UNIVERSITAT DE BARCELONA

Computational modelling to study medium effect over molecules and biopolymers of interest in biomedicine

Haruna Luz Barazorda Ccahuana

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (**www.tdx.cat**) i a través del Dipòsit Digital de la UB (**diposit.ub.edu**) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (**www.tdx.cat**) y a través del Repositorio Digital de la UB (**diposit.ub.edu**) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (**www.tdx.cat**) service and by the UB Digital Repository (**diposit.ub.edu**) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

UNIVERSITAT DE BARCELONA

Computational modelling to study medium effect over molecules and biopolymers of interest in biomedicine

Haruna Luz Barazorda Ccahuana

A dissertation submitted to the Universitat de Barcelona in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in Theoretical Chemistry and Computational Modelling.

Supervisor: Dr. Sergio Madurga Diez
Co-Supervisor: Dr. Francesc Mas Pujadas

Barcelona
2023

Dedication

To my wonderful parents, sisters, brother, and my beautiful nephew, who have selflessly helped me pursue my dreams.

Acknowledgments

I would like to express my deepest gratitude to Dr. Francesc Mas and Dr. Sergio Madurga, for their unlimited guidance, attention and support throughout my PhD study.

I want to thank my friends, especially Sabrina and Daniel. I feel extremely lucky to have shared many joys with them during the course of my doctorate.

Last, but not least, I would like to thank Mauricio for his unconditional love and support in this final stage of my PhD study.

Table of Contents

Chapter 1

Chapter 2

Chapter 3

Computational Methods **7**

Chapter 4

Effect of pH on the Supramolecular Structure of Helicobacter pylori Urease by Molecular Dynamics Simulations **20**

Chapter 5

Chapter 6

Chapter 9

List of Tables

List of Figures

Resumen

Las tecnologías de la simulación computacional facilitan la resolución de problemas complejos de la biomedicina al ayudar a los investigadores a predecir lo qué sucederá en un sistema natural en respuesta a diversas condiciones externas. Actualmente existen escasos análisis por simulación computacional sobre los efectos del pH en estructuras macromoleculares. Por ejemplo, la ureasa de *Helicobacter pylori* permite la supervivencia del patógeno a condiciones de pH ácidos. Otro caso reciente es la importancia de considerar el pH en blancos terapéuticos de SARS-CoV-2, debido a que algunos mecanismos ocurren a pHs ácidos o pHs alcalinos. Otro problema en biomedicina es la agregación de compuestos aromáticos y la formación de fibras tóxicas que conllevan a grandes problemas de salud, como es el caso de la fenilcetonuria. Además de las aplicaciones que permiten comprender las macromoléculas de importancia en biomedicina, existen compuestos orgánicos como los ácidos tricarboxilicos que son utilizados en el diseño de nanopartículas con aplicaciones médicas y el diseño depende de la incorporación de soluciones de estos ácidos tricarboxílicos y otros disolventes.

Este proyecto de tesis tuvo como objetivo analizar por modelamiento computacional el efecto del medio sobre moléculas y biopolímeros de interés en biomedicina. Se consideró estudiar cinco diferentes escenarios de interés en biomedicina. El primer objetivo fue analizar la resistencia a la desnaturalización de la estruc-

tura macromolecular de la ureasa de *Helicobacter pylori* a pHs ácidos, por lo que se buscó analizar el comportamiento a siete pHs desde 2 hasta 7.5. El segundo objetivo estuvo enfocado en analizar la estabilidad estructural de la proteasa principal de SARS-CoV-2 a pHs bajos y la unión al fármaco PF-00835231 de Pfizer. Por otro lado, el tercer objetivo fue analizar la union de cinco variantes de la Spike-S1 de SARS-CoV-2 a ACE2 teniendo en cuenta cinco pHs (4,5, 6,7 y 8). El cuarto objetivo fue analizar la congregación de monómeros de fenilalanina en temperatura fisiológica. El ultimo objetivo de esta tesis fue estudiar por simulaciones atomísticas el citrato de sodio como un modelo de ácido tricarboxílico sobre diferentes cantidades de los disolventes dietilenglicol y etilenglicol.

La metodología computacional utilizada en este trabajo de tesis, fue usar el procedimiento Semi-Grand-Canonical Monte Carlo para asignar el estado de carga para cada residuo de las macromoleculas propuestas según el pka individual obtenido por programa PROPKA. Posteriomente, se aplicaron simulaciones de dinámica molecular de todos los átomos con solvente explicito para todos los sistemas propuestos y finalmente se calculó la energía libre de union por MM/GBSA para los sistemas de cinco variantes de la Spike-S1 frente al ACE2.

Se adaptó a partir de un código casero un programa informático capaz de predecir los estados de protonación de residuos titulables de proteínas en diferentes condiciones de pH. La nueva adaptación se puede encontrar en el enlace `https://github.com/smadurga/Protein-Protonation`. Este programa fue diseñado considerando el método Semi-Grand Canonical Monte Carlo para diferentes pH.

Las simulaciones computacionales nos permitieron analizar el comportamiento de proteínas a pHs fijos. En principio se tuvo en cuenta los estados de protonación para siete pHs de la estructura supramolecular de la ureasa de *Helicobacter pylori* teniendo en cuenta la carga fija del centro activo y su coordinación con los iones de níquel. Se observó que existe una alta estabilidad estructural de la ureasa a pHs ácidos, debido a la compleja conformación que posee esta enzima. Por otro lado, se observó que los doce sitios activos de la ureasa están cubiertos por regiones móviles que actúan como aletas las cuales aumentan su movilidad a pH ácidos. Sin embargo, las condiciones ácidas extremas hacen que la ureasa tenga el menor número de interacciones estabilizadoras entre sus subunidades proteicas, dando inicio a su desnaturalización.

La proteasa principal de SARS-CoV-2 analizada a pHs demostró que los residuos del sitio activo y los residuos que permiten la dimerización de la proteasa principal no se vieron afectados por los cambios de pH. Sin embargo, los residuos de unión al sustrato de la proteasa principal se alteraron a pH bajos, lo que permitió el aumento del volumen del centro activo. Otro resultado, demostró que la distribución de los solventes alrededor de los residuos catalíticos Cys145 e His41 mostraron una baja y alta afinidad por el agua a pH ácido, respectivamente. Lo que podría ser crucial en el mecanismo catalítico de SARS-CoV-2 Mpro a pH bajos. Además, el análisis de las interacciones la proteasa principal y el fármaco de Pfizer PF-00835231 muestra una excelente afinidad a diferentes pH.

El análisis de las variantes, alfa, beta, delta, kappa, y omicron de la Spike-S1 de SARS-CoV-2 frente a la proteína ACE2 fueron analizadas a pH 4, 5, 6, 7 y 8, resultó en que las mutaciones localizadas en las zonas de unión de la Spike-S1 son capaces de contribuir con una mejor unión al ACE2. Donde, la variante omicron posee mas mutaciones en la zona de union que las otras variantes y esta variante es capaz de unirse fuertemente a pHs ácidos y alcalinos.

Los resultados de las simulaciones de dinámica molecular de los monómeros de fenilalanina a temperatura fisiológica nos ayudaron a comprender el autoensamblaje de la fenilalanina mediada por interacciones no covalentes como puentes de hidrogeno y enlaces π-π entre los anillos aromáticos. Este fuerte apilamiento se organiza para crear un núcleo hidrofílico formado por terminales zwitteriónicos, donde se sugiere que la vía de agregación de la fenilalanina inicia en forma de escalera y a partir de 200 monómeros de fenilalanina se presenta la conformación en zig-zag que posteriormente involucra la formación de una fibra que difícilmente puede disgregarse.

Se demostró que la conformación de citrato en cada uno de los disolventes mantuvo una orientación $gauche^{+}$ o $gauche^{-}$, mientras que la dispersión del Na^{+} alrededor del citrato típicamente ocuparon una orientación cercana a su anión carboxilato central. Paralelamente, se observaron más enlaces de hidrógeno entre el citrato con dietilenglicol puro que con el etilenglicol puro. El citrato tiene un momento dipolar alto cuando se combina con dietilenglicol puro, donde numerosas conexiones de enlaces de hidrógeno entre ambas moléculas causan este momento dipolar significativo. El blindaje del citrato aumenta con el alto número de enlaces de hidrógeno, lo que reduce la constante dieléctrica del disolvente.

En conclusión, podemos ver que el uso de las herramientas computacionales es de gran utilidad para lograr comprender diferentes ambientes biológicos donde se lograron nuevos alcances sobre el comportamiento de moléculas y biopolímeros de importancia en biomedicina. En principio del estudio de la ureasa de *Helicobacter pylori*, se logró demostrar que la ureasa es una estructura macromolecular altamente compleja que posee doce centros activos y que estos presentan un colgajo que se abre a condiciones de pH ácidos permitiendo el acceso del sustrato. En cuanto al estudio de la proteasa principal de SARS-CoV-2 se pudo comprender que es una estructura estable a pH 5 y que es capaz de unirse al fármaco PF-00835231 a las condiciones de pH analizados. El estudio del pH de un sistema interactuante como es el caso de cinco variantes de la Spike-S1 de SARS-CoV-2 y ACE2 nos mostró que la variante omicron logró la mejor energía de unión al ACE2, lo cual podría demostrar el rápido contagio de esta variante. Las conclusiones que resultan del análisis del autoensamblaje de los monómeros de fenilalanina demuestran la influencia de la interacción no covalente como enlaces de hidrógeno y π-π, los cuales influyen en el apilamiento de estos monómeros. Así mismo, las fibras de fenilalanina formadas son capaces de atraer otro tipo de aminoácidos, como es el caso de la alanina. Finalmente, los hallazgos finales de las simulaciones atomísticas sobre el comportamiento del citrato sódico en solventes orgánicos a una temperatura relevante (473.15 K) fue de gran importancia para la creación de nanopartículas de importancia en biomedicina.

Abstract

Computer simulation technologies facilitate solving complex biomedical problems by helping researchers predict what will happen in a natural system in response to various external conditions. Currently, there are few computational simulation analyses on the effects of pH on macromolecular structures. For instance, *Helicobacter pylori*'s urease enables the pathogen to survive in acidic pH environments. Another recent example shows how crucial it is to consider the pH of SARS-CoV-2 treatment targets because some mechanisms only take place at acidic or alkaline pHs. Another problem in biomedicine is the aggregation of aromatic compounds and the formation of toxic fibres that lead to significant health problems, such as phenylketonuria. In addition to the applications that allow understanding macromolecules of importance in biomedicine, there are organic compounds such as tricarboxylic acids that are used in the design of nanoparticles with medical applications, and the design depends on the incorporation of solutions of these tricarboxylic acids and other solvents.

This thesis project aimed to analyze by computational modelling the effect of the medium on molecules and biopolymers of interest in biomedicine. It was considered to study five different scenarios of interest in biomedicine. The first goal was to examine the urease from *Helicobacter pylori*'s resistance to denaturation at acidic pHs; hence it was desired to examine the behaviour at seven different pHs

between 2 and 7.5. The second goal was to examine the structural stability of the SARS-CoV-2 major protease at low pHs and its binding to the Pfizer medication PF-00835231. The third goal was to examine the binding of five SARS-CoV-2 Spike-S1 variants to ACE2 while accounting for five pHs (4, 5, 6, 7, and 8). Analyzing the coaggregation of phenylalanine monomers at physiological temperature was the fourth goal. The final goal of this thesis was to investigate the effects of various concentrations of diethylene glycol and ethylene glycol solvents on sodium citrate, a model for tricarboxylic acid, using atomistic simulations.

The computational methodology used in this thesis work used the Semi-Grand-Canonical Monte Carlo procedure to assign the charge state for each residue of the proposed macromolecules according to the individual pka obtained by the PROPKA program. Then, molecular dynamics simulations of all the atoms with explicit solvents were used for each of the suggested systems. Finally, the binding free energy using MM/GBSA was determined for the systems of the five Spike-S1 versus ACE2 variations.

A computer program capable of predicting the protonation states of titratable protein residues under different pH conditions was adapted from a homemade code. The new adaptation can be found at the link `https://github.com/smadurga/Protein-Protonation`. This program was designed considering the Semi-Grand Canonical Monte Carlo method for different pHs.

Computer simulations allowed us to analyze the behaviour of proteins at fixed pHs. In principle, the protonation states for seven pHs of the supramolecular structure of *Helicobacter pylori* urease were taken into account, considering the active centre's fixed charge and its coordination with nickel ions. It was observed that urease has high structural stability at acidic pHs due to this enzyme's complex conformation. On the other hand, it was observed that the twelve urease active sites are covered by mobile regions acting as flaps, increasing their mobility at acidic pH. However, extreme acid conditions cause urease to have the least number of stabilizing interactions between its protein subunits, starting its denaturation.

The SARS-CoV-2 main protease analyzed at pHs demonstrated that the active site residues and residues allowing dimerization of the significant protease were unaffected by pH changes. However, the substrate-binding residues of the main protease were altered at low pH, allowing for increased volume of the active site. Another result showed that the distribution of solvents around the Cys145 and His41 catalytic residues showed low and high affinity for water at acidic pH, which could be crucial in the catalytic mechanism of SARS-CoV-2 Mpro at low pH. In addition, analysis of the main protease and Pfizer drug interactions PF-00835231 shows excellent affinity at different pHs.

The analysis of the variants, alpha, beta, delta, kappa, and omicron of the Spike-S1 of SARS-CoV-2 against the ACE2 protein was analyzed at pH 4, 5, 6, 7 and 8; it resulted in that each mutation located in the Spike-S1 binding sites are capable of contributing to a better binding to ACE2. The omicron variant has more mutations in the binding zone than the other variants and can bind firmly to acid and alkaline pHs.

The results of the molecular dynamics simulations of phenylalanine monomers at physiological temperature helped us to understand the self-assembly of phenylalanine mediated by non-covalent interactions such as hydrogen bonds and π-π interactions between aromatic rings. This strong stacking is organized to create a hydrophilic nucleus formed by zwitterionic terminals, where it is suggested that the phenylalanine aggregation pathway begins in the form of a ladder and from 200 phenylalanine monomers the zig-zag conformation is presented that later involves the formation of a fiber that can hardly be disintegrated.

It was shown that the citrate conformation in each of the solvents maintained a *gauche*$^+$ or *gauche*$^-$ orientation, whereas the Na$^+$ dispersion around the citrate typically occupied an orientation close to its carboxylate anion central. In parallel, more hydrogen bonds were observed between the citrate with pure diethylene glycol and pure ethylene glycol. Citrate has a high dipole moment when combined with pure diethylene glycol, where numerous hydrogen bond connections between the two molecules cause this significant dipole moment. The citrate shielding increases with the high number of hydrogen bonds, which reduces the dielectric constant of the solvent.

In conclusion, using computational tools is very useful for understanding different biological environments where new insights on the behaviour of molecules and biopolymers of importance in biomedicine were achieved. At the beginning of the study of *Helicobacter pylori* urease, it was possible to demonstrate that urease is a highly complex macromolecular structure with twelve active centres. These present a flap that opens at acidic pH conditions allowing access to the substrate. Regarding the study of the main protease of SARS-CoV-2, it was possible to understand that it is a stable structure at pH 5 and capable of binding to the drug PF-00835231 at the pH conditions analyzed. The study of the pH of an interacting system, such as the case of five variants of Spike-S1 of SARS-CoV-2 and ACE2 showed us that the omicron variant achieved the best binding energy to ACE2, which could demonstrate rapid contagion of this variant. The conclusions resulting from the analysis of the self-assembly of the phenylalanine monomers demonstrate the influence of non-covalent interactions, such as hydrogen bonds and π-π interactions, which influence the stacking of these monomers. Likewise, the phenylalanine fibres formed can attract other types of amino acids, such as alanine. Finally, the final findings of the atomistic simulations on the behaviour of sodium citrate in organic solvents at a relevant temperature (473.15 K) were of great importance for creating nanoparticles in biomedicine.

Chapter 1

Introduction

Biomedicine is a multidisciplinary field that consider chemistry, biology, mathematics, computer science, and others. For some years, advances in biomedicine have been notable, from deciphering the genetic code to designing new treatments for known or little-explored diseases.

Recent advances in computational tools allow exploration in biomedical sciences to predict molecules' behaviour in different biological environments. Unlike experimental methods, computer-aided models are cheaper and highly reproducible. This method's application allows the identification of potential ligands towards receptor macromolecules that are becoming increasingly valuable and accurate. However, to better understand the behaviour of different compounds, it is imperative to consider some conditions of the medium, such as temperature, ion concentrations, pH, among others.

Beyond theories, considerable research has been addressed to study atomistic simulations. The atomistic simulations feature is the individual description of the atoms that make up molecular structures and allows us to study systems of thousands of atoms, such as proteins. There are two main categories of atomistic simulations; the first is molecular dynamics simulation methods, and the second category focuses on Monte Carlo simulations.

In this doctoral thesis, classical molecular dynamics simulation techniques, also called deterministic simulations, were used, which produce dynamic trajectories for a system composed of N particles by integrating Newton's law of motion. The powerful ability to treat thousands of atoms provides us with quantitative and qualitative data of interest in biomedicine against external factors such as temperature, pH and other medium properties regarding the problems and high cost of experimental tests.

The molecular dynamics simulations have some limitations, such as the preparation of the initial conditions of some structures. For instance, consideration of proteins' protonation/deprotonation states must be addressed from the outset. Nowadays, several strategies allow us to consider the constant pH of proteins. In this work, we have used the Semi-Grand Canonical Monte Carlo (SGCMC) methods to define the protonation/deprotonation states from acidic and basic amino acids for a specific pH considering the experimental titration data.

This thesis is divided into five chapters. The first chapter referred to the general introduction of the importance of computational tools to predict the behaviour of molecules of importance in biomedicine. The objectives are shown in the second chapter. The third chapter tells us the computational theory methods used. Chapter 4 presents the results, where the first two are accepted articles, and the other three are completed works ready to be submitted for publication.

The first result shows us the approximations of the protonation/deprotonation states by SGCMC technic for six different pHs in the *Helicobacter pylori* urease enzyme. Urease is a metalloprotein composed of twelve heterodimeric subunits, where each dimer contains the active centre with two Nickel cations (Ni^{+2}). The importance of studying the behaviour of this enzyme at low pHs is of importance in public health because *Helicobacter pylori* infection in its early stage is silent and normally, patients become aware of this problem in a complicated stage, as well the treatment they follow leads to the use of a set of medications. The results were published in the MDPI journal Polymers.

Since this thesis was developed during the season of COVID-19 pandemic. The behaviour of the enzyme SARS-CoV-2 Main protease has been analysed with different protonation/deprotonation states for a particular pH. The pH studies were linked to the fact that this enzyme is usually in the late endosome, where the pH conditions are altered. Likewise, we have observed that a drug explored in the clinical phase shows a feasible action at pH 4. These results were specified in a paper for publication in the MDPI journal Polymers.

Currently, mutations in the SARS-CoV-2 virus turn out spontaneously during replication. Therefore, analyzing the different protonated/deprotonated states from mutations in a specific target of SARS-CoV-2 motivated us to include this study. The SGCMC methodology analyzed six mutations of the SARS-CoV-2 Spike Protein S1 to determine the protonation/deprotonation states for five pHs. Molecular dynamics and MMGBSA analyzed the alpha mutation of SARS-CoV-2 Spike Protein S1 with ACE2 receptor. The preliminary results are shown in Chapter 6.

On the other hand, inborn errors of metabolism (IEM) represent a group of inherited diseases that are receiving tremendous attention in biomedicine. The IEM in phenylketonuria leads to blockage of the phenylalanine metabolic pathway, resulting in a enzymatic dysfunction. The atomistic simulations allowed us to evaluate the phenylalanine accumulation. The results showed that each phenylalanine stacks orderly, forming beta strands, where the main bond between them is usually dominated by hydrogen bonds between their amino and carboxyl groups. As the number of phenylalanines in the medium increases, these unions manage to generate fibres that are difficult to denature and that can attract other compounds, subsequently causing toxicity to the cellular environment. The results are shown in Chapter 7.

In Chapter 8, the analysis of different organic solvents on the behaviour of sodium citrate is shown. This study investigated the effect of water, polyethylene glycol, and ethylene glycol solvent over sodium citrate. The results showed us the importance of the concentration of organic solvents on the structural conformation of sodium citrate, which is used in the design of magnetic nanomaterials such as Fe_3O_4.

Finally, the last chapter shows the most outstanding conclusions of this thesis, which include recommendations and future perspectives.

Chapter 2

Objetives

2.1 General Objetive

To study by computational modelling the medium effect over molecules and biopolymers of interest in biomedicine.

2.2 Specific Objetives

- To design a computational protocol to predict the protonation/deprotonation state.
- To determine the protonation/deprotonation state of macromolecules by Semi-Grand Canonical Monte Carlo (SGCMC) for an specific pH.
- To analyzed by atomistic simulation the biological macromolecules at different pHs.
- To evaluated the the binding free energy by Molecular Mechanics with Generalised Born and Surface Area solvation (MM/GBSA) of receptor-ligand.
- To evaluated the accumulation of phenylalanine as study of amino acid disorders.
- To study the importance of different concentration of organic solvents over sodium citrate structure.

Chapter 3

Computational Methods

The behaviour of macromolecules is particularly interesting in biomedical sciences and can be approached at a multiscale level. Molecular dynamics (MD) has frequently been utilized to examine structural function connections since biomolecular behaviour is complex and difficult to identify using conventional studies. A branch of physics known as statistical mechanics uses probability theory to infer specific hypotheses about the atoms or particles that make up macroscopic physical systems and their reciprocal interactions, as well as the behaviour of those systems' statistically significant number of equivalent constituents.

Three methods were used in this research project, Semi-Grand Canonical Monte Carlo (SGCMC) simulations, Molecular Dynamics (MD) simulations, and Gibb free energy analysis by Molecular Mechanics Generalized Born Surface Area (MM/GBSA) methods. The SGCMC method was used to determine the probability of the partial charge of sensitive residues to be protonated or deprotonated at a given pH. The classical molecular dynamics method was applied in all the chapters of this project in order to know the structural behavior of molecules and macromolecules under different biological conditions. Finally, the MM/GBSA method allowed us to relate the free energy between the interacting systems considering molecular dynamics simulation trajectories. The following are the three methods used in this project:

3.1 Semi-Grand Canonical Monte Carlo (SGCMC)

The non-deterministic Monte Carlo (MC) method is the ancestor of the Semi-Grand Canonical Monte Carlo (SGCMC) technique. Based on various probabilities, this approach calculates a quantifiable macroscopic property by averaging a representative sample of degrees of freedom. MC methods are used to calculate thermodynamic properties without providing information regarding the rate of change or kinetics of molecular processes because they efficiently sample the configuration space of a molecular structure without accounting for a time dependence.

It is important to utilize a Monte Carlo model for systems based on the Grand Canonical ensemble, which is a theory for open systems. The Grand Canonical Monte Carlo (GCMC) simulation differs in that the new configuration corresponds to both the insertion or removal of a particle within the system and a random displacement of the molecule within the system, with different acceptance criteria applied sequentially in each of these cases.

A few attempts combine a GCMC for the ion binding with a Canonical Monte Carlo (CMC) for the other ions in the supporting electrolyte, known as SGCMC simulations. While the Grand Canonical formalism is required to examine chemical events, such as protonation and deprotonation processes occurring when the polyelectrolytes include binding sites, introducing the CMC technique enables the explicit description of the electrolyte solution. Combining the two methods makes it possible to accurately describe the equilibrium between the charged polyelectrolyte and the electrolyte solution.

In this thesis we work with macromolecular structures, therefore an advantage of using the SGMCM method is demonstrated in finding the different conditions of deprotonation and protonation states at different pHs of the amino acids that are part of these macromolecular structures.

Since amino acids, as well as peptides and proteins, incorporate both acidic and basic functional groups, the predominant molecular species present in an aqueous solution will depend on the pH of the solution. The homemade developed code was used to calculate protonation states and was adapted for macromolecular proteins. The new adaptation of the code is available for free on GitHub `https://github.com/smadurga/Protein-Protonation`. In which the free energy associated with the protonation and deprotonation of each titrable group is calculated using Equation (3.1) [1, 2].

$$\Delta G = \pm k_{\mathrm{B}} T (ln(10)(\mathrm{pH} - \mathrm{p}K_{a,i}) \tag{3.1}$$

The positive sign of expression is used for protonation and the negative sign for deprotonation. $K_{\mathrm{a},i}$ is the constant of the acid/base equilibrium corresponding to each protonable i residue of proteins determined by PROPKA program.

3.2 Molecular Dynamics Simulation

Molecular dynamics simulations aim to describe a molecular system of a large number of atoms as a function of time, where a classical approach is used to integrate Newton's equations of motion. In this thesis, we worked with atoms that make up biological structures. The systems were evaluated at the atomistic level, and structural and energetic information was obtained.

3.2.1 Force Field

A force field is a mathematical expression describing the dependence of the energy of a system on the coordinates of its particles. The parameters are typically obtained either from *ab initio* or semi-empirical quantum mechanical calculations or by fitting to experimental data such as neutron, X-ray and electron diffraction, NMR, infrared, Raman and neutron spectroscopy, etc.

There are many force fields available in the literature, a typical expression for a force field may look like this:

$$
\begin{aligned}
U(R) = & \sum_{bonds} k_r (r - r_{eq})^2 \\
& + \sum_{angles} k_\theta (\theta - \theta_{eq})^2 \\
& + \sum_{dihedrals} k_\phi (1 + cos[n\phi - \gamma]) \\
& + \sum_{impropers} k_\omega (\omega - \omega_{eq})^2 \\
& + \sum_{i<j}^{atoms} \epsilon_{ij} \left[\left(\frac{r_m}{r_{ij}} \right)^{12} - 2 \left(\frac{r_m}{r_{ij}} \right)^{6} \right] \\
& + \sum_{i<j}^{atoms} \frac{q_i q_j}{4\pi\epsilon_0 r_(ij)}
\end{aligned}
\tag{3.2}
$$

Where the first four terms refer to intramolecular or local contributions to the total energy (bond stretching, angle bending, dihedrals, and impropers), and the last two terms describe the intermolecular interactions.

3.2.1.1 Intramolecular terms or Bonded Interactions

Intramolecular terms refer to bond stretching terms, angle bending terms, dihedral or torsional terms and improper dihedrals. The bond stretching and angle bending contribute to the potential energy as harmonic oscillators as a function of the bond length and valence angle, respectively. In this approximation, the parameters are needed for each bond and angle: the reference or equilibrium value ($r_{eq}$and θ_{eq}) and the force constant (k_r and k_θ). The torsional energy or dihedral angle is represented by a sum of cosine functions with multiplicities and amplitudes k_ϕ, this torsion angle is the angle of rotation about the covalent bond between the middle two atoms and the potential. Finally, the improper torsion potentials are defined for groups of 4 bonded atoms where the central atom is connected to the 3 peripheral atoms, in which the energy functions include an additional out-of-plane term, usually an improper dihedral, where the potential energy is harmonic as a function of the out-of-plane angle ω.

3.2.1.2 Intermolecular terms or Non-bonded Interactions

The intermolecular term or non-bonded interactions describe non-electrostatic and electrostatic interactions between all pairs of atoms.

The Lennard-Jones model is a pair potential that describes the essential features of interactions between atoms. The Lennard-Jones potential or non-electrostatic interaction is given by the following equation:

$$V(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right] \tag{3.3}$$

Where r is the distance of separation between both particles, ϵ is the well depth and a measure of how strongly the two particles attract each other,σ is the distance at which the intermolecular potential between the two particles is zero, and V is the intermolecular potential between the two atoms or molecules.

The electrostatic interactions are represented by a Coulomb energy, which utilizes atomic charges q_i to represent chemical bonding ranging from covalent to polar covalent and ionic bonding. The electrostatic potential is described with the Coulomb's law:

$$V_{Elec} = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \tag{3.4}$$

where r_{ij} is the distance between the pair of atoms, q_i and q_j are the charges on the atoms i and j,ϵ_0 is the permittivity of vacuum and ϵ_r is the relative permittivity.

3.2.2 Energy minimization

The most standard problem of molecular mechanics applied to proteins is the following: what is the best (optimum) structure of the studied protein? The common assumption is that the minimization of the energy (or sometimes other thermodynamic functions such as free energy) will give the answer. It is believed that such a structure is a good approximation to the native structure of the protein of interest.

Energy minimization is essential to determining the proper molecular arrangement in space since the drawn chemical structures are not energetically favorable. The potential energy of a molecule contains different energy components like stretching, bending, and torsion; hence, when an energy minimization program is run, it will immediately reach a minimum local energy value, and it might stop if the employed program is not exhaustive. In other words, an energy minimization might stop after it finds the first stable conformer that is structurally closest to the starting molecular arrangement. Several energy minimization algorithms are available: Steepest descent (robust but slow), conjugate gradient (aster than steepest descent), Newton–Raphson (Rapid, but computationally expensive), and Dynamic quenching (Freeze minimization).

3.2.2.1 Steepest descent

In thiw work, we apply the steepest descent or saddle-point method that is an extension of Laplace's method for approximating an integral, where one deforms a contour integral in the complex plane to pass near a stationary point (saddle point), in roughly the direction of steepest descent or stationary phase. The saddle-point approximation is used with integrals in the complex plane, whereas Laplace's method is used with real integrals.

3.2.3 Molecular Dynamics in Different Ensembles

A Molecular Dynamics (MD) simulation can be carried within different conditions often referred to as ensembles, a fancy term from statistical mechanics. A thermodynamic ensemble provides a way to derive the thermodynamic properties of a system through the laws of classical and quantum mechanics. The mathematical methods needed to understand how the thermodynamic quantities are obtained will not be discussed in too many details. For now, we simply show the existing ensembles, underlining what type of system they aim to simulate. The main idea is that different ensembles represent systems with different degrees of separation from the surrounding environment, ranging from completely isolated systems (i.e., microcanonical ensemble) to completely open ones (i.e., grand canonical ensemble). The ensemble you choose depends on the specific problem you are trying to solve and the conditions under which you want to simulate the system.

3.2.3.1 NVT ensemble

The canonical ensemble is considered representative of classical systems, described by continuous conjugate canonical variables. The system is in thermal equilibrium with an external bath at temperature T. The other variables that define the macroscopic state in a hydrostatic system are the volume V and the number of particles N. The thermodynamic potential associated with this ensemble is the Helmholtz free energy F(N,V,T).

3.2.3.2 NPT ensemble

In the isothermal-isobaric ensemble, energy can transfer across the boundary, but matter cannot. The volume of the system can change such that the internal pressure of the system matches the pressure exerted on the system by its surroundings. Similar to the canonical ensemble, we can describe the isothermal-isobaric ensemble as a system immersed in a heat bath at a temperature (T), where the heat bath is much larger than the system. No amount of heat put off by the system will significantly increase the temperature of the surroundings.

The isothermal-isobaric ensemble is a statistical mechanical ensemble that maintains a constant total number of particles, and constant temperature (T) and pressure (P), typically abbreviated NPT. This ensemble plays an important role in chemistry since the majority of important chemical reactions are carried out under constant pressure conditions. The isothermal-isobaric ensemble is also important when attempting to describe the Gibbs free energy of a system, which is the maximum amount of work a system can do at constant pressure (P) and temperature (T).

3.3 Molecular Mechanics with Generalised Born and Surface Area solvation (MM/GBSA)

Most molecular-level biological processes include molecule binding and unbinding. Nearly all biological functions revolve around protein-protein interactions. Understanding the biological function of protein-protein interactions requires a thorough understanding of the structure of protein-protein complexes and their binding affinities. The related binding free energy of protein-protein interactions determines the stability of association and the prerequisites for complex formation.

Numerous methods have successfully figured out how the Gibbs free energy change of protein-protein binding changes. The binding free energy is frequently calculated using molecular dynamics and related techniques. Furthermore, because molecular dynamic simulations can reveal atomic-level information on the dynamics of interactions between proteins, ligands, and solvent molecules, structural flexibility and entropic effects are explicitly considered in free energy estimates.

A continuous solvent model and full atomic resolution molecular mechanics methods have recently been used to assess protein-protein complexes. After deleting the explicit water molecules using an implicit solvent model, the trajectory is reanalyzed due to the significant energy fluctuations of the explicit solvent molecules.

A common way for determining the free energy of the binding of biological macromolecules is to combine the molecular mechanic energies with the Poisson-Boltzmann or Generalized Born and surface area continuum solvation (MM/PBSA and MM/GBSA) methodologies.

The pairwise electrostatic Coulomb interactions and electrostatic (polar) solvation contributions calculated using the finite-difference Poisson-Boltzmann (PB) or Generalized Born (GB) equations make up the mean interaction energy. In order to create a cavity in a solvent where the solute resizes, and solute-solvent dispersion interactions occur, the nonpolar component of the solvation free energy is connected with attractive short-range dispersion interactions. A hydrophobic term, a modification of the van der Waals interaction between proteins and solvent, and a cavity that depends on surface area can also be used to split the nonpolar component further. A solvent grid representation of the complex or a surface integral technique can be used to calculate the latter contribution.

In MM/PBSA or MM/GBSA, the binding free energy (ΔG_{bind}) between a ligand (L) and a receptor (R) to form a complex RL is calculated as:

$$\Delta G_{bind} = \Delta H - T\Delta S \simeq \Delta E_{MM} + \Delta G_{sol} - T\Delta S \tag{3.5}$$

$$\Delta E_{MM} = \Delta E_{internal} + \Delta E_{electrostatic} + \Delta E_{vdw} \tag{3.6}$$

$$\Delta G_{solv} = \Delta G_{PB/GB} + \Delta G_{SA} \tag{3.7}$$

where ΔE_{MM}, ΔG_{sol} and -$T\Delta S$ are the changes of the gas phase MM energy, the solvation free energy, and the conformational entropy upon binding, respectively. ΔE_{MM} includes $\Delta E_{internal}$ (bond, angle, and dihedral energies), $\Delta E_{electrostatic}$ (electrostatic), and ΔE_{vdw} (van der Waals) energies. ΔG_{solv} is the sum of electrostatic solvation energy (polar contribution), $\Delta G_{PB/GB}$, and the nonelectrostatic solvation component (nonpolar contribution), ΔG_{SA}. The polar contribution is calculated using either the GB or PB model, while the nonpolar energy is estimated by solvent accessible surface area (SASA). The conformational entropy change -$T\Delta S$ is usually computed by normal-mode analysis on a set of conformational snapshots taken from MD simulations.

Chapter 4

Effect of pH on the Supramolecular Structure of Helicobacter pylori Urease by Molecular Dynamics Simulations

The effect of pH on the supramolecular structure of *Helicobacter pylori* urease was studied by means of molecular dynamics simulations at seven different pHs. Appropriate urease charge distributions were calculated using a semi-grand canonical Monte Carlo (SGCMC) procedure that assigns each residue's charge state depending on the assigned individual pK_a obtained by PROPKA. The effect of pH on protein stability has been analyzed through root-mean-square deviation (RMSD), radius of gyration (RG), solvent-accessible surface area (SASA), hydrogen bonds (HB) and salt bridges (SB). Urease catalyses the hydrolysis of urea in 12 active sites that are covered by mobile regions that act like flaps. The mobility of these flaps is increased at acidic pHs. However, extreme acidic conditions cause urease

to have the least number of stabilizing interactions. This initiates the process of denaturalization, wherein the four $(\alpha\beta)_3$ subunits of the global structure $((\alpha\beta)_3)_4$ of urease start to separate.

---

[1] The article derived from this chapter has the title *"Effect of pH on the Supramolecular Structure of Helicobacter pylori Urease by Molecular Dynamics Simulations"* and was published in the journal Polymers of the MDPI[3].

4.1 Introduction

Helicobacter pylori (HP) is a Gram-negative helical bacterium, infects about half of the world's population, inducing gastritis, peptic ulcers and increased risk of gastric cancer [4, 5]. Interestingly, HP is highly adapted to the medium found in the stomach. First, HP stays in the lumen with a pH of about 2.0, and then colonizes the mucus layer which has a pH that ranges from 4.5 to 6.5 [6]. HP is a spiral-shaped Gram-negative bacterium, microaerophilic and highly motile due to the presence of 4–8 flagella [7]. Cheng-Yen Kao et al. described four fundamental steps for colonization, which consist of survival, motility, adhesion and toxicity [8]. One of the key characteristics that makes HP to survive at acidic conditions is the generation of large quantities of the urease enzyme.

Urease is an enzyme that catalyses the hydrolysis of urea to ammonia and carbamate. Carbamate further decomposes to ammonia and carbonic acid. The protonation of ammonia and dissociation of carbonic acid causes an increase in the pH, allowing HP to survive in the pH conditions of the stomach. Thus, this important role in the survival of HP bacteria makes urease an excellent target for inhibitors.

Some studies have been performed to elucidate the mechanism of action of the enzyme and to identify suitable inhibitors for HP urease using quantum mechanical (QM), molecular dynamics (MD) and docking methodologies. QM/MM methodology was applied to study the mechanism of inhibition of hydroxamic acids acting on *Helicobacter pylori* urease [9]. Mechanistic pathways were also explored in combination with MD simulations based on hydrogen-bonding network of the catalytic center [10]. Molecular dynamics (MD) studies reveal the mobility of regions called flaps that open and close the access to the active center of the urease [11, 12].

Docking calculations have been performed to identify potential inhibitors of urease that could interact with the active center [13] or could lock the flexible loop in open conformation [14].

The first crystallized urease was jack bean urease [15]. It was observed that urease has a complex structure requiring nickel ions for its activity. Nam Chul-ha et al. analyzed the supramolecular assembly of HP urease, identifying two functional subunits named α and β of 26.5 and 61.7 kDa, respectively [5]. The α and β subunits constitute a heterodimeric unit. Then, three heterodimeric units of $\alpha\beta$ ($(\alpha\beta)_3$) form the tetrahedral complex $((\alpha\beta)_3)_4$ which turns out to be the structure of urease [5]. Additionally, the macromolecular complex of urease presents an internal hollow with a diameter of 1.3 nm. HP urease has 12 active sites, each being coordinated with two Ni^{2+} ions. It is one of the most important regions of the enzyme that could protect it from the acidic environment by urea hydrolysis. The active site is covered by a protein region, denominated ”flap,” having high flexibility that could change from closed to open state in order to promote the access of urea at the dinickel center [12, 14].

For a long time, the pH dependence in structural protein stability has been considered. Decreased stability at more acidic pH is explained by the increase of unfavourable electrostatic interactions due to the high increase of distributed positive charges, and the contrary occurs with negative charges at a highly basic pHs [16]. Therefore, the pKa values of ionizable groups in proteins play essential roles in their stability and functions [17].

The optimum pH of urease is known to be approximately 7.5, and below a pH of 4.5 the enzyme has been reported to be inactive [18]. However, the effect of pH on the mechanism of urease is still not known. For this reason, in this work we determined the different protonation states at seven pHs by semi-grand canonical

Monte Carlo (SGCMC) [1, 2, 19, 20]and the behaviour of each system through molecular dynamic simulations.

4.2 Computational Methods

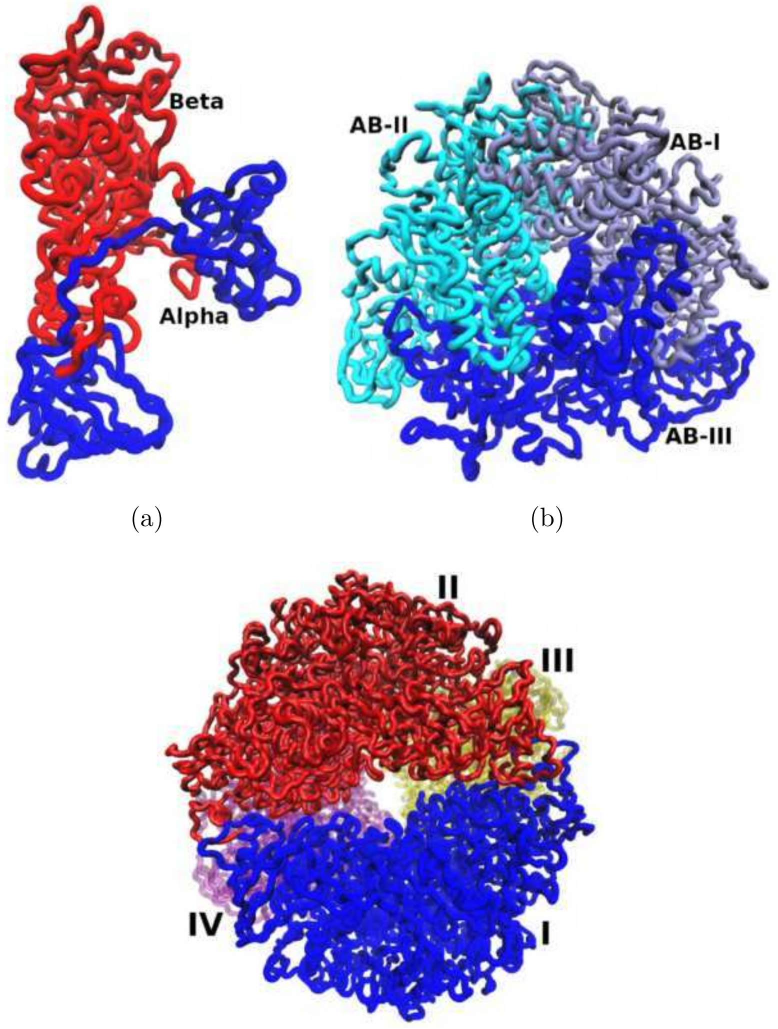
4.2.1 Multimeric protein preparation

Urease is a protein with a dodecameric $((\alpha\beta)_3)_4$ architecture with spherical assembly. The fundamental unit is a heterodimer with two chains labeled α and β. The crystal structure obtained from the Protein Data Bank (PDB) with the access code 1E9Z was used as initial structure. Urease has 12 active sites located on the β chains. The residues of this chain that participate in the active site are a set of four histidines (His136, His138, His248, and His274), a carbamylated lysine (KCX219), an aspartic acid (Asp362), two nickel ions (Ni^{2+}) and a hydroxide ion. The non-natural carbamylated lysine was parameterized using the atom types and parameters corresponding to the similar fragments of lysine and aspartic acid.

HP urease has an interesting structural arrangement with different levels of chain assembly. Figure 4.1 shows three different levels of organization of urease protein. The fundamental unit is a heterodimer $(\alpha\beta)$ which interacts with two other heterodimers of the $(\alpha\beta)_3$ assembly, and finally, this assembly interacts with three other $(\alpha\beta)_3$ assemblies, completing the full structure $((\alpha\beta)_3)_4$ [5].

4.2.2 Protonation states for different pHs

In order to determine the charge state of each titrable residue of urease, and to determine how it changes with the simulated pH, it is required to have the pK_a values of all titrable residues of the protein. The urease structure was evaluated



(a) (b)

(c)

Figure 4.1: Structural arrangement of HP urease at pH 7.5. a.) Heterodimer ($\alpha\beta$). b.) Arrangement of one $(\alpha\beta)_3$) subunit. c.) Entire urease structure where each $(\alpha\beta)_3$) is shown in a different color and labeled with I, II, III and IV.

using PROPKA [21], which predicts the pK_a values of ionizable groups based on the 3D structure. PROPKA uses the terms indicated in Equation (6.1) to estimate the pK_a value of each titrable residue:

$$\mathrm{p}K_\mathrm{a} = \mathrm{p}K_\mathrm{model} + \Delta\mathrm{p}K_\mathrm{DS} + \Delta\mathrm{p}K_\mathrm{HB} + \Delta\mathrm{p}K_\mathrm{CC} \tag{4.1}$$

where $\mathrm{p}K_{\mathrm{model}}$ is the $\mathrm{p}K_{\mathrm{a}}$ of the ionizable groups in an ideal system; $\mathrm{p}K_{\mathrm{DS}}$ takes into account the penalty of desolvation; $\mathrm{p}K_{\mathrm{HB}}$ considers the effect of hydrogen bonds of the side chain and backbone; and $\mathrm{p}K_{\mathrm{CC}}$ takes into account the effect of the interaction of the charge groups.

We have designed a homemade *c* program to calculate the different protonation states of Asp, Glu, Arg, Lys and His residues, and the C- and N- termini for seven pHs (2, 3, 4, 5, 6, 7 and 7.5) using the semi-grand canonical Monte Carlo (SGCMC) ensemble [1, 2, 19, 20, 22]. A characteristic feature of semi-grand canonical Monte Carlo (SGCMC) simulations is that the total charge is not constant but it fluctuates as a natural result of the dynamic proton binding/unbinding. The required inputs of the program are the pH, number of trial protonation/deprotonation steps and the list of acid/base constants in the format given by the summary of Propka. Initially, all protonable residues are assigned the protonated stated. The system arrives quickly to the equilibrium, as the $\mathrm{p}K_{\mathrm{a}}$ values are considered to be independent, being unaffected by the protonated state of neighbour residues. The histidines that coordinate the Ni2+ ions are fixed to be in the appropriate deprotonated state to correctly coordinate the metal ions. The program calculates the total charge, the distribution of protonated/desprotonated forms for each residue and generates a output file with the format that the *pdb2gmx* command of Gromacs requires for the interactive selection of charged states.

