nature structural biology
molecular form & function
november 2000
vol. 7 no. 11

Breadth plus depth

We are pleased to present a supplement on Structural Genomics to accompany this issue of *Nature Structural Biology*. This supplement outlines plans formulated over the past few years for a major new structural biology initiative — a move away from primarily hypothesis-driven research (in which the general biological function of a protein is known before its structure is solved) toward a plan to obtain representative structures of every fold family (without necessarily having any prior knowledge of their biological functions). Structural Genomics researchers believe that this strategy will allow biologists to address more problems in depth — that the structural data that they generate will serve as a solid foundation for further research on the mechanisms and actions of proteins and nucleic acids, such as the kind of structure-function work featured in every issue of *Nature Structural Biology*.

The phrase 'Structural Genomics' is often loosely defined, as the reviews in the supplement show. For example, some tend to think of it as primarily the development of high throughput technology for structure determination, others as a strategy for determining enough structures to cover all of 'fold space', others as a plan to determine all of the structures of proteins encoded within a single genome, and still others as the means to solve all structures related to a biological process (such as signal transduction or cancer). In reality, the term encompasses all of these goals, but regardless of how it is defined, to most of its participants Structural Genomics is a logical follow-up to the large scale sequencing projects of the last decade. Those projects have provided a wealth of new genes encoding proteins with unknown structures and functions. However, it should not be forgotten that it will cost a great deal more to solve 10,000 structures than to sequence 10,000 genes.

Although small pilot projects, which aimed to demonstrate the feasibility of structural genomics, have been ongoing for several years, researchers have now received strong financial backing, enough to jump-start the field in earnest. In late September, the National Institute of General Medical Sciences in the USA announced a program to fund seven centers for Structural Genomics, giving ~$4 million to each center for the first of five years. Japan has also allocated a large amount of money (more than $40 million per year) for structural and functional genomics. It has been said that these allocations are new funds — that these new initiatives will not cut into financial support for traditional structural biology research programs. If this is indeed the case, then over the next five years, more money than ever will be flowing into structural biology labs, which should be an exciting prospect for structural biologists.

Ten years ago, these high throughput structure determination plans would have been hard to imagine. But over the past decade researchers have begun, with great success, to routinely use a variety of techniques and tools — such as synchrotron radiation, cryofreezing techniques, and multiwavelength anomalous dispersion (MAD) methods — to increase the speed at which their research is done. Funding should continue for 'traditional' structural biology, which has served the community well for many decades, as well for the new Structural Genomics centers. Strong support for the different approaches should facilitate new technological developments, a decline in cost per structure, and the investigation of more difficult research problems.

Key issues

As the large scale Structural Genomics projects around the world get started, there are numerous important issues to address, not the least of which are: international cooperation, the relationship between public and private initiatives, and deposition policies. The reviews in the accompanying supplement raise these concerns (and in some cases tackle them). However, none of these issues is clearly resolved at this point.

Right now, there is no defined international agreement to cooperate on Structural Genomics. It is imperative that cooperation occurs, to avoid unnecessary duplication of effort and wasted resources in solving thousands of protein structures. Researchers in North America and Japan are communicating and advising one another, but in general are pursuing independent plans; efforts in Europe are still in their infancy. Recently, a meeting was held specifically to discuss this issue (see *http://www.nigms.nih.gov/news/meetings/hinxton.html*), and its message was encouraging, but no specific agenda about how to manage international cooperation was put forth. Regular meetings will be necessary in the future to keep this issue at the forefront.

Efforts in the private sector will also have to be taken into account to maximize the potential of this field. Already, several biotechnology companies, such as Syrrx and Structural GenomiX, are planning to focus on high throughput structural determination. This may be good news for the structural biology job market (as these companies will provide new alternatives to structural positions in academia or large pharmaceutical companies), but it is not obvious how these young companies will fare. It is widely believed that the future of drug discovery will be guided by a combination of combinatorial chemistry and structure based drug design, and thus these companies hope to make their mark by selling structural information to the pharmaceutical industry. They may take up the model offered by genomics companies such as Incyte and Celera — of selling access to databases, in this case ones containing a large amount of structural data on potential drug targets. Or, they may try to sell access to individual structures.

Time will tell if private companies can indeed turn high throughput structure determination into a profit-making enterprise. It seems that they may have their task cut out for them. Structures are expensive to generate (and thus will presumably have to be sold for rather handsome sums to make determining them worthwhile to the private sector). Related to this, it is not obvious what kinds of intellectual property rights will be available to Structural Genomics researchers, in the private or the public sector. Moreover, in addition to the government funding of Structural Genomics, there is a movement afoot to organize a consortium of large pharmaceutical companies to invest heavily in structure determination, with the intention of adding structures to the public domain without restriction. Nevertheless, the leaders of the young companies focused on high throughput structure determination, some of whom also have ties to the publicly funded Structural Genomics efforts, seem confident that their approach will be valuable and complementary to the public efforts. For example, they note that their companies may focus more on depth than breadth — on determining the structures of target pathogenic proteins and all of the related human proteins, for example, rather than on solving a single representative member of a fold family. We should know within the next few years if their efforts will be successful, and if so, whether the publicly funded efforts will feel the need to respond.

Another question that should be tackled soon is: when should data from the publicly funded Structural Genomics projects be deposited into the Protein Data Bank, as has been agreed by those receiving grants? The information produced by these projects should be available as soon as possible, both to benefit communication between Structural Genomics centers, and to benefit biological research in general. Some proponents of Structural Genomics have advocated a statistical approach, such as automatic deposition when the R_{free} value reaches a certain level for a structure determined by X-ray crystallography. But most feel that an investigator must decide when the quality of a structure is good enough for deposition, since quality is not always a black-and-white issue, and since the release of poor quality structural information could be very misleading. For now, it has been generally agreed that coordinates will be released immediately after refinement (admittedly a fuzzy definition), and that publication should not delay release.

Our interest in Structural Genomics

At *Nature Structural Biology*, we are keenly interested in Structural Genomics, as shown by our coverage of this emerging field in the accompanying supplement. However, we do not intend to publish every Structural Genomics paper that we receive. We will apply the same editorial standards to papers coming from Structural Genomics centers; we will look for a desirable combination of novel results, interest to a wide audience, and insight into a biological process. The journal will continue to follow the development of this field, and we hope that both the structural data and the new methods that are generated will make it easier to address a wide variety of biological questions — such as the ones presented on the pages of this issue — in even greater depth.