Why computational methods for the study of biological macromolecules and their effectors?

Rachid C. Maroun

To cite this version:

Rachid C. Maroun. Why computational methods for the study of biological macromolecules and their effectors?. Boletín médico del Hospital Infantil de México, 2016, 73 (6), pp.363-364. 10.1016/j.bmhimx.2016.10.005 . hal-02174490

HAL Id: hal-02174490
https://univ-evry.hal.science/hal-02174490

Submitted on 18 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bol Med Hosp Infant Mex. 2016;**73(6)**:363–364

Boletín Médico del
Hospital Infantil de México
www.elsevier.es/bmhim

EDITORIAL

Why computational methods for the study of biological macromolecules and their effectors?

¿Por qué utilizar métodos computacionales para el estudio de las macromoléculas biológicas y sus efectores?

Perhaps we should start this editorial by raising the following question: Why appeal to theoretical approaches for studying biological macromolecules and the effectors with therapeutic potential? Let us take the time to look at some statistics concerning the biological databases. As of June 2016, there were around 9000 completed and published genome sequences (archaea, prokaryotes, and eukaryotes) including strains or varieties of the same species (GOLD database, https://gold.jgi.doe.gov/).

The exponential growth of sequenced genomes has kept steady over the past 15 years. On another hand, the growth of the databanks containing protein sequences doubles in less than 1.5 years. As a result, there are more than 15 million protein sequences at 50% sequence identity in the UniProt database (http://www.uniprot.org/). However, the growth of the databank of the 3D structures of proteins, PDB (http://www.rcsb.org/pdb/), is several orders of magnitude smaller (10^{-3}). As it is, the number of non-redundant sequences in the PDB database at 30% sequence identity is of approximately only 26000; not to mention that the overall number of membrane protein structures (pharmacological targets par excellence) remains small in spite of recent experimental advances in the determination of their 3D structure; it only reaches less than 300 redundant structures in the PDB. Accordingly, there is an increasing need to develop and apply *in silico* methods to obtain reliable 3D molecular models, given that the experimental determination of the 3D structures of all sequences is impossible on the one hand, and useless on the other hand. Indeed, nowadays there are computational methods that can generate reliable 3D models in a reasonable amount of time for a large number of protein sequences, without having to perform the experimental structure determination. This theoretical approach leads to saving time and money. For example, the IDC market research firm indicated that for $1 invested in modeling and simulation software, $3 to $9 were returned in incremental revenue and costs savings.

But why is it so important to determine the 3D structure of a macromolecule such as a protein? Answer: the way in which the structure is linked to the sequence and the function is of fundamental importance. This is what we may call the structural biology dogma. Thus, a protein that is not in its native conformation will in general not show its expected biological activity or function. On the other hand, a given protein sequence will always fold into the same conformation under the same conditions.

Nevertheless, beyond the architecture of a single protein, proteins interact with partners, such as other bio-macromolecules and ligands to exert their functions. Specific molecular recognition of proteins as targets, and small molecules as ligands adopts a special interest given the possibility to modify, suppress or modulate the activity of a given protein through non-endogenous, artificial ligands.

A druggable target is one that is capable of binding to the drug and whose activity can be modulated by it. The target may be known or may need to be predicted. Important properties in the recognition of a druggable target by a ligand are the affinity and the selectivity (the ability of the ligand to differentiate between different acceptors). Molecular recognition uses principles of hydrophobicity, steric, physicochemical, and electrostatic complementarity. In addition, internal dynamics and mutually induced conformational changes may take place upon the binding between the target and the drug molecules that adopt ''bioactive'' conformations. In order to improve then the drug discovery process, the internal motions of the partners involved in the interaction need to be described and taken into account, including allosteric effects. All this knowledge contributes to the understanding of the molecular basis of disease.

It is useful to know that the chemical space of potential pharmacologically active molecules may contain around 10^{62} molecules. The known chemicals are in the order of 10^{7}. The known drugs occupy thus 10^{-55} of the ''active'' chemical space. A long way to go, in any case! Now, there

http://dx.doi.org/10.1016/j.bmhimx.2016.10.005
1665-1146/© 2016 Hospital Infantil de México Federico Gómez. Published by Masson Doyma México S.A. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