The free energy associated with the protonation/deprotonation of each titrable group is calculated using Equation (6.2) [1, 2].

$$\Delta G = \pm k_{\mathrm{B}} T (ln(10)(\mathrm{pH} - \mathrm{p}K_{a,i}) \tag{4.2}$$

The positive sign of expression of Equation (6.2) is used for protonation and the negative sign for deprotonation. $K_{\mathrm{a},i}$ is the constant of the acid/base equilibrium corresponding to each protonable i residue of urease determined by PROPKA. A total of 30,000 steps for protonation and deprotonation trials for residues of urease were performed with the SGCMC simulation program at each pH. As the protonation states are obtained from pKa values calculated with PROPKA using the crystal structure of urease, these protonation states will correspond correctly to the assigned pHs if the global structure of the urease does not change significantly. Furthermore, the effect of the explicit Na or Cl ions included in the simulations to neutralize the systems is not taken into account in the selection of the protonated or deprotonated states of the titratable residues.

4.2.3 Molecular dynamic simulations

All molecular dynamics simulations were performed with Gromacs (Groningen Machine for Chemical Simulations) version 2019.2 software [23, 24] using the OPLS-AA force field. Seven different initial structures of urease were prepared while attending to the different charge states of the titrable amino acids obtained from the SGCMC calculations corresponding to each simulated pH. In each case, the urease structure was located in the center of a box with dimensions about 18 nm^3 having a minimum separation of 1.5 nm to the box edges. Distances of the Ni^{2+} with the coordinated residues of the active site were restrained at the values of the crystal structure because there are not available bonding parameters for nickel metal ions in the used force field. The protein was solvated with TIP4P [25] waters and the system was neutralized with the addition of the appropriate quantity of Na$^+$ or Cl$^-$ (Table A.1 from Appendix A). It is worth knowing that the salt concentration of ions required to neutralize the system is higher than the normal physiological salt concentrations (0.154 M) for the more acidic pHs, being, for example, 0.4 and 0.35 M for pHs 3 and 4, respectively. However, although the survival of *Helicobacter pylori* has been reported to be optimal at physiological salt concentrations, it survives also at high concentrations, as the simulated in this study [26].

Initially, the systems were relaxed with an energy minimization using the steepest descent algorithm with 20,000 steps while applying 0.001 nm as a maximum step size. Periodic boundary conditions were applied in all directions. Short-range interactions of van der Waals (vdW) and electrostatic interactions were considered with a cut-off of 0.9 nm. Long range electrostatic interactions were considered with the particle-mesh Ewald (PME) [27] method.

Preparation of the systems was performed with MD in a canonical ensemble with V-rescale [28] thermostat (309.65 K) during 10 ns. Finally, an isobaric-isothermal ensemble with V-rescale thermostat (309.65 K) and Parrinello-Rahman [29] barostat (1.0 bar) were performed for a simulation length of 100 ns. Molecular visualization of the different systems was performed using Visual Molecular Dynamics version 1.9.4. [30] and Chimera UCSF version 1.12.0 [31]. All plot analysis were performed with Gnuplot version 5.3 software.

Simulations corresponding to pHs 2 and 7.5 were prolonged for 100 ns in order to check that the simulation length was appropriate for obtaining converged results. In Table A.2 from Appendix A shows the statistics of the poor and favoured rotamers and of the Ramachandran outliers and favoured angles after the NVT and NPT simulations. It can be seen that for all pH, the 10 ns NVT simulations improved the initial quality of the structure. Ramachandran outliers were reduced significantly, and residues in Ramachandran favoured regions achieved the stability value. Subsequently, the 100 ns NPT simulations were able to obtain a greater number of rotamers in favoured regions. The results obtained from the 100 ns NPT continuation simulations indicate that longer simulations are not required to obtain equilibrated rotamers and Ramachandran angles.

4.3 Results

4.3.1 Urease charge as a function of pH

The appropriate distribution of charged residues over the proteins will confer structural stability. This distribution is a function of the pH of the medium. In Figure 4.2, a monotonic decrease of the positive total charge of the urease from pH 2 until pH 7 can be seen. After this pH, the urease changes the total charge from a positive to a negative sign, indicating this point, 7.0, as the isoelectric point (IP) of urease. Experimental IPs of urease range from 5.4 [32] to 6.1 [33]. Differences with calculated IP could be attributed to precision of calculated intrinsic constants of titrable residues and/or the effects of condensed counterions in the internal cavity of the urease protein that could distort the deduced protein charge obtained by usual techniques.

In Table 5.2 the charge as a function of pH obtained from SGCMC simulations for the Asp, Glu, Arg, Lys and His titrable amino acids and for the entire structure of the urease can be seen. It can be seen that the greatest contributions to the variation of the charge are because of the Asp and Glu acid residues. From pH 2 to pH 7.5, the charges change from -7 and -33 to -575 and -653 for Asp and Glu, respectively. In contrast, the effect of the variation of the charge of Arg and Lys basic residues is very reduced as the average pK_a of these residues (Table 4.2) is located outside of the studied pH range (average pK_a of 12 and 10 for Arg and Lys,

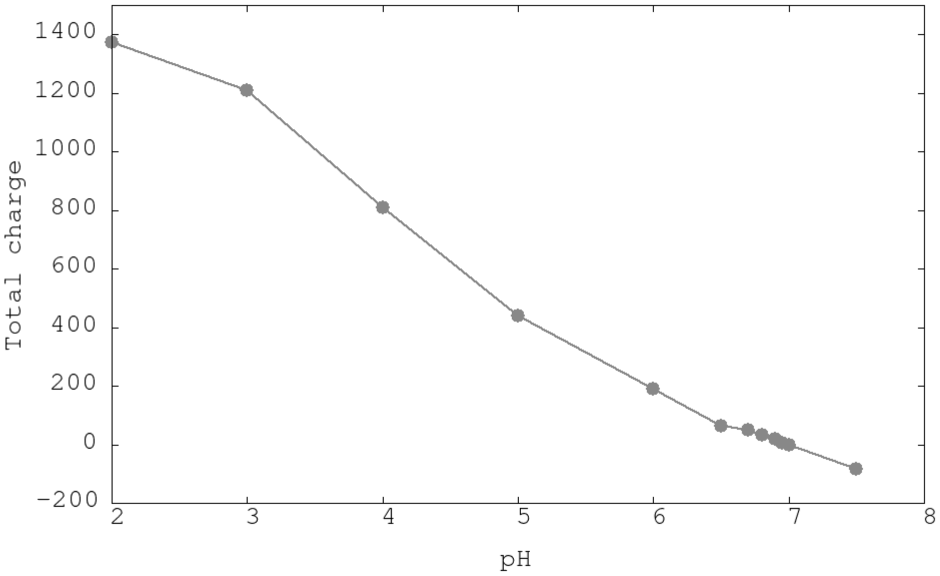


Figure 4.2: Total urease charge obtained from SGCMC simulations of titrable amino acids at different pHs.

Table 4.1: Charge as a function of pH for titrable amino acids and for the entire structure of the urease.

System	**ASP**	**GLU**	**ARG**	**LYS**	**HIS**	**Urease**
pH2	−7	−33	360	792	237	1373
pH3	−73	e99	360	792	204	1208
pH4	−246	−281	360	791	162	810
pH5	−415	−464	360	789	144	438
pH6	−505	−566	360	778	98	189
pH7	−554	−627	360	757	37	−3
PH7.5	−575	−653	359	744	18	−83

respectively). It is worth noting that the His residues show the wider distribution of pK_a (4 ± 3). This could be the result of a more variable environment for residues of this kind.

Table 4.2: Number an average pK_a values of titrable amino acids of urease.

AA	**Number**	**pK_a**
ASP	624	4.7 ± 1.6
GLU	696	4.6 ± 1.6
ARG	360	12.3 ± 1.5
LYS	792	10.0 ± 1.2

4.3.2 Structural changes at different pHs: RMSD, RG and SASA.

The temporal evolution of the root-mean-square deviation (RMSD) of the backbone atoms of the entire structure of urease is shown in Figure 4.3a. Two patterns can be distinguished, depending on the simulated pH conditions. Simulations from pHs of 4 to 7.5 show a plateau in the RMSD starting approximately around 40–50 ns. However, simulations at pHs 2 and 3 show a linear increase of RMSD with time after 20 ns. The fact that the RMSD profile does not arrive at a constant value with more time indicates that the urease structure is not stable at the two more acidic pHs. To better quantify this effect, the average RMSD values through the last 20 ns were computed (Table 4.3). The average RMSD values show a differentiated behavior for the two more acidic pHs (about 0.8 nm) with respect to the other cases (0.32–0.45 nm). It has to be noted that pH 4 and pH 5 have greater RMSD values than the more basic pHs, indicating more structural flexibility of the protein for these intermediate pHs with respect to the neutral cases.

The effect of structural distortion could be analyzed for the different simulated pHs by computing the radius of gyration (RG) Figure 4.3b. As in the analysis of global RMSD, two patterns could be distinguished depending on the simulated pH conditions. pHs 2 and 3 showed a linear increase of RG, whereas pHs 4 to 7.5 show a plateau. Thus, the two more acidic conditions presented a structure

Table 4.3: Root-mean-square deviation (RMSD), radius of gyration (RG), solvent-accessible surface area (SASA), hydrogen bonds (HB) and salt bridges (SB) averaged over the last 20 ns or in the last frame of MD simulations at different pHs.

pH	**RMSD(nm)**[a]	**RG(nm)**[b]	**int-RG**[c]	**ext-RG**[d]	**SASA(nm^2)**[e]	**HB**[f]	**SB**[g]	**Exp.Activity**[h]
2	0.81±0.02	6.56±0.01	2.69±0.02	8.34±0.02	3589±13	5853±40	96	0
3	0.83±0.05	6.60±0.03	2.79±0.04	8.40±0.04	3482±28	6044±56	131	0
4	0.45±0.01	6.29±0.01	2.46±0.01	8.10±0.01	3182±11	6518±41	183	0
5	0.38±0.01	6.24±0.01	2.41±0.01	8.07±0.01	3052±11	6973±42	179	0.38
6	0.33±0.01	6.20±0.01	2.40±0.01	8.02±0.01	2893±11	7287±44	182	1.75
7	0.35±0.01	6.22±0.01	2.42±0.01	8.03±0.01	2852±9	7549±47	169	4.5
7.5	0.32±0.01	6.18±0.01	2.35±0.01	8.01±0.01	2828±11	7460±41	179	5.5

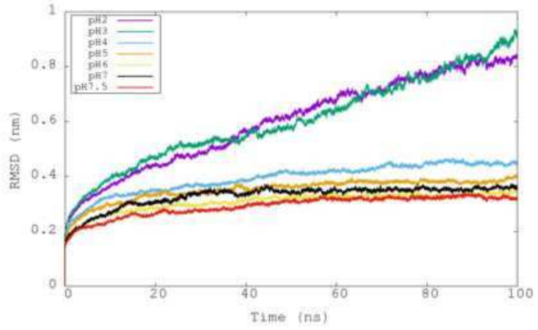
[a] Root-mean-square deviation of the backbone atoms of the entire structure of urease,

[b] radius of gyration of the entire urease, [c] radius of gyration of internal Glu505,

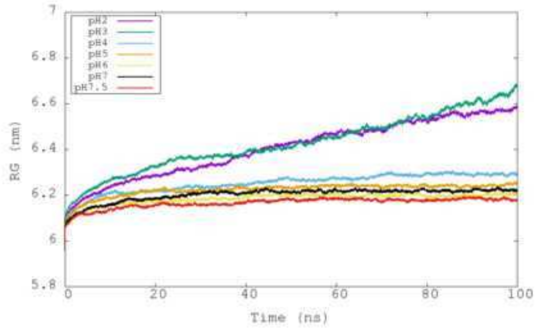
[d] radius of gyration of external Glu101, Glu116 and Glu386, [e] solvent-accessible surface area of the entire urease,

[f] total number of intramolecular hydrogen bonds of urease, [g] Intramolecular urease salt bridges in the last snapshot of the simulations, [h] experimental urease activity in μmols/min/mg.

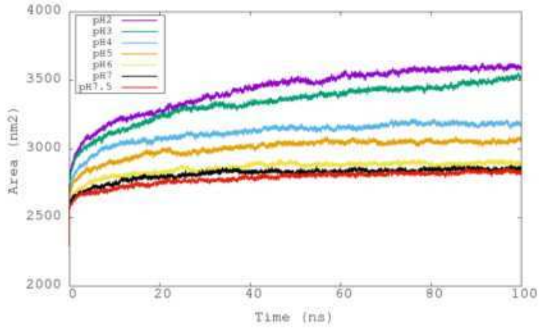
that is increasing in size during time. The other conditions showed again a stable structure which is structurally more compact. In the last 20 ns, the RG values of the two more acidic conditions were about 0.4 nm greater that those corresponding to the other pH conditions (Table 4.3). Thus, the analysis of the radius of gyration of the entire structure of the ureases allows us to identify the increase of the size at acid conditions. In order to distinguish if this growth is performed on the external surface, and if the inner part of urease is more separated or more compact, two new calculations of radius of gyration were performed while selecting only the internal Glu505 (to analyze the size variation of the internal cavity) and the external Glu101, Glu116 and Glu181 (to analyze the size variation of the external surface). The obtained values indicate that both internal and external RG increase at acidic conditions in the same range (Table 4.3). Differences between pHs 2 and 3 seem to be attributable to statistical fluctuations. Thus, these values indicate that urease is increasing in size at the same time that the dimensions of its internal cavity increase.



(a)

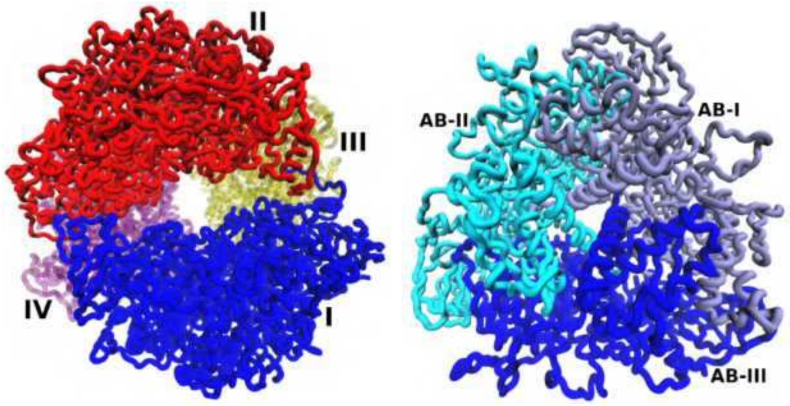


(b)



(c)

Figure 4.3: Representation of the temporal evolution of structural properties of urease in the molecular dynamics simulations at pH 2, 3, 4, 5, 6, 7 and 7.5. a.) Root-mean-square deviation (RMSD) for the total *Helicobacter pylori* urease. b.) Radius of gyration (RG). c.) Solvent accesible surface area (SASA).



(a) (b)

Figure 4.4: Representation of urease at pH 2. a.) $((\alpha\beta)_3)_4$ multimeric structure. Each $(\alpha\beta)_3)$ is shown in a different color and indicated with I, II, III and IV. The separation of cluster I and cluster II increases at pH 2. b.) Similar structural arrangement for the $(\alpha\beta)_3$ subunit.

The variation of the solvent-accessible surface area (SASA) with time for the different pHs is shown in Figure 4.3c. The greatest SASA values were obtained for conditions of pH 2 and 3. This behaviour is related to the observed increased of RG for these pHs. The more compact structures in terms of surface were obtained for the more basic pHs. The increase of size at acidic pHs can be also noticed in Figure 4.4a where an increase in the separation between I (blue) and II (red) $(\alpha\beta)_3$ subunits in the final frame of the simulation at pH 2 can be seen. This disaggregation phenomenon is in correspondence with mass spectrometry studies that observe that the dodecamer disassembles readily into $(\alpha\beta)_3$ subunits in solution and under controlled collision-induced dissociation in the gas phase [34].

4.3.3 Analysis of stability at different pHs: HB and SB

The number of the total intramolecular hydrogen bonds is a measure of the internal stability of the protein. Gromacs tools were used to analyze the hydrogen bonds

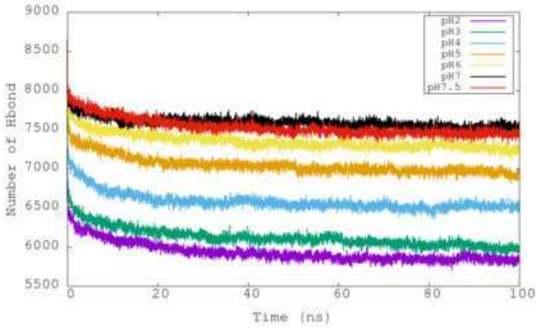


Figure 4.5: Representation of the temporal evolution of the number of total intramolecular hydrogen bonds of urease in molecular dynamics simulations at pH 2, 3, 4, 5, 6, 7 and 7.5.

considering all possible donor and acceptor atoms with a geometric criterion of a distance of less than 3.5 Åand an angle of $180^\circ \pm 30^\circ$.

In Figure 4.5, the variations of the intramolecular hydrogen bonds of urease with respect to time at different pHs can be seen. In contrast to the observed pattern for RMSD and RG properties, the total number of hydrogen bonds shows a more gradual variation among the simulated pHs. Simulations at pH 7 and 7.5 presented the largest numbers of intramolecular hydrogen bonds (around 7500), and then they diminished progressively with lower pHs to arrive at about 5900 for pH 2 (Table 4.3). The reduced number of hydrogen bonds at more acidic conditions could explain the reduced structural stability and the increase in size of the urease with time.

An important feature of urease is its tetrahedral arrangement of four subunits of $(\alpha\beta)_3$. The interactions of $(\alpha\beta)$ dimers were analyzed at different pHs by means of hydrogen bonds. Each subunit $\alpha\beta$ establishes hydrogen bond interactions with two neighbour subunits $\alpha\beta$ (of the same $(\alpha\beta)_3$), and to a lesser extent, with the other $\alpha\beta$ subunit of another $(\alpha\beta)_3$ arrangement. All the hydrogen bond interactions established between atoms of different subunits are referred as interdimer hydrogen

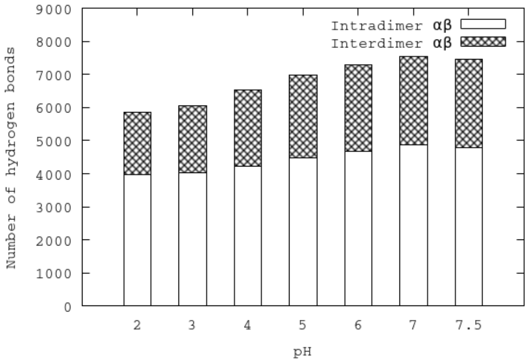


Figure 4.6: Numbers of average inter- and intradimer hydrogen bonds in urease at different pHs.

bonds. In contrast, all the hydrogen bonds established between atoms of the same dimer are referred as intradimer hydrogen bonds. In Figure 4.6, the average of the last 50 ns of the inter- and intradimer $\alpha\beta$ hydrogen bonds for each pH is shown. It can be seen that the numbers of both inter- and intradimer $\alpha\beta$ hydrogen bonds diminish at acidic pHs. The separation of $(\alpha\beta)_3$ subunits at acidic pHs could be attributed to the reduction of interdimer hydrogen bonds. On the other hand, the reduction of the number of intradimer hydrogen bonds could make the individual subunits more flexible and mobile.

Another property that provides information about the structural stability of urease is the number of established intramolecular salt bridges (SB). SB were calculated using the salt bridges plugin of VMD based on a cut-off distance of 3.2 Åbetween any oxygen atom of an acidic residue and nitrogen atom of a basic residue. It can be seen in Table 4.3 that the numbers of SB corresponding to the last frame of pHs 2 and 3 (96 and 131, respectively) are significantly lower than in the other pH conditions.

It is worth noting that the more acidic conditions that are shown to be less structurally stable correspond to the conditions where there is not experimental activity of the urease (Table 4.3). At pH 7.5, it is found to have the greatest experimental activity, and the simulation at this pH shows the lowest values of SASA and RG. It seems that the compactness of the protein is related to high enzymatic activities.

4.3.4 Mobile flap at different pHs

Urease has 12 binding pockets with two Ni^{+2} each. In the crystal structure of urease, it can be seen that these binding pockets are hidden by fragments composed of residues 304–347 of the beta monomers. These regions, named flaps, have been reported to be flexible and mobile, allowing the entrance of substrates to the binding pockets [12]. The average RMSD of the backbone atoms of each flap is represented as a function of the simulated pH in Figure 7.5. The temporal evolution of RMSD of the 12 flaps of urease enzyme at different pHs can be also seen in Figure A.1 from Appendix A. It can be seen that highest RMSD values were obtained for the most acid condition, indicating the greatest distortion with respect to the close state of the flaps. The lowest RMSD values are for those flaps simulated at pHs from 6 to 7.5, whereas the more acidic conditions correspond to states having more flaps in a more open state. Mean RMSD values and their corresponding dispersion are greater for the more acidic pHs, indicating a less stable urease structure. A similar pattern found for flap RMSD was found for the calculated RMSD values of dimers (Figure A.2 from Appendix A). pHs 2 and 3 have greater RMSD values and a greater dispersion among dimers, whereas RMSD values of dimers of pHs 4 to 7.5 have similar and more stables values.

In addition, the process of the opening of the flap could be analyzed by means of the distance of residue Cys321 to one of the Ni^{+2} ions (Figure 5.6). As Cys321 is a residue located in the central part of the flap, an increase of the distance with respect to the nickel ion of the binding pocket implies an opening of the flap. It can be seen that the fluctuation of these distances among the flaps has the general tendency to increase with the reduction of the pH. The greatest difference was observed for the more acidic condition having two flaps with the greatest average distances among all simulations. In Figure 4.9a we show the Cys321–Ni^{+2} distance corresponding to one of the flaps at pH 2 (3.1 nm) being significantly greater than corresponding to the flap at pH 7.5 (0.6 nm) in closed state (Figure 4.9b). This important fluctuation in conjunction with the results obtained from RMSD of flap residues indicates that the pH is an important factor that modulates the mobility of the flaps. It can be seen that the highest mobilities are obtained for the more acidic pHs.

The activity of urease could be related to the ability of reactants to access each of the 12 active centers. This access seems to be regulated by the region that acts as a flap. The effect of reduction of pH is to increase the mobility of this flap. Thus, a reduction of pH is expected to increment the protein's activity. However, at extremely acidic pHs, the most important effect is the significant reduction of the stabilizing interactions (intermolecular and intramolecular non-bonded interactions). This important reduction of stabilizing interactions in combination with the absence of stabilization of the RMSD of the entire urease seems to indicate a denaturation process at very high pHs. However, at moderate acidic pHs, the increase of mobility of these flaps could facilitate urea's access to the active site, increasing the pH in the environment of urease proteins in order to avoid their denaturation.

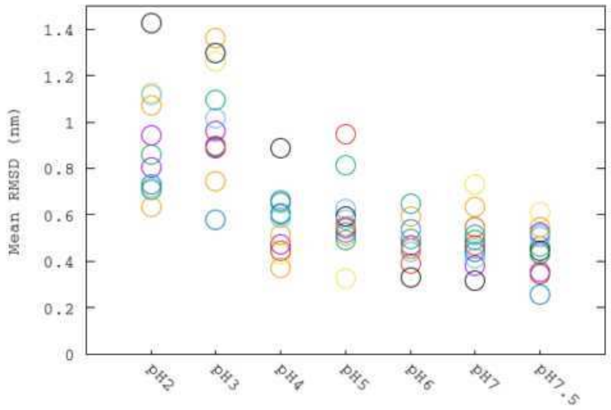


Figure 4.7: Average root-mean-square deviation (RMSD) values of backbone atoms of each flap at different pH simulations. Each flap is aligned with respect to the backbone atoms of its corresponding dimer. Each flap is displayed with a different color.

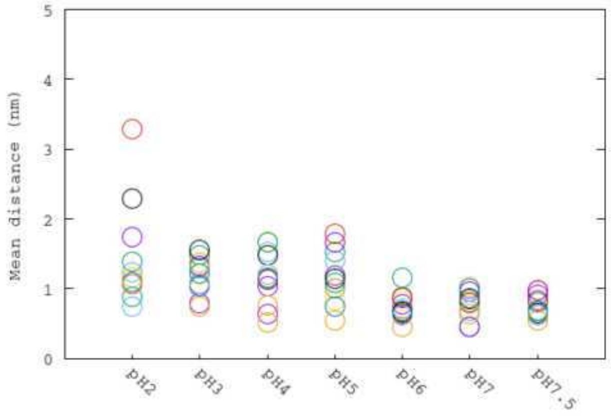
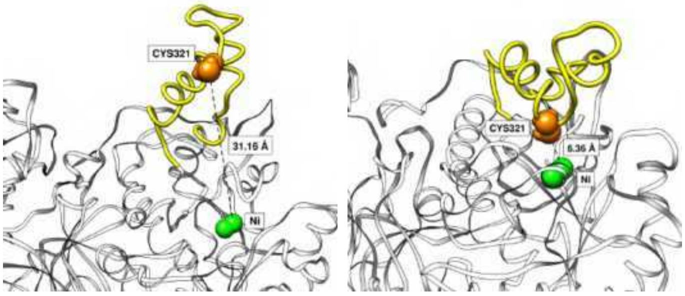


Figure 4.8: Average nickel–Cys321 distances of each subunit for all pH simulations. Same color code as Figure 7.5.

4.4 Conclusion

The effect of pH on the supramolecular structure of *Helicobacter pylori* urease has been studied with a set of molecular dynamics simulations. The different protonation states of titrable residues of urease have been determined by semi-grand canonical Monte Carlo (SGCMC) at different pHs.

The SGCMC procedure used to obtain the urease charge distribution seems



(a) (b)

Figure 4.9: Final snapshots of one urease flap (yellow) in the simulations at a.) pH 2 in open conformation and b.) pH 7.5 in closed conformation. Ni^{+2} ions are represented in green. Distances of Ni^{+2} to Cys321 (orange) are also indicated.

to be an efficient method to generate different charge states to simulate proteins under a wide range of pH conditions [1, 2, 19, 20]. This procedure assigns charges to monomers of the protein that are not identical but are statistically compatible with the simulated pH.

Molecular dynamics simulations performed until 100 ns allow one to understand the activity of urease in a wide range of pHs. At more acidic conditions represented with simulations at pH 2 and 3, the subunits $(\alpha\beta)_3$ start to separate. This would lead to the disaggregation of the protein at these extreme pH conditions. Protein stability could be also analyzed by means of inter- and intra-dimer hydrogen bonds and salt bridges. Extreme acidic conditions cause urease to have the least number of these stabilizing interactions. This behaviour correlates well with the experimental activity of the urease observed only in acidic media greater than pH 5 [35, 18].

The measured RMSD, RG and SASA properties and HB and SB analysis seem to be appropriate indicators to study protein stability at different pHs through molecular dynamics simulations.

Chapter 5

Unveiling the Effect of Low pH on the SARS-CoV-2 Main Protease by Molecular Dynamics Simulations

Main Protease (Mpro) is an attractive therapeutic target that acts in the replication and transcription of the SARS-CoV-2 coronavirus. Mpro is rich in residues exposed to protonation/deprotonation changes which could affect its enzymatic function. This work aimed to explore the effect of the protonation/deprotonation states of Mpro at different pHs using computational techniques. The different distribution charges were obtained in all the evaluated pHs by the Semi-Grand Canonical Monte Carlo (SGCMC) method. A set of Molecular Dynamics (MD) simulations was performed to consider the different protonation/deprotonation during 250 ns, verifying the structural stability of Mpro at different pHs. The present findings demonstrate that active site residues and residues that allow Mpro dimerisation was not affected by pH changes. However, Mpro substrate-binding residues were altered at low pHs, allowing the increased pocket volume. Additionally, the results of the solvent distribution around Sγ, Hγ, Nδ1 and Hδ1 atoms of the catalytic residues Cys145 and His41 showed a low and high-water affinity at acidic pH, respectively. It which could be crucial in the catalytic mechanism of SARS-CoV-2 Mpro at low pHs. Moreover, we analysed the docking interactions of PF-00835231 from Pfizer in the preclinical phase, which shows excellent affinity with the Mpro at different pHs. Overall, these findings indicate that SARS-CoV-2 Mpro is highly stable at acidic pH conditions, and this inhibitor could have a desirable function at this condition.

---

[2] The article derived from this chapter has the title *"Unveiling the Effect of Low pH on the SARS-CoV-2 Main Protease by Molecular Dynamics Simulations"* and was published in the journal Polymers of the MDPI[36].

5.1 Introduction

The COVID-19 pandemic is still an ongoing major health threat for the whole world, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [37, 38, 39]. The coronavirus family enters the host by fusion with the plasma or endosomal membranes [40, 41]. This mechanism has been used by many positive-strand RNA viruses like the influenza A virus (IAV) [42, 43], the human immunodeficiency virus (HIV) [44, 45], coronaviruses and others. Among the therapeutical targets of these types of viruses are proteases, classified according to their catalytic site as cysteine proteases [46], serine proteases [47] aspartic proteases [48, 49] and metalloproteases.

In the SARS-CoV-2 entry mechanism, the coronavirus spike protein binds to ACE2 receptors found on the surface of many human cells. The virus enters the cytoplasm by endocytosis, then its genetic material, a single-stranded RNA, is released into the cytoplasm. This material encoded two polyproteins (pp1a and pp1ab) which are necessary for viral replication and transcription [50, 51, 52]. The proteolytic process was carried out by the SARS-CoV-2 Main Protease (Mpro), a cysteine protease with a catalytic dyad that compromises two principal residues, Cys145 and His41 [53, 54, 55, 56, 57, 58]. This protein is essential in the life cycle of the virus and is considered a relevant target against the replication of the coronavirus. The catalytic mechanism among coronavirus cysteine proteases involves either using a catalytic dyad to perform a nucleophilic attack that covalently links the protease to the substrate protein, releasing the first half of the product. This covalent acyl-enzyme intermediate is then deacylated by an active water molecule releasing the second half of the product [59, 60].

SARS-CoV-2 Mpro is active in its dimeric conformation, but the SARS-CoV-1

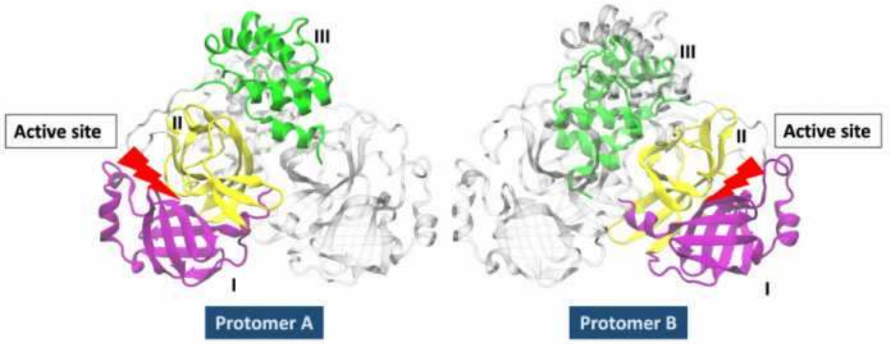


Figure 5.1: SARS-CoV-2 Mpro, PDB ID: 5RE4. Domain I in purple colour, domain II in yellow colour and domain III in green colour.

Mpro is found as a mixture of monomer and dimers in solution [61]. SARS-CoV-1 Mpro has been crystallised in different pH-dependent conformations suggesting flexibility, particularly around the active site [62]. The protease is highly conserved in all coronaviruses, including SARS-CoV-2, so targeting either dimerisation or enzymatic activity may give rise to drugs that can target multiple coronaviruses, known and yet unknown [63, 64]. In Figure 5.1, the three-dimensional structure representation of SARS-CoV-2 Mpro (approx. 33kDa of total weight) shows two identical protomers (homodimer conformation). Each protomer has three domains, domain I (residue 10 to 99), domain II (residue 100 to 182), and domain III (residue 198 to 303); a cluster of five α-helices characterises this last domain. Domain II is connected to domain III via a longer loop.

Many macromolecular properties in the coronavirus family are pH-dependent [40, 65]. As a starting point, it is important to mention that there are studies on the effect of pH in SARS-CoV-1 Mpro, showing the high proteolytic activity of SARS-CoV-1 Mpro around the physiological pH (7.0–7.6) [66, 67, 62, 68, 69]. The probability of the pH-activity profile of the SARS-CoV-1 Mpro was determined by protonation of His163 (inactivation at acidic pH) and deprotonation of His172 [62].

This behaviour is different for conformational stability; some viral proteases have shown their biological role at low pHs, as shown in the endogenous viral proteases from porcine coronavirus are active in a low-pH environment [70].

The SARS-CoV-2 Mpro activity is thought to exhibit a bell-shaped pH-rate profile as determined for the SARS-CoV-1 Mpro. For SARS-CoV-1, fitting kinetic data to model equations provides rate constants and two pKa values (6.2–6.4, 7.7–8.3) that presumably correspond to the catalytically competent His41 and Cys145 side chains [71, 72]. Further understanding of pH effects can be gained from theoretical pKa estimations for all the titratable amino acids of Mpro, which have been also employed to determine their most likely protonation states [67]. Another recent work has demonstrated that the structure of SARS-CoV-2 Mpro shows pH dependence. At basic pH, the Mpro structure was less stable, and as pH becomes more acidic, the protein gets destabilised. Moreover, the SARS-CoV-2 Mpro showed maximum stability at neutral pH but in the other pHs analysed other than neutral pH, dimerisation and S1 binding pocket were found to be affected [73].

Moreover, due to the high homology between SARS-CoV-1 Mpro and SARS-CoV-2 Mpro, similar behaviour is expected. Indeed, the coronavirus family is characterised as an enveloped virus that may or may not be low pH dependent [62, 74, 75]. It has been reported that higher levels of ACE2 in tissues could be the consequence of some of the comorbidities associated with severe SARS-CoV-2 infection [76]. They showed that the acid pH associated with Barrett's oesophagus patients might drive an increased expression of ACE2. In addition, they studied human monocytes cultured in acid pH, observing an increased ACE2 expression and higher viral load upon SARS-CoV-2 infection [76]. A recent review shows that acidosis is generally associated with the severity of SARS-CoV-2. Lactic acidosis produces an excess of the levels of lactate, which is associated with a reduction of pH [77]. The associated enzyme, lactate dehydrogenase (LDH), which transforms pyruvate to lactate, could be selected as one biomarker, as high LDH levels have been associated with worse outcomes in patients with the viral infection [78]. The excess of lactate, a stronger acid than pyruvate, leads to a decrease in the pH of the medium. Herein, we investigate the effect of acidic conditions on the pH-dependent conformational properties of SARS-CoV-2 Mpro, which may be involved in inhibition processes associated with severe SARS-CoV-2 infections. Currently, the literature shows us several studies of the effect of pH by docking calculations. These studies have shown that the pH dependence of the enzyme activity may reflect structural changes in the catalytic dyad [79, 80, 81, 82, 83, 84, 85]. For instance, Drazic et al. tested the influence of different pH values on substrate cleavage velocity and found no significant impact in the pH range from 7.0 to 8.0 [86]. Several publications described the influence of pH on the enzyme activity of the SARS-CoV-1 Mpro, which reported a peak of substrate cleavage at pH 7.0 [87, 62], whereas other studies reported the highest processing at around pH 7.5 [67], or pH 8.0 [88, 89].

The first analysis by molecular dynamic simulation at pH 7 of SARS-CoV-2 Mpro demonstrated the active site's flexibility and effects induced by substrate binding [71]. On the other hand, SARS-CoV-2 Mpro showed high thermodynamic stability in a wide pH range and low stability in the presence of salts [81]. With respect to the catalytic site, the most likely hydrolysis mechanism assisted by the coronavirus Mpro, the reactive events would involve the proton transfer from Cys145(SγH) to His41(Nϵ) and the nucleophilic attack of the Sγ atom to the carbonyl group of a Gln. Thus, the general mechanism considers that Cys145 should be protonated and His41 acts as a base, as in the case of the neutral pH [71].

Verma et al. analysed the proton-coupled dynamics of SARS-CoV-2 Mpro that represented the first step toward the mechanistic understanding of the proton-coupled structure dynamics function at pH 5 to 9 [90]. Likewise, new advances in pH-dependence have shown the importance of appropriate histidine protonation states of SARS-CoV-2 Mpro in the coupling with inhibitors [91]. In the last two years, different SARS-CoV-2 inhibitors have been studied, such as synthetic and natural compounds [92, 93, 94, 95, 96, 97, 98, 64, 99, 100, 101, 102, 103, 104, 105, 106, 107].

Our study provides evidence about the relation between flexibility and pH-dependent Mpro, which was observed at the low pH conditions. We applied Semi-Grand Canonical Monte Carlo (SGCMC) method [1, 2, 19, 20] to assign different protein charge distributions, considering the protonation/deprotonation states for each pH. The homemade developed code within our group makes it possible to define different protonation/deprotonation microstates for a particular pH as shown in the work of Barazorda-Ccahuana et al. [3]. One step further was made by analysing the observed conformational changes due to protonation/deprotonation states by atomistic molecular dynamics simulations.

Along with the explored protonation/deprotonation impact presented in an atomistic level for the Mpro in the frame of molecular dynamics simulations, also we stressed the effect of interaction with PF-00835231 as a potent inhibitor of SARS-CoV-2 Mpro by docking calculations at different pHs. The finding of the therapeutic effect of the PF-00835231 has demonstrated promising data or outcomes in Pfizer preclinical studies [55, 108, 109, 110, 111, 112, 113]. Indeed, the SARS-CoV-2 Mpro and PF-00835231 coupling analysis could serve as a model to study other molecules of natural origin targeting SARS-CoV-2.

5.2 Computational Methods

5.2.1 Protonation/Deprotonation States by SGCMC

We have used the SARS-COV-2 Mpro crystal structure reported in the Protein Data Bank (`https://www.rcsb.org`, accessed on 10 June 2020) by the access code PDB ID: 5RE4. It was determined by the X-ray diffraction method with a resolution of 1.88 Å. The first step was to determine the pKa values of each ionisable residue with the PROPKA v. 3 program [21, 114]. Indeed, the summary of pKa values were used for the calculation of protonation states of Asp, Glu, Arg, Lys and His residues, and the C-terminal and N-terminal ends.

In this work, we use the homemade code (freely available at GitHub `https://github.com/smadurga/Protein-Protonation`) that calculates the protonation and deprotonation states by Semi-Grand Canonical Monte Carlo (SGCMC) procedure based on the free energy associated with the pKa of each i titratable residue, this is calculated using the Equation (6.1) or Equation (6.2) [1, 2, 19, 20]

$$\Delta G = +k_B T(ln(10)(pH - pKa_i)) \qquad \text{Protonation} \tag{5.1}$$

$$\Delta G = -k_B T(ln(10)(pH - pKa_i)) \qquad \text{Deprotonation} \tag{5.2}$$

In this work, we applied 200,000 steps for protonation/deprotonation of SARS-CoV-2 Mpro residues generating ten independent microstates for five pHs (3 to 7).

5.2.2 Molecular Dynamics Simulation Details

Molecular Dynamics (MD) simulations have been performed for four microstates for each condition of pHs with the Gromacs (Groningen Machine for Chemical Simulations) v. 2019.1 software [23]. AMBER99SB-ILDN [115] force fields have been used in the simulations of SARS-CoV-2 Mpro. The periodic boundary conditions (PBC) [116] were considered to minimise edge effects in a finite system. For each system, the protein was in the centre of a cubic box with a minimum distance of 1.5 nm to the faces of the simulation box. The box system was solvated with TIP5P [117] water model. As well as, the total charge of each microstate was neutralised with Cl^- or Na^+ to attain equilibration. Energy minimisation was carried out using the steepest-descent algorithm with 200,000 steps of simulation. WE performed 10 ns MD simulations of equilibrium in the canonical ensemble NVT with position restraint, and the temperature was regulated with the V-rescale thermostat at 309.65 K. Furthermore, we performed the production of simulation in the isothermal-isobaric ensemble with the Parrinello–Rahman barostat with a reference pressure of 1 bar, V-rescale thermostat with 309.65 K and 250 ns time of the simulation. To further corroborate our findings, the conformational ϕ and ψ angles of residues in Mpro were analysed. This analysis of the quality of the structure was solved with the Ramachandran plot in the Molprobity server [118].

5.2.3 Molecular Docking

Molecular Docking calculations were performed using the AutoDock Vina software [119] integrated into the SAMSON molecular design platform as a SAMSON extension available at `http://samson-connect.net`. The extension provides additional functionality to prepare receptors and ligands easily, dock ligand libraries, analyse and export docking results. We used SAMSON and Vina extensions to configure calculations, export input files, and run docking calculations.

The number of modes was set to 100 with an energy range = 3 kcal/mol (default value). The energy range (kcal/mol) is a maximum energy difference between the best binding mode and the unfavourable one displayed. The energy (affinity) that differs more than 3 kcal/mol from the best mode are not saved among results. In the configuration file, the parameter called "exhaustiveness" was set to 8. This parameter controls how comprehensive will be the search space.

5.2.4 Simulation Data Analysis

The RMSD (Root-Mean Squared Deviation), RMSF (Root-Mean Squared Fluctuation), RG (Radius of Gyration), HB (Number of Hydrogen Bonds), SASA (Solvent Accessible Surface Area) and RDF (Radial Distribution Function) analysis were calculated by Gromacs tools.

Additionally, the Salt Bridges (SB) were analysed, considering the last frame of MD for each microstate with the "Salt Bridges" utility of the Visual Molecular Dynamic (VMD [30]) program. Finally, the plots were made in Gnuplot v. 5.4 [120] and the represented in the manuscript images were created with VMD.

5.3 Results

5.3.1 Protonation/Deprotonation states of SARS-CoV-2 Mpro

The molecular structure of SARS-CoV-2 Mpro is a homodimeric structure with approximately 36% of acid/base residues (Asp, Glu, Arg, Lys, His), and almost 8% corresponds to Cys residue distributed throughout the protein structure. For this reason, the analysis of protonation and deprotonation of titratable residues could affect the total charge of Mpro. In this work, different charge distributions are used to assign the protonation state of Mpro acid/basic residues for each pH. For instance, in the case of acidic residues (Glu and Asp) of each microstate, these can be found in protonated or deprotonated mode for pH studied; as well as for basic residues (His, Lys, and Arg). In the case of the residues of the catalytic dyad (His41 and Cys145), we considered Cys145 protonated and His41 in neutral state.

Thus, as conventional MD simulations are used, the atomic charges and the associated protonation states are constant throughout the simulation. The effect of charge fluctuation on a pH is approximated by averaging different simulations with different charge distributions. This approach should be a good approximation of the pH effect, as no overall structural changes of the protein are observed that could significantly change the intrinsic acidity constants.

For instance, the Figure 7.2 shows the Mpro at pH 3 and pH 7, where the beads of blue colour represent the protonated residues and the red colour the deprotonated residues. We notice that at pH 3 there is a low concentration of red beads, the microstates are more protonated at this condition. In contrast, at pH7 a more equilibrated distribution of red and blue beads are seen in the structure of the Mpro. Whereas, in both states (pH 3 and pH 7) the yellow (Cys145) and green (His41) beads represented the residues from the active site with a neutral charge.

Indeed, in Table 5.1 we evaluate the total charge of 10 microstates generated by the SGCMC method, where the average total charge was +(41±2) (pH 3) and -(8.4±0.7) (pH 7), and zero value around pH 5, as shown in Figure 7.3. These results allowed us to find the estimated computationally isoelectric point (pI) of SARS-CoV-2 Mpro (pI=5.5) similar to other studies [121].

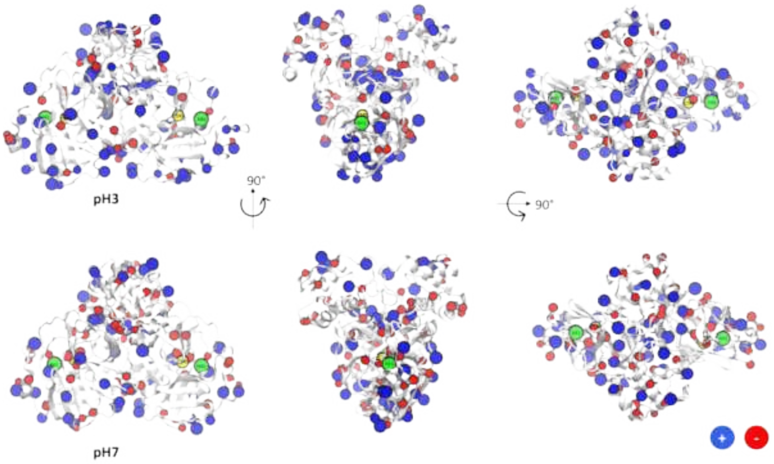


Figure 5.2: SARS-CoV-2 Main Protease (Mpro) with titratable group in beads representation. Yellow beads (Cys145), green beads (His41), blue beads (basic amino acids), and red beads represent (acidic amino acids).

Table 5.1: Total charge of Mpro as a function of pH for the different microstates (MicroS).

