J Comput Aided Mol Des (2012) 26:87–90
DOI 10.1007/s10822-011-9493-2

PERSPECTIVE

Is there a future for computational chemistry in drug research?

Gerald M. Maggiora

Received: 25 October 2011 / Accepted: 8 November 2011 / Published online: 19 November 2011

Abstract Improvements in computational chemistry methods have had a growing impact on drug research. But will incremental improvements be sufficient to ensure this continues? Almost all existing efforts to discover new drugs depend on the classic one target, one drug paradigm, although the situation is changing slowly. A new paradigm that focuses on a more systems biology approach and takes account of the reality that most drugs exhibit some level of polypharmacology is beginning to emerge. This will bring about dramatic changes that can significantly influence the role that computational methods play in future drug research. But these changes require that current methods be augmented with those from bioinformatics and engineering if the field is to have a significant impact on future drug research.

Keywords Computational chemistry · Biological reductionism · Emergent properties · Systems biology · Polypharmacology · Drug design

Since computational chemistry touches numerous areas of chemistry and biology, the road ahead has many branches. Here we will follow the branch corresponding to chemical biology and drug research, since work in these areas is routinely published in the *Journal of Computer-Aided Molecular Design*. The key issue is in what direction this branch should be heading. To address this, it is first necessary to provide some background on the current state of research in computational chemistry, biology, and drug research.

Computational chemistry and its close companion chemical informatics have made substantial progress over the past several decades, driven in large measure by an explosive growth in the speed and affordability of computer power. This has enabled the handling of massive datasets containing several million molecules and their associated biological and physico-chemical properties. Mathematical models for computing many of these properties have improved significantly. Considerable progress has also been made in quantum and molecular mechanical computations of structure and properties. Computed structures now rival experimentally determined ones for many organic molecules and even some biomacromolecules. And a number of properties difficult or impossible to measure experimentally can now be obtained with improved accuracy.

In addition to the more static properties of molecules, the capability of realistically simulating a number of complex mesoscopic and macroscopic systems has improved tremendously. For example, simulations of a number of membrane bound proteins in realistic aqueous environments have been carried out over the last several years. And it is safe to say that even more complex systems will be tackled in the near future.

Such dramatic improvements along with the development and refinement of new methods suggests that computational chemistry will continue, as it has in the past, to enhance our understanding of the molecular components that make up biological systems, but what about its role in drug research? Will a changing drug discovery paradigm bring about a shift away from computational chemistry methods? And, if so, how must computational chemistry change to meet this challenge?

G. M. Maggiora (✉)
University of Arizona College of Pharmacy, Tucson, AZ, USA
e-mail: gerry.maggiora@gmail.com

G. M. Maggiora
Translational Genomics Research Institute, Phoenix, AZ, USA

Suppose, for the sake of argument, that within the next 10–15 years we could compute most of the biologically relevant properties (e.g., structure and dynamics) of molecules to a level of accuracy that is equal to or better than can be determined experimentally. And suppose that we could simulate in molecular detail complex systems such as membranes with their associated proteins, or the supramolecular machines responsible for protein synthesis, within realistic aqueous environments to an accuracy also rivaling experiment. Would we understand living systems better? Our understanding of the molecular components of these systems would certainly be improved, but the same may not said for the overall systems, which contain the essential components of living systems.

What then are some of the impediments to a fuller understanding of the complex systems found in biology? One impediment is the tradition of biological reductionism that pervades much of the research in the biological sciences. Biological reductionism seeks to resolve the workings of biological systems into smaller and smaller components. Research in biology has been extremely successful in this venture. X-ray crystallography has dramatically increased our knowledge of the structures of the molecules that are the fundamental building blocks of living systems. Research in molecular and cell biology has seen equally dramatic advances. All of these examples reinforce the notion underlying biological reductionism that understanding the subsystems will ultimately lead us to an understanding of the complete system. But this is similar to assuming that a knowledge of an automobile's isolated subsystems will tell one how they act in a fully integrated fashion as an automobile. Although understanding the detailed features of the individual subsystems at some level is important if we want to build them, understanding how they function in concert with one another is absolutely crucial to an overall understanding of how an automobile works.

