



PERSPECTIVE

From computational quantum chemistry to computational biology: experiments and computations are (full) partners

To cite this article: Buyong Ma and Ruth Nussinov 2004 *Phys. Biol.* 1 P23

View the [article online](#) for updates and enhancements.

Related content

- [The origami of life](#)
Timothy R Lezon, Jayanth R Banavar and Amos Maritan
- [Applications of ab initio atomistic simulations to biology](#)
Matthew D Segall
- [Concepts and schemes for the re-engineering of physical protein modules](#)
Carlos Alemán, David Zanuy, Ana I Jiménez et al.

Recent citations

- [Jordi Valverdú](#)
- [An Epistemological Analysis of QSPR/QSAR Models :](#)
Jordi Valverdú
- F. Apollonio *et al*

PERSPECTIVE

From computational quantum chemistry to computational biology: experiments and computations are (full) partners

Buyong Ma¹ and Ruth Nussinov^{1,2}

¹ Basic Research Program, SAIC-Frederick, Inc., Laboratory of Experimental and Computational Biology, NCI-Frederick, Frederick, MD 21702, USA

² Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

E-mail: mab@ncifcrf.gov and ruthn@ncifcrf.gov

Received 11 October 2004

Accepted for publication 21 October 2004

Published 16 November 2004

Online at stacks.iop.org/PhysBio/1/P23

doi:10.1088/1478-3967/1/4/P01

Abstract

Computations are being integrated into biological research at an increasingly fast pace. This has not only changed the way in which biological information is managed; it has also changed the way in which experiments are planned in order to obtain information from nature. Can experiments and computations be *full* partners? Computational chemistry has expanded over the years, proceeding from computations of a hydrogen molecule toward the challenging goal of systems biology, which attempts to handle the entire living cell. Applying theories from *ab initio* quantum mechanics to simplified models, the virtual worlds explored by computations provide replicas of real-world phenomena. At the same time, the virtual worlds can affect our perception of the real world. Computational biology targets a world of complex organization, for which a unified theory is unlikely to exist. A computational biology model, even if it has a clear physical or chemical basis, may not reduce to physics and chemistry. At the molecular level, computational biology and experimental biology have already been partners, mutually benefiting from each other. For the perception to become reality, computation and experiment should be united as full partners in biological research.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

It is currently increasingly recognized that to understand biological systems at their various levels of complexity, ranging from structure and dynamics of a single molecule to cellular networks and organisms, the traditionally more quantitative fields of physics, chemistry, computer science and other mathematics-based disciplines are essential. Here we advocate an equal-partnership, full integration of experiments and computations. Traditionally, biological studies have been dominated by experiments. Computations have been

a tool used by a theoretical observer to obtain information from nature [1]. Here, we ask whether such a partnership proposition can be realized.

The goal of *computational chemistry* is to model chemistry as closely as possible to reality using calculations rather than experiments. According to the National Institutes of Health (<http://www.bisti.nih.gov/>), the overall aim of *computational biology* is ‘the development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, behavioral, and social systems’. Yet,

whereas in studies of biological systems, computational chemistry and computational biology share a common boundary, the behavioral and social systems are far from the chemical world. Still, one may argue that a biologically critical molecule such as DNA serves as a bridge between chemistry and the ‘behavioral and social systems’. Hence, the question arises—can computational biology be achieved by a ‘bottom-up’ approach, where we start by practising computational chemistry and eventually compute the social behavior which, at the end, is a consequence of the DNA molecule?

Apart from the chemical nature of the DNA molecule and its beautiful double helical structure, its most prominent feature is its digital code [2]. The DNA specifies the genes which encode proteins and the gene regulatory networks dictating genomic behavior. Within this framework, computations may be divided into three categories: those related to the physico-chemical and the structural behavior of the DNA itself; those addressing the behavior of RNA and proteins; and systems biology, the biology of the regulatory networks. Eventually, the goal of computational biology is to compute the components of the networks and their inter-relationships. To explore whether this goal is reachable or too ambitious, we examine the performance of computational chemistry and computational biology over the past thirty years.