	MicroS-1	**MicroS-2**	**MicroS-3**	**MicroS-4**	**MicroS-5**	**MicroS-6**	**MicroS-7**	**MicroS-8**	**MicroS-9**	**MicroS-10**	**Average value**
pH3	38	40	41	42	43	43	42	41	39	41	**41±2**
pH4	24	27	20	22	23	20	26	22	22	22	**23±2**
pH5	7	4	8	6	2	5	1	5	2	7	**5±2**
pH6	-6	-5	-4	-4	-5	-6	-8	-7	-4	-6	**-(5.5±1.4)**
pH7	-8	-9	-9	-8	-8	-7	-8	-9	-9	-9	**-(8.4±0.7)**

Table B.1 from Appendix B shows the distribution of the total charge of the protein among the different titratable residues of the protein. It could be seen that the charge state of acidic residues (Asp and Glu) was the most affected by pH in the studied range, where a large part of these residues at pH 3 was in their protonated state, generating an average of -(9±1) (Asp) and -(4±1) (Glu) partial charge. For pH 4, there are values of -(23±2) (Asp) and -(7±1) (Glu). For the case of pH 5, pH 6, and pH 7, they have had the similar number of Asp and Glu deprotonated. The basic residues (Lys and Arg) were always charged independently of the studied pH. Contrary, the average total charge of His was +10, +(9±1), +(6±1), +(2±1), and +(1±1) for pH 3, pH 4, pH 5, pH 6, and pH 7, respectively (as shown in Table B.1 from Appendix B). Additionally, we focused on the charge state of two important

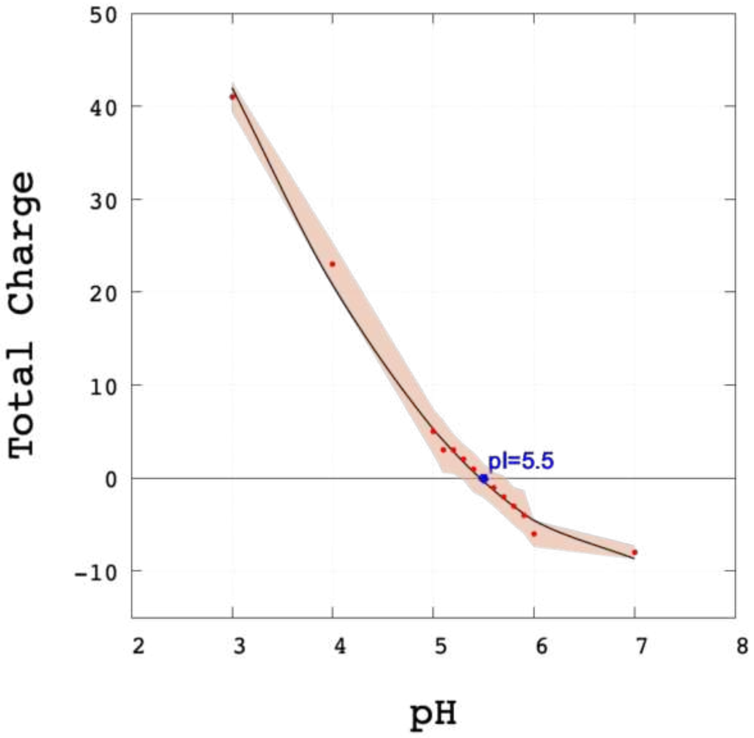


Figure 5.3: Total charge of Mpro at different pHs. The representation of smooth (color) indicates the average and the error for each microstate, the estimated computationally pI=5.5 is in blue color, and the points were fitted to a polynomial regression line (black line).

histidines of the active site, His41 and His172. In general, these two histidines were protonated at pH 3 and pH 4 and deprotonated at pH 5, pH 6, and pH 7 (Table B.2 from Appendix B).

5.3.2 Analysis of Molecular Dynamic simulation

To envisage the role and the importance of conformational flexibility of the explored cases 20 MD simulations were conducted (4 independent microstates for each of 5 pHs considered). In Figure 7.4a the RMSD diagram indicates that the Mpro in explicit solvent TIP5P water model starts to be relatively stable around 50 ns at the different conditions of pHs.

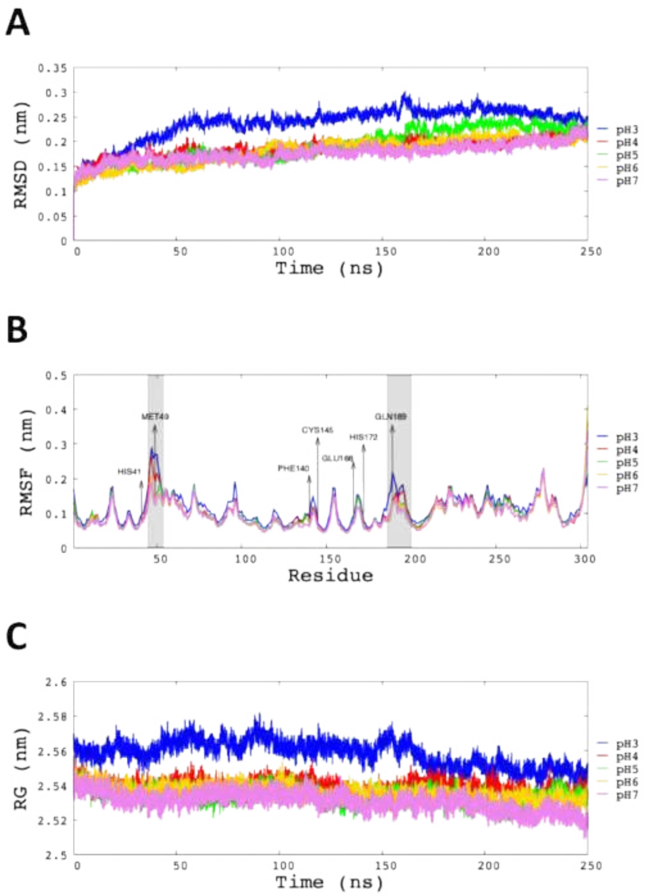


Figure 5.4: RMSD, RMSF and RG graph of backbone for each pH modelled. a.) RMSD presented as a function of time. b.) RMSF *per residue* of C-α, highlighting the most important residues of Mpro, during last 100 ns of MD simulations. c.) RG analysis of structural compactness.

It could be seen in Table 5.2 that the average value of the RMSD at pH 3 was (0.26±0.06) nm; this was the highest value among all simulated pHs. In the range of pHs 4 to 7, the average RMSD value was very similar. In order to analyse the dispersion of RMSD values among different microstates, individual RMSD values are given in Table B.3 from Appendix B.

The RMSF for C-α atoms *per residue* of Mpro were computed at different pHs (Figure 7.4b). In all cases, the behaviour seems to be similar for all pHs; for example, the catalytic dyad's residues (His41 and Cys145) presented low fluctuation

Table 5.2: Values of the MD was fitted against the average structure of the last 100 ns of the dynamics was used.

pH	**RMSD (nm)**	**RG (nm)**	**SASA (nm^2)**	**Hbond (protein-protein)**	**Hbond (protein-waters)**
pH3	0.26±0.06	2.553±0.010	268±2	401±1	(107±4)x10
pH4	0.20±0.02	2.538±0.002	261±2	437±10	(110±2)x10
pH5	0.21±0.04	2.537±0.005	259±3	448±3	(114±8)x10
pH6	0.20±0.02	2.532±0.005	260±2	456±6	(114±12)x10
pH7	0.19±0.02	2.526±0.007	259±1	461±1	(114±12)x10

independently of the studied pH. On the other hand, the average RMSF of the regions that compromise residues 45 to 50 (Thr45, Ser46, Glu47, Asp48, Met49, Leu50) and 186 to 190 (Val186, Asp187, Arg188, Gln189, Thr190) showed us high fluctuation at pH 3 and pH 4 (Figure 7.4b). Regarding the Glu166 residue, an important residue that participates in the protomers dimerisation, its degree of fluctuation is not affected by the studied pHs. Therefore, the relative stability of global RMSD in conjunction with the relative low values of RMSF of the main part of the amino acids implies that the global structure of Mpro does not disaggregate at low pH conditions, dimeric conformation is stable.

The RG values determined the compactness of the Mpro and were plotted as a function of time. It could be seen in Table 5.2 that at pH 3 and pH 4 are obtained the highest RG value 2.55±0.01 and 2.54±0.01, respectively. In contrast, at pH 5, pH 6 and pH 7 the stability and compactness of the protein were similar (approx. 2.53 nm). The SASA analysis of each pH studied is given in Table 5.2. The average SASA value at pH 3 exhibited increased compared to the other systems ($SASA_{pH3}$ = (268±2) nm^2). Therefore, the SASA value increase in average as pH decreases. The temporal evolution of the SASA values at different pHs are shown in Figure B.1a from Appendix B.

However, significant differences could be observed in the number of protein-

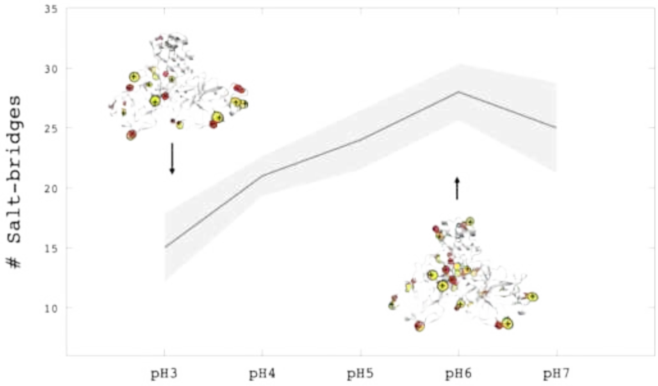


Figure 5.5: Salt-bridges of whole Mpro structure at different conditions of pHs. At low pHs, Mpro is more protonated than pH 6 and pH 7, this graphic show less salt-bridges at pH 3. The line-tendency (black) represent the average value and the smooth tendency (gray) represent the statistical error.

protein and protein-waters hydrogen bonds. In particular, the average number of hydrogen bonds protein-protein and protein-waters increase in average as pH increase (Figure B.1b and B.1c from Appendix B show additional information of each pH studied). In summary, these results show that Mpro at pH 3 presented a lower quantity of hydrogen bonds (protein-protein) than other pHs analysed. The fact that Mpro has many aspartic residues led to constraints for hydrogen bonds formation when the residues were protonated at low pH.

Another parameter that determines the stability of Mpro, was the number of salt-bridges which depend on the binding of positively and negatively group charges. [122] In this work, the change in the protonated and deprotonated states of Mpro determined that at low pHs exist a loss of salt-bridges. Usually, since the salt-bridges depend on the pH, we saw that at pH 3, there is a lower number of salt-bridges than the other pHs (Figure 7.5). Since these systems are further from the isoelectric point (pI = 5.5), the positive and negative charge is unbalanced, so a lower amount of salt bridges is expected.

The flexibility of residues close to the active site was the object of interest of interest and for other studies dedicated to the same stated problem. The S1 pocket was crucial in the protein dimerisation and residues involved in the substrate-binding sites of SARS-CoV-2 Mpro. For these reasons, we focused on the residues of the catalytic dyad (Cys145 and His41), two residues of the S1 pocket (His172 and Phe140) proposed by Verma et al. [90] and additionally the most fluctuating residues of RMSF analysis (Met49 and Gln189). In Figure 5.6, the relative distance of His41 and Met49 to Cys145 was analysed for the last frame of each microstate. The most remarkable difference is observed at pHs 3 and pH 4, where Met49 is located with the most significant distance to Cys145 (in both protomers). Despite this behaviour in Met49, the other residues (His41, Cys145, Phe140, and His172) were not altered by changing the pH.

The calculation of the average distance of the last 100ns between Cα atoms showed us a constant distance value between Cys145-His41 and His172-Phe140; therefore, the active centre and pocket S1 of Mpro is not significantly altered by pH effect. However, the average distance between Met49 and Gln189 shows significant differences at pH 3 (see Table B.4 and Figure B.2 in the Appendix B for more details). This final result indicates that the pocket increase its volume as there is a greater distance between these residues at pH 3 and pH 4. Therefore, the size of the pocket can compromise the selectivity of the substrate.

Two main steps are identified in the catalytic dyad: the acylation (acyl-enzyme complex) and deacylation (acyl-enzyme hydrolysed). Świderek and Moliner analysed the deacylation step mediated by a water molecule activated by His41. [123] Therefore we have evaluated how the pH affects the solvation of the reactive atoms of this catalytic residue.

The radial distribution function (RDF) is the probability distribution to find a particle at a distance r away from a given reference particle. In this work, the RDF

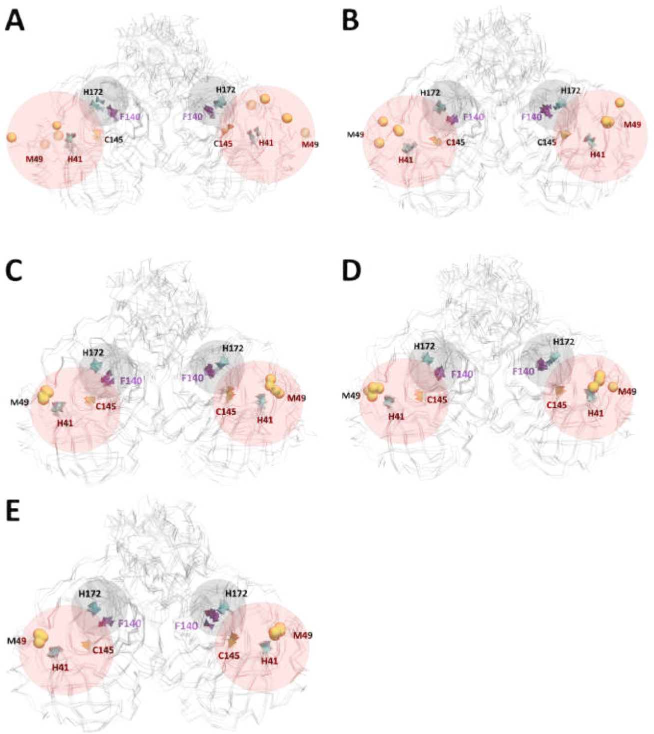


Figure 5.6: Snapshots of the last frame of all microstates in Mpro, highlighting the active site (pink area) and pocket S1 (gray area). a.), b.) pH 3 and pH 4 did not show a constant Met49-Cys145 distance, which was also analysed in the RMSF, where the Met49 residue showed high fluctuations at these pHs. c.), d.), and e.) pH 5, pH 6, and pH 7 with a constant distance between Met49-Cys145, His41-Cys145, and His172-Phe140. The His41, His172, Cys145, and Phe49 residues (polyhedral graphic representation) and Met49 (bead graphic representation) are shown.

of specific atoms of Cys145 (Sγ, Hγ), His41 (Nδ1, Hδ1), Met49 (Sδ), water sites (Hω, Oω), and ions, were analysed. In Figure 7.8, the high of the first peak-height of Sγ–Hω and Sγ–Oω appear at r distance of 0.36 nm and 0.34 nm, respectively. The most striking variation is seen at pH 7 and pH 6, followed by pH 5 and lower peaks at pH 4 and pH 3. In general, the RDF of the solvent distribution on Sγ and Hγ shows a low water affinity at low pHs, which could be crucial to understand the role of environmental conditions in the catalytic mechanism of SARS-CoV-2 Mpro. Adittionally, the RDF analysis of ions around Sγ from Cys145, shows the peak-height at r distance of 0.5 nm for pH 3 and pH 4. It is worth noting that

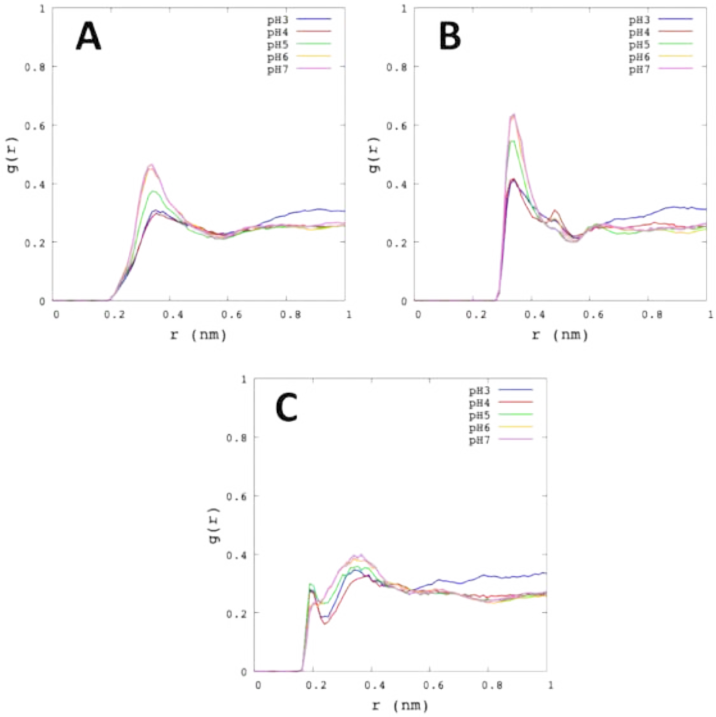


Figure 5.7: Radial distribution function of Cys145. The graphs show the average value of the Mpro protomers and the different microstates for each pH. a.) Sγ–Hω. b.) Sγ–Oω. c.) Hγ–Oω.

there is a significatively presence of mobile ions close to the active site (see Figure B.3a from Appendix B).

His41 reported more affinity for water at pH 3 than pH 7 (Figure 7.9). Indeed, Nδ1–Hω and Nδ1–Oω show the high first of the peak-height at pH 3 with r=0.33 nm and r=0.28 nm, respectively. Likewise, the Hδ1–Oω showed a large peak at r distance of 0.18 nm, where the highest was pH 3. Therefore, with this result we can observe that the Nδ1 and Hδ1 atoms from His41 are more related to water at low pHs. Furthermore, His41 with ions showed a similar behaviour of Cys145 (see Figure B.3b from Appendix B).

The RDF analysis of Met49 is seen in Figure 7.10. The Sδ–Hω obtained the peak-height at r distance of 0.49 nm, similar to Sδ–Oω at pH 3. Although Met49

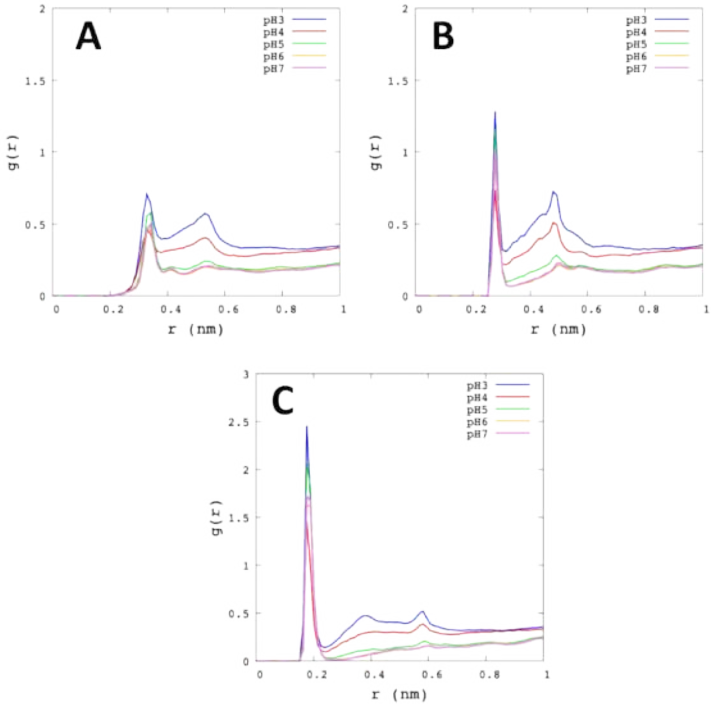


Figure 5.8: Radial distribution function of His41. The graphs show the average value of the Mpro protomers and the different microstates for each pH. a.) Nδ1–Hω. b.) Nδ1–Oω. c.) Hδ1–Oω.

is highly fluctuating at pH 3, this does not influence the early interaction with water molecules.

In summary, all findings of MD simulations allowed us to understand the high stability of Mpro at low conditions of pHs. The conformational stability of Mpro during the simulation time remains stable at the explored pHs scales. The significant number of residues at low pHs led to increased SASA. The impact of this effect was translated and to hydrogen bonds formation (intramolecular and intermolecular), which was confined. Although these results are not unexpected, they show that the ensemble of the Mpro dimer remains stable and that the most acidic pH conditions seem to be unable to disaggregate the enzyme. Furthermore, we have shown that the distance between the residues of the active site and the S1

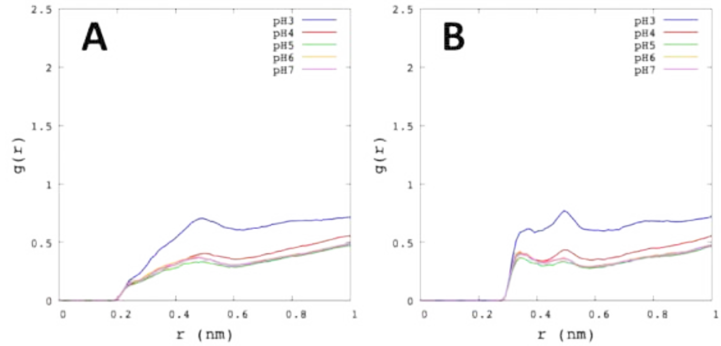


Figure 5.9: Radial distribution function analysis of Met49. The graphs show the average value of the Mpro protomers and the different microstates for each pH. a.) Sδ–Hω. b.) Sδ–Oω.

pocket of the Mpro remained constant during the simulation time, which indicated that the structure of Mpro could not disaggregate at low pH conditions. On the contrary, the residues of the zone of binding-substrate supported conformational changes at pH 3, which increased in the pocket volume. This result has related to the high fluctuations are presented in the RMSF diagram. Taking into account the importance of the solvent in the reaction mechanism of SARS-CoV-2 Mpro. In the RDF analysis, we have found that the solvent showed a high and low affinity towards the atoms of His41 and Cys145, respectively. Finally, the last frame of each trajectory was utilized to analyze the accessible values of the torsion angles ϕ and ψ create a Ramachandran plot. This analysis allowed us to validate the protein structures of each pH. Mpro at different pHs conditions (3 to 7) had a similar percentage of residues in allowed (∼99%) and favoured (∼92%) regions (more details are in Table B.5 from Appendix B).

5.3.3 Molecular Docking analysis

To analyse the effect of environmental pH in the interaction with protease inhibitors, a recent ligand of Mpro protease has been selected in the docking study.

Molecular docking calculations with PF-00835231 ligand [55] and the 40 receptors (last frame of five pHs x five microstates x two monomers) in different conformations according to the protonation states have been conducted (See Figure B.4 from Appendix B, where PF-00835231 presents an estimated pKa equal to 12.76±0.10 predicted using ACDlabs software v11.02, which allows it to be neutral and does not suffer alteration with the change in pH.). Appropriated protonation states were preserved to both receptors and ligand. Both the number of flexible side chains and the size of the search domain were different for all the cases because of the receptor's conformation (i.e. chain orientation, the position of residues). On average, there were about 30 flexible side chains (from about 27 to about 35 flexible side chains). Table B.6 from Appendix B represents the data for the SARS-CoV-2 Mpro with different pH: sizes of the search domain, the volume of the search domain, and the number of the flexible side chains.

The search space was defined by a docking box that wraps the space around the receiver for each system with 40 different docking boxes. Active pocket amino acid residues were used to centre the docking box. Therefore, the pocket was placed around Cys145.

In AutoDock Vina the electrostatic interactions were mainly determined through hydrogen bonding terms. Figure 5.10a shows the residues of Mpro that establish hydrogen bonds with the PF-00835231 drug at different pHs (Figure B.5 from Appendix B shows the 2D representation of the best coupling of each pH). Furthermore, at pH3, an additional pattern of interaction could be seen that shows the interaction of the drug with residues Met82, Asn84, and Cys85. For the other hand, Figure 5.10b shows the binding affinity of the best docking poses for each microstate. The binding mode of the drug close to the catalytic Cys145 shows a similar range of affinity (-7 to -6 kcal/mol) among all studied pHs. However, the

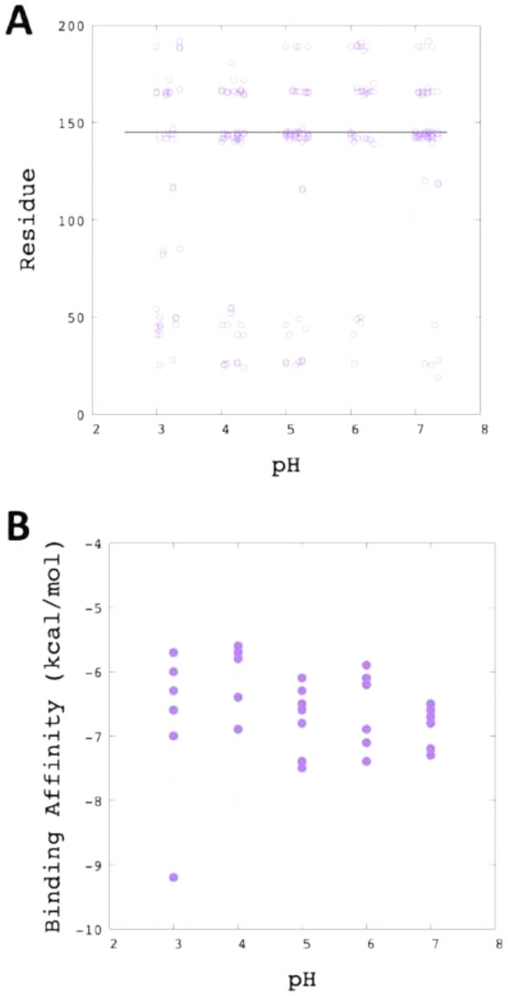


Figure 5.10: Binding analysis. a.) Mpro residues that interact with the drug PF-00835231 through hydrogen bonds at different pH (3, 4, 5, 6 and 7). For each pH, the best docking of each of the 8 microstates is analysed. To distinguish between the different docking poses, each is shifted slightly on the pH axis. A horizontal line is also depicted at 145 to help visualize the interaction with the catalytic Cys145. b.) Predicted binding affinity of drug PF-00835231 with Mpro at different pH. For each pH, the best binding affinity of the 8 different microstates are shown.

more acidic pH shows an additional mode of interaction, more separated from the catalytic Cys145 with the highest affinity.

According to the docking results, the Figure 5.11 reports the best coupling energies for each system studied. Where the His41 and Cys145 residues remain

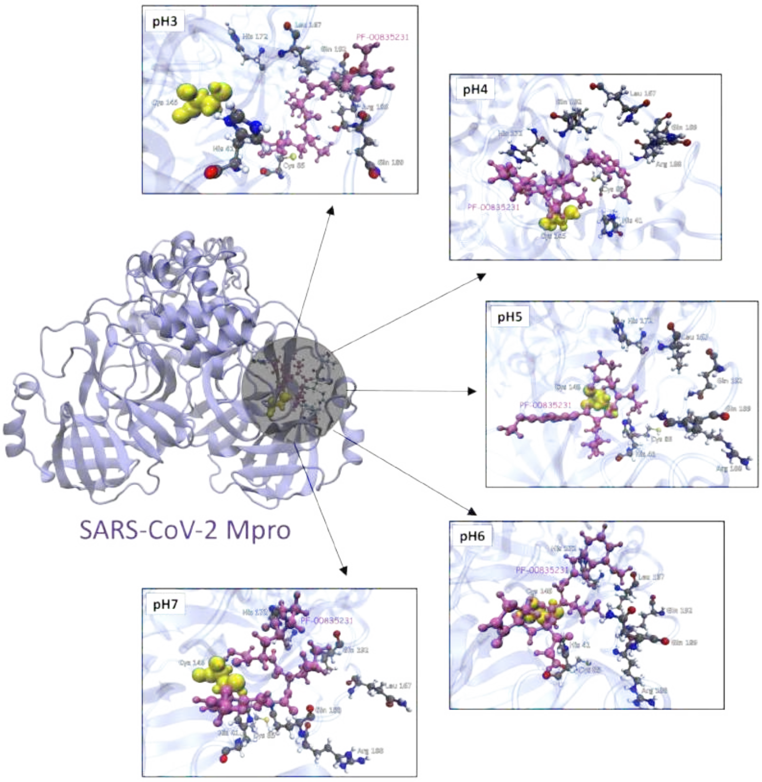


Figure 5.11: Docking snapshot of the best binding affinity at different pHs between PF-0085231 and SARS-CoV-2 Mpro. The Cys145 is yellow highlighted and the PF-00835231 drug is pink highlighted. The ligand is located close to the SARS-CoV-2 Mpro active site at different conditions of pHs.

very close to PF-00835231 at different pH conditions.

5.4 Conclusion

The SARS-CoV-2 Mpro is an exciting target gaining the attention of many research groups globally. Its primary importance lies in the replication of the virus, and its inhibition could be the key to stopping the virus from spreading. However, one of the parameters in which is compromised the activity of this enzyme is the pH-dependence.

Although from the point of view of experimental analysis, the study of proteins under low pH conditions is a great challenge. The computational chemistry approach looks the way to study the effect of pH on many important mechanisms such as explored herein.

The program was developed to define a total charge distribution for a specific pH, developed under the SGCMC method. This work showed the distribution of protonation/deprotonation states of the titratable residues with a total net positive charge in pH 3 until pH 5 and a total net negative charge at pH 6 and 7. Moreover, this analysis was able to estimate the isoelectric point (pI = 5.5) computationally. The conventional all-atom MD showed us that Mpro has a stable at low pH structure.

Indeed, Mpro at pH 4 to 7 showed high structural stability of the whole protein, active site, permitting the promoters' dimerisation. For the case of microstates at pH 3, slight conformational changes were observed. This pH-dependent computational approach revealed that Met49 showed the most significant fluctuations at pH 3 and increased the Mpro pocket volume. Identifying these mechanisms could be used for target selectivity validation.

The last part of the paper elucidates the druggability interface for protonation-dependent Mpro binding sides for a drug candidate in the clinical trial phase. The protein surface was explored using a molecular docking-based screen for a new drug candidate (PF-00835231). The predicted binding modes and affinity for the drug candidate were found in the range from pH 4 to pH 7. While at pH 3, an additional mode of binding was observed; however, more studies could be necessary to elucidate its relevance in Mpro inhibition at very acidic pH.

Chapter 6

Computational study of pH effect on different SARS-CoV-2 mutations in the S1 region of the Spike protein against the ACE2 receptor

The spike glycoprotein mediates the entry to the host cell in the numerous SARS-CoV-2 variations, which have placed the entire globe on high alert about a potential spread. Likewise, the importance of pH in the binding of spike glycoprotein S-1 (Spike-S1) with the angiotensin-converting enzyme 2 (ACE2) is also becoming increasingly apparent. This study aimed to assess the binding of Spike-S1 of five SARS-CoV-2 variants with ACE2 at five pHs. Here, the charges of the titratable residues at five pHs were defined using the Semi-Grand Canonical Monte Carlo approach. The structural stability of the complex was next examined using all-atom molecular simulation methods, and the estimation of the binding energy between the proteins by MM/GBSA was assessed. The results show that among the five variants of SARS-CoV-2, omicron was the one that presented the best binding free energy against ACE2, suggesting that this variant has a high tolerance for acidic and basic pHs.

6.1 Introduction

At the beginning of 2020, the World Health Organization (WHO) declared the outbreak of Coronavirus disease 2019 (COVID19) as an international public health emergency [52, 124]. The new virus called Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is a variety of the coronavirus family. Although the clinical management of COVID19 has achieved significant advances, the different mutant variants of the virus have managed to adapt to the human host, which keeps the world on alert regarding the spread of the virus [125, 126].

To prioritise global monitoring and research, the WHO characterises the specific Variants of Interest (VOIs) and Variants of Concern (VOCs). In September 2020, the B.1.1.7 (Alpha) variant was reported in the United Kingdom; this variant showed a significant increase in transmission [127, 128, 129, 130]. Subsequently, the B.1.351 (Beta) variant was reported, which appears to be less sensitive than the alpha variant [131, 132, 133]. The lineage B.1.617 is divided into three sublineages B.1.617.1 (Kappa) [134, 135], B.1.617.2 (Delta) [136, 137], and B.1.617.3. Among them, Delta caused the resurgence of COVID19 in early 2021, which kept countries that initially had COVID19 outbreaks under control on alert [138, 139].

In mid-November 2022, 99.2% of sequences reported globally correspond to variant B.1.1529 (Omicron) [140, 141, 142]. This variant of SARS-CoV-2 continues to be a Variant of Concern (VOC) considered as such due to its rapid transmissibility and its subvariants. The WHO emphasizes that these descendant lineages should be monitored as distinct lineages by public health authorities, and comparative assessments of their virus characteristics should be undertaken.

In general, it has been possible to identify different therapeutic targets of SARS-CoV-2 that are important for its spread. Among them are the spike glycoprotein

(S) [143, 144, 145], a small envelope protein (E) [146, 147, 148], membrane proteins (M) [149], and nucleocapsid (N) [150]. The spike glycoprotein protruding from the virion surface plays a critical role in the initiation of viral infection by facilitating the binding of the coronavirus to the host cell surface receptor and the fusion of the viral and host cell membranes. The spike glycoprotein consists of two functional subunits, S1 and S2, and the receptor-binding domain (RBD) resides within the S1 subunit. RBD binds to the peptidase domain of angiotensin-converting enzyme 2 (ACE2), initiating virus attachment to the host cell surface [151, 152, 153, 154, 155]. Fusion and genome penetration into the host cell require exposure of the virion to an acidic medium ranging from pH 6.2 to 6.8. This event is closely related to the conditions of the respiratory mucosa that could facilitate the rapid entry of SARS-CoV-2 [156].

In order to understand how coronaviruses enter the host cell, it is crucial to understand the impact that pH plays in these processes. Several studies have examined how pH levels can change how these viruses' target proteins fold [3, 157, 36]. Within these, the pH is important in folding protein structures. In the Main protease of SARS-CoV-1 and SARS-CoV-2, the structural stability at low pH conditions was observed by computational simulation methods in which the protonated and deprotonated states of titratable residues have a desirable impact at different pHs [62, 36]. On the other hand, a study by Yixin Xie et al. related to the stability of SARS-CoV-2 Spike-S1 at pH values ranging from 7.5 to 9 [158]. To date, few investigations have involved the study of pH conditions in SARS-CoV-2. Therefore, considering this condition is relevant to understanding the binding mechanisms between the virus and the host cell. Thus, although the SARS-CoV-2 virus is controlled in different countries, the world's concern is not lowering its guard against the different variants that can come to light. This work aim is to analyse by computational simulation of five variants of SARS-CoV-2 (Alpha, Beta, Delta, Kappa, and Omicron) and how the pH could influence the binding with the therapeutic target SARS-CoV-2 Spike-S1 against ACE2.

6.2 Computational Methods

This study analyzed the structures of the spike-S1 protein of five SARS-CoV-2 variants. These were obtained from the Protein Data Bank (PDB) database under the following access codes: alpha:7LVW, beta:7LYO, kappa:7V7E, delta:7V7Q, and omicron:7T9K.

The design of the coupled systems was based on aligning the different mutations taking as a fixed point the Spike-S1 wild previously linked to ACE2. This test made it possible to obtain six coupled systems considering the structural model of ACE2 reported in the Spike-S1 wild.

The Semi-Grand Canonical Monte Carlo (SGCMC) method was used to calculate the protonation/deprotonation states [1, 2, 19, 20]. In principle, the experimental pKa of each atom that makes up the two protein models was considered. The pKa value was based on the use of PROPKA v3 [159] program. The output file provided us with a list of experimental pKas values. Later, this file was used as the input file in the homemade program designed by our group (available at the link `https://github.com/smadurga/Protein-Protonation`). The protonation/deprotonation state is calculated based on the following Equations:

$$\Delta G = +k_B T(ln(10)(pH - pKa_i)) \qquad \text{Protonation} \tag{6.1}$$

$$\Delta G = -k_B T(ln(10)(pH - pKa_i)) \qquad \text{Deprotonation} \qquad (6.2)$$

A thousand cycles of calculations were applied, and ten microstates of protonation/deprotonation states were obtained at five pHs (4, 5, 6, 7, and 8) for each system (Spike-S1 mutated vs ACE2).

The atomistic simulations of the interacting systems were carried out in the free access program GROMACS v. 2021 [23]. For this work, the CHARMM27 [160] force field was selected for all simulations. Herein, the protonation/deprotonation states for each titratable residue were fixed in the topology preparation. Each system was placed in a cubic box with a distance of 1.5 from the box's edge towards the macromolecule's surface. In addition, each system was solvated with water molecules, considering the TIP5P [161] water model. Likewise, the NA and CL ions were to added to neutralize the systems. Periodic boundary conditions (PBC) [116] were considered in all systems. Three molecular dynamics simulation steps were carried out; the first step consisted in relaxed the structure through energy minimization with the steep-descent algorithm, the second step was the equilibrium simulation conducted under an NVT ensemble (Number of particles, Volume, and Temperature constant) with the V-rescale [28] thermostat at 309.65 K, and the third step was the production simulation conducted under an NPT ensemble (Number of particles, Pressure, and Temperature constant) at 50 ns with the V-rescale [28] thermostat at 309.65 K and the Parrinello-Rahman [29] barostat at constant pressure (1 bar).

Additionally, the binding energy between the different Spike-S1 mutated against ACE2 was evaluated. The Molecular Mechanics Generalised Born and Surface Area Solvation (MM/GBSA) [162] method was adequate to evaluate the best energy contribution based on the following Equations:

$$\Delta G_{bind} = G_{complex} - (G_{protein} + G_{lig}) \tag{6.3}$$

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{GB} + \Delta G_{SA} - T\Delta S \tag{6.4}$$

$$\Delta G_{bind} = \Delta E_{vdw} + \Delta E_{ele} + \Delta G_{GB} + \Delta G_{SA} - T\Delta S \tag{6.5}$$

Where ΔE_{MM} is the variation between the minimized energy of the van der Waals (ΔE_{vdw}) and electrostatic (ΔE_{ele}) interactions of the Spike-S1 mutated against ACE2; ΔG_{GB} and ΔG_{SA} are the polar and non-polar component of solvation free energy; and $-T\Delta S$ refers to the contribution of entropy to temperature (T).

For this analysis, it was used the gmx_MMPBSA [163] tool based on the AmberTools22 [164] program aiming to calculate the free energy with GROMACS output files.

6.3 Results

6.3.1 Sequence-based alignment of SARS-CoV-2 Spike-S1 variants

Five SARS-CoV-2 Spike-S1 mutations were investigated. The first step in this study was to align the amino acid sequences of the proposed Spike-S1 mutations using the Spike-S1 alpha sequence as a reference. The results of the alignment helped us to identify the residues that stand out in the different mutations, the differences in Spike-S1 mutation sequence alignments are shown in Figure 6.1. Despite the high homology between them, they demonstrate how point mutations can alter the look of characters in an alignment column. Numerous secondary structural elements, such as beta strands (green shadow) and alpha helices (yellow shadow), have been identified. Moreover, a red box that encircles each sequence's alterations is emphasized.

The sequences imply three major hot spot locations (blue shadow), with the first hot spot enclosing residues at positions 416 to 418. The second hot spot commits residues from the mutated Spike-S1 from position 439 to 455. The third hot spot considers from position 493 to 506 of Spike-S1 mutations. In the second and third hot spots, delta and kappa showed two alterations with the R450 and N501 mutations. A mutation that distinguishes delta (K478) from kappa (T478) is found in a separate region of the sequence. The only mutation N417 in beta is around the first hot spot residues. The hot spot locations in the omicron sequence show higher recombination rates than the alpha (wild Spike-S1) sequence expectation. In the three hot spot zones, the Spike-S1 omicron exhibits six alterations (N417, K440, S446, R493, S496, and R498). Also noted are three omicron alterations (K440, R493, and R498) that are titratable residues.

6.3.2 Protonation/deprotonation states by Semi-Grand Canonical Monte Carlo

Proteins are polyelectrolytes made up of protonable amino acids [165, 166]. The acidic and basic amino acids (Asp, Glu, Arg, Lys, and His), which can carry a net charge, comprise most of a protein's electrostatic potential. The charge distribution will change with pH due to some neutral protein residues, which is crucial to remember. The Spike-S1-ACE2 complex was considered; Figure 6.2 shows where each protein's titratable residues are located. This study used the Semi-Grand

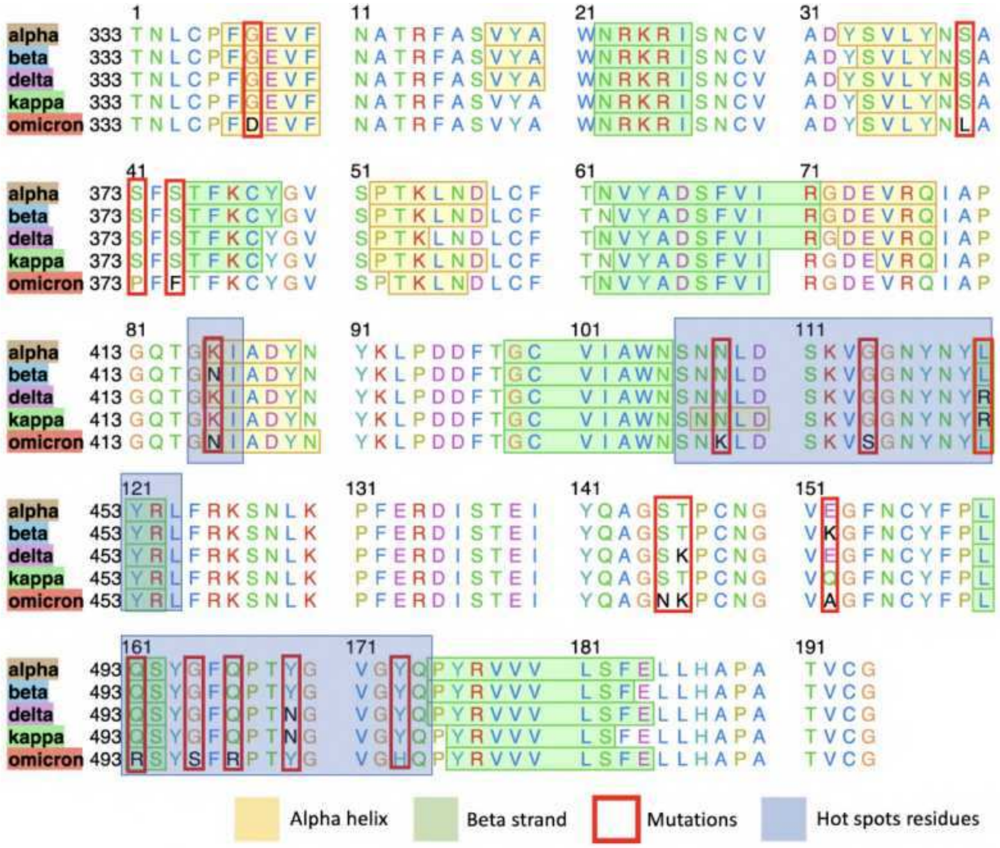


Figure 6.1: A multiple sequence alignment demonstrating how the Spike-S1 mutations used in this research are comparable. The sequences are very homologous to each other. The alpha helix, beta strand, specific mutations, and hot spot residues are depicted by the yellow, green, red box, and blue shadows, respectively.

Canonical Monte Carlo (SGCMC) method to determine the protonation state of the titratable residues at five pHs (4, 5, 6, 7, and 8). SGCMC simulations are naturally distinguished by the overall charge changes due to the dynamic proton binding and unbinding. The homemade software needs the pH, the number of trial protonation/deprotonation steps, and the list of acid/base constants in the format specified by Propka's summary as inputs. The protonated state is initially allocated to each protonable residue. The system reaches equilibrium quickly because the pK_a values are thought to be independent of one another and unaffected by the protonation state of nearby residues.

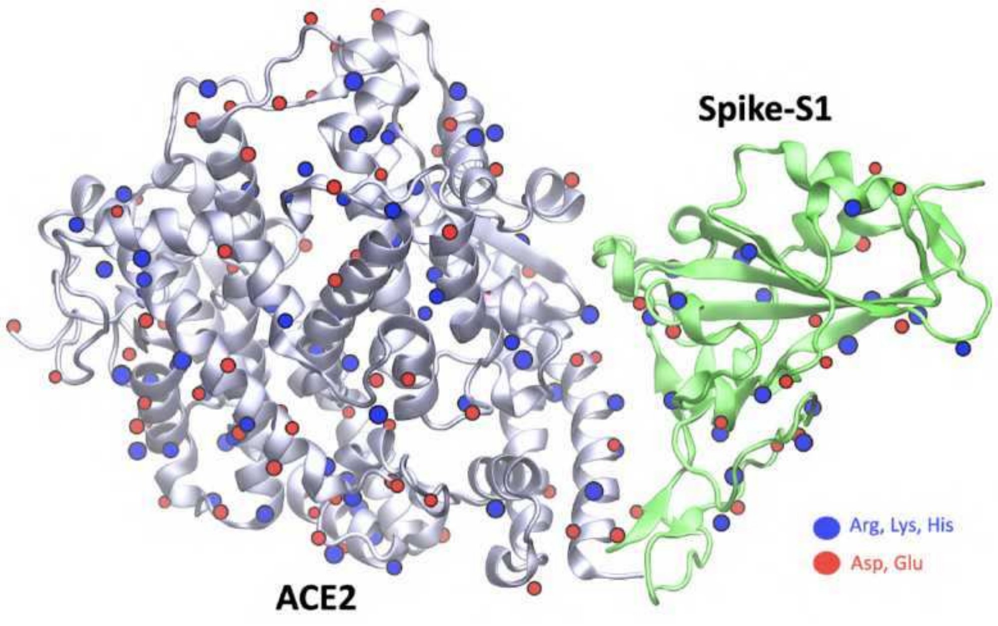


Figure 6.2: Highlighting the titratable residues in complex ACE2-Spike-S1.

