International Journal of Pharmaceutical Research and Applications
Volume 8, Issue 4 July-Aug 2023, pp: 1410-1413 www.ijprajournal.com ISSN: 2249-7781

To review on the drug discovery

Rohit Hanumant Mane, Shubham Dada Walke, Rakhi Sanjay Mane

Submitted: 25-07-2023 Accepted: 05-08-2023

ABSTRACT

Drug discovery aims at finding new compounds with specific chemical properties for the treatment of diseases. In the last years, the approach used in this search presents an important component in computer science with the skyrocketing of machine learning techniques due to its democratization. With the objectives set by the Precision Medicine initiative and the new challenges generated, it is necessary to establish robust, standard and reproducible computational methodologies to achieve the objectives set. Currently, predictive models based on Machine Learning have gained great importance in the step prior to preclinical studies. This stage manages to drastically reduce costs and research times in the discovery of new drugs. This review article focuses on how these new methodologies are being used in recent years of research. Analyzing the state of the art in this field will give us an idea of where cheminformatics will be developed in the short term, the limitations it presents and the positive results it has achieved.

Keywords: Machine LearningDrug DiscoveryCheminformaticsQSARMolecular DescriptorsDeep Learning

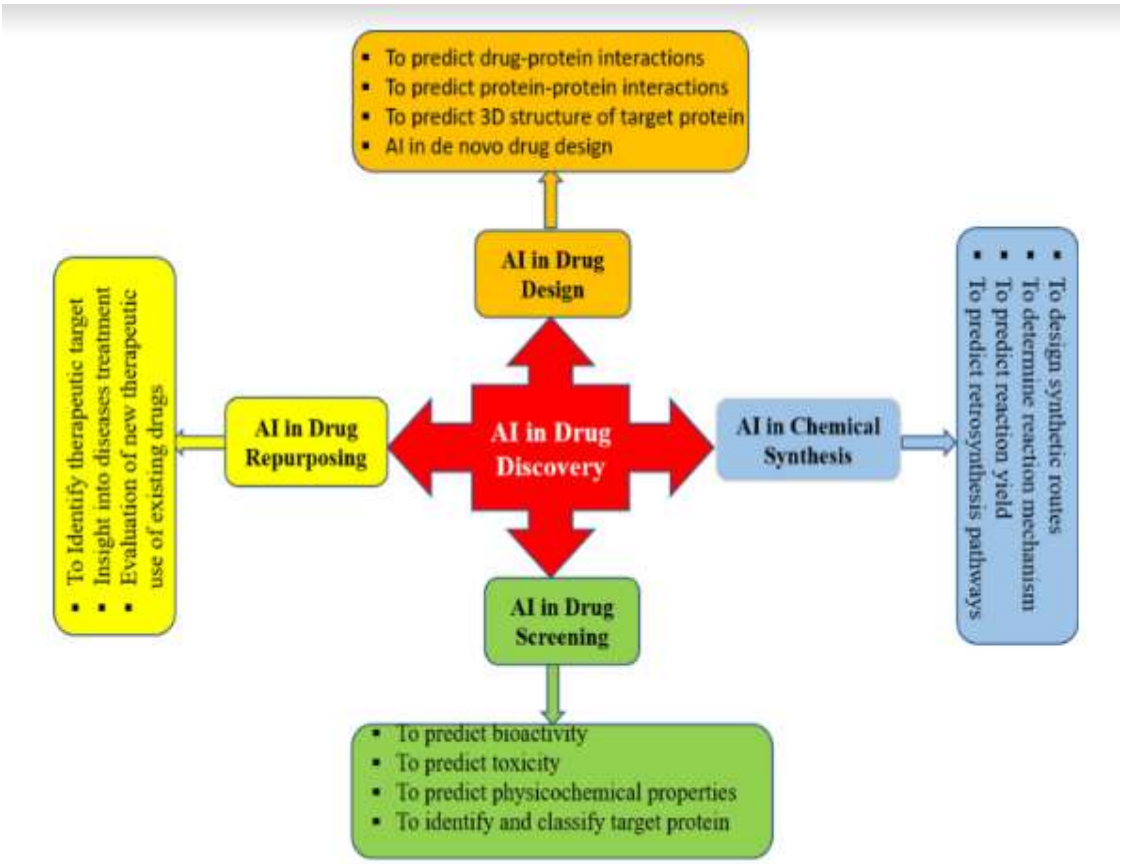
I. INTRODUCTION

According to the Precision Medicine Initiative, precision medicine is "an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment and lifestyle for each person" [1]. This new approach allows physicians and researchers to increase accuracy in predicting disease treatment and prevention strategies that will work for particular groups of people. This approach contrasts with the "one-size-fits-all" approach, more widely used until relatively recently, in which the strategies mentioned above are developed with the average person in mind, regardless of differences between individuals.