2. Computation and experiment at the molecular level

The modern era of computational chemistry started in 1970 [3]. Between 1962 and 1970 there was essentially a universal scientific agreement that the methylene molecule was linear in its triplet ground state, as concluded by the brilliant spectroscopist Gerhard Herzberg (the father of modern spectroscopy) from experiments described in his Nobel Prize citation. In 1970, the theoretical treatment by Bender and Schaefer applied rigorous quantum mechanics to the triatomic molecule. Previously, the method had been applied only to atoms and diatomic molecules. Their theoretical result, which predicted that the methylene molecule was bent by 135° , contradicted experiment. Yet, indirect experimental evidence for such a highly bent methylene molecule came out quickly, followed by a reinterpretation of the spectroscopic studies confirming the bent geometry predicted by theory. The reliability of a molecular quantum mechanical model for chemistry and a new role for theory, ‘full partner with experiment’ was charted. Since then, computational quantum chemistry advanced to a stage allowing us to reliably use computed results to calibrate or to substitute experiments in many small molecule systems.

Standing back and looking at the current state of computational biology, we observe that many studies, including some from our own lab (for example, [4, 5]), focus on explaining experimental results. This reminds us of the warning statement in 1970 about the practice of quantum chemistry. Even Mullican claimed that his initial work in quantum mechanics ‘interpreted’ rather than ‘discovered’ chemical facts [6]. Alberte Pullman commented in 1970, ‘While it is certainly indispensable that theoretical chemist

constantly try to improve the values of the size they calculated and more and more approach exact energy values . . . quantum chemistry risk giving the impression that its essential goal is *reproducing by uncertain methods known results*, in contrast to all other sciences whose goal is to use well-defined methods for the research of unknown truths’ [6] (our emphasis). Thus, the question arises as to whether computational biology today is in the similar state to that of computational chemistry 30 years ago: namely *interpret* rather than *discover* biological facts.

Computational design of novel globular protein folds and of protein–protein interfaces at atomic-level accuracy is no longer a dream [7, 8]. David Baker’s lab has successfully computationally designed a novel global protein fold with a remarkably high accuracy. The RMSD (root mean squared deviations) between the designed model and the solved crystal structure is only 1.17 Å. Using a ‘computational second-site suppressor’ strategy, the same group has further redesigned a DNase-inhibitor protein-pair interface, also confirmed by high-resolution x-ray crystallographic analysis. The designed switch in specificity was observed in *in vitro* binding and in functional assays. Computational design of protein function also enjoyed impressive results, achieving a high degree of control in some biological and biosensing activities (even a TNT molecule) [9]. A structure-based computational method was used to construct soluble receptors that bind trinitrotoluene, L-lactate or serotonin with high selectivity and affinity. These engineered receptors were also incorporated into synthetic bacterial signal transduction pathways, regulating gene expression in response to extracellular trinitrotoluene or L-lactate. In yet another remarkable feat, Kaplan and DeGrado designed a protein sequence using a computational method that not only considered the stabilization of the desired fold, but in addition the destabilization of likely alternatives [10]. The catalytic function of the designed protein was confirmed by subsequent experiments.

Our group also achieved a great synergy with experimental studies. The amyloid structure for the A- β peptide has long been sought by structural biologists. To control Alzheimer’s disease, it is essential to understand how the peptide aggregates. Based on extensive molecular dynamics simulations, we proposed that the A- β amyloid proto-fibril should have a bent sheet structure, with the bent region stabilized by a salt bridge [11]. Encouragingly, simultaneously, a similar model was proposed based on experimental solid state NMR studies [12]. Recently, our study of the ribosomal release factors has led to a computational proposition: release factors eRF1 terminate protein synthesis by recognizing stop codons on the mRNA via their conserved amino acid motifs (NIKS), and by the conserved tripeptide (GGQ) interactions with the ribosomal peptidyl-transferase center (figure 1). Crystal structures of eRF1 (figure 1(a)) do not fit their ribosomal binding pocket (~ 73 Å). We found that the conformational transitions and dynamics of the eRF1 between the free and ribosome-bound states are controlled by the protonation of histidines. For eRF1, the distance between the NIKS and GGQ motifs shrinks from 97.5 Å in the crystal

