Chapter 12
Computational Modeling in Virus Infections and Virtual Screening, Docking, and Molecular Dynamics in Drug Design

Rachel Siqueira de Queiroz Simões, Mariana Simões Ferreira, Nathalia Dumas de Paula, Thamires Rocco Machado, and Pedro Geraldo Pascutti

Abstract  Computer modeling is an area of broad multidisciplinary knowledge that includes the study of various biological systems. This chapter will describe the molecular aspects of viral infections and molecular modeling techniques applied to drug discovery with examples of applications in protein activity inhibition in several pathologies. The first part will cover topics of computational chemistry methods, DNA technologies, structural modeling of virus proteins, molecular biology, viral vectors, virus-like particles, and pharmaceutical bioprocess with application in some specific viruses such as papillomavirus, hepatitis B virus, hepatitis C virus, Coronavirus, and Zika Virus. The second part will deal with methods in Virtual Screening for the drug design based on ligands and on the structure of target macromolecules. Molecular docking in drug design, its search algorithms, and scoring functions will be covered in the third part. Finally, a protocol of the Molecular Dynamics technique for studies of protein-ligand complexes and analysis of free energy of binding will be exposed in the last part.
The emergence of computers in the twentieth century allowed new advances in the field of life sciences because it allowed scientists to overcome limitations when exploring molecules in their experimental investigations. However, the term that represents this approach, involving simulations and the manipulation of a large amount of data, computational biology, was only used after the human DNA sequencing project in the 1990s. Since then, several approaches have used computational technology to solve problems in biology, which include, among others, the treatment of nucleic acid and protein sequences, description of biological systems networks, and modeling and simulation of the interactions of macromolecules and their ligands [1].

Advancements in computational technologies have also enabled improvements in the computational chemistry field of knowledge, such as in the development of new drugs. This way, computer software have become present in the areas of medicine, pharmacy, biology, chemistry, among others related to the development of biological processes or systems, investigating more complex processes at atomic-molecular scale each day at a much faster pace [2].

The investigation of molecular interactions is of great importance for biological understanding and allows observing and studying various processes involved in infectious diseases such as enzymatic action, binding of DNA repair proteins, and the formation of viral capsids. Moreover, molecular interactions also play a central role in the design of new drugs in the industry and a large part of the efforts are currently carried out using computational biology techniques, known as computer-aided drug design approaches. Thus, research has been developed through computer simulations in the investigation of molecular structures, protein-protein and protein-ligand complexes, biological membranes, and protein folding, with some of them being addressed throughout this chapter.

12.1 Computational Modeling in Viral Infections

In general, numerous conformations in protein folding have been documented by simulated annealing techniques through the application of Molecular Dynamics and Statistical Mechanics methods. These methods can simulate, for example, conformational changes in the antiviral proteins present in host cells, called viral restriction factors such as IFITM, SAMHD1, APOBEC3G and BST-2, when activated by the innate immune system during a viral infection and hinder the replication of viruses such as Influenza, Western Nile, Zika, HIV, and Hepatitis C virus [3].

Innate immune response, also known as natural or nonspecific response, acts mainly during viral infections in dendric cells (CD) and Natural Killer cells (NK), which have the ability to destroy virus-infected cells given its cytotoxic activity. They
contribute to an antiviral defense action by secreting cytokines such as IFN capable of interfering with viral replication like IFN-α and IFN-β and by killing inhibitory receptors. In the acquired immune response, T helper lymphocytes (Th) and T cytotoxic lymphocytes (Tc) generate a specific cellular response while the antibodies produced by B lymphocytes generate a humoral response. The three-dimensional representation of protein structures by conformational analysis is obtained by Molecular Modeling, a branch of Computational Chemistry using mathematical models. Several computer programs are used, as described in this chapter, which allows the construction and identification of small molecules linked to proteins and nucleotides by analyzing protein databases such as Protein Data Bank (PDB) (Table 12.2). For example, the Osiris Property Explorer program is capable of signaling the molecule’s behavior through two measures: Drug likeness and Drug score [2].

Computational Chemistry strategies are divided into the methods of Molecular Mechanics (MM) and Quantum Mechanics (QM). The MM approaches are based on Newtonian mechanics, in which the molecule is interpreted as a set of spheres connected by springs and can be used in calculations of very large molecules. In this method, the parameters associated with the atoms of the molecule are reasonably constant between different structures, as long as the hybridization of the atoms is maintained [2].

The QM methods, on the other hand, consider the movement of electrons, assuming fixed the nucleus, and represent molecular orbitals as a linear combination of atomic orbitals. They can be divided into strategies semi-empirical or ab initio. The semi-empirical approaches are developed based on empirical parameters or parameters already calculated from Schrodinger’s equation, whereas the ab initio ones are based on the laws of quantum mechanics applying approximations for directly solving Schrodinger’s equation that enable solutions for a wide range of molecules [2].

In Computational Chemistry, nevertheless, studies of new compounds have been applied regarding many infectious and viral diseases. An example is to analyze the activity of DNA polymerase of cytomegalovirus (CMV) against the physical-chemical properties of molecules in order to find potential candidates such as anti-CMV agents, in strategies such as Virtual Screening (VS), later described in this chapter [2]. It is an important investigation since in recent years there has been an increase in the number of opportunistic infections to human CMV, occurring mainly in transplant patients and in immunocompromised individuals.

Another example applies the molecular modeling of the interaction between human Respiratory Syncytial Virus (RSV) Fusion Protein (hRSV F protein) and Viral Action Inhibitors. RSV is responsible for outbreaks causing inflammatory disease in the lower respiratory tract, pneumonia in children, and cardiopulmonary disease in adults and immunodeficient patients. Currently, through the Molecular Docking technique, explained in detail further ahead in this chapter, possible sites of interaction of hRSV F protein with flavonoids have been proposed [2].
12.1.1 Viral Vectors

Experimental approaches in gene therapy commonly use viruses as vectors for inserting the genetic material into the cell. It applies the knowledge on the complex interaction between a virus and a host cell, known to include several steps. They include interactions with cell membrane receptors (adsorption), viral entry into the cell (penetration), partial or total denudation of the viral capsid, and release of viral nucleic acid for replication inside the cell, as well as early and late proteins synthesis, assembly and release by exocytosis, viral budding, cell lysis or even syncytia formation. Each step will be exemplified throughout the chapter and viral vector designs and applications will also be addressed as it progresses in Biotechnology. Currently used vectors in human gene therapy suffer from a number of limitations with respect to safety and reproducibility through the use of animal models, toxicology, and biodistribution studies. Therefore, other strategies such as nonviral and hybrid vectors are being explored trying to overcome these issues [4].

When it comes to the introduction of genes into mammalian cells, methodologies require the use of chemical, physical, and biological tools. Briefly, chemical methods may include DNA-calcium phosphate, DNA-DEAE dextran, DNA-lipide (liposomes not integrated into the genome of the host cell), DNA-protein and HCAs (artificial chromosomes). Even so, there are physical methods with high transfection rates such as direct microinjection, electroporation, high-pressure plasmid injection, and DNA ballistic injection [4].

Viral vectors for gene therapy can be enveloped as retrovirus, lentivirus and baculovirus, and non-enveloped, like adenovirus. The key point of human gene therapy clinical trials is the production of a therapeutic vector, which can be viruses of the *Retroviridae* family, mainly gammaretroviruses and lentiviruses, such as the Moloney-murine leukemia virus (Mo-MLV) and the chimeric Moloney-Human lentiviral (prototype HIV-1). Moreover, adenoviruses, adeno-associated virus (AAV), chimeric-AAV, and vaccinia virus are other types of biological vectors possible. There are a number of additional viral vectors based on Epstein–Barr virus, herpes simplex virus type 1 (HSV-1), simian virus 40, hepatitis virus, and papillomavirus, which are still being studied for application in clinical gene therapy [4].

12.1.2 Virus-like Particles

Virus-like particles (VLPs), structures that resemble viruses without the infectious capability of them since there is no genetic material present, show differences according to the presence of envelopment or not. In the case of non-enveloped VLPs, examples are the single protein hepatitis B core antigen and the two proteins of papillomavirus L1 and L2. As for the enveloped VLPs, there is Influenza virus with proteins hemagglutinin and neuraminidase. Some viral particles have the ability to self-assemble into VLPs regardless of the viral genome. HBV subviral particles
(HBsAg) have been used as a vector for the presentation of several antigen epitopes. The epitopes presented have included poliovirus capsid proteins, HCV envelope proteins, and human immunodeficiency virus (retrovirus) envelope proteins, among others [5].

VLPs can stimulate both innate and adaptive immune responses. Therefore, the particulate nature of VLPs generally induces an immune response that is more efficient than that generated by soluble proteins, and can trigger both a humoral and cellular response, without the need for an adjuvant. In addition, they are produced in large quantities, can be easily purified, and are not infectious. Currently, VLPs are a powerful tool in the presentation of immunogenic epitopes and the generation of recombinant VLPs, carrying relevant antigens, opening the way for the development of bivalent vaccines [6].

The incorporation of heterologous sequences of several viral structures into VLP particles has caused the development of chimeric VLPs. The use of this system as a platform for expression and presentation of exogenous epitopes favors the optimization of the immunogenicity of these peptide sequences. Numerous characteristics of VLPs allow the use of these particles as carriers of epitopes, the main one being that these particles are not infectious, since they do not carry the genetic material of the infectious agent, and still present the epitope expressed on its surface in a multi-sequential form [5, 7].

Several recombinant protein-based virus-like protein vaccines have been used as Human papillomavirus (HPV), Hepatitis B virus (HBV), and Hepatitis E vaccine. Others under development are the Influenza vaccine candidates, Ebola and Marburg virus vaccine, Hepatitis C virus, and Human Immunodeficiency Virus (HIV) [5].

### 12.1.3 Pharmaceutical Bioprocess

Clustering algorithms and multivariate analysis are statistical data mining, classified in nonparametric methods. In a large scale, Bayesian networks and stoichiometric modeling for metabolic pathway analysis are also applied in Systems Biology. In addition, differential equations and several estimation algorithms are parametric modeling methods. So, pharmaceutical products are very different in size and complexity as small molecules, antibodies, viruses and viral vectors, and cells [8].