The program SGCMC allowed us to analyze ten microstates, which showed similar values at each pH. In Table 6.1, the average of the total net charge for each pH is displayed. The Spike-S1 mutations exhibit different total charges at different pHs. The average charge at pH4 and pH5 typically implies a positive total net charge. In contrast, the total net charge at pH6, pH7, and pH8 are negative.

Table 6.1: Description of the specific total charge for ten microstates for each Spike-S1 mutated at

Spike-S1 variant	pH	MicroS-1	MicroS-2	MicroS-3	MicroS-4	MicroS-5	MicroS-6	MicroS-7	MicroS-8	MicroS-9
alpha	pH4	38	37	32	36	44	36	40	37	41
	pH5	3	4	6	3	6	7	0	1	12
	pH6	-9	-9	-13	-16	-9	-11	-9	-13	-12
	pH7	-19	-20	-18	-21	-22	-19	-18	-19	-24
	pH8	-23	-25	-22	-22	-25	-24	-24	-24	-25
beta	pH4	42	39	39	34	33	36	41	39	38
	pH5	6	8	5	7	6	6	5	11	10
	pH6	-16	-12	-15	-12	-9	-7	-10	-11	-12
	pH7	-20	-21	-20	-20	-22	-16	-20	-18	-20
	pH8	-25	-24	-25	-25	-24	-24	-24	-23	-24
delta	pH4	42	42	41	40	39	36	40	43	46
	pH5	4	11	4	6	4	9	8	9	5
	pH6	-10	-9	-10	-13	-11	-7	-11	-8	-13
	pH7	-17	-20	-18	-18	-20	-18	-19	-16	-16
	pH8	-22	-22	-20	-22	-21	-21	-20	-22	-20
kappa	pH4	41	42	39	33	35	40	43	41	36
	pH5	12	5	11	4	8	7	15	9	6
	pH6	-7	-9	-13	-10	-8	-10	-11	-11	-11
	pH7	-19	-19	-18	-19	-20	-20	-18	-17	-18
	pH8	-24	-22	-23	-22	-22	-21	-22	-20	-23
omicron	pH4	43	33	52	44	42	45	39	45	49
	pH5	11	7	13	7	10	7	15	10	8
	pH6	-8	-13	-6	-10	-12	-6	-12	-5	-9
	pH7	-15	-18	-17	-16	-17	-14	-17	-18	-18
	pH8	-20	-21	-24	-20	-21	-21	-21	-21	-22

6.3.3 Molecular Dynamics Simulation Analysis

From the ten microstates that the SGCMC software generated, two microstates were chosen for molecular dynamics (MD) simulation. The interactions from Spike-S1 binding into ACE2 were examined using MD simulations; each model was evaluated using explicit water models and ions to balance the system. In addition, the protonation/deprotonation states for the titratable residues at five pHs by the SGCMC method were assigned in the Gromacs pdb2gmx module.

The MD simulation optimized the position of the residues that facilitate the interaction between Spike-S1 and ACE2. Root-mean-square deviations (RMSD) and root-mean-square fluctuations (RMSF) of the Spike-S1 regions were examined. The RMSD allows us to measure the difference between the reading frames of a protein from its initial conformation to its final conformation during a simulation period. The structural stability is determined based on the deviations generated, whereas a lower deviation indicates that the protein structure is stable. In Figure 6.3, the RMSD plots of the different mutations at each proposed pH are observed, where we highlight omicron (blue line). Herein, omicron forms a more stable conformer at all pHs than the other mutations.

When comparing the results at the various pHs, we find that the alpha and beta variants act in a very unstable form at pH 4, where they appear unable to achieve equilibrium at 50 ns. In contrast, the omicron displays an approximately average value of 0.3 nm. Alpha, beta, delta and kappa mutations at pH5 present a more altered behaviour than omicron. Beta and omicron are more stable than the other mutations at pH6. It’s interesting to note that all mutations showed similar behaviour at pH7. Concerning the investigations at pH8, we found that Spike-S1 mutations are more stable than other pHs, especially omicron standing out. The

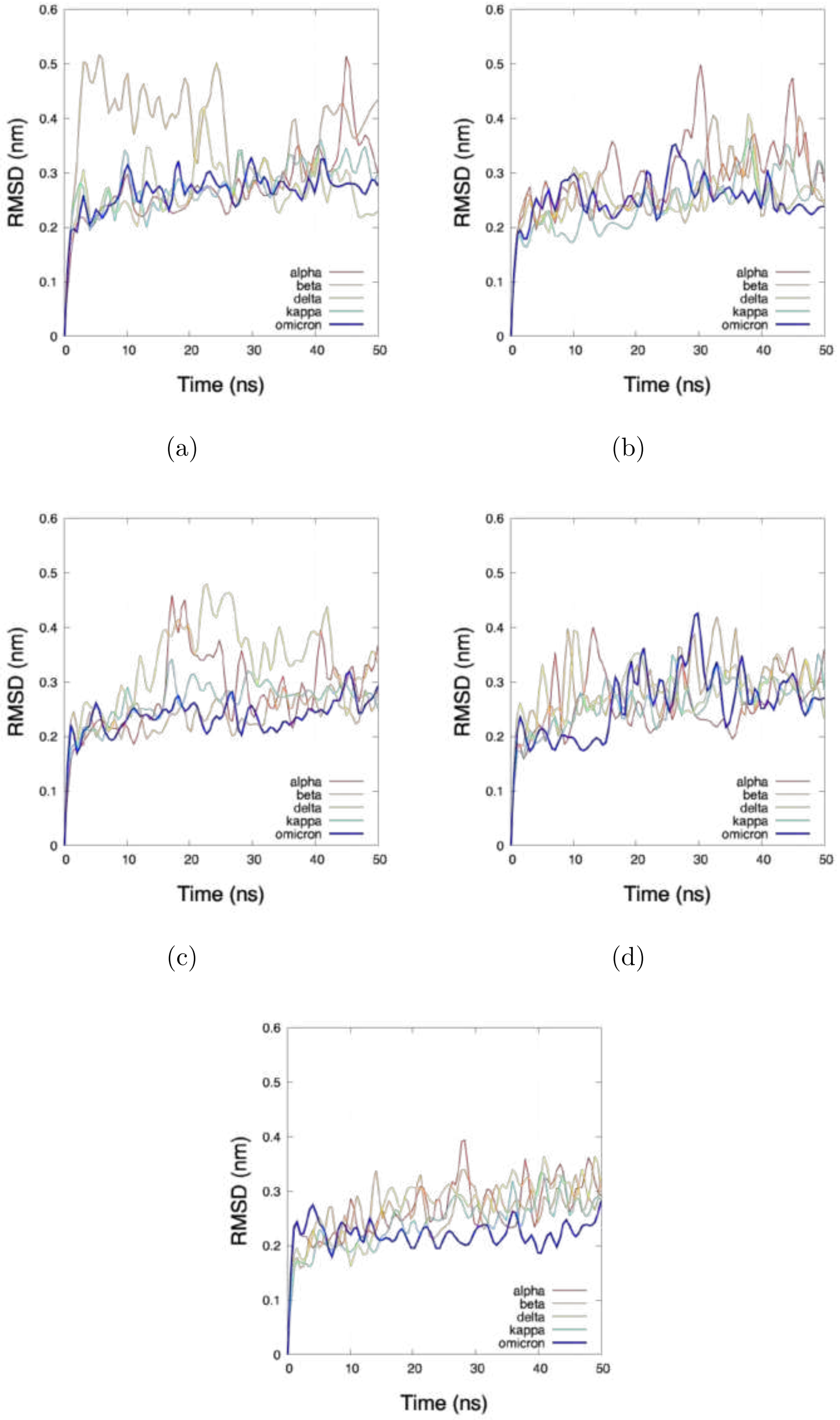
Spike-S1 variations generally exhibit more unstable behaviour at pH values below 8, which suggests that these structures behave better at alkaline pHs.

Subsequently, the flexibility of the Spike-S1 residues was tested by analyzing the RMSF of the backbone *per residue*. This allowed us to verify the low Spike-S1 residues fluctuation in contact with ACE2, identifying the most effective contact to the ACE2 binding area. Figure 6.4 shows the RMSF of Spike-S1 over the last 5 ns, where the hot spots residues region was identified with a grey shadow (first hot spot enclosing residues at positions 416 to 418, the second hot spot residues from position 439 to 455, and the third hot spot considers from position 493 to 506 of Spike-S1 mutations).

Herein, the low fluctuations in the hot spots residues region are shown concerning the other residues that are part of the Spike-S1 surface. Likewise, in all cases, the low value is obtained by omicron. This analysis reveals that this area forms a stable complex with ACE2. Evaluating the alpha variant reveals that the hot spot area exhibits significant changes in all pHs. The beta, delta, kappa, and omicron variants behave similarly at various pHs in the hot spot zone. Also, it is highlighted that the range from 493 to 506 position residue of Spike-S1 exhibits little fluctuation in all the systems examined.

The protein–protein interaction plays a crucial role in the biological system [167]. The interactions that are analyzed from the atomistic field are non-covalent interactions, which are weak interactions between the atoms of proteins; within these are electrostatic forces, hydrophobic effects, hydrogen bonds, π-effects, van der Waals forces, among others. Hence, we analyzed this intriguing complexation between ACE2 and Spike-S1 by analyzing hydrogen bonds and salt-bridges.

The hydrogen bonds are interactions at the boundary between chemical and non-covalent interactions. They occur between pairs of atoms only if one is a pro-



(a)

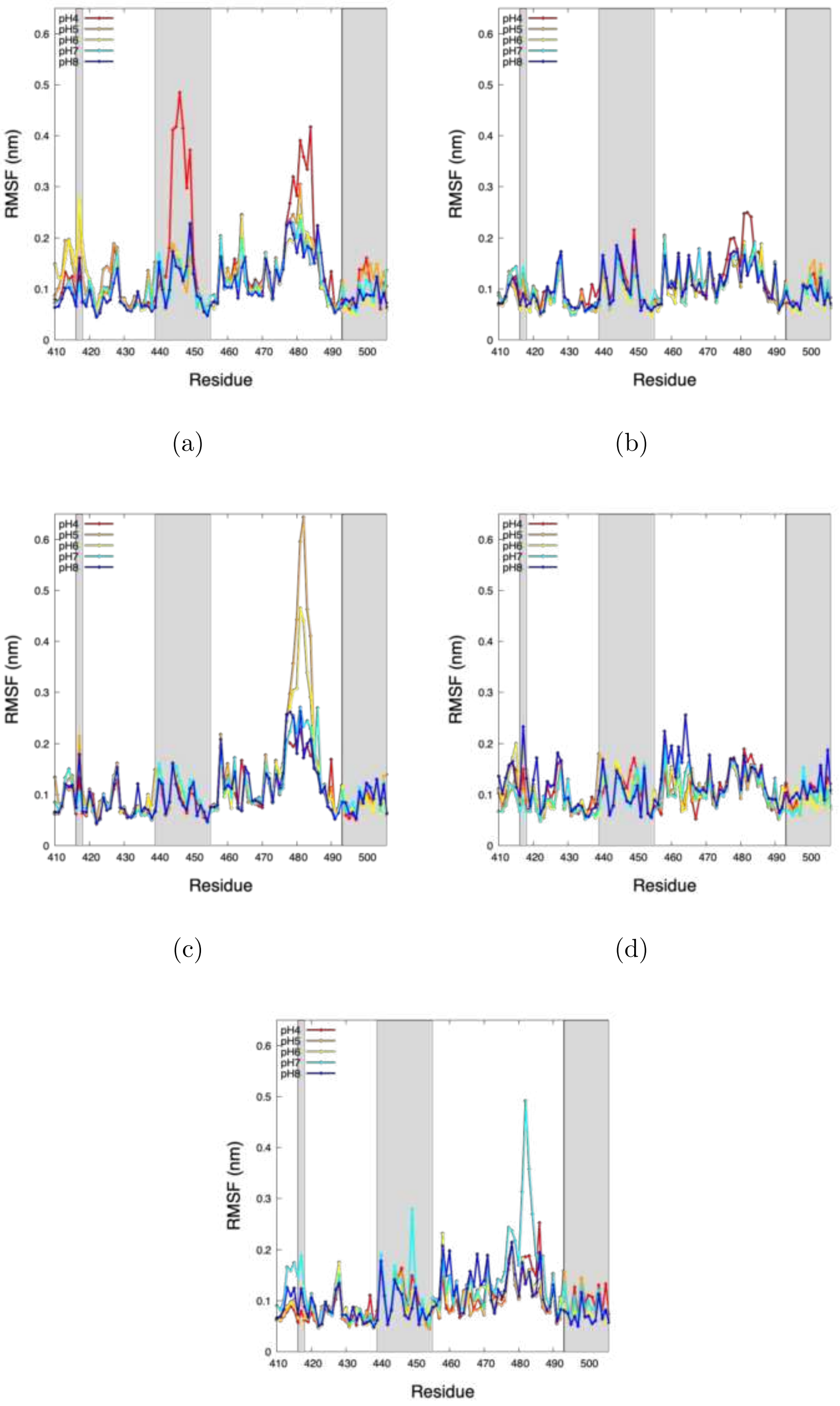
(b)

(c)

(d)

(e)

Figure 6.3: RMSD plot for the C-alpha backbone to the whole protein of systems Spike-S1 with ACE2 during the simulation. a.) pH4. b.) pH5. c.) pH6. d.) pH7. e.) pH8.



(a)

(b)

(c)

(d)

(e)

Figure 6.4: RMSF of mutated Spike-S1 at each pH of the last 5 ns of the MD trajectory, where we highlight the hot spot areas in colour gray. a.) alpha. b.) beta. c.) delta. d.) kappa. e.) omicron.

ton donor and the other is a proton acceptor. This type of interaction is often associated with the driving force of complexation, so it is important to explore the relative contributions of these effects and review differences between the analyzed mutations. On the other hand, salt-bridges in proteins are bonds between oppositely charged residues that are close enough to each other to experience electrostatic attraction. They contribute to the structure of the protein and to the specificity of the protein-protein interaction, where it is not necessary to increase the unfolding free energy of a protein.

Table 6.2 shows the number of hydrogen bonds generated during the molecular dynamics simulation time and the salt-bridges of the last frame. Hydrogen bonds at pH4 were higher in omicron > kappa > delta > alpha > beta. For pH5, the trend from highest to lowest was in omicron, kappa, delta, beta, and alpha. At pH6, omicron remains on the enzyme, followed by kappa, delta, beta, and alpha. At pH7, the trend increases from beta, alpha, delta, kappa, and omicron. Regarding pH8, omicron had the highest number of hydrogen bonds. This Table shows that omicron generally maintains more hydrogen bonds with ACE2 than other mutations. Another result that can be seen in the Table 6.2 is the number of salt-bridges in the last frame.

For the alpha, we have found that the salt-bridge is formed between two groups of opposite charge of Arg403(Spike-S1) and Glu37(ACE2), and the second salt-bridge by Lys417(Spike-S1) and Asp30(ACE2), estos salt-bridges are only observed at pH8. In the case of the beta, the formation of salt-bridge was observed at pH6 (Glu75(ACE2) with Lys484(Spike-S1)) and pH8 (Glu23(ACE2) with Lys458(Spike-S1)). For delta, the salt-bridges were shown at pH6, pH7 and pH8; in all cases, they were formed by Asp30(ACE2) with Lys417(Spike-S1). For the kappa, the presence of salt-bridges occur in all pH conditions, where at pH4

Table 6.2: The number of hydrogen Bonds and salt-bridge of the contacts and interacting residue of ACE2 and Spike-S1

	alpha		beta		delta		kappa		omicron	
	#Hbond	#SB	#Hbond	#SB	#Hbond	#SB	#Hbond	#SB	#Hbond	#SB
pH4	5	0	3	0	5	0	6	1	7	2
pH5	4	0	4	0	5	0	7	2	7	3
pH6	5	0	5	1	5	1	6	1	8	1
pH7	5	0	3	0	6	1	6	1	8	2
pH8	5	2	5	1	6	1	5	1	9	1

we find a salt-bridge formed by Glu23(ACE2) with Lys458 (Spike-S1). Kappa at pH5 presents two salt-bridges formed by Glu37(ACE2)-Arg403(Spike-S1) and Asp30(ACE2)-Lys417(Spike-S1), while for pH6, pH7, and pH8 only one salt-bridge formed by Asp30(ACE2)-Lys417(Spike-S1) is observed.

The findings for omicron were different; more salt-bridges were visible in all the pHs examined, emphasizing the union of Glu35(ACE2)-Arg493(Spike-S1). At pH4 and pH7, Glu35(ACE2)-Arg493(Spike-S1) and Asp38(ACE2)-Arg493(Spike-S1) created two salt-bridges. Three salt-bridges, Glu35(ACE2)-Arg493(Spike-S1), Asp38(ACE2)-Arg493(Spike-S1), and Glu23(ACE2)-Lys458(Spike-S1), are formed for omicron at pH5. The union of Glu35(ACE2)-Arg493(Spike-S1) for omicron at pH6 and pH8 only indicates the presence of one salt-bridge. In summary, the residues necessary for the union of ACE2 with Spike-S1 are often glutamic acids. As a result, the binding of these proteins will be improved if the Spike-S1 binding region contains basic residues. Interestingly, the omicron mutation displays altered residues in the ACE2 binding area compared to the other mutations of SARS-CoV-2.

6.3.4 Binding free-energy estimation by MM/GBSA

Based on the last 10 ns of MD simulation, the protein-protein MM/GBSA calculation was utilized to estimate the free energy of binding of Spike-S1-ACE2. This approach uses the continuous solution method to compute the binding energy while considering the polar desolvation term.

Figure 6.5 shows the average total binding energy between the different mutations. Here, the best binding energy occurs in omicron (blue colour bar) except at pH5, where the best energy occurs with kappa. In this Table, the energy values of omicron were -43.04 kcal/mol (pH4), -38.41 kcal/mol (pH5), -49.54 kcal/mol (pH6), -48.70 kcal/mol (pH7), and -55.29 kcal/mol (pH8). Additionally, here we report the Table 6.3 of the average value of each energy contribution. ΔEGB provides the best energy input at pH4 in all mutations, while ΔEEL produces the worst outcomes. This finding suggests that the ΔEEL can negatively impact the system due to the wide variety of titratable residues in their protonatable state.

Omicron has a much lower free energy of binding than other mutations in all mutations at pH4. Kappa has the highest interaction energy when we look at the pH5 values, followed by omicron, delta, beta, and alpha. From pH6 to pH8, the pattern in the various mutations is often similar, with ΔEEL contributing the most energy. At these pHs, it is also noted that omicron has the best interaction energy with -43.04 kcal/mol, -38.41 kcal/mol, -49.54 kcal/mol, -48.70 kcal/mol, and -55.29 kcal/mol for pH4, pH5, pH6, pH7, and pH8 respectively.

In order to determine the energy contribution of each amino acid involved in the binding to ACE2, the free energy decomposition analysis was also carried out. Remainders are displayed in our analysis from positions 417 to 506 of Spike-S1 (See Figure 6.6).

Table 6.3: MM/GBSA calculations and individual energy terms showed that omicron Spike-S1 had considerable binding free energy in each pH. All values are in kcal/mol.

	pH4	pH5	pH6	pH7	pH8
Spike-S1 alpha vs ACE2					
ΔVDWAALS	-71.79	-66.44	-78.98	-71.29	-71.67
ΔEEL	2095.15	-159.55	-579.89	-815.81	-690.60
ΔEGB	-2042.29	212.83	636.07	867.31	745.77
ΔESURF	-10.71	-9.49	-11.17	-10.31	-10.30
ΔGGAS	2023.36	-225.98	-658.87	-887.10	-762.28
ΔGSOLV	-2053.00	203.34	624.90	857.00	735.47
ΔTOTAL	-29.64	-22.64	-33.97	-30.10	-26.81
Spike-S1 beta vs ACE2					
ΔVDWAALS	-71.89	-69.60	-75.64	-70.48	-79.97
ΔEEL	2323.33	74.26	-604.46	-662.80	-830.58
ΔEGB	-2265.28	-19.80	663.56	722.61	892.62
ΔESURF	-10.23	-9.58	-10.61	-9.80	-10.95
ΔGGAS	2251.44	4.66	-680.10	-733.28	-910.55
ΔGSOLV	-2275.51	-29.38	652.95	712.81	881.67
ΔTOTAL	-24.07	-24.72	-27.15	-20.46	-28.88
Spike-S1 delta vs ACE2					
ΔVDWAALS	-70.78	-73.54	-74.61	-72.32	-75.24
ΔEEL	1510.69	340.47	-862.38	-617.88	-705.62
ΔEGB	-1459.48	-290.25	912.38	670.23	760.94
ΔESURF	-9.99	-10.58	-10.79	-10.77	-11.03
ΔGGAS	1439.91	266.94	-936.99	-690.20	-780.86
ΔGSOLV	-1469.47	-300.83	901.59	659.46	749.90
ΔTOTAL	-29.55	-33.89	-35.40	-30.74	-30.96
Spike-S1 kappa vs ACE2					
ΔVDWAALS	-68.40	-78.71	-78.95	-79.32	-73.29
ΔEEL	2180.61	42.04	-579.02	-1052.53	-1125.04
ΔEGB	-2130.60	4.91	627.43	1103.97	1182.16
ΔESURF	-10.33	-11.85	-11.74	-12.01	-10.76
ΔGGAS	2112.21	-36.66	-657.96	-1131.85	-1198.33
ΔGSOLV	-2140.93	-6.94	615.69	1091.96	1171.40
ΔTOTAL	-28.72	-43.61	-42.27	-39.89	-26.93
Spike-S1 omicron vs ACE2					
ΔVDWAALS	-67.05	-72.77	-81.02	-82.95	-81.15
ΔEEL	2931.89	19.20	-776.23	-1132.82	-1357.34
ΔEGB	-2897.62	26.14	819.99	1179.65	1395.69
ΔESURF	-10.25	-10.99	-12.28	-12.58	-12.50
ΔGGAS	2864.83	-53.57	-857.25	-1215.77	-1438.48
ΔGSOLV	-2907.87	15.16	807.71	1167.07	1383.19
ΔTOTAL	-43.04	-38.41	-49.54	-48.70	-55.29

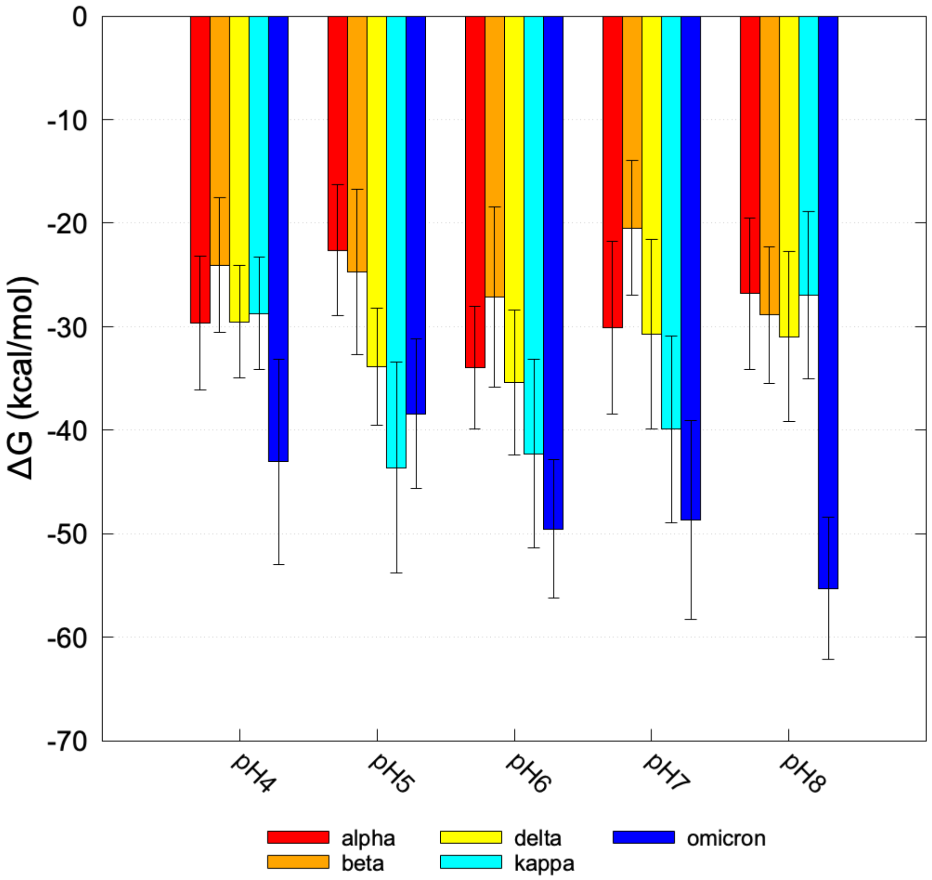


Figure 6.5: The MM/GBSA calculations for Spike-S1-ACE2 of the last ten ns reveal the average total binding free energy in kcal/mol. Where the blue bar shows the total average binding free energy of omicron at pH 4, pH6, pH7, and pH8.

Figure 6.6a depicts the energy contribution in the alpha variant, where the major residues involved in ACE2 binding were Lys417, Tyr453, Leu455, Gln493, Gly496, Gln498, Thr500, Tyr501, Gly502, Val503, and Tyr505. The most notable residues in the beta variant (Figure 6.6b) were Leu455, Gln493, Gln498, Thr500, Tyr501, Gly502, Val503, and Tyr505, with Tyr505 having the highest energy contribution.

On the other hand, the analysis of the energetic contribution in the delta variant

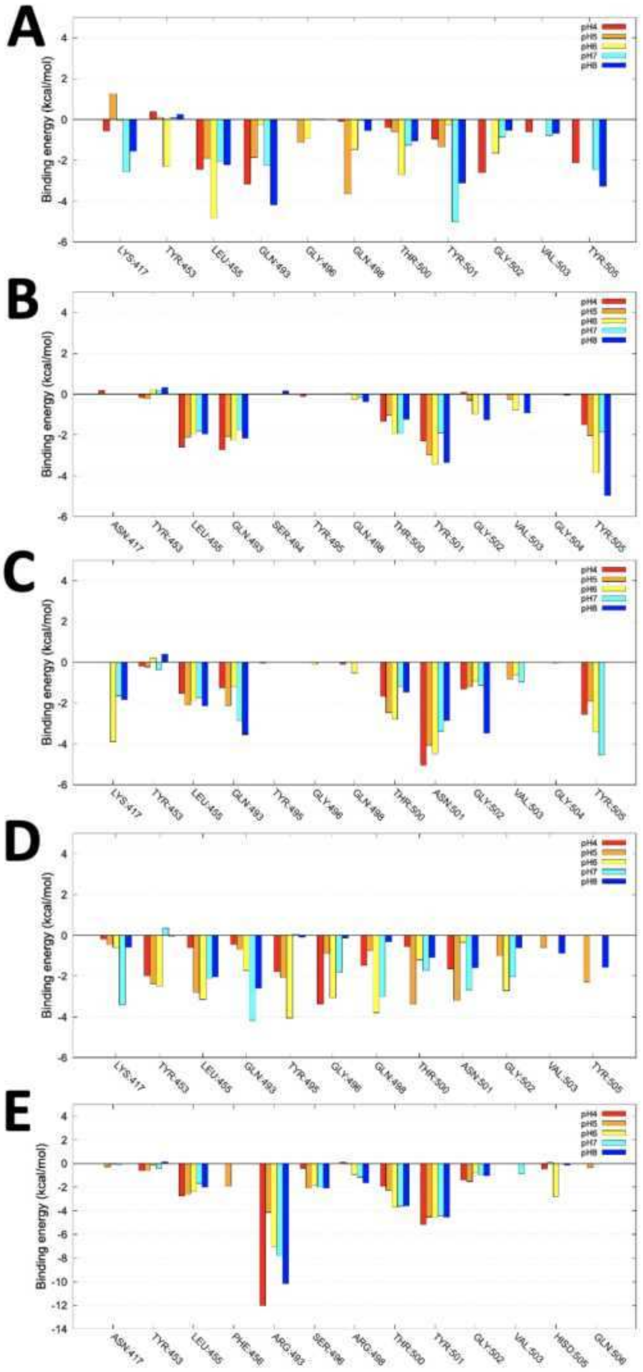


Figure 6.6: The residue decomposition plot MM/GBSA of the Spike-S1 hot spots representing the binding free energy contribution per residues. a.) alpha. b.) beta. c.) delta. d.) kappa. e.) omicron.

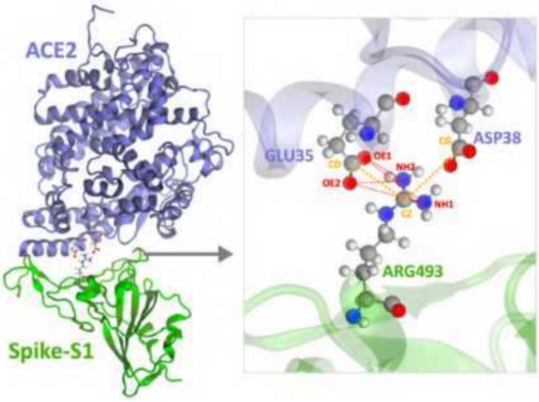


Figure 6.7: The best binding energy between omicron Spike-S1 and ACE2 at pH8 is represented. The Glu35 and Asp38 residues from ACE2 are highlighted against the Arg493 residue from Spike-S1 omicron.

shows us that the residues Lys417, Leu455, Gln493, Thr500, Asn501, Gly502, Val503 and Tyr505 were the ones that had a strong interaction with ACE2, where it is observed that Asn501 and Tyr505 had the highest energy (See Figure 6.6c).

The residues Lys417, Tyr453, Leu455, Gln493, Tyr495, Gly496, Gln498, and Tyr505 all exhibit an energetic contribution in the kappa variant (See Figure 6.6d). In Figure 6.6e, the ARG493 residue of the omicron variant made a significant energetic contribution above -4 kcal/mol at all pH levels. Therefore, we believe that the union of omicron Spike-S1 to ACE is greatly impacted by this mutation.

In light of this significant energy contribution of the omicron variant, a study of the principal residues that interact with ACE2 was conducted. We can see that Glu35 and Asp38 are seen interacting with Arg493 for pH values of 4, 5, and 7. While, Arg493 and Glu35 are seen to interact for pH values of 6 and 8. The result indicates that Arg493 has a strong affinity for Glu35 and Asp38 of ACE2 at pH8; this representation is observed in Figure 6.7, where the formation of hydrogen bonds between the carboxyl groups of the ACE2 residues and the amino group of omicron is shown.

6.4 Conclusion

SARS-CoV-2 continues to be one of the world's challenges, mainly caused by the several varieties that develop as the virus develops resistance. The omicron variation was one of the variants that attained substantial concern over the past two years. In this investigation, we were able to comprehend the causes of the apparent differences in the Spike-S1 therapeutic target of this variation. Alpha, beta, delta, kappa, and omicron were the five varieties whose linear structures were briefly compared. In the hot spot of the omicron variant, we discovered about five residues that underwent mutation and became titratable residues. The charge distribution was examined for pH values of 4, 5, 6, 7, and 8 to see if the pH's impact may enhance the binding of ACE2 to Spike-S1. The net total charge that we discover in the various complexes reveals a net positive total charge for pHs lower than pH6, and a net negative total charge for pHs higher than this pH. The molecular dynamics stability study was performed in 50 ns with the various charge distributions for a certain pH. The union of the omicron Spike-S1 with the ACE2 at the various pHs produced the most remarkable complex. The binding free energy estimation by MM/GBSA shows that the Arg493 from omicron contributes the maximum energy with the Glu35 and Asp38 from ACE2 in all pHs. Overall, this study clarifies the significance of taking protonation/deprotonation states into account for macromolecular structures at a certain pH and how it may aid in developing future drugs for COVID-19 patients.

Chapter 7

Unveiling the Phenylalanine Coaggregation Mechanism for a Deep Understanding of Phenylketonuria Disease

Abnormal accumulation of the amino acid phenylalanine is an essential cause of the metabolic disorder phenylketonuria (PKU). In an effort to understand the mechanism of phenylalanine coaggregation, it was important to analyze the self-assembly of phenylalanine. In this work, the mechanisms associated with the coaggregation of phenylalanine at physiological temperature (309.65 K) were analyzed using all-atom molecular dynamics. The results show that the phenylalanine monomers are organized to create a hydrophilic nucleus made of zwitterionic terminals, where it is suggested that the phenylalanine aggregation pathway starts in a ladder-like structure and from 200 phenylalanine monomers, the zig-zag conformation becomes present. Our finding demonstrates the influence of non-covalent interaction as hydrogen bonding and π-stacking of phenylalanine monomers for self-assembly formation.

7.1 Introduction

In 1938 the biochemist Asbjorn Folling's observations on mentally disabled children led him to conclude that this was due to a defect in phenylalanine metabolism. Later this disease has attributed the name phenylketonuria. Phenylketonuria (PKU) is mainly due to mutations in the enzyme phenylalanine hydroxylase (PAH), which is responsible for metabolizing L-phenylalanine (PHE) into L-tyrosine [168, 169, 170, 171, 172]. The problem with the non-processing of phenylalanine is that it is impossible to generate essential amino acids such as tyrosine, a dopamine precursor. The high blood and brain PHE concentrations can impair neuropsychological function through several mechanisms [173].

The PHE self-assembly is linked to amyloid formation toxicity in PKU, [174] because this type of aromatic amino acid shows an intrinsic ability in self-assembly [175, 176]. The self-assembly of various biological molecules is generated by external factors [177]. Further, the amino acid self-assembly has demonstrated the development of clearly defined, well-ordered self-assembled architects [178].

Generally, the self-assembly mechanism is governed by specific non-covalent interactions facilitated by hydrophobic, electrostatic interaction, aromatic π-π interaction, van der Waals, and hydrogen bonding. [179, 180, 181, 182] These interactions play a key role in the formation of various nanoscopic physical morphologies, for which it is necessary to gain a better understanding of these interactions and their role during the bottom-up self-assembly process [183]. The non-covalent interaction analysis in biological systems is fundamental to rationalising the structure and getting insights into the functions and dynamics of macromolecules [184].

Computer simulations increasingly provide us with exciting approximations of the behaviour of molecules of biological importance. Analyzing the self-assembly of aromatic amino acids such as tyrosine, tryptophan and phenylalanine at different temperatures by molecular dynamics allowed exploring the self-assembly that allows the formation of fibril-like aggregates [185, 186]. Another study by Thanh et al. [187] examined how phenylalanine oligomers self-assembled under various pH conditions. Unstable interactions drive phenylalanine monomers at high pH, while aggregation and compactness are more pronounced in the systems at neutral pH.

Considering the importance of phenylalanine fibres in generating aggregates and the limited information on the mechanisms that lead to their toxicity, we seek to analyze phenylalanine fibrils in solution and how it drives the aggregation of alanine amino acids. For this, we use computational strategies of molecular dynamics simulations with different concentrations of phenylalanine in its zwitterionic form at 309.65 K.

7.2 Computational Methods

Molecular Dynamics (MD) simulations have been performed with the Gromacs (Groningen Machine for Chemical Simulations) v. 2019.1 software [23] and OPLSAA [188] force field. The periodic boundary conditions (PBC) [116] to minimize edge effects in a finite system were considered. For each system, the protein was located in the centre of a cubic box with a minimum distance of 1.5 nm to the faces of the simulation box. The box system was solvated with TIP5P [117] water model. As well as, the total charge of each microstate was neutralized with Cl^- or Na^+ to attain equilibration.

Energy minimization was carried out using the steepest-descent algorithm with 200000 simulation steps. To equilibrate the system, 10 ns MD simulations were carried out in the canonical ensemble NVT with position restraint and the temperature was regulated with the V-rescale thermostat at 309.15 K. To determine the average properties of the systems, an MD simulation was performed in isothermal-isobaric ensemble with the Parrinello-Rahman barostat with a reference pressure of 1 bar, V-rescale thermostat with 309.15 K and a simulation length of 500ns.

To further corroborate our findings, the RMSD (Root-Mean Squared Deviation), SASA (Solvent Accessible Surface Area), and RDF (Radial Distribution Function) analysis were calculated by Gromacs tools. We also analyze the non-covalent interaction (NCI) using the Multiwfn program [189] to reveal intra and intermolecular interaction between phenylalanine monomers. Finally, the plots were made in Gnuplot v. 5.4 [120] and the manuscript images were created and rendered by Visual Molecular Dynamics (VMD) [30] software.

7.3 Results

7.3.1 Molecular Dynamics Simulation Analysis

This research evaluated the zwitterionic phenylalanine model during 500 ns of trajectory. Samples in water at 100, 200, 300, 400, and 500 monomers of phenylalanine monomers were used. During the equilibration phase, the monomers quickly consolidated into aggregates (within 100 ns) after starting from random beginning positions and orientations. At this point, we analyzed the behaviour through the MD simulation time.

The analysis of the Root-mean-square deviation (RMSD) shows us the similarity in three-dimensional structures along the MD from its initial structural conformation to its final position. The variations during the MD simulation can be used to assess the stability of the aggregate systems concerning their conformation,

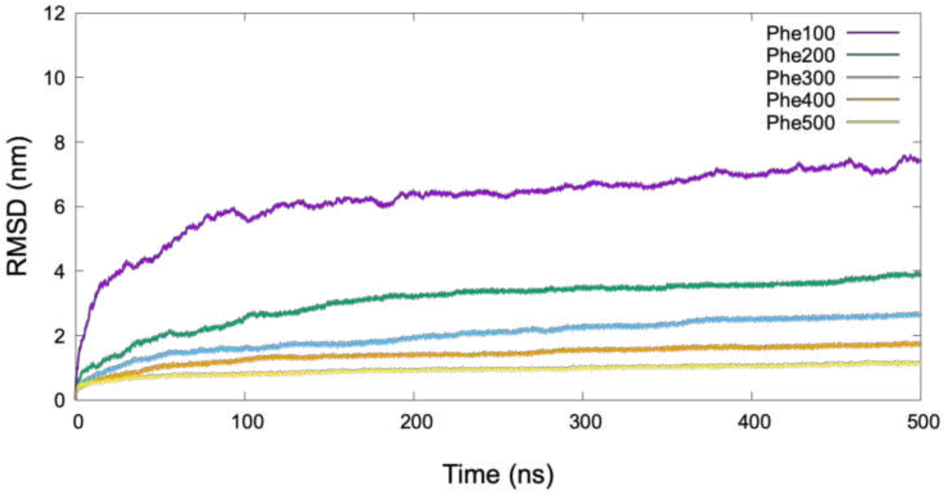


Figure 7.1: RMSD plot calculated for phenylanine monomers during the 500 ns of molecular dynamic simulation.

Table 7.1: Average values of RMSD and SASA of the last 100 ns of MD simulation for each system.

Concentration	RMSD (nm)	SASA *per residue* (nm^2)
Phe100	7.20±0.15	1.6±0.9
Phe200	3.70±0.11	1.1±0.7
Phe300	2.56±0.04	0.7±0.6
Phe400	1.68±0.03	0.6±0.5
Phe500	1.10±0.03	0.5±0.4

where the smallest deviations reflect the stability of the new conformer. Figure 7.1 shows the plot of the RMSD for each proposed system. The RMSD results of the different systems were obtained, where the lower concentrations showed a higher average RMSD than the other concentrations (the average values are shown in Table 7.1).

The Solvent Accessible Surface Area (SASA) is the surface area of a biomolecule that is accessible to a solvent. The average SASA value analyzed *per residue* is presented in Table 7.1. Where, Phe100 was higher than Phe200, Phe300, Phe400, and Phe500 SASA values. Figure 7.2 shows us the SASA *per residue* for each

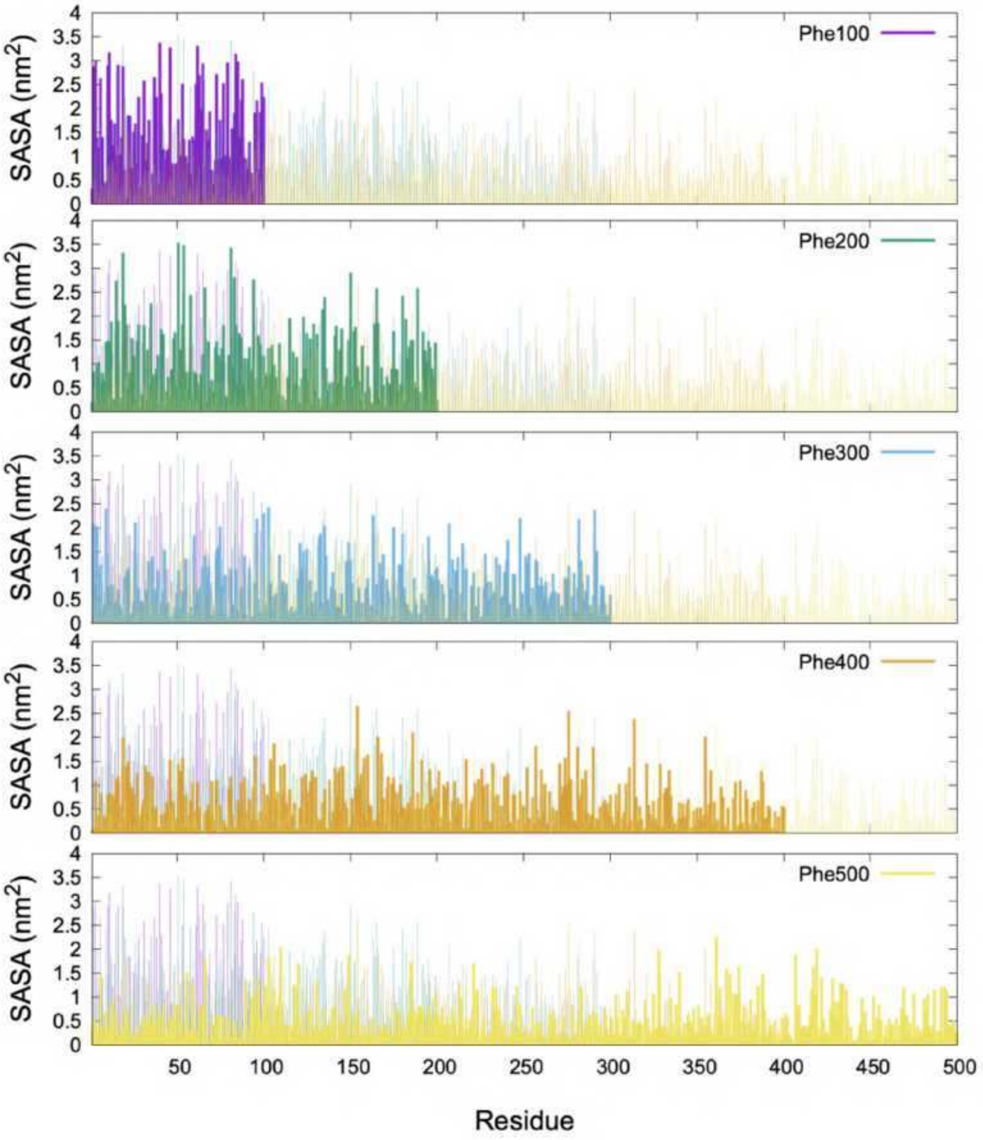


Figure 7.2: SASA *per residue* over the trajectory for each system. Phe100 is in solid purple colour, Phe200 is in solid green color, Phe300 is in solid sky blue color, Phe400 is in solid orange color, and Phe500 is in solid yellow color.

system analyzed during the last 100 ns. It should be noted that the highest SASA occurs in 100 and 200 phenylalanine monomers.

