The Missing Crystallography Data

Some disgruntled researchers are mounting a campaign to force crystallographers to make available key data when they publish the structure of complex molecules

SHOULD A SCIENTIST be allowed to publish only the conclusions of a piece of research while keeping secret the data and results on which those conclusions are based? That's what some say is happening in the field of x-ray crystallography, where an increasingly vocal group of disgruntled researchers are accusing colleagues of sitting on vital information and thus slowing the progress of science.

"Publication means making public," says crystallographer Richard Dickerson of the University of California, Los Angeles. "I don't know of any other field of science where you are required to make public neither your data nor your results, only your commentaries."

Biochemist Howard Schachman, at the University of California, Berkeley, gripes: "[People] publish, they get the scientific prestige, and then they don't make the real results available, which are needed to assess the validity of the work, and allow people to extend it."

What has scientists like Schachman and Dickerson upset is that journals allow x-ray crystallographers to publish the structures of macromolecules without making available the primary results on which the conclusions were based—the three-dimensional coordinates describing the position of every atom in the molecule under study. Without those coordinates a structure article is nearly useless to many researchers who may want to verify or build on the work.

In an effort to change the system, 2 years ago Yale crystallographer Frederic Richards began a crusade to get journals to force authors to make their primary data public. He was joined by 173 colleagues, who cosigned a letter imploring journals to take a stand. And indeed, some have changed their policies, now requiring crystallographers to make coordinates available within a reasonable amount of time following publication of their papers. But some researchers are still not doing so, prompting Dickerson 2 months ago to fire off a "dear colleague" letter in which he warned his colleagues that "we are on our way to developing a miniature scandal."

That's not how every crystallographer sees it. Although few will publicly advocate withholding coordinates indefinitely, many argue that there are valid reasons not to release them immediately—to further improve their accuracy or to reap scientific rewards after the years invested in obtaining them. And, in the case of structure laboratories springing up in many drug companies, there can be a strong incentive to keep information secret in order to cash in on its commercial value.

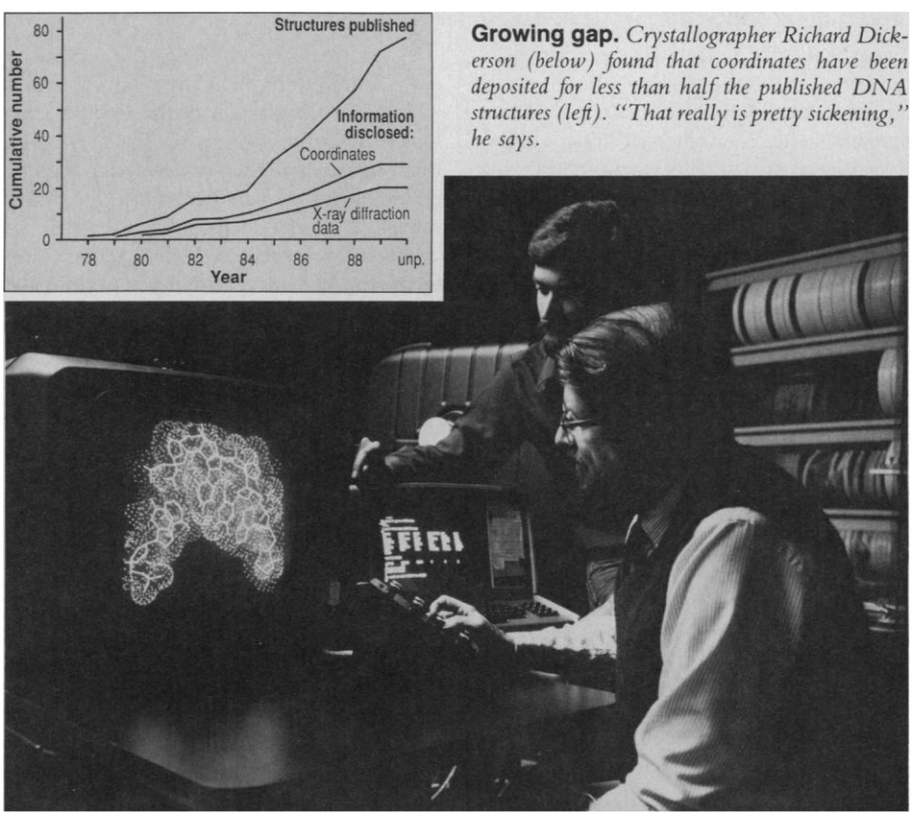
While the urge to withhold vital information can tempt a researcher in any field, the nature of crystallography provides a convenient excuse for not sharing. The raw x-ray diffraction data generated during structure studies of proteins and nucleic acids, and the atomic coordinates derived from them, consist of tables of numbers too voluminous to fit easily into a journal article. So researchers have traditionally left them out.

