Big Data to Expand the Antimicrobial Therapeutic Arsenal:
De Novo Discovery and Drug Repurposing

B

Antonio Tarín-Pelló
https://orcid.org/0000-0003-3781-844X
Instituto de Ciencias Biomédicas, Mexico

Beatriz Suay-García
Universidad CEU Cardenal Herrera, Spain

Antonio Falcó
Universidad CEU Cardenal Herrera, Spain

María Teresa Pérez-Gracia
https://orcid.org/0000-0001-5083-9267
Universidad CEU Cardenal Herrera, Spain

INTRODUCTION

The current Big Data era implies the presence of vast and intricate amounts of data that cannot be analyzed with traditional tools and analysis techniques (Gandomi A, *et al.*, 2015). Between these data, we can find those related to biomedicine, which can be used for synthesis and compound selection in the discovery and design of drugs (Kim S, *et al.,* 2021). Some examples of these databases include ChemSpider (more than 26 million registered molecules), ChEMBL (data on bioactive molecules and their pharmacological properties), ZINC (commercial chemical compounds especially prepared for virtual screening), BindingDB (intermolecular binding data such as protein-ligand interactions) and Protein Data Bank (PDB) (protein and nucleic acid tridimensional structures).

The high throughput screening (HTS) has been one of the most important processes in the discovery of biologically active molecules (Pinzi L, *et al.,* 2019). Improved computational techniques used for HTS, such as artificial intelligence (AI) and machine learning (ML), also contribute to improved data management by accelerating and refining the process (Gupta R, *et al.,* 2021). Another relevant process in the search for new drugs, and that AI and ML techniques (along with databases such as PDB) have helped to optimize, is the determination of target molecules. Predicting and identifying the drug-target interaction (DTI) is a crucial step in the discovery and design of new drugs, as it reduces the costs of experimental validation (Thafar MA, *et al.,* 2021). The new drugs are designed based on the analysis and observation of the binding of ligands to the three-dimensional structure of molecules such as amino acids and their sequences, proteins, DNA or RNA (Robichaux JP, *et al.,* 2021). This translates into faster discovery of effective drugs, with a higher success rate and a reduction in computational costs related to traditional methods, such as molecular docking or virtual screening (VS) (Gupta R, *et al.*, 2021). An example of this is found in recent studies on computational analysis for the identification of effective

DOI: 10.4018/978-1-6684-7366-5.ch027

drugs against the SARS-CoV-2 coronavirus, where active molecules against this pathogen were identified (Pérez-Moraga R, *et al.,* 2021).

Another interesting factor in the design and discovery of drugs is the development of relationships between chemical structures and their physicochemical properties with biological activities. Mathematical prediction models, which characterize the structural, physical, chemical, and biological properties of molecules, have become an essential tool for this issue, becoming the key to the success of ML models targeting both drug design and repurposing (Suay-Garcia B, *et al.,* 2020a). Drug repositioning is the generation of new clinical opportunities for molecules already known and/or approved, providing a new therapeutic indication different from the usual one (Suay-Garcia B, *et al.,* 2019). The repositioning of drugs that have undergone extensive toxicological and pharmacological analysis is an effective method to reduce the time, cost, and risks of de novo synthesis, moving directly to preclinical testing and clinical trials (Liu Y, *et al.,* 2021) (Figure 1). This method has proven useful for identifying a new clinical use against different diseases in molecules already known or commercialized (Suay-Garcia B, *et al.,* 2020b). In addition, new molecules have been detected through in silico homology studies that could be reused as lead compounds from which to obtain new molecules with greater efficiency (Troeman DPR, *et al.,* 2019). An important feature of this method is that it can re-evaluate molecules considered as failed in previous studies, adding value to a lost investment by providing new indications for these drugs (Natalie KB, *et al.,* 2021). In addition, trials that can be conducted by repositioning drugs could reveal new therapeutic targets and improve knowledge of known therapies.

Figure 1. Phase and time difference between drug development from de novo synthesis and the drug repurposing method. Created with BioRender.com