DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

National Institutes of Health
Bethesda, Maryland 20892

www.nih.gov

AUG 2 2 2005

Dr. William F. Carroll, Jr.
President
American Chemical Society
1155 Sixteenth St., N.W.
Washington, DC 20036

Dear Dr. Carroll:

I want to thank you for your willingness to conduct frank discussions about PubChem and for clarifying your concerns about its possible impact on the CAS Registry®. As we have discussed, I think we have come a long way in trying to resolve this matter. We have given a great deal of thought to your questions from our June 24 meeting, your follow up e-mail from July 2, and the issues discussed in our teleconference on July 27. I especially thank you for your recent letter of August 9, which has stimulated additional thinking here at NIH. I am sorry for the delay in responding due to the many different levels of reviews inherent to Federal Government rules and procedures and my having to take leave for personal reasons. In this letter, based on our many discussions, I would like to outline a multipart proposal to resolve this matter, which I hope you will find satisfactory.

In your letter of August 9, you put forward the interesting proposal that ACS would build, manage, and make available for free a "PubChem" database through its CAS division. The offer to staff this for a minimum of five years is most generous, and we have given this proposal very serious consideration. However, we are concerned that some of the most critical aspects of PubChem would be lost in such a model. In particular, the integration of PubChem with other public biomedical databases, including protein structures, genome information, and the biomedical research literature, is a prime driving force behind the need for PubChem. Given their intimate familiarity with biomedical research data, the staff of the National Center for Biotechnology Information are in an ideal and unique position to create this integrated view. With respect, we do not think that CAS, an organization that quite appropriately focuses on chemistry, would be able to provide that same sort of seamless connectivity that biomedical researchers need in this new era.

We are also concerned about how decisions would be made regarding entries into a CAS-operated version of PubChem. We would all agree that small molecules with available bioassay data should be included. But, as you know from our prior discussions, NIH feels strongly that the database should not be limited to that set of information, or its utility will be greatly constrained. Your proposal indicates that ACS and NIH would

work together to develop mutually agreed upon protocols for disseminating any additional compound or other data, but based on our discussions so far, we are concerned that this will be difficult to agree upon. NIH's position is that any molecule of potential biomedical relevance is appropriate for inclusion in the database. As you know, a central purpose of the Molecular Libraries Roadmap Initiative, of which PubChem is a part, is to discover which chemical compounds, of the vast number that are available or can be made, are able to affect gene and cell functions involved in health and disease. To make these discoveries, biomedical researchers must have unfettered access to the millions of compounds to test—the "libraries" in "Molecular Libraries." Such libraries are being tested in the Molecular Libraries Screening Centers Network, and the first results are now in PubChem. Equally important for biomedical research is the ability to test compounds for activity via computer algorithms. This computer-based testing can be done in many biomedical researchers laboratories and will contribute greatly to the Molecular Libraries Initiative's impact on biomedical science, but only if researchers have access to large computer-based chemical libraries. Several such libraries, of millions of compounds each, have been contributed to PubChem by NIH-funded researchers. For example, an NIH grantee from the University of California, San Francisco, has recently requested that PubChem display the entire content of his ZINC database of over two million chemicals, collated from commercial suppliers of chemical compounds with their permission. There are no bioassay data on many of these chemicals, but since they can be tested for biomedical activities by computer-based screening, their availability to the biomedical research community in PubChem is critical to the goals of the NIH Molecular Libraries Roadmap Initiative. The recent peer-reviewed funding of the ZINC database further documents the interest of the biomedical research community in having that information collected in an accessible place.

Lastly, even if these serious concerns did not exist, NIH would not be able to enter into an exclusive bilateral relationship with ACS without such an opportunity being made available to other private sector suppliers of chemical information, per the requirements of the Federal Acquisition Regulation.

Considering all of these issues, and building on all of our prior discussions, we would like to propose an alternative structure for resolving these issues. This proposal has six parts:

1) We recognize that CAS is a highly valuable resource, and that its manually curated content will be of great interest to biomedical researchers (many of whom are not currently using it to full advantage). We would like to maximize the interactiveness of CAS with PubChem. Accordingly, we propose that NIH work with CAS to validate or assign registry numbers for all PubChem structures. We also propose that NIH provide reasonable financial compensation to CAS for this service, consistent with Government requirements on competitive acquisition. Furthermore, to avoid unnecessary duplication, PubChem will not disseminate information on chemical reactions, measured properties, methods, patents and

applications, markush structures, or conference information (except when then conference was funded by NIH).

This cooperative relationship would ensure that PubChem is directly and accurately linked to CAS and should create greater synergy while limiting duplication.

2) We recognize ACS's serious concerns about PubChem including molecules that are irrelevant to biomedical research. You had asked us to further define "biomedical relevance" so as to set boundaries on the kinds of chemicals that would be appropriate for deposition into PubChem. As noted above, the current lack of knowledge in the field requires us to have a broad and inclusive definition of biomedical relevance. As a possible alternative, however, we would be willing to consider a retrospective process based upon the recommendations of an outside group (see below), wherein compounds that were judged as not to have biomedical relevance could be retrospectively removed from PubChem.

3) We also want to be responsive to the concerns raised in your e-mail about the provenance of the data entered into PubChem. NCBI currently assesses the legitimacy of all submissions to PubChem by requiring each submitter to sign a submission statement in which they certify that the data are reliable and accurate, and by reserving the right of the NLM to withdraw the data from PubChem if they are found to be "erroneous or in violation of intellectual property rights." We are willing to consider additional measures to assure the provenance of the data.

4) We believe that the private sector has a great deal to offer in terms of expertise about many PubChem matters. Given this, we plan to form a new working group of outside experts that can advise on PubChem as it develops. This group would be subject to The Federal Advisory Committee Act (FACA) as a subgroup of the NCBI Board of Scientific Counselors and will include representation from a diverse group of interested providers of chemical information. All members of such a working group would be required to disclose their potential conflicts. We hope that the American Chemical Society will send a representative to participate in this working group. This group would not debate the limits of PubChem, but would advise the NCBI Board of Scientific Counselors on such issues as:

- Establishing a process for retrospective evaluation of the biomedical relevance of compounds entered into PubChem
- Ensuring the provenance of the data (i.e., whether private data are being improperly deposited in PubChem)
- Ensuring the high quality of data in PubChem
- Monitoring the effect of PubChem on scientific progress
- Improving/integrating interactions with commercial information providers
- Avoiding unnecessary duplication with commercial information providers.

5) This new working group of outside experts would be separate from the existing PubChem Advisory Board, which provides advice to NCBI about details of the operation of the PubChem database and also reports to the NCBI Board of Scientific Counselors. We appreciate your suggestion that the PubChem Advisory Board would also benefit from representation from the community of private sector data providers. We think that this might be best accomplished by having a liaison member from the new working group as part of the PubChem Advisory Board. Your suggestion of Dr. Matthew Toussant is appreciated, and we will consider him along with others as we provide an opportunity for other interested chemical database providers to also become involved.

6) In addition, we are still very much open to the idea of having an independent outside group, such as the National Academy of Sciences, conduct a workshop or study on the issue of the impact of PubChem on scientific progress and on private sector chemical databases. An independent group of scientists and database experts could be pulled together to look at these issues and could make recommendations to NIH as to how to proceed with PubChem.

Please let me know your thoughts on this proposal. I am very pleased with the progress we have made on these matters during our many meetings on this subject. We are determined to move forward in our efforts to work with the private sector to ensure that PubChem develops in a way that enhances all of biomedical research without having a negative impact on the private sector.

Sincerely,

Elias A. Zerhouni, M.D.
Director