new centers established (*21*).

Building consensus among the public, international organizations, academics, industry, governments, nongovernmental organizations, and the media will be difficult but essential to address different value orientations and develop wise public policy. It is possible that a commission on genomics and global health could serve as a platform to raise awareness, mobilize resources, and bring stakeholders together to focus on their common interest in the health of people in developing countries and close gaps in health equity. Commissions can occasionally be effective. The Commission on Health Research for Development (the Evans Commission) galvanized the health research community (*22*) with the concept of the "10/90 gap": that 90% of research expenditure is dedicated to the health problems of 10% of the world's population. An early consensus-building effort is now underway on a regional basis. On 8 August 2001 in Nairobi, Kenya, the First Roundtable on Africa, Science, and Technology in the Age of Globalization (Fig. 1) resolved to establish a regional process to develop science and technology strategies aimed at closing the digital and genome-related biotechnology gaps with the rest of the world. The Roundtable appointed John Mugabe, Director of the African Centre for Technology Studies, as interim secretary. Participants included 38 leading policy-makers and scientists, including permanent secretaries and directors of science and technology policy bodies, from 11 African countries. This process provides an opportunity to pursue biotechnological advances in the context of the New African Initiative, which is on the G8 agenda next year.

The voices of those in developing countries must be heard as the health biotechnology revolution unfolds. Those protesting in Genoa are not the ones who are sick in Africa. We need to develop a mechanism to tap the views of opinion leaders in developing countries on important policy questions and in real time.

Finally, it will be necessary to create innovative financing mechanisms to channel large investments into promising scientific ideas targeted on health problems of developing countries. One major project established this year by the United Nations, the Global Health Fund, set a goal of raising $7 billion to $10 billion, but only about $1.4 billion had been pledged by early August 2001 (*23*). The fund is an important development, but this result may indicate fatigue on the part of developed-country governments for donations. A possible investment model is the one developed by Globalegacy (*24*), a United Kingdom–based organization working to create long-term social and economic growth through commercial ventures with deprived urban communities. An investment fund based on similar principles but focusing on health genomics and biotechnology in developing countries could channel needed investment to undercapitalized scientific ideas. The business model would optimize health improvement in developing countries but would also provide economic return on investment. If one or more developed-country government invested just 10% of the 0.7% of gross domestic product target for official development assistance to such a fund for only 1 year, and this investment was matched by the private sector, the fund would have sufficient capital to pursue its work.

We will know that these efforts are successful when the G8 take up this challenge in Kananaskis, when we see more examples like the Cuban meningitis vaccine, and when we ultimately see decreased inequities in life expectancy and other indicators of global health equity. Perhaps the best indicator of success will be if there is no World Bank report in 2010 on the health genomics divide!

References and Notes

1. B. R. Bloom, D. D. Trach, *Br. Med. J.* **322**, 1006 (2001).
2. See www.g7.utoronto.ca/g7/summit/2001genoa/africa.html.
3. World Bank, *World Development Report 1998/99: Knowledge for Development* (Oxford Univ. Press, Oxford, 1998).
4. P. A. Singer, A. S. Daar, *Nature Biotechnol.* **18**, 1225 (2000).
5. E. Harris, A. Belli, N. Agabian, *Biochem. Edu.* **24**, 3 (1996).
6. E. Harris, *A Low-Cost Approach to PCR: Appropriate Transfer of Biomolecular Techniques*, N. Kadir, Ed. (Oxford Univ. Press, New York, 1998).
7. See www.pugwash.org/reports/ees/ees8d.htm.
8. K. Carr, *Nature* **398** (6726 suppl.), A22 (1999).
9. *IAVI Press Release*, 27 January 2001 (see www.iavi.org/press/46/kenya_trial.htm).
10. L. J. Richter, Y. Thanavala, C. J. Arntzen, H. S. Mason, *Nature Biotechnol.* **18**, 1167 (2000).
11. See www.celera.com/genomics/news/articles/07_00/vaccines_trees.cfm.
12. See www.malariavaccines.org/files/MVI-India-pr.htm.
13. See http://dbtindia.nic.in/.
14. See www.iitb.ac.in/latest/biotech/biotech.html.
15. H. Jomaa *et al.*, *Science* **285**, 1573 (1999).
16. E. Schaeffeler *et al.*, *Lancet* **358**, 383 (2001).
17. See www.who.int/director-general/speeches/2001/english/20010514_wha54.html.
18. A. Daar, J.-F. Mattei, "Medical genetics and biotechnology: implications for public health" (document WHO/EIP/GPE/00.1; annex 1 of *Report of the Informal Consultation on Ethical Issues in Genetics, Cloning and Biotechnology: Possible Future Directions for WHO*, December 1999). The report can be obtained from WHO by writing to T. Pang (e-mail pangt@who.int).
19. P. A. Singer, S. R. Benatar, *Br. Med. J.* **322**, 747 (2001).
20. See www.state.gov/documents/organization/4426.doc.
21. C. Juma, paper delivered at the United Nations University/Institute for Natural Resources in Africa Annual Lectures, 1999 (see www.unu.edu/inra/pub/juma/AL99.html).
22. L. C. Chen, concluding reflections at the International Conference on Health Research for Development, Bangkok, Thailand, 10 to 13 October 2000 (located at www.rockfound.org).
23. See http://allafrica.com/stories/200107310376.html (1 August 2001).
24. See www.globalegacy.com.

