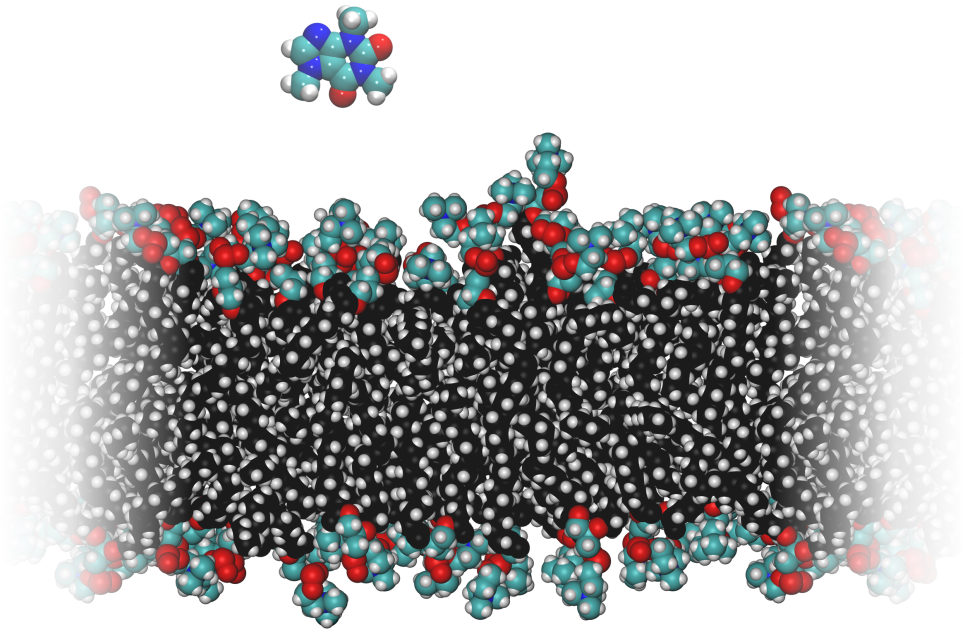


Development of a computational protocol to study passive diffusion through biological membranes

Rui Pedro Fernandes Ribeiro
Mestrado em Bioquímica
Departamento de Química e Bioquímica
2015

Orientador
Doutor Pedro Fernandes, Professor Associado com Agregação,
Faculdade de Ciências da Universidade do Porto

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

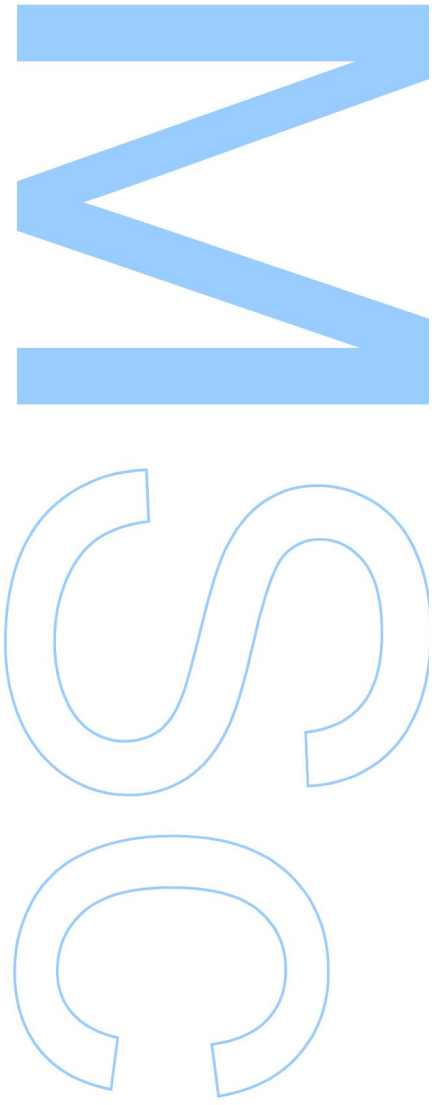
U.PORTO

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____



'Only those who will risk going too far can possibly find out how far one can go.'

T.S. Eliot

Acknowledgments

Agradecimentos

Ao Professor Doutor Pedro Fernandes, meu orientador, gostaria de agradecer a oportunidade de estagiar no seu grupo de investigação, a confiança e motivação depositada em mim, bem como os conhecimentos transmitidos e a sua enorme disponibilidade.

Ao Mestre João Coimbra agradeço pela amizade, simpatia e enorme paciência e pelo constante apoio na elaboração deste projeto. Quero agradecer-lhe por tudo o que aprendi durante estes largos meses: as técnicas, as dicas e os conselhos sábios que levo comigo e me ajudarão na minha carreira científica.

A todos os elementos do grupo de investigação agradeço pelo companheirismo e excelente ambiente de trabalho e os grandes momentos de descontração que se misturaram com o trabalho.

Aos meus amigos, por toda amizade que aqui é impossível descrever, e aos meus pais, avó, e restante família, por me acompanharem todos os dias da minha vida, deixo aqui o meu sentido agradecimento. Obrigado pelo apoio e força incondicionais que me motivam todos os dias a ser mais e melhor.

A todos vocês, obrigado por terem entrado na minha vida, me terem inspirado e iluminado com a vossa presença!

Abstract

During drug discovery and development, the routes of administration are normally chosen for practical reasons. Commonly, the most desirable strategy is oral administration, since it is more convenient, painless, and allows easy self-administration. Therefore, oral bioavailability is one of the most important properties in drug discovery and development. On one hand, a high oral bioavailability allows reducing the amount of an administered drug necessary to achieve a desired pharmacological effect, and therefore reducing the risk of side-effects and toxicity. On the other hand, the low oral bioavailability in clinical trials is a major reason for drug candidates failing to reach the market.

Because the absorption of an orally administered drug occurs mostly by passive diffusion, membrane permeation is recognized as a common requirement for oral bioavailability, and consequently, a number of eminent researchers, such as C. A. Lipinski and D. F. Veber, have defined rules that described the most relevant molecular properties required for a molecule to be bioavailable.

In this work, it was pretended, using a computational approach, to understand the conceptual and theoretical framework of Lipinski and Veber's rules, and to establish the molecular properties that are essential for the passive diffusion of compounds through cell membranes.

For this purpose, we started by selecting a small library of compounds, and their free energies in water and in 1-octanol were calculated in order to validate the force field that was being used. Next, the passive diffusion process for caffeine, nelfinavir, and atorvastatin in a DOPC/water membrane model were simulated, and Gibbs energy profiles were obtained. Lastly, nelfinavir and atorvastatin's rotatable bonds and PSA were characterized. During this stage the umbrella sampling protocol was carefully optimized, once it looks to be the most protocol suitable method for our studies.

Key words: drug discovery, molecular dynamics, potential of mean force, free energy profile, umbrella sampling.

Resumo

Durante a descoberta e desenvolvimento de um novo fármaco, a indústria farmacêutica dá preferência a uma formulação que possa ser administrada oralmente, uma vez que, para além de um fármaco administrado por via oral ser mais cómodo e indolor, permite uma autoadministração, o que aumenta a adesão à terapêutica.

Porém, a biodisponibilidade de um fármaco oral é muito mais reduzida do que a de um fármaco administrado por outras vias, tal como por via intravenosa, visto que os fármacos orais necessitam de atravessar membranas biológicas. Quando não há transportadores, os fármacos orais atravessam as membranas por difusão passiva, e alguns autores, como Lipinski e Veber, definiram regras que descrevem as propriedades moleculares necessárias para que uma molécula atravesse uma membrana biológica por difusão passiva.

Através de uma abordagem computacional, pretendeu-se neste projeto: compreender o fundamento molecular das regras de Lipinski e de Veber, e estabelecer, de forma clara, as propriedades moleculares que são fundamentais para a permeação de membranas celulares por moléculas de pequena dimensão, através de difusão passiva.

Para tal, numa primeira fase, foram estabelecidas as moléculas que serviram de base de estudo, bem como a parametrização e cálculo das suas energias livres de hidratação e solvatação (em octanol) para validar o campo de forças. Numa segunda fase, simulou-se o processo de difusão passiva de cada três moléculas em estudo: cafeína, nelfinavir e atorvastatina, tendo-se obtido perfis de energia de Gibbs para a sua difusão através de um modelo de membrana celular. Finalmente, numa terceira fase, procedeu-se à caracterização de ligações rotáveis e de área superficial polar para o nelfinavir e atorvastatina.

Palavras-chave: descoberta de fármacos, dinâmica molecular, perfil de energia livre, umbrella sampling.

Contents

List of figures

List of tables

Abbreviations

A

AMBER	Assisted Model Building with Energy Refinement

B

BAR	Bennett Acceptance Ratio

C

CPU	central processing unit

D

DOPC	1,2-di-(9E-octadecenoyl)-sn-glycero-3-phosphocholine

F

FDA	Food and Drug Administration
FF	Force Field

G

GAFF	General Amber Force Field
GPU	graphics processing unit
GROMACS	GROminogen Machine for Chemical Simulation
GSK	GlaxoSmithKline

L

LINCS	LINear Constraint Solver

M

MD	Molecular Dynamics
MM	Molecular Mechanics

N

NME	New Molecular Entities
NPT	isothermal-isobaric ensemble
NVT	canonic ensemble

P

PBC	Periodic boundary conditions
PME	Particle Mesh-Ewald method
PMF	Potential of Mean Force
PSA	Polar Surface Area

Q

QM	Quantum Mechanics

R

RESP	Restrained ElectroStatic Potential

S

SLIPID	Stockholm Lipids

W

WHAM	Weighted Histogram Analysis Method

CHAPTER 1

Introduction to Drug Discovery

'The process of drug discovery is not only of scientific interest, it entails a fascinating interplay among a variety of economic, social, and political institutions.' [1]
Donald Kennedy

Discovery and developing a new drug is a long, costly, and risky process, which can take 12-20 years from the original idea to generate a marketed product. Working together medicinal chemists, pharmacologists, biochemists, molecular biologists, computer scientists, medical doctors, and even lawyers and economists make of drug research an interdisciplinary science. When a new drug is planned, its route of administration is chosen carefully. Overwhelmingly, oral route is the most convenient; therefore, one of the most important goals for drug discovery and development is to understand the molecular properties that limit oral bioavailability. The pharmaceutical industry has been using computational methods to calculate those properties; however, such rules were based on simple and empirical observations, and they need to be understood if we want to refine and improve them.

1.1 A brief history of drug discovery

Drug research is not much older than a century, and it began in an era when chemistry had been capable of applying its principles and methods to problems outside chemistry itself, and when pharmacology had started to stand up as a scientific discipline. [2] In the beginning, organic chemists had the responsibility to create compounds, which

were designed individually and in large quantities, for being tested in live animals by pharmacologists. [3]

The history of drug research was shaped and enriched not only by new technologies and new discoveries [2], but also by the '*paradigm shift'* concept, which was introduced by Thomas Kuhn in his book '*The Structure of Scientific Revolutions*'. Applying Thomas Kuhn's concept to drug discovery means drug discovery was predominantly conducted by organic chemists in the beginning, but later, in the middle of the 20th century, chemists, biologists, and pharmacologists have started to dialogue and work together; furthermore, biochemists and molecular biologists have also joined them in order to help to bring drug discovery into a new level that had never been achieved before. [4]

Biochemistry brought into drug discovery the knowledge of enzymes, receptors, and the basic skills required for monitoring routes of drug metabolism, pharmacokinetic analysis and safety testing, whereas molecular biology brought the knowledge that allowed scientists to understand disease processes at the molecular and genetic levels, as well as to determine the optimum molecular targets for drug interference. [2]

Moreover, computer science and bioinformatics also influenced drug research in many ways; most of the drugs discovered from the end of the 20th century have been developed with the help of computational methods. [5]

In fact, drug discovery is no longer solely based on chemical synthesis, and is driven instead by an interdisciplinary knowledge.

1.2 The process of drug discovery

When people buy medicines, they hardly imagine developing a new medicine is a complex, long, costly, and a high risky process. According to *Pharmaceutical Research and Manufacturers of America*, *PhRMA*, the research and development process of a new drug takes more than 10 years from the starting project to the approval from the FDA (US Food and Drug Administration). In addition, not only the average cost to develop a new drug has been increasing over the years (with current costs approaching $2.6 billion, see Fig. 1), but also less than 12% of drug candidates (compounds with strong therapeutic potential, and whose activity and specificity have been optimized) will approved by the FDA.

Drug discovery is a multi-step process; see Fig. 2, beginning with the identification of a medical need, when there are no satisfactory medicines for a disease or clinical

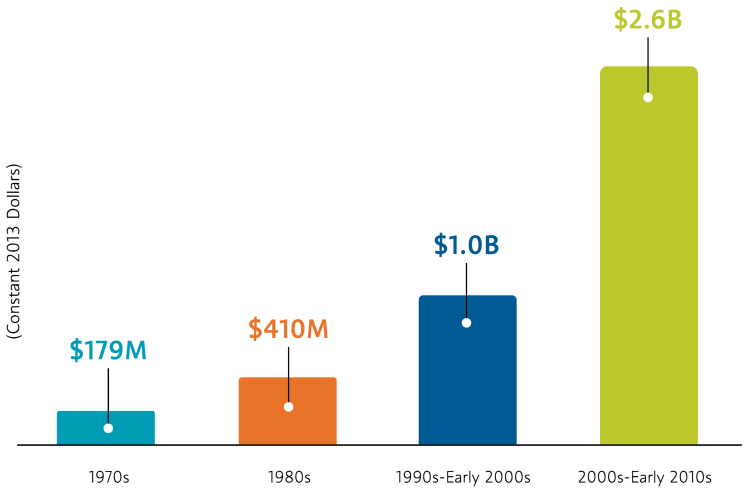


Fig. 1 - The increase in average cost to develop a new medicine (including the cost of failures) over the years. Adapted from *phrma.org*.

condition. Target identification and validation is one of the most important steps in the earlier phase of drug discovery; commonly, initial research starts in academia where the disease's mechanisms are established allowing researchers to select specific targets. Such potential biological targets need to be efficacious and safe, and they are normally called 'drugable targets', which means the target needs to bind a drug molecule with high affinity. [3, 6]

Once the target has been identified, it needs to be validated in order to show its relevance as a target in a disease process. Target validation is performed by multiple approaches, and it is based on several criteria: different gene expression in healthy and diseased tissue, proteomic studies, and studies in animal models on the disease-related phenotype. There are, nowadays, a range of techniques and methods available to perform these studies, which include: gene chips, neutralizing antibodies, knockout or transgenic mice, etc. [4, 6]

The next step in drug discovery consists of exposing the target to a large number of molecular compounds in *in vitro* and *in silico* assays. When a biologically active compound exceeds a certain activity threshold in a given assay it will be referred to as a HIT compound.

A variety of compound screening assays exist, and the most used today is the automated high-throughput screening, HTS, which involves the screening of several million molecules in an automatic manner. The use of HTS allows the pharmaceutical industry to reduce the time and the cost of HIT identification. [2, 3, 6]

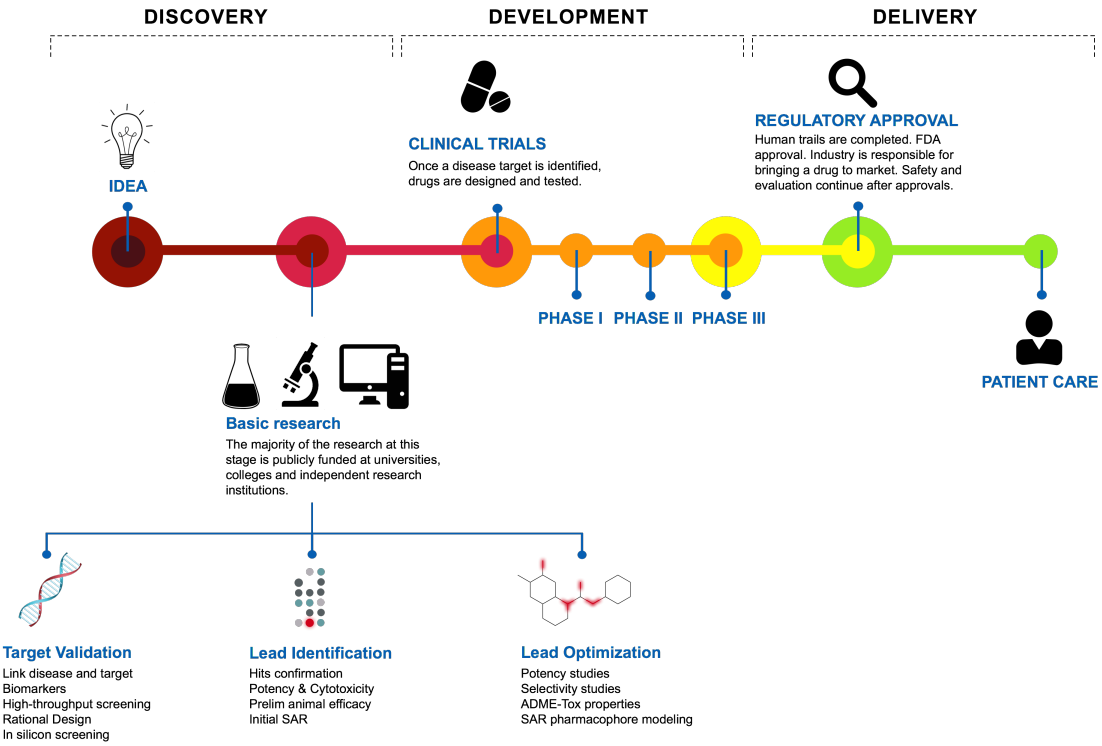


Fig. 2 - This infographic provides a summary of the sequential steps that are necessary for a drug to progress through the research and development pipelines. Adapted from *researchamerica.com*.

After HIT identification by HTS, those molecules are evaluated and undergo limited optimization strategies to find out the most promising compounds, ***lead*** compounds, to work with in more advanced stages of the drug discovery pipeline. Once a ***lead*** has been defined, it is important to understand precisely which structural features are responsible for its biological activity; typically, analog compounds are synthesized to explore structure-activity relationship (SAR) around each core compound structure, '*pharmacophore*', in an effort to maximize the desired activity. [3, 6, 7]

However, the HIT compound's pharmacokinetics must be validated through *in vivo* disease models, and this might be one of the most complex tasks to accomplish, because it mainly consist of evaluating several characteristics: absorption, distribution, metabolism, excretion, and toxicity. These characteristics are often abbreviated as ADMET.

After the complete process of lead optimization, the molecule is declared as a preclinical candidate, and the next phase will serve to determine whether a compound is suitable for human testing. If the compound fails the preclinical tests, researchers con-

tinue to synthesize and explore analogous compounds in order to produce back up molecules. [6]

Finally, before the FDA approval, the candidate medicine undergoes the clinical trials, a process that occurs in several phases:

- **Phase I (safety and dosage)** - the new drug is tested in a small group of healthy volunteers to determine the safety and its metabolites in the body;
- **Phase II (efficacy and side effects)** - the compound is tested in volunteer patients who have the diseases that are targeted by designed compound. In this phase the effectiveness and possible side effects are also determined;
- **Phase III (efficacy and monitoring of adverse effects)** - the compound is tested in a large group comprising thousands of volunteer patients, which allow generating statistically information about safety and efficacy.

To sum up, developing a new drug, from its original idea to generate a marketed product, is a complex process which can take at least 10 years, and it cost billions; moreover, approximately only 1 out of 20-25 drug candidates is approved, and eventually leads to a new marketed medicine. [*Phrma.org*]

Despite all the available technology, the insights provided by in genomic research and the growing resources and founding, the number of new molecular entities (NMEs), that is, a medication containing an active ingredient that has not been previously approved for marketing in any form, had been decreasing in the early 2000s; probably, due to several facts, such as clinical failures, increasing demands by regulatory authorities, and the disconnection between business and research goals. [3, 4, 8]

Fortunately, some new prospects are surfacing for the drug discovery process. In the last 5 years the number of NMEs approved by FDA has increased. More specifically, 51 new drugs were approved in 2014 (mostly for immunotherapy in cancer, diabetes, and hepatitis C); the highest number since 1996, see Fig. 3.

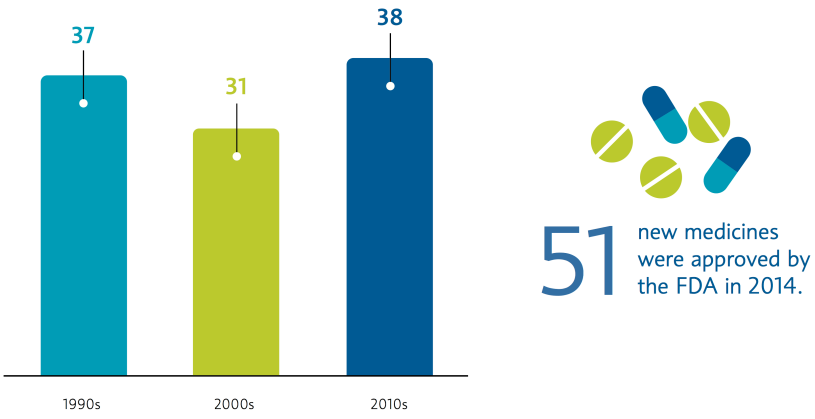


Fig. 3 - Average number of new medicines approved by the FDA per year- Adapted from *phrma.org*.

1.3 Computational drug discovery

The advent of computers and computational methods has changed all aspects of drug research. The reality is that the use of computers reduced the time, the cost, and the risk of failure of drug research by up to 50%, allowing potent HITs be obtained in a matter of weeks. [5, 9]

Different computational tools exist, and they can be used in various stages of drug discovery, from target identification to preclinical tests, which included: molecular docking, *de novo* design, quantitative structure-activity relationships (QSAR), pharmacophore modeling, similarity searching, sequence-based virtual screening, etc. [5, 9]

Moreover, algorithms to calculate molecular properties have been used in order to predict intestinal absorption and drugability, since oral ingestion is the most preferred route of drug administration, and these analyses avoid costly late-stage preclinical and clinical failures. [10]

1.4 Orally delivered drugs

In drug development, new drugs are designed and tested in a dosage form that is administered by a specific route. If the most promising drug is unable to reach its target organ at a concentration sufficient to have a therapeutic effect, it will fail in clinical trials; thus, a successful drug must be able to overcome physical, chemical, and biological barriers that limit the access of foreign substances to the body.

