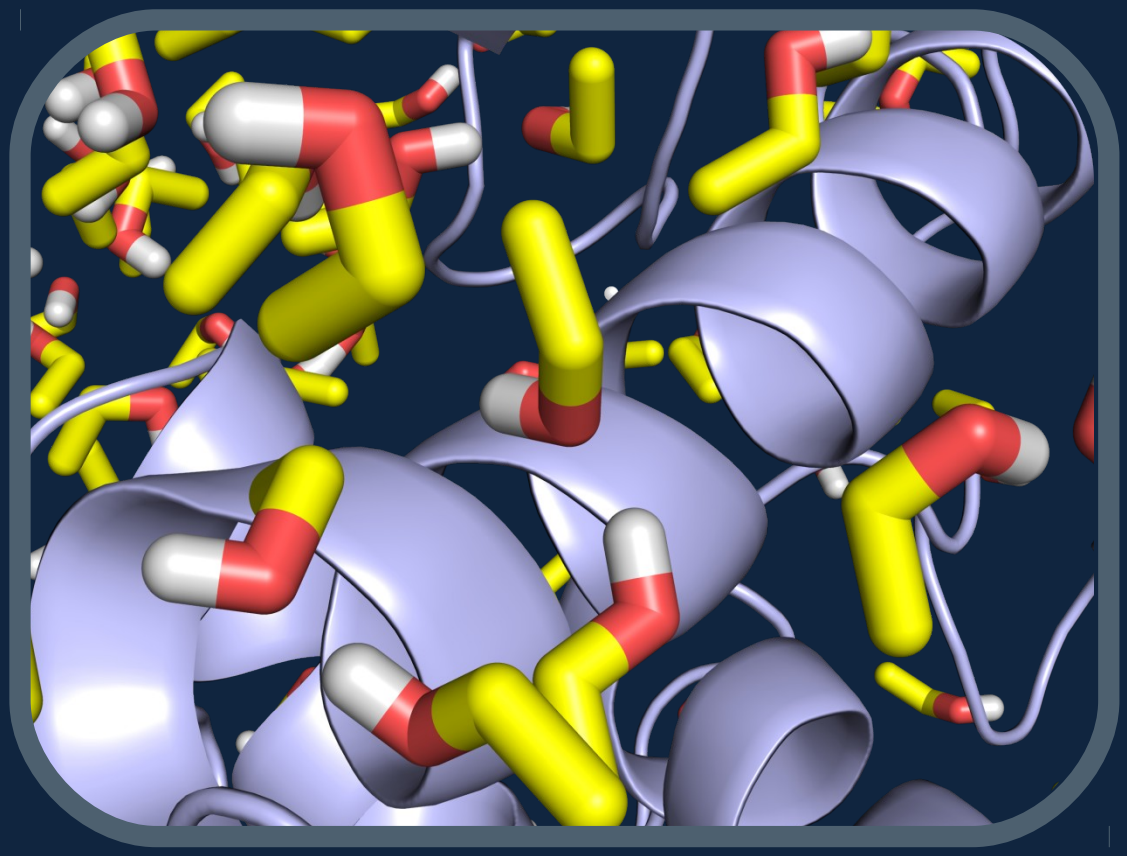
Molecular determinants of nonaqueous biocatalysis

A computational analysis

Diana Andreia Pereira Lousa



Dissertation presented to obtain the Ph.D degree in Biochemistry

Instituto de Tecnologia Química e Biológica | Universidade Nova de Lisboa

Oeiras,
March, 2013

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL

Knowledge Creation

Molecular determinants of nonaqueous biocatalysis

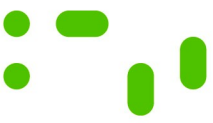
A computational analysis

Diana Andreia Pereira Lousa

Dissertation presented to obtain the Ph.D degree in Biochemistry

Instituto de Tecnologia Química e Biológica | Universidade Nova de Lisboa

Oeiras, March, 2013



INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL

Knowledge Creation

Molecular determinants of nonaqueous biocatalysis

A computational analysis

Diana Andreia Pereira Lousa

Supervisors: Professor Cláudio M. Soares and Doctor António M. Baptista

Dissertation presented to obtain the Ph.D degree in Biochemistry

The work presented in this thesis was financed by Fundação para a Ciência e a Tecnologia through grant SFRH/BD/28269/2006, with the support from the European Social Fund.

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

Contents

Acknowledgments

First, I would like to thank my supervisors, Prof. Cláudio M. Soares and Dr. António M. Baptista, for teaching me everything I know about science, and for their support and friendship. I have to thank them for making me believe, even when I couldn't see the light in the end of the tunnel. I am convinced that this is one of the most important qualities of a supervisor.

I am grateful to my colleagues from the Protein Modeling and Molecular Simulation groups for all their help and friendship, and for making this fun. I am proud to be part of the most "eccentric" (a.k.a. nerd) group of ITQB.

I also want to thank my parents, who always supported my decisions. I always felt that I could choose to be whatever I wanted (although most of the time I didn't know what that was).

I am thankful to my brothers for making me realize, early in life, that my athletic skills were so bad that I could only become an intellectual. By the time I was eight, after three consecutive last places in running events, I was pretty sure sports weren't my future. My brothers, of course, made sure I would stay on the right path, by constantly reminding me of my impressive record of three last places in a row.

I want to thank my friends and family for being there for me when I needed them.

Finally, I acknowledge Instituto de Tecnologia Química e Biológica for the excellent working conditions and Fundação para a Ciência e a Tecnologia for funding through grant SFRH/BD/28269/2006.

List of publications

Papers presented in this thesis

Lousa D., Cianci M., Helliwell J. R., Halling P. J., Baptista A. M., Soares C. M. (**2012**), "Interaction of counterions with subtilisin in acetonitrile: insights from molecular dynamics simulations", *Journal of Physical Chemistry B*, vol. 116, pp 5838-5848 (doi: http://dx.doi.org/10.1021/jp303008g)

Lousa D., Baptista A. M., Soares C. M. (**2012**), "Analyzing the molecular basis of enzyme stability in ethanol/water mixtures using molecular dynamics simulations", *Journal of Chemical Information and Modeling*, vol. 52, pp 465-473 (doi: http://dx.doi.org/10.1021/ci200455z)

Lousa D., Baptista A. M., Soares C. M. (**2011**) "Structural determinants of ligand imprinting: A molecular dynamics simulation study of subtilisin in aqueous and apolar solvents", *Protein Science*, vol. 20, pp 379-386 (doi: http://dx.doi.org/10.1002/pro.569)

Abstract

Over the last thirty years, the tremendous biotechnological potential of nonaqueous biocatalysis has boosted research efforts in this area. Numerous studies have tried to elucidate how enzymes work in these nonconventional media and many properties are now well understood. However, when this thesis was initiated, some aspects of this field were poorly characterized at the molecular level. In particular, the molecular determinants of protein-ion interactions, enzyme stability, and molecular memory, are important issues which were lacking a thorough molecular analysis. These three subjects are herein investigated using molecular simulation methodologies.

In chapter 3, we present a molecular dynamics (MD) simulation study of enzyme-ion interactions in a nonaqueous solvent. Because organic solvents are less polar than water, it is generally accepted that, in these media, ions bind very strongly to charged and polar groups, stabilizing the protein. However, it is hard to analyze enzyme-ion interactions with molecular detail using current experimental methods. Although this issue has been addressed in other simulation studies[1, 2], it was not the main subject of these studies and many questions remained opened. The present study was prompted by intriguing results from our experimental collaborators in the U.K. and Germany. They have determined the X-ray structure of subtilisin Carlsberg soaked in CsCl and acetonitrile, where several ion binding sites were clearly detected[3]. Given the inherent limitations of the crystallographic analysis then performed (crystallographic contacts, constrained protein, artificial electrostatic environment, lack of explicit dynamics, use of heavy cesium ions), we decided to pursue this characterization by simulating the behaviour of the protein and ions in acetonitrile solution. Towards this end, multiple MD

simulations in different conditions were performed, using the X-ray structure obtained by our co-workers[3] as a starting point. We observed that chloride ions tend to stay close to the protein surface, while cesium ions often move away. Similar cation distributions are found when sodium is used instead of cesium, which may thus be a reasonable model for more physiological ions. Nevertheless, sodium forms stronger ionic pairs with chloride, leading to a decreased interaction of the anion with the protein, which may explain the experimentally observed cation-dependent catalytic rate. Our simulations show that the crystal environment promotes the electrostatic stabilization of ion binding sites to an extent absent from the protein in solution. This study has therefore been able to provide useful insights into the interaction between subtilisin and its counterions in acetonitrile solution, which could not be obtained just by analysing the X-ray structure.

The fact that most enzymes are less stable in nonaqueous solvents than in water is a serious drawback of nonaqueous biocatalysis. In order to overcome this limitation, one needs to have a deep understanding into the molecular causes underlying this behaviour. In chapter 4, we have addressed this issue, by analysing the molecular determinants of enzyme stability in ethanol/water mixtures. Once again, MD simulations were used to gain detailed atomic insight into previous experimental findings. These experimental studies have shown that pseudolysin (PSL) is considerably more stable in ethanol/water solutions than other proteases, including thermolysin (TLN), which is very similar at the structural level[4]. Experiments also found that the C30-C58 disulfide bridge of PSL is important for its stability[5]. However, a molecular explanation for the observed behaviour was lacking. Towards this end, we performed μs-long MD simulations of PSL, TLN and the C58G mutant of PSL, in ethanol/water (25% v/v) and in pure water. Five independent replicates were

used for each condition, in order to have statistically meaningful results. Our results were in good agreement with the previous experimental findings. PSL was considerably more stable than TLN in the presence of ethanol and the abolishment of the disulfide bridge of PSL resulted in a stability decrease. Additionally, our simulations revealed that thermolysin has a higher tendency to interact with ethanol molecules (especially through van der Waals contacts) than pseudolysin. Ethanol molecules can, therefore, break intra-protein hydrophobic contacts, which in turn leads to the unfolding of TLN. The C58G mutant of pseudolysin undergoes larger conformational changes than the wild type enzyme. The mutant protein opens during the simulations, becoming more permeable to ethanol molecules, which accumulate in its interior and ultimately result in protein denaturation. These simulations were able to provide a molecular explanation for the previously observed stability difference between PSL and TLN in ethanol/water mixtures. Our findings may, therefore, be useful in the rational development of enzymes with increased stability in these media.

Chapter 5 describes a molecular analysis of ligand imprinting, a curious phenomenon observed in nonaqueous enzymology. This phenomenon was first reported by Klibanov and co-workers, who observed that the activity of subtilisin Carlsberg in apolar solvents could be enhanced by lyophilising the enzyme in the presence of competitive inhibitors[6]. Given that the ligand was removed before the enzyme was transferred to the nonaqueous media, the biocatalyst must have some sort of "memory" of the ligand-induced state. Ligand imprinting can be a useful strategy to increase enzyme activity and selectivity in apolar solvents. However, this requires a molecular explanation for this phenomenon, which was not available when we set out to do this work. Our aim was to elucidate the molecular determinants of ligand

imprinting using a molecular dynamics simulation methodology. In order to replicate the wet-lab experiments, we started by docking an inhibitor in the active site of subtilisin. The enzyme-ligand complex was placed in water and multiple MD simulations, extending for 20 ns, were carried out. The ligand was then removed and the resulting structure was subjected to 10 ns of MD simulations in hexane and in water. As a control, we performed multiple MD simulations in the same solvents, but starting from a "ligand untreated" structure of subtilisin. Our results show that pre-treating the enzyme with the inhibitor increases the probability of finding an open active site in hexane. In water the active site of the enzyme in the "pre-treated" simulations is indistinguishable from the control case. Analyzing the protein fluctuations, we could conclude that the observed behaviour reflects the fact that subtilisin is considerably more rigid in hexane than in water. Therefore, in the apolar solvent the enzyme tends to get "locked" in the conformation induced by the ligand. The molecule is, therefore, in an appropriate conformation to receive the incoming substrate (which is similar to the inhibitor) and this explains the observed rate enhancement. In water this phenomenon is not observed because the enzyme is so flexible that it rapidly deviates from the ligand-induced conformation. The findings of this study can contribute to the development of bioimprinting strategies to increase enzyme activity in apolar solvents.

Resumo

Ao longo dos últimos 30 anos, o enorme potencial biotecnológico da biocatálise em solventes não aquosos tem impulsionado a investigação nesta área. Vários estudos têm procurado elucidar como os enzimas funcionam nestes meios não convencionais e muitas das suas propriedades foram, entretanto, clarificadas. No entanto, quando esta tese foi iniciada, alguns aspectos da enzimologia não aquosa estavam insuficientemente caracterizados do ponto de vista molecular. Em particular, os determinantes moleculares das interacções entre o enzima e os iões, a estabilidade enzimática e a memória molecular são questões importantes, que necessitavam de uma análise molecular mais detalhada. Estes três problemas são aqui investigados, utilizando metodologias de simulação molecular.

No capítulo 3, apresentamos um estudo baseado em simulações de dinâmica molecular (MD) das interacções entre um enzima e os seus contra-iões, num solvente não aquoso. Uma vez que os solventes orgânicos são menos polares do que a água, é geralmente aceite que nestes meios os iões se ligam muito fortemente a grupos carregados e polares, estabilizando a proteína. No entanto, é difícil analisar interacções enzima-ião, com detalhe atómico, usando métodos experimentais. Apesar de outros estudos já se terem debruçado sobre este assunto[1, 2], este não era o ponto central destes estudos e muitas questões permaneceram abertas. O presente estudo foi motivado por resultados intrigantes obtidos pelos nossos colaboradores experimentalistas no Reino Unido e Alemanha, que determinaram a estrutura de raios-X da enzima subtilisina Carlsberg, *soaked* em CsCl e acetonitrilo[3]. Nessa estrutura, os locais de ligação dos iões na superfície da proteína são claramente visíveis. Dadas as limitações inerentes à análise cristalográfica

então realizada (contactos cristalográficos, proteína constrangida, ambiente electrostático artificial, dificuldade de caracterização das propriedades dinâmicas, uso de iões pesados), decidimos alargar este estudo, simulando o comportamento da proteína e dos iões numa solução de acetonitrilo. Para este fim, foram realizadas múltiplas simulações de MD, em diferentes condições, partindo da estrutura de raios-X obtida pelos nossos colaboradores[3]. Observou-se que os iões cloreto tendem a ficar perto da superfície da proteína, enquanto os iões césio muitas vezes se afastam. Os catiões sódio e césio distribuem-se de forma semelhante na superfície da proteína, pelo que o césio pode ser considerado um modelo razoável para iões com maior relevância biológica. No entanto, o ião sódio forma interacções mais fortes com o ião cloreto, levando a uma diminuição da interacção deste anião com a proteína, o que pode explicar o efeito de diferentes catiões na actividade catalítica, observado experimentalmente. As nossas simulações indicam que o ambiente cristalino promove a estabilização electrostática dos iões na superfície da proteína, criando locais de ligação artificiais, que não são observados em solução.

O facto de a maioria dos enzimas serem menos estáveis em solventes não aquosos do que em água é um dos grandes inconvenientes da biocatálise não aquosa. Para superar esta limitação, é necessário conhecer as causas moleculares deste comportamento. No capítulo 4, abordamos esta questão, analisando os determinantes moleculares da estabilidade da enzimática em misturas de etanol e água. Tal como no capítulo anterior, foram usadas simulações de MD para obter uma compreensão atómica detalhada de observações experimentais prévias. Estes estudos experimentais mostraram que a protease pseudolisina (PSL) é consideravelmente mais estável em misturas de etanol e água que outras proteases, incluindo a termolisina

(TLN), que é muito semelhante a nível estrutural[4]. Também foi demonstrado que a ponte persulfureto C30-C58, presente na PSL, é importante para a sua estabilidade[5]. No entanto, quando iniciámos este trabalho, as causas moleculares deste comportamento não eram conhecidas. Esta questão foi analisada por nós, usando simulações de MD, na escala dos μs, das enzimas PSL, TLN e do mutante C58G da PSL, em etanol/água (25% v/v) e em água pura. Foram usadas cinco réplicas independentes para cada condição, de forma a obter resultados estatisticamente significativos. Os resultados obtidos são concordantes com os resultados experimentais anteriores. Nas nossas simulações, a PSL mostrou-se consideravelmente mais estável do que a TLN na presença de etanol e a supressão da ponte persulfureto da PSL resultou numa diminuição da sua estabilidade. Adicionalmente, as simulações mostraram que a termolisina tem uma tendência maior para interagir com as moléculas de etanol (especialmente através de contactos de van der Waals) do que a pseudolisina. As moléculas de etanol podem, assim, destruir os contactos hidrofóbicos da TLN, levando ao seu *unfolding*. O mutante C58G da pseudolisina sofre maiores alterações conformacionais do que a enzima nativa. A proteína mutante abre durante as simulações, tornando-se mais permeável às moléculas de etanol, que se acumulam no seu interior e levam à sua desnaturação. As nossas simulações permitiram compreender as causas moleculares subjacentes à diferença de estabilidade entre a PSL e a TLN em misturas de etanol e água. Este conhecimento pode ser útil para o desenvolvimento racional de enzimas estáveis nestes meios.

O capítulo 5 descreve a análise microscópica da memória molecular, um fenómeno curioso observado em enzimologia não aquosa. Este fenómeno foi relatado pela primeira vez por Klibanov e seus colaboradores, que observaram que a actividade do enzima subtilisina Carlsberg, em solventes

apolares, pode ser aumentada liofilizando-o na presença de inibidores competitivos[6]. Dado que o inibidor é removido antes do enzima ser transferido para o meio não aquoso, o biocatalisador deve ter algum tipo de "memória" do estado induzido pelo ligando. A memória molecular pode ser explorada para aumentar a actividade e a selectividade de enzimas em solventes apolares. No entanto, isto requer uma explicação molecular para este fenómeno, que não existia na altura em que decidimos conduzir este estudo. O nosso objectivo era compreender os determinantes moleculares da memória molecular, utilizando uma metodologia baseada em simulações de dinâmica molecular. Com o intuito de replicar as experiências laboratoriais, começámos por encaixar um inibidor no centro activo da subtilisina, usando um método de *docking* molecular. O complexo enzima-ligando foi colocado em água e múltiplas simulações de MD, com uma extensão de 20 ns, foram realizadas. O ligando foi então removido e a estrutura resultante foi usada em 10 ns de simulações de MD, em hexano e em água. Como controlo, corremos simulações de MD nos mesmos solventes, mas partindo de uma estrutura da subtilisina que não teve contacto prévio com o ligando. Os nossos resultados mostram que o pré-tratamento do enzima com o inibidor aumenta a probabilidade de encontrar um centro activo aberto, em hexano. Em água, as simulações em que houve contacto com o inibidor são indistinguíveis do teste de controlo. Analisando as flutuações da proteína, pode-se concluir que o comportamento observado reflecte o facto de esta ser consideravelmente mais rígida em hexano do que em água. Assim, no solvente apolar o enzima tende a ficar "bloqueado" na conformação induzida pelo ligando. Quando o substrato (que é semelhante ao inibidor) é adicionado, a proteína encontra-se numa conformação apropriada para o receber, o que explica o aumento de actividade observado. Em água, este

fenómeno não é observado, porque o enzima é tão flexível que rapidamente se desvia da conformação induzida pelo ligando. Os resultados deste estudo podem contribuir para o desenvolvimento de estratégias de *bioimprinting*, que serão usadas para aumentar a actividade enzimática em solventes apolares.

List of symbols and abbreviations

Abbreviations

CALB	*Candida antarctica* lipase B
CD	circular dichroism
CE	continuum electrostatic
DSC	differential scanning calorimetry
FEP	free energy perturbation
FF	force field
fs	Femtoseconds
FTIR	Fourier transform infrared
GS	ground state
IL	ionic liquid
MC	Monte Carlo
MD	molecular dynamics
MM	molecular mechanics
µs	microseconds
nm	nanometers
NMR	nuclear magnetic resonance
NPT	isothermic-isobaric ensemble
ns	nanoseconds
PMF	potential of mean force
PSL	pseudolysin
QM	quantum mechanics
rdf	radial distribution function
rmsd	root mean square deviation
rmsf	root mean square fluctuation
$scCO_2$	super critical carbon dioxide
TI	tetrahedral intermediate
TLN	thermolysin

TS	transition state
v/v	ratio of volume

Latin symbols

$\boldsymbol{a}$	acceleration vector
A	Helmoltz free energy
G	Gibbs free energy
H	Hamiltonian function
k	Boltzmann constant
K	kinetic energy
k_{cat}	catalytic rate constant
K_d	dissociation constant
K_m	Michaelis constant
n	vector with all protonation states (n_1, n_2,…, n_s)
n_i	protonation state of titrable site i
P	pressure
q_i	partial atomic charge of atom i
$\boldsymbol{r}$	position vector
r_{ij}	interatomic distance between atoms i an j
S	entropy
T	absolute temperature
T	time
$\boldsymbol{v}$	velocity vector
V	potential energy
z	total charge of a protein molecule in protonic units
z_i	charge of a titrable site i in protonic units

Greek symbols

ε_o	electric permittivity of vacuum

ε_r	relative dielectric constant
κ	Debye inverse length
λ	coupling parameter

Chapter 1

Introduction

1.1 Biomolecular catalysis: How do enzymes work?

Enzymes are one of the most important and fascinating biological molecules. These biomolecular catalysts mediate almost all biological processes and without them life as we know it would not be possible. The most striking feature of enzymes is their ability to drastically increase the rates of the reactions they catalyze – rate enhancements of 10^6 to 10^{12}, relative to the uncatalyzed reaction, are commonly observed[7]. It is, therefore, not surprising that enzymes have gained so much attention from the scientific community over the last century. However, although tremendous progresses in this field have been achieved, the origin of the enormous catalytic power of enzymes is not yet fully understood.

1.1.1 Historical perspective

Although it is difficult to determine exactly when the history of enzymology started, some argue that it dates back to the mid 19th century, when Louis Pasteur proposed that fermentation was carried out by yeasts. He did not, however, recognize that this process was mediated by enzymes and thought there was a “vital force” (which he named “ferments”) that enabled yeasts and other living organisms to perform fermentation[7]. The term enzyme, which means "in leaven", in Greek, was introduced by Wilhelm Kühne, in 1878, to designate the entity responsible for the fermentation process observed by Pasteur[7]. To some researchers, vitalism, i.e., the idea that living organisms possess a vital force that is absent from inanimate substances, seemed absurd. One of the greatest opponents of Pasteur and his theory was Justus von Liebig, who believed that fermentation was nothing more than a complex chemical process[7]. The Liebig-Pasteur dispute was only settled in 1897, after both of them had died, when Eduard Büchener discovered that

fermentation could be carried out by yeast extracts without the presence of the actual cells. He proposed that the fermentative process was carried out by a soluble protein, that he designated zymase. Büchener's observations showed that the truth lied somewhere in the middle: fermentation is, indeed, a complex chemical process, as Liebig postulated, but the molecules which are responsible for this process are produced by living organisms, as Pasteur defended.

Büchener's experiment proved that enzymes (or zymases, as he called them) were essential in biological processes, like fermentation, but he did not elucidate how these molecules work. The first author that aimed to explain the mechanism of enzyme catalysis was Emil Fischer, who proposed that the substrate is able to bind to the active site of the enzyme because they have complementary shapes, like a "lock and key"[8]. It is now accepted that this hypothesis can explain the selectivity of enzymes, but not their efficiency. In 1932, Haldane suggested that "the key does not fit the lock perfectly but exercises a certain strain on it"[9], introducing a concept which is now known as ground-state destabilization. A different hypothesis was postulated by Linus Pauling, in 1946, according to which the enzyme is not complementary to the substrate molecules, but rather to the activated complex for the reaction[10]. Nowadays, this theory is called transition state stabilization. In 1958, Daniel Koshland introduced the "induced fit" model, which was based on the idea that the substrate may cause an appreciable change in the structure of the active site[11].

1.1.2 Current perspective(s)

Like all catalysts, enzymes are able to accelerate the reactions they catalyze because they decrease the activation barrier relative to the uncatalyzed reaction in solution. To elucidate the origin of the tremendous catalytic power of enzymes, one has to understand how they lower the reaction activation free energy. Figure 1.1 shows a schematic representation of the main free energy changes that occur during an enzymatic reaction.

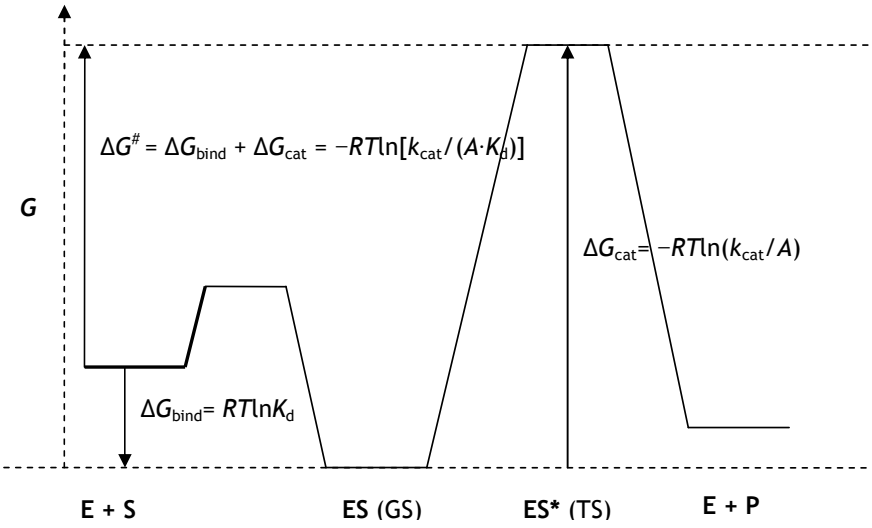


Figure 1.1. Schematic diagram illustrating the free energy changes associated with the main steps of an enzymatic reaction. The activation free energy ($\Delta G^{\#}$) is the sum of the free energy of substrate binding (ΔG_{bind}) plus the free energy associated with the conversion of the ground state into the transition state (ΔG_{cat}). In the equations displayed in the figure, k_{cat} is the rate constant, K_d is the dissociation constant, A is the pre-exponential factor in the Arrhenius equation, R is the gas constant and T is the temperature.

As can been seen in this figure, enzymes can increase the catalytic rate by increasing their binding affinity for the substrate (reducing K_d) and/or by reducing the activation free energy of going from the enzyme-substrate

complex (ES) to the activated complex (ES*) (reducing ΔG_{cat}). It is now generally accepted that many enzymes have evolved by optimizing the k_{cat}/K_d ratio. The effect of reducing K_d, as well as the mechanisms that are used to accomplish this reduction, are now fairly well understood. The real puzzle is to understand how an enzyme can decrease $\Delta G_{cat.}$

Turning once again to fig. 1.1, it becomes obvious that there are two different strategies that an enzyme can use to decrease ΔG_{cat}: destabilize the ground state (GS) or stabilize the transition state (TS).

The TS stabilization model assumes that the ΔG_{cat} of a given enzymatic reaction is smaller than the corresponding ΔG of the same reaction in solution because the TS is more stabilized by the enzyme than by the solvent. A possible explanation for the catalytic power of enzymes, which is sustained by Arieh Warshel and co-workers, is that they stabilize the transition state through electrostatic effects[12]. These effects would include protein charges, permanent and induced dipoles, and the solvation by bound water molecules[12]. One of the first studies to demonstrate the importance of the electrostatic effect in protein catalysis was carried out by Warshel and Levitt in 1976[12]. Using a theoretical approach, these authors showed that the carbonium ion intermediate, formed in the cleavage of a glycosidic bond by lysozyme, is mainly stabilized by electrostatic interactions with a charged aspartate residue. Since then, a large number of experimental and theoretical studies have provided evidences that support the importance of electrostatic effects in enzyme catalysis (an overview of these studies can be found in ref. 13). According to this hypothesis, in the active site of the enzyme, the dipoles (associated with polar and charged groups, and water molecules) are partially preorganized toward the TS charge distribution[13, 14].

On the other hand, in solution, the solvent dipoles need to reorganize in order to stabilize the TS, which has a large energetic cost[13, 14].

Some authors argue that enzymes work by ground-state destabilization. This hypothesis states that the substrate is not perfectly accommodated in the active site, but is under some kind of stress or strain, which is relieved when the reactant-state is converted into the transition-state. Haldane was the first to propose this model, when he stated that "Using Fischer's lock and key simile, the key does not fit the key perfectly but exercises a certain strain on it"[9]. This strain can arise from an unfavourable electrostatic interaction between the substrate and active site residues, a distortion of the substrate (or enzyme) bonds or bond angles, or a loss of substrate conformational freedom upon binding. William Jencks was one of the enzymologists who defended the importance of ground-state destabilization in enzyme catalysis. He proposed a model, which he named the "Circe effect" (after the Greek goddess), according to which the enzyme uses part of the substrate binding energy to destabilize the reactive group[15]. In 2000, Wu *et al* performed a combined crystallographic and theoretical study with a very proficient enzyme (orotidine monophosphate decarboxylase (ODCase)) and concluded that the catalytic power of this enzyme is mainly due to ground-state destabilization of the substrate[16]. They proposed that the enzyme uses the very strong binding energy of the phosphate and ribose groups of the substrate to destabilize the reacting orotate group, and considered this an example of the "Circe effect"[16]. However, later on, Warshel et al. refuted this interpretation and attributed the efficiency of ODCase to a high TS stabilization, on the basis of their own calculations (in which they used a different electrostatic treatment and considered a proton transfer from Lys72 to the substrate)[17].

A modified version of the "Circe effect", that was proposed by Menger[18], considers that both substrates and enzymes have a binding site and a reactive site. The interaction energy between the enzyme and the substrate is given by the sum of the interactions at the two distinct sites ($ES = ES_B + ES_R$). According to this "split-site" model (as the author named it), the interaction between the enzyme and the substrate is stabilized at the binding site (negative interaction free energy) and destabilized at the reactive site (positive interaction free energy). The interaction at the binding site remains unchanged as the reaction proceeds, whereas the interaction at the reactant site is altered, because chemical changes are taking place at this site. The effect of changing ES_B and ES_R on the reaction rate will depend on the profile of the reaction, i.e., on the relation between [S] and K_m (for a detailed analysis see ref. 18).

According to Bruice et al., enzymes increase the reaction rate by facilitating the formation of near attack conformations (NACs)[19-21]. NACs correspond to structures in which two reacting atoms are at a distance of van der Waals contact and at an angle resembling the one to be formed in the TS. It has been shown that, for some enzymes, the stabilization of NACs is the factor that gives the highest contribution to catalysis. However, the relative contribution of different catalytic factors varies between enzymes[19]. Some opponents of this model argue that the stabilization of NACs in some enzymatic reactions is a consequence of TS stabilization, rather than the cause of the catalytic power of enzymes[13].

One question that has been extensively debated, and is still open, is whether dynamic effects are important for enzyme catalysis. The controversy around this question is in part due to an unclear definition of what is meant by "dynamical effects on enzyme catalysis". In a strict sense, it means that the

enzyme has directional motions (different from the random thermal fluctuations that are also present in solution), which are directly coupled to the chemical reaction. Currently, there seems to be no experimental or theoretical evidence that this occurs[22]. However, there is no question that enzymes undergo large conformational changes (and in that sense they are "dynamic") that, in many cases, facilitate the catalytic process[23].

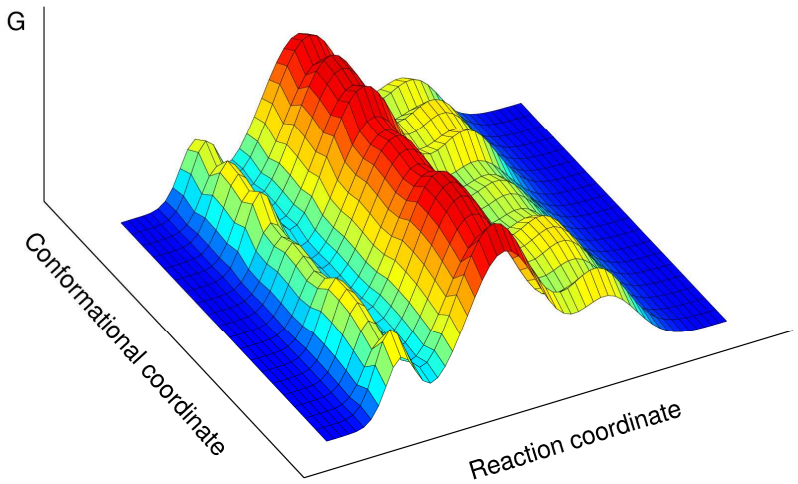


Figure 1.2. Schematic representation of the free-energy landscape of an enzyme reaction, according to the catalytic network model. Each intermediate of the catalytic process (reaction coordinate axis) is composed by an ensemble of conformations (conformational coordinate axis), forming a catalytic network with multiple alternative reaction paths.

Nowadays, it is commonly accepted that enzyme catalysis should be represented by a very rugged 3-dimensional free-energy landscape, with multiple minima and transition states (see fig. 1.2). According to this view, each intermediate of the chemical process is, in fact, an ensemble of conformational states, leading to several alternative reaction paths which,

collectively, form a catalytic network[24]. Before the reaction initiates, the free enzyme has multiple conformers in equilibrium and the most appropriate ones are selected by the incoming substrate upon binding[24, 25]. The binding of the substrate, in turn, can shift the populations of the different reaction intermediates[24-26]. This conformational dynamics enables the enzyme to stabilize the transition state(s) (even for enzymes catalyzing multiple chemical steps) and can also facilitate product release[24-26].

1.2 Enzymatic catalysis in nonaqueous media

Water is the most abundant solvent on our planet and the major constituent of living cells. Therefore, most enzymes have evolved in an aqueous environment and the common belief was that they could not catalyze reactions in nonaqueous media. However, this notion was proven wrong when it was shown that enzymes could function in a biphasic system comprised by water and a water-immiscible organic solvent[27]. This finding paved the way for nonaqueous enzymology, which emerged as a promising field.

Since the 1980s, several studies have shown that many enzymes could not only function in organic solvents but also display interesting novel properties[28-30]. These properties include altered substrate specificity and enantioselectivity, suppression of unwanted hydrolysis side-reactions, increased stability, and “molecular memory”. Additionally, solvents which are less polar than water can solvate hydrophobic substrates and/or products. Organic solvents are also considerably less prone to microbial contamination than water. Due to all these advantages, nonaqueous enzymology has a large technological potential and has been applied in industrial processes.

In addition to its biotechnological potential, nonaqueous enzymology is also interesting from a fundamental point of view. When analyzing the behavior of proteins, scientists often pay little attention to the role played by the solvent (in most cases water). Nevertheless, enzyme properties (e.g. native fold, stability, flexibility, activity, protonation and redox state, and interaction with counterions) are strongly influenced by the environment that surrounds them. Studying enzymes in nonaqueous solvents enables the comparison of their properties in different media and the analysis of how these properties are dependent of the characteristics of the solvent (polarity, hydrophilicity, density, viscosity, etc). Additionally, in this type of media, the amount of water present can be controlled, and by varying this property, one can determine the role of water in enzyme structure and function [31].

In a broad sense, the term "nonaqueous" can be applied to any media containing a considerable percentage of solvent(s) other than water and having properties distinct from those of pure water. A large number of nonaqueous solvents, with very different characteristics, have been used in enzyme catalysis, and they can be divided in three main classes: organic solvents, ionic liquids and supercritical fluids.

Organic solvents are carbon containing compounds, which exist in liquid form at room temperature and are usually volatile. They can be further subdivided into polar and apolar according to their dielectric constant. Polar organic solvents can be either protic or aprotic. Protic solvents have a labile hydrogen atom attached to a strongly electronegative atom (O or N), which can be donated in a hydrogen bond. Additionally, these O-H or N-H bonds can serve as a source of protons, although some protic solvents (e.g., ethanol) are very weak acids. Aprotic solvents, on the other hand, lack O-H or N-H bonds and, therefore, do not act as hydrogen donors in hydrogen bonds or

behave as acids. Another property that is used to classify organic solvents is hydrophilicity, which refers to the tendency of a molecule to be solvated by water.

Ionic liquids are salts which are liquid at, or close to, room temperature. These salts are usually composed by an organic cation and an inorganic anion. Through the combination of different cations and anions the properties of the liquid can be tuned. This makes ionic liquids very powerful solvents. Additionally, they have negligible vapour pressure and are non-flammable, which makes them more environmentally friendly than common organic solvents.

Supercritical fluids are substances which are above their critical temperature and pressure values. Under these conditions, they combine liquid and gas properties. These fluids are very good solvents, because they are able to diffuse like gases and dissolve solutes like liquids. Supercritical fluids like $scCO_2$, which are nontoxic and non-flammable, are emerging as promising new solvents and are being used in several industrial applications, including enzymatic processes[32].

Given the diversity of nonaqueous solvents and their divergent properties, enzymes display different behaviours in different types of nonaqueous solvents. The biotechnological potential of these catalysts can be increased by taking advantage of this fact, i.e., one can choose the most appropriate solvent (and control other reaction conditions) to obtain the desired enzyme properties (medium engineering).

1.2.1 Structural and dynamical properties of enzymes in nonaqueous solvents

Several studies have tried to elucidate how the solvent influences the structural properties of enzymes, including X-ray crystallographic studies of proteins soaked in organic solvents[33-42]. Given that it is very hard to crystallize proteins in these media, the crystals were grown in aqueous solution (most were cross-linked afterwards) and then soaked with the solvent of interest. The general conclusion of these studies was that no major conformational changes were observed when compared with the structures of the same proteins in aqueous solution. This is not surprising, because if there were major conformational changes after the soaking procedure, the crystal packing would not be maintained and the crystals would not diffract well and could even break. The structures obtained might not be identical to the structures that are found in an unconstrained organic solution, because the crystal packing (and in some cases the cross-linking procedure) can constrain the protein. Therefore, one should be cautious when using these structures to analyze the effect of the solvent on the structural properties of the enzyme.

Other techniques, such as nuclear magnetic resonance (NMR)[43], Fourier transform infrared (FTIR)[44-49], circular dichroism (CD)[50-53], differential scanning calorimetry (DSC)[48] and fluorescence[51-53] have been used to analyze protein structures in organic solvents. In a very interesting pioneering work, Griebenow and Klibanov showed, through FTIR experiments, that enzymes were denaturated in aqueous-organic mixtures but not in pure organic solvents[44]. Their results indicate that this is a kinetic effect, i.e., although proteins have a thermodynamic tendency to unfold in pure organic solvents, this does not happen because they loose conformational mobility and get trapped in the aqueous conformation. Similar observations were obtained by

other groups[50, 51, 53]. In another study, it was found that the there is a correlation between the stability of a protein structure in a given organic solvent and the activity of the dissolved enzyme[45]. This correlation was observed for enzyme solutions but not for enzyme suspensions[45]. Later on, the same authors found that there is another factor controlling the activity of an enzyme in organic solvents: flexibility. Using FTIR, they observed that the highest enzyme activity was obtained when both the structure and the conformational mobility of the enzyme were similar to the ones found in water[46]. In another study, carried out by the same group, glycosilation was used to modulate the structural dynamics of α-chymotrypsin, which was followed by H/D exchange kinetics. This analysis showed that at higher glycosilation levels the protein is more constrained, becoming more thermostable and shifting its optimum activity to higher temperatures[54].

Simulation studies have also been very useful in the analysis of the structural and dynamical properties of enzymes in nonaqueous solvents. The major findings of these studies are discussed below (see section 1.3).

1.2.2 Enzyme activity and selectivity in nonaqueous solvents

One of the major drawbacks of nonaqueous biocatalysis is the fact that enzymatic activity in these media is usually much lower than in water. This idea is, however, too simplistic, given that in many cases the reaction rates in nonaqueous solvents can be greatly improved by choosing the appropriate reaction conditions[31, 55]. The major causes underlying the decrease in enzyme activity upon transfer to nonaqueous solvents are structural changes[45, 50, 52, 53], active site blockage or distortion by solvent molecules[56-59], unfavourable substrate desolvation[59, 60], transition state destabilization[61-63] and restricted

conformational mobility[46, 62, 64]. Many studies have shown that the water content of the media has a very strong influence on the reaction rate (see ref. 31 for a review). In anhydrous conditions enzymes are very rigid and therefore their activity is low. As the water content increases the enzyme becomes more flexible and its activity increases, but at a certain point the enzyme has enough conformational mobility to start unfolding and, therefore, its activity diminishes. The optimum water percentage depends on the enzyme and the solvent[65]. In apolar solvents, the activity is higher at low water concentrations, whereas, in polar solvents, higher amounts of water result in a better performance[65]. This is due to the fact that polar solvents have a much higher ability to strip water molecules from the enzyme surface[65]. The study of Valivety et al., where five solvents with distinct polarities were tested, showed that if enzyme activity is analyzed as a function of water activity (a_w), instead of water concentration, it has an optimum at a_w around 0.5 in all the solvents analysed[65].

Several strategies have been used to increase enzyme activity and stability in nonaqueous solvents (see ref. [66] for a review), including chemical modification (e.g. PEGylation and attachment of hydrophobic moieties to lysine residues), protein engineering (through site-directed mutagenesis or directed evolution), and immobilization in solid matrices. Lyophilisation in the presence of excipients, such as salts, crown ethers, and cyclodextrins, has also proven to be a good strategy to prevent the destruction of the enzyme during the lyophilisation process and therefore enhance its catalytic power. In addition to the above mentioned approaches, which target the enzyme itself, some researchers have focused on engineering the solvent[66].

One of the reasons why nonaqueous enzymology is so attractive from the industrial viewpoint is the fact that the selectivity (including substrate,

enantiomeric, prochiral, regio and chemoselectivity) of an enzyme can be altered by the solvent. The most valuable type of enzymatic selectivity for synthetic applications is stereoselectivity, which includes enantioselectivity and prochiral selectivity. Therefore, the discovery that these properties can be strongly influenced by the solvent caused great excitement[67, 68]. Klibanov and co-workers were the first to realize this fact, when they observed that the stereoselectivity of subtilisin in the preparative synthesis of several peptides was completely altered when it was transferred from water to *tert*-amyl alcohol[69]. In another study, it was shown that subtilisin's enantioselectivity in the hydrolysis of 2-chloroethyl esters of *N*-acetyl-L- and D-amino acids decreases with the hydrophobicity of the solvent[70]. The same group has also shown that the preference of this enzyme for the *S* enantiomer in the transesterification of *N*-acetylalanine is inversely proportional to the polarity of the solvent. Their results indicate that this is probably a consequence of the lower enzyme flexibility in apolar solvents, which prevents the *R* enantiomer from accommodating properly in the binding pocket[71]. The relationship between enzyme flexibility and enantioselectivity was addressed in several studies and different conclusions were reached. Broos et al. observed that both the flexibility and the enantioselectivity of subtilisin increase with solvent polarity[64]. They proposed that the enzyme needs to have conformational mobility in order to maximize favourable interactions with the preferred enantiomer. Different results were obtained by Rariy et al., using the same enzyme but a different strategy and a different ester as substrate[72]. These authors found that the enantioselectivity of the enzyme decreases with increasing water content (which has been shown to result in higher enzyme mobility). They hypothesized that when the enzyme becomes more flexible it is able to bind and stabilize the two

enantiomers equally well, becoming less selective. As Broos pointed out[73], the discrepant results of the two groups are probably due to the fact that one group used a specific substrate, whereas the other group used an unspecific one. Other studies have found that (at least in some cases) the dependence of enzyme selectivity on the water content of the medium has a bell-shaped behaviour with the optimum enantioselectivity located at low water percentages[74, 75]. An MD simulation study performed in our group showed that cutinase displays a higher selectivity at water percentages in which the enzyme has structural and dynamic properties similar to the ones found in water[74]. It was also observed that the interaction of the catalytic histidine with the tetrahedral intermediate (which varies with the hydration level) is determinant for the selectivity of the enzyme. The studies described in this section and others[75-84] indicate that the mechanisms through which the solvent alters the enzyme selectivity are very diverse and cannot be generalized. The ability of biocatalysts to discriminate enantiomers can be greatly enhanced by the use of ionic liquids as reaction media[85] and this is emerging as a promising approach to obtain high selectivities.

1.2.3 The role of counterions

The notion that ions selectively bind to proteins dates back to the 19th century, when Franz Hofmeister observed that egg white proteins could be solubilised ("salted in") or precipitated ("salted out") by the addition of salt to the solution[86]. This effect is dependent on the composition and concentration of the salt. Ions can be ordered according to their ability to precipitate proteins and this classification is known as the "Hofmeister series".

Although the Hofmeister effect has been known for more than a century, its molecular determinants are only beginning to be understood (see [87] for a

review). The initial belief was that this effect was dependent on the ions' ability to alter the hydrogen bonding network of water. According to this criterion, ions are considered kosmotropes (structure makers) if they are strongly hydrated, stabilizing water structure, and chaotropes (structure breakers) if they are weakly hydrated, disrupting water structure. Kosmotropes are generally small and have a high charge density, which results in a high salting out ability, whereas chaotropes are usually large, with a low charge density, and are not effective in precipitating proteins. Studies over the last decade have shown that ions do not alter the hydrogen bonding network of water outside their direct vicinity (see, e.g., refs. [88, 89]). Additionally, thermodynamic analysis also indicates that there is no correlation between the stabilizing effects of solutes and their tendency to make or break water structure[90]. Therefore, the hypothesis that the ions effect on proteins arises from their influence on the structure of water is becoming discredited. The current view is that these effects are a result of the direct interaction between the ions and the protein surface and its first hydration shell.