Phenylalanine is the only amino acid known to self-assemble into toxic fibrillar aggregates [187, 174]. Similar to experimental studies,[171, 190] our molecular dynamics results over the monomers organised to create a hydrophilic core made

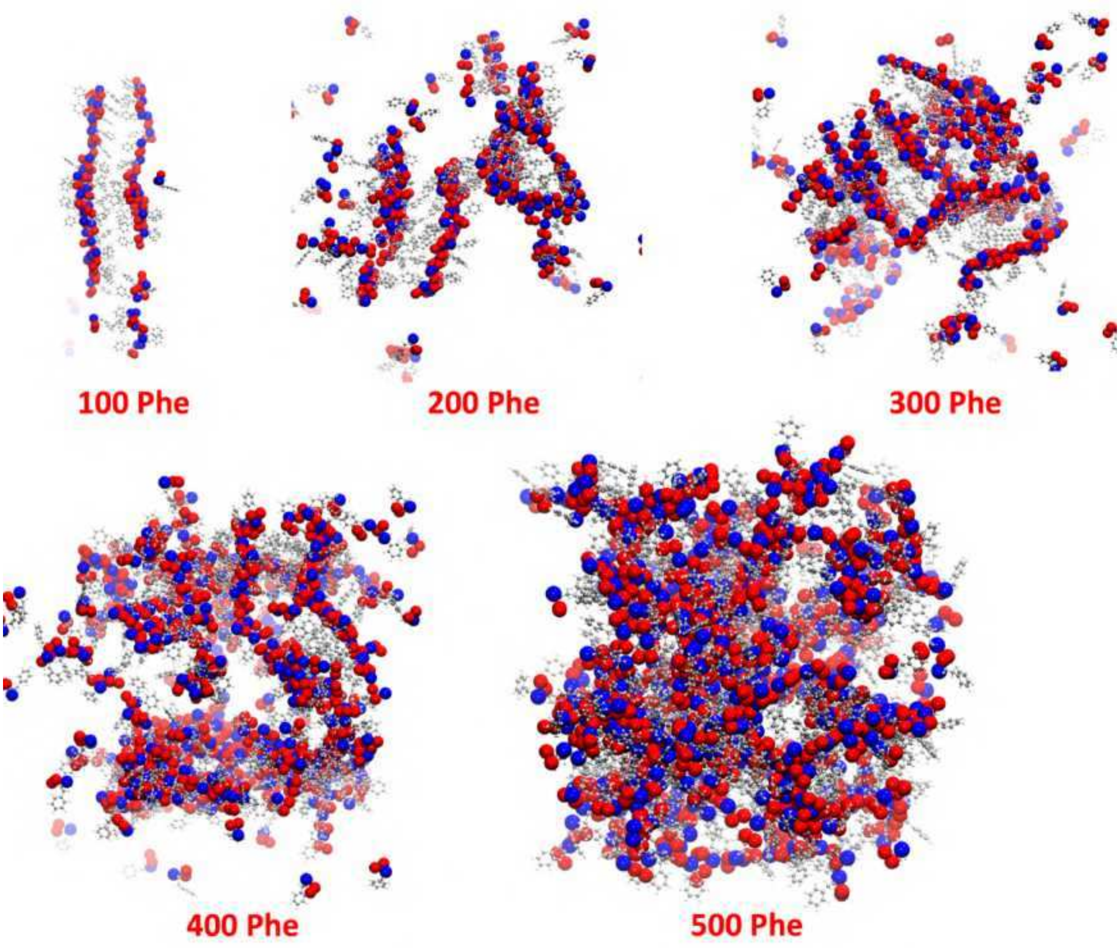


Figure 7.3: Last frame of the molecular dynamics simulation time for the different phenylalanine systems.

of zwitterionic terminals and expose the hydrophobic aromatic side chains to the outside are shown in Figure 7.3.

7.3.2 Non-covalent interaction analysis

Designing and comprehending supramolecular self-assembly processes requires a thorough grasp of intermolecular interactions. Electrostatic contacts, hydrogen bonds, π-cationic interactions, π-anionic interactions, π-π interactions, and van der Walls are examples of typical non-covalent interactions. This work identified

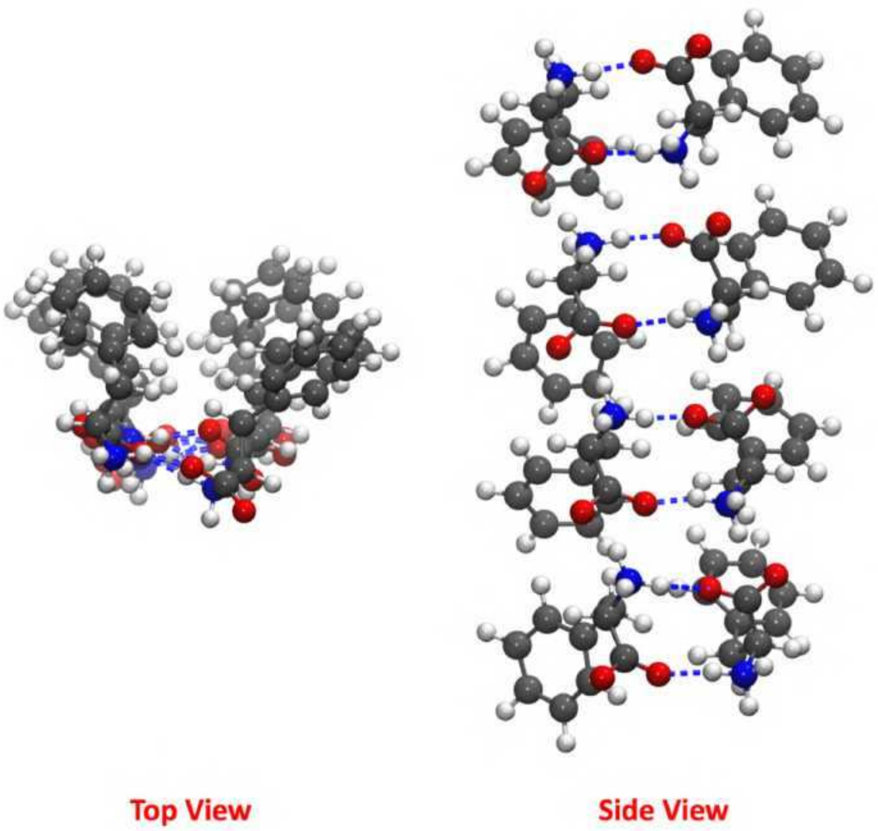


Figure 7.4: Representation of the top and side view of the analysis of the hydrogen bonds formed by two phenylalanine monomers, last frame taken from the study system of 100 phenylalanine units.

two types of non-covalent interactions; hydrogen bonding and the π-π interactions.

The hydrogen bond can be considered as a particular type of dipole-dipole interaction. The Lewis base claims this is a desirable interaction between a hydrogen bond donor and an acceptor. Hydrogen atoms bound to electronegative atoms like oxygen or nitrogen are the most common sources of hydrogen donors. Figure 7.4 shows the best model of the ladder-shaped structure formed by six phenylalanine monomers of the Phe100 system in which the amino and carboxyl groups joined to generate hydrogen bonds. This result shows us the most straightforward way for two phenylalanine monomers to initiate the formation of the first core.

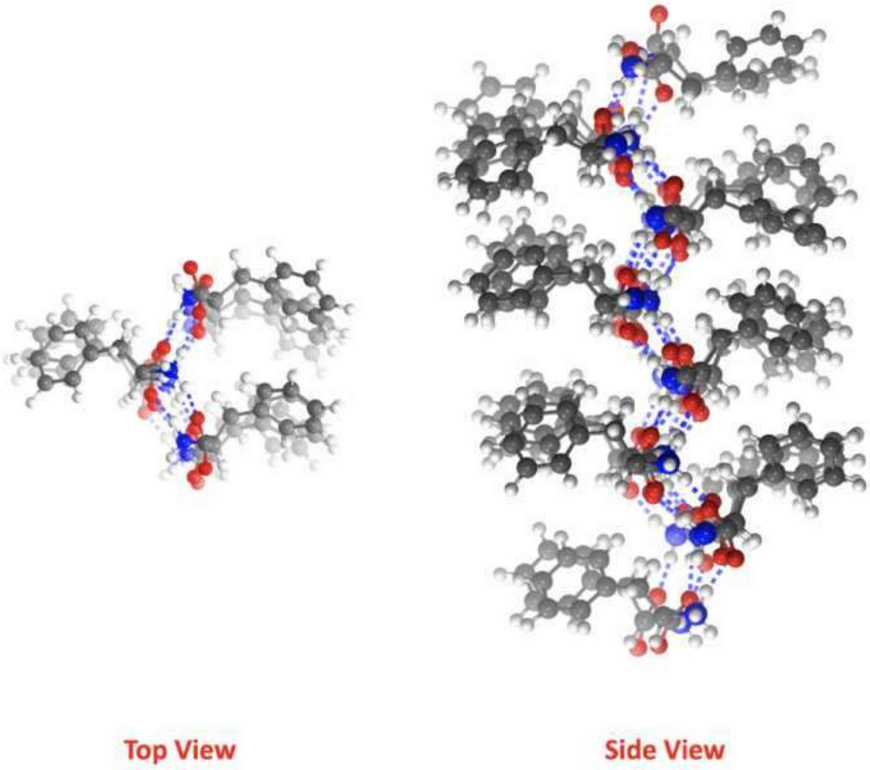


Figure 7.5: Molecular patterns detected by all-atom simulations from phenylalanine monomers. Representation of the top and side view of the last frame taken from the system of 200 phenylalanine units.

Figure 7.5 shows the system with 200 phenylalanine monomers, where a zigzag pattern between the monomers is detected. Like Phe100, the hydrogen bonds mediate this conformation between each phenylalanine monomer's carboxyl and amino groups. This stacking allows the formation of a fibre.

Additionally, we captured the last frame of the system with 300 phenylalanine monomers. Here we find that the fibre is stable, and the union between the phenylalanine monomers is given by hydrogen bonds formed by the zwitterionic endings. This representation is observed in the Figure 7.6 where the green box focuses on the hydrogen bonds formed by the zwitterionic terminations.

Furthermore, another non-covalent interaction mediated by π-π interaction be-

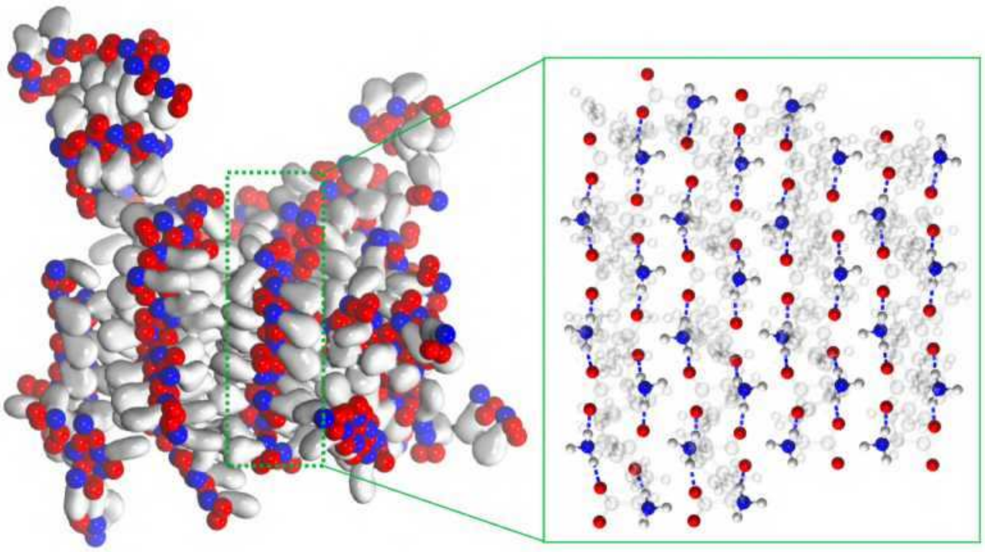


Figure 7.6: Conformation of the phenylalanine fibre taken from the system with 300 phenylalanine monomers; the green contour represents the zone of union between the monomers mediated by hydrogen bonds.

tween the aromatic rings of phenylalanines is observed. The π-π interactions are weak between two π systems; usually, one π is deficient, and the other π is surplus. These interactions combine electrostatic forces, hydrophobic effects, dipole-dipole or charge transfer between an electron donor and acceptor, repulsive forces between clouds or solvatophobic effects.

In the association between two aromatic rings, two types of π interactions can occur, the first is face-face interaction, where the aromatic rings are located parallel to each other, and the second is the T-shaped arrangement, where the aromatic rings are located perpendicularly creating a T-shaped geometry. In this work, we found the face-face interaction of the phenylalanine aromatic rings. Figure 7.7 illustrates this type of interaction between the aromatic rings in green.

In order to verify the interactions formed between the phenylalanine monomers, the analysis of the radial distribution function (RDF) was used. The g(r) between the alpha carbons of phenylalanine (CA) was investigated in order to clarify the

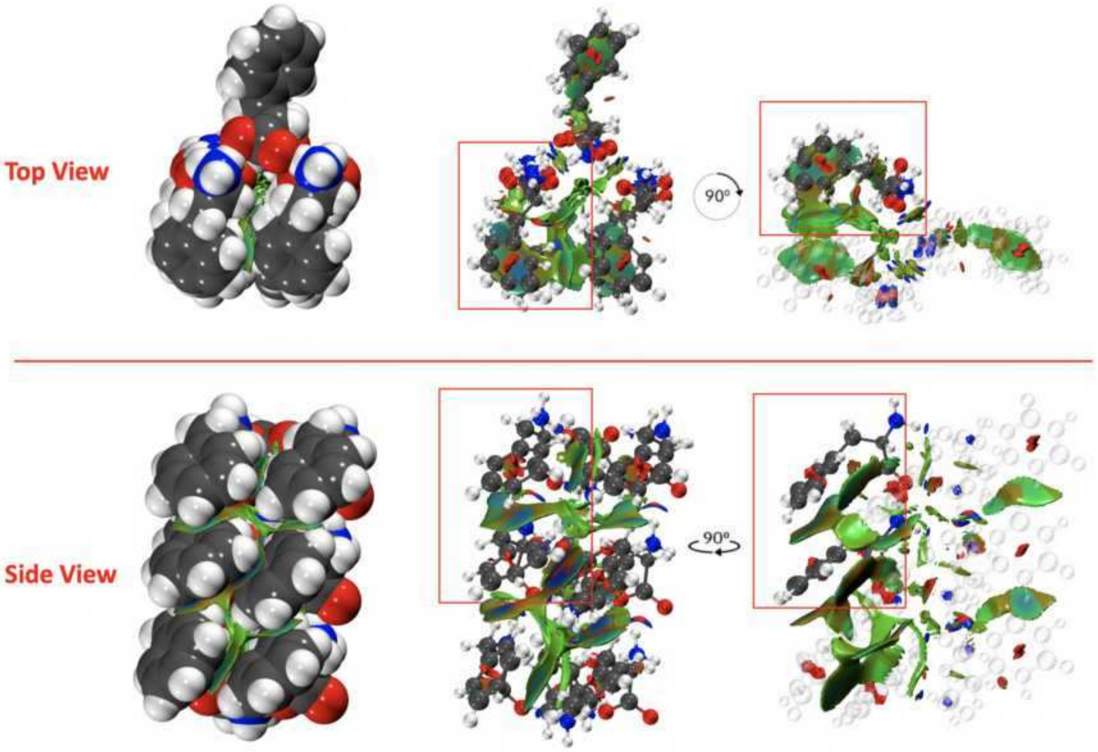


Figure 7.7: Non-covalent interaction π-π interactions.

difference in the coordination state of each zwitterionic termini.

Figure 7.8a shows the analysis of the g(r) of CA, in which we see two large symmetrical peaks. The first peak occurs at a distance of 0.5 nm and the second at 0.56 nm. This result indicates that CA atom interacts with two CA atoms (two phenylalanine monomers).

Additionally, the g(r) of CZ of the phenylalanine ring was determined. The distance of the largest peak is given at 0.5 nm for the case of the CZ atoms of the phenylalanine rings (see Figure 7.8b). This result indicates that phenylalanine rings are shown one above and one below of each phenylalanine monomer.

In this work, we can see that the final shape of the physical morphology can be regulated based on the intermolecular interactions involved in the self-assembly process of phenylalanine. Therefore, a high concentration of zwitterionic pheny-

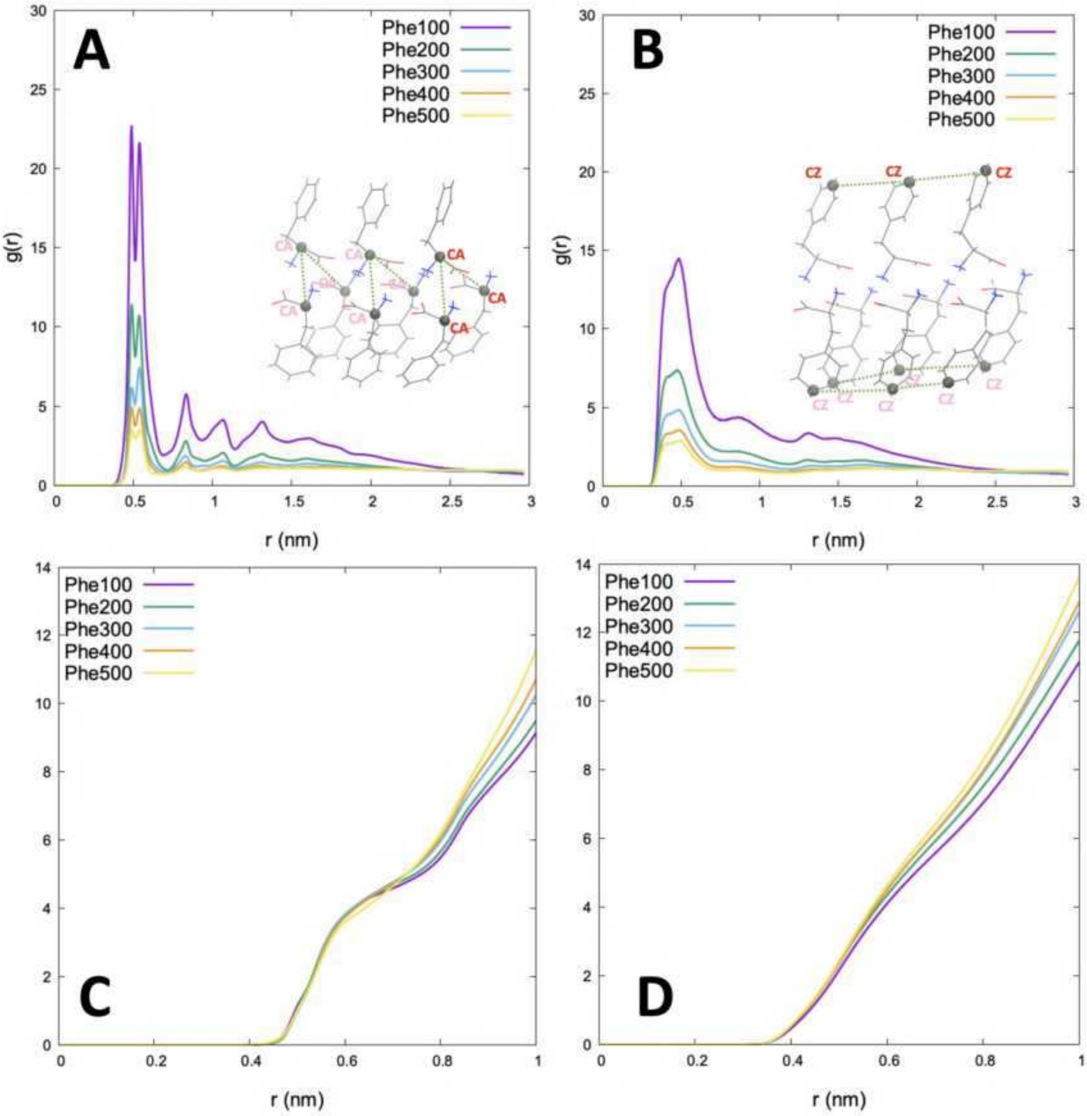


Figure 7.8: Radial distribution function of CA-CA of zwitterionic termini and CZ-CZ of phenylalanine aromatic ring. a.) g(r) of CA-CA. b.) g(r) of CZ-CZ. c.) g(r) integer of CA-CA. d.) g(r) integer of CZ-CZ

lalanine manages to form a fibre.

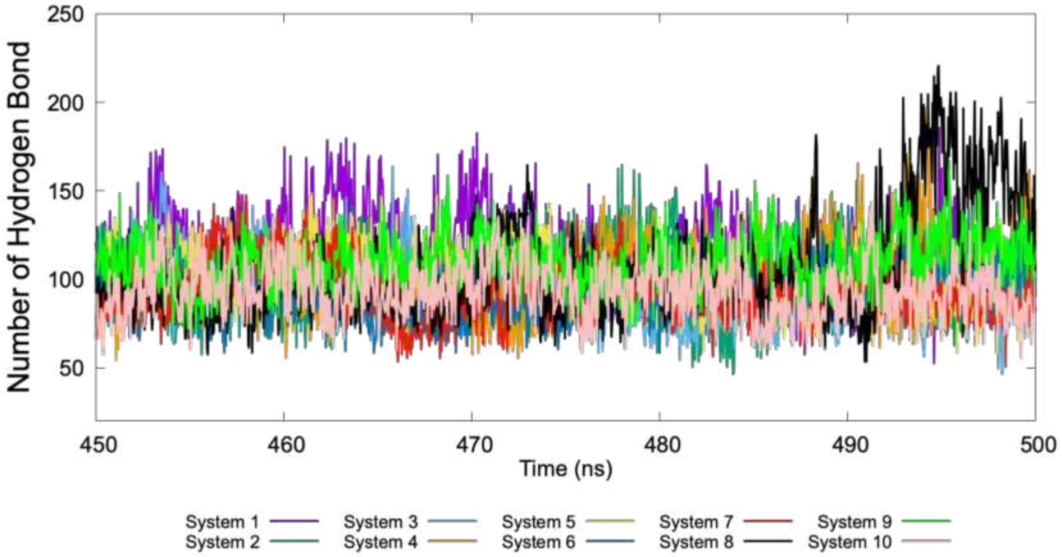


Figure 7.9: Number of hydrogen bonds of the phenylalanine fiber with respect to the alanine monomers analyzed in the last 50 ns.

7.3.3 Co-aggregation of alanine to phenylalanine fibres

Previous studies show the spontaneous aggregation of amino acids towards phenylalanine fibres[175]. Therefore, alanine aggregation was analysed by taking the phenylalanine fibres obtained from 200 monomers as a model. Ten systems were designed, where 100 alanine monomers were introduced into the phenylalanine fibres and evaluated by molecular dynamics simulation for 500 ns.

The results showed us that in ten repetitions of calculations, the behaviour of the alanines against the phenylalanine fibres was mediated by hydrogen bonds. For this reason, the number of hydrogen bonds between the alanine monomers against phenylalanine fibre is analysed. Figure 7.9 plots the number of hydrogen bonds during the last 50 ns of the simulation time. Here, the alanines incorporated into the system were attracted by the phenylalanine fibre, forming approximately 100 hydrogen bonds between them.

Figure 7.10 shows the final snapshot of phenylalanine fibres against the ala-

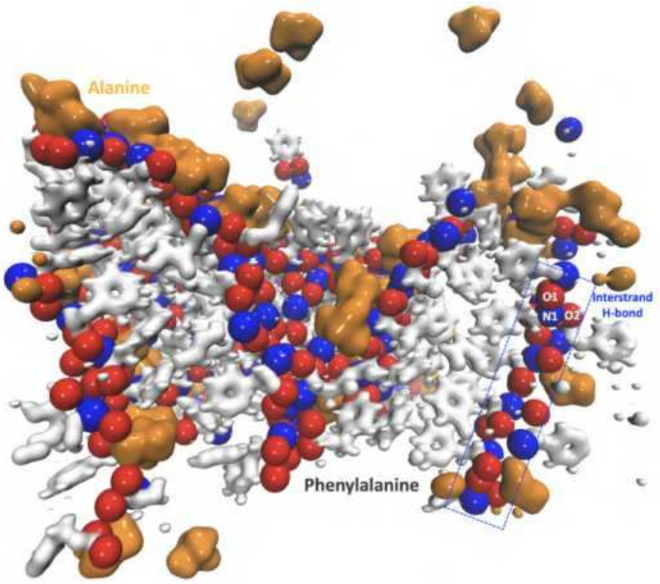


Figure 7.10: The final illustration of the co-aggregation of alanine in fibres of 200 phenylalanines, where the alanines are depicted in orange, the blue sphere represents the nitrogen atom of the amino group of the phenylalanines, and the red spheres indicate the oxygen atoms of the group phenylalanine carboxyl.

nine monomers. The co-aggregation of alanine (orange coloured surface) is shown attached to the terminal region of phenylalanine (blue and red coloured surface corresponds to the amino and carboxyl group, respectively) by hydrogen bonds. The aromatic rings of phenylalanine are shown as a white van der walls surface, linked by hydrophobic interactions.

The results show that the fibres allow the contact of the carboxyl and amino groups towards the lateral part of the fibre, which makes this area vulnerable to joining other residues.

7.4 Conclusion

In summary, this research analyses five systems with different amounts of zwitterionic phenylalanine monomers (100, 200, 300, 400 and 500 monomers) using all-atom molecular dynamics for a time of 500 ns. Likewise, it should be noted that we have focused on a system at 309.65 K with explicit water molecules. It was observed that when working with 100 phenylalanine units, these are joined in a ladder-like structure by hydrogen bonds. At the same time, the systems of 200 or more phenylalanine units presented a dominant aggregation in the form of stairs and zig-zag conformation, which seems to influence the formation of the fibres. Additionally, the role of non-covalent interactions such as hydrogen bonding and π-π interaction influence the stacking of these monomers. It is worth mentioning that the fibres formed are capable of attracting another type of amino acid. Finally, we use the fibre obtained from 200 phenylalanine monomers with 100 alanine monomers considering doing ten repetitions for a time of 500 ns. It was concluded that alanine could bind to phenylalanine fibre through the zwitterionic end zone, which could be related to the toxicity of phenylalanine fibres in biological systems.

Chapter 8

Influence of polar solvent mixtures in the conformation of polycarboxylates: Case study of citrate in ethylene glycol and diethylene glycol mixtures

Aspects of solvation dynamics play an essential role in the strong intramolecular hydrogen bonds between polar solvents and tricarboxylic acids. This study used atomistic simulations to explore sodium citrate as a tricarboxylic acid model about different quantities of two solvents (diethylene glycol and ethylene glycol). Explicit solvent models were used in all-atom molecular dynamics simulation. It was demonstrated that the citrate conformation in each of the solvents maintained a *gauche*$^{+}$ or *gauche*$^{-}$ orientation, whereas the dispersion of the Na^{+} around the citrate typically occupied an orientation close to its central carboxylate anion. In

parallel, more hydrogen bonds between citrate with pure diethylene glycol were observed over pure ethylene glycol. Citrate has a high dipole moment when combined with pure diethylene glycol, where numerous hydrogen bonding connections between both molecules cause this significant dipole moment. The shielding of the citrate increases with the high number of hydrogen bonds, which lowers the solvent's dielectric constant. In conclusion, these findings show the importance of the solvent against a tricarboxylic acid by atomistic simulations at a high constant temperature (473.15 K).

8.1 Introduction

Carboxylic acids are organic acids widely present in nature. The leading functional group is the carboxyl group (-COOH), and depending on the number of groups present in the structure, these are classified as dicarboxylic acid, tricarboxylic acid, and others. These carboxylic acids have a high boiling point and are weak acids in an aqueous solution, and they can interact intermolecularly with water and alcohols. Carboxylic acids wide range of applications including use in many synthetic reactions, for example as alternative carbon structures to build various derivatives of carboxylic acids as connecting sites [191]. For instance, synthetic substitutes for many halides and organometallic reagents can be also made using carboxylic acids.

One of the benefits of using carboxylic acids in the construction of nanoparticles (NPs) is that they act as stabilizing agents that limit the expansion of the particles. The sodium citrate ($Na_3(C_6H_5O_7)$) is a tricarboxylic acid of natural and synthetic origin, non-toxic, and eco-friendly chemical, forming biocompatible NPs [192] and is the most abundant form in the nature. The citrate's carboxyl groups attach to iron oxide and create a stable magnetised NP dispersion in an aqueous solution [193]. To effectively bind to many metallic NPs, sodium citrate must work in the presence of other solvents, which must be considered [194, 195, 196]. For example, the work developed by Wentao Wang et al. shows an experimental procedure of synthesis of magnetite NPs using mixed solvent system with diethylene glycol (DEG) and ethylene glycol (EG), obtaining a highly water-dispersible superparamagnetic NP of size-controlled [197]. The superparamagnetic properties of these NPs are broadly used for diagnosis, drug delivery, and therapy. Furthermore, it has been seen that DEG and EG behave as solvents and reducing agents, which affects the overall size of the NPs [198, 199, 200, 201, 202].

Computational calculations have been proof to be a good approach to study molecular solvation. However, when dealing with large or complicated molecules or when there are two or more functional groups close to one another, these approaches are not good enough because they exclude the effect of stereochemistry. As a result, molecular dynamics simulations can be used as an alternative to compute the molecular solubility in any mixture of solvents [203]. These analyses have the benefit of providing predictions based on the surrounding environment of the molecule. On the basis of an understanding of the effect of solvation in conformational properties in mixtures of DEG in solvents, investigations of DEG using molecular dynamics simulation methods were reported [204, 205]. The same trend has been noted for EG, which exhibits diverse conformational and structural behaviour [206, 207].

This work aimed to use all-atom molecular dynamics simulation techniques to examine the conformational properties of citrate in the presence of mixtures of organic solvents at a temperature of 473.15 K. This temperature has been taken into consideration for this study because it is the one employed in several experiments for the construction of magnetized nanoparticles, including ferrite nanoparticles [208, 209, 210] magnetite magnetic nanoparticles [211, 212], metallic nanoparticles [213, 214, 215], and others. During the molecular dynamics simulation period, the behaviour of the citrate population was examined, and it was discovered that *gauche* conformers dominate. Additionally, it was determined that the distribution of the molecular dipole moments and estimation of the dielectric constant of solvents.

8.2 Computational Methods

Six different systems of solvents were prepared: diethylene glycol with 200 molecules ($DEG_{100\%}$), ethylene glycol with 200 molecules ($EG_{100\%}$), diethylene glycol:ethylene glycol with 50:150 molecules ($DEG_{75\%}$), diethylene glycol:ethylene glycol with 100:100 molecules ($DEG_{50\%}$), diethylene glycol:ethylene glycol with 150:50 molecules ($DEG_{25\%}$), and water with 621 molecules ($WAT_{100\%}$). We employed PMEMD version of Amber 16 package [216]. Subsequently, we used an Antechamber [217] generator of topology files for use with the AmberTools to assign a set of point charges by semiempirical method AM1-BCC [218] for citrate, DEG, and EG molecules. Each system had one citrate molecule with a total net charge of -3 and three sodium atoms (Na^+). Additionally, we added organic solvents in the concentration proposed, in which the total net charge of diethylene glycol and ethylene glycol was equal to zero. For the case of citrate in water, the water model TIP3P was considered.

We applied four steps in molecular dynamics calculations. The first step was the energy minimization using 100 minimization cycles, switched from steepest descent to conjugate gradient. A cut-off of 8 Å and periodic boundary condition (PBC) [116] in all directions were considered. Additionally, the systems were heated from 0 K to 300 K during 0.5 ns in the canonical (NVT) ensemble. Then, the systems are equilibrated with molecular dynamics calculations in the isobaric-isothermal (NPT) ensemble with a reference pressure of 1 atm at 300 K, pressure relaxation time of 2.0 ps, and isotropic position scaling during 100 ns. The production MD for data collection was simulated in the canonical (NVT) ensemble with a cut-off distance of 8 Å during 100 ns at 473.15 K. The visualization and the graphs were plotted in Visual Molecular Dynamics v. 1.9.4 (VMD) [30] and

Gnuplot v. 5.4 packages, respectively.

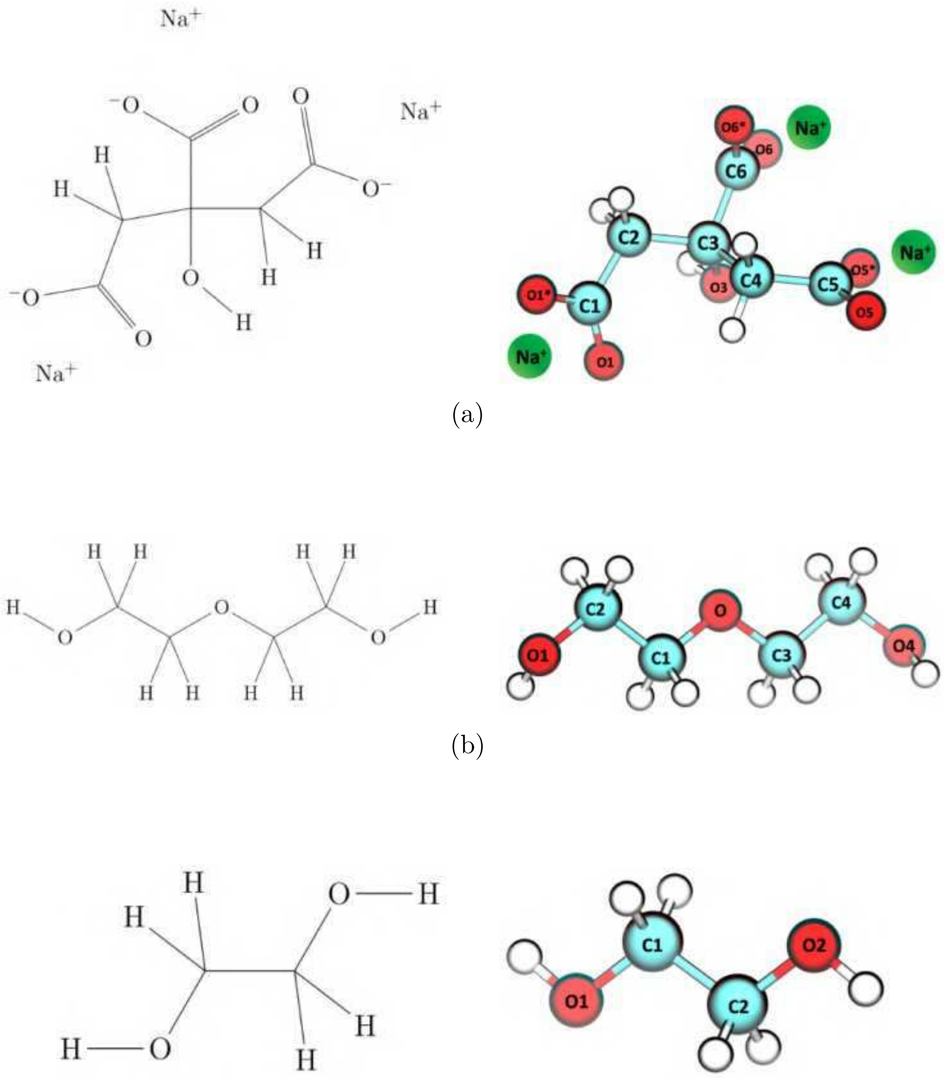
8.3 Results

8.3.1 Role of solvent on sodium citrate conformation

In general, carboxylic acids are completely miscible with water. However, different solvation properties could be expected in organic solvents, as dipole forces become less significant and dispersion forces become more significant. Therefore, sodium citrate's conformation may differ depending on the solvent and the proportion of solvent mixtures. Herein, we explored the conformation of sodium citrate in the solution of diethylene glycol (DEG), ethylene glycol (EG), mixtures of DEG and EG, and in pure water (WAT) using molecular dynamics (MD) simulation. For the preparation of the solution, a sodium citrate molecule was considered in a solution of $DEG_{100\%}$, $DEG_{75\%}/EG_{25\%}$, $DEG_{50\%}/EG_{50\%}$, $DEG_{25\%}/EG_{75\%}$, $EG_{100\%}$, and $WAT_{100\%}$. The chemical structures of sodium citrate, diethylene glycol, and ethylene glycol are depicted in 2D and 3D in Figure 8.1.

In this illustration, sodium citrate is shown to have three carboxylate anions (COO^-); the carbon atoms were given the designations C1, C5, and C6, while the oxygen atoms were given the designations O1, O1*, O5, O5*, O6, and O6*. The carbon atoms C4 and C2 are joined to the carboxyl group's carbons (C5 and C1), the central carbon (C3), and two hydrogen atoms. After that, the central carbon (C3) is linked to one hydroxyl molecule and carbon (C6) with two oxygens (O6 and O6*)

The citrate conformation was conveniently described via two dihedral angles along the main chain that connect the central carboxylate anion with the terminal



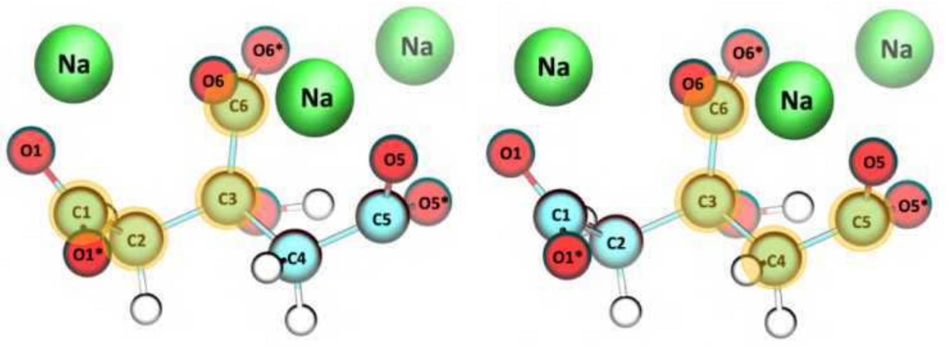
(a)

(b)

(c)

Figure 8.1: Pictorial drawing of 2D (left) and 3D (right) representation. a.) Sodium citrate. b.) Diethylene glycol. c.) Ethylene glycol.

carboxylate anions (see Figure 8.2). The measurement of the dihedral angles was one of the analyses that allowed us to assess the most consistent structural conformation throughout molecular dynamics simulation. To define a specific dihedral angle it is required to select four atoms. In this study, the sodium citrate dihedral



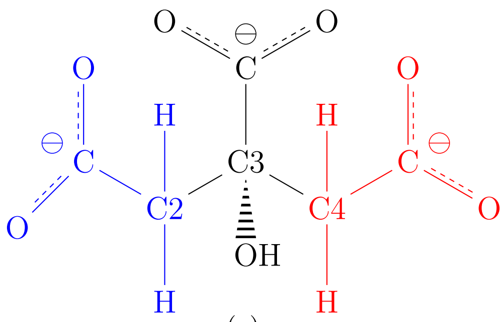
(a) **D1** (b) **D2**

Figure 8.2: Representation of dihedral angles of citrate analysed in the simulation. a.) Dihedral angle 1 around C2—C3 bond (D1). b.) Dihedral angle 2 around C3—C4 bond (D2).

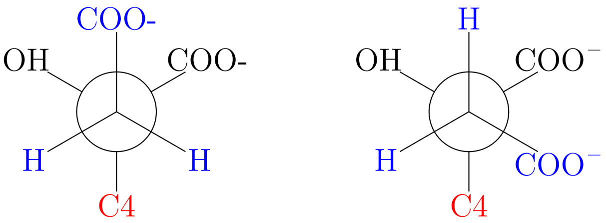
angles selected were formed by C1—C2—C3—C6 atoms (dihedral angle 1 = D1) and C5—C4—C3—C6 atoms (dihedral angle 2 = D2).

Citrate's molecular dynamics in various solvents are carried out to analyze the conformational properties over time in which the values of the sodium citrate dihedral angles (D1 and D2) served as a dynamic variable that defined the motion of the molecules during the MD simulation. In general, this study reports the dihedral angle distributions of citrate having the position of the maximum conformational probability at -50°and 50°. These dihedral angles correspond to the *gauche* conformation because there are close to its canonical value (±60°). The dihedral angles analyzed are well illustrated in the Newman projection (Figure8.3). It should be emphasized that citrate dihedral angle distribution is symmetrical, where the average value 50°and -50°corresponded to the $gauche^{+}$ and $gauche^{-}$ conformation, respectively. It differs from the structure of citric acid, which exhibits a second conformation known as *trans* conformation.[219]

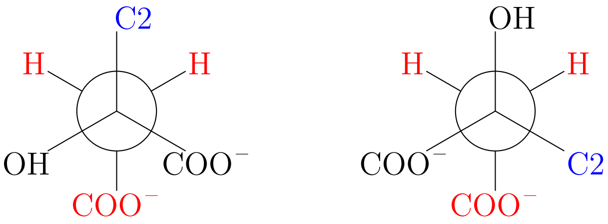
The systems with $\mathrm{DEG}_{100\%}$, $\mathrm{DEG}_{75\%}$, $\mathrm{DEG}_{50\%}$, and $\mathrm{WAT}_{100\%}$ had the position of the highest maximum in D1 and D2 dihedral angles at -50°and 50°, respectively.



(a)



(b) g^+ (C2–C3) (c) g^- (C2–C3)



(d) g^+ (C3–C4) (e) g^- (C3–C4)

Figure 8.3: Newman projections of two conformational minima obtained from the rotation of the C2–C3 and C3–C4 central bonds of sodium citrate.

While, $\text{DEG}_{25\%}$ and $\text{EG}_{100\%}$, the highest peak of D1 is shown at 50°and the highest peak of D2 is shown at -50°(see Figure 8.4).

We can see that the population of $\text{DEG}_{100\%}$, $\text{DEG}_{75\%}$, $\text{DEG}_{50\%}$, and the $\text{WAT}_{100\%}$ presented the tendency $gauche^- > gauche^+$ on D1 and $gauche^- < gauche^+$ on D2 when we examine each of the systems separately. While $\text{DEG}_{25\%}$ and $\text{EG}_{100\%}$ displayed a population distribution for D1 and D2 with a population of $gauche^{-} < gauche^{+}$ and $gauche^{-} > gauche^{+}$, respectively. In all systems studied, according to the MD simulation the citrate molecule prefers the $gauche^{+}$ or $gauche^{-}$ conformations over the *trans* conformation.

8.3.2 Principal Component Analysis

In this study, we acquire a sizable amount of data from MD simulations, which can be processed using different techniques like principal component analysis (PCA) that extract essential trajectory data. A mathematical approach known as PCA [220] reduces the dimensionality of data while retaining the majority of the data set's variation. The PCA analysis was based on Cartesian coordinate of the citrate analysis during simulation time [221], where the Cartesian coordinates describe the position of each sodium citrate atom in each frame. To achieve an optimal analysis, the center of mass of each citrate generated in the simulation steps was aligned in all frames. The results that we observe in Figure 8.5 show us the Cartesian PCA analysis of the entire trajectory in which we can visually evaluate the similarities and differences between each frame at 473.15 K.

Citrate's behaviour in relation to the DEG and EG solvents differs noticeably from that of citrate in water (Figure 8.5f). This happens as a result of a larger disturbance in the medium, which modifies the conformational properties of citrate. It should be noted that each point on the graphs represents a particular citrate conformation. On the graph, two conformations are similar if the corresponding points are close and conformations are different if the corresponding points are separated by a large distance.

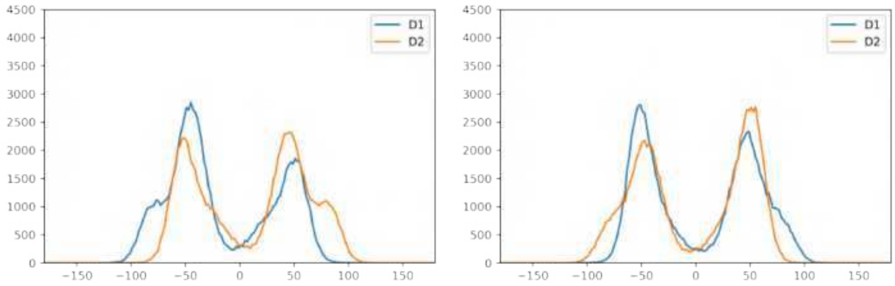
In all cases, it is observed that the color that represent the last frames of the simulations cover all the explored map (yellow dots). In light of the fact that the yellow dots cover the whole conformational space in the D1 and D2 planes rather of being distributed solely in a specific area, this result suggests that simulations were sufficiently long to sample all relevant conformations.

8.3.3 Dispersion of Na^+ counterions around citrate's carboxyl groups

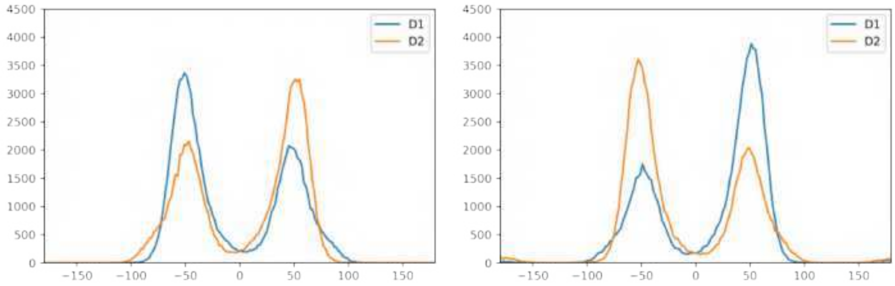
We computed the radial distribution function (RDF) of Na^+ counterions around the carboxylate anions and compared them. The analyzed oxygens were named according to their location in each carboxylate anions, in which the oxygens that were close to the C1 side carbon were designated as O1 and O1*, while for the other C5 side carbon their oxygen atoms were named as O5 and O5*, and as for the central carbon C6, its oxygen atoms had the denomination O6 and O6*. In general, the results of the distribution of the Na^+ counterions around the different types of oxygen revealed the presence of a unique peak, where the peaks were similar for the oxygen groups that corresponded to each carboxylate anions. Figure 8.6 displays the RDF of Na^+ counterions around the oxygens, where we have designated them as O1 (average value of O1 and O1*), O5 (average value of O5 and O5*), and O6 (average value of O6 and O6*).