The routes of administration are normally chosen to take advantage of transport molecules that allow the drug to enter body tissues. From all the available routes, oral administration is usually the most desirable strategy, since it is more convenient, painless, and allows self-administration. However, an orally administrated drug must resist the digestive tract degradation, it must be stable during its absorption across the gastrointestinal track epithelium, and it must resist to the first passage effect.

All human cells and its organelles are surrounded by a lipid membrane, which is a dynamic and flexible structure consisting of phospholipids, glycolipids, and sterols. The membrane assumes a structure with a hydrophobic core and two hydrophilic surfaces due to the amphiphilic nature of membrane lipids. Biological membranes also contain a variety of proteins varying between 20 and 80% (w/w) that may span the membrane, or be exposed at the extracellular or intracellular membrane surface. [11-13]

Although the hydrophobic core of a biological membrane consists of a barrier to drug transport, there are several known mechanisms of intestinal drug absorption: pas-

sive diffusion (paracellular and transcellular transport), carrier-mediated transport, and active transport.

Normally, the absorption of orally administered drugs occurs in the duodenum and small intestine; mainly, by passive diffusion, that is driven by a concentration gradient, and can occur by i) paracellular transport, when molecules cross the membrane between cell junctions, or ii) by transcellular transport, in which molecules cross the cytoplasm. [11, 12, 14]

The type and rate of transport depends on molecules' properties; for instance, lipophilic drugs, such as steroids, can cross the lipid bilayer easily, whereas hydrophilic drugs, such as mannitol, cannot cross the membrane, but can pass through gaps between cells of the intestinal mucosa. [14] Moreover, the drug's acid dissociation constants (pK_a) and the pH gradient across the membrane may also affect drug's diffusion through biological membranes, and because of that, drugs only cross the lipid membrane in its neutral form; thus, weakly acids are better absorbed in the acidic environment of stomach, while weakly basic drugs are preferentially absorbed in duodenum, which has a basic environment. [15]

However, drug permeability cannot be accurately estimated based solely on physicochemical factors like pH, lipophilicity, or solubility. Over the past 15 years, many researchers have attempted to understand which relevant molecular properties directly influence oral the drug-like behavior of compounds. Lipinski *et al.* were the first to point out trends between physicochemical and structural properties, and the absorption of known drugs, when he came out with the so-called "rule of 5". Lipinski's rules were derived from a statistical analysis of a selected library of 2.245 compounds with superior physicochemical absorption properties that were believed to have entered phase II clinical trials. On the whole, the 'rule of 5' associates four parameters with drugs' permeability: molecular weight, logP, the number of H-bond donors, and the number of H-bond acceptors. [16-18]

Veber *et al.* have brought up another important contribution into this field with the analysis of trends in the properties of 1.100 drug candidates studied at GlaxoSmithKline, GSK. Veber's rules correlate drug's low-bioavailability with high polarity, expressed as polar surface area or total hydrogen bond count, and high flexibility, expressed as measured by the number of rotatable bonds.

Apart from Lipinski and Veber, other researchers have carried out analysis of the properties of marketed drugs and compounds under clinical investigation, and a thorough summary of their works can be found in a review by Lajiness. [19]

1.5 Molecular Properties

1.5.1 Molecular Weight

One of the most important physicochemical parameters that has long been thought to disfavor compound's permeability is molecular weight, which has been increasing over the years, see Fig. 4. [8, 20] Moreover, it was demonstrated drug candidate leads tend to be heavier than the corresponding lead. [21]

Lipinski *et al.* have shown molecular weights of compounds, in their selected library, were lower than those in the whole *World Drug Index* (WDI) data set, where his library was selected from, concluding good absorption or permeation are more likely when molecular weight is under 500 Da. [17] Apart from good permeation, a low molecular weight reduces drug's clearance, because such molecules are less likely to be conjugated, particularly with glucuronate, and eliminated by biliary excretion, which increase its bioavailability. [20]

This rule seems to work very well for most cases; however, there are a few exceptions: renin inhibitors and HIV protease inhibitors are part of the bulk of drugs currently marked with molecular weights above 500 Da, see Fig. 5. Such out-of-rage behavior may result in part from conformation flexibility. [20]

1.5.2 Octanol/Water partition coefficient - LogP

Passive diffusion is highly influenced by lipophilicity, as previously stated. The lipophilicity of a substance can be quantified by the octanol/water partition coefficient (logP), which corresponds to the negative logarithm of the ratio of the concentration of the substance between two solvents, in the hydrophobic and aqueous phases, respectively. The predictions of logP have been studied for more than 40 years by computational methods; therefore, huge databases of logP values are available. [8] The computational methodologies to work logP values out will be described later.

A lower logP (hydrophilic) limits drug's absorption and permeation; however, it has been reported that highly lipophilic compounds tend to have a higher incidence of non-specific toxicity. [8, 13] Actually, according to Lipinski *et al.*, only about 10% of compounds in their selected library have a logP over 5. [17] In addition, they reported an increase in logP values of synthesized compounds in drug discovery programs at *Pfizer*, just as the molecular weight trend. [8, 16] As a matter of fact, attention has been given to changes in logP values of drugs over time; see Fig. 4. [8]

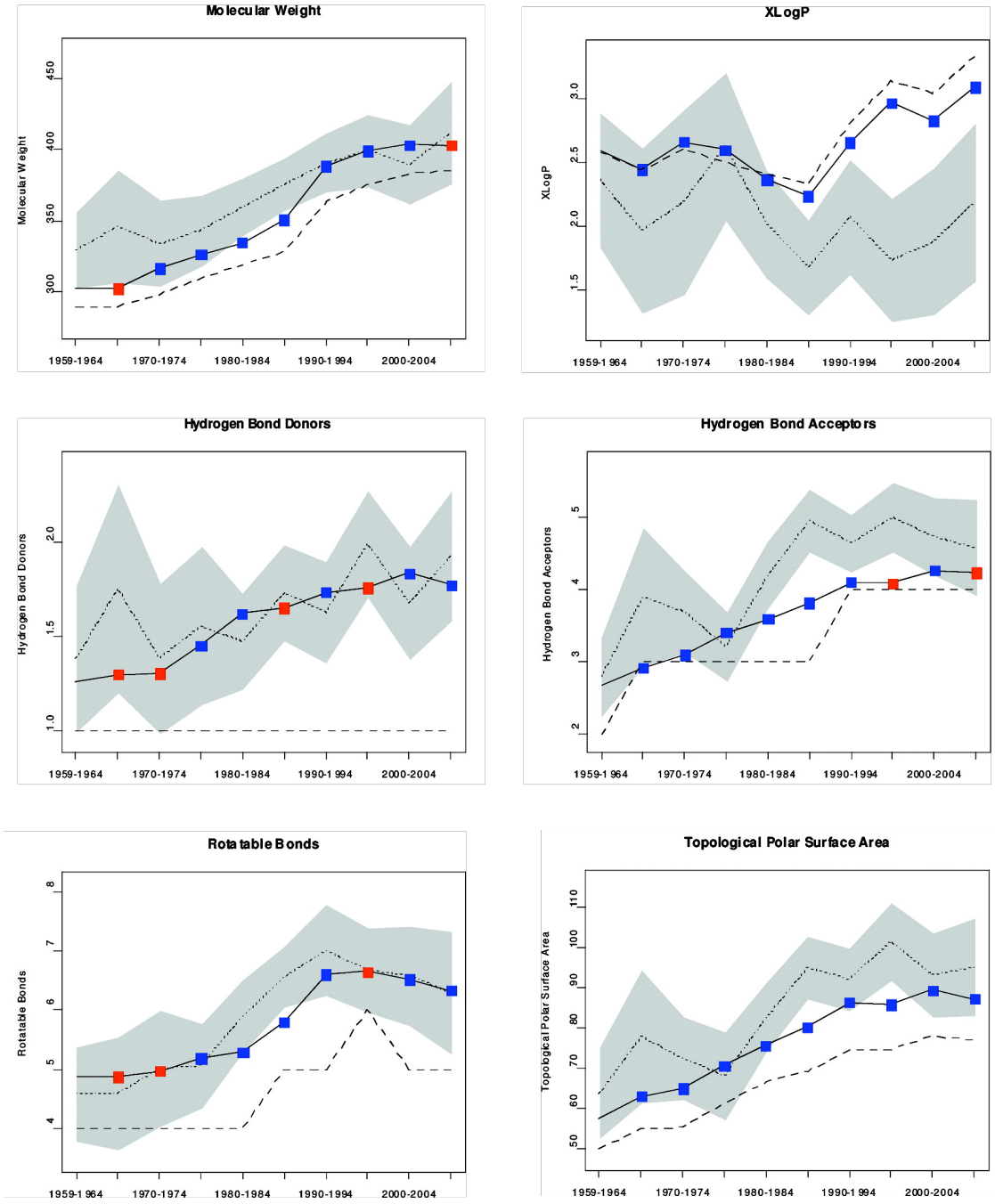


Fig. 4 - Mean values for molecular weight, LogP, number of hydrogen bond donors and acceptors, number of rotatable bonds, and polar surface area for molecules published in Journal of Medicinal Chemistry in 1959-2009 are shown as solid lines with squares. Blue squares indicate a statistically significant difference between the mean for a particular 5-year interval and the preceding interval. Red squares indicate that the difference was not statistically significant. Dashed lines indicate the median value for each property during the designated interval. The gray regions represent the 95% confidence interval for drugs launched during the same time period, with the dotted line representing the mean value for drugs during each 5-year period. Adapted from Walters et al.[8]

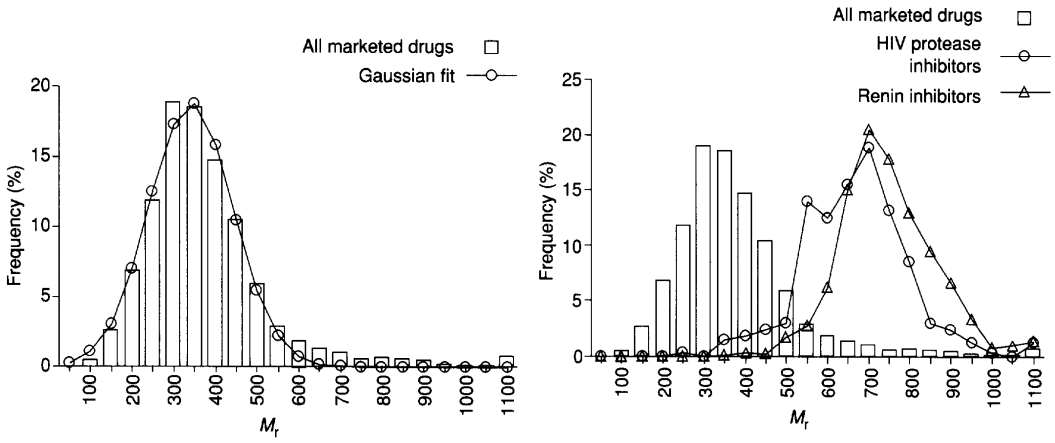


Fig. 5 - Distributions for compounds of therapeutic interest at earlier stages of development. (left) Mr distribution of a set of marketed drugs. (right) Certain classes of compounds, including the renin and HIV protease inhibitors, fall outside the Gaussian region of the distribution of Mr Values. Adapted from Navia et al. [20]

The big issue in lipophilicity is the interaction between a drug and the hydrophobic bilayer core: hydrophilic compounds have weak interactions with the bilayer's hydrophobic core, whereas hydrophobic compounds have strong interactions, and tend to pass the membrane easier than hydrophilic compounds.

1.5.3 Hydrogen Bonding Groups

Hydrogen bond count is a parameter that directly influence molecule's lipophilicity and structure; consequently, its permeability. The total number of hydrogen bonds is the sum of hydrogen donors and hydrogen acceptors, which are defined as any heteroatom with at least one bonded hydrogen, and any heteroatom without a positive formal charge, respectively. [22]

An excessive number of hydrogen donors or hydrogen acceptors hinders molecules' permeability. Lipinski *et al.* noted 92% of compounds in their selected library had 5 or fewer hydrogen bond donors, and 88% of compounds had 10 or less hydrogen acceptors, [17] whereas Veber *et al.*, in their turn, reported a sum of 12 or less hydrogen bond donors and acceptors for an ideal oral drug. [22]

In addition, the mean number of hydrogen bonds had increased for 30 years following 1960, but it appears to have stabilized since then, see Fig. 4. [8]

1.5.4 Rotatable Bonds

Molecular flexibility is directly related to rotatable bonds, which can be defined as any single non-ring bond attached to a non-terminal heavy-atom. The amide C-N bonds are not counted because they have peptide-like bond characteristics due their high rotational energy transition barrier.

In medicinal chemistry, flexible molecules are often rigidified in order to reduce its conformational entropy, which allows an increase in its binding affinity and selectivity. When molecules are flexible there is a loss of entropy on binding or on membrane crossing due to conformational restriction; for this reason, if a molecule is rigidified, less entropy will be lost

According to Veber's analysis, for a rat oral bioavailability higher than 20-40% the number of rotatable bond needs to be 12 or less. [22] An increase in the average number of rotatable bonds over time was also noted; see Fig. 4. [8]

1.5.5 Polar Surface Area

The polar surface area, PSA, is defined as the area occupied by nitrogen, oxygen, phosphorus, and sulfur atoms, and hydrogen atoms attached to these heteroatoms. [23] Generally, calculated PSAs are correlated closely with the total hydrogen bond count; so, it has been reported by many authors a correlation between PSA and drug's absorption. [24, 25]

Palm *et al.* found drugs with a PSA under 63 $Å^2$ were highly absorbed (>90%), while drugs with a PSA greater than 139 $Å^2$ were poorly absorbed (<10%). [24]

Subsequently, Veber *et al.* observed a PSA lower than or equal to 140 $Å^2$ for molecules with rat oral bioavailability higher than 20-40%. [22]

A consistent increase in both the mean and median PSA over time was observed, once again, see Fig. 4. [8]

1.5.6 Lipinski and Veber's rules implementation

To sum up, at this point, five parameters were described associated with drugs' solubility and permeability: molecular weight, octanol/water partition coefficient (logP), the number of hydrogen bonds, rotatable bonds, and polar surface area (PSA).

Lipinski's "rule of 5" states a high probability of poor oral bioavailability for a molecule when its molecular weight is over than 500 Da, logP is over 5, the number of hydrogen bond donors is more than 5, and the number of hydrogen bond acceptors is

more than 10. Furthermore, compounds classes that are substrates of biological transporters are exceptions to the rule.[17]

Conversely, Veber's rules described only two criteria for drugs with high rat oral bioavailability: 10 or fewer rotatable bonds, and polar surface area less than or equal to 140 $Å^2$, or 12 or fewer hydrogen bond donors and acceptors. [22]

Although Lipinski and Veber's rules have had a huge impact on how the pharmaceutical industry evaluates drug's behavior, caution should be taken. Despite its success, the rule of 5 and Veber's rules are derived from simple statistical analysis. In addition, observations made by Veber and his co-workers were based on an internal dataset from GSK; consequently, other datasets should be evaluated, and those rules should be refined and improved. [19]

In any case, pharmaceutical industry has been using computational methods to calculate molecular properties stated by Lipinski and Veber's rules in order to predict intestinal absorption before investment, because they are often quicker and cheaper than the current experimental methods.

1.6 What is this thesis about?

Once drugs' permeability is one of the major factors to decide whether a drug candidate could make the development continue, it appeared reasonable to attempt to understand the conceptual and theoretical framework of Lipinski and Veber's rules; as well as to clearly establish the molecular properties that are essential for the passive diffusion of compounds through cell membranes. In addition, new and more predictive rules than current ones would be interesting to establish, in order to anticipate the prediction of drugs candidates' diffusion properties.

In this regard, a few oral drugs, for which chemical properties could be extracted from the literature, and for which our group has a vast experience, were selected, and their passive diffusion's process were simulated. Not only did we obtain a complete structural characterization of the permeation process (including relevant interaction and drugs' orientation through the bilayer), but also Gibbs free energy profiles for the diffusion through cellular membrane models.

CHAPTER 2

Computational Methodology

'So in short, what we develop is a way, which require a computer, to take the structure of a protein and then to eventually understand how exactly it does, what it does… you could use it, for example, to design drugs or just, like in my case, to satisfy your curiosity.'
Arieh Warshel,
Nobel Prize in Chemistry, 2013

Computational chemistry can be used to investigate the energetics associated with changes in both conformation and chemical structure. Recent advances in computational techniques, that make use of quantum mechanics (QM) or molecular mechanics (MM), now allow for a detailed understanding of the microscopic events, at the atomic level, of chemical or biochemical processes. MM uses empirical energy functions, which include relatively simple terms to describe the physical interactions that dictate the structure and dynamic properties of molecules. On the other hand, applying other computational techniques, macroscopic events can also be understood. Molecular dynamics (MD) methods, for instance, calculates the time dependent behavior of a molecular system, offering a great insight on how the systems works.

2.1 Introduction

Computational chemistry has emerged as a subfield of theoretical chemistry where mathematical methods are combined with fundamental laws of physics to solve chemically related problems. The real strength of computational chemistry is the ability to generate data, and thereby, predict structures and energetics of chemical systems, including large-scale biological systems. In applying theoretical ideas using computers,

computational chemistry can help us to better understand the dynamic of chemical and biochemical events, some of them difficult to interpret until now.

Currently, there are three popular computational classes of methods to approach hypothetical chemistry problems: i) quantum mechanics (QM), ii) molecular mechanics (MM), and iii) hybrid quantum mechanics/molecular mechanics (QM/MM) methods. Quantum chemistry is based on wave function determination by *ab initio* and semi-empirical numerical techniques, or based on the electronic density-by-density functional theory (DFT); QM is primarily concerned with the numerical computation of molecular electronic structures. *Ab initio* (Latin term for 'from the beginning') methods are derived directly from theoretical principles, with no inclusion of experimental data. They are based on wave function and try to resolve the non-relativist Schrödinger's equation to get the most accurate results for any kind of system; however, since they take enormous amounts of computer CPU time, memory, and disk space, they can be only used for small systems (typically less than one hundred atoms). Semi-empirical techniques are based on the Hartree-Fock's formalism, but use approximations from empirical or experimental data to provide input into the mathematical models. The good side of semi-empirical calculations is that they are much faster than the *ab initio* calculations, and allow studying larger systems than *ab initio* methods (typically up to 1 thousand atoms). In its turn, DFT methods are based on electronic density, and they are designed to determine the properties of a many-electron system by using density functionals. They are less computational demanding and less accurate than *ab initio* methods of the same accuracy, but more demanding and accurate than semi-empirical techniques. [26] Similarly to *ab initio* methods, only small systems (less than three hundred atoms) can be studied by DFT methods.

When a system is too large to be studied by *ab initio*, semi-empirical, or DFT methods, it is still possible to simulate its behavior by Molecular Mechanics (MM), that uses the laws of classic mechanics and a force field (FF) based on empirical parameters to explain and interpret the atoms and molecules' behavior. [26]

The work described in this dissertation is based on molecular mechanics; hence, in this thesis we will focus only on the description of MM-based notions and techniques, and not on quantum, semi-empirical or hybrid approaches. Indeed, the systems we purpose to study are too large and complex (water/membrane/drug) to be treated by QM-based approaches. However, quantum methods were used in systems' parameterization, especially to ascribe atomic point-charges to molecules. For this reason, only a small description of the QM-based procedure will be given in section 3.1.

2.2 Molecular Mechanics

Many of the chemical or biochemical systems are too large to be considered by QM-based methods, as was stated before. Fortunately, molecular mechanics can be applied to large-scale systems because it is a purely empirical method, which ignores the electronic degrees of freedom, and calculates the energy of a system as a function of the nuclear positions only. This conception greatly reduces the computational burden of an MM-based calculation. Although, in some cases, molecular mechanics can estimate physical-chemical quantities as accurately as quantum mechanical calculations, it cannot provide properties that depend upon changes in the electronic distribution in a molecule. [26]

The use of MM to study chemical or biochemical systems is possible due to the Born-Oppenheimer's approximation, which assumes the motion of atomic nuclei and electrons in a molecule can be separated; therefore, is possible to consider energy as a function of the nuclear coordinates, which are stated by the laws of classical mechanics. As a result, each atom is simulated as a single spherical particle, which is assigned a radius (typically the van der Waals radius), and a constant net charge, generally derived from *ab initio* methods. Moreover, bond interactions like stretching bonds and angle bending are treated by harmonic potentials with an equilibrium distance equal to the experimental or calculated bond length, see Fig. 6. [26]

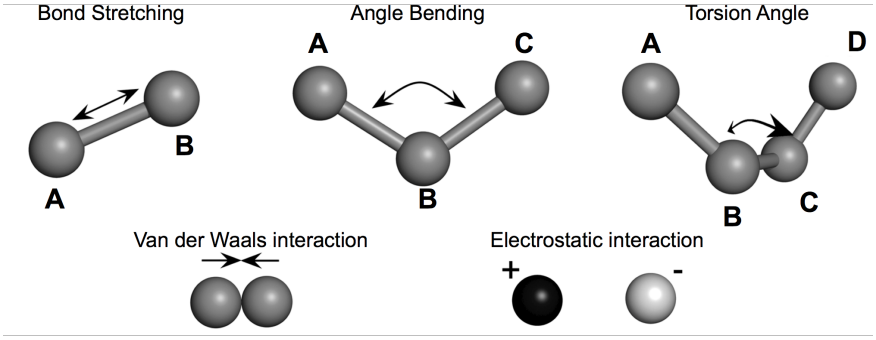


Fig. 6 - Schematic illustration of the most important interactions that contribute to a molecular mechanics force field: bond stretching, angle bending, torsional terms, and van der Waals and electrostatics interactions.

