Virtual Screening for Drug Discovery; Hurdles to Overcome for Better Drug Prediction

Anton Sameera Perera

Department of Chemistry, University of Kentucky, USA

The cost to develop a new drug that would enter the market is found to be $2.6 billion, and only a percentage of less than 12% of new drug candidates that would enter clinical trials would obtain FDA approval as a prescriptible medication.[1] The rational molecular design could potentially change this drastically and save a lot of money and time by eliminating candidates that fail in the process of only selecting candidates with a chance of being ultimately successful. Virtual screening is one such method where computational chemistry simulations are used to screen molecules instead of using conventional biochemical assays. Antiviral drug prediction has recently become a hot topic in science due to the COVID-19 outbreak, where scientists in the whole world are challenged to develop a cure within the shortest period of time in history. In such an endeavor, computational prediction, if correctly executed, could become an ultimate deal-breaker.

However, the question remains as to how far computer-aided drug design (CADD) can bring us in terms of drug discovery. A biological system, in my opinion, is the most complex entity a computational chemist/biologist will ever try to simulate. Any computational model in its core is a type of mathematical expression or correlation to a physical system or phenomena in the real world. A biological system per se would ideally comprise of a countless number of independent/interdependent variables. It is highly unlikely that scientists would be able to address all of them in the near future, even with state-of-the-art computational resources. However, possibly the dawn of commercial level quantum computing would be the next most significant step in technological evolution where pure analytical solutions as such would be a reality. Therefore, computational efforts in this regard are often simplified to overcome the difficulty with the cost of computation. Thus, in order to understand these issues, we should dig into some basics of computational simulations.

Among the many different methods used in Virtual screening, one of the most popular methods is molecular docking. Molecular docking of small molecules to protein binding sites was initiated in the early 1980s, yet continues to be a highly active area of research.[2,3] When

merely the structure of a target protein/enzyme and its active or binding site is located, docking is primarily used as a hit- identification tool. I presume that it is quite reasonable to start addressing certain limitations of the field using this familiar example among many chemists and biologists in Sri Lanka. Even though there may be plenty of other issues pertaining to the field, I would narrow it down to the most compelling three limitations, in my opinion, that one should be aware of when making use of this great toolset. Nevertheless, these can technically be applied to most of the other techniques used in virtual screening in general. The first and foremost is the intrinsic restrictions that are not yet resolved in theoretical chemistry. Second would be the level of accuracy and applicability of the method to the question of interest and its ability to capture the expected experimental outcome. Finally, the limited search space confined to the molecule libraries used, which is more specific to the screening of larger data sets. Thus, in order to understand these issues, we should dig into some basics of computational simulations.

Computational simulations of chemical systems have made its way of becoming its own subdiscipline in chemistry while having a substantial effect on other subdisciplines in chemistry within the past few decades.[4] These simulations have made possible the prior prediction and rational explanation of chemical and physical properties in simple di-atomic systems to extremely complicated biological systems. A chemical system can be modeled using multiple approaches, such as with an analytical method or with a machine-learning method based on empirical data. An analytical approach can either be constructed on simple Newtonian mechanics or rigorous quantum mechanics. Quantum mechanical calculations are proceeded either from an *ab-initio* method or using the electron density as a function of chemical structure with a DFT (Density functional theory) calculation.

Nevertheless, quantum mechanics (QM)-based methods are built on solving Schrödinger's equation while making assumptions. The electronic level nature of QM methods makes them superior in accuracy, thus, they closely resemble the experimental observation. However, solving Schrödinger's equation is quite computationally expensive even for systems with only a limited number of atoms, thus making it unrealistic to be performed on large systems.

In the light of the challenge with the high computational demand, the best remedy is to use classical mechanics to simulate these larger systems, given that the electronic structure properties are sufficiently captured through a forcefield. Even though they are computationally less expensive, the accuracy suffers from not being able to access the systems at the electronic level, unlike QM. This hurdle in using molecular mechanics (MM) is overcome by using a molecular forcefield that helps this simulation to mimic the electronic level properties.[5] Therefore, the accuracy of properties extracted through an MM based simulation vastly relies on the forcefield used. A forcefield cannot be universally applied to any system with the expectation of benchmark performance. This is due to the fact that the forcefield parameters are built upon calculations and experiments performed on dissimilar systems with specific intentions. Simply the forcefield parameters for atoms C, N, and O are not the same when comparing a forcefield made to simulate biological systems and another forcefield made for inorganic polymers. The credibility of such simulations heavily depends on the forcefield parameters used to represent the electronic structure governing effects since MD simulations are not based on quantum mechanics. One such issue that profoundly affects the accuracy is the use of non-polarizable forcefields, which is a common choice in performing MM based simulations due to their low computational cost.

