All committee meetings were conducted by teleconference or Web meeting. When discussions were held with PSI investigators or outside experts, the committees participated jointly. Most deliberative meetings were conducted separately, but each committee had access to the notes

from all meetings to maximize feedback between the two groups. As recommendations were being developed, notes from each committee's meetings were forwarded to the counterpart committee so that they could work together iteratively.

Recommendations

The recommendations of the committees are presented in two sections: addressing the preservation of PSI resources deemed valuable by the committees and priorities for the future of structural biology. The recommendations are expressed at a conceptual level. While the committees refrained from suggesting specific funding levels or grant mechanisms, they did address major questions such as the appropriateness of grant or contract approaches, whether funds should be set aside for certain activities or whether there exist special considerations that should be taken into account in peer review. The committees tried to describe the defining features of potential programs in enough detail for them to be understood.

I. Preserving PSI Resources

Resources for protein expression: PSI supported the development of several robust, high-capacity protein expression pipelines, as well as some specialized or even unique approaches to difficult expression problems (e.g., anaerobic protein expression). Development of these resources was driven by internal program goals, such as the experimental determination of large numbers of protein structures. Some of these capabilities have been refined and disseminated to the degree that they are now widely available to nonspecialists outside the PSI. However, not all resources developed in the PSI are technologically mature or are freely available to the community. PSI protein expression resources have primarily been applied to specific collaborative projects with PSI-funded biological consortia. PSI protein expression capabilities not otherwise widely available should be considered for further support, with the clear understanding that they will be made widely available.

As macromolecular structural biology research becomes more accessible to nonspecialists, access to sophisticated, reliable protein expression resources is critical. The biochemistry of some individual proteins, complexes or pathways makes their expression challenging. Very frequently, an active, highly iterative collaboration between the group initiating a study and the protein expression specialists is essential for success. These undertakings often require a collaborative research project rather than a routine service. Most investigators do not have access to these resources within their institution (e.g., 96-well robotic cloning stations, high-

capacity fermenters, etc.). As technologies continue to evolve and expertise matures, some of this may change. For now, however, certain aspects of current PSI protein expression resources represent an important potential resource for the biomedical research community.

Both prokaryotic and eukaryotic expression resources are important. Eukaryotic protein expression is often complex and challenging, requiring innovation and problem solving. High-throughput prokaryotic expression is also important for co-expression of multiple subsets of proteins participating in large complexes or for expression of libraries of proteins (e.g., truncation scanning). However, simply continuing to support present PSI resources would also maintain the orientation of these facilities toward the original PSI program goals, which are no longer relevant. To be effective, resources for the community need to be developed and administered with an outward-facing orientation and by researchers with a strong collaborative orientation.

It is recommended that a modest number of protein expression research resource centers be supported using existing funding mechanisms. This recommendation concurs with that made in the 2013 program evaluation. The new centers would work with the research community to develop appropriate approaches to challenging expression problems. One possible working mechanism would be for NIH-funded investigators outside the center to submit short proposals to the center detailing how and why center technologies are critical for overcoming a significant production or crystallization bottleneck on a high-interest biological problem. To avoid flooding the centers with trivial problems, investigators using the centers would be required to defray some portion of the costs associated with using the center, thereby ensuring they have "skin in the game." Because some NIGMS resource programs, such as the Biomedical Technology Research Resources, support a diverse collection of technologies and resources, it is possible that no specific funding opportunity announcement (FOA) will be necessary for the solicitation of protein expression resources. Alternative approaches are discussed at the conclusion of this section.

The PSI Materials Repository: The PSI Materials Repository (PSI-MR) is presently supported by a U01 cooperative agreement that extends through the summer of 2016. The PSI-MR is responsible for archiving and distributing plasmids for the expression of proteins that have been the subject of PSI research efforts. The PSI-MR performs quality control analyses of the plasmids submitted. The PSI-MR exists within the physical and institutional infrastructure of the DNASU plasmid repository at Arizona State University. The PSI-MR represents a substantial portion of the support for and depositions to DNASU.

The PSI-MR is a well-designed, thoughtfully administered resource. It includes a substantial degree of automation and underlying informatics, allowing large numbers of plasmids to be either submitted or requested. The latter facilitates the development of custom libraries for

systems-level experiments. This is a unique capability. Blanket material transfer agreements are used to ensure that plasmids are freely distributable. User fees largely support the cost of distribution, but not the core functions of the repository.

