Short Communication

A Point of View about Computational Challenges for the Design of New Drugs for Combating Malaria

Raúl Isea*

Fundación Instituto de Estudios Avanzados, Hoyo de la Puerta, Baruta, Venezuela

*Corresponding Author: Raúl Isea, Fundación Instituto de Estudios Avanzados, Baruta, Venezuela.

Abstract

The development of new drugs is a problem with a difficult solution for overcoming from a computational point of view. This is so due to the lack of programs that could manage great volumes of data in many different formats from the entire world. As a consequence, it is crucial to solve this digital divide in order to perform *in silico* experiments for the design of new medicines in a short period of time. To visualize these aspects, it will be emphasized in the case of the malaria problem with special attention on *Plasmodium falciparum*.

Keywords: Docking; In Silico Experiment; Docking; Malaria, Plasmodium falciparum

Introduction

Currently, the development of new drugs is a challenge from the computational point of view and for most of the cases without a solution. As an example, we can mention that in the last twenty years, there has been an exponential increase in the number of available genomic sequences for Academia and Industry due to advances in the automation of the biotechnological processes. In this sense, if we take into account the volume of information that is now stored in the National Center for Biotechnology Information database, we see that the total number of sequences provided by several international institutions sum up to 203.953.682 [1]. Such a collection is composed by 244.914.705.468 nucleotides (October 2017). In order to have an idea of what such a volume of data represents, we can cite the work by the National Human Genome Research Institute "*Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project*" [2], which is written with 119.465 characters (not taking into account the blanks) in 18 pages. If we think of a kind of illustrative equivalence in a way that a character would represent a nucleotide, we could infer that looking for a specific calculation in the NCBI database would mean to do so in more than thirty nine millions of different pages.

Even more, in the last years and because of the huge amounts of data that is available, several metadata databases have been created and can be consulted by the scientific community. For example, the Entrez database provides access to 37 different databases where we can read a total of 528.176 books, 4.066.155 full text journal articles, and so on as of September, 2017.

Perhaps you feel that there is little information that has been studied worldwide with a particular emphasis on the malaria problem. Remember that malaria is a tropical disease caused by the protozoan parasite *Plasmodium*, and human malaria is caused by five different species with the most dangerous being *Plasmodium falciparum* where half of the world's population is at risk of malaria according to the World Health Organization.

The first step for the design of new antimalarial medicines by computational methods is the identification of the proteins that should be considered as the target of an interaction with a specific drug. In doing so, a bibliographic search must be done for all the works related to that disease. The NCBI database is usually the most common reference since it maintains a huge catalogue of publications; thus, searching only by the "*drug+design*" term, almost five hundred thousand results appear as full text journal articles and more than sixteen thousand books in the Entrez database. If we combine it with a specific disease, for example malaria, we find 20.068 full text journal articles until December, 2017. If one reads an average of five articles per day, we need to spend eleven years to read all these articles. At this point, many of the publications are not going to be of interest. The point is that it requires a considerable amount of time to review them.

Once the bibliographic search has concluded, the next step is to identify all the cited proteins in the bibliography which is crucial for the final success of the research. Thus, *Plasmodium falciparum* has been selected for having the highest viral impact where more than a hundred of proteins have been identified as possible targets for interacting with a drug, available in TRD Targets database [3].

For each sequence, it is necessary to find a similar relationship and for do that we employing Blast algorithm. This step is required for both the stored space and power computing. To understand the last comment, we mention the calculation that was performed at the Argonne National Laboratory with a supercomputer where more than one thousand processors were simultaneously employed [4] and where one of their objectives was sequence searching in the entire microbial genome database against itself that could be similar, so the identification of those still uncharacterised could be done. This comparison generated one Petabyte of data which is the equivalent to two thousand hard disks of 500 GB each. In that sense, it is important to innovate in the development of computational techniques that could deal with any kind of information irrespective of its volume in an efficient way.

The final step for the selection of potential drugs that could inhibit or activate a biological response in the malaria case is to use the *in silico* Docking technique. This methodology consists in quantifying the interaction between the drug and the protein by means of calculating the common interaction energy. In doing so, a wide range of possible random positions of the drug around the protein must be simulated and any of the relevant energies are calculated. After such a complex calculation, the position with the lowest energy is selected. Of course, these calculations are limited by the selection of the drug or a family of them; in order to overcome such an issue, a drug database such as ZINC (available at http://zinc.docking. org) is used. Thus, the discovery of new useful medicines against a specific disease is easier due to more than six millions of drugs per target protein are simulated.