To rationalize ion-specific effects, including the Hofmeister effect, Collins introduced a new concept: the *law of matching water affinities*[91-93]. According to this theory, ions tend to pair with oppositely charged ions of similar surface charge density. The rationale is that small ions with high charge densities (kosmotropes) have more negative hydration free energies than large ions with lower surface charge densities (chaotropes). Therefore, it is thermodynamically unfavourable for a kosmotrope to pair with a chaotrope, because the cost of desolvating the kosmotrope is higher than the gain of forming the ion pair. On the other hand, when two kosmotropes pair, their interaction is so strong (due to their high charge density) that it

compensates the cost of desolvation. In the case of an ion pair between two chaotropes, it is favourable because the desolvation penalty is low. Carboxilate, the anionic group present in aspartate and glutamate residues, is a kosmotrope, whereas the cationic protein groups (lysine ε-ammonium, arginine δ-guanidinium, and histidine imidazolium) are chaotropes. Therefore, based on the law of matching water affinities, chaotropic anions and kosmotropic cations have high affinities towards the protein charged residues. An interesting example of this effect is the different affinities of sodium and potassium ions towards protein surfaces. Hofmeister found that Na^+ has a higher "salting out" ability than K^+. This can be explained by Collins' law of matching water affinities. The charge density of carboxilate matches Na^+ better than K^+, and thus, the former ion binds the protein more effectively, leading to protein precipitation. Several theoretical and experimental[94-97] studies have indeed shown that Na^+ has a higher affinity towards the protein charged groups than K^+. This can explain why the intracellular concentration of sodium is considerably lower than that of potassium. The low concentration of sodium inside the cell prevents it from "salting out" proteins[93].

Counterions play a very important role in nonaqueous enzymology, because most organic solvents are less polar than water. Thus, the interaction between counterions and proteins is stronger in these media and protein charged groups will pair with oppositely charged ion, unless they can form intramolecular ion pairs. By binding to charged groups, counterions increase the enzyme stability and activity in nonaqueous solvents. Additionally, counterions can influence the protonation state of ionisable residues[98].

Several studies have shown that lyophilizing enzymes in the presence of high salt concentrations dramatically enhances their activity in nonaqueous

solvents[99-108]. It was also found that the best results are obtained when kosmotropic anions are used[101, 102, 104] especially when these are combined with chaotropic anions[102].

The mechanisms through which salts activate enzymes in nonaqueous solvents remain unclear. One possibility is that the salts (especially those containing kosmotropic anions) protect the enzyme during the lyophilisation process, by a mechanism of preferential hydration of the protein in the salt matrix[101]. Salts can also prevent the enzyme molecules from aggregating, facilitating the diffusion of substrates and products[108]. Another hypothesis is that the active site polarity increases due to the presence of the ions or water molecules recruited by them, stabilizing polar transition states, which are usually present in the reactions catalysed by proteases and lipases[100, 104]. A study using deuterium spin relaxation showed that lyophilizing subtilisin Carlsberg in the presence of CsF increases the mobility of water molecules bound to the enzyme, when it is immersed in hexane[107]. The authors propose that the increase in water mobility makes the enzyme more flexible and therefore more active.

Chapter 3 describes an extensive molecular dynamics simulation study of the interaction between subtilisin and counterions in acetonitrile, which complements a previous crystallographic analysis of the same system. The combination of these two studies provides a detailed molecular picture of the role of counterions in nonaqueous enzymology.

1.2.4 pH effects

Zacks and Klibanov found that the catalytic activity of porcine pancreatic lipase in an organic solvent is highly dependent on the pH of the aqueous

solution from which the enzyme was precipitated[28]. They observed that the enzyme activity as a function of the previous aqueous pH has a marked peak, which is coincident with the optimum pH of the enzyme in water. Given that most nonaqueous solvents are much less efficient in exchanging protons than water, and also due to the fact that enzymes are more rigid in these media, the enzyme ionisable groups tend to retain their original protonation state. This phenomenon is known as "pH memory", because the enzyme behaviour in the nonaqueous solvent depends on the previous history of the enzyme preparation. This effect has been observed for other systems[57]. However, some enzymes show a different behaviour. The optimum pH for alcohol dehydrogenase in water is 7.5, but the highest activity in heptane was obtained when the enzyme was lyophilized from a solution with a pH of 2[109]. This indicates that pH memory of enzymes in nonaqueous solvents is not as general as initially thought. Halling's group has shown that ions which are volatile or can be dissolved in a given nonaqueous media can override the pH memory of enzymes[110]. The authors describe various mechanisms that can lead to changes in the ionization state of the protein groups upon transfer to the nonaqueous media. Proton transfer can occur between neighbour ionisable residues. If the counterions present in the media are weakly acidic or basic, they will be able to exchange protons with the protein. The two mechanisms described above are reversible upon rehydration. On the other hand, if the media contains volatile weakly basic or acidic counterions, the protonation state of the protein will be irreversible changed.

1.2.5 Ligand imprinting

When an enzyme is placed in an anhydrous apolar solvent after being in contact with a competitive inhibitor, it appears to retain the state induced by

the ligand. This curious phenomenon is known as ligand imprinting, bioimprinting or ligand-induced enzyme memory, and was first reported by Russell and Klibanov in 1988[6]. They observed that the activity of subtilisin in *n*-octane was enhanced when it was lyophilized from a solution containing competitive inhibitors (that were subsequently removed). Moreover, these authors found that, in water, the pre-treatment with the ligand has no effect.

It has also been shown that pre-treating α-chymotrypsin with a given ligand increases the enzyme stereo-[111, 112] and substrate selectivity[112] in 2-propanol. In another study, the authors found that bioimprinting can also be applied to enzymes in aqueous media, if the enzyme is cross-linked with ethylene glycol dimethacrylate after the ligand treatment[113]. Recently, bioimprinting was combined with other strategies to enhance the activity and operational stability of *Geotrichum sp*. lipase[114]. These researchers were able to enhance the esterification rate of the enzyme more than 18 times and considerably increase its lifetime. This shows that ligand imprinting is a promising strategy to enhance enzyme activity.

The molecular determinants of ligand imprinting were investigated by us and the results of that study are described in chapter 5 of the present thesis.

1.3 Simulation studies of enzymes in nonaqueous solvents

In order to explore the biotechnological potential of nonaqueous enzymology, we need to have a detailed molecular picture of the interactions between the enzyme, the substrate, the organic solvent and other molecules present in the media (e.g. water, ions, co-solvents). It is also crucial to obtain a molecular understanding of how the solvent affects enzymatic properties,

such as activity, stability, water-dependence and molecular memory. Computer simulations can be very useful in this respect, because they enable a detailed atomic analysis of chemical/biochemical processes. MD simulations, in particular, are a valuable tool, because they model the time dependent conformational behaviour of molecules. Therefore, they can be used to explore the conformational space of the system under study. Additionally, they can give insights into the dynamical properties of the system.

MD simulation is now a well established methodology and there are a vast number of studies where this technique was used to analyse the molecular properties of proteins. In the great majority of these studies, water was used as a solvent (either explicitly or implicitly). This is not surprising because most proteins operate in aqueous environment and most in-vitro studies are also performed in these conditions. Therefore, much effort has been devoted to the development of simulation protocols and improvement of MD force fields, in order to accurately reproduce the properties of proteins in aqueous solution. On the other hand, simulations of proteins in nonaqueous solvents are much less common. The first reason for this is the fact that nonaqueous enzymology, despite being a promising field, is still marginal compared with its aqueous counterpart. Additionally, only a small number of nonaqueous solvents have been parameterized for MM force fields and the validity of these parameterizations has not been as extensively tested as in the case of water.

1.3.1 Setup challenges

Simulating enzymes in nonaqueous solvents is more challenging than performing simulations of proteins in water. One of the challenges is the fact that proteins are not soluble in most organic solvents and, even when they are, they tend to be unstable or inefficient. Therefore, in experimental studies and technological applications of enzymes in nonaqueous solvents, the biocatalysts are found in many different forms: suspended, adsorbed to a solid surface, immobilized in a solid matrix, or in the form of cross-linked enzyme crystals (CLECs). These enzyme preparations are difficult to simulate, mainly because they are not fully characterized at the molecular level. Additionally, some of the substances which are used in these preparations require specific force field parameters, which in many cases are not available. Therefore, the most common strategy is to simulate the free enzyme, surrounded by solvent molecules. Although this is not identical to the experimental conditions, it can still give valuable insights into the molecular-level effect of the solvent on enzyme properties.

Another issue that cannot be easily addressed when simulating enzymes in nonaqueous solvents is the determination of the protonation states of ionisable groups. As has been mentioned above, it has been shown that enzymes in nonaqueous solvents tend to retain the protonation state found in the aqueous solution from which they were lyophilized (pH memory). However, this is not always true and pH memory can disappear under certain conditions[98, 110]. Given that the protonation states of ionisable groups in nonaqueous solvents are very hard to predict, the simplest approach is to assume pH memory and use the protonation states obtained for the protein in water.

The initial placement of water molecules and counterions in nonaqueous simulations also deserves special attention. In apolar media, water molecules and counterions tend to be firmly attached to the protein surface. If they were randomly placed in the simulation box, it would take a long time for them to find the most attractive binding sites. One way to overcome this limitation is to start with the water molecules and ions attached to the protein surface. The water binding sites can be found by performing a short MD simulation of the protein in water, followed by a simulated annealing step[1]. In this step, the temperature of the system is progressively reduced, allowing the water molecules to find the most favourable binding sites.

In aqueous solutions, counterions tend to be dispersed in solution and they are not firmly attached to the protein. In apolar solvents, on the other hand, ions tend to form strong interactions with protein charged groups. Thus, they stabilize charged residues that cannot form intra-protein salt bridges. One approach that can be used to find the positions of counterions is to dock them on the protein surface (using a molecular docking algorithm) until all the protein charges have been neutralized. A methodology that has been used to do this was developed in our group and is explained in detail in ref. 1 and in chapter 3 of this thesis.

1.3.2 Protein structure

The limitations of experimental methods make it difficult to obtain a detailed molecular description of the structural properties of proteins in a nonaqueous solution (see above). MD simulations have proven to be very useful in the elucidation of this question[1, 2, 115-121].

The structural properties of a protein in MD simulations are usually analysed by measuring properties such as the root mean square deviation (rmsd) from the X-ray structure, radius of gyration, solvent accessible surface, and secondary structure content. Using these measures, several simulation studies have shown that protein structures in pure organic solvents are usually different from the ones observed in water[1, 115-117, 121]. Their radius of gyration and solvent accessible surface (especially hydrophobilic) tend to be smaller in apolar than in aqueous environments[1, 115, 116, 119]. Hartsough and Merz observed that the secondary structure of bovine pancreatic trypsin inhibitor (BPTI) is retained both in water and chloroform simulations[115]. However, the simulation times that were used were very short[115] (for current standards) and probably were not sufficient to observe secondary structure changes, which occur on longer time scales. Simulations of cutinase and ubiquitin, performed in our group, revealed that the secondary structure of these proteins in water and pure hexane shows some differences[1].

Polar solvents also influence the structural properties of proteins, because they can bind the protein through hydrogen bonding. The structural properties of subtilisin Carlsberg in water and acetonitrile have been studied using MD simulations[120] The authors observed that the enzyme undergoes large conformational changes in acetonitrile simulations, when compared with the control simulations in water[120]. The simulations indicate that acetonitrile molecules can replace protein–bound water molecules and this leads to the disruption of the protein structure[120]. The protein becomes more open and acetonitrile is then able to penetrate into the protein core, further disrupting the structure[120].

1.3.3 Protein flexibility

The first MD simulation study of a protein using an explicit nonaqueous solvent was performed by Hartsough and Merz, who compared the behaviour of bovine pancreatic trypsin inhibitor (BPTI) in water and chloroform[122]. Using ps-long MD simulations at different temperatures, they observed that the protein flexibility (especially of side-chains) is higher in water than in chloroform, even when the temperature of the organic solution is 50 ºC higher than that of the aqueous medium[122].

Several other simulation studies have corroborated the notion that enzymes are less flexible in apolar solvents than in water[1, 116, 119, 121]. It has been shown that the structures of cutinase and ubiquitin have smaller fluctuations in hexane than in water, indicating that the enzymes are less flexible in the organic solvent[1]. A study on triosephosphate isomerase also showed that its conformational mobility in pure decane is reduced relative to the aqueous situation[121]. Similar findings were made by Pleiss and co-workers using *Candida antarctica* lipase B (CALB) in apolar solvents[119].

In a simulation study of subtilisin Carlsberg, the protein displayed higher fluctuations in pure acetonitrile than in water[120]. This result led the authors to conclude that the enzyme is very flexible in acetonitrile[120], showing an atypical behaviour for an enzyme in a nonaqueous solvent. However, the large fluctuations observed in acetonitrile are located in very specific regions of the protein. Given that the protein had large conformational changes in this solvent, this could influence the calculation of the fluctuations, which, thus, may not represent increased protein flexibility.

1.3.4 Formation of salt bridges and intra-protein hydrogen bonds

As has been described in the previous sections, MD simulations have shown that the structural and dynamical properties of enzymes in nonaqueous solvents are different from the ones observed in water. MD simulations have also helped to understand the molecular determinants of this behaviour and several studies have shown that salt bridges and hydrogen bonds are two of the most important factors. These studies have shown that the hydrophilic solvent accessible surface of proteins is larger in water that in apolar solvents[1, 115-117, 119]. This happens because, in water, the polar surface side-chains like to be in contact with the solvent, whereas in apolar media they avoid being exposed, forming polar interactions with other side chains or folding back onto the surface of the protein[1, 115-117, 119]. The first simulation studies of a protein in an explicit organic solvent showed that the intramolecular hydrogen bonding network in water and in apolar solvents was quite different [115, 117], with the number and persistence of hydrogen bonds being much higher in apolar solvents than in water. Additionally, in apolar media, there are numerous hydrogen-bonds linking charged side-chains with other polar side chain or main chain groups, which explains the reduced flexibility and the higher thermal stability commonly observed for proteins in pure apolar solvents. Numerous simulation studies have also shown that the number and persistence of intra-protein hydrogen bonds decreases with increasing water concentration in apolar media[1, 2, 115, 121, 122].

1.3.5 Protein-solvent interactions

One of the goals of the study of enzymes in nonaqueous solvents is the characterization of protein-solvent interactions. MD simulations, like the ones described below, have helped to elucidate this question.

Hartsough and Merz found that both water and chloroform molecules are considerably less mobile when they are in contact with BPTI relative to the bulk solution[115]. Whereas for water this is not surprising, because it can form strong interactions with the protein, in the case of chloroform the explanation for the reduced mobility near the protein is not obvious. The authors proposed that the protein blocks the solvent molecules from moving in certain directions, thus reducing its mobility[122]. The structure of the solvents around selected residues was also evaluated in this study, using radial distribution functions (rdf)[122]; and, as expected, water accumulated near charged residues. Chloroform, on the hand, did not have a clear structure around most residues, although there are some cases in which distinct peaks were observed in the rdf. None of the solvents displayed an ordered structure around nonpolar residues. Their analysis also showed that the water molecules that were present in chloroform simulations never abandoned the protein surface. Other MD simulation studies have also shown that the residence times of protein-bound water molecules is higher in the presence of organic solvents than in pure water[1, 2, 121].

In a recent study, the protein-solvent interactions of triosephosphate isomerase in water/decane mixtures was analysed through MD simulations [121]. These simulations indicate that the distribution of water molecules around the protein surface was more specific than the distribution of decane molecules[121]. Nevertheless, the authors were able to identify several

hydrophobic regions of the protein surface in which the organic solvent tends to accumulate[121].

Polar solvents, like DMSO and acetonitrile, establish stronger interactions with proteins than apolar ones. The protein-solvent interactions of cytochrome P450 BM-3 in a DMSO/water mixture (14% v/v) have been analysed using MD simulations[118]. It was observed that the concentration of DMSO is larger at the protein surface than in the bulk solution, especially at protein concave regions. An accumulation of DMSO molecules close to the substrate binding site was observed. However, DMSO molecules were not able to penetrate into the active site during the simulation. In a subsequent study, the same authors analysed the behaviour of the F87A mutant of P450 BM-3 in the presence of (14% v/v) DMSO. This mutation is known to severely reduce the enzyme activity in this medium[123]. Once again, DMSO molecules were not able to diffuse into the active site cavity during the 15 ns of simulation[123]. The authors postulated that, given the nature and the small size of DMSO molecules, they should be able to penetrate into the active site and argue that on longer time scales this would occur[123]. To test the effect of DMSO on the enzyme activity, they introduced one or three molecules of this solvent directly into the active site cavity of both the wild type and mutant protein[123]. In the wild type enzyme, all the DMSO molecules remained far from the catalytic heme iron, whereas in the mutant some molecules approached the Fe atom[123]. Therefore, it seems that the F87 residue is important for preventing the access of DMSO molecules to the heme group[123]. The DMSO molecules that approach the heme disturb the interaction between the Fe atom and a coordinated water molecule, which can explain the reduced activity of the mutant in these conditions[123].

The effect of the interaction between subtilisin and the polar solvent, acetonitrile, have also been analysed using MD simulations[120]. The authors of this study found that, during the simulation, acetonitrile strips water molecules from the protein surface and then starts interacting with the enzyme[120]. The protein-acetonitrile interactions lead to conformational changes, which enable more acetonitrile molecules to penetrate into the enzyme and make it more unstable.[120]

1.3.6 Effect of water concentration and solvent polarity

Several experimental studies have shown that the water content and the polarity of the organic solvent have a dramatic influence on the properties of enzymes in nonaqueous solvents[65, 124-126]. However, the molecular causes of these effects are difficult to analyse using experimental techniques. Therefore, researchers have turned to simulation and, over the last decade, a considerable number of MD simulation studies have shed light on this issue.

In a study performed by our group, the behaviour of cutinase and ubiquitin in hexane, at different hydration conditions, was investigated using a large number of ns-long MD simulations[1]. It was observed that hydration has a profound effect on protein stability and flexibility. There is an optimum water concentration (~10% w/w) at which the enzyme properties are similar to the ones found in pure water[1]. At lower water concentrations the enzyme is very rigid and when the water amount is high the enzyme starts to unfold. These results explain the bell-shaped dependence of enzymatic activity on hydration. When the medium is too dry, the enzyme lacks flexibility and therefore cannot efficiently catalyse the reaction. As the water content increases, the protein becomes more flexible and its activity increases. After

the optimum water concentration has been reached, the protein starts to unfold and its activity decreases again[1]. Consistently with other MD simulation studies[115, 117, 121, 122], it was observed that the number of intra-protein hydrogen bonds decreases when the amount of water increases, which explains the reduced of flexibility observed in the absence of water[1]. The location and behaviour of water molecules was also analysed and it was found that they tend to stay bound to the protein surface and are mainly located near polar groups[1]. When the number of water molecules in the simulation is large, some of them start to detach from the protein and temporarily migrate to the solvent, either as isolated molecules or in clusters[1].

In another study performed in our laboratory, the hydration mechanisms of an enzyme in organic solvents with different polarities have been compared[2]. The serine protease cutinase was used as a model and five solvents with different polarities, at different hydration conditions, were analysed. It was observed that the nature of the solvent strongly influences the behaviour of water molecules and counterions. In apolar solvents, water molecules and counterions remained firmly attached to the protein throughout the simulations, whereas in more polar solvents their interaction with the protein is much weaker and they frequently migrated to the solution bulk. Polar solvents are able to replace water molecules at the protein surface and have a much higher ability to strip water molecules from the enzyme. To obtain the same amount of water bound to the enzyme that is found in apolar solvents, it is necessary to add very high amounts of water. These results are in agreement with another MD study, in which subtilisin BPN' was simulated in three organic solvents (octane, tetrahydrofuran, and acetonitrile)[127]. The residence time of water molecules on the protein surface also varies with the

nature of the solvent[2]: in apolar solvents, water molecules have longer residence times. It was also observed that the solvents have different effects on the protein structure and that the optimum amount of water for protein stability also depends on the solvent[2]. The more apolar the solvent, the less water is required to stabilize the structure[2]. This reflects the fact that more apolar solvents do not strip water from the protein surface[2]. Interestingly, despite these differences, it was observed that water tends to bind in the same regions of the protein surface in polar and apolar solvents, forming clusters around the polar and charged groups of the protein[2]. The size and number of these clusters depends on the nature of the solvent and on the water content of the media[2].

The effect of solvents with different polarity on protein structure and dynamics has also been investigated by Pleiss and co-workers[119]. Simulations in water, methanol, chloroform, isopentane, toluene and cyclohexane were performed, using CALB as a model enzyme. They observed that the enzyme structure was maintained throughout the 2.5 ns of simulation, in all the solvents[119]. Their results also show that the flexibility of the enzyme increases with increasing solvent polarity. The decreased flexibility observed in apolar solvents seems to be a consequence of the formation of a network of water molecules with long residence times at the protein surface[119].

Given that the amount of water that can associate with a protein depends on the nature of the solvent, it is more relevant to describe the system in terms of the water activity than of water concentration[31]. Although this parameter is not easy to determine in MD simulations, some studies have addressed this issue. Molecular dynamics simulations of CALB in the gas phase have been used to analyse the hydration mechanism and the effect of water activity (a_w) on protein structure and dynamics[128]. In a first step, the authors obtained

experimental water adsorption/desorption isotherms for desalted and deglycosylated CALB[128]. They observed that the isotherms can be divided in two distinct regions[128]. In the first region (a_w <= 0.40) the water adsorbed to the protein increases linearly, which corresponds to the formation of the first hydration layer[128]. At higher values of a_w, the amount of bound water molecules increases sharply, corresponding to the formation of additional water layers around the protein[128]. In a second step, MD simulations of CALB surrounded by gaseous argon at different water activities were performed[128]. The simulations were in good agreement with the experimental isotherms[128]. The analysis of these simulations showed that at low water activities the water molecules form clusters around hydrophilic binding sites on the protein surface[128]. These clusters grow until the whole hydrophilic surface of the protein is covered (this happens at a_w ~ 0.50)[128]. At higher a_w values, the number of protein-bound water molecules increases sharply as the water starts forming a multilayer around the protein[128]. The simulations also indicate that the stability of the protein structure is independent of the water activity[128]. The flexibility of the enzyme is affected by the water activity, but no general trend was found[128]. Some regions of the protein become more flexible when the a_w increases, while others loose flexibility[128].

A different approach to analyse the dependence of enzyme behaviour on water activity has been used in a recent study[129]. The lipase CALB was used once again as a model enzyme and simulations were performed in five different solvents (water, methanol, tert-butyl alcohol, methyl tert-butyl ether, and hexane)[129]. The hydration of the protein as a function of water activity was analysed. Water activity was calculated by determining the bulk water concentration and activity coefficient[129]. To calculate the activity coefficient the authors combined alchemical free energy perturbations (to

obtain the free energy of solvation of water in each of the solvents) and simulations of water/organic solvent mixtures analyzed using Kirkwood-Buff theory[129]. Interestingly, they found that the hydration level of the protein at a given water activity is very similar for all the solvents tested[129]. These results are consistent with the common assumption that the amount of water adsorbed on the enzyme surface is proportional to the water activity of the system[31].

1.3.7 The role of counterions

Counterions play an important role in enzyme catalysis, particularly when the reactions take place in nonaqueous solvents. The role of counterions has been analysed in several simulation studies performed by our group[1, 2, 130]. In these studies, the starting positions of counterions were chosen using a docking methodology that is described in detail in references 1 and [130]. This methodology places a counterion near each charged group that cannot be neutralized by establishing a salt bridge with an oppositely charged group. In our first study, the ions remained firmly attached to the protein charged groups during the simulations in hexane, contrary to what happened in water[1], which is in agreement with other simulation studies[115]. Control simulations without ions were analysed and it was observed that the protein structure is much less native-like in these conditions, which indicates that counterions play a fundamental role in stabilizing the protein in a nonaqueous environment[1]. It was also observed that, when large amounts of water are present in the media, counterions tend to form ion clusters[1]. This leads to large protein conformational changes, because the protein charge groups tend to move along with these ion clusters when they migrate to the solvent[1].

The behaviour of chloride and sodium ions in solvents with different polarities at different hydration conditions has been compared[2]. As expected, ions are preferentially bound to the enzyme at low hydration levels[2]. In more polar solvents, like acetonitrile, counterions can migrate to solution, even when the amount of water in the media is low[2].

The role of counterions in nonaqueous enzymology has been extensively analysed in an MD simulation study, which is described in chapter 3 of the present thesis.

1.3.8 Enzyme activity and enantioselectivity

One of the first simulation studies of enzyme enantioselectivity using an explicit organic solvent was performed by Colombo et al.[131]. These authors analysed the enantioselectivity of subtilisin in the transesterification reaction of 1-phenylethanol by vinyl acetate, in dimethylformamide (DMF)[131]. The formation of the tetrahedral intermediate is believed to be the rate limiting step of this reaction, and it is also generally assumed that the structure of the transition state is very similar to the tetrahedral intermediate itself[132]. Therefore, the free energy difference between the *R* and *S* tetrahedral intermediates corresponds to the difference in the activation energy between the two enantiomers. Colombo et al. used a free energy perturbation (FEP) methodology to calculate the free energy difference between the *R* and *S* tetrahedral intermediates[131]. The charges of the two tetrahedral intermediates in the active site of the enzyme were calculated using a QM/MM methodology. Interestingly, the authors found that the two enantiomers have considerably different charge distributions in the active site of the enzyme[131]. Using these charges sets, they were able to obtain a

good agreement between the calculated free energy and the experimentally obtained value[131].

The enantioselectivity of cutinase in hexane at different hydration conditions has been analysed by our group, using MD simulations[74]. The transesterification reactions between vinyl butyrate and two aromatic alcohols (1-phenylethanol and 2-phenyl-1-propanol) were analysed[74]. Free energy calculations were used to determine the free energy difference between the *R* and *S* enantiomer of the reaction tetrahedral intermediates (TIs)[74]. The free energy calculations were in good agreement with the experimentally observed preference of the enzyme towards the *R* enantiomer of the two substrates analysed[74]. The enantioselectivity of the enzyme was influenced by the water content of the media, being higher when the amount of water was in the range of 5-10% (v/v)[74], where the structure and dynamical properties of cutinase resemble the ones observed in water[1]. These observations indicate that the discriminative power of the enzyme is correlated with its structural and dynamical properties. A more detailed analysis showed that the dependence of the enantioselectivity on the amount of water is a consequence of the hydrogen bond pattern between the catalytic histidine and the reaction TI. The simulations revealed that the interaction between the catalytic histidine and the *R* TI is stabilized when the water content of the media is 5-10%[74], which explains why the enzyme has a higher preference for this enantiomer in these hydration conditions[74].

The stability and activity of CALB in solvents with different polarities has been analysed by molecular dynamics and QM/MM simulations[133]. The enzyme was stable in all the solvents tested[133]. However, the active site suffered larger conformational changes in polar than in apolar solvents[133]. The MD simulations revealed that polar solvents frequently interact with the active

site residues[133], and this interaction destabilizes the hydrogen bond between the catalytic serine and histidine residues, which decreases the activity of the enzyme[133]. The QM/MM calculations show that the activation energy is higher in polar than in apolar solvents, indicating that the enzyme activity is higher in apolar media[133].

1.3.9 Lipase interfacial activation

Interfacial activation is an interesting property displayed by some lipases in nonaqueous solvents, which has been analysed using simulation methodologies[134-138]. Most lipases have a mobile element at the surface (lid), which covers the active site. It is known that this lid opens upon binding to a hydrophobic surface or when it is in contact with a nonpolar solvent - interfacial activation. To analyse this behaviour, Trodler et al. performed MD simulations of *Burkholderia cepacia* lipase (BCL) in water and toluene[137]. Two different structures were used as starting points of the simulations, with the lid either in the open or in the closed state. The solvent had a strong influence on the behaviour of the enzyme. In water, the lid had a tendency to be in a closed state, independently of the structure that was used to initiate the simulations. In toluene, the lid gradually opened during the first 10-15 ns of simulation, when the closed structure was used as a starting point. When the lid was initially open, it remained opened throughout the simulations in toluene. The open state of the lid was stabilized in toluene by the formation of a network of intra-protein hydrogen bonds, and the residue D130 was crucial in the formation of this network[137].

MD simulations have also been able to shed light on the behaviour of *Pseudomonas aeruginosa* lipase[136]. The results indicate that this lipase has

two lids, instead of one, as is commonly observed in other lipases[136]. These lids tend to close in water simulations and open in water/octane simulations[136]. The authors found that the movement of lid1 (according to their nomenclature) is triggered by lid2[136]. Simulations using *in silico* mutants revealed that hydrophobic interactions between the two lids play an important role in the opening and closing processes[136].

1.3.10 Simulation studies of enzymes in ionic liquids

Ionic liquids (ILs) and supercritical fluids are emerging as promising alternative media for biocatalytic processes. In spite of the large number of experimental studies on this subject, the molecular details of the interaction between proteins and these unconventional media remain unclear. In the last five years, MD simulation studies have helped to shed light on this matter.

The first MD simulation of an enzyme solvated by ionic liquids was performed by our group[139]. The serine protease cutinase was simulated in two different ionic liquids ([BMIM][PF_6] and [BMIM][NO_3]), which were previously parameterized in our laboratory[140]. For each system, two different temperatures (298 and 343 K) and several water concentrations were tested[139]. In accordance with previous experimental reports, the type of anion present in the ionic liquid had a strong influence on enzyme stability[139]. The protein was considerably more stable in the presence of PF_6^- than when NO_3^- was present[139]. The latter anion has a stronger ability to bind to the protein, because it is smaller and has a higher charge density. Therefore, during the simulations, several NO_3^- anions formed hydrogen bonds with amide groups of the main chain, disrupting main chain hydrogen bonds, which are essential for the stability of the secondary and tertiary structure. The

simulations also revealed that the water content of the media had a strong influence on enzyme stability[139]. In [BMIM][PF_6], the enzyme is more stable when the water content of the media is in the range of 5-10% v/v. At lower and higher water concentrations, the protein showed larger deviations from the X-ray structure[139]. This trend was not observed in [BMIM][NO_3], where the enzyme was more stable at very low or very high water concentrations[139]. The analysis of the high temperature simulations showed that [BMIM][PF_6] increases the protein thermal stability (relative to the aqueous control), when the water content of the media is low. In [BMIM][NO_3], on the other hand, the enzyme had a very low thermal stability. The results also showed that both ionic liquids strip most water molecules from the protein, similarly to what is observed in polar organic solvents[139].

Recently, the solvation of CALB by eight distinct ILs was analysed using MD simulations[141]. The ionic liquids tested are composed by imidazolium or guanidinium based cations paired with nitrate, tetrafluoroborate or hexafluorphosphate anions[141]. The authors estimated the enthalpy of solvation of CALB in these solvents, by calculating the different energy terms that contribute to the enthalpy, from equilibrated MD simulations[141]. They found that the interaction between the protein and ILs is dominated by electrostatic interactions with anions[141]. Cations also play a relevant role, although their interaction with the protein is mostly mediated by van der Waals contacts, being weaker than in the case of anions[141]. Smaller anions with high surface charge were found to bind more strongly to the protein. For the same cation, the interaction increases in the order PF_6^- < BF_4^- < NO_3^-. These results are in agreement with experimental evidences and the previous MD simulation study performed in our laboratory[139]. The calculations indicate that the decreased solubility of CALB in ILs when compared to water is mainly

due to the large energies that are spent to form solute cages in ILs[141]. In a following paper, the same authors analysed the stability of CALB in the same ILs in high temperature simulations [142]. In line with previous experimental and simulation evidences, their simulations showed that anions have a stronger influence than cations on enzyme stability. The conservation of the enzyme secondary and tertiary structure in ILs follows the order PF_6^- > BF_4^- > NO_3^-, which is in accordance with previous observations, including the MD simulation study performed by our group[139]. They also found that there are two major causes for protein destabilization in ILs[142]. On one hand, strong electrostatic interactions (mainly with IL anions) lead to conformational changes on the enzyme surface, characterized by the unravelling of surface α-helices and an increase in the protein gyration radius and surface area[142]. Additionally, van der Waals interactions between ILs and the protein core also disrupt its structure[142]. In most cases, these contacts are established with cations that have long alkyl chains which are able to penetrate into the protein core[142]. In other cases, ILs cause major conformational changes that expose the protein interior[142].

1.3.11 Simulation studies of enzymes in supercritical fluids

Supercritical fluids, such as supercritical carbon dioxide ($scCO_2$), offer many advantages over organic solvents as reaction media for enzymatic catalysis, because they combine the solvation ability of liquids with gas-like transport properties[32]. $scCO_2$ is considered a green solvent, because it is nontoxic and non-flammable[32]. It is also easy to obtain, has the ability to solvate both hydrophobic and hydrophilic solutes and has been shown to be a good medium for some enzymatic reactions[32].

MD simulation studies have been used to obtain a molecular picture of the behaviour of enzymes in supercritical fluids. In a recent study, the stability of CALB in a $scCO_2$/liquid water biphasic mixture has been analysed using MD simulations[143]. Several sets of MD simulations were performed, at different hydration conditions[143]. The authors observed that the addition of water to the medium increases the protein stability relative to the situation found in pure $scCO_2$[143]. Interestingly, only a small amount of water is necessary to maintain the protein structure, and increasing the water content does not have a significant effect[143]. The simulations also show that CO_2 and water molecules are heterogeneously distributed on the enzyme surface[143]. The authors argue that this observation can explain the high activity displayed by CALB in this type of medium[143]. They claim that the heterogeneous arrangement of the solvents around the protein may facilitate the diffusion of hydrophobic solutes, favouring the lipase activity[143]. In another study, a mutant of CALB (cp283Δ7) was simulated in anhydrous scCO2 as well as in pure water [144]. In agreement with the study by Silveira et al.[143], the protein was considerably more stable in water than in pure $scCO_2$[144]. The same authors have also compared the behaviour of CALB in a $scCO_2$/ionic liquid biphasic mixture and pure $scCO_2$, using MD simulations[145]. Although all ionic liquid molecules were initially randomly distributed, they accumulated around the protein during the simulation[145]. All the analysis performed (rmsd, radius of gyration, amount of secondary structure) showed that the enzyme is more stable in the biphasic system than in pure $scCO_2$[145]. The results indicate that the ionic liquid forms a protective layer around the protein, which decreased the ability of CO_2 to interact with the protein and destabilize it[145].

1.4 Scope of the present thesis

Obtaining a solid fundamental knowledge of the key aspects of nonaqueous biocatalysis is essential for exploring its technological potential. The main goal of the present thesis is to contribute to a deeper understanding of the molecular determinants of enzyme behaviour in nonaqueous solvents. We decided to focus on three important subjects, which were poorly characterized at the molecular level: protein-ion interactions, enzyme stability, and molecular memory. They were investigated using molecular simulation methods, which enable the analysis of the structural and dynamic properties of the systems under study, with atomic detail.

Protein-ion interactions play a very important role in enzyme catalysis, especially when the reactions take place in organic solvents. These media are less polar than water, which leads to the formation of strong salt-bridges between the ions and the protein. Numerous studies have shown that salts can enhance enzyme reactions in nonaqueous solvents[99-108, 146]. In order to gain a molecular insight into the role of counterions in nonaqueous enzymology, our collaborators in the U.K. and Germany have determined the X-ray structure of subtilisin Carlsberg soaked in acetonitrile and $CsCl$[3]. Although several chloride and cesium ions could be detected in this structure, their locations in the crystal environment might be influenced by crystallographic contacts, an artificial electrostatic environment, and protein cross-linking. Additionally, the X-ray structure represents an average over the conformations present in the crystal and cannot capture the dynamics of protein-ion interactions. In order to complement the previous X-ray analysis, we used an MD simulation approach to simulate the dynamic behaviour of the protein and ions in acetonitrile solution. This study is described in chapter 3.

Enzyme stability in nonaqueous solvents is usually lower than in water, which limits the industrial potential of nonaqueous biocatalysis. In order to overcome this limitation, it is important to have a molecular understanding of the solvent effects on enzyme stability. We have addressed this issue by comparing the behaviour of pseudolysin (PSL) and thermolysin (TLN) in ethanol/water mixtures. We chose to use these enzymes as models because, although they have very similar structures, the former is considerably more stable than the latter in these media[4]. Experiments have also shown that a disulfide bridge between cysteines 30 and 58 of PSL is important for the stability of this enzyme[5]. In order to understand the molecular causes of this behaviour, we performed μs-long MD simulations of TLN, wild type PSL and the C58G mutant of PSL, in water and ethanol/water mixtures. The results of this study are described in chapter 4.

The behaviour of an enzyme in an apolar solvent depends on its past history (molecular memory). In particular, it has been found that lyophilizing an enzyme in the presence of competitive inhibitors (which are subsequently removed) enhances its activity in apolar solvents[6]. This curious phenomenon is known as ligand imprinting and has been shown to be useful strategy to increase enzyme activity and specificity in nonaqueous solvents. In order to take full advantage of this phenomenon, it is crucial to know its molecular determinants. Given that, when this thesis was initiated, there was no clear molecular explanation for ligand imprinting, we decided to analyse this phenomenon using an MD simulation approach, which is described in detail in chapter 5.

Chapter 2

Theory and methods

2.1 Biomolecular modelling and simulation

Molecular modelling can be described as the use of a simplified description (model) of a molecular system. This model is used to perform simulations that enable one to study and make predictions about the behaviour of the system.

To fully understand biological processes and phenomena, one has to characterize their molecular determinants. Although experimental techniques can be used to analyse biological systems at the molecular level, these approaches almost always involve averaging over time and/or space and do not provide detailed information on the distribution of conformations that characterize the system. Computer simulations, on the other hand, can be used to explore the conformational space of biomolecules. Additionally, they offer the possibility of analysing the behaviour of molecules under conditions that cannot be tested experimentally. Simulations are also used to make preliminary predictions, which can be tested in the laboratory. For instance, docking simulations are commonly used in preliminary steps of drug design projects, saving laborious and expensive experimental work.

There are several types of molecular models, which use different types of approximations. These models can be classified according to the level of detail that is used to represent the system, the energy function that is used to model the system, the method that is used to sample the configuration space and the treatment of boundaries and external conditions[147]. The choice of the model depends on the question that is being addressed. If the properties that are being analysed depend on the electronic degrees of freedom of the system (e.g. formation/break of chemical bonds) one has to use quantum mechanical (QM) models. Although these models have a very high resolution, they are very demanding from the computational viewpoint

and, thus, cannot be applied to large systems or to simulate long time scales. Some properties can be studied without considering the degrees of freedom of electrons. In this case, the system can be modelled by a molecular mechanics (MM) force field, in which the energy of the system is only a function of nuclear positions and velocities. The advantage of these methods is that they are much faster than quantum mechanical methods and can, therefore, be used to model systems containing a large number of atoms. In some cases (e.g., when studying redox properties of a molecule), continuum electrostatic (CE) models can be used. These methods are less detailed than MM or QM methods, but are also less computationally demanding. Three different simulation methods have been used in the scope of the current PhD thesis: molecular dynamics (MD), molecular docking, and continuum electrostatics/Monte Carlo (CE/MC). The main characteristics of these methods are described in table 2.1.

Table 2.1 Simulation methods used in the scope of this thesis

Method	Level of detail	Energy function	Treatment of the solvent	Sampling method
MD simulation	Atomic	Empirical force field	Explicit	Integration of Newton's laws of motion
Molecular docking simulation	Atomic	Empirical force field	Implicit	Genetic and Monte Carlo simulated annealing algorithms
CE/CM simulation	The protein is modelled as a low dielectric region with partial atomic charges	Based on Poisson-Boltzmann equation	Implicit	Monte Carlo simulations

2.2 Molecular mechanics

Molecular mechanics (MM) methods use potential energy functions (force fields) to model molecular systems. Unlike quantum mechanical methods, in which electronic degrees of freedom are explicitly taken into account, in molecular mechanics the energy of the system is only a function of nuclear positions and velocities. This simplification is based on the assumption that the electronic and nuclear motion of atoms can be separated (Born-Oppenheimer approximation). Due to this approximation, MM methods cannot be used to study properties which depend on the electronic distribution of a molecule. They can only be used to study the conformational behaviour of molecules, usually in their ground state.

2.2.1 Molecular mechanics force fields

In molecular mechanics, the potential energy function that is used to model the interactions between the atoms of the system is called force field (FF)[148-150]. The potential energy (V) is a function of the positions ($\boldsymbol{r}$) of the particles that compose the system. Most molecular mechanics FFs are defined as the sum of a set of terms, which account for bonded (covalent) and nonbonded (noncovalent) interactions between atoms. A typical force field for biomolecules is represented by the following expression:

$$V(r) = V_{\text{bonds}} + V_{\text{angles}} + V_{\substack{\text{improper}\\\text{dihedrals}}} + V_{\substack{\text{proper}\\\text{dihedrals}}} + V_{\substack{\text{van der}\\\text{Waals}}} + V_{\text{electrostatics}} \qquad (2.1)$$

The bonded contributions include energy penalties that arise from the deviations of bond lengths and angles from their equilibrium value. The bonded part of the FF also comprises terms which account for the torsion of bonds (proper dihedrals) and terms that maintain the geometry of the

molecules (improper dihedrals). There are two important nonbonded terms, which account for the van der Waals and electrostatic interactions.

The GROMOS 53A6 FF[151] was used in all the simulation studies described in this thesis. This FF uses a united atom approach, in which all atoms are represented explicitly, except nonpolar hydrogens. The nonpolar hydrogens are merged with the heavy atoms to which they are attached, in order to increase the speed of the calculations. The energy terms of the GROMOS FFs are given by the following equations:

$$V_{\text{bonds}} = \sum_{\text{bonds}} \frac{1}{4} k_b (b^2 - b_0^2)^2 \tag{2.2}$$

$$V_{\substack{\text{bond}\\\text{angles}}} = \sum_{\substack{\text{bond}\\\text{angles}}} \frac{1}{2} k_\theta (\cos\theta - \cos\theta_0)^2 \tag{2.3}$$

$$V_{\substack{\text{improper}\\\text{dihedrals}}} = \sum_{\substack{\text{improper}\\\text{dihedrals}}} \frac{1}{2} k_\xi (\cos\xi - \cos\xi_0)^2 \tag{2.4}$$

$$V_{\substack{\text{proper}\\\text{dihedrals}}} = \sum_{\substack{\textit{proper}\\\textit{dihedrals}}} k_\varphi [(1 + \cos(\delta)\cos(m\varphi))] \tag{2.5}$$

$$V_{\substack{\text{van der}\\\text{Waals}}} = \sum_{\substack{\textit{pairs}\\i,j}} 4\varepsilon_{i,j} \left[\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{6} \right] \tag{2.6}$$

$$V_{\text{electrostatics}} = \sum_{\substack{\text{pairs}\\i,j}} \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r} \frac{1}{r_{i,j}} \ , \tag{2.7}$$

in which, $K_{b,}$ $K_{\Theta,}$ $K_{\xi,}$ K_{φ} are the force constants for bond stretching, bond-angle bending, improper dihedral-angle and proper dihedral-angle, respectively; b, Θ, ξ, φ are the bond length, bond-angle, improper and proper dihedral-angle, respectively; b_0, Θ_0, and ξ_0 are the reference values for bond length, bond-angle and improper dihedral-angle, respectively; δ and m are the phase shift and multiplicity of proper dihedral angles, respectively; $\varepsilon_{i,j}$ is the depth of

the Lennard-Jones potential well between atoms *i* and *j*; $\sigma_{i,j}$ is the distance at which the Lennard-Jones potential between atoms *i* and *j* is zero; $r_{i,j}$ is the distance between atoms *i* and *j*; q_i and q_j are the atomic charges of atom *i* and *j*, respectively; ε_0 and ε_r are vacuum permittivity and the relative permittivity of the medium, respectively.

2.2.2 Bonded interactions

As shown in eq. 2.2, the stretching of covalent bonds is modelled by a quadratic potential, which keeps the bond lengths (b) close to their reference values (b_0). The stiffness of the bond is controlled by the force constant (K_b).

The bending of the angle between two adjacent covalent bonds is modelled by a harmonic potential function (eq. 2.3), which increases when the angle (Θ) deviates from the reference value, (Θ_0). The energy penalty due to deviations from the reference angle is defined by the force constant (K_Θ).

The improper dihedral term shown in eq. 2.4 is used to maintain the chirality of tetrahedral centres and the planarity of planar groups. This term is also an harmonic potential, which penalizes deviations from the reference improper dihedral-angle (ξ_0), according to the force constant K_ξ.

The energy associated with the rotation about a chemical bond (proper dihedral) is taken into account by the potential shown in eq. 2.5. According to this equation, proper dihedrals are modelled by a periodic function, characterized by the force constant (K_φ) (which controls the amplitude of the curve), the dihedral multiplicity (m), and the phase shift (δ).

2.2.3 Nonbonded interactions

Atoms that are separated by three or more covalent bonds, or that are not part of the same molecule, interact through nonbonded forces. There are two main types of nonbonded interactions: van der Waals and electrostatic.

van der Waals interactions account for the deviation from ideal gas behaviour, which is found in real gases and was first observed by Johannes van der Waals. They include attractive forces, which are due to the instantaneous dipoles that arise from fluctuations in the electron clouds, and repulsive forces that can be explained by Pauli's exclusion principle. Although van der Waals forces are considerably weaker than electrostatic interactions, their correct modelling is crucial in molecular mechanics FFs. These forces are usually described by a Lennard-Jones potential (eq 2.6), which has an attractive term that varies with the sixth power of the distance between the interacting particles (r) and a repulsive term, which decays with the twelfth power of r. Given their short-ranged nature, it is usually reasonable to ignore van der Waals forces between particles whose distance is greater than a certain cut-off (e.g. 1.4 nm), saving a significant amount of computational time.

