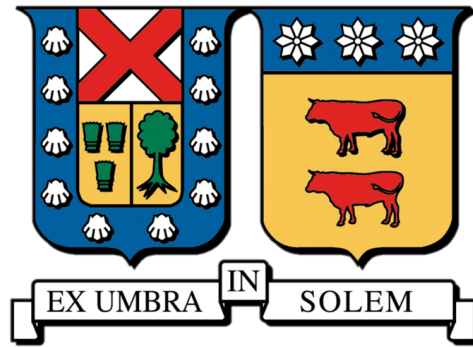


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

DEPARTAMENTO DE INGENIERÍA MECÁNICA
SANTIAGO, CHILE



ESTUDIO DEL COMPORTAMIENTO CINÉTICO Y
TERMODINÁMICO DE LLENADO DE NANOTUBOS
ESTRECHOS DE CARBONO SUJETOS A CAMPOS
ELÉCTRICOS, A TRAVÉS DE SIMULACIONES DE
DINÁMICA MOLECULAR CLÁSICA

IGNACIO TOMÁS MUÑOZ QUIROZ

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL MECÁNICO

Profesor Guía: PhD. Christopher Cooper

Profesor Coreferente: PhD. Harvey A. Zambrano

Profesor Coreferente: PhD. José Antonio Gárate

MES - AÑO

Contents

1	Introduction	3
1.1	Life is based on chemical interactions	3
1.2	Molecular Structure	3
1.2.1	Atomic Structure: Bohr's Model	4
1.2.2	Atomic Structure: Quantum Mechanics	4
1.3	Computer Simulation of Molecular Systems	6
1.4	Objectives	7
2	Molecular Dynamics	8
2.1	Degrees of freedom	9
2.2	Potential Energy Function: The Force Field	11
2.2.1	Force-Field Bonded Interactions	12
2.2.2	Force-Field Non-bonded Interactions	15
2.3	Configurations of Atoms	18
2.4	Boundary Conditions	19
2.5	General Molecular Dynamics Algorithm	20
2.5.1	Numerical Integration	22
3	Implicit Solvent Models	24
3.1	Implicit model: Solvation Free Energy	24
3.1.1	Solvent-Accessible Surface Area (SASA) Model	26
3.2	Non-polar Solvation Free Energy Estimation	26

3.2.1	Implicit Solvent Model for ΔG_{cav} and ΔG_{disp}	27
4	Free Energy	30
4.1	Thermodynamic Integration	31
4.1.1	Gibbs Free Energy	33
4.1.2	Soft-Core Potential Energy	35
5	Methods	37
5.1	Structure Preparation	37
5.2	About GROMOS software	37
5.2.1	Building a topology	38
5.2.2	Generating atomic cartesian coordinates	39
5.2.3	Energy minimisation of the peptide	39
5.2.4	Solvating the peptide	40
5.2.5	Thermalization and equilibration	41
5.3	Molecular Dynamics Simulation	42
5.4	Free Energy Calculations	43
5.4.1	Thermodynamic Integration's Numerical Error	45
6	Results and Discussion	46
6.1	Energy Interaction Results	46
6.2	Free energy Results	48
6.2.1	Statistical Error	49
6.3	Comparison Between Models	52

6.3.1	Adjust Models constants	55
6.4	Conclusions	56

List of Tables

1	Acronyms used in the text and their definitions.	viii
2	Mass and Charge of individual protons, neutrons and electrons.	4
3	Examples of levels of modeling in computational biochemistry and molecular biology. Table from from Van Gunsteren et al. [1]	10
4	Summarize of the 20 system's dimensions and the number of atoms to be simulated.	42
5	Distribution of 32 λ points between 0 and 1.	45
6	Summarize of the 20 free energies calculations with statistical error.	50
7	Comparison of ΔG_{np} between standard SASA model and the Implicit model proposed.	54
8	Comparison of ΔG_{np} between standard SASA model and the Implicit model with the re-parametrization.	55

List of Figures

1	Schematics showing the general shapes of s, p, d, and f orbitals.	5
2	Four basic choices in the definition of a model for molecular simulation. Figure from Van Gunsteren et al. [1]	9
3	Illustration of the various terms included in an empirical potential energy function. Figure adapted from C. Chipot [2]	14
4	Illustration of the sequence of events that generates the Van der Waals forces.	16
5	Graphic representation of the attractive and repulsive terms of the Lenard Jones Potential	17
6	The Lennard-Jones Potential curve, calculated for $\epsilon = 0.0103$ and $\sigma = 3.8$.	18
7	Three types of spatial boundary conditions used in molecular simulation. Figure from Van Gunsten et al [1]	20
8	Molecular dynamics basic algorithm. Figure from Hospital <i>et al.</i> [3]	22
9	The leap-frog integration scheme. Figure from Van Gunsteren: Gromos Man- ual Vol.2	23
10	Representation of the solvation process for any molecule in solution.	25
11	Thermodynamic cycle for solvation energy. Figure taken from [4]	27
12	Sketch of the capacitor model. Figure taken from [4]	28
13	Thermodynamic cycle for the determination of solvation free energies . . .	32
14	Transition between the states A and B with according to coupling parameter λ	35
15	Graphic representation of both penta-Alanines (helix and extended) config- urations generated with VMD.	38
16	Penta-alanine in box full of water molecules.	41
17	The kinetic energy during the equilibration process of the helix penta-alanine.	43

18	Graphic representation of the process of Thermodynamic Integration where a molecules is turn into a dummy molecule according to the couple parameter λ . The transition goes froms State A to State B.	44
19	Evolution of interactions between peptide and solvent while the transition of the λ parameter.	46
20	Evolution of van der Waals interactions between the solvent with itself while the transition of the λ parameter.	47
21	Evolution of coulomb interactions between the solvent with itself while the transition of the λ parameter.	48
22	TI curve for hepta-alanine in helix configuration.	49
23	Solvation free energies of the extended and helix alanine peptides plotted as functions of the number (N) of residues in the peptide.	50
24	Gibbs free energy results from the three models described in this work for extended configuration, per alanine peptide.	52
25	Gibbs free energy results from the three models described in this work for helix configuration, per alanine peptide.	53
26	Differences between SASA factor generated with GROMOS software and msms program.	54
27	Results after adjustment of parameters $\gamma = 0.025$, $b = -4.1$ for SASA standard model, and $\epsilon_{shell} = 3.1$ for the implicit model presented in this work.	55

Acronymus

Acronym	Definition
CM	Classical mechanics
GROMOS	GRoningen MOlecular Simulation
MC	Monte Carlo
MD	Molecular dynamics
PDB	Protein Data Bank
QM	Quantum mechanics
SASA	Solvent-Accessible Surface Area
TI	Thermodynamic Integration
VMD	Visual Molecular Dynamics
Z	Atomic number

Table 1: Acronyms used in the text and their definitions.

Abstract

The free energy of solvation ΔG^{solv} is the energy associated with the process of reorganizing solvent molecules, like water, around a solvated molecule, i.e, the process of moving the solute from vacuum to its position into the solvent. This process will be thermodynamically favorable only if the overall Gibbs energy of the solution is decreased, compared to the Gibbs energy of the separated solvent and solute. It's also function as an indicator of how probable will be a molecular system, in terms of its configuration, compare to others.

ΔG^{solv} can be separated in two components: ΔG_{polar} and $\Delta G_{non-polar}$. The first is define as the free energy of creating the solute's charge distribution inside a pre-existing solute cavity, and the second term, correspond to work required to place an uncharged solute inside the solvent and generates a dry cavity on it.

In this work, Molecular Dynamics technique (explicit model) was perform using GROMOS simulation software to calculate the van der Waals ΔG_{vdw} cavity-formation free energy for 20 alanine peptides divided in extended and helix configurations, ranging from 1 to 10, and compare the results with two implicit solvent models: the standard model based on the molecular surface of the solute, and a new methodology which has a closer physical meaning modeling the solute as an uncharged capacitor.

An extended overview of the MD method is provided with the methodology for implementing GROMOS simulation software not just to replicate these results, but also for the calculation of the relative free energies of any structure or compound of interest.

Resumen

La energía libre de solvatación ΔG^{solv} es la energía asociada al proceso de reorganización de las moléculas de disolvente y soluto, en moléculas solvatadas. Este proceso es termodinámicamente favorable solo si la energía libre de Gibbs de la solución es menor, comparada con la energía libre de Gibbs de solvente y molécula por separado. Este parámetro también funciona como un indicador de qué tan probable podría ser una configuración de

un sistema molecular, en comparación a otros.

ΔG^{solv} puede separarse en dos componentes: ΔG_{polar} y $\Delta G_{no-polar}$. El primer término se define como la energía libre necesaria para crear la distribución de cargas del soluto en una cavidad pre-existente, y el segundo término corresponde al trabajo requerido para ubicar una molécula de soluto no-cargado en el solvente generando una cavidad seca en este último.

En este trabajo, se usó la técnica de Dinámica Molecular junto al software de simulación GROMOS para calcular la energía libre de van der Waals ΔG_{vdw} necesaria para generar la cavidad para 20 péptidos de alanina divididos en dos tipos de estructuras: hélice y extendida, cuya composición va desde 1 a 10 monómeros. Los resultados fueron contrastados con dos modelos: modelos implícitos tradicionales, y una nueva metodología que tiene una fundación más profunda y un significado físico más cercano a la realidad. Se presenta también una descripción general de la metodología MD junto con el flujo de información usado por el software GROMOS, no solo para replicar los resultados de este trabajo, sino que también para el cálculo de energías libres de cualquier estructura molecular de interés.

1 Introduction

1.1 Life is based on chemical interactions

Understanding the basis of any biological phenomena is inseparable from understanding the principles that govern the behavior of atoms and molecules. All biological functions depend on events that occur at the molecular level. These events are directed, modulated, or detected by complex biological machines, which are themselves large molecules or clusters of molecules. Included are proteins, nucleic acids, carbohydrates, lipids, and complexes of them. Many areas of biological science focus on the signals detected by these machines or the output from these machines.

Most of biological processes happens in solution, mostly water, with tightly regulated concentrations of ions such as sodium and chloride. These solvent molecules strongly influence the structure and function of molecules, affecting protein folding and stability, as well as the affinity and specificity with which bio-molecules bind.

1.2 Molecular Structure

As we know, molecules are composed of atoms, which are the units of matter that correspond to the elements in the periodic table. For example, only a few of these elements are abundant in cells. In fact, the vast majority of biological matter, about 99% is made of just six elements: carbon, hydrogen, nitrogen, oxygen, sulfur, and phosphorus [5]. Molecules such as proteins, carbohydrates, lipids, and nucleic acids, are made out of carbon as it's main backbone.

Atoms, in turn, are composed of particles known as protons, neutrons, and electrons.

The properties of each particle is shown in Table 2. Protons and neutrons reside in the atomic nucleus and account for almost all of the mass of the atom. The number of protons present in an atom's nucleus, its atomic number (Z), determines the identity of that atom as an element.

Particle	Mass [kg]	Relative mass [amu]	Charge [Coulomb]	Charge
Proton (p^+)	1.6726×10^{-27}	1.007	1.6022×10^{-19}	+1
Neutron (n)	1.6749×10^{-27}	1.008	0	0
Electron (e^-)	9.1094×10^{-31}	5.489×10^{-4}	1.6022×10^{-19}	-1

Table 2: Mass and Charge of individual protons, neutrons and electrons.

1.2.1 Atomic Structure: Bohr's Model

Presented by Niels Bohr and Ernest Rutherford in 1913, the Bohr's Model offers a representation of how electrons orbits around the nucleus of an atom. The electron orbits into shells: The first shell contains 2 electrons, while the next shell can hold 8 electrons. The arrangement of electrons around the atomic nucleus is complex, and electrons do not simply orbit the nucleus as a planet would orbit a star. Broadly speaking, electrons are located in concentric shells that surround the nucleus. The shells that are closer to the nucleus are generally lower in energy than the shells that are farther from the nucleus, meaning that placing an electron in a shell that is closer to the nucleus is more stable and more favorable than placing an electron in a shell that is farther from the nucleus. Because of this, electrons fill shells from the inside to the outside. Electrons in atoms with few electrons are in shells close to the nucleus, whereas atoms with many electrons first utilize the shells that are closer to the nucleus and then those that are farther from the nucleus once the inner shells are filled.

1.2.2 Atomic Structure: Quantum Mechanics

Electrons don't travel in neat 2D circular orbits as the Bohr's model suggest, in fact we can't even determine the position and the velocity of an electron. Instead, we can make predictions about electron's general locations in 3D space. These regions that we

call **orbitals** are the locations in space that has about 90% certainty that an electron is located somewhere within that region [6]. In Figure 1 are represented the shapes that orbitals can take around the atom. Each orbital now can hold two electrons. In the case of carbon, we focus on the "*s*" orbital, which are spherical in shape with the nucleus of the atom at the center, and "*p*" orbital, which have the shape of the number eight, or the infinity symbol with the nucleus also at the center.

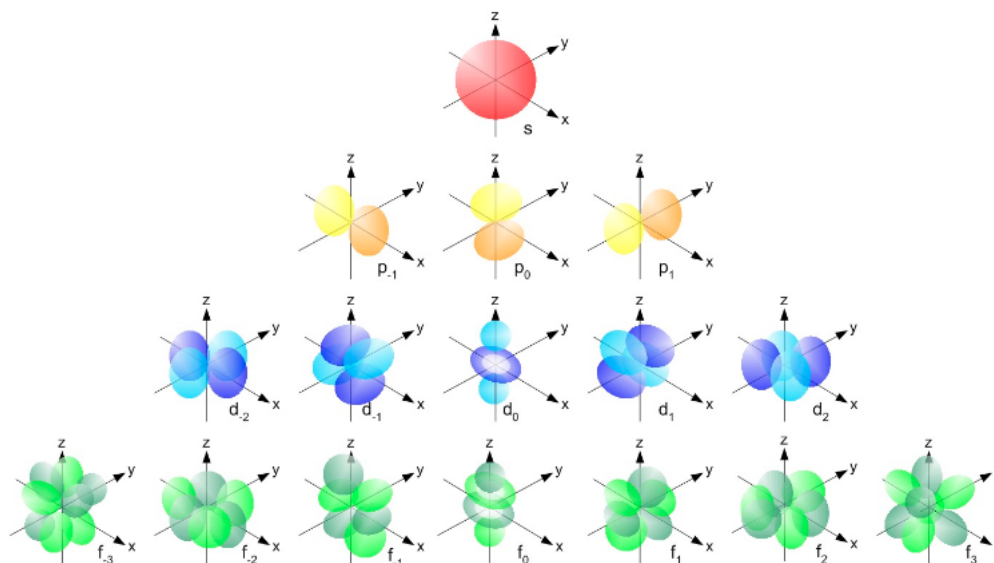


Figure 1: Schematics showing the general shapes of *s*, *p*, *d*, and *f* orbitals.

In the ground state, electrons will occupy the lowest energy orbitals first. In the case of the element **carbon**, which contains 6 electrons on its outer shell, the first orbital that fills is the $1s$, which can hold 2 electrons. Next, is the $2s$ -orbital which is a larger sphere, and can also hold 2 electrons. Finally, we have $3p$ -orbital: each one aligned along the x , y and z -axis, each capable of 2 electrons so they are filled with one electron in the px -orbital and one in the py -orbital. To be stable, carbon wants to fill these three p orbitals with 2 electrons each.

The interaction between different atom's orbitals results in "connections" between them and can be thought of as forces of attraction (or in some cases repulsion). There are two types of interactions: **bonded** and **non-bonded** interactions. The first is known as chemical bonds, and there are three types: (1) *Ionic bond*, formed when one atom donates valence electrons to another atom. (2) *Covalent bond*, when both atoms share pairs of

valence electrons, and (3) *Metallic bond*, formed between a cloud of free electrons and the positively charged ions in a metal. interact. For example, a water molecule is made of one oxygen atom connected to two hydrogen atoms through single covalent bonds. However, some covalent bonds involve the sharing of more than one pair of electrons between atoms. Double and triple bonds involve the sharing of four and six valence electrons, respectively. This versatility allows carbon to create many different kinds of molecules which gives this particular element, a vast number of properties.

The second interaction, **non bonded**, is explained in detail in Section 2.2 because is the main objective of this work. Study the behavior of a particular component of this type of interaction.

1.3 Computer Simulation of Molecular Systems

The role of computation in biology, biological chemistry, and biophysics has shown a steady increase over the past few decades. The continuing growth of computing power (in particular in the context of personal computers) has made it possible to analyze, compare, and characterize large and complex data sets that are obtained from experiments on biomolecular systems. [1]. Chemical systems are generally too complex to be treated by analytical theoretical methods so it becomes a necessity to implement numerical analysis to study and predict some properties and configurations.