Molecular associations as gene expressions, analysis by DNA microarrays, PCR techniques, and kinetic experiments on molecular states are some examples of experimental methods. Modeling requires knowledge of mathematics, physics, biology, and information technology to develop processes for manufacturing therapeutic proteins in the biopharmaceutical field. It can be used, for example, to establish a cell culture process with higher productivity and lower costs [8].
12.1.4 Papillomavirus

The advancement of stem cell research provides tools to model diseases, test new drugs, and develop effective therapies for use in regenerative medicine and cell therapy. The mapping of the stem cells can generate biomedical implications with prospects for production in vitro of induced pluripotent stem cells (iPS). These pluripotent stem cells are reprogrammed from somatic cells using retroviral vectors containing many genes [8].

In the near future, the keratinocyte reprogramming will be possible in tissue engineering laboratories as creating new models for the study of tumors associated with HPV as anogenital cancers and oropharyngeal carcinoma. So, the expression of genes and the role of proteins involved in DNA damage repair pathways in cell lines as primary human keratinocytes (PHK), HPV-positive (SiHa—HPV-16 and HeLa—HPV-18), and HPV-negative (C33A) human cervical carcinoma cell lines, as also in immortalized keratinocyte cell lines (HaCaT, not tumor control), have been described as possible prognostic markers of cervical cancer [9].

Some studies have investigated the ability of the cytokine to inhibit the proliferation in vitro of normal and HPV infected keratinocytes, as well as the expression of E6 and E7 oncogenes. L1 capsids can trigger innate immune responses in some types of antigen-presenting cells, including dendritic cells. Thus, the dendritic cell control tissue immunity is a field with important implications in both basic and clinical immunology. So, the cytokine inhibition ability has been demonstrated [9].

Research has been carried out on the Molecular Modeling and Dynamics of the E6 oncoprotein of human papillomavirus type 18. This molecule is an essential protein for the formation of cervical cancer, since it is responsible for preventing cell apoptosis by binding in the tumor suppressor p53, promoting cell immortalization, and inducing carcinogenesis. In order to refine the generated E6 model, the N terminal version was modeled ab initio and part of the E6 model was done by homology based on the 4 GIZ protein [10].

New biotechnologies in Molecular Biology have been applied in papillomavirus research and are still being investigated for the application in gene therapy. For analysis of the transcriptome of genomes with high resolving power, RNA sequencing (RNA-seq) monitors the cell expression and allows quantifying the transcript levels using the Illumina platform for sequencing, which is a useful tool for gene mapping and identification of transcribed regions. Furthermore, examples of new techniques include the multiply primed rolling-circle amplification (RCA) method, which has been used to identify novel viruses, and the DNA sequencing strategy, based on real-time pyrophosphate.

The constant development of new biotechnologies is followed by the description of new therapeutic targets, for example, regarding HPV vaccination. They include DNA-based vaccines, recombinant proteins, nanoparticles, synthetic peptides, viral and nonviral vectors, and expressed chimeric proteins self-assembled into virus like particles (VLP) from L1 major capsid proteins to finally produce the HPV vaccines [7, 11].
12.1.5 Hepatitis B Virus (HBV)

Two billion people are infected and more than 350 million people are chronic carriers of the hepatitis B virus worldwide, according to the World Health Organization (WHO). Moreover, there are approximately six hundred thousand cases of HBV-related deaths, with 4.5 million of new infections around the world each year. It is estimated that 25% of this new infection will result in the development of liver diseases, meaning HBV is a major causative agent of chronic liver disease, presenting high morbidity and mortality rates. Therefore, it is necessary to study such virus at epidemiological and molecular levels, through the scope of the genomic variability analysis of HBV. Currently, it brings the classification of the genome into eight different genotypes named A–H that diverge by less than 8% from all the nucleotide sequence. Lastly, this virus is classified as being part of the family Hepadnaviridae, divided into the two genera Orthohepadnavirus (include viruses responsible for infections in mammals) and Avihepadnavirus (comprises bird infective viruses) [12–15].

The genome of this virus is composed of a double-stranded DNA with 3,200 base pairs (bp) having a partially incomplete smaller strand with positive polarity and a larger strand with negative polarity and complementary to the viral produce during replication. Four open reading frames (ORFs) are present in this genome: Pre-S/S, Pre-C/C, P, and X. The Pre-S/S ORF corresponds to the HBV surface gene (HBsAg). The Pre-S1, Pre-S2, and S regions have three initiation codons in the same reading phase which, after being translated, give rise to L “Large” (400 aa), M “Middle” (281 aa), and S “Small” proteins (226 aa), respectively [13, 16]. The HBsAg protein is present in the non-glycosylated (p24) and glycosylated (p27) forms in the S region, a glycoprotein (gp32) or two glycoproteins (gp36) in the M region, and the non-glycosylated (p37) and glycosylated (p39) in the region L. HBsAg hybrid particles, also called chimeras, have been shown in different immunization experiments to be very efficient proteins in the presentation of viral epitopes [12, 13].

Complete particles (42 nm), spherical, infectious, and containing the HBV genome, as well as subviral particles (22 nm), spherical or filamentous, noninfectious and composed exclusively of HBsAg, are produced during HBV infection. Subviral particles are in large quantities during infection [16].

HBV S protein is a glycoprotein, which is anchored in the endoplasmic reticulum (ER) membrane probably through four transmembrane (TM) segments. The first segment (TM1), located at the amino-terminal end of the protein (amino acids 4–24), is followed by a cytosolic loop (amino acids 24–80) comprising epitopes for T cells, a TM2 segment (amino acids 81–100), and an antigenic loop (amino acids 101–164), located in the luminal portion of the ER, which includes the main epitope for B cells (the determinant a, amino acids 124–147). The topology of the carboxy-terminal domain (amino acids 165–226) is not precisely known. It has been suggested that this domain contains two transmembrane segments (TM3 and TM4) at positions 173–193 and 202–222, respectively, being separated by a second “loop” (amino acids 194–201) [12, 15, 16].
The HBV S protein has the property of self-organizing into VLPs and carries the information necessary for its secretion mediated by mammalian cells. Each HBsAg particle is composed of 100–150 subunits of the S protein. Due to the high immunogenic potential, hybrid HBsAg particles, also called chimeras, have been shown in different immunization experiments to be very efficient proteins in the presentation of viral epitopes and antigenic sequences of various infectious agents, such as *Clostridium tetani*, Herpes simplex glycoprotein D, poliovirus capsid protein antigens derived from the malaria parasite, dengue virus, HCV, and HIV, which have been successfully fused to HBsAg, maintaining the antigenicity and immunogenicity of their regions after immunization in experimental animals [17, 18].

HBsAg has an antigenic region called a determinant (residues 124–147), which has the ability to induce a protective immune response against HBV. Neutralizing antibodies are directed to this region. Mutations in the determinant justify the development of new vaccines against HBV, among other factors that influence the vaccine response. New technologies such as DNA vaccines and bifunctional vaccines (for immunization against two vaccine targets simultaneously) have been a reality in the study and development of new vaccines [17, 19].

The first HBV vaccines, produced between 1981 and 1982, were created from the HBV surface antigen, HBsAg, from the plasma of chronic carriers of the virus. In the mid-1980s, with the introduction of recombinant DNA technology, this antigen started to be expressed in yeasts transfected by the HBsAg gene. This new technology offered unlimited production potential, which allowed the HBV vaccine to become more widely used worldwide. The development of recombinant vaccines composed exclusively of HBsAg (Engerix-B, SmithKline and Recombivax, Merck & Co) was possible with the advance of genetic engineering. So, the HBV vaccine was the first recombinant vaccine licensed and produced from yeast expression [15, 19].

**12.1.6 Hepatitis C Virus (HCV)**

Hepatitis C virus (HCV) is a chronic infection that affects more than 2% of the global population and is a leading cause of hepatocellular carcinoma and end-stage liver diseases. The progression from the viral infection to hepatocarcinoma is a worldwide public health problem that grows each year, supported by the lack of a vaccine against it to date. This virus belongs to the order *Nidovirales*, family *Flaviviridae*, genus *Hepacivirus* [14], and it is classified into seven genetically distinct genotypes, which differ approximately 30% at the nucleotide level. It is an enveloped virus, strand of positive polarity RNA, containing approximately 9,400 nucleotides. HCV is the virus that most infects humans among the viruses of clinical importance, being one of those with the highest mutation rate [20, 21].

The envelope glycoproteins (E1 and E2) have highly variable domains in their sequences, a large N-terminal ectodomain, and a C-terminal transmembrane portion. At the N-terminal end of the E2 protein, there is a region characterized by a high degree of variability called the hypervariable region 1 (HVR1), important in the
neutralization of HCV. However, the high variability of this antigenic fragment plays a fundamental role in the viral escape mechanism of the host’s immune response and represents the greatest obstacle in the development of an HCV vaccine [22, 23]. Despite this high variability, there are some residues in the HVR1 region that are conserved. Studies have shown that a HCV clone defective for HVR1 proved to be infectious in assays performed on chimpanzees, but with a highly attenuated infectivity, supporting the functional role of this domain in the entry of the virus.

E1 and E2 have 6 and 11 glycosylation sites, respectively, many of which are extremely conserved among HCV strains. It has been shown that these regions can play an important role in the entry of the virus into the cell and because of this, envelope glycoproteins have been the subject of extensive studies for the development of antiviral molecules, which prevent the entry of the virus into the hepatocyte [18, 24, 25].

The development of a chimeric vaccine, a prototype of heterologous antigens based on HBV envelope proteins, including HCV antigen, based on immunogenicity and antigenicity tests obtained from the collection and analysis of sera from animals laboratory immunized with synthetic peptides was assayed [26]. Production of specific antibodies in rabbits against conserved and potentially immunogenic peptides of the HCV envelope glycoprotein E2 was investigated. Possible cross-reactivity between the peptides and pre-immune sera of five rabbits that were further immunized was tested by in house Elisa. The sera at high concentrations showed a high level of cross-reactivity with the exception of anti-HVR1 antibody, which shows beyond the specific reaction, a nonspecific reaction with BSA. Thus, all the designed peptides were able to generate anti-E2 specific antibodies in rabbits at relatively high titles indicating candidate vaccine against HCV [15, 18, 25, 26].