Most molecular mechanics FFs model the charge distribution of molecules through the assignment of a partial charge to each atom. The electrostatic interaction between two atoms is calculated using Coulomb's law (eq 2.7), where q_i and q_j are the atomic partial charges, r_{ij} is the distance between the atoms, ε_o is the permittivity of vacuum and ε_r is the relative permittivity of the medium. Like van der Waals interactions, electrostatic interactions can be truncated, making their calculation significantly faster. However, given that these interactions are very strong and long-ranged (they decay with $1/r$), they can not be abruptly truncated. There are two main strategies used

to account for the long-range effect of electrostatic interactions: lattice sum methods and continuum methods. Lattice sum methods, like the Particle Mesh Ewald (PME) method[152], are used to compute long-range contributions to the potential energy in periodic systems. These methods calculate the interactions of each particle with all the other particles in the simulation box and their periodic images. In these algorithms, long-range interactions are computed in reciprocal space, using a Fourier transform, which considerably accelerates the calculations.

Continuum methods, such as the Reaction Field method[153], model the effect of long range dipole-dipole interactions. In this method, each molecule is surrounded by a sphere with a radius defined by a given cut-off. The interactions between a given molecule and the particles which are within the sphere are calculated explicitly. The medium beyond a certain cut-off is modelled as a homogenous environment with a given dielectric constant (ε_r). The molecule induces polarization in this media, which in turn creates a reaction field.

2.3 Energy minimization

The potential energy of the different configurations of a molecular system can be mapped in a potential energy surface (also known as hypersurface)[150]. This surface is a complicated, multidimensional function of the particles coordinates. We can picture it as a rugged landscape, in which there are mountains and valleys. The valleys, or energy minima, correspond to stable configurations of the system. The minimum with the lowest potential energy is called the global minimum, whereas the others are known as local minima. Sometimes it is useful to have the molecular system in a stable configuration,

which corresponds to a minimum in the potential energy landscape. This can be done through a process called energy minimization, using one of the many algorithms that have been developed for this purpose[150]. Mathematically, a minimum of a function can be described as a point in which all partial first derivatives are zero and whose Hessian matrix is positive definite[150]. Therefore, one strategy that is implemented in some minimization algorithms consists in finding the derivatives of the energy with respect to the system's coordinates[150]. These methods are called derivative minimization algorithms and can use either analytical or numerical derivatives[150]. Derivative methods can be classified as first-order or second-order methods, according to the highest order derivative used[150]. First-order algorithms, such as the *steepest descents* and the *conjugate gradient*, iteratively change the coordinates of the atoms, pushing the system towards the closest local minimum[150]. In all the studies performed in this thesis, the steepest descent algorithm, which is implemented in the GROMACS package[154], was used to minimize the energy of the system. This algorithm moves the system downhill, i.e., in a direction parallel to the net force, at each iteration, the forces are calculated and the positions are updated using the expression:

$$\boldsymbol{r}_{n+1} = \boldsymbol{r}_n + \frac{\boldsymbol{F}_n}{max\left(\left|\boldsymbol{F}_n\right|\right)} h_n \ , \tag{2.8}$$

where h_n is the maximum displacement, $\boldsymbol{F}_n$ is the force and max($|\boldsymbol{F}_n|$) is the largest absolute value of the force components. The process stops after a fixed number of iteration steps or when the gradient of the energy is smaller than a given predefined value[150].

2.4 Molecular dynamics simulations

Although we usually picture molecules as having a defined, rigid structure, that is just an idealized picture that does not correspond to reality. Molecules and molecular systems are dynamic entities, whose behaviour can only be fully understood in the light of statistical mechanics theory. According to this theory, a physical system can be described by an ensemble, which represents all the possible configurations of the system and their probabilities of occurrence. The experimentally measured value of a macroscopic property (e.g., pressure) is an ensemble average of all the possible microscopic values of that property[150]. In order to calculate the value of a given property from a simulation, the simulation method needs to be able to appropriately sample the configurational space of the system[150]. One way to do this is to use molecular dynamics (MD) simulation. This technique not only provides an ensemble of configurations, but also gives a temporal relation between them, i.e., it generates a trajectory[148-150]. It is a deterministic method, because the state of the system at any future (or past) time can (ideally) be predicted (or traced back) from its current state[148-150]. If the MD simulation is sufficiently long to enable a proper sampling of the conformational space, it can be used to analyse structural, dynamic and thermodynamic properties of the system, which can be compared with experimental results[148-150]. This is possible because, in the light of the ergodic hypothesis, the time average of a property is equivalent to its ensemble average, for a system at equilibrium[148]. Given that, with current computational resources, large systems cannot be simulated for very long times, a common strategy to obtain a proper sampling consists in using several replicates, with identical conditions, but different starting points (e.g., by using different initial velocities).

In MD simulations, the evolution of the positions and momenta of particles over time is calculated by integrating Newton's second law of motion[148-150], which states that the acceleration (***a***) of a body of mass *m* is parallel and directly proportional to the net force (**F**) acting on that body:

$$\boldsymbol{F} = m\,\boldsymbol{a} \tag{2.9}$$

For a molecular system of N interacting particles, the force experienced by a given particle (*i*) is proportional to the particle's acceleration times its mass:

$$\boldsymbol{F}_i = m_i \frac{d^2 \boldsymbol{r}_i}{dt^2}, i = 1 \ldots \mathrm{N} \tag{2.10}$$

The force $\boldsymbol{F}_i$ acting on each particle at a given time instant *t* is given by the negative gradient of the potential energy (*V*) with respect to the particle's position (r_i):

$$\boldsymbol{F}_i = -\frac{\partial V}{\partial \boldsymbol{r}_i} \tag{2.11}$$

The particle's acceleration corresponds to the first and second time derivative of its velocity ($\boldsymbol{v}_i$) and position ($\boldsymbol{r}_i$), respectively. Therefore, these quantities can be calculated by the following expressions:

$$\frac{d\boldsymbol{v}_i}{dt} = \frac{\boldsymbol{F}_i}{m_i} \tag{2.12}$$

$$\frac{d\,r_i}{dt} = \boldsymbol{v}_i \tag{2.13}$$

The potential energy (*V*) is a function of the atomic coordinates and is given by the empirical FFs that have been described above (eq 2.1 - 2.7). The kinetic energy (*K*) is a function of the particles masses and velocities:

$$K = \frac{1}{2} m v^2 \tag{2.14}$$

The total energy of the system, or halmiltonian (*H*), is the sum of the potential (*V*) and kinetic (*K*) energies:

$$H = V + K \tag{2.15}$$

The equations described above can be used simulate a molecular system, in order make predictions about its conformational behaviour, as well as estimate relevant thermodynamic properties.

One of the most important applications of MD simulations is the study of the molecular behaviour of proteins. This technique can provide useful insights into the structural and dynamic properties of these biomolecules, which are known to play a crucial role in many of their biological functions (e.g., catalysis, transport, ATP synthesis). The first MD simulation of a protein was performed in 1977, in Martin Karplus' laboratory[155] and comprised 9.2 ps of a small protein in vacuum. Due to the tremendous increase in computational power and the developments in simulation algorithms, MD simulations of proteins have come a long way since then[147, 156, 157]. Most current simulations include explicit solvent and cover hundreds, and in some cases thousands, of ns. Modelling proteins in membrane or nonaqueous environments has also become possible and several studies have been conducted in these media. MD simulation studies have contributed to the advancement of several areas of protein science, such as folding, catalysis, transport and energy transduction[156, 157].

2.4.1 Integration algorithms

There are several algorithms that can be used to integrate Newton's laws of motion and update the positions and velocities of particles over time. One of these algorithms updates these variables alternately, in such a way that they 'leapfrog' over each other[149]. This method is, therefore, called the Leapfrog algorithm and uses the following equations[149]:

$$\boldsymbol{v}\left(t+\frac{\Delta t}{2}\right)=\boldsymbol{v}\left(t-\frac{\Delta t}{2}\right)+\frac{\boldsymbol{F}(t)}{m}\Delta t \tag{2.16}$$

$$\boldsymbol{r}(t+\Delta t)=\boldsymbol{r}(t)+\boldsymbol{v}\left(t+\frac{\Delta t}{2}\right)\Delta t \tag{2.17}$$

At time t, each particle (i) is in a given position ($\boldsymbol{r}_i$) and the potential energy of the system can be calculated using the potential energy function described by eqs. 2.1-2.7. The force acting on each particle at time t is then obtained through equation 2.11. Equations 2.16 and 2.17, which are the numerical equivalents of equations 2.12 and 2.13, respectively, are used to obtain the particles' velocities and positions alternately. This cycle is repeated at each step, until the simulation is completed.

The choice of the time step (Δt) is crucial in MD simulations. In principle it should not be larger than the frequency of the fastest motions of the system, which are usually the vibrations of bonds involving hydrogen atoms[149, 150]. On the other hand, the time step should be as large as possible, in order to speed up the simulation, enabling it to sample a significant fraction of the system's configurational space. A strategy that is frequently used to increase the integration time step is to constrain bond lengths to their reference values, using constraint algorithms, such as SHAKE[158] and LINCS[159]. These methods remove the fast vibrations of the system, which are usually not

coupled with relevant slower movements, enabling the use of larger time steps.

2.4.2 MD simulations with periodic boundary conditions

In a macroscopic system, the fraction of molecules which interact with the walls of the container is usually very small. On the other hand, due the restricted size of a typical MD simulation box, the number of molecules that are within the influence of the walls is very large[150]. In order to circumvent this limitation and simulate the behaviour of molecules in the bulk solution, molecular modellers have developed strategies to treat boundary conditions. One of the most common strategies is the use of periodic boundary conditions (PBC)[150]. The simulation box is replicated in all directions and molecules can move between adjacent box replicates[150]. Although this strategy removes boundary artefacts, it is not perfect, because the periodicity itself can cause errors in simulations of non-periodic systems[150].

In MD simulations using PBC, a molecule should only have short-range nonbonded interactions with the nearest neighbour of each particle (minimum image convention)[150]. This means that simulation boxes have to be large enough to avoid unwanted interactions. Given that very large simulation boxes make the calculations very slow, one has to find a compromise between these two conditions. Simulation boxes can have different geometries, e.g., cubic, rhombic dodecahedron, truncated octahedron. When simulating solutes with approximately spherical shapes, like most proteins, it is more efficient to use box geometries which are close to a sphere (e.g. truncated octahedron or rhombic dodecahedron). This reduces the number of

solvent molecules that is needed to fill the box, given a minimum distance between periodic images.

2.4.3 MD simulations at constant temperature and/or pressure

MD simulations can be performed under different macroscopic environmental constraints, which lead to different types of ensembles. If the simulation is performed at constant energy and volume, it is said to be in the microcanonical (NVE) ensemble[148-150]. In this ensemble, the system is isolated from its surroundings and all the microstates that compose the ensemble have the same energy and, therefore, the same probability of occurring[160]. Given that most real systems are not thermally isolated from their surroundings, it is more realistic to perform the simulations in NVT or NPT ensembles. In NVT simulations, the system is coupled to a heat bath and the only constraint that is imposed is that the total energy of the system plus reservoir remains constant[150]. The probability distribution of the microscopic states of the system is then given by the Boltzmann distribution[160]:

$$P_i(N,V,T) = \frac{e^{-E_i(N,V)/kT}}{\sum_j e^{-E_j(N,V)/kT}} \qquad (2.18)$$

Where E_i is the energy of state i, k is the Boltzmann constant and T is the absolute temperature. Nowadays, most biomolecular MD simulations are performed at constant temperature. This is important when one wants to compare the MD simulation results with in vivo or in vitro experiments performed at a given temperature. By performing constant temperature simulations at different temperature values, one can also analyse how the behaviour of the system is influenced by this parameter.

The temperature of the system is a measure of its average kinetic energy, which in turn, is a function of the particles velocities. The most common way to control the temperature in a MD simulation is to couple the system to an external heat bath, which acts as an energy reservoir (supplying or removing energy from the system)[150]. An example of a temperature coupling algorithm is the Berendsen bath[161], in which the deviation of the system's temperature (T) from the bath's temperature (T_{bath}) is slowly corrected[161]. This correction guarantees that the rate of temperature change is proportional to the difference between the bath and system temperatures[161]:

$$\frac{dT(t)}{dt} = \frac{T_{bath} - T(t)}{\tau} \tag{2.19}$$

The coupling parameter (τ), determines the strength of the coupling between the system and the bath (the smaller the τ, the stronger the coupling). In practice, at each time step, the velocities are multiplied by a time-dependent scaling factor (λ), given by[161]:

$$\lambda = \left[1 + \frac{\Delta t}{\tau}\left\{\frac{T_{bath}}{T(t)} - 1\right\}\right]^{1/2} \tag{2.20}$$

In some cases, it is desirable to perform the simulation in an *NPT* ensemble, i.e., at constant temperature and pressure. Just like temperature, the pressure can be controlled by coupling the system to a pressure bath. For instance, the Berendsen pressure algorithm[161] scales the coordinates and box vectors at every time step, so that the pressure can relax towards a given reference pressure (P_{bath})[161]:

$$\frac{dP(t)}{dt} = \frac{P_{bath} - P(t)}{\tau_P} \tag{2.21}$$

To rescale the coordinates and box vectors, the algorithm uses a scaling factor (μ), given by the following expression (in which β is the isothermal compressibility of the system)[161]:

$$\mu = 1 - \frac{\beta \Delta t}{3\tau_P}(P_{bath} - P) \tag{2.22}$$

2.4.4 Free energy calculations using MD simulations

Free energy is one of the most important quantities in physical chemistry. It tells us, e.g., if a given reaction will occur spontaneously. If the system under study is at constant temperature and volume (*NVT* ensemble), the free energy is expressed by the Helmholtz function (*A*). If the pressure is kept constant (*NPT* ensemble), the free energy is given by Gibbs function (G)[150].

The calculation of absolute free energies of systems containing a large number of degrees of freedom is not accessible to current computer simulations. Fortunately, in most cases, what is relevant is not the absolute free energy, but the free energy difference between two states. One approach that can be used to calculate this difference is to use the free energy perturbation (FEP) method, which was devised by Zwanzig, in 1954[162]. Based on statistical mechanics theory, Zwanzig showed that the free energy difference between two states, A and B, can be expressed by following equation[162]:

$$\Delta A = -kT \ln \left\langle \exp\left(\frac{H_B - H_A}{kT} \right) \right\rangle_A \tag{2.23}$$

where H_A and H_B are the Hamiltonians of state A and B, k is the Boltzmann constant and T is the absolute temperature. The angular brackets represent

an average over state A. In practice, one runs a normal simulation at state A (e.g. using molecular dynamics), but for each configuration, the energy of state B is also computed[150]. The problem of this approach is that it requires that the states A and B overlap in phase space[150]. Otherwise, one cannot adequately sample the phase space of state B, when simulating state A, or vice-versa[150]. The approach that is commonly used to circumvent this limitation is to define several intermediate states in progressing form A to B[150, 163]. A coupling parameter (λ), which varies between 0 and 1, is used to connect the initial and final states. As λ is changed from 0 to 1, the Hamiltonian of the system varies from H_A to H_B[163]:

$$H(\lambda) = \lambda H_B + (1-\lambda)H_A \tag{2.24}$$

A simulation is performed at different λ values and the free energy difference between states A and B is calculated by summing the free energy changes for the various values of λ.

An alternative way to calculate the free energy difference between the initial and final state is to use the thermodynamic integration (TI) method, which is represented by the expression bellow[163]:

$$\Delta A = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial H(\lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \tag{2.25}$$

The function $H(\lambda)$ in eq. 2.25 can be a nonlinear function and, therefore, it can be different from the function $H(\lambda)$ in equation 2.24. In practice, in the TI method, simulations are performed at discrete λ values between 0 and 1. For each intermediate state (λ_i), an ensemble average of the derivative of the Hamiltonian with respect to λ is obtained. The total free energy

difference is then calculated by integrating the curve corresponding to $\langle \partial H/\partial \lambda \rangle_\lambda$ as a function of λ.

2.5 Molecular docking

The aim of molecular docking methods is to predict the binding mode and the strength of interaction between two molecules, which form a complex. Common problems that can be addressed by these methods include protein-protein and protein-ligand interactions. Given that many drugs work by binding to a given receptor, the use of molecular docking has become a useful tool in drug discovery projects.

In practice, a docking simulation uses the three dimensional structures of the molecules which are going to be docked (e.g. a protein and a ligand). During the simulation, the program finds the most stable ligand conformation(s) and orientation(s) (which is commonly referred to as *docking pose*), and estimates the binding energy[164]. To do this, docking methods use search algorithms and scoring functions. In order to be efficient, docking methods commonly use several approximations. The most common approximation is to exclude solvent molecules and account for their effect implicitly (by introducing solvation and other terms in the scoring function). Additionally, most methods reduce the number of degrees of freedom included in the conformational state, by maintaining the protein rigid or only partially flexible, allowing it to be represented by affinity grids.

2.5.1 Docking algorithms

There are several approaches that can be used in docking methods to search for the most stable docking pose(s), which can be divided in two main classes: systematic methods, and random or stochastic methods[164]. The first class of methods tries to systematically explore all the degrees of freedom of the ligand, e.g. by sequentially rotating all the rotatable bonds of the ligand. Other systematic search algorithms use a technique called fragmentation, in which the ligand is divided in fragments, which are docked separately. The systematic search can also be done using databases containing pre-generated collections of conformations.

In random search algorithms, the conformational space is explored by performing random alterations in the conformation and orientation of a ligand or population of ligands. At each step, a tentative change is accepted or rejected based on a predefined criterion. Examples of random search algorithms, which use different acceptance criteria, include Monte Carlo (MC) simulated annealing methods, and Genetic Algorithms (GA). Monte Carlo methods use a Boltzmann probability function to determine the acceptance of a given change. Genetic algorithms are inspired by Darwin's theory of biological evolution. In the beginning of each GA simulation, a population comprising several different conformations (defined by a set of state variables called genes) of the ligand is generated[165]. Genetic processes, such as mutations, crossovers and migration, are applied to the population, which evolves until a predefined fitness function is optimized[165].

2.5.2 Scoring functions

Docking methods use scoring functions to rank the stability of different binding modes found during the search procedure and to estimate their binding energy. Some docking methods use scoring functions based on molecular mechanics force fields (which have been described above). In many cases, the force field-based functions are combined with implicit solvent models, which account for solvent effects. Other methods use empirical energy functions, described by a small number of terms, which account for the most relevant types of interactions (hydrophobic contacts, hydrogen bonds, torsional potentials, etc) and are designed to reproduce experimental data. These functions are calibrated using linear regression against training sets, comprising experimentally determined binding constants of several protein-ligand complexes. Another type of scoring functions, called knowledge-based scoring functions, uses very simple atomic interaction potentials. These potentials are derived from the frequency of occurrence of an interaction between a given pair of atoms in a large dataset of protein-ligand complexes.

In the work presented in this thesis, docking simulations were performed using the software AutoDock4[166], which implements a semi empirical free energy function. The free energy of binding corresponds to the difference between the free energy of the protein (P)-ligand (L) complex and the free energy of the unbound molecules[167]. This free energy includes the potential energy difference between the bound and unbound state and an entropic term[167]:

$$\Delta G = \left(V_{\text{bound}}^{\text{L–L}} - V_{\text{unbound}}^{\text{L–L}}\right) + \left(V_{\text{bound}}^{\text{P–P}} - V_{\text{unbound}}^{\text{P–P}}\right) + \left(V_{\text{bound}}^{\text{P–L}} - V_{\text{unbound}}^{\text{P–L}}\right) + \Delta S_{\text{conf}} \qquad (2.26)$$

The potentials correspond to the sum of dispersion/repulsion, hydrogen bonding, electrostatics, and desolvation terms, whose weights (*W*) are calibrated using large datasets of protein-complexes with known binding constants[167]:

$$V = W_{\text{wdW}} \sum_{i,j} \left(\frac{A_{i,j}}{r_{i,j}^{12}} - \frac{B_{i,j}}{r_{i,j}^{6}} \right) + W_{\text{hbond}} \sum_{i,j} E(t) \left(\frac{C_{i,j}}{r_{i,j}^{12}} - \frac{D_{i,j}}{r_{i,j}^{10}} \right) + W_{\text{elec}} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{i,j}) r_{i,j}} + W_{\text{sol}} \sum_{i,j} \left(S_i V_j + S_j V_i \right) exp \left(\frac{-r_{i,j}^2}{2\sigma^2} \right) \quad (2.27)$$

The first term of eq. 2.27 accounts for the energy associated van der Waals interactions, with $A_{i,j}$ and $B_{i,j}$ being the attractive and repulsive parameters for the interaction between atoms *i* an *j*. The second term describes the potential associated with hydrogen bonds. The parameters *C* and *D* are specific for different types of hydrogen bonds (O H and N H, and S H) and control the location and depth of the energy minimum, and $E(t)$ is a directional weight based on the angle (t) between the probe and the target atom. The third term is a screened Coulomb potential, which represents electrostatic interactions between atoms *i* and *j*, where q_i and q_j are the atomic charges of atom *i* and *j*, respectively; $r_{i,j}$ is the distance between these atoms; and $\varepsilon(r_{i,j})$ is the effective dielectric permittivity[168]. The last term is a desolvation potential, where *V* is the volume of the atoms surrounding a given atom; *S* is a solvation parameter; $r_{i,j}$ is the distance between atoms *i* an *j*; and σ is a distance weighting factor.

The term for the loss of torsional entropy upon binding (ΔS_{conf}), which appears in eq. 2.26, is given by:

$$\Delta S_{\text{conf}} = W_{\text{conf}} N_{\text{tors}} \ , \quad (2.28)$$

where W_{conf} is a weighting factor and N_{tors} is the number of torsional degrees of freedom of the ligand.

2.4 Prediction of protonation states using continuum electrostatics and Monte Carlo simulations

The protonation state of the ionisable groups of a protein strongly influences its structure and function. However, it is very difficult to perform experimental pH-titrations of individual protein groups and researchers have therefore turned to computational methods. In all studies reported in this thesis, the prediction of the protonation states of ionisable groups was done using computational titrations based on the Poisson–Boltzmann (PB) model[169].

The scheme shown in fig. 2.1 represents the possible protonation states of a protein with *N* ionisable groups. Given that each ionisable group can be either protonated or deprotonated, the number of possible protonation states of a protein is 2^N.

Figure 2.1. Schematic representation of the protonation states available to a protein with 5 ionisable residues

Each object in fig. 2.1 corresponds to a different protonation state $\mathbf{n} = (n_1,n_2,\ldots,n_n)$, where n_i represents the occupancy of site *i* and can take the values 0 (if the site is deprotonated) or 1 (if the site is protonated).

The free energy of a given protonation state, relative to a reference state (e.g. with all sites deprotonated), can be represented by the thermodynamic cycle shown in fig. 2.2, where the upper part corresponds to the protonation equilibrium of the same groups in water.

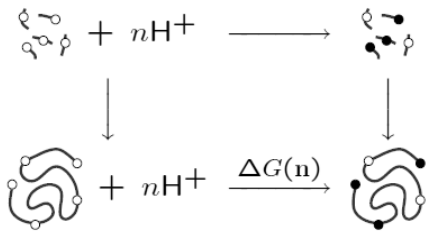


Figure 2.2. Thermodynamic cycle associated with a given protein protonation state, relative to a reference state. The upper part of the cycle corresponds to the protonation equilibrium of the same groups in water.

Each protonable site, *i*, contributes with individual and pairwise terms for the free energy (ΔG_n^0) of a given protonation state **n**, relative to a reference state, according to:

$$\Delta G_n^0 = -2.3RT\sum_i n_i \mathrm{p}K_i^{\mathrm{int}} + \sum_i \sum_{j<i} (n_i n_j + n_i z_j^0 + n_j z_i^0) W_{ij}\,, \tag{2.29}$$

where $\mathrm{p}K_i^{\mathrm{int}}$ is the intrinsic pK_a of site *i*, n_i is the occupancy of site i in state **n**, z_i^0 is the charge of site *i* when it is deprotonated and W_{ij} is the interaction free energy between sites *i* and *j*.

The $\mathrm{p}K_i^{\mathrm{int}}$ of a titrable site, which appears in eq. 2.29, is defined as the p*K*a of the site when all other sites are kept neutral, and is given by the following expression[170]:

$$pK_i^{int} = pK_a(\text{model}) + \frac{1}{2.303k_BT}\left[\Delta G_{s,p}(A) - \Delta G_{s,p}(AH)\right] = pK_a(\text{model}) + \frac{1}{2.303k_BT}\left[\Delta G_p(AH,A) - \Delta G_s(AH,A)\right], \quad (2.30)$$

where the subscripts s and p represent the solvent and the protein, respectively, the pK_a(model) is obtained from experimental titrations of model compounds and the ΔG_s(AH,A) is calculated using a model compound with the same conformation as the titrating group in the protein[169, 170].

The free energy terms related with solvation and with interactions between charges, which appear in eqs. 2.29 and 2.30, can be calculated from the electrostatic potentials, which can be done using a Poisson–Boltzmann (PB) model. In this model, the protein is represented as a low dielectric region with partial atomic charges at the positions of the nuclei, surrounded by an environment with a different dielectric value (solvent). A 3D grid is then constructed and the electrostatic potential in each grid point is calculated using a finite difference method to solve the linearized Poisson–Boltzmann equation (LPBE)[169, 170]:

$$\nabla \cdot [\varepsilon(\boldsymbol{r})\nabla\phi(\boldsymbol{r})] - \kappa^2(\boldsymbol{r})\varepsilon(\boldsymbol{r})\phi(\boldsymbol{r}) = -4\pi\rho(\boldsymbol{r}) \quad (2.31)$$

Where $\boldsymbol{r}$ is the position vector, $\phi(\boldsymbol{r})$ is the electrostatic potential at $\boldsymbol{r}$, ε is the dielectric constant of the medium, $\rho(\boldsymbol{r})$ is the charge density at $\boldsymbol{r}$, and κ is given by:

$$\kappa(\boldsymbol{r}) = \begin{cases} \left(\frac{8\pi e^2 I}{\varepsilon_{out} kT}\right)^{1/2} & \text{if } \boldsymbol{r} \text{ is in an ion accessible region} \\ 0 & \text{otherwise} \end{cases} \quad (2.32)$$

In which ε_{out} is the dielectric constant of the solvent, k is the Boltzmann constant, T is the absolute temperature and I is the ionic strength.

Using eq. 2.29, one can calculate the free energy (ΔG_n^0) of a given protonation state (**n**) of a protein, relative to some reference state. According to statistical mechanics, the probability of occurrence of that state at a given pH is, then, given by[169, 170]:

$$p(\boldsymbol{n}) = \frac{e^{-\beta \Delta G_n^0 - 2.3 z_n pH}}{\sum_j e^{-\beta \Delta G_j^0 - 2.3 z_j pH}} \qquad (2.33)$$

Where $\beta = 1/(kT)$ and z_n is the net charge corresponding to the protein in a given protonation state.

Given that most proteins have a considerable number of titrable sites, it would take too long to calculate the probabilities of all possible states analytically. This can be overcome by performing Monte Carlo simulations, in which random states are generated and accepted or rejected according to a Metropolis criterion[171]. The probability distributions at different pH values can be used to construct the titration curves of the ionisable groups, as well as the global titration curve of the protein. The pK_a, which corresponds to the pH of half titration, can be obtained from the titration curves.

Histidine is the only protein residue that has two alternative protonable sites (tautomers) in solution. However, in PB calculations, molecules are treated as rigid objects and alternative proton conformers cannot interconvert (as they do in the "real world"). Thus, even sites that do not display proton isomerism in solution should be considered tautomeric[172]. One way to deal with the lack of flexibility of PB calculations is to split the titrable site into several pseudosites (corresponding to alternative tautomers)[172], which are

then titrated using the methodology described above. The inclusion of proton isomerism makes PB calculations considerably more realistic and is therefore used in all pK_a predictions performed in our laboratory.

Chapter 3

Interaction of counterions with subtilisin in acetonitrile: Insights from molecular dynamics simulations

This work has been published in the following paper:

Lousa D., Cianci M., Helliwell J. R., Halling P. J., Baptista A. M., Soares C. M. **(2012)**, Interaction of counterions with subtilisin in acetonitrile: insights from molecular dynamics simulations, *Journal of Physical Chemistry B*, vol. 116, pp 538-548 (doi: http://dx.doi.org/10.1021/jp303008g)

Contributions of the author of the present thesis to this work:

The author of the present thesis has participated in the design of all the *in silico* experiments presented in this chapter, and executed all the simulations and analysis described herein.

3.1 Abstract

A recent X-ray structure has enabled the location of chloride and cesium ions on the surface of subtilisin Carlsberg in acetonitrile soaked crystals[3]. To complement the previous study and analyze the system in solution, molecular dynamics (MD) simulations, in acetonitrile, were performed using this structure. Additionally, $Cl^-$and Cs^+ ions were docked on the protein surface and this system was also simulated. Our results indicate that chloride ions tend to stay close to the protein, whereas cesium ions frequently migrate to the solvent. The distribution of the ions around the enzyme surface is not strongly biased by their initial locations. Replacing cesium by sodium ions showed that the distribution of the two cations is similar, indicating that Cs^+ can be used to find the binding sites of cations like Na^+ and K^+, which, unlike Cs^+, have physiological and biotechnological roles. The Na^+Cl^- is more stable than the Cs^+Cl^- ion pair, decreasing the probability of interaction between Cl^- and subtilisin. The comparison of water and acetonitrile simulations indicates that the solvent influences the distribution of the ions. This work provides an extensive theoretical analysis of the interaction between ions and the model enzyme subtilisin in a nonaqueous medium.

3.2 Introduction

Although most enzymes have evolved to catalyze reactions in an aqueous environment, the use of nonaqueous solvents as reaction media is not only feasible but can also be advantageous[30]. The biotechnological potential of nonaqueous biocatalysis has attracted the attention of researchers over the last three decades. These researchers have addressed and elucidated questions such as the structure, dynamics and stability of enzymes in

nonaqueous solvents [1, 35, 40-42, 115, 117, 120, 122, 139, 173-175], the role played by water[1, 2, 31, 74, 124, 176-178], the influence of the solvent and reaction conditions (e.g. amount of water) on enzyme selectivity[74, 131, 179-181], and the interesting phenomena of pH memory[182, 183] and ligand imprinting[6, 112, 184].

Another issue that has gathered the attention of researchers in this field is the effect of ions and other compatible solutes on enzyme activity, stability, and enantioselectivity[103, 185-187]. Several studies have shown that lyophilizing the enzyme in the presence of salts enhances its catalytic activity in organic solvents[101, 104-106]. Moreover, it has been observed that the degree of activation depends on the nature of the salt, and that the combination of a kosmotropic anion and a chaotropic cation usually gives the best results[102]. Despite the large number of studies in this area, the mechanism through which the ions stabilize and enhance the activity of enzymes remains elusive. In some cases, it seems that the ions change the mobility of the water molecules that surround the protein[107].

The interaction between the enzyme molecule and its counterions is strongly dependent on the properties of the solvent[98]. In solvents with high dielectric constants, like water, the ions tend to be dispersed in solution and generally would not be expected to form stable interactions with the protein. In apolar solvents, on the other hand, the ions are expected to form very strong salt bridges with the charged groups of the protein, playing a very important role in stabilizing the enzyme. A more complex problem is to determine how the counterions interact with the enzyme when the reaction takes place in a moderately polar solvent like acetonitrile.

Subtilisin Carlsberg is a serine protease secreted by some strains of the Gram-positive bacterium *Bacillus subtilis*. Like all serine proteases, it has a

catalytic triad, composed by a serine, a histidine, and an aspartate. It converts a large number of substrates and is able to perform both hydrolytic and esterification reactions, being more active under alkaline conditions, where the catalytic triad tends to have a global negative charge. Subtilisin Carlsberg is active in a large number of nonaqueous solvents and is often used as a model enzyme in nonaqueous enzymology. Several structures of this enzyme in the presence of organic solvents have been obtained[3, 37, 38, 188] (by soaking the cross-linked crystals with the solvent), and its catalytic behavior in nonaqueous media has been extensively studied (see e.g. ref. 30).

In a recent work, the X–ray structure of subtilisin Carlsberg soaked in acetonitrile and cesium chloride was obtained to determine the positions of chloride and cesium ions in the conditions found in the crystal environment[3]. This was important to the understanding of the interaction between enzymes and counterions in nonaqueous media. However, X-ray crystallography techniques generally give us a static perspective of reality, or certainly an average of many possible conformations in the crystal, and yet a dynamic picture is essential to gain a deeper knowledge of this problem. Additionally, the crystal environment, despite being similar to solution, is not exactly a free solution and crystal contacts can constrain the structure of the protein and influence the distribution of counterions. Herein, we complement the previous study by using a molecular dynamics (MD) simulation approach to characterize the interaction of counterions with subtilisin in acetonitrile solution and to obtain a dynamic perspective of this interaction.

3.3 Materials and methods

3.3.1 Calculation of the potentials of mean force (PMFs)

Before analyzing the interaction of counterions with subtilisin, we wanted to know how chloride ions interact with cesium and sodium ions in solvents with different polarities. Towards this end, we calculated the potentials of mean force between the anion (Cl^-) and the cation (Cs^+ or Na^+) in three different solvents: water, acetonitrile, and hexane. The calculation of the potentials of mean force was done using a methodology based on constrained MD simulations[189, 190]. This methodology was implemented by performing MD simulations with the application of distance constraints between the cation and the anion at many different distance values. The mean force at each constrained distance is then given by the negative of the average constraint force, and the PMF is obtained by integrating the mean force over the ion separation distances[189, 190].

To calculate the PMF between an ion pair in a given solvent, the cation and anion were placed in the center of a dodecahedron box that was filled with the solvent of interest. The distance between ions was constrained to values between 0.2 and 1.2 nm in intervals of 0.02 nm, using the LINCS algorithm[159], and for each distance, a MD simulation of 1 ns was performed. The MD simulations were performed as described below (see *Setup for MD simulations*). Note that, contrary to previous implementations of the methodology[189, 190], we used a reaction field correction to treat long-range electrostatic interactions.

3.3.2 MD simulations

Although the recent determination of the X-ray structure of subtilisin Carlsberg in acetonitrile in the presence of CsCl has enabled the determination of the counterions' binding sites in the crystal conditions, several questions remain unanswered. In order to have a complete picture of the interaction between subtilisin and counterions in acetonitrile solution, we have performed an extensive study, comprising a large number of MD simulations in different conditions. In this study, we have addressed several questions and used different sets of MD simulations (which are summarized in table 3.1) to tackle these questions.

3.3.3 Protein structures used in the MD simulations

The crystallographic structure of subtilisin Carlsberg obtained by Cianci *et al* at 2.24 Å resolution, after soaking the crystals with CsCl (PDB ID: 2WUV)[3], was used in the simulations where the ions were kept in their crystallographic binding sites. In the simulations where the ions were placed according to our docking methodology, we used the structure determined by the same authors at 2.23 Å resolution, in the absence of CsCl (PDB ID: 2WUW)[3]. All water and acetonitrile molecules found in the X–ray structure were used in our simulations.

3.3.4 Modeling protein protonation equilibrium

The determination of the pK_a of each titrable site in the protein was performed using a methodology developed by us, based on continuum electrostatics and Monte Carlo sampling of protonation states that has been

explained in detail before[171, 172] (the details of the protocol that was used can be found in the Appendix A).

Table 3.1. Overview of the MD simulations performed to tackle the main questions of this work

Question	Protonation state of $N\varepsilon_2$ of H64	Solvent	Cations	Anions	Method used to place ions	Number of replicates	Simulation time
1. Do the ions in solution occupy the same sites as in the X-ray structure?	Protonated	Acetonitrile	11 Cs^+	8 Cl^-	X-ray structure	20	10 ns
2. Are the simulations biased by the initial locations of the ions?	Protonated	Acetonitrile	4 Cs^+	6 Cl^-	Docking-based method	20	10 ns
3. How does the solvent influence the behavior of the ions?	Protonated	Water	11 Cs^+	8 Cl^-	X-ray structure	20	10 ns
4. What is the effect of replacing Cs^+ with Na^+?	Protonated	Acetonitrile	11 Na^+	8 Cl^-	X-ray structure (Cs^+ positions)	20	10 ns
5. What is the influence of the protonation state of H64?	Deprotonated	Acetonitrile	11 Cs^+/Na^+	8 Cl^-	X-ray structure	20	10 ns
6. How are ions distributed around the protein in water?	Protonated	Water	271 Cs^+	273 Cl^-	Random[1]	5[2]	50 ns

[1] The method that was used to randomly distribute the ions in these simulations is described in the Appendix A.

[2] To make sure that the ions that were initially randomly distributed in solution (far from the protein) would have enough time to reach the protein surface and explore a large number of binding sites, and given that in water simulations the enzyme is stable, 5 replicates of 50 ns were used, amounting to a total of 250 ns, which enables a good sampling.

3.3.5 Setup for MD simulations

The methodology used in the molecular dynamics simulations is similar to the one that we have used in many previous studies and is explained in detail elsewhere[1]. MD simulations were performed with the GROMACS package[154], version 4.0[191], using the GROMOS 53A6 force field[151]. Water was modeled with the simple point charge (SPC) model[192], the parameters from Gee *et al*[193] were used for acetonitrile and hexane was treated as a flexible united atom model using the GROMOS 53A6 parameters for alkanes[151]. The parameters of Reif et al.[194] were used for sodium, cesium and chloride ions. Bond lengths of the solute, acetonitrile and hexane molecules were constrained with LINCS[159], and SETTLE[195] was used for water. The simulations were performed at constant temperature and pressure. Temperature coupling was implemented using the Berendsen thermostat[161] with a reference temperature of 300 K. For the simulations carried out in acetonitrile and hexane, the protein, ions, and water were coupled to the same heat bath and the solvent was coupled to a separate heat bath. For the aqueous simulations, the protein and ions were coupled to the same heat bath and water was coupled to a separate heat bath. The pressure control was done by applying the Berendsen algorithm[161] with an isotropic pressure coupling, using a reference pressure of 1 atm and a relaxation time of 0.5, 1.3, and 1.5 ps for water, acetonitrile, and hexane simulations, respectively. An isothermal compressibility of 4.5×10^{-5} bar^{-1} was used for all the solvents. Nonbonded interactions were calculated using a twin-range method with short and long range cutoffs of 0.8 and 1.4 nm, respectively[196]. A reaction field correction for electrostatic interactions was applied[153, 197], considering a dielectric constant of 54[198] and 35.84[199] for water and acetonitrile, respectively. The preparation of the systems to run the

production MD simulations can be found in the Supporting Information, available in Appendix A.

3.4 Results and discussion

3.4.1 Potentials of mean force between the cations, Cs^+ and Na^+, and the anion, Cl^-, in solvents with different polarities

Before studying the interaction between ions and subtilisin, we considered that it was relevant to analyze how the isolated cations, Cs^+ and Na^+, each separately interact with the anion, Cl^-, in solvents with different polarities, namely, water, acetonitrile, and hexane. Towards this end, we calculated the potentials of mean force (PMFs) between the anion and the cation in these solvents (the results are shown and discussed in detail in the Supporting Information, available in Appendix A). Our results indicate that in hexane the interaction between oppositely charged ions is very strong and the ions form highly stable complexes that are never broken at room temperature (see fig. A2, Appendix A). In contrast, in water, the ions tend to be dispersed and do not form stable complexes at room temperature (as can be observed in the same figure). The PMF analysis displayed in fig. A2 (available in Appendix A) indicates that in acetonitrile the ions form stable associations, which was confirmed by unconstrained MD simulations (see fig. A3 , Appendix A). Our PMF analysis not only provides a description of how sodium and cesium interact with chloride in different media, but more generally, it gives us an idea of how oppositely charged particles interact in these media. On the basis of these results, one can expect the interaction between these ions and protein charged groups in acetonitrile to be stable.

3.4.2 Determination of the protonation state of ionisable residues at pH 6.5

Before starting a MD simulation of a protein, one needs to determine the protonation states of all the ionisable residues at the pH of interest. The solution that was used to soak the subtilisin crystals had a pH of 6.5, and therefore, this was the pH value that was considered when assigning the protonation states. The determination of the pK_a values of all the titrable residues of subtilisin was performed using a methodology based on continuum electrostatics, and the results are available in Appendix A. Our results indicate that, at the pH of interest (6.5), the protonated fraction of the catalytic histidine (H64) is around 70% (fig. A4, Appendix A). This means that both states (fully protonated and partially deprotonated) are expected to coexist at this pH. Although, according to our calculations, the fully protonated state is the predominant one, it is believed that this residue must be partially deprotonated in order to accept the proton from serine 221 during the catalytic process. Therefore, both states were considered in our MD simulations.

3.4.3 Stability of the simulations

The temporal evolution of the root mean square deviation (rmsd) from the X-ray structure and secondary structure content can be used to analyze the stability of a protein during a MD simulation. It is clear from figs. 3.1 and figs. A5 and A6 (available in Appendix A) that subtilisin is much more unstable in acetonitrile than in water simulations. These results are in line with previous MD simulation studies which show that subtilisin undergoes large conformational changes in acetonitrile[120]. The fact that the X–ray structure obtained shows a fold in acetonitrile very similar to the one obtained in

water is probably a consequence of the glutaraldehyde cross-linking performed before washing the crystals with acetonitrile. When enzymes are industrially used in acetonitrile, they will almost always be in some constrained solid state, immobilized on a surface, cross-linked, or even in crystals, all prepared initially in aqueous media and, therefore, the X–ray structure provides a good model of the interaction of counterions with subtilisin in these conditions. Our results indicate that during the simulation a considerable fraction of intra-main-chain hydrogen bonds are replaced by hydrogen bonds with acetonitrile molecules (data not shown). This can, at least partially, account for the loss of secondary structure that is observed in the MD simulations.

The analysis described above indicates that it is risky to prolong the simulations in acetonitrile for more than 10 ns, because the protein structure starts to show signs of unfolding. Therefore, although it would be desirable to have longer simulations, we decided to stop them at 10 ns. At this point, the enzyme structure is still reasonably similar to the X-ray structure (see fig A5 in Appendix A), which means that the ion distributions will not be affected by large protein conformational changes. Given that we cannot have longer simulations, to achieve more sampling, we used a large number of replicates (20) for each condition.

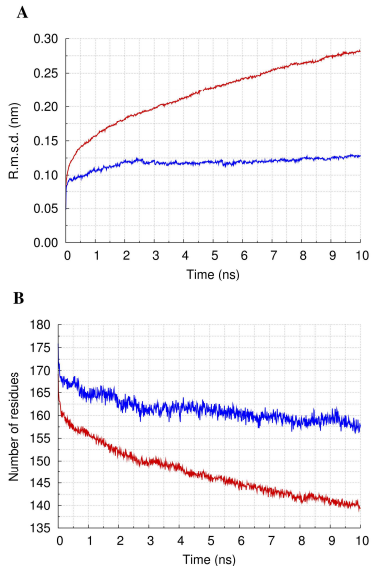


Figure 3.1. Temporal evolution of the average root mean square deviation (r.m.s.d.) of Cα atoms from the X-ray structure (A) and the average secondary structure content (B). The averages were calculated using all the simulations performed in each solvent (80 and 25 for acetonitrile and water, respectively (see table 3.1)). The secondary structure content was computed as the sum of the number of residues that are part of α-helices, β-sheets, β-bridges, or turns, according to DSSP criterion[200]. The blue and red lines correspond to the simulations performed in water and acetonitrile, respectively.

3.4.4 Comparison of X-ray and docking ion binding sites

As was described above (see table 3.1), we used two different strategies to find the initial locations of counterions. In the first approach, we used the previously determined X-ray structure with bound Cs^+ and Cl^- ions[3]. Additionally, we also placed ions using a docking protocol previously developed by us[1], which provides another reference to compare the ions' behavior in MD simulations using different initial positions. This unbiased protocol (which is described in Appendix A) docks ions on the enzyme surface until all the protein side chains are neutral.

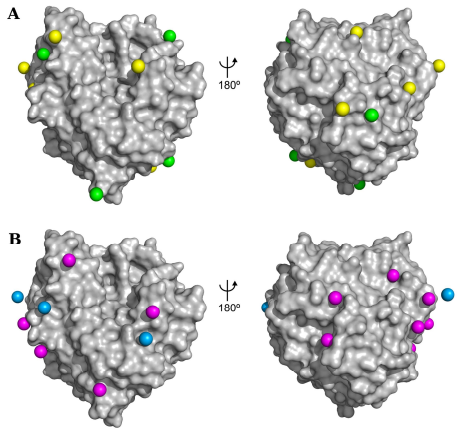


Figure 3.2. Comparison of X−ray and docking binding sites for (A) the chloride ions, yellow and green respectively, and (B) the cesium ions, magenta and blue, respectively.

In fig. 3.2, the locations of counterions obtained in the X−ray structure and using our docking methodology are compared. In the upper part of the figure, the binding sites of chloride ions are shown. The occupancies of the 8 crystallographic Cl^- sites sum to a total of 4.65 which is close to the number of docked chlorides (6). Three of the six Cl^- binding sites found with our docking methodology are close to crystallographic binding sites, although they are not interacting with the same residues. Intriguingly, the region that was found to be the most attractive site for Cl^- ions by our docking methodology (a very positively charged region formed by the N−terminus and a calcium ion) did not contain any Cl^- ion in the X−ray structure.

The number of cesium positions found by our docking methodology (4) is considerably smaller than in the X-ray structure (11). However, the sum of the X-ray crystal structure derived occupancies of the Cs^+ ions is 2.90, which indicates that there are various sites with the capacity to bind Cs^+, but only a

fraction can be occupied around a given protein molecule, in part because of repulsion between the Cs^+. In the crystals, different sites are occupied on different molecules, leading to the observed partial occupancy. Similarly to what was observed for chloride ions, three approximately coincident binding sites were found with the two approaches.

There are two constraints that account for the observed differences in the locations of the ions found by the two methodologies. The first one is the fact that contacts between adjacent molecules in the crystal can create artificial binding sites that are not found in solution. Indeed, in this specific case, there are a large number of ions coordinated by two or more distinct protein molecules in the X-ray crystal structure. The second reason that can explain the observed differences is the fact that, in our docking methodology, the positive and negative ions are docked separately, whereas, obviously, in the soaking solution both ions are present simultaneously. This explains why Cl^- and Cs^+ ions never form ion pairs in the docking methodology, contrary to what is observed in the X-ray crystal structure.