There is currently a wide portfolio of computational tools that allow such calculations; among them are AutoDock (http://autodock. scripps.edu/), ZDOCK – RDOCK (http://zlab.bu.edu/zdock/index. shtml), DOCK (http://dock.compbio.ucsf.edu/), DOT (http://www. sdsc.edu/CCMS/DOT/), FTDock-RPScore MultiDock (http://www. sbg.bio.ic.ac.uk/docking/), GRAMM (http://vakser.bioinformatics. ku.edu/resources/gramm/gramm1/), Hex (http://www.csd.abdn. ac.uk/hex/), and so on. The selection of one of these applications is a research line itself. We will mention three different cases with their computational used resources as examples.

First, the calculation performed in the IBM supercomputer Blue Gene with the DOCK application where 16384 processors PowerPC 440 dual-core (1 GB RAM) were used. The objective was to identify all the possible effective drugs against HIV-1 [5]. Second, the calculation performed in the San Diego Supercomputer SDSC for the identification of new medicines against the avian flu [6]. There were 27 possible drugs similar to Tamiflu that were determined using of the AutoDock software in the 128 PowerPC processors.

In the third case, the data challenge was devoted to fight against malaria promoted by the *Wide In-Silico Docking Of Malaria* (WISDOM) initiative [7]. Even when such a calculation was performed in several hundreds of processors, the consumed CPU time lasted for more than 76 days in a computational platform distributed around the world.

Although the main problem was to process the volume of information that is generated by such a methodology (WISDOM produced 1600 GB of data), the discovery of new drugs by computational techniques is possible. This management of information has also a limitation in that the high number of solutions resulting from the computational point of view is correct. However from the biological, it is not. There are false-positives. They are produced because the energetic interaction between the drug and the protein is not a correct solution since both elements are too close to each other so there is no difference between the derived Docking medicines.

As a consequence, an energetic function which minimizes the number of false-positives must be implemented. Besides, and to the knowledge of the author, new physical and chemical parameters have to be taken into account that could simulate the answer produced by the metabolic lattice of the targeted protein. This supposition is because the selected protein is not isolated and it might obey a cellular determined reaction.

Conclusion

This work emphasizes the necessity of innovation in the development of computational platforms that integrate the available information from different research centres. At the same time, new methodologies to reduce the number of false-positives created by the Docking technique must be implemented. It is also necessary to develop computational techniques that allow us to optimize the amount of information that is generated when the sequences are aligned and integrated with the scientific literature. Finally, a step beyond is the interdisciplinary cooperation that generates new computational approaches as that of WISDOM-2 [7], which produces a huge volume of data, but can also reduce the number of *in vitro* experiments (the only ones that validate the previous *in silico* results).

Conflict of Interest

The author declares no conflicts of interest in this article.

Acknowledgment

The author wishes to express his sincere thanks to Professor Karl E Lonngren for your unconditional help and the comments concerning this manuscript.

Bibliography

1. NCBI Resource Coordinators. "Database Resources of the National Center for Biotechnology Information". *Nucleic Acids Research* 4.45 (2017): D12-D17.

2. ENCODE Project Consortium. "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project". *Nature* 447.7146 (2007): 799-816.

3. María P Magariños., *et al*. "TDR Targets: a chemogenomics resource for neglected diseases". *Nucleic Acids Research* 40 (2012): D1118-D1127.

4. Balaji P., *et al*. "Distributed I/O with ParaMEDIC: Experiences with a worldwide supercomputer". International Supercomputing Conference (2008): 1-10.

5. Moustakas DT., *et al*. "Development and validation of a modular, extensible docking program: DOCK 5". *Journal of Computer-Aided Molecular Design* 20.10-11 (2006): 601-619.

6. Cheng LS., *et al*. "Ensemble-Based Virtual Screening Reveals Potential Novel Antiviral Compounds for Avian Influenza Neuraminidase". *Journal of Medicinal Chemistry* 51.13 (2008): 3878-3894.

7. Vinod Kasam., *et al*. "WISDOM-II: Screening against multiple targets implicated in malaria using computational grid infrastructures". *Malaria Journal* 8 (2009): 88.