12.1.7 Coronavirus

In the context of emerging diseases, a useful tool to investigate evolutionary history and viral origin from animals to humans is the phylogenetic analysis of molecular sequences of the viral genome sharing a common ancestor, since it may show the evolutionary relationships of novel viruses. The current outbreak of a coronavirus associated acute respiratory disease called coronavirus disease 19 (COVID-19), for example, has been designated as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2 or 2019-ncov) based on phylogeny and in the viral taxonomy among clusters to the prototype human and bat viruses [27, 28].

From that, it was possible to identify the differences between SARS-CoV-2 and two other zoonotic coronaviruses, the SARS-CoV and the MERS-CoV, classified in the Coronaviridae family. As it occurs with other RNA viruses, high genetic variability leads to gene recombination in coronaviruses, because they have similar and nonidentical genome sequences generating variants of the same virus. So, enveloped positive-sense RNA viruses with a 27 kb genome have the region of the
most conserved replicative proteins encoded in the open reading frames 1a and 1b (ORF1a/1b) of the coronavirus genome [27–29].

In the envelope of SARS-CoV-2, there is a protein called spike (S), which has a receptor-binding domain (RBD) that binds to the cell surface and has the appearance of a crown when a viral particle is seen in great magnification in electron microscopy [27]. The three-dimensional structure of the spike protein of this new coronavirus has been recently obtained by computational modeling. A 3D map of the protein or blueprint was created using the cryoelectron microscopy technique. The discovery of this structure was only possible thanks to the recent complete sequencing of the genome of this coronavirus strain. The Pasteur Institute in France, responsible for monitoring the respiratory virus, sequenced the entire genome of the new coronavirus [28]. The coronavirus S glycoprotein is surface-exposed and mediates entry into host cells; for this reason, it is the main target of neutralizing antibodies (Abs) upon infection and the focus of therapeutic and vaccine design.

The pandemic COVID-19 has accelerated molecular studies using genomic tools with Next Generation Sequencing (NGS) for the characterization of viral samples and genetic engineering for the development of vaccines. In this regard, Israeli researchers are developing an oral vaccine based on one that targets infectious bronchitis caused by other coronavirus which affects birds. GSK announced a collaboration with the Chinese firm Clover Biopharmaceuticals to develop a COVID-19 subunit vaccine consisting of a trimerized SARS-CoV-2 S protein using their patented Trimer-Tag® technology. Other countries such as the USA, Brazil, England, and China are also developing possible candidates for vaccines against SARS-CoV-2 but with different strategies, using mRNA synthetic molecules and virus-like particles (VLP), for example.

With the global emergence of this new virus, several diagnostic tests such as real-time PCR and rapid kit have been developed for its early detection. Moreover, the implementation of stem cell therapy in the treatment of patients infected with COVID-19 capable of repairing the damage caused to the lungs, liver, and other organs has been successful, as well as the use of blood plasma from patients recovered from the infection. To date, Remdesivir, a drug that was also used in Ebola and Marburg virus treatment, has demonstrated antiviral activity against SARS-Cov-2. In addition, this drug has also been found to show antiviral activity against other coronaviruses including MERS and SARS. Other drugs are also being studied against the 2019-ncov, such as favilavir and chloroquine. More effective treatment has been achieved by association with antibiotic therapy [29]. Furthermore, new studies are constantly being carried out. An example of it is the action of the antiparasitic called Ivermectin, widely used in veterinary medicine, that was tested in cell culture, and showed activity on reducing the replication of the virus by decreasing the concentration of viral proteins present in the nucleus as well as other therapeutic targets investigated.
12.1.8 Zika Virus

The Zika virus (ZIKV) belongs to the family *Flaviviridae*, genus *Flavivirus*, which includes more than 70 viruses, as the Dengue virus (DENV), Yellow fever virus (YFV), West Nile virus (WNV), and Japanese encephalitis virus (JEV). In 1947, Zika fever was first diagnosed in a sentinel rhesus monkey in Africa, the Republic of Uganda, in the forest called Zika, which gave the name for the virus [4, 27, 30].

Regardless of the viral family, but because they have the RNA genome, both the coronavirus and the Zika has been detected by molecular biology using reverse transcription of polymerase chain reaction (RT-PCR) in several biological specimens, such as samples of amniotic fluid, blood, tissues (brain, liver, spleen, kidney, lung, and heart), saliva, urine, and semen [27]. It has already been confirmed on the relationship between Zika virus infection and the occurrence of microcephaly (brain malformation affecting babies at birth), including neurological disorders, the Guillain-Barré syndrome, causing muscle paralysis with special attention to respiratory muscles [27]. Another similarity is that both, to date, have vaccines in development and clinical trials in progress to find the least toxic antiviral drugs [30]. It has been tested that a DNA vaccine called pcTPANS1 against the dengue virus NS1 protein (N-linked glycoprotein conserved) induces the production of antibodies and activation of CD4+ lymphocytes acting in the infected cells tested in animal models as target studies for flavivirus infections [30].

Studies have performed computational modeling of the NS5 protein of the Zika virus complexed with the SAH cofactor. Viable alternative treatments have been investigated targeting proteins that are essential for viral replication. Nonstructural protein 5 (NS5) is an important pharmacological target, being the most conserved among flaviviruses because it contains a methyltransferase (MTase) domain in the N-terminal region and an RNA polymerase-dependent RNA domain (RdRp) in the C portion terminal. The inhibition of methyltransferase is lethal for virus replication, thus, it is considered an excellent target for drug design [31, 32]. Finally, with computational methods, such as Molecular Dynamics (MD) simulations, technique thoroughly exemplified in the last part of this chapter, there are studies that were able to map binding sites and search for inhibitors based on the interaction of NS5 in its complex with SAH [32].

12.2 Virtual Screening in Drug Design

Virtual Screening (VS) is an in silico technique that consists in applying algorithms to filter from a library of small compounds, the ones with the potential to bind a target molecule. The surging of this methodology represented a major increase in the capability of designing and developing new drugs. It has allowed researchers to find, among millions of molecules in a dataset, lead compounds for binding a biological
target in a very short amount of time, while saving costs, in comparison to experimental approaches such as High Throughput Screening (HTS), in which hundreds of compounds are experimentally tested. This is possible through the constant development and optimization of faster computers and robust algorithms that continuously diminish the time such analysis takes, while increasing the accuracy of the interactions predicted. The VS method can be combined with other techniques described in this chapter, such as MD and Docking and, more recently, with machine learning algorithms. With them, it is possible to predict the binding and check the profile of the protein-ligand interaction through time and molecular flexibility, representing the selection of a more assertive set of lead compound candidates. Currently, VS is the first step in a drug development process, narrowing down the investigation on effective binding molecules before continuing exploration through in vitro assays (Fig. 12.1) [33, 34].

Fig. 12.1  Virtual screening workflow. From different perspectives, LBDD or SBDD, it is possible to obtain compounds for in vitro tests
12.2.1 Computer-Aided Drug Design (CADD)

Structure-based drug design (SBDD) and ligand-based drug design (LBDD) are two types of computer-aided drug design (CADD) approaches. SBDD methods analyze three-dimensional structural information from the macromolecular target, usually from proteins or RNA, to identify important locations and interactions that are fundamental to their respective biological functions. LBDD methods concentrate on known ligands in order to establish a relationship between their physical-chemical properties and biological activity, known as the structure-activity relationship (SAR). This information can be used to optimize drugs or guide the design of new ones [35].

The virtual screening technique can be approached through these two general knowledge-driven strategies. They both fight the problem from different, but complementary, perspectives. While the LBDD takes a group of compounds and searches for similarity to a known ligand of the biological target, the SBDD explores a specific site or the whole 3D structure of the target searching for compounds that are able to bind it [35].

12.2.1.1 Ligand-Based Drug Design

In the lack of reliable biological target structure, or in the presence of ligands known to have activity against the target, a better strategy is the ligand-based approach. It consists in taking a known ligand, such as a natural substrate of an enzyme, and screening in a given dataset for other small molecules that share similarity at structural or physicochemical level, which must present lower binding energy [35].

In this regard, LBDD is related to the ligand structure-activity relationship (SAR) studies, for example, the chemical group position, binding affinity, logP, number of rotatable chemical bonds, and number of donors and acceptors of hydrogen bonds. A strategy for calculation uses quantitative analysis of structure-activity relationship (QSAR), in which the molecule descriptors can be quantified. It is also possible for the association with the exploration of the tridimensional structure of the compounds (3D-QSAR). Results when applying such strategies are more energy accurate by taking the geometry of the compounds in the complex into account. Beyond tridimensional descriptors, many strategies involve ligand exploration in many more variables, from multiple conformations to solvation profile. An example is 4D-QSAR, that, by exploring molecular mechanics (MM) or even quantum mechanics (QM), adds the representation of the multiple conformations of the same ligand as another variable. The selection of descriptors in either SAR methodology must be carefully done based on each investigation, excluding highly correlated ones, for avoiding overlapping data, that can be identified by multiple linear regression (MLR) analysis, or even principal component analysis (PCA) [35–37].

Another assessment in LBDD is proposing a pharmacophore model to find other effective ligands. The modeling requires analyzing the functional groups, from hydrogen bond donors and acceptors to aromatic or hydrophobic groups in a known
ligand or set of ligands superimposed, and proposing the minimum of electrostatic energy and spatial distribution of some groups as required for effective activity against the aimed target (Fig. 12.2). Closely related to the 3D-QSAR technique, investigating a pharmacophore model, is an extremely powerful way to screen for new active compounds as well as a way to optimize the known ones. It is important to address that from analysis of a certain complex, there is not a unique answer, which means many different pharmacophore models can be proposed, and will be valid as well. Superimposing of compounds structures to the pharmacophore model comprehends the next steps on this process, since it will be able to identify the ones that present the desirable structure. Software to be used in such strategy can be Discovery Studio Visualizer, LigandScout, and the visualizing platforms for analyzing the structures are in Table 12.1 [37].

