Investigating functional aspects and inhibition of the human carnosinase CN1 enzyme by computational methods

Von der Fakultät fur Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Dipl.-Chem. Matic Pavlin

aus Ljubljana, Slowenien

Berichter:	Univ.-Prof. Dr. Paolo Carloni
	Univ.-Prof. Dr. rer. nat. Marc Spehr
Tag der mündlichen Prüfung:	13.12.2016

I hereby declare that I have created this work completely on my own and used no other sources or tools than the ones listed, and that I have marked any citations accordingly.

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstsändig verfasst und keine anderen als die angegebenen Quellen und Hilfsmitel benutzt sowie Zitate kenntlich gemacht habe.

Matic Pavlin

Aachen, July 2016

Bibliographic information

Part of this thesis has been published:

Pavlin, M., Rossetti, G., De Vivo, M., Carloni, P. Carnosine and Homocarnosine Degradation Mechanisms by the Human Carnosinase Enzyme CN1: Insights from Multiscale Simulations. *Biochemistry* **2016**, 55(19), 2772-2784.

Abstract

Alzheimer's and Parkinson's disease (AD and PD) are two of the most common neurodegenerative disorders that, together, affect more than 50 million people worldwide. The pathophysiological picture of these diseases is still unclear. This makes their pharmacological treatment more difficult. One of the molecules with possible therapeutic potential is dipeptide carnosine (β-alanine-L-histidine), which has shown neuroprotective effects in animal models of these diseases. Unfortunately, the transferability of carnosine's neuroprotective ability to humans is limited. This is due to the human serum carnosinase enzyme (hCN1), absent in animal serum. In humans, hCN1 was found to have an age-dependent equilibrium between its monomeric and dimeric form, with the latter being the most active and predominant form in adults. In fact, in its active form, hCN1 hydrolyzes carnosine in humans, thus preventing its beneficial pharmacological effects. Intriguingly, homocarnosine (γ-aminobutyryl-L-histidine), a close analog of the substrate carnosinase, inhibits hCN1. The reasons for such an inhibitory effect of hCN1 by homocarnosine were still unknown.

The final scope of this thesis was to design novel non-hydrolysable carnosine derivatives that could prove beneficial for the treatment of AD and PD. To achieve this, I have used computational techniques to (i) characterize structural properties of carnosine *vs* homocarnosine, (ii) predict the structure of their Michaelis complex with hCN1 and (iii) design novel non-hydrolysable carnosine derivatives able to block hCN1 function.

i. By using classical molecular dynamics, I have shown that both ligands are predominantly present in their zwitter-ionic form and mainly adopt extended conformations. The quality of the simulations was validated against experimental ^{1}H and ^{13}C chemical shifts.
ii. By using molecular docking in conjunction with classical molecular dynamics and hybrid quantum mechanics / molecular mechanics simulations, I predicted for the first time the structure of hCN1 Michaelis complex with either carnosine or homocarnosine. Based on the obtained data, I have explained why carnosine is a hCN1's substrate, while homocarnosine leads to hCN1 inhibition. Additionally, I propose a mechanistic explanation for the more elevated activity of hCN1 in its dimeric form.
iii. Finally, based on the information obtained from the structures of Michaelis complexes, I screened a few large libraries of small molecules compounds to search novel non-hydrolysable carnosine-like derivatives. These compounds may be used to start drug discovery projects for the design of novel therapeutic agents to treat neurodegenerative diseases. A few selected compounds are currently being tested in the group of Professor Enrico Rizzarelli, University of Catnia, Italy.

Anyone looking for goal will remain empty when it will be reached, but whoever finds a way, will always carry the goal inside.

Nejc Zaplotnik

Acknowledgment

First of all, I would like to thank my supervisor Prof. Paolo Carloni for his invaluable help and support in the past three years. He patiently guided me on my first steps into scientific world and supported me in my professional growth.

I would like to express my sincere gratitude to Jun. Prof. Giulia Rossetti for all constructive discussions, help and great support.

Special thanks goes to Dr. Emiliano Ippoliti for many fruitful discussions and Dr. Marco De Vivo for his collaboration and support.

I would like to thank the whole Computational Biophysics group for their help, scientific discussions and many nice moments that we spent together.

I also thank Ms. Elke George, Ms. Elisa Polese and all the administration staff for their patience and help with administrative problems.

I would like to thank my friends, Anna, thanks for your friendship; Thomas and Luca P., thanks for helping me to live according to the *mens sana in corpore sano* principle; Clara, Andrea, Menica, Massimo, Gloria, Fabrizio, Luca M., Rodrigo and all others for all the amazing moments that we spent together. Thank you guys!

I would like to thank Dr. Janez Mavri for the support and many useful advices.