Computer simulations on models of physical systems have been carried out for more than 65 years now [7]. Many developments have turned molecular dynamics simulations into a valuable tool, complementary to experimental investigation, to probe into structure, dynamics, and activity of large biologically relevant molecules. The thermodynamic information that can be obtained from computer simulations allows analysis and understanding of molecular processes and prediction of molecular properties, highly valued information in medical chemistry and drug design.

1.4 Objectives

- The main objective of this work, is to calculate the free energy of solvation (non-polar) of 20 peptides using classical Molecular Dynamics Simulation method.
- Compare the results of the explicit MD model with the calculations given by the implicit solvent models available by measure the accuracy of them.
- Become familiar with the use of GROMOS++ software for the calculation of free energies using Thermodynamic Integration.
- Study the behavior of the van der Waals interaction and its contribution to the solvation process.
- Simulate 20 alanine-peptides in a pre-equilibrated water box to preform the free energy calculations.

This document is organized as follow: Section 2, presents an overview of the Molecular Dynamics (MD) method, which is the tool to preform the calculations of the free energies of solvation described in the abstract and the criteria and requirements needed to build this method. It also provides a characterization of the force-fields whit a special attention to describe the nature of non-bonded interaction which is the main subject treated in this work. In Section 3, implicit solvent models are explain with they advantages and disadvantages and also the reasons to use this methods as an alternative to MD. In particular, Section 3.2 described a novel way to calculate ΔG_{solv} , sharing the same ideas as the implicit solvent models, but with a clearer physical meaning than current available models. Next, Section 4 presents the underlying thermodynamic and statistical mechanics theory that had been develop to calculate free energies, with the mathematical construction of these models. Section 5 talks about the methodology used to compute the free energy results, starting with the preparation of the molecular structures (Section 5.1), following by a review of the scope and functionality of GROMOS simulation software, explaining how it was used to preform such calculations step by step. Finally, in Section 6 the results and comparison between the models explained above are presented.

2 Molecular Dynamics

Computation based on molecular models is playing an increasingly important role in biology, biological chemistry, and biophysics. Since only a very limited number of properties of bio-molecular systems is actually accessible to measurement by experimental means, computer simulation can complement experiment by providing not only averages, but also distributions and time series of any definable quantity, for example, conformational distributions or interactions between parts of systems. As it's mentioned in section 1.3, computer modeling of bio-molecular systems has become a standard technique to study and describe the properties and behaviour of such systems in terms of interactions between atoms or electrons of atoms.

Molecular dynamics (MD) is a **computer simulation method** for analyzing the physical movements of atoms and molecules. The atoms and molecules are allowed to interact for a fixed period of time, giving a view of the dynamic “evolution” of the system. Computer simulations can then be used to test and validate theories and models or to replicate and compare experimental results.

There are some key elements that must be taken into account when developing any simulation exercise. The model approximations must be considered to ensure applicability to the system of interest, and appropriate statistical analyses must be performed (for which sufficient sampling is required).

In general, to set-up a particle based simulations, there are four main choices to consider for the model that will be used which will be discussed in the next four sections as it's shown in Figure 2: The degrees of freedom to be considered, interactions that occur between the chosen degrees of freedom, how they are to be sampled, and how the boundaries and/or external forces of the system are to be considered [1].

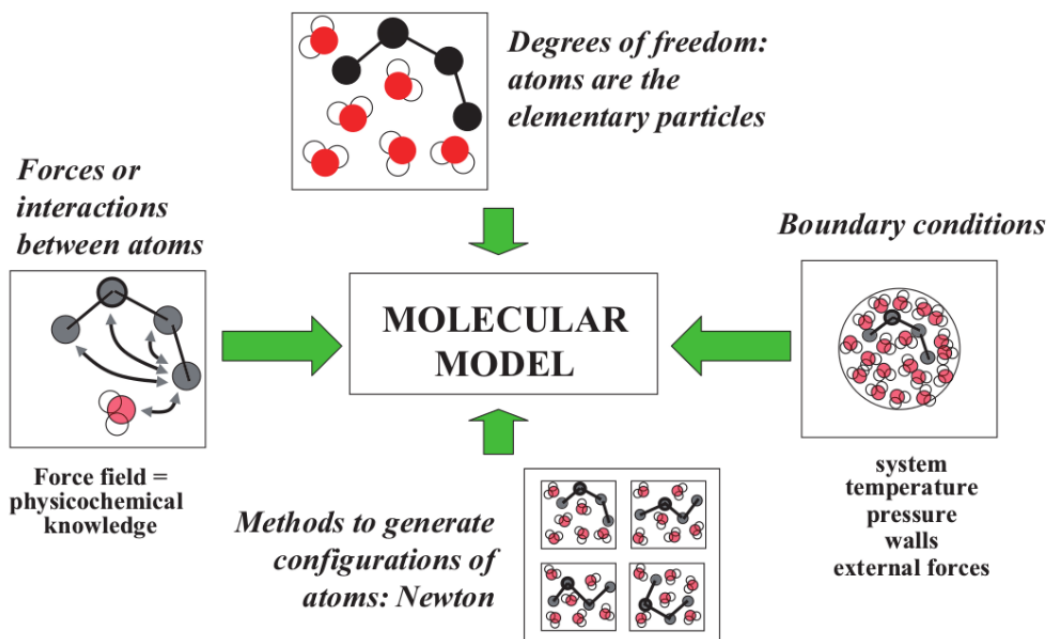


Figure 2: Four basic choices in the definition of a model for molecular simulation. Figure from Van Gunsteren et al. [1]

2.1 Degrees of freedom

This criteria is oriented to determine which molecular description are explicitly considered in the model and is strictly related to the resolution of the system of interest. The level of modeling chosen to describe a particular bio-molecular process depends on the type of process. This will determine the theory employed, ranging from quantum to classical mechanics. In this particular case of study, we are interest on evaluate the interaction between the solvent and the solute which is driven by weak, non-bonded inter-atomic interactions. Therefore, these processes are most promisingly modeled at the atomic or molecular level. In Table 3 are organized by level of detail, the methods for different molecular descriptions.

As it is shown in Table 3, the models were categorized by order of approximation, ranging from quantum descriptions in the form of quantum mechanics (QM) based models which take into account the interaction between atom's orbitals as is described in Section 1.2 with multiple degrees of freedom, to approximations for heavy and slow particles on a

Methods	Degrees of freedom	Properties, processes	Time scale
quantum dynamics	atoms, nuclei, electrons	excited states, relaxation, reaction dynamics	picoseconds
quantum mechanics	atoms, nuclei, electrons	ground and excited states, reaction mechanisms	no time scale
classical statistical mechanics (MD, MC, force fields)	atoms, solvent	ensembles, averages, system properties, folding	nanoseconds
statistical methods (database analysis)	groups of atoms, amino acid residues, bases	structural homology and similarity	no time scale
continuum methods (hydrodynamics and electrostatics)	electrical continuum, velocity continuum etc.	rheological properties	supramolecular
kinetic equations	populations of species	population dynamics, signal transduction	macroscopic

Table 3: Examples of levels of modeling in computational biochemistry and molecular biology. Table from Van Gunsteren et al. [1]

larger scale were the processes involved are very well described by classical mechanics. The cost of the energy and force evaluation is primarily determined by two factors: the degrees of freedom that are considered and the functional form of the Hamiltonian (Section 2.2). One can move from a quantum-mechanical description of the system where the electronic degrees of freedom are modeled explicitly, to a classical description where one atom is treated as one particle, to a coarse-grained description where groups of atoms are merged into one particle. A reduction in the number of degrees of freedom can, however, also be obtained by reducing the system size or treating parts of the system (e.g., the solvent) as a continuum [8].

The choice of which degrees of freedom are modeled explicitly depends on the system of interest and the property one wishes to estimate. If a certain degree of freedom is

believed to have no effect on the property of interest, it can be omitted and the computing time gained invested in sampling the relevant degrees of freedom more extensively.

2.2 Potential Energy Function: The Force Field

A bio-molecular force field generally consists of potential energy terms representing covalent interactions between atoms (such as bond-stretching, bond-angle bending, improper and proper dihedral-angle torsion), and non-bonded interactions between atoms in different molecules and between atoms in a molecule that are separated by more than two or three covalent bonds [1]. Consider a system that contains N particles that are treated explicitly in the model. Often, these particles are atoms, but also groups of atoms can be treated as a single particle, such as hydrocarbons like CH , CH_2 , CH_4 groups. In classical simulations, the system is fully defined by the positions “ \mathbf{q} ”, and the conjugate momenta, “ \mathbf{p} ”, of the individual particles, where r and p represent 3-N dimensional vectors. In this work, we will use q_i and p_i for the three-dimensional vectors describing the position and momentum of the particle “ i ”. The potential energy function that determine the system, the **Hamiltonian**, is written as:

$$H(p, q) = K(p) + V(q) \quad (1)$$

where $K(p)$ is the kinetic energy which can be

$$K(p, m) = \sum_{i=1}^N \frac{p_i^2}{2m_i} = \sum_{i=1}^N \frac{1}{2} m_i v_i^2 \quad (2)$$

$V(r)$ in equation 1 is the potential energy, describing the interaction between the particles in the system and possibly external influences on the system. It’s a function of the particle position “ \mathbf{q} ”. The functional form and parameters describing $V(r)$ is called **force field**. There are several well-known force fields for biomolecular simulation described in the literature, such as AMBER [9], CHARMM [10], ECEPP/3 [11], and GROMOS [12][13], and it takes different forms depending on the description of the system, and is usually a sum of multiple potential energy terms which describe different contributions corresponding to physical (or special) interactions. The potential energy is usually written as a sum over

different contributions, separated in physical interactions and any special (unphysical, such as distance restraints between atoms and positional restraints) interaction one might want to apply to the system

$$V(q) = V^{phys}(q) + V^{special}(q) \quad (3)$$

The first term of Equation 3 is the mathematical description of the sum of various terms describing the bonded (bon) and nonbonded (nonb) interactions between particles. Bonded potential energy can be divided further to account for bonds, bond angles, harmonic (improper) dihedral angles and torsional or proper dihedral angles (Equation 4), and not bonded interaction are the sum of electrostatic (Coulomb) and van der Waals (Lennard-Jones potential) interactions between pairs of atoms (Equation 5).

$$V^{bon}(q) = V^{bond}(q) + V^{angle}(q) + V^{imp}(q) + V^{tor}(q) \quad (4)$$

$$V^{nonb}(q) = V^{vdw}(q) + V^{elec}(q) \quad (5)$$

2.2.1 Force-Field Bonded Interactions

In the following sections, V^{bon} terms are explained according to the GROMOS force field which is used in this work.

Bond stretching force-field term

The potential energy term associated with bond-stretching interactions is the term $V^{bond}(q)$ in Equation 4 and it's the function describing the potential energy associated with the stretching of a single bond and has the form of an harmonic oscillator given by:

$$V^{bond}(q) = \sum_{n=1}^N \frac{1}{2} k_b (b_n - b_n^0)^2 \quad (6)$$

where the quantities k_b and b_n represent force-field parameters, force constant and reference length, respectively, characteristic for the specific bond n , and N the total number of covalent bonds.

Bond-angle bending force-field term

The potential energy term associated with bond angle bending interactions is the term $V^{angle}(q)$ in Equation 4. Each definable covalent bond angle associated with one and only one bending term in the GROMOS force and this term represent the function describing the potential energy associated with the bending of a single bond angle. It's given by:

$$V^{angle}(q) = \sum_{n=1}^N \frac{1}{2} k_{\theta} (\theta_n - \theta_n^0)^2 \quad (7)$$

where k_{θ} and θ_n represent force-field parameters, force constant and reference bond angle, respectively, characteristic for the specific bond angle n.

Improper dihedral-angle bending force-field term

The potential energy term associated with improper dihedral-angle bending interactions, typically controlling out-of-plane or out-of-tetrahedron distortions, is the term $V^{imp}(q)$. Improper dihedral angles are used to select the correct geometry or chirality of atoms. The mathemtaical expression is given by:

$$V^{imp}(q) = \sum_{n=1}^N \frac{1}{2} k_{\xi} (\xi_n - \xi_n^0)^2 \quad (8)$$

where k_{ξ} and represent the force-field force constant for each ξ_n represents the value of improper dihedral angle n in the given system configuration, *i.e.*, the dihedral angle formed by the four atoms.

Proper dihedral-angle torsion force-field term

The potential energy term associated with torsional dihedral-angle bending interactions, is typically controlling, together with non-bonded interactions, the rotational barriers around covalent bonds. $V^{tor}(q)$ is the function describing the potential energy contribution of the

term to the torsion of the corresponding proper dihedral angle.

$$V^{tor}(q) = \sum_{n=1}^N k_{\phi}(1 + \cos(n\phi - \delta)) \quad (9)$$

where k_{ϕ} is the force constant define by the force field, δ the phase shift (which corresponds to the minimum energy angle), n is the multiplicity (number of minima of the potential energy as it rotates a full 360 degrees), and ϕ is the dihedral angle.

In Figure 3 are represented each of the potential energy terms described above and the main interaction that the potential describes.

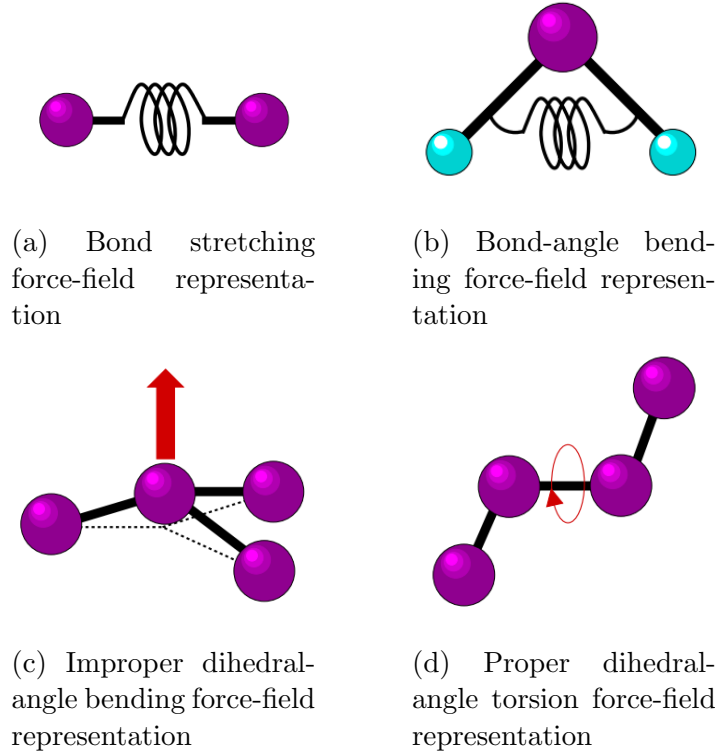


Figure 3: Illustration of the various terms included in an empirical potential energy function. Figure adapted from C. Chipot [2]

As it is presented in Section 1.2, the main objective of this work is to evaluate the non-bonded interactions between the solvent and the solute, particularly the van der Waals interaction. To formulate a complete description of the non-bonded interaction described

in Equation 3, the following section is focus on describe the detail mathematical formulation of the terms involved in Equation 5.

2.2.2 Force-Field Non-bonded Interactions

In addition to the bonded interactions between atoms described above, force fields also contain non-bonded interactions. Non-bonded interactions act between atoms in the same molecule and those in other molecules. Force fields usually divide non-bonded interactions into two: Electrostatic interactions and van der Waals interactions.

Electrostatic Interactions

Electrostatic or Coulomb potential describes the interactions potential between pairs of partial charges and it's given by de Coluomb's law.

$$V^{elec}(q) = \frac{1}{4\pi\epsilon_0\epsilon_1} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{q_i q_j}{r_{ij}} \quad (10)$$

where ϵ_0 is the dielectric permittivity of vacuum, ϵ_1 the relative permittivity of the medium, q_i and q_j are the charges of atoms i and j involved in the interaction, and r_{ij} the distance between the atoms.

Van der Waals Interactions. The Lenard-Jones Potential

The van der Waals force is a weak, short-range force that arises from temporal fluctuations of the charge distribution. This force can attract ideal gas atoms together. Ideal gas atoms are electrically neutral (Figure 4a) so there is no Coulomb attraction between them. The average dipole moment of an ideal gas atom is zero but the charge fluctuates around its average position and can temporarily create a dipole moment (Figure 4b). This dipole moment induces a response in the neighboring atom such that there is a net attractive force between the atoms (Figure 4c). The average over the attractive forces caused by the charge fluctuations is the van der Walls force.