![Fig. 12.2](image) Hydrogen bonds (on the left, dashed lines) and hydrophobic contacts between receptor and ligand (on the right, dashed line) in the enzyme catalytic site. Visualizing the interactions between the ligand and protein residues helps the researcher to propose pharmacophoric groups. Images obtained by Discovery Studio Visualizer

| Table 12.1 Platforms for structural visualization |
|-----------------------------------------------|
| Softwares | Web link |
| **Free** | |
| Avogadro | https://avogadro.cc/ |
| PyMOL | https://pymol.org/2/ |
| VMD | https://www.ks.uiuc.edu/Research/vmd/ |
| Mercury | https://www.ccdc.cam.ac.uk/Community/csd-community/freemercury/ |
| Discovery Studio Visualizer | https://www.3dsbiovia.com/products/collaborative-science/biovia-discovery-studio/ |
| **Commercial** | |
| Molecular Operating Environment (MOE) | https://www.chemcomp.com/Products.htm |
| OpenEye | https://www.eyesopen.com/ |
| Spartan Pro | https://www.wavefun.com/spartan |
12.2.1.2 Structure-Based Drug Design

In contrast to LBDD, structure-based drug design requires the knowledge of the macromolecular structure in order to predict the binding of a library of compounds. Once defined the biological target, information on its structure must be gathered at several different repositories (Table 12.2), the most important one being the Protein Data Bank (PDB), where macromolecular structures solved experimentally by X-ray crystallography diffraction, nuclear magnetic resonance (NMR), or cryoelectron microscopy (Cryo-EM) are found. In the absence of the structure solved by experimental techniques, the alternative is to get an already proposed model or to create one based on homology using software such as the robust Modeler (https://salilab.org/modeller/). Once the structure of the biological target is obtained, the definition of the binding sites to have the screening compounds docked there should be the next step. If well-established, the binding site may even be identified with structures in the presence of ligands on repositories (Fig. 12.3). But, when exploring a new biological target or when studying other possible allosteric sites in the structure, investigation can occur through software such as MDpocket [38], where it can discover transitional pockets and potential target sites, in combination. Once the binding site of the biological target is defined and clustered, the following step is the docking of the compounds in it [36, 39].

12.2.2 The Virtual Screening Process

After defining the starting strategy (either LBDD or SBDD) to undergo virtual screening, the next steps are as follows. The first and mandatory one is the preprocessing of the compounds database (Table 12.3). It comprises especially identifying and obtaining the tridimensional geometry of such structures, if it is not available in the database chosen. Several software (Schrodinger's LigPrep, OpenBabel) can be used to convert the 2D into the 3D structures. More confident structures may be obtained by Cambridge Structure Database (CSD) that stores the crystal structures of compounds. Before heading to the next steps, checking the protonation and tautomeric state is also mandatory to allow confident and accurate interactions in docking them into the target. Moreover, regarding the biological target preparation, beyond correcting the structure for a given pH, an interesting optional step can be to cluster the binding site among an ensemble of structures, either from different

| Table 12.2 Free repository of protein structures |
|-----------------------------------------------|
| Repositories | Data       | Web link                                      |
| Protein Data Bank | Experimental | https://www.rcsb.org/                      |
| SWISS-MODEL | Model       | https://swissmodel.expasy.org/repository      |
| ModBase | Model       | https://modbase.compbio.ucsf.edu/modbase-cgi/index.cgi |
Fig. 12.3 Example of Virtual Screening preparation. In the absence of experimental structure, the protein vivapain-4 (surface representation in gray and its catalytic site in red) of *Plasmodium vivax* was modeled from homology toward the PDB structure ID:3BPM (falcipain-3 from *Plasmodium falciparum*) with the presence of Leupeptin (represented in 2D and 3D at the catalytic site of vivapain-4), a known inhibitor for cysteine-proteases

| Table 12.3 Free database of compounds |
|--------------------------------------|
| Repositories                         | Data                                | Web link                          |
| ZINC                                 | Commercially available compounds     | https://zinc.docking.org/          |
| Cambridge Structure Database (CSD)   | Crystal structures of compounds (3D) | https://www.ccdc.cam.ac.uk/        |
| ChEMBL Database                      | Bioactive molecules                  | https://www.ebi.ac.uk/chembl/      |
| ChemSpider                           | Chemical structures                  | http://www.chemspider.com/         |
| DrugBank                             | Drugs library (2D and 3D)            | https://www.drugbank.ca/           |
| PubChem                              | Small compounds library (2D)         | https://pubchem.ncbi.nlm.nih.gov/  |
| DEKOIS 2.0                            | Decoys                              | http://www.dekois.com/             |
| DUD-E                                | Decoys                              | http://dude.docking.org/           |
models or structures found in repositories or from MD simulations, for finding a more accurate pose in docking. Therefore, clustering can be done using software of MD such as Amber’s cpptraj or GROMACS’s clustering function [33].

Compounds in investigation are usually docked in a grid assigned at one of the many docking software (Table 12.4). During the docking process, carefully explained in the next section of this chapter, the degrees of freedom of the molecules are explored, and software do that using different strategies, from systematic methods such as conformational search, to stochastic ones such as Monte Carlo. This step gives scores based on Van der Waals and electrostatic interactions at the binding site for the predicted poses, which will be used to rank and choose the best potential ligands further along. It is an extremely important step that helps us reduce the number of molecules analyzed, through scoring and observing the compounds that are capable of binding and the ones that do not interact in a stable manner at the biological binding site [40]. After evaluating the energy of the complex protein-ligand, the evaluation of pharmacokinetics properties is the last step before entering in vitro assays with the set of compounds left [37].

Scoring functions for ranking the poses from docking, however, are still subject to improvement. That is, the predicted binding affinity of the compounds does not necessarily represent that the active compounds are the ones located at the top of the rank exclusively. Errors in the energy of the systems are one of the issues compromising the analysis. To overcome that, the most effective techniques developed so

| Table 12.4 Docking softwares to be used in a VS process |
|-----------------|-----------------|-------------------|
| Software        | Algorithm           | Web                           |
| **Free**        |                  |                                |
| AutoDock4.2     | Genetic Algorithm | http://autodock.scripps.edu/   |
| AutoDock Vina   | Broyden–Fletcher–Goldfarb–Shanno (BFGS) Method | http://vina.scripps.edu/       |
| DOCK 6          | Geometric Algorithm | http://dock.compbio.ucsf.edu/DOCK_6/index.htm |
| DockThor (Server) | Genetic Algorithm | https://dockthor.lncc.br/v2/   |
| PLANTS          | Stochastic search algorithm | https://uni-tuebingen.de/fakultaeten/mathematisch-naturwissenschaftliche-fakultaet/fachbereiche/pharmazie-und-biochemie/pharmazie/pharmazeutische-chemie/pd-ddr-exner/research/plants/ |
| **Commercial**  |                  |                                |
| Glide           | Systematic search algorithm | https://www.schrodinger.com/glide |
| GOLD            | Genetic Algorithm  | https://www.ccdc.cam.ac.uk/solutions/csd-discovery/Components/Gold/ |
far are machine learning and deep learning. Machine learning consists of an algorithm that is able to recognize patterns (chemical descriptors) in a dataset. It can be a training dataset and applied to the list of compounds in the test, or it can be the list of compounds itself, in which the algorithm tries to identify a pattern without previous exposure. Examples of algorithms available for such assessment can be the K-Nearest Neighbor (KNN), Naive Bayes (NB), and Random Forest (RF). Deep learning, on the other hand, is considered a class of machine learning approaches which works with neural networks. It comprises layers of multiple nonlinear arrangements of processing units to solve hierarchically the issue. Even though it represents a powerful way to process many descriptors quickly by presenting intrinsic flexibility in its connections, it still requires reproducibility improvements [34, 41].

Any model should be first tested on a training set of complexes, to be certain of the prediction scores on your own system. The selection is tested using algorithms in a training set of compounds, with known activity against a specific target. If no in house library of reference compounds is available for training the VS model developed, useful dataset training for different types of the target can be obtained at Database of Useful Decoys—Enhanced (DUD-E) repository, for example, where already known active compounds and their targets are found. Training evaluation may be accessed by specificity \((SP)\) for identifying true non-active compounds, selectivity \((SE)\) for correctly identification of true active compounds and through a combined prediction \((Q)\), in equations as follows:

\[
\begin{align*}
(a) \quad SP &= \frac{TN}{TN + FP} \\
(b) \quad SE &= \frac{TP}{TP + FN} \\
(c) \quad Q &= \frac{TP + TN}{TP + TN + FP + FN}
\end{align*}
\]

Equations for validation scoring: (a) Specificity \((SP)\), (b) selectivity \((SE)\), and (c) total prediction of the system \((Q)\). Compounds in the test will be scored as true positive \((TP)\), true negative \((TN)\), false positive \((FP)\), or false negative \((FN)\) [42].

The question “does my model truly separate active from inactive compounds?” can be solved by analysis of the receiver operating characteristics (ROC) curve. It helps visualize the profile of prediction throughout the compound list scores, from descriptors selected. Area under the curve (AUC) must be closer to one in order of the model in question to be considered effective, which means the model is effective in distinguishing an active compound as true positive and a compound not active as true negative [43].

### 12.2.3 Repurposing of Drugs

In a new disease outbreak, or in fighting the constant new drug resistant-diseases, it is extremely desirable to find compounds that have passed tests and are already in the market, that present activity against the new problem, in a strategy called drug repurposing. It is important because when redirecting an already commercial
This is of particular interest and use in the recent Ebola, Zika, and Covid-19 viral outbreaks, since developing a new drug takes a very long time and money to be available for the public. The repurposing process involves the assessment of databases of already commercial compounds, such as ZINC, followed by the standard steps in a VS process. Urgency is required in an outbreak event to reduce the spread of the disease and its consecutive deaths, and potent hardware and algorithms are, therefore, necessary for this combat. The most recent example of it is the pandemic Covid-19, which pushed scientists to their limits on finding the fastest alternatives to contain its spread, that is, by developing, in a very short amount of time, the screening of 1000 hits among 1.3 billion already regulated compounds against the main protease of such virus [44–47].