The opportunity for the creation of new treatments offered by precision medicine generates at the same time great difficulties in the development of new methodologies. For this reason, in recent years a large amount of biomedical data has been generated, coming from very diverse sources: from small individual laboratories to large international initiatives. These data, known mostly as omic data (genomic, proteomic, metabolomic, pharmacogenomic, etc.), are an inexhaustible source of information for the scientific community, which allows stratifying patients, obtaining specific diagnoses or generating new treatments.

Diagnostic tests are frequently performed in some disease areas, as they allow immediate identification of the most effective treatment for a specific patient through a specific molecular analysis. With this, the practice of trial and error medicine, which is often frustrating and considerably more expensive, is often avoided. In addition, drugs created from these molecular characteristics usually improve treatment results and reduce side effects. One of the most common examples can be found in the treatment of patients with breast cancer. A significant percentage of patients with this type of tumor are characterized by overexpression of human epidermal growth factor receptor 2 For these patients, treatment with the drug trastuzumab (Herceptin) in addition to chemotherapy treatment can reduce the risk of recurrence to more than 50%

International Journal of Pharmaceutical Research and Applications
Volume 8, Issue 4 July-Aug 2023, pp: 1410-1413 www.ijprajournal.com ISSN: 2249-7781



The importance of input data in Machine Learning predictions

A critical step in the training of the model depends on the representation of the molecules by descriptors that are capable of capturing their properties and structural characteristics. Hundreds of molecular descriptors have been reported in the literature ranging from simple properties of the molecules to elaborate three-dimensional and complex molecular fingerprint formulations, stored in vectors of hundreds and/or thousands of elements.

Quantitative structure–activity relationship

Under the premises "the structure of a molecule defines its biological activity" and "structurally similar molecules have a similar biological activity", the models of quantitative relationship between structure and activity (QSAR), which numerically relate the chemical structures of the molecules with their biological activity, allow, through mathematical systems, to predict the physicochemical and biological fate properties that a new compound will have from the knowledge of its chemical structure and from existing experimental studies.

QSAR models integrate computer and statistical techniques in order to make a theoretical prediction of biological activity that allows the theoretical design of possible future new drugs, avoiding the trial and error process of organic synthesis. As it is a science that exists only in a virtual environment, it allows dispensing with certain resources such as equipment, instruments, materials and laboratory staff. With a focus on the relationships between chemical structure and biological activity, the design of candidates for new drugs is much cheaper and faster. Modeling studies such as QSAR is one of the most effective methods to perform compound prediction when there is a lack of adequate experimental data and facilities [20].

To carry out a QSAR study, three types of information are needed [21]:

1.) Molecular structure of different compounds with a common mechanism of action

2.) Biological activity data of each of the ligands included in the study.

International Journal of Pharmaceutical Research and Applications
Volume 8, Issue 4 July-Aug 2023, pp: 1410-1413 www.ijprajournal.com ISSN: 2249-7781

3.) Physicochemical properties, which are described from a set of numerical variables, obtained from the molecular structure virtually generated by To pj techniques.

Molecular descriptors

Molecular descriptors (MD) play a key role in many areas of research. They can be defined as numerical representations of the molecule that quantitatively describe its physicochemical information. But not all the information contained in a molecule, but only a part, can be extracted through experimental measurements. In recent decades there has been an increasing focus on how to capture and convert, in a theoretical way, the information encoded in the molecular structure into one or more numbers that are used to establish quantitative relationships between structures and properties, biological activities and other experimental properties. In this way, MDs have become a very useful tool to carry out the search for similarities in molecular repositories, since they can find molecules with similar physicochemical properties according to their similarity to the values of the calculated descriptors.

The molecular descriptors can be divided into two main categories. Experimental measurements, such as log P, molar refractivity, dipole moment, polarisability and, in general, additive physical–chemical properties and theoretical molecular descriptors, which are derived from a symbolic representation of the molecule and can be further classified according to the different types of molecular representation. Theoretical ones, in turn, are classified into:

1.) Constitutional: reflect general properties of molecular nature
2.)Topological: its calculation is done through graph theory
3.) Geometric: are derived from empirical schemes and encode the ability of the molecule to participate in different types of interactions.
4.) Electronics: refer to the electronic properties
5.) Physicochemicals: define the behaviour of the molecule in the face of external reactions

Biological problems asses by Machine Learning in drug discovery

A drug can be defined as a molecule that interacts with a functional entity in the organism, called a therapeutic target or molecular target, modifying its behaviour in some way. Known drugs act on known targets, but the discovery of new ones that can modify the course of a disease or improve the effectiveness of existing treatments is one of the main objectives of research in the field of chemistry and biology.