25. We thank A. J. Ivinson for editing this commentary, A. Smith and S. Nast for researching the examples of health genomics and biotechnology, and E. Dowdeswell for discussions about global governance. Grant support was provided by the Program in Applied Ethics and Biotechnology (supported by the Ontario Research and Development Challenge Fund, GlaxoSmithKline, Merck and Co., Pfizer, Sun Life Financial, the University of Toronto, the Hospital for Sick Children, Sunnybrook and Women's College Health Sciences Centre, and the University Health Network), and the Canadian Program on Genomics and Global Health (supported by Genome Canada). P.A.S. is supported by an investigator award from the Canadian Institutes of Health Research.

VIEWPOINT

Global Efforts in Structural Genomics

Raymond C. Stevens,[1] Shigeyuki Yokoyama,[2] Ian A. Wilson[1]

A worldwide initiative in structural genomics aims to capitalize on the recent successes of the genome projects. Substantial new investments in structural genomics in the past 2 years indicate the high level of support for these international efforts. Already, enormous progress has been made on high-throughput methodologies and technologies that will speed up macromolecular structure determinations. Recent international meetings have resulted in the formation of an International Structural Genomics Organization to formulate policy and foster cooperation between the public and private efforts.

A major international structural genomics effort is now in progress, with the goal of obtaining three-dimensional (3D) protein structures on an equivalent scale to the genome sequencing projects. During the past 2 years alone, more than half a billion dollars of additional public and private funds have been invested worldwide in structural genomics projects. Most of this effort is focused on protein structure determinations that will finally delineate the total repertoire of protein folds and provide representative structures for each of the individual protein families (*1*).

[1]Joint Center for Structural Genomics, Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA. [2]RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan.

These new structures can then be added to the existing structural database, which would allow homology modeling to fill in the structures for other protein family members, as well as to provide active-site geometries for drug design. [See (*2*–*6*) for some representative recent reviews on various aspects of structural genomics.]

New issues have surfaced with the sheer scale of proteome-wide structure analysis projects. Although the number of genes in the genomes sequenced to date is much smaller than originally anticipated, the number of actual proteins in the various proteomes will almost certainly be much larger as a result of splice variants, protein modifications, etc. This will require structural genomics efforts to focus on large-scale structure determination projects. In addition, the biophysical characteristics of proteins vary widely in comparison to the conservative properties displayed by nucleic acids, making this protein structure determination project orders of magnitude more challenging than the various genome sequencing projects. Also, in contrast to the genome projects, the endpoint is not as clear: It is unlikely that even the smallest genome will have its complete complement of structures determined. As more gene sequences become available, the number of protein structures that could be determined will obviously increase; however, smaller and smaller percentages of new sequences would be expected to represent new folds or new family members. Hence, once the current structural genomics efforts have generated a fruitful number of 3D protein structures, future structure determinations should become more routine, or unnecessary in light of the promise of more reliable homology-modeling routines.

Structural genomics requires a large number of process steps to convert sequence information into a 3D structure (*2*). A high percentage of proteins coded in the genomes sequenced so far have unknown function and minimal or undetectable sequence homology to proteins of known structure. Thus, the majority of new protein structure determinations would remain very labor-intensive using conventional methods. However, high-throughput (HT) technological advances are now changing this facet of structural biology (*2*). Automation, miniaturization, and parallelization of process steps can deliver increasingly higher rates of protein structure determinations. New technologies, such as cell-free protein production and nanovolume crystallization, can facilitate HT preparation of protein samples. The present goals are to obtain targeted structural information reliably within a 6- to 12-month time period from cDNA to structure and to reduce the cost per structure by 90%. Current cost estimates for a protein structure determination range widely throughout the world from $50,000 to $200,000 per novel structure, not including membrane proteins, depending on the degree of difficulty of producing and processing the protein under investigation. Currently, HT technologies are mainly being developed to increase the structure determination throughput rate. However, less emphasis has been placed so far on decreasing the cost; this issue needs to be addressed quickly if we are to tackle HT structure determination in a financially acceptable manner.

