

Comment

## An idea whose time has gone

Gregory A Petsko

Address: Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02454-9110, USA.  
Email: [petsko@brandeis.edu](mailto:petsko@brandeis.edu)

Published: 2 July 2007

*Genome Biology* 2007, **8**:107 (doi:10.1186/gb-2007-8-6-107)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2007/8/6/107>

© 2007 BioMed Central Ltd

When I was in college, I had a roommate who was a little weird. OK - more than a little weird. One night, at about 2 am, the campus police woke me up to tell me that he was sitting in the middle of the main street of our college town, naked, with a blanket over his head. They wanted me to go and bring him home, so they didn't have to arrest him. I got dressed, walked into town, and sure enough, there he was, sitting naked in the middle of the street with a blanket over his head. I walked over to him, called his name, and said, "Warren, why are you sitting naked in the middle of Prospect Street with a blanket over your head at 2 o'clock in the morning?" He peered out from under the blanket, looked up at me, and said, "It seemed like a good idea at the time."

My guess is that whoever was responsible for the Protein Structure Initiative (PSI) must have felt the same way. The PSI is a fancy name for the US 'Structural Genomics' effort. The stated aim of Structural Genomics is determination of the three-dimensional structures of all proteins. Its members and proponents claim that this aim can be achieved in four steps: first, organizing known protein sequences into families; second, selecting family representatives as targets; third, solving the three-dimensional structures of targets by X-ray crystallography or NMR spectroscopy; and fourth, building models for other proteins by homology to solved three-dimensional structures.

The Initiative currently funds ten large centers scattered around the United States (similar efforts exist in Europe and Japan). They are supported for a five-year period to the tune of about \$300 million total. The PSI website [<http://www.nigms.nih.gov/Initiatives/PSI/>] states that "Expected benefits from the PSI include: structural descriptions to help researchers discover the functions of proteins, design experiments, and solve other key biomedical problems; faster identification of promising new structure-based medicines; better therapeutics for treating both genetic and infectious diseases; and development of

technology and methodology for protein production and crystallography." The National Institute of General Medical Sciences (NIGMS), the main branch of the US National Institutes of Health (NIH) funding the PSI, is currently engaged in an assessment of the PSI. I know this because I was asked to provide my views on the initiative. I'm afraid they weren't very complimentary.

The PSI actually has had two incarnations. The specific goals of PSI-1 (which existed from 2000-2005) were to develop methodology and technology to increase success rates and lower costs of structural determination; to construct and automate the protein production and structural determination pipeline; and, finally, to determine unique protein structures. Lots of unique protein structures. By 2005, it became apparent that most of these goals weren't being met, nor were they likely to be met in the near future.

One of the many great scenes in the wonderful old Errol Flynn movie 'The Adventures of Robin Hood' is the archery contest. The finest archers from all over the kingdom are gathered in Nottingham to compete for a gold arrow. After many rounds, only two competitors are left: Robin Hood, disguised as a tinker, and one of Sir Guy of Gisborne's archers. After they both shoot and both hit the bulls-eye, Robin Hood asks that the target be moved back, "to a fit distance for men to shoot at." When it looked like the PSI wasn't going to be able to meet its goals, what did it do? It moved the target in. The specific goals of PSI-2 (funded from 2005-2010) are now to increase the number of sequence families with structural representatives, including families with high biological impact; to continue methodology and technology development, especially for challenging classes of proteins such as membrane proteins; and to facilitate the use of structures by the broad scientific community. These goals are so squishy, it would almost be impossible not to meet them. Or for it to matter much if they were.

But as I considered my assessment, it became clear to me that even if the goals of the original PSI-1 could be met, I wouldn't care. Nor, I think, should most anyone else.

Do we really need a catalog of structures? What will that teach us? We already know that proteins are composed of beta sheets and alpha helices, interspersed with loops. Filling the fold catalog might be of interest to bioinformaticists, but why should they drive the science that others do?