German Research School for Simulation Sciences GmbH is acknowledged for the financial support.

Finally, I would like to thank my family, especially my parents, for their patience, full support and trust in me in all these years.

Contents

1 Introduction 1
2 Biological introduction 4
2.1 Carnosine 4
2.2 Carnosinase enzymes 11
3 Methods 16
3.1 Molecular dynamics 16
3.2 Protein-ligand docking 45
3.3 Ligand-based virtual screening 52
4 Structural characterization of carnosine and homocarnosine 53
4.1 Introduction 53
4.2 Methods 54
4.3 Results and discussion 57
4.4 Conclusions 64
5 Michaelis complexes of carnosine and homocarnosine with hCN1 65
5.1 Introduction 65
5.2 Methods 70
5.3 Results 75
5.4 Discussion 89
5.5 Conclussions 90
6 Ligand screening for non-hydrolyzable carnosine derivatives 92
6.1 Introduction 92
6.2 Methods 93
6.3 Results and discussion 94
6.4 Conclusions 99
7 Conclusions 100
A1 Appendix 1 – Proposed mechanism of dipeptide hydrolysis 102
A2 Appendix 2 – Supporting information 103
QM/MM simulations of **CT-I•**hCN1β and **HT-I•**hCN1β Michaelis complexes ... 103
8 References 127
List of figures 150
List of tables 156

1 Introduction

The presence of the endogenous dipeptide carnosine (β-alanine-L-histidine) in mammals was discovered more than a century ago [86]. Carnosine is mainly present in skeletal muscles but it can be also found in central nervous system [30, 62, 122]. The physiological role of the peptide is still not fully understood [120]. We know that the dipeptide may be beneficial for many diseases, such as diabetes, cancer, ischemia damage and ocular diseases [1, 13, 15, 43, 64, 140, 207, 224, 259, 265]. Moreover, in animal models carnosine features also neuroprotective properties [52, 93, 203, 240]. These characteristics are likely to be linked to its capability of acting as a mobile pH buffer, as a free readicals and a reactive oxygen species scavenger, and as a heavy metal ions chelator [61, 123, 206].[a] Unfortunately, however, its rapid enzymatic hydrolysis in human body hampers most of its therapeutic applications [248]. This is performed by tissue and serum carnosinases [142, 145, 233]. The former one has a broader range of substrates, while the latter is highly specific for carnosine.[b] In the human bloodstream, serum carnosinase hampers most of carnosine's beneficial effects. The enzyme exists as a dimer and as a monomer and the equilibrium between the two forms is age-dependent, with the dimeric form shown to be the more active one in adults [110, 142, 197]. Structural information is available only for the dimeric form of

[a] Carnosine is indeed commercially available as dietary supplement [216].

[b] Interestingly, the serum carnosinase is present only in primates and not in all mammals, unlike tissue carnosinase [110].

the enzyme (PDB ID: 3DLJ[c]): it shows that each monomeric unit has a di-zinc co-catalytic active site located on the catalytic domain. The active site is fully conserved among the other members of M20 family of metallopeptidases [208], consisting of two zinc ions, each coordinated by acidic and histidine residue and bridging aspartate and water molecule that coordinates both ions [25] (Fig. 1.1; see Appendix 1 for details of the currently proposed enzymatic reaction mechanism). However, the molecular basis of the different activity between the two physiological forms is not known. Inhibiting this enzyme is a very promising strategy for the treatment of pathological conditions using carnosine itself. Instead of using carnosinase inhibitors, an alternative beneficial strategy would be to use non-hydrolysable carnosine derivates. Unfortunately, despite the effort to design and to use non-hydrolysable carnosine derivatives [22-24, 27, 248, 249], none of them is currently in clinical trials. New strategies to identify potent and efficient ligands should therefore be sought.

Here I have used computational methods to address key issues regarding the human serum carnosinase activity. First, I have investigated structural differences between the monomeric and dimeric forms of the enzyme and described how are these differences reflected to the substrate binding. QM/MM simulations show that upon dimerization, the loop region of one monomeric unit is entangled in the active site of the cognate monomeric unit (Fig. 1.1), leading to the formation of an extra hydrogen bond with the peptide oxygen of the substrate. This affects the polarization of the peptide bond and makes it more prone to hydrolysis. This may be consistent with experimental evidences showing that the dimer is more active with respect to the monomer [197].