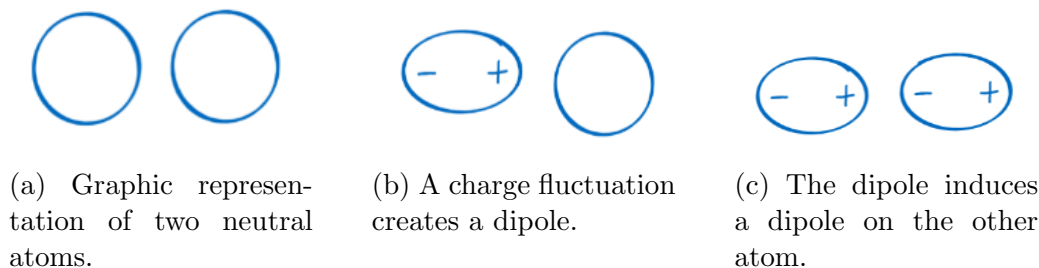


Figure 4: Illustration of the sequence of events that generates the Van der Waals forces.

The Van der Waal bond potential is often approximated by a Lennard-Jones model which consist of two components: a steep repulsive term, $(\sigma/r_{ij})^{12}$ and a smoother attractive term, $(\sigma/r_{ij})^6$ as it is shown in Figure 11, which respectively denote repulsive and attractive force. The mathematical description of this potential is given by:

$$V^{vdw}(r_{ij}) = 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right] \quad (11)$$

where ϵ is the well depth and a measure of how strongly the two particles attract each other, and σ is the distance at which the inter-molecular potential between the two particles is zero as it's shown in Figure 5. It gives a measurement of how close two non-bonding particles can get and r_{ij} is the distance of separation between both particles measured from the center of one particle to the center of the other particle.

Like the bonding potential energy, the stability of an arrangement of atoms is a function of the Lennard-Jones separation distance. As the separation distance decreases below equilibrium, the potential energy becomes increasingly positive (indicating a repulsive force). Such a large potential energy is energetically unfavorable, as it indicates an overlapping of atomic orbitals. However, at long separation distances, the potential energy is negative and approaches zero as the separation distance increases to infinity (indicating an attractive force). This indicates that at long-range distances, the pair of atoms or molecules experiences a small stabilizing force. Lastly, as the separation between the two particles reaches a distance slightly greater than σ , the potential energy reaches a minimum value (indicating zero force) which is represented by r_{min} in Figure 6. At this point, the pair of particles is most stable and will remain in that orientation until an external force is exerted upon it.

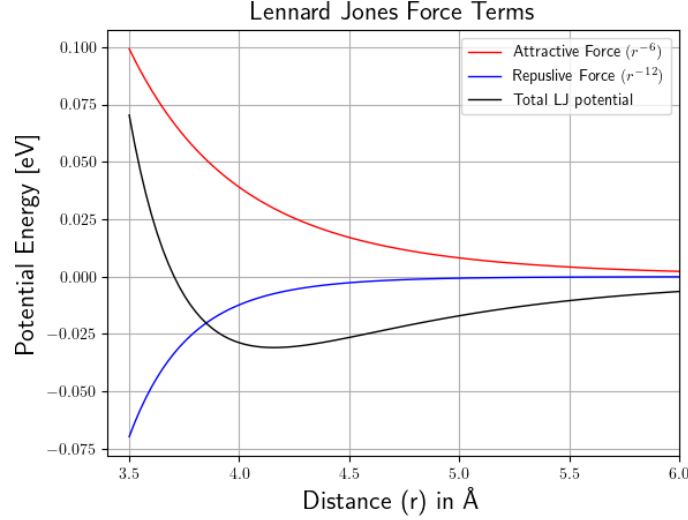


Figure 5: Graphic representation of the attractive and repulsive terms of the Lenard Jones Potential

$$r_{min}(ij) = 2^{1/6}\sigma \quad (12)$$

There are several different ways to formulate the Lennard-Jones potential besides Equation 11, the most common form in the literature is the **AB potential** and is frequently used in implementations of simulation software as it is computationally favorable. The Lennard-Jones potential can be written as:

$$V^{LJ}(r_{ij}) = \frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^6} \quad (13)$$

where

$$A = 4\epsilon\sigma^{12} \quad B = 4\epsilon\sigma^6 \quad \sigma = \sqrt[6]{\frac{A}{B}} \quad \epsilon = \frac{B^2}{4A}$$

In GROMOS, the van der Waals (vdw) interaction term V^{vdw} in Equation 5 is represented by a Lennard-Jones function, and the term is partitioned as a sum of two contributions. It's given by:

$$V^{vdw}(r_{ij}) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left[\frac{C_{12}(i,j)}{r_{ij}^6} - C_6(i,j) \right] \cdot \frac{1}{r_{ij}^6} \quad (14)$$

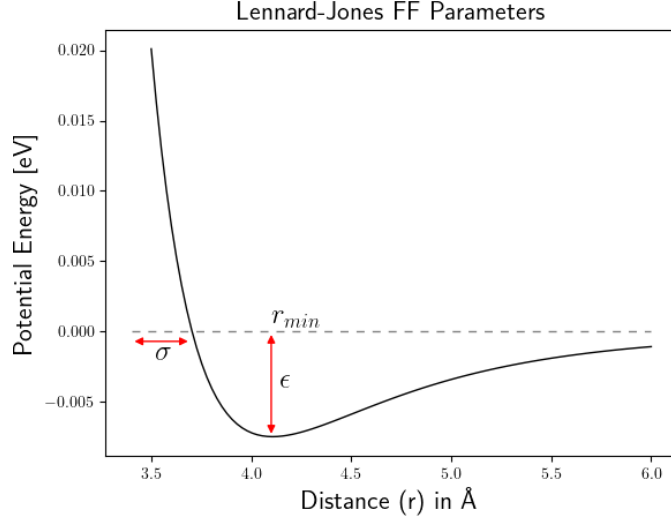


Figure 6: The Lennard-Jones Potential curve, calculated for $\epsilon = 0.0103$ and $\sigma = 3.8$.

and every $C_{12}(i, j)$, $C_6(i, j)$ is easily extracted from the force-field parameters file in the GROMOS software. More about how GROMOS work is discussed in Section 5.2. Finally, the total potential energy involved in the therm $V(q)$ from Equation 1 it's determined by the sum of every potential energy described in this section, for the pair of atoms (i, j) .

2.3 Configurations of Atoms

The search of configuration space can be expanded by applying the molecular dynamics (MD) computer simulation technique. Applying MD the classical Newtonian equations of motion of all atoms in the system are solved. These are for a set of N atoms with masses N_i and Cartesian position vectors $r_i(i = 1, 2, \dots, N)$. The expression is given by:

$$\frac{d^2 r_i(t)}{dt^2} = m_i^{-1} F_i(r_1(t), r_2(t), \dots, r_N(t)) \quad (15)$$

where the force applied to each atom of the system is obtain by:

These equations can be numerically integrated using small ($10^{15}s$) time-steps Δt producing a trajectory (atomic positions as a function of time t) of the system [14].

Regarding the method to generate configurations, the formulation of the energy of the system (Equation 1) which is determined by the molecular system by the force-field, gives the mathematical description of the potential energy described in Equation ??.

$$\frac{\partial H}{\partial p_i} = \frac{dq_i}{dt} \quad (16)$$

$$\frac{\partial H}{\partial q_i} = -\frac{dp_i}{dt} \quad (17)$$

These two derivation from the Hamiltonian gives the velocity (Equation 16) and as the kinetic energy of the system is not explicitly dependent on positions, Equation 17 is an alternative way to express Equation ??.

These equations describe the physical basis for molecular dynamics (MD) simulations, which integrate these equations as a function of time, for each degree of freedom (for N particles there are 3 position and 3 momentum coordinates, for a total of 6N degrees of freedom). A common alternative is the Monte Carlo (MC) method which is based on random non physical moves to sample an energy distribution. However, MC simulations can be problematic when an explicit solvent is used, as when solvent molecules are simulated along with the molecules of interest problems arise as the solvent restricts movement within traditional MC methods leading to poor sampling when compared to MD methods [15].

2.4 Boundary Conditions

The last aspect to consider in a MD simulation is the boundary conditions, understanding this as how the physical limits of the system or the spatial boundaries and external forces are modeled, and the properties of where , in reality, biologically interesting molecules are found in solvents and cells. When it comes to selecting boundary conditions, there are many options to consider with their own advantages, disadvantages, and computational cost (see Figure 7). The simplest form of boundary conditions is to just simulate the molecule(s) of interest in a vacuum, however this is not representative of molecules in biological conditions and causes many problems in the surface of the molecule. The next level of detail, and the boundary condition selected for this work because of the property

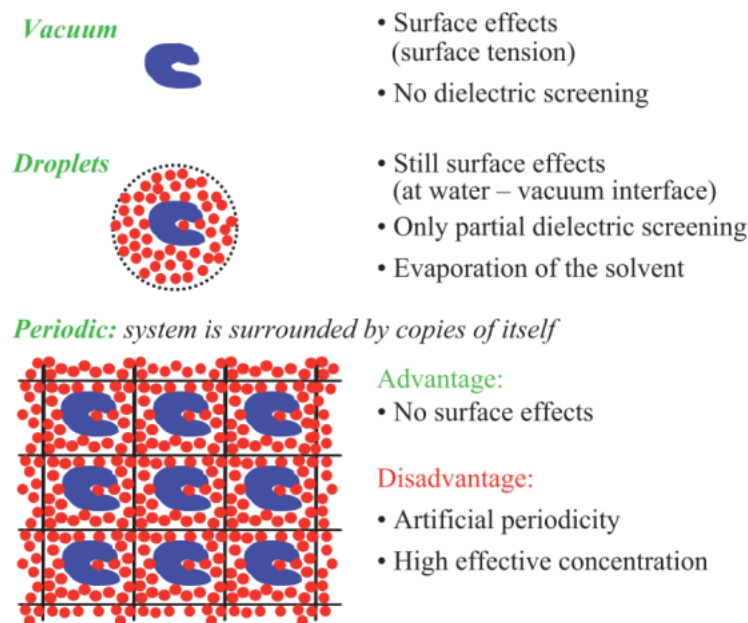


Figure 7: Three types of spatial boundary conditions used in molecular simulation. Figure from Van Gunsten et al [1]

that is evaluated, is when a molecule can be solvated in a droplet of solvent with a specific geometry. However, this transfers complications from the solute-vacuum interactions to solvent-vacuum interactions, and a “wall” must be created by fixing the position of the external solvent molecules, so that they don’t evaporate. The method usually considered most appropriate for biological simulations is periodic boundary conditions, where the system is surrounded in all directions by copies of itself, removing surface effects and emulating bulk conditions.

2.5 General Molecular Dynamics Algorithm

Starting with a given number of atoms “ N ”, one wants to be able to simulate or predict how those atoms will be moving according to time. What will be the displacement or trajectory of these atoms at a different time. The algorithm works as follows:

1. System Initial Configuration $\vec{r}_i(t=0)$; $\vec{v}_i(t=0)$

Given a system of “ N ” atoms, initial position vectors \vec{r}_i and initial velocity vectors \vec{v}_i at $t = 0$, with $i = \{1, 2, 3, \dots, N\}$, are generated depending on the system configuration. In Section 5.2 is explained how GROMOS software generates data sets for both initial position and initial velocities by minimisation and thermalization processes respectively.

2. Total Potential Energy U_{ij}

Given the initial configuration of the system, the inter-atomic potential energy between atoms now can be determined. Each pair of atoms potentially will have a given energy between them U_{ij} because of the forces applied \vec{F}_{ij} . This interaction is determined by electrostatic (Equation 10) and Van der Waals (Equation 11) forces and is intrinsically dependent on the distance between them r_{ij} . The total potential energy is equal to the sum of each pair interaction between atom i and j .

$$U_{tot} = \sum_{i \neq j}^N U_{ij}$$

3. Determine Total Forces Applied

With the potential energy U_{ij} as a function of the distance r_{ij} between two atoms, the force applied by a given atom i on atom j can be calculated by the gradient of the energy.

$$F_{i \rightarrow j} = -\vec{\nabla} U_{ij}$$

$$\vec{F}_i = \sum_{i \neq j}^N F_{i \rightarrow j}$$

Therefore, the resultant force on an atom i must be the sum of all forces between it and all other atoms $j \neq i$.

4. Acceleration on each atom.

Applying Newton’s equations of motion and knowing the forces and the masses of each individual atom, the acceleration applied on each atom is given by:

$$\vec{a}_i = \frac{\vec{F}_i}{m_i}$$

5. Prediction of Motion $r(t + dt)$; $v(t + dt)$

Knowing the initial position and initial velocity of an atom i , and knowing it's acceleration one can predict how the atoms will be moving after a small increment of time dt . Now, one can calculate the new velocity $v(t + dt)$ and new position $r(t + dt)$ after a small increment in time. This relies on the numerical integration. Knowing the acceleration, it can be integrating in order to get the position and how the position is changing according to time.

Figure 8 summarizes the algorithm described above.

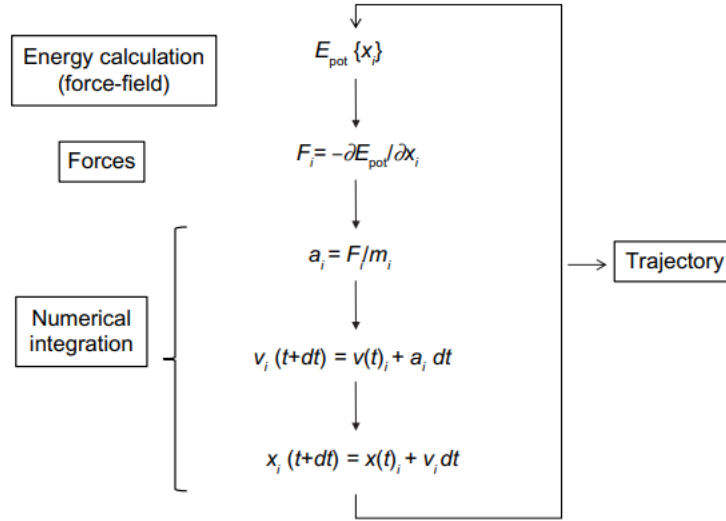


Figure 8: Molecular dynamics basic algorithm. Figure from Hospital *et al.* [3]

2.5.1 Numerical Integration

Multiple methods to integrate equations of motion have been developed. Among these is the **leap-frog algorithm** [16] which was used for the simulations carried out in this work. Figure 9 is a schematic representation of how the variables for velocities and positions are compute throw the mathematical scheme [17]. When a Taylor expansion of the velocity $v_i(t_n - \Delta t/2)$ at $t = t_n$ is subtracted from an expansion of $v_i(t_n + \Delta t/2)$ at $t = t_n$, neglecting

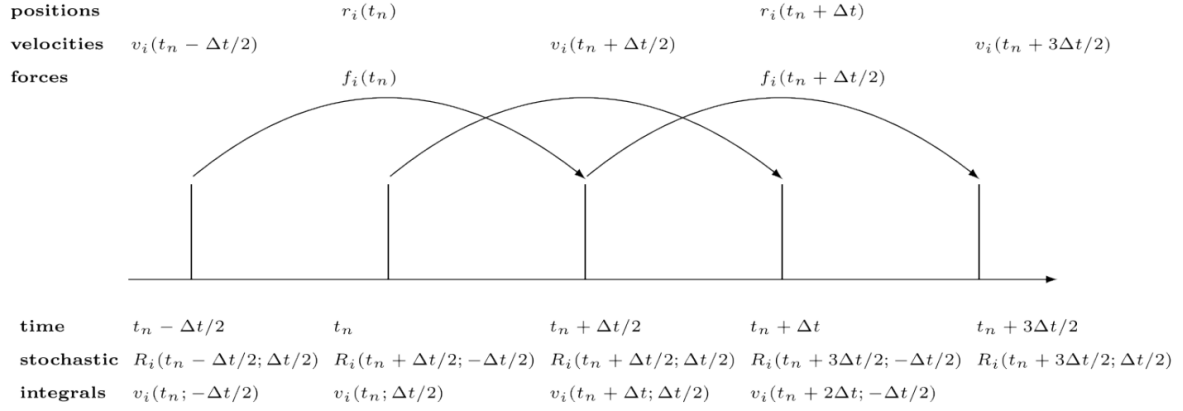


Figure 9: The leap-frog integration scheme. Figure from Van Gunsteren: Gromos Manual Vol.2

terms of third and higher order for simplicity, results:

$$v_i(t_n + \Delta t/2) = v_i(t_n - \Delta t/2) + \frac{F_i(t_n)}{m_i} \Delta t \quad (18)$$

And similarly, when a Taylor expansion of the position $r_i(t_n)$ at $t = t_n + \Delta t/2$ is subtracted from an expansion of $r_i(t_n + \Delta t)$ at $t = t_n + \Delta t/2$ results (again neglecting terms of third and higher order for simplicity):

$$r_i(t_n + \Delta t) = r_i(t_n) + v_i(t_n + \Delta t/2) \Delta t \quad (19)$$

Equations 18 and 19 are the leap-frog velocity and position formulas, respectively, which together form the leap-frog scheme. This provides a simple method where a MD algorithm alternates between calculating (and updating) positions and velocities.

