

PREFACE

Rational and Irrational Exuberance in Structural Genomics

Structural Genomics is all the rage these days. Much of the drive towards high throughput structure determination comes from the need to process the influx of DNA sequence that continues to gush out at an alarming rate. There is nothing like having a complete description of a genome to make you realize that what you really needed was a complete description of the proteome, or metabolome, or even a whole cell. The incompleteness of the information at hand is driving the acceleration and industrialization of biology. Twenty years ago the cloning and sequencing of a single (albeit interesting) gene was considered sufficient to earn a PhD, whereas nowadays determining a crystal structure may well be part of an undergraduate project.

The growing popularity in structural biology has pushed the simplification of the process (as has much more powerful computers and software). The influx of many people to the field who see x-ray diffraction techniques as a means of generating vast amounts of information about a biological molecule, rather than an end in itself, has broadened the background and expertise in structural biology. Thus, the reason that many people go into crystallography is to understand biology at atomic resolution, not just to learn the art of crystallography. In the near future, we expect that the vast majority of structural biologists trained will not be crystallographers *per se* but biologists asking questions about function and mechanisms of action of given proteins. Given this, we should expect a rational exuberance about the field of structural genomics, as it means that in the foreseeable future there may be a structure of your favorite protein available, even before you know you need it. This will inevitably speed up the discovery process in biology, as the drudgery associated with structure determination (setting up crystallization trials, waiting for your turn on the area detector, etc.) does little to advance our knowledge.

There is also exuberance, indubitably rational, that significant funding and effort is going into structural biology at this critical time to boost our knowledge of structure/ function relationships. There is fear, almost certainly irrational, that this signals the end of pure crystallography. “What is happening to our noble science?” is the lament of some of the old school crystallographers who fear that the industrialization of the field will come at the expense of intellectual rigor. A tool is more useful if it is easy to use. Although there needs to be a handful of hard-core crystallographers to push the envelope of the technique, and indeed to solve the really difficult structures, the vast majority of the end users would prefer just to have the information. Is knowing the difference between $P2_12_12$ and $P2_12_12_1$, really something to Bragg (*sic*) about?

By increasing resources targeted to large scale structure determinations we can expect to have many of the simple structures done in the next two to five years. Also, the techniques that necessarily must be developed for high-throughput crystallography will enable the whole community to solve structures more efficiently. We can see this by looking at the large scale sequencing projects-how many labs are still using P^{32} to label their sequencing reactions? The trivialization of routine structure determinations means that attention will turn towards the challenging problems that even now are starting to crack: the recent ribosome structures at atomic resolution by various groups are exciting examples of this [1, 2, 3].

Just as Henry Ford took a very complicated process done by experts and broke it down into simple steps, structural genomics will change the way we do structural biology. However, we can't apply all of the tenets of assembly line manufacturing to structural biology. In a manufacturing process the end product is

predetermined, whereas structural biology is a discovery process rife with surprises. Still, to meet the output goals of the new structural genomics initiatives, both public and private, the assembly lines must be developed now. The goal of structural genomics-the breakdown of the process into automatable steps-is anathema to academia, where the aim is to provide an education for students and post-docs. This dictates that the structural genomics projects are intrinsically unsuitable for individual academic labs. The question that all the structural genomics projects face is: how best to do this? Should we be taking current techniques and methodologies and automating them? Or should we be reviewing the whole process so that it is optimized (where possible) for automation? It may happen that we can achieve the near-term goals of structure solution fastest by the first path. However, we may then be stuck in a QWERTY-type situation, where the process itself becomes a rate-limiting anachronism. We must think hard about all of the steps in the process of structure determination and re-arrange the process intelligently, instead of just being clever within a tired framework.

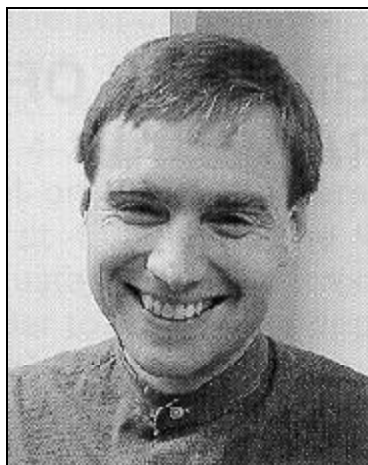
At the end of the day, greater understanding of biology will lead to better medicines, agricultural products and industrial catalysts. As other authors have pointed out, various methods are currently used by groups across the country to develop new pharmaceuticals. The development of new drugs, such as the second generation HIV protease inhibitors and neuraminidase inhibitors, has proven rational drug design a viable technique for compound design. Combinatorial chemistry, sometimes unfairly referred to as irrational drug design, has also been shown to be a strong approach for finding some chemical compounds that affect the activity of proteins. Rational drug design has been successful for some, others have found combinatorial chemistry more fruitful. We believe that the strongest long-term strategy will be the integration of this 'rational' and 'irrational' exuberance.

Because different organisms and pathways are being targeted by the different public effort consortia, we hope that each will produce new and different protein structures. These diverse efforts, along with the industrial efforts, should give the biology community both new structures and new technologies. In addition, even if some of the same (or related) sets of proteins are structurally determined, we will learn a great deal by making comparisons between proteins and studying how different sequences give common folds.

Databases of structures will become increasingly important tools for biologists as comparative biology becomes more powerful. As more data flows into these databases, modeling and other predictive techniques will become increasingly accurate. Eventually this will lead to the ability to design de novo proteins for novel functions. This will lead to a true renaissance in medicine, industrial biocatalysis and agriculture. We still believe however, that there are decades of work to be done and the industrialization of this one aspect of biology will not answer all of our questions. Meanwhile, our growing ability to produce more structures, to make more compounds and to understand more biology will continue to make this a very interesting era in science.

References

- [1] Ban, N. *et al.*, *Science* **289**: 905-920 (2000).
- [2] Cate, J.H. *et al.*, *Science* **285**: 2095-2104(1999).
- [3] Clemons Jr., W.M. *et al.*, *Nature* **400**:833-840(1999).



Tom Peat
Structural GenomiX
10505 Roselle St
San Diego, CA, 92121
tel. (+ 1 858) 558 4850
fax. (+1 858) 558 4859
email: tom@stromix.com



Janet Newman
Structural GenomiX
10505 Roselle St
San Diego, CA, 92121
tel. (+ 1 858) 558 4850
fax. (+1 858) 558 4859
email: janet-newman@stromix.com