As a next step, I have designed new ligands that fit nicely in the enzymatic catalytic site, capitalizing on the information gained by the simulations. First, I have compared the binding modes of the substrate carnosine with that of a competitive inhibitor, which has much slower rate of hydrolysis [197, 233]. I focus here on carnosine's homologue homocarnosine (γ-aminobutyryl-L-histidine) [197, 233]. This has allowed me to clarify the molecular basis for the fact that carnosine is a substrate of the enzyme and homocarnosine is, instead, a competitive inhibitor. In addition, my calculations show that homocarnosine forms more hydrogen bonds with the enzyme than carnosine. This may be consistent with its lower experimental K_m value [233]. Based on this structural information, I could design novel carnosine derivatives that might retain most of the carnosine's beneficial properties, while being less prone to be hydrolyzed. Structure-

[c] Article was never published.

based virtual screening techniques lead to the determination of 25 potential candidates, currently under enzymatic essays in the laboratory of Professor Enrico Rizzarelli at University of Catania, Italy.

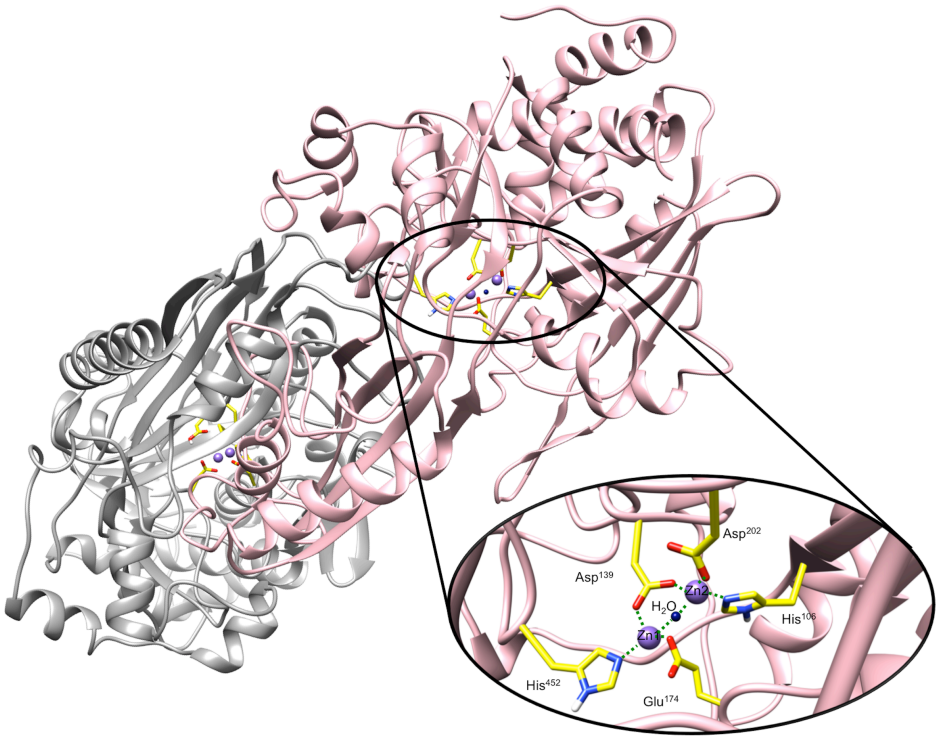


Figure 1.1: hCN1 homodimer structure shown in ribbons representation, where two monomers are shown in pink and grey, respectively. The active site residues are colored in yellow and Zn^{2+} ions in dark violet. Oxygen atoms are colored in red, nitrogen atoms in blue and bridging water molecule in dark blue. Coordination bonds are shown as dashed green lines. Non-polar hydrogen atoms are not shown due to clarity reasons.

2 Biological introduction

2.1 Carnosine

L-carnosine is an endogenous dipeptide, formed from β-alanine and L-histidine.[d] It was discovered already in 1900 by Gulewitsch and Amiradžibi in Liebig's meat extract [86], and named accordingly after the tissue, where it was discovered.[e] Its most common natural derivatives are anserine, ophidine -known also as balenine (both are methylated on the imidazole ring)- and homocarnosine (β-alanine is substituted by γ-aminobutyric acid, GABA). The other naturally occurring carnosine derivatives include acetylcarnosine (β-alanine is acetylated) and carcinine (L-histidine is replaced by histamine) [30]. Schematic representation of carnosine and its connection with its derivatives is shown in Figure 2.1.

[d] D-carnosine is its stereoisomer, however only L-carnosine is an endogenous dipeptide and was more extensively studied in the presented work. Therefore, throughout the thesis term carnosine is used for L-carnosine.

[e] The root *carn-* in Latin means "meat" or "flesh".

