Check for updates

Quantum mechanics guided by simplicity

Dr Arieh Warshel, distinguished professor of chemistry at the University of Southern California and 2013 Nobel laureate in chemistry, discusses with *Nature Computational Science* past and current research, his Nobel Prize, and the benefits and challenges of using computational modeling in his work.

How did you first become interested in computational chemistry?

I first became interested in enzymes while working on an undergraduate project, in which I studied fast enzyme kinetics using nuclear magnetic resonance. Of course, I had no idea that one day I would study enzymes using computers. Around 1965, I had a summer project in which I had the task of using number crunchers — the machines before computers — to perform massive calculations, and then realized that crunchers/computers indeed could do the job much better. When I started my PhD, I was very interested in working with my soon-to-be mentor, as the two of us had similar backgrounds and we were both from a kibbutz. It was not clear to me at first that my work would be in computational chemistry, but my assignment was to use computers in conformational analysis. So, really, it was by chance that I started working in computational chemistry.

Credit: Max Gruber

How did the idea of developing multi-scale methods, such as quantum mechanics/molecular mechanics (QM/MM), come to be?

It started with trying to model medium-size molecules during my PhD. First, this was done by modeling the molecule as balls and springs, where the former represents atoms and the latter represents bonds between atoms. This modeling approach works very well to describe structural changes, but the problem is that it does not describe well the breaking of chemical bonds: when you break the spring, this does not reflect the reality of bond breaking in chemistry, in which the electrons move from one atom to another. To describe the electrons themselves and in this breaking process, you need what is called quantum mechanics. Calculating chemical processes using quantum mechanics is a very expensive task, even now, and it was far more expensive in the mid-1970s, when it would be impossible to represent a whole enzyme or a large protein by quantum mechanics. So, we came up with a very simple idea to restrict the quantum descriptions only to the bonds where the chemistry occurred. The crucial piece was to tell these bonds what the electrostatic field from the rest of the system is by adding charges to the rest of the system, and then by letting those charges influence the quantum mechanical portion. Essentially, we described a small part of the system using quantum mechanics and the rest of the system with a simpler ball-and-spring-plus-charges approach. In this way, we could describe the chemistry of very large systems by using different descriptions for the central system and its surroundings. This is what made our modeling approach multi-scale.

What were the challenges that you faced while developing QM/MM?

The main challenge was really just figuring out how to approach the system, because we were asking questions about things that we did not understand well. In addition, even very big computers in the early 1960s and late 1970s were much less powerful than any small iPhone is today, and storage was a major problem. The code had to be written in a way such that we could run calculations despite having very limited storage. Another challenge, which was eventually overcome, was that the quantum treatment was not accurate at all. For example, if you tried to describe a bond breaking quantum mechanically, and if you did not include the effect of the protein, the bond would never break. It took some time for us to realize this point, but this was key to our understanding of how to best treat the system. We were always guided by whether or not we could reproduce reality. So, when I saw that I could not break a bond in an enzyme, or in any other system, without including the electrostatic effect of the environment, this was a major step forward in QM/MM.

Later developments focused on calibrating the quantum mechanics portion to reproduce reality in the solution, such as in water, and not in the protein. Then, the calibrated quantum regions would be transferred to the protein without changing it. This is the version of QM/MM, which is called the empirical valence bonds (EVB), that looked, in some regards, like high-school chemistry because it was very simple. In fact, many of our approaches looked too simple and it led to a lot of criticism and rejections from the research community, which was challenging. For example, a typical criticism used to be that, since quantum mechanics could not reproduce the energy of an H_2 molecule, then obviously an enzyme could not be modeled correctly either. However, the reason why this approach worked for larger molecules, such as enzymes, was that we did not try to use the quantum mechanics for exact energy calculations; instead, we had a simple quantum part that we forced to reproduce experimental reality in water, and then we had a consistent way to move this quantum part to the protein. What captivated the public relatively fast was the simple idea of separating QM and MM, rather than a way to make it more accurate. There was a very large degree of pushback at first, but by the mid-1990s, more people were using these approaches, and then there was less resistance.

What do you see as the main challenges related to studying and understanding enzyme behavior?

Designing enzymes correctly and accurately is one of the main challenges, and I have found that it is harder to predict what will happen when you design new mutants than it is to take a known mutant with a known structure and reproduce its effect.