3 Implicit Solvent Models

As it is mentioned in Section 2, interactions between a solute molecule and a surrounding solvent are of fundamental importance across chemistry, biology, and associated fields of science and engineering. Fully atomistic, explicit-solvent MD simulations provide an accurate and chemically detailed understanding of these interactions therefore they can be highly demanding in the consumption of computational resources. This cost is associated with the large number of solvent molecules required to model a bulk solution. In practical situations, a large fraction of the time is spent calculating a detailed trajectory of the solvent molecules, even though it is primarily the solute behavior that is of interest.

Because of this, it is desirable to develop different approaches in which the influence of the solvent is incorporated **implicitly**. Approximate schemes treating the solvent implicitly can provide useful quantitative estimates and remain computationally inexpensive [18] [19].

3.1 Implicit model: Solvation Free Energy

Solvation is the process of reorganizing solvent and solute molecules into solvation systems. Solvation involves bond formation, hydrogen bonding, and van der Waals forces. Solvation of a solute by water is called hydration [20]. The solvation process describes the interaction between the solute and solvent molecules. This process will be thermodynamically favored only if the overall Gibbs energy of the solution is decreased, compared to the Gibbs energy of the separated solvent and solid (or gas or liquid). Regarding Equation 29, a negative Gibbs energy indicates a spontaneous process but does not provide information about the rate of dissolution.

This process involves multiple steps with different energy consequences: First, a cavity

must form in the solvent to make space for a solute (Figure 10a to Figure 10b). This is both entropically and enthalpically unfavorable, as solvent ordering increases and solvent-solvent interactions decrease. Stronger interactions among solvent molecules leads to a greater enthalpic penalty for cavity formation. Next, a particle of solute must separate from the bulk. This is enthalpically unfavorable since solute-solute interactions decrease, but when the solute particle enters the cavity, the resulting solvent-solute interactions are enthalpically favorable.

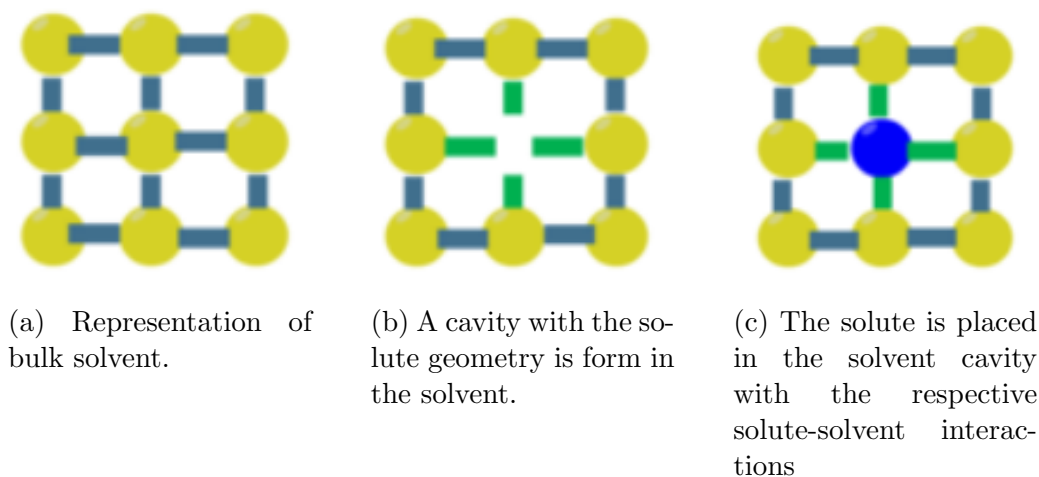


Figure 10: Representation of the solvation process for any molecule in solution.

The solvation free energy, process described in Figure 10, is frequently decomposed as a sum of *polar* (ΔG_{polar}) and *non-polar* (ΔG_{np}) terms [19].

$$\Delta G^{solv} = \Delta G_{polar}^{solv} + \Delta G_{np}^{solv} \quad (20)$$

The polar term is the free energy of creating the solute's charge distribution inside a pre-existing solute cavity, Assuming a charge distribution given by a molecular mechanics force field, the polar solvation contribution can be modeled with macroscopic continuum dielectric theory, e.g. the Poisson-Boltzmann equation. This works focuses on the **non-polar** component which is define as the work required to place an uncharged solute inside the solvent and generates a dry cavity on it. This contribution is often estimated from the solvent accessible surface area (SASA) of the molecule using a uniform surface tension coefficient obtained from a fit of experimental [19].

3.1.1 Solvent-Accessible Surface Area (SASA) Model

Implicit-solvent models often treat ΔG_{np} as a weighted combination of the solute’s SASA and other geometric measures given by:

$$\Delta G_{np} = \gamma(SASA) + b \tag{21}$$

where γ is usually interpreted as the liquid-vapor surface tension, and b is a scaling constant. Geometrically, *Solvent Accessible Surface Area* is related to the region of the molecule surface exposed enough to be able to interact with solvent molecules. It is usually defined as a surface built by the delineation drawn by the center of a sphere (rough representation of a solvent molecule, usually of 1.4 [Å] radii; i.e. a water molecule) rolling over the molecular surface. Nevertheless, the accurate prediction of the solvation free energy remains a very challenging issue and several inaccuracies with this method have been presented by some authors [21] [22], suggesting that an empirical parametrization of the hydration free energy and the free energy of association of apolar species based only on the solvent accessible surface area is insufficient. Successful parametrization should contain a surface area term to reproduce effects due to entropy loss and solvent reorganization energy and terms that depend on the number, location, and type of atomic interaction centers to reproduce effects due to the solute-solvent dispersion interactions which are only weakly correlated with surface area.

3.2 Non-polar Solvation Free Energy Estimation

The non-polar term in Equation 20 can itself be decomposed into a sum of two free energies [19], ΔG_{cav} the creation of a solute-shaped cavity by turning on the potential short-range **repulsive interactions**, and ΔG_{disp} , in which one turns on the longer-range **attractive** solute-solvent dispersion interactions.

$$\Delta G_{np} = \Delta G_{cav} + \Delta G_{disp} \tag{22}$$

Traditionally, ΔG_{cav} is highly correlated with the SASA factor previously described, however, the value of γ does not agree with the surface tension coefficient and works as a

adjustable parameter which can change the errors. On the other hand, ΔG_{disp} can be described with a continuum model, where the solute-solvent dispersion interaction potential is integrated over the solvent, which in turn can be formulated as an volume integral over the solvent accessible surface [23]. It is been study that while repulsive and attractive components of the non-polar part both scale roughly with surface area (or volume) of the solute, the total non-polar free energy does not scale with the solute surface area or volume. [24].

3.2.1 Implicit Solvent Model for ΔG_{cav} and ΔG_{disp}

A new approach to estimate ΔG_{cav} and ΔG_{disp} [] proposes an implicit solvent model with a clearer physical meaning than current available models. Regarding Equations 20 and 22, the non-polar solvation free energy can be expressed as (See Figure 11):

$$\Delta G_{solv} = \Delta G_{cav} + \Delta G_{disp} + \Delta G_{polar} \quad (23)$$

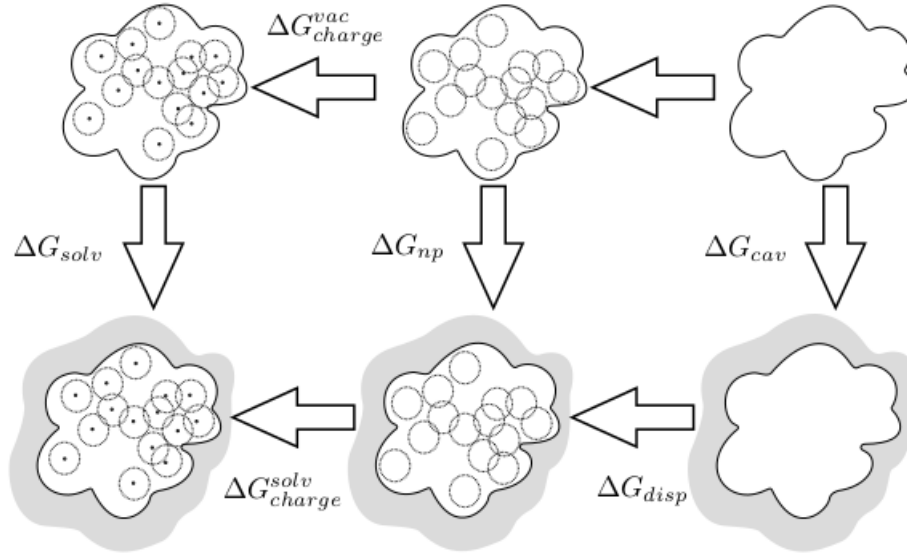


Figure 11: Thermodynamic cycle for solvation energy. Figure taken from [4]

Using a capacitance-based approach for ΔG_{cav} , the potential inside the solute is model as a constant ϕ_{static} , defining the solute boundary as the solvent-excluded surface (SES) [25]. In the infinite solvent region, the potential is model as $\phi = 0$, and bound the region using the boundary of the first solvation shell (the solvent-accessible surface, SAS [25]. These surfaces bound a shell, which is model as a macroscopic dielectric with the uniform relative permittivity ϵ_{shell} , and where the electrostatic potential obeys the Laplace equation. Fixing the potentials at both boundaries implies that the dielectric constants of the solute and bulk solvent are irrelevant.. It is used ϵ_{shell} as a fitting parameter, but it has clear physical bounds and significance. The energy associated with this boundary-value problem can be solved using a pair of coupled boundary-integral equations for unknown charge densities on the two surfaces by:

$$\begin{bmatrix} V_{diel} & V_{diel} \\ V_{shell} & V_{shell} \end{bmatrix} \begin{bmatrix} \sigma_{diel} \\ \sigma_{shell} \end{bmatrix} = \begin{bmatrix} \phi_{static} \\ 0 \end{bmatrix} \quad (24)$$

where $V_{diel}(\phi_{shell}) = \oint_{shell} \frac{\phi(r')}{4\pi|r_{diel}-r'|}$ denotes the potential induced at the inner surface (the dielectric boundary) by the surface charge distribution on the outer boundary (the solvent-accessible surface). Figure 12 represent this idea. Having found the surface charge

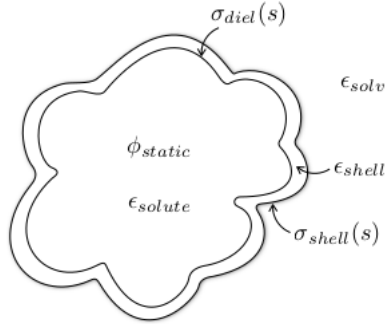


Figure 12: Sketch of the capacitor model. Figure taken from [4]

distributions it is possible to compute the work (and hence, the free energy) required to charge the resulting capacitor with standard electrostatic theory

$$\Delta G_{cav} = \int_{\Omega} \phi(r) \rho(r) dr \quad (25)$$

being ρ the charge distribution, in this case, concentrated on the SAS (σ_{shell}) and SES

(σ_{diel}) Noting that the potential on the outer surface is zero, we can rewrite Equation 25 as:

$$\Delta G_{cav} = \oint_{diel} \phi_{static} \sigma_{diel}(r) dr \quad (26)$$

The previous ΔG_{cav} component is combine with a modification of a continuum integral method [26] for calculating ΔG_{disp} , by integrating the Lennard-Jones potential outside the solvent-accessible surface which can be formulated as a surface integral using the divergence theorem by:

$$\Delta G_{disp} = \sum_i \oint_{shell} \rho_w \frac{\partial}{\partial n} \left(\frac{A_i}{90 |r - r_i|^{10}} - \frac{B_i}{12 |r - r_i|^4} \right) dr \quad (27)$$

where ρ_w is considered to be the bulk number density, A_i and B_i re related to the Lennard-Jones parameters for the water oxygen and atom i, the sum is over the solute's atoms, and the unit vector \mathbf{n} points into the solvent. Finally, a solvation-layer interface condition (SLIC) continuum electrostatics method [27] is used to compute ΔG_{polar} (or ΔG_{charge}^{solv} in Figure 11). (Notice that regarding the thermodynamic cycle represented in Figure 11, ΔG_{solv} is equivalent to $\Delta G_{charge}^{solv} - \Delta G_{charge}^{vac}$).

4 Free Energy

The free energy is a highly desirable quantity to compute. It is in essence the factor that determines how a process will proceed and the probability that a system will adopt a given state. The ability to calculate free energies from molecular simulations not only allows one to understand the underlying processes on an atomic level but also to probe states of a system not accessible experimentally. In particular, if precise and accurate estimates of the free energy of a system could be obtained directly from numerical simulations, the need to measure thermodynamic properties of a system, such as ligand-binding constants [28], by experiment would be greatly reduced. This is why, for example, free energy calculations have attracted much interest in areas such as rational drug design and material science. Free energy calculations can be classified into two categories. The calculation of conformational free energies aims at evaluating the relative free energies of relevant conformational states of a given molecular system. The calculation of alchemical free energies on the other hand aims at evaluating the relative free energies of different molecules, which is the main objective of this work.

Free energy is an extensive property, meaning that its magnitude depends on the amount of a substance in a given thermodynamic state, and it's expressed in two forms depending on the conditions that it is determined: The relative (Helmholtz) free energy F is obtained under isochoric-isothermic conditions (NVT), whereas the relative (Gibbs) free enthalpy G is obtained under isobaric-isothermic conditions (NPT). Both expressions are given by:

$$F = U - TS \tag{28}$$

$$\begin{aligned} G &= U + PV - TS \\ G &= H - TS \end{aligned} \tag{29}$$

From an MD trajectory, described in Section 2, the statistical equilibrium averages

can be obtained for any desired property of the molecular system for which a value can be computed at each point of the trajectory. Examples of such properties are the potential or kinetic energy of relevant parts of the system, structural properties and fluctuations, electric fields, diffusion constants, etc. A number of thermodynamic properties can be derived from such averages. However, two important thermodynamic quantities, the entropy and the (Gibbs or Helmholtz) free energy, generally cannot be derived from a statistical average. They are global properties that depend on the extent of phase (or configuration) space accessible to the molecular system [14]. Fortunately, with statistical mechanics methods one can evaluate the *relative* free energy differences.

4.1 Thermodynamic Integration

The aim of thermodynamic integration is to compute the difference in a thermodynamic property (usually the free energy) of the system between some reference state and the state of interest. It make use of the fact that the free energy changes related to small perturbations of a molecular system can be determined during a simulation. Taking that fact into account, the free energy difference between two states A and B of a system can be determined from an MD simulation in which the potential energy function $V(q)$ is slowly changed such that the system slowly changes from state A to state B over a reversible path.

Since the absolute value for free energy of solvation is difficult to calculate directly, we use the fact that the free energy is a **state function**. From the thermodynamic cycle presented in Figure 13 it follows that:

$$\Delta F_{solute}^{solvation} = \Delta F_{dummies,solute}^{vacuo} + \Delta F_{dummies}^{solvation} - \Delta F_{dummies,solute}^{solvent} \quad (30)$$

The solvation free energy $\Delta F_{solute}^{solvation}$ is the work required to transfer a molecule from the gas phase into solute solution. $\Delta F_{dummies,solute}^{vacuo}$ is the work required to remove all the internal nonbonded interactions in the solute in vacuo, which is achieved by gradually mutating all atoms in a solute molecule into dummy atoms. A dummy atom is an atom for which the nonbonded interactions, i.e. Lennard-Jones and electrostatic interactions, with all other atoms are set to zero while the bonded interactions within the molecule and

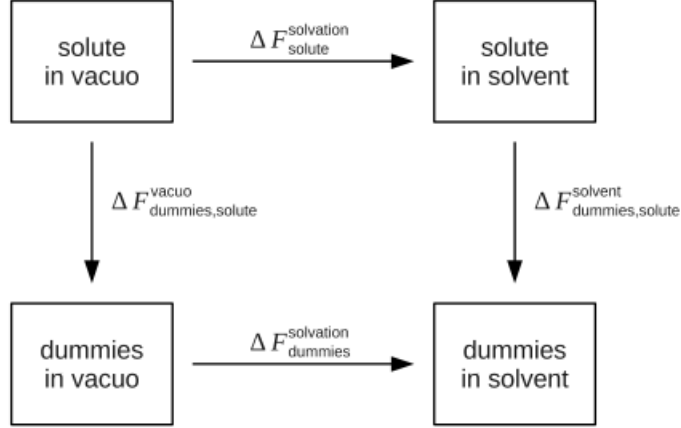


Figure 13: Thermodynamic cycle for the determination of solvation free energies

the masses of individual atoms are kept unchanged. $\Delta F_{dummies}^{solvation}$ is the work required to transfer the dummy solute molecule from vacuum to the solvated phase. As the dummy molecule does not interact with the rest of the system this term is also equal to zero. In order to determine the solvation free energy of any peptide with this method, only the free energy $\Delta F_{dummies,solute}^{solvent}$ must therefore be calculated. Since $\Delta F_{dummies,solute}^{solvent}$ is the work required to remove the solute-solvent and solute-intermolecular interactions. for this particular case, as we are interesting only in the solute-solvent interaction (Van der Waals) the intermolecular forces between the peptide atoms remain constant.