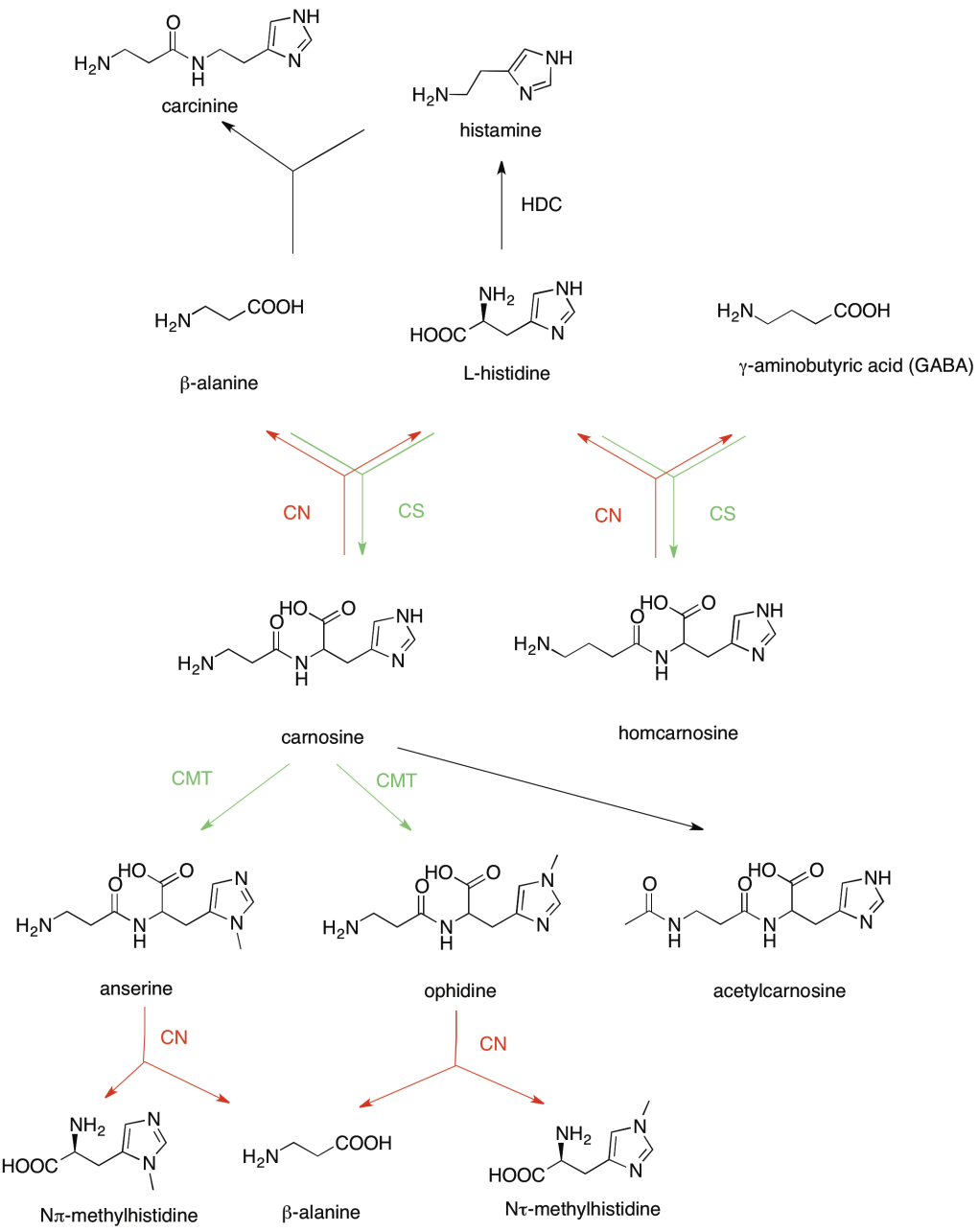


Figure 2.1: Metabolic pathways of carnosine and its naturally occurring derivatives. The enzymatic reactions that are further discussed in the thesis are colored in green (synthesis) and in red (degradation). All compounds are shown in their non-ionized state. CN stands for carnosinase, CS for carnosine synthase, CMT for carnosine-*N*-methyltransferase and HDC for L-histidine decarboxylase.

2.1.1 Distribution of carnosine and its natural analogs

Carnosine and its natural derivatives have been detected only in muscles of vertebrates, while they are almost completely absent in other living organisms. Normally in amphibians, reptiles, birds and fishes only one of the analogs is present, namely carnosine in amphibians, anserine in birds and ophidine in reptiles, while in fishes either anserine, ophidine or just free L-histidine were detected [2, 3, 10, 11, 29, 30, 45, 53, 113]. In contrast with other vertebrates mammal's muscles contain carnosine and either anserine or ophidine [10, 11, 29, 30, 45, 48, 53, 66, 91, 109, 121, 169, 205, 227]. While in the majority of land-living mammals studied[f] anserine was detected, ophidine is usually found in marine mammals [10, 11, 30, 45, 48, 53, 66, 91, 109, 169, 205, 227, 262]. There are exceptions such as some odd- and even-hoed ungulates, whose muscles contain carnosine together with both of its methylated analogs, though carnosine is the predominant compound [30]. Interestingly however is the fact that *Homo sapiens* is one of only two mammalian species where only carnosine is present in their muscles [30].