are several approaches and levels for the rational discovery, design, and development of *de novo* pharmaceuticals. Structure-based drug design applies when the structure of the biological target is known. For that purpose, mining of drugs is accomplished through access to libraries of drug-like chemical compounds stored in public and corporate databases. These databases represent an enormous diversity of tens of millions of potential drug molecules (enzyme activators and inhibitors, receptor agonists and antagonists, ion channel openers or blockers, modulators that bind to secondary sites) and their effects. However, it is still necessary to estimate their ADME-Tox pharmacokinetic properties from the chemical structure (Absorption/solubility), Distribution, Metabolism, Excretion – Toxicity (carcinogenicity, mutagenicity, oral LD50, developmental toxicity potential, skin sensitization) in order to filter out those compounds that do not possess the appropriate properties. The searched increase in specificity allows the drug to bind to the desired target and binding pocket(s), reducing adverse drug reactions. Subsequently, virtual screening with computers of the results of data mining through ligand docking leads to the formation of the protein-ligand complexes and an estimation of their affinity. Of course, binding site identification must come previously to the docking of the ligand to it. Medicinal chemists play a fundamental role in the optimization of an initial compound from hit to lead since their knowledge points to compounds that can be realistically synthesized. Another aspect that must be taken into consideration more often is that it is the metabolite(s) of the drug that actually binds to the intended target molecules and not the administered parent molecule.

When the structure of the receptor molecule is unknown, then ligand-based design is used based on the compounds binding to the biological target of choice. In such situation, the search for a pharmacophore, i.e. a schematic model of the compound with the structural and physicochemical properties needed to bind and exert the desired effect on the target, is required. The ligand-based design may also lead to mapping the receptor binding site. In complement to the ligand-based approach, a quantitative structure-activity relationship (QSAR) may also be obtained to conceive and predict the activity of new analogs. Considering these variables helps to increase enrichment in the hit-to-lead process. However, room for improvement exists in the docking simulations between protein and ligand. As we know, water molecules play essential structural and functional roles in biology. Protein-ligand complexes will contain water molecules mediating the interaction of both partners. These water molecules contribute to the formation and stability of the complex. Until very recently, the complexes obtained through computation excluded the presence of the aqueous solvent. In order to improve the determination of the thermodynamic Gibbs free energy of complex formation in solution required for obtaining the protein-ligand binding affinity, a solvated docking approach with an explicit treatment of water molecules needs to be generalized.

With the advent and progress of computer power and powerful software, the drug design cycle has been enormously optimized. Even though present computational methods provide mostly qualitative results, they contribute by eliminating many cycles in the drug design process. Successful stories of computer-assisted rational drug design abound nowadays and have led to applications in the field of chemotherapy in cancer, antivirals, agonists and antagonists of membrane receptors and proteins (antipsychotics, antidepressants). In this issue of the *Boletín Médico del Hospital Infantil the México* dedicated to cancer, the article by Prada-Gracia et al. illustrates the applications of these methods in drug design and the implications for novel cancer therapies.[1] Moreover, an updated review article by Moreno-Vargas and Prada-Gracia examines the computational approach to drug discovery in the physiopathology of cancer.[2]

I have witnessed the different fields of computer-assisted drug design, structural bioinformatics, and cheminformatics go a long way since the early 1990s when I was doing postdoctoral work at the Faculté de Pharmacie of the Université de Paris V–René Descartes. Despite reticence, doubts, and skepticism, these fields have finally passed the tests of experience and proven worthy. It is true that translational medicine (i.e. bringing a drug from bench to bedside) remains a costly, complex and time-consuming process with no guarantee of success, taking a drug at least five years to make it to the marketplace at the cost of several billion dollars. Nevertheless, the computational approach will play an ever increasing role in overcoming those bottlenecks. Then, pharmacogenomics and precision medicine will contribute to more efficient drug development and therapies destined to improve the life and health of patients.

Undeniably, the multidisciplinary rational drug discovery approach involving biologists, chemists, physicists, computer scientists and medical doctors does not keep creativity and serendipity from playing a role, reminding us that ultimately ''There is no logical path to these laws; only intuition, resting on sympathetic understanding of experience, can reach them'' (Albert Einstein).

References

1. Prada-Gracia D, Huerta-Yépez S, Moreno-Vargas L. Application of computational methods for anticancer drug discovery, design, and optimization. Bol Med Hosp Infant Mex. 2016;73:411–23.
2. Moreno-Vargas LM, Prada-Gracia D. New perspectives on the computational characterization of the kinetics of binding-unbinding in drug design: implications for novel therapies. Bol Med Hosp Infant Mex. 2016;73:424–31.

Rachid C. Maroun
Laboratoire Structure et Activité des Biomolécules
Normales et Pathologiques, Université d'Évry-Val
d'Essonne, Évry, France
E-mail address: rmaroun@gmail.com