

nature Structural biology

june 1998
vol. 5 no. 6

Structure coordinate straw poll

The Web version of the Editorial has links to many of the comments themselves, as well as further data on, and analysis of, the vote; see <http://structbio.nature.com/survey/results/>



855 to 410: these are the results of web-based vote, hosted by *Nature Structural Biology*, on arguments put forward in a petition by Alex Wlodawer, regarding the release of structure coordinates upon publication. Briefly, the argument put forward in the petition was: should coordinates deposited in the Protein Data Bank (PDB) be released immediately upon publication, rather than be held for one year by depositors? 855 people voted in favor of the petition, 406 voted against (four 'no' votes were tests, or duplicates from the same e-mail address). The figures would suggest a strong feeling that structure coordinates should be released immediately on publication (Tables 1, 2). But is this what the field should do? And how does the result of this poll fit into the debate?

The vote is best described as a straw poll: "an unofficial vote (as taken at a chance gathering or by letters of inquiry) to indicate the relative strengths of opposing candidates or issues"¹. Of the voters, some one hundred and fifty people provided comments (of these approximately 40 people sponsored Wlodawer's petition to the International Union of Crystallography²). The comments fall into a number of categories, many of which I will try to summarize here (some of which have been rehearsed in previous editorials³⁻⁵), as they provide a means for putting the vote in context (with the caveat that there would be a stronger motivation for those with opinions contrary to the petition to write in).

There were concerns about the poll itself. One point raised was the lack of distinction between those voters who are depositors in the structure data bases and those who use the structure data (modelers, biochemists, teachers, etc). Should votes be weighted differently, depending on who casts them? Some structural biologists would argue that the originators of the data should have a greater say in the 'regulations' that apply to the data. They after all, are the ones who take the risks and invest their time to generate the information: Peter Moore (Yale University, USA) points out that "the user community, which is probably much larger than the producer community, will, of course, vote 100% for the elimination of the hold. It costs them nothing and may gain them a lot."

But who's to say who should benefit from such results, once they have been published? Shouldn't all scientists be able to take published data and use them to advance their own research, to teach the next generation of structural biologists, and so on? The claim that publication without release of the coordinates still provides valuable information is certainly true, but it is a practice antithetical to presently accepted scientific philosophy.

The pressure to publish ahead of a thorough analysis of the structure — because of competition, grants, job searches, and so on — is the present rationale for the hold among some academic researchers, along with the fact that solving (arguably the more interesting) structures is still a non-trivial exercise. "[O]btaining a structure is much easier now than in the past,

Table 1 Breakdown of votes cast in the NSB poll on coordinate release

YES VOTE	855	(67.8%)	Academic	621	(73.6% of academics)
			Industry	67	(53.2% of industry)
			Government	59	(85.5% of government)
			Not identified	108	
NO VOTE	406	(32.3%)	Academic	223	(26.4% of academics)
			Industry	59	(46.8% of industry)
			Government	10	(14.5% of government)
			Not identified	114	
Voter total	844	(66.9%)	Total Academic		
	126	(10%)	Total Industry		
	69	(5.5%)	Total Government		
	222	(17.6%)	Total Not identified		
Total	1261	(989 respondents provided e-mail addresses, 272 did not)			

editorial

Table 2 Breakdown of voter e-mail addresses by domain

Address	Total	Yes	No
USA (.edu)	408	328	80
(.gov)	57	55	2
Sub-total	465	383	82
UK (.uk)	109	66	43
Commercial (.com)	97	48	49
Germany (.de)	61	34	27
France (.fr)	36	29	7
Canada (.ca)	26	19	7
Sweden (.se)	26	16	10
Non-profit (.org)	21	20	1
Australia (.au)	18	11	7
Switzerland (.ch)	16	12	4
Netherlands (.nl)	13	9	4
Japan (.jp)	12	12	0
Spain (.es)	12	8	4
Denmark (.dk)	10	7	3
Others	67	60	7
Total	989	734	255
(%)		(74.2)	(35.8)

but difficult crystal structures can take years to solve if twinning, ambiguous space groups, or poor diffraction are present. Until the phase problem is solved for proteins, this situation will remain the same", argues Kurt Krause (University of Houston, Texas, USA).