3.4.5 Occupancy of the ion binding sites during MD simulations

In order to evaluate the affinity of chloride and cesium ions for the binding sites that were found in the X–ray structure or using our docking methodology, we calculated the occupancy of each of these binding sites throughout the last 8 ns of simulation. This occupancy corresponds to the fraction of time during which the binding site is occupied by a Cl^- or Cs^+ ion.

In fig. 3.3, ions are colored according to the occupancies of the original binding site. It is clear from fig. 3.3A that the affinity of Cl^- ions for the X–ray

binding sites in acetonitrile simulations displays a high variability. Half of the binding sites have occupancies greater than 0.5, and the occupancies range from very low (there is one binding site with an occupancy less than 0.1) to high (one of the binding sites is occupied more than 80% of the time). In the simulations performed with the docked ions, five of the six chloride binding sites have an occupancy superior to 0.9 (see fig 3.3C), indicating that these sites are very attractive locations for Cl^- ions. Cesium ions exhibit a very different behavior. As can be observed in fig. 3.3B and D, most Cs^+ ions have occupancies lower than 0.1. These cations do not spend much time in any of the binding sites, irrespective of whether these sites are the crystallographic ones or the ones obtained with our docking methodology.

One of our aims was to analyze the effect of replacing cesium by sodium ions. Comparing figs. 3.3B and F, we can see that there are two binding sites which are considerably more populated by Na^+ than by Cs^+ (blue and cyan spheres in fig. 3.3F). As can be found by comparing figs. 3.3A and E, replacing the cations influences the behavior of some of the Cl^- anions. Four of the chloride binding sites which had an occupancy superior to 0.6 in the presence of cesium (cyan and blue spheres in fig. 3.3A) are less populated when cesium is replaced by sodium. Visual analysis of the trajectories indicates that the chloride ions often form ion pairs with sodium ions and these ion pairs can (temporarily) migrate to the solvent. This behavior is not observed when the cation used is cesium, which makes sense in the light of our previous results that showed that the Na^+Cl^-ion pair is stronger than its Cs^+Cl^- counterpart (see the potential of mean force analysis, above).

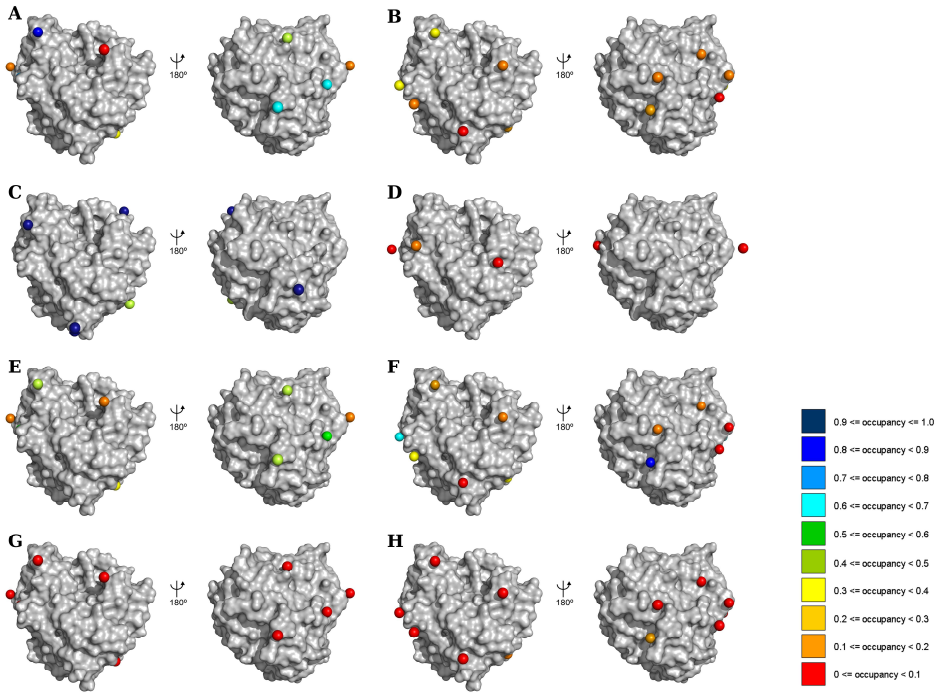


Fig. 3.3: Occupancies of the original ion binding sites during MD simulations. The spheres represent the chloride (left side) and cesium or sodium ions (right side) placed in the initial binding sites and colored according to the respective occupancy (see scale). A and B correspond to the X-ray binding sites; C and D correspond to the locations found with the docking method; E and F correspond to the simulations where the X-ray cesium ions were replaced by sodium ions; and G and H correspond to the simulations in water using the X-ray positions. To calculate the occupancy, we first found the residues which comprise each binding site and then counted the number of frames in which the minimum distance between these residues and the corresponding ion was smaller than 0.4 nm (this cutoff was chosen after inspecting the histogram of the minimum distance between the ions and binding site residues) and then divided this value by the total number of frames.

In order to elucidate the role played by the solvent in the observations described above, we performed control MD simulations in water. Figure 3.3G and H shows that, in these simulations, both Cl^- and Cs^+ have low occupancies (in most cases, lower than 0.1). This is not surprising, if we think that, in

aqueous media, ions are generally found in the bulk solution rather than on the enzyme surface (except in the case of high affinity binding sites). This is also consistent with our potential of mean force PMF analysis (see above) that shows that cesium and chloride tend to be dissolved in water.

3.4.6 Distribution of counterions on the enzyme surface in acetonitrile simulations

In order to determine which regions of the enzyme are more populated by counterions during our simulations, we calculated the probability density maps of the ions during the last 8 ns of simulation. The limited temporal extent of our acetonitrile simulations (which is a consequence of the poor stability of the enzyme in this media (see the analysis of the Stability of the Simulations, above)) and the low number of ions used could compromise our sampling and bias the distribution of the ions. To avoid this, we used a large number of replicates (20) and tested the convergence of our probability density maps. For each set of simulations, we have divided our sample into two subsets of 10 replicates and calculated the ion probability density maps for each subset. We observed that the maps obtained in the two subsets of simulations are similar (results not shown), which indicates that our sampling is good and that our probability density maps are reliable.

Comparing the maps obtained in the simulations of the X–ray and docked Cl^- ions (fig. 3.4A and D) we can see that these maps are similar. The similarity between the maps obtained using two distinct methodologies for the initial placement of the ions indicates that the behavior of these ions during the course of the simulations is not strongly biased by the choice of the initial binding sites. Moreover, these results show that the docking methodology

that we have been using to place ions in our simulations of proteins in nonaqueous media enables a good prediction of Cl^- binding sites.

Comparing the probability density maps of Cs^+ obtained for the simulations with the X–ray and docked ions (fig. 3.4B and E), we can see that there are overlapping regions, although there are areas that are populated in the simulations performed with the X–ray ions and not in the simulations where ions were placed according to docking predictions. This is probably a consequence of the fact that the number of cesium positions found in the crystal structure is considerably higher than the number of ions found through our docking methodology. However, these cesium positions are not fully occupied in the X–ray structure and would not be expected to be occupied at the same time. Therefore, the probability density maps that were obtained for Cs^+, in the simulations where the ions were initially placed in the crystallographic positions, are biased by the fact the number of cesium ions used is not realistic (although the sum of the X–ray derived occupancies is reasonable).

As has been mentioned above, one of the aims of this study is to evaluate the consequences of replacing the cesium ions that were found in the X–ray crystal structure by sodium ions. The probability density maps obtained for Cs^+ and Na^+ (fig. 3.4B and H, respectively) are similar, which means that the two cations populate approximately the same regions of the protein surface. These results support the hypothesis that the Cs^+ binding sites found in the X–ray crystal structure may be occupied by Na^+ or K^+ in biological conditions, as has been previously proposed[3], supporting that it is valid to soak crystals with Cs^+ (which is easier to distinguish from water than smaller cations) to identify the positions of Na^+ and K^+ ions. Comparing parts A and G of fig. 3.4, we can

see that the chloride density around the protein surface is considerably lower in the presence of Na^+ than in the presence of Cs^+, which is a consequence of the tendency of Cl^- to pair with Na^+ and migrate to the bulk solution. Some difference in the behavior of Cs^+ and Na^+ ions is consistent with the difference in catalytic activity between crystals soaked with these different salts[1].

With the purpose of analyzing the behavior of chloride, cesium and sodium ions during the time course of the simulations, we looked at the evolution of the probability density maps of the ions. In our analysis, we divided the simulations in 10 windows of 1 ns each and calculated the probability density map for each window. All the replicates were included in the calculation, and therefore, the maps represent the average probability density. In Movie_A1 (see Supporting Information), we can see that the crystallographic chloride ions, which are concentrated around their original binding sites in the beginning of the simulations, tend to get more dispersed as the simulation progresses, and occupy a larger portion of the protein surface. The analysis of the trajectories of the MD simulations indicates that most anions explore large regions around the initial position but rarely move to distant areas or abandon the protein surface. Interestingly, it can be observed that ions that were found in the bottom right area of the protein in the crystal structure migrate slightly down and to the center of the enzyme and end up occupying a region that was found by our docking methodology to have a strong interaction with Cl^-.

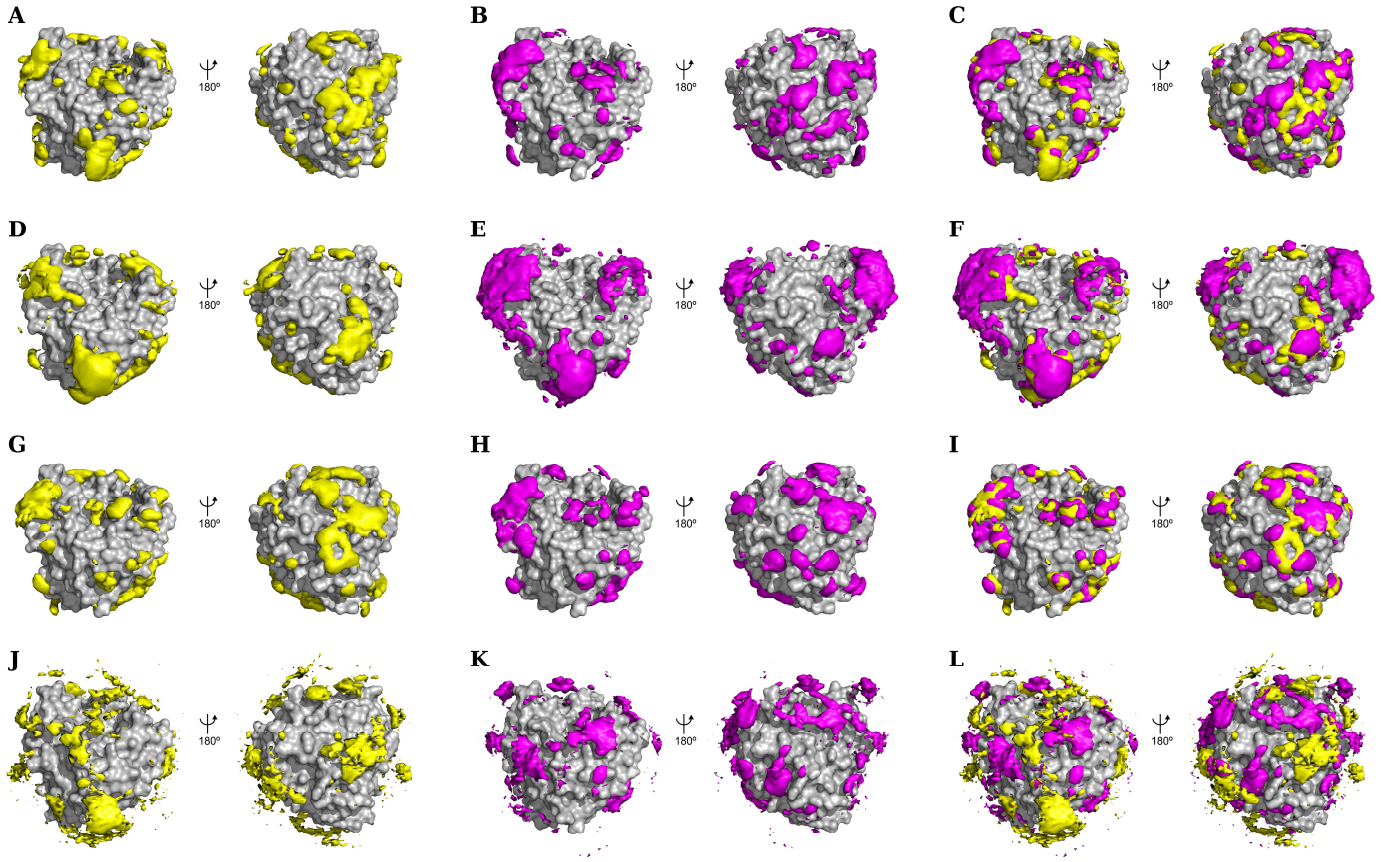


Figure 3.4. Average probability density maps of chloride (yellow surfaces) and cesium or sodium (magenta surfaces) in the last 8 ns of simulation. The contours enclose regions with a probability density above 2×10^{-5} Å^{-3} and 6×10^{-6} Å^{-3} for acetonitrile and water simulations, respectively. The left-hand column shows the map for Cl^-, the middle column shows the map for Cs^+ or Na^+, and the right-hand column shows the two maps together. A, B, and C correspond to the simulations with the X–ray ions in acetonitrile. D, E ,and F correspond to the simulations with the docked ions in acetonitrile. G, H, and I correspond to the simulations where the X-ray cesium ions were replaced by sodium ions. J, K, and L correspond to the water simulations with 1.5 M of CsCl.

The docked Cl^- ions are more stable than their X-ray crystal structure counterparts and, in most cases, remain concentrated around their original binding sites throughout the simulations (Movie_A3 in Supporting Information).

As can be seen in movies Movie_A2 and Movie_A4 (available in Supporting Information), cesium ions are very rapidly dispersed, both in the simulations performed with the X–ray ions and in the ones where the ions were placed according to our docking methodology. It is clear from these movies that these cations are very dynamic.

As has been mentioned above, we analyzed the effect of replacing cesium by sodium ions. Looking at Movie_A6 (available in Supporting Information), which shows the behavior of sodium ions, and comparing this movie with Movie_A2 (available in the Supporting Information), showing the behavior of cesium ions, we see that the sodium ions are less mobile than the cesium ions. Comparing the behavior of chloride ions in the presence of cesium (Movie_A1) and sodium (Movie A5) shows that the chloride ions get more dispersed in the presence of sodium.

To complete our study of the interaction between subtilisin and counterions, we analyzed the ions' tendency to remain close to the protein surface, by measuring the temporal evolution of the distance between the ions and the protein (results not shown). This analysis shows that, as could be inferred from the previous results, in the simulations using CsCl, Cl^- ions tend to be close to the protein surface and almost never go to the bulk of the solution (the percentage of time that the ions spend in the solvent is ≈3%). Cesium ions, on the other hand, move in and out from the protein surface to the solvent and spend around 40% of the

time in solution, showing once again a very dynamic behavior. When the cation used is Na^+ instead of Cs^+, the situation changes considerably. In these simulations, chloride ions spend a larger percentage of time (≈11%) in solution and sodium ions are less frequently found in the solution bulk (≈24% of the total simulation time) than cesium ions. Interestingly, we observed that the Cl^- and Na^+ ions tend to go into the bulk solution as ion pairs and not as isolated ions. The percentage of time that isolated ions pass in solution is around 1% for Cl^- (comparable to the simulations with Cs^+) and 11% for Na^+ ions (considerably lower than the 40% value obtained for Cs^+). In the simulations using Cs^+ ions, we did not observe this behavior and the ions that go into the bulk solution, in the great majority of cases, migrate alone and not as ion pairs.

The higher tendency of cesium and sodium ions to move into the solution bulk when compared with chloride ions can be explained on the basis of the evidence showing that acetonitrile (and polar aprotic solvents in general) solubilize positive ions considerably better than negative ions[201-206]. This difference is most likely due to the fact that, in aprotic solvents, the negative end of the dipole is concentrated in a small, accessible part of the molecule, whereas the positive end of the dipole is distributed over a large and difficult to access region[203, 204]. Although Na^+ has been shown to have a more negative absolute free energy of solvation in acetonitrile than Cs^+ [206], we observed that cesium ions spend more time in the bulk solution than sodium ions. This finding can be attributed to the higher tendency of sodium ions to interact with the protein's charged or polar groups and is in line with the results obtained in the PMF analysis which showed that the Na^+Cl^- ion pair was more stable than the Cs^+Cl^-pair. This observation is also consistent with previous studies[94, 96, 97], where it was found that Na^+ binds

more strongly to protein surfaces than K^+. This finding was attributed to the fact that cations tend to pair with anions of similar surface densities[91], and sodium matches carboxylate anions (found in glutamate and aspartate residues) better than potassium. The difference between Na^+ and Cs^+ is expected to be even more pronounced, because cesium has a much lower surface charge density. It is worth noting that the former results were obtained in water, and this is a solvent-dependent effect, i.e., the tendency to form an ion pair depends on a delicate balance between the cost of desolvating the ions and the benefit of forming the ion pair. Given that the differences between the solvation free energies of Na^+ and Cs^+ in water and acetonitrile are similar[206], the same trend should be observed in the two solvents, which is in agreement with our results.

3.4.7 Distribution of counterions on the enzyme surface in water simulations

In order to analyze how counterions interact with subtilisin in water, we have performed MD simulations in which the enzyme was placed in an aqueous solution containing 1.5 M CsCl. In the beginning of these simulations, the ions were randomly distributed in the most external region of the water box, far from the protein. Not surprisingly, the ions do not form very stable interactions with the enzyme and spend 98.2% (for Cl^-) and 95.5% (for Cs^+) of the time in solution. Nevertheless, there are some regions of the protein surface where the ions accumulate. Looking at fig. 3.4, we can see that, although there are clear differences between the maps obtained in water (fig. 3.4J and L) and acetonitrile (fig. 3.4A and C), there is some overlap between them. This indicates that, although the nature of the solvent influences the interaction

between counterions and the protein, some binding sites are conserved in different solvents. Additionally, fig. 3.4C and L show that Cl^- and Cs^+ ions tend to be close to each other in acetonitrile (the yellow and magenta surfaces often overlap) but not in water.

The crystallographic structure of subtilisin, crystallized in aqueous conditions and soaked with a solution containing 1.5 M CsCl (Cianci *et al.* (to be published)), has recently been determined at 2.28 Å resolution. This structure enabled the identification of a significant number of ion binding sites in this aqueous crystal environment, which were compared with the distributions of the ions in the MD simulations in water. Figure 3.5 shows the probability density maps obtained in the aqueous simulations with 1.5 M CsCl and the positions of the ions in the X-ray structure obtained in aqueous conditions (Cianci et al. (to be published)). In fig. 3.5A, we can see that there is almost no overlap between the map obtained for Cl^- and the positions that were found in the X–ray structure. The distribution of cesium ions in our simulations shows some agreement with the binding sites found in the X–ray structure (fig. 3.5B). Examining the locations of chloride and cesium ions in the crystal structure, we observed that two chloride and two cesium ions are interacting with more than one molecule in the crystal. These crystal contacts can create artificial binding sites that will not be found in solution, and this can explain why the ions did not populate these sites in our simulations. However, for the binding sites which are not formed by more than one enzyme molecule, there has to be an alternative explanation for the disagreement between the X–ray and MD simulation results.

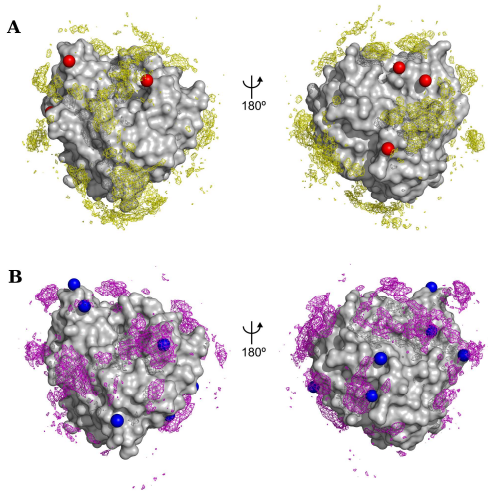


Figure 3.5. Comparison of the probability density maps (represented using a mesh) for chloride (A) and cesium ions (B) obtained in the MD simulations in water using 1.5 M of CsCl with the positions of the chloride (red spheres) and cesium (blue spheres) ions in the crystal structure obtained in aqueous conditions (Cianci et al. (to be published)). The contours enclose regions with a probability density above 6×10^{-6} $Å^{-3}$.

In an attempt to understand these differences, we calculated the electrostatic potential in the crystal and in solution. This calculation was done using the *potential* tool, available in the Mead package[207], version 2.2.5, and assigning dielectric constants of 2.0 and 80.0 to the protein interior and solvent, respectively. To simulate the crystal environment, we reconstructed the neighboring asymmetric units from the PDB file using the software PyMOL[208] (www.pymol.org). All water molecules and chloride and cesium ions were removed from the structure, in order to investigate the potential created by the protein alone. Our calculations indicate that the crystal environment strongly influences the electrostatic potential on the protein surface (see fig. A7, available in Appendix A). In the crystal, a large fraction of the protein surface has a positive

potential, whereas in solution this is not observed. Looking at the potential in the chloride X–ray binding sites, we observed that in most cases it is clearly positive in the crystal environment and becomes more negative in solution. This explains why the chloride binding sites found in the X–ray structure are not very populated during the simulations. In what concerns the potential in the cesium binding sites, we observed that, in general, it is more negative in solution than in the crystal and, therefore, in this case we have a better agreement between the theoretical and experimental results.

3.4.8 Analyzing the effect of different cations on the activity of subtilisin

In a previous work, it was observed that the type of countercation used influences the catalytic activity of subtilisin crystals in acetonitrile: using a larger cation gives a larger rate enhancement[3]. In order to rationalize this observation, we decided to compare the distribution of counterions in the active site of subtilisin, in the simulations using CsCl and NaCl. As was mentioned above, our pK_a calculations indicate that, at a pH of 6.5, the catalytic histidine can be either fully protonated or have only one proton (with probabilities of ≈70 and ≈30%, respectively). Therefore, for each salt used, we have performed MD simulations considering the two possible protonation states of H64 and analyzed how the ions are distributed around the active site of subtilisin in the four sets of simulations. In fig. 3.6, we can observe that, in the simulations where H64 is protonated and therefore positively charged, there is an accumulation of chloride ions very close to this residue when the cation used is Cs^+ (fig. 3.6A).

This is not observed when the histidine is neutral (fig. 3.6B). Curiously, when the cation used is Na^+, we do not observe such a high concentration of chloride in the vicinity of the catalytic histidine (fig 3.6C). The radial distribution functions displayed in fig. A8 (see Appendix A), show that Cl^- has a considerably higher probability of forming an ionic interaction with H64 when the cation used is Cs^+ compared with Na^+, which is probably a consequence of the higher tendency of Cl^- to bind to Na^+ than to Cs^+. From these results, one would expect that in the presence of Cs^+ the charged state of H64 would be more stabilized (due to the higher concentration of Cl^- in the vicinity of H64) than in the presence of Na^+. Therefore, one would predict that subtilisin would be more active when Na^+ is used instead of Cs^+, because it is accepted that the catalytic histidine needs to be in the neutral state in order to be active.

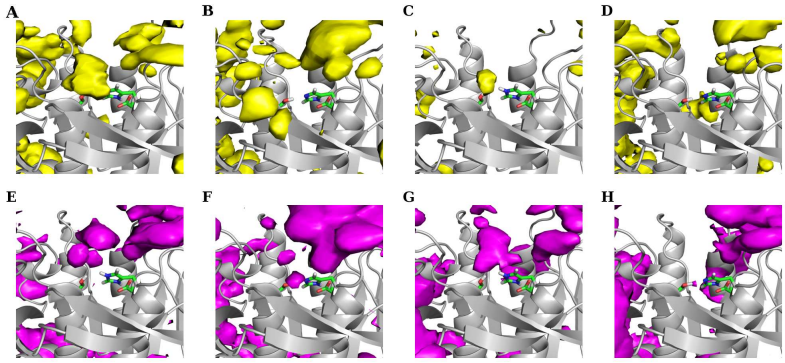


Figure 3.6. Probability density maps of chloride (yellow surfaces) and cesium or sodium (magenta surfaces) ions in the active-site of subtilisin in acetonitrile simulations. The contours enclose regions with a probability density above 2×10^{-5} $Å^{-3}$. A and E correspond to the simulations with CsCl and charged H64; B and F correspond to the simulations with CsCl and neutral H64; C and G correspond to the simulations with NaCl and charged H64; D and H correspond to the simulations with NaCl and neutral H64.

However, this is inconsistent with the experimental observations, which indicate that subtilisin is more active in the presence of larger cations. In the light of these observations, we propose an alternative explanation, which is to consider that the Cl^- ion could accept the proton from H64, stabilizing the catalytically active neutral state. Although this may seem counterintuitive because we are used to thinking in aqueous conditions, it is possible that in a moderately polar medium like acetonitrile the equilibrium represented in eq. 3.1 is shifted towards the right side. A good indication that this hypothesis is plausible is the fact that a value of 10.3 has been determined for the p*K*a of HCl in acetonitrile[209].

$$\mathrm{His-H^+ + Cl^- \leftrightarrow His + HCl} \tag{3.1}$$

Given that there are more chloride ions available when the cation used is Cs^+ than when Na^+ is used (because the Na^+Cl^- is more stable than the Cs^+Cl^- pair), the neutral state of H64 would be more stabilized in the presence of Cs^+ and this would explain why subtilisin is more active in the presence of CsCl.

We emphasize that the proposed explanation described above is just one hypothesis. We do not have enough evidences to confirm it and we do not exclude that other factors might contribute to the cation-dependence of subtilisin activity. However, we think our explanation is plausible, and it opens the door for future studies, which may clarify this question.

3.5 Conclusions

In this work, we used molecular dynamic simulations to complement the X-ray crystallographic analysis of the interaction between subtilisin and counterions in acetonitrile, performed in a previous work[3]. In order to analyze the interaction between subtilisin and counterions in acetonitrile and to characterize the dynamic behavior of the system, we performed two different sets of simulations in acetonitrile. In the fist set, the initial positions of cesium and chloride ions were the ones available in the X–ray crystal structure determined after soaking with CsCl. In the second set, we used a methodology based on docking simulations to find the ion binding sites, and then started the MD simulations with the ions placed in those locations. Our results indicate that some of the chloride binding sites found in the X–ray crystal structure are highly populated during the simulations, whereas others are rarely occupied. The Cl^- binding sites determined using the docking methodology have high occupancies in the MD simulations. Cesium binding sites have low occupancies independently of the method that was used to define their initial binding sites. We also observed that chloride ions tend to stay close to the protein, whereas cesium ions are considerably more dynamic and frequently move into the bulk solution. Comparing the distribution of the ions in the two sets of simulations, we observed that they are reasonably similar, which indicates that the simulations are not strongly biased by the initial locations of the ions.

Additionally, we performed simulations in which the crystallographic cesium ions were replaced by sodium ions. The distribution of sodium and cesium ions around the protein surface is similar, indicating that the Cs^+ binding sites found in the X–ray crystal structure may be occupied by Na^+ or K^+ in biological conditions, as

previously proposed[3]. Therefore, using soaking with Cs^+ as a method to identify the position of Na^+ and K^+ is here validated. We also observed that Cl^- and Na^+ ions frequently form ion pairs and move into the bulk solution together. This leads to a decrease in the concentration of chloride ions bound to the protein. Most interestingly, this is observed in the vicinity of the catalytic histidine, when this residue is positively charged. We propose that, in acetonitrile, chloride can accept the proton from the charged H64, moving the equilibrium towards the catalytically active neutral state, which can explain the previous experimental observations showing that subtilisin is more active when the cation present is Cs^+ or choline when compared with smaller cations[3].

In addition to the acetonitrile simulations, we performed simulations in water, using 1.5 M of CsCl. The analysis of the probability density maps showed that there are some differences in the distribution of the ions around the enzyme surface in water and acetonitrile, although the maps have some overlapping regions. Additionally, we observed that in water the ions are much more frequently found in the bulk solution than in acetonitrile. These results indicate that the solvent influences the interaction between the ions and the protein. Comparing the probability density maps obtained in our simulations with the positions of the ions in the X–ray crystal structure obtained in an aqueous medium (Cianci et al. (to be published)), we observed that there is some agreement in the case of cesium, but not in the case of chloride ions. The difference between the results obtained in the simulations and the chloride binding sites found in the crystal structure can be explained by the fact that the crystal lattice can generate an electrostatic potential which is very different from the one found in solution.

3.6 Acknowledgements

The authors acknowledge Dr. Nuno Micaêlo, Dr. Susana Barreiros and Dr. André Melo for helpful discussions, and the financial support from Fundação para a Ciência e a Tecnologia, Portugal, through grants SFRH/BD/28269/2006, POCTI/BIO/57193/2004 and PEst-OE/EQB/LA0004/2011. MC, PJH and JRH are grateful to the Synchrotron Radiation Source at STFC Daresbury Laboratory for X-ray beamtime; details are as described in ref 3. The EMBL Hamburg (MC), the University of Manchester (JRH) and Strathclyde University (PJH) are thanked for general support; details are again as described in ref 3.

Chapter 4

Analyzing the molecular basis of enzyme stability in ethanol/water mixtures using molecular dynamics simulations

This work has been published in the following paper:

Lousa D., Baptista A. M., Soares C. M. **(2012)**, Analyzing the molecular basis of enzyme stability in ethanol/water mixtures using molecular dynamics simulations, *Journal of Chemical Information and Modeling*, vol. 52, pp 465-473

doi: http://dx.doi.org/10.1021/ci200455z

Contributions of the author of the present thesis to this work:

The author of the present thesis has participated in the design of this study and executed all the simulations and analysis described in this chapter.

4.1 Abstract

One of the drawbacks of nonaqueous enzymology is the fact that enzymes tend to be less stable in organic solvents than in water. There are, however, some enzymes that display very high stabilities in nonaqueous media. In order to take full advantage of the use of nonaqueous solvents in enzyme catalysis, it is essential to elucidate the molecular basis of enzyme stability in these media. Towards this end, we performed μs-long molecular dynamics simulations using two homologous proteases, pseudolysin and thermolysin, which are known to have considerably different stabilities in solutions containing ethanol[4]. The analysis of the simulations indicates that pseudolysin is more stable than thermolysin in ethanol/water mixtures and that the disulfide bridge between C30 and C58 is important for the stability of the former enzyme, which is consistent with previous experimental observations[4, 5]. Our results indicate that thermolysin has a higher tendency to interact with ethanol molecules (especially through van der Waals contacts) than pseudolysin, which can lead to the disruption of intra-protein hydrophobic interactions and ultimately result in protein unfolding. In the absence of the C30-C58 disulfide bridge, pseudolysin undergoes larger conformational changes, becoming more open and more permeable to ethanol molecules which accumulate in its interior and form hydrophobic interactions with the enzyme, destroying its structure. Our observations are not only in good agreement with several previous experimental findings on the stability of the enzymes studied in ethanol/water mixtures but also give an insight on the molecular determinants of this stability. Our findings may, therefore, be useful in the rational development of enzymes with increased stability in these media.

4.2 Introduction

The application of organic solvents in enzyme catalysis is of great technological and fundamental interest, because enzymes in these media can display novel properties[30], such as the capacity to catalyze reactions that are not feasible in water[28], different substrate specificity and enantioselectivity[71, 74, 76, 79, 210], and molecular "memory[6, 112, 183, 184]. Computational tools for the understanding of enzyme mechanisms, both at the kinetic level (see refs.[211-214] for recent examples) as well as at the atomic level[1, 2, 74, 120, 137, 139, 184], have proven to be important for a deeper understanding of enzyme catalysis[24, 215-217] and, in particular, enzyme catalysis in nonaqueous solvents[66, 218].

Despite its great technological potential, the use of enzymes in nonaqueous solvents has limitations and one of the most serious is the fact that enzymes in organic solvents are usually less stable than in water. Several strategies have been used to overcome this limitation, including chemical modification, enzyme immobilization, protein engineering, and directed evolution[219, 220]. Another promising approach is the search for enzymes that are naturally stable in organic solvents[221, 222]. Using the latter strategy, Ogino et al. found that the stability of the protease pseudolysin (PSL) in solutions containing hydrophilic solvents is higher than in pure water[4]. Moreover, they observed that pseudolysin is more stable in these solutions than other proteases, namely subtilisin Carlsberg, α-chymotrypsin and thermolysin (TLN)[4].

Pseudolysin, also known as *Pseudomonas elastase*, is a zinc metalloprotease secreted by Pseudomonas aeruginosa that belongs to the protein family M4. Although its precise biological function is not completely clear, it is known that this enzyme plays a role in the infectious process of P. aeruginosa[223-225], and that

PSL can degrade elastin (hence the name elastase)[226], as well as collagen,[227] human IgG[228] and other important human proteins and peptides. Thermolysin is a thermostable enzyme secreted by *Bacillus thermoproteolyticus*. This protease is the prototypical enzyme of the M4 family of zinc metalloproteases, being a neutral endopeptidase that specifically hydrolyzes peptide bonds containing hydrophobic residues[229]. The different stability displayed by pseudolysin and thermolysin in ethanol/water mixtures is curious, given that they share 28% sequence identity, a similar fold, and a conserved catalytic center (composed by a zinc atom tetrahedrally coordinated by a glutamate, two histidines and a water molecule). The main difference between the structures of the two proteases is the presence of two disulfide bonds in pseudolysin (between Cys-30 and Cys-58, and between Cys-270 and Cys-297) that are absent from thermolysin (see fig. 4.1). It has been shown that the disulfide bond located in the C-terminal domain is essential for the protein activity, whereas the bond between Cys-30 and Cys-58 is very important for the solvent stability of PSL[5].

Our aim is to gain a deeper understanding of the molecular determinants underlying the different stability displayed by pseudolysin and thermolysin in solutions containing ethanol. Additionally, we intend to elucidate the role played by the disulfide bond between C-30 and C-58 in maintaining the stability of pseudolysin. With this objective, we have performed μs-long molecular dynamics (MD) simulations of PSL and TLN, in pure water and in an ethanol/water mixture, and of the C58G mutant of PSL, in ethanol/water. The behavior of the enzymes in our simulations is consistent with the previous experimental observations[4, 5], and the analysis of the protein-ethanol interactions enabled us to unravel the molecular causes of this behavior.

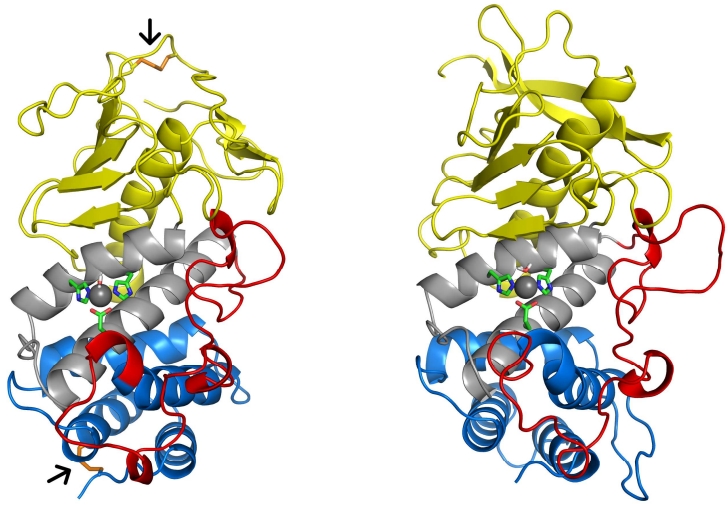


Figure 4.1. X-ray structures of pseudolysin (left) and thermolysin (right). Both proteins are composed by three domains: N-terminal domain (yellow), active site domain (gray) and C-terminal domain (blue and red). The red color is used to highlight the loop comprising residues 180 to 224 and 181 to 229 in PSL and TLN, respectively, that is very mobile (see the Results section). The residues of the catalytic center are shown in sticks, and the two arrows indicate the disulfide bridges of PSL, which are displayed using orange sticks.

4.3 Materials and methods

We performed five sets of MD simulations, as summarized in table 1. Given that we are trying to capture a slow phenomenon, i.e. loss of protein stability, the simulations were run for 1 μs. Although 1 μs is a short period of time compared with the time-scale of unfolding, our aim was to capture early signs of stability. Additionally, in order to obtain a good sampling, five replicates were calculated for each system under study.

Table 4.1. Description of the systems analyzed in this work

Short description	Enzyme	Solvent
PSL in water	Wild type pseudolysin	Water
PSL in eth/water	Wild type pseudolysin	Ethanol + water (25% v/v)
PSL-C58G in eth/water	C58G mutant of pseudolysin	Ethanol + water (25% v/v)
TLN in water	Wild type thermolysin	Water
TLN in eth/water	Wild type thermolysin	Ethanol + water (25% v/v)

For thermolysin, the X-ray structure determined by Holland et al. at 1.70 Å resolution (PDB code: 1LNF)[230] was used. In the pseudolysin simulations, we used the X-ray structure obtained by Thayer et al. at 1.50 Å resolution (PDB code: 1EZM)[231]. The mutant C58G of pseudolysin was obtained by removing all the side-chain atoms of Cys-58, transforming this residue into a glycine.

The determination of the protonation state of each titrable site in the protein at pH 7 was performed using a methodology developed by us, based on continuum electrostatics and Monte Carlo sampling of protonation states, that has been explained in detail before[171, 172] (further details can be found in Appendix B).

MD simulations were performed with the GROMACS package[154], version 4.0[191], using the GROMOS 53A6 force field[151]. Water was modeled with the simple point charge (SPC)[192] model. Bond lengths of the solute and ethanol molecules were constrained with LINCS[159] and for water molecules, the SETTLE[195] algorithm was used. The temperature and pressure were kept constant during the simulations. Temperature coupling was done using the Berendsen thermostat[161] with a temperature coupling constant of 0.1 ps and a reference temperature of 300 K.

The protein and solvent (water or ethanol/water) were coupled to separate heat baths. The pressure was controlled by applying the Berendsen algorithm[161] with an isotropic pressure coupling, using a reference pressure of 1 atm, a relaxation time of 0.5 ps and an isothermal compressibility of 4.5×10^{-5} bar^{-1}. Nonbonded interactions were calculated using a twin-range method with short- and long-range cutoffs of 8 Å and 14 Å[196], respectively. A reaction field correction for electrostatic interactions was applied[153, 197], using dielectric constants of 78 and 67 for water and the ethanol/water mixture (25% v/v), respectively[232]. The preparation of the systems to run the production MD simulations can be found in the Appendix B.

4.4 Results and discussion

4.4.1 Structural stability of the proteins in water and ethanol/water simulations

A common way to analyze the structural stability of a protein in a MD simulation is to monitor the root mean square deviation (rmsd) from the initial structure along the simulation. The rmsd from of all the systems under study is shown in fig. 4.2A. It is clear from these plots that after 1 μs of simulation all the structures are considerably distinct from the X-ray structures that were employed as the starting point of the simulations.

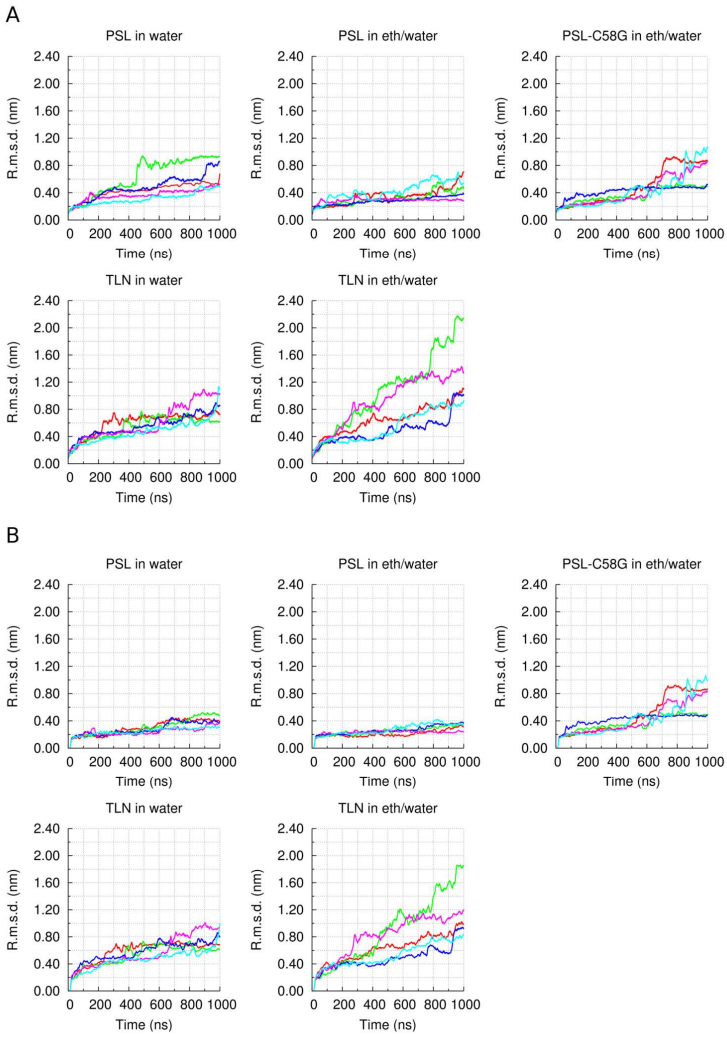


Figure 4.2. Moving average of the rmsd from the X-ray structure calculated using all the protein Cα atoms (panel A) and excluding the loop comprising residues 180 to 224 and 181 to 229 for PSL and TLN, respectively (panel B). The different replicates are represented by lines with different colors (replicate 1: red, replicate 2: green, replicate 3: blue, replicate 4: magenta, and replicate 5: cyan).

In the case of wild-type pseudolysin, the structure is more conserved in ethanol/water (average rmsd in the last 100 ns = 0.45 nm) than in the pure water simulations (average rmsd in the last 100 ns = 0.65 nm). This is in agreement with previous experimental findings that show that the half-life of pseudolysin in an ethanol/water solution (25% v/v) exceeds 100 days, whereas in aqueous solution it is around 9 days[4]. Studies of a closely related protease, elastase strain K, which has an identity of 99% with PSL, have also shown that this enzyme is more stable in ethanol/water mixtures (25% v/v) than in aqueous solution[233]. The pseudolysin mutant C58G in ethanol/water deviates more from the X-ray structure than the native enzyme (average rmsd in the last 100 ns = 0.72 nm), which is also in accordance with site-directed mutagenesis experiments, where it was found that the C58G mutant has a considerably lower half-life (~5 days)[5] than the wild-type[5]. Thermolysin is quite unstable in both media analyzed, although it clearly undergoes considerably larger conformational changes in ethanol/water simulations than in pure water (the average r.m.s.d. values in the last 100 ns of simulation are 1.25 and 0.80 nm, for ethanol/water and water, respectively). Once again, this is consistent with the experimentally measured half-lives of TLN in ethanol/water and pure water, which are 3 and 10.8 days, respectively[4]. These results are also in line with another study, where it was observed that the thermal stability of thermolysin is severely decreased in the presence of 50% (v/v) of n-propanol, relative to aqueous solution[234].

Visual inspection of the trajectories obtained in the simulations of pseudolysin gave us an indication that the largest conformational changes took place in the loop comprising residues 180 to 224 (highlighted in red in fig. 4.1), located in the

C-terminal domain. Therefore, we performed new rmsd calculations without including this loop. Comparing the plots in figs. 4.2A and B, it can be seen that the rmsd of the wild-type pseudolysin is considerably lower when this loop is not included, especially in water simulations. This means that the high rmsd values observed in the wild-type pseudolysin simulations are mainly caused by the conformational changes in one of its loops and do not necessarily represent protein unfolding. In what concerns the pseudolysin mutant C58G, the rmsd obtained with or without including the loop is very similar, indicating that the structural changes are not localized in this loop and probably reflect global protein unfolding. In the simulations of thermolysin in water, the rmsd calculated without including the loop comprising residues 181 to 229 is very similar to the one obtained when the loop is included. In ethanol/water simulations, the r.m.s.d. of thermolysin is slightly lower when the loop is not included but remains higher than the one obtained in water.

Additionally, visual analysis led us to suspect that there were rigid body motions between the protein domains (especially in the case of TLN). This is in accordance with what has been previously observed in experimental and simulation studies[235-237], where it was found that there is a hinge-bending motion between the N-terminal and C-terminal domains of thermolysin. The structure obtained by Hausrath et al.[236] revealed that in the absence of a substrate, the enzyme adopts an open conformation, which is not observed when the enzyme-ligand complex is formed. Although the TLN structure that was used in this study was obtained in the absence of inhibitors, it was subsequently found to have electron density in the active site that probably corresponds to a bound dipeptide which could be the result of autolysis during protein purification or

crystallization[237], suggesting that the closed conformation of the enzyme found in this structure is induced by the presence of the dipeptide, explaining why the protein opens during the simulations (where the peptide was not included). In order to investigate if there are rigid body motions between the protein domains, we calculated the rmsd of each domain separately (the results are shown and discussed in more detail in Appendix B). This analysis shows that the rmsd values obtained for each separate domain of thermolysin are lower than the global rmsd, both in water and in ethanol/water simulations, which indicates that there are in fact rigid motions. In the case of pseudolysin, we only found significant interdomain movements in two replicates of the C58G mutant in ethanol/water.