12.3 Molecular Docking in Drug Design

Molecular Docking, together with VS, is one of the initial stages of investigating molecules as potential drugs and aims to predict the conformation and orientation of ligands at the target macromolecule binding site [48]. It is based on the premise that the energy of the receptor-ligand complex is lower than that of the separate parts, being, therefore, favorable to the complex formation [35]. Complexes formed by protein-protein and protein-ligand are the most studied currently in drug discovery.

Molecular Docking can be divided into two main parts:

I. the investigation of the orientation and conformation of a ligand at the receptor binding site carried out by search methods;
II. and the prediction of the binding free energy, or affinity, of a receptor-ligand complex through a scoring function.

In order to perform Molecular Docking, the macromolecule receptor and the ligand molecule structures are needed. To apply the technique, some steps for preparing the system components are necessary (Fig. 12.4).

Initially, one should obtain the three-dimensional structure of the receptor, usually a protein, ideally from experimental data that can be acquired in libraries such as the Protein Data Bank (PDB). PDB has the coordinates of the atoms of individual molecules and resolved complexes by experimental methods such as X-ray diffraction crystallography and nuclear magnetic resonance (NMR). For proteins that have not yet been solved, some computational techniques can be used as alternatives. One of these is the Comparative Modeling, in which the protein to be discovered has its amino acid sequence aligned with the sequences of other proteins with previously known structures [49]. Ab initio modeling can also be performed using theoretical calculation for structural prediction [50].
The binding site at the receptor to be analyzed must be selected manually and delimited by specifying the coordinates or selected automatically using coordinates of ligands already bonded. In cases where the site is not known, prediction algorithms can be used to find possible cavities. Different algorithms are used for this purpose, from the ones that use small probes to map possible interaction sites on the protein surface [51] to algorithms in which druggability scores are predicted after training and testing a target library [52].

Computational mapping methods of co-solvent based on Molecular Dynamics (MD) are also being employed with the objective of searching at the entire surface of the protein, encompassing orthosteric and allosteric sites. In this mapping, a set of conformations of the target proteins, sampled by MD and fragments of small molecules, are used to identify and characterize binding sites [53, 54]. In addition, most software used for Docking have the option of “Blind Docking”, where the search for potential interaction sites is done over the entire surface of the receptor structure, thus presenting a high computational cost compared to those previously described [55].

For the standard molecular docking calculation, to reduce computational costs, the preparation of the receptor involves the construction of a grid, which can be understood as a box containing a cubic network of points where the interaction energy is calculated, delimiting the binding site [55]. The receptor potential energy is pre-calculated for each point of the grid and, subsequently, the interaction electrostatic and Lennard-Jones energies of the ligand are calculated with each point in this grid [56]. The choice of box size influences the correct results and virtual screening time. For the Vina software, higher precision was observed when the dimensions of the search space are about 3 times greater than the ligand’s rotation radius [57]. Some
software perform the calculation of the grid previously, as occurs in AutoDock, and others perform in real time producing a faster calculation, such as Vina [55].

In addition, it is important to investigate the protonation states of the amino acids that perform interactions at the receptor binding site. Some residues, such as Aspartate, Glutamate, Histidine, and Cysteine, are not commonly found in their usual protonation states due to the electrostatic environment in which they are inserted. In this case, it is possible to evaluate these amino acids in other complexes of the target protein and proteins homologous to several solved ligands, as well as using programs that determine the pka of each residue, such as the PROPKA software and the H++ server.

The ligands can have their structures obtained from public databases such as PubChem and Zinc [58, 59], virtual compounds, or organic synthesis. To build the input files with the three-dimensional coordinates, specific programs for the molecular draw can be used, such as MarvinSketch, ChemDraw, and Avogadro [60]. Some points must be taken into account when preparing the ligand file. These include charges, stereochemical geometry, and just like the receptor, the protonation states of chemical groups [54]. To determine the protonation states of the ligand groups, not only the pH should be considered, but also the interaction with the macromolecule binding site, which makes it a slightly more complex task. However, in some cases Docking can be performed with several protonation states of the ligand [61]. The charges are usually distributed as partial charges to the constituent atoms of the molecule, from the calculated net charge. Most Docking programs assume that both charges and protonation states are fixed and do not vary between on and off states [56].

Some important conformational changes can occur in the molecular recognition of ligands that contain nonaromatic cyclic structures. Software such as LigPrep [62] can generate representative structures of these changes that can be used in Molecular Docking with the receptor. The type of Docking to be calculated must be defined, which directly influences the results obtained. The ligand and the receptor can be considered rigid or flexible in the calculations. In the case of the flexible ligand, some additional terms for energy must be included, which means that the potential is “smoothed out” to allow the flexibility of the residue and its interaction with the ligand [55]. The degree of complexity in the calculation of flexible receptors and ligands is higher in relation to rigid ones, since the number of degrees of freedom to be evaluated is higher. Some software that perform docking with flexible receptors and ligands are based on the induced-fit theory proposed by Koshland, where both must suffer small changes to the binding that may occur [63].

12.3.1 Theory of Molecular Docking

As described before, the Molecular Docking method is divided into two parts: Search Algorithm and scoring functions, which involves a conformational search and the selection of best pose by the scoring functions (Fig. 12.5).
12.3.1.1 Search Algorithm

The exploration of different orientations and possible conformations for a ligand into a receptor binding site by the Docking software should find the best solution, which means to find the global minimum of energy. The function of energy needs to reproduce the entropic and enthalpic effects associated with the free energy of the system so that the global minimum of energy corresponds to the mode of binding described experimentally. However, due to the approximations introduced in the molecular interaction model, the global minimum does not always achieve this correspondence [61]. The selection of Search Algorithm is extremely important to reliable results of conformation and placement of the ligand at the receptor binding site. This search can be systematic or stochastic.

Systematic Search comprises a large sampling of conformations and a combination of structural parameters exploring each ligands’ degree of freedom, including the rotational, translational, and conformational (rotatable dihedral angles) movements, which can generate a combinatorial explosion in the number of attempts. This search method uses more resources and spends significantly more time to generate the conformation poses and assesses them individually. By building the ligand from different fragments, selecting one as an anchor and sequentially adding combinations of remaining fragments, this problem is avoided. Some Docking algorithms also apply pharmacophoric data from proteins and ligands and try to match the distances between each point of them. Stochastic search is made randomly, developing different conformations based on bond rotations as degrees of freedom. Some implementations
are necessary to guarantee the convergence to the best solution. Genetic Algorithm, Monte Carlo, Swarm Optimization, and Tabu Search are commonly implemented methods [55].

### 12.3.1.2 Scoring Functions

A large range of conformations is generated by the searching algorithms, although just a moiety of them are biologically relevant. To evaluate the quality of these conformations and to order according to the binding affinity to the receptor, it’s necessary to apply scoring functions. Scoring functions are mathematical models, usually linear functions, composed of terms related to physicochemical properties as intermolecular interactions, desolvation, electrostatic, and entropic effects, involved at interactions of the small molecule and the binding site of the receptor.

There are different functions of scoring, that may vary mainly at the number and types of terms, mathematical complexity, and the form of parametrization. These functions can be applied according to the objective and the stage of the study of molecular docking. In the first stages of the docking, simpler functions can be used to evaluate the conformations generated by the search method, to reduce computational costs. At the final stages, more sophisticated and complex functions are employed to obtain greater accuracy in predicting the correct binding mode and affinity for the receptor [63]. Scoring functions can be classified as force field-based, empirical, and knowledge-based.

Force field-based functions are a sum of terms from a classic molecular force field, as MMFF94, CHARMM, and AMBER, including experimental data or quantum calculations as parameters. The terms are divided by bonding interactions, as energy associated with the torsional chemical bonds, and non-bonding interactions, as associated with Van der Waals interactions, electrostatics and hydrogen bonds. To include additional effects as solvation effects and hydrophobic contacts, some terms are incorporated.

Empirical scoring functions are based on the idea that the binding free energy can be related through the sum of uncorrelated variables. This function is developed from tridimensional structures and affinities of known receptor-ligand complexes, on which the terms are adjusted to reproduce experimental data more accurately possible. Therefore, these functions are limited, as they are derived from heterogeneous data in training sets. Compared to the force field-based functions, the empirical ones are computed much faster.

Knowledge-based functions are based on the statistical mechanics of simple fluid systems, which employ the potentials of mean force (PMF). They are built from the statistical analysis between the atom pairs of experimentally solved receptor-ligand complexes. Thereby, like the empirical functions, these functions are differentiated by the size of the training set and the type of receptor-ligand interactions considered in parameterization. A combined approach using more than one type of scoring function can be also applied to improve accuracy.
12.3.2 Challenges of Molecular Docking

One of the best ways to evaluate a protein-ligand complex is to use a dynamic system approach. However, the protein flexibility remains a challenge in Molecular Docking and most current flexible ligand Docking programs treat the receptor as rigid. Some methods apply different tools to solve that. The Induced-fit method considers the neighboring residues of the binding site as flexible. The backbone movement affects the side chains, which leads to a higher order of magnitude in terms of the number of degrees of freedom [63]. These flexible docking algorithms predict the binding mode of a molecule, and also its binding affinity relative to other compounds, more accurately than rigid body ones.

Protein Energy Landscape Exploration (PELE) is a Monte Carlo method to sample flexibility combining protein and ligand perturbations. This technique is based on a steered localized perturbation followed by side-chain sampling and minimization cycles [64]. It acts as a useful refinement, in particular, if the binding site of a protein is not known. Compared to Molecular Dynamics simulations, PELE has a smaller computational cost associated with the exploration. Machine Learning (ML) functions are also recently applied to simulate the flexibility of the receptor in the scoring functions for Molecular Docking.

An interesting strategy to improve the performance with respect to compound scoring and pose prediction is to combine two or more Docking methods, as a Consensus Method [65]. These methods showed better results than the single ones that compose it. They can vary in how conformations are obtained, the scoring functions selected, and the algorithm to achieve the consensus. Even though Molecular Docking has been developed for about 10 years, it remains being updated and new features are incorporated. Currently, the Virtual Screening of thousands of compounds is performed by using the background of the Molecular Docking technique.