This transition, from a particular atom to a dummy particle, works as follows: **(1)** The Hamiltonian $H(p, q)$ is made a function of a coupling parameter λ , such that $H(p, q, \lambda)$ characterises the two different sates of interest, normal and dummy particle.

$$H(p, q, \lambda) = \sum_{i=1}^N \frac{p_i^2}{2m_i(\lambda)} + V(q, \lambda) \quad (31)$$

with

$$m_i(\lambda) = (1 - \lambda)m_i^A + \lambda m_i^B \quad (32)$$

and

$$\begin{aligned}
V(q, \lambda) = & \sum_{bonds, b} V(b, \lambda) + \sum_{angles, \theta} V(\theta, \lambda) + \sum_{torsion, \xi} V(\xi, \lambda) + \\
& + \sum_{dihedrals, \xi} V(\psi, \lambda) + \sum_{nonb(i, j)} V(r_{ij}, \lambda)
\end{aligned} \tag{33}$$

Equation 33 represents each one of the potential energy terms described in Equations 4 and 5 with the couple parameter λ . In this work, only the non bonded interaction will be modify with the coupling parameter and study the transition between the twp states. The mathematical formulation will be presented in following sections.

4.1.1 Gibbs Free Energy

The Gibbs free energy of the system becomes a function of λ and is given by:

$$G(\lambda) = -kT \ln \Delta\lambda \tag{34}$$

where k denotes Boltzmann's cosntant, T is the temperature of the system and $\Delta\lambda$ is the isobaric partition function given by:

$$\Delta\lambda = \frac{1}{h^{3N} N!} \int \int \int \exp [(-H(p, q, \lambda) + PV)/kT] dV dp dq \tag{35}$$

where P and V are the pressure and the volume of the system respectively, p and q are the momentum and Cartesian coordinates of the N atoms.

The free energy difference ΔG_{BA} is givne by:

$$\Delta G_{BA} = G(\lambda_B) - G(\lambda_A) = -kT \ln \left(\frac{\Delta\lambda_B}{\Delta\lambda_A} \right) \tag{36}$$

which can be expressed as an ensemble average by:

$$\begin{aligned}\Delta G_{BA} &= -kT \ln \\ &\left\{ \frac{\int \int \int \exp [(-H(p, q, \lambda_B) - H(p, q, \lambda_A))/kT] \exp [(-H(p, q, \lambda_A) + PV)/kT] dV dp dq}{\int \int \int \exp [(-H(p, q, \lambda_A) + PV)/kT] dV dp dq} \right\} \\ &= -kT \ln \{ \langle \exp [(-H(p, q, \lambda_B) - H(p, q, \lambda_A))/kT] \rangle \} \end{aligned} \quad (37)$$

where the $\langle \dots \rangle_\lambda$ mean an ensemble average over the volume V , the coordinates q and momentum p at the value λ . Equation 37 is called the perturbation formula, since only yield accurate results when state B is close to state A. If this difference is large, that change from A to B must be split up into a number of steps between intermediate states that are close enough to allow the use of Equation 37 [14]. Meeting this condition, ΔG_{BA} is just the sum of the individual ΔG for all intermediate steps. The differentiation of Equation 34 with respecto to λ at constant temperature and pressure yields:

$$\begin{aligned} \left(\frac{\partial G(\lambda)}{\partial \lambda} \right)_{T,P} &= - \left(\frac{kT}{\Delta(\lambda)} \right) \left[\frac{\partial \Delta(\lambda)}{\partial \lambda} \right]_{T,P} \\ &= \frac{\int \int \int [\partial H(p, q, \lambda)/\partial \lambda] \exp [-(H(p, q, \lambda) + PV)/kT] dV dp dq}{\int \int \int \exp [-(H(p, q, \lambda) + PV)/kT] dV dp dq} \\ &= \left\langle \frac{\partial H(p, q, \lambda)}{\partial \lambda} \right\rangle_\lambda \end{aligned} \quad (38)$$

The free-energy difference (ΔG) between two states, A and B, can be readily calculated using the TI formula:

$$\Delta G_{BA} = G(\lambda_B) - G(\lambda_A) = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \quad (39)$$

where $H(\lambda)$ is a combined potential energy function connecting the potential energy functions with a coupling parameter λ between states $A(H_A; \lambda = 0)$ and state $B(H_B; \lambda = 1)$.

In this methodology, state A is represent the peptide with the normal peptide-solvent interactions, while state B refers to a non-interacting dummy particle described in the previous paragraph. In the practice, the result of Equation 39 is computed a every λ point and then numerically integrated to obtain the free energy difference between the two states. Figure 14 represent this idea.



Figure 14: Transition between the states A and B with according to coupling parameter λ

If λ is being changed very slowly from λ_A to λ_B during a MD simulation, the integration of Equation 39 can be carried out in the course of the MD run. In this way ΔG_{BA} can be obtained for rather different states A and B, as long as the continuous change in λ is so slow that the system remains essentially in equilibrium for each intermediate value of λ .

4.1.2 Soft-Core Potential Energy

A practical problem affect the predicted free energy difference and that is the presence of a "singularity" in the function for which an ensemble average is to be determined, in this case, the Hamiltonian. When the ensemble average is obtained by MD simulation such singularities may also lead to numerical instabilities in the integration of the equations of motion [29]. In general, a singularity problem arises in simulations when an observable, for which the ensemble average is to be calculated, contains regions high-energy regions of the Hamiltonian. Depending on the pathway chosen, the ensemble average of this observable may diverge. The van der Waals interaction and the electrostatic interaction between charges of equal sign (repulsive forces) become infinitely large for r_{ij} approaching zero.

In the practice, the singularity problem is often encountered when non-bonded interactions sites are to be removed, added or modify. A soft-core potential energy function is employed by GROMOS software to avoid singularities near the end-states of the TI

simulations. The Lennard-Jones potential (LJ) energy function, Equation 14, between atoms i and j at a state X is then defined as:

$$H_{i,j,\lambda}^{VdW} = \left[\frac{C_{12}^X(i,j)}{\alpha_{LJ}\lambda^2 C_{12,6}^X(i,j) + (r_{ji})^6} - C_6(i,j) \right] \cdot \frac{1}{\alpha_{LJ}\lambda^2 C_{12,6}^X(i,j) + (r_{ji})^6} \quad (40)$$

The singularity at $r_{ij} = 0$ is removed and the interaction function is smoothened for $\lambda = 0$ and α_{LJ} is the softness parameter for the LJ interaction. A similar transformation is used to couple the λ parameter to electrostatic potential (Equation 10) and it's given by:

$$H_{i,j,\lambda}^{elec} = \frac{q_i^X q_j^X}{4\pi\epsilon_0\epsilon_1} \left[\frac{1}{[\alpha_{CRF}(i,j)\lambda^2 + r_{ij}^2]^{\frac{1}{2}}} - \frac{\frac{1}{2}C_{rf}r_{ij}^2}{[\alpha_{CRF}(i,j)\lambda^2 + R_{rf}^2]^{\frac{3}{2}}} - \frac{(1 - \frac{1}{2}C_{rf})}{R_{rf}} \right] \quad (41)$$

where q_i^X is the partial charge of atom i and C_{rf} and R_{rf} are parameters of the reaction-field method [30]. α_{CRF} is the softness parameter for the electrostatic interactions, similar to α_{LJ} described before. Finally, the non-bonded potential energy for particles i, j in a given state (from A to B) is calculated as:

$$H_{i,j,\lambda}^{NB} = H_{i,j,\lambda}^{VdW} + H_{i,j,\lambda}^{elec} \quad (42)$$

The combined non-bonded potential energy function connecting states A and B is now:

$$H(r_{ij}; \lambda) = \lambda H^{NB}(r_{ij}; B; (1 - \lambda)) + (1 - \lambda) H^{NB}(r_{ij}; A; \lambda) \quad (43)$$

Because the focus of the present study is on calculating non-polar contributions to the solvation process, the alanine molecules were treated as neutral without any partial charges on the atoms for simplicity and to avoid any electrostatic contributions to solvation. This was accomplished by set to 0 all the atomic charges of every atom in the ***.top** file.

5 Methods

5.1 Structure Preparation

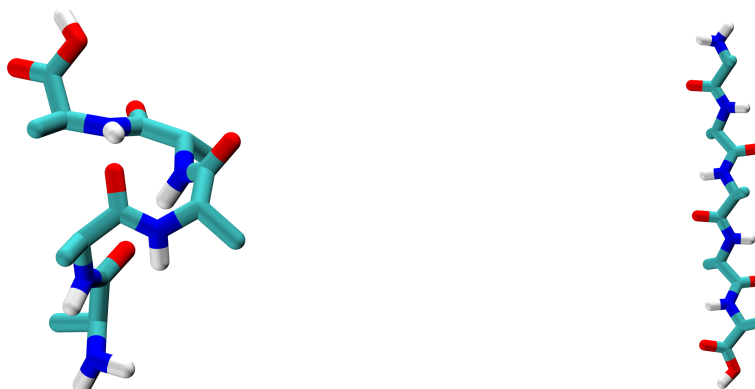
In order to study the behavior of non-polar interaction in the solvation process, the free energy calculation, and how these two properties are related to the volume (size) of the molecules, a total of 20 structures were prepared to preform MD simulations using VMD-**molefacture** plugin [31]. This tool has been designed to facilitate the construction and parameterisation of small molecules. It additionally provides a simple interface to prepare structures and files. These structures classifies as follows:

- 10 helix *n*-Alanines: These correspond to the “natural” configurations of the alanines with curvatures present on they structures. (Figure 15a)
- 10 extended *n*-Alanines: They conformation where fixed to maintain a straight configuration along an axis. (Figure 15b)

with n ranging from 1 (ALA_1) to 10 (ALA_{10}) in both configurations. The peptide terminals where capped with with neutral amino-terminal (NH_2) and C-terminal ($COOH$). From this point, every step was preform using GROMOS molecular simulation software [32].

5.2 About GROMOS software

GROMOS is an acronym of the **GR**Oningen **MO**lecular **S**imulation computer program. It’s a software package used for computer simulations of molecular systems like proteins, inorganic and organic chemical compounds, DNA, etc, developed since 1978 and it is written in C++ programming language. In order to preform MD simulations with



(a) Representation of helix-(ALA_5) configuration.

(b) Representation of extended-(ALA_5) configuration.

Figure 15: Graphic representation of both penta-Alanines (helix and extended) configurations generated with VMD.

GROMOS software, three main elements must be generated and used in every step of the simulation. They are:

- Topology of the system (***.top** file)
- Spatial coordinates of the system (***.cnf** file)
- Input file (***.imd** file) with the parameters for the MD simulation

5.2.1 Building a topology

The first task is to generate a molecular topology file containing the topological and force-field data concerning the molecular system under consideration. The topology information is referred to the data on the covalent structure, atomic masses, charges, van der Waals parameters, atom-atom distance restraints specification, restraints specification, local-elevation dihedral angles specification, etc. Every constant and parameter described in section 2.2 it's contained in the ***.top** file. In GROMOS, a molecular topology is generated from molecular topology building blocks which carry the topological information of molecules like amino acid residues, nucleotides, etc. The molecular topology building

blocks can be linked together into a complete molecular topology file using the GROMOS++ program `make_top`. In this work, the **GROMOS-54a7** force-field was selected to perform the molecular dynamics simulations.

5.2.2 Generating atomic cartesian coordinates

The next step is to generate the configurational information regarding to the atomic coordinates and atomic coordinate dependent or related quantities, such as velocities and forces, atom-atom distances, dihedral angles, energies, size of the computational box, etc. These two types of information (topological and configurational) are generally stored in separate files, since configurations change continuously during a simulation, whereas molecular topological data generally do not change.

Coordinates for biomolecules are often available from X-ray or Nuclear magnetic resonance spectroscopy (NMR) experiments and can be obtained in Protein Data Bank (PDB) format, which can be converted to GROMOS format using the GROMOS++ program `pdb2g96`. In this case, the ***.pdb** files for both alanines configurations (extended and helix) are generated using VMD's plugin **Molefacture** which is a user interface to create and edit molecules. This includes the ability to add, delete, or manipulate their structure at an atomic level, and to build new components from a library of common fragments. As the coordinates for hydrogen atoms are not present in ***.pdb** files (usually) they have to be generated using the GROMOS++ program `gch`. In the standard GROMOS force fields, part of the hydrogen atoms (polar, aromatic) are explicitly treated, whereas other hydrogen atoms (aliphatic, some aromatic) are implicitly treated by incorporation into the (carbon)atom to which they are attached. Program `gch` calculates optimal positions for hydrogen main structure atoms.

5.2.3 Energy minimisation of the peptide

Before putting any peptide into a box of solvent, its configuration has to be relaxed by energy minimisation. The use of this technique with an empirical energy function such as Equation 3 is a widely used tool for obtaining low-energy configurations of a molecular system. In this work, the *steepest descent method* is implemented for this

propose [14]. When applying energy minimization algorithms one searches for a minimum energy configuration of a system by moving (approximately) along the direction of the gradient of the potential energy through configuration space.

$$\Delta r_i \sim -\frac{\delta}{\delta r_i} V(r_1, r_2, \dots, r_n) \quad (44)$$

where Δr_i is the change in position, V the potential energy function given by Equation 3, and r_i the position vector of a particle i .

Since all configurations used for simulations were not directly obtained from physical data, but instead generated with VMD, these artificial structures can possess high amounts of potential energy so they were minimized with a maximum minimization time of 5 [ps] in 2 [fs] steps (maximum time because the simulation also stops if the potential energy difference between two steps is less than a certain minimum, in this case 0.1 [kJ/mol])

5.2.4 Solvating the peptide

Once the structure is minimized, the next step is to put the peptide in a box of solvent using the GROMOS program `sim_box` which can solvate the solute in a pre-equilibrated box of solvent molecules. There are three types of box's shapes (rectangular, truncated octahedron and cylindrical), and a water model must be determined. In this work, the rectangular box shape and a SCP-water model was selected to perform the simulations.

In order to relax the unfavorable atom-atom contacts between the solute and the solvent, energy minimisation of the solvent should be performed while keeping the solute positionally restrained (i.e. connecting the atom to a reference position by a spring). The restraining is achieved by a harmonic force term with a force constant determined in the simulation input file (`*.imd`).

At this point, two **energy groups** are defined: One is the solute and the other the solvent. If one want to add, for example, counter ions to the simulation, it will be helpful to assign them other groups number. This is a key component because later, for the energy analysis, one can preform individual analysis between energy groups.

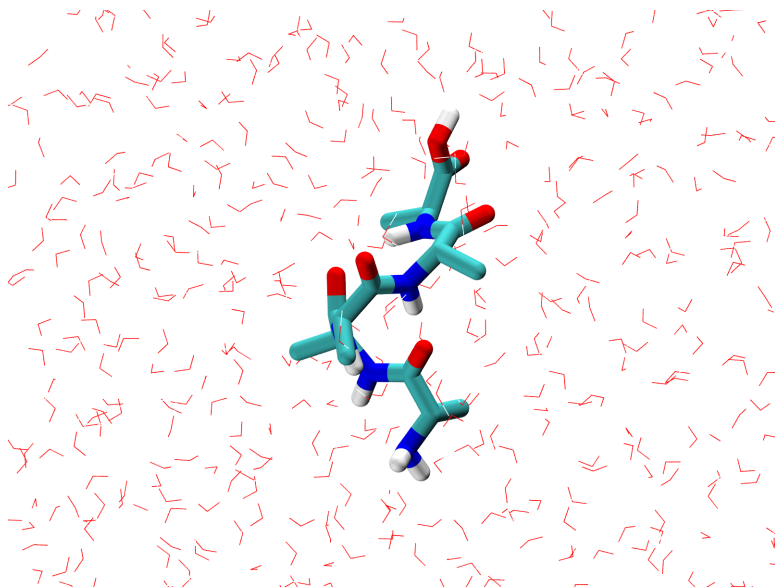


Figure 16: Penta-alanine in box full of water molecules.

5.2.5 Thermalization and equilibration

After generated the topology and initial coordinates for the system, the next step is to generate initial velocities. Since not all molecules from the system travel at the same speed, initial velocities can be randomly generated from a Maxwell-Boltzmann distribution at a low temperature and the system is slowly heated up to the final production simulation temperature:

$$f(v) = \left(\frac{2}{\pi}\right)^{1/2} \left(\frac{m}{kT}\right)^{3/2} v^2 \exp\left[-\frac{mv^2}{2kT}\right] \quad (45)$$

where $f(v)$ is the probability distribution of the velocity (v) for a particle of mass m and kT is the product of Boltzmann's constant and the thermodynamic temperature [33].