Essentially, a force field (FF) refers to an equation that uses those simple terms and parameters from experiment or more accurate methods, to calculate the energy of a molecule. Different force fields can be found in literature, and each one is specific for different kinds of molecules; however, most of them share the same mathematical terms, see equation 1:

$$V_t = V_{bond} + V_\theta + V_\omega + V_{vdw} + V_{el} + V_{cross} \qquad (1)$$

The first three terms, V_{bond}, V_θ, and V_ω describe the energy associated to the bond stretching between two atoms, the angle bending between three atoms, and the torsion energy for dihedral angles, respectively. These terms are only used to describe covalently connected atoms. The fourth and fifth terms, V_{vdw} and V_{el} are descriptive of through-space interactions through non-covalent terms of van der Waals and electrostatics interactions, respectively. Finally, the last term, V_{cross}, is related to the correlation of the bond-terms (V_{bond}, V_θ, and V_ω); however it is not often included in FFs because it is computational demanding. Indeed, V_{cross} only was found to be important in FFs designed to predict vibrational spectra, and it is only included in a FF when an optimal performance must be achieved. [26]

Force field applications have been extensively based on force field parameters' transferability. The term 'transferable' implies that a set of parameters developed and tested on a relatively small number of cases should be transferable between different molecules with similar chemical structure, and could be applied to a much wider range of problems. Moreover, not only the force field parameters should be transferable between different molecules, but also it should be also applicable in systems in different physical conditions like pressure, temperature, or composition. In addition, a transferable force field should be properly validated for different thermodynamic properties. [26, 27]

The successful application of computational methods depending on the accuracy of the force field model has long been recognized. The most commonly used FFs for biomolecular simulations are those included in the AMBER (Assisted Model Building with Energy Refinement), CHARMM (Chemistry at HARvard Macromolecular Mechanics), and GROMOS (GROningen Molecular Simulation) bundles. The AMBER FFs were initially designed for simulation of biomolecules like proteins, nucleic acids, and even carbohydrates. The first AMBER FF version used an 'united-atom' model; nonetheless, the latest versions make use of 'all-atom' representations, in which all the atoms in the

system, including hydrogen's, are parameterized. Similarly, CHARMM FFs also use 'all-atom' representations, but they are often seen as more flexible than AMBER FFs, since they were developed for a vast range of molecules, from small molecules to complex and solvated biological macromolecules. In its turn, GROMOS FFs are taking hold, once they are not only as flexible as CHARMM FFs, but also faster. [28-30]

One of the FFs used in this work, which is compatible with the AMBER FFs, is a more generic force field for organic molecules, especially designed for the automated analysis of diverse libraries of a variety of organic and pharmaceutical molecules, called GAFF (General AMBER force field). The equation 2 is the potential energy function of GAFF, and each term will be described in next subchapter. [30, 31]

$$V_t = \sum_{bonds} k_1(l - l_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 +$$
$$\sum_{torsions} k_\phi[\cos(n\omega - \delta) + 1] + \sum_{nonbonds}\left(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\varepsilon_{ij} r_{ij}}\right) \quad (2)$$

Although CHARMM and GAFF include explicit hydrogen atoms, they have had problems to describe lipid bilayers in the isothermal−isobaric ensemble (NPT), which is the desired ensemble to perform simulations of lipid bilayers. Lipid molecules have an amphiphilic nature, and it has been difficult to study lipid bilayers in full atomistic detail, because if no surface tension was applied, or if the area of the membrane plane was not kept constant, the lipid bilayers often ended up in the gel phase instead of the expected fluid phase.

In order to do reliable and realistic modeling of lipid bilayers, Jämbeck et al. have systematically derived and validated a force field, which they called Slipids (Stockholm lipids), which can reproduce a number of properties measured experimentally. Slipids includes parameters, for example, for unsaturated phosphatidylcholine and phosphatidylethanolamine lipids, sphingomyelin, phosphatidylglycerol, phosphatidylserine lipids, and cholesterol. [32-34]

Furthermore, others FFs for more specific situations exist, like the MM (Molecular Mechanics) series (MM1, MM2, MM3, MM4) and OPLS (Optimized Potential for Liquid Simulations), which are often used in organic solvents' simulations. [26]

2.3 Force Field's Parameters

In this point, the individual terms of molecular mechanics force field it will be discussed in some detail, giving a selection of the various functional forms that are in common use.

2.3.1 Bond stretching

The potential energy curve for a typical bond is well described by the Morse potential, which has the form of equation 3:

$$V_l = D_e(1 - e^{-a(l-l_0)})^2 \quad , \tag{3}$$

where l is the bond length, D_e is the depth of the potential energy minimum, and $a = \omega\sqrt{\mu/2D_e}$ where μ is the reduced mass, and ω is the frequency of a particular bond's vibration.

Although the Morse potential is very accurate in describing a wide range of behaviors, from the strong equilibrium behavior to dissociation, is also computationally demanding; consequently, simpler expressions, like Hook's law formula, are often used. Therefore, bond stretching is described as a harmonic potential, in which the energy varies with the square of the displacement from the reference bond length l_0, as displayed in equation 4:

$$V_l = \frac{k}{2}(l - l_0)^2 \quad , \tag{4}$$

in which l_0 corresponds to the equilibrium bond length, the value that the bond adopts when all others terms in the force field are set to zero, and k is the force constant.

The use of Hook's law is possible because is a reasonable approximation to the shape of the potential energy curve at the bottom of the Morse potential, at the distances that correspond to bonding in ground state molecules, see Fig. 7.

However, the harmonic potential is less accurate in regions far from equilibrium, and does not describe the best chemical behavior if bonds stretch are too far from the reference length. To overcome this problem, cubic and higher terms can be included and the bond stretching potential can be written as a truncation form from the Taylor expansion, equation 5, where V_{l_0} and $\frac{\partial V_l}{\partial l}$ are zero.

$$V_l = V_{l_0} + \frac{\partial V_l}{\partial l}(l - l_0) + \frac{1}{2!}\frac{\partial^2 V_l}{\partial l}(l - l_0)^2 + \frac{1}{3!}\frac{\partial^3 V_l}{\partial l}(l - l_0)^3 \ldots \tag{5}$$

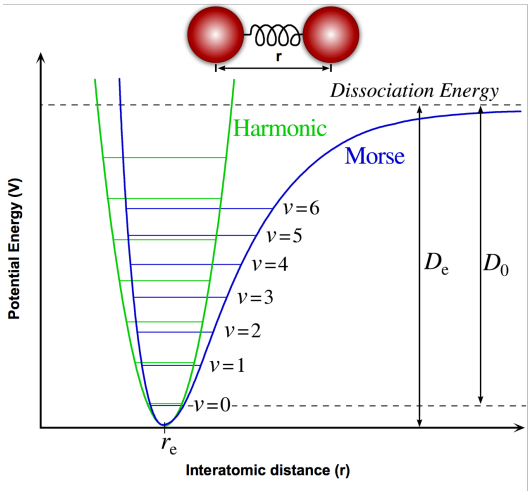


Fig. 7 - Comparison of the simple harmonic potential (Hook's law) with the Morse potential.

2.3.2 Angle bending

The angles' deviation from its reference value is also frequently described using the Hook's law or harmonic potential, like bond stretching, by:

$$V_\theta = \frac{k}{2}(\theta - \theta_0)^2 \quad , \tag{6}$$

were θ represents the angle between the atoms A, B, and C (see Fig. 7), θ_0 is the reference value for the same angle, and k is the force constant.

Once more, the accuracy of the force field can be improved using higher-order terms:

$$V_\theta = V_{\theta_0} + \frac{\partial V_\theta}{\partial \theta}(\theta - \theta_0) + \frac{1}{2!}\frac{\partial^2 V_\theta}{\partial \theta}(\theta - \theta_0)^2 + \frac{1}{3!}\frac{\partial^3 V_\theta}{\partial \theta}(\theta - \theta_0)^3 \ldots , \tag{7}$$

where V_{l_0} and $\frac{\partial V_l}{\partial l}$ are zero.

Nevertheless, when biological systems are being studied the use of harmonic terms for bond stretching and angle bending is sufficient, not only because higher-order terms required more computational power, but also because such simulations are performed in the vicinity of room temperature; typically, at room temperature bonds and angles stay close to their equilibrium. [28]

2.3.3 Dihedral torsion

The torsion angle or dihedral angle (ω) is the angle between the two planes defined by the atoms ABC and BCD, see Fig. 6. The angle ω can vary between [0º, 360º] or [-180º, +180º], and its energy is given by equation 8:

$$V_\omega = \frac{1}{2}\sum_i k_{i_{ABCD}}[1 + \cos(n_i\omega_i - \gamma_i)] \quad , \qquad (8)$$

where ω represents the torsion angle between AB and CD bonds, k_{ABCD} is the force constant associated with the rotation, n is the multiplicity (the number of minima of the function as the bond is rotated through 360º), γ is the phase that defines the minimums' position, and i the number of dihedral angles of the system. The torsional energy variation as a function of the dihedral angle for the butane molecule can be observed in Fig. 8.

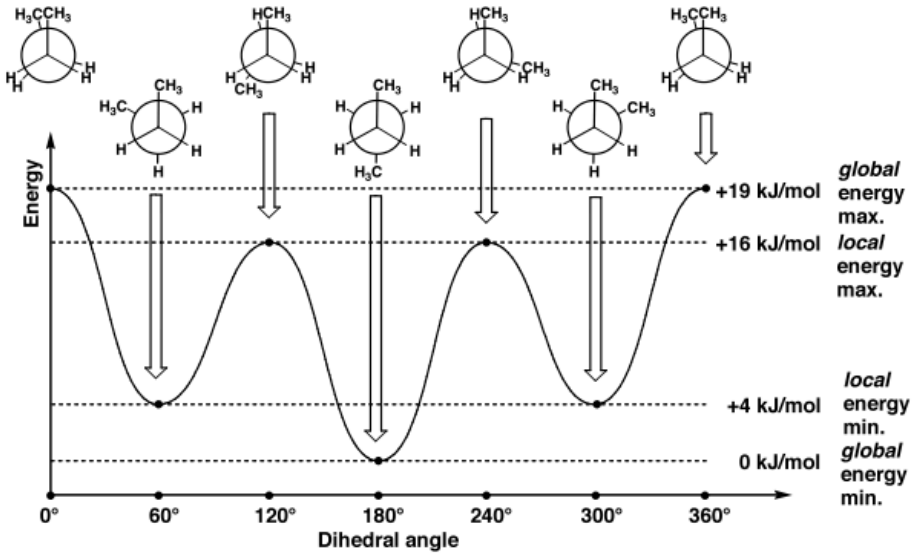


Fig. 8 - Torsional energy as a function of the rotation of the dihedral angle for the butane molecule. A schematic illustration of the occurrence of the periodic structures, depending on the torsional angle is also presented.

2.3.4 Van der Waals energy

The van der Waals energy, V_{vdw}, is described as the sum of attractive and repulsive forces between non-covalently interacting atoms. The attractive forces, also called dispersive forces, or London forces, are relevant only at short and medium ranges, and V_{vdw} tends quickly to zero for high inter-atomic distances. On the other hand, the repul-

sive forces or Pauli's repulsion are very short-range, and they are higher when two atoms come very close, and there is overlapping of their electronic densities (Pauli exclusion principle). Moreover, for intermediate distances there are a certain attraction between electronic clouds of the atoms, promoted by instant dipole - induced dipole interactions, see Fig. 9.

The van der Waals energy term can be mathematically described by Lennard-Jones potential for a pair of atoms *i* and *j*,

$$V_{vdw}(i,j) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad , \qquad (9)$$

where $-4\varepsilon_{ij}\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}$ is the attractive term, $4\varepsilon_{ij}\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}$ is repulsive term, ε_{ij} corresponds to the energy minima of the function, σ is the finite distance at which the inter-particle potential is zero, and *r* is the distance between the particles.

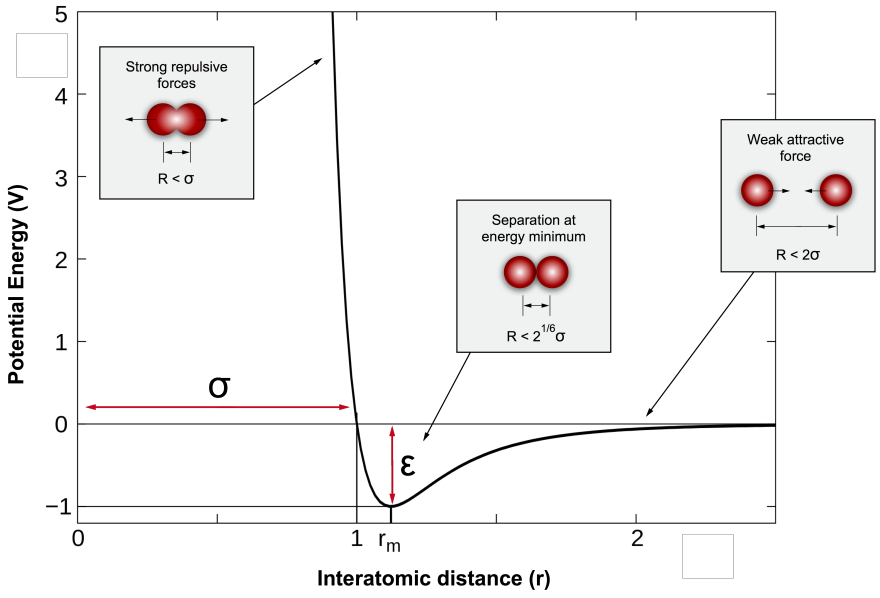


Fig. 9 - Graphical representation of the Lennard-Jones potential, which describes attraction and repulsion between two particles. σ is the interatomic distance for which the potential energy is zero. The deeper the well depth (*ε*), the stronger the interaction between the two particles.

2.3.5 Electrostatic energy

The electrostatic energy, V_{el}, describes the non-covalently bonded interactions arising from the presence of permanent atomic charges, and its value is obtained according to Coulomb's law, equation 10 and Fig. 10:

$$V_{el}(i,j) = \frac{1}{4\pi\varepsilon}\frac{q_i q_j}{r_{ij}} \quad , \qquad (10)$$

in which q_i and q_j represent point charges of the atoms i and j, ε is the dielectric constant of the medium, and r_{ij} the distance between atoms.

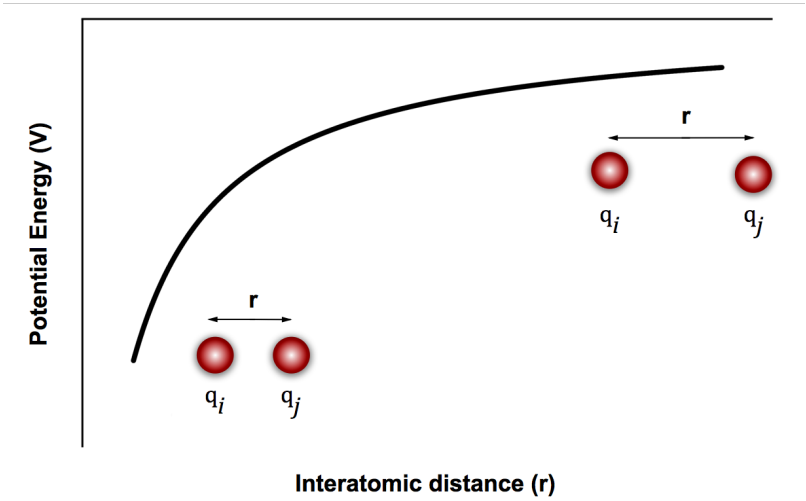


Fig. 10 - Graphical representation of the Coulomb potential of particles with opposite charges.

Differently from the other terms, V_{el} is not transferable between systems, because it is more variable, and cannot be reliably determined based on the type of atom, since point-charges greatly depend on the environment. [35] As a consequence, different methods to calculate V_{el} have been reported in literature, like the Mulliken population analysis, the CHarges from Electrostatic Potentials using a Grid-based (CHELPG) method, the Restrained ElectroStatic Potential (RESP) method, Merz-Kollman (MK) method, etc. [26]

2.4 Molecular Dynamics

Molecular dynamics simulation (MD) is one of the principal tools in theoretical studies of biological molecules. This computational method simulates the dynamic behavior of a molecular system, from which detailed information on molecular properties can be calculated.

The first MD simulation resulted from the pioneering applications to the dynamics of liquids by Alder and Wainwright, in the late 1950s. Due to the revolutionary advances in computer technology and algorithmic improvements, MD has subsequently become a valuable tool in many areas of physics and chemistry, and MD methods are, nowadays, routinely used to investigate the structure, dynamics, and thermodynamics of biological molecules and their complexes. [36]

In MD simulations, the time evolution of a set of atomic positions are derived in sequence by applying Newton's equations of motion, equation 11, where for a particle i, r_i, is its position vector, F_i, is the force acting upon i at time t, and m_i is the particle's mass.

$$F_i = m_i \frac{\partial^2 r_i(t)}{\partial t^2} \qquad (11)$$

Particles move in straight lines at constant acceleration, and after a time step, the new velocities, positions, and the instantaneous forces acting on particles are calculated by integrating equation 11 using the principle of conservation of linear momentum. However, due to the many-body nature of the problem, the equations of motion are discretized and solved numerically; therefore, the positions and velocities are propagated with a finite time interval using numerical integrators. That is to say, the equations of motion are integrated by breaking the calculation into a series of very short time steps. At each step, the forces on the atoms are computed and combined with the current positions and velocities to generate new positions after the time interval. The resultant force, which acts on each atom, is also calculated according to the new atomic positions, and this process is repeated until a desirable number of microstates are computed. In this way, MD generates a trajectory that describes how coordinates and momenta change with time, and the dynamic properties and events, which may influence the functional properties of the system, can be directly traced at the atomic level making MD especially valuable in molecular biology. [26, 36]

Simulation's trajectory is nothing but a set of microstates sharing the same thermodynamic properties, which is known as an ensemble. An ensemble is defined according to thermodynamic state functions that have values explicitly defined and fixed, and different ensembles exist: the canonic ensemble (NVT), where the number of particles (N), volume (V), and temperature (T) are fixed; the isothermal-isobaric ensemble (NPT), in which N, T, and pressure (P) are fixed; the microcanonic ensemble (NVE), with fixed N, V, and total energy (E); and the big canonic (μVT) in which the V, T, and the chemical potential (μ) are also fixed. In MD, ensemble's choice depends on what we want to simulate. If we want to emulate a physiological condition, NPT would be the most realistic choice, since pressure and temperature of a closed system tend to be constant due to atmospheric pressure and physiological/room temperature of the surroundings. The choice for the integration time interval (Δt) is also essential for any molecular dynamics simulation, because a lower value for Δt results in a more accurate simulation; unfortunately, a lower value for Δt also results in a higher computational demand. Typically, the chosen value for Δt corresponds approximately to one-tenth the time of the shortest period of motion, or smaller. In flexible molecules, the highest-frequency vibrations are due to bond stretches containing a hydrogen atom. For instance, a C-H bond vibrates with a frequency of approximately 10 fs; therefore, a value of 1 fs for Δt is commonly used. However, these high-frequency motions are usually of relatively little interest, and have a minimal effect on the overall system's behavior. In order to decrease the computational burden, such vibrations can be restraint to their equilibrium value using algorithms such as SHAKE or LINCS, and a 2 fs value for Δt can be used without decreasing the simulation's accuracy. [26]

To perform realistic simulations, it has to be accounted that an aqueous solution is the typical environment for biological macromolecules, and an atomically detailed representation of the solvent (explicit solvent) is preferably used. However, because of limited computer memory and also to speed up the calculations, only a finite sample of an extended (infinite) system can be represented explicitly in a computer model. Thus, due to the finite size and surface effects of the system, caution should be taken in non-bonding interactions' calculation.

Short-range forces (van der Waals interactions) are described according to Lennard-Jones terms, which decay proportionally to r^{-6} at longer distances, whereas long-range forces (electrostatic forces) vary with the distance proportional to r^{-1}. Long-range forces are treated through periodic boundary conditions (PBC), which are imposed on

a system to deal with its finite size and surface effects, or the Particle Mesh-Ewald method (PME).

PBC method allows representing simulation's cell as single cell in an infinite network of its replicas, see Fig. 11, which needs to be large enough for molecules to not interact significantly with their replicas.

Finally, PME method has been considered the best method to determine long-range interactions in periodic systems. Essentially, the electrostatic problem is solved in two parts, and uses also a cut-off value as the PBC method. The difference is that outside the cut-off value the potential is not set to zero, but is solved using an approximate method, and inside the cut-off value the potential energy is calculated using the Poisson's equation. [26]

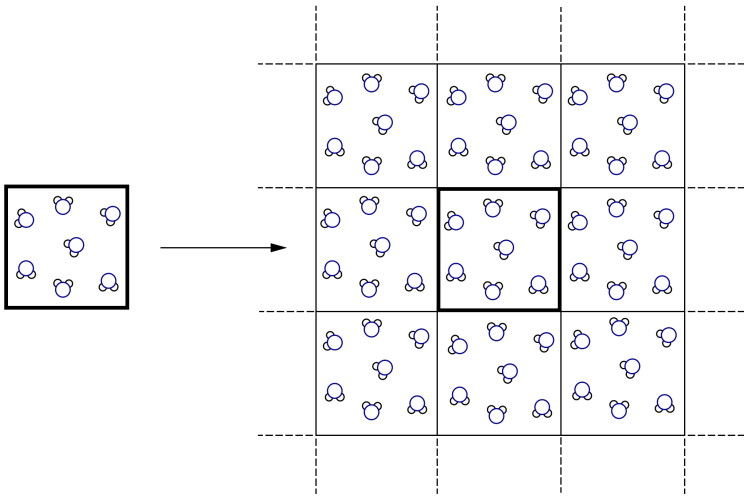


Fig. 11 - Illustration of the PBC's principle: a particle, which goes out from the simulation box by one side, is reintroduced in the box by the opposite side (in the 3 dimension of space). Image created using *Bohemian Coding* (2014) SKETCH (version 3.1.1) 'software'.