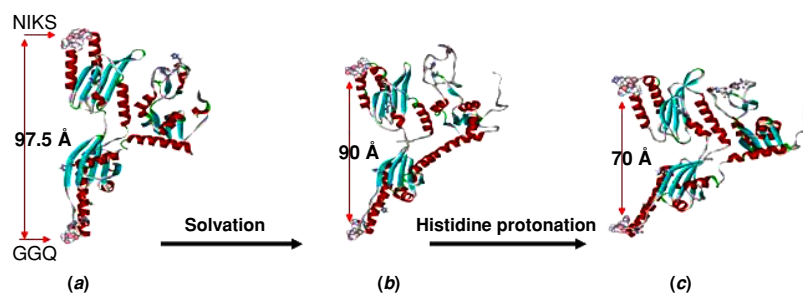


Figure 1. A computation-based proposition for the release factor eRF1 conformational change from the crystal structure to the active form. Solvation relaxes the open conformation a little and then histidine protonations drive the conformation to its active closed state.

to 70–80 Å. *eRF1* functions like a molecular machine, fueled by histidine protonation [13]. This is not surprising, since the eRF1 binds to the ribosome, the master molecular machine. Our proposition can be tested experimentally by monitoring the conformational changes of eRF1 as a function of changes in the pH.

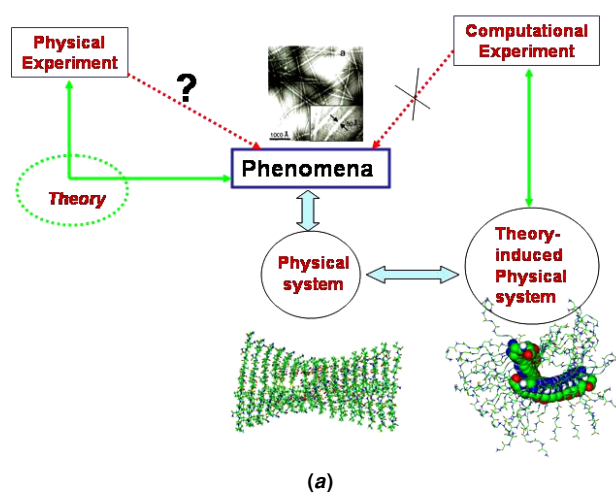
These results indicate that computational biology can (and should be) an equal partner to experiment. Computational approaches can be used as ‘experimental tools’ to obtain information from nature. The philosophical significance of such an approach is illustrated in figure 2(a). Physical experiments do not record phenomena in a completely objective manner; rather, they *already involve at least some theory*. At the same time, computational experiments provide complementary tools to physical experiments. Conceptually, scientific theories used in computation are mere models to describe the phenomena: since the physical systems are used to characterize the phenomena in terms of a few parameters abstracted from the phenomena, they are abstract replicas of the actual phenomena. What we study in our computational virtual world is the theory-induced physical system, describing a real-world phenomenon. For the perception to become reality, computation and experiment should be full partners in biological research. The success of computational quantum chemistry and of some computational biological studies at the molecular level indicates the advantages in full synergy between computation and experiment.