Another key aspect of biological systems is that they, like all complex systems, possess *emergent properties*, which are properties that are not obviously derivable from knowledge of the system components themselves. For example, by observing the components of a deconstructed clock it is not obvious that their function is to measure time. The notion of analogs also plays an important role in this regard. Analogues are subsystems that carry out the same overall function, but are made up of different components organized in different ways. Both notions are relevant here.

The issue now becomes whether when confronted with a set of components of some sort, can we reconstruct the system from which they came? The answer in some but not all cases is yes, but with the important proviso that some sort of theory must exist that describes the underlying behavior of the systems. An example, taken from chemical thermodynamics provides an illustration. Systems of interest in chemical thermodynamics are typically made up of on the order of 10^{26} molecules. For these systems, in many cases, it is their macroscopic properties such as temperature, vapor pressure, heats of formation, free energies of reaction, etc. that are typically determined. However, with the development of statistical thermodynamics it is possible, in principle at least, to compute the values of these thermodynamic variables from the properties of their individual chemical components.

Such activities represent a "bottom-up" approach, which requires some form of theory (vide supra) if the higher-order behavior of the system is to be described based on knowledge of its lower-order components. This is, however, difficult to do in the case of living systems since, unlike the case of statistical mechanics applied to physicochemical systems (vide supra), detailed and computable molecular theories that deal with the basic molecular components of living systems are not available. Phenomenological theories can sometimes be used in limited circumstances, but even these theories are currently in short supply.

Alternatively, a "top-down" approach can be used. In such cases, the focus is typically on the functional aspects of the system under study, eschewing molecular detail in many cases. Here the system can be visualized as a black box or collection of interacting black boxes. Each box has a well-defined set of inputs and outputs, but the underlying mechanism(s) that transform the inputs into the observed outputs may be unknown. Nor in many cases does the mechanism need to be known. The only requirement is that a prescribed set of inputs is correctly transformed into an appropriate set of outputs. This is the essence of "reverse engineering", a well-known technique in the development of computer systems. In such cases, the nature of a subsystem residing in the black box must accurately mimic the behavior(s) of the "true" subsystem, but it need not be identical to the true subsystem since it is an analog.

As an example, consider the regulation of blood pressure, which does not necessarily require detailed knowledge of the structures of the biomolecules and biomacromolecules involved in the process. However, information on some of the molecular components involved in the process is beneficial, especially if one is trying to discover new drugs to influence the process. Even in the latter instance, as demonstrated by many early drug discoveries, detailed knowledge of the molecular components is not absolutely required. In fact, one might argue, based on the mounting number of drug recalls, that information on individual drug targets works against rather than for the discovery of new drugs. While this may appear somewhat contradictory, it makes sense from the following standpoint. By dissecting living

systems into more and more basic molecular components, the fundamental features of living systems are "edited out" so that the resulting subsystems do not faithfully capture enough of the essential features of the living system for which drugs are being developed. Unless sufficient information on the "biological environment" in which targets reside is considered in the drug discovery process, we can expect more downstream drug failures.

This is just another aspect of the aphorism "novel drug target". Because if the target, which us usually a biomacromolecule of some sort, is novel very little is, almost by definition, known about its biochemical, physiological, and pharmacological behavior within the context of a living system. This contradicts the myth that "novel targets" will by their very nature lead to beneficial new drug therapies.

In addition, the emerging field of *polypharmacology* is providing firmer evidence of the fact, which was known but largely ignored in the past, that most drugs and bioprobes interact with multiple targets, albeit to differing degrees. Thus, the issue of drug discovery is further confounded by the fact that subsystems of living systems may be affected directly through their regulatory and metabolic pathways and indirectly through adventitious interactions with many other macromolecular cellular components giving rise to so-called "off target effects." Thus, it is not surprising that drugs developed under the narrow single target, single drug paradigm (vide infra) nowadays tend to fail (i.e., are recalled).

So what does all of this mean for computational chemistry's future in drug research? Or more specifically as the title of this article proclaims, "Is there a future for computational chemistry in drug research? My answer is yes, but with some important caveats.

All scientific fields evolve, some faster than others, punctuated by occasional, dramatic paradigm shifts. Drug research is in the initial phase of one such shift, a shift that is based on reversing the prevalent reductionist trend with a more holistic systems biology approach. As this change progresses, information on the quasi-isolated molecular systems (viz. drugs and targets) studied in in vitro experiments today will play a diminishing role in future drug research. Emphasis will be placed on experiments in higher-level systems with more fully functioning biology, including whole animals.