2.5 Free Energy

As was already mentioned in chapter 1, drug's permeability limits the oral availability of drugs via cellular membranes, and is a major factor to decide whether a drug candidate could continue through development. As for many other chemical and biological processes, permeability is governed by the change in free energy when drugs cross cell membranes. The ability to predict free energies, along selected degrees of freedom, offers a great insight how passive diffusion works, and this is achieved by the calcula-

tion of the free energy differences between two states of the system. Besides, it may also be interesting to know how the free energy changes as a function of some inter- or intramolecular interactions, such as the distance between two atoms, or the torsion angle of a bond within a molecule, which is achieved by Potentials of Mean Force's calculations (PMF). [26]

2.5.1 The Calculation of Free Energy Differences

The Gibbs free energy of solute transfer between water and 1-octanol can be calculated from the partition coefficient between the 1-octanol and the water phases (the so-called logP):

$$\log P^{(o/w)} = \frac{\Delta_{hyd}G - \Delta_{solv}G}{2.303RT} \quad , \qquad (12)$$

where $\Delta_{hyd}G$ is the hydration free energy, $\Delta_{solv}G$ is the Gibbs free energy of solvation in 1-octanol, R is gas constant, and T is the temperature. The solvation and hydration free energy is the work required to transfer a molecule from the gas phase into 1-octanol and water, respectively. Both are calculated according to the thermodynamic cycle shown in Fig. 12, and following equation 13 and equation 14:

$$\Delta G_{solv} = \Delta G_1 + \Delta G_4 - \Delta G_3 \quad , \qquad (13)$$

$$\Delta G_{hyd} = \Delta G_1 + \Delta G_4 - \Delta G_2 \quad , \qquad (14)$$

where ΔG_1 is the work required to remove all the internal non-bonded interactions in the compound in vacuum, ΔG_2 and ΔG_3 are the work required to remove solute-solvent and the solute-intramolecular interactions, respectively. This is achieved by gradually mutating all atoms in a given compound (state A) into 'dummy' atoms (state B). In practice, a 'dummy' atom has its non-bonded interaction parameters (Lennard-Jones and electrostatics) set to zero. ΔG_4, which appears twice in the cycle, is the work required to transfer the 'dummy' particle from vacuum to water or 1-octanol; however, the term is zero, because the molecule does not interact with the rest of the system. [37, 38]

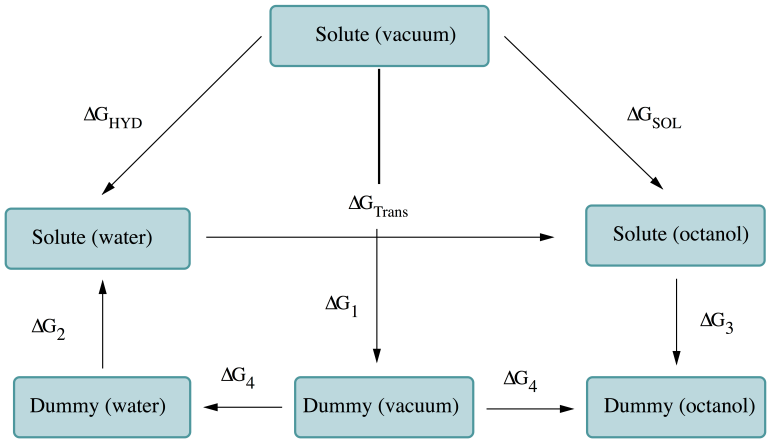


Fig. 12 - Thermodynamic cycle for the determination of solvation free energies. ΔG_{HYD} and ΔG_{SOLV} are calculated using a non-physical path involving ΔG_1, ΔG_2, ΔG_3, and ΔG_4. The transfer free energy, ΔG_{TRANS}, can be obtained from the difference between ΔG_{HYD} and ΔG_{SOLV}.

Different methods have been proposed for calculating the difference in free energy of the system between state A and state B, like Thermodynamic integration (TI), Free Energy Perturbation (FEP), Bennett acceptance ratio (BAR), Multistate Bennett acceptance ratio (MBAR), etc.

The BAR method might be less intuitive and more difficult to implement; however, it is less computational demanding than FEP, and it has been shown to be more efficient than TI.[39] BAR is a perturbation method, using the overlap between individual states, that was developed by Bennett[40] in 1976. Bennett showed the free energy difference change between state *A* and *B* is a function of a coupling parameter λ, which indicates the level of change that has taken place between the two states. At the end of the conducted simulations conducted at different values of λ, a $\partial \mathrm{H}/\partial\lambda$ curve is plotted, from which ΔG_{AB} is derived. The free energy difference is then obtained according to:

$$\Delta G_{AB}^{BAR} = -kT\ ln\frac{N_A}{N_B} \quad , \qquad (15)$$

where k is the Boltzmann constant, T is the temperature, and $\mathrm{N_A}$ and $\mathrm{N_B}$ represent the number of coordinate frames at λ_A and λ_B respectively.[41]

2.5.2 Potentials of Mean Force

Differently from BAR method, the potential of mean force (PMF) method is one approach to describe how the free energy changes as a particular coordinate is varied.

Formally, a conventional method for calculating the free energy, $G(\xi)$, as a function of a reaction coordinate, ξ, relies on obtaining the probability density function, $P(\xi)$, of finding the system in the states corresponding to different values of ξ, and to exploit the relation:

$$G(\xi) = -kT \ lnP(\xi) + C \quad , \qquad (16)$$

where k is the Boltzmann constant, T is the temperature, and C is an arbitrary constant often chosen so that the most probable distribution corresponds to a free energy of zero. Unfortunately, this approach may be inefficient if there are considerable free energy barriers along ξ; however, this method can be improved on its efficiency and accuracy by modifying the Hamiltonian of the system by a biasing potential chosen such that sampling of ξ becomes more uniform. [42, 43]

2.5.3 Umbrella Sampling

Developed by Torrie and Valleau[44], umbrella sampling is the most established technique to determine PMF values along a reaction coordinate by adding a biased force. To ensure sampling in all regions, the reaction coordinate is split into a number of windows, and a bias function is applied to keep the system close to the reference point for each window, which are run in an independent simulation. Often, a simple harmonic potential is used, which allows for oscillation within each window that results in a set of overlapped biased probability distributions; see Fig. 13. [45] Rosta *et al.* have analyzed the difference between the average distances of structures, and conclude the path to be continuous if the difference did not exceed twice the sum of the standard deviations of distances in the two neighboring windows. [45, 46]

As soon as all windows have been simulated, the free-energy differences are combined to generate a global free energy profile. Using techniques like Weight Histogram Analysis Method (WHAM), the effect of the bias potential can be taken out, and the results combined to minimize statistical error, resulting in an unbiased probability distribution, which gives the PMF, see Fig. 14. Overlap between windows is required by WHAM analysis, and the sampling's quality is as better as well the phase space in sampled in each window. If the reaction coordinate misses important structural chang-

es it can lead to artificial lowering or raising of the result obtained by umbrella sampling.

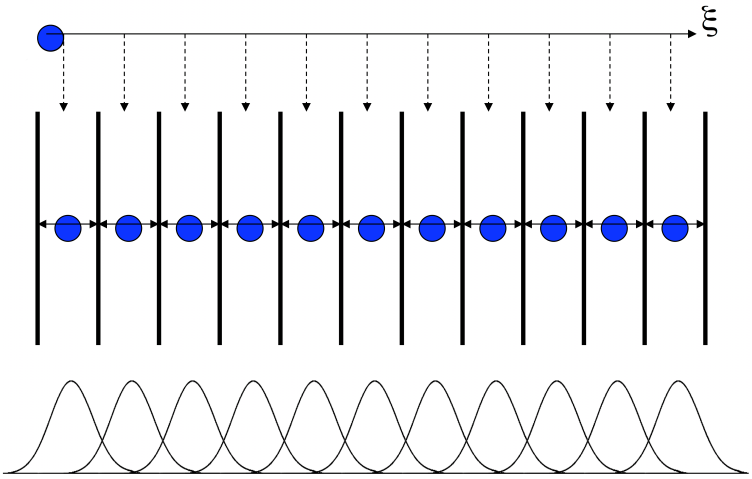


Fig. 13 - The figure illustrates the umbrella sampling windows' principle. A pulling simulation allows for a generation of a series of microstates in specific points along the reaction coordinate (ξ). These microstates are extracted after the simulation is complete (dashed arrows), and independent simulations are conducted within each sampling window, with the molecule center of mass restrained in that window by a harmonic biasing potential. At the bottom is shown the ideal result as a histogram of configurations, with neighboring windows overlapping, such that a continuous energy function can later be derived from these simulations. Adapted from 'Umbrella Sampling Tutorial' from *Justin Lemkul*.**[47]**

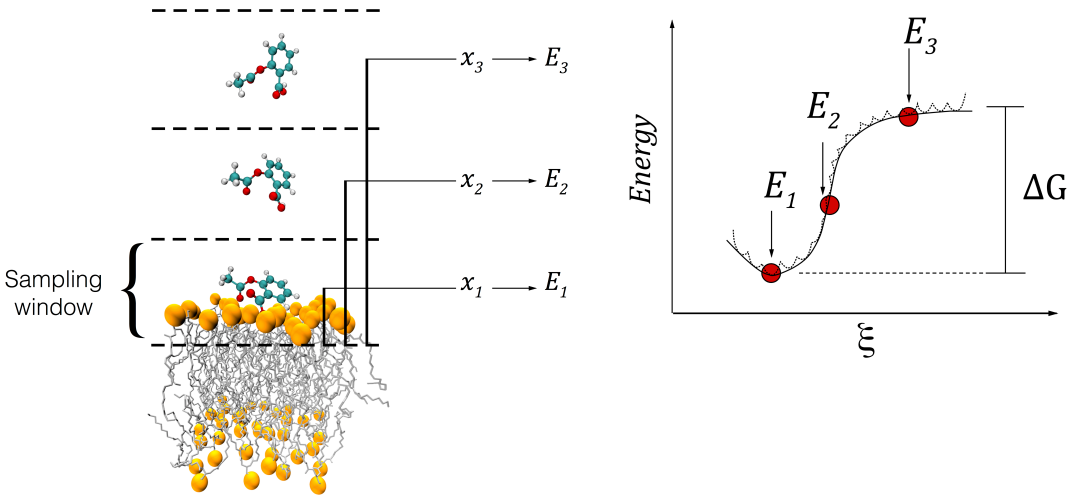


Fig. 14 - Umbrella sampling is a method that confines the molecule of interest into a set of configurations (represented by x_1, x_2, and x_3) along the reaction coordinate (ξ) (on the left). After the simulations' completion, the free energy curves (dashed lines) are combined into a global free energy curve (solid line) (on the right). E_1, E_2, and E_3 represent the PMF of configuration x_1, x_2, and x_3, respectively. Adapted from 'Computation of Potentials of Mean Force in GROMACS' from *Justin Lemkul*.**[48]**

CHAPTER 3

Line of Work

In general, the work described in this thesis took place in three phases. Firstly, a small library of compounds was selected and parameterized, and free energy calculations, in water and in 1-octanol, were computed in order to validate the force field that was being used. Secondly, a DOPC/water membrane model was constructed, and umbrella sampling simulations were performed for three molecules: caffeine, nelfinavir, and atorvastatin. These umbrella sampling simulations helped us to attain the free energy profile for the drugs' permeation through the membrane. Thirdly, nelfinavir and atorvastatin's rotatable bonds and PSA were determined. In this section, the protocol for each phase will be detailed.

3.1 Parameterization

A small library of compounds, for which solvation free energies could be extracted from the literature, was set up, and free energy calculations for each compound (neutral form), in water and in 1-octanol, were performed. Therefore, a few compounds, for which the chemical structures are shown in Fig. 15, were chosen from Nicholls et al. [49], and their primordial structures (mol2 files) were extracted from the *ZINC database* (zinc.docking.org).

After the compounds' selection, they were gas-phase parameterized, using the *Antechamber* and *Gaussian 03* programs. Antechamber is an auxiliary module in AMBER ("Assisted Model Building with Energy Refinement") designed to be used with GAFF in order to identify bond and atom types, judge atomic equivalence, generate residue topology files, and find missing FF parameters and propose reasonable alternatives. [50] In its turn, *Gaussian 03* is an electronic structure program, based on the fundamental

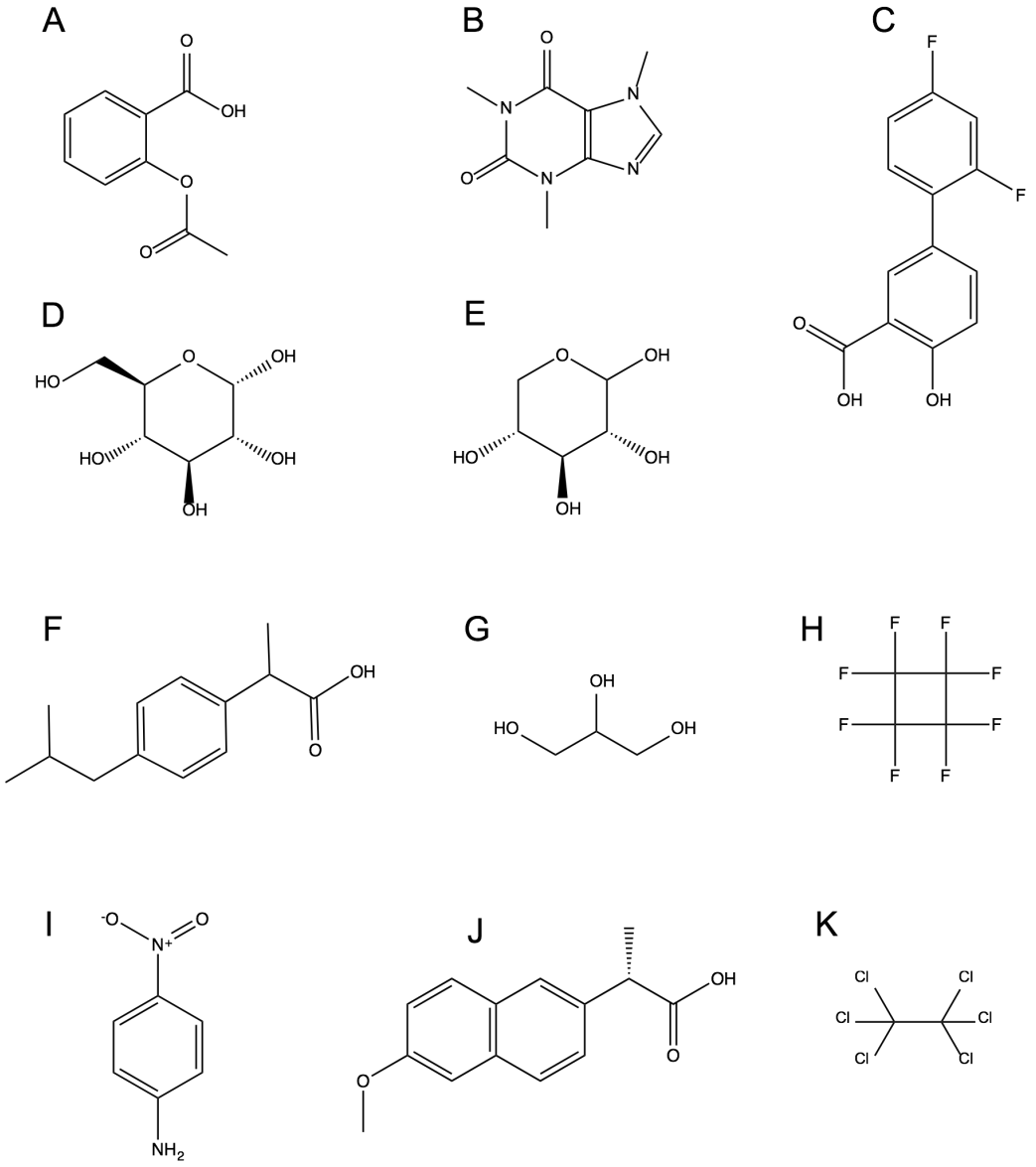


Fig. 15 - Squematic representaiton of the compounds that were selected to make part of the compound's library for validation: (A) aspirin; (B) caffeine; (C) diflunisal; (D) D-glucose; (E) D-xylose; (F) ibuprofen; (G) glycerol; (H) octafluorocyclobutene; (I) 4-nitroaniline; (J) naproxen; (K) hexachloroethane.

laws of Quantum Mechanics (QM), and it is used by scientists to predict the energies, molecular structures, and vibrational frequencies of molecular systems, along with numerous molecular properties derived from these basic computation types.

Gaussian was then used to perform a geometry optimization on each molecule in order to determine the minimum energy structure, in a process that was carried out using the HF/6-31G(d), a Hartree-Fock theory that uses a medium-sized basis set. The

RESP method to derive a set of atom-centered point charges, which best reproduces the electrostatic potential of the molecule, was also used. [51]

After compounds' parameterization, the structure and topology files were converted, using the ACPYPE toll, into other files compatible with GROMACS (Groningen MAchine for Chemical Simulation) program. ACYPYPE is an ANTECHAMBER-based toll that calculates partial charges and generates topology and parameters in different formats for different MM programs.[52] GROMACS is a very fast program for molecular dynamics simulation; although, it does not have a force field of its own, it is compatible with commonly FFs like AMBER or GROMOS. [53] Moreover, the Slipids force field, which is used in this work for the simulation of biological membranes, was designed exclusively in GROMACS format. Indeed, in this work, all molecular simulations were performed using the GROMACS 4.6.1 package.

3.2 Free energy calculations

For each parameterized molecule, a cubic water box of at least 1.5 nm between any atom of the molecule and the edge of the box was built, using *editconf* and *genbox* tools from GROMACS (see Fig. 16). The TIP3P[54] water model was chosen for this regard.

As it was discussed before, free energy calculations describes the transformation from state A (λ=0) to state B (λ=1); so, the first step in planning free energy calculations

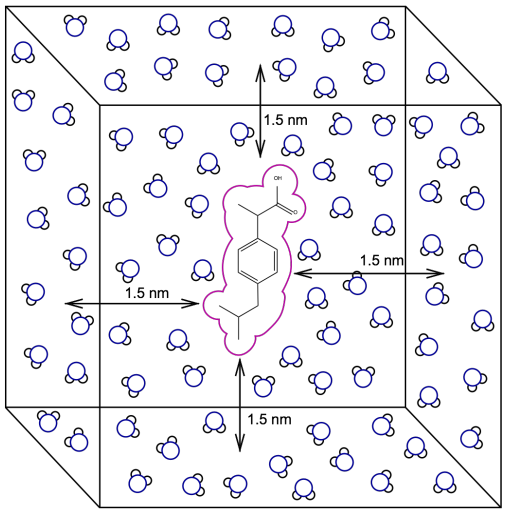


Fig. 16 - Squematic representation of a cubic water box. Image created using *Bohemian Coding* (2014) SKETCH (version 3.1.1) 'software'.

is how many λ point will be used. In this work, Coulombic and van der Waals' interactions were decoupled, and an equidistant λ spacing of 0.05 was used, that is to say, 20 λ were used.

For each value of λ, in which λ=0 refers to a fully interacting solute and λ=1 to a noninteracting solute, a complete workflow was conducted: steepest descents minimization of 5000 steps, L-BFGS minimization[55] of 5000 steps, NVT equilibration of 100 ps, NPT equilibration of 100 ps, and MD production under the NPT ensemble of 5 ns. Two energy minimization procedures were used, since the results are more accurate when they are used together.

The integration of Newton's equations of motion was carried out using an accurate leap-frog stochastic dynamics (SD) integrator[56] with a time step of 2 fs. Short-range non-bonded interactions were cut off at 1.0 nm, with long-range electrostatics calculated using the particle mesh Ewald (PME) algorithm. Dispersion correction was applied to energy and pressure terms to account for truncation of van der Waals terms. Periodic boundary conditions were applied in all directions. Hydrogen bonds were constraint using the LINCS (LINear Constraint Solver) algorithm[57]. Temperature coupling was implicitly handled by de SD integrator with a time constant of 1 ps, and a reference temperature of 298.15 K. For constant pressure simulations the Parrinello-Rahman[58] barostat with a time constant of 0.5 ps, a reference pressure of 1 bar and an isothermal compressibility of 4.5×10^{-5} bar $^{-1}$ was used to enforce pressure coupling.

The use of the ‘soft-core' potentials is important to highlight. In the transformation process between states with different λ values, the λ dependence of the Lennard-Jones potential is interpolated between the neighboring states via this 'soft-core' expression, which allows to eliminate singularities in the calculation as the Lenard-Jones interactions are turned off. Consequently, a 'soft-core' parameter of 0.5 and a 'soft-core' σ value of 0.3 nm were used. [37]

Finally, data analysis was conducted using the GROMACS tool g_bar, which uses the BAR method for calculating free energy differences. The final error estimates were obtained from multivariate analysis of multiblock analysis over 500 blocks.

