



nature Structural biology

molecular form & function

march 2001

vol. 8 no. 3

Making data available

Finally — after the fanfare surrounding the joint announcement of completion of working drafts of the human genome — we can read the relevant papers, which, when this issue of *Nature Structural Biology* becomes available, will have just been published in *Nature* and *Science*. There is no need to go on about the significance of this wealth of data in these pages, as the supporting material in *Nature* and *Science* will surely satisfy. However, here it is worth discussing a controversial issue that is peripheral to the science, but of considerable interest to the scientific community.

The controversy, in a nutshell, is this. The Human Genome Project is making their data freely available to the community through GenBank, a publicly accessible database. However, Celera will gain similar prestige through publication but is not being held to the same generally accepted standard — they will not be required to deposit their data into GenBank (or any other publicly maintained database, for that matter). Instead, Celera's sequence data will remain on Celera's website, and limited access will be granted to the information.

According to a press release¹ distributed by *Science*, the journal that is publishing Celera's work, the restrictions on access to Celera's data seem fairly severe, and they vary depending on whether the intended user is in academics or industry. *Science* will hold a copy of the genome at the time of publication in escrow, presumably to ensure that Celera fulfills the promise of at least limited access. At the time of this printing, there were only a few concrete details about many aspects of the agreement between *Science* and Celera, but upon publication, we hope that the precise terms will be crystal clear.

A gut reaction of some may be that this makes sense. The Human Genome Project was funded with public funds, while Celera used private capital to accomplish its goals. Thus, Celera should be entitled to hold on to its results. However, the issue here is criteria for publication, not the right to proprietary interest. Celera has the right to keep its database strictly proprietary with restricted access, but in that case, some argue that they just should not attempt to publish it. Why should Celera get to reap the rewards of a high-profile publication without adhering to accepted practice by releasing the data? After all, other industrial scientists face a similar dilemma — they cannot both publish a discovery in a timely fashion and patent it, in part because upon publication they would have to make their findings, data and reagents available to the community. In the press release describing the agreement, *Science* maintains that their policy says nothing about making data available in one particular database or another, implying that posting on Celera's website and their terms of access are sufficient.

Not surprisingly, there has been a fair amount of negative outcry. Some have termed *Science's* decision to openly grant Celera an exception to standard practice as an heretical choice, others as a step onto a very slippery slope, and still others as a sensible alternative, given the inevitable trend toward commercialization of biological results. The latter group argues that without some such agreement, the public would be given no access at all to the sequence without paying hefty fees, as Celera has a strong interest in capitalizing on its investment. They also argue that additional deals of this type may encourage more companies to release their data in the future.

The latter argument is not unprecedented. In fact, for some time, a majority of the structural community posed a similar case for supporting a hold of up to one year on structural coordinate files deposited in the Protein Data Bank. A major difference between these two situations, however, is that even though a hold on a coordinate file may be in place, the data are still lodged in a public database and will eventually be accessible to the public.

The structural community supported a hold in part because many believed that without it as an option, companies would never publish their work or deposit their data at all. But this sentiment has been changing in the community over the past few years, as it has become easier to determine structures. For example, recently, the International Union of Crystallography endorsed a no-hold policy on both coordinate and structure factor files², and the Structural Genomics community is also in favor of rapid release of data upon publication³. Consistent with these views, we have decided to make a change in the policies of *Nature Structural Biology*. While for some time we have allowed a hold period of up to six months after publication, we will no longer continue this practice. For all manuscripts submitted on or after May 1, 2001, we will expect coordinate and structure factor files to be deposited before publication and released at the time of publication.

No matter which side you stand on with respect to the agreement between *Science* and Celera, the view to the bottom line is clear: commercial interests are beginning to exert a stronger effect in many arenas, including publishing. These are tough issues for practicing scientists and for scientific journals. Clearly there will be pros and cons to both sides of every issue, and as commercial interests become even more prevalent, it may be difficult to sort out the most appropriate measures to take if the ultimate goal is to ensure easy access to important scientific information.

1. <http://www.sciencemag.org/feature/data/announcement/genomesequenceplan.shl>
2. *Acta Crystallogr. D* **56**, 2 (2000).
3. Burley, S.K. *Nature Struct. Biol.* **7**, 932–934 (2000).