12.4 Molecular Dynamics (MD)

Molecular Dynamics (MD) is an in silico technique based on Newton’s laws that describes the variation of atomic positions in a molecule as a function of time. Therefore, it is different from the docking technique, which is mainly stochastic. With the use of MD, it is possible to evaluate characteristics such as content of secondary structure, orientation of side chains, conformation of loops, and the energy of interaction between different molecules, such as protein and ligands, over time. For this reason, it presents similar results to experimental methods such as Nuclear Magnetic Resonance (NMR), but with the advantage of more reduced costs, MD can be applied to bigger molecular systems, has better reproduction of biological systems, and costs less time [61].
MD is part of the so-called classical methods, or methods of Molecular Mechanics (MM), and it is based on the solution of Newton’s second law, according to the equation:

\[ F_{xi} = \frac{d^2 x_i}{dt^2} m_i = \frac{\Delta v_i}{\Delta t} m_i = a_i m_i \]

where \( F_{xi} \) is the force applied to atom \( i \) at position \( x_i \), \( t \) is time, \( v \) the velocity, and \( a_i \) the acceleration of atom \( i \).

### 12.4.1 Force Fields

In Molecular Mechanics (MM), molecules are described as a set of connected atoms, instead of nuclei and electrons as in quantum methods. Since these atoms are linked to other atoms, they are subject to intermolecular and intramolecular forces. A complete set of interaction potentials between particles is referred to as a “force field”. An empirical force field, as it is known, is a mathematical function that allows the calculation of total potential energy of a system, \( V(r) \), from the coordinates of its particles. In molecular systems, \( V(r) \) is described as the sum of several energy terms, including terms for bonded atoms (lengths and angles of bonding, dihedral angles) and terms for nonbonded atoms (Van der Waals and Coulomb interactions) [66]. In most force fields, the potential energy of the molecules can be represented by the equation:

\[ V(r) = \sum E_l + \sum E_a + \sum E_t + \sum E_{nl} \]

where \( \sum \) is a sum of the energy terms as a function of the bonding length \( E_l \), the bonding angles \( E_a \), the torsion of the angles \( E_t \), and the interactions of nonbonded atoms \( E_{nl} \) (Coulombian and Van der Waals).

The existing force fields have been developed independently and with all specific parameter sets. Some include other terms to specifically describe hydrogen bonds or to couple oscillations between bond angles and lengths, in order to achieve better agreement with vibrational spectra. The reliability of the results is based on the elaboration of a force field with well-defined parameters. The choice of the force field depends, to a large extent, on the system to be studied and on the properties that will be investigated. For example, while a type of force field can describe proteins with high fidelity, it can be quite limited in reproducing the geometry of carbohydrates or nucleic acids. In the case of biomolecular systems, the most used force fields are CHARMM, GROMOS, AMBER, OPLS, among others [61, 66].

There are different levels of simplification in the description of atoms in a force field depending on the type of system to be studied. Force field can describe all atoms in the system (all-atom), which requires a higher computational cost and may be unfeasible for large systems depending on the computational power available. The
first simplification would be to join the hydrogen atoms attached to carbon atoms as a pseudo-atom, representing the properties of CH, CH2, or CH3 groups. Another possibility for simplification is through the method called coarse-grained (CG). In this force field, several atoms can be aggregated into a single particle, analogous to the pseudo-atom of the joined atom model. This model is useful for systems such as lipid membranes, allowing long simulation times, since they are much faster to perform. This model is not indicated in studies where changes in secondary structure content are expected and for studies involving systems of proteins and organic molecules, such as drugs [61].

The simulations can also consider the solvent explicitly, cases wherein the water molecules are physically included in the simulations, or implicitly, in which the water molecules are not included in the simulations, only the dielectric properties of the solvent are represented, considerably reducing the time of the simulation. Another difference between the force fields is the description of the water molecules, an example of a widely used model is the TIP3P, used in the force field of AMBER, OPLS, and GROMOS [61].

The programs currently available for performing MD differ in relation to the force field used, computational cost and access; some of the most used are AMBER, GROMACS, and NAMD. These differences need to be evaluated by the researcher according to the system to be studied.

12.4.2 MD Simulations

MD simulation involves several steps, from obtaining the protein structure either by accessing the repositories of experimental or modeled structures (Table 12.2) to parameterization, minimization, equilibration, simulation, and analysis. The diagram below demonstrates the basic steps for a protein MD simulation (Fig. 12.6) [67].

12.4.2.1 Energy Minimization and Equilibration

Before performing the molecular dynamics simulation, it is necessary to perform an energy minimization of the system, which aims to eliminate bad contacts between atoms. During this calculation, the geometry of the system is optimized causing a reduction of its overall energy, reaching a more stable conformation, which is a state of energy minimum. This conformation will be the starting point for molecular dynamics. A widely used method is the steepest descent, which uses the first derivative to determine the direction to the minimum. It is a robust technique used to initially optimize a structure that is far from a minimum point [66]. A new structure optimization is then performed with a second derivative minimization of energy method to improve the result.
After minimizing the energy, it is necessary to heat the system at constant volume, bringing it to a temperature of interest, which is usually 300 K. After heating, equilibration is performed, with the function of controlling the system pressure (protein + water + ions), and stabilizing the system properties. In this stage, it is common to restrict the movement of the atoms of the protein and crystallographic water molecules from PDB, which allows the free water molecules and ions to organize and tend to come into balance, with the position restriction gradually decreasing until all the components of the system have obtained total freedom of movement. The equilibration period is variable and depends on the system under study. Generally, it is considered finished when the thermodynamic balance is reached. From that point, you can then generate the MD trajectories and calculate the different properties for the system of interest.

12.4.2.2 Periodic Boundary Conditions

The use of the explicit solvent in Molecular Dynamics is what demands the most computational cost, therefore, it is necessary to optimize the number of water molecules in order to obtain a faster simulation. For this reason, boxes are used, which are the three-dimensional space where the biomolecule and solvent are placed; the size and shape of these boxes are important in defining how many water molecules will be used in the simulation. The most common forms are cubic, octahedral, and dodecahedral. The form to be used will depend on the system and the computer power available for such simulation, but in the case of proteins, the cubic form is widely used [61].

Unlike biological conditions, in MD we have a vacuum around the simulation box, which could cause the solvent to evaporate. To prevent the system’s molecules from exceeding the box’s limits, a strategy called periodic boundary conditions was
created. In this strategy, the box is replicated in all its directions from space, making it so that if a molecule leaves the central box, one of its images enters from the opposite side, representing a continuity of the simulation, in a closer way to the experimental conditions [61].

12.4.2.3 Ensemble

The ensemble is the group of settings and properties kept constant during the simulation, representing the state of the system. Examples of widely used ensembles are the canonical or NVT (with a constant number of particles, volume, and temperature), generally used for the equilibration stage, and the isothermal-isobaric or NPT (with a constant number of particles, pressure, and temperature), normally used during MD simulation.

12.4.2.4 Sampling

The ability of a simulation to describe the expected behavior of a system is related to sampling. It is necessary that the simulation has enough time to observe the phenomena of interest, which is a great challenge in the simulation of biomolecules, since the computational cost to obtain a sample close to the ideal is very high. The longer the MD simulation, the greater the sampling, however, the number of atoms in the system has a great influence on how many conformations can be adopted. An all-atom simulation would need a larger time scale to obtain the same sampling as a coarse-grained simulation, for example. The type of problem to be analyzed also has a great influence on sampling and, consequently, on the necessary simulation time. The analysis of domain movement, for example, requires a longer simulation time, on the microseconds scale, whereas in the analysis of protein-ligand interaction, usually, the nanoseconds scale is sufficient.

12.4.3 Analysis of MD Simulations

Analyzing MD simulations is considered by many to be the most difficult part of MD. The first and most basic analysis to be performed is the stability of properties such as temperature, volume, pressure, density, and total energy throughout the simulation. If these were stable, you can move on to the rest of the analyses.

Among the most common analyses to be carried out, mainly for proteins, there is the root mean square deviation (RMSD), which allows evaluating how much the protein structure varied along the MD simulation in comparison with an initial structure, which is usually the crystallographic structure or the minimized structure before MD. In this analysis, the RMSD starts from 0 and rises until it reaches a point of equilibrium, where it remains until the end of the simulation (Fig. 12.7). Root
mean square fluctuation (RMSF) evaluates the variation of each protein residue in relation to an average structure along the molecular dynamics. More flexible regions, such as loops, will show greater variation (Fig. 12.8).

**12.4.4 MD Applications**

Molecular Dynamics has numerous applications, such as the prediction of protein structures, analysis of protein-protein, protein-lipid, and protein-ligand interactions, virus studies, and drug design for the most diverse diseases.
12.4.4.1 Drug Design and Free Energy Calculation

Among a variety of approaches used for structure-based drug design, MD simulation combined with a free energy calculation method is a great tool, as it can provide detailed information about protein-ligand interactions and consider the environment with solvent and protein flexibility [68].

The calculation of binding free energy can be performed using the Molecular Mechanics energies combined method with the Poisson–Boltzmann or Generalized Born and surface area continuum solvation (MM/PBSA and MM/GBSA). The purpose of this method is to calculate the difference in free energy between the bound and unbound states of two solvated molecules. However, in this calculation, most of the energy contributions would come from solvent-solvent interactions, making the calculation take a long time. Thus, in the AMBER program, an effective method is used, which is the division of the calculation according to the following equation [69]:

\[ \Delta G_{\text{bind, solv}} = \Delta G_{\text{bind, vacuum}} + \Delta G_{\text{solv, complex}} - (\Delta G_{\text{solv, ligand}} + \Delta G_{\text{solv, receptor}}) \]

where \( \Delta G_{\text{bind, solv}} \) is the free energy of binding in aqueous medium, \( \Delta G_{\text{bind, vacuum}} \) is the free energy of binding in vacuum, \( \Delta G_{\text{solv, complex}} \) is the free energy of the solvated complex, \( \Delta G_{\text{solv, ligand}} \) is the free energy of the solvated ligand, and \( \Delta G_{\text{solv, receptor}} \) is the free energy of the solvated protein [68].