The committees recognize that the PSI-MR has demonstrated impact beyond the PSI consortium and beyond the structural biology community. It is clear that there is substantial variation in the community's interest in the plasmids held in the repository. Many may never be requested. A critical review of the content of the repository should be undertaken to determine whether significant cost savings would result from the elimination of plasmids that are not utilized. However, utilization of the resource outside the structural biology community is growing quickly and represents a cost-effective means of providing access to important biomedical research resources for the academic community. This discussion is based on current knowledge. It is possible that at some point in the near future, the time and resources necessary to reliably recapitulate needed plasmids may become competitive with that of maintaining them in a resource.

It is recommended that support for this resource be continued, in some form, by either an assistance mechanism (grant or cooperative agreement) or a contract. This support should be at a level that would be adequate to support the submission and full characterization of up to 10,000 submissions per year. The virtue of a contract is that it will allow research activities to be de-emphasized in favor of straightforward access to resources. However, under current federal regulations, contracts cannot allow the collection of user fees to offset costs. If this continues to be the case, a cooperative agreement similar to the arrangement presently in place might be the best option.

The Structural Biology Knowledgebase (SBKB): The Structural Biology Knowledgebase (SBKB) could provide several resources that might be useful to the general biology research community, but the committees found that it is not clear that the SBKB is presently meeting this goal. This concern is consistent with those expressed in PSI evaluations in 2007 and 2013. The SBKB began late in the 15-year span of the PSI, and the broader community has been largely unaware of what it offered or that it could be valuable to their research.

The SBKB is a heterogeneous collection of modular resources. The committee concluded that none appear to be heavily utilized by the broader community. However, some of these resources appear to have strong potential or represent unique resources that cannot be easily recapitulated if lost. The committees feel that, given this uncertainty, a well-executed plan for support of a slimmed-down SBKB during a pilot period, with clear milestones for measuring impact, might be appropriate. The components of the SBKB deemed potentially most useful to the community are suggested for inclusion in a pilot:

1. The protein modeling portal (http://www.proteinmodelportal.org)
2. The links to multiple external databases with information on protein sequences (http://beta.sbkb.org/page/show/protein-structure-function-relationships)
3. The BioSync synchrotron database (http://biosync.sbkb.org)
4. The TargetTrack database (http://sbkb.org/tt)

The BioSync database holds summary information about all synchrotron beamlines as well as all structures solved at those facilities. It is a very useful resource for structural biologists worldwide.

The TargetTrack database is still actively used to track structural genomics targets from the NIAID structural genomics effort (two centers) and the joint Wellcome Trust-industry Structural Genomics Consortium (in Canada, the UK and Sweden). The committee thought that perhaps one or both of those groups might be interested in participating in the funding of any continuation of the TargetTrack database. However, despite the longstanding availability of TargetTrack for use by the broader structural biology community, it has not been widely adopted by those outside the PSI and other large programs referenced above.

The simplest solution for continued access to SBKB resources, particularly databases, would be to roll them into the Protein Data Bank (PDB). The rationale is that the two resources presently share personnel and physical infrastructure at Rutgers and are intended to serve similar user communities. However, it is important to recognize that the SBKB and the PDB differ substantially in their governance. The PDB is subject to international agreements regarding the nature of its contents. For the SBKB to be incorporated into the PDB would represent a significant shift in the mission of the PDB. Merging would necessitate modifying the international agreements drawn in 2005 that presently define its scope. However, this expansion in scope might be of interest to the PDB.

If the SBKB is supported in some form, the committees strongly encourage NIGMS to ensure that there is substantial emphasis placed on outreach to potential users, particularly those presently in training at the graduate and postdoctoral levels. Acceptance by these potential users would seem to be critical if SBKB resources are ever to be adopted widely.

Close-out and bridge funding to prevent loss of infrastructure during transition to other programs: There are significant PSI investments in topic areas consistent with the committees' recommendations for future investment. The committees do not recommend administrative or competitive extension of funding explicitly for these PSI activities other than for orderly ramp-down or to keep infrastructure and personnel intact during the process of receiving new applications. The committees feel that research and resource activities in those areas designated for continued support should be organized differently from the current PSI model.