The RDF of Na–O6, Na–O5, and Na–O1 showed the position of the highest peak at 2.275 Å. Figure 8.6a indicates a trend where the highest peak occurs with the solvent $DEG_{100\%}$, followed by $DEG_{75\%}$, $DEG_{50\%}$, $DEG_{25\%}$, $EG_{100\%}$, and $WAT_{100\%}$. Where $DEG_{100\%}$ shows integration of the radial distribution function at the high peak of 1.696 (Na–O6), 1.562 (Na–O5), and 1.585 (Na–O1), respectively. In the case of the oxygens found in the lateral carboxylate anions (O1 and O5) the behaviour is similar between them (Figure 8.6b and Figure 8.6c). It is interesting to note that the peak was moved to lower distances in water (red line), and that the RDFs in non-polar solvents and water differ considerably. As a result, we can see that the central carboxylate anion has a high value relative to the lateral carboxylate anions in terms of the number of counterions integrated up until the first minimum in the RDFs.

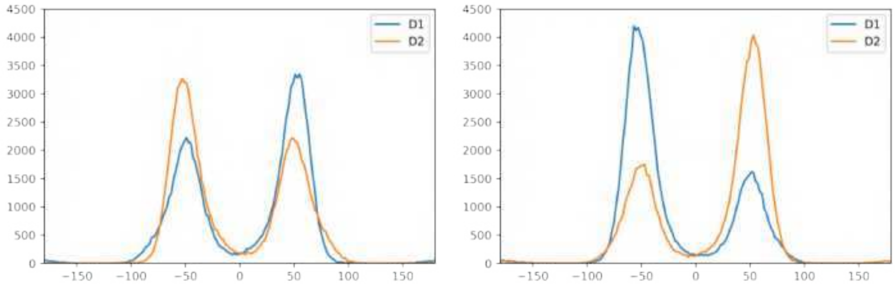
In addition, the RDF of Na^+ around the carbon atoms of the carboxylate anions of the citrate (C1, C5, and C6) was calculated, which behaved similarly to the oxygen atoms (the results can be seen in Figure D1 from Appendix D). In this result, two clear peaks exist in the first solvation shell at $r \approx 2.7$ Å and ≈ 3.2 Å. The first has a shorter distance peak and the second has a longer distance peak. As expected from the results obtained from the radial distribution functions with respect to the oxygen of citrate, the carboxylate anions of citrate with Na^+ in $DEG_{100\%}$ has the highest peaks compared to the other solvents.



(a) **DEG**$_{100\%}$ (b) **DEG**$_{75\%}$

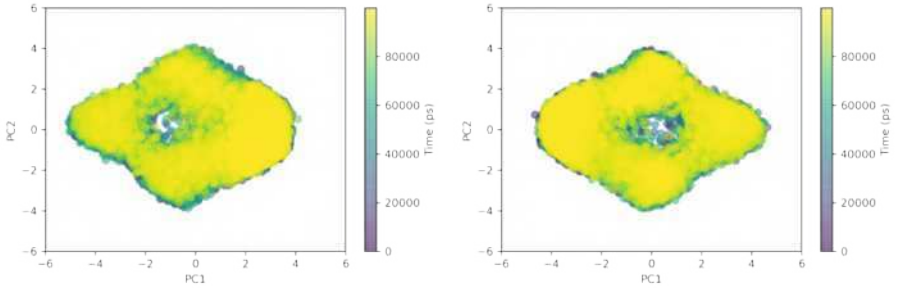


(c) **DEG**$_{50\%}$ (d) **DEG**$_{25\%}$

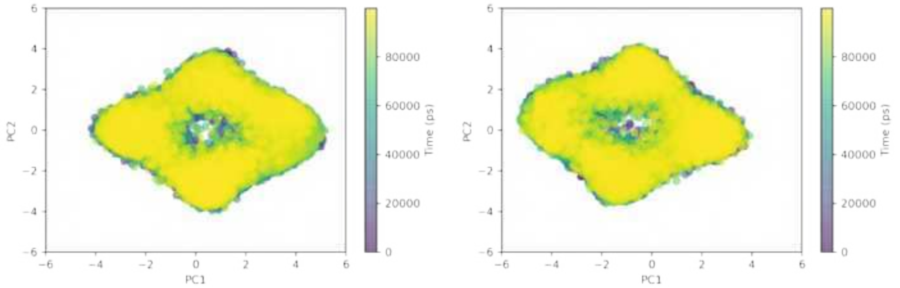


(e) **EG**$_{100\%}$ (f) **WAT**$_{100\%}$

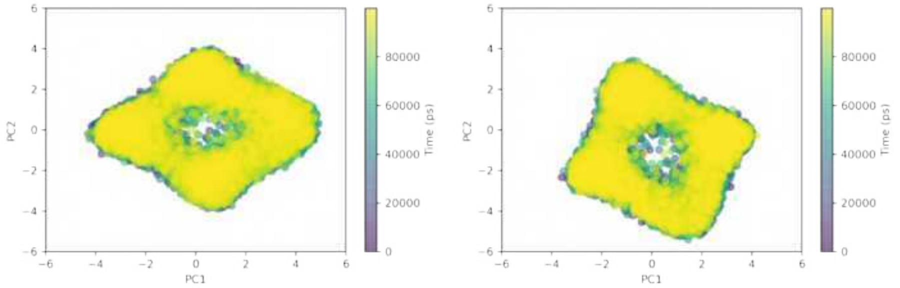
Figure 8.4: Population estimation histograms of citrate dihedral angle throughout MD simulation in different solvents at 473.15K. Blue line and orange line define the dihedral angle 1 (D1) and dihedral angle 2 (D2), respectively. a.) Diethylene glycol 100%. b.) Diethylene glycol 75% with Ethylene glycol 25%. c.) Diethylene glycol 50% with Ethylene glycol 50%. d.) Diethylene glycol 25% with Ethylene glycol 75%. e.) Ethylene glycol 100%. f.) Water 100%.



(a) **DEG**$_{100\%}$ (b) **DEG**$_{75\%}$

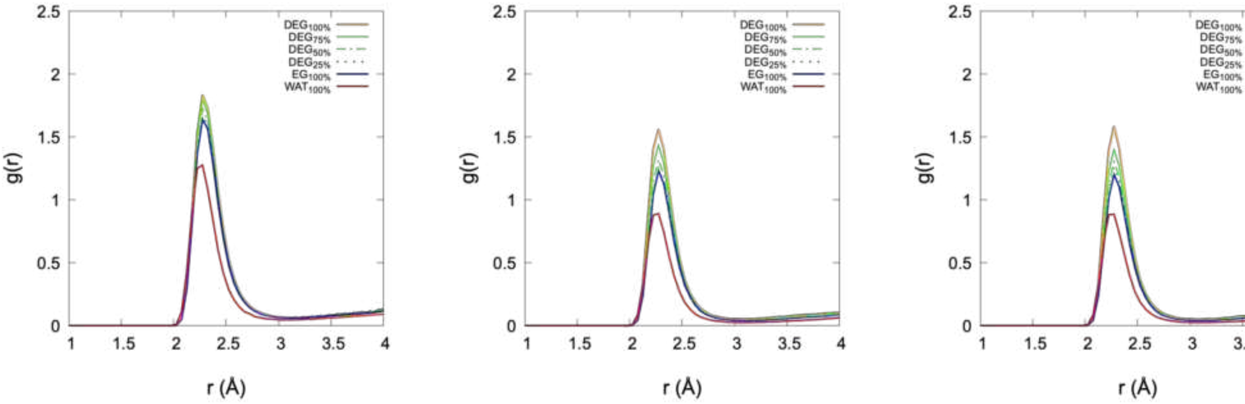


(c) **DEG**$_{50\%}$ (d) **DEG**$_{25\%}$

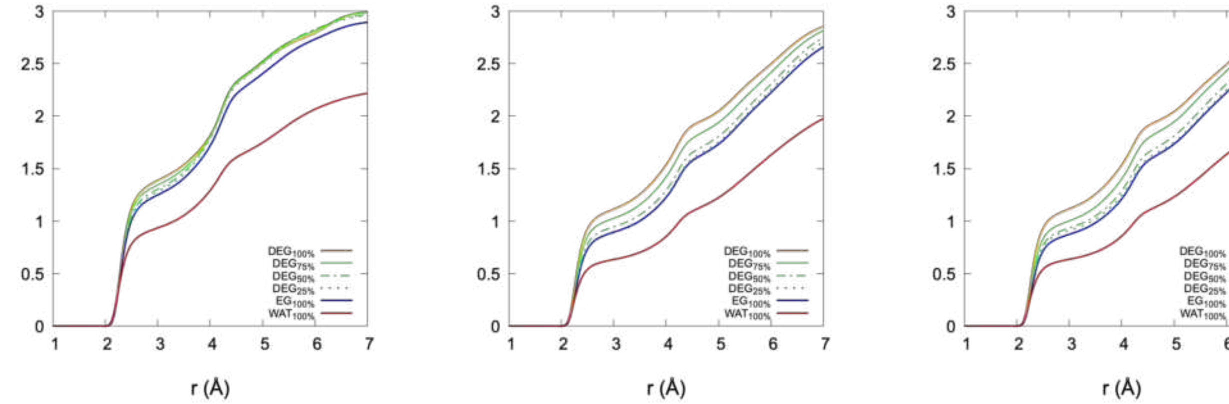


(e) **EG**$_{100\%}$ (f) **WAT**$_{100\%}$

Figure 8.5: Cartesian PCA analysis of citrate during the MD simulation at 473.15 K in which Citrate in the different systems is balanced, covering all conformations (dots in yellow colour). a.) Diethylene glycol 100%. b.) Diethylene glycol 75% with Ethylene glycol 25%. c.) Diethylene glycol 50% with Ethylene glycol 50%. d.) Diethylene glycol 25% with Ethylene glycol 75%. e.) Ethylene glycol 100%. f.) Water 100%.



(a) **O6** (b) **O5** (c) **O1**



(d) **O6** (e) **O5** (f) **O1**

Figure 8.6: Radial distribution function between Na^+ counterions with oxygen O6, O5, and O1 from ca of citrate at 473.15 K. a.), b.), c.) Plot of the radial distribution function, g(r) vs the distance r(Å). integral of radial distribution function values plot.

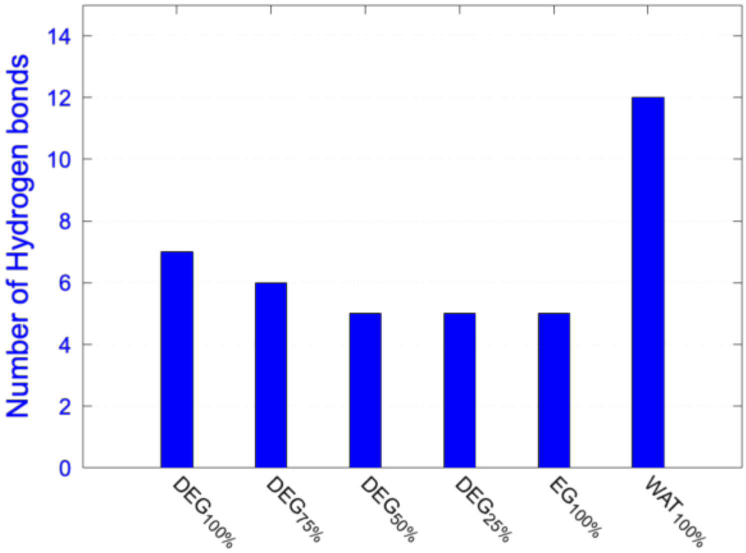


Figure 8.7: Representation of the average of the solvent-citrate number of hydrogen bonds in each solvent composition.

8.3.4 Hydrogen bond analysis

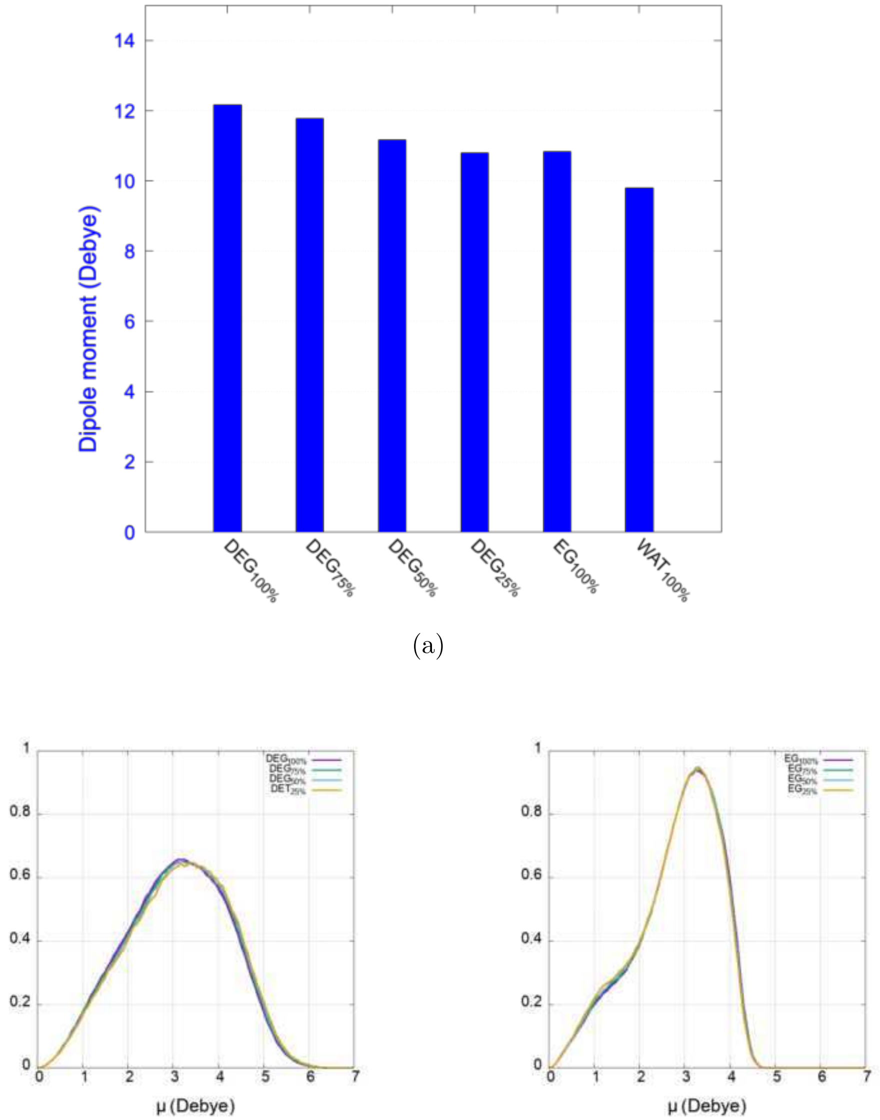
Measurements of the number of solute-solvent hydrogen bonds were used to analyze the effects of the various solvents on the sodium citrate molecule in the simulations. Hydrogen bonds (HB) have a dynamic process with frequent unions and breaks caused by the molecules' quick vibrational motions. It could be seen that the highest number of HB occurred among water-citrate, followed by $DEG_{100\%}$ and $DEG_{75\%}$. The systems that showed similar number of HB for $DEG_{50\%}$, $DEG_{25\%}$ and $EG_{100\%}$ (see Figure 8.7). In the DEG/EG mixtures, a decrease in the number of HB is observed with the proportion of DEG in the mixture. Regarding the MD simulation of citrate in water, a greater capacity of water molecules to establish HB with citrate is observed.

8.3.5 Dipole moment and dielectric constant analysis

Each system's dipole moments were determined from its trajectory and displayed as probability distributions. In this effort, we converted the Amber output files to Gromacs output files for the dipole moment calculus. Figure 8.8a illustrates the average dipole moments of citrate in each solvent. Citrate in $DEG_{100\%}$ had the highest value, followed by combinations of $DEG_{25\%}$, $DEG_{50\%}$, and $DEG_{75\%}$, then $EG_{100\%}$, and citrate in $WAT_{100\%}$ which has the lowest value. This is because the water molecules on citrate provide a stronger shielding effect, which results in a lower dipole moment.

On the other hand, when we review the dipole moment for each solvent, we notice that the DEG conformer tends to have a lower distribution dipole moment than the EG conformer (Figure 8.8b and Figure 8.8c). The value of the dipole moment for pure DEG is shown with a 3.1 Debye (Table 8.1). While the values obtained for $DEG_{75\%}$, $DEG_{50\%}$, and $DEG_{25\%}$ were similar with 3.1 Debye average value. The dipole moment distribution of systems with EG had values of 2.8 Debye, whereas the greatest value for the pure EG solvent was recorded at 2.9 Debye. The fact that the dipole moment of DEG or EG does not change as a function of the mixing ratio suggests that there haven't been any major conformational changes caused by the composition. The mean value for the magnitude of water dipole moments shows the high hydration of citrate at 473.15 K in the NVT simulation. In general, the citrate dipole moment is conserved regardless of the concentration of DEG or EG. In contrast to water, citrate's dipole moment is decreased by water molecules nearby.

The dielectric constant is yet another physicochemical characteristic of liquids that enables the examination of citrate-solvent interactions. In this study,



(a)

(b)

(c)

Figure 8.8: Graph plot of the dipole moment distribution. a.) The average value of dipole moment of citrate in different solvents. b.) Distribution of dipole moment of DEG at 100%, 75%, 50%, and 25%. c.) Distribution of dipole moment of EG at 100%, 75%, 50%, and 25%.

Table 8.1: The weighted mean value of probability distribution functions of the magnitude of the calculated solvents dipole moments.

Solvents	Dipole Moment (Debye)
$DEG_{100\%}$	3.1±1.1
$DEG_{75\%}$	3.1±1.1
$DEG_{50\%}$	3.1±1.1
$DEG_{25\%}$	3.1±1.1
$EG_{25\%}$	2.8±0.9
$EG_{50\%}$	2.8±0.9
$EG_{75\%}$	2.8±0.9
$EG_{100\%}$	2.9±0.9
$WAT_{100\%}$	3.7±0.9

molecular dynamics simulations were used to examine how the solvents' dielectric constants changed over time and in relation to their composition (see Figure 8.9). As the water molecules approach the citrate, we observe an increase in the relative permittivity ($WAT_{100\%}$). Moreover, comparing the results obtained for all studied systems we can observe that the value of dielectric constant of water was higher than that of other solvents. The tendency observed in DEG and EG for the dielectric constant was (from highest to lowest values): $EG_{100\%}$, $DEG_{25\%}$, $DEG_{50\%}$, $DEG_{75\%}$, and $DEG_{100\%}$.

In this inquiry, the values obtained from simulations are shown in Table 8.2, the average dielectric constant value displays a tendency from greater to smaller, considering the following order: water ($\epsilon = 45.5$), $EG_{100\%}$ (ϵ=15.3), $DEG_{25\%}$ (ϵ=13.4), $DEG_{50\%}$ (ϵ=11.9), $DEG_{75\%}$ (ϵ=10.7), and $DEG_{100\%}$ (ϵ=9.4). All these values have been obtained at 473.15 K at constant volume. As in the experiments, the volumes correspond to those the system would have at standard pressure and a temperature of 298.15 K. At the moment, the dielectric constant value of EG was reported from 243 K to 453 K,[222] where the dielectric constant reported by Lifanova et al.[223] was ϵ=18.37 at 453 K. While the value of the dielectric constant of diethylene

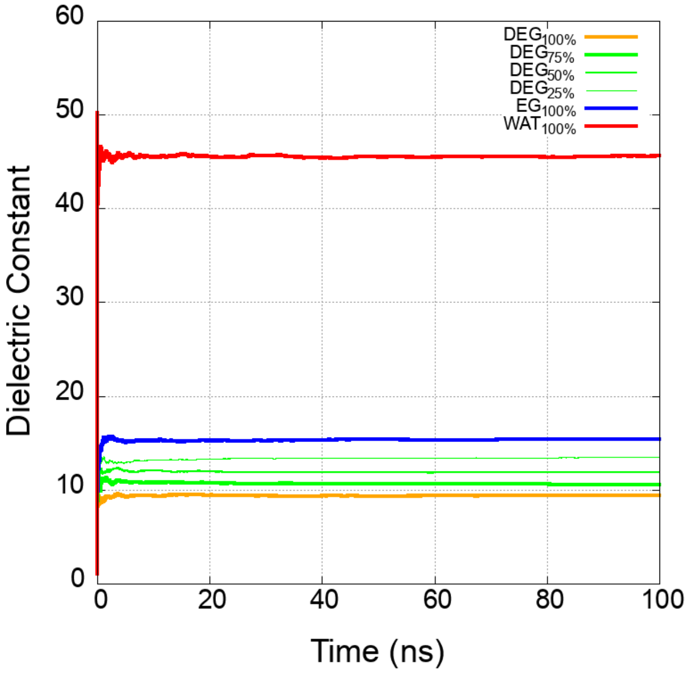


Figure 8.9: Dielectric constant during the MD simulations for each solvent.

Table 8.2: Theoretical dielectric constant values at 473.15 K and constant volume.

Solvents	Dielectric constant (ϵ)
$\mathrm{DEG}_{100\%}$	9.4±0.1
$\mathrm{DEG}_{75\%}$	10.7±0.1
$\mathrm{DEG}_{50\%}$	11.9±0.1
$\mathrm{DEG}_{25\%}$	13.4±0.2
$\mathrm{EG}_{100\%}$	15.3±0.3
$\mathrm{WAT}_{100\%}$	45.5±0.4

glycol at 423.2 K was ϵ=15.5, so far the information of the dielectric constant of temperatures from 253.2 K to 423.2 K has been reported.[224]

Finally, this research revealed a specific distribution of Na^+ counterions in the space with respect to the orientation of the central carboxylate anion (See Figure 8.10). If the central carboxylate anion is oriented upwards, the surrounding Na^+ counterions are located only in the upper plane. As a result, the Na^+ counterions are not arranged in the opposite direction of the central carboxylate anion.

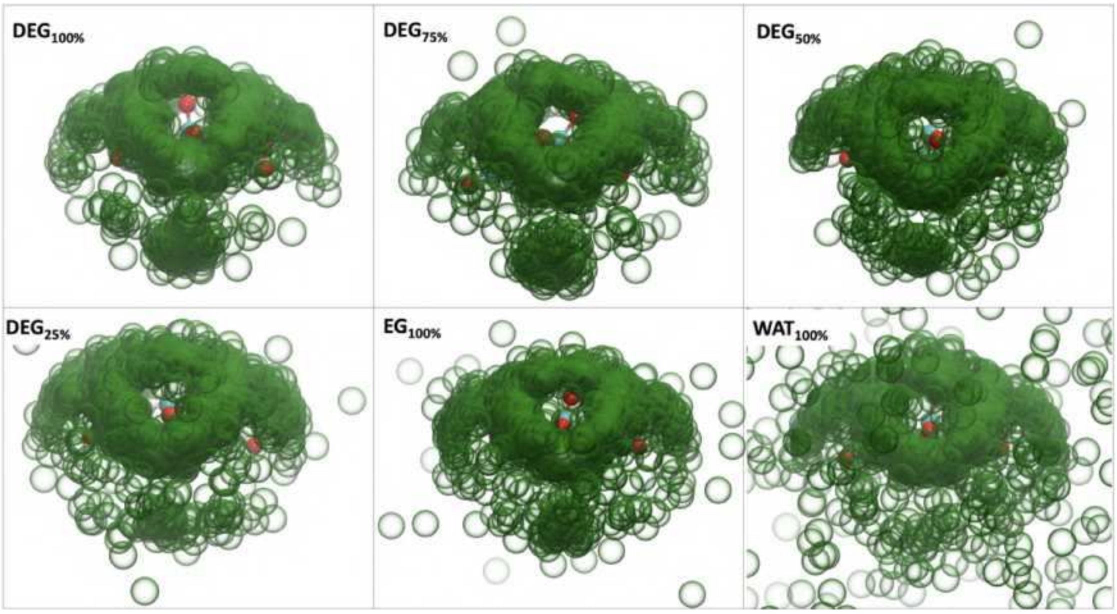


Figure 8.10: Distribution of Na$^+$ counterions in the superficial layer of citrate.

This is the key feature of the citrate that best illustrates the dispersion of the Na$^+$ counterions.

8.4 Conclusion

In this study, we examined simulations of molecular dynamics for citrate in organic solvents. Using this methodology, we were able to identify the solvents water, diethylene glycol, ethylene glycol, and various mixtures of these solvents as having high structural stability for citrate at 473.15 K. The examination of the distribution of the dihedral angle revealed that the primary conformations of the citrate were *gauche*$^+$ or *gauche*$^-$. As a result, we also stated that the Na$^+$ counterions were closer to the central and lateral carboxylate anions of citrate, with the central O6 (corresponding to carbon C6) having a stronger attraction to the Na$^+$ than the lateral O5 and O1 (corresponding to C1 and C5) in all systems examined. It should be noted that the values of the dipole moments of just the solvents showed a trend consistent with prior experimental research, with the maximum dipole moment occurring in pure water, then ethylene glycol, and finally diethylene glycol. In conclusion, this research contributes to our understanding of citrate behaviour in organic solvents at a temperature relevant to the creation of various nanoparticles.

Chapter 9

Conclusions

1. A program based on the Semi-Grand Canonical Monte Carlo method was designed to calculate the protonation states of a protein's titratable residues at different pH and generate ten microstates for each type of protein submitted.

2. The protonation and deprotonation states of titratable residues were determined at different types of pH, taking into account the experimental pka for the urease protein of *Helicobacter pylori*, and the main protease and Spike-S1 of SARS-CoV-2.

3. The structural stability of macromolecular systems was analysed using all-atom molecular dynamics simulations, considering different protonation states for a fixed pH. We have found that the high stability of *Helicobacter pylori* urease at extreme pH conditions is linked to its complex structural conformation, organised by twelve heterodimeric subunits and twelve active centres, which allow its stability at acidic pHs. Regarding the structural stability of the main protease of SARS-CoV-2 at acidic pHs, the high stability at pH 5 demonstrates its survival in the late endosome when SARS-CoV2 infects its host. On the other hand, analysing the pH for five Spike-S1 mutations of SARS-CoV-2 helped us identify the very stable omicron variant at pH 8.

4. The analysis of the binding energy of interacting systems (Spike-S1 and ACE2) showed us that the mutations present in five variants of SARS-CoV-2 were very important to compare because it was possible to observe that the mutations in the omicron variant allow better coupling to ACE2.

5. All-atom computational simulation methods evaluated the self-assembly of phenylalanine; this amino acid is directly linked to the disease phenylketonuria, and indeed our results demonstrated that the non-covalent interac-

tions between phenylalanines allow stacking in the ladder-shaped structure and zig-zag, allowing the formation of a fibre.

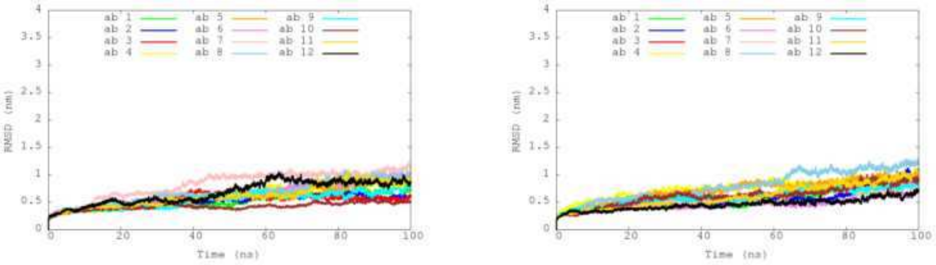
6. The behaviour of sodium citrate in different diethylene glycol and ethylene glycol solvent mixtures was analyzed by computational simulation methods, where we found that the temperature that is typically used in experimental laboratories contributes to the control of the citrate form, which favours the design of magnetized nanoparticles.

Appendix

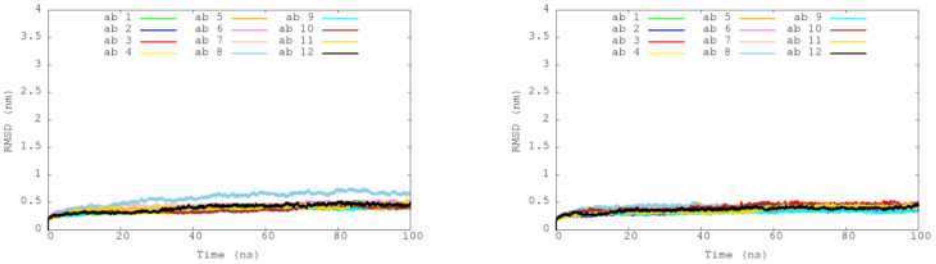
Appendix A: Effect of pH on the Supramolecular Structure of Helicobacter pylori Urease by Molecular Dynamics Simulations

Table A.1: Number of molecules in the system for each pH

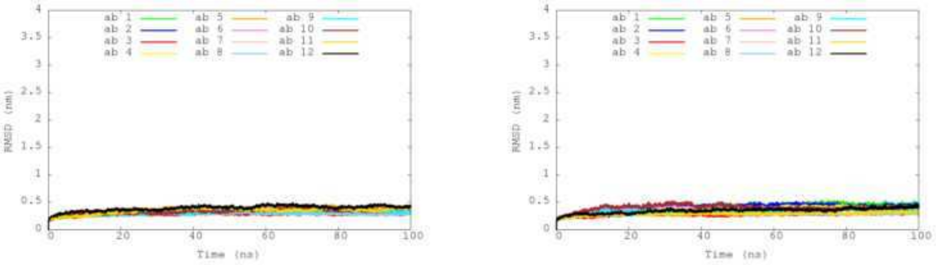
pH	TIP4P Water	Ions
2	148944	1373 Cl
3	149121	1208 Cl
4	149573	810 Cl
5	149987	438 Cl
6	150261	189 Cl
7	150465	3 Na
7.5	150391	83 Na



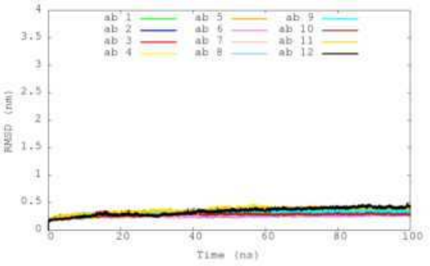
(a) pH2 (b) pH3



(c) pH4 (d) pH5

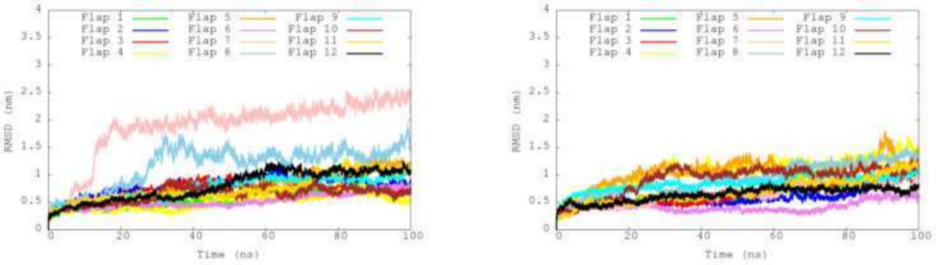


(e) pH6 (f) pH7



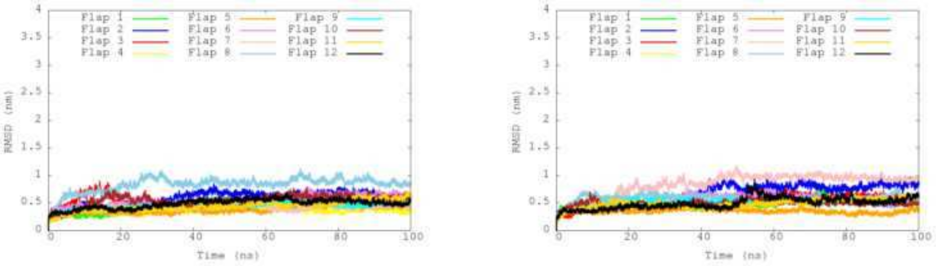
(g) pH7.5

Figure A.1: MD analysis of each protomer at different pHs.



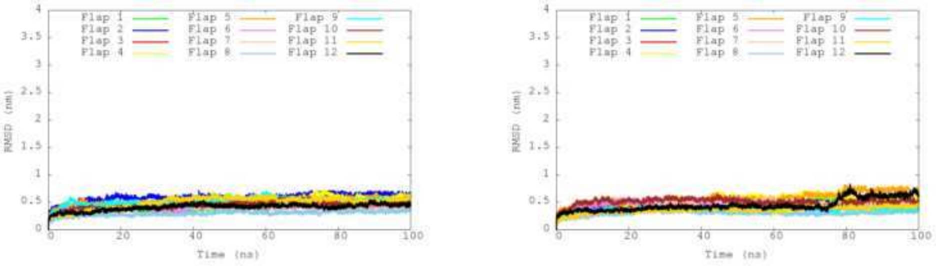
(a) pH2

(b) pH3



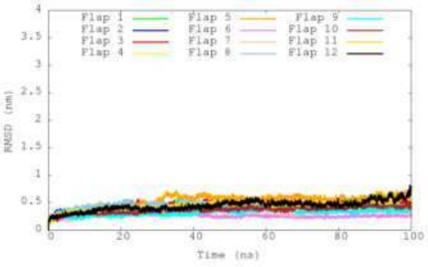
(c) pH4

(d) pH5



(e) pH6

(f) pH7



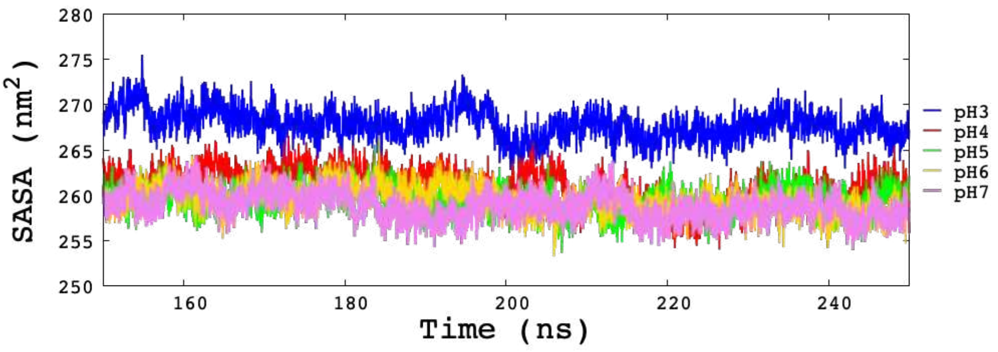
(g) pH7.5

Figure A.2: MD analysis of each flap at different pHs.

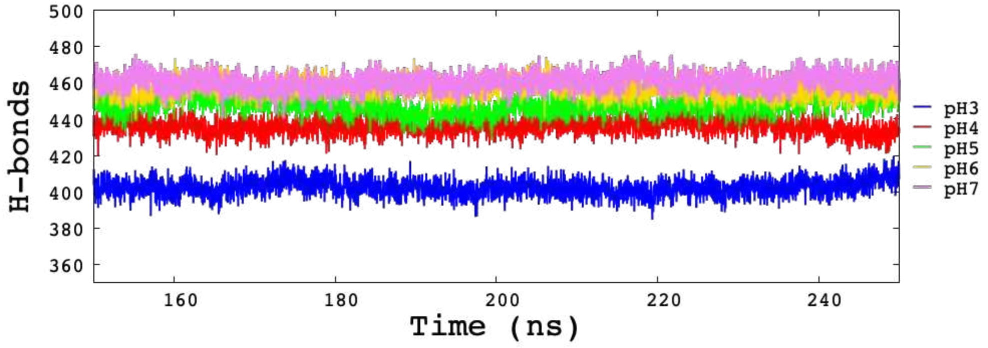
Table A.2: Structural properties of Ramachandran analysis

Structural model	**Poor rotamers**	**Favored rotamers**	**Ramachandran outliers**	**Ramacha**
Crystal	9.90%	77.60%	7.10%	7
pH2	4.20%	86.40%	1.90%	8
pH3	3.90%	86.50%	2.30%	8
pH4	4.20%	85.50%	2.10%	8
pH5	4.10%	86.40%	2.10%	8
pH6	4.80%	84.80%	2.00%	8
pH7	4.90%	85.30%	2.10%	8
pH7.5	4.60%	85.40%	1.80%	8

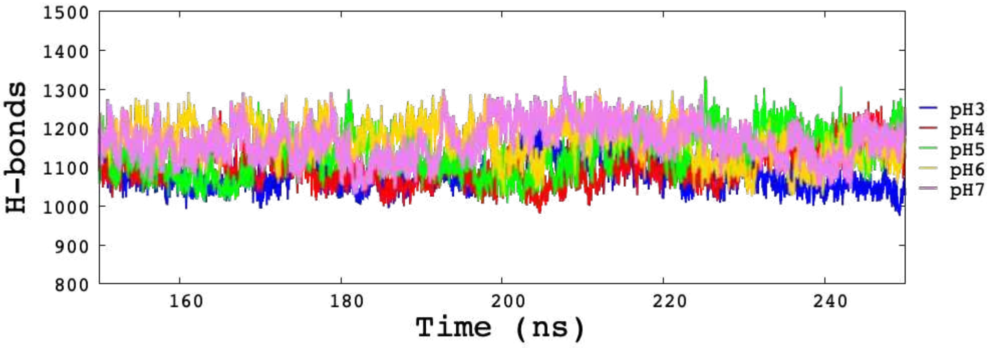
Appendix B: Unveiling the Effect of Low pH on the SARS-CoV-2 Main Protease by Molecular Dynamics Simulations



(a)

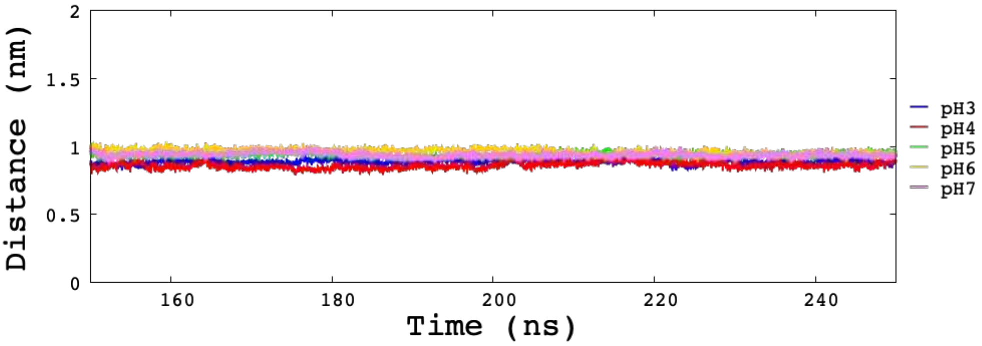


(b)

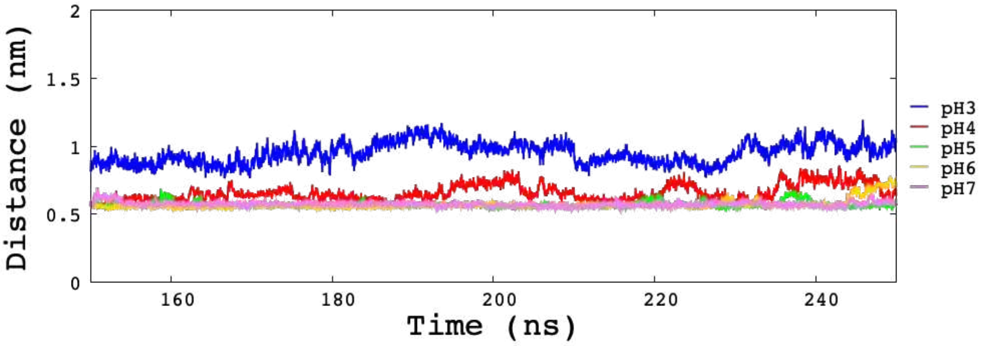


(c)

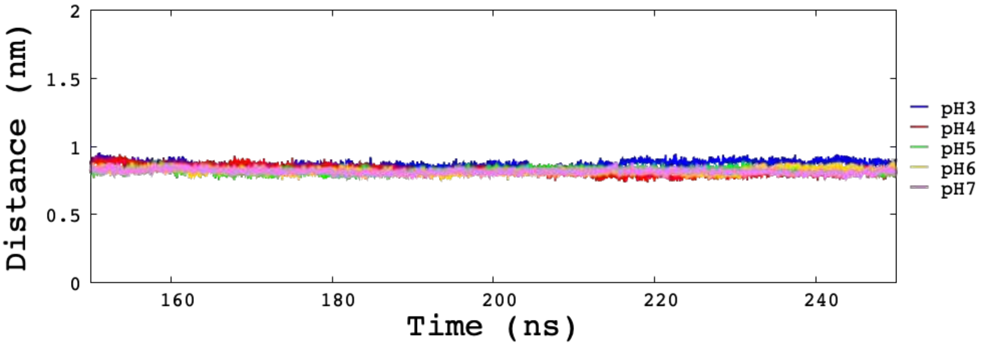
Figure B.1: Plot of hydrogen bonds and SASA during the time of MD simulations at different pHs. (a) Solvent Accessible Surface Area. (b) Intramolecular Hydrogen bonds between protein-protein. (c) Intermolecular Hydrogen bonds between protein-waters.



(a)

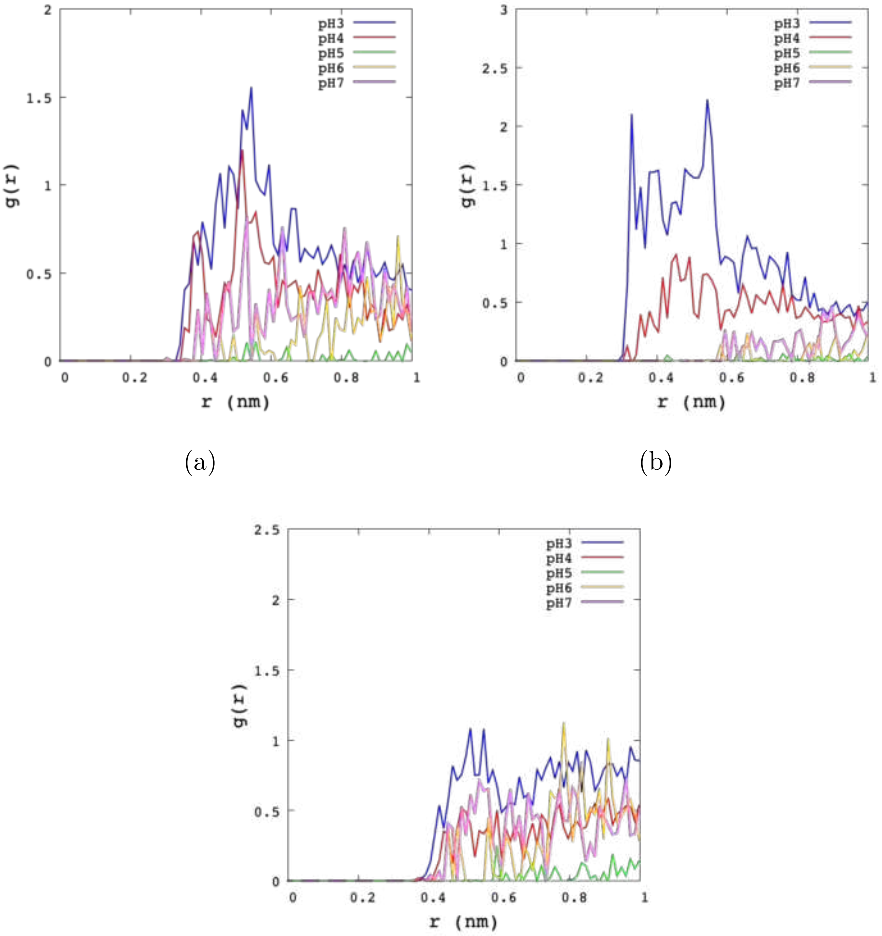


(b)



(c)

Figure B.2: Distances between C-α atoms from residues of the active site (Cys145 and His41) and S1 pocket in Mpro. (a) Average distance Cys145-His41. (b) Average distance of Met49-Gln189. (c) Average distance of His172-Phe140.



(a)

(b)

(c)

Figure B.3: Radial Distribution of active site residues against Mpro. (a) Sγ–ions of Cys145. (b) Nδ1–ions of His41. (c) Sγ–ions of Met49.