Although the rmsd is a good measure of the degree of conservation of a structure, it is still limited, and other analysis, such as the radius of gyration and secondary structure content, can bring further insight. As can be observed in fig. 4.3, native PSL has approximately the same radius of gyration in pure water as in ethanol/water simulations (the average values in the last 100 ns of simulation are 1.99 and 2.02 nm for water and ethanol/water, respectively), whereas thermolysin is considerably more open in the ethanol/water mixture than in pure water (the average values in the last 100 ns of simulation are 2.14 and 2.35 nm for water and ethanol/water, respectively). The C58G mutant of pseudolysin is considerably less compact (average radius of gyration in the last 100 ns ≈ 2.10 nm) than the wild-type enzyme in our ethanol/water simulations. The loss of compactness of TLN and the PSL mutant, in the ethanol/water mixture, indicates that these enzymes are unfolding. These results are consistent with our rmsd analysis and the previous experimental findings[4].

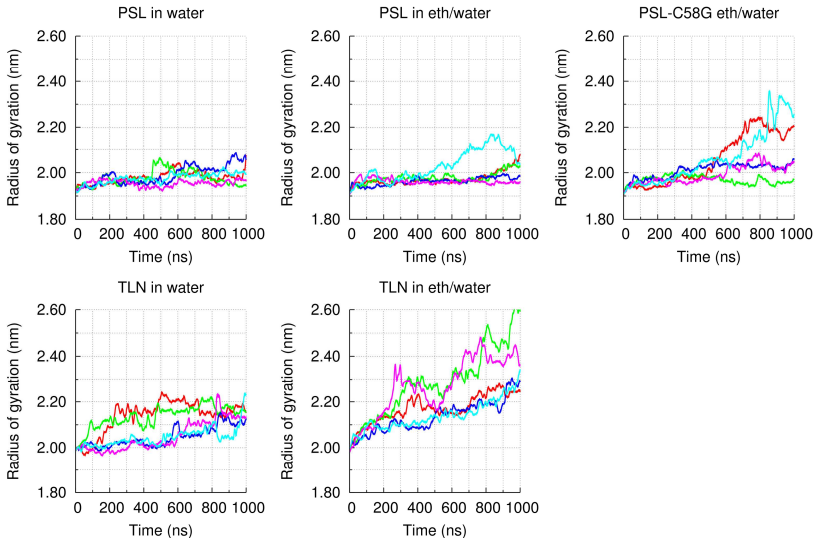


Figure **4.3.** Moving average of the radius of gyration. The lines with different colors represent different replicates, as in fig. 4.2.

The analysis of the secondary structure content of the proteins studied, in the two media used (fig. 4.4), indicates that the loss of secondary structure by pseudolysin is slightly more pronounced in water (average loss of secondary structure content ≈ 11%) than in ethanol/water simulations (average loss of secondary structure content ≈ 10%). Although our rmsd and radius of gyration analysis indicates that the mutant of pseudolysin is considerably more unstable than the wild-type enzyme in the ethanol/water mixture, this does not correspond to a clear difference in what concerns the loss of secondary structure, except in replicate 1 of the mutant, where there is a greater loss of secondary structure (note that this replicate is the one that has a higher rmsd – see fig. 4.2 and fig. B1 (available in Appendix B)). There are two possible explanations for this fact: either the disulfide bridge plays a role in maintaining

the enzyme tertiary structure but not its secondary structure or our sampling is not sufficient to distinguish between the two forms of the enzyme. The latter hypothesis is supported by the fact that in one of the mutant replicates we do observe a marked loss of secondary structure.

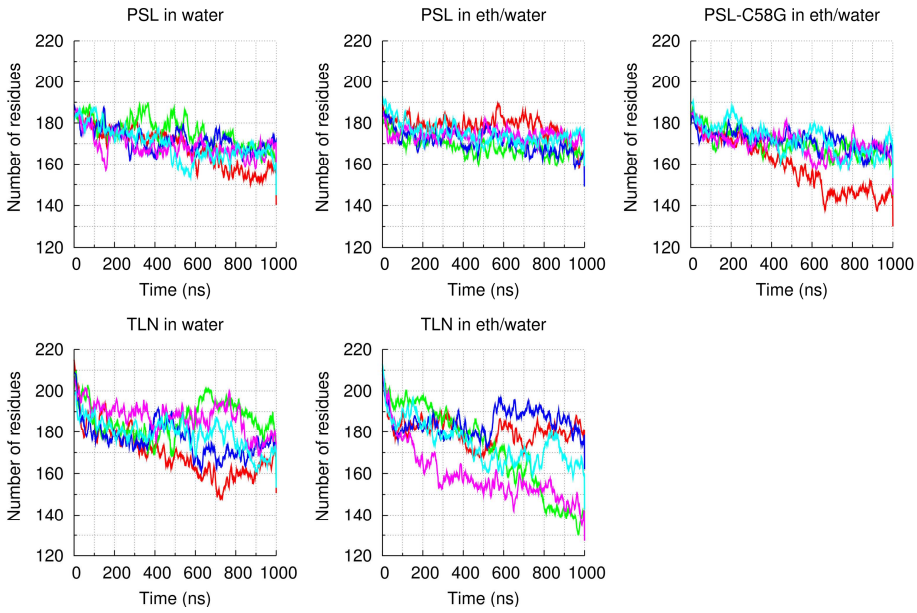


Figure 4.4. Moving average of the total secondary structure content, computed as the sum of the number of residues that are part of α-helices, β-sheets, β-bridges or turns, according to DSSP criterion[200]. The lines with different colors represent different replicates, as in fig. 4.2.

Thermolysin suffers a higher loss of secondary structure content in ethanol/water (average loss of secondary structure content ≈ 21%) than in pure water simulations (average loss of secondary structure content ≈ 16%), which is consistent with the results discussed above. Our secondary structure analysis is in line with a previous study, in which the authors determined the CD spectra of

PSL and TLN in the presence and in the absence of methanol, and found that the secondary structure of pseudolysin was more conserved in the presence of methanol, whereas the opposite was observed for thermolysin[238].

4.4.2 Protein-ethanol interaction

With the aim of understanding the molecular determinants underlying the stability of PSL (wild-type and C58G mutant) and TLN, in media containing ethanol, we analyzed the contact area between the enzymes and the alcohol. Figure 4.5 shows that the protein-ethanol contact area reaches higher values at the end of the simulation in the C58G mutant of pseudolysin than in the wild-type (157 vs 147 nm^2 for the mutant and wild-type, respectively), and in the mutant, it is still increasing after 1 μs of simulation, whereas the native pseudolysin-ethanol contact area appears to reach a plateau after the first 200 ns of simulation. These results indicate that the mutant PSL has a greater tendency to interact with ethanol than the wild-type enzyme. As can be seen in fig. 4.5, thermolysin has a strong propensity to interact with ethanol molecules, given that the contact area between this protein and ethanol increases sharply in the first 500 ns of simulation (reaching a value of 159.51 nm2 in the last 100 ns of simulation).

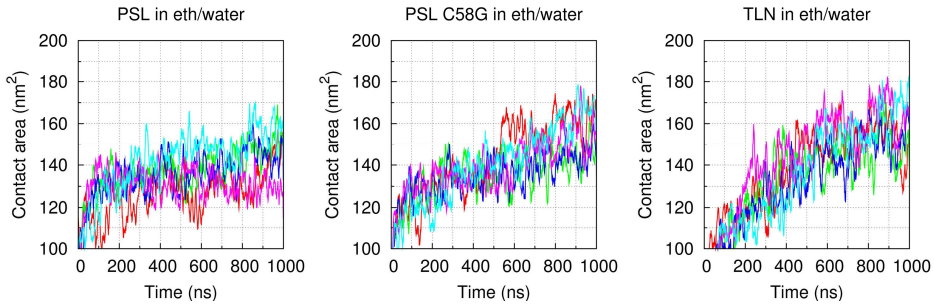


Figure 4.5. Moving average of the contact area between ethanol molecules and the protein. The protein-ethanol contact area is given by the following expression: $CA_{prot\text{-}eth} = SAS_{prot} + SAS_{eth} - SAS_{prot+eth}$, where $CA_{prot\text{-}eth}$ is the protein-ethanol contact area, SAS_{prot} is the solvent accessible surface of the protein, SAS_{eth} is the solvent accessible surface of ethanol, and $SAS_{prot+eth}$ is the solvent accessible surface of the protein-ethanol system. The lines with different colors represent different replicates, as in fig. 4.2.

As a control we also measured the contact area between the protein and the water. Figure B2 (see Appendix B), shows that the wild-type and mutant pseudolysin have similar contact areas with water (average values of 190.16 and 191.03 nm^2 for the wild-type and mutant, respectively). The contact area between thermolysin and water is larger than in the case of pseudolysin (average value of 215.59 nm^2), which is not surprising, given that TLN has a larger solvent accessible surface than PSL.

To complete our analysis of the interaction between the proteins under study and ethanol, we investigated if this interaction is mainly polar or hydrophobic. The distribution of ethanol molecules around the protein surface is displayed in fig. 4.6. In this plot, there are two clearly distinct peaks, the first one is centered at ~0.2 nm and corresponds to ethanol molecules that form hydrogen bonds with the protein, and the second one is centered at ~0.35 nm and

corresponds to van der Waals interactions between ethanol molecules and the protein. The area of the second peak is larger than the area of the first peak for the three proteins analyzed (see table B1 in Appendix B), meaning that the majority of the interactions between the protein and ethanol are van der Waals interactions. Comparing the areas of the peaks of the three proteins (table B1 in Appendix B), we can see that thermolysin has a higher number of interactions (which are mainly hydrophobic) with ethanol than pseudolysin. The peaks of the mutant pseudolysin are slightly larger than the peaks of the wild-type enzyme (especially the second peak), meaning that the mutant has a higher number of hydrophobic interactions with ethanol, which is in agreement with the results obtained in the analysis of the protein-ethanol contact surface (see above). As a control, we analyzed the distribution of water molecules around the proteins. Fig. B3 and table B2 (available in Appendix B) show that the distribution of water molecules is very similar for the three proteins, in the simulations performed in ethanol/water mixtures.

To further elucidate the nature of the interactions between the protein and the ethanol, we analyzed the distributions of the alcohol (OH) and alkyl (CH_2CH_3) moieties of the molecule, separately. In fig. B4A, Appendix B, we can see that the OH group interacts with the protein through hydrogen bonding (first peak) and van der Waals interactions (second peak). The areas of the two peaks (table B3, Appendix B) indicate that the former is the predominant type of interaction for all the proteins analyzed. Both moieties of the ethanol molecule have a peak centered at ~0.35 nm that corresponds to van der Waals interactions with the protein. This means that both moieties contribute to the second peak observed

in fig. 4.6, although the contribution of alkyl moiety is larger than that of the alcohol moiety.

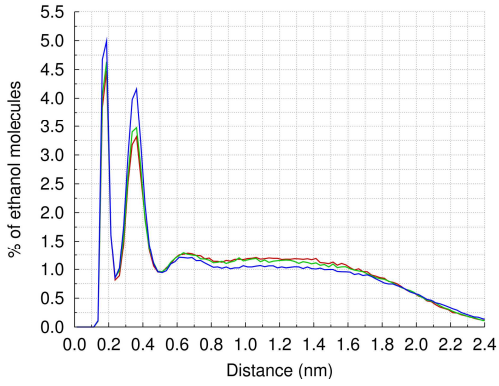


Figure 4.6. Distribution of the ethanol molecules around the protein in the last 100 ns of the simulations performed in the ethanol/water mixture. The red, green, and blue lines correspond to wt PSL, mutant C58G of PSL, and TLN, respectively.

The observation that thermolysin tends to form hydrophobic interactions with alcohol molecules is in agreement with a previous crystallographic analysis, in which thermolysin crystals were soaked with isopropanol. In that study, 12 different binding sites for isopropanol were identified, and most of these binding sites were located in hydrophobic pockets[174]. Interestingly, 10 of the twelve isopropanol crystallographic binding sites correspond to residues which interact very frequently with ethanol molecules in our simulations (see fig. B5, Appendix B). The remaining two binding sites found in the X-ray structure are formed by more than one molecule in the crystal lattice, and therefore it is not surprising that they are not very populated during the simulations. The agreement between our results and the experimental findings is a very good indicator that our observations are realistic. Our results are also in line with a biochemical study

where it was observed that more hydrophobic solvents cause a more severe decrease in the thermal stability of thermolysin than more hydrophilic ones[234], which indicates that the solvent destabilizes the enzyme because it can bind to hydrophobic pockets and distort its tertiary structure. Additionally, another study has shown that there is a negative correlation between the polarity index of a solvent and its power to irreversibly denature enzymes[239], which is most likely due to the fact that apolar solvents can bind to hydrophobic regions of the proteins and lead to their irreversible unfolding.

4.4.3 Comparing the behavior of wild type and C58G mutant of pseudolysin

One of the aims of this work is to elucidate the role played by the disulfide bridge between C30 and C58 in maintaining the stability of pseudolysin in ethanol/water mixtures. Towards this end, we compared the behavior of the wild-type and the C58G mutant of pseudolysin in the simulations performed in the ethanol/water mixture. In order to analyze the major conformational changes that occur during the simulations, we divided each trajectory in 10 ns windows and calculated the average structure in each window. The results are displayed in Movie_B1 and Movie_B2 (see Supporting Information). These movies show that, in agreement with the results discussed above, the mutant enzyme suffers larger conformational changes than the wild-type. Focusing on the region where the residues C30 and (C/G)58 are located (highlighted in magenta), we can see that these loops undergo larger conformational changes in the mutant than in the wild type simulations, indicating that the disulfide bridge constrains the motion of these loops. In the absence of this bond, the loops are free to

move and consequently become more unstable. Nevertheless, it is clear from Movie_B2 (see Supporting Information), that the large conformational changes of the C58G mutant of pseudolysin are not restricted to these loops. One example of this is the behavior of replicate 1, where we can observe large conformational changes in the C-terminal domain (located in the opposite side of the protein). Analyzing the distribution of ethanol around the enzyme in this simulation (see Movie_B3 in Supporting Information), we can see that after about 500 ns of simulation, ethanol starts to accumulate in the interior of the protein, surrounding the α-helices located in the C-terminal domain, and at the same time, this region of the enzyme starts to unfold until it gets completely destroyed. These observations indicate that in the absence of the disulfide bridge between the loops, they become more flexible. It is possible that the conformational changes of the loops are then propagated to other regions of the enzyme, which can become more open and, therefore, more permeable to ethanol molecules. The accumulation of ethanol in these regions will substitute essential intra-protein hydrophobic interactions, leading to unfolding.

4.5 Conclusion

Using a MD simulation approach, we were able to obtain a molecular picture that explains the experimentally observed difference in stability of pseudolysin and thermolysin in ethanol/water solutions. In accordance with the previous experimental findings[4], pseudolysin is more stable than thermolysin in the simulations performed in ethanol/water media. The analysis of the interaction between the proteins and ethanol showed that the contact surface between

thermolysin and the alcohol is larger than that of pseudolysin. Our results also indicate that the nature of the interaction between the proteins and ethanol is mainly hydrophobic, and therefore, the alcohol molecules that reach the interior of thermolysin will replace the native intra-protein hydrophobic interactions, leading to the unfolding of the enzyme.

We also found that, in agreement with site-directed mutagenesis experiments[5], the absence of the C30-C58 disulfide bond makes pseudolysin more unstable. The investigation of the protein-ethanol interaction showed that the mutant C58G has a larger protein-ethanol contact surface than the wild-type enzyme. Our results indicate that the disulfide bridge constrains the motion of the loops that it connects. In the mutant (which lacks this bridge) the loops undergo higher conformational changes than in the wild-type. We think that these conformational changes can propagate to other regions of the enzyme, causing it to open and enabling ethanol molecules to penetrate. Analogously to what happens in the case of thermolysin, these ethanol molecules can disrupt essential intra-protein interactions, which explains the low stability of the mutant in ethanol/water mixtures.

The results presented here are in good agreement with several experimental studies, which shows that simulation studies can mimic what is observed experimentally concerning the stability of enzymes in solutions containing organic solvents. Additionally, this study complements the previous experimental works by providing a molecular explanation for their observations and may be used in the prediction and engineering of optimized enzymes for this type of media.

4.6 Acknowledgements

The authors acknowledge Dr. Nuno Micaêlo for helpful discussions and the financial support from Fundação para a Ciência e a Tecnologia, Portugal, through grants SFRH/BD/28269/2006, POCTI/BIO/57193/2004 and PEst-OE/EQB/LA0004/2011.

Chapter 5

Structural determinants of ligand imprinting: A molecular dynamics simulation study of subtilisin in aqueous and apolar solvents

This work has been published in the following paper:

Lousa D., Baptista A. M., Soares C. M. **(2011)**, Structural determinants of ligand imprinting: A molecular dynamics simulation study of subtilisin in aqueous and apolar solvents, *Protein Science*, vol. 20, pp 379-386

(doi: http://dx.doi.org/10.1002/pro.569)

Contributions of the author of the present thesis to this work:

The author of the present thesis has participated in the design of all the *in silico* experiments presented in this chapter and executed all the simulations and analysis described herein.

5.1 Abstract

The phenomenon known as “ligand imprinting” or “ligand-induced enzyme memory” was first reported in 1988, when Russell and Klibanov observed that lyophilizing subtilisin in the presence of competitive inhibitors (that were subsequently removed) could significantly enhance its activity in an apolar solvent.[6] They further observed that this enhancement did not occur when similar assays were carried out in water. Herein, we shed light on the molecular determinants of ligand imprinting using a molecular dynamics (MD) approach. To simulate the effect of placing an enzyme in the presence of a ligand before its lyophilization, an inhibitor was docked in the active site of subtilisin and 20 ns MD simulations in water were performed. The ligand was then removed and the resulting structure was used for subsequent MD runs using hexane and water as solvents. As a control, the same simulation setup was applied using the structure of subtilisin in the absence of the inhibitor. We observed that the ligand maintains the active site in an open conformation and that this configuration is retained after the removal of the inhibitor, when the simulations are carried out in hexane. In agreement with experimental findings, the structural configuration induced by the ligand is lost when the simulations take place in water. Our analysis of fluctuations indicates that this behavior is a result of the decreased flexibility displayed by enzymes in an apolar solvent, relatively to the aqueous situation.

5.2 Introduction

Enzymatic catalysis in anhydrous solvents has attracted the attention of biotechnologists and biochemists for more than two decades. In nonaqueous media, enzymes can display several novel and valuable properties,[30] such as the capacity to catalyze reactions that are not feasible in water,[28] interfacial activation,[137, 240] increased thermostability,[241] and an altered substrate specificity as well as enantiomeric selectivity.[71, 76, 79, 210] It is now clear that reactions in nonaqueous media are strongly dependent on the water content of the solvent. The amount of water can influence enzymatic properties like activity,[65] structure,[1, 74, 139] dynamics [1, 74, 139] and enantioselectivity [71, 74], and can thus be used to control the catalytic process. Despite their great potential, reactions in nonaqueous solvents are often limited by a drastic reduction in enzyme activity when compared with their aqueous counterparts.[55] This raises an obvious question: How can the activity of enzymes in organic solvents be enhanced?

In 1988, Russell and Klibanov observed that the enzymatic activity of the serine protease subtilisin, in anhydrous *n*-octane, could be enhanced by previously lyophilizing the enzyme in the presence of competitive inhibitors.[6] In their study, the ability of five different inhibitors to enhance the rate of transesterification reactions was tested. The authors reported an increase of up to ~100 fold in enzyme activity relatively to the enzyme lyophilized in the absence of inhibitors. This was the first description of a curious phenomenon known as "ligand-induced enzyme memory" or "ligand imprinting". Interestingly, when the same assays were carried out in water, there was no difference between the enzyme preparations lyophilized in the presence and in the absence of inhibitors, indicating that the enzyme looses its "memory" in water. Moreover, the authors

found a clear correlation between the percentage of water retained in the organic solvent and the observed rate enhancement: the larger the water content, the smaller the rate enhancement. In an attempt to explain this behavior, they speculated that the competitive inhibitor causes a conformational change in the enzyme that is retained in anhydrous apolar solvents, even after the removal of the ligand, because the enzyme is rigid in the absence of water and thus it gets kinetically trapped in the conformation induced by the inhibitor: the enzyme behaves as if it has a "memory". As the water content increases, the protein becomes more flexible and rapidly "forgets the ligand imprinted state".

In another study, Stähl et al showed that the substrate specificity and seteroselectivity of α-chymotrypsin in anhydrous organic media could be tuned by using an enzyme preparation obtained by precipitation with different inhibitors.[112] These results show that the activation increases as the similarity between the substrate and the inhibitor used for "imprinting" increases, indicating that the effect is very specific and located in the active site.

The application of molecular imprinting has been extended by Rich and Dordick to the activation of subtilisin-catalyzed acylation of nucleosides. [242] The authors complemented their experimental findings with a molecular dynamics study and concluded that the activation of enzymes by imprinting is caused by structural changes of the catalytic triad.

The molecular determinants of the observations reported above remain unclear. In the present work, we addressed this question by mimicking the effect of lyophilizing subtilisin in the presence and in the absence of an inhibitor and then performing MD simulations using the resulting structures, both in hexane and in

water. Our results indicate that the inhibitor induces an open conformation of the S1 pocket that is maintained after the removal of the ligand in anhydrous, but not in aqueous, simulations. Our analysis of fluctuations suggests that this behavior is caused by the decreased flexibility exhibited by subtilisin in hexane.

5.3 Materials and methods

5.3.1 Protein structure selection

The structure of Subtilisin Carlsberg covalently bound to the inhibitor L-[(1*R*)-1-acetamido-2-(1-naphthyl)ethyl]boronic acid, refined at 2.65 Å (PDB code: 1AV7 [243]) was used in our studies. This structure was selected because it contains an inhibitor that structurally resembles the ligand that we intended to dock and thus we expected the active site to be in a proper configuration to accommodate the ligand of interest.

The inhibitor L-[(1*R*)-1-acetamido-2-(1-naphthyl)ethyl]boronic acid was withdrawn from the structure before the subsequent steps of this study.

5.3.2 Determination of protonation states

The determination of the protonation state of each titrable site in the protein was performed using a methodology developed by us, based on continuum electrostatics and Monte Carlo sampling of protonation states, that has been explained in detail before.[171, 172] Only water molecules with a relative accessibility inferior or equal to 0.5 were included in the calculations of the protonation equilibrium. The electrostatic energy terms were calculated by

solving the Poisson-Boltzmann equation, using the MEAD package.[207, 244] The program PETIT,[172] that implements a Monte Carlo procedure, was used to sample the protonation states at different values of pH, using the energy terms calculated by MEAD.

5.3.3 Docking of the inhibitor

The inhibitor *N*-acetyl-L-tryptophan amide was docked in the active site of subtilisin, using the software AutoDock, version 3.0.[165] This ligand was chosen because it displayed the largest rate enhancement in Russell and Klibanov's experiments.[6] The inhibitor structure was built using PyMOL.[208] All waters were removed from the structure of subtilisin. Kollman united-atom partial charges were assigned to the protein and the ligand. Only polar hydrogens were considered. Solvation parameters and fragmental volumes were assigned using AutoDockTools

(http://AutoDock.scripps.edu/resources/adt/index_html). This tool was also employed to determine the ligand's rotatable bonds. The program AutoGrid was used to define grid maps of 70 × 70 × 60 points, in the *x*, *y* and *z* directions, respectively, with a 0.375 Å spacing and centered at the active site. The docking was performed using the Lamarckian genetic algorithm, with a population of 300 random individuals, a maximum number of 2.5×10^6 energy evaluations, a maximum number of 27000 generations, an elitism value of 1, a mutation rate of 0.02 and a crossover rate of 0.80. The pseudo-Solis and Wets method was used for local search, having a maximum of 300 iterations per search and a probability of performing local search on an individual in the population of 0.06; the

maximum number of consecutive successes or failures before doubling or halving the local search step size was 4 and the local search was terminated when the local search step size reached a value equal or lower than 0.01. Five hundred docking runs were performed and the results were processed using cluster analysis with a root mean square deviation (rmsd) tolerance of 1.0 Å. The best docking solution was selected and used as a starting point for the MD simulations.

5.3.4 Molecular dynamics simulations

The general methodology used in the molecular dynamics simulations of proteins in nonaqueous media was developed by our group and is explained in detail elsewhere.[1] MD simulations were performed with the GROMACS package,[154] version 4.0,[191] using the 53A6 force field. [151] Water was modeled with the simple point charge (SPC) model[192] and hexane was treated as a flexible united atom model using the 53A6 alkane parameters.[245] Bond lengths of the solute and hexane molecules were constrained with LINCS[159] and SETTLE[195] was used for water. The simulations were performed at constant temperature and pressure. For the simulations carried out in hexane, the protein, ions and water were coupled to the same heat bath and hexane was coupled to a separate heat bath. For the aqueous simulations, the protein and water were coupled to two separate heat baths. Temperature coupling was implemented using the Berendsen thermostat[161] with a temperature coupling constant of 0.1 ps and a reference temperature of 300 K. The pressure control was done by applying the Berendsen algorithm with an isotropic pressure coupling, using a reference pressure of 1 atm and a relaxation time of 0.5 ps and 1.5 ps for water and

hexane simulations, respectively. An isothermal compressibility of 4.5×10^{-5} bar^{-1} was used both for water and hexane. Nonbonded interactions were calculated using a twin-range method with short and long range cutoffs of 8 Å and 14 Å, respectively.[196] In water simulations, a reaction field correction for electrostatic interactions was applied,[153, 197] considering a dielectric constant of 54 for water (the dielectric constant of SPC water).[198]

The necessity of using multiple replicates, in molecular dynamics simulations, has been highlighted in previous studies conducted in our group.[1, 246] It was clear in both works that a unique simulation does not capture the characteristics of the ensemble that ideally one whishes to sample. This reflects the fact that protein molecules have very complex conformational energy landscapes, with multiple minima where the system may become trapped during the simulation. To circumvent these sampling difficulties, in this study we have used several replicates, as indicated in fig. 5.1.

5.3.5 Hydration conditions in hexane simulations

In their experiments, Russell and Klibanov found that the larger the water content the smaller the ligand imprinting effect. This is most likely due to the fact that there is a positive correlation between the amount of water present in an apolar organic solvent and protein flexibility.[1, 62, 64, 126, 247] Our aim was to test the two extreme cases: anhydrous vs. aqueous conditions. Completely anhydrous conditions (0% water) are very rarely found and removing all the waters from the protein could be drastic to its stability. It has been shown that in apolar media like hexane, the water molecules are in close contact with the protein and for

low water percentages, the amount of water located beyond 0.25 nm away from the enzyme surface is negligible.[2] Thus, in our hexane simulations, we decided to keep only water molecules with a relative accessibility inferior or equal to 0.1.

5.3.6 Selection of counterion positions

The selection of counterion positions was done using an approach based on docking simulations, similar to the one applied before by us.[1] A detailed description of this methodology can be found in Appendix C, in the section *1. Protocol for selecting counterion positions.*

5.4 Results and discussion

The hypothesis analyzed in this study is that a ligand in complex with an enzyme induces conformational changes in the active site that can be maintained once the ligand is removed and the protein is immersed in an anhydrous apolar solvent. On the other hand, if the protein is immersed in water, its conformation rapidly deviates from the ligand-induced one.

To test this hypothesis, we used the strategy summarized in the fluxogram represented in figure 5.1. As the fluxogram shows, we performed two distinct sets of simulations, the first set will be referred to as "ligand-treated" simulations and the second set will be called "ligand-untreated" simulations.

In the ligand-treated simulations, we started by docking an inhibitor in the active site of subtilisin. We then placed the enzyme-ligand complex in water and performed 30 independent MD simulations of 20 ns each. The purpose of these

simulations was to adapt the active site to the ligand. In 16 out of the 30 simulations carried out the ligand remained in the catalytic pocket and the final structures of these 16 simulations were used in the subsequent steps of the methodology. The next step consisted in the removal of the inhibitor. Finally, we conducted 10 ns of MD in *n*-hexane (which is similar to *n*-octane) and in water, using as a starting point the conformations obtained in the previous step.

As a control, we performed a set of 16 ligand-untreated simulations, in which we began by running 20 ns of MD simulations in water, starting from the x-ray structure of subtilisin. We then used the final conformations of these simulations to carry out 10 ns of MD in hexane and in water.

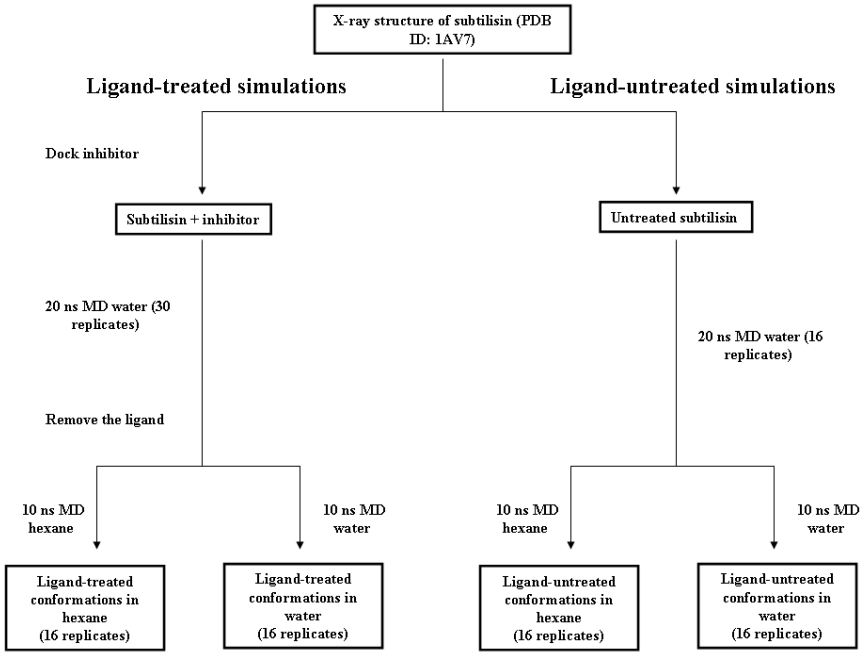


Figure 5.1. Overview of the simulation methodology.

5.4.1 Docking of the inhibitor

Given that the S1 pocket, which is the specificity subsite in serine proteases, is known to accommodate hydrophobic substrates [243] (as it is the case for our inhibitor), we restricted our docking searches to the area surrounding this site. The best docking solution found is displayed in figure 5.2.

This solution was reached in 116 out of 500 runs and has the lowest docked energy of all the solutions found. As can be seen in the figure, the ligand is docked very near the catalytic triad, preventing substrates from binding. Therefore, this docking position is compatible with the competitive character of the inhibition of subtilisin by *N*-acetyl-L-tryptophan amide. These results, led us to choose the enzyme-ligand complex shown in figure 5.2 as the starting point for all the subsequent steps of this study.

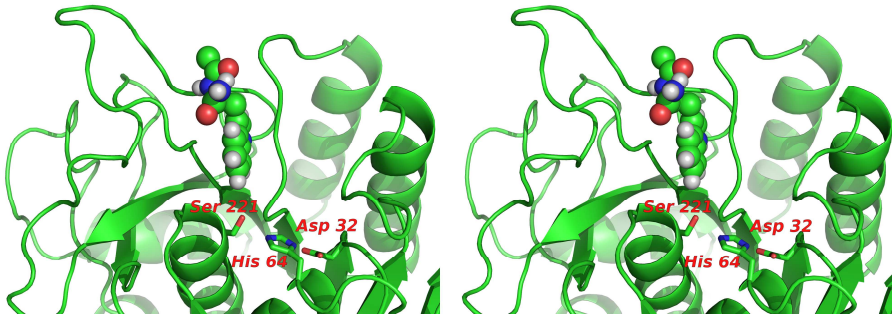


Figure 5.2. Stereo view of the best docking solution of the inhibitor *N*-acetyl-L-tryptophan amide in the S1 site of subtilisin. The atoms of the ligand are represented using spheres. The residues that compose catalytic triad are represented with sticks and labeled.

5.4.2 Stability of the simulations

The temporal evolution of the root mean square deviation (rmsd) from the x-ray structure provides information on the stability of a simulation. Figure C5.2 (in Appendix C) shows the rmsd from the crystal structure, for the final 10 ns of all the systems studied (see fig. 5.1). The plots show that in general the simulations carried out in hexane (figs. C5.2A and C5.2B, Appendix C) stabilized slower than the simulations that were performed in water (figs. C5.2C and C5.2D, Appendix C). This observation can be explained by the fact that the starting points of these simulations are the final conformations obtained after 20 ns of simulations in water (see fig. 5.1). In the case of hexane simulations, the protein is placed in a new medium and the system has to reach a new equilibrium state. On the other hand, in the simulations carried out in water, the protein is kept in the same environment and there is no adaptation phase. Looking at the plots in figure C5.2 (Appendix C), we can also observe that in hexane simulations the protein deviates more from the crystal structure than in water simulations, which probably reflects the fact that hexane is an unnatural medium for proteins, that leads them to adopt a different conformational arrangement. In the great majority of the simulations, the rmsd stabilizes before 5 ns of simulation time; thus we considered that the MD simulations were equilibrated after that period.

5.4.3 Effect of pretreating the enzyme with the ligand: hexane vs. water simulations

As mentioned above, the ligand was docked in the S1 pocket. It is therefore relevant to analyze the behavior of this pocket after the removal of the ligand

and compare this behavior with the ligand-untreated case. From the analysis of our simulations we observed that the S1 pocket can adopt three different states, that we named “closed”, “intermediate” and “open”. These states are illustrated in figure 5.3.

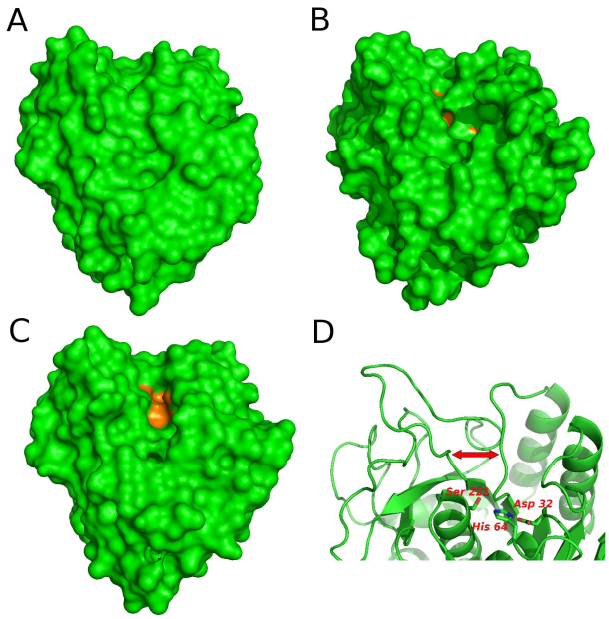


Figure 5.3. Illustration of the three distinct states adopted by the S1 pocket. **A.** Example of a closed conformation (last configuration of replicate number 3 of the ligand-untreated simulations in hexane). **B.** Example of an intermediate conformation (last configuration of replicate number 4 of the ligand-untreated simulations in water). **C.** Example of an open conformation (last configuration of replicate 3 of the ligand-treated simulations in hexane).**D.** The minimum distance between the two loops (represented by a red arrow) can be used to analyze the state of the pocket. This measurement corresponds to the minimum distance between all the atoms of residues 125 to 127 and residues 153 to 155.

To have a quantitative measurement that could capture the state of the S1 pocket during the simulations, we calculated the minimum distance between the two loops surrounding the pocket (see the illustration in figure 5.3D). The histograms in figure 5.4 represent the distributions of these distances in the last 5 ns of simulation (when the simulations were considered equilibrated). In figure 5.4A we can see that the distributions of the ligand-treated and the ligand-untreated simulations in hexane are clearly distinct, which indicates that the inhibitor has an effect on the structural arrangement of the S1 pocket. The distribution of the ligand-treated simulations is more extended and there is a considerably larger number of conformations with an open pocket and a smaller number of conformations presenting a closed or intermediate pocket.

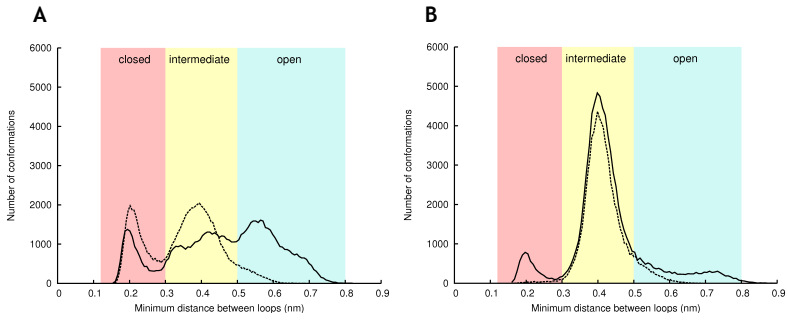


Figure 5.4. Distributions of the minimum distance between the loops surrounding the S1 pocket during the last 5 ns of simulation. The three distinct areas highlighted in the plots, in different tones of grey, correspond to three different sates of the pocket (closed, intermediate and open, respectively). **A.** Ligand-treated (solid line) and ligand-untreated (dashed line) simulations in hexane. **B.** Ligand-treated (solid line) and ligand-untreated (dashed line) simulations in water.

The plots provided in fig. C5.3A (Appendix C) indicate that the inhibitor induces an open state of the S1 pocket that in many cases is retained when the enzyme is placed in hexane after the removal of the ligand. This behavior is illustrated in Movie_C1 (Supplementary Information).

When the enzyme has no contact with the ligand, the pocket tends to have a more closed configuration that is maintained or even accentuated when the enzyme is placed in hexane (fig. C5.3B, Appendix C). As a consequence, pretreating the enzyme with a competitive inhibitor increases the probability of finding an open S1 pocket.

The broad distribution exhibited by ligand-treated simulations (fig. 5.4A) can be explained by the fact that during the accommodation step, when the enzyme-ligand complex is simulated in water, the ligand adopts several distinct conformations, which influence the structure of the active site. Therefore, at the beginning of the ligand treated simulations, each replicate displays a different conformation of the active site, which tends to be retained in hexane, giving rise to the extend distribution shown in the plot.

In opposition to what was described for hexane, in water, the distributions of the ligand-treated and ligand-untreated simulations are very similar (fig. 5.4B). The plots in figure C5.3C (Appendix C) show that when the enzyme is kept in water after the removal of the ligand, the S1 pocket tends to deviate from the open state induced by the ligand and reach an intermediate state, in which the loops are separated by a distance between 0.4 and 0.5 nm. The behavior of the S1 pocket in the ligand-treated simulations in water is illustrated in the Movie_C2 (Supporting Information).

When the enzyme is placed in water with no a priory contact with the inhibitor, the pocket remains in the same intermediate conformation (fig. C5.3D, Appendix C). The homogeneity between the ligand-treated and control simulations indicates that in water the ligand has no imprinting effect.

These observations give an insight into the molecular determinants that are on the basis of the experimental findings made by Russell and Klibanov.[6] Our results indicate that, in hexane, the active site of subtilisin tends to be more open when the enzyme is pretreated with a competitive inhibitor. It is easier for a substrate to bind to an open active site and this explains the fact that the enzymatic activity, towards substrates that are similar to the inhibitor, increases when the enzyme is lyophilized from a solution containing a competitive inhibitor. In water, when subtilisin is pretreated with an inhibitor and then washed, the pocket tends to rapidly deviate from the conformation induced by the ligand and adopt a configuration that is similar to the one found in enzyme molecules that had no contact with the ligand. This accounts for the lack of rate enhancement observed when the reactions were performed in water, after placing the enzyme in the presence of a competitive inhibitor and then removing the ligand.

5.4.4 Why does ligand imprinting occur in hexane but not in water?

As it was discussed above, we observed that pretreating subtilisin with an inhibitor has an effect on the conformation of the S1 pocket in hexane but not in water. What are the physical determinants of these observations? It is generally accepted that enzymes are less flexible in apolar anhydrous solvents than in

water and therefore they may get kinetically trapped in metastable states.[1, 62, 64, 126, 247] In order to analyze the protein mobility in water and in hexane, we measured its root mean square fluctuations (rmsf).

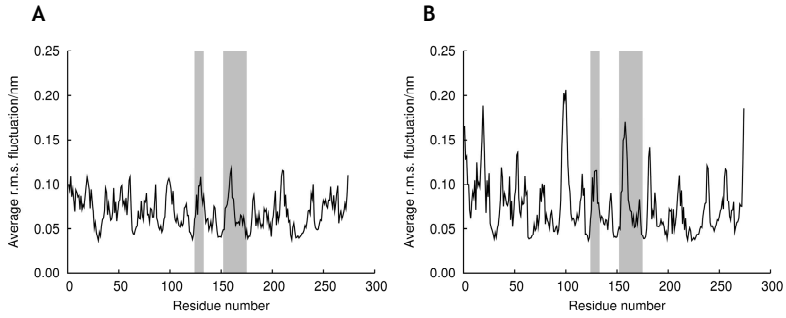


Figure 5.5. Average root mean square fluctuation during the last 5 ns of the simulations in hexane (A) and water (B). The value was obtained by averaging the rmsf per residue of the simulations in the corresponding solvent. The residues that correspond to the loops surrounding the S1 pocket are highlighted in grey.

Looking at the plots in figure 5.5, we can see that the rmsf values in hexane simulations are considerably lower than in water simulations. The most significant difference in average rmsf corresponds to the loop formed by residues 94 to 102 which is located near the active site. These observations indicate that, in accordance with what has been observed previously, subtilisin is more flexible in water than in anhydrous hexane. In particular, the loops that surround the S1 pocket (highlighted in grey in fig. 5.5) have a higher mobility in water than in hexane simulations.

In the light of these results it is reasonable to think that, due to this reduced flexibility, in hexane the S1 pocket tends to retain the conformation induced by the competitive inhibitor. This facilitates the subsequent binding of the substrate. In water the enzyme is mobile and does not retain the configuration induced by the ligand, therefore ligand imprinting is not observed.

5.5 Conclusion

This work sheds light on the molecular determinants underlying the phenomenon known as ligand imprinting. Our simulation results indicate that the inhibitor *N*-acetyl-L-tryptophan amide induces an open conformation in the active site of subtilisin. We observed that in hexane simulations the active site remained open even after the removal of the ligand. When the same assays were carried out in water, the enzyme showed a very different behavior. In this case, the structure of the S1 pocket in the ligand-treated simulations was almost indistinguishable from its structure in the ligand-untreated simulations.

Our rmsf analysis supports the hypothesis that the different behavior observed in the two solvents reflects differences in protein flexibility. Enzymes in water are highly mobile and therefore rapidly "loose memory" of the ligand-induced state. This accounts for the fact that no activation is observed in reactions that take place in water when the enzyme is lyophilized in the presence of competitive inhibitors and then washed. In anhydrous apolar solvents subtilisin is rigid and therefore more likely to get "locked" in the ligand-imprinted conformation. When the reaction is carried out in an anhydrous apolar solvent, after lyophilizing the enzyme from a solution containing a competitive inhibitor, there

is a higher probability that the reacting substrate will find an open S1 pocket, which would then explain the rate enhancement observed in *n*-octane [6].

5.6 Acknowledgements

The authors acknowledge Dr. Nuno Micaêlo for helpful discussions and the financial support from Fundação para a Ciência e a Tecnologia, Portugal, through grants SFRH/BD/28269/2006 and POCTI/BIO/57193/2004.

Chapter 6

Final discussion

The technological potential of nonaqueous enzymology has been recognized for more than thirty years. In order to take full advantage of this potential, numerous studies have tried to elucidate the unusual properties of enzymes in nonaqueous media. Most of the pioneering studies analysed the influence of the solvent and other reaction conditions on the enzyme macroscopic properties. Yet, a deep understanding of nonaqueous enzymology requires a microscopic analysis of the molecular determinants underlying the observed behaviour.

In the last decade, several experimental and computational studies have focused on a molecular-level analysis and currently many properties displayed by enzymes in nonaqueous solvents are well understood, both from macroscopic and microscopic viewpoints (see, e.g., refs. 30, 31). Nevertheless, at the beginning of the PhD work described in this thesis, some aspects of this field were poorly characterized at the molecular level. The main objective of the current thesis is to contribute to fill this gap. Towards this end, molecular simulations were used to address some of the unsolved questions of biocatalysis in nonaqueous solvents. We focused on three different subjects, which were lacking a more detailed molecular characterization: protein-ion interactions, enzyme stability in aqueous/nonaqueous mixtures, and molecular memory.

6.1 Protein-ion interactions in nonaqueous solvents

In apolar or moderately polar environments, ions are expected to form strong associations with protein charged or polar groups, playing an important role in the catalytic process. Thus, a very important issue of nonaqueous enzymology is the molecular characterization of protein-ion interactions.

A thorough molecular characterization of protein-ion interactions in solution requires a method that is able to capture these interactions with atomic detail. X-ray crystallography can provide this description, but biologically relevant cations, like Na^+ and K^+, are difficult to detect using standard crystallographic techniques, because their electronic densities are similar to that of water oxygen. Additionally, ions tend to be quite mobile, which makes them hard to resolve. Despite these difficulties, using a heavier cation (Cs^+) and X-ray crystallography with anomalous dispersion, our collaborators were able to determine the X-ray structure of subtilisin soaked in acetonitrile and cesium chloride[3]. The ion binding sites were clearly visible in the X-ray structure, which brought a valuable molecular insight into the issue of protein-ion interactions in nonaqueous solvents[3]. Nevertheless, the X-ray structure represents an average of the conformations found in the crystal and does not capture the dynamical behaviour of the molecules in solution. Additionally, the crystal environment and crystallographic contacts can create artificial binding sites. To circumvent these limitations, the X-ray characterization was complemented with an MD simulation study, performed in the scope of the present thesis and described in chapter 3. In that study we performed molecular dynamics (MD) simulations in acetonitrile, using the previously determined X-ray structure[3]. Additionally we used a docking methodology to predict the ions' binding sites and performed similar simulations with this system. The purpose of this second set of simulations was to avoid biasing our results by the initial placement of the ions.