Because of the amount, complexity, and hierarchical nature of much of the data gained in these experiments, methods for its capture, storage, retrieval, organization, and analysis are essential. The increasing amount of relevant molecular data being supplied by computation methods will further exacerbate the need for these methods. As the methods associated with the first three areas are related to well-known database technologies in existence today, their implementation is more straightforward, albeit non-trivial. Methods for dealing with organizing and analyzing the data, however, represent additional challenges.

Basically, there are two broad categories of data of importance in the description of biological systems, relational and dynamical. The former captures generally pairwise but sometimes higher-order relationships among various molecular and biological entities. The recent network-based analyses of drug-target and other similar relationships are prime examples that illustrate the complex inter-relationships that are commonplace in living systems. Add to this the familiar regulatory and metabolic pathways, which depict directional relations among molecular entities, and the magnitude of the number and complexity of the relationships becomes clearly manifest. But this so to speak is only the tip of the iceberg. What is missing is some account of the time-dependent or sequential behavior of the many biological processes, which derive from the dynamical processes in living systems.

Handling this tsunami of data and information is impossible without computers, but computers are not in themselves sufficient. What is critically needed are relational and dynamical models, from simple to complex, that capture key features of living systems that are relevant to the discovery of new pharmaceutical agents ("drugs"). This provides a golden opportunity for computational chemists to extend their already substantial capabilities for dealing with molecular entities into new directions that are more in sync with the requirements of the "new age" of drug discovery. This will require that computational chemists learn new mathematics in addition to that normally taught in the physical sciences. Bioinformatics and computational biology are helping to chart the path in this regard, although most practitioners in those fields typically lack the level of "molecular expertise" possessed by computational chemists.

Recently, the new field of *synthetic biology* has emerged that combines principles from biology, chemistry, and engineering. The use of engineering tools and principles should further enrich the study of biological systems, one of the consequences of which being the advancement of drug research. The American Chemical Society has recognized this emerging trend and that of systems biology by launching in 2012 the new journal *Synthetic Biology*, which is dedicated to the publication of papers covering synthetic biology and systems bioscience.

The power that integration of all of these fields brings to the study of biological systems suggests that the training of future computational chemists, at least those interested in chemical biology and drug research, should be expanded to include some training in all of these related fields. Is this a realistic goal given the complexities of each field? I believe it is, but only if one views the challenge from the perspective of the fundamental concepts that underlie all of

the fields. It must be borne in mind that while the nature of these concepts may be the same in each of the fields, their realizations can be quite different. This is an example of what mathematicians call an isomorphism.

To provide integrated training in computational chemistry will require careful thought as to what the most important underlying concepts are, and how to present them in a way that shows their generality and at the same time provides information that illustrates their application within the different disciplines. Thus, a balance will have to be struck between descriptions of the general principles underlying all of the disciplines and the amount of information explicitly associated with each of them. In addition, what can loosely be called computational chemical biology is not primarily about sophisticated and elegant computational methods for interrogating biological systems, rather it is about enhancing our understanding of these systems at all levels of resolution. Thus, future computational chemists must be able to converse intelligently with biologists, chemists, and engineers involved in drug research. Admittedly this is a tall order for anyone, but it is a goal that is worthy to aspire to.

As an example, consider the concepts of "system" and "structure." The concept of system provides a framework that encompasses almost all aspects of modern biology, chemistry, and engineering and many human endeavors for that matter. The concept of structure is taken here in its most general sense as the arrangement of and relations between the parts or elements of a given entity. Thus, structure in this more general sense refers not only to molecular structure but also to the structure of regulatory and metabolic pathways, the structure of relational networks of drugs and their targets, and the structure of cells, organs, and tissues, to name a few. Such an overarching view can eventually lead to more comprehensive approaches to many biological problems that merge chemistry and engineering methods with those from biology. This should enhance our ability to not only move within various layers of the biological hierarchy but also between layers, a capability that will be essential to success in future drug research projects. Plus it would enable us to have a deeper understanding of the level of description needed in the study of these systems.

All of this points to the fact that the way computational chemists are trained, especially those interested in the study of chemical biology and drug research, needs to be rethought. An understanding at some level of the mechanisms that underlie the functions of biological systems and the way in which drugs modulate these functions is essential to a deeper understanding of biology in general and drug design specifically. Training of future computational chemists must reflect this reality.