The development of a new drug can take up to 12 years and it is estimated that its average cost, until it reaches the market, is approximately one billion euros. The time and costs involved are largely associated with the large number of molecules that fail at one or more stages of their development, as it is estimated that only 1 in 5,000 drugs finally reach the market.

The previous statistics show that the discovery and development of new drugs is a very complex and expensive process. This process has been carried out for a long time using exclusively experimental methods. The technological advances of the last few decades have promoted the birth of the term in silico, a term that is now common in biology laboratories, and which designates a type of experiment that is not done directly on a living organism (these are called in vivo experiments) or in a test tube or other artificial environment outside the organism (experiments called in vitro), but is carried out virtually through computer simulations of biological processes.

Administration, distribution, metabolism, elimination and toxicity

The concept of drug similarity, established from the analysis of the physicochemical properties and structural characteristics of existing or candidate compounds, has been widely used to filter out compounds with undesirable properties in terms of administration, distribution, metabolism, elimination and toxicity

International Journal of Pharmaceutical Research and Applications
Volume 8, Issue 4 July-Aug 2023, pp: 1410-1413 www.ijprajournal.com ISSN: 2249-7781

(ADMET) [65]. The study of the ADME phases that a drug undergoes after being administered to an individual is another of the fundamental tasks in the development of new compounds [66]. Alteration in a patient of any of these stages (for example, excretion problems due to some type of renal failure, increased volume of distribution in obese people, absorption problems due to gastrointestinal pathology or problems in the metabolism of the drug due to deterioration of liver function) may influence the final plasma concentration of the drug modifying the expected response of the organism, thus requiring a decrease or increase in the dose of the drug in each case.Therefore, it is essential in the early stages of research to estimate the behaviour of the pharmacokinetic properties of a compound, and new tools have been developed to improve and speed up this phase of development. This is the example of Chemi-Net [67], for the prediction of ADME properties, which increases the accuracy over another tool with the same purpose already in existence.

The company Bayer Pharma has implemented a platform for absorption, distribution, metabolism and excretion ADMET in silico with the aim of generating models for a wide variety of useful pharmacokinetic and physicochemical properties in the early stages of drug discovery, but these tools are accessible to all scientists within the company

II. CONCLUSIONS

The latest advances in the design of new algorithms in the field of Artificial Intelligence have offered the opportunity to solve problems in different disciplines. In cheminformatics, and more specifically in drug discovery, the use of these models has greatly benefited the pharmaceutical industry. Previously, the only tool was the use of descriptors generated from the structure of small molecules or peptides. More recently, artificial neuron networks were adapted to model directly the molecules represented by graphs. Today, molecular descriptors are still widely used in the industry, but the rise of graph-based models is obtaining results that surpass the more conventional models in certain contexts.

REFERENCES

[1]. F.S. Collins, H. VarmusA new initiative on precision medicineNew England J Med, 372 (9) (2015), pp. 793-795View in ScopusGoogle Scholar

[2]. C. Curtis, S.P. Shah, S.-F. Chin, G. Turashvili, O.M. Rueda, M.J. Dunning, D. Speed, A.G. Lynch, S. Samarajiwa, Y. Yuan, et al.The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroupsNature, 486 (7403) (2012), pp. 346-352View article CrossRefView in ScopusGoogle Scholar

[3]. E.H. Romond, E.A. Perez, J. Bryant, V.J. Suman, C.E. Geyer Jr, N.E. Davidson, E. Tan-Chiu, S. Martino, S. Paik, P.A. Kaufman, et al.Trastuzumab plus adjuvant chemotherapy for operable her2-positive breast cancerN Engl J Med, 353 (16) (2005), pp. 1673-1684View in ScopusGoogle Scholar

[4]. J.L. Blanco, A.B. Porto-Pazos, A. Pazos, C. Fernandez-LozanoPrediction of high anti-angiogenic activity peptides in silico using a generalized linear model and feature selectionSci Rep, 8 (1) (2018), pp. 1-11View in ScopusGoogle Scholar

[5]. C.R. Munteanu, E. Fernández-Blanco, J.A. Seoane, P. Izquierdo-Novo, J. Angel Rodriguez-Fernandez, J. Maria Prieto-Gonzalez, J.R. Rabunal, A. PazoslDrug discovery and design for complex diseases through qsar computational methodsCurrent Pharmaceutical Des, 16 (24) (2010), pp. 2640-2655