Carnosine is mainly found in skeletal muscles (5-8 mmol/L in wet weight or 20-30 mmol/kg in dry weight) [62], where over 99% of all body carnosine is present. The only other location, with a milimolar order of magnitude of concentration, is the olfactory bulb [122], Other tissues like brain regions and body fluids also contain carnosine, however in 10- to 1,000 fold lower concentrations compared to the carnosine's concentration in skeletal muscles [122]. These data comes from mice, since, unfortunately, information about human carnosine distribution is so far lacking [30]. Due to the high concentrations and ease of measurement the carnosine content in human was measured only in the skeletal muscle [157, 184]. Homocarnosine on the other hand is mainly found in the human brain, whereas its concentration elsewhere is negligible [198].

Since the scope of the present thesis lies in understanding the role of carnosine in human body and its possible therapeutic applications in the treatment of various

[f] Not all mammalian orders have been investigated so far for their carnosine content.

neurodegenerative diseases, further discussion on carnosine and its analogs is mainly limited on their distribution and role in human body.

2.1.2 Biochemical properties of carnosine and its role in human body

Carnosine in human body is mainly found in its zwitterionic form [30]. Due to the presence of the imidazole ring, which has pK_a value of 6.72 [251], it is a suitable compound to act as a pH buffer. Even though it provides a relatively small contribution to the total buffering capacity (between 4.5% and 9.4%) [157], its advantage lies in the mobility. In contrast with proteins, which are localized buffers, carnosine is dissolved in the myoplasm and therefore has a role of a mobile buffer [30, 119].

Another important characteristic of carnosine is its metal ion chelation ability. Studies were reported for its affinity toward Cu^{2+}, Co^{2+}, Ni^{2+}, Cd^{2+} and Zn^{2+} [16, 30, 63, 71, 146, 168, 237, 238], though the most biologically relevant are complexes with Cu^{2+} and Zn^{2+} as their (dis)homeostasis is important in many neurological disorders, such as Alzheimer's disease (AD) and Parkinson's disease (PD) [125, 175]. Both ions form at least four different complexes with carnosine *in vitro* and their presence is pH-dependent [237, 238]. For human biology the most important complexes are those formed at physiological pH, at which both, Cu^{2+} and Zn^{2+} ions, form two different complexes with carnosine, monomeric and dimeric ones [237, 238]. However, due to the rather low concentrations of free metals in human body, the formation of dimeric complexes *in vivo* is unlikely [38, 223]. The metal chelation property of carnosine could be used in therapeutic purposes aimed in the regulation of Zn^{2+} and Cu^{2+} ions homeostasis in order to prevent neurodegenerative pathological states [190]. Additionally, the imidazole ring of carnosine can also act as a free radical (*e.g.* peroxyl radicals) and reactive oxygen species (ROS) scavenger [61, 129, 266]. Carnosine's ability to scavange free radicals and ROS and to chelate heavy metal ions in turn all contribute to its antioxidant activity [31-33].

2.1.3 Effects of carnosine on different pathological conditions and its possible therapeutic use

Most of the research on human diseases is performed in animal models. However, in contrast with humans, where carnosine is rapidly degraded by human serum carnosinase enzyme, animal serum shows only low carnosinase activity and therefore many carnosine's beneficial and therapeutic effects could be at least partially attributed to the elevated carnosine concentrations in their bloodstream [69, 260]. Carnosine showed beneficial effects in diabetes (carnosine increased diet showed hypoglycemic effect [140]), ocular diseases (it inhibited α-crystallin aggregation [13, 15, 224, 259]),

ischemia damage (mainly through its antioxidant activity [1, 64, 207, 265]), cancer (it is able to deplete glycolysis intermediates by its carbonyl quenching ability and therefore reduce the generation of ATP, which is predominantly glycolytic in tumor cells, while in differentiated fibroblasts it is predominantly of aerobic and mitochondrial origin [43, 257]) and wound healing (as a L-histidine precursor it regulates histamine synthesis during trauma [73] and on the other hand as a source of β-alanine stimulates collagen biosynthesis [176]). Most importantly, it also showed protective ability against AD and PD as well as it slows down the aging process [34, 79, 162, 261]. The role of carnosine against neurodegeneration is further described in the next paragraphs.