The atoms of the solute are positionally restrained and these restraints are loosened while heating up. With the help of these restraints you make sure that the random initial velocities do not disrupt the initial conformation too much. In this work, the equilibration process of every peptide was divided in 5 steps, ranging from 60 [K] to 298.15 [K]. In every step, the temperature of both, peptide and solvent, were increased by 60 [K] in a simulation window of 100 [ps]. Figure 17 shows the result of the process.

n-ALA	Helix		Extended	
	Box size [nm^3]	N^o atoms	Box size [nm^3]	N^o atoms
1	2.879	2310	2.921	2421
2	3.136	2985	3.067	2814
3	3.166	3078	3.028	2706
4	3.075	2820	3.183	3138
5	3.148	3000	3.545	4386
6	3.119	2937	3.901	5820
7	3.299	3.119	4.263	7551
8	3.568	4443	4.620	9537
9	3.736	5106	4.981	12099
10	3.511	4260	5.269	14202

Table 4: Summarize of the 20 system’s dimensions and the number of atoms to be simulated.

5.3 Molecular Dynamics Simulation

The equilibration period already produced a short simulation at constant temperature and volume. At this point we want to elongate the simulation to any time window defined by the user, under constant temperature and pressure.

All MD simulations run for this work had a total length of 100 [ns] with a timestep of 2 [fs] using the leapfrog algorithm described in Section 2.5.1. The SHAKE algorithm [34] was employed to constrain all bonds to their reference values with a relative tolerance of $10^5[kJ/mol - nm^2]$. In order to study only the effects of generating the dry cavity, the solute must be **positionally restrain** using the same method, to minimizes the changes in entropy related to configurations that can generates the movement of the peptide across the water box. This is accomplished by generated two extra configuration files, in which the positionally restrained (***.pos**) atoms and the reference coordinates(***.rpr**) are specified, and are generated from the origin coordinate file (***.cnf**).

Periodic boundary conditions with a cubic box were applied according to the information provided in Table 4. Non-bonded interactions were computed using a triple range cut-off. Interactions within a short-range cut-off of $0.8nm$ were computed every time-step, from a pair-list that was generated every 5 steps. At these time points, interactions

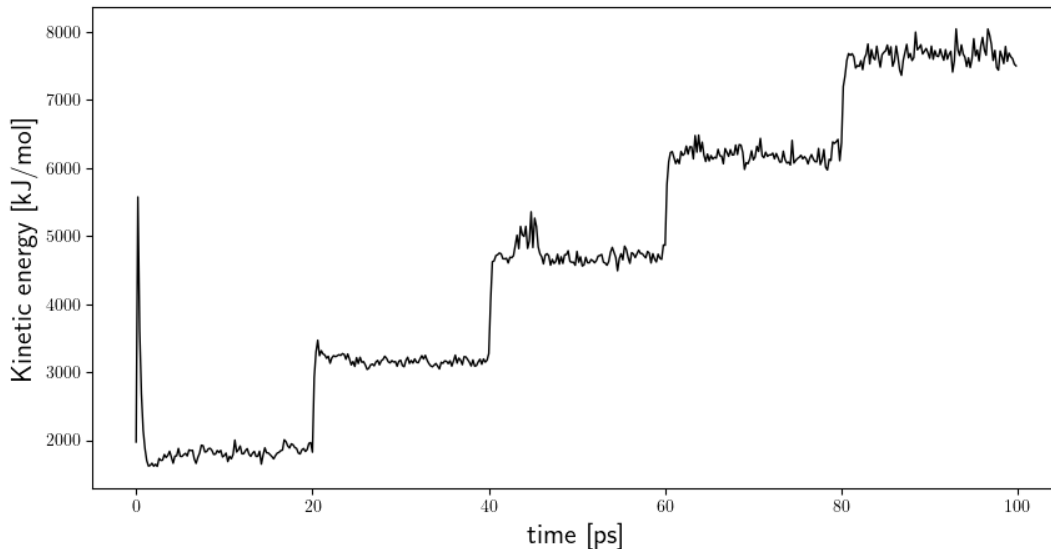


Figure 17: The kinetic energy during the equilibration process of the helix penta-alanine.

between 0.8 and 1.4nm were also computed which were kept constant between these updates. A reaction-field contribution was added to electrostatic interactions approximating for a homogeneous medium outside the 1.4nm cut-off, employing the relative permittivity of SPC water ($\epsilon_w = 61$) [30]. Finally, all simulations were simulated using the isobaric isothermic ensemble (NPT) using the weak-coupling (Berendsen thermostat and barostat) scheme for temperature (298.15[K]) and pressure (1[atm]) control [35].

5.4 Free Energy Calculations

As it is mentioned in Section 4.1, The free energy difference (ΔG) between two states A and B, can be calculated using Equation 39. This equation includes all the potential interactions present in the system. In this case, there are two sources of energy interactions: peptide-peptide (P-P), peptide-solvent (P-S) and solvent-solvent (S-S) interactions. (Notice that the coulomb interactions does not contribute to the Hamiltonian because each

atom's charge present in the peptide, were set to 0).

$$\left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda}^{vdw} = \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{P-P}^{vdw} + \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{P-S}^{vdw}. \quad (46)$$

Both terms in Equation 46 contributes to the total numerical result of the free energy change between the states A and B, but, the main objective of this work is to calculate only the **free energy of forming the dry cavity** generated by the non-polar interaction (ΔG^{vdw}) of the solvated solute. Taking this into account, the therm $\left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{P-P}^{vdw}$ does not contribute to generate the cavity because only consider the interaction between it self, so it must be unchanged during the λ -transition, i.e, the van der Waals **peptide-peptide** interactions must remain unchanged during the process. This was obtain by adding an extra block (LAMBDA) to the molecular dynamics input file (***.imd**) which control the interaction between the coupling parameter λ , and the groups previously define (Sec. 5.2.4). This allows to determine that the parameter only affects the peptide-solvent interaction, groups (1-2). Figure 18 represent this phenomenon.

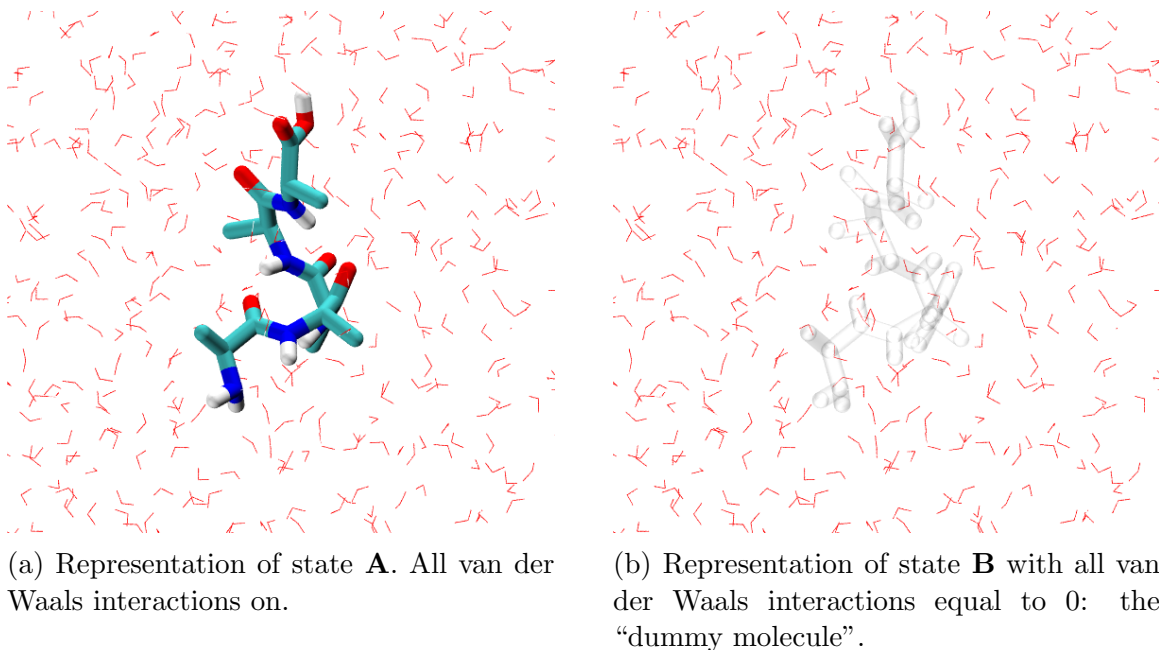


Figure 18: Graphic representation of the process of **Thermodynamic Integration** where a molecules is turn into a dummy molecule according to the couple parameter λ . The transition goes froms State A to State B.

5.4.1 Thermodynamic Integration’s Numerical Error

To compute the free energy between state A and B, the integrate curve over all λ -steps must be calculated. For each value of λ , every $\frac{\partial H}{\partial \lambda}$ in the soft-core potential form (Sec. 4.1.2) is computed. To avoid numerical errors related to the smoothness of the curve generated, and further numerical integration, the transition over the λ -points must assure that the function generated by Equation 39 is smooth enough to prevent this. This is accomplish by simulate more λ -points in between other to generate more $\frac{\partial H}{\partial \lambda}_i$ values for the curve. In this case, five versions of TI were preformed to determine the total number of λ -steps: [10, 17, 20, 23, 32]. Al five versions are attached to this work in 6.4. As a result of that, the λ space, between 0 and 1 were distribute in 32 discrete points:

0.000	0.400	0.600	0.690	0.715	0.735	0.800	0.880
0.100	0.450	0.625	0.700	0.720	0.740	0.820	0.900
0.200	0.500	0.650	0.705	0.725	0.760	0.840	0.950
0.300	0.550	0.675	0.710	0.730	0.780	0.860	1.000

Table 5: Distribution of 32 λ points between 0 and 1.

For each window, 0.2 [ns] of equilibration was performed, followed by 15 [ns] of production MD, from which the free energy differences between adjacent λ values were computed with the methods outlined above, and the total free energies were calculated by integrate the curve over all the λ -steps.

6 Results and Discussion

Here, the results of the free energies of disappear the van der Waals interactions for 20 $n - alanines$ with n ranging from 1 to 10, in both configurations (helix and extended) are presented under the methods highlighted above.

6.1 Energy Interaction Results

First, the discrete transition between states A and B acopled with the λ -parameter, explained in Section 4.1 is presented in Figure 19.

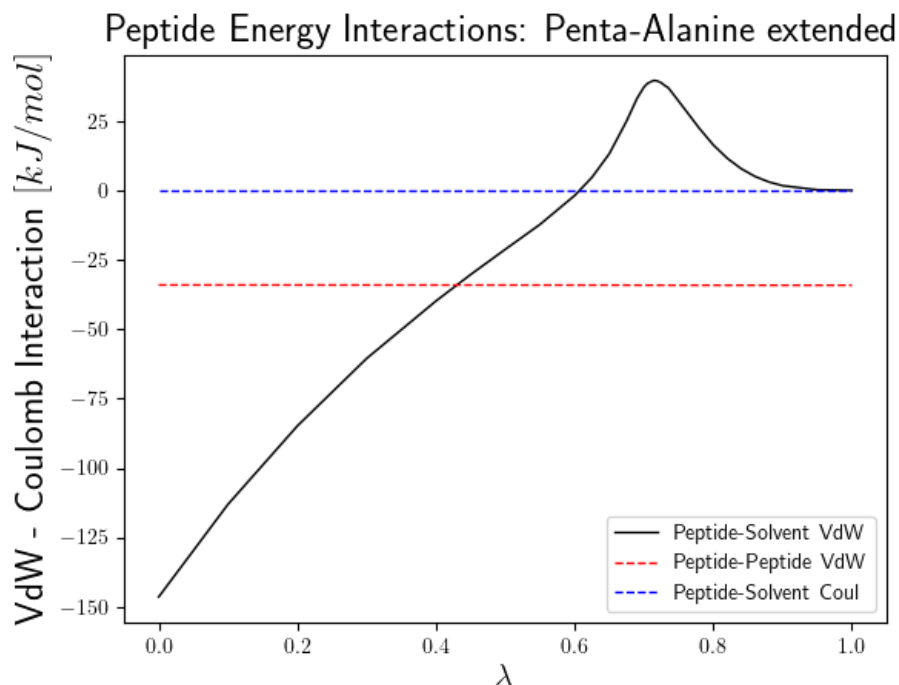


Figure 19: Evolution of interactions between peptide and solvent while the transition of the λ parameter.

The energy interactions between the peptide and the solvent is changed, using the soft core potential function (Equation 41), from the initial value calculating with the force-fields parameters and the system configuration, to zero. Here, two important points must

be notice: (1) the coulombic interactions is equal to zero during all the transition. This is because as it was said before, the atomic charges were manually set to zero (Section 5.1), and (2), the peptide-peptide interaction remains constant during the transition. This two factors, assures that the free energy change ΔG corresponds to the free energy required to turn-off the van der Waals interactions between the peptide and the solvent, i.e, the free energy of **generating the dry cavity on the solvent**.

Figure 19 is a sample of the penta-alanine energy interactions in extended configuration, of all the 20 results. In addition, Figure 20 shows how the solvent van der Waals interactions evolves with itself during the transition of the λ -parameter.

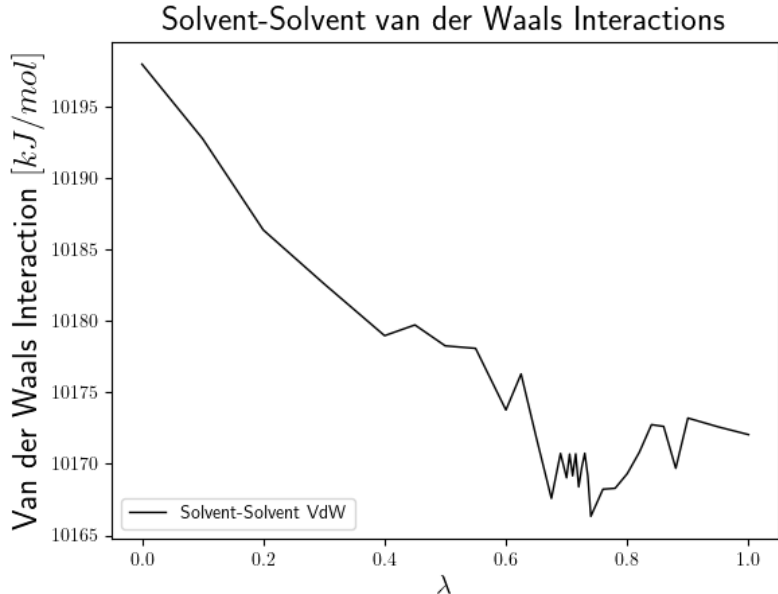


Figure 20: Evolution of van der Waals interactions between the solvent with itself while the transition of the λ parameter.

Finally, the coulomb interaction between the solvent is presented in Figure 21. Notice that both interactions becomes more negative, i.e, more favorable during the transition because in the lack of interaction between molecules, or the presence of a dummy atom, allows the connection between more solvent molecules. We can think of this like the normal molecule is acting as a screen for the solvent molecules that are at the other side. in the process of turn-off the peptide-solvent interactions, one can see how the solvent molecules

that wasn't interacting with themselves, starts to contribute to the total value of these interactions.

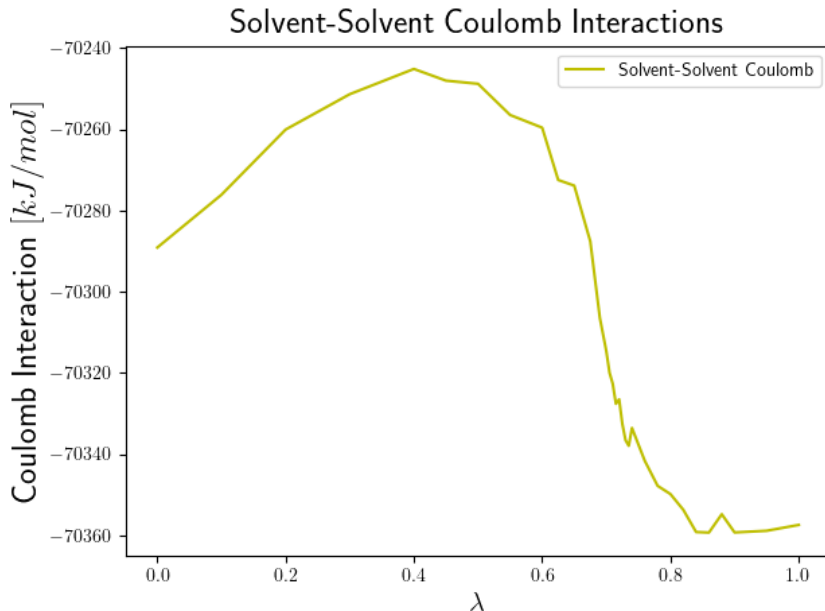


Figure 21: Evolution of coulomb interactions between the solvent with itself while the transition of the λ parameter.