Many virtual screening methods including docking, mostly if not entirely depend on MM that suffer from all aforementioned limitations. In many efforts that use docking for virtual screening, do not involve validating and optimizing forcefield parameters to meet the need of the specific system (enzyme/protein and ligands) we are interested in. Different docking algorithms make use of different scoring functions to grade the small molecules in terms of potency to interact with a defined site on the enzyme. Therefore, the results produced are often subjected to the scoring function that is used, and the order of potency could substantially change between scoring functions. One way to overcome this is again to validate the scoring function with experimental results of similar systems. However, it is well comprehended that validation and finetuning of forcefields and scoring functions are not always feasible or realistic when dealing

with extensive libraries of compounds or experimental access is limited. Thus, one should always be cautious about making statements based on any *in-silico* method used in screening.

Moreover, many virtual screening methods including docking does not take into account the flexibility of the macromolecule in full. However, it should be noted that addressing the side-chain flexibility at least on an explicitly specified binding site is of utmost importance. Similarly, docking is mostly static, in its nature where stability and possible alternative configurations in a time-dependent manner are often neglected. The common practice of overcoming this is accomplished through coupling docking with molecular dynamics simulations. In the account of the first two limitations as mentioned at the beginning, another important fact when using virtual screening is that not every system can be modeled using a method such as docking, especially where molecular interactions alone cannot capture conformational changes that take place after binding.

The most significant advantage of virtual screening is that it allows researchers to screen a massive number of compounds by utilizing molecular libraries, which otherwise would be extremely costly, impractical, or even be rather impossible to be done using *in-vitro* or *in-vivo* methods even with access to sophisticated high throughput automated screening facilities. Nevertheless, it is not rational to assume that the entire chemical search space is screened, although the virtual screening was done with all available commercial and non-commercial databases. It is always essential to be aware of the fact that the search space will always be limited unless a machine learning approach is used.

All these bring us to the ultimate question of "Can virtual be really useful in drug discovery, and are there any other methods without these limitations?" The simple answer in my perspective for the first part is "Yes". Virtual screening is extremely useful in this hunt for new drugs. Molecular docking itself has been able to discover new drugs, predict binding modes, understand binding mechanisms and study the effect of mutations in diseases such as cancer, influenza, Zika, Malaria, and HIV etc.[6] However like any other tool used in science it has to be utilized with intuition and understanding of its limitations to yield meaningful results. The answer to the second question is, "Yes, but not entirely". In terms of accuracy, *ab-inito* molecular dynamics simulations (AIMD), does provide a better solution. Specifically, the capability to mimic intermolecular interactions such as π-π stacking and π-cation interactions presents with a high level of confidence, require AIMD or use of polarizable forcefields that are computationally expensive. Many academic and research institutions worldwide, however, have limited computational resources. Thus, a method such as AIMD is not computationally affordable in general. Being able to perform a simulation of the same caliber with less computationally expensive methods such as charge renormalization provides a great opportunity to many researchers.[5] The next solution would be through machine learning (ML). This is nothing new to the field; the theoretical basis of ML is the same as quantitative structure-activity relationships (QSAR) that have been used for virtual screening well before docking has been introduced. The impact of "Big Data" analysis in the modern world, almost in every aspect of human life has become substantial. Global enterprises such as healthcare, education, marketing, business, finance, and economics are heavily dependent on the insights they obtain through ML with big data.[7] Lately, many virtual screening efforts are being directed through ML. Yet again, ML methods are not perfect, but that would be a topic for a separate article. However, in general, the way to go would be through experimental validation of any virtual screening method of interest before implementing a specific project.

In conclusion, virtual screening has served the field of drug discovery for decades and will continue to do so in the future, although methods of virtual screening will change and evolve over time. It is highly doubtful that all these limitations would cease to exist in the near future. There will always be new hurdles to overcome, but researchers will keep on overcoming them as science progresses. After all, is not overcoming hurdles is what science is all about?

References

1. DiMasi, J. A.; Grabowski, H. G.; Hansen, R. W. Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs. *J. Health Econ.* **2016**.
2. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to

Macromolecule-Ligand Interactions. *J. Mol. Biol.* **1982**, *161* (2), 269–288.

3. Bitencourt-Ferreira, G.; de Azevedo, W. F. Molegro Virtual Docker for Docking. In *Methods in Molecular Biology*; 2019.
4. Seddon, G.; Lounnas, V.; McGuire, R.; Van Den Bergh, T.; Bywater, R. P. P.; Oliveira, L.; Vriend, G. Drug Design for Ever, from Hype to Hope. *J. Comput. Aided. Mol. Des.* **2012**, *26* (1), 137–150.
5. Li, Z.; Robertson, L. A.; Shkrob, I. A.; Smith, K. C.; Cheng, L.; Zhang, L.; Moore, J. S.; Z, Y. Realistic Ion Dynamics through Charge Renormalization in Nonaqueous Electrolytes. *J. Phys. Chem. B* **2020**.
6. Phillips, M. A.; Stewart, M. A.; Woodling, D. L.; Xie, Z.-R. Has Molecular Docking Ever Brought Us a Medicine? In *Molecular Docking*; 2018.
7. Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **2018**, *4* (2), 268–276.