In the beginning, the majority of new structures may indeed come from the so-called "low-lying fruit" in the genome sequences, such as small stable proteins from thermophilic bacteria. But more intractable protein targets, such as membrane proteins and large macromolecular assemblies—and, indeed, human proteins in general—could also have their structures determined by applying a "learning factory" approach to structural genomics (*2*). A learning factory is established by collecting, archiving, and analyzing results for all protein structure determinations that are attempted, and then feeding these data back into the process pipelines to improve efficiency and reduce the number of failures (*2*). In this approach, correlations can be made between favorable and unfavorable process steps, and trends in results (both positive and negative outcomes) can be analyzed with expert systems—currently using manual analysis, but eventually using automated learning systems—to expand our knowledge base.

The sheer magnitude of this challenge in determining proteome-wide structures has necessitated the current global initiative in the academic and industrial structural genomics communities. Over the past 2 years, structural genomics consortiums have sprung up all over the world. The RIKEN Structural Genomics/Proteomics Initiative (RSGI) has begun a HT analysis of protein 3D structures and molecular functions, using cell-free protein synthesis and a combination of nuclear magnetic resonance (NMR) at the RIKEN Genomic Sciences Center in Yokohama and x-ray crystallography at the Harima Institute SPring-8 facility (*7*). Another project at the Biological Information Research Center in Tokyo targets membrane proteins. The Protein Structure Factory, located in Berlin, was started by the German Human Genome Project (DHGP) to enable structural biologists from the Berlin area to work on the structural characterization of proteins encoded by the genes or cDNAs available from DHGP (*8*). A nascent consortium is developing in the European Union, where proposals are being solicited for "high-throughput structural genomics related to human health" (*9*). The British Medical Research Council has funded a Protein Production Facility for Structural and Functional Genomics in Oxford (*10*). On the other hand, the Wellcome Trust (*11*) has been considering forming an industrial rather than academic consortium, similar to their highly successful single-nucleotide polymorphism consortium. In Canada, structural proteomics was initiated at the Clinical Genomics Centre in Toronto using a combination of NMR and crystallography (*12*). In the United States, the Protein Structure Initiative (PSI), sponsored by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH), has a primary aim of fully populating the protein structure space (*13*). According to program director John Norvell, the main NIGMS mission is to provide structures of unique, nonredundant proteins that are representative of all the protein sequence families. The NIH effort is completely genome-driven and is the only project so far that has made it a specific goal to provide a large database of structures for use by the entire community. In October 2000, the seven NIH pilot project structural genomics consortiums (*14*) were awarded 5-year research grants to pursue the long-term goal of determining 10,000 novel protein structures over 10 years; two new centers will be supported this year. Preliminary structural genomics initiatives are also under way in many other countries, including Sweden, France, Italy, China, and Brazil.

In the United States, the PSI is based on the premise that the goal of fully characterizing protein fold space can be achieved. First, protein sequences can be organized into fold families, and then family representatives not currently present in the Protein Data Bank (PDB) (*15*) can be selected as targets. By solving the structure of these selected targets using x-ray crystallography or NMR spectroscopy, the extent of our coverage of protein family and fold space can be rapidly increased (Fig. 1). This vastly enhanced structural database will enable more reliable computations of structural models using homology modeling for the vast number of remaining protein sequences. However, from a global perspective, there is not yet any organized plan for target selection. Targets are being selected by completely different criteria—some on the basis of protein families, others according to their presence in intact organisms (such as thermophiles and infectious microorganisms), and yet others on the basis of function or potential as drug targets. However, efficient coverage of protein family representatives and protein fold space is not likely to be optimally achieved unless tar-

gets are strategically selected (*16*). It is estimated that reliance on a random selection of targets would require up to seven times as many structures to be determined to achieve 90% coverage of protein fold representation (*16*). As opposed to the random target selection choice, if the target selection process were globally coordinated to optimize the choice of structures determined, it is estimated that as few as 16,000 carefully selected structures would be a sufficient basis for constructing reasonable models for almost all proteins (*16*). Hence, the NIH effort by itself could populate much of this protein family space in 10 years.