To analyse our simulations, we built spatial probability density maps of the ions. Based on these maps, we observed that the distribution of the ions around the enzyme surface is not strongly biased by their initial locations. We also analysed

the occupancy of the original binding sites during the simulations, as well as the time spent by ions in the bulk solution. These analyses indicate that chloride ions tend to stay close to the protein, whereas cesium ions frequently migrate to the solvent.

In order to test the realism of using heavier cations, like Cs^+, to probe the locations of biologically relevant cations, such as Na^+ and K^+, we replaced the crystallographic cesium by sodium ions and subjected this system to MD simulations. These simulations revealed that the distribution of the two cations is similar, indicating that Cs^+ can, indeed, be used as a probe to find relevant cation binding sites.

Additionally, we performed MD simulations of subtilisin in an aqueous solution containing 1.5 M of CsCl (which was the concentration of the solution used to soak the crystals). The probability density maps of chloride and cesium ions obtained in water were compared with an X-ray structure determined by our collaborators, using a typical aqueous crystal, which was soaked with an aqueous CsCl solution (unpublished results). We observed that there is some agreement in the case of cesium, but not in the case of chloride ions. In an attempt to explain difference between the results obtained in the simulations and the chloride binding sites found in the crystal structure, we calculated the electrostatic potential in the crystal and in solution and mapped this potential on the enzyme surface. The analysis of the electrostatic potential maps showed that the crystal lattice generates an electrostatic potential which is very different from the one found in solution, altering the distribution of the ions. This explains the disagreement between the simulation and experimental results. Additionally, the probability density maps of chloride and cesium ions obtained in water were

compared to the ones obtained in acetonitrile. The maps obtained in the two solvents show some differences, which indicates that the solvent influences the distribution of the ions around the protein.

This study also found a possible explanation for the cation-dependent activation of subtilisin, which had been observed in the previous crystallographic study[3]. That study showed that the activity of the enzyme is higher in the presence of larger cations, such as Cs^+ and choline, when compared with Na^+ and K^+ cations[3]. MD simulations revealed that chloride ions form stronger ion pairs with Na^+ than with Cs^+. Therefore, when the cation present is Cs^+, Cl^- is more available to bind to the protein and tends to accumulate more around the protonated catalytic histidine (H64). Given that H64 needs to be deprotonated for the reaction to proceed, we hypothesize that the Cl^- ion can abstract a proton from this residue, when the reaction takes place in a moderately polar solvent, like acetonitrile. The higher availability of chloride found in the presence of cesium explains the observed rate enhancement observed when this cation is used.

This collaboration work has provided an important contribution to the elucidation of the role of counterions in nonaqueous biocatalysis, which is a long standing issue, with important applications in the rational development of enzymes that can efficiently work in these environments. We are, currently, extending these studies to other enzymes, ions, and solvents, which will enable us to obtain a general picture of protein-ion interactions in non-conventional media.

6.2 Protein stability in ethanol/water mixtures

The effect of the solvent on enzyme stability is another important aspect of nonaqueous catalysis, given that enzymes need to be stable in order to be useful for industrial processes. The molecular determinants of enzyme stability in aqueous/nonaqueous solutions have been analyzed in the scope of the present thesis (chapter 4). We compared the behaviour of two homologous proteases, pseudolysin (PSL) and thermolysin (TLN), which have similar structural properties, but considerably different stabilities in ethanol/water mixtures[4]. PSL is more stable than TLN in this medium[4] and this stability seems to be related with the presence of a disulfide bridge between cysteines 30 and 58 of the former enzyme[5]. Our goal was to analyse the molecular causes underlying this behaviour. Towards this end, we performed μs-long MD simulations of the two wild type enzymes and of a mutant of PSL, in which C58 was replaced by a glycine, abolishing the C30-C58 disulfide bridge. The proteins were immersed in two different solutions: water and ethanol/water (25% v/v), which correspond to the experimental conditions. We used μs-long simulations, in order to be able to detect appreciable conformational changes, which would not be visible on shorter time scales.

The stability of the proteins in the simulations was analyzed using the measures that are commonly employed for this purpose: rmsd, radius of gyration and secondary structure content. The results obtained were consistent with the previous experimental results. We observed that PSL is at least as stable in the alcohol/water mixture as in pure water. Thermolysin, on the other hand, suffers larger conformational changes in the presence of ethanol than in pure water. The C58G mutant of PSL is considerable less stable than the wild type protein in

the presence of ethanol, which corroborates the previous experimental findings, indicating that the C30-C58 disulfide bridge plays an important role in the stability of this enzyme[5].

To elucidate the causes underlying this behaviour we analysed the protein-solvent interactions during the simulations. The analysis of the contact surface between the proteins and ethanol molecules proved to be particularly insightful. This analysis showed that TLN was considerably more exposed to the nonaqueous solvent, during the simulations, than PSL. Interestingly, we also found that the C58G mutant of PSL became more exposed to ethanol, during the course of the simulations, than the wild type enzyme. These results indicate that protein-ethanol interactions are probably the driving force of the unfolding observed for TLN and the C58G mutant of PSL. These two proteins are more exposed to ethanol than wt PSL and, thus, are less stable in the presence of this alcohol.

This analysis has contributed to the elucidation of the molecular factors underlying enzyme stability in aqueous/nonaqueous mixtures. Given that enzyme stability is very important in technological applications, the findings of this study may be useful for the rational development of enzymes with increased industrial value. Due to the different characteristics and protein-solvent interactions displayed by different nonaqueous solvents, it would be interesting to perform similar studies using other solvents and enzymes.

6.3 Ligand imprinting

Molecular memory is a very curious enzyme property, which is only observed in apolar solvents. It was first recognized when Klibanov and co-workers observed

that the activity of subtilisin in apolar media can be enhanced by lyophilising the biocatalyst in the presence of competitive inhibitors, which are removed before transferring the protein to the nonaqueous solvent[6]. This observation led them to conclude that the enzyme "remembers" the ligand-induced state, given that its activity in the apolar media is affected by the previous contact with the ligand[6]. This phenomenon is also known as ligand-imprinting, bioimprinting or ligand-induced enzyme memory. Given that one of the drawbacks of nonaqueous enzymology is the reduced activity observed in these media, this finding opened the way for a new strategy to increase the catalytic efficiency in these solvents. Studies with other enzymes have shown that this property is not specific of subtilisin.

The molecular determinants of ligand imprinting were not clear when this thesis was initiated. Given that this phenomenon is quite interesting from a technological and fundamental viewpoint, we decided to investigate it, using MD simulations (see chapter 5). Our approach was to mimic Russel and Klibanov's experiment[6] *in silico*, in order find the molecular causes of this phenomenon. We started by docking in the active site of subtilisin one of the inhibitors that had been experimentally tested, using a molecular docking methodology. The enzyme-ligand complex was simulated in water for 20 ns, so that the active site could adapt to the ligand. The ligand was then removed and the enzyme structure obtained after these 20 ns was used to perform simulations in water and hexane. As a control, we performed simulations in both solvents, using the X-ray structure without contact with the ligand. This corresponds to the control simulations performed experimentally, in which the enzyme was lyophilised in the absence of the ligand[6].

We observed that the prior contact with the ligand has a large effect in the behaviour of the protein in hexane. The inhibitor induces an open conformation of the active site which is retained in hexane simulations. When the enzyme is not pre-treated with the ligand, the active site is considerably more closed. These results can explain the rate enhancement observed by Russell and Klibanov upon treatment with competitive inhibitors[6]. In accordance with the previous experimental results[6], we found that in water, the enzyme behaviour is not affected by the previous contact with the ligand. The simulations revealed that this phenomenon is a consequence of the decreased flexibility of the enzyme in the apolar solvent. Due to this low conformational mobility, the enzyme is trapped in the metastable conformation induced by the inhibitor. Given that this state is appropriate to receive the incoming substrate, this leads to the observed rate enhancement.

This work is not only in line with the previous experimental findings, but also provides a molecular perspective and elucidates the structural determinants of ligand-imprinting. These results are useful for the rational development of enzyme reactions with increased activity and specificity towards particular substrates.

Overall, this thesis has successfully contributed to a broader understanding of the molecular determinants of nonaqueous enzymology. In particular, it has shed light on three distinct aspects of this field, which were poorly characterized at the molecular level: protein-ion interactions, enzyme stability and ligand-imprinting. These and other simulation studies (see section 1.3) show that,

despite the inherent challenges and limitations of nonaqueous MD simulations, they can provide valuable molecular insights into the behaviour of these systems.

Appendix A

Supporting information for chapter 3

A.1.1 Protocol for selecting counterion positions using molecular docking

Figure S1 illustrates the approach used for selecting counterion positions using a docking based methodology previously developed by us[1]. As can be observed in the figure, cations (Cs^+) and anions (Cl^-) were docked independently from each other and two sets of docking simulations were performed for each type of ion. In the first set, we docked the ions directly on the X-ray structure (see the left-hand side of fig. S1), performing 15 sequential docking simulations for each type of ion 15 and (note that each docked ion was added to the protein structure before the next docking simulation). This guarantees that all the protein charged side-chains are neutralized by ions. In the second set of simulations, we performed simulated annealing before docking the ions (see the right-hand side of fig. S1). In the simulated annealing procedure, the temperature was linearly decreased from 300 K to 0 K in 30 ps, with the purpose of relaxing the protein side chains, so that neighbor negative and positive residues could form salt bridges (and therefore avoid the necessity of charge neutralization by counter ions), therefore, the resulting structure is named "relaxed structure". In the simulated annealing simulation, the Cα atoms of the protein were restrained using a force constant of 10^5 kJ mol^{-1} nm^{-2} in the *x*, *y* and *z* directions. The ions were then docked, applying the same methodology described above for the X-ray structure, however, in this case some of the original binding sites were no longer available, because some residues were able to form salt bridges with each other

during the annealing process. The ions that occupy the same locations in the X-ray and relaxed structures were considered essential and, therefore, maintained. We used X-ray structure with the essential ions as a starting point for subsequent MD simulations

The docking simulations were performed using the software AutoDock, version 4.0[166, 167]. Given that the Monte Carlo simulated annealing algorithm is more efficient in the docking of molecules with no rotatable bonds than the other algorithms implemented in AutoDock, we chose this algorithm, using an initial RT value of 41.84 $kJmol^{-1}$ and performing 100 cycles with an annealing temperature reduction factor of 0.92 per cycle. The number of maximum accepted or maximum rejected Monte Carlo steps was 20000. The initial translation step was 0.1 nm and was reduced by a factor of 0.9702 in each cycle. Thirty independent runs were performed for placing each ion, and the lowest energy solution was selected; in most cases, this solution was found many times. All waters were removed from the protein structure, following the general setup of the docking procedure implemented in this methodology. We used Kollman united-atom partial charges for the protein and added only polar hydrogens, using distance dependent dielectric constant was used for electrostatic interactions[248, 249].

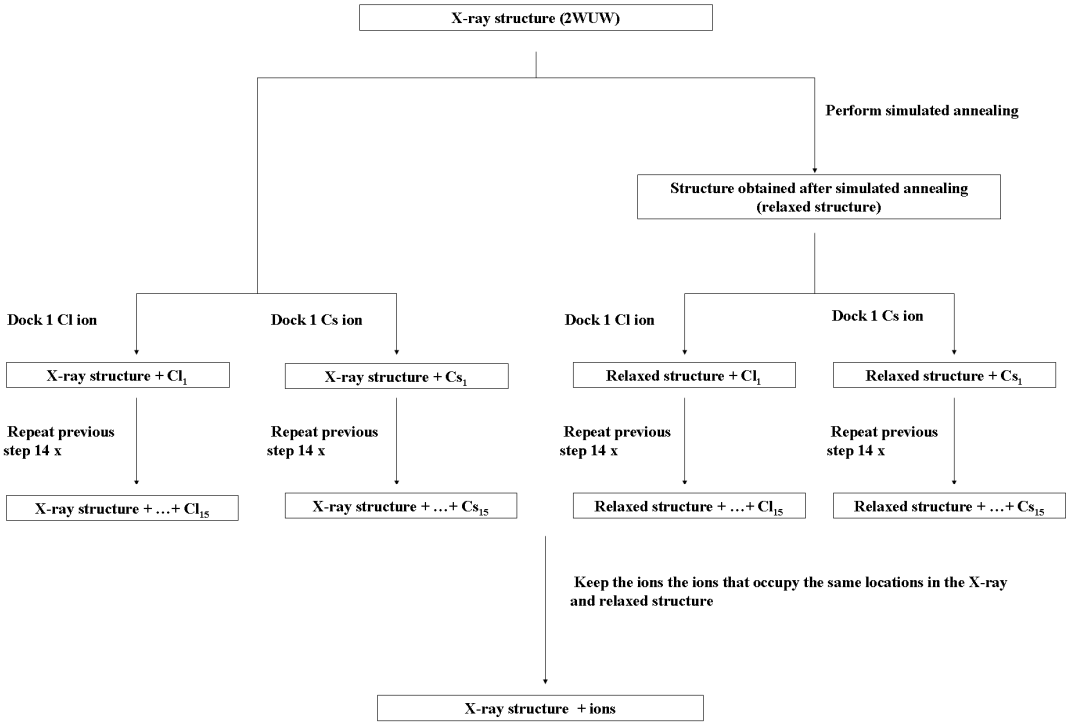


Figure A1. Protocol for selecting counterion positions using molecular docking

A.1.2 Methodology used to randomly distribute Cs^+ and Cl^- ions in the simulations performed in water with 1.5 M of salt

In order to analyze how the ions interact with subtilisin in an aqueous solution containing 1.5 M of CsCl, we randomly inserted the corresponding amount of ions in the simulation box, far from the protein, which assures that our results are not biased by the starting positions of the ions. Subtilisin was initially placed in a center of dodecahedral box, setting the minimum distance between the protein and the box walls to 1.2 nm; and this box was then filled with water. Cs^+ and Cl^- ions were added to the solvated protein, using the tool *genion*, available in the GROMACS package[154], until a 3M concentration was reached and assuring that the system was neutral. All the water molecules were kept in the box. Then, we removed the ions which were closer to the protein, leaving only half of the ions in the box. In the end, the system had the desired concentration of 1.5 M of salt and the ions were randomly distributed in the most external part of the simulation box, having no contact with the protein. This procedure was only used in water simulations, because unfortunately, it can not be used in acetonitrile due to the very low solubility of CsCl in this medium.

A.1.3 Protocol for modeling protein protonation equilibrium

The determination of the pK_a of each titrable site in the protein was performed using a methodology developed by us, based on continuum electrostatics and Monte Carlo sampling of protonation states that has been explained in detail before[171, 172]. This methodology, besides considering the tautomeric and pseudo-tautomeric states of ionizable groups, also considers pseudo-tautomers for alcohol groups and water molecules. This is an attempt to improve over the rigid picture of the protein imposed by the continuum electrostatic method. Only water molecules with a relative accessibility inferior or equal to 0.5 were included in these calculations. The electrostatic energy terms were calculated by solving the Poisson-Boltzmann equation, using the MEAD package[207, 244]. The program PETIT[172] that implements a Monte Carlo procedure, was used to sample the protonation states at different values of pH, using the energy terms calculated by MEAD.

A.1.4 System preparation for MD simulations

Before starting the simulations in acetonitrile, the protein was placed in a dodecahedral box, keeping all the crystallographic waters and leaving 1.2 Å between the protein and the box walls. The box was then filled using a cubic box of acetonitrile that had been previously equilibrated at the experimental density at 300 K and 1 atm. The solvent was relaxed by performing 2000 steps of energy minimization, using the steepest descent algorithm and applying restraints in all the protein heavy atoms, ions and water molecules, followed by 2000 steps of energy minimization with restraints in the Cα atoms of the protein and ions. After the minimization procedure, we performed four initialization steps. In the first step, velocities were assigned according to a Maxwell–Boltzmann distribution and 50 ps of MD we carried out in the NVT ensemble, with a temperature coupling constant of 0.025 and restraints on all the protein heavy atoms, ions and water molecules. In the second step, we performed 50 ps of MD, in the NPT ensemble, using coupling constants of 0.05 and 8 ps, for the temperature and pressure, respectively and keeping all the heavy atoms, ions and water molecules restrained. In the third step of the initialization, the system was simulated for 200 ps in the NPT ensemble, using temperature and pressure coupling constants of 0.1 and 1.3 ps, respectively, and position restraints in the Cα atoms, ions and water molecules. Finally, we performed 500 ps of MD simulation, maintaining all the conditions, except the position restraints on the ions and water molecules, which were removed.

The solvation and minimization procedures for water simulations were identical to the ones used for acetonitrile simulations. In the first initialization step, velocities were assigned according to a Maxwell–Boltzmann distribution and 50

ps of MD were carried out in the NVT ensemble, with a temperature coupling constant of 0.025 ps and restraints in all the protein heavy atoms and ions. In the second step, we performed 50 ps of MD, in the NPT ensemble, using coupling constants of 0.025 and 0.5 ps, for the temperature and pressure, respectively, and keeping all the heavy atoms and ions restrained. In the third step of the initialization, the system was simulated for 50 ps in the NPT ensemble, using the same pressure coupling constant, a temperature coupling constant of 0.05 ps and position restraints in the Ca atoms and ions. Finally, we performed 50 ps maintaining all the conditions, except the temperature coupling constant, which was changed to 0.1 ps.

A.2 Results and discussion

A.2.1 Potentials of mean force between the cations, Cs^+ and Na^+, and the anion, Cl^-, in solvents with different polarities

In order to analyze how a cation (Cs^+ or Na^+) and an anion interact in solvents with distinct polarities (water, acetonitrile and hexane), we calculated the potentials of mean force (PMFs) between the two oppositely charged ions in these media. The plots in fig. A.2 show the PMFs obtained. We can see that in a very apolar medium like hexane the PMFs of both Na^+Cl^- and Cs^+Cl^- have very deep minima at distances of around 0.25 and 0.3 nm, respectively. This means that, as expected, in very apolar solvents, the interaction between oppositely charged ions is very strong and the ions form highly stable complexes that are never broken at room temperature.

The situation in water is very different from the one found in hexane. The plots in fig. A.2 show that in aqueous solution the minima are very shallow, which indicates that the ions tend to be dispersed and do not form stable complexes at room temperature. Acetonitrile is much more polar than hexane, but not as polar as water and therefore one would expect that the interaction between opposite charges would be much weaker in acetonitrile than in hexane, but stronger than in water. Indeed, looking at the plots in figure A.2B, we can clearly distinguish minima in the PMFs of Na^+Cl^- and Cs^+Cl^- in acetonitrile. The depth of these minima is around -50 kJ and -25 kJ for Na^+Cl^- and Cs^+Cl^-, respectively. This means that the complexes formed between the cation and the anion in acetonitrile are strong. To further elucidate the strength of the interaction between the ions in acetonitrile and water, we performed unconstrained MD simulations where the ions were initially placed 1 nm apart. In fig. A.3, we can see that the ions tend to form stable associations in acetonitrile, which once formed were not broken during

the 10 ns of simulation. In contrast, the unconstrained simulations in water do not show stable associations between the ions.

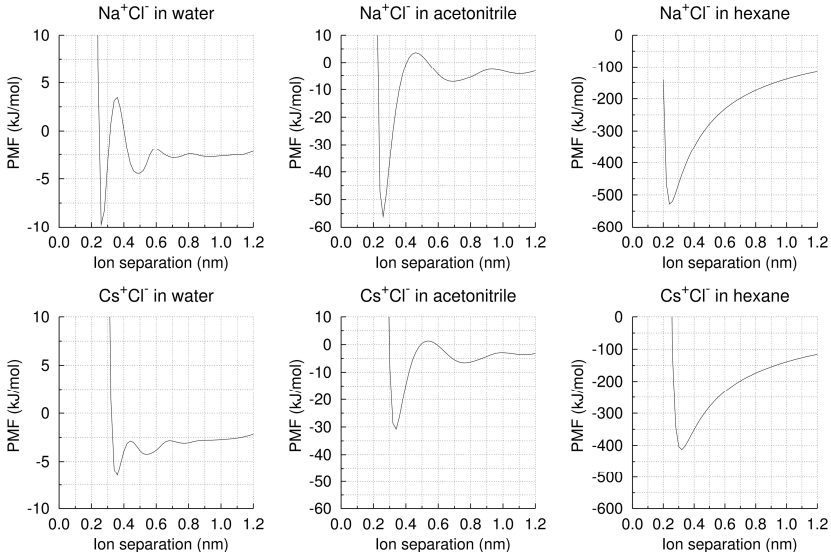


Figure A2. Potentials of mean force between the cations Cs^+ and Na^+ in water, acetonitrile and hexane. Different scales were used for different solvents to enable a clear visualization of the peaks.

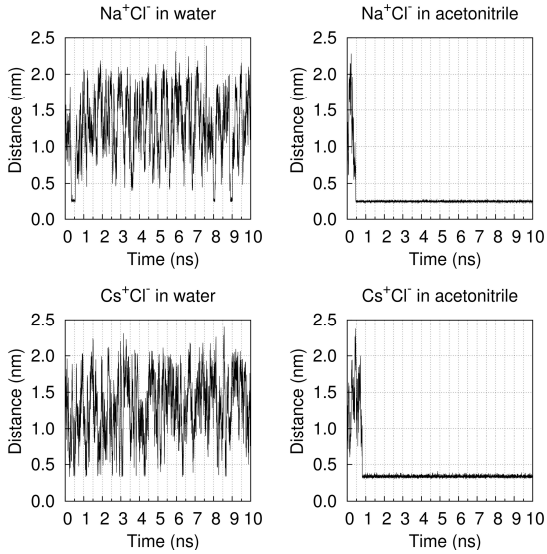


Figure A3. Temporal evolution of the distance between cation and anion during unconstrained MD simulations in water and acetonitrile

A.2.2 Determination of protonation of ionizable residues at pH 6.5

The determination of the pK_a values of all the titrable residues of subtilisin was performed using a methodology based on continuum electrostatics. As is usually the case, at pH 6.5 all the acidic residues of subtilisin, including the C-terminal, were found to be deprotonated and all its basic residues were found to be protonated. The N-terminus has a pK_a around 7 and was therefore considered to be protonated at pH 6.5. One of the most relevant residues of subtilisin is histidine 64, which is part of its catalytic triad. It is generally accepted that this residue acts both as an acid and a base during the course of the catalytic process and, therefore, its pK_a should be close to 7. As can be seen in fig. A.4, at the pH of interest (6.5), the protonated fraction of the catalytic histidine is around 70%. This means that both states (fully protonated and partially deprotonated) are expected coexist at this pH. Although, according to our calculations, the fully protonated state is the predominant one, it is believed that this residue must be partially deprotonated in order to accept the proton from serine 221 during the catalytic process. Therefore, both states were considered in our MD simulations.

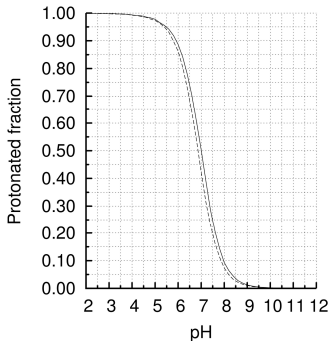


Figure A4. Titration curves of the catalytic histidine. The solid line corresponds to the structure obtained in absence of CsCl (PDB ID: 2WUW) and the dashed line corresponds to the structure obtained in presence of CsCl (PDB ID: 2WUV).

A.2.3 Evolution of the protein structure in acetonitrile and water simulations

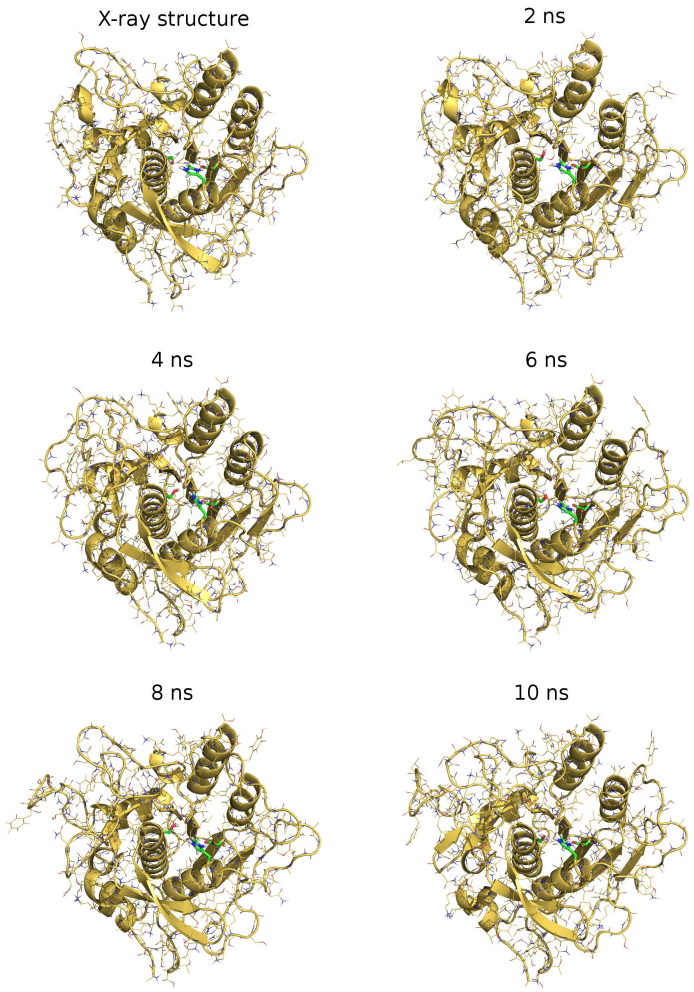


Figure A5. Evolution of subtilisin's structure during a typical simulation in acetonitrile. The figure shows the X-ray structure that was used as a starting point of the simulations and snapshots obtained at several time points of a trajectory (as indicated in the figure labels). The snapshots were taken from a simulation (replicate 17 of the simulations performed in acetonitrile with docked ions) which has an r.m.s.d. profile similar to the average and is, therefore, representative of acetonitrile simulations. The catalytic triad residues are shown using sticks with the carbon atoms colored in green.

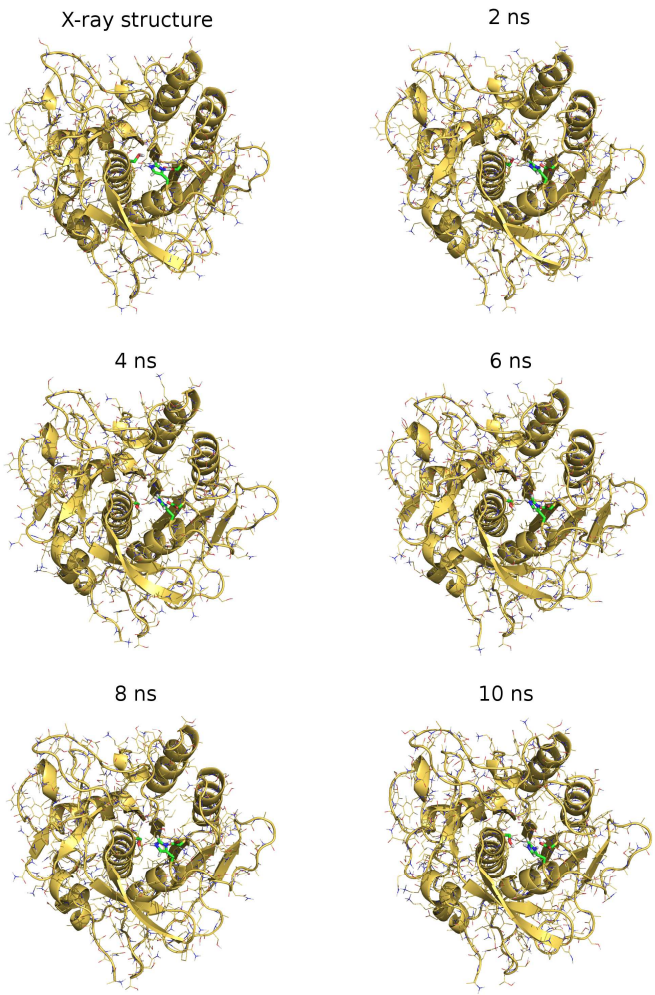


Figure A6. Same as figure S5, but for water simulations. The snapshots were taken from replicate 1 of the simulations performed in water with crystallographic ions.

A.2.4 Electrostatic surface maps of subtilisin in the crystal environment and in solution

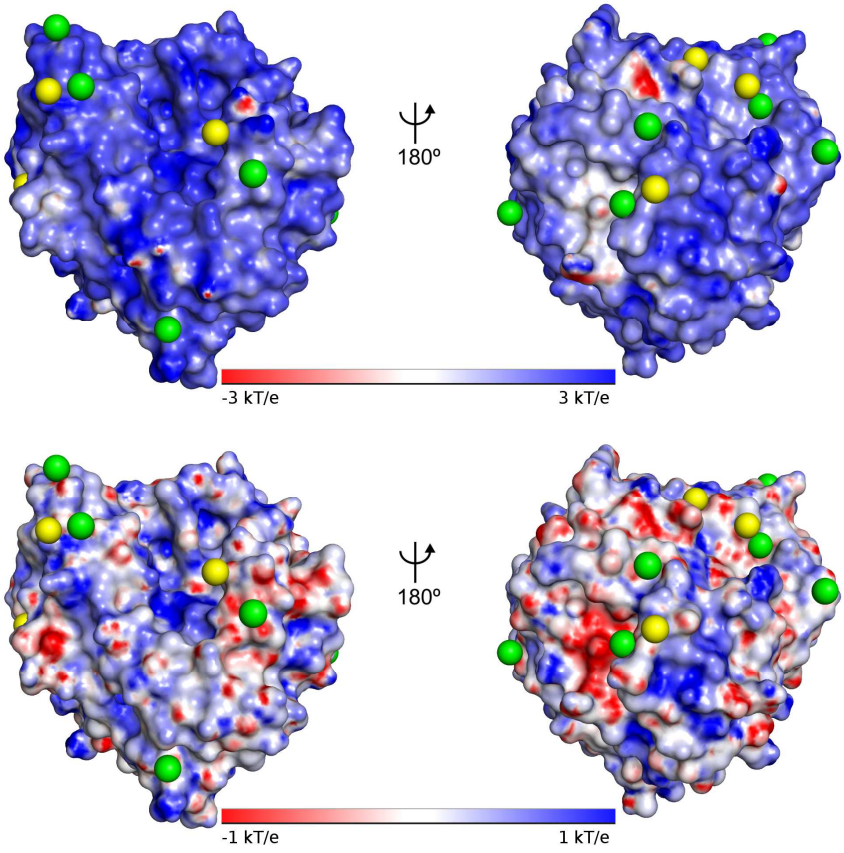


Figure A7. Electrostatic surface maps of subtilisin in the aqueous crystal environment (A) and in aqueous solution (B). The cesium (green spheres) and chloride (yellow spheres) ions are shown in the locations found in the crystal structure obtained in aqueous conditions (Cianci et al. (to be published)). Note that the scale in A and B is different.

A.2.5 Radial distribution function of Cl^- around the $N\varepsilon_2$ of H64

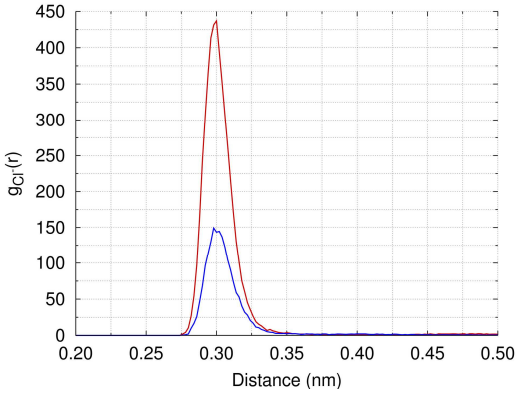


Figure A8. Radial distribution function of Cl^- around the $N\varepsilon_2$ of H64. The red line refers to the simulations with CsCl and the blue line corresponds to the simulations with NaCl. In both cases H64 is protonated.

A.3 Movies

Movie_A1. Temporal evolution of the probability density maps for Cl^- in acetonitrile simulations with CsCl. The initial positions of the ions were the ones available in the X-ray structure.

Movie_A2. Same as Movie_A1 but for Cs^+.

Movie_A3. Same as Movie_A1, but the initial positions of the ions were the ones determined using the docking methodology.

Movie_A4. Same as Movie_A3 but for Cs^+.

Movie_A5. Same as Movie_A1 but for the simulations where Cs^+ was replaced by Na^+.

Movie_A6: Same as Movie_A1 but for Na^+, in the simulations where NaCl was used instead of CsCl.

Appendix B

Supporting information for chapter 4

B.1. Methods

B.1.2 System preparation for MD simulations

Before performing the equilibrium MD simulations, the systems under study had to be prepared. In the case of water simulations, the protein was placed in a dodecahedral box, keeping crystallographic waters with a solvent accessibility lower or equal to 0.5 and leaving 0.9 nm between the protein and the box walls. The box was then filled using a water solution that had been previously equilibrated. Water was relaxed, by performing 2000 steps of energy minimization using the steepest descent algorithm and applying restraints in all the protein heavy atoms, followed by 2000 steps of energy minimization with restraints in the Cα atoms of the protein and, finally, 2000 steps with no restraints. After the minimization procedure, we performed four initialization steps. In the first step, velocities were assigned according to a Maxwell–Boltzmann distribution and 50 ps of MD were carried out in the NVT ensemble, with a temperature coupling constant of 0.025 and restraints on all the protein heavy atoms. In the second step, we performed 50 ps of MD, in the NPT ensemble, using coupling constants of 0.025 and 0.5, for the temperature and pressure, respectively, and keeping all the heavy atoms restrained. In the third step of the initialization, the system was simulated for 50 ps in the NPT ensemble, using the same pressure coupling constant, a temperature coupling constant of 0.05 ps and position restraints in the Cα atoms. Finally, we performed 50 ps maintaining all the conditions of the

previous step, except the temperature coupling constant, which was changed to 0.1 ps.

The first step of the simulations carried in ethanol/water mixture concerned the preparation and equilibration of the ethanol/water solution. This was done by randomly placing 100 molecules in a cubic box with 3.365 nm^3 and then filling the box with water. In order to maintain the right proportion between ethanol and water, the water in excess was removed and, in the end, the box contained 972 water molecules. Next, we performed 2000 steps of energy minimization with the steepest descent algorithm and then conducted three initialization steps, starting by assigning velocities according to a Maxwell–Boltzmann distribution and performing 50 ps of MD in the NVT ensemble, using a temperature coupling of 0.025 ps, followed by 50 ps in the NPT ensemble with the same temperature coupling and a pressure coupling of 0.5 ps, and finally 50 ps in the NPT ensemble with a temperature coupling of 0.05 ps and a pressure coupling of 0.5 ps. The system was then equilibrated for 10 ns, in the NPT ensemble, using a temperature coupling of 0.1 ps and a pressure coupling of 0.5 ps. The protein was placed in the center of a dodecahedral box with a distance of 1.2 nm between the protein and the box walls. The box was then filled with the ethanol/water solution that had been previously equilibrated and water molecules were removed until the fraction of ethanol in solution reached 0.25 v/v. The minimization procedure was identical to the one applied in water simulations. The initialization protocol was also similar to the one used in aqueous simulations, with the exception of the second step, where 500 ps were used (instead of 50), to enable the box volume to adjust, due to the removal of water molecules.

B.1.3 Methodology used in the determination of protonation states

The determination of the protonation state of each titrable site in the protein at pH 7 was performed using a methodology developed by us, based on continuum electrostatics and Monte Carlo sampling of protonation states, that has been explained in detail before[171, 172]. Only water molecules with a relative accessibility inferior or equal to 0.5 were included in the calculations of the protonation equilibrium. The electrostatic energy terms were calculated by solving the Poisson-Boltzmann equation, using the MEAD package[207, 244]. The program PETIT[172], that implements a Monte Carlo procedure, was used to sample the protonation states at different values of pH, using the energy terms calculated by MEAD.

B.2 Results

B.2.1 Analysis of rigid body motions between the domains of the proteins under study

In order to have a quantitative measure of the rigid body motions that occur in our simulations, we calculated the r.m.s.d. for each domain separately. In figure S1, we can see that the three domains are quite stable in the simulations of native pseudolysin, both in water and in the ethanol/water mixture. In replicate 1 of the C58G mutant of pseudolysin, the C-terminal domain displays a large r.m.s.d., which is responsible for its large global r.m.s.d (see fig. 2). In replicates 4 and 5, the r.m.s.d of the three individual domains is considerably lower than the r.m.s.d. of the whole protein, which means that in these replicates there are rigid body motions. In the simulations of thermolysin in water, the three domains have similar r.m.s.d. values, which are lower than the global r.m.s.d., indicating that there are rigid body motions. In the ethanol/water mixture, the active-site and C-terminal domains of thermolysin display larger r.m.s.d values than the N-terminal domain. As we suspected, the value obtained for the global r.m.s.d of thermolysin in ethanol/water is considerably higher than the ones obtained for the domains separately, confirming that there are interdomain rigid-body motions.

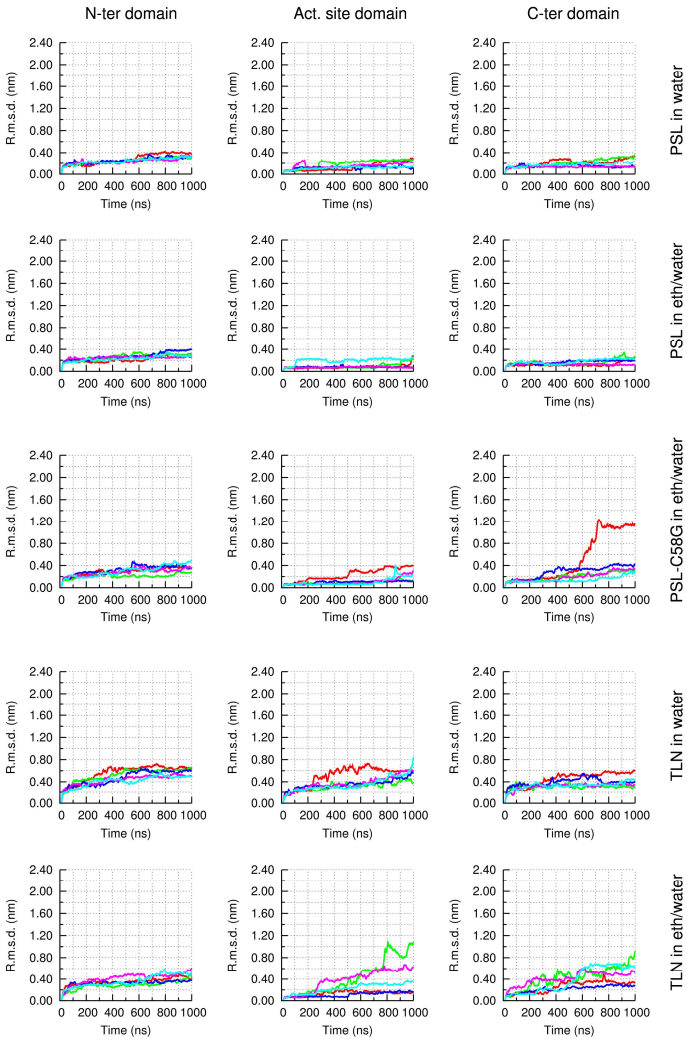


Figure B1. Moving average of the r.m.s.d from the X-ray structure calculated separately for each domain and excluding the same loop as in fig. 2. The plots in the 1st, 2nd and 3rd columns correspond to the Cα atoms in the N-terminal, active-site and C-terminal domains, respectively. Each replicate is represented by a line with a different color, as in fig. 2.

B.2.2 Contact area between water molecules and the protein

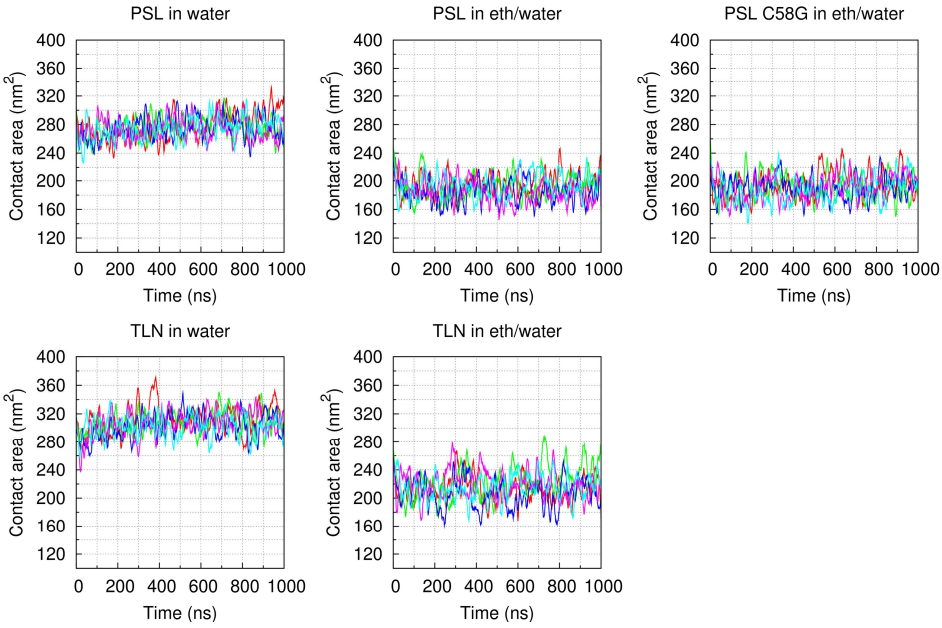


Figure B2. Moving average of the contact area between water molecules and the protein (calculated as in fig. 5). The lines with different colors represent different replicates, as in fig. 2.

B.2.3 Distribution of the water molecules around the protein

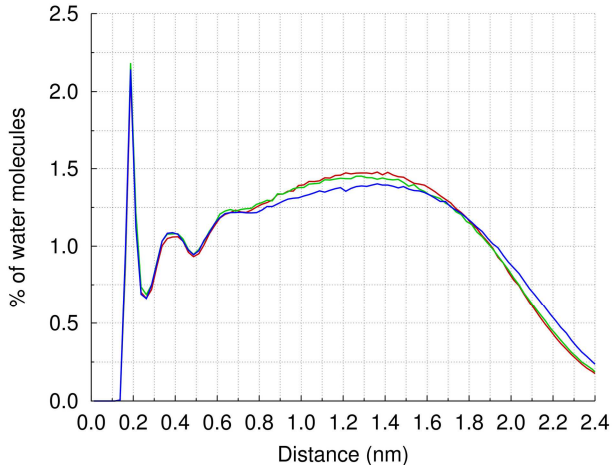


Figure B3. Distribution of the water molecules around the protein in the last 100 ns of the simulations performed in water. Each line represents a different protein, as in fig. 6.

B.2.4 Distributions of the alcohol and alkyl moieties of the ethanol molecule around the protein

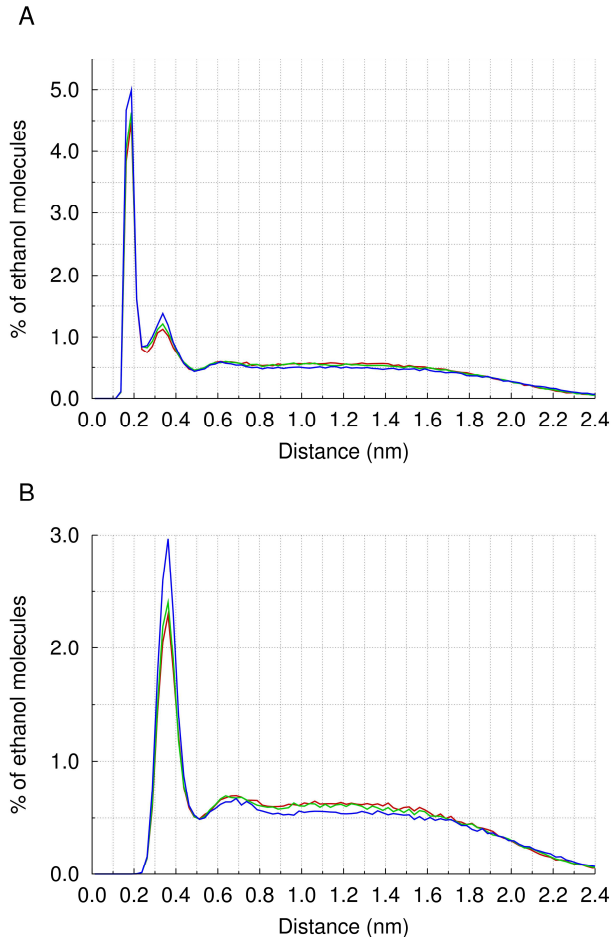


Figure B4. Distributions of the OH (A) and CH_2CH_3 (B) moieties of the ethanol molecule around the protein in the last 100 ns of the simulations performed in the ethanol/water mixture. Each line represents a different protein, as in fig. 6.