If we have the structure of the mutant and the sequence, we can do quite well in reproducing the effect of the enzyme and also in explaining where the catalysis comes from. However, when we try to reproduce directed evolution and move from one mutant to the next one without knowing the structure a priori, this becomes a very difficult, challenging task. There are similar challenges when predicting binding, in which current methods are not accurate enough. These problems are harder because you have to model structural changes in which sometimes the energy landscape is very complicated. Recently, we have played around using artificial intelligence for this problem, and we were able to find correlations that could help in enzyme design, not necessarily by further understanding the process, but by being able to predict the behavior. One day, I believe that it will be possible for people to design enzymes without needing to understand anything.

Another challenge is to convince people that computers are the only way to definitely understand how enzymes work, because you cannot physically go into the enzyme with 'tiny' people and look around: with computational approaches, you can test many different hypotheses and show that they are wrong and in which quantities. There is still a large resistance by some experimentalists who really believe that there are factors that you cannot model, which is incorrect.

What do you think is the best approach for answering questions in science using computational techniques?

It is important to realize that using the best computers may not always be the answer. For instance, for enzymes, having more powerful computers is not necessarily the main issue, and in the example of ion channels, very powerful computers could predict the trajectories of the ions, but they still cannot give the free energies or the nature of the values. So, we need to reach a stage in which we can get relatively reasonable free energy landscapes and then learn from this; for instance, how does the result change with different software?

There is also the issue of running calculations on a machine without thinking about what is happening. To me, the best approach would require a lot of training in universities, teaching students to focus on the problem and not to just run anything that takes more than three days, otherwise they lose track of what the problem was. Even if you use the computer as an experimental tool and reproduce the observed results, you still have to analyze the results. To make the analysis more approachable, you need simpler models that are going to produce the same results. Personally, I believe in thinking about the results and decomposing the problem into different, smaller factors, and not necessarily just using bigger, more powerful computers just to get an answer.

Your molecular dynamics simulations of the first step in vision represented the first use of such simulations in biology, and around 30 years later, the 'bicycle pedal' motion that you predicted was confirmed by ab initio studies. What are your thoughts on this study?

Miraculously, our calculations predicted everything exactly, such as the time of 100 femtoseconds, which appeared to be correct, and the very large quantum yield. Since it was such a fast process, a lot of people wondered about the correctness of the results. When I run the simulations, the kinetic energy of moving downhill in the excited state dissipated to all other degrees of freedom. To me, it was clear back then that this is what happens when you have a realistic simulation with enough degrees of freedom. Some researchers argued strongly that this could not be correct because it was known, experimentally, that the quantum yield did not depend on how much light energy was being put into the system. If a system does not lose kinetic energy, then you should have different results for different light energies, and therefore my prediction would be incorrect. However, we showed that the kinetic energy dissipated very fast, exploiting the advantage of having a realistic model rather than just an idea of how the system 'should be'. This demonstrated that one should choose systems that can be modeled with the tools of that time rather than trying to use, for example, ab initio techniques without any dynamics. Including the complete system in a less accurate way was apparently better than trying to do everything accurately without the resources to do so.

Putting the work that contributed to your Nobel Prize aside, what aspects of your research are you most proud of?

I am most proud of the enzyme work, even though I have not succeeded in convincing too many people about it! One of the Nobel Prize committee members actually told me during the ceremony that they did not give me the Nobel Prize for the EVB work (which allowed me to study enzymes in an accurate way), but for the basic QM/MM multi-scale idea.

What did winning a Nobel Prize mean to you?

First of all, I was extremely happy. It is a great thing that basically changes your life. I cannot say that I thought about winning a Nobel Prize throughout all my life, like a lot of other scientists who are very eager to get it do. It makes you very happy, but it takes time to sink in that it is real. Overall, it is a very nice validation of what you have done, despite the fact that most of my papers were rejected in the first round!

What do you think could be the next achievement in computational science to be recognized with a Nobel Prize?

It could be quantum computing, but this could also happen in another 100 years or so. Personally, my opinion when it comes to quantum computing is that I will believe it when I see it. There could also be a prize for very fast and powerful computers that allow you to do calculations orders of magnitude faster. It could also be something that neither of us have any idea what it is! But what I tell young students and researchers is that they should work on brain science. Essentially, if somebody develops a meaningful simulation of the brain, it could lead to a Nobel Prize, even though there would be a lot of objections from the experimental community. So, I would say that, one day, a Nobel Prize could be given for modeling the brain.

Interviewed by Kaitlin McCardle

Published online: 26 September 2022
https://doi.org/10.1038/s43588-022-00285-2