In order to achieve sufficient throughput within a reasonable time frame, the current structural genomics consortiums, as well as a number of privately funded enterprises (Syrrx, Structural Genomix, Astex, Integrative Proteomics, Plexxikon), have developed or are about to complete HT process pipelines for determining macromolecular 3D structures. The technologies developed by these structural genomics initiatives will usher us into the era of industrial structural biology, where the research focus will be shifted back to biologists (to understand the workings of the cell) or chemists (who will prioritize drug discovery programs based on a rational process of deciding what targets are likely to lead to marketable drugs).

Important collaborations have already been initiated within the public and private sectors of the structural genomics community. For example, the Genomics Institute of the Novartis Research Foundation and the La Jolla–based biotech company Syrrx are collaborating with the NIH-funded Joint Center for Structural Genomics (JCSG) to accelerate developments in technology and improve efficiencies in the processing steps for gene targets, while still allowing these different partners to focus on their independent goals. In addition, JCSG is in the process of arranging global collaborations with other international structural genomics consortiums to advance HT technology (e.g., SPring-8, which is focusing on beamline automation).

To avoid some of the problems that developed during the public and private genome sequencing efforts, several recent international meetings in structural genomics have been the focus of intense discussions between representatives from the public and private sectors. In April 2000, the First International Structural Genomics Meeting was held in Hinxton, U.K., sponsored by NIGMS and the Wellcome Trust, to discuss policy for the international structural genomics efforts and to set up task forces to recommend guidelines. In November 2000, the first full-scale international scientific meeting, the International Conference on Structural Genomics 2000, was held in Yokohama, Japan, and reported progress on various structural genomics efforts from an expanding number of countries. In April 2001, the Second International Structural Genomics Meeting was held in Airlie, Virginia (the "Airlie Center meeting"), where delegates from four continents met to discuss international policy as well as to address current bottlenecks in structural genomics (*13, 17*).

At the Airlie Center meeting, a general consensus was reached that a global effort to disseminate data and share technology would provide the best opportunity to accelerate progress in attaining the goals of the public and private structural genomics efforts. Participants agreed that emphasis should be placed on fostering cooperation and collaboration between the public and private efforts. An International Structural Genomics Organization (ISGO) was formed to coordinate and promote these goals. Three representatives from the global community in structural genomics (Tom Terwilliger, United States; Udo Heinemann, Germany; Shigeyuki Yokoyama, Japan) were elected to a committee to help direct and formulate policy until the next International Conference on Structural Genomics in Berlin in October 2002. In the coming years, it will be imperative for the ISGO to coordinate and maintain a bioinformatics infrastructure that will be freely accessible to the public, as well as to facilitate the timely transfer of novel methods and new results to the research community.

At Airlie, the ISGO committee on information exchange strongly recommended that all structural targets be openly disclosed and their current status posted. Currently, the NIH-funded consortiums are developing and maintaining their own Web-based databases that list targets and tracking information on the status of any structure determinations (*14*). Participants at the Airlie Center meeting also considered the need for a central target registry; indeed, NIH has since begun developing a target registry at the PDB for its own research centers. This Web site could serve as a template for the international effort. The committee also encouraged the PDB to store experimental protocols and data as well as software, particularly as this information may no longer be easily accessible if the new