There should be an explicit break with the present structure and priorities. However, the committees do recognize that the present investments constitute in some cases a nucleus for development of new projects and resources. NIGMS is encouraged to make programmatic determinations regarding the appropriateness of bridge or ramp-down funding for PSI activities. While new awards should be the result of open competition, it is important to recognize that on a practical level, these existing investments in infrastructure and expertise represent the most likely starting point for new activities in those areas.

Dissemination of data and software resources: PSI investigators have developed numerous discrete, relatively simple resources. Examples include protocols, data acquisition parameters for NMR experiments and data analysis tools. Consistent with NIH data and resource sharing policies, these resources should be disseminated to the extent reasonably possible. NIGMS program staff should be attentive to this process and work with PSI grantees to facilitate dissemination of those resources that may be useful beyond the sunsetting of PSI.

II. High Priorities for the Future of Structural Biology

Elucidation of biologically relevant structures remains highly important for biomedical research. Thus, it is important to maintain the technologies that make structural investigations possible at the most advanced level and to ensure that these resources are accessible to NIH-funded researchers. The major contributing technology to three-dimensional structures solved through the PSI is macromolecular crystallography (MX), followed by NMR. Because MX remains the most widely used technology for high-resolution structure determination and because cryo-electron microscopy (cryo-EM) is on the cusp of being able to provide similar resolution structures of large- and medium-sized assemblies, support for both methods is explicitly recommended. Other contributing technologies of importance include NMR and small-angle X-ray scattering. Indeed, it is combinations of these and other technologies and approaches that enable structure determination of the most complex and perhaps highest-impact macromolecules and assemblies.

Continue to support synchrotron beamlines for macromolecular crystallography: Advances in synchrotron MX technologies over the last decade have made it accessible to a wide range of researchers. These essential resources are now augmented by automation and user-friendly software that can guide strategic data collection, processing and even electron density map interpretation. More than 90 percent of all macromolecular structures that are currently deposited in the PDB were solved with synchrotron data. Over 4,000 scientists from the biological community use synchrotrons in the United States annually. As portions of PSI funding

contribute to the operation of some highly productive MX beamlines, it will be important to ensure that any support they have received through PSI is continued, commensurate with the demand for these critical resources. About $4 million per year from PSI has supported access to MX resources at the Stanford Synchrotron Radiation Lightsource (SSRL), Advanced Photon Source (APS) and National Synchrotron Light Source (NSLS) in addition to other NIGMS support for beamlines at these and other locations.

Meet the need for modern cryo-electron microscopy resources: Recent innovations in direct electron detection technology have led to revolutionary advances in the capabilities of cryo-EM. Direct electron detectors with high frame rates and correction for beam-induced specimen motion have enabled major improvements in resolution, approaching that of MX for specimens of MW greater than 150 KDa. Thus, it is important to anticipate the need for these resources for the elucidation of three-dimensional structure by an even broader research community. These advanced technologies are expensive and currently only easily accessible by cryo-EM experts, which limit the practicality of having them in many individual investigator laboratories. Therefore, it is recommended that NIGMS support regionally shared resources with the aim of eventually providing access to any researcher whose project would benefit from structure determination of a macromolecular assembly.

An important distinction of cryo-EM is that it places specialized demands on sample preparation. For instance, samples must be freshly prepared biochemically and freezing protocols generally need to be optimized on the fly. Thus, in addition to supporting regional centers with cryo-EM research expertise and extensive wet-lab equipment and on-site sample preparation support, significant investment in local university and research institute cryo-EM infrastructure will need to be established nationwide to maximize the use of high-end national facilities. In addition, the development of publicly accessible software and the provision of necessary computing resources to process the large amount of image data generated from direct electron detectors will be needed to make cryo-EM readily accessible to the broader research community.

NMR resources: The PSI has driven progress in the development of methodologies for protein NMR data collection and signal assignments. Some of these resources have already migrated to commercial instrumentation, and some are now available at national NIH-funded NMR centers such as the National Magnetic Resonance Facility at Madison (NMRFAM). It is recommended that efforts be made to ensure that all NMR pulse sequences and processing/data analysis software developed as part of the PSI be made fully available through these other existing centers. In addition, we recommend that NIH continue to support competitively reviewed NMR centers that function as national resources for NMR methods development, provide access to

advanced instrumentation that is not routinely available, serve as repositories for NMR data (e.g., the Biological Magnetic Resonance Bank) and provide training at levels appropriate to both novices and experts.