The analysis of free energy is important mainly to compare the interaction of different ligands with the same protein performing an MD simulation, as more negative values indicate that that compound could be more promising, as it performed more relevant intermolecular interactions with the target. Programs such as AMBER also allow the decomposition of the binding free energy by aminoacid residue, which is very important to evaluate which residues have the best contribution to protein-ligand interaction.

The analysis of the prevalence of hydrogen bonds over time is another important application of MD for drug design, as it allows to evaluate which residues made stable hydrogen bonds with the ligand and to compare with the existing data in the literature for the protein under study, allowing greater reliability of the results obtained (Table 12.5).

| Ligand | Protein   | Prevalence (%) |
|--------|-----------|----------------|
| N1H1   | GLU71_OE2 | 53.8           |
| O1     | ASP168_NH | 65.4           |
| N2H2   | GLU71_OE2 | 22.5           |
| O3     | MET109_H  | 43.7           |
| N1H1   | GLU71_OE1 | 37.0           |

Table 12.5 Prevalence of hydrogen bonds between p38 MAPK and an inhibitor candidate in a 50 ns MD simulation.
Another interesting approach to drug design in Molecular Dynamics is the solvent mapping technique, which allows the simulation of proteins in different solvents, such as ethanol/water, to identify hotspots that could be important in the interaction with possible inhibitors, through observations of intermolecular interactions between residues of the studied protein and solvent molecules. This technique is an in silico version of the multiple solvent crystal structures (MSCS) approach, but with the advantage of being much faster and cheaper. Water represents polar interactions well, while the ethanol molecule is highly miscible in water and represents polar interactions, through hydroxyl, and nonpolar interactions, through the methyl group [70]. Structural clustering algorithms can be used to extract representative conformations from MD trajectories to understand different interaction patterns between the ligand, in this case water and ethanol molecules, and the protein that contributes to binding (Fig. 12.9).

**12.4.4.2 Study of Viruses**

Another interesting application of Molecular Dynamics is the study of viruses. In 2006, Schulten and co-workers reported the first all-atom simulation of a complete non-enveloped virus, called the satellite tobacco mosaic virus (STMV). Regarding enveloped viruses, the first capsid simulation of an immature HIV virus was carried out in 2010 by Ayton and Voth. However, as the viral particles are large for all-atom simulations, most studies are directed to simulations of viral proteins and capsids. The limitation for this area is the study of more complex processes such as replication, membrane fusion, virus maturation, and nuclear entry. However, the study of viral
proteins provides a lot of important information, also allowing the drug design for viral diseases through MD [71, 72].

Among the most recent examples of MD applications in viruses is the study of the Ebola virus by Pappalardo and co-workers, wherein they determined that VP24 protein of the Ebola virus is a key protein for pathogenicity, through the observation of its interactions with the human protein KPNA5 [73]. Nasution and co-workers also conducted a drug design study to combat the Ebola virus through Molecular Dynamics, finding promising compounds [74]. Another example is the study of influenza B virus, by Zhang and Zheng, wherein they provided structural and dynamic details of the effects of serine triad on proton conduction in the tetrameric channel of influenza B channel M2 (BM2), important in the virus life cycle, at the atomic level [75]. Studies of ZIKA virus are also being carried out, such as by Bowen and co-workers, wherein a collection of more than 7 million commercially and freely available compounds from the ZINC15 database were subjected to a virtual screening procedure consisting of consensus-based docking followed by MD simulation and binding energy calculations in order to identify promising potential inhibitors of the Zika NS2B-NS3 [76].