B.2.5 Comparison of the thermolysin residues that interact most frequently with ethanol in our simulations with the binding sites of isopropanol determined in a previous X-ray study

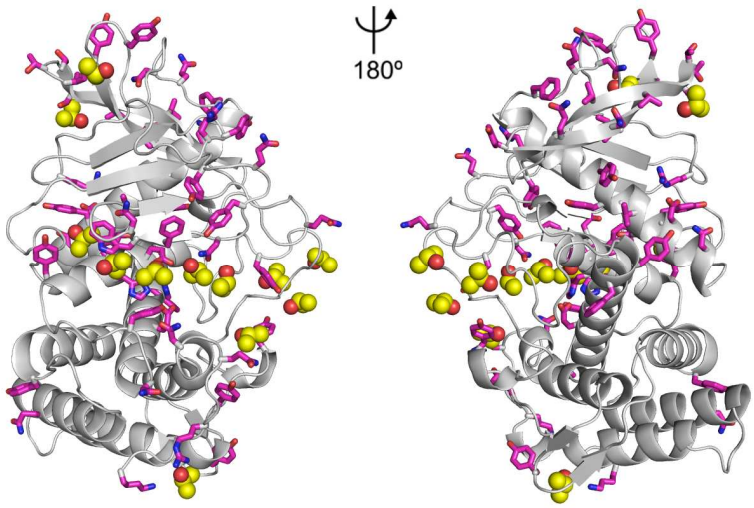


Figure B5. Comparison of the residues that interact most frequently with ethanol in our simulations with the binding sites of isopropanol in a previously determined X-ray structure. The structure of thermolysin is shown in grey, using a cartoon representation. The residues that interact with ethanol at least 99% of the time in our simulations are highlighted using sticks, with the carbons colored in magenta. The isopropanol molecules that were found in the X-ray structures 7TLI and 8TLI[174] are represented using spheres, with the carbons colored in yellow. The figure shows that there is a good agreement between our simulations and the previous experimental study.

B.2.6 Areas of the histogram peaks

Table B1. Areas of the peaks observed in the distribution of ethanol molecules around the protein in the last 100 ns of the simulations performed in the ethanol/water mixture (fig. 6).

System	Area of the 1st peak	Area of the 2nd peak
PSL in eth/water	0.261	0.432
PSL-C58G in eth/water	0.273	0.454
TLN in eth/water	0.295	0.518

Table B2. Areas of the peak observed in the distribution of the water molecules around the protein in the last 100 ns of the simulations performed in the ethanol/water mixture (fig. B.3).

System	Area of the 1st peak	Area of the 2nd peak
PSL in eth/water	0.113	0.214
PSL-C58G in eth/water	0.117	0.219
TLN in eth/water	0.114	0.218

Table B3. Areas of the peak observed in the distribution of the OH moiety of the ethanol molecule around the protein in the last 100 ns of the simulations performed in the ethanol/water mixture (fig. B.4A).

System	Area of the 1st peak	Area of the 2nd peak
PSL in eth/water	0.260	0.160
PSL-C58G in eth/water	0.270	0.180
TLN in eth/water	0.295	0.191

Table B4. Areas of the peak observed in the distribution of the CH_2CH_3 moiety of the ethanol molecule around the protein in the last 100 ns of the simulations performed in the ethanol/water mixture (fig. B.4B).b

System	Area of the peak
PSL in eth/water	0.266
PSL-C58G in eth/water	0.276
TLN in eth/water	0.329

B.2.7 Comparing the behavior of wild type and C58G mutant of pseudolysin

Movie_B1. This movie shows the structural changes that occur during the simulations of the wild type pseudolysin in the ethanol/water mixture (25% v/v). The simulations were divided in 10 ns windows and each frame of the movie represents the average structure of the enzyme in the corresponding time window. The structures are shown using a cartoon representation. The side chains of residues C30 and C58G are represented by green sticks and the loops where these residues are located are colored in magenta. All the replicates are displayed in the movie sequentially.

Movie_B2. This movie shows the structural changes that occur during the simulations of the C58G mutant of pseudolysin in the ethanol/water mixture (25% v/v). See legend of Movie S6 for further details.

Movie_B3. This movie shows the temporal evolution of the distribution probability density of ethanol in one of the simulations (replicate 1) of the C58G mutant pseudolysin in the ethanol/water mixture (25% v/v). The simulation was divided in 10 ns windows and each frame of the movie represents the average distribution probability density in the corresponding time window. The contours enclose regions with a probability density above 9 $\times 10^{-6}$ $Å^{-3}$. The average structure of the enzyme in each time window is shown using a cartoon representation.

Appendix C

Supplementary information for chapter 5

C.1 Methods

C.1.1 Protocol for selecting counterion positions

Figure S1 illustrates the approach used for selecting counterion positions. We used sodium ions to neutralize negative exposed side chains and chloride ions to neutralize positive exposed side chains. Sodium and chloride ions were docked independently from each other. As can be observed in figure 2, we performed two sets of docking simulations for each type of ion, one where the ions were docked on the starting structure of a given hexane simulation ("original structure") and another where this structure was previously subjected to simulated annealing ("relaxed structure"). The left side of figure 2 summarizes the protocol used for the original structure. We started by performing 15 successive docking simulations for each type of ion. Each docked ion was added to the protein structure before the next docking simulation. This places ions in all the locations for which they have affinity. The protocol applied for the relaxed structure is shown in the right side of figure 2. We started with a simulated annealing procedure, where the temperature was linearly decreased from 300 K to 0 K in 30 ps. Our aim was to relax the protein side chains, so that neighbor negative and positive residues could form salt bridges. In the simulated annealing simulation, the Cα atoms of the protein were restrained using a force constant of 10^5 kJ mol^{-1} nm^{-2} in the *x*, *y* and *z* directions. During the first part of the simulation, when the system was subjected to high temperatures, the side chains were able to explore the conformational space and eventually come into contact with

opposite charged side chains. As the temperature was decreased, the system started to freeze, reaching a low energy state. This procedure enabled the formation of salt bridges between opposite charged residues that were close enough to interact. We then applied the same docking methodology described above for the original structure. In the case of the relaxed structure the ions did not tend to dock near the residues that were able to form salt bridges. The side chains that were neutralized by ions both in the original and in the relaxed structures were considered essential ion sites. Finally, ions were placed in the essential ion sites of the original structure.

The docking simulations were performed using the software AutoDock, version 4.0 [167]. All waters were removed from the protein structure. Kollman united-atom partial charges were used. Only polar hydrogens were considered. A distance dependent dielectric constant was used for electrostatic interactions [248, 249]. A Monte Carlo simulated annealing algorithm was used, starting with an RT value of 1000 kcal mol^{-1} and performing 100 cycles with an annealing temperature reduction factor of 0.92 per cycle. We chose this algorithm because it is more efficient in the docking of molecules with no rotatable bonds than the other algorithms implemented in AutoDock. The number of maximum accepted or maximum rejected Monte Carlo steps was 20000. The initial translation step was 1.0 Å and was reduced by a factor of 0.9702 in each cycle. Thirty independent runs were performed for placing each ion, and the lowest energy solution was selected; in most cases, this solution was found many times.

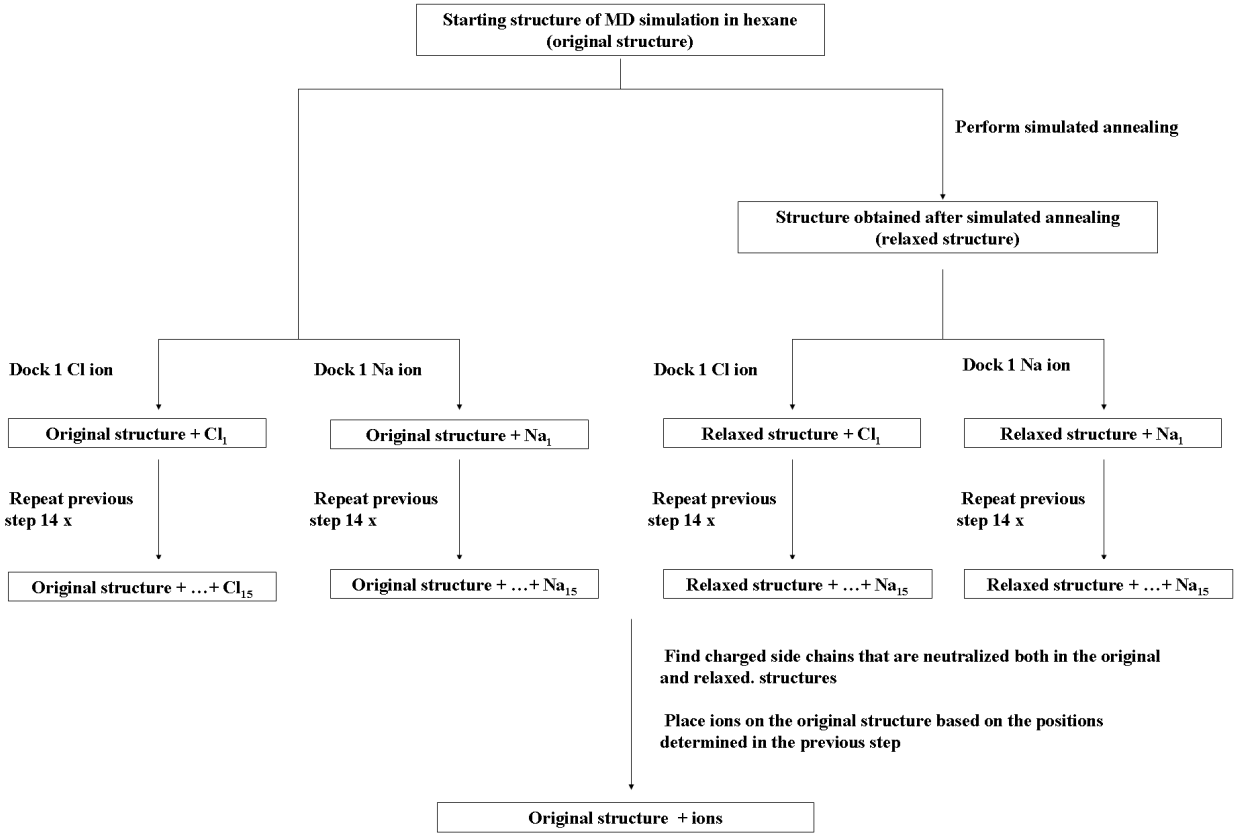


Figure C1. Protocol for selecting counterion positions.

C.2 Results

C2.1 Protein stability

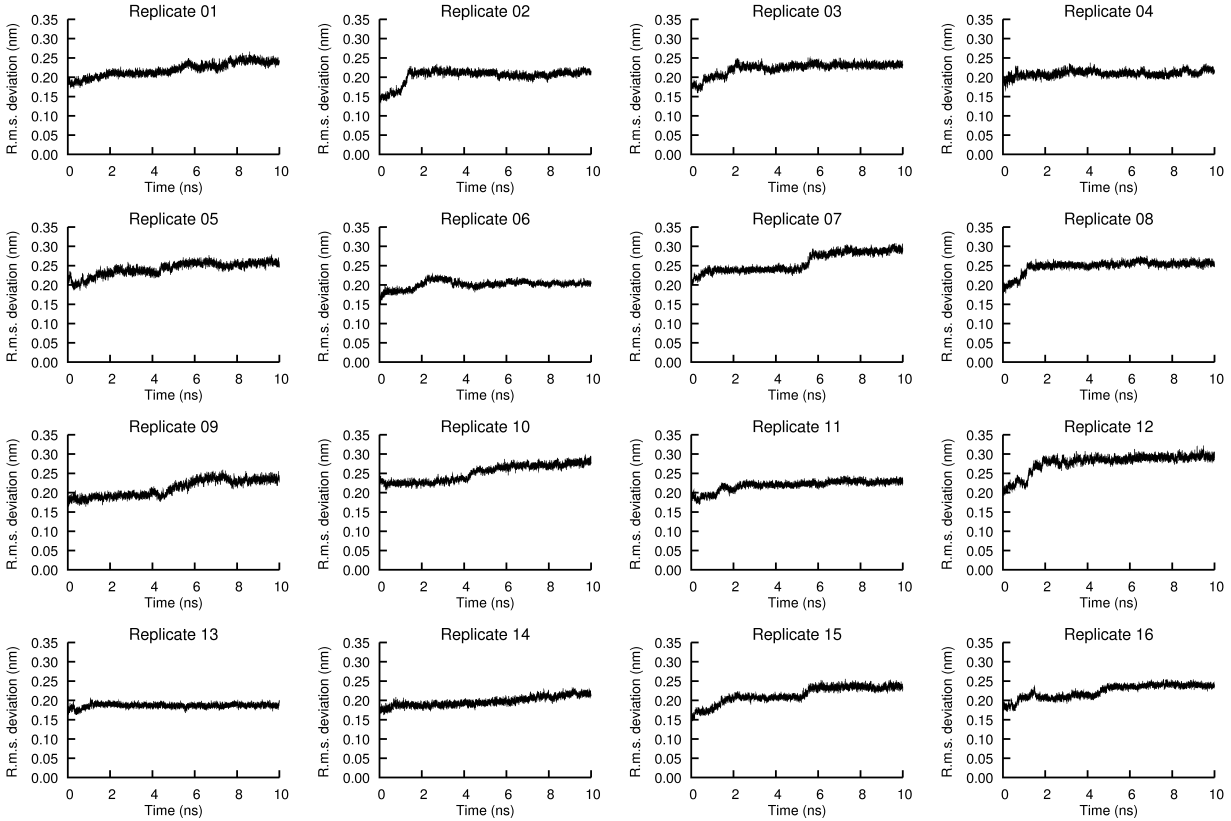


Figure C2.1. Root mean square deviation of Cα atoms from the x-ray structure. of ligand-treated simulations in hexane.

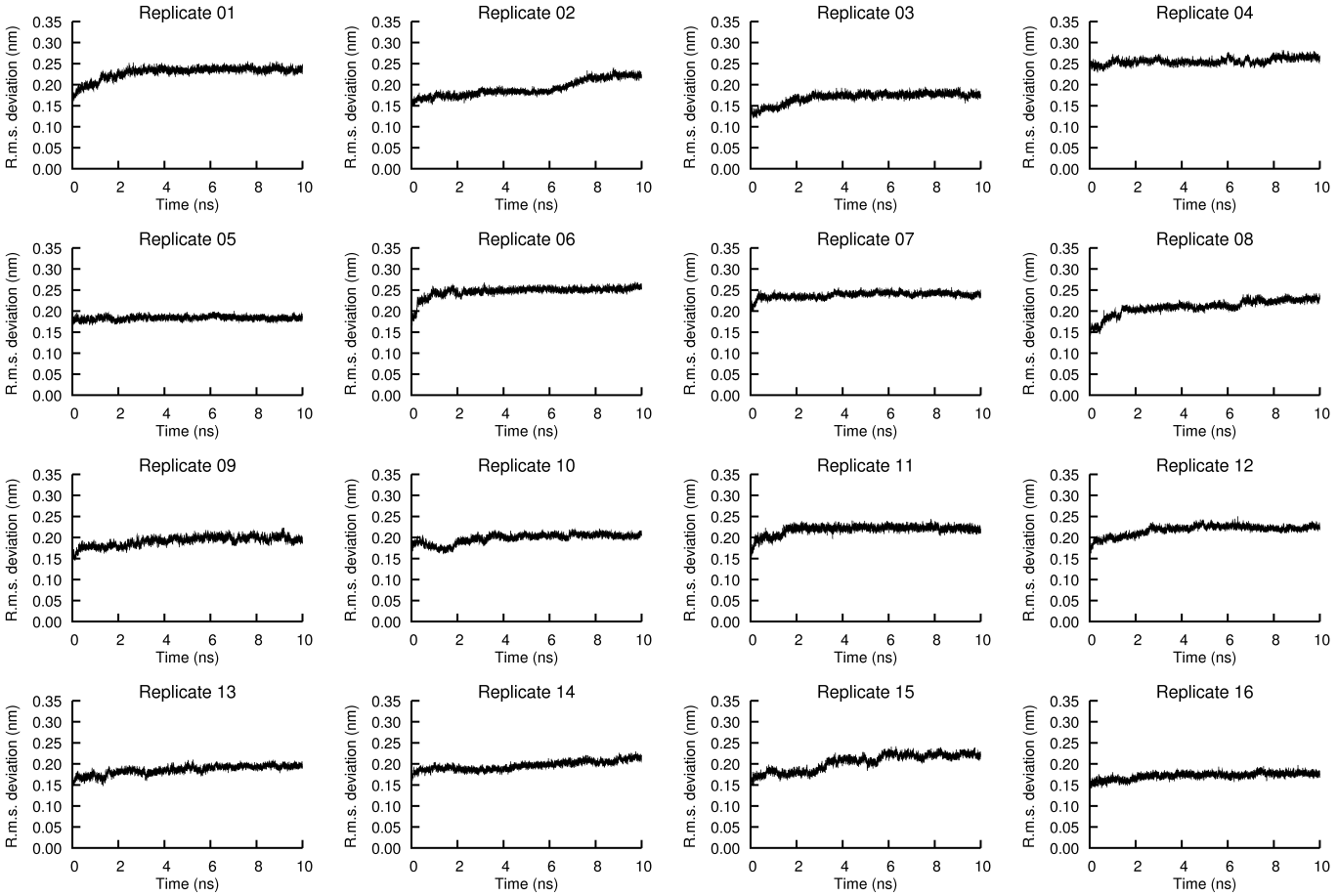


Figure C2.2. Same as figure C.2.1 but for igand-untreated simulations in hexane.

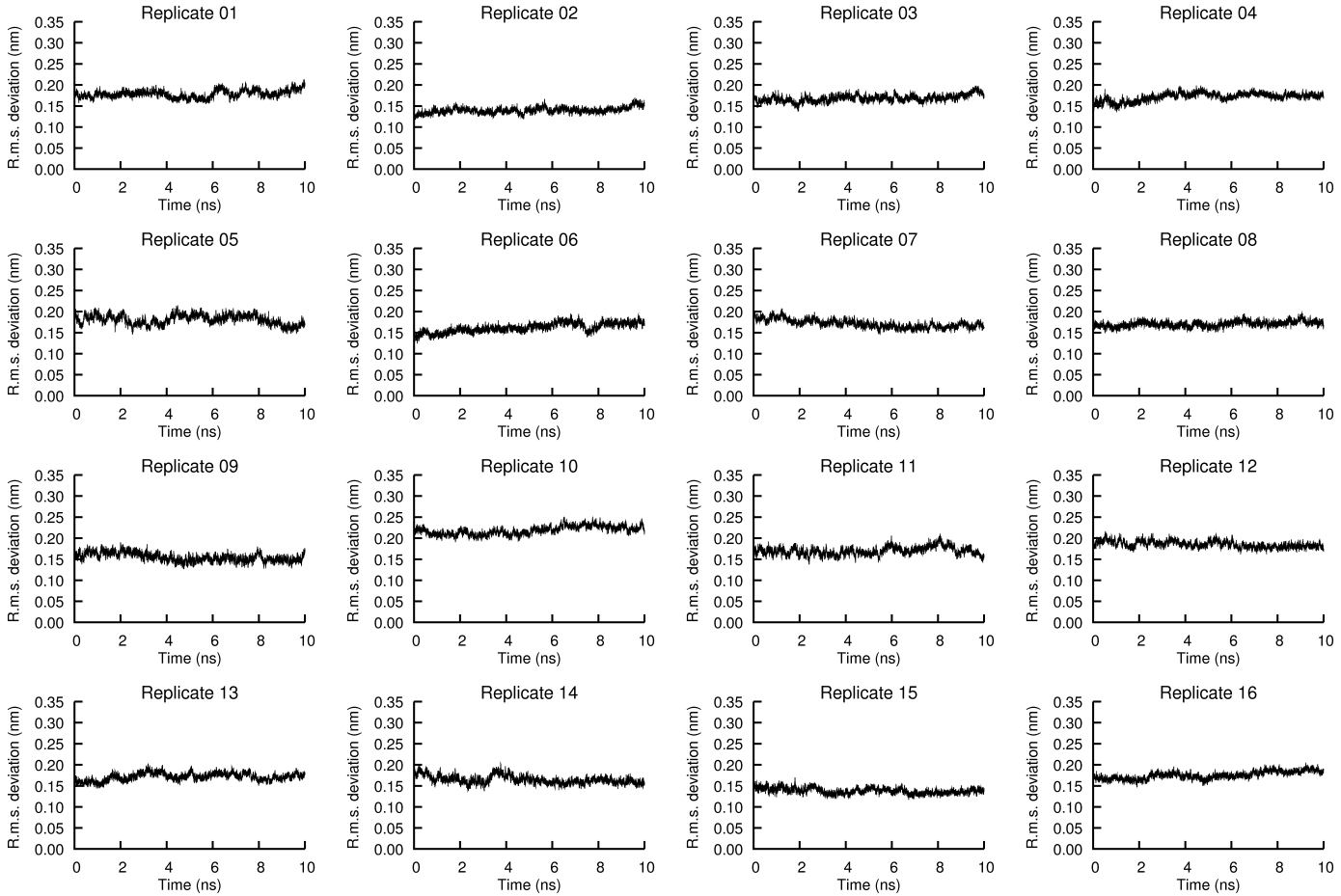


Figure C2.3. Same as figure C.2.1 but for ligand-treated simulations in water. D. Ligand-untreated simulations in water

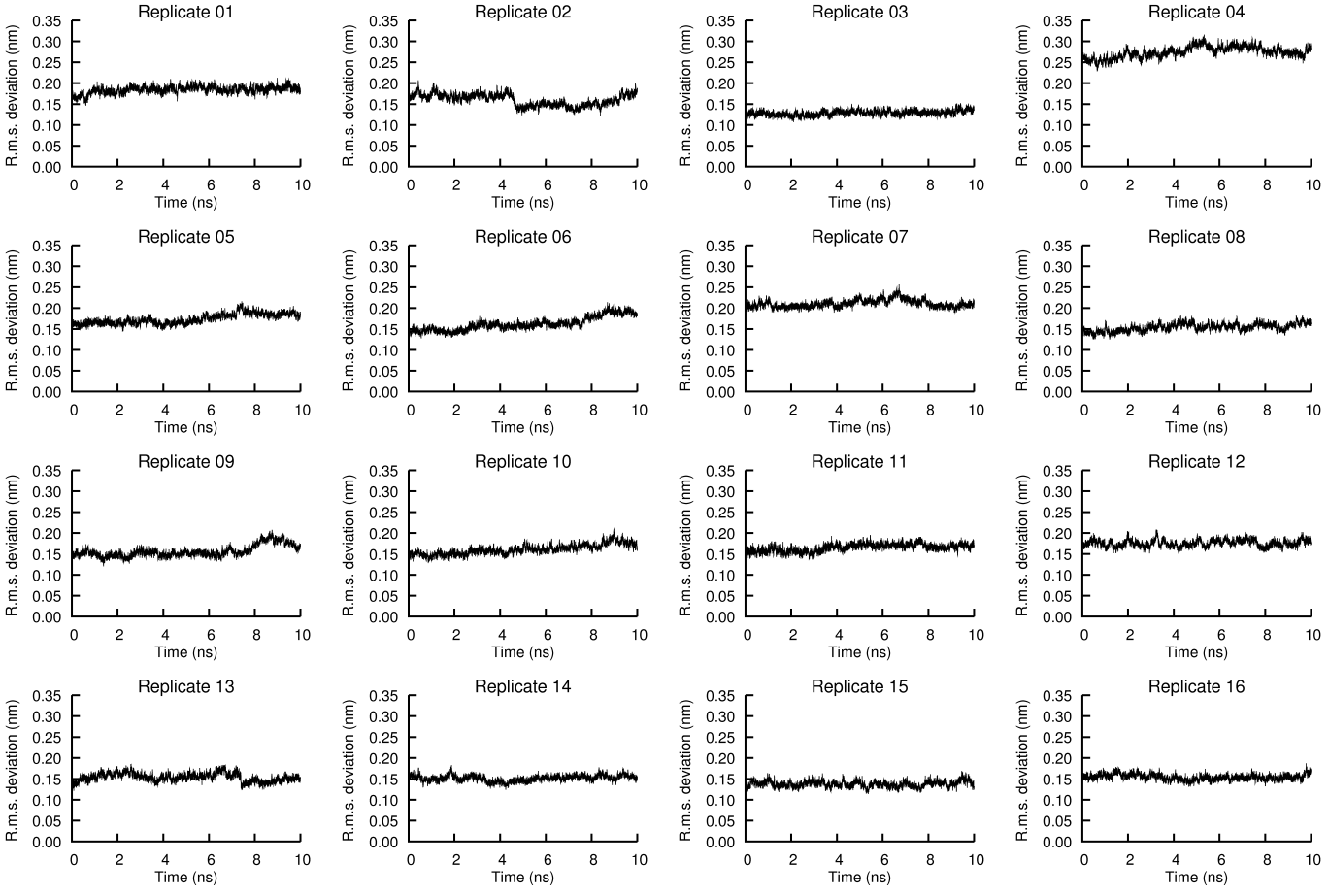


Figure C2.4. Same as figure C.2.1 but for ligand-untreated simulations in water

C.2.2. Behavior of the loops surrounding the S1 pocket

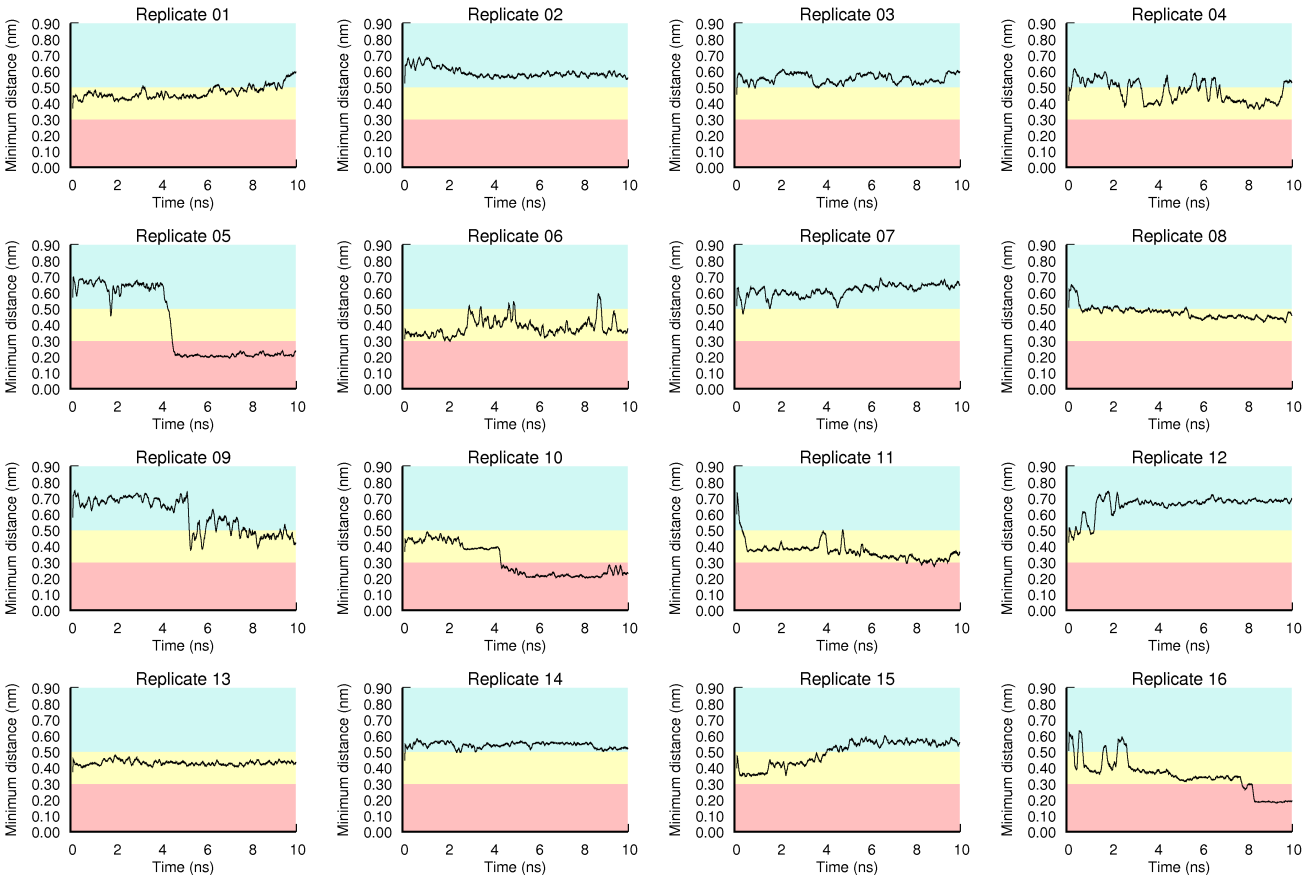


Figure C3.1. Moving averages of the minimum distance between the loops surrounding the S1 pocket in ligand-treated simulations in hexane.

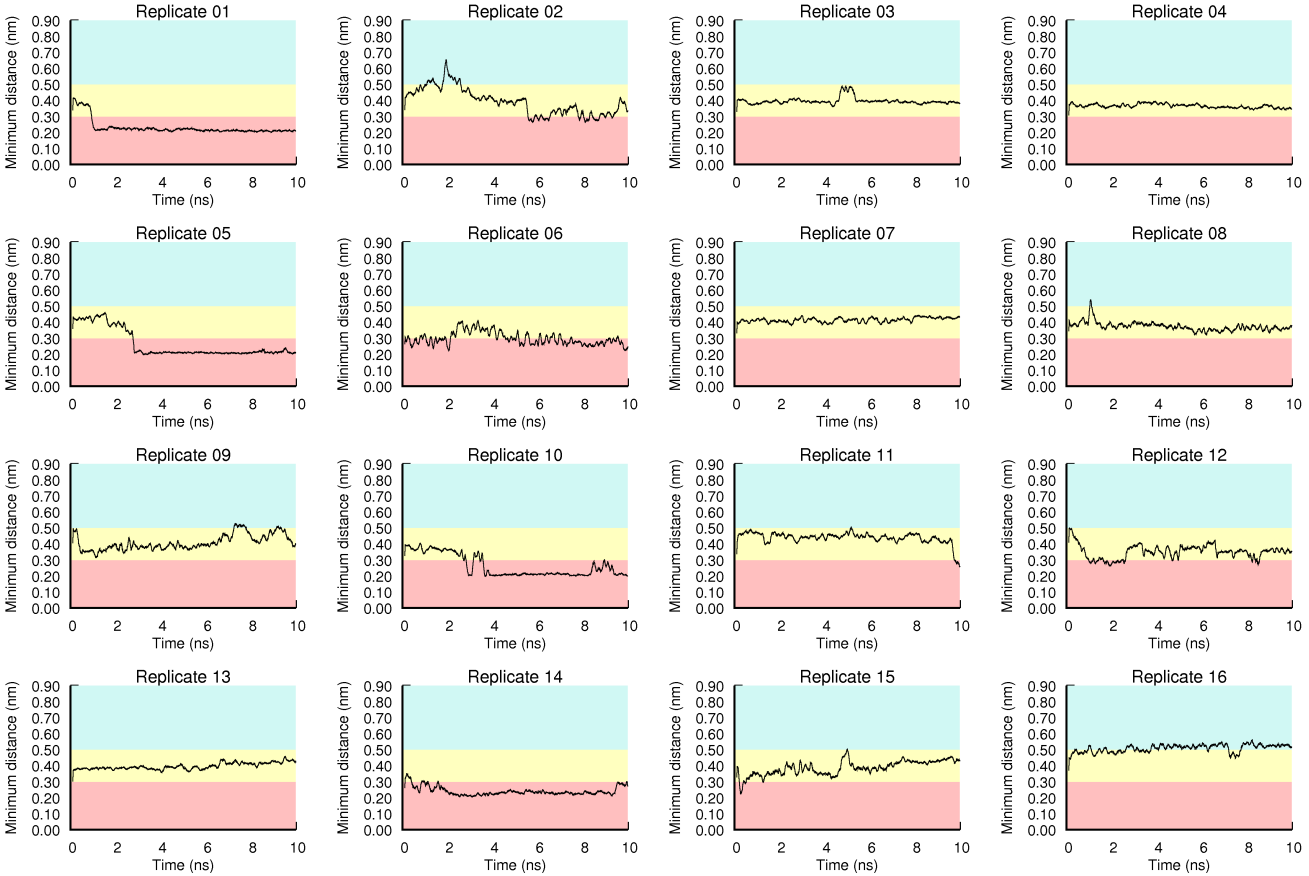


Figure C3.2. Same as fig. C.3.1. but for ligand-untreated simulations in hexane.

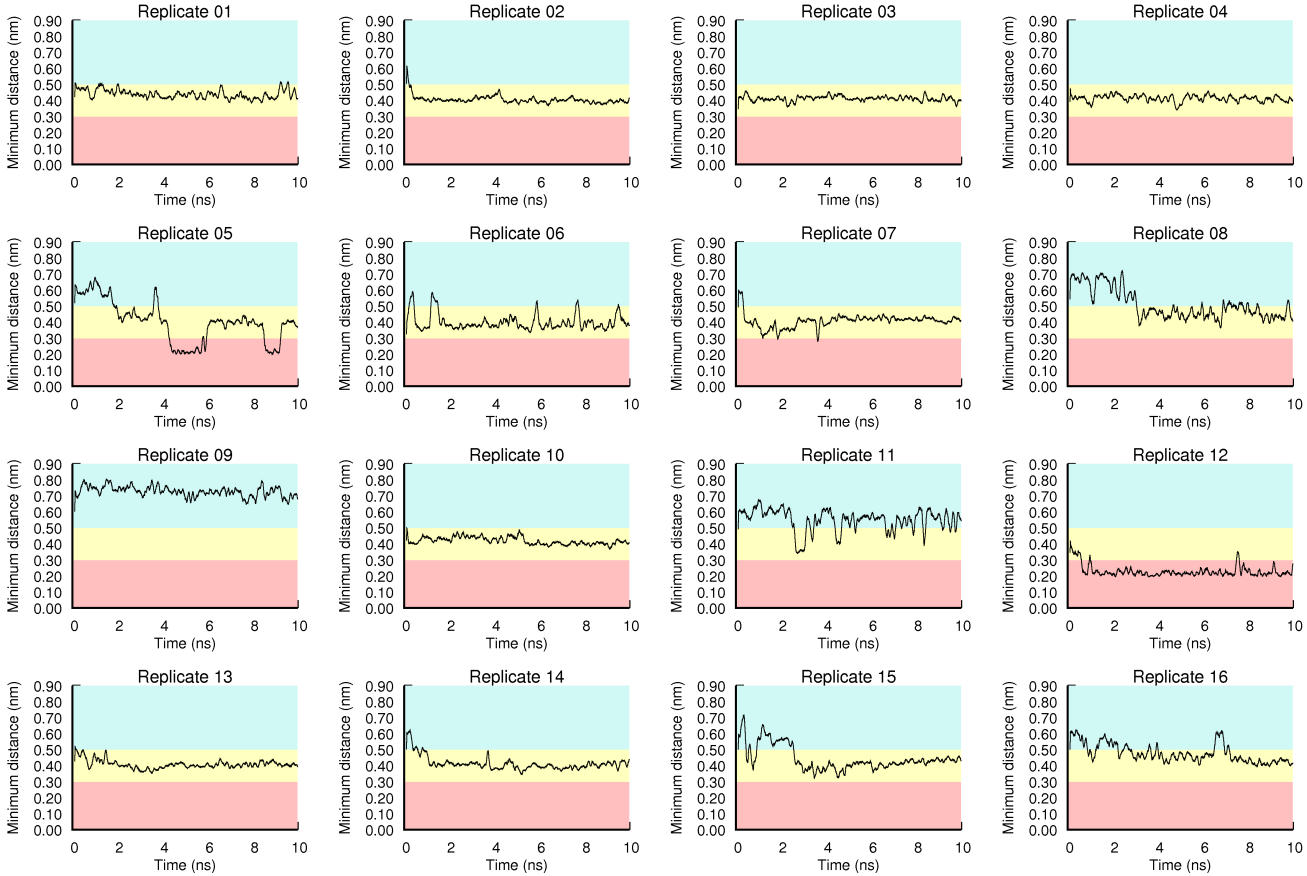


Figure C3.3. Same as fig. C.3.1 but for ligand-treated simulations in water.

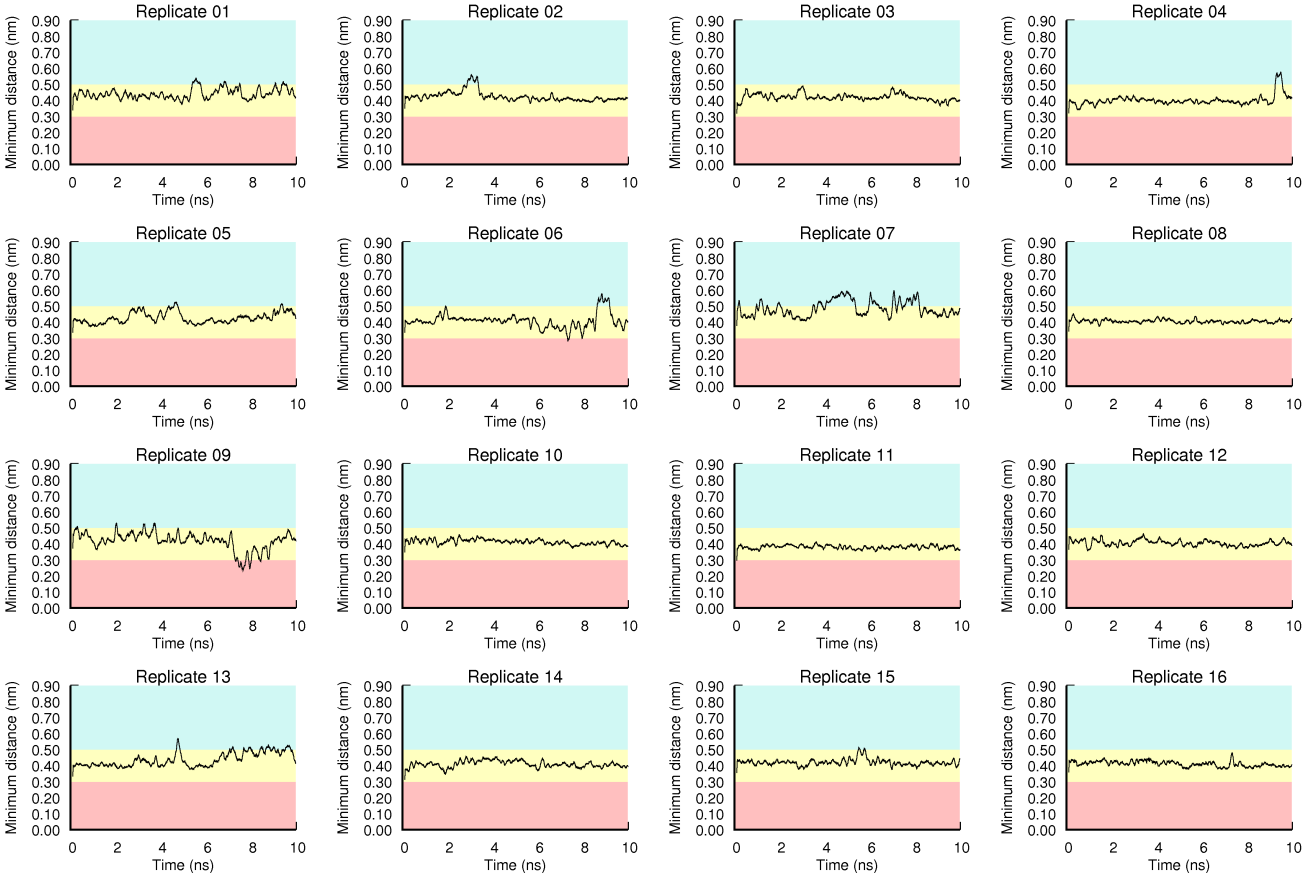


Figure C3.4. Same as fig. C.3.1 but for ligand-untreated simulations in water.

C.3 Movies

Movie_C1. Illustration of the behavior of the S1 pocket in the ligand-treated simulations in hexane. The movie shows the movement of the pocket during the 10 ns of replicate 2 of the ligand-treated simulations in hexane. The frames were recorded with an interval of 100 ps.

Movie_C2. Illustration of the behavior of the S1 pocket in the ligand-treated simulations in water. The movie shows the movement of the pocket during the 10 ns of replicate 2 of the ligand-treated simulations in water. The frames were recorded with an interval of 100 ps.

Bibliography

[1] Soares C. M., Teixeira V. H. and Baptista A. M. (2003), Protein structure and dynamics in nonaqueous solvents: Insights from molecular dynamics simulation studies, Biophys. J. 84: 1628-1641.

[2] Micaelo N. M. and Soares C. M. (2007), Modeling hydration mechanisms of enzymes in nonpolar and polar organic solvents, FEBS J. 274: 2424-2436.

[3] Cianci M., Tomaszewski B., Helliwell J. R. and Halling P. J. (2010), Crystallographic analysis of counterion effects on subtilisin enzymatic action in acetonitrile, J. Am. Chem. Soc. 132: 2293-2300.

[4] Ogino H., Watanabe F., Yamada M., Nakagawa S., Hirose T., Noguchi A., Yasuda M. and Ishikawa H. (1999), Purification and characterization of organic solvent-stable protease from organic solvent-tolerant Pseudomonas aeruginosa PST-01, J. Biosci. Bioeng. 87: 61-68.

[5] Ogino H., Uchiho T., Yokoo J., Kobayashi R., Ichise R. and Ishikawa H. (2001), Role of intermolecular disulfide bonds of the organic solvent-stable PST-01 protease in its organic solvent stability, Appl. Environ. Microbiol. 67: 942-947.

[6] Russell A. J. and Klibanov A. M. (1988), Inhibitor-induced enzyme activation in organic-solvents, J. Biol. Chem. 263: 11624-11626.

[7] Voet D. and Voet J. G. (1995), Biochemistry, 2nd ed., Wiley and Sons: New York.

[8] Fischer E. (1894), Einfluss der configuration auf die wirkung der enzyme, Ber Dtsch Chem Ges 27: 2985–2993.

[9] Haldane J. B. S. (1930), Enzymes, Longmans, Green and Co.: London.

[10] Pauling L. (1946), Molecular architecture and biological reactions, Chemical and Engineering News 24: 1375-1377.

[11] Koshland D. (1958), Application of a theory of enzyme specificity to protein synthesis, Proc. Natl. Acad. Sci. U. S. A. 44: 98-104.

[12] Warshel A. and Levitt M. (1976), Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of Lysozyme, J. Mol. Biol. 103: 227-249.

[13] Warshel A., Sharma P. K., Kato M., Xiang Y., Liu H. B. and Olsson M. H. M. (2006), Electrostatic basis for enzyme catalysis, Chem. Rev. 106: 3210-3235.

[14] Warshel A. (1978), Energetics of enzyme catalysis, Proc. Natl. Acad. Sci. U. S. A. 75: 5250-5254.

[15] Jencks W. P. (1981), On the Attribution and Additivity of Binding-Energies, Proc. Natl. Acad. Sci. U. S. A. 78: 4046-4050.

[16] Wu N., Mo Y. R., Gao J. L. and Pai E. F. (2000), Electrostatic stress in catalysis: Structure and mechanism of the enzyme orotidine monophosphate decarboxylase, Proceedings of the National Academy of Sciences of the United States of America 97: 2017-2022.

[17] Warshel A., Florian J., Strajbl M. and Villa J. (2001), Circe effect versus enzyme preorganization: What can be learned from the structure of the most proficient enzyme?, ChemBioChem 2: 109-111.

[18] Menger F. M. (1992), Analysis of ground-state and transition-state effects in enzyme catalysis, Biochemistry 31: 5368-5373.

[19] Hur S. and Bruice T. C. (2002), The mechanism of catalysis of the chorismate to prephenate reaction by the Escherichia coli mutase enzyme, Proc. Natl. Acad. Sci. U. S. A. 99: 1176-1181.

[20] Hur S. and Bruice T. C. (2003), Comparison of formation of reactive conformers (NACs) for the Claisen rearrangement of chorismate to prephenate in water and in the E-coli mutase: The efficiency of the enzyme catalysis, J. Am. Chem. Soc. 125: 5964-5972.

[21] Hur S. and Bruice T. C. (2003), Enzymes do what is expected (chalcone isomerase versus chorismate mutase), J. Am. Chem. Soc. 125: 1472-1473.

[22] Kamerlin S. C. L. and Warshel A. (2010), At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis?, Proteins: Struct., Funct., Bioinf. 78: 1339-1375.

[23] Nashine V. C., Hammes-Schiffer S. and Benkovic S. J. (2010), Coupled motions in enzyme catalysis, Curr. Opin. Chem. Biol. 14: 644-651.

[24] Benkovic S. J., Hammes G. G. and Hammes-Schiffer S. (2008), Free-energy landscape of enzyme catalysis, Biochemistry 47: 3317-3321.

[25] Benkovic S. J. and Hammes-Schiffer S. (2003), A perspective on enzyme catalysis, Science 301: 1196-1202.

[26] Ma B. and Nussinov R. (2010), Enzyme dynamics point to stepwise conformational selection in catalysis, Curr. Opin. Chem. Biol. 14: 652-659.

[27] Klibanov A. M., Samokhin G. P., Martinek K. and Berezin I. V. (1977), A new approach to preparative enzymatic synthesis, Biotechnol. Bioeng. 19: 1351-1361.

[28] Zaks A. and Klibanov A. M. (1985), Enzyme-catalyzed processes in organic solvents, Proc. Natl. Acad. Sci. U. S. A. 82: 3192-3196.

[29] Klibanov A. M. (1989), Enzymatic catalysis in anhydrous organic solvents, Trends Biochem. Sci. 14: 141-144.

[30] Klibanov A. M. (2001), Improving enzymes by using them in organic solvents, Nature 409: 241-245.

[31] Halling P. J. (2004), What can we learn by studying enzymes in non-aqueous media?, Philos. Trans. R. Soc. Lond. B Biol. Sci. 359: 1287-1296.

[32] Beckman E. J. (2004), Supercritical and near-critical CO2 in green chemical synthesis and processing, J. Supercrit. Fluids 28: 121-191.

[33] Yennawar N. H., Yennawar H. P. and Farber G. K. (1994), X-Ray Crystal-Structure of Gamma-Chymotrypsin in Hexane, Biochemistry 33: 7326-7336.

[34] Yennawar H. P., Yennawar N. H. and Farber G. K. (1995), A structural explanation for enzyme memory in nonaqueous solvents, J. Am. Chem. Soc. 117: 577-585.