2.1.4 Carnosine in Alzheimer's and Parkinson's disease

AD is an incurable and progressive neurodegenerative diseases [182] and is the most common form of dementia. It affects more than 46 million people worldwide and this number is expected to almost triple by 2050 [204]. The most notable changes in people suffering from the AD are the loss of memory, behavioral and physical disability and cognitive decline, which at final instance lead to death [28, 161, 163]. The two main histopathological hallmarks of AD are extracellular depositions of fibrillar peptides (*i.e.* senile plaques) and intraneuronal fibrillar tangles [125]. Three main theories about the disease pathogenesis exist so far, (i) amyloid cascade hypothesis, (ii) metal ion hypothesis and (iii) oxidative stress hypothesis [125]. It is worth mentioning that these three hypotheses are not excluding each other but rather complementing.

(i) *The amyloid cascade hypothesis*. The focus of amyloid cascade hypothesis lies in the amyloid imbalance and formation of senile plaques. β-amyloids (Aβ), ~40-residues long fragments formed when Aβ precursor protein (APP) is cleaved by two aspartyl proteases [87, 226, 230], accumulate to form soluble oligomers, protofibrils and extracellular insoluble senile plaques, from which the soluble oligomers are believed to be the toxic form [107, 124, 147]. They can directly damage neurons and cause their death, destroy electrochemical signaling by (*i.e.* by forming small membrane-soluble channels that cause ions dishomeostasis, specially Ca^{2+}) and accumulate in mitochondria, disturbing the respiratory chain, which in turn causes oxidative stress and in final instance neuronal death.

(ii) *The metal dishomeostasis hypothesis*. Observations have shown heavy metal dishomeostasis in AD patients [133, 153]. Namely, higher zinc concentrations have been reported in blood and hippocampus of AD patients [55, 215], while on the other hand the zinc levels in cerebrospinal fluid (CSF) of AD patients

was lowered [112]. Globally, the overall zinc levels in the brain are reported to be either unchanged or lowered [185, 222]. Moreover, copper levels were found to be depressed in the AD brain tissue [106, 209, 221, 222]. For both, zinc and copper ions, it is believed that they catalyze the aggregation of Aβ [228]. Finally, the increased iron concentrations were also reported in AD neuropils [154, 218, 263] as well as very high concentrations of iron were found in Aβ plaques [154].

(iii) *The oxydative stress hypothesis*. Finally, the oxidative stress hypothesis originates from the effects of ROS and reactive nitrogen species (RNS) on the neurodegeneration. Both, ROS and RNS, have the ability of lipid peroxidation, protein oxidation and DNA oxidation, which were all observed in AD brains [131, 180, 200, 202].

Carnosine can possibly play a protective role in AD development through at least two different mechanisms. First, it can directly inhibit Aβ polymerization and aggregation by forming weak interactions with the charged groups surrounding the central hydrophobic cluster of Aβ [14, 203]. The second mechanism could be related to carnosine's heavy metal ion chelating ability, more precisely Zn^{2+} chelating ability. Elevated concentrations of Zn^{2+} ions were found in the post-mortem brains of AD patients, especially in the hippocampus [39, 49, 190]. Since it was shown that Zn^{2+} may induce oligomerization and aggregation of β-amyloid, carnosine's ability to chelate free Zn^{2+} ions may prove neuroprotective. It was shown *in vivo* in animal models that carnosine was indeed able to counteract β-amyloid aggregation [52]. Additionally, physiologically relevant carnosine concentrations provide a protective effect against Zn^{2+}- and Cu^{2+}-induced toxicity in cultured neurons [105]. Moreover, one study reports significantly lower carnosine concentrations in plasma in AD patient compared to the healthy controls [74], further suggesting carnosine's neuroprotective properties.

Another widely spread common neurodegenerative disorder is Parkinson's disease (PD), though in comparison with AD it affects an order of magnitude less people worldwide (~6 million). Its development starts in several brain regions, from which the most affected is the substantia nigra that controls balance and movement [136, 181]. Usually the first symptom is the trembling or shaking of a limb or one part of the body [201].

The pathophysiological picture of the disease is still not completely understood [175], though it is believed that the following two neuropathologies are the main cause of PD: (i) loss of pigmented dopaminergic neurons in the substantia nigra pars compacta [4, 156] and (ii) presence of Lewy bodies and Lewy neuritis [254]. The protective role of carnosine in PD is suggested to be due to its ability to inhibit α-synuclein

oligomerization [118, 126], which is one of the major constituents of Lewy bodies, and its ability to suppress some other changes in animals models of PD [240].