Since 1971, Brookhaven National Laboratory has maintained a data bank for storage and distribution of x-ray data and coordinates. But researchers deposit their data in the bank on an entirely voluntary basis, and many do not volunteer. The data bank currently has 435 sets of coordinates available, many of which are multiple sets for the same macromolecule. But data bank director Thomas Koetzle says there are more than 70 molecules for which structures have been published but no coordinates deposited. In some cases, years have gone by since publication.

UCLA's Dickerson has found what is to him particularly depressing evidence of this. While writing a review on DNA structure recently, he discovered that, of the 78 DNA structures published, the coordinates for only 29 had been deposited at Brookhaven (see figure). Some of the structures without coordinates were nearly a decade old. "That really is pretty sickening," says Dickerson.

To Yale's Richards, journal editors are the key. "If the journals took a stand, then [the withholders] would be stuck," he says. "If they're going to publish, they have to follow the rules, and if they don't want to follow the rules, they don't publish."

But Richards' effort to persuade journals to establish rules has drawn a mixed response. Of about 40 journals he contacted, he says eight, including the *Journal of Biological Chemistry (JBC)*, *Biochemistry*, and *Science*, have changed their policies, and at least seven more, including the *Proceedings of the*



Growing gap. *Crystallographer Richard Dickerson (below) found that coordinates have been deposited for less than half the published DNA structures (left). "That really is pretty sickening," he says.*

National Academy of Sciences (PNAS), are considering doing so. "People argue about whether a journal should be in the enforcement business," says *PNAS* editor Igor Dawid. Nevertheless, he reports that *PNAS* will have a statement in January's instructions to authors, although its wording has yet to be determined.

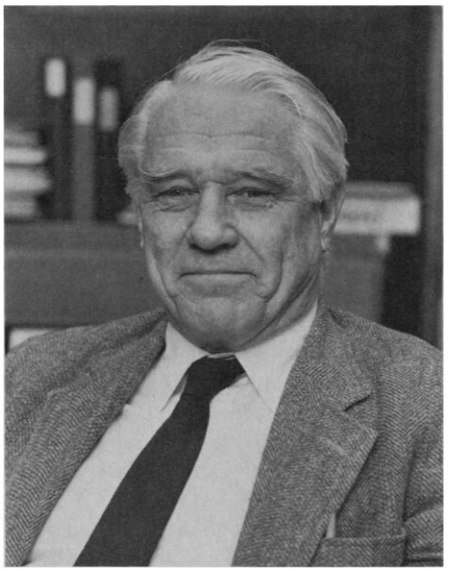
Some journals, such as *Biochemistry* and *JBC*, require proof of deposition before the paper is published. *Science*'s newly adopted policy requires authors to state in their papers that coordinates have been deposited. *Science* editor Daniel Koshland says it is not practical to require proof of deposition before publication. Rather, he says he will trust authors to do as they say, dealing with noncompliers only if complaints occur.

Nature, on the other hand, has decided to impose no new requirements. "Our position is that we put as few impediments as we can between authors and publication," says *Nature* editor John Maddox. According to Maddox, the journal encourages contributors to deposit information in the appropriate data banks but will go no further.

Those who depend on coordinates include theoreticians who want to analyze interactions within the molecule, crystallographers trying to solve the structure of a related molecule, biochemists interested in interactions between an enzyme and its substrate, drug designers who want to create inhibitors for an enzyme, and genetic engineers who want to alter the molecule and change its function.

Abbott Laboratories researcher Jonathan Greer says that the utility of a structure paper without coordinates depends on what the authors choose to reveal about the specifics of the structure. "It may be useful to know that a protein has two domains, and the active site lies between them. . . . That may help eliminate alternative hypotheses," he says. However, to be most useful to readers who want to build on the work, he says the coordinate information must be available.

"Publishing without [coordinates] is not publishing—it's useless," snaps Purdue University researcher Michael Rossmann. Frustration over difficulty obtaining coordinates drove Rossman to devise and publish, in 1980, a computer program designed to extract coordinate information from the stereo drawings that are often included in structure papers. But this may have only escalated the data tug-of-war. Complained one crystallographer, the advent of such programs has driven some researchers to obscure their results even further, generating what he called the "Brillo pad" version of molecular structure drawings, too small and snarled to be easily interpreted.