In addition, ‘reverse’ simulations for caffeine and D-glucose were performed. In this context, ‘reverse’ means a coupling free energy calculation instead, in which the transformations occurs from a non-interacting solute to a fully interacting solute. Discrepancies between forward and reverse transformations yield the reaction's hysteresis, which constitutes a measure of systematic errors in the free energy calculation. Moreover, an

extended free energy calculation of 20 ns *per* λ for D-glucose was performed, because hysteresis tends to increase with the increase in the simulation's time. [59]

The same process used to calculate free energies in water, was also used to calculate free energies in 1-octanol. First, the 1-octanol molecule was parameterized according to the parameterization protocol described earlier; secondly, a system in which a drug is solvated by 1-octanol was built by the Packmol (Packing Optimization for Molecular Dynamics Simulations) [60, 61] program. Packmol packs molecules in defined regions of space, and guarantees that short repulsive interactions do not disturb the initial stage of simulation. Basically, through Packmol, a single 1-octanol molecule was replicated along the x and y axes until a rectangular grid of 300 1-octanol's molecules was obtained [62]

Before free energy calculations, a geometry optimization was performed in order to eliminate bad contacts, through a complete workflow: steepest descent minimization of 2000 steps, NVT equilibration of 250 ps, NPT equilibration of 500 ps, and MD production under NPT ensemble of 1 ns. The integration of Newton's equations of motion was carried out using a leap-frog MD integrator with a time step of 2 fs. Short-range non-bonded interactions were cut off at 1.0 nm, with long-range electrostatics calculated using the particle mesh Ewald (PME) algorithm. Dispersion correction was applied to energy and pressure terms to account for truncation of van der Waals terms. Periodic boundary conditions were applied in all directions. Hydrogen bonds were constraint using the LINCS (LINear Constraint Solver) algorithm[57]. Temperature coupling was handled by V-rescale thermostat (a modified Berendsen thermostat[63]) with a time constant of 1 ps, and a reference temperature of 298.15 K. For constant isotropic pressure simulations the Parrinello-Rahman[58] barostat with a time constant of 1 ps, a reference pressure of 1 bar and a isothermal compressibility of 7.64×10^{-5} bar^{-1} [64] was used to enforce pressure coupling. After that, free energy calculations, according to the same simulation conditions for the free energy calculations in water, were performed, and data analysis was also conducted through the *g_bar* tool.

3.3 Setting up of the membrane system

The lipid membrane system was built in several steps. First, the lipid bilayer was built using a Membrane Builder web-based graphical user interface, called CHARM-GUI[65-68]. The Membrane Builder generates a series of CHARMM inputs necessary to build a protein/membrane complex for molecular dynamics simulations. Therefore, a DOPC

lipid membrane with 100 DOPC's molecules per leaflet and a water box thickness of 8 nm was built, see Fig. 17. Next, the drug was included in the system, 5 nm away from the DOPC bilayer, and after that, using the *genbox* tool. TIP3P[54] water molecules filled the 8 nm water box, thus establishing a solvated bilayer model. Moreover, Na^+ and Cl^- ions were added, using *genion* tool, to satisfy a saline concentration of 0.154 M of NaCl, in order to mimetize physiological conditions.

Three compounds were studied in this part of the work: caffeine, nelfinavir, and atorvastatin. Caffeine was chosen to optimize protocols since its PMF profile can be found in the literature, whereas nelfinavir and atorvastatin were chosen because they have many rotatable bonds, which makes them an interesting case of study. Moreover, these molecules have different permeability: caffeine has high permeability, nelfinavir has moderate permeability, and nelfinavir has low permeability. Besides, our group has

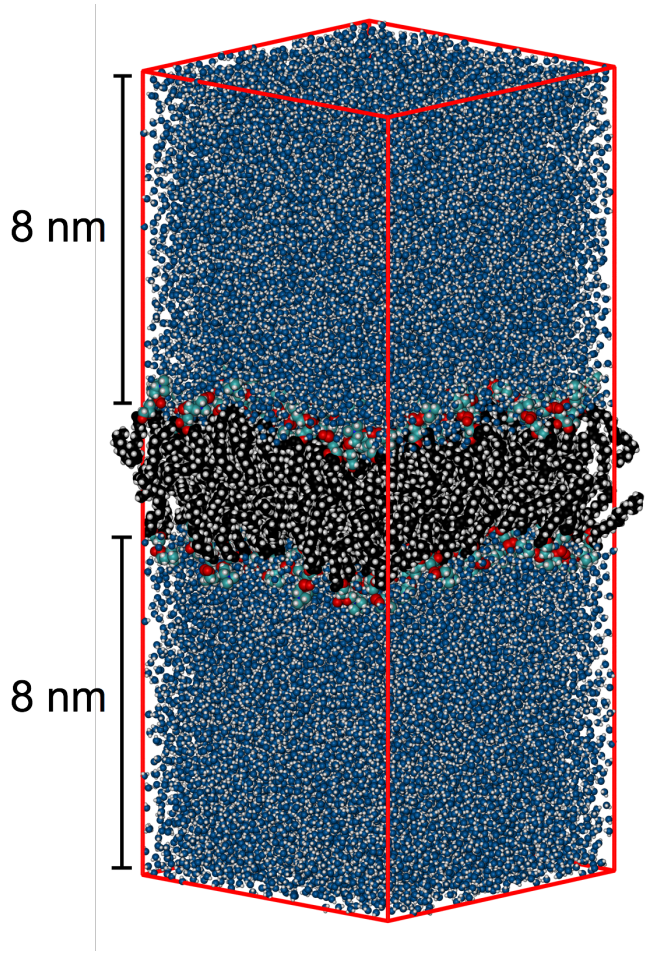


Fig. 17 - DOPC's membrane representation, with 100 DOPC's molecules per leaflet and a water's thickness of 8 nm.

extensive experience in working with nelfinavir[69, 70]. The chemical structure of caffeine is represented in Fig. 15.B, and the chemical structure of nelfinavir and atorvastatin is represented in Fig.18. The chemical and physical properties of these molecules, which are important for Lipinski and Veber's rules, are shown in Table I.

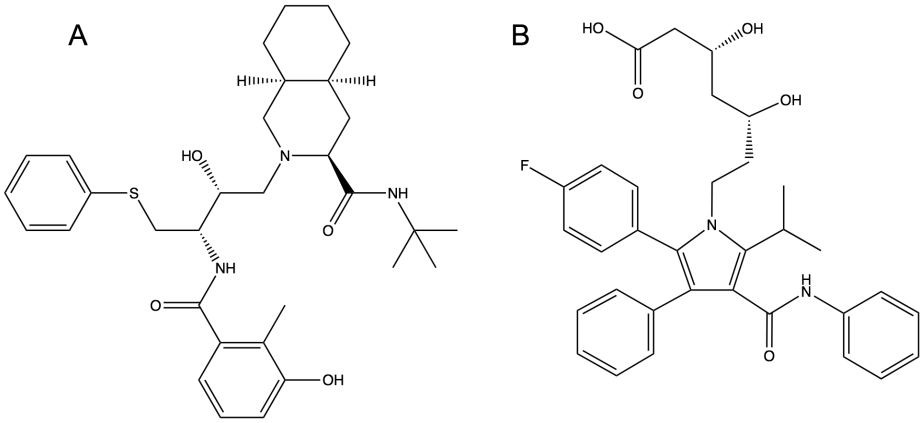


Fig. 18 - Nelfinavir (A) and Atorvastatin's (B) chemical representations.

The system was then equilibrated through a complete workflow: steepest descents minimization of 2000 steps, NVT equilibration of 250 ps, NPT equilibration of 1 ns, and MD production under the NPT ensemble for 40 ns. The integration of Newton's equations of motion was carried out using a leap-frog MD integrator with a time step of 2 fs. Short-range non-bonded interactions were calculated using the Verlet cut-off scheme[71] and a cut off distance of 1.2 nm, with long-range electrostatics calculated using the PME algorithm. The Verlet cut-off scheme is a requirement to use GPU accelerators for running MD simulations, allowing the nonbonded interactions to be calculated on the GPU, whereas bonded and PME interactions are calculated on standard CPU hardware. Dispersion corrections were applied to energy and pressure terms to account for truncation of van der Waals terms. Periodic boundary conditions were applied in all directions. Hydrogen bonds were constraint using the LINCS algorithm[57]. Temperature coupling was handled by Nosé-Hoover[72, 73] thermostat with a time constant of 0.5 ps, and a reference temperature of 310 K, whereas semiisotropic (separate pressure coupling in xy and z separately) pressure coupling was handled by the Parrinello-Rahman[58] barostat with a time constant of 10.0 ps, a reference pressure of

Table I - Relevant chemical and physical properties for Lipinski and Veber's rules extracted from *chemicalize.org by ChemAxon.*

	Molecular weight / g•mol^{-1}	**log *P***	**Hydrogen bond donor count**	**Hydrogen bond acceptor count**	**Rotatable bond count**	**PSA / Å^2**
Caffeine	194.19	-0.07	0	3	0	58.44
Nelfinavir	567.78	4.72	4	6	10	101.90
Atorvastatin	558.64	5.39	4	6	12	111.79

1 bar, and an isothermal compressibility of 4.5 ×10^{-5} bar $^{-1}$ (same as for water) was used.

Next, to verify whether the systems were well equilibrated several parameters were analyzed: the deuterium order parameters of the acyl chains, the density of the membrane environment, the area per lipid's headgroup, and the bilayer's thickness. These will be thoroughly described and presented in the results section of this thesis.

3.4 Calculations of Potentials of Mean Force

As it was described in section 2.10.3, to conduct umbrella sampling simulations it is necessary to generate a series of configurations along a reaction coordinate ξ, which is accomplished by 'pull' simulations. During a pull simulation forces or constraints are applied between the center of mass of one or more pairs of groups of atoms. Thus, to generate the configurations we had to choose and defined a series of options for different settings that GROMACS offer us to perform pulling simulations. In this work it is desired that a molecule cross the membrane; so, technically, the pulled group (the drug) has to cross the plane of the reference group (the DOPC membrane); consequently, one of the most important settings in pulling simulations is pull geometry, that is to say the path (direction and final position) that the drug has to follow. GROMACS offers us different pull geometries, and two of them were used in this work: the *position* geometry, and the *cylinder* geometry. For both pull geometries used, several simulations, varying the simulation's time (0.5-1 ns), the force constant for pulling (10 -100 kcal•mol^{-1}), and the pull rate (0.1, 0.01, or 0.001 nm•s^{-1}), were performed until a complete crossing through the membrane had occurred. The pull simulation was run with the same MD setup used in section 3.3.

After that, frames for defining the umbrella sampling windows, with 0.1 nm spacing were extracted from the pull's trajectory files, for every 0.1 nm of the drug's progression. Some of these frames were then submitted to a MD simulation with only a NPT equilibration of 1 ns, and MD production under NPT ensemble of 5 ns. Many of the simulation's parameters were the same as in the pull simulation, with the exception of the 'pull rate', which was set to zero to hinder the configuration's movement along the reaction coordinate.

At the end of each all umbrella sampling simulation, the PMF and the configurations' histograms were extracted using the WHAM method, which is included in GROMACS as the *g_wham* utility.

3.5 Rotatable bonds and polar surface area characterization

To characterize rotatable bonds of nelfinavir and atorvastatin, a 10 ns simulation for each molecule in a water box was performed, using the same simulation's setup described in section 3.3. The value of each dihedral angle was calculated using the *g_angle* GROMACS tool for the last 5 ns of each simulation. Along with rotatable bonds, the PSA was calculated from that simulation using the *g_sas* GROMACS tool. From *g_sas*, we got a file containing all atoms' area, but only oxygen, nitrogen, and sulfur were taken into account.

CHAPTER 4

Results and Discussion

4.1 Free energy

Hydration and 1-octanol solvation free energies were estimated for a set of small compounds (Fig. 15 in section 3.1), in order to validate the FF's parameters. The purpose of the validation process is to test whether the FF is capable of reproducing these thermodynamic properties. If that happens, the FF must be also capable of reproducing the PMF profiles accurately, since the water is the same and the 1-octanol is analogous to the hydrophobic core of the membrane. Hence, we compared our computed free energy values with those that made part of the SAMPL experiment, which was designed as a blind challenge, where the actual experimental values of the compounds are withheld from participants until after predictions have been made.[74] The hydration free energy values that were determined using the BAR method are shown in Table II. The standard deviation (sd) values for ΔG_{AB} are important to highlight; these values were estimated from multivariate analysis of a multiblock analysis of 500 blocks, assuming the blocks are independent, and taking into account time correlation.

Although, the reverse process (coupling) should be the correspondent symmetrical value of the decoupling process (assuming reversibility), coupling free energy calculations for caffeine and D-glucose were performed to find the reaction's hysteresis value, which constitutes a measure of the error in the free energy calculation. Mathematically, the hysteresis is the difference between coupling and decoupling processes, and if its value is markedly larger than the estimated statistical errors, it is usually indicative of ergodicity issues during the transformations. [59] As can be noted when looking at Table II, the hysteresis values were almost zero, which means the systematic error of our free energy calculations was very small; however, because hysteresis tends to increase with the increase in the simulation's time[59], an extended free energy calcula-

tion of 20 ns for D-glucose was performed. In the end, both simulations yielded the same result. Although, the hysteresis values obtained were negligible, and statistical errors were small, this does not imply that the calculated free energy difference is accurate, because it may be burdened with other systematic errors, due, for example, to unsuitable potential energy functions. Furthermore, the system may have been confined to a sub-region in the physical space. [59]

To confirm whether free energy calculations are accurate, it is necessary to compare the results obtained with experimental data extracted from literature. A difference of 1.15 kcal•mol^{-1}, on average, between experimental values and the values obtained computationally can be observed also from Table II. This average value, which was obtained by the arithmetic mean calculation of the values for all compounds except

Table II - Estimated values of hydration free energy, experimental free energy values, and difference between computed and experimental values.

Molecule	$\Delta G_{Hyd\ (Calc)}$ / kcal•mol^{-1}	$\Delta G_{Hyd\ (Expt[49])}$ / kcal•mol^{-1}	$\lvert \Delta G_{Hyd\ (Calc)} - \Delta G_{Hyd\ (Expt)} \rvert$ / kcal•mol^{-1}
Caffeine	-14.46 ± 0.03	-12.82	1.64
	14.43 ± 0.03 [b]		
D-glucose	-23.78 ± 0.08	-25.44	1.66
	-23.88 ± 0.05 [a]		
	23.84 ± 0.13 [b]		
	23.90 ± 0.05 [a,b]		
D-xylose	-19.68 ± 0.08	-20.50	0.82
Glycerol	-14.01 ± 0.04	-13.92	0.09
Hexachoroethane	0.57 ± 0.03	-0.28	0.85
Naxoprene	-11.60 ± 0.04	-10.35	1.25
Uracil	-15.32 ± 0.03	-16.06	0.74
4-Nitroaniline	-12.44 ± 0.04	-9.89	2.55
Aspirin	-12.94 ± 0.03	-12.33	0.61
Ibuprofen	-8.93 ± 0.04	-7.00	1.93
Octafluorocyclobutanol	-2.94 ± 0.02	-3.40	0.46
Diflunisal	-11.78 ± 0.05	-7.63	4.15

[a] simulations of 20 ns
[b] reverse simulation

Diflunisal, which was statistically considered an outlier (see Fig. 19). Indeed, Diflunisal has two atoms of fluorine, and since polarization effects are important when dealing with halogens and since these are not accurately described by classical, fixed-charged FFs, it could justify the outlier characteristics of Diflunisal.[75] However, this average value was considered acceptable, which supports the accuracy of the GAFF force field in describing hydration free energies for this subset of compounds. Obviously, this is a reduced dataset for assessing on the accuracy of a force field. Hence, we are working to expand the dataset of compounds to be tested. Furthermore, though it would be interesting to calculate hydration and 1-octanol solvation free energies for nelfinavir and atorvastatin, such calculations would take too long, and there are no experimental data available in the literature.

Regarding the free energy calculations in 1-octanol, the sd values for ΔG_{AB} were also estimated from multivariate analysis of a multiblock analysis of 500 blocks, assuming the blocks are independent, and taking into account time correlation. In Table III the results of 1-octanol solvation free energy calculations in 1-octanol and the experimental 1-octanol solvation free energy values are shown. These experimental values were calculated according to equation 12 (see section 2.10.1), and using the logP and hydration free energy experimental values. The logP values were also obtained according to the same equation, and they can be found in Table IV. Comparing the obtained 1-octanol solvation free energy values to the experimental ones, there is a discrepancy on average of 2.33 kcal•mol^{-1} between them. Regarding the logP values, we obtained consistent values for the two molecules with the highest logP (caffeine and D-xylose), and for the two molecules with lower logP values, though there is a discrepancy between these last two molecules. We assume that this could be attributed to the short time of simulation (5 ns / λ) and/or to the parameters for 1-octanol. Thus, to improve the free energy calculation in 1- octanol, it must be necessary to increase the simulation's time in 20 ns or more per λ, and/or improve the parameters for 1-octanol. Moreover, it would be necessary to increase the number of compounds to be tested. As a

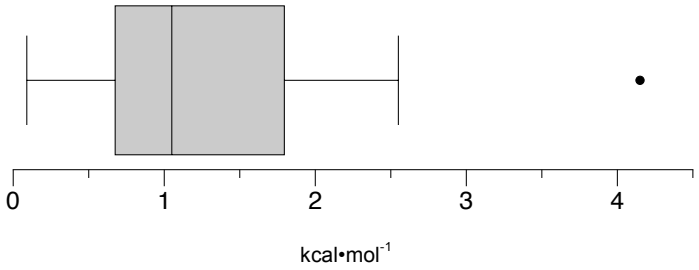


Fig. 19 - Hydration free energy values distribution. The median value is 1.05.

Table III - Estimated values of ΔG_{solv} in 1-octanol, calculated experimental ΔG_{solv} in 1-octanol values, and the difference between them.

Molecule	$\Delta G_{solv\ (Calc)}$ / kcal•mol^{-1}	$\Delta G_{solv\ (Expt)}$ / kcal•mol^{-1}	$\lvert \Delta G_{solv\ (Calc)} - \Delta G_{solv\ (Expt)} \rvert$ / kcal•mol^{-1}
Caffeine	-17.47 ± 0.05	-12.72	4.75
D-glucose	-22.50 ± 0.32	-20.99	1.51
D-xylose	-18.28 ± 0.08	-19.03	0.75
Uracil	-15.33 ± 0.04	-13.01	2.32

Table IV - Estimated logP values, experimental logP values, and the difference between them.

Molecule	**logP** $_{(Calc)}$	**logP** $_{(Expt)}$	**logP**$_{(Calc)}$ **- logP**$_{(Expt)}$
Caffeine	2.192 ± 0,008	-0.07	2.26
D-glucose	-0.932 ± 0,004	-3.24	2.31
D-xylose	0.0738 ± 2 x10^{-5}	-1.07	1.08
Uracil	-1.019 ± 0,006	-2.22	1.20

matter of fact, free energy calculations in 1-octanol were only performed for four compounds due to the computational demand required for each free energy calculation.

Despite the discrepancy between experimental and computed logP values, that needs to be explored for a larger number of compounds, we proceeded with the permeation studies through the membrane. These less promising results could be indicative that GAFF is unable to accurately describe the free energy for different phases. However, we cannot markedly state such hypothesis without an extended validation considering numerous compounds. Numerous aspects led us to pursue with GAFF: 1) the computed hydration free energies were good, 2) the SLIPID parameters have been extensively validated, contrarily to 1-octanol parameters; and 3) this less good results for logP computation could be indicative of not enough sampling. We have then decided to continue with GAFF for the subsequent evaluations. Another main reason is due to the compatibility between parameters; by using other MM parameters and/or energy functions, from other philosophies, this could lead to increased errors due to FF incompatibility. Normally, the combination between different FF is not advisable without proper validation.

4.2 Setting up of the membrane system

Cell membranes have very complex compositions, and for mammalian cell membranes, phosphatidylcholine (PC) glycerophospholipids are the most abundant. In the present work, we carefully choose to employ a dioleoylphosphatidylcholine (DOPC) bilayer in our observations, because it had been one of the most studied glycerophospholipidic bilayer systems. Moreover, DOPC has a transition temperature of -16.5 ºC, which means that DOPC is in the fluid phase state at room temperature, and is ideal for studies at physiologic temperature.[76] To confirm whether the simulated membrane systems were well equilibrated, several parameters were analyzed and compared with values found in literature: bilayer thickness, carbon–deuterium order parameters, and area per lipid.

Bilayer thickness can be obtained from the density profile considering the hydrated system, by calculating the peak-to-peak distance between phosphate groups' density (see Fig. 20) [33, 77]. The bilayer thickness' values for each membrane system over the last 10 ns of the MD runs are shown in Table V. When we compare these values to the experimental one, 3.69 nm (30ºC) [78], we see that these are very similar and that minor differences could arise from different temperature settings.

In the liquid crystalline phase, the lipid tails are conformationally disordered and fluid. To verify whether the membrane models had entered into a gel phase during the simulation, order parameters for the acyl chains were determined. They measure the disorder of the acyl chains and can be related to experimental carbon–deuterium order parameter (S_{CD}) of the deuterated carbons in the hydrocarbon chains (representing the relative orientation of the carbon–deuterium bonds). [77] The calculated values for the sn-1 and sn-2, over all lipid acyl chains and for each membrane system, are plotted in Fig. 21 as a function of carbon atom number, starting from the atom closest to the headgroup, averaged over the last 10 ns of the MD runs. In all membrane systems the S_{CD} values are smaller than 0.30, which is characteristic of a disordered hydrocarbon region. In addition, it can be observed, in Fig. 21, a dip in the values that correspond to the unsaturated carbon atoms present in the DOPC acyl chains in positions 9 and 10. [77]

Finally, the area per lipid is another useful quantity to evaluate the bilayer systems' equilibration. This parameter was determined by dividing the area of the xy plane, which covers the lipid's nitrogen atoms, by the total number of lipids in one layer (36 for caffeine membrane system, and 100 for nelfinavir and atorvastatin membrane systems). [62] Looking at Table V, in which the area per lipid average value for each

membrane system over the last 10 ns of the MD runs is presented, we observe a high consistency with experimental data, 0.691/0.674 nm^2 (30ºC)[79, 80].

Table V - Bilayer thickness and area per lipid's values averaged for the last 10 ns of the MD runs.