2.1.5 Carnosine metabolism

Carnosine is formed from β-alanine and L-histidine, where β-alanine is located on the N-terminal and L-histidine on the C-terminal part of the dipeptide. Though there were attempts from 1950 onwards to characterize the protein responsible for carnosine synthesis, it was only in 2010 when carnosine synthase (CS) was molecularly identified [65]. It is believed that the CS is a homotetramer with molecular mass of ~430 kDa and uses ATP in carnosine synthesis [65]. Based on this data it was classified as an ATP-grasp superfamily member, more specifically, as an ATPGD1 (ATP-grasp domain containing protein 1) [65].

The tissues with highest presence of CS include skeletal and heart muscles and certain brain regions [65], where the highest CS expression in the latter is in olfactory neurons. This is completely in agreement with the high carnosine concentration in those tissues [103, 158]. CS is located intracellularly and studies indicate that it is a cytosolic enzyme [65, 90, 177]. Except for the suggestion that the enzyme is potentially redox regulated, little is known about regulatory mechanism of CS expression and activity [65].

While it assumed that CS is the sole enzyme responsible for the synthesis of homocarnosine in brain, methylated carnosine analogs are synthesized either through carnosine N-methyltransferase, which catalyzes the methyl group transfer from S-adenozylmethionine to carnosine, forming either anserine or ophidine, or through enzymatic condensation of β-alanine with N-methylhistidine by CS [30].

Carnosine can be transported across cellular membranes using oligopeptide transporters 1 and 2 (PEPT1 and PEPT2) and peptide/histidine transporters 1 and 2 (PHT1 and PHT2), all of them belonging to the proton-coupled oligopeptide family of transporters [214, 258]. PEPT1 and PEPT2 have a very broad specificity of dipeptide and tripeptide targets, however they are not able to transport single amino acids or longer peptides [258]. On the other hand, PHT1 and PHT2 are able to transport also L-histidine itself, however their specificity is not clearly known.

Absorption of carnosine in small intestine to enterocytes, where it is partially hydrolyzed by the tissue carnosinase enzyme (CN2), is PEPT1 mediated [18, 80]. The remaining carnosine that is not hydrolyzed enters the bloodstream, where it is hydrolyzed by the serum carnosinase enzyme (CN1) [18, 80]. PEPT2 is responsible for the renal uptake of carnosine to epithelial cells [234, 235], where it is either

accumulated or hydrolyzed by carnosinases and its constituent amino acids are then transported to the blood across basolateral membrane using amino acids transporters. Moreover, PEPT2 is also responsible for more than 90% of the carnosine uptake in choroid plexus from CSF, suggesting that PEPT2 plays an important role of peptides homeostasis in CSF [234].

PHT1 and PHT2 on the other hand are primarily expressed in skeletal muscle, spleen, placenta, lung, leukocytes and heart [35]. Since the majority of the carnosine in human body is present in skeletal muscles, one could assume that also carnosine transporters are functionally expressed there. Indeed, mRNA transcripts of PHTs were recently found in mouse and human skeletal muscles [68].

Since for this work the most important enzymes in the carnosine metabolism are carnosinases, carnosine degrading enzymes are described in more details in the following paragraphs.

2.2 Carnosinase enzymes

While carnosine degrading enzyme in animals have been observed already in late 1940s by Hanson and Smith [89] and later studied by Rosenberg [210, 211], it was only twenty years later when carnosinase activity in humans was first described by Perry *et al*. [195], who studied carnosine and anserine degradation in patients with carnosinemia, and purified from human placenta [267]. Initially, two forms of carnosine-degrading activity were observed from human sera and tissue [172]. Therefore it was assumed that the enzyme present in human serum and tissue is actually the same enzyme in two different forms and similar to hog kidney carnosinase [145]. It was only 10 years later when Lenney *et al*. distinguished that in fact carnosinase activity in the human tissue and serum comes from two isoenzymes rather than two forms of the same enzyme [142, 145].[g] Based on their distribution in human

[g] Their assumption was based on the different properties of both isoenzymes purified from human tissue and serum, such as their specificity for other di- and tripeptides (tissue carnosinase hydrolyzed much wider range of di- and tripeptides), optimum pH range for carnosine hydrolysis (8.5 and 9.5 in serum and tissue carnosinase, respectively), isoelectric point (4.7 compared to 5.6 in serum and tissue carnosinase), molecular weight (160 and 90 kDa), activity towards homocarnosine (hydrolyzed only

body they were named human serum carnosinase (EC 3.4.13.20) and human tissue carnosinase EC 3.4.13.18)

Tissue carnosinase is present in the tissue of all mammals, while serum carnosinase was found only in higher primates (its presence was confirmed in the serum of humans, chimpanzees, gibbons, gorillas, orangutans and pygmy chimpanzees). On the other hand, in non-primate mammals, with exception of Syrian golden hamster, the carnosinase enzyme was absent in the serum [110].

