Modeling Complex Materials: Do We Need All of the Atoms?

ROB PHILLIPS
Division of Engineering and Applied Science
California Institute of Technology
Pasadena, California

PROBLEMS WITH MULTIPLE SCALES: A TERABYTE-SIZED CHALLENGE

The advent of computers ushered in a new era in which a host of complex problems ranging from weather prediction to the microstructural evolution of multiphase alloys to the analysis of the DNA/protein interactions that mediate gene regulation could be modeled explicitly. Indeed, some say that the physical sciences are now based on a triumvirate of experiment, theory, and simulation. Clearly, the use of computation to understand and even design complex materials is one of the major challenges that will make it possible to replace enlightened empiricism with rational design. Similar anticipation concerning the role of simulation touches many other fields as well. On the other hand, despite their promise, simulations in many of these settings have generated enormous quantities of information (i.e. terabyte data sets) without necessarily delivering the promised concomitant increase in understanding.

The terabyte data sets engendered by such simulations represent a staggering quantity of information. A simple estimate reveals that the entire ten floors worth of books in the Caltech Millikan Library corresponds to roughly two terabytes of information. More impressively, the genome of many viruses have an information content that can be stored comfortably on a 256 megabyte memory stick alongside the genomes of even more complex organisms from bacteria to yeast. Indeed, even organisms as complex as humans have genomes that are much smaller than a terabyte. And yet, our computers are overflowing with terabyte data sets and worse yet,

discussion of petabyte datasets is even becoming routine. As an example, a molecular dynamics calculation on a 100,000 atom system if run for 10 nanoseconds, woefully inadequate for accessing most materials processes, already generates a terabyte worth of data. Clearly there is a mismatch between the quality of information generated in our simulations and that present in genomes and libraries.

The question of how to build quantitative models of complex systems with many interacting degrees of freedom is not a new one. Indeed, one of the great threads in the history of physics, namely the development of continuum theories, resulted in two of the most compelling examples of this kind of theory, namely, elasticity and hydrodynamics. What these theories share is the idea of smearing out the underlying discreteness of matter with continuum field variables such as the displacement or velocity fields. Similarly, the underlying interactions between molecules are subsumed in material parameters such as the elastic moduli and the viscosity that are introduced to capture the effects of the microscopic physics that is not treated explicitly. One of the messages to be drawn from these examples is the idea that "multiscale modeling" is not the exclusive domain of those who carry out computational model building. Indeed, in the deepest sense, the sentiment that animates all efforts at model building, whether analytical or computational, is that of finding a minimal, but predictive description of the problem of interest.

One of the most intriguing responses to the unbridled proliferation of simulational data has been the search for streamlined models in which there is variable resolution. In particular, many of the most interesting problems currently being tackled in arenas ranging from molecular biology to atmospheric science are those in which structures or processes at one scale influence the physics at another scale. As a response to these challenges, modelers have begun to figure

out how to construct computational models in which the microscopic physics is maintained only where needed. We describe several examples of this type of thinking in what follows.

A CASE STUDY IN MULTISCALE MODELING: THE QUASICONTINUUM METHOD

As noted above, one of the computational responses to problems involving multiple scales is multiresolution models that attempt to capture several of these scales at the same time. There has been great progress along these lines in recent years and presently we will consider one such example, namely, the quasicontinuum method. The main idea of this method is to allow for atomic-level detail in regions where interesting physical processes such as dislocation nucleation, dislocation intersections, and crack propagation are occurring, while exploiting a more coarse-grained description away from the key action. The numerical engine that permits such a development is the use of finite-elements that allow for nonuniform meshes and that introduce geometric constraints on atomic positions through the presence of interpolation functions (so-called finite element shape functions).

To be concrete, consider a crystalline solid subjected to external loading. The quasicontinuum philosophy is to discretize the system in such a way that there is full atomic-resolution in the vicinity of defects such as dislocations and crack tips and to select only a representative subset of atoms to serve as nodes of the finite element regions where all-atom resolution is surrendered. At each node in the system, a displacement is defined and the displacements between the nodes are determined by simple interpolation. In order to follow the system through some particular loading history, it is necessary to know the forces on the nodes and then to move them either in accordance with some dynamics scheme or use an energy minimizer. One of the elegant features of this method is the use of atomic-level force fields to

determine the forces on *all* of the nodes. In particular, using the interpolated atomic positions, a neighborhood of atoms around each node is constructed and the energies and forces are then computed using standard atomistic techniques. The reason we argue that there is an elegance to this prescription is that it insures that the material response is strictly determined by the underlying microscopic physics without making any *ad hoc* material assumptions.

The main point of describing this example is to reveal the kind of thinking that is now being put forth to greet complex problems such as material deformation. As is exemplified by the quasicontinuum method, the underlying microscopic physics of bond stretching and bond breaking is treated explicitly where needed, and only approximately elsewhere.

THE PROBLEM OF LIVING MATERIALS

Understanding the workings of complex biological systems presents an equally compelling set of multiscale challenges. One of the most pressing challenges to have grown from the stunning successes of structural biology is the study of assemblies such as viruses and the many *"SOMES,"* such as the nucleosome, ribosome, the proteosome, and the assemblies that mediate gene expression. Models for the function of these assemblies will remain out of reach of traditional atomic-level techniques for the foreseeable future. To demonstrate this concretely, consider the process of translation mediated by the ribosome. Even if we very generously assumed that a new amino acid is added once every millisecond, this means that a molecular dynamics simulation of translation would have to be run for 10^{12} time steps to see the addition of even a single amino acid to the nascent polypeptide. The number of atoms (including the surrounding water) engaged in this process is well in excess of 100,000, implying a whopping 10^{17} numbers corresponding to the positions of all of the atoms during the entirety of the

molecular dynamics trajectory.