6.2 Free energy Results

Figure 22 illustrates the $\frac{\partial H}{\partial \lambda}$ values for each λ -points define in Section 5.4.1. This particular result correspond to the hepta-alanine (ALA_7) in helix configuration, and the total result of the free energy of solvation correspond to the integrate of all the points over the λ space. This exercise was preform for every configuration and the results are presented as follows: Table 6 summarizes each numerical result for each configuration, both extended and helix, and Figure 23 illustrates the free energies. All TI curves are attached to this work in the Appendix 6.4.

As it it seen in Figure 23, a clear linear relationship between the size of the peptide that it is solvated, and the free energy required to generate the cavity it's seen.

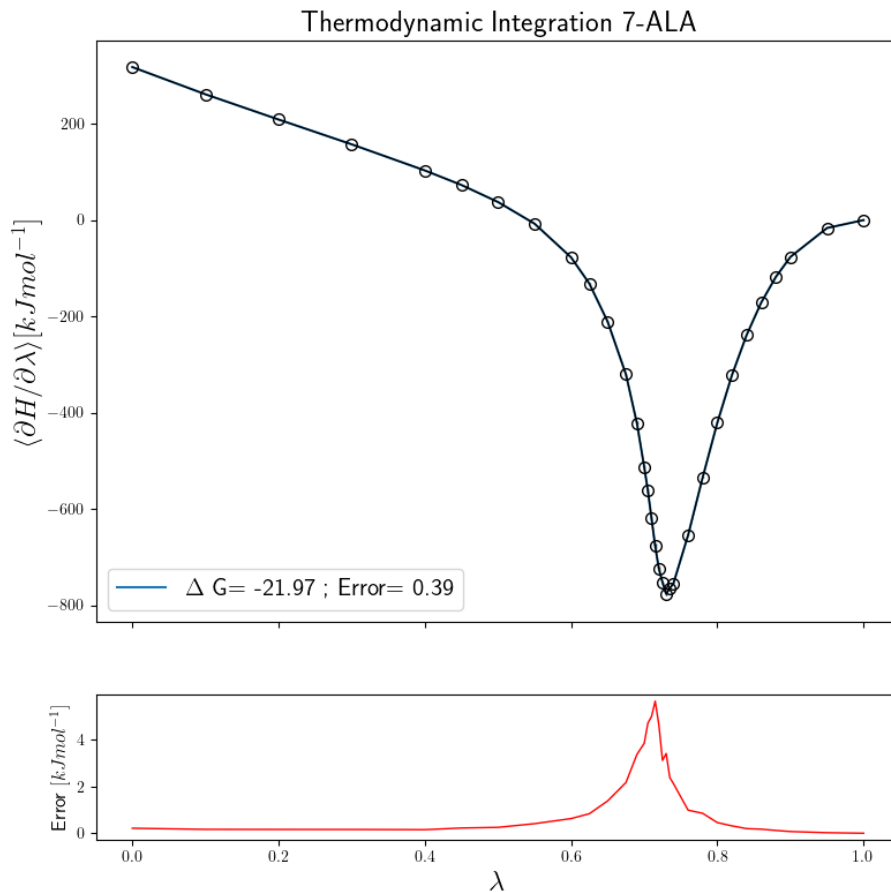


Figure 22: TI curve for hepta-alanine in helix configuration.

6.2.1 Statistical Error

As it is seen in Figure 22, there are error values related to each λ -point. One can think that as a simulation is preform, there is no experimental measure to compare the results to estimate an error. For this, GROMOS software calculate **averages** and **standard deviation** by blocks for to determine a *statistical error* to each result. This is: all teh simulation samples are group in several blocks and for each block, the mean of the propertie of interest, is calculated. Then the mean of the mean of the blocks is reported and the standard deviation of this mean. This is because what we measure in general behaves

n-ALA	Helix		Extended	
	$\Delta G_{cav} [kJ/mol^{-1}]$	Error $[kJ/mol^{-1}]$	$\Delta G_{cav} [kJ/mol^{-1}]$	Error $[kJ/mol^{-1}]$
1	-6.350	0.153	-7.752	0.152
2	-10.198	0.207	-10.488	0.213
3	-12.709	0.261	-12.641	0.242
4	-16.198	0.291	-14.020	0.291
5	-16.524	0.333	-15.767	0.323
6	-20.223	0.361	-18.007	0.356
7	-21.969	0.391	-19.996	0.376
8	-23.102	0.422	-21.915	0.383
9	-24.663	0.433	-24.103	0.420
10	-26.584	0.460	-25.523	0.477

Table 6: Summarize of the 20 free energies calculations with statistical error.

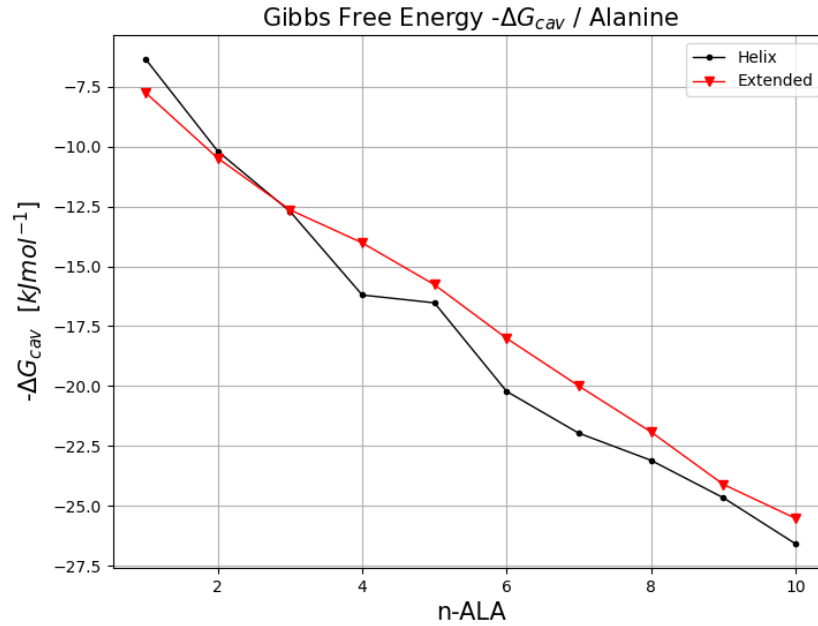


Figure 23: Solvation free energies of the extended and helix alanine peptides plotted as functions of the number (N) of residues in the peptide.

like normal distributions, so the deviation does not tell us the error, but tells us other properties. But if we make blocks we can estimate as if we had made independent mea-

surements, and from each one we have taken a single value (the average). With this, it's possible to report an error [36].

To estimate the precision of Equation 39, the following formula can be used for the standard deviation on the mean [37]:

$$\sigma \left(\left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda_n} \right) = \left(\frac{S_{\lambda_n}}{N_{conf}} \right)^{\frac{1}{2}} \sigma \left(\frac{\partial H}{\partial \lambda} \right) \quad (47)$$

where N_{conf} denotes the number of configurations over which the ensemble average at $\lambda = \lambda_n$ is taken, and $\sigma \left(\frac{\partial H}{\partial \lambda} \right)$ is calculated as the square root of the variance over the ensemble.

$$\sigma^2 \left(\frac{\partial H}{\partial \lambda} \right) = \frac{1}{N_{conf}} \sum_{t=1}^{N_{conf}} \left[\frac{\partial H}{\partial \lambda} - \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda_n, N_{conf}} \right]^2 \quad (48)$$

To calculate the statistical inefficiency, S_{λ_n} , the total simulation time at $\lambda = \lambda_n$ is divided into M blocks of length b , and N_b configurations are taken from each block such that $M \cdot N_b = N_{conf}$ [36]. The variance of the mean is then calculated for each possible block length b as:

$$\sigma^2 \left(\left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda_n, b} \right) = \frac{1}{M} \sum_{m=1}^M \left[\left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda_n, N_b, m} - \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_{\lambda_n, N_{conf}} \right]^2 \quad (49)$$

where $\langle \dots \rangle_{\lambda_n, N_b, m}$ denotes the ensemble average at $\lambda = \lambda_n$ over N_b configurations of the m -th block. S_{λ_n} is then calculated as:

$$S_{\lambda_n} = \lim_{N_b \rightarrow \infty} \frac{N_b \sigma^2 \left(\left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda_n, b} \right)}{\sigma^2 \left(\frac{\partial H}{\partial \lambda} \right)} \quad (50)$$

It is the limiting ratio of the observed variance of an average to the limit expected on the assumption of uncorrelated Gaussian statistics. Having estimated S_{λ_n} , we can write Equation 48. This statistical error is presented in Table 6 and in the bottom graph in Figure 22. This error is corrected by extending the simulation time for the TI because it generates more configuration to sample. This is the main reason why in Section 5.3 the

simulation time for the TI was determined at 15[ns].

6.3 Comparison Between Models

For the standard SASA model, described in Equation 20, ΔG_{cav} is computed with $\gamma = 2.51[kJ/mol/nm^2]$ or $0.06[kcal/mol/\text{\AA}^2]$ and $b = -12.55[kJ/mol]$ or $-3[kcal/mol]$. For the propose implicit solvent model, the following constants were used according to Equation 27: $\Phi_{static} = 10.7[kcal/mol/e]$, $\epsilon_{shell} = 7.75$, and $\rho_{shell} = 1.8\rho_w$, with $\rho_w = 0.0336\text{\AA}^{-3}$ the bulk water density. All this parameters were selected to calibrate the model after compared the results with the Mobley's explicit-solvent free energy perturbation (FEP) calculations [24]. Figures 25 and 24 shows these results.

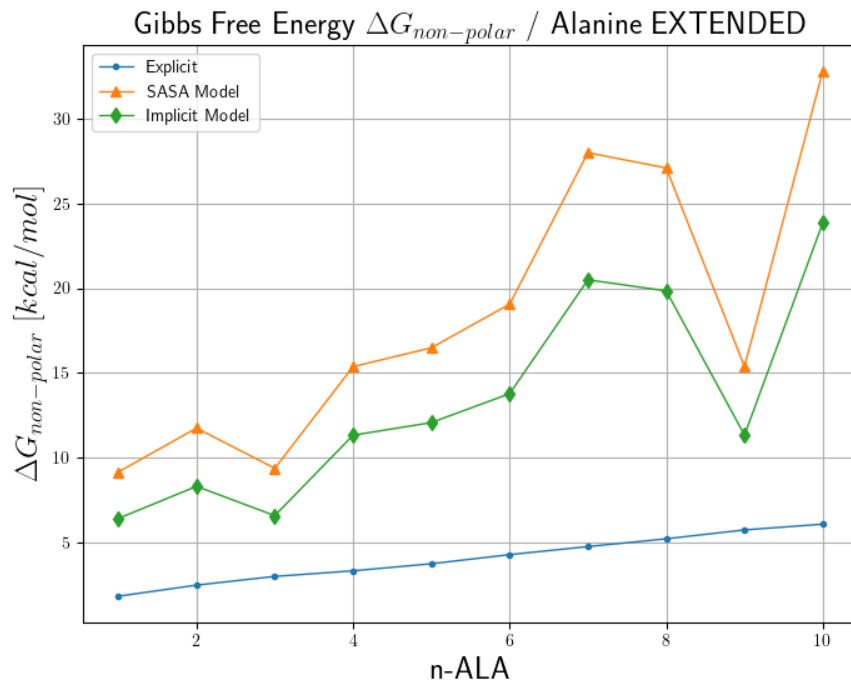


Figure 24: Gibbs free energy results from the three models described in this work for extended configuration, per alanine peptide.

The use of parametrizations, like the constants described in Section 6.3, or the force-field use for the simulations, implies an intrinsic source of uncertainty specially because

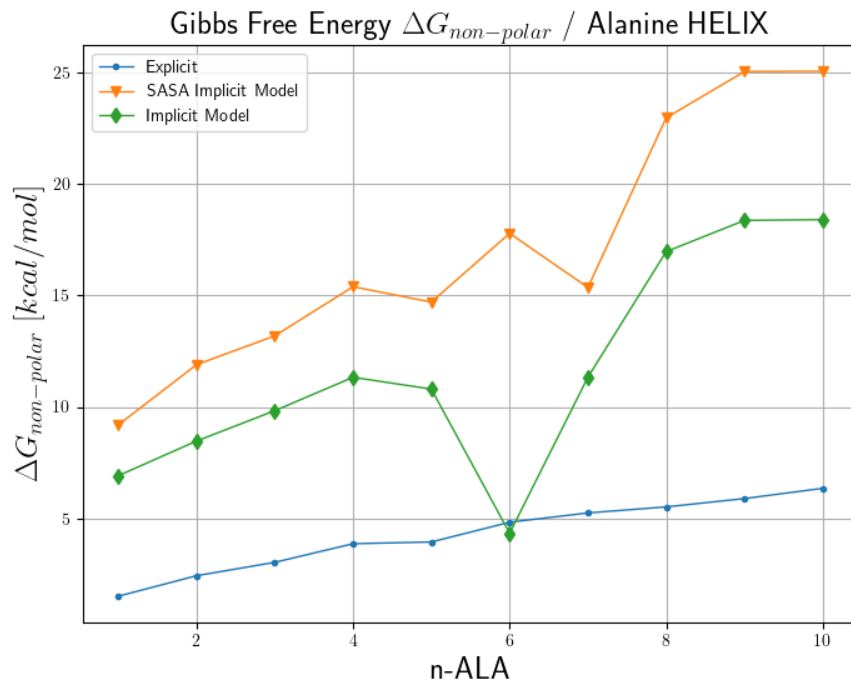


Figure 25: Gibbs free energy results from the three models described in this work for helix configuration, per alanine peptide.

the lac of empirical measurements related to the free energy calculations.

As it's seen in Equation 20 and 27, both models depends strongly on the SASA parameter. The standard SASA model can be interpreted as a linear correlation dependent on the SASA factor, and the Implicit solvent model, uses this surface to integrate the potential on it (Eq. 26). Depending on the method used to generate the SASA factor, both models can carry inaccuracies on the calculations. Figure 26 present the results of tho methods to obtain the SASA factor, one with GROMOS program `sasa` and the other generated with `msms` and used in the implicit model.

The calculation with implicit and standard sasa model (Figure 25 and 24), shows “anomalies” in the free energy values in ($ALA_9 - extended$) and ($ALA_6 - helix$, $ALA_7 - helix$). This is correlated with the same behavior in SASA factors showed in Figure 26 which generates the free energy results. To correct this, the method to generate the mesh around the peptide coordinates must be fixed, for example, using a probe with higher

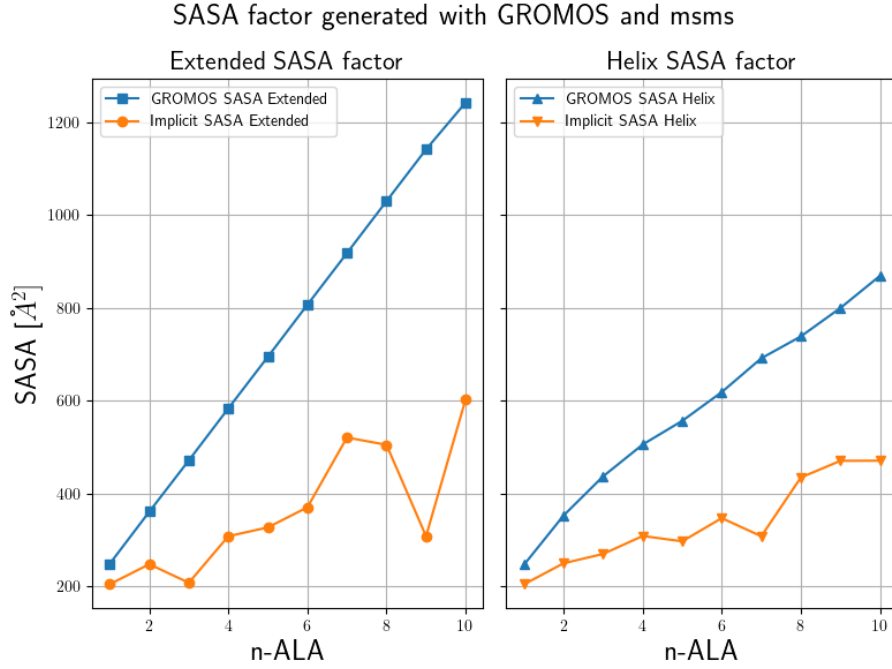


Figure 26: Differences between SASA factor generated with `GROMOS` software and `msms` program.

molecule radius.

To measure the accuracy of the prediction for ΔG_{np} of both models, Pearson coefficient ρ_p , root mean square of the difference (RMSD) and the correlation coefficient R^2 are used and presented in Table 7.