Crusading crystallographer. *Frederic Richards of Yale persuaded 173 colleagues to cosign a letter imploring journals to take a stand.*

To those researchers who withhold their coordinates, this is a bread and butter issue. What many crystallographers will privately argue is that a scientist who has put what often amounts to years of work into obtaining crystals and solving a structure deserves some time to reap the benefits of that work before it is released to the public.

"The counterargument is that you don't publish until you have used the coordinates in whatever way you want," says University of Oregon crystallographer Brian Matthews. But, he warns, that can hurt young investigators, who may be under pressure to publish results as quickly as possible because of an impending tenure decision.

On a more practical level, many crystallographers argue that newly determined coordinates are not reliable enough to be released into the public domain. While they may allow the researcher to draw a three-dimensional structure of the protein, they contain inherent uncertainties that are gradually diminished with further work, by a mathematical process called refinement. "There will be some regions which may be quite ambiguous at the time one first solves the structure," says Matthews. Even when warnings of the uncertainty are included in the data bank entry, Matthews says some researchers may use the coordinates inappropriately, leading to misunderstandings and wasted time.

In the corporate world, the views on withholding coordinates are most sharply polarized. "Naturally, we have to look out for our interest," says Monsanto spokeswoman Deb DeGraff, going on to frankly state that although Monsanto recently published the structure of bovine growth hormone, it has no plan to deposit the coordinates in the data bank.

Yale's Richards retorts: "If they do the structure and want to sit on [the coordinates] as a proprietary right, fine, but don't publish it." That's exactly what will happen if journals adopt strict publication rules, predicts Genentech vice president for medicinal and biomolecular chemistry, Michael Ross. "We believe in publication," he says, "but if for some reason journals were to decide that in order to publish crystallography data, you had to put [coordinates] in the database, we wouldn't publish. That would be very stupid on the journals' part."

Not all companies take this view though. Greer of Abbott says his research group is about to publish the structure of porcine pepsin, and plans to deposit the coordinates upon publication. "We believe that, overall, the dissemination of this information is to everyone's advantage," he says. "We can only solve so many structures, and the release of coordinates by other people benefits us, just as our release of our few structures benefits others."

Greer notes the recent case of Merck, Sharpe & Dohme's publication of a structure for the HIV protease and its deposition of the coordinates for the alpha carbons, which trace the protein's backbone. Part of the structure was wrong, but the coordinates aided Alexander Wlodawer and his colleagues at the National Cancer Institute in finding the flaw. "It's a classic case of why it's a good thing to do," says Greer. "You can't always be right in this world, and [sharing coordinates] keeps the quality up. [The Merck scientists] were basing a lot of their theories on the parts that weren't correct."

To try to avert civil war in the crystallography community, the Commission on Biological Macromolecules of the International Union of Crystallographers (IUCr) began 2 years ago to hammer out a policy that all crystallographers can live with. The going was slow at first, because of the range of viewpoints represented by commission members, but they settled on a guideline that requires authors to deposit both x-ray data and coordinates in the data bank when they publish a crystal structure. But the proposed guideline would allow release of the coordinates to be delayed for up to 1 year and the x-ray data could be withheld for 4 years.

Most crystallographers claim to support the compromise. Says Oregon's Matthews, who participated in the IUCr commission: "My feeling at the start was that the investigator ought to be able to ask for essentially as long a delay as he or she liked, but I've become convinced that a period of 1 year

wouldn't be unreasonable."

Will the compromise work? Data bank director Koetzle says he will go along with the IUCr policy in applying the requested delays. But the cooperation of either journals or funding agencies will be required to force researchers to deposit the information in the first place.

Even crystallographers who claim to support the IUCr guidelines admit that their laboratories have not always deposited coordinates promptly—or at all. DNA researcher Alexander Rich of the Massachusetts Institute of Technology was a cosigner of Richards' letter, but he acknowledges that some coordinates from his lab were not deposited due to oversight. "With active labs, not everything gets nailed down," he says.

Slip-ups, innocent or otherwise, can lead to valuable information being lost forever. "I know of one case where they just never got around to [depositing the coordinates] and then they literally lost the numbers," says Helen Berman of Rutgers University, who participated in the IUCr commission. "It was a structure that other people were interested in." UCLA's Dickerson, while sympathetic to the notion that innocent oversight can occur, warns that there are repeat offenders, who hide behind the excuse of inadvertent error.