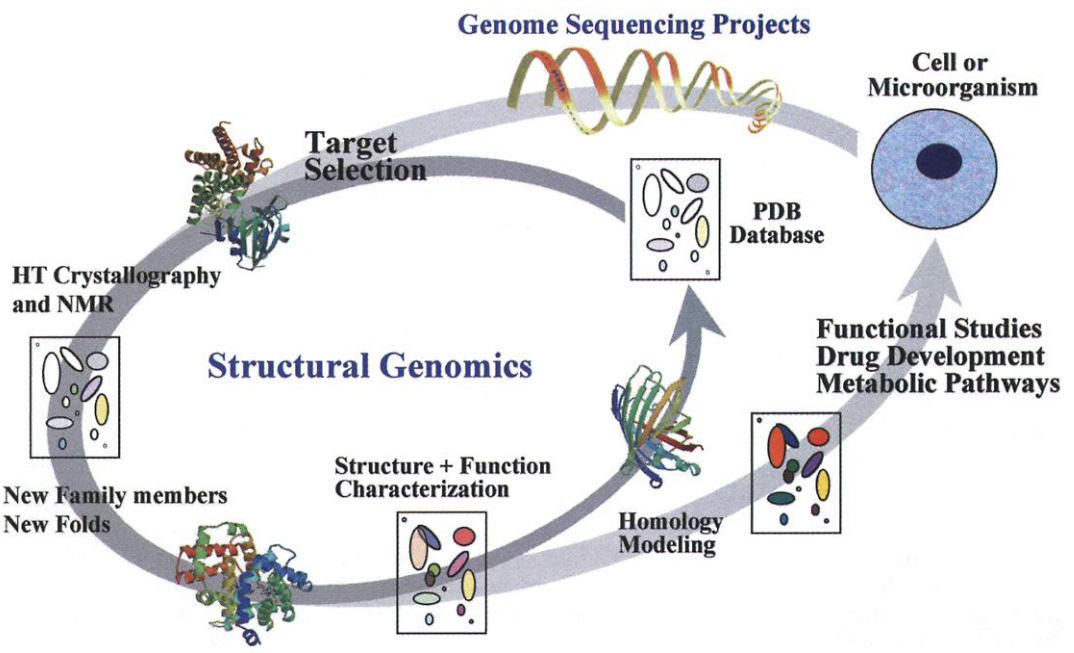

Fig. 1. Global structural genomics efforts will be major players in completing the protein family and fold landscape. The rectangular panels represent our current knowledge of the set of protein sequence families, showing whether they contain any 3D structural examples (colored encircled regions) or not (white encircled regions). The amount of color increases as more structures are determined experimentally. Only a small fraction of the protein families may not contain a known 3D structure (small circles), but the majority of the fold landscape will be represented, permitting homology modeling of most of the remaining and new gene sequences.

structures are not all published. The committee also suggested establishing organized access to protein target-related materials (cDNA clones, expression vectors, expression constructs, and purified proteins). Again, NIH is in the process of developing such a central resource for storage of materials for its own research centers. The PDB has always been the depository of biological macromolecular structure data (*15*), and it has been a constant challenge for curators at the PDB to keep up with exponential increases in data submission. This data curation will become an even bigger challenge with the development of structural genomics programs. Perhaps the biggest challenge of all will be the maintenance of the present high quality and reliability of these data. However, there are no current plans for triggering automatic structure deposition into the PDB, an issue that was hotly contested and ultimately rejected at the Airlie Center meeting by the data curation task force.

Another ISGO task force is compiling a comprehensive list of new data items to be collected in the PDB. Final structural information will continue to be deposited in the PDB, but the question remains as to what to do with the "unfinished" structures that either are incomplete or cannot pass the stringent analysis conducted on depositions to the PDB. Currently, in the United States, these "unfinished" structures will be kept in the individual NIH-funded PSI structural genomics center sites and will be available for public access. Although NMR and x-ray crystallography–based structure determinations will provide analogous atomic coordinate output files, additional data outputs, such as *B* value analysis and local density of NMR restraint values, will require different formats.

At the Airlie Center meeting, after much discussion and debate, a consensus was reached that most structure depositions should follow completion of refinement in a short time, but in some cases a 6-month maximum time lag could be allowed before public release of structural results. However, NIH has decided that the time lag should be much less for the PSI pilot centers, allowing just 4 to 6 weeks from completion of a protein structure to deposition in the PDB and public release of the coordinates. Thus, NIH will require that the PSI consortiums quickly publish and rapidly file any patent applications, when appropriate, to adhere to this time restriction. The patent systems in the United States are different from those in other countries, and individual policy decisions on coordinate release will likely vary depending on national location. However, the Japanese delegation argued strongly that 6 months would be necessary to fulfill their obligation to the Japanese government for securing patent protection, based on public funding for their structural genomics projects and an expected return on that financial investment.

Airlie Center meeting participants were well aware of the financial gains for holders of patents on pharmaceutically and agriculturally important structural targets. While recognizing the importance of protecting the inherent value of structural information and new technological innovations, the ISGO task force recommended a policy of open information exchange, encouraging international cooperation in the structural genomics community for both the public and private efforts, with the caveat of limited delay to protect favorable patent acquisition. In February 2000, the U.S. Patent and Trademark Office started issuing a number of 3D crystal structure patents. One of the main issues concerning disclosure stems from the expectation that most of the new structures will likely not have a known function at the time of structure determination. Without functional data, it would be difficult to identify the "utility" of the protein coordinates by themselves, making the validity of any patent application questionable. The ISGO task force is currently working on these issues.