References

1. Hagen JB (2000) The origins of bioinformatics. Nat Rev Genet 1(3):231–236
2. Mesquita APR (2014) Modelagem molecular de compostos anti-citomegalovírus. Trabalhos de Conclusão de Curso (Universidade Federal Fluminense) 1–52
3. Vale G, Silva T, Ferreira A, Bou-Habib D, Siqueira M, Lopes TM, Miranda M (2020) Inibição da replicação do influenza através da modulação de fatores restritivos pelos ligantes dos receptores CCR5 e CXCR4. Resumos Caderno Simpósio de Virologia (Universidade Federal do Rio de Janeiro) 53
4. Simões RSQ, Barth OM (2015) Papillomavirus: viral vectors in the gene therapy and new therapeutic targets. Int J Biomed Res 6(10):763–768
5. Jain NK, Sahni N, Kumru OS, Joshi SB, Volkin DB, Middaugh CR (2015) Formulation and stabilization of recombinant protein based virus-like particles vaccine. Adv Drug Deliv Rev 93(1):42–45
6. Huber B, Schellenbacher C, Shafti-Keramat S, Jindra C, Christensen N, Kirnbauer R (2017) Chimeric L2-based virus-like particle (VLP) vaccines targeting cutaneous human papillomaviruses (HPV). Plos One 1–27
7. Simões RSQ, Barth OM (2017) Immunological and structural analysis of HPV-positive cervical carcinoma cell lines and bovine papillomavirus virus-like particles (BPV-VLP). Int J Adv Res 5:1003–1009
8. Lavine BK, Mirjankar N (2012) Clustering and classification of analytical data. Encyl Anal Chem
9. Lenz P, Day PM, Pany YYS, Frye SA, Jensen PN, Lowy DR, Schiller JT (2020) Papillomavirus-like particles induce acute activation of dendritic cells. J Immunol 166:5346–5355
10. Nagib NRC (2017) Modelagem e dinâmica molecular da oncoproteína E6 do vírus do papiloma humano (HPV) tipo 18. Trabalho de Conclusão de Curso (Universidade Federal de Uberlândia) 1–37
11. Simões RSQ, Barth OM (2018) Papillomavirus (PV)—associated skin diseases in domestic and wild animals: animal nucleotide sequence identity of PV types to their closest related PV and HPV sequences deposited in the gen bank. Int J Curr Microbiol Appl Sci 6:938–951
12. Dehesa-Violante M, Nunez-Nateras R (2007) Epidemiology of hepatitis virus B and C. Arch Med Res 38(6):606–611
13. Heermann KH, Goldmann U, Schwartz W, Seyffarth T, Baumgarten H, Gerlich WH (1984) Large surface proteins of hepatitis B virus containing the pre-S sequence. J Virol 52:396–402
14. Arauz-Ruiz P, Norder H, Robertson BH, Magnus LO (2002) Genotype H: a new American genotype of hepatitis B virus revealed in Central America. J Gen Virol 83:2059–2073
15. Vieira MB (2010) Estudos de antigenicidade e imunogenicidade de vetores HBsAg carreadores de epitope do HCV. Dissertação de Mestrado (Fundação Oswaldo Cruz) 1–145
16. Seeger C, Mason W (2000) Hepatitis B virus biology. Microbiol Mol Biol Rev 64:51–68
17. Delpeyroux F, Chenciner N, Lim A, Malpie Y, Blondel B, Crainic R et al (1986) A poliovirus neutralizing epitope expressed on hybrid hepatitis B surface antigen particles. Science 233:472–475
18. Netter HJ, Macnaughton TB, Woo W, Tindle R, Gowans E (2001) Antigenicity and immunogenicity of novel chimeric hepatitis B surface antigen particles with exposed hepatitis C virus epitopes. J Virol 75:2130–2141
19. Patient R, Hourioux C, Vaudin P, Pages JC, Roingeard P (2009) Chimeric hepatitis B and C viruses envelope proteins can form subviral particles: implications for the design of new vaccine strategies. New Biotechnol 25(4):226–234
20. Forns X, Bukh J, Purcell RH (2002) The challenge of developing a vaccine against hepatitis C virus. J Hepatol 37:684–695
21. Major MM, Vivitski L, Schleef M, Whalen RG, Trepo C (1995) DNA-based immunization with chimeric vectors for the induction of immune responses against the hepatitis C virus nucleocapsid. J Virol 69:5798–5805
22. Drazan KE (2000) Molecular biology of hepatitis C infection. Liver Transpl 6:396–406
23. Geissler M, Gesein A, Tokushige K, Wands JR (1997) Enhancement of cellular and humoral immune responses to hepatitis C virus core protein using DNA-based vaccines augmented with cytokine-expressing plasmids. J Immunol 158:1231–1237
24. Yam R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q (2020) Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE-2. Science
25. Liu J, Cao R, Xu M, Wang X, Zhang H, Hu H, Li Y, Hu Z, Zhong W, Wang M (2020) Hydroxychloroquine, a less toxic derivative of chloroquine is effective in inhibiting SARS-CoV-2 infection in vitro. Cell Discov 6:16
26. Simões RSQ, Barth OM (2019) Emerging and reemerging virus. In: Human and veterinary virology, vol 1, pp 317–24
27. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q (2020) Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE-2. Science
28. Liu J, Cao R, Xu M, Wang X, Zhang H, Hu H, Li Y, Hu Z, Zhong W, Wang M (2020) Hydroxychloroquine, a less toxic derivative of chloroquine is effective in inhibiting SARS-CoV-2 infection in vitro. Cell Discov 6:16
29. Magnani DM et al (2017) Neutralizing human monoclonal antibodies prevent Zika virus infection in macaques. Sci Transl Med 9:8184
33. Sliwoski G, Kothiwale S, Meiler J, Lowe EW Jr (2013) Computational methods in drug discovery. Pharmacol Rev 66:334–395
34. D’Souza S, Prema KV, Balaji S (2020) Machine learning in drug–target interaction prediction: current state and future directions. Drug Discov Today
35. Yu W, Mackerel AD Jr (2017) Computer-aided drug design methods. Methods Mol Biol Antibiot 1520:85–106
36. Macalino SJY, Gosu V, Hong S, Choi S (2015) Role of computer-aided drug design in modern drug discovery. Arch Pharmacal Res 38:1686–1701
37. Shim J, Mackerell AD Jr (2011) Computational ligand-based rational design: role of conformational sampling and force fields in model development. MedChemComm 2:356–70
38. Schmidtke P, Bidon-Chanal A, Luque FJ, Barril X (2011) MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. Bioinformatics 27:3276–3285
39. Anderson AC (2003) The Process of Structure-Based Design. Cell Chem Biol 10:787–797
40. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discovery 3:935–949
41. Wang D, Cui C, Ding X, Xiong Z, Zheng M, Luo X, Jiang H, Chen K (2019) Improving the virtual screening ability of target-specific scoring functions using deep learning methods. Front Pharmac 10:924
42. Ren J-X, Zhang R-T, Zhang H (2020) Identifying novel ATX inhibitors via combinatory virtual screening using crystallography-derived pharmacophore modelling, docking study, and QSAR analysis. Molecules 25:1107
43. Swift RV, Jusoh SA, Offutt TL, Li ES, Amaro RE (2016) Knowledge-Based methods to train and optimize virtual screening ensembles. J Chem Inf Model 56:830–842
44. Zheng W, Sun W, Simeonov A (2018) Drug repurposing screens and synergistic drug-combinations for infectious diseases. Br J Pharmacol 175:181–191
45. Schuler J, Hudson ML, Schwartz D, Samudrala R (2017) A systematic review of computational drug discovery, development, and repurposing for ebola virus disease treatment. Molecules 22:1777
46. Santos F, de Nunes DAF, Lima WG, Duvyt D, Santos LL, Taranto AG, Maria Siqueira Ferreira J (2019) Identification of Zika virus NS2B-NS3 protease inhibitors by structure-based virtual screening and drug repurposing approaches. J Chem Inf Model
47. Ton AT, Gentile F, Hsing M, Ban F, Cherkasov A (2020) Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds. Mol Inf
48. Morris GM, Lim-Wilby M (2008) Molecular docking
49. Webb B, Sali A (2016) Comparative protein structure modeling using MODELLER. Curr Protoc Protein Sci 86:2.9.1–2.9.37
50. Barreiro EJ, Rodrigues CR (1997) Modelagem molecular: Uma Ferramenta Para O Planejamento Racional De Fármacos Em Química Medicinal. Química Nova 20(1)
51. Kozakov D, Grove LE, Hall DR, Bohnuud T, Mottarella SE, Luo L, Xia B, Beglov D, Vajda S (2015) The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. Nat Protoc 10(5):733–755
52. Volkamer A, Kuhn D, Grombach T, Rippmann F, Rarey M (2012) Combining global and local measures for structure-based druggability predictions. J Chem Inf Model 52(2):360–372
53. Sayyed-Ahmad A (2018) Hotspot identification on protein surfaces using probe-based MD simulations: successes and challenges. Curr Top Med Chem 18(27):2278–2283
54. Arcon JP, Defelipe LA, Modenutti CP, Lopez ED, Alvarez-Garcia D, Barril X, Turjanski AG, Martí MA (2017) Analyzing the molecular basis of enzyme stability in ethanol/water mixtures using molecular dynamics simulations. J Chem Inf Model 57:846–863
55. Prieto-Martínez FD, Arciniega A, Medina-Franco JL (2018) Molecular docking: current advances and challenges. TIP Rev Esp Cienc Quím Biol 21:1–23
56. Torres PHM, Sodero ACR, Jofily P, Silva FP Jr (2019) Key topics in molecular docking for drug design. Int J Mol Sci 20:4572019
57. Feinstein WP, Brylnski M (2015) Calculating an optimal box size for ligand docking and virtual screening against experimental and predicted binding pockets. J Cheminform 7:18
58. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J, Bolton EE (2019) PubChem 2019 update: improved access to chemical data. Nucleic Acids Res 47
59. Sterling T, Irwin JJ (2015) ZINC 15—ligand discovery for everyone. J Chem Inf Model 55(11):2324–2337
60. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J Cheminform 4:17
61. Verli H (2014) Bioinformática: da Biologia à Flexibilidade Molecular. Sociedade Brasileira de Bioquímica e Biologia Molecular
62. LigPrep, Schrödinger, LLC, New York, NY, 2020. https://www.schrodinger.com/ligprep. Accessed 29 March 2020
63. Pagadala NS, Syed K, Tuszynski J (2017) Software for molecular docking: a review. Biophys Rev. 9:91–102
64. Borrelli KW, Vitalis A, Alcantara R, Guallar V (2005) PELE: Protein Energy Landscape Exploration, a novel monte carlo based technique. J Chem Theory Comput 1(6):1304–1311
65. Ren X, Shi YS, Zhang Y, Liu B, Zhang LH, Peng YB, Zeng R (2018) Novel consensus docking strategy to improve ligand pose prediction. J Chem Inf Model 58(8):1662–1668
66. Namba AM, Da Silva VB, Da Silva CHTP (2008) Dinâmica molecular: Teoria e aplicações em planejamento de fármacos. Eclet Quim 33(4):13–24
67. Yu W, Jr ADM (2016) Computer-aided drug design methods. Methods Mol Bio Antibiot 85–106
68. Yang Y, Shen Y, Liu H, Yao X (2011) Molecular dynamics simulation and free energy calculation studies of the binding mechanism of allosteric inhibitors with p38 α MAP Kinase. J Chem Inf Model 3235–3246
69. Walker R (2020) Amber advanced tutorials–tutorial 3–MM-PBSA–introduction. http://ambermd.org/tutorials/advanced/tutorial3/. Accessed 27 March 2020
70. Arcon JP, Defelipe LA, Modenutti CP, López ED, Alvarez-Garcia D, Barril X, Martí MA (2017) Molecular dynamics in mixed solvents reveals protein-ligand interactions, improves docking, and allows accurate binding free energy predictions. J Chem Inf Model 57(4):846–863
71. Huber RG, Marzinek JK, Holdbrook DA, Bond PJ (2017) Multiscale molecular dynamics simulation approaches to the structure and dynamics of viruses. Prog Biophys Mol Biol 128:121–132
72. Perilla JR, Goh BC, Cassidy CK, Liu B, Bernardi RC, Rudack T, Schulten K (2015) Molecular dynamics simulations of large macromolecular complexes. Curr Opin Struct Biol 31:64–74
73. Pappalardo M, Collu F, Macpherson J, Michaelis M, Fraternali F, Wass MN (2017) Investigating Ebola virus pathogenicity using molecular dynamics. BMC Genomics 18(Suppl 5)
74. Nasution MAF, Toepak EP, Alkaff AH, Tambunan USF (2018) Flexible docking-based molecular dynamics simulation of natural product compounds and Ebola virus Nucleocapsid (EBOV NP): a computational approach to discover new drug for combating Ebola. BMC Bioinform 19(Suppl 14)
75. Zhang Y, Zheng QC (2019) What are the effects of the serine triad on proton conduction of an influenza B M2 channel? an investigation by molecular dynamics simulations. Phys Chem Chem Phys 21(17):8820–8826
76. Bowen LR, Li DJ, Nola DT, Anderson MO, Heying M, Groves AT, Eagon S (2019) Identification of potential Zika virus NS2B-NS3 protease inhibitors via docking, molecular dynamics and consensus scoring-based virtual screening. J Mol Model 25(7)
77. Brady GP Jr, Stouten PF (2000) Fast prediction and visualization of protein binding pockets with PASS. J Comput Aided Mol Des 14:383–401
78. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242
79. Wang R, Fang X, Lu Y, Yang CY, Wang S (2005) The PDBbind database: methodologies and updates. J Med Chem 48(12):4111–4119
80. Szklarczyk D, Santos A, von Mering C, Jensen LJ, Bork P, Kuhn M (2016) STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. Nucleic Acids Res 44(D1):D380–D384
81. ChemDraw (2020) PerkinElmer Informatics. https://www.perkinelmer.com/category/chemdraw
82. MarvinSketch (2020) (version 20.9, calculation module developed by ChemAxon). http://www.chemaxon.com/products/marvin/marvinsketch
83. ACD/ChemSketch (2019) Advanced Chemistry Development, Inc., Toronto, On, Canada. www.acdlabs.com
84. Pymol (2020) Schrödinger, New York, NY. Version 2.3. https://pymol.org/2/
85. Maestro (2020) Schrödinger, LLC, New York, NY. https://www.schrodinger.com/maestro
86. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779–815
87. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER suite: protein structure and function prediction. Nat Methods 12(1):7–8
88. Zheng W, Zhang C, Wuyun Q, Pearce R, Li Y, Zhang Y (2019) LOMETS2: improved meta-threading server for fold-recognition and structure-based function annotation for distant-homology proteins. Nucleic Acids Res 47:W429–W436
89. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumieny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46:W296–W303
90. Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 32:W526–W531
91. Olsson MHN, Sondergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. J Chem Theory Comput 525–537
92. Ananda Krishnan R, Aguilar B, Onufriev AV (2012) H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulation. Nucleic Acids Res 40(W1):W537–W541
93. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 27:2875–2791
94. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612
95. Villar S, Cozza G, Moro S (2008) Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery. Curr Top Med Chem 8(18):1555–1572
96. Radoux CJ, Olsson TSG, Pitt WR, Groom CR, Blundell TL (2016) Identifying interactions that determine fragment binding at protein hotspots. J Med Chem 59:4314–4325
97. dos Santos KB, Guedes IA, Karl ALM, Dardenne L (2020) Highly flexible ligand docking: benchmarking of the DockThor program on the LEADS-PEP protein-peptide dataset. J Chem Inf, Model
98. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267:727–748
99. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multitasking. J Comput Chem 31:455–461
100. Korb O, Stützle T, Exner TE (2009) Empirical scoring functions for advanced protein-ligand docking with PLANTS. J Chem Inf Model 49(1):84–96
101. Allen WJ, Balias TE, Mukherjee S, Brozell SR, Moustakas DT, Therese Lang P, Case DA, Kuntz ID, Rizzo RC (2015) DOCK 6: impact of new features and current docking performance. J Comput Chem 36(15):1132–1156

102. Kramer B1, Rarey M, Lengauer T (1999) Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. Proteins 37(2):228–241

103. Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD (2015) Molecular docking and structure-based drug design strategies. Molecules 20