Model	Helix			Extended			Total
	R^2	RMSD	ρ_p	R^2	RMSD	ρ_p	
SASA Model	0.853	22.339	-0.949	0.651	24.216	-0.807	23.296
Implicit Model	0.503	16.939	-0.709	0.661	18.791	-0.813	17.889

Table 7: Comparison of ΔG_{np} between standard SASA model and the Implicit model proposed.

With the parameters described at the beginning of this section, the implicit solvent model presents an improvement of 22.53 % in predicting the free energy of solvation ΔG_{cav} over the standard SASA model.

6.3.1 Adjust Models constants

To improve the accuracy of both implicit models, one can adjust the constants to fit the explicit MD curve and improve the overall forecast result. Taking this into account, modifications to both models were made: First, related to the standard SASA model, we propose the values of $\gamma = 0.025$ and $b = -4.1$ in contrast to the values presented above [4]. Second, we propose a value of $\epsilon_{shell} = 3.1$ in contrast to the value reported in [4] of $\epsilon_{shell} = 7.75$ as a fitting parameter. The result of this consideration are presented in Figure 27. Table 8 summarizes the results with the new parametrization.

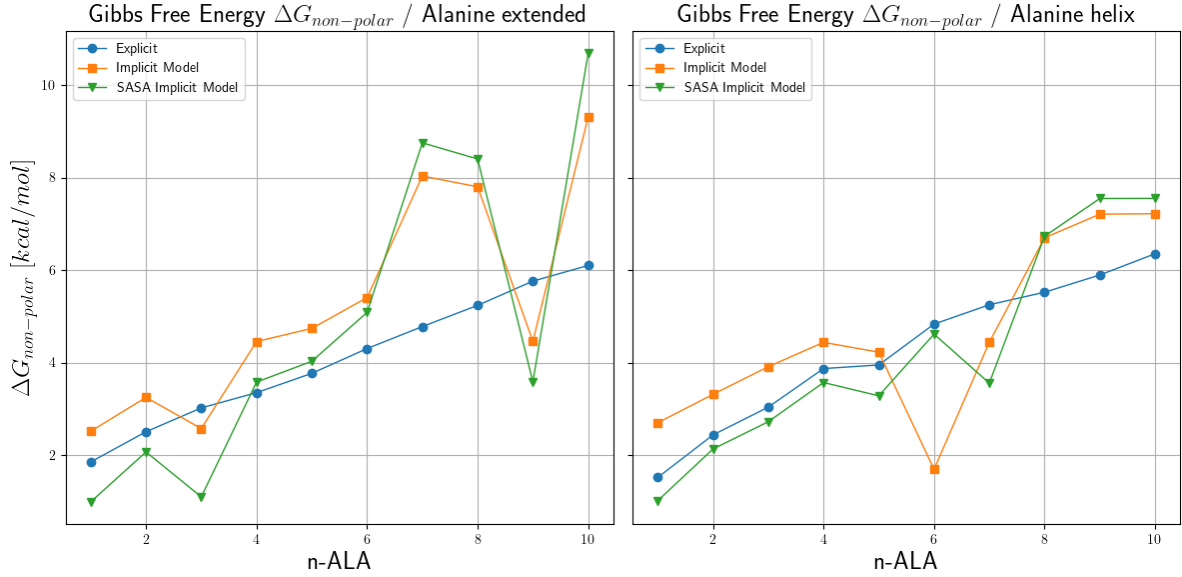


Figure 27: Results after adjustment of parameters $\gamma = 0.025$, $b = -4.1$ for SASA standard model, and $\epsilon_{shell} = 3.1$ for the implicit model presented in this work.

Model	Helix			Extended			Total
	R^2	RMSD	ρ_p	R^2	RMSD	ρ_p	RMSD
SASA Model	0.852	9.269	-0.949	0.653	9.905	-0.808	9.592
Implicit Model	0.501	9.364	-0.708	0.662	9.935	-0.814	9.654

Table 8: Comparison of ΔG_{np} between standard SASA model and the Implicit model with the re-parametrization.

6.4 Conclusions

In this work, we studied the contribution of van der Waals interactions in the solvation free energy using the Molecular Dynamics (MD) method. This computer simulation methodology allows us to analyze the physical movements of atoms and molecules which are interacting between them for a fixed period of time and used to extract macro-thermodynamic properties like, in this case, free energies. The solvation free energy is the energy cost of the process of reorganizing solvent and solute molecules into solvation systems. The solvation process describes the interaction between the solute and solvent molecules and it is govern by electrostatic and van der Waals interactions. In this case, we focus on the free energy which is calculated by the second component. To analyze how the property changes with the size of the molecule, 20 alanine-peptide structures were prepared ranging from 1 to 10, and in helix and extended configuration. Finally, free energy results were calculated using the Thermodynamic Integration (TI) method during a MD run.

From all the 20 TI, it was clear the linear relation between the volume (size) of the alanine peptides, represented in Figure 23, which is an expected result. If one want to solvate a bigger molecule, higger free energies will result from these solvation processes. In the other hand, as we add identical residues to the next alanine-peptide, from ALA_1 to ALA_2 , from ALA_2 to ALA_3 and so on, we expect a linear constant characteristic from the alanine peptide.

We also analyze two implicit methods to estimate this free energy change. One standard method that relates the Solvent-Accessible Surface Area (SASA) with some fitting constants ($\gamma = 0.06[kcal/mol/\text{\AA}^2]$ and $b = -3[kcal/mol]$) that have shown high accurate results, and a new method that relies in a clearer physical meaning and formulation which uses a capacitor model to emulate the interaction distribution around the solvated peptide.

To compare the accuracy of the implcit models, the calculations generated with the explicit-MD simulation was used as the “real/observed” result, and contrast them with the implicit results. It was found the new method for calculating the solvation free energy, and with the parameters described in this work, improves the estimation of the thermodynamic property by **22.53%** according to RMSD parameter.

It is also proposed a re-parametrization of the fitting constants according to the results

presented in this work of $\gamma = 0.025$ and $b = -4.1$ in the case of the standard SASA model, and $\epsilon_{shell} = 3.1$ for the propoused implicit solvent model.

References

- [1] W. F. Van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glättli, P. H. Hünenberger, *et al.*, “Biomolecular modeling: goals, problems, perspectives,” *Angewandte Chemie International Edition*, vol. 45, no. 25, pp. 4064–4092, 2006.
- [2] C. Chipot, “Numerical methods for molecular dynamics simulations of biological systems,” 2010.
- [3] A. Hospital, J. R. Goñi, M. Orozco, and J. L. Gelpí, “Molecular dynamics simulations: advances and applications,” *Advances and applications in bioinformatics and chemistry: AABC*, vol. 8, p. 37, 2015.
- [4] C. D. Cooper and J. P. Bardhan, “A simple electrostatic model for the hard-sphere solute component of nonpolar solvation,” *arXiv preprint arXiv:2005.13019*, 2020.
- [5] U. Museums, G. Commission, *et al.*, “An introduction to materials,” *Conservation science teaching series. Science for conservators*, 1, 1992.
- [6] W. H. Brown and T. Poon, *Introduction to organic chemistry*. John Wiley & Sons, 2016.
- [7] C. Oostenbrink, M. van Lipzig, and W. F. van Gunsteren, “Applications of molecular dynamics simulations in drug design,” *Comprehensive Medicinal Chemistry II*, vol. 4, pp. 651–668, 2007.
- [8] C. D. Christ, A. E. Mark, and W. F. Van Gunsteren, “Basic ingredients of free energy calculations: a review,” *Journal of computational chemistry*, vol. 31, no. 8, pp. 1569–1582, 2010.
- [9] D. A. Case, T. Darden, T. E. Cheatham III, C. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, D. A. Pearlman, M. Crowley, *et al.*, “Amber 9,” *University of California, San Francisco*, vol. 45, 2006.
- [10] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, *et al.*, “Charmm general force field: A force field

- for drug-like molecules compatible with the charmm all-atom additive biological force fields,” *Journal of computational chemistry*, vol. 31, no. 4, pp. 671–690, 2010.
- [11] Y. A. Arnautova, A. Jagielska, and H. A. Scheraga, “A new force field (ecepp-05) for peptides, proteins, and organic molecules,” *The Journal of Physical Chemistry B*, vol. 110, no. 10, pp. 5025–5044, 2006.
 - [12] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen, “Gromacs: fast, flexible, and free,” *Journal of computational chemistry*, vol. 26, no. 16, pp. 1701–1718, 2005.
 - [13] N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark, and W. F. van Gunsteren, “Definition and testing of the gromos force-field versions 54a7 and 54b7,” *European biophysics journal*, vol. 40, no. 7, pp. 843–856, 2011.
 - [14] W. F. van Gunsteren, “The role of computer simulation techniques in protein engineering,” *Protein Engineering, Design and Selection*, vol. 2, no. 1, pp. 5–13, 1988.
 - [15] F. Wang, S. J. Stuart, and R. A. Latour, “Calculation of adsorption free energy for solute-surface interactions using biased replica-exchange molecular dynamics,” *Biointerphases*, vol. 3, no. 1, pp. 9–18, 2008.
 - [16] A. R. Leach and A. R. Leach, *Molecular modelling: principles and applications*. Pearson education, 2001.
 - [17] M. A. Cuendet and W. F. van Gunsteren, “On the calculation of velocity-dependent properties in molecular dynamics simulations using the leapfrog integration algorithm,” *The Journal of chemical physics*, vol. 127, no. 18, p. 184102, 2007.
 - [18] B. Roux and T. Simonson, “Implicit solvent models,” *Biophysical chemistry*, vol. 78, no. 1-2, pp. 1–20, 1999.
 - [19] M. Zacharias, “Continuum solvent modeling of nonpolar solvation: Improvement by separating surface area dependent cavity and dispersion contributions,” *The Journal of Physical Chemistry A*, vol. 107, no. 16, pp. 3000–3004, 2003.
 - [20] N. Campbell, *Chemistry - California Edition*. Pearson Prentice Hall, 2006.

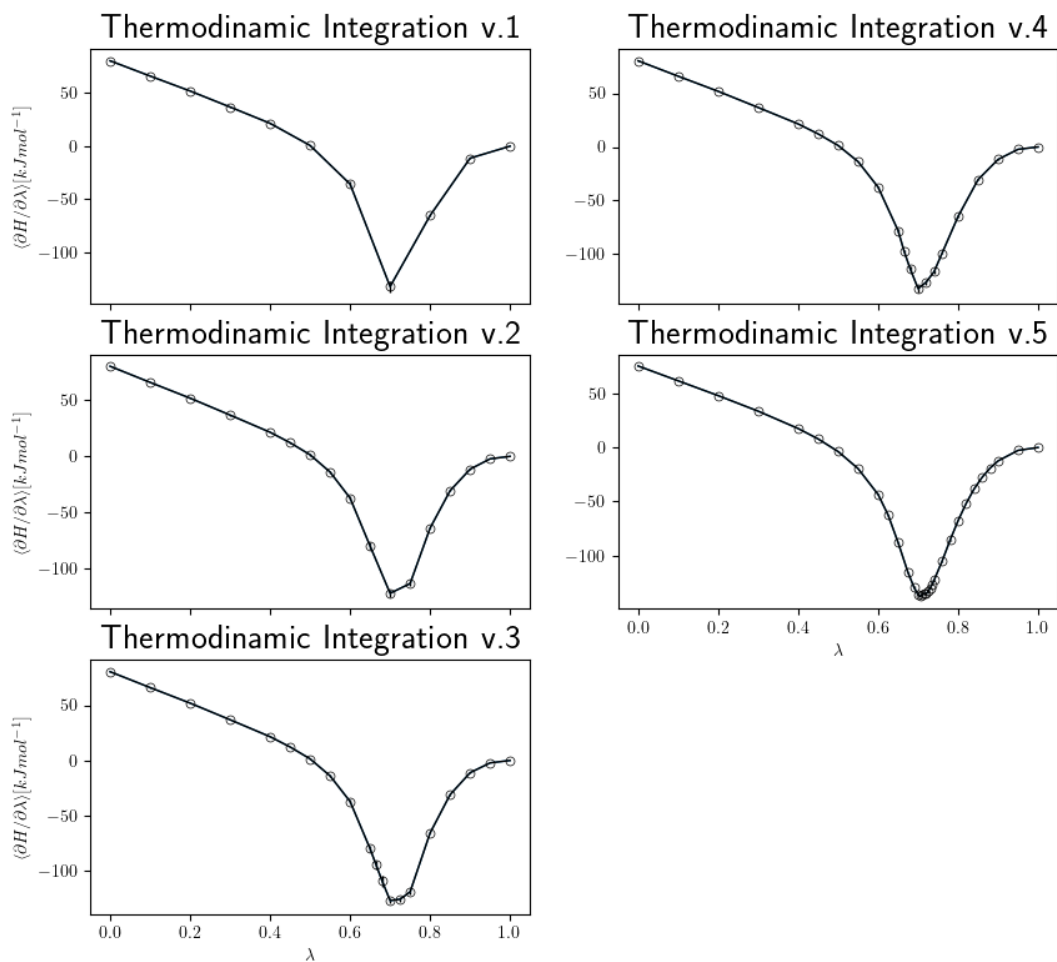
- [21] E. Gallicchio, M. Kubo, and R. M. Levy, “Enthalpy- entropy and cavity decomposition of alkane hydration free energies: Numerical results and implications for theories of hydrophobic solvation,” *The Journal of Physical Chemistry B*, vol. 104, no. 26, pp. 6271–6285, 2000.
- [22] B. Wang, C. Wang, K. Wu, and G.-W. Wei, “Breaking the polar-nonpolar division in solvation free energy prediction,” *Journal of computational chemistry*, vol. 39, no. 4, pp. 217–233, 2018.
- [23] C. Tan, Y.-H. Tan, and R. Luo, “Implicit nonpolar solvent models,” *The Journal of Physical Chemistry B*, vol. 111, no. 42, pp. 12263–12274, 2007.
- [24] D. L. Mobley, C. I. Bayly, M. D. Cooper, M. R. Shirts, and K. A. Dill, “Small molecule hydration free energies in explicit solvent: an extensive test of fixed-charge atomistic simulations,” *Journal of chemical theory and computation*, vol. 5, no. 2, pp. 350–358, 2009.
- [25] M. L. Connolly, “Analytical molecular surface calculation,” *Journal of applied crystallography*, vol. 16, no. 5, pp. 548–558, 1983.
- [26] R. M. Levy, L. Y. Zhang, E. Gallicchio, and A. K. Felts, “On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute- solvent interaction energy,” *Journal of the American Chemical Society*, vol. 125, no. 31, pp. 9523–9530, 2003.
- [27] A. Mehdizadeh Rahimi, A. Molavi Tabrizi, S. Goossens, M. G. Knepley, and J. P. Bardhan, “Solvation thermodynamics of neutral and charged solutes using the solvation-layer interface condition continuum dielectric model,” *International Journal of Quantum Chemistry*, vol. 119, no. 1, p. e25771, 2019.
- [28] C. D. Christ, A. E. Mark, and W. F. Van Gunsteren, “Basic ingredients of free energy calculations: a review,” *Journal of computational chemistry*, vol. 31, no. 8, pp. 1569–1582, 2010.
- [29] T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. Van Gunsteren, “Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations,” *Chemical physics letters*, vol. 222, no. 6, pp. 529–539, 1994.

- [30] I. G. Tironi, R. Sperb, P. E. Smith, and W. F. van Gunsteren, “A generalized reaction field method for molecular dynamics simulations,” *The Journal of chemical physics*, vol. 102, no. 13, pp. 5451–5459, 1995.
- [31] W. Humphrey, A. Dalke, and K. Schulten, “Vmd: visual molecular dynamics,” *Journal of molecular graphics*, vol. 14, no. 1, pp. 33–38, 1996.
- [32] Z. Basharat and A. Yasmin, “Structure, interaction and post-translational modification study of arsenic reduction system in bifidobacterium longum,” *bioRxiv*, p. 066324, 2016.
- [33] H. J. Muller-Kirsten, *Basics of statistical physics*. World Scientific Publishing Company, 2013.
- [34] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, “Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes,” *Journal of computational physics*, vol. 23, no. 3, pp. 327–341, 1977.
- [35] H. J. Berendsen, J. v. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, “Molecular dynamics with coupling to an external bath,” *The Journal of chemical physics*, vol. 81, no. 8, pp. 3684–3690, 1984.
- [36] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*. Oxford university press, 2017.
- [37] X. Daura, P. H. Hünenberger, A. E. Mark, E. Querol, F. X. Aviles, and W. F. Van Gunsteren, “Free energies of transfer of trp analogs from chloroform to water: comparison of theory and experiment and the importance of adequate treatment of electrostatic and internal interactions,” *Journal of the American Chemical Society*, vol. 118, no. 26, pp. 6285–6294, 1996.

Appendices

Thermodynamic Evolution

Thermodynamic evolution for mono-alanine helix configuration



Thermodynamic Integration Curves