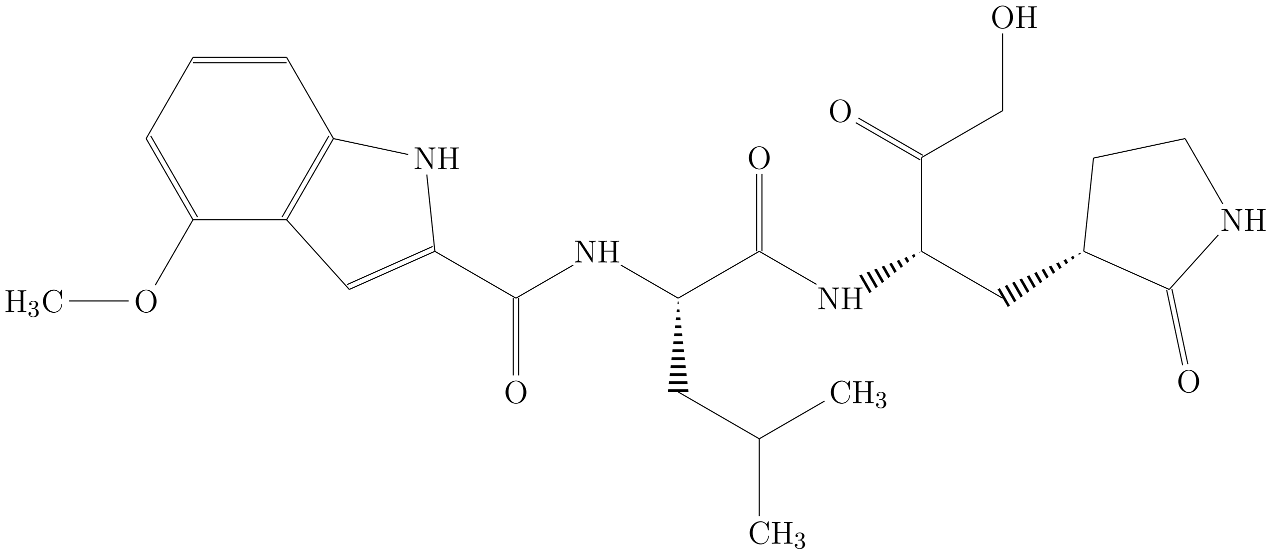
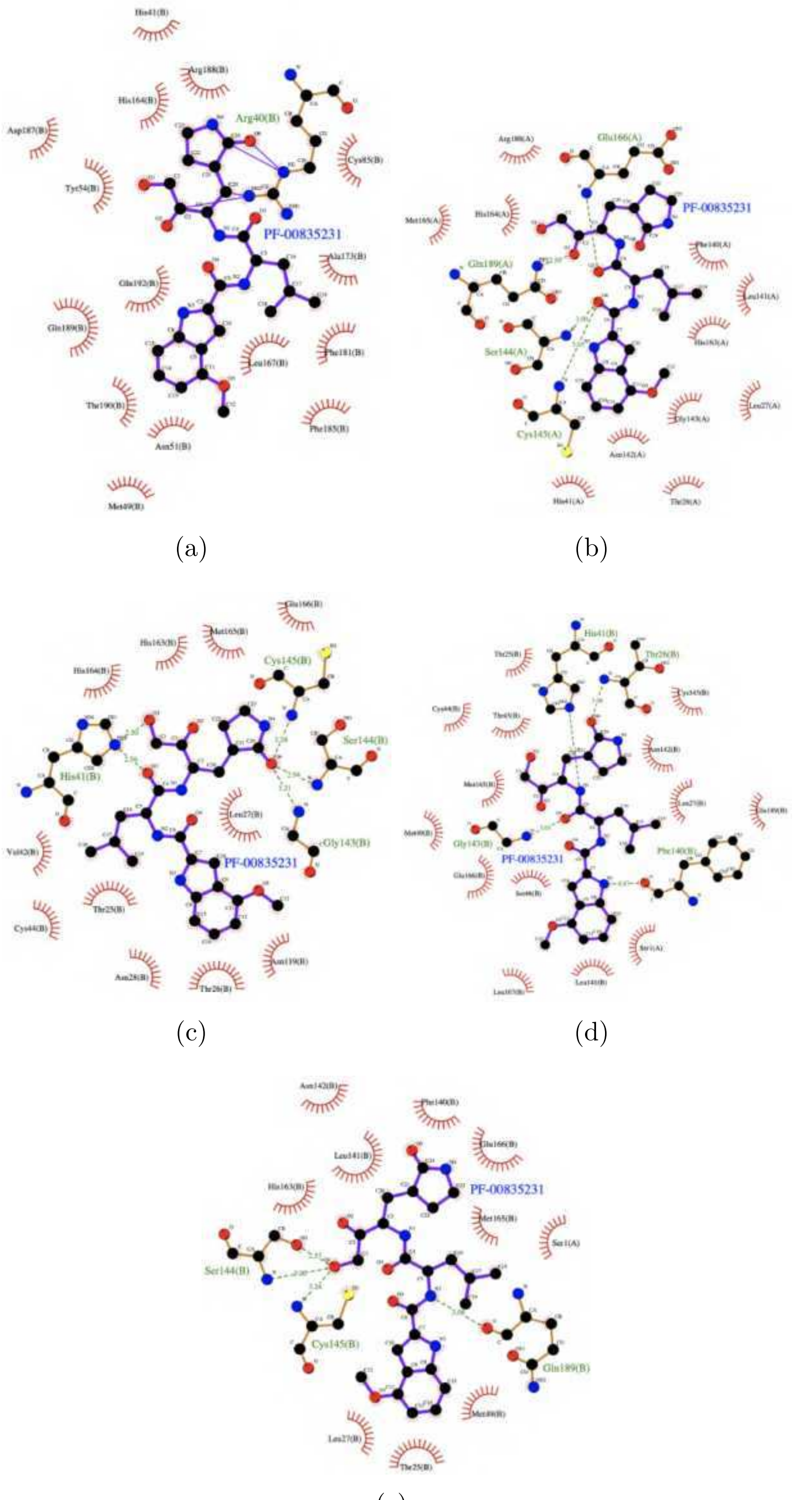


Figure B.4: 2D representation of PF-00835231.



(a) (b)

(c) (d)

(e)

Figure B.5: 2D representation of the best binding affinity Mpro–PF-00835231 . (a) pH3. (b) pH4. (c) pH5. (d) pH6. (e) pH7.

Table B.1: Total charge of each titratable amino acids.

pH3

Residues	**MicroS-1**	**MicroS-2**	**MicroS-3**	**MicroS-4**	**MicroS-5**	**MicroS-6**	**MicroS-7**	**MicroS-8**	**MicroS-9**	**MicroS-**
ASP (34)	-11	-9	-9	-8	-8	-8	-9	-7	-10	-8
GLU (18)	-5	-5	-3	-4	-3	-3	-3	-6	-5	-5
ARG (22)	22	22	22	22	22	22	22	22	22	22
LYS (22)	22	22	22	22	22	22	22	22	22	22
HIS (14)	10	10	9	10	10	10	10	10	10	10

pH4

Residues	**MicroS-1**	**MicroS-2**	**MicroS-3**	**MicroS-4**	**MicroS-5**	**MicroS-6**	**MicroS-7**	**MicroS-8**	**MicroS-9**	**MicroS-**
ASP (34)	-21	-21	-27	-22	-22	-24	-20	-26	-22	-24
GLU (18)	-6	-6	-6	-9	-8	-8	-6	-6	-7	-6
ARG (22)	22	22	22	22	22	22	22	22	22	22
LYS (22)	22	22	22	22	22	22	22	22	22	22
HIS (14)	9	10	9	9	9	9	9	10	8	8

pH5

Residues	**MicroS-1**	**MicroS-2**	**MicroS-3**	**MicroS-4**	**MicroS-5**	**MicroS-6**	**MicroS-7**	**MicroS-8**	**MicroS-9**	**MicroS-**
ASP (34)	-32	-30	-34	-33	-31	-33	-30	-34	-33	-32
GLU (18)	-9	-13	-7	-9	-12	-11	-15	-10	-12	-10
ARG (22)	22	22	22	22	22	22	22	22	22	22
LYS (22)	22	22	22	22	22	22	22	22	22	22
HIS (14)	6	5	7	5	3	7	4	7	5	7

pH6

Residues	**MicroS-1**	**MicroS-2**	**MicroS-3**	**MicroS-4**	**MicroS-5**	**MicroS-6**	**MicroS-7**	**MicroS-8**	**MicroS-9**	**MicroS-**
ASP (34)	-34	-34	-34	-33	-33	-34	-34	-34	-34	-33
GLU (18)	-16	-16	-14	-16	-15	-16	-17	-17	-15	-16
ARG (22)	22	22	22	22	22	22	22	22	22	22
LYS (22)	22	22	22	22	22	22	22	22	22	22
HIS (14)	2	3	2	3	1	2	1	2	3	1

pH7

Residues	**MicroS-1**	**MicroS-2**	**MicroS-3**	**MicroS-4**	**MicroS-5**	**MicroS-6**	**MicroS-7**	**MicroS-8**	**MicroS-9**	**MicroS-**
ASP (34)	-34	-34	-34	-34	-34	-34	-34	-34	-34	-34
GLU (18)	-17	-17	-17	-16	-17	-17	-16	-17	-18	-18
ARG (22)	22	22	22	22	22	22	22	22	22	22
LYS (22)	22	22	22	22	22	22	22	22	22	22
HIS (14)	1	0	0	0	1	2	0	0	1	1

Table B.2: Protonate/deprotonate microstate of His41 and His172. The symbol P and D define the deprotonated states, respectively.

Residue	MicroS	pH3	pH4	pH5	pH6	
His41	1	chainA(P)/chainB(P)	chainA(P)/chainB(D)	chainA(D)/chainB(D)	chainA(D)/chainB(D)	cha
	2	chainA(P)/chainB(P)	chainA(P)/chainB(P)	chainA(D)/chainB(D)	chainA(D)/chainB(D)	cha
	3	chainA(D)/chainB(P)	chainA(D)/chainB(P)	chainA(P)/chainB(P)	chainA(D)/chainB(D)	cha
	4	chainA(P)/chainB(P)	chainA(P)/chainB(P)	chainA(D)/chainB(D)	chainA(D)/chainB(D)	cha
His172	1	chainA(P)/chainB(P)	chainA(P)/chainB(P)	chainA(P)/chainB(P)	chainA(D)/chainB(D)	cha
	2	chainA(P)/chainB(P)	chainA(P)/chainB(P)	chainA(D)/chainB(D)	chainA(D)/chainB(D)	cha
	3	chainA(P)/chainB(P)	chainA(P)/chainB(P)	chainA(D)/chainB(D)	chainA(D)/chainB(D)	cha
	4	chainA(P)/chainB(P)	chainA(P)/chainB(P)	chainA(D)/chainB(D)	chainA(D)/chainB(D)	cha

Table B.3: Values of the MD was fitted against the average structure of the last 100 ns of the dynamics was used.

pH	**MicroS**	**RMSD (nm)**	**RG (nm)**	**SASA (nm^2)**	**Hbond (protein-protein)**	**Hbond (protein-waters)**
pH3	pH3-1	0.28±0.02	2.543±0.009	269±3	404±9	(111±9)x10
	pH3-2	0.21±0.02	2.561±0.011	269±3	409±10	(103±5)x10
	pH3-3	0.32±0.05	2.543±0.019	266±5	393±10	(110±8)x10
	pH3-4	0.24±0.02	2.562±0.008	267±3	405±9	(109±9)x10
	Average	0.26±0.06	2.553±0.010	268±2	401±1	(107±4)x10
pH4	pH4-1	0.22±0.02	2.537±0.008	261±3	438±10	(109±10)x10
	pH4-2	0.21±0.02	2.538±0.008	262±3	427±9	(110±10)x10
	pH4-3	0.18±0.02	2.536±0.008	259±3	447±10	(109±9)x10
	pH4-4	0.21±0.02	2.541±0.009	263±4	432±10	(112±10)x10
	Average	0.20±0.02	2.538±0.002	261±2	437±10	(110±2)x10
pH5	pH5-1	0.25±0.03	2.532±0.009	258±3	448±9	(122±15)x10
	pH5-2	0.24±0.02	2.542±0.008	262±3	450±10	(109±9)x10
	pH5-3	0.24±0.02	2.518±0.008	256±3	445±10	(107±9)x10
	pH5-4	0.18±0.02	2.537±0.008	262±3	449±9	(115±11)x10
	Average	0.21±0.04	2.537±0.005	259±3	448±3	(114±8)x10
pH6	pH6-1	0.18±0.01	2.534±0.007	261±3	462±9	(102±2)x10
	pH6-2	0.22±0.01	2.527±0.007	258±3	450±9	(122±10)x10
	pH6-3	0.20±0.02	2.534±0.008	259±3	452±9	(126±12)x10
	pH6-4	0.21±0.02	2.537±0.008	259±3	458±8	(116±11)x10
	Average	0.20±0.02	2.532±0.005	260±2	456±6	(114±12)x10
pH7	pH7-1	0.16±0.02	2.533±0.010	259±3	460±10	(102±2)x10
	pH7-2	0.21±0.02	2.526±0.008	258±3	462±9	(126±10)x10
	pH7-3	0.21±0.03	2.528±0.008	260±3	460±9	(118±12)x10
	pH7-4	0.19±0.01	2.519±0.006	258±3	461±10	(121±12)x10
	Average	0.19±0.02	2.526±0.007	259±1	461±1	(114±12)x10

Table B.4: Distances between amino acids from the catalytic dyad for each microstate.

pH	**MicroS**	**His41-Cys145**	**Met49-Cys145**	**His172-Cys145**
pH3	pH3-1	0.74±0.10	2.50±0.09	0.94±0.10
	pH3-2	0.71±0.11	1.59±0.18	0.93±0.16
	pH3-3	0.65±0.06	1.39±0.04	0.67±0.09
	pH3-4	0.50±0.03	1.37±0.05	0.85±0.04
	Average	0.65±0.04	1.71±0.05	0.85±0.05
pH4	pH4-1	0.61±0.08	1.62±0.14	1.20±0.04
	pH4-2	0.68±0.09	1.31±0.08	1.13±0.13
	pH4-3	0.57±0.07	1.53±0.19	1.21±0.06
	pH4-4	0.60±0.07	1.39±0.05	1.11±0.14
	Average	0.62±0.04	1.46±0.07	1.16±0.05
pH5	pH5-1	0.55±0.04	1.41±0.06	0.91±0.06
	pH5-2	0.59±0.06	1.41±0.06	0.72±0.08
	pH5-3	0.53±0.06	1.43±0.07	0.68±0.12
	pH5-4	0.54±0.07	1.43±0.06	0.87±0.12
	Average	0.55±0.03	1.42±0.03	0.79±0.05
pH6	pH6-1	0.54±0.08	1.43±0.06	1.06±0.07
	pH6-2	0.56±0.08	1.41±0.06	0.70±0.10
	pH6-3	0.54±0.06	1.41±0.07	0.89±0.10
	pH6-4	0.48±0.05	1.49±0.05	1.00±0.06
	Average	0.53±0.03	1.43±0.03	0.91±0.04
pH7	pH7-1	0.52±0.06	1.40±0.06	0.97±0.11
	pH7-2	0.50±0.06	1.46±0.07	0.93±0.08
	pH7-3	0.50±0.08	1.39±0.07	0.94±0.06
	pH7-4	0.54±0.03	1.38±0.06	0.81±0.05
	Average	0.51±0.03	1.41±0.03	0.91±0.04

Table B.5: Numerical values of Ramachandran plot of the last frame of each microstate determined by Molprobity server.

	MicroS	**Poor rotamers**		**Favored rotamers**		**Ramachandran outliers**		**Ramachandran favored**		**Favored**	**Allowed**
pH3	pH3-1	9	1.80%	452	90.40%	6	0.99%	563	93.21%	93.20%	99.00%
	pH3-2	13	2.60%	456	91.20%	5	0.83%	564	93.38%	93.40%	99.20%
	pH3-3	13	2.60%	456	91.20%	7	1.16%	565	93.54%	93.50%	98.80%
	pH3-4	10	2.00%	452	90.40%	6	0.99%	567	93.87%	93.90%	99.00%
pH4	pH4-1	11	2.20%	455	91.00%	6	0.99%	565	93.54%	93.50%	99.00%
	pH4-2	10	2.00%	450	90.00%	5	0.83%	568	94.04%	94.00%	99.20%
	pH4-3	13	2.60%	445	89.00%	4	0.66%	558	92.38%	92.40%	99.30%
	pH4-4	10	2.00%	449	89.80%	4	0.66%	567	93.87%	93.90%	99.30%
pH5	pH5-1	15	3.00%	450	90.00%	2	0.33%	552	91.39%	91.40%	99.70%
	pH5-2	14	2.80%	449	89.80%	2	0.33%	562	93.05%	93.00%	99.70%
	pH5-3	14	2.80%	454	90.80%	7	1.16%	553	91.56%	91.60%	98.80%
	pH5-4	11	2.20%	459	91.80%	2	0.33%	555	91.89%	91.90%	99.70%
pH6	pH6-1	19	3.80%	439	87.80%	8	1.32%	556	92.05%	92.10%	98.70%
	pH6-2	13	2.60%	454	90.80%	6	0.99%	564	93.38%	93.40%	99.00%
	pH6-3	8	1.60%	451	90.20%	8	1.32%	554	91.72%	91.70%	98.70%
	pH6-4	12	2.40%	451	90.20%	5	0.83%	554	91.72%	91.70%	99.20%
pH7	pH7-1	13	2.60%	448	89.60%	3	0.50%	554	91.72%	91.70%	99.50%
	pH7-2	17	3.40%	457	91.40%	0	0.00%	562	93.05%	93.00%	100.00%
	pH7-3	10	2.00%	450	90.00%	5	0.83%	554	91.72%	91.70%	99.20%
	pH7-4	7	1.40%	466	93.20%	5	0.83%	556	92.05%	92.10%	99.20%

Table B.6: Docking results of each microstate at different pHs.

Name	Size x (Å)	Size y (Å)	Size z (Å)	Search Domain Volume (A^3)	Number of flexible side chains
ph3-2-a	39.2	31.4	31.7	39018.896	32
ph3-2-b	40	33.2	32	42496	32
ph3-3-a	34.3	31.6	37.7	40862.276	32
ph3-3-b	38.4	33.8	27.8	36082.176	33
ph3-4-a	30.8	28.7	35.1	31026.996	32
ph3-4-b	31.6	33	31.9	33265.32	28
ph3-a	29.4	44	33.9	43853.04	27
ph3-b	42.7	37.9	33.6	54375.888	30
ph4-2-a	29.9	36.4	30.4	33086.144	28
ph4-2-b	31	29.5	29.8	27252.1	29
ph4-3-a	27.8	34.5	30.5	29252.55	27
ph4-3-b	28.9	30	33.3	28871.1	28
ph4-4-a	27.9	33.9	27.5	26009.775	29
ph4-4-b	29.8	29.2	37.7	32805.032	29
ph4-a	40	38	43	65360	31
ph4-b	37	40	48.1	71188	32
ph5-2-a	35	31.6	32	35392	30
ph5-2-b	32.1	36.2	35	40670.7	30
ph5-3-a	30.6	26.1	32.5	25956.45	31
ph5-3-b	28	32.2	37.8	34080.48	30
ph5-4-a	35.2	34.8	27.5	33686.4	30
ph5-4-b	25.8	32.5	31.8	26664.3	31
ph5-a	37.9	37.1	47.5	66789.275	30
ph5-b	37.2	37.8	37.8	53152.848	30
ph6-2-a	28.3	28.5	34.8	28067.94	30
ph6-2-b	30.6	32.2	30.5	30052.26	29
ph6-3-a	30.8	34.9	33.2	35687.344	28
ph6-3-b	36.8	30.9	30.7	34909.584	29
ph6-4-a	30.2	28.5	43	37010.1	30
ph6-4-b	30.2	31.5	32.4	30822.12	30
ph6-a	36	37.6	36.8	49812.48	30
ph6-b	34	32.1	46.6	50859.24	31
ph7-2-a	30	28.3	33.4	28356.6	30
ph7-2-b	33.9	37.7	28.8	36807.264	29
ph7-3-a	30	26.4	37.6	29779.2	30
ph7-3-b	35.3	28.1	26	25790.18	31
ph7-4-a	25.9	34.8	30.8	27760.656	31
ph7-4-b	32.8	29.3	32.5	31233.8	31
ph7-a	37	34.5	45.7	58336.05	30
ph7-b	45.2	41	35	64862	31

Appendix C: Computational study of pH effect on different SARS-CoV-2 mutations in the S1 region of the Spike protein against the ACE2 receptor

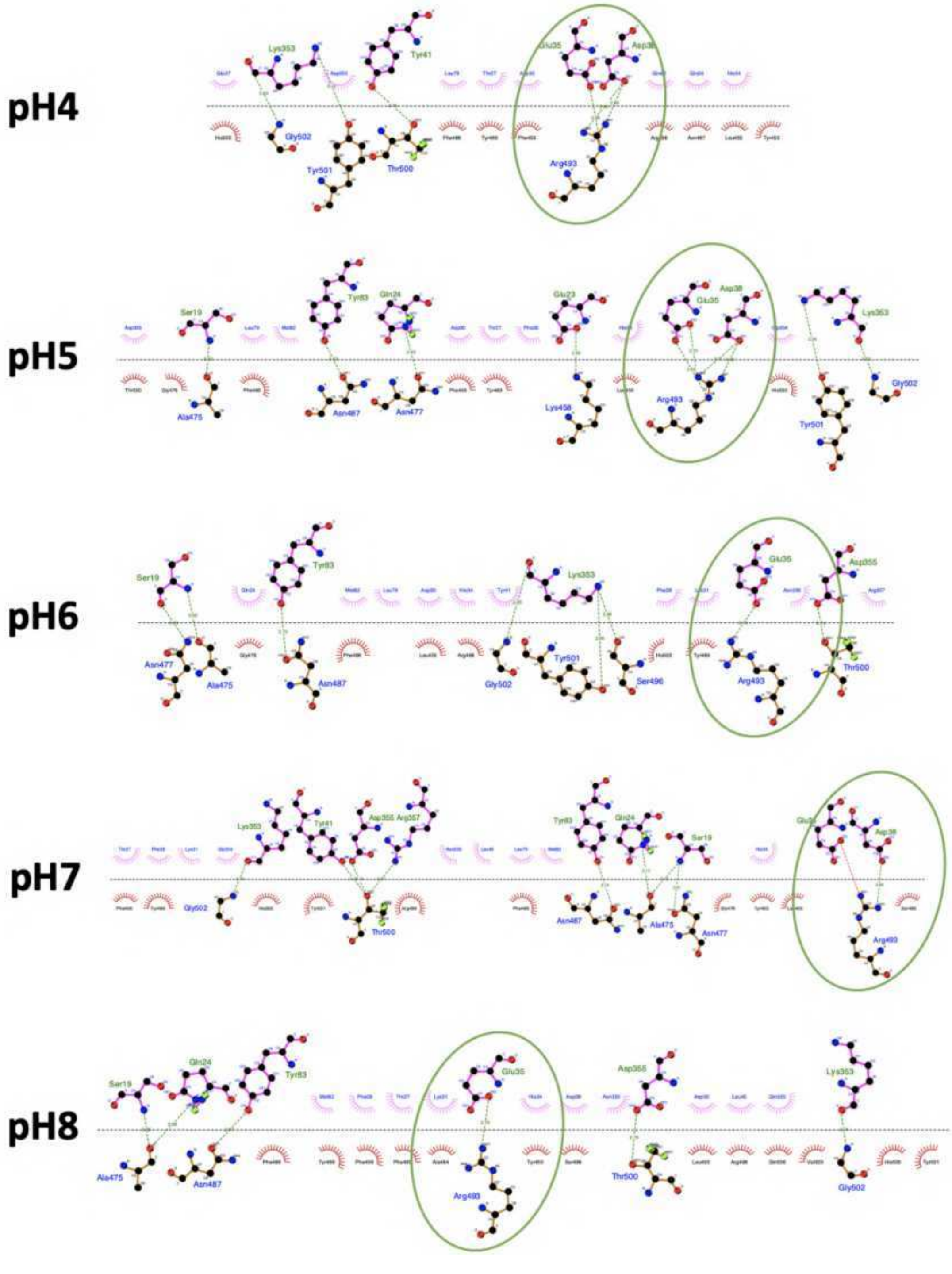
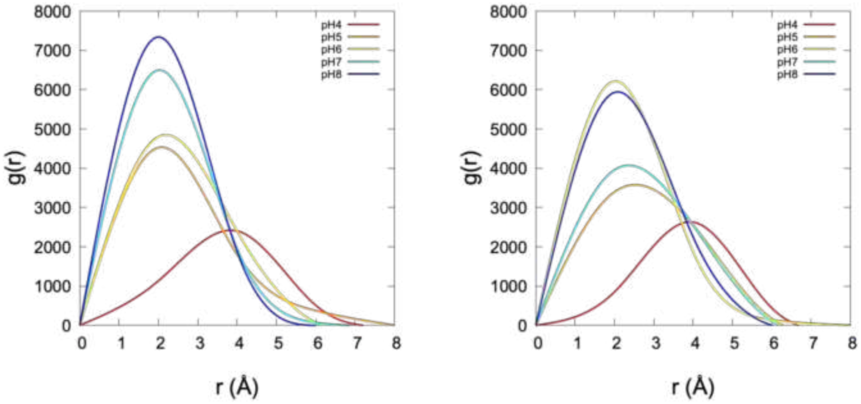
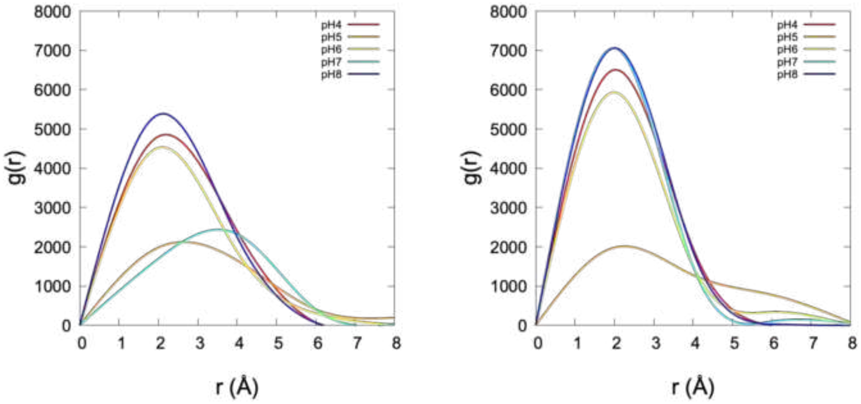


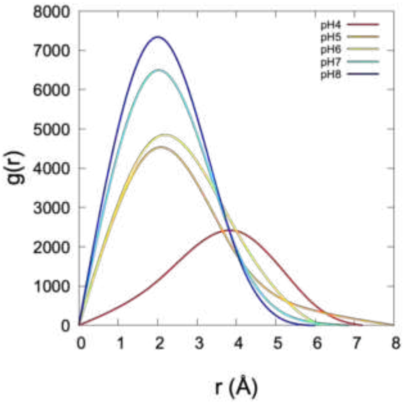
Figure C.1: Ligplot diagram elucidating the Spike-S1 omicron vs ACE2 interactions. Red denotes the Spike-S1 omicron the principal hot spots residues, while purple denotes the ACE2 aminoacids. The green and red dotted lines, respectively, depict the hydrogen bond and salt bridge interactions between aminoacids.



(a) ARG493(NH1)-GLU35(OE1) (b) ARG493(NH1)-GLU35(OE2)



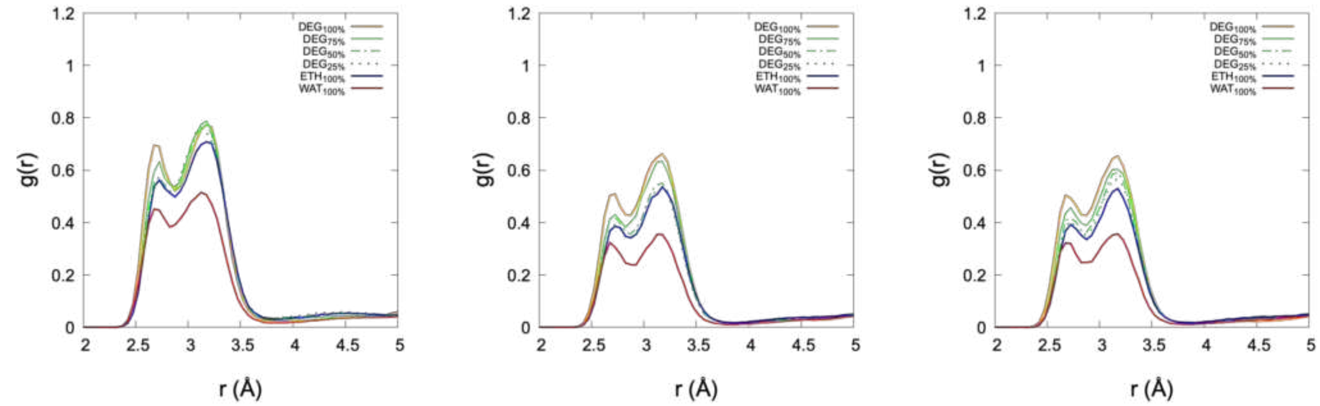
(c) ARG493(NH2)-GLU35(OE1) (d) ARG493(NH2)-GLU35(OE2)



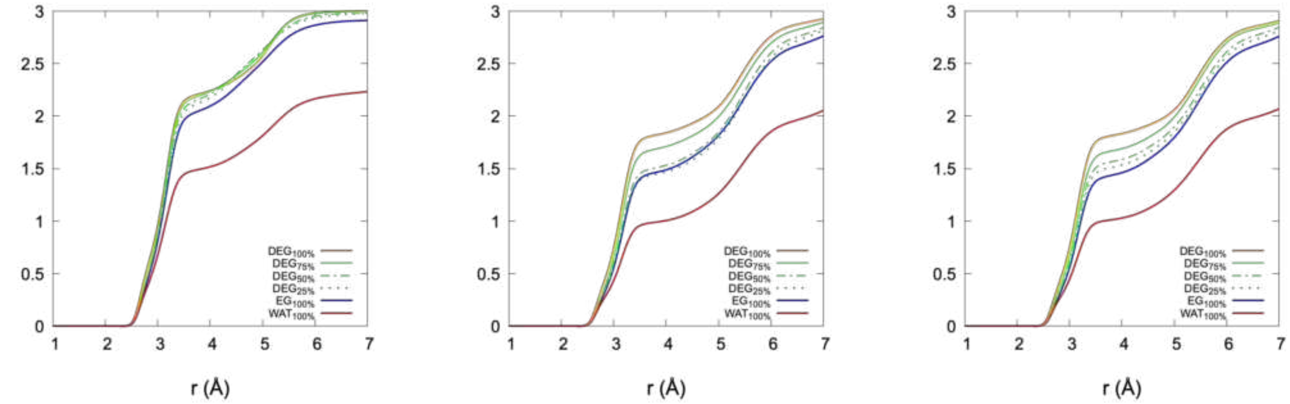
(e) ARG493(CZ)-GLU35(CD)

Figure C.2: Radial Distribution function of ARG493 with GLU35

Appendix D: Influence of polar solvent mixtures in the conformation of polycarboxylates: Case study of citrate in ethylene glycol and diethylene glycol mixtures



(a) **C6** (b) **C1** (c) **C5**



(d) **C6** (e) **C1** (f) **C5**

Figure D.1: Radial distribution function between Na ions with carbon from carboxilic group of citrate.
473.15 K

Bibliography

[1] Sergio Madurga, Josep Lluís Garcés, Encarnació Companys, Carlos Rey-Castro, José Salvador, Josep Galceran, Eudald Vilaseca, Jaume Puy, and Francesc Mas. Ion binding to polyelectrolytes: Monte carlo simulations versus classical mean field theories. *Theoretical Chemistry Accounts*, 123(1):127–135, 2009.

[2] Sergio Madurga, Carlos Rey-Castro, Isabel Pastor, Eudald Vilaseca, Calin David, Josep Lluís Garcés, Jaume Puy, and Francesc Mas. A semi-grand canonical monte carlo simulation model for ion binding to ionizable surfaces: Proton binding of carboxylated latex particles as a case study. *The Journal of chemical physics*, 135(18):184103, 2011.

[3] Haruna L Barazorda-Ccahuana, Badhin Gómez, Francesc Mas, and Sergio Madurga. Effect of ph on the supramolecular structure of helicobacter pylori urease by molecular dynamics simulations. *Polymers*, 12(11):2713, 2020.

[4] Shamshul Ansari and Yoshio Yamaoka. Survival of helicobacter pylori in gastric acidic territory. *Helicobacter*, 22(4):e12386, 2017.

[5] Nam-Chul Ha, Sang-Taek Oh, Jae Young Sung, Kyeung Ah Cha, Mann Hyung Lee, and Byung-Ha Oh. Supramolecular assembly and acid resistance of helicobacter pylori urease. *Nature structural biology*, 8(6):505–509, 2001.

[6] Süren Schreiber, Roland Bucker, Claudia Groll, Marina Azevedo-Vethacke, Désirée Garten, Peter Scheid, Susanne Friedrich, Süren Gatermann, Christine Josenhans, and Sebastian Suerbaum. Rapid loss of motility of helicobacter pylori in the gastric lumen in vivo. *Infection and immunity*, 73(3):1584–1589, 2005.

[7] Karen M Ottemann and Andrew C Lowenthal. Helicobacter pylori uses motility for initial colonization and to attain robust infection. *Infection and immunity*, 70(4):1984–1990, 2002.

[8] Cheng-Yen Kao, Bor-Shyang Sheu, and Jiunn-Jong Wu. Helicobacter pylori infection: An overview of bacterial virulence factors and pathogenesis. *Biomedical journal*, 39(1):14–23, 2016.

[9] Richa Arora, Upasana Issar, and Rita Kakkar. In silico study of the active site of helicobacter pylori urease and its inhibition by hydroxamic acids. *Journal of Molecular Graphics and Modelling*, 83:64–73, 2018.

[10] Guillermina Estiu and Kenneth M Merz. Catalyzed decomposition of urea. molecular dynamics simulations of the binding of urea to urease. *Biochemistry*, 45(14):4429–4443, 2006.

[11] Benjamin P Roberts, Bill R Miller III, Adrian E Roitberg, and Kenneth M Merz Jr. Wide-open flaps are key to urease activity. *Journal of the American Chemical Society*, 134(24):9934–9937, 2012.

[12] Mona S Minkara, Melek N Ucisik, Michael N Weaver, and Kenneth M Merz Jr. Molecular dynamics study of helicobacter pylori urease. *Journal of chemical theory and computation*, 10(5):1852–1862, 2014.

[13] Yong-Peng Xu, Jie Qin, Shi-Min Sun, Tong-Tong Liu, Xiao-Lin Zhang, Shao-Song Qian, and Hai-Liang Zhu. Synthesis, crystal structures, molecular docking and urease inhibitory activity of nickel (ii) complexes with 3-pyridinyl-4-amino-5-mercapto-1, 2, 4-triazole. *Inorganica Chimica Acta*, 423:469–476, 2014.

[14] Lee Macomber, Mona S Minkara, Robert P Hausinger, and Kenneth M Merz Jr. Reduction of urease activity by interaction with the flap covering the active site. *Journal of chemical information and modeling*, 55(2):354–361, 2015.

[15] James B Sumner. The isolation and crystallization of the enzyme urease preliminary paper. *Journal of Biological Chemistry*, 69(2):435–441, 1926.

[16] D Eric Anderson, Wayne J Becktel, and Frederick W Dahlquist. ph-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of t4 lysozyme. *Biochemistry*, 29(9):2403–2408, 1990.

[17] Donald Bashford and Martin Karplus. pka's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry*, 29(44):10219–10225, 1990.

[18] David R Scott, David Weeks, Charlie Hong, Stefan Postius, Klaus Melchers, and George Sachs. The role of internal urease in acid resistance of helicobacter pylori. *Gastroenterology*, 114(1):58–70, 1998.

[19] Pablo M Blanco, Sergio Madurga, Francesc Mas, and Josep L Garcés. Coupling of charge regulation and conformational equilibria in linear weak polyelectrolytes: Treatment of long-range interactions via effective short-ranged and ph-dependent interaction parameters. *Polymers*, 10(8):811, 2018.

[20] Pablo M Blanco, Sergio Madurga, Claudio F Narambuena, Francesc Mas, and Josep L Garcés. Role of charge regulation and fluctuations in the conformational and mechanical properties of weak flexible polyelectrolytes. *Polymers*, 11(12):1962, 2019.

[21] Mats HM Olsson, Chresten R Søndergaard, Michal Rostkowski, and Jan H Jensen. Propka3: consistent treatment of internal and surface residues in empirical p k a predictions. *Journal of chemical theory and computation*, 7(2):525–537, 2011.

[22] Abraham López, Marta Vilaseca, Sergio Madurga, Monica Varese, Teresa Tarragó, and Ernest Giralt. Analyzing slowly exchanging protein conformations by ion mobility mass spectrometry: study of the dynamic equilibrium of prolyl oligopeptidase. *Journal of Mass Spectrometry*, 51(7):504–511, 2016.

[23] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and Herman JC Berendsen. Gromacs: fast, flexible, and free. *Journal of computational chemistry*, 26(16):1701–1718, 2005.

[24] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015.

[25] Hans W Horn, William C Swope, Jed W Pitera, Jeffry D Madura, Thomas J Dick, Greg L Hura, and Teresa Head-Gordon. Development of an improved four-site water model for biomolecular simulations: Tip4p-ew. *The Journal of chemical physics*, 120(20):9665–9678, 2004.

[26] AP West, MR Millar, and DS Tompkins. Effect of physical environment on survival of helicobacter pylori. *Journal of Clinical Pathology*, 45(3):228–231, 1992.

[27] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An n·log (n) method for ewald sums in large systems. *The Journal of chemical physics*, 98(12):10089–10092, 1993.

[28] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1):014101, 2007.

[29] Michele Parrinello and Aneesur Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, 52(12):7182–7190, 1981.

[30] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.

[31] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13):1605–1612, 2004.

[32] Bruce E Dunn, Gail P Campbell, GI Perez-Perez, and MJ Blaser. Purification and characterization of urease from helicobacter pylori. *Journal of Biological Chemistry*, 265(16):9464–9469, 1990.

[33] Gavin R Turbett, PB Høj, Robert Horne, and Brian J Mee. Purification and characterization of the urease enzymes of helicobacter species from humans and animals. *Infection and immunity*, 60(12):5259–5266, 1992.

[34] Martijn WH Pinkse, Claudia S Maier, Jung-In Kim, Byung-Ha Oh, and Albert JR Heck. Macromolecular assembly of helicobacter pylori urease investigated by mass spectrometry. *Journal of mass spectrometry*, 38(3):315–320, 2003.

[35] David R Scott, Elizabeth A Marcus, David L Weeks, and George Sachs. Mechanisms of acid resistance due to the urease system of helicobacter pylori. *Gastroenterology*, 123(1):187–195, 2002.

[36] Haruna Luz Barazorda-Ccahuana, Miroslava Nedyalkova, Francesc Mas, and Sergio Madurga. Unveiling the effect of low ph on the sars-cov-2 main protease by molecular dynamics simulations. *Polymers*, 13(21):3823, 2021.

[37] Ensheng Dong, Hongru Du, and Lauren Gardner. An interactive web-based dashboard to track covid-19 in real time. *The Lancet infectious diseases*, 20(5):533–534, 2020.

[38] Chen Wang, Peter W Horby, Frederick G Hayden, and George F Gao. A novel coronavirus outbreak of global health concern. *The Lancet*, 395(10223):470–473, 2020.

[39] Alexander E. Gorbalenya, Susan C. Baker, Ralph S. Baric, Raoul J. de Groot, Christian Drosten, Anastasia A. Gulyaeva, Bart L. Haagmans, Chris Lauber, Andrey M. Leontovich, Benjamin W. Neuman, Dmitry Penzar, Stanley Perlman, Leo L. M. Poon, Dmitry V. Samborskiy, Igor A. Sidorov, Isabel Sola,

John Ziebuhr, and Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-ncov and naming it sars-cov-2. *Nature Microbiology*, 5(4):536, 2020.

[40] Miyuki Kawase, Kazuya Shirato, Shutoku Matsuyama, and Fumihiro Taguchi. Protease-mediated entry via the endosome of human coronavirus 229e. *Journal of virology*, 83(2):712–721, 2009.

[41] Naidi Yang and Han-Ming Shen. Targeting the endocytic pathway and autophagy process as a novel therapeutic strategy in covid-19. *International journal of biological sciences*, 16(10):1724, 2020.

[42] Sarah Stauffer, Yuehan Feng, Firat Nebioglu, Rosalie Heilig, Paola Picotti, and Ari Helenius. Stepwise priming by acidic ph and a high k+ concentration is required for efficient uncoating of influenza a virus cores after penetration. *Journal of virology*, 88(22):13029–13046, 2014.

[43] Sai Li, Christian Sieben, Kai Ludwig, Chris T Höfer, Salvatore Chiantia, Andreas Herrmann, Frederic Eghiaian, and Iwan AT Schaap. ph-controlled two-step uncoating of influenza virus. *Biophysical journal*, 106(7):1447–1456, 2014.

[44] David C Chan and Peter S Kim. Hiv entry and its inhibition. *Cell*, 93(5):681–684, 1998.

[45] Kosuke Miyauchi, Yuri Kim, Olga Latinovic, Vladimir Morozov, and Gregory B Melikyan. Hiv enters cells via endocytosis and dynamin-dependent fusion with endosomes. *Cell*, 137(3):433–444, 2009.

[46] Hans-Hartwig Otto and Tanja Schirmeister. Cysteine proteases and their inhibitors. *Chemical reviews*, 97(1):133–172, 1997.

[47] Lizbeth Hedstrom. Serine protease mechanism and specificity. *Chemical reviews*, 102(12):4501–4524, 2002.

[48] Jordan Tang and Ricky NS Wong. Evolution in the structure and function of aspartic proteases. *Journal of cellular biochemistry*, 33(1):53–63, 1987.

[49] Pal B Szecsi. The aspartic proteases. *Scandinavian Journal of Clinical and Laboratory Investigation*, 52(sup210):5–22, 1992.

[50] Subramanian Boopathi, Adolfo B Poma, and Ponmalai Kolandaivel. Novel 2019 coronavirus structure, mechanism of action, antiviral drug promises and rule out against its treatment. *Journal of Biomolecular Structure and Dynamics*, 39(9):3409–3418, 2021.

[51] Ashish M Kanhed, Dushyant V Patel, Divya M Teli, Nirav R Patel, Mahesh T Chhabria, and Mange Ram Yadav. Identification of potential mpro inhibitors for the treatment of covid-19 by using systematic virtual screening approach. *Molecular diversity*, 25(1):383–401, 2021.

[52] Jyoti Choudhary, Shrivardhan Dheeman, Vipin Sharma, Prashant Katiyar, Santosh Kumar Karn, Manoj Kumar Sarangi, Ankit Kumar Chauhan, Gaurav Verma, and Nitin Baliyan. Insights of severe acute respiratory syndrome coronavirus (sars-cov-2) pandemic: a current review. *Biological procedures online*, 23(1):1–22, 2021.

[53] Linlin Zhang, Daizong Lin, Xinyuanyuan Sun, Ute Curth, Christian Drosten, Lucie Sauerhering, Stephan Becker, Katharina Rox, and Rolf Hilgenfeld. Crystal structure of sars-cov-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science*, 368(6489):409–412, 2020.

[54] Wenhao Dai, Bing Zhang, Xia-Ming Jiang, Haixia Su, Jian Li, Yao Zhao, Xiong Xie, Zhenming Jin, Jingjing Peng, Fengjiang Liu, Chunpu Li, You Li, Fang Bai, Haofeng Wang, Xi Cheng, Xiaobo Cen, Shulei Hu, Xiuna Yang, Jiang Wang, Xiang Liu, Gengfu Xiao, Hualiang Jiang, Zihe Rao, Lei-Ke Zhang, Yechun Xu, Haitao Yang, and Hong Liu. Structure-based design of antiviral drug candidates targeting the sars-cov-2 main protease. *Science*, 368(6497):1331–1335, 2020.

[55] Robert L Hoffman, Robert S Kania, Mary A Brothers, Jay F Davies, Rose A Ferre, Ketan S Gajiwala, Mingying He, Robert J Hogan, Kirk Kozminski, Lilian Y Li, et al. Discovery of ketone-based covalent inhibitors of coronavirus 3cl proteases for the potential therapeutic treatment of covid-19. *Journal of medicinal chemistry*, 63(21):12725–12747, 2020.

[56] Daniel W Kneller, Gwyndalyn Phillips, Hugh M O'Neill, Robert Jedrzejczak, Lucy Stols, Paul Langan, Andrzej Joachimiak, Leighton Coates, and Andrey Kovalevsky. Structural plasticity of sars-cov-2 3cl m pro active site cavity revealed by room temperature x-ray crystallography. *Nature communications*, 11(1):1–6, 2020.

[57] Sven Ullrich and Christoph Nitsche. The sars-cov-2 main protease as drug target. *Bioorganic & Medicinal Chemistry Letters*, page 127377, 2020.

[58] Rudolf A Römer, Navodya S Römer, and A Katrine Wallis. Flexibility and mobility of sars-cov-2-related protein structures. *Scientific reports*, 11(1):1–13, 2021.

[59] Annette Hegyi and John Ziebuhr. Conservation of substrate specificities among coronavirus main proteases. *Journal of general virology*, 83(3):595–599, 2002.

[60] Anjana Sharma and Satya P Gupta. Fundamentals of viruses and their proteases. In *Viral Proteases and Their Inhibitors*, pages 1–24. Elsevier, 2017.

[61] Kanchan Anand, John Ziebuhr, Parvesh Wadhwani, Jeroen R Mesters, and Rolf Hilgenfeld. Coronavirus main proteinase (3clpro) structure: basis for design of anti-sars drugs. *Science*, 300(5626):1763–1767, 2003.