If journals don't formulate and enforce strict policies, the funding agencies may adopt their own remedies, Dickerson warns. Marvin Cassman, acting deputy director of the National Institute of General Medical Sciences, agrees. Although he said he knows of no cases so far in which nondeposition of coordinate data has influenced a decision on funding, he said it is conceivable that such considerations could have an influence in the future. But Cassman said he hopes things won't come to that. "[Crystallographers] are acting in a responsible way. There is a lot of debate going on. It's not a trivial issue. I really sincerely hope the scientific community will be able to come to some kind of agreement . . . because NIH has never really acted as a regulatory agency for this kind of thing, and I certainly hope that won't be necessary." ■ **MARCIA BARINAGA**

Help Wanted: Director, NIH

"This is not the time for the general to leave the battlefield to go back to the Pentagon," AIDS chief Anthony S. Fauci told *Science*. Fauci, at the top of the list of candidates for the directorship of the National Institutes of Health, has formally withdrawn from consideration in order to continue his basic research on AIDS.

Members of the NIH search committee who learned of Fauci's decision from *Science* expressed two consistent sentiments: "I'm very disappointed. I was sure Tony would take the job," said one, adding "but I have no doubt that he made the right decision."

Fauci's decision to stick with research now forces the Administration to face a troublesome question: Will someone else on the search committee's list take the job? Health and Human Services Secretary Louis Sullivan responded with a discernible pause when asked that by *Science* at a press breakfast last week. "Good people have a number of options," Sullivan said. "We will need to convince one of those people that the NIH director's job would be as challenging and rewarding as what they're already doing."

Those "good people," whose names have been sent to Sullivan unranked after a 3-month search, are William H. Danforth, chancellor of Washington University, St. Louis; P. Roy Vagelos, chief executive officer of Merck & Company; Philip Leder, chairman of genetics at Harvard Medical School; and Leon E. Rosenberg, dean of Yale University School of Medicine.

None of the four has as yet been told by either Sullivan or James O. Mason, HHS assistant secretary for health, that he is on the short list, according to the candidates. Nor, apparently, have people on the search committee's list of additional candidates. They include Institute of Medicine president Samuel O. Thier; Harvard dean Daniel C. Tosteson; and Bernadine P. Healy, chairman of the Research Institute at the Cleveland Clinic Foundation (*Science*, 8 September, p. 1046).

A Merck spokesman says Vagelos is not leaving. Leder says simply, "I am a working scientist with a very full research agenda." Rosenberg believes that the crucial element in the search is to find someone "who understands the importance of the NIH director's job and who has a real feel for the issues it faces." Thier has made a commitment to continue at the Institute of Medicine. Tosteson appears to be thriving at Harvard.

The NIH director's job, which ought to be the pinnacle in biomedical science, has lost a lot of its luster during the past decade as the director's authority has been eroded. For instance, Congress had elevated the heart and cancer directorships to the status of presidential appointment, in effect giving these two posts independent power. Because each of the 11 NIH institutes' budgets is handled separately by Congress, the NIH director has little to say there. And, because the top NIH post carries no significant financial independence, the director lacks authority to take initiatives on his own.

As one observer noted, the only reason to accept the post unless its scope is changed is "as a genuine act of public service."

On that score, Danforth is at the top of the list. A member of a conservative Republican family that has made a fortune in the Ralston Purina company, Danforth is known to have a strong belief that public service is a duty. And he has good political connections, including his brother, U.S. Senator John C. Danforth, who is ranking minority member on the Committee on Commerce, Science, and Transportation. Given his family wealth, Danforth may be the only NIH candidate who could readily afford to take the job, which pays less than $100,000 and would entail real financial sacrifice for many of those on the list.

Money, of course, did not play a role in Fauci's decision to withdraw since he has been living with an NIH salary for more than 20 years and does not intend to leave the institutes at this point for more lucrative pastures.

Fauci has managed to lead a very productive research laboratory despite his administrative responsibilities at the allergy institute and has been in the public spotlight both because of his research and his role of AIDS coordinator for all of the National Institutes of Health. It was through this route that he came to know then Vice President George Bush during the past couple of years, leading Bush to make his now famous campaign declaration that Fauci is one of his heroes. Had Fauci decided to leave the lab, he would have been one of the few NIH directors in a quarter of a century to have real contact with the President.

The next step in the search process rests with Mason and Sullivan at Health and Human Services. For all practical purposes, the search committee appears to be out of the loop now, having fulfilled its mandate to submit a list of names—unranked—for the politicians to deal with.

It is said that the only one who can make a difference is George Bush himself. ■ **BARBARA J. CULLITON**