System	Bilayer thickness / nm	Area per lipid / nm^2
Caffeine	3.81	0.6436
Nelfinavir	3.80	0.6645
Atorvastatin	3.78	0.6592

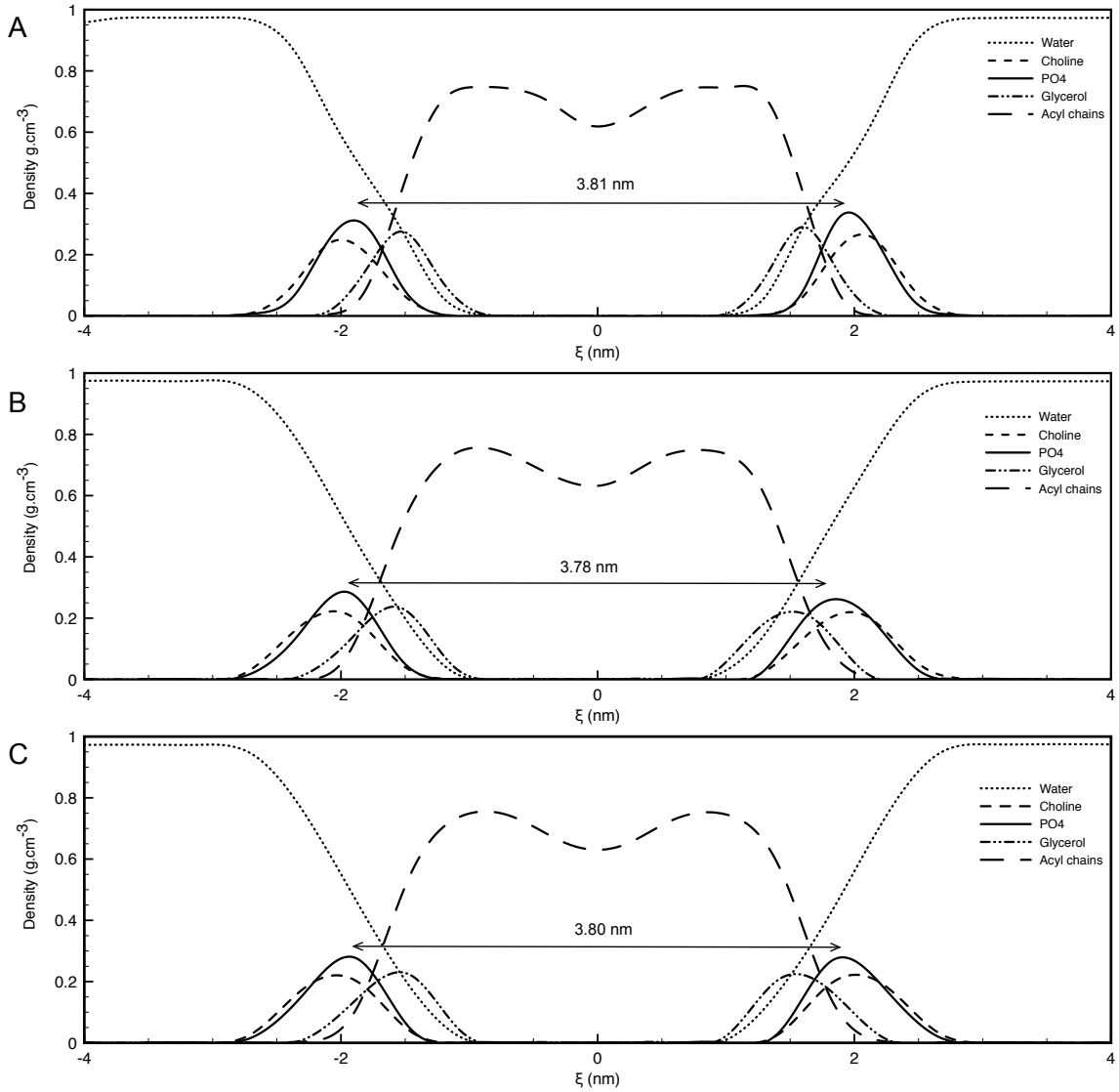


Fig. 20 - Graphical representation of membrane groups' densities of (A) caffeine, (B) nelfinavir, and (C) atorvastatin membrane systems, over the last 10 ns of the MD runs.

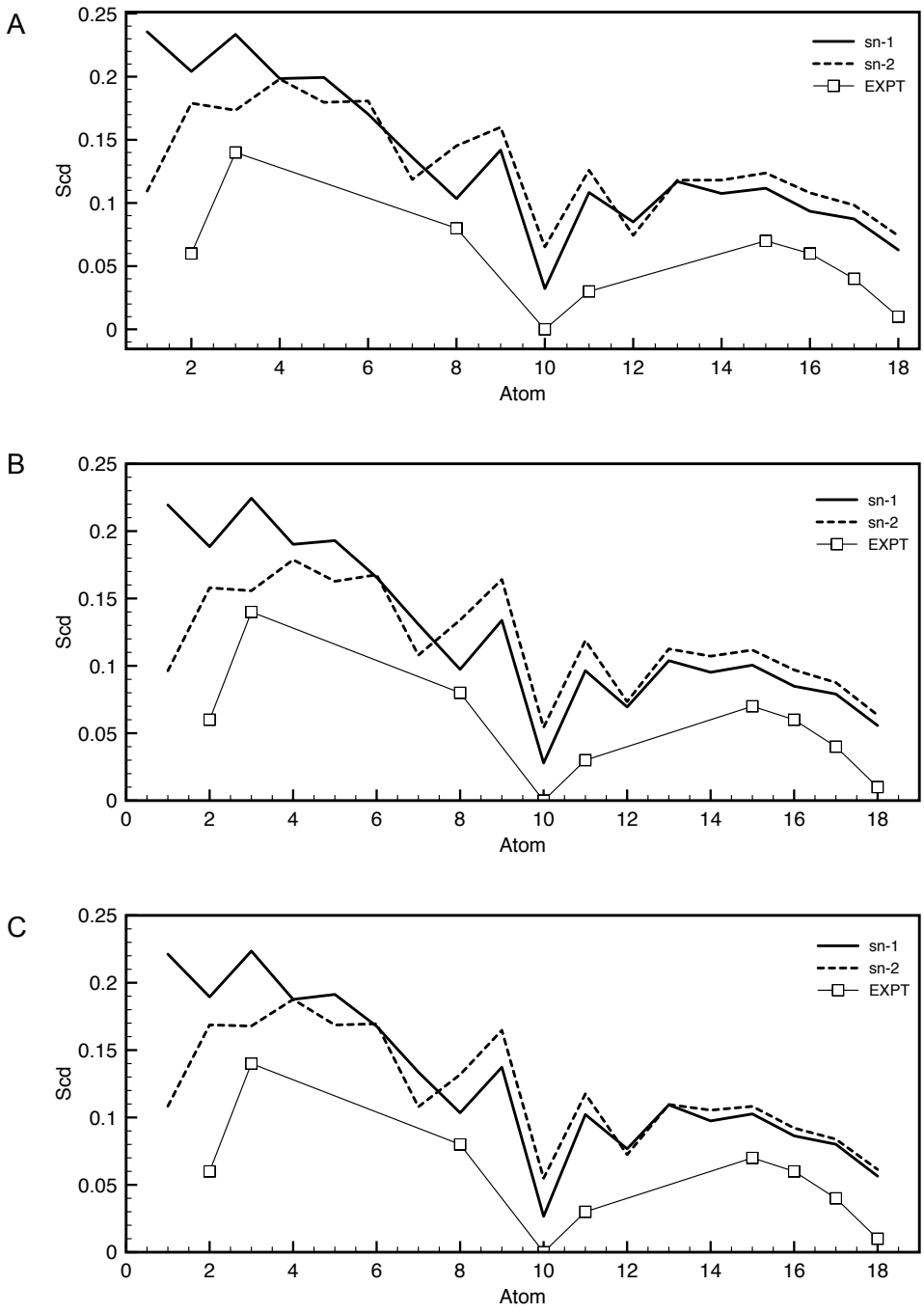


Fig. 21 - Graphical representation of deuterium order parameters for the sn-1 and sn-2 acyl chains of (A) caffeine, (B) nelfinavir, and (C) atorvastatin membrane systems, over the last 10 ns of the MD runs. Experimental values extracted from Waschawski *et al*.**[81]**

4.3 Protocol Adapting

To perform umbrella sampling calculations on GROMACS, the 'Umbrella Sampling' tutorial written by Justin Lemkul was followed. However, in this tutorial, Lemkul and his co-workers had used the umbrella sampling to calculate the free energy of the dissociation of a single peptide from the growing end of an Aβ42 protofibril. [82] Therefore, it was necessary to adapt the Lemkul's tutorial to the system that we wanted to study, and in this regard, umbrella sampling was primarily performed for caffeine. This was due to the smaller size of this compound, which results in faster simulations.

As stated earlier (see section 2.5), the accuracy of the PMF profile depends largely on the width and on the overlap of the sampling windows, and to achieve an accurate PMF profile, one has to conduct several simulations while varying the harmonic force constant. In Table VI the number of configurations calculated and total simulation time for all umbrella sampling simulations performed are shown, though only some of them will be discussed.

Firstly, we conducted an umbrella sampling simulation for the first 25 ξ points with a space of 0.2 nm between them, and 6 ns (1 ns of NPT and 5 ns of a production simulation) per ξ, which results in 150 ns on the whole. A harmonic force constant of 500 $kJ•mol^{-1}•nm^{-2}$ was used. In Fig. 22 the PMF and the umbrella histograms are represented, and an exceeding overlap of umbrella windows in some regions is observed, which though they do not impart the accuracy of the results, they required more simulations to cover the entire reaction coordinate. We thought at first that by increasing the harmonic constant force we would improve the overlap between neighboring windows.

Table VI - Overview of umbrella sampling simulations performed for method validation.

Harmonic force constant / $kJ•mol^{-1}•nm^{-2}$	**Pull geometry**	**No. of ξ points**	**Time per ξ point /** ns	**Simulation's time /** ns
500	position	25	6	150
800	position	26	6	150
1000	position	25	6	150
2000	position	46	6	276
2000	position	20	21	420
2000	cylinder	30	6	180
3000	position	30	6	180
4000	position	30	6	180

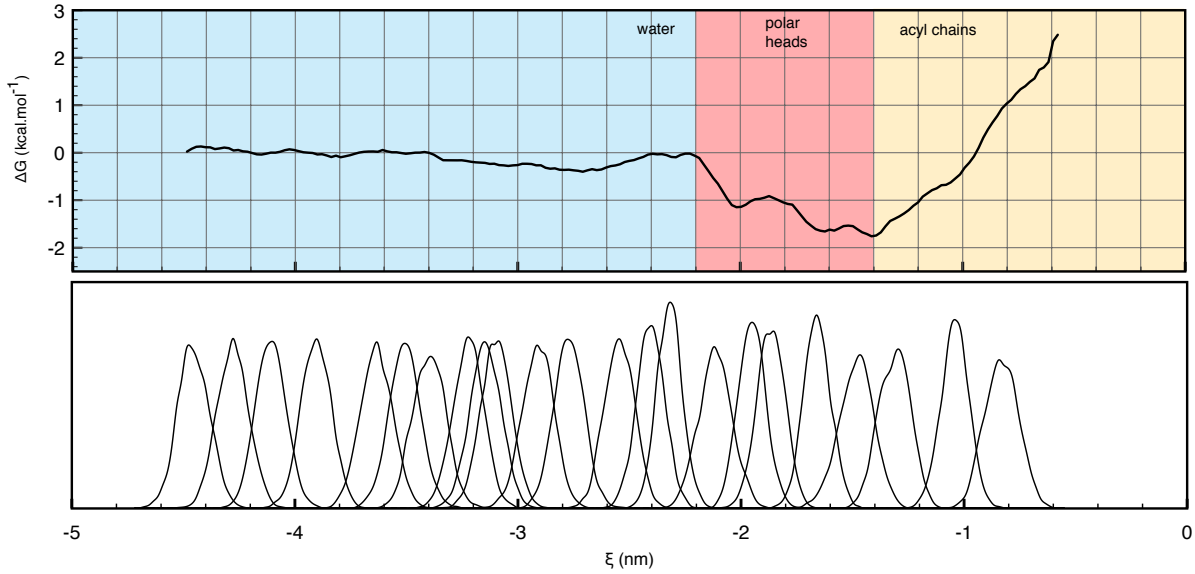


Fig. 22 - PMF profile and umbrella histograms for the validation of the umbrella sampling protocol, considering the caffeine membrane system's simulations. Umbrella sampling was conducted following a 'position' geometry protocol and a harmonic force constant of 500 kJ•mol^{-1}•nm^{-2} for 25 configurations spaced by 0.2 nm. The free energy in water was set to be zero.

Thus, an equal umbrella sampling simulation, but with a harmonic force constant of 1000 kJ•mol^{-1}•nm^{-2}, was performed. The results are shown in Fig. 23, where an enhanced PMF profile, and a refined and smoother distribution of umbrella histograms can be observed. However, not only there are still some regions with an excess overlap, but also there were some regions in which there was no overlap. Once again, to improve the histogram distribution, a third harmonic force constant of 2000 kJ•mol^{-1}•nm^{-2} was tested. In this third situation, more sampling configurations were included to overcome the lack of sampling windows in some regions; however, there was still overlapped sampling windows, as can be seen in Fig. 24. Although not shown, umbrella sampling simulations with harmonic force constants of 800, 3000, and 4000 kJ•mol^{-1}•nm^{-2} were performed.

While we were trying to improve the method, by conducting new simulations varying the harmonic force constant, and adding more sampling configurations, we were faced with different pull geometry protocols on GROMACS that seem to be designed specifically for a layered system, as for instance, for a lipid bilayer in order to minimize the effects of membrane undulations. While 'position' geometry allows us to calculate the PMF of a drug as function of its distance from the whole lipid bilayer, 'cylinder' geometry defines the reaction coordinate more locally not using all the atoms of the refer-

ence group, but instead dynamically only those within a cylinder around the pull vector going through the pull group, see Fig. 25.

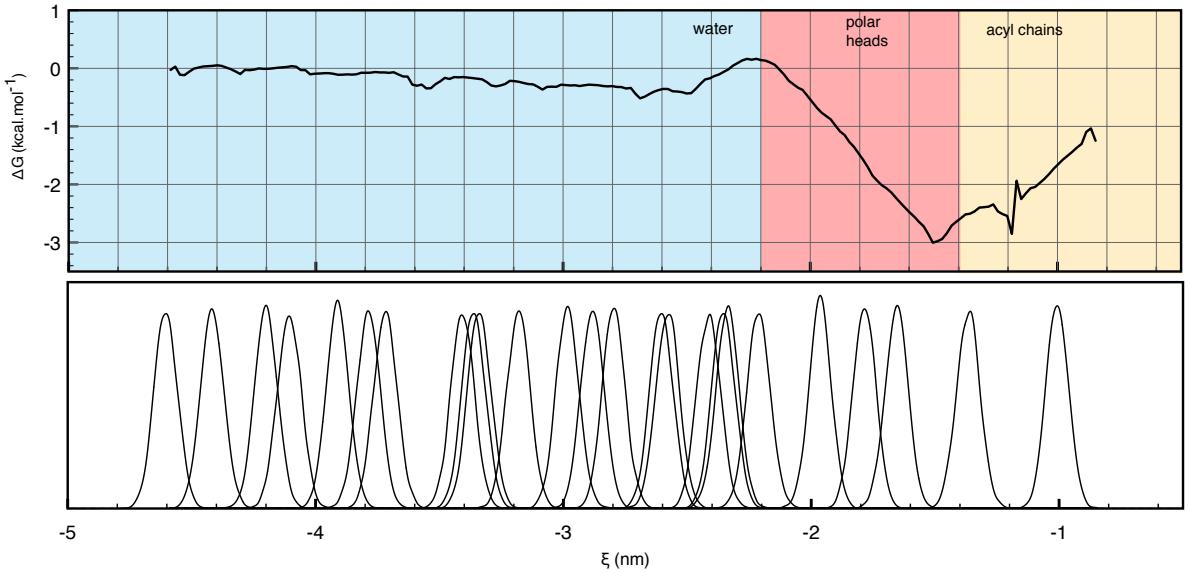


Fig. 23 - PMF profile and umbrella histograms for the validation of the umbrella sampling protocol, considering the caffeine membrane system's simulations. Umbrella sampling was conducted following a 'position' geometry protocol and a harmonic force constant of 1000 kJ•mol^{-1}•nm^{-2} for 25 configurations spaced by 0.2 nm. The free energy in water was set to be zero.

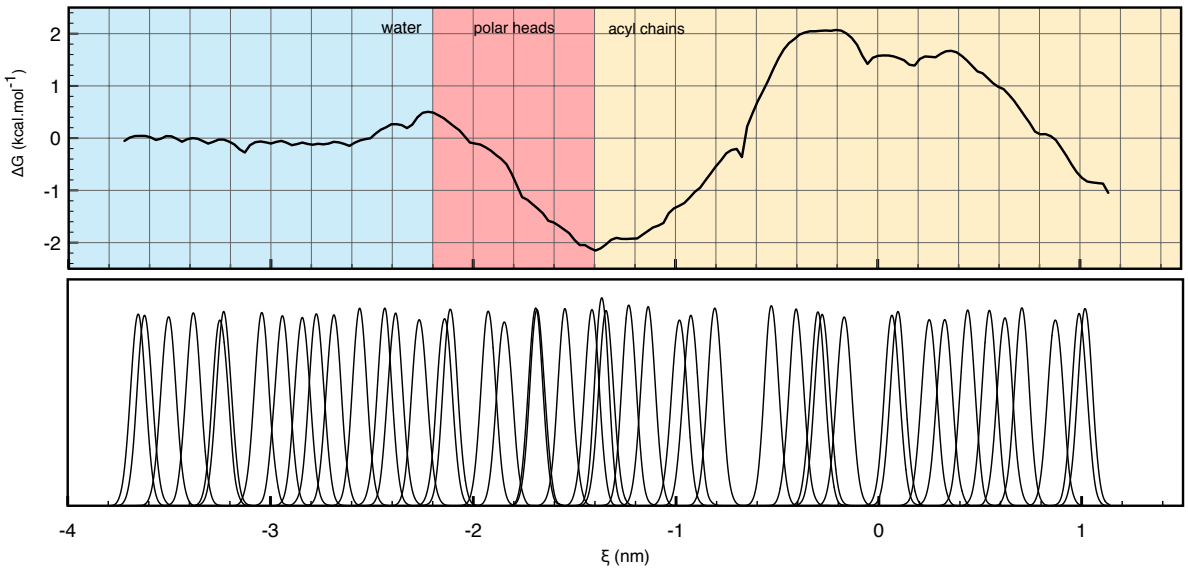


Fig. 24 - PMF profile and umbrella histograms for the validation of the umbrella sampling protocol, considering the caffeine membrane system's simulations. Umbrella sampling was conducted following a 'position' geometry protocol and a harmonic force constant of 2000 kJ•mol^{-1}•nm^{-2} for 46 configurations. The free energy in water was set to be zero.

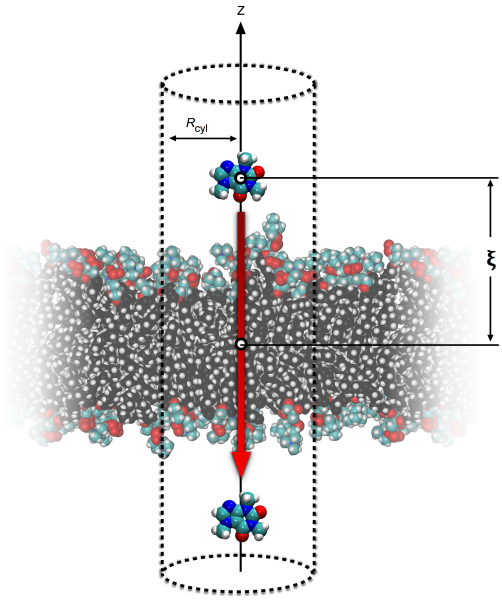


Fig. 25 - Representation of a 'cylinder' geometry' setting. The z-axis is perpendicular to the bilayer. The molecule is confined within a cylinder of radius R_{cyl}. The z coordinate of the compound's center of mass, relatively to the center of membrane defines the reaction coordinate ξ. Adapted from Gordon *et al*.[83]

To apply the 'cylinder' geometry we had to create a new protocol from scratch, since there is no available protocol suitable for our studies. Hence, several umbrella sampling calculations, with 'cylinder' geometry, varying the harmonic force constants, were conducted. Harmonic forces ranging from 500 to 4000 $kJ.mol^{-1}.nm^{-2}$ were tested, and the same problems founded in umbrella sampling with 'position' geometry were found with the 'cylinder' geometry. For several cases, the umbrella histograms presented a large overlap between some windows or the lack of sampling for other regions of the reaction coordinate, as can be seen in Fig. 26. Moreover, displacements of some sampling windows along the reaction coordinate were also detected, that is to say, the position in the umbrella histograms in some sampling windows did not correspond to the reference distance at t=0. By carefully analyzing the literature and the feedback from forums of the GROMACS simulation package, we detected that there could possibly exist an error with the 'cylinder' geometry, in which pull group departs from its initial position, not getting fixed. Such problem was overcome by defining individually the distance (ξ, see Fig. 25) of the pull group to the reference group for each sampling configuration, which allowed for an improvement of the distribution of the umbrella histograms, as can be noted in Fig. 27.A, and the PFM profile shown in Fig. 28.