[35] Schmitke J. L., Stern L. J. and Klibanov A. M. (1998), Comparison of x-ray crystal structures of an acyl-enzyme intermediate of subtilisin Carlsberg formed in anhydrous acetonitrile and in water, Proc. Natl. Acad. Sci. U. S. A. 95: 12918-12923.

[36] Gag X. G., Maldonado E., Perez-Montfort R., Garza-Ramos G., De Gomez-Puyou M. T., Gomez-Puyou A. and Rodriguez-Romero A. (1999), Crystal structure of triosephosphate isomerase from Trypanosoma cruzi in hexane, Proc. Natl. Acad. Sci. U. S. A. 96: 10062-10067.

[37] Fitzpatrick P. A., Steinmetz A. C. U., Ringe D. and Klibanov A. M. (1993), Enzyme crystal-structure in a neat organic solvent, Proc. Natl. Acad. Sci. U. S. A. 90: 8653-8657.

[38] Fitzpatrick P. A., Ringe D. and Klibanov A. M. (1994), X-Ray crystal-structure of cross-linked subtilisin Carlsberg in water vs acetonitrile, Biochem. Biophys. Res. Commun. 198: 675-681.

[39] Allen K. N., Bellamacina C. R., Ding X. C., Jeffery C. J., Mattos C., Petsko G. A. and Ringe D. (1996), An experimental approach to mapping the binding surfaces of crystalline proteins, J. Phys. Chem. 100: 2605-2611.

[40] Zhu G. Y., Huang Q. C., Wang Z. M., Qian M. X., Jia Y. S. and Tang Y. Q. (1998), X-ray studies on two forms of bovine beta-trypsin crystals in neat cyclohexane, Biochim. Biophys. Acta 1429: 142-150.

[41] Zhu G. Y., Huang Q. C., Zhu Y. S., Li Y. L., Chi C. W. and Tang Y. Q. (2001), X-ray study on an artificial mung bean inhibitor complex with bovine beta-trypsin in neat cyclohexane, Biochim. Biophys. Acta 1546: 98-106.

[42] Mattos C., Bellamacina C. R., Peisach E., Pereira A., Vitkup D., Petsko G. A. and Ringe D. (2006), Multiple solvent crystal structures: Probing binding sites, plasticity and hydration, J. Mol. Biol. 357: 1471-1482.

[43] Matsubara T., Fujita R., Sugiyama S. and Kawashiro K. (2006), Stability of protease in organic solvent: Structural identification by solid-state NMR of lyophilized papain before and after 1-propanol treatment and the corresponding enzymatic activities, Biotechnol. Bioeng. 93: 928-933.

[44] Griebenow K. and Klibanov A. M. (1996), On protein denaturation in aqueous-organic mixtures but not in pure organic solvents, J. Am. Chem. Soc. 118: 11695-11700.

[45] Griebenow K. and Klibanov A. M. (1997), Can conformational changes be responsible for solvent and excipient effects on the catalytic behavior of

subtilisin Carlsberg in organic solvents?, Biotechnology and Bioengineering 53: 351-362.

[46] Griebenow K., Vidal M., Baez C., Santos A. M. and Barletta G. (2001), Nativelike enzyme properties are important for optimum activity in neat organic solvents, J. Am. Chem. Soc. 123: 5380-5381.

[47] Sirotkin V. A. (2005), Effect of dioxane on the structure and hydration-dehydration of alpha-chymotrypsin as measured by FTIR spectroscopy, Biochim. Biophys. Acta 1750: 17-29.

[48] Sirotkin V. A., Zinatullin A. N., Solomonov B. N., Faizullin D. A. and Fedotov V. D. (2001), Calorimetric and Fourier transform infrared spectroscopic study of solid proteins immersed in low water organic solvents, Biochim. Biophys. Acta 1547: 359-369.

[49] Castillo B., Sola R. J., Ferrer A., Barletta G. and Griebenow K. (2008), Effect of PEG modification on subtilisin Carlsberg activity, enantioselectivity, and structural dynamics in 1,4-dioxane, Biotechnol. Bioeng. 99: 9-17.

[50] Simon L. M., Garab M. K. G. and Laczkó I. (2001), Structure and activity of a-chymotrypsin and trypsin in aqueous organic media, Biochem. Biophys. Res. Commun. 280: 1367-1371.

[51] Szabo A., Kotorman M., Laczko I. and Simon L. M. (2006), Spectroscopic studies of stability of papain in aqueous organic solvents, Journal of Molecular Catalysis B-Enzymatic 41: 43-48.

[52] Secundo F., Fiala S., Fraaije M. W., de Gonzalo G., Meli M., Zambianchi F. and Ottolina G. (2011), Effects of water miscible organic solvents on the activity and conformation of the Baeyer-Villiger

monooxygenases from Thermobifida fusca and Acinetobacter calcoaceticus: A comparative study, Biotechnol. Bioeng. 108: 491-499.

[53] Simon L. M., Kotorman M., Szabo A., Nemcsok J. and Laczko I. (2007), The effects of organic solvent/water mixtures on the structure and catalytic activity of porcine pepsin, Process Biochem. 42: 909-912.

[54] Sola R. J. and Griebenow K. (2006), Influence of modulated structural dynamics on the kinetics of alpha-chymotrypsin catalysis - Insights through chemical glycosylation, molecular dynamics and domain motion analysis, FEBS J. 273: 5303-5319.

[55] Klibanov A. M. (1997), Why are enzymes less active in organic solvents than in water?, Trends Biotechnol. 15: 97-101.

[56] Belyaeva E. A., Gra D. V. and Eremeev N. L. (2002), On the mechanism of interaction of organic solvents with the active site of alpha-chymotrypsin, Biochemistry-Moscow 67: 1032-1036.

[57] Zaks A. and Klibanov A. M. (1988), Enzymatic catalysis in nonaqueous solvents, J. Biol. Inorg. Chem. 263: 3194-3201.

[58] Wangikar P. P., Carmichael D., Clark D. S. and Dordick J. S. (1996), Active-site titration of serine proteases in organic solvents, Biotechnol. Bioeng. 50: 329-335.

[59] Schmitke J. L., Wescott C. R. and Klibanov A. M. (1996), The mechanistic dissection of the plunge in enzymatic activity upon transition from water to anhydrous solvents, J. Am. Chem. Soc. 118: 3360-3365.

[60] Borole A. P., Cheng C. L. and Davison B. H. (2004), Substrate desolvation as a governing factor in enzymatic transformations of PAHs in aqueous-acetonitrile mixtures, Biotechnol. Prog. 20: 1251-1254.

[61] Xu Z. F., Affleck R., Wangikar P., Suzawa V., Dordick J. S. and Clark D. S. (1994), Transition-state stabilization of subtilisins in organic media, Biotechnol. Bioeng. 43: 515-520.

[62] Affleck R., Xu Z. F., Suzawa V., Focht K., Clark D. S. and Dordick J. S. (1992), Enzymatic catalysis and dynamics in low-water environments, J. Am. Chem. Soc. 89: 1100-1104.

[63] Hudson E. P., Eppler R. K., Beaudoin J. M., Dordick J. S., Reimer J. A. and Clark D. S. (2009), Active-site motions and polarity enhance catalytic turnover of hydrated subtilisin dissolved in organic solvents, J. Am. Chem. Soc. 131: 4294-4300.

[64] Broos J., Visser A. J. W. G., Engbersen J. F. J., Verboom W., van Hoek A. and Reinhoudt D. N. (1995), Flexibility of enzymes suspended in organic solvents probed by time-resolved fluorescence anisotropy. Evidence that enzyme activity and enantioselectivity are directly related to enzyme flexibility, J. Am. Chem. Soc. 117: 12657-12663.

[65] Valivety R. H., Halling P. J. and Macrae A. R. (1992), Reaction rate with suspended lipase catalyst shows similar dependence on water activity in different organic solvents, Biochim. Biophys. Acta. 1118: 218-222.

[66] Hudson E. P., Eppler R. K. and Clark D. S. (2005), Biocatalysis in semi-aqueous and nearly anhydrous conditions, Curr. Opin. Biotechnol. 16: 637-643.

[67] Wescott C. R. and Klibanov A. M. (1993), Solvent variation inverts substrate specificity of an enzyme, J. Am. Chem. Soc. 115: 1629-1631.

[68] Carrea G., Ottolina G. and Riva S. (1995), Role of solvents in the control of enzyme selectivity in organic media, Trends Biotechnol. 13: 63-70.

[69] Margolin A. L., Tai D. F. and Klibanov A. M. (1987), Incorporation of D-amino acids into peptides via enzymatic condensation in organic solvents, J. Am. Chem. Soc. 109: 7885-7887.

[70] Sakurai A., Margolin A. L., Russel A. J. and Klibanov A. M. (1988), Control of enzyme enantioselectivity by the reaction medium, J. Am. Chem. Soc. 110: 7236.

[71] Fitzpatrick P. A. and Klibanov A. M. (1991), How can the solvent affect enzyme enantioselectivity?, J. Am. Chem. Soc. 113: 3166-3171.

[72] Rariy R. V. and Klibanov A. M. (2000), On the relationship between enzymatic enantioselectivity in organic solvents and enzyme flexibility, Biocatal. Biotransform. 18: 401-407.

[73] Broos J. (2002), Impact of the enzyme flexibility on the enzyme enantio-selectivity in organic media towards specific and non-specific substrates, Biocatal. Biotransform. 20: 291-295.

[74] Micaelo N. M., Teixeira V. H., Baptista A. M. and Soares C. M. (2005), Water dependent properties of cutinase in nonaqueous solvents: A computational study of enantioselectivity, Biophys. J. 89: 999-1008.

[75] Ueji S., Fujino R., Okubo N., Miyazawa T., Kurita S., Kitadani M. and Muromatsu A. (1992), Solvent-induced inversion of enantioselectivity in lipase-catalyzed esterification of 2-phenoxypropionic acids, Biotechnol. Lett. 14: 163-168.

[76] Nakamura K., Takebe Y., Kitayama T. and Ohno A. (1991), Effect of solvent structure on enantioselectivity of lipase-catalyzed transesterification, Tetrahedron Lett. 32: 4941-4944.

[77] Bovara R., Carrea G., Ferrara L. and Riva S. (1991), Resolution of (+/-)-trans-sobrerol by lipase PS-catalyzed transesterification and effects of organic-solvents on enantioselectivity, Tetrahedron-Asymm. 2: 931-938.

[78] Tawaki S. and Klibanov A. M. (1992), Inversion of enzyme enantioselectivity mediated by the solvent, J. Am. Chem. Soc. 114: 1882-1884.

[79] Secundo F., Riva S. and Carrea G. (1992), Effects of medium and of reaction conditions on the enantioselectivity of lipases in organic-solvents and possible rationales, Tetrahedron-Asymm. 3: 267-280.

[80] Fitzpatrick P. A., Ringe D. and Klibanov A. M. (1992), Computer-assisted modeling of subtilisin enantioselectivity in organic-solvents, Biotechnol. Bioeng. 40: 735-742.

[81] Wescott C. R., Noritomi H. and Klibanov A. M. (1996), Rational control of enzymatic enantioselectivity through solvation thermodynamics, J. Am. Chem. Soc. 118: 10365-10370.

[82] Anthonsen T. and Jongejan J. A. (1997), Solvent effect in lipase-catalyzed racemate resolution, Methods Enzymol. 286: 473-495.

[83] Colombo G., Ottolina G., Carrea G., Bernardi A. and Scolastico C. (1998), Application of structure-based thermodynamic calculations to the rationalization of the enantioselectivity of subtilisin in organic solvents, Tetrahedron-Asymm. 9: 1205-1214.

[84] Griebenow K., Laureano Y. D., Santos A. M., Clemente I. M., Rodriguez L., Vidal M. W. and Barletta G. (1999), Improved enzyme activity and enantioselectivity in organic solvents by methyl-beta-cyclodextrin, J. Am. Chem. Soc. 121: 8157-8163.

[85] Ulbert O., Frater T., Belafi-Bako K. and Gubicza L. (2004), Enhanced enantioselectivity of Candida rugosa lipase in ionic liquids as compared to organic solvents, Journal of Molecular Catalysis B-Enzymatic 31: 39-45.

[86] Hofmeister F. (1888), Zur Lehre von der Wirkung der Salze, Arch. Exp. Pathol. Pharmakol. 24: 247-260.

[87] Pazhang M., Khajeh K., Ranjbar B. and Hosseinkhani S. (2006), Effects of water-miscible solvents and polyhydroxy compounds on the structure and enzymatic activity of thermolysin, J. Biotechnol. 127: 45-53.

[88] Omta A. W., Kropman M. F., Woutersen S. and Bakker H. J. (2003), Negligible effect of ions on the hydrogen-bond structure in liquid water, Science 301: 347-349.

[89] Soper A. K. and Weckstrom K. (2006), Ion solvation and water structure in potassium halide aqueous solutions, Biophys. Chem. 124: 180-191.

[90] Batchelor J. D., Olteanu A., Tripathy A. and Pielak G. J. (2004), Impact of protein denaturants and stabilizers on water structure, J. Am. Chem. Soc. 126: 1958-1961.

[91] Collins K. D. (1997), Charge density-dependent strength of hydration and biological structure, Biophys. J. 72: 65-76.

[92] Collins K. D. (2004), Ions from the Hofmeister series and osmolytes: effects on proteins in solution and in the crystallization process, Methods 34: 300-311.

[93] Collins K. D. (2006), Ion hydration: Implications for cellular function, polyelectrolytes, and protein crystallization, Biophys. Chem. 119: 271-281.

[94] Vrbka L., Vondrasek J., Jagoda-Cwiklik B., Vácha R. and Jungwirth P. (2006), Quantification and rationalization of the higher affinity of sodium over potassium to protein surfaces, Proc. Natl. Acad. Sci. U. S. A. 103: 15440-15444.

[95] Hess B. and van der Vegt N. F. A. (2009), Cation specific binding with protein surface charges, Proc. Natl. Acad. Sci. U. S. A. 106: 13296-13300.

[96] Uejio J. S., Schwartz C. P., Duffin A. M., Drisdell W. S., Cohen R. C. and Saykally R. J. (2008), Characterization of selective binding of alkali cations with carboxylate by x-ray absorption spectroscopy of liquid microjets, Proc. Natl. Acad. Sci. U. S. A. 105: 6809-6812.

[97] Aziz E. F., Ottosson N., Eisebitt S., Eberhardt W., Jagoda-Cwiklik B., Vacha R., Jungwirth P. and Winter B. (2008), Cation-specific interactions with carboxylate in amino acid and acetate aqueous solutions: X-ray absorption and ab initio calculations, J. Phys. Chem. B 112: 12567-12570.

[98] Halling P. J. (2000), Biocatalysis in low-water media: understanding effects of reaction conditions, Curr. Opin. Chem. Biol. 4: 74-80.

[99] Khmelnitsky Y. L., Welch S. H., Clark D. S. and Dordick J. S. (1994), Salts dramatically enhance activity of enzymes suspended in organic solvents, J. Am. Chem. Soc. 116: 2647-2648.

[100] Ru M. T., Dordick J. S., Reimer J. A. and Clark D. S. (1999), Optimizing the salt-induced activation of enzymes in organic solvents: Effects of lyophilization time and water content, Biotechnol. Bioeng. 63: 233-241.

[101] Ru M. T., Hirokane S. Y., Lo A. S., Dordick J. S., Reimer J. A. and Clark D. S. (2000), On the salt-induced activation of lyophilized enzymes in

organic solvents: Effect of salt kosmotropicity on enzyme activity, J. Am. Chem. Soc. 122: 1565-1571.

[102] Lindsay J. P., Clark D. S. and Dordick J. S. (2004), Combinatorial formulation of biocatalyst preparations for increased activity in organic solvents: Salt activation of penicillin amidase, Biotechnol. Bioeng. 85: 553-560.

[103] Yu H. W., Chen H., Yang Y. Y. and Ching C. B. (2005), Effect of salts on activity, stability and enantioselectivity of Candida rugosa lipase in isooctane, J. Mol. Catal. B: Enzym. 35: 28-32.

[104] Bilanicova D., Salis A., Ninham B. W. and Monduzzi M. (2008), Specific anion effects on enzymatic activity in nonaqueous media, J. Phys. Chem. B 112: 12066-12072.

[105] Morgan J. A. and Clark D. S. (2004), Salt-activation of nonhydrolase enzymes for use in organic solvents, Biotechnol. Bioeng. 85: 456-459.

[106] Altreuter D. H., Dordick J. S. and Clark D. S. (2002), Nonaqueous biocatalytic synthesis of new cytotoxic doxorubicin derivatives: Exploiting unexpected differences in the regioselectivity of salt-activated and solubilized subtilisin, J. Am. Chem. Soc. 124: 1871-1876.

[107] Eppler R. K., Komor R. S., Huynh J., Dordick J. S., Reimer J. A. and Clark D. S. (2006), Water dynamics and salt-activation of enzymes in organic media: Mechanistic implications revealed by NMR spectroscopy, Proc. Natl. Acad. Sci. U. S. A. 103: 5706-5710.

[108] Triantafyllou A. O., Wehtje E., Adlercreutz P. and Mattiasson B. (1997), How do additives affect enzyme activity and stability in nonaqueous media?, Biotechnol. Bioeng. 54: 67-76.

[109] Guinn R. M., Skerker P. S., Kavanaugh P. and Clark D. S. (1991), Activity and flexibility of alcohol-dehydrogenase in organic solvents, Biotechnol. Bioeng. 37: 303-308.

[110] Zacharis E., Halling P. J. and Rees D. G. (1999), Volatile buffers can override the "pH memory" of subtilisin catalysis in organic media, Proc. Natl. Acad. Sci. U. S. A. 96: 1201-1205.

[111] Stahl M., Mansson M. O. and Mosbach K. (1990), The synthesis of a D-amino acid ester in an organic media with a-chymotrypsin modified by a bio-imprinting procedure, Biotechnol. Lett. 12: 161-166.

[112] Staahl M., Jeppsson-Wistrand U., Maansson M. O. and Mosbach K. (1991), Induced stereo- and substrate selectivity of bioimprinted alpha-chymotrypsin in anhydrous organic media, J. Am. Chem. Soc. 113: 9366-9368.

[113] Peibker F. and Fischer L. (1999), Crosslinking of imprinted proteases to maintain a tailor-made substrate selectivity in aqueous solutions, Bioinorgan. Med. Chem. 7: 2231-2237.

[114] Yan J. Y., Yan Y. J., Yang J. K., Xu L. and Liu Y. (2009), Combined strategy for preparation of a bioimprinted Geotrichum sp lipase biocatalyst effective in non-aqueous media, Process Biochem. 44: 1128-1132.

[115] Hartsough D. S. and Merz K. M. (1993), Protein dynamics and solvation in aqueous and nonaqueous environments, J. Am. Chem. Soc. 115: 6529-6537.

[116] Norin M., Haeffner F., Hult K. and Edholm O. (1994), Molecular dynamics simulations of enzyme surrounded by vaccum, water, or a hydrophobic solvent, Biophys. J. 67: 548-559.

[117] Toba S., Hartsough D. S. and Merz K. M. (1996), Solvation and dynamics of chymotrypsin in hexane, J. Am. Chem. Soc. 118: 6490-6498.

[118] Roccatano D., Wong T. S., Schwaneberg U. and Zacharias M. (2005), Structural and dynamic properties of cytochrome P450BM-3 in pure water and in a dimethylsulfoxide/water mixture, Biopolymers 78: 259-267.

[119] Trodler P. and Pleiss J. (2008), Modeling structure and flexibility of Candida antarctica lipase B in organic solvents, BMC Struct. Biol. 8: 1-10.

[120] Cruz A., Ramirez E., Santana A., Barletta G. and Lopez G. E. (2009), Molecular dynamic study of subtilisin Carlsberg in aqueous and nonaqueous solvents, Mol. Simul. 35: 205-212.

[121] Diaz-Vergara N. and Pineiro A. (2008), Molecular dynamics study of triosephosphate isomerase from Trypanosoma cruzi in water/decane mixtures, J. Phys. Chem. B 112: 3529-3539.

[122] Hartsough D. S. and Merz K. M. (1992), Protein flexibility in aqueous and nonaqueous solutions, J. Am. Chem. Soc. 114: 10113-10116.

[123] Roccatano D., Wong T. S., Schwaneberg U. and Zacharias M. (2006), Toward understanding the inactivation mechanism of monooxygenase P450BM-3 by organic cosolvents: A molecular dynamics simulation study, Biopolymers 83: 467-476.

[124] Zaks A. and Klibanov A. M. (1988), The effect of water on enzyme action in organic media, J. Biol. Inorg. Chem. 263: 8017-8021.

[125] Affleck R., Haynes C. A. and Clark D. S. (1992), Solvent dielectric effects on protein dynamics, Proc. Natl. Acad. Sci. U. S. A. 89: 5167-5170.

[126] Partridge J., Dennison P. R., Moore B. D. and Halling P. J. (1998), Activity and mobility of subtilisin in low water organic media: hydration is more important than solvent dielectric, Biochim. Biophys. Acta 1386: 79-89.

[127] Yang L., Dordick J. S. and Garde S. (2004), Hydration of enzyme in nonaqueous media is consistent with solvent dependence of its activity, Biophys. J. 87: 812-821.

[128] Branco R. J. F., Graber M., Denis V. and Pleiss J. (2009), Molecular mechanism of the hydration of candida antarctica lipase B in the gas phase: Water adsorption isotherms and molecular dynamics simulations, ChemBioChem 10: 2913-2919.

[129] Wedberg R., Abildskov J. and Peters G. H. (2012), Protein dynamics in organic media at varying water activity studied by molecular dynamics simulation, J. Phys. Chem. B 116: 2575-2585.

[130] Lousa D., Cianci M., Helliwell J. R., Halling P. J., Baptista A. M. and Soares C. M. (2012), Interaction of counterions with subtilisin in acetonitrile: Insights from molecular dynamics simulations, J. Phys. Chem. B 116: 5838-5848.

[131] Colombo G., Toba S. and Merz K. M. (1999), Rationalization of the enantioselectivity of subtilisin in DMF, J. Am. Chem. Soc. 121: 3486-3493.

[132] Warshel A., Naray-Szabo G., Sussman F. and Hwang J. K. (1989), How do serine proteases really work?, Biochemistry 28: 3629-3637.

[133] Li C., Tan T. W., Zhang H. Y. and Feng W. (2010), Analysis of the conformational stability and activity of Candida antarctica lipase B in organic solvents. Insight from molecular dynamics and quantum mechanics simulations, J. Biol. Chem. 285: 28434-28441.

[134] Jensen M. O., Jensen T. R., Kjaer K., Bjornholm T., Mouritsen O. G. and Peters G. H. (2002), Orientation and conformation of a lipase at an interface studied by molecular dynamics simulations, Biophys. J. 83: 98-111.

[135] James J. J., Lakshmi B. S., Seshasayee A. S. N. and Gautam P. (2007), Activation of Candida rugosa lipase at alkane-aqueous interfaces: A molecular dynamics study, FEBS Lett. 581: 4377-4383.

[136] Cherukuvada S. L., Seshasayee A. S. N., Raghunathan K., Anishetty S. and Pennathur G. (2005), Evidence of a double-lid movement in Pseudomonas aeruginosa lipase: Insights from molecular dynamics simulations, PLoS Comput. Biol. 1: 182-189.

[137] Trodler P., Schmid R. D. and Pleiss J. (2009), Modeling of solvent-dependent conformational transitions in Burkholderia cepacia lipase, BMC Struct. Biol. 9: 38-50.

[138] Santini S., Crowet J. M., Thomas A., Paquot M., Vandenbol M., Thonart P., Wathelet J. P., Blecker C., Lognay G., Brasseur R., Lins L. and Charloteaux B. (2009), Study of Thermomyces lanuginosa lipase in the presence of tributyrylglycerol and water, Biophys. J. 96: 4814-4825.

[139] Micaelo N. M. and Soares C. M. (2008), Protein structure and dynamics in ionic liquids. Insights from molecular dynamics simulation studies, J. Phys. Chem. B 112: 2566-2572.

[140] Micaelo N. M., Baptista A. M. and Soares C. M. (2006), Parametrization of 1-butyl-3-methylimidazolium hexafluorophosphate/nitrate ionic liquid for the GROMOS force field, J. Phys. Chem. B 110: 14444-14451.

[141] Klahn M., Lim G. S., Seduraman A. and Wu P. (2011), On the different roles of anions and cations in the solvation of enzymes in ionic liquids, Phys. Chem. Chem. Phys. 13: 1649-1662.

[142] Klahn M., Lim G. S. and Wu P. (2011), How ion properties determine the stability of a lipase enzyme in ionic liquids: A molecular dynamics study, Phys. Chem. Chem. Phys. 13: 18647-18660.

[143] Silveira R. L., Martinez J., Skaf M. S. and Martinez L. (2012), Enzyme microheterogeneous hydration and stabilization in supercritical carbon dioxide, J. Phys. Chem. B 116: 5671-5678.

[144] Housaindokht M. R., Bozorgmehr M. R. and Monhemi H. (2012), Structural behavior of Candida antarctica lipase B in water and supercritical carbon dioxide: A molecular dynamic simulation study, J. Supercrit. Fluids 63: 180-186.

[145] Monhemi H., Housaindokht M. R., Bozorgmehr M. R. and Googheri M. S. S. (2012), Enzyme is stabilized by a protection layer of ionic liquids in supercritical CO2: Insights from molecular dynamic simulation, J. Supercrit. Fluids 69: 1-7.

[146] Khmelnitsky Y. L. and Rich J. O. (1999), Biocatalysis in nonaqueous solvents, Curr. Opin. Chem. Biol. 3: 47-53.

[147] van Gunsteren W. F., Bakowies D., Baron R., Chandrasekhar I., Christen M., Daura X., Gee P., Geerke D. P., Glattli A., Hunenberger P. H., Kastenholz M. A., Ostenbrink C., Schenk M., Trzesniak D., van der Vegt N. F. A. and Yu H. B. (2006), Biomolecular modeling: Goals, problems, perspectives, Angewandte Chemie-International Edition 45: 4064-4092.

[148] Frenkel D. and Smit B. (2002), Understanding molecular simulation: From algorithms to applications, 2nd ed., Academic Press: San Diego.

[149] Allen M. P. and Tildesley D. J. (1987), Computer Simulation of Liquids, 1st ed., Oxford University Press: New York.

[150] Leach A. R. (1996), Molecular Modelling: Principles and Applications, 1st ed., Addison Wesley Longman: Essex.

[151] Oostenbrink C., Villa A., Mark A. E. and Van Gunsteren W. F. (2004), A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6, J. Comput. Chem. 25: 1656-1676.

[152] Darden T., York D. and Pedersen L. (1993), Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems, J. Chem. Phys. 98: 10089-10092.

[153] Tironi I. G., Sperb R., Smith P. E. and van Gunsteren W. F. (1995), A generalized reaction field method for molecular-dynamics simulations, J. Chem. Phys. 102: 5451-5459.

[154] Berendsen H. J. C., van der Spoel, D., van Drunen, R. (1995), GROMACS: A message-passing parallel molecular-dynamics implementation, Comput. Phys. Commun. 91: 43-56.

[155] McCammon J. A., Gelin B. R. and Karplus M. (1977), Dynamics of folded proteins, Nature 267: 585-590.

[156] Dror R. O., Dirks R. M., Grossman J. P., Xu H. F. and Shaw D. E. (2012), Biomolecular simulation: A computational microscope for molecular biology, Annu. Rev. Biophys. 41: 429-452.

[157] Karplus M. and McCammon J. A. (2002), Molecular dynamics simulations of biomolecules, Nat. Struct. Biol. 9: 646-652.

[158] Ryckaert J., Ciccotti G. and Berendsen H. J. C. (1977), Numerical integration of the cartesian equations of motion of a system with constrains: Molecular dynamics of n-alkanes, J. Comput. Phys. 23: 327-341.

[159] Hess B., Bekker H., Berendsen H. J. C. and Fraaije J. G. E. M. (1997), LINCS: A linear constraint solver for molecular simulations, J. Comput. Chem. 18: 1463-1472.

[160] Hill T. L. (1960), An introduction to statistical thermodynamics, Dover Publications: New York.

[161] Berendsen H. J. C., Postma J. P. M., van Gunsteren W. F., Dinola A. and Haak J. R. (1984), Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81: 3684-3690.

[162] Zwanzig R. W. (1954), High-temperature equation of state by a pertubation method. I. Nonpolar gases, J. Chem. Phys. 22: 1420-1426.

[163] Beveridge D. L. and Dicapua F. M. (1989), Free-energy via molecular simulation - applications to chemical and biomolecular systems, Annu Rev Biophys Bio 18: 431-492.

[164] Sousa S. F., Fernandes P. A. and Ramos M. J. (2006), Protein-ligand docking: Current status and future challenges, Proteins-Structure Function and Bioinformatics 65: 15-26.

[165] Morris G. M., Goodsell D. S., Halliday R. S., Huey R., Hart W. E., Belew R. K. and Olson A. J. (1998), Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, J. Comput. Chem. 19: 1639-1662.

[166] Morris G. M., Huey R., Lindstrom W., Sanner M. F., Belew R. K., Goodsell D. S. and Olson A. J. (2009), AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, J. Comput. Chem. 30: 2785-2791.

[167] Huey R., Morris G. M., Olson A. J. and Goodsell D. S. (2007), A semiempirical free energy force field with charge-based desolvation, J. Comput. Chem. 28: 1145-1152.

[168] Mehler E. L. and Eichele G. (1984), Electrostatic effects in water-accessible regions of proteins, Biochemistry 23: 3887-3891.

[169] Bashford D. (2004), Macroscopic electrostatic models for protonation states in proteins, Front. Biosci. 9: 1082-1099.

[170] Bashford D. and Karplus M. (1990), pKa's of ionizable groups in proteins: Atomic detail from a continuum electrostatic model, Biochemistry 29: 10219-10225.

[171] Baptista A. M., Martel P. J. and Soares C. M. (1999), Simulation of electron-proton coupling with a Monte Carlo method: Application to cytochrome c3 using continuum electrostatics, Biophys. J. 76: 2978-2998.

[172] Baptista A. M. and Soares C. M. (2001), Some theoretical and computational aspects of the inclusion of proton isomerism in the protonation equilibrium of proteins, J. Phys. Chem. B 105: 293-309.

[173] Deshpande A., Nimsadkar S. and Mande S. C. (2005), Effect of alcohols on protein hydration: crystallographic analysis of hen egg-white lysozyme in the presence of alcohols, Acta Cryst. D Biol. Crystallogr. 61: 1005-1008.

[174] English A. C., Done S. H., Caves L. S. D., Groom C. R. and Hubbard R. E. (1999), Locating interaction sites on proteins: The crystal structure of thermolysin soaked in 2% to 100% isopropanol, Proteins Struct. Funct. Bioinformat. 37: 628-640.

[175] Gao X. G., Maldonado E., Perez-Montfort R., Garza-Ramos G., De Gomez-Puyou M. T., Gomes-Puyou A. and Rodriguez-Romero A. (1999),

Crystal structure of triosephosphate isomerase from Trypanosoma cruzi in hexane, Proc. Natl. Acad. Sci. U. S. A. 96: 10062-10067.

[176] Bell G., Halling P. J., Moore B. D., Partridge J. and Rees D. G. (1995), Biocatalyst behaviour in low-water systems, Trends Biotechnol. 13: 468-473.

[177] Bell G., Janssen A. E. M. and Halling P. J. (1997), Water activity fails to predict critical hydration level for enzyme activity in polar organic solvents: Interconversion of water concentrations and activities, Enzyme. Microb. Technol. 20: 471-477.

[178] Halling P. J. (1990), High-affinity binding of water by proteins is similar in air and in organic solvents, Biochim. Biophys. Acta 1040: 225-228.

[179] Ke T. and Klibanov A. M. (1998), Insights into the solvent dependence of chymotryptic prochiral selectivity, J. Am. Chem. Soc. 120: 4259-4263.

[180] Ke T., Tidor B. and Klibanov A. M. (1998), Molecular-modeling calculations of enzymatic enantioselectivity taking hydration into account, Biotechnol. Bioeng. 57: 741-746.

[181] Ke T., Wescott C. R. and Klibanov A. M. (1996), Prediction of the solvent dependence of enzymatic prochiral selectivity by means of structure-based thermodynamic calculations, J. Am. Chem. Soc. 118: 3366-3374.

[182] Ke T. and Klibanov A. M. (1998), On enzymatic activity in organic solvents as a function of enzyme history, Biotechnol. Bioeng. 57: 746-750.

[183] Klibanov A. M. (1995), Enzyme memory - What is remembered and why, Nature 374: 596-596.

[184] Lousa D., Baptista A. M. and Soares C. M. (2011), Structural determinants of ligand imprinting: A molecular dynamics simulation study of subtilisin in aqueous and apolar solvents, Protein Sci. 20: 379-386.

[185] Serdakowski A. L. and Dordick J. S. (2008), Enzyme activation for organic solvents made easy, Trends Biotechnol. 26: 48-54.

[186] Der A. (2008), Salts, interfacial water and protein conformation, Biotechnol. Biotec. Eq. 22: 629-633.

[187] Zhao H. (2005), Effect of ions and other compatible solutes on enzyme activity, and its implication for biocatalysis using ionic liquids, J. Mol. Catal. B: Enzym. 37: 16-25.

[188] Schmitke J. L., Stern L. J. and Klibanov A. M. (1997), The crystal structure of subtilisin Carlsberg in anhydrous dioxane and its comparison with those in water and acetonitrile, Proc. Natl. Acad. Sci. U. S. A. 94: 4250-4255.

[189] Hess B., Holm C. and van der Vegt N. (2006), Modeling multibody effects in ionic solutions with a concentration dependent dielectric permittivity, Phys. Rev. Lett. 96: 164509.

[190] Hess B., Holm C. and van der Vegt N. (2006), Osmotic coefficients of atomistic NaCl (aq) force fields, J. Chem. Phys. 124: 147801.

[191] Hess B., Kutzner C., van der Spoel D. and Lindahl E. (2008), GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation, J. Chem. Theory Comput. 4: 435-447.

[192] Hermans J., Berendsen H. J. C., van Gunsteren W. F. and Postma J. P. M. (1984), A consistent empirical potential for water-protein interactions, Biopolymers 23: 1513-1518.

[193] Gee P. J. and Van Gunsteren W. F. (2006), Acetonitrile revisited: a molecular dynamics study of the liquid phase, Mol. Phys. 104: 477-483.

[194] Reif M. M. and Hunenberger P. H. (2011), Computation of methodology-independent single-ion solvation properties from molecular

simulations. IV. Optimized Lennard-Jones interaction parameter sets for the alkali and halide ions in water, J. Chem. Phys. 134: 144104.

[195] Miyamoto S. and Kollman P. A. (1992), SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models, J. Comput. Chem. 13: 952-962.

[196] van Gunsteren W. F. and Berendsen H. J. C. (1990), Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry, Angew. Chem., Int. Ed. 29: 992-1023.

[197] Barker J. A. and Watts R. O. (1973), Monte-Carlo studies of dielectric properties of water-like models, Mol. Phys. 26: 789-792.

[198] Smith P. E. and Pettitt B. M. (1994), Modeling solvent in biomolecular systems, J. Phys. Chem. 98: 9700-9711.

[199] Barthel J., Kleebauer M. and Buchner R. (1995), Dielectric-relaxation of electrolyte-solutions in acetonitrile, J. Solution Chem. 24: 1-17.

[200] Kabsch W. and Sander C. (1983), Dictionary of protein secondary structure - Pattern-recognition of hydrogen-bonded and geometrical features, Biopolymers 22: 2577-2637.

[201] Coetzee J. F. and Campion J. J. (1967), Solute-solvent interactions II. Relative activities of anions in acetonitrile and water, J. Am. Chem. Soc. 89: 2517-2521.

[202] Coetzee J. F. and Campion J. J. (1967), Solute-solvent interactions I. Evaluations of relative activities of reference cations in acetonitrile and water, J. Am. Chem. Soc. 89: 2513-2517.

[203] Davidson W. R. and Kebarle P. (1976), Ionic solvation by aprotic-solvents - Gas-phase solvation of alkali ions by acetonitrile, J. Am. Chem. Soc. 98: 6125-6133.

[204] Yamdagni R. and Kebarle P. (1972), Solvation of negative-ions by protic and aprotic solvents - Gas-phase solvation of halide ions by acetonitrile, J. Am. Chem. Soc. 94: 2940-2943.

[205] Marcus Y. (1983), Thermodynamic functions of transfer of single ions from water to non-aqueous and mixed-solvents: Part 1 - Gibbs free-energies of transfer to non-aqueous solvents, Pure Appl. Chem. 55: 977-1021.

[206] Kelly C. P., Cramer C. J. and Truhlar D. G. (2007), Single-ion solvation free energies and the normal hydrogen electrode potential in methanol, acetonitrile, and dimethyl sulfoxide, J. Phys. Chem. B 111: 408-422.

[207] Bashford D. (1997), An object-oriented programming suite for electrostatic effects in biological molecules. In Scientific Computing in Object-Oriented Parallel Environments, Y. Ishikawa, Oldehoeft, R. R., Reynders, J. V. W., Tholburn, M., Ed. Springer: Berlin, Vol. 1343, pp 233-240.

[208] DeLano W. L. (2003), The PyMOL Molecular Graphics System, DeLano Scientific: Palo Alto.

[209] Kutt A., Rodima T., Saame J., Raamat E., Maemets V., Kaljurand I., Koppel I. A., Garlyauskayte R. Y., Yagupolskii Y. L., Yagupolskii L. M., Bernhardt E., Willner H. and Leito I. (2011), Equilibrium acidities of superacids, J. Org. Chem. 76: 391-395.

[210] Zaks A. and Klibanov A. M. (1986), Substrate-specificity of enzymes in organic-solvents vs. water is reversed, J. Am. Chem. Soc. 108: 2767-2768.

[211] Bevc S., Konc J., Stojan J., Hodoscek M., Penca M., Praprotnik M. and Janezic D. (2011), ENZO: A web tool for derivation and evaluation of kinetic models of enzyme catalyzed reactions, PLoS One 6: e22265.

[212] Johnson K. A. (2009), Fitting enzyme kinetic data with kintek global kinetic explorer, Methods Enzymol. 467: 601-626.

[213] Kuzmic P. (2009), DynaFit - A software package for enzymology, Methods Enzymol. 467: 247-280.

[214] Walsh R., Martin E. and Darvesh S. (2010), A method to describe enzyme-catalyzed reactions by combining steady state and time course enzyme kinetic parameters, Biochim. Biophys. Acta 1800: 1-5.

[215] Kamerlin S. C. L. and Warshel A. (2010), The EVB as a quantitative tool for formulating simulations and analyzing biological and chemical reactions, Faraday Discuss. 145: 71-106.

[216] McGeagh J. D., Ranaghan K. E. and Mulholland A. J. (2011), Protein dynamics and enzyme catalysis: Insights from simulations, Biochim. Biophys. Acta 1814: 1077-1092.

[217] Acevedo O. and Jorgensen W. L. (2010), Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions, Acc. Chem. Res. 43: 142-151.

[218] Roccatano D. (2008), Computer simulations study of biomolecules in non-aqueous or cosolvent/water mixture solutions, Curr. Protein Pept. Sci. 9: 407-426.

[219] Gupta M. N. and Roy I. (2004), Enzymes in organic media - Forms, functions and applications, Eur. J. Biochem. 271: 2575-2583.

[220] Iyer P. V. and Ananthanarayan L. (2008), Enzyme stability and stabilization - Aqueous and non-aqueous environment, Process Biochem. 43: 1019-1032.

[221] Gupta A. and Khare S. K. (2009), Enzymes from solvent-tolerant microbes: Useful biocatalysts for non-aqueous enzymology, Crit. Rev. Biotechnol. 29: 44-54.

[222] Doukyu N. and Ogino H. (2010), Organic solvent-tolerant enzymes, Biochem. Eng. J. 48: 270-282.

[223] Jensen S. E., Fecycz I. T. and Campbell J. N. (1980), Nutritional factors controlling exocellular protease production by Pseudomonas aeruginosa, J. Bacteriol. 144: 844-847.

[224] Morihara K., Tsuzuki H. and Oda K. (1979), Protease and elastase of Pseudomonas aeruginosa: inactivation of human-plasma alpha 1-proteinase inhibitor, Infect. Immun. 24: 188-193.

[225] Pavlovskis O. R. and Wretlind B. (1979), Assessment of protease (elastase) as a Pseudomonas aeruginosa virulence factor in experimental mouse burn infection, Infect. Immun. 24: 181-187.

[226] Morihara K. (1964), Production of elastase and proteinase by Pseudomonas aeruginosa, J. Bacteriol. 88: 745-757.

[227] Heck L. W., Morihara K., Mcrae W. B. and Miller E. J. (1986), Specific cleavage of human type-III and IV collagens by Pseudomonas aeruginosa elastase, Infect. Immun. 51: 115-118.

[228] Holder I. A. and Wheeler R. (1984), Experimental studies of the pathogenesis of infections owing to Pseudomonas aeruginosa: elastase, an Igg Protease, Can. J. Microbiol. 30: 1118-1124.

[229] Matsubara H., Sasaki R., Singer A. and Jukes T. H. (1966), Specific nature of hydrolysis of insulin and tobacco mosaic virus protein by thermolysin, Arch. Biochem. Biophys. 115: 324-331.

[230] Holland D. R., Hausrath A. C., Juers D. and Matthews B. W. (1995), Structural analysis of zinc substitutions in the active site of thermolysin, Protein Sci. 4: 1955-1965.

[231] Thayer M. M., Flaherty K. M. and Mckay D. B. (1991), 3-Dimensional structure of the elastase of Pseudomonas aeruginosa at 1.5 A resolution, J. Biol. Chem. 266: 2864-2871.

[232] Akerlof G. (1932), Dielectric constants of some organic solvent-water mixtures at various temperatures, J. Am. Chem. Soc. 54: 4125-4139.

[233] Rahman R. N. Z. R. A., Salleh A., Basri M. and Wong C. F. (2011), Role of alpha-helical structure in organic solvent-activated homodimer of elastase strain K, Int. .J. Mol. Sci. 12: 5797-5814.

[234] Pazhang M., Khajeh K., Ranjbar B. and Hosseinkhani S. (2006), Effects of water-miscible solvents and polyhydroxy compounds on the structure and enzymatic activity of thermolysin, J. Biotechnol. 127: 45-53.

[235] van Aalten D. M. F., Amadei A., Linssen A. B. M., V.G.H. E., Vriend G. and Berendsen H. J. C. (1995), The essential dynamics of thermolysin: Confirmation of the hinge-bending motion and comparison of simulations in vacuum and water, Proteins Struct. Funct. Bioinformat. 22: 45-54.

[236] Hausrath A. C. and Matthews B. W. (2002), Thermolysin in the absence of substrate has an open conformation, Acta Crystallogr. Sect. D Biol. Crystallogr. 58: 1002-1007.

[237] Holland D. R., Tronrud D. E., Pley H. W., Flaherty K. M., Stark W., Jansonius J. N., Mckay D. B. and Matthews B. W. (1992), Structural comparison suggests that thermolysin and related neutral proteases undergo hinge-bending motion during catalysis, Biochemistry 31: 11310-11316.

[238] Ogino H., Gemba Y., Yutori Y., Doukyu N., Ishimi K. and Ishikawa H. (2007), Stabilities and conformational transitions of various proteases in the presence of an organic solvent, Biotechnol. Prog. 23: 155-161.

[239] Gupta M. N., Batra R., Tyagi R. and Sharma A. (1997), Polarity index: The guiding solvent parameter for enzyme stability in aqueous-organic cosolvent mixtures, Biotechnol. Prog. 13: 284-288.

[240] Derewenda Z. S. (1994), Structure and function of lipases, Adv. Protein Chem. 45: 1-52.

[241] Zaks A. and Klibanov A. M. (1984), Enzymatic catalysis in organic media at 100-degrees C., Science 224: 1249-1251.

[242] Rich J. O. and Dordick J. S. (1997), Controlling subtilisin activity and selectivity in organic media by imprinting with nucleophilic substrates, J. Am. Chem. Soc. 119: 3245-3252.

[243] Stoll V. S., Eger B. T., Hynes R. C., Martichonok V., Jones J. B. and Pai E. F. (1998), Differences in binding modes of enantiomers of 1-acetamido boronic acid based protease inhibitors: Crystal structures of gamma-chymotrypsin and subtilisin Carlsberg complexes, Biochemistry 37: 451-462.

[244] Bashford D. and Gerwert K. (1992), Electrostatic calculations of the pka values of ionizable groups in bacteriorhodopsin, J. Mol. Biol. 224: 473-486.

[245] Daura X., Mark A. E. and van Gunsteren W. F. (1998), Parametrization of aliphatic CHn united atoms of GROMOS96 force field, J. Comput. Chem. 19: 535-547.

[246] Oliveira A. S. F., Teixeira V. H., Baptista A. M. and Soares C. M. (2005), Reorganization and conformational changes in the reduction of tetraheme cytochromes, Biophys. J. 89: 3919-3930.

[247] Hutcheon G. A., Parker M. C. and Moore B. D. (2000), Measuring enzyme motility in organic media using novel H-D exchange methodology, Biotechnol. Bioeng. 70: 262-269.

[248] Mehler E. L., Solmajer, T. (1991), Electrostatic effects in proteins: comparison of dielectric and charge models, Protein Eng. 4: 903-910.

[249] Solmajer T., Meheler, EL (1991), Electrostatic screening in molecular dynamics simulations, Protein Eng. 4: 911-917.

ITQB-UNL | Av. da República, 2780-157 Oeiras, Portugal
Tel (+351) 214 469 100 | Fax (+351) 214 411 277

www.itqb.unl.pt