

**Second International Structural Genomics Meeting
Sponsored by NIGMS, RIKEN/MEXT, and the Wellcome Trust**

Airlie Center, Virginia, USA

April 4-6, 2001

Agreed Principles and Procedures

Coordination of International Programs in Structural Genomics

This document reports the principles agreed at the April 4-6, 2001 meeting of representatives of the structural genomics community. Its purpose is to generate further co-operation in the structural biology and general scientific communities.

This Airlie Agreement builds on the agreement produced following the first international meeting in Hinxton, UK, in April 2000. The broad overall goals and principles are unaltered. Policy extensions and more detailed definitions are based on the initial reports of the five task forces that were established following the first meeting, and discussions at the second meeting. The amended reports of the task forces form appendices to this document.

The field of structural genomics continues to evolve very rapidly, and it is expected that further policy revisions in many areas will be made at subsequent meetings of the community.

I. Introduction

Success of the genome sequencing projects and major advances in methods of protein structure determination have led the structural biology community to propose the large scale mapping of protein structure space. This structural genomics initiative aims at the discovery, analysis and dissemination of three-dimensional structures of protein, RNA and other biological macromolecules representing the entire range of structural diversity found in nature. Such a complete knowledge will facilitate fundamental understanding and applications in biology, agriculture and medicine. The three-dimensional structures will be crucial for rational drug design, for advancing catalysis in chemistry and biotechnology, and for diagnosis and treatment of disease, as well as for advancing basic principles of biology. A broad collection of structures will provide valuable biological information beyond that which can be obtained from individual structures.

This opportunity is made possible by rapid progress in several related key technologies. These include the construction of synchrotrons and high-field NMR instruments, the MAD method of phase determination, high throughput cloning and recombinant expression, a flood of

information from genome sequencing projects, and bioinformatic methods for fold assignment, model building, and prediction of function.

The following document outlines issues related to achieving this expansion of knowledge. The goal is to encourage harmonious cooperation among a broad range of public and private sector institutions in the international effort to characterize macromolecular structures in living organisms on a pan-genomic scale.

II. Goals

A. Specific goals

1. Large scale determination and analysis of three-dimensional structures.
 - 1.a To determine by experimental methods a representative set of macromolecular structures, including medically important human proteins and proteins from important pathogens and model organisms.
 - 1.b To provide models based on sequence similarity to significantly extend the coverage of structure space.
 - 1.c To derive functional information from these structures by experimental and computational methods.
2. Development of methods for Structural Genomics.
 - 2.a Methods of selecting representatives of protein families based on enhancement of structure space coverage, or functional significance.
 - 2.b High-throughput methods for production of target proteins suitable for structure determination.
 - 2.c Methods for high throughput data collection.
 - 2.d Methods for automated determination, validation, and analysis of 3D structures.
 - 2.e Methods for homology-based modeling, related methods and validation of modeled structures.
 - 2.f Informatics systems to optimize and support the process of structure determination.
 - 2.g Bioinformatics methods for assessing biological function based on structure and other linked biological information sources.
 - 2.h Methods for more challenging problems of production and structure determination such as those involving membrane proteins and multimolecular complexes.

B. Programs needed.

1. Financial and organizational support for structural genomics projects.
 - 1.a International network to co-ordinate and promote efficient application of resources and rapid dissemination of methods and results; to coordinate policies, standards, and formats; and to promote access to unique resources such as synchrotron and high field NMR facilities. To this end, an international organization shall be formed to advance the interests of the structural genomics community. The best long term form for this is not yet clear. As the first step in this direction, one representative

from each of the three principal constituencies has been selected by the community to form an executive committee. The committee is charged with organizing international affairs until the next meeting, including further work by the Task Forces. The committee may co-opt others, as it sees fit. Further evolution of the organization is expected to follow. The following have been elected to serve on the Executive Committee: Tom Terwilliger (USA), Shigeyuki Yokoyama (Japan) and Udo Heinemann (Europe).

- 1.b Support for the collection, archiving and dissemination of detailed structural information, including atomic co-ordinates, as well as experimental data, protocols, and materials.

III. Cooperation

A. Public funding agencies can cooperate:

1. By implementing the agreed policies for deposition, release, quality standards, and formats.
2. By providing sustainable support for public programs in structural genomics.
3. By encouraging and supporting appropriate international collaborative programs.

B. Information and Material Release in the National Structural Genomics Programs

1. The primary impetus for structural genomics is to obtain a base of freely available structural information and tools that will support advances in wide areas of biology and medicine. Free exchange of data and materials is essential to the success of this effort, including the timely deposition of coordinates, data, and protocols.
2. The community agrees to work to maximize the pool of structures available to the public in all countries, as a basis for both academic research and commercial use.
3. For the structural genomics programs with public funding, the following guidelines for release of structural data should be supported:
 - 3.a. The community agrees to work toward the timely release to the public of all basic structural data. The promptness of data release is expected to improve over time.
 - 3.b. Structural genomics laboratories with public funding are expected to deposit their structure co-ordinates and other agreed mandatory data in the PDB immediately on completion of structure determination. In most cases, data release to the public will follow in a short time. It is recognized that in some cases release can be delayed by up to six months after deposition. This should be sufficient for investigators, for example, to assess intellectual property prospects and to file a patent application if desired.
 - 3.c. All structural genomics laboratories with public funding will fully adopt these deposition and release policies no later than April 2002.