The ISGO task force has also recommended that short communications of deposited structural results be published in peer-reviewed journals, preferably in electronic form; publication of full-length papers was also encouraged. Clearly, publication could become a severe bottleneck in the release of structural results as structures are more rapidly determined. New journals such as the *Journal of Structural and Functional Genomics* (Kluwer Academic Publishing) have sprung up to meet the anticipated need for fast review and increased demand for publication space. Other existing journals, such as *Acta Crystallographica* and *Proteins*, are also expected to cater to the new genre of structural genomics papers. Along with structural information, the task force also recommended a policy of disclosing methods that are relevant to the submitted structural data—for example, any new technology development and implementation that was critical in the HT structure determination process pipeline.

Although the field of structural genomics is still in its infancy, a quite astonishing amount of new technology development and progress on worldwide policy issues has already emerged. This field is expected to continuously evolve as more HT technology comes online and as more targets are carried through the structure determination pipeline. The initial guidelines and policies will certainly suffice for now, but they will need constant reassessment and refinement as the technology advances and the database of target structural information expands. Global structural genomics efforts have gotten off to a very good start, have attempted to set reasonable policies that can be adhered to, and have identified problems and challenges that need resolution in the immediate future. The new technology developments—whether in protein expression, crystallization, robotics at the beamline, or bioinformatics and data manipulation—will provide invaluable new tools for all structural biologists. However, the real winners will be the scientific community at large, who will be provided with a new and expanding database of structures that can be more quickly and freely accessed than anyone previously thought possible.

References and Notes

1. A. Bateman *et al.*, *Nucleic Acids Res.* **27**, 263 (2000).
2. R. C. Stevens, I. A. Wilson, *Science* **293**, 519 (2001).
3. S. K. Burley *et al.*, *Nature Genet.* **23**, 151 (1999).
4. S. H. Kim, *Nature Struct. Biol.* **5** (synchrotron suppl.), 643 (1998).
5. T. C. Terwilliger, *Nature Struct. Biol.* **7** (suppl.), 935 (2000).
6. G. T. Montelione, S. Anderson, *Nature Struct. Biol.* **6**, 11 (1999).
7. The RIKEN Genomic Sciences Center is located at the Yokohama Institute, Japan (www.rsgi.riken.go.jp). SPring-8 is operated by the Japan Synchrotron Radiation Research Institute (www.spring8.or.jp).
8. The Protein Structure Factory is located in Berlin, Germany (http://userpage.chemie.fu-berlin.de/~psf).
9. See the committee report on integrated projects in functional genomics relating to human health (www.cordis.lu/life/generic/integ_proj.htm).
10. The Wellcome Trust Centre for Human Genetics, Division of Structural Biology, is located at the University of Oxford (www.strubi.ox.ac.uk).
11. See the Welcome Trust "News Online" article on structural genomics (www.wellcome.ac.uk/en/1/awtpubnwswnoi26ana5.html).
12. See the Clinical Genomics Centre: Proteomics Web site (www.uhnres.utoronto.ca/proteomics).
13. See the NIGMS Web site for the Protein Structure Initiative (Structural Genomics) (www.nigms.nih.gov/funding/psi.html). The Airlie Center agreement and reports can also be found at this Web site.
14. There are seven NIH-sponsored pilot centers funded in the Protein Structure Initiative: the Midwest Center for Structural Genomics (www.mcsg.anl.gov), Northeast Structural Genomics Consortium (www.nesg.org), New York Structural Genomics Research Consortium (www.nysgrc.org), Southeast Collaboratory for Structural Genomics (www.secsg.org), Berkeley Structural Genomics Center (www.strgen.org), *Mycobacterium tuberculosis* Structural Genomics Consortium (www.doe-mbi.ucla.edu/TB), and Joint Center for Structural Genomics (www.jcsg.org).
15. H. M. Berman *et al.*, *Nucleic Acids Res.* **28**, 235 (2000).
16. D. Vitkup, E. Melamud, J. Moult, C. Sander, *Nature Struct. Biol.* **8**, 559 (2001).
17. E. Marshall, *Science* **292**, 188 (2001).

18. We thank J. Norvell for helpful discussion and comments, M. Elsliger and M. Pique for graphics, and M. Patch for manuscript preparation. Supported by NIH grant P50 GM62411 (I.A.W., R.C.S.), the Skaggs Institute for Chemical Biology (I.A.W.), and the RIKEN Genomic Sciences Center (S.Y.).