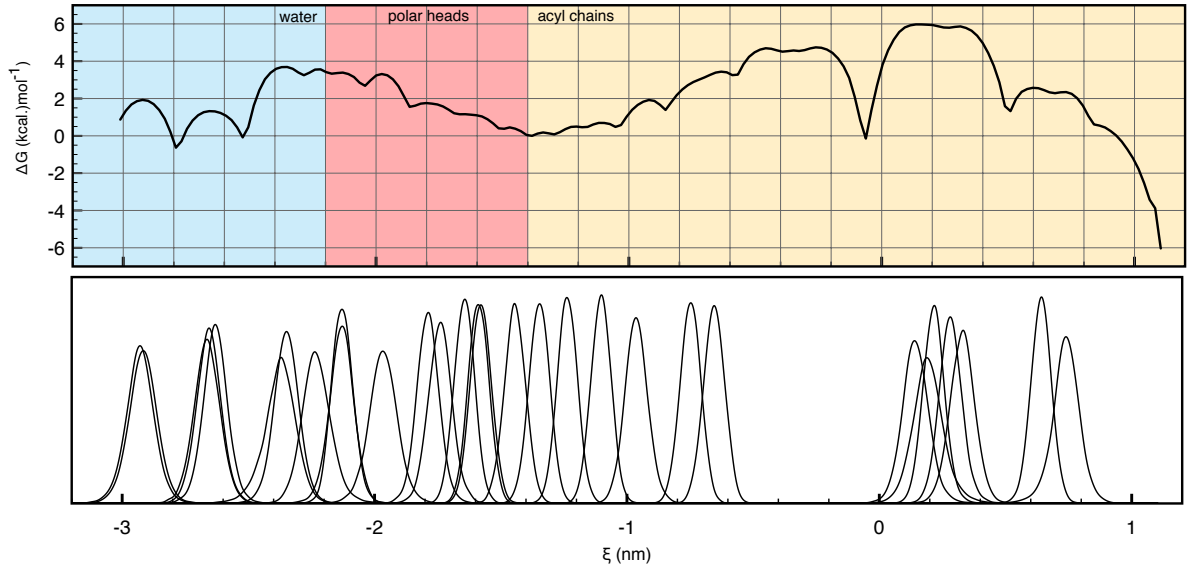


Fig. 26 - PMF profile and umbrella histograms for the validation of the umbrella sampling protocol, considering the caffeine membrane system's simulations. Umbrella sampling was conducted following a 'cylinder' geometry protocol and a harmonic force constant of 2000 kJ•mol^{-1}•nm^{-2}. The free energy in water was set to be zero

This new protocol was then applied to nelfinavir and atorvastatin membrane systems, resulting in sampling windows well overlapped in both cases, Fig. 27.B and C. The PMF profiles are shown in Fig. 29.A and 11.B and will be discussed, in the subsequent chapter, despite not being possible to conclude all sampling windows simulation for the nelfinavir membrane system in time for this thesis' deadline.

4.4 PMF characterization

The free energy profiles for caffeine, nelfinavir, and atorvastatin are shown in Fig. 28, 29.A and 29.B, respectively, and the number of configurations calculated for each molecule and simulation's time are shown on Table VII. As can be seen from Fig. 28, the PMF profile of caffeine shows two minimums with similar values, -0.94 and -0.66 kcal•mol^{-1}, located at -1.37 nm and 1.36 *nm*, respectively (corresponding to symmetric positions within the bilayer). Energy minimum values represent the most energetically favorable positions of the molecule on the bilayer. Both minimums obtained are located in the polar head regions of the bilayer, and this is an expected behavior due to the polar character of caffeine. According to data found in literature, our results are consistent with that data, since it has been described the minimum energy of caffeine is located at the water/lipid interface.[84] The water/lipid barrier, ΔG^{wat}, reflects the affinity of the

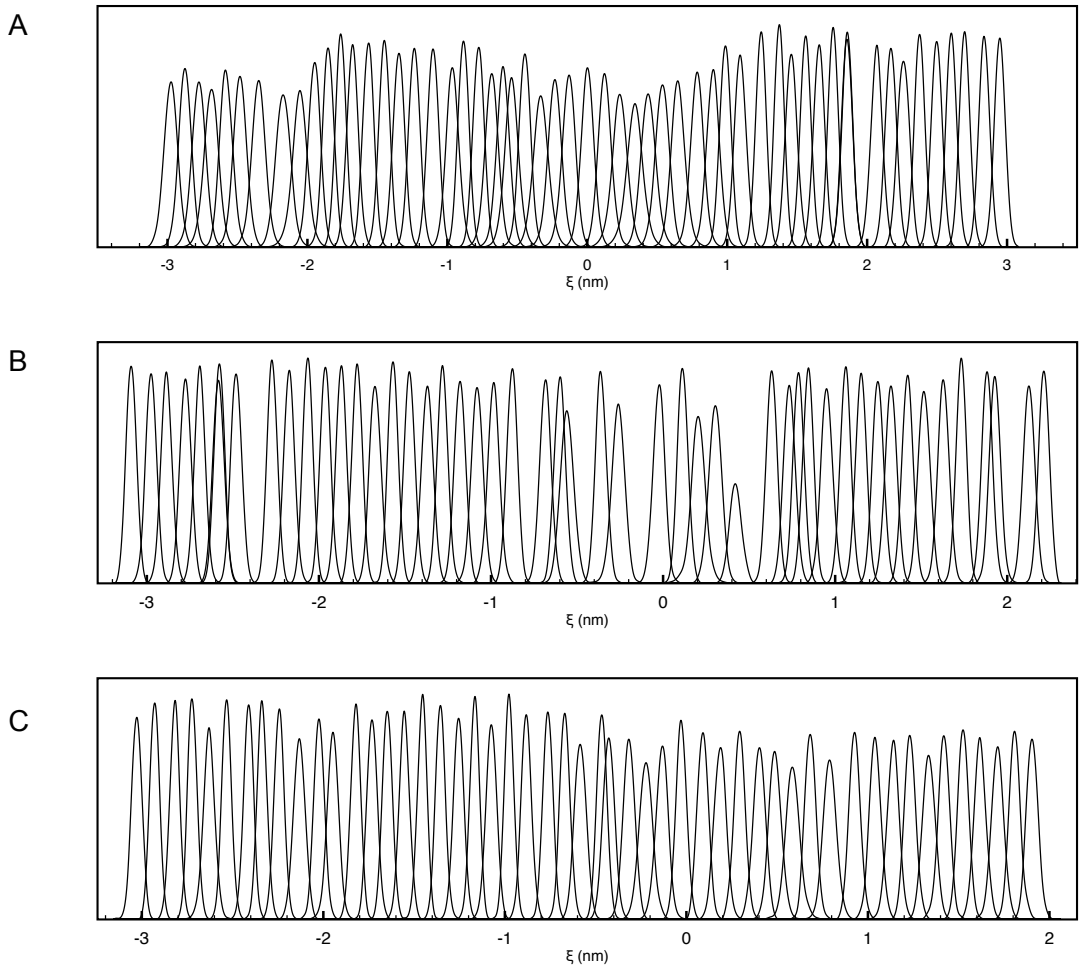


Fig. 27 - Umbrella histograms for the validation of the umbrella sampling protocol, considering the (A) caffeine, (B) nelfinavir, and (C) atorvastatin membrane system's simulations. Umbrella sampling simulations were conducted following a 'cylinder' geometry protocol and a harmonic force constant of 2000 kJ•mol^{-1}•nm^{-2}.

drug to the bilayer in comparison to the water environment, and according to literature[84], a gradual decrease in ΔG^{wat} was expected, but an energy peak at -2.11 nm (water/membrane interface) is observed instead. However, at the opposite site, at 2.11 nm (opposite water/membrane interface), no energy peak is detected, which means that our PMF profile is not symmetrical as it was supposed.

The shape of free energy profiles for nelfinavir and atorvastatin are different from that for caffeine. Looking at PMF profile for nelfinavir in Fig. 29.A, the energy decreases with nelfinavir coming from bulk water to DOPC membrane until it reaches a minimum (-2.82 kcal•mol^{-1} at -1.82 nm (polar heads)). Then, the energy starts to increase as it approaches the bilayer center, and continues to rise till +0.72 nm (acyl chains), declining further to reach a new minimum at 1.33 nm (polar heads/acyl chains interface) (19.25 kcal•mol^{-1}).

Table VII - Number of configurations and total simulations' time, taking into account 6 ns per configuration.

Molecule	No. of ξ points	Simulation's time / ns
Caffeine	56	336
Atorvastatin	50	300
Nelfinavir	51	306

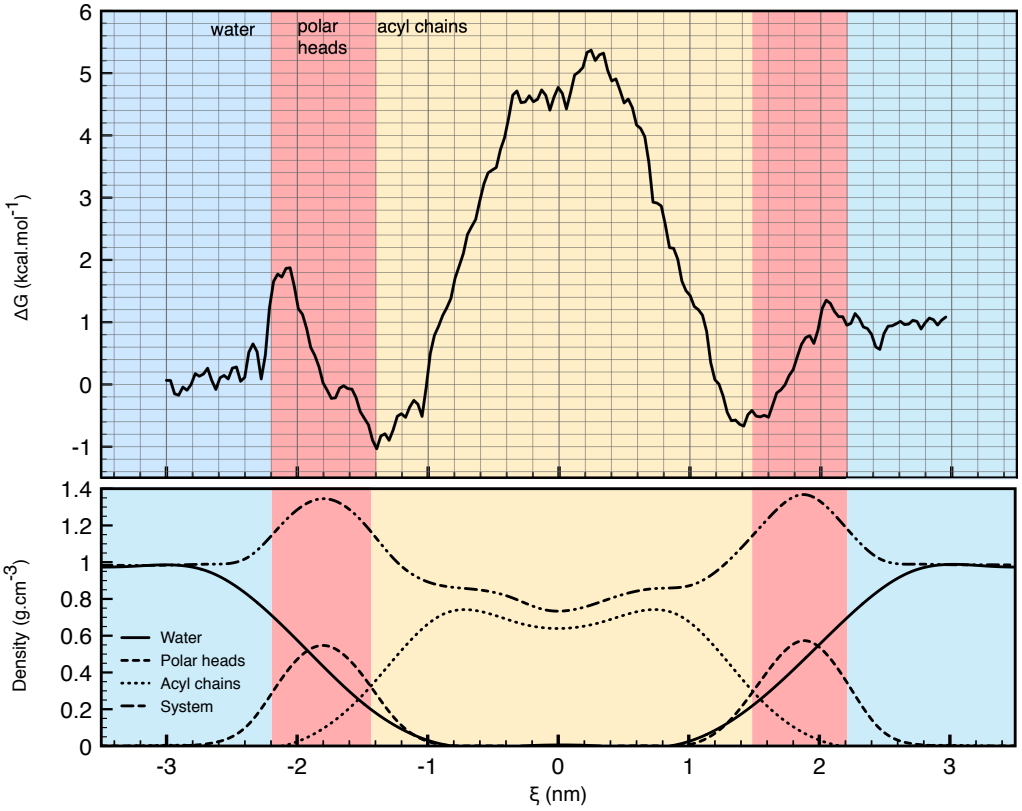


Fig. 28 - PMF profile of caffeine membrane system's simulation. Umbrella sampling was conducted following a 'cylinder' geometry protocol and a harmonic force constant of 2000 kJ•mol^{-1}•nm^{-2}. Density profiles for different membrane groups are also shown. The free energy in water was set to be zero.

Similarly to nelfinavir, in the PMF profile for atorvastatin in Fig. 29.B we can observe a slightly decrease in energy as atorvastatin comes from the bulk water to DOPC membrane until it reaches a minimum (-0.77 kcal•mol^{-1}at -1.62 nm (polar heads)). Then, the energy also starts to increase as it approaches the bilayer center, and continues to rise till the +0.85 nm (acyl chains) position.

Nevertheless, a decrease in the free energy after the membrane center for nelfinavir and atorvastatin was expected. The continuous increase in energy after the membranes' center might be due to several factors. A trajectory analysis of the nelfinavir

umbrella sampling simulation revealed the drug assumes the same orientation along the reaction coordinate: when nelfinavir enters the membrane, the more apolar groups are facing the acyl chains, whereas when it is coming out of the other side of the membrane the same groups are facing the water. This can naturally result in a less stable orientation, which can lead to an increase in free energy; however, we must understand whether this effect is real or a misleading result. In addition, nelfinavir is a bulky molecule and due to the membrane's viscosity it tends to move more difficulty and more slowly inside the membrane. Thus, a simulation time of 5 ns per configuration seems to be not enough so nelfinavir can explore all conformations, and we must increase the simulation time. In the case of atorvastatin, when its trajectory was analyzed, we noted the orientation of atorvastatin changes along the reaction coordinate. However, because atorvastatin has a higher MW and it might be stiffer than nelfinavir, it becomes even tougher to move around the membrane. Once again, we must increase the simulations time in order to improve the free energy profiles.

Drug's permeability depends largely on the bilayer center penetration barrier, ΔG^{pen}, which is related to the velocity of transfer of the molecule to the other monolayer (bilayer leaflet), and is calculated by the difference between the minimal free energy and the maximal free energy. The values of ΔG^{pen} for caffeine, nelfinavir, and atorvastatin, as well as the experimental permeability values are shown in Table VIII, where an inverse correlation between ΔG^{pen} and permeability can be observed. Caffeine, which is the most permeable compound from the three tested, has the lowest ΔG^{pen}, whereas nelfinavir, which is the less permeable, has the highest ΔG^{pen}. Despite the non-perfect profiles that we obtained for these compounds we could still correlate the ΔG^{pen} with the compounds' permeability. Seeing that one of the main goals of this work is establish the molecular properties that are essential for the passive diffusion of compounds through cell membranes, or even predict new ones, these are very promising results that could change the way molecular diffusion is characterized. Obviously, the protocol needs to be extensively validated and refined for us to achieve such promise.

Table VIII - The calculated water/lipid barrier, bilayer center penetration barrier, and experimental permeability.

Molecule	ΔG^{wat} / kcal•mol^{-1}	ΔG^{pen} / kcal•mol^{-1}	Permeability [a] / cm•s^{-1}
Caffeine	0.94 / 1.60	6.18	50.5 x10^{-6}
Atorvastatin	0.77	22.31	4.9 x10^{-6}
Nelfinavir	2.82	25.80	1.39 x10^{-6}

[a] high permeability ≥ 10 x10^{-6} cm•s^{-1}; 2.5 x10^{-6} ≤ moderate permeability < 10 x10^{-6} cm•s^{-1}; low permeability < 2.5 x10^{-6} cm•s^{-1}

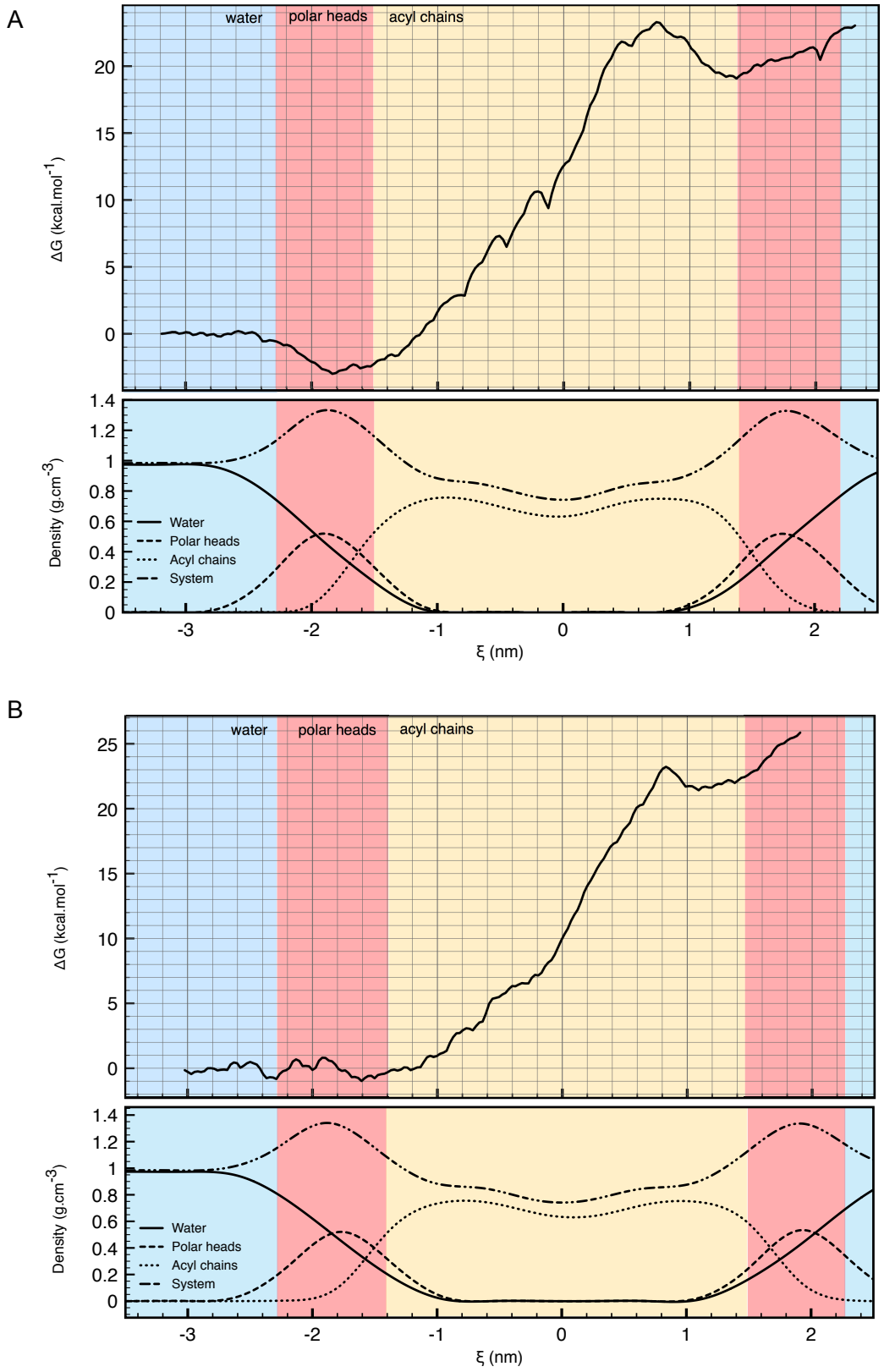


Fig. 29 - PMF profile of (A) nelfinavir and (B) atorvastatin membrane system's simulations. Umbrella sampling was conducted following a 'cylinder' geometry protocol and a harmonic force constant of 2000 kJ•mol^{-1}•nm^{-2}. Density profiles for each membrane are also shown. The free energy in water was set to be zero.

4.5 Rotatable bonds and PSA characterization

Nelfinavir and atorvastatin have 10 and 12 rotatable bonds, respectively, and for both molecules the rotatable bonds are shown in Fig30. From the figure it may be possible to predict if nelfinavir is more flexible than atorvastatin; as can be seen, nelfinavir has a carbon atom that makes part of 3 rotatable bonds (see Fig. 30.A), which mean that this atom is a kind of 'rotatable center'. A visual analysis of nelfinavir simulation allows us to observe not only a different molecular groups orientation over time as was expected, but also a change in its shape, as in can be seen in Fig. 31. Therefore, we asked our-

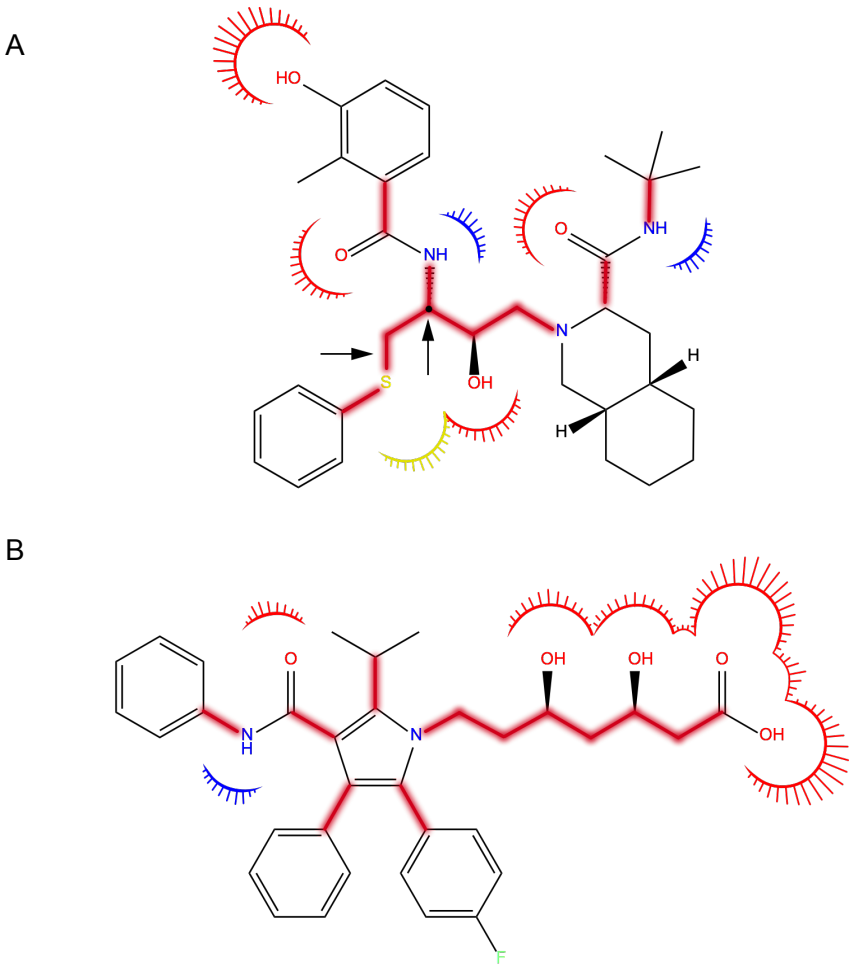


Fig. 30 - Nelfinavir (A) and Atorvastatin's (B) chemical representations with the rotatable bonds highlighted in red and the polar surface area depicted as the incomplete spheres surrounding the colored hetero-atoms: nitrogen (blue), oxygen (red)(and any hydrogen atom attached to these atoms), and sulfur (yellow). The vertical arrow is pointing the 'rotatable center' and the horizontal one is pointing the most relevant rotatable bond for nelfinavir's conformation (see text for more details).

selves whether the changes in nelfinavir shape could result in PSA changes. In order do answer this question, first we calculated the average value for PSA from a simulation of nelfinavir inside a water box; secondly, we have verified the most relevant rotatable bonds for shape flickers; and thirdly, we run simulations with those rotatable bonds constricted to evaluate the PSA.