Small laboratories and young investigators are often cited among the comments as those most susceptible to such a change in policy. "The big problem for any young crystallographer who solved a structure is that a biocomputing person like me can harvest the coordinates very fast, thereby leaving the crystallographer empty handed", Gert Vriend (EMBL, Heidelberg, Germany) points out. Indeed, part of the problem for those opposed to immediate release is the instant availability of the data describing the information-rich structure, a situation rarely encountered in other areas of research. But had the hold not existed, wouldn't smaller structural biology research groups have evolved procedures (collaboration being the most obvious) to ensure their success, like smaller research groups in other fields that do not protect their data once it has been published?

The arguments regarding commercial/industrial interests were raised a number of times. One anonymous contributor stated that "without the hold on coordinates, industrial crystallographers, in competition with other companies, would either never publish their structures or wait until they were far less relevant. This diminishes not only their contributions to structural biology but also to the whole structural database. In order to justify the expense of solving the structures, industrial companies have to be able to recover some of those costs by holding the coordinates for use during the drug design process."

Those working in academia and who collaborate with industry are also concerned, as Michael Blaber (Florida State University) points out: "For some of us, the current funding situation is extremely difficult and, in fact, career threatening. Thus, every avenue of funding is important to pursue. This includes funding opportunities in the private sector. Protection of intellectual property maximizes the potential for private sector funding. This may include delaying release of coordinates". Another (anonymous) researcher thinks that perhaps "the one year hold should be abolished for all except industrial laboratories". Yet, the only fair policy would be one applied uniformly: how would editors judge the case for or against laboratories funded partially by industry?

More importantly, to what extent should commercial/industrial concerns influence scientific practice? Some might say that to think it is possible to ignore such forces is, in this day and age, naive. Of course, this can be turned on its head: the question would then be, how important is structural biology to industry? If biotechnology and pharmaceutical companies want to attract and maintain the best scientists, or collaborate with the best laboratories, they will surely have to foster an environment that allows structural biologists to thrive, and this must include the ability to publish competitively. The poll would suggest that this is what many of those working in industry want, with ~50% supporting immediate release ("Industry", Table 1, or ".com", Table 2).

A number of alternatives to the one year hold or immediate release were mentioned. An interim hold period, of 3–6 months from publication, was one of the less radical options suggested by a number of people (related to this, there seemed to be a strong feeling that structure factors/NMR restraints should be available at the same time that the coordinates are).

Some voters believe that improving present deposition practices would eliminate many of complaints and concerns. "A more serious problem, in my opinion, is the situation where authors fail to submit their coordinates to the PDB, even when required to do so. If this problem could be eliminated, then I think the present arrangements would work very well", comments Guy Vigers (Amgen, USA). Nevertheless, Harold Varmus, the Director of the National Institutes of Health (NIH) in the USA, in a letter circulated to the editors of journals that publish structural studies, has urged them to adopt a policy of immediate release. Indeed, the *Journal of Biological Chemistry* and the *Proceedings of the National Academy of Science of the USA*⁶ have already made the decision to require immediate release of structure coordinates for papers published in their pages. Should the rest of the journals (and granting agencies) follow suit?

Many recognized the potential dangers of a policy that requires immediate release. For example, Donald Boyd (Indiana University-Purdue University at Indianapolis, USA): "The downside is that people might delay publishing a structure if they want to make money from it. Or they might release backbone atomic coordinates, but not the full set". John Tainer (Scripps Research Institute, USA) goes further, saying that the "major problem in structural biology remains getting scientists to take the big risks for the important problems, as evidenced by the fact that only about 10% of the structures in the Protein Data Bank are really novel", and that immediate release "will strongly favor exploitation over innovation". On the other hand, those such as Mark Trumbore (NIH, USA), believe that "without a doubt prompt release of deposited structures will help push forward all areas of biomedical science": this, ultimately, is the point of releasing data.

The question is, will the potential loss of future structural data that a general policy of immediate release of coordinates may cause — either temporarily or permanently — be outweighed by the progress it may fuel?



1. Webster's Third New International Dictionary (G.&C. Merriam Co., MA, USA; 1976)
2. IUCr Newsletter, 6, 6 (1998)
3. Editorial, *Nature Struct. Biol.* 5, 83–84 (1998)
4. Editorial, *Nature Struct. Biol.* 5, 165–166 (1998)
5. Editorial, *Nature Struct. Biol.* 5, 245–246 (1998)
6. Editorial, *Proc. Natl. Acad. Sci. USA* 95, iii (1998)