SCIENCE AND SOFTWARE

The foundations of our science are empirical, i.e., based on direct observation, to the extent where that is possible. For science to move forward and build on other discoveries, it is necessary that any group of experimenters be able to reproduce the results of any other group to within recognizable experimental error. This requirement is so critical that it is virtually an unspoken rule that any scientific paper published must provide sufficient information to permit another researcher to reproduce the published results. Were this practice violated too severely, there would soon be little that is scientific about our science.

It seems unlikely that many would quarrel with this view of empirical science. There do appear to be quite a few, however, who are willing to support a different rule—even a completely different set of rules—with regard to the publication of scientific papers based essentially on computation.

In chemistry we are beginning to accumulate a set of tools and equipment which may well permit us to obtain vast amounts of chemical information by means of computation. This is the same information which only a few years ago could be obtained exclusively by experiment. Thus we now have two nearly parallel approaches to the practice of chemistry: the empirical and the computational. It is quite natural that both groups of researchers be interested in publishing, thereby sharing the results of their efforts with their colleagues.

All would be well were these two approaches to chemistry identical in all aspects. They are not; and therein lies a major problem which could ultimately prove a threat to our science. This problem seems especially acute in chemistry, for reasons not wholly obvious at first glance.

In experimental chemistry, for instance, someone runs a specific reaction using a specific piece of equipment to measure something. (This is obviously simplistic, but it will serve the purpose of the discussion.) The results are then published and subjected to the scrutiny of other chemists. The results are thereby verified or refuted on solid, reproducible grounds. In computational chemistry we are facing an element which could render impossible the simple scenario I have described, one which is strong enough to threaten the process of science which has served us so well. This element is profit—and its power cannot be ignored.

Computational chemistry essentially began in the early 1960s with the availability of large-scale solid-state computers. Computer access was still quite limited, but interested researchers began to experiment with many different concepts, and within a decade people had some fairly dependable ideas on what approaches would work and what sort of directions to take. By the early 1980s, the colossal advances in computer hardware capabilities and the virtual collapse in computing costs put this computing power within reach of any interested researcher—power significant enough to practice computational chemistry. This ready availability of hardware naturally created a vast market for computational chemistry software where one had never really existed before. Chemists have historically been entrepreneurs and, recognizing the opportunity for profit, many of them brought commercial software products to market to meet this need.

It was quickly discovered, however, that computer software is very different, as a property,

0920-654X/$ 3.50

from a piece of instrumentation. The problem of protecting one's proprietary interests in computer software is in no way comparable to protecting one's interest in the design of a piece of equipment or hardware. Hardware is, and always has been, protected successfully by patents. Everyone understands the rules; most abide by them. Computer software, however, is in one sense an intellectual property, yet in another sense it may be considered analogous to a hardware circuit design. There was no established and reliable way to protect a developer's interests. Recognizing the dilemma, the developers have taken the obvious step to protect their interests by releasing little or no information about the actual details of the products they offer the chemistry market. They sell computational chemistry software products in executable code only. The vital source code which would permit a researcher to assess the quality of the science (and, admittedly, permit a thief to misappropriate the science) is withheld. The chemist cannot take the cover off *this* instrument to figure out what it is doing!

Many of these commercial packages are widely available. Heavily promoted, they have been put into the hands of scientists who have no true concept of their capabilities or areas of applicability. In the case of highly proprietary packages, there is no way they *could* know these things.

Yet scientists have systematically been using these tools, and it is only natural that they wish to publish the results of their work. When such results are submitted to a journal, however, no editor in good conscience can do other than hesitate and, frequently, refuse to publish. Why? Because these results cannot be readily reproduced. Further, they cannot even be intelligently criticized, as nothing can be known in detail about how the results have been achieved. Indeed, such results can only be regarded as some sort of divine revelation—or the nearest approach to magic we have seen in modern times.

Despite this obvious problem, such results are still being publicized with sincere profundity at such places as American Chemical Society (ACS) meetings in the United States. As often as not, one finds that the people reporting these results are receiving "considerations" from the purveyors of the software in question. The problem, sadly, appears to be escalating rather than abating.

What is at stake here is the integrity of computational chemistry as a useful tool. Given that such practices are springing up at professional society meetings, the integrity of the entire field may be called into question. We need to seek some middle ground that protects both the individual and science itself.

Richard W. Counts
QCPE
Indiana University
Bloomington, IN 47405
U.S.A.