[62] Jinzhi Tan, Koen H.G. Verschueren, Kanchan Anand, Jianhua Shen, Maojun Yang, Yechun Xu, Zihe Rao, Janna Bigalke, Burkhard Heisen, Jeroen R. Mesters, Kaixian Chen, Xu Shen, Hualiang Jiang, and Rolf Hilgenfeld. ph-dependent conformational flexibility of the sars-cov main proteinase (mpro) dimer: molecular dynamics simulations and multiple x-ray structure analyses. *Journal of molecular biology*, 354(1):25–40, 2005.

[63] Xiaoyu Xue, Hongwei Yu, Haitao Yang, Fei Xue, Zhixin Wu, Wei Shen, Jun Li, Zhe Zhou, Yi Ding, Qi Zhao, et al. Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. *Journal of virology*, 82(5):2515–2527, 2008.

[64] Bhupesh Goyal and Deepti Goyal. Targeting the dimerization of the main protease of coronaviruses: a potential broad-spectrum therapeutic strategy. *ACS combinatorial science*, 22(6):297–305, 2020.

[65] Kemper Talley and Emil Alexov. On the ph-optimum of activity and stability of proteins. *Proteins: Structure, Function, and Bioinformatics*, 78(12):2699–2706, 2010.

[66] Haitao Yang, Maojun Yang, Yi Ding, Yiwei Liu, Zhiyong Lou, Zhe Zhou, Lei Sun, Lijuan Mo, Sheng Ye, Hai Pang, George F. Gao, Kanchan Anand, Mark Bartlam, Rolf Hilgenfeld, and Zihe Rao. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proceedings of the National Academy of Sciences*, 100(23):13190–13195, 2003.

[67] Changkang Huang, Ping Wei, Keqiang Fan, Ying Liu, and Luhua Lai. 3c-like proteinase from sars coronavirus catalyzes substrate hydrolysis by a general base mechanism. *Biochemistry*, 43(15):4568–4574, 2004.

[68] Hui-Ping Chang, Chi-Yuan Chou, and Gu-Gang Chang. Reversible unfolding of the severe acute respiratory syndrome coronavirus main protease in guanidinium chloride. *Biophysical journal*, 92(4):1374–1383, 2007.

[69] Rolf Hilgenfeld. From sars to mers: crystallographic studies on coronaviral proteases enable antiviral drug design. *The FEBS journal*, 281(18):4085–4096, 2014.

[70] Peng Yuan, Zhou Yang, Han Song, Kai Wang, Yang Yang, Luyi Xie, Shilei Huang, Jia Liu, Lin Ran, and Zhenhui Song. Three main inducers of alphacoronavirus infection of enterocytes: Sialic acid, proteases, and low ph. *Intervirology*, 61(2):53–63, 2018.

[71] Dimas Suárez and Natalia Díaz. Sars-cov-2 main protease: A molecular dynamic study. *Journal of Chemical Information and Modeling*, 60(12):5815–5831, 12 2020.

[72] Natalia Díaz and Dimas Suárez. Influence of charge configuration on substrate binding to sars-cov-2 main protease. *Chemical Communications*, 57(43):5314–5317, 2021.

[73] Shilpa Sharma and Shashank Deep. ph effect on the dynamics of sars-cov-2 main protease (mpro). *bioRxiv*, 2020.

[74] Judith M White and Gary R Whittaker. Fusion of enveloped viruses in endosomes. *Traffic*, 17(6):593–614, 2016.

[75] Nabab Khan, Xuesong Chen, and Jonathan D Geiger. Role of endolysosomes in severe acute respiratory syndrome coronavirus-2 infection and coronavirus disease 2019 pathogenesis: implications for potential treatments. *Frontiers in pharmacology*, 11:1739, 2020.

[76] Leandro Jimenez, Ana Campos Codo, Vanderson de Souza Sampaio, Antonio ER Oliveira, Lucas Kaoru Kobo Ferreira, Gustavo Gastão Davanzo, Lauar de Brito Monteiro, João Victor Virgilio-da Silva, Mayla Gabriela Silva Borba, Gabriela Fabiano de Souza, et al. Acid ph increases sars-cov-2 infection and the risk of death by covid-19. *Frontiers in Medicine*, page 1358, 2021.

[77] Yury D Nechipurenko, Denis A Semyonov, Igor A Lavrinenko, Denis A Lagutkin, Evgenii A Generalov, Anna Y Zaitceva, Olga V Matveeva, and Yegor E Yegorov. The role of acidosis in the pathogenesis of severe forms of covid-19. *Biology*, 10(9):852, 2021.

[78] Brandon Michael Henry, Gaurav Aggarwal, Johnny Wong, Stefanie Benoit, Jens Vikse, Mario Plebani, and Giuseppe Lippi. Lactate dehydrogenase levels predict coronavirus disease 2019 (covid-19) severity and mortality: a pooled analysis. *The American journal of emergency medicine*, 38(9):1722–1726, 2020.

[79] Chunlong Ma, Michael Dominic Sacco, Brett Hurst, Julia Alma Townsend, Yanmei Hu, Tommy Szeto, Xiujun Zhang, Bart Tarbet, Michael Thomas Marty, Yu Chen, et al. Boceprevir, gc-376, and calpain inhibitors ii, xii

inhibit sars-cov-2 viral replication by targeting the viral main protease. *Cell research*, 30(8):678–692, 2020.

[80] Olga Abian, David Ortega-Alarcon, Ana Jimenez-Alesanco, Laura Ceballos-Laita, Sonia Vega, Hugh T Reyburn, Bruno Rizzuti, and Adrian Velazquez-Campoy. Structural stability of sars-cov-2 3clpro and identification of quercetin as an inhibitor by experimental screening. *International journal of biological macromolecules*, 164:1693–1703, 2020.

[81] Juliana C Ferreira and Wael M Rabeh. Biochemical and biophysical characterization of the main protease, 3-chymotrypsin-like protease (3clpro) from the novel coronavirus sars-cov 2. *Scientific reports*, 10(1):1–10, 2020.

[82] Khadim Sheikh Hamdullah, Arshad Tanzila, Mohammad ZainaSher, Arshad Iqra, and Hassan Mohtasheemul. ph dependent differential binding behavior of prtotease inhibitor molecular drugs for sars-cov-2. 2020.

[83] Erol C Vatansever, Kai S Yang, Aleksandra K Drelich, Kaci C Kratch, Chia-Chuan Cho, Kempaiah Rayavara Kempaiah, Jason C Hsu, Drake M Mellott, Shiqing Xu, Chien-Te K Tseng, et al. Bepridil is potent against sars-cov-2 in vitro. *Proceedings of the National Academy of Sciences*, 118(10), 2021.

[84] Thi Thanh Hanh Nguyen, Jong-Hyun Jung, Min-Kyu Kim, Sangyong Lim, Jae-Myoung Choi, Byoungsang Chung, Do-Won Kim, and Doman Kim. The inhibitory effects of plant derivate polyphenols on the main protease of sars coronavirus 2 and their structure–activity relationship. *Molecules*, 26(7):1924, 2021.

[85] Mohamed Hamdi, Hend Mohamed Abdel-Bar, Enas Elmowafy, Ahmed El-Khouly, Mai Mansour, and Gehanne AS Awad. Investigating the internalization and covid-19 antiviral computational analysis of optimized nanoscale zinc oxide. *ACS omega*, 6(10):6848–6860, 2021.

[86] Tonko Dražić, Nikos Kühl, Mila M Leuthold, Mira AM Behnam, and Christian D Klein. Efficiency improvements and discovery of new substrates for a sars-cov-2 main protease fret assay. *SLAS DISCOVERY: Advancing the Science of Drug Discovery*, page 24725552211020681, 2021.

[87] Keqiang Fan, Ping Wei, Qian Feng, Sidi Chen, Changkang Huang, Liang Ma, Bing Lai, Jianfeng Pei, Ying Liu, Jianguo Chen, et al. Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3c-like proteinase. *Journal of Biological Chemistry*, 279(3):1637–1642, 2004.

[88] Vito Graziano, William J McGrath, Ann Marie DeGruccio, John J Dunn, and Walter F Mangel. Enzymatic activity of the sars coronavirus main proteinase dimer. *FEBS letters*, 580(11):2577–2583, 2006.

[89] Shuai Chen, Li-li CHEN, Hai-bin LUO, Tao Sun, Jing Chen, Fei Ye, Jian-hua CAI, Jing-kang SHEN, Xu Shen, and Hua-liang JIANG. Enzymatic activity characterization of sars coronavirus 3c-like protease by fluorescence resonance energy transfer technique 1. *Acta Pharmacologica Sinica*, 26(1):99–106, 2005.

[90] Neha Verma, Jack A Henderson, and Jana Shen. Proton-coupled conformational activation of sars coronavirus main proteases and opportunity for designing small-molecule broad-spectrum targeted covalent inhibitors. *Journal of the American Chemical Society*, 142(52):21883–21890, 12 2020.

[91] Anna Pavlova, Diane L Lynch, Micholas Dean Smith, Jeremy D Smith, and James C Gumbart. Inhibitor binding influences the protonation state of histidines in sars-cov-2 main protease. *Biophysical Journal*, 120(3):204a–205a, 2021.

[92] Shiyong Fan, Dian Xiao, Yanming Wang, Lianqi Liu, Xinbo Zhou, and Wu Zhong. Research progress on repositioning drugs and specific therapeutic drugs for sars-cov-2. *Future medicinal chemistry*, 12(17):1565–1578, 2020.

[93] Zhenming Jin, Xiaoyu Du, Yechun Xu, Yongqiang Deng, Meiqin Liu, Yao Zhao, Bing Zhang, Xiaofeng Li, Leike Zhang, Chao Peng, et al. Structure of m pro from sars-cov-2 and discovery of its inhibitors. *Nature*, 582(7811):289–293, 2020.

[94] Camila Coelho, Gloria Gallo, Claudia B Campos, Leon Hardy, and Martin Würtele. Biochemical screening for sars-cov-2 main protease inhibitors. *PLoS One*, 15(10):e0240079, 2020.

[95] Željko Reiner, Mahdi Hatamipour, Maciej Banach, Matteo Pirro, Khalid Al-Rasadi, Tannaz Jamialahmadi, Dina Radenkovic, Fabrizio Montecucco, and Amirhossein Sahebkar. Statins and the covid-19 main protease: in silico evidence on direct interaction. *Archives of medical science: AMS*, 16(3):490, 2020.

[96] Aleix Gimeno, Júlia Mestres-Truyol, María José Ojeda-Montes, Guillem Macip, Bryan Saldivar-Espinoza, Adrià Cereto-Massagué, Gerard Pujadas, and Santiago Garcia-Vallvé. Prediction of novel inhibitors of the main protease (m-pro) of sars-cov-2 through consensus docking and drug reposition. *International journal of molecular sciences*, 21(11):3793, 2020.

[97] Neeraj Kumar, Amardeep Awasthi, Anchala Kumari, Damini Sood, Pallavi Jain, Taru Singh, Neera Sharma, Abhinav Grover, and Ramesh Chandra. Antitussive noscapine and antiviral drug conjugates as arsenal against covid-19: a comprehensive chemoinformatics analysis. *Journal of Biomolecular Structure and Dynamics*, pages 1–16, 2020.

[98] Carlos Moneriz and Cristian Castro-Salguedo. Fármacos prometedores y potenciales para el tratamiento de covid-19. *Revista chilena de infectología*, 37(3):205–215, 2020.

[99] Huihui Yang and Jinfei Yang. A review of latest research on mpro targeted sars-cov inhibitors. *RSC Medicinal Chemistry*, 2021.

[100] Ting Xiao, Mengqi Cui, Caijuan Zheng, Ming Wang, Ronghao Sun, Dandi Gao, Jiali Bao, Shanfa Ren, Bo Yang, Jianping Lin, et al. Myricetin inhibits sars-cov-2 viral replication by targeting mpro and ameliorates pulmonary inflammation. *Frontiers in Pharmacology*, 12:1012, 2021.

[101] Aweke Mulu, Mulugeta Gajaa, Haregewoin Bezu Woldekidan, et al. The impact of curcumin derived polyphenols on the structure and flexibility covid-19 main protease binding pocket: a molecular dynamics simulation study. *PeerJ*, 9:e11590, 2021.

[102] Lucia Silvestrini, Norhan Belhaj, Lucia Comez, Yuri Gerelli, Antonino Lauria, Valeria Libera, Paolo Mariani, Paola Marzullo, Maria Grazia Ortore, Antonio Palumbo Piccionello, et al. The dimer-monomer equilibrium of sars-cov-2 main protease is affected by small molecule inhibitors. *Scientific reports*, 11(1):1–16, 2021.

[103] Hylemariam Mihiretie Mengist, Tebelay Dilnessa, and Tengchuan Jin. Structural basis of potential inhibitors targeting sars-cov-2 main protease. *Frontiers in Chemistry*, 9, 2021.

[104] Caterina Vicidomini, Valentina Roviello, and Giovanni N Roviello. In silico investigation on the interaction of chiral phytochemicals from opuntia ficus-indica with sars-cov-2 mpro. *Symmetry*, 13(6):1041, 2021.

[105] Valentina Roviello, Domenica Musumeci, Andriy Mokhir, and Giovanni N Roviello. Evidence of protein binding by a nucleopeptide based on a thymine-decorated l-diaminopropanoic acid through cd and in silico studies. *Curr. Med. Chem*, 28, 2021.

[106] Luis Daniel Goyzueta-Mamani, Haruna Luz Barazorda-Ccahuana, Karel Mena-Ulecia, and Miguel Angel Chávez-Fumagalli. Antiviral activity of metabolites from peruvian plants against sars-cov-2: An in silico approach. *Molecules*, 26(13):3882, 2021.

[107] Bruno Rizzuti, Fedora Grande, Filomena Conforti, Ana Jimenez-Alesanco, Laura Ceballos-Laita, David Ortega-Alarcon, Sonia Vega, Hugh T Reyburn, Olga Abian, and Adrian Velazquez-Campoy. Rutin is a low micromolar inhibitor of sars-cov-2 main protease 3clpro: implications for drug design of quercetin analogs. *Biomedicines*, 9(4):375, 2021.

[108] Britton Boras, Rhys M Jones, Brandon J Anson, Dan Arenson, Lisa Aschenbrenner, Malina A Bakowski, Nathan Beutler, Joseph Binder, Emily Chen, Heather Eng, et al. Discovery of a novel inhibitor of coronavirus 3cl protease as a clinical candidate for the potential treatment of covid-19. *BioRxiv*, pages 2020–09, 2021.

[109] Maren de Vries, Adil Mohamed, Rachel Anne Prescott, Ana Valero-Jimenez, Ludovic Desvignes, Rebecca O'Connor, Claire Steppan, Joseph Cooper Devlin, Ellie Ivanova, Alberto Herrera, et al. A comparative analysis of sars-cov-2 antivirals in human airway models characterizes 3clpro inhibitor pf-00835231 as a potential new treatment for covid-19. *bioRxiv*, pages 2020–08, 2021.

[110] Dafydd R Owen, Charlotte MN Allerton, Annaliesa S Anderson, Lisa Aschenbrenner, Melissa Avery, Simon Berritt, Britton Boras, Rhonda D Cardin, Anthony Carlo, Karen J Coffman, et al. An oral sars-cov-2 mpro inhibitor clinical candidate for the treatment of covid-19. *medRxiv*, 2021.

[111] Matteo Pavan, Giovanni Bolcato, Davide Bassani, Mattia Sturlese, and Stefano Moro. Supervised molecular dynamics (sumd) insights into the mechanism of action of sars-cov-2 main protease inhibitor pf-07321332. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 36(1):1646–1650, 2021.

[112] Varada Anirudhan, Hyun Lee, Han Cheng, Laura Cooper, and Lijun Rong. Targeting sars-cov-2 viral proteases as a therapeutic strategy to treat covid-19. *Journal of Medical Virology*, 93(5):2722–2734, 2021.

[113] Mohammad Hassan Baig, Tanuj Sharma, Irfan Ahmad, Mohammed Abohashrh, Mohammad Mahtab Alam, and Jae-June Dong. Is pf-00835231 a pan-sars-cov-2 mpro inhibitor? a comparative study. *Molecules*, 26(6):1678, 2021.

[114] Hui Li, Andrew D Robertson, and Jan H Jensen. Very fast empirical prediction and rationalization of protein pka values. *Proteins: Structure, Function, and Bioinformatics*, 61(4):704–721, 2005.

[115] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John L Klepeis, Ron O Dror, and David E Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins: Structure, Function, and Bioinformatics*, 78(8):1950–1958, 2010.

[116] G Makov and MC Payne. Periodic boundary conditions in ab initio calculations. *Physical Review B*, 51(7):4014, 1995.

[117] Michael W Mahoney and William L Jorgensen. Diffusion constant of the tip5p model of liquid water. *The Journal of Chemical Physics*, 114(1):363–366, 2001.

[118] Vincent B Chen, W Bryan Arendall, Jeffrey J Headd, Daniel A Keedy, Robert M Immormino, Gary J Kapral, Laura W Murray, Jane S Richardson, and David C Richardson. Molprobity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 66(1):12–21, 2010.

[119] Oleg Trott and Arthur J Olson. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2):455–461, 2010.

[120] Jeff Racine. gnuplot 4.0: a portable interactive plotting utility. *Journal of Applied Econometrics*, 21(1):133–141, 2006.

[121] Christin Scheller, Finja Krebs, Robert Minkner, Isabel Astner, Maria Gil-Moles, and Hermann Wätzig. Physicochemical properties of sars-cov-2 for drug targeting, virus inactivation and attenuation, vaccine formulation and quality control. *Electrophoresis*, 41(13-14):1137–1151, 2020.

[122] Liqun Zhang and Matthias Buck. Molecular simulations of a dynamic protein complex: role of salt-bridges and polar interactions in configurational transitions. *Biophysical journal*, 105(10):2412–2417, 2013.

[123] Katarzyna Świderek and Vicent Moliner. Revealing the molecular mechanisms of proteolysis of sars-cov-2 m pro by qm/mm computational methods. *Chemical Science*, 11(39):10626–10630, 2020.

[124] Ben Hu, Hua Guo, Peng Zhou, and Zheng-Li Shi. Characteristics of sars-cov-2 and covid-19. *Nature Reviews Microbiology*, 19(3):141–154, 2021.

[125] Marco Cascella, Michael Rajnik, Abdul Aleem, Scott C Dulebohn, and Raffaela Di Napoli. Features, evaluation, and treatment of coronavirus (covid-19). *Statpearls [internet]*, 2022.

[126] Lalitha Guruprasad. Human sars cov-2 spike protein mutations. *Proteins: Structure, Function, and Bioinformatics*, 89(5):569–576, 2021.

[127] Public Health England. Investigation of sars-cov-2 variants of concern: technical briefings. *GOV. UK*, 2020.

[128] A Sarah Walker, Karina-Doris Vihta, Owen Gethings, Emma Pritchard, Joel Jones, Thomas House, Iain Bell, John I Bell, John N Newton, Jeremy Farrar, et al. Tracking the emergence of sars-cov-2 alpha variant in the united kingdom. *New England Journal of Medicine*, 385(27):2582–2585, 2021.

[129] Francesca Rovida, Irene Cassaniti, Stefania Paolucci, Elena Percivalle, Antonella Sarasini, Antonio Piralla, Federica Giardina, Josè Camilla Sammartino, Alessandro Ferrari, Federica Bergami, et al. Sars-cov-2 vaccine breakthrough infections with the alpha variant are asymptomatic or mildly symptomatic among health care workers. *Nature communications*, 12(1):1–7, 2021.

[130] Bo Meng, Steven A Kemp, Guido Papa, Rawlings Datir, Isabella ATM Ferreira, Sara Marelli, William T Harvey, Spyros Lytras, Ahmed Mohamed, Giulia Gallo, et al. Recurrent emergence of sars-cov-2 spike deletion h69/v70 and its role in the alpha variant b. 1.1. 7. *Cell reports*, 35(13):109292, 2021.

[131] Hiam Chemaitelly, Roberto Bertollini, and Laith J Abu-Raddad. Efficacy of natural immunity against sars-cov-2 reinfection with the beta variant. *New England Journal of Medicine*, 385(27):2585–2586, 2021.

[132] Kaiming Tao, Philip L Tzou, Janin Nouhin, Ravindra K Gupta, Tulio de Oliveira, Sergei L Kosakovsky Pond, Daniela Fera, and Robert W Shafer. The biological and clinical significance of emerging sars-cov-2 variants. *Nature Reviews Genetics*, 22(12):757–773, 2021.

[133] S Momsen Reincke, Meng Yuan, Hans-Christian Kornau, Victor M Corman, Scott van Hoof, Elisa Sánchez-Sendin, Melanie Ramberger, Wenli Yu, Yuanzi Hua, Henry Tien, et al. Sars-cov-2 beta variant infection elicits potent lineage-specific and cross-reactive antibodies. *Science*, 375(6582):782–787, 2022.

[134] Matthew McCallum, Alexandra C Walls, Kaitlin R Sprouse, John E Bowen, Laura E Rosen, Ha V Dang, Anna De Marco, Nicholas Franko, Sasha W Tilles, Jennifer Logue, et al. Molecular basis of immune evasion by the delta and kappa sars-cov-2 variants. *Science*, 374(6575):1621–1626, 2021.

[135] Alexander Wilhelm, Tuna Toptan, Christiane Pallas, Timo Wolf, Udo Goetsch, Rene Gottschalk, Maria JGT Vehreschild, Sandra Ciesek, and Marek Widera. Antibody-mediated neutralization of authentic sars-cov-2 b. 1.617 variants harboring l452r and t478k/e484q. *Viruses*, 13(9):1693, 2021.

[136] Matan Levine-Tiefenbrun, Idan Yelin, Hillel Alapi, Rachel Katz, Esma Herzel, Jacob Kuint, Gabriel Chodick, Sivan Gazit, Tal Patalon, and Roy Kishony. Viral loads of delta-variant sars-cov-2 breakthrough infections after

vaccination and booster with bnt162b2. *Nature medicine*, 27(12):2108–2110, 2021.

[137] Mengxin Zhang, Ying Liang, Dongsheng Yu, Bang Du, Weyland Cheng, Lifeng Li, Zhidan Yu, Shuying Luo, Yaodong Zhang, Huanmin Wang, et al. A systematic review of vaccine breakthrough infections by sars-cov-2 delta variant. *International journal of biological sciences*, 18(2):889, 2022.

[138] Petra Mlcochova, Steven A Kemp, Mahesh Shanker Dhar, Guido Papa, Bo Meng, Isabella ATM Ferreira, Rawlings Datir, Dami A Collier, Anna Albecka, Sujeet Singh, et al. Sars-cov-2 b. 1.617. 2 delta variant replication and immune evasion. *Nature*, 599(7883):114–119, 2021.

[139] Ying Liu and Joacim Rocklöv. The reproductive number of the delta variant of sars-cov-2 is far higher compared to the ancestral sars-cov-2 virus. *Journal of travel medicine*, 2021.

[140] World Health Organization et al. Covid-19 weekly epidemiological update, edition 115, 26 october 2022. 2022.

[141] Yusha Araf, Fariya Akter, Yan-dong Tang, Rabeya Fatemi, Md Sorwer Alam Parvez, Chunfu Zheng, and Md Golzar Hossain. Omicron variant of sars-cov-2: genomics, transmissibility, and responses to current covid-19 vaccines. *Journal of medical virology*, 94(5):1825–1832, 2022.

[142] Dandan Tian, Yanhong Sun, Huihong Xu, and Qing Ye. The emergence and epidemic characteristics of the highly mutated sars-cov-2 omicron variant. *Journal of Medical Virology*, 94(6):2376–2383, 2022.

[143] Liangwei Duan, Qianqian Zheng, Hongxia Zhang, Yuna Niu, Yunwei Lou, and Hui Wang. The sars-cov-2 spike glycoprotein biosynthesis, structure, function, and antigenicity: implications for the design of spike-based vaccine immunogens. *Frontiers in immunology*, 11:576622, 2020.

[144] Anastassios C Papageorgiou and Imran Mohsin. The sars-cov-2 spike glycoprotein as a drug and vaccine target: structural insights into its complexes with ace2 and antibodies. *Cells*, 9(11):2343, 2020.

[145] Daniel Martínez-Flores, Jesús Zepeda-Cervantes, Adolfo Cruz-Reséndiz, Sergio Aguirre-Sampieri, Alicia Sampieri, and Luis Vaca. Sars-cov-2 vaccines based on the spike glycoprotein and implications of new viral variants. *Frontiers in immunology*, 12, 2021.

[146] Venkata S Mandala, Matthew J McKay, Alexander A Shcherbakov, Aurelio J Dregni, Antonios Kolocouris, and Mei Hong. Structure and drug binding

of the sars-cov-2 envelope protein transmembrane domain in lipid bilayers. *Nature structural & molecular biology*, 27(12):1202–1208, 2020.

[147] Jin Chai, Yuanheng Cai, Changxu Pang, Liguo Wang, Sean McSweeney, John Shanklin, and Qun Liu. Structural basis for sars-cov-2 envelope protein recognition of human cell junction protein pals1. *Nature Communications*, 12(1):1–6, 2021.

[148] Mustafa Yalcinkaya, Wenli Liu, Mohammad N Islam, Andriana G Kotini, Galina A Gusarova, Trevor P Fidler, Eirini P Papapetrou, Jahar Bhattacharya, Nan Wang, and Alan R Tall. Modulation of the nlrp3 inflammasome by sars-cov-2 envelope protein. *Scientific reports*, 11(1):1–12, 2021.

[149] Martina Bianchi, Domenico Benvenuto, Marta Giovanetti, Silvia Angeletti, Massimo Ciccozzi, and Stefano Pascarella. Sars-cov-2 envelope and membrane proteins: structural differences linked to virus characteristics? *BioMed Research International*, 2020, 2020.

[150] Anas Shamsi, Taj Mohammad, Saleha Anwar, Samreen Amani, Mohd Shahnawaz Khan, Fohad Mabood Husain, Md Tabish Rehman, Asimul Islam, and Md Imtaiyaz Hassan. Potential drug targets of sars-cov-2: From genomics to therapeutics. *International Journal of Biological Macromolecules*, 177:1–9, 2021.

[151] Yi Zhang and Tatiana G Kutateladze. Molecular structure analyses suggest strategies to therapeutically target sars-cov-2. *Nature communications*, 11(1):1–4, 2020.

[152] Ramesh Kumar Paidi, Malabendu Jana, Sumita Raha, Mary McKay, Monica Sheinin, Rama K Mishra, and Kalipada Pahan. Eugenol, a component of holy basil (tulsi) and common spice clove, inhibits the interaction between sars-cov-2 spike s1 and ace2 to induce therapeutic responses. *Journal of Neuroimmune Pharmacology*, 16(4):743–755, 2021.

[153] Jong-Hwan Lee, Minsuk Choi, Yujin Jung, Sung Kyun Lee, Chang-Seop Lee, Jung Kim, Jongwoo Kim, Nam Hoon Kim, Bum-Tae Kim, and Hong Gi Kim. A novel rapid detection for sars-cov-2 spike 1 antigens using human angiotensin converting enzyme 2 (ace2). *Biosensors and Bioelectronics*, 171:112715, 2021.

[154] Ramesh K Paidi, Malabendu Jana, Rama K Mishra, Debashis Dutta, and Kalipada Pahan. Selective inhibition of the interaction between sars-cov-2 spike s1 and ace2 by spidar peptide induces anti-inflammatory therapeutic responses. *The Journal of Immunology*, 207(10):2521–2533, 2021.

[155] Manisha Prajapat, Nishant Shekhar, Phulen Sarma, Pramod Avti, Sanjay Singh, Hardeep Kaur, Anusuya Bhattacharyya, Subodh Kumar, Saurabh Sharma, Ajay Prakash, et al. Virtual screening and molecular dynamics study of approved drugs as inhibitors of spike protein s1 domain and ace2 interaction in sars-cov-2. *Journal of Molecular Graphics and Modelling*, 101:107716, 2020.

[156] Alex JB Kreutzberger, Anwesha Sanyal, Anand Saminathan, Louis-Marie Bloyet, Spencer Stumpf, Zhuoming Liu, Ravi Ojha, Markku T Patjas, Ahmed Geneid, Gustavo Scanavachi, et al. Sars-cov-2 requires acidic ph to infect cells. *Proceedings of the National Academy of Sciences*, 119(38):e2209514119, 2022.

[157] Cristian Privat, Sergio Madurga, Francesc Mas, and Jaime Rubio-Martinez. Unravelling constant ph molecular dynamics in oligopeptides with explicit solvation model. *Polymers*, 13(19):3311, 2021.

[158] Yixin Xie, Wenhan Guo, Alan Lopez-Hernadez, Shaolei Teng, and Lin Li. The ph effects on sars-cov and sars-cov-2 spike proteins in the process of binding to hace2. *Pathogens*, 11(2):238, 2022.

[159] Michał Rostkowski, Mats HM Olsson, Chresten R Søndergaard, and Jan H Jensen. Graphical analysis of ph-dependent properties of proteins predicted using propka. *BMC structural biology*, 11:1–6, 2011.

[160] Alexander D MacKerell Jr, Nilesh Banavali, and Nicolas Foloppe. Development and current status of the charmm force field for nucleic acids. *Biopolymers: original Research on biomolecules*, 56(4):257–265, 2000.

[161] Thomas J Dick and Jeffry D Madura. A review of the tip4p, tip4p-ew, tip5p, and tip5p-e water models. *Annual Reports in Computational Chemistry*, 1:59–74, 2005.

[162] Samuel Genheden and Ulf Ryde. The mm/pbsa and mm/gbsa methods to estimate ligand-binding affinities. *Expert opinion on drug discovery*, 10(5):449–461, 2015.

[163] Mario S Valdés-Tresanco, Mario E Valdés-Tresanco, Pedro A Valiente, and Ernesto Moreno. gmx_mmpbsa: a new tool to perform end-state free energy calculations with gromacs. *Journal of chemical theory and computation*, 17(10):6281–6291, 2021.

[164] David A Case, Robert E Duke, Ross C Walker, Nikolai R Skrynnikov, Thomas E Cheatham III, Oleg Mikhailovskii, Carlos Simmerling, Yi Xue, Adrian Roitberg, Sergei A Izmailov, et al. Amber 22 reference manual. 2022.

[165] D Horn and C-Chr Heuck. Charge determination of proteins with polyelectrolyte titration. *Journal of Biological Chemistry*, 258(3):1665–1670, 1983.

[166] Zoë Folchman-Wagner, Jennica Zaro, and Wei-Chiang Shen. Characterization of polyelectrolyte complex formation between anionic and cationic poly (amino acids) and their potential applications in ph-dependent drug delivery. *Molecules*, 22(7):1089, 2017.

[167] Susan Jones and Janet M Thornton. Principles of protein-protein interactions. *Proceedings of the National Academy of Sciences*, 93(1):13–20, 1996.

[168] Savio LC Woo, Alan S Lidsky, Flemming Güttler, T Chandra, and Kathryn JH Robson. Cloned human phenylalanine hydroxylase gene allows prenatal diagnosis and carrier detection of classical phenylketonuria. *Nature*, 306(5939):151–155, 1983.

[169] Nenad Blau, Francjan J Van Spronsen, and Harvey L Levy. Phenylketonuria. *The Lancet*, 376(9750):1417–1427, 2010.

[170] John J Mitchell, Yannis J Trakadis, and Charles R Scriver. Phenylalanine hydroxylase deficiency. *Genetics in medicine*, 13(8):607–617, 2011.

[171] Lihi Adler-Abramovich, Lilach Vaks, Ohad Carny, Dorit Trudler, Andrea Magno, Amedeo Caflisch, Dan Frenkel, and Ehud Gazit. Phenylalanine assembly into toxic fibrils suggests amyloid etiology in phenylketonuria. *Nature chemical biology*, 8(8):701–706, 2012.

[172] Marte I Flydal and Aurora Martinez. Phenylalanine hydroxylase: function, structure, and regulation. *IUBMB life*, 65(4):341–349, 2013.

[173] Robert Surtees and Nenad Blau. The neurochemistry of phenylketonuria. *European journal of pediatrics*, 159(2):S109–S113, 2000.

[174] Virender Singh, Ratan Kumar Rai, Ashish Arora, Neeraj Sinha, and Ashwani Kumar Thakur. Therapeutic implication of l-phenylalanine aggregation mechanism and its modulation by d-phenylalanine in phenylketonuria. *Scientific reports*, 4(1):1–8, 2014.

[175] Bibin G Anand, Kriti Dubey, Dolat S Shekhawat, and Karunakar Kar. Intrinsic property of phenylalanine to trigger protein aggregation and hemolysis has a direct relevance to phenylketonuria. *Scientific reports*, 7(1):1–9, 2017.

[176] Cong Guo, Zohar A Arnon, Ruxi Qi, Qingwen Zhang, Lihi Adler-Abramovich, Ehud Gazit, and Guanghong Wei. Expanding the nanoarchitectural diversity through aromatic di-and tri-peptide coassembly: Nanostructures and molecular mechanisms. *ACS nano*, 10(9):8316–8324, 2016.

[177] Weilin Chen, Zhiqiang Zhu, Chang Yin, Yibao Li, Yi Liu, Yuting Zhang, Yulan Fan, and Xiaolin Fan. Water-induced formation of a chiral phenylalanine derivative supramolecule. *Physical Chemistry Chemical Physics*, 20(6):4144–4148, 2018.

[178] Prabhjot Singh, Surinder K Brar, Manish Bajaj, Nikesh Narang, Venus S Mithu, Om P Katare, Nishima Wangoo, and Rohit K Sharma. Self-assembly of aromatic α-amino acids into amyloid inspired nano/micro scaled architects. *Materials Science and Engineering: C*, 72:590–600, 2017.

[179] Kevin E Riley and Pavel Hobza. Noncovalent interactions in biochemistry. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(1):3–17, 2011.

[180] Yasmine S Al-Hamdani and Alexandre Tkatchenko. Understanding non-covalent interactions in larger molecular complexes from first principles. *The Journal of chemical physics*, 150(1):010901, 2019.

[181] Marcos Juanes, Rizalina T Saragi, Walther Caminati, and Alberto Lesarri. The hydrogen bond and beyond: Perspectives for rotational investigations of non-covalent interactions. *Chemistry–A European Journal*, 25(49):11402–11411, 2019.

[182] Subhrakant Jena, Juhi Dutta, Kiran Devi Tulsiyan, Akshay Kumar Sahu, Shubhranshu Shekhar Choudhury, and Himansu S Biswal. Noncovalent interactions in proteins and nucleic acids: Beyond hydrogen bonding and π-stacking. *Chemical Society Reviews*, 2022.

[183] Tong Li, Xian-Mao Lu, Ming-Rong Zhang, Kuan Hu, and Zhou Li. Peptide-based nanomaterials: Self-assembly, properties and applications. *Bioactive Materials*, 11:268–282, 2022.

[184] David Arias-Olivares, Erna K Wieduwilt, Julia Contreras-García, and Alessandro Genoni. Nci-elmo: A new method to quickly and accurately detect noncovalent interactions in biosystems. *Journal of Chemical Theory and Computation*, 15(11):6456–6470, 2019.

[185] Sahin Uyaver, Helen W Hernandez, and M Gokhan Habiboglu. Self-assembly of aromatic amino acids: a molecular dynamics study. *Physical Chemistry Chemical Physics*, 20(48):30525–30536, 2018.

[186] Sahin Uyaver. Tyrosine, phenylalanine, and tryptophan undergo self-aggregation in similar and different manners. *Atmosphere*, 13(9):1448, 2022.

[187] Thanh D Do, William M Kincannon, and Michael T Bowers. Phenylalanine oligomers and fibrils: the mechanism of assembly and the importance of tetramers and counterions. *Journal of the American Chemical Society*, 137(32):10080–10083, 2015.

[188] George Kaminski and William L Jorgensen. Performance of the amber94, mmff94, and opls-aa force fields for modeling organic liquids. *The Journal of Physical Chemistry*, 100(46):18010–18013, 1996.

[189] Tian Lu and Feiwu Chen. Multiwfn: A multifunctional wavefunction analyzer. *Journal of computational chemistry*, 33(5):580–592, 2012.

[190] Deepak Tomar, Shilpi Chaudhary, and Kailash Chandra Jena. Self-assembly of l-phenylalanine amino acid: electrostatic induced hindrance of fibril formation. *RSC advances*, 9(22):12596–12605, 2019.

[191] Lukas J Gooßen, Nuria Rodríguez, and Käthe Gooßen. Carboxylic acids as substrates in homogeneous catalysis. *Angewandte Chemie International Edition*, 47(17):3100–3120, 2008.

[192] Bailin Xiang, Nguyen Thuy Vu Truong, Liyuan Feng, Tianzi Bai, Chao Qi, and Qingxia Liu. Study of the role of sodium citrate in bitumen liberation. *Energy & Fuels*, 33(9):8271–8278, 2019.

[193] Nikola Kallay and Egon Matijevic. Adsorption at solid/solution interfaces. 1. interpretation of surface complexation of oxalic and citric acids with hematite. *Langmuir*, 1(2):195–201, 1985.

[194] Yongpeng Chen, Jianguo Zhang, Zhixin Wang, and Zunning Zhou. Solvothermal synthesis of size-controlled monodispersed superparamagnetic iron oxide nanoparticles. *Applied Sciences*, 9(23):5157, 2019.

[195] Haider Bayat, Hashim Alhmoud, Mohammad Raoufi, Nicolas H Voelcker, and Holger Schönherr. Geometrical constraints of poly (diethylene glycol methyl ether methacrylate) brushes on spherical nanoparticles and cylindrical nanowires: Implications for thermoresponsive brushes on nanoobjects. *ACS Applied Nano Materials*, 3(4):3693–3705, 2020.

[196] Jhelai Sahadevan, R Sojiya, N Padmanathan, K Kulathuraan, MG Shalini, P Sivaprakash, and S Esakki Muthu. Magnetic property of fe2o3 and fe3o4 nanoparticle prepared by solvothermal process. *Materials Today: Proceedings*, 58:895–897, 2022.

[197] Wentao Wang, Bingtao Tang, Benzhi Ju, and Shufen Zhang. Size-controlled synthesis of water-dispersible superparamagnetic fe 3 o 4 nanoclusters and their magnetic responsiveness. *RSC advances*, 5(92):75292–75299, 2015.

[198] RJK Udayana Ranatunga, Robert JB Kalescky, Chi-cheng Chiu, and Steven O Nielsen. Molecular dynamics simulations of surfactant functionalized nanoparticles in the vicinity of an oil/water interface. *The Journal of Physical Chemistry C*, 114(28):12151–12157, 2010.

[199] Nan Xu, Hao Yan, Xu Jiao, Le Jiang, Ranran Zhang, Jingyun Wang, Zhongqun Liu, Ziyu Liu, Yingying Gu, Fangli Gang, et al. Effect of oh-concentration on fe3o4 nanoparticle morphologies supported by first principle calculation. *Journal of Crystal Growth*, 547:125780, 2020.

[200] HE Rizk and NE El-Hefny. Synthesis and characterization of magnetite nanoparticles from polyol medium for sorption and selective separation of pd (ii) from aqueous solution. *Journal of alloys and compounds*, 812:152041, 2020.

[201] Maryam Heidariramsheh, Mohammad Mahdi Dabbagh, Seyed Mohammad Mahdavi, and Ali Beitollahi. Morphology and phase-controlled growth of cuins2 nanoparticles through polyol based heating up synthesis approach. *Materials Science in Semiconductor Processing*, 121:105401, 2021.

[202] Joelle Medinger, Miroslava Nedyalkova, and Marco Lattuada. Solvothermal synthesis combined with design of experiments—optimization approach for magnetite nanocrystal clusters. *Nanomaterials*, 11(2):360, 2021.

[203] NH De Leeuw and JAL Rabone. Molecular dynamics simulations of the interaction of citric acid with the hydroxyapatite (0001) and (0110) surfaces in an aqueous environment. *CrystEngComm*, 9(12):1178–1186, 2007.

[204] Sanjoy Bandyopadhyay and Jnanojjal Chanda. Monolayer of monododecyl diethylene glycol surfactants adsorbed at the air/water interface: a molecular dynamics study. *Langmuir*, 19(24):10443–10448, 2003.

[205] Yusuke Hayashi, Yuka Nakajima, and Hirokazu Sugiyama. Computational screening of cryoprotective agents for regenerative medical products using quantum chemistry and molecular dynamics simulations. *Cryobiology*, 100:101–109, 2021.

[206] Anjali Gaur and Sundaram Balasubramanian. Liquid ethylene glycol: prediction of physical properties, conformer population and interfacial enrichment with a refined non-polarizable force field. *Physical Chemistry Chemical Physics*, 24(18):10985–10992, 2022.

[207] Anjali Gaur and Sundaram Balasubramanian. Conformer selection upon dilution with water: The fascinating case of liquid ethylene glycol studied via molecular dynamics simulations. *ChemistryOpen*, page e202200132, 2022.

[208] Kebede K Kefeni, Titus AM Msagati, and Bhekie B Mamba. Ferrite nanoparticles: synthesis, characterisation and applications in electronic device. *Materials Science and Engineering: B*, 215:37–55, 2017.

[209] Boris I Kharisov, HV Rasika Dias, and Oxana V Kharissova. Mini-review: Ferrite nanoparticles in the catalysis. *Arabian Journal of Chemistry*, 12(7):1234–1246, 2019.

[210] S Swathi, R Yuvakkumar, P Senthil Kumar, G Ravi, and Dhayalan Velauthapillai. Annealing temperature effect on cobalt ferrite nanoparticles for photocatalytic degradation. *Chemosphere*, 281:130903, 2021.

[211] Prashant Kumar, Himani Khanduri, Saurabh Pathak, Arjun Singh, GA Basheed, and RP Pant. Temperature selectivity for single phase hydrothermal synthesis of peg-400 coated magnetite nanoparticles. *Dalton Transactions*, 49(25):8672–8683, 2020.

[212] Armin Rezanezhad, Abdollah Hajalilou, Farhad Eslami, Elahe Parvini, Ebrahim Abouzari-Lotf, and Bagher Aslibeiki. Superparamagnetic magnetite nanoparticles for cancer cells treatment via magnetic hyperthermia: effect of natural capping agent, particle size and concentration. *Journal of Materials Science: Materials in Electronics*, 32(19):24026–24040, 2021.

[213] Mohammad Zaki Ahmad, Sohail Akhter, Gaurav Kumar Jain, Mahfoozur Rahman, Shadab Ahmad Pathan, Farhan Jalees Ahmad, and Roop Krishen Khar. Metallic nanoparticles: technology overview & drug delivery applications in oncology. *Expert opinion on drug delivery*, 7(8):927–942, 2010.

[214] Adam Schröfel, Gabriela Kratošová, Ivo Šafařík, Mirka Šafaříková, Ivan Raška, and Leslie M Shor. Applications of biosynthesized metallic nanoparticles–a review. *Acta biomaterialia*, 10(10):4023–4042, 2014.

[215] Harish Kumar, Nagasamy Venkatesh, Himangshu Bhowmik, Anuttam Kuila, et al. Metallic nanoparticle: a review. *Biomed. J. Sci. Tech. Res*, 4(2):3765–3775, 2018.

[216] David A Case, Thomas E Cheatham III, Tom Darden, Holger Gohlke, Ray Luo, Kenneth M Merz Jr, Alexey Onufriev, Carlos Simmerling, Bing Wang, and Robert J Woods. The amber biomolecular simulation programs. *Journal of computational chemistry*, 26(16):1668–1688, 2005.

[217] Junmei Wang, Wei Wang, Peter A Kollman, and David A Case. Antechamber: an accessory software package for molecular mechanical calculations. *J. Am. Chem. Soc*, 222(1), 2001.

[218] Araz Jakalian, Bruce L Bush, David B Jack, and Christopher I Bayly. Fast, efficient generation of high-quality atomic charges. am1-bcc model: I. method. *Journal of computational chemistry*, 21(2):132–146, 2000.

[219] Sergio Madurga, Miroslava Nedyalkova, Francesc Mas, and Josep Lluís Garcés. Ionization and conformational equilibria of citric acid: delocalized proton binding in solution. *The Journal of Physical Chemistry A*, 121(31):5894–5906, 2017.

[220] Ian T Jolliffe. *Principal component analysis for special types of data.* Springer, 2002.

[221] Florian Sittel, Abhinav Jain, and Gerhard Stock. Principal component analysis of molecular dynamics: On the use of cartesian vs. internal coordinates. *The Journal of Chemical Physics*, 141(1):07B605_1, 2014.

[222] Ch. Wohlfarth. Dielectric constant of ethane-1,2-diol: Datasheet from landolt-börnstein - group iv physical chemistry · volume 17: “supplement to iv/6” in springermaterials (https://doi.org/10.1007/978-3-540-75506-7_59). Copyright 2008 Springer-Verlag Berlin Heidelberg.

[223] NV Lifanova, TM Usacheva, and MV Zhuravlev. Equilibrium and relaxation dielectric properties of 1, 2-ethanediol. *Russian Journal of Physical Chemistry A*, 81:820–828, 2007.

[224] M.D. Lechner and C. Wohlfarth. *Static Dielectric Constants of Pure Liquids and Binary Liquid Mixtures: Supplement to Volume IV/17.* Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology - New Series. Springer Berlin Heidelberg, 2015.