3. Future development at the systems level

Currently, systems biology is a major challenge in computational biology. The aim of systems biology is to characterize the network of intermolecular interactions and their regulation. Systems biology is the next level up, progressing from molecules to cells. Undertaking systems biology is essential if we are to fully understand and predict the biological system. The significance of the challenge is not only in the complexity of the cellular system; rather, it is the shift in the perception of the physical and the informational sciences, leading to their deep involvement in this crucial endeavor.

p53, the tumor suppressor protein, may be taken as an example. p53 is a crucial protein not only against cancer as a tumor suppressor (50% of cancers relate to p53 mutations),

Physical Experimentation versus Theoretical Experimentation



Solution to the challenge: integration

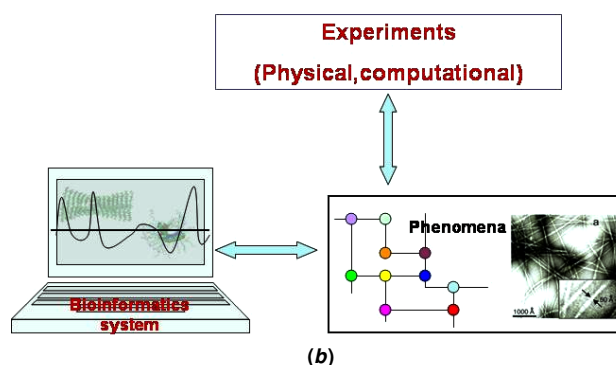


Figure 2. An illustration of the current relationship between computational experimentation and physical experimentation and the suggested integration to understand the biological system.

but also for systems biology. It is one of the most connected hubs in the cell [14], regulating more than 160 genes and maintaining genome stability. Conventional physical and computational approaches can address problems such as mutations and the structural stability of the p53 molecule

or its interaction with DNA. However, new approaches are essential in order to investigate the *informational content* of the p53 interaction networks, how its interactions affect the DNA transactivation and how genome stability changes as a result of the altered pattern of interactions.

The new approach must fully integrate physical and computational experiments. Figure 2(b) sketches our thoughts on such a united strategy, as compared with the conventional indirect interactions between the two disciplines. The main point of our suggested approach is that computations should not only be used as bioinformatics tools; rather, they should also build the system for the physical experiment to test, modify and alter as needed. The new territory and practice of systems biology is being defined as research progresses. The physical and informational sciences and the integrated wet and computational experiments combine to meet this emerging challenge.

Acknowledgments

The projects described here have been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number NO1-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names,

commercial products, or organizations imply endorsement by the US Government.

References

- [1] Ma B 1995 *PhD Dissertation* University of Georgia
- [2] Hood L and Galas D 2003 *Nature* **421** 444–8
- [3] Schaefer H F 1986 *Science* **231** 1100–7
- [4] Pan Y, Ma B, Keskin O and Nussinov R 2004 *J. Biol. Chem.* **279** 30523–30
- [5] Ma B and Nussinov R 2003 *Protein Eng.* **16** 561–75
- [6] Nye M J 1993 *From chemical philosophy to theoretical chemistry* (Berkeley, CA: University of California Press)
- [7] Kuhlman B, Dantas G, Ireton G C, Varani G, Stoddard B L and Baker D 2003 *Science* **302** 1364–8
- [8] Kortemme T, Joachimiak L A, Bullock A N, Schuler A D, Stoddard B L and Baker D 2004 *Nat. Struct. Mol. Biol.* **11** 371–9
- [9] Looger L L, Dwyer M A, Smith J J and Hellinga H W 2003 *Nature* **423** 185–90
- [10] Kaplan J and DeGrado W F 2004 *Proc. Natl Acad. Sci. USA* **101** 11566–70
- [11] Ma B and Nussinov R 2002 *Proc. Natl Acad. Sci. USA* **99** 14126–31
- [12] Petkova A T, Ishii Y, Balbach J J, Antzutkin O N, Leapman R D, Delaglio F and Tycko R 2002 *Proc. Natl Acad. Sci. USA* **99** 16742–7
- [13] Ma B and Nussinov R 2004 *J. Biol. Chem.* (Published online 18 October)
- [14] Vogelstein B, Lane D and Levine A J 2000 *Nature* **408** 307–10