The rotatable bond pointed out in Fig. 30.A was the one that showed a greater relevance for nelfinavir's conformation. As can be seen in Fig. 32, the dihedral angle can be at -70°, +55º, or even at +180º. Thus, three simulations, in which this rotatable bond was constrained at each of those angle values, were performed. The resulting PSA values calculated from each simulation, as well as the value calculated for the free simulation in water, are shown in Table IX. As it can be observed, the difference between the PSA values in each case is negligible, which means that a relation between the dihedral angle and PSA cannot be deduced for now. As a matter of fact, this was a preliminary study, and other studies will be needed. Nonetheless, the PSA value obtained for nelfinavir is somehow concordant with the value found in literature.

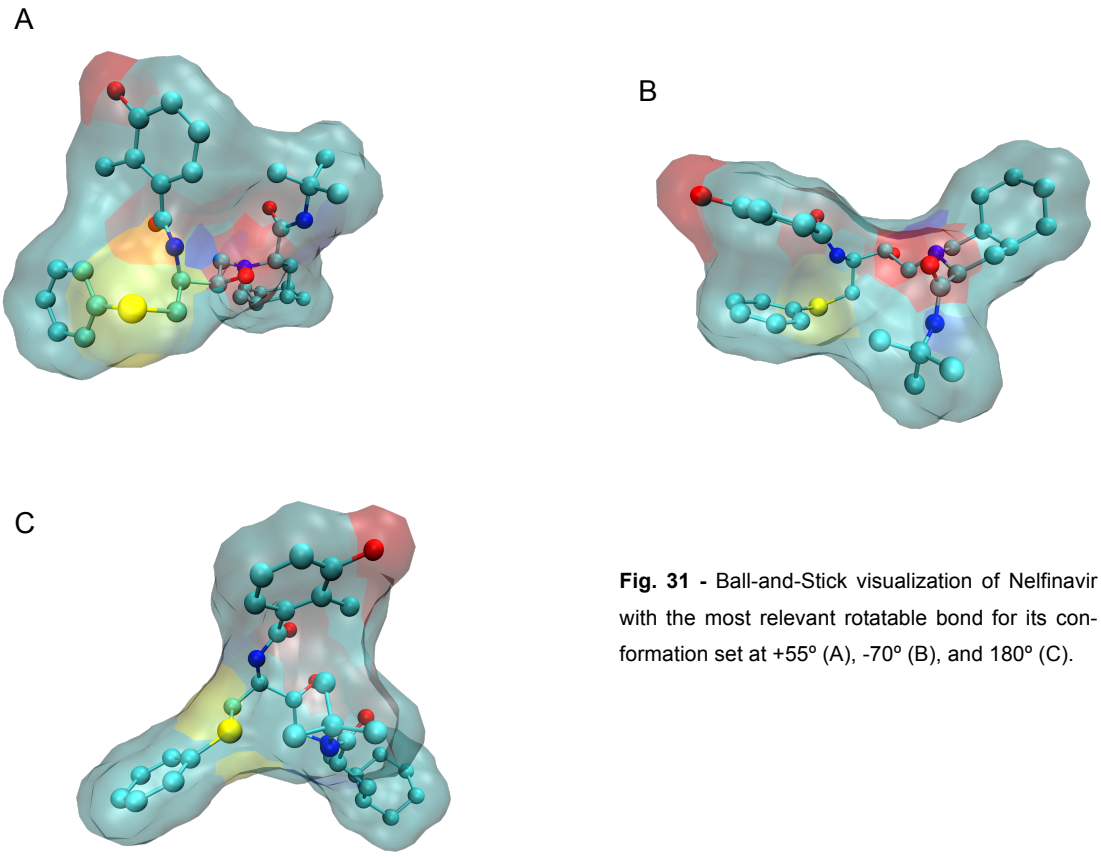


Fig. 31 - Ball-and-Stick visualization of Nelfinavir with the most relevant rotatable bond for its conformation set at +55º (A), -70º (B), and 180º (C).

Relatively to atorvastatin, due to a matter of time, it was only calculated its PSA value, yielding 110.68 $Å^2$, which is similar to the value found in the literature (see Table I in section 3.3).

In order to try to understand Veber's rules: why rigidity facilitates the diffusion process, and how rotatable bonds influence the molecular flexibility; as well as how rotatable bonds also influence the PSA, umbrella sampling simulations in a membrane system, in which rotatable bonds are constrained, will be performed prospectively.

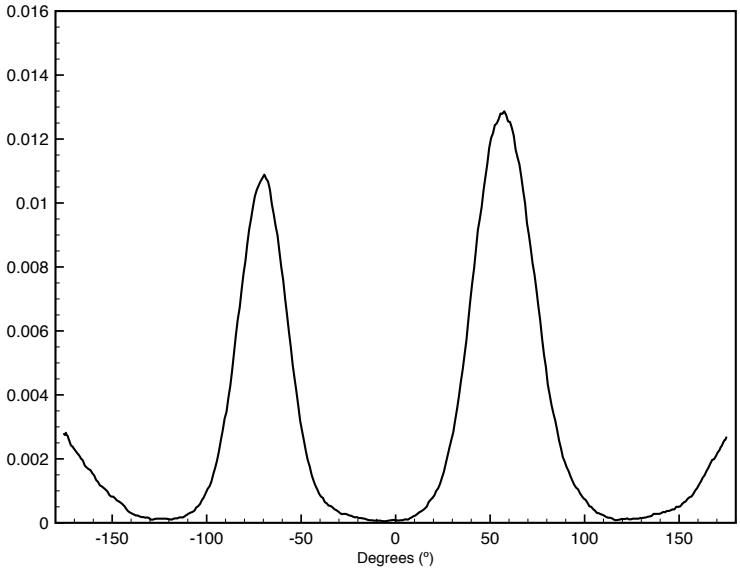


Fig. 32 - Dihedral angle distribution of the most relevant rotatable bond for Nelfinavir's conformation (see text for more details).

Table IX - PSA values obtained for nelfinavir.

Constraint / Degrees (º)	**Polar Surface Area /** $Å^2$
No constraint	105.40
+55	99.16
-70	108.00
+180	96.17

CHAPTER 5

Conclusions

In this thesis umbrella sampling simulations have been used to study the fundamental processes that occur during passive diffusion of drugs through biological membranes.

We began our study by selecting a small library of compounds, whose free energies in water and in 1-octanol were calculated in order to validate the force field that was being used. Regarding free energy calculations in water, we obtained an acceptable error of 1.15 kcal•mol^{-1}, on average, between our computationally determined values and the experimental ones, which supports the accuracy of the force field that was being used in describing hydration free energies for this subset of compounds. Concerning free energies calculations in 1-octanol, a discrepancy of 2.33 kcal•mol^{-1} between the theoretical and experimental ΔG_{solv} values was observed, which is an indicator of the force field's accuracy limits. To improve this result, we should working on 1-octanol parameters refinement. Nevertheless, such error cannot impart the results obtained for membrane studies, since DOPC's parameters have been extensively validated.

In a second stage of the work, we built a DOPC/water membrane model, in which a molecule of caffeine was included, and we started to optimize the umbrella sampling protocols. To do that, we tested a series of important settings: simulation time, harmonic force constant, pull geometry, and pull rate. From all pull geometries in GROMACS, only two of them were used: *position* and *cylinder*. In the end, the *cylinder* geometry has revealed itself as the best pull geometry option for these systems; not only it was specifically designed for layered systems, allowing for the minimization of membrane undulation effects, but also it yielded the best sampling results. The resulting PMF for caffeine, with the cylinder geometry was very similar to the one found in literature.

In the next stage, umbrella sampling simulations for nelfinavir and atorvastatin were performed. Although the obtained PMF profiles had shown that we must increase the simulation time, since a simulation time of 5 ns per configuration seemed to be not

enough so nelfinavir could explore all conformations, we could still observe an inverse correlation between the ΔG^{pen} with permeability: caffeine, which is the most permeable compound from the three tested, has the lowest ΔG^{pen}, whereas nelfinavir, which is the less permeable, has the highest ΔG^{pen}. Consequently, these promising results may be a turning point in the way molecular diffusion is characterized. However, extensive studies will be needed to achieve such promise.

Lastly, nelfinavir and atorvastatin's rotatable bonds and PSA were characterized. We observed that although nelfinavir has less two rotatable bonds than atorvastatin, it has a 'rotatable center', that is to say, it has a carbon atom that makes part of 3 rotatable bonds, which probably makes it more flexible than atorvastatin. Seeing that molecular flexibility hinds the passive diffusion, such observation indicates nelfinavir has a bad permeability, which is confirmed by experimental data. As a result, molecular flexibility can be no longer defined exclusively by the quantity of rotatable bonds, but also by its distribution in the molecule. However, some interesting questions arose: 1) why molecular flexibility hinds passive diffusion; 2) whether stiffness is important at any part of the molecule; 3) how many conformational states does the molecule have in solution; 4) how PSA changes with the rotatable bonds count and rotatable bounds distribution.

To answer this question, we propose, in the future, to constrain rotatable bonds to decrease the flexibility, and perform umbrella sampling calculations. We also propose a previous rotatable bonds' characterization in the water/lipid interface and in the core of the membrane, in order to see in which values rotatable bonds should be constrained.

To conclude, in this work, we developed a new protocol to study the molecular properties required for drugs' passive diffusion, which can be subsequently applied to a large number of molecules.

References

1. Kennedy D (2004) Drug Discovery. Science 303:1729.

2. Drews J (2000) Drug Discovery: A Historical Perspective. Science 287:1960–1964.

3. Lombardino JG, Lowe JA (2004) The role of the medicinal chemist in drug discovery--then and now. Nat Rev Drug Discov 3:853–862.

4. Wess G, Urmann M, Sickenberger B (2001) Medicinal Chemistry: Challenges and Opportunities. Angew Chem Int Ed Engl 40:3341–3350.

5. Jorgensen WL (2004) The Many Roles of Computation in Drug Discovery. Science 303:1813.

6. Hughes JP, Rees S, Kalindjian SB, Philpott KL (2011) Principles of early drug discovery. Br J Pharmacol 162:1239–1249.

7. Keserű GM, Makara GM (2006) Hit discovery and hit-to-lead approaches. Drug Discovery Today 11:741–748.

8. Walters WP, Green J, Weiss JR, Murcko MA (2011) What do medicinal chemists actually make? A 50-year retrospective. J Med Chem 54:6405–6416.

9. Ou-Yang S, Lu J, Kong X, et al. (2012) Computational drug discovery. Acta Pharmacologica Sinica 33:1131–1140.

10. Clark DE, Pickett SD (2000) Computational methods for the prediction of “drug-likeness.” Drug Discovery Today 5:49–58.

11. Lodish H, Berk A, Kaiser CA, et al. (2012) Molecular Cell Biology. W. H. Freeman

12. Nelson DL, Cox MM (2013) Lehninger Principles of Biochemistry. W.H. Freeman

13. Balaz S (2012) Does transbilayer diffusion have a role in membrane transport of drugs? Drug Discovery Today 17:1079–1087.

14. Li AP (2001) Screening for human ADME/Tox drug properties in drug discovery. Drug Discovery Today 6:357–366.

15. Rang HP, Ritter JM, Flower RJ, Henderson G (2014) Rang & Dale's Pharmacology. Elsevier Health Sciences UK

16. Lipinski CA (2000) Drug-like properties and the causes of poor solubility and poor permeability. J Pharmacol Toxicol Methods 44:235–249.

17. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews 46:3–26.

18. Lipinski CA (2004) Lead- and drug-like compounds: the rule-of-five revolution. Drug Discov Today Technol 1:337–341.

19. Lajiness MS, Vieth M, Erickson J (2004) Molecular properties that influence oral drug-like behavior. Curr Opin Drug Discov Devel 7:470–477.

20. Navia MA, Chaturvedi PR (1996) Design principles for orally bioavailable drugs. Drug Discovery Today.

21. Hann MM, Leach AR, Harper G (2001) Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* **2001,** *41,* 856-864

22. Veber DF, Johnson SR, Cheng HY (2002) Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 45:2615–2623

23. Ertl P, Rohde B, Selzer P (2000) Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. J Med Chem 43:3714–3717.

24. Palm K, Stenberg P, Luthman K, Artursson P (1997) Polar molecular surface properties predict the intestinal absorption of drugs in humans. Pharm Res 14:568–571.

25. Kelder J, Grootenhuis P, Bayada DM, et al. (1999) Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs. Pharm Res 16:1514–1519.

26. Leach AR (2001) Molecular Modelling: Principles and Applications. Prentice Hall

27. Sato F, Hojo S, Sun H (2003) On the transferability of force field parameters with an ab initio force field developed for sulfonamides. The Journal of Physical Chemistry A 107:248–257.

28. Becker OM, MacKerell AD, Roux B, Watanabe M (2001) Computational Biochemistry and Biophysics. Taylor & Francis

29. Ponder JW, Case DA (2003) Force fields for protein simulations. Adv Protein Chem 66:27–85.

30. Case DA, Cheatham TE, Darden T, et al. (2005) The Amber biomolecular simulation programs. J Comput Chem 26:1668–1688.

31. Wang JM, Wolf RM, Caldwell JW, et al. (2005) Development and testing of a general amber force field (vol 25, pg 1157, 2004). J Comput Chem 26:114–114.

32. Jämbeck J, Lyubartsev AP (2012) An extension and further validation of an all-atomistic force field for biological membranes. J. Chem. Theory Comput. 2012, 8, 2938−2948

33. Jämbeck JPM, Lyubartsev AP (2012) Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids. J Phys Chem B 116:3164–3179.

34. Jämbeck J, Lyubartsev AP (2012) Another piece of the membrane puzzle: extending slipids further. J. Chem. Theory Comput. 2013, 9, 774−784

35. Burger SK, Schofield J, Ayers PW (2013) Quantum Mechanics/Molecular Mechanics Restrained Electrostatic Potential Fitting. J Phys Chem B 117:14960–14966.

36. Meller JA (2001) Molecular Dynamics. ENCYCLOPEDIA OF LIFE SCIENCES, John Wiley & Sons, Ltd

37. Garrido NM, Queimada AJ, Jorge M (2009) 1-octanol/water partition coefficients of n-alkanes from molecular simulations of absolute solvation free energies. *J. Chem. Theory Comput., Vol. 5, No. 9, 2009*

38. Villa A, Mark AE (2002) Calculation of the free energy of solvation for neutral analogs of amino acid side chains. J Comput Chem 23:548–553.

39. de Ruiter A, Boresch S, Oostenbrink C (2013) Comparison of thermodynamic integration and Bennett's acceptance ratio for calculating relative protein-ligand binding free energies. J

Comput Chem 34:1024–1034.

40. Bennett CH (1976) Efficient estimation of free energy differences from Monte Carlo data. Journal of Computational Physics 22:245–268.

41. Christ CD, Mark AE, van Gunsteren WF (2010) Basic ingredients of free energy calculations: a review. J Comput Chem 31:1569–1582.

42. Darve E, Pohorille A (2001) Calculating free energies using average force. J Chem Phys 115:9169.

43. Darve E, Wilson MA, Pohorille A (2002) Calculating free energies using a scaled-force molecular dynamics algorithm. Molecular Simulation 28:113–144.

44. Torrie GM, Valleau JP (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. Journal of Computational Physics 23:187–199.

45. Kästner J (2011) Umbrella sampling. Wiley Interdisciplinary Reviews: Computational Molecular Science

46. Rosta E, Woodcock HL, Brooks BR, Hummer G (2009) Artificial reaction coordinate “tunneling” in free-energy calculations: the catalytic reaction of RNase H. J Comput Chem 30:1634–1641.

47. Lemkul JA Umbrella Sampling.
http://www.bevanlab.biochem.vt.edu/Pages/Personal/justin/gmx-tutorials/umbrella/index.html

48. Lemkul JA (2013) Computing Potential of Mean Force in GROMACS.
http://www.gromacs.org/@api/deki/files/208/=pmf_tutorial_slides.pdf

49. Nicholls A, Wlodek S, Grant JA (2010) SAMPL2 and continuum modeling. Journal of computer-aided molecular

50. Wang J, Wang W, Kollman PA, Case DA (2001) Antechamber: an accessory software package for molecular mechanical calculations. J. Am. Chem. Soc.

51. BAYLY CI, CIEPLAK P, CORNELL WD, Kollman PA (1993) A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges - the Resp Model. Journal of Physical Chemistry 97:10269–10280.

52. Sousa da Silva AW, Vranken WF (2012) ACPYPE - AnteChamber PYthon Parser interfacE. BMC Res Notes 5:367.

53. van der Spoel D, Lindahl E, Hess B, et al. (2005) GROMACS: fast, flexible, and free. J Comput Chem 26:1701–1718.

54. Jorgensen WL, Chandrasekhar J, Madura JD, et al. (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926–935.

55. Liu DC, Nocedal J (1989) On the limited memory BFGS method for large scale optimization. Mathematical Programming 45:503–528.

56. Van Gunsteren WF, Berendsen HJC (1988) A Leap-frog Algorithm for Stochastic Dynamics. Molecular Simulation 1:173–185.

57. Hess B, Bekker H, Berendsen H, Fraaije J (1997) LINCS: A linear constraint solver for molecular simulations. J Comput Chem 18:1463–1472.

58. PARRINELLO M, RAHMAN A (1981) Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. Journal of Applied Physics 52:7182–7190.

59. Chipot C, Pohorille A (2007) Free Energy Calculations: Theory and Applications in Chemistry and Biology. Springer Berlin Heidelberg

60. Martinez L, Andrade R, Birgin EG, Martinez JM (2009) PACKMOL: A Package for Building Initial Configurations for Molecular Dynamics Simulations. J Comput Chem 30:2157–2164.

61. Martínez JM, Martínez L (2003) Packing optimization for automated generation of complex system's initial configurations for molecular dynamics and docking. J Comput Chem 24:819–825.

62. Coimbra JTS, Sousa SF, Fernandes PA, et al. (2014) Biomembrane simulations of 12 lipid types using the general amber force field in a tensionless ensemble. J Biomol Struct Dyn 32:88–103.

63. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126:014101.

64. Yaws CL (2014) Thermophysical Properties of Chemicals and Hydrocarbons. Elsevier Science

65. Jo S, Kim T, Im W (2007) Automated builder and database of protein/membrane complexes for molecular dynamics simulations. PLoS ONE 2:e880. doi:10.1371/journal.pone.0000880

66. Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem 29:1859–1865.

67. Jo S, Lim JB, Klauda JB, Im W (2009) CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. Biophysical Journal 97:50–58.

68. Wu EL, Cheng X, Jo S, et al. (2014) CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. J Comput Chem 35:1997–2004.

69. Fernandes PA, Araujo MLAC, Barbosa AJM, et al. (2006) Lead optimisation: Improving the affinity of the antiretrovirals Nelfinavir and Amprenavir for HIV-1 protease. Letters in Drug Design & Discovery 3:383–389.

70. Perez MAS, Fernandes PA, Ramos MJ (2007) Drug design: New inhibitors for HIV-1 protease based on Nelfinavir as lead. Journal of Molecular Graphics and Modling 26:634–642.

71. VERLET L (1967) Computer Experiments on Classical Fluids .I. Thermodynamical Properties of Lennard-Jones Molecules. Physical Review 159:98

72. Nosé S (1984) A unified formulation of the constant temperature molecular dynamics methods. J Chem Phys 81:511–519.

73. Hoover WG (2011) Canonical dynamics: Equilibrium phase-space distributions. Phys. Rev. A 31, 1695.

74. Guthrie JP (2009) A blind challenge for computational solvation free energies: introduction and overview. J Phys Chem B 113:4501–4507.

75. Jämbeck JPM, Lyubartsev AP (2014) Update to the general amber force field for small solutes with an emphasis on free energies of hydration. J Phys Chem B 118:3793–3804.

76. ULRICH AS, SAMI M, WATTS A (1994) Hydration of Dopc Bilayers by Differential Scanning Calorimetry. BBA - Biomembranes 1191:225–230.

77. Rosso L, Gould IR (2008) Structure and dynamics of phospholipid bilayers using recently developed general all-atom force fields. J Comput Chem 29:24–37.

78. Nagle JF, Tristram-Nagle S (2000) Structure of lipid bilayers. Biochim Biophys Acta 1469:159–195.

79. Pan J, Tristram-Nagle S, Kucerka N, Nagle JF (2008) Temperature dependence of structure, bending rigidity, and bilayer interactions of dioleoylphosphatidylcholine bilayers. Biophysical Journal 94:117–124.

80. Kucerka N, Nagle JF, Sachs JN, et al. (2008) Lipid bilayer structure determined by the simultaneous analysis of neutron and X-ray scattering data. Biophysical Journal 95:2356–2367.

81. Warschawski DE, Devaux PF (2005) Order parameters of unsaturated phospholipids in membranes and the effect of cholesterol: a 1H-13C solid-state NMR study at natural abundance. Eur Biophys J 34:987–996.

82. Lemkul JA, Bevan DR (2010) Assessing the stability of Alzheimer's amyloid protofibrils using molecular dynamics. J Phys Chem B 114:1652–1660.

83. Gordon D, Chen R, Chung S-H (2013) Computational Methods of Studying the Binding of Toxins From Venomous Animals to Biological Ion Channels: Theory and Applications. Physiological Reviews 93:767–802.

84. Paloncýová M, Berka K, Otyepka M (2013) Molecular insight into affinities of drugs and their metabolites to lipid bilayers. J Phys Chem B 117:2403–2410.