4. Public information on progress of projects. A primary mechanism for encouraging compliance with the guideline of timely release will be openness of progress tracking for projects.
 - 4.a Structural genomics laboratories with public funding shall adopt a policy of open exchange of target information, in order to facilitate target selection and to avoid unnecessary duplication of effort. It is recognized that publication of these data may have some disadvantages, but on balance these are out-weighed by the advantages.
 - 4.b As recommended by the task force on data tracking, each laboratory will maintain a public site, listing target sequences and the status of the work, using a simple, standardized format. The current standards are listed in the appendix. An ongoing working group is responsible for the implementation and operation. This system shall become operational by June 1, 2001.
 - 4.c The future need for a central registry should be considered further. In particular, international laboratories should evaluate the registry being developed by the NIH.
5. Short scientific papers.
 - 5.a Ensuring high quality of released structures is a priority. In order to help achieve this, structures released by members of the public programs may be accompanied by a short, peer-reviewed paper. These papers could be similar in format and content to the publications of small molecule crystal structures in Acta Cryst. C. Detailed procedures are outlined in the Task Force report. Electronic publication is encouraged. Papers should maximize the inclusion of relevant structural and functional information. Criteria for acceptance of publications will include those recommended by the Task Force on Numerical Criteria.
 - 5.b The key requirement is that the whole process of publication be completed rapidly.
 - 5.c Full-length publication is of course also possible. Any publication prior to the end of the maximum six month delay period will trigger data release, following accepted practice in structural biology.
6. Technology exchange between structural genomics laboratories. The community recognizes that there is much to be gained by an open exchange of the new technologies being developed in each structural genomics laboratory. Therefore it adopts a policy of open exchange of information on emerging technologies. In particular, we encourage exchange of protocols and software. This policy will greatly reduce duplication of effort, and greatly speed progress in the field. To these ends, a central clearing house for this information shall be established by the international organization.
7. Assurance of Data Quality.

- 7a. The community recognizes that the production of high quality structures is an integral part of any high throughput structural genomics operation. That is, quality is not to be sacrificed in the interests of quantity.
 - 7b. The community accepts the report of the task force on numerical criteria for assessing structure quality. For the time being, the numerical criteria recommended by the task force are adopted, as a minimum set of measures to be associated with all released structures. Experience using these criteria, together with new developments in the field, will make periodic reassessment necessary.
 - 7c. Structure depositions will be accompanied by experimental data in a defined format. For crystallographic studies, these will include structure factor amplitudes, and un-merged, un-scaled integrated intensities for all data sets used in the structure determination. For NMR studies, these will include time domain data sets. It is desirable to move as far as possible towards archiving the raw data, as the data management technologies permit.
8. Curation and Data Archiving.
- 8a. The community endorses the key recommendation of the Task Force on curation and deposition, namely, that overall objective is to capture the level of detail presented in the material and methods section in a good journal paper. Among other benefits, these data will provide the basis for rapid publication. To this end, an appropriate comprehensive set of data items will be collected in a consistent format. Progress in establishing this set is described in the report of the Task Force.
 - 8b. The Task Force is asked to continue its work, as outlined in the report, with the goal of completing all data item definitions and establishing recommended procedures by April, 2002. It is anticipated that small workshops will be held to speed progress towards this goal. It is desirable that a template be provided for structural genomics laboratories to use in preparing their depositions.
 - 8c. In the long run, it is also desirable to collect information on abandoned target studies, and methods for accomplishing this should be developed.
 - 8d. Organized access to material such as clones, cell lines, and protein samples is also encouraged, provided that satisfactory procedures can be put in place for archiving, storage, and dissemination.
- C. Relationship to industrial activities.
- 1. The public structural genomics community should explore productive relationships with industrial partners to further the goals of structural genomics.
 - 2. International efforts should be made to facilitate the eventual deposition of structures determined in the private sector, and to promote harmonious cooperation and exchange between the public and private sectors.

IV. Intellectual Property Rights

Raw fundamental data on the shape of natural protein molecules, including 3D positional coordinates, should be made freely available to researchers everywhere. However, intellectual property protection for inventions based on these can play an important role in stimulating the development of important new health care projects.

Public funding for structural genomics has varying degrees of support for fundamental science and for potential commercial exploitation. The data release policy described earlier has been designed to accommodate these differences, at the same time optimizing the speed of release of data as much as possible under all circumstances.

Fundamental research underpins all practical uses and applications. Policy makers are urged to preserve and promote the free access and exchange of scientific information among scientists engaged in basic research. This community welcomes efforts around the world to harmonize patent law.

We also encourage efforts to strengthen the utility requirement for patentability. This community is concerned about the implications of the granting of patents based solely on the submission of three-dimensional structural co-ordinates, without any identified non-trivial utility.

V. Future Meetings

Further meetings of representatives of the structural genomics community are anticipated for the continued reexamination of these issues and to further develop these principles and guidelines as the field expands and evolves. The next meeting will occur in Berlin, Germany in October 2002.

These principles were supported by the participants in the Second International Structural Genomics Meeting in the Airlie Center, Virginia, USA, April 4-6, 2001.