And I reject categorically the notion that enough structures will allow us to build homology models for every sequence. First of all, the methods for recognizing which fold a sequence belongs to aren't that robust. False positives seem to be fairly rare, but false negatives abound. Second, homology models aren't very accurate when the sequences are less than about 50% identical, which happens most of the time. For drug discovery and understanding biochemistry, accurate models are essential. Nor is it clear to me that when you have a structure, you necessarily have learned all that much about the function of the protein. The coupling between sequence, overall fold, and function is rather loose. Even when a fold has an accurately annotated function, which is not as often as it ought to be, there is a high probability that a homolog with less than about 50% sequence identity will have a different biochemical and cellular function. My guess is that a large catalog of structures will just lead to even more missannotation of function by homology - one of the greatest problems in genomics today.

It's also clear, I think, that it isn't enough to have the structure of a protein; you need to have the right structure. Small changes in sequence can lead to big changes in oligomerization state, which in turn can lead to changes in active site geometry and function (a good example can be found in a recent study by Wei *et al.*: *Identification of functional subclasses in the DJ-1 superfamily proteins*. *PLoS Comput Biol* 2007, **3**:e10). We don't have any method for predicting the oligomerization state of a protein from its sequence or its homology to a protein of known structure when such changes occur. And what about changes in conformation? It did Novartis no good to have the structure of the Bcr-Abl tyrosine kinase for the design of the anti-leukemia drug Gleevec. The kinase exists in two structural forms, open and closed, and when Gleevec was being developed the only existing structure was of the open form. Gleevec binds to the closed state.

As for the stated goal of the PSI to develop technology to make protein crystal structure determinations easier, I'm afraid I have to say, so what? Solving structures isn't really the rate-determining step for most good structural biology projects - the bottlenecks are usually biochemical. The PSI's focus on high-throughput methods of expression, purification and crystallization means that it isn't really furnishing

solutions for most of those problems. To be perfectly selfish, I have to say that it hasn't made any contribution to my own work, and I'd be willing to bet that it hasn't contributed much to yours either.

Another problem I have is with the entire mindset of such an initiative. It is focused on cranking stuff out as fast as possible, with little attention to whether the structures that it's determining are worth determining. I also reject categorically the notion that all protein structures are worth having. Structures have value when they are part of a larger effort to understand the biochemical and biological functions of the protein in question. Doing them in isolation has no more intellectual content than does assembling a car. As a structural biologist, I want to train people who use structure determination as part of what they do. It is not the end in itself, nor should it be, not any more. When we knew almost nothing about the universe of protein structures, every structure had value. But Adam Smith's law of supply and demand works in science just as it does in economics: with the supply of structures already in the tens of thousands, the value of any new structure in and of itself is likely to be rather small.

The argument can probably be made - almost certainly will be made - that it will be useful for pharmaceutical and biotechnology companies to have structures of all proteins from various pathogens and from certain human disease tissues. Maybe, though I doubt it - there's a big difference between a potential target and a validated one. And if such structural information is of value to the private sector, why shouldn't the private sector fund it? \$300 million over five years is petty cash for a consortium of drug companies, but I don't see them lining up to pay even that pittance for this information.

But what's a drop in the bucket to drug companies is life and death to academic research. The \$60 million a year in public money that is being spent - I would say, wasted - on the PSI is enough to fund approximately 100-200 individual investigator-initiated research grants. These hypothesis-driven proposals are the lifeblood of the scientific enterprise, and as I have discussed recently in other columns, they are being sucked dry by, among other things, an increasing trend to fund large initiatives at their expense. That \$60 million a year would raise the payline at a typical NIH institute by about 6 percentile points, enough to make a huge difference to peer review and to the continuance of a lot of important science.

I simply can't see the justification, in a time when budgets are so tight, for continuing a program that has produced little useful information, has not furnished many widely disseminated technologies or methods, and has minimal intellectual content. Regular readers of this column (all five of you) will know that I am not a disparager of big science

*per se*. Many such initiatives make a lot of sense, in large part because their information drives good small science. But I don't believe that the PSI has, or that it will.

So my overall assessment of the PSI is that it is an idea whose time has gone. Given its ability to change its shape (that is, reformulate its goals) so as to continue to suck blood - I mean funding - from the NIH, I think it isn't going to be enough to recommend that it be phased out. It should have a stake driven through its heart, and then it should be buried in a coffin filled with its native soil so that it can't rise again with the next full moon. If that seems harsh, then on its tombstone, if you like, we could engrave the words of my erstwhile roommate: "It seemed like a good idea at the time."