Support integration of methods for structural biology: It is important to ensure that hybrid approaches--the application of more than one structural approach--are adequately supported as technologies for structural biology continue to advance. This includes but is not limited to crystallography, cryo-EM, scattering, and NMR, and the integration of the resulting data. Such an effort will inevitably require an increased focus on integration of both experimental approaches and data analysis. It is impractical to co-locate the multiple infrastructure-intensive technologies described above. Encouraging technology developers and early adopters to maximize the compatibility of these approaches should be a priority. Support for improvement and open dissemination of computational tools that enable the integration of data from multiple platforms is essential. The committees asked European colleagues associated with the Instruct program about emerging needs in structural biology, and they concurred about the need for integrative approaches to structural biology.

Enable collaborative, multi-investigator efforts in membrane protein and large macromolecular assembly structure determination: Membrane proteins and large macromolecular assemblies are important classes of targets for basic science and drug discovery, and their characterization presents unique technical challenges. Recent investments through the NIH Common Fund and the PSI have contributed to progress in their expression and structure determination. Advances are also being made by individual laboratories. However, successes are isolated, and it is difficult to generalize the solutions to technical challenges encountered in individual projects. In general, membrane protein and large macromolecular assembly structure determination benefits from multi-investigator collaboration.

It is recommended that NIGMS support multi-investigator grants that focus on the development of improved tools for membrane protein and large macromolecular assembly expression and structure determination and application of emerging methods to critical biomedical research problems. Multi-investigator projects will likely require substantially higher budgets than conventional R01 research project grants, and therefore applications should be solicited using program announcements. Because this science is still emerging and the approaches are somewhat arcane, review in special emphasis panels (SEPs) will help to ensure that the appropriate expertise is available in considering applications that are likely to be large, complex and highly specialized.

III. Implementing the Recommendations

In closing, the recommendations in this report focus on the resource needs of the biomedical research community rather than any imperative to preserve current implementations of those resources. Downstream support of current resources should be directed to the best approaches to addressing those future needs, determined through rigorous peer review in open competitions in which non-PSI investigators have an equal opportunity with PSI incumbents.

Solicitation of applications for support of these resources: It is inevitable that NIGMS will need to describe its interests in these areas through some vehicle. One possibility is the use of notices in the *NIH Guide* or the NIGMS Web site. Applications in response to these expressions of interest would be submitted through existing FOAs. NIGMS should also consider publishing new FOAs that are consistent with existing program requirements but describe in detail the new resources desired. This approach has two advantages: It would provide the opportunity to be very clear regarding program requirements, and it would also allow a thorough discussion of the important ways in which new structural biology resources supported by NIGMS will differ from any predecessor PSI activities. The recommendations in this report discuss resources developed in PSI that the committees feel will be important for the community in the future. However, in essentially every instance, it is recommended that the future of these resources, both in terms of continued development and their relationship to the community, be very different from their instantiation in PSI. To the degree that NIGMS can emphasize these differences, prospective applicants and the community will benefit.

Appendices

Charge to the Internal and External Committees

The PSI:Biology program is being sunsetted. Current funding obligations will be met, but it is not anticipated that new funding opportunity announcements will be issued under PSI, and existing awards will not be eligible for competing renewal under the existing program. PSI: Biology has developed a number of novel technologies and a substantial infrastructure to address challenging problems in structural biology. NIGMS must now make determinations regarding what elements of the program should be preserved, and in what manner that is best done, in order to provide unique capabilities and resources that are not available elsewhere. These may include elements of the PSI Knowledgebase, the PSI Materials Repository, ongoing technology development, and the pipelines in the centers for expression construct design, gene synthesis, protein expression and purification.

Technology development in all areas of structural biology, including crystallography, is still an important area for investment by NIGMS and NIH. The committees should articulate broad goals and specific objectives that can help to define the continued development of infrastructure and new directions for research. Areas of focus may include: access to existing resources, continuity of support for technology development with a high potential for broad biomedical impact and emerging opportunities. A critical element of these discussions should be the needs of the broader biomedical research community served by structural biology. This process is not intended to lead to a specific program with dedicated funding, and any investments emerging from this process will be considerably more modest in scope than the PSI. It will also be important to understand and articulate the anticipated long-term trajectory of any such investments.