As with the problems involving deformation of materials described in the previous section, the hunt is on to find methods that will permit the simulation of processes of biological relevance involving assemblies of diverse molecular actors such as proteins, lipids, and DNA. One example of great importance is the *lac* operon, which has served as the "hydrogen atom" of gene regulation. That is, the study of the gene regulatory network that controls the digestion of the sugar lactose in bacteria has served as the cornerstone of the development of our modern picture of how genes are regulated. The basic idea is that it is only when a bacterium is deprived of glucose and has a supply of lactose that the enzymes needed to digest lactose are synthesized by the bacterium. The "decisions" made by the bacterium are mediated by molecules such as lac repressor, which is a protein that sits on the DNA and prevents the genes responsible for lactose digestion from being expressed. In particular, lac repressor binds to several sites in the vicinity of the promoter for the genes responsible for lactose digestion and prohibits expression of those genes while simultaneously creating a looped out region of DNA between the two repressor binding sites.

One recent multiscale response to the ambition of simulating the interaction between DNA and lac repressor uses a mixed atomistic/continuum scheme in which the lac repressor and the surrounding complement of water molecules are treated explicitly. This repressor molecule is bound to a DNA molecule that is treated via a continuum mechanics solution in which DNA is treated as an elastic rod. The power of this approach is that it permits the DNA to present an appropriate boundary condition to the lac repressor simulation without having to pay the price of a full atomistic simulation of both the DNA and the protein. Figure 1 shows an example of the simulation box as well as the elastic rod treatment of DNA.

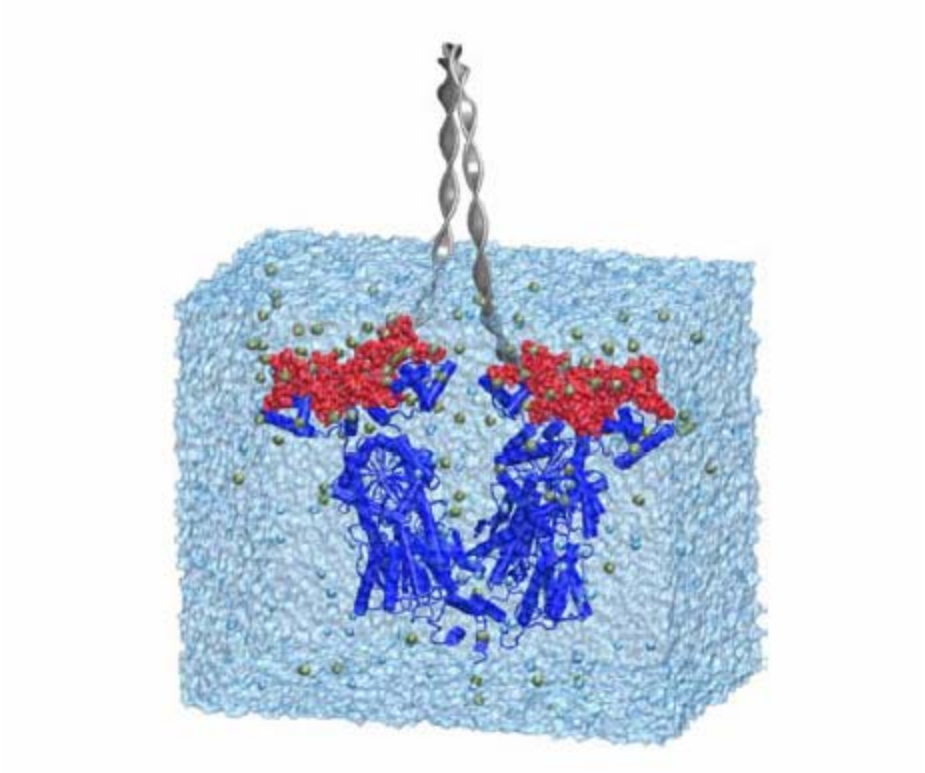

Figure 1. Illustration of a mixed atomistic/continuum description of the interaction of lac repressor protein with DNA. Courtesy of Klaus Schulten and Elizabeth Villa.

SUMMARY

The emergence of the field of "multiscale modeling" reveals the fact that even with ever-increasing computational power, there are a host of important problems that will remain out of reach of strictly brute-force approaches. Several of the problems highlighted here are those to be found in the analysis of the material world, whether the complex, rigid metallic structures that are used to construct our cities or the soft, squishy materials that make up the organisms that populate them. In both cases, there are situations in which the key action takes place at the level of individual atoms—whether we consider the bond breaking at a crack tip or the active site of an enzyme. On the other hand, many of the atoms that are part of these processes are interlopers whose job seems to be little more than serving to provide boundary conditions for the atoms in the active region. On a final note, though multiscale computational models garner the majority

of both effort and attention, the view subscribed to here is that the hunt should not be given up for *analytic* models which can capture the key features of complex materials.

ACKNOWLEDGMENTS

I am grateful to Michael Ortiz, Ellad Tadmor, Ron Miller, Vijay Shenoy, Vivek Shenoy, David Rodney, Klaus Schulten, and Laurent Dupuy for collaboration and conversation. All of them have played a key role in the development of my understanding of multiscale modeling.