While the entire charge applies to both committees, the work of the two committees will be complementary rather than duplicative. The external committee's primary focus will be on the development of a strategic vision for the future of structural biology at NIGMS. The internal committee will focus on the practical aspects of how to proceed with an orderly shutdown of the current program, including any continued support of existing resources that the committees conclude is needed. The internal committee will make recommendations to the external committee and the NIGMS director for their consideration, and will serve as a resource for the external committee. The internal committee will also begin the implementation of the vision that emerges from this process. The chair of the internal committee will act as liaison.

Committee Rosters

EXTERNAL COMMITTEE

Leemor Joshua-Tor, Co-chair
HHMI/W.M. Keck Structural Biology Laboratory
Cold Spring Harbor Laboratory
1 Bungtown Road
Cold Spring Harbor, NY 11724

Rick Horwitz*, Co-chair
University of Virginia
P.O. Box 400301
Charlottesville, VA 22904

Brenda Bass
University of Utah
Department of Biochemistry
15 N Medical Drive East, Rm 4800
Salt Lake City, UT 84112-5650

James Berger
Johns Hopkins University School of Medicine
725 N. Wolfe Street, 713 WBSB
Baltimore, MD 21205

Wah Chiu
Baylor College of Medicine
The Verna and Marrs McLean Department of Biochemistry
One Baylor Plaza, Room: BCMN-N420
Mail Stop: BCM125
Houston, TX 77030

Adrian Ferré-D'Amare
National Heart, Lung, and Blood Institute
Laboratory of RNA Biophysics and Cellular Physiology
Biochemistry and Biophysics Center
Building 50, Room 2140
50 South Drive, MSC 8012
Bethesda, MD 20892-8012, USA

Lila M. Gierasch
Depts. of Biochemistry and Molecular Biology and Chemistry
Life Sciences Laboratories Room N331
240 Thatcher Way
University of Massachusetts, Amherst
Amherst, MA 01003-9364

Janet L. Smith
Life Sciences Institute
Department of Biological Chemistry
210 Washtenaw Avenue
Ann Arbor, MI 48109-2216

Michael F. Summers
Department of Chemistry and Biochemistry
University of Maryland Baltimore County
1000 Hilltop Circle
Baltimore, MD 21250

*National Advisory General Medical Sciences Council (NAGMSC) member

INTERNAL COMMITTEE

Douglas Sheeley, Chair
Division of Biomedical Technology,
Bioinformatics, and Computational Biology
National Institute of General Medical Sciences
45 Center Drive
Bethesda, MD 20892-6200

Christine Colvis
National Center for
Advancing Translational Sciences
BG 1DEM RM 994
6701 Democracy Boulevard
Bethesda, MD 20817

Valentina Di Francesco
Division of Genome Sciences
National Human Genome Research Institute
5635 Fishers Lane, Room 4111
Rockville, MD 20892
Formerly:
NIAID Office of Genomics and Advanced Technologies

Charles Edmonds
Division of Cell Biology and Biophysics
National Institute of General Medical Sciences
45 Center Drive
Bethesda, MD 20892-6200

Adam Felsenfeld
Division of Genome Sciences
National Human Genome Research Institute
5635 Fishers Lane, Room 4115
Rockville, MD 20892

Sue Haynes
Division of Genetics and Developmental Biology
National Institute of General Medical Sciences
45 Center Drive
Bethesda, MD 20892-6200

J. Randy Knowlton
National Cancer Institute
Structural Biology and Molecular Applications Branch (SBMAB)
9609 Medical Center Drive, Room 6W334
Rockville, MD 20850-9747

Pamela Marino
Division of Pharmacology, Physiology, and Biological Chemistry
National Institute of General Medical Sciences
45 Center Drive
Bethesda, MD 20892-6200

Brian Pike
Office of Scientific Review
National Institute of General Medical Sciences
45 Center Drive
Bethesda, MD 20892-6200

Michael Sesma
Division of Training, Workforce Development, and Diversity
National Institute of General Medical Sciences
45 Center Drive
Bethesda, MD 20892-6200

Ward Smith**
Division of Cell Biology and Biophysics
National Institute of General Medical Sciences
45 Center Drive
Bethesda, MD 20892-6200

Amy Swain
Division of Biomedical Technology,
Bioinformatics, and Computational Biology
National Institute of General Medical Sciences
45 Center Drive
Bethesda, MD 20892-6200

** PSI:Biology Network Director