Finally, in 2003 Teufel *et al*. identified and characterized the nucleotide sequence of two novel genes coding human serum carnosinase (hCN1) and human tissue carnosinase (hCN2), showing 49% sequence identity between both of them [233]. Both enzymes have been identified and characterized in humans as well as their tissue counterpart in mice (mCN2) [183, 233]. High sequence identity between the hCN2 and mCN2 (91% identity) suggests that the primary structure is highly conserved between the tissue carnosinase enzymes from different mammals. Furthermore, properties discriminating between serum and tissue isoenzymes described by Lenney *et al*. [142] were confirmed in the study of Teufel *et al*. [233]. Namely, pI had values of 4.4 and 5.6 for hCN1 and hCN2, respectively. The optimum pH for enzymatic activity towards carnosine was the same as reported before, namely in the range between 7.5 and 8.5 for hCN1, whereas it had more narrow range around 9.5 for hCN2. The molecular mass of native forms of both enzymes was 167 kDa and 90 kDa for hCN1 and hCN2. hCN1 showed catalytic activity for homocarnosine while this was not the case for hCN2, despite its broader substrate range. The K_m values for carnosine were 1.27 mM and 15 mM for hCN1 and hCN2 and 0.2 mM for homocarnosine in hCN1. For comparison between properties of hCN1 and hCN2 see also Table 2.1. Both, hCN1 and hCN2 were classified as metallopeptidases of M20 family in the MEROPS database (http://merops.sanger.ac.uk) [208].

---

by the serum enzyme), affinity for carnosine (serum carnosinase had affinity of 4 mM, while this value increases to 10 mM in tissue carnosinase) and inhibitory effect of different compounds (while addition of metal chelators, such as EDTA inhibited both enzymes, bestatin inhibited only the tissue carnosinase).

Table 2.1: Differences between hCN1 and hCN2 as reported by Lenney *et al*. [142] and Teufel *et al*. [233].

Property	Humans serum carnosinase		Human tissue carnosinase	
	Lenney *et al*. [142]	Teufel *et al*. [233]	Lenney *et al*. [142]	Teufel *et al*. [233]
Specificity	-	Narrow	-	Broad
Optimum pH	8.5	7.5 – 8.5	9.5	9.5
pI	4.7	4.4	5.6	5.6
M_r (of homodimer)	160 kDa	167 kDa	90 kDa	90 kDa
K_m				
Carnosine	4 mM	1.27 mM	10 mM	15 mM
Homocarnosine	active	0.2 mM	not active	-
Inhibitors	EDTA	1,10-*o*-phenantroline	EDTA, bestatine	*p*-hydroxy-mercuri-benzoate, bestatine

2.2.1 Homocarnosine degradation enzymes

An enzyme with the homocarnosine degrading property was found in the hog kidney and was named homocarnosinase [143]. However, its counterpart was not found in the humans. So far the only enzyme with known homocarnosine degrading ability is hCN1. Based on its ability to hydrolyze also homocarnosine it is believed that hCN1 in human body takes place of homocarnosinase in other mammal species [144]. Since the hCN1 is responsible for degradation of carnosine and homocarnosine in human blood and CSF, its interaction with both endogenous substrates was further considered in our work.

2.2.2 Distribution of hCN1 in humans

hCN1 in human is mainly found in the serum and cerebrospinal fluid (CSF). The hCN1 activity detected in other tissues was found to be proportional to the ratio of the blood

in the tissue extracts [233]. Moreover, it was proven that carnosinase enzymatic activity in CSF comes from the hCN1, which is believed to be secreted to CSF from brain cells [233]. The role of CN1 in brain is believed to be the regulation of homocarnosine levels, while in the blood stream it may act as regulator of carnosine and anserine concentrations [144].

Already the first study of hCN1 in human sera by Perry *et al*. reports an absence of hCN1 from sera of newly born infants and its concentration gradually increases to reach a physiological concentration in sera of healthy adults. The same conclusions were confirmed in studies of Lenney *et al*. [142], Jackson *et al*. [110] and Teufel *et al*. [233]. However, recently Peters *et al*. reported that also sera of children and newly born infants does contain hCN1, however in its less active monomeric form [197]. The content of the dimeric form in humans seems to increase with age [197]. Since the activity of the monomer is lower than that of the dimer, this leads us to suggest that a possible rearrangement of the structure upon dimerization may affect the enzymatic function [149].

2.2.3 hCN1's substrates and its role in blood and CSF

It was shown that human serum carnosinase is a dipeptide-degrading enzyme with very narrow range of substrates. Apart for its carnosine-hydrolyzing activity, it shows activity, however lower, for homocarnosine, anserine, ophidine etc. While enzymatic activity (k_{cat}/K_m ratio) is known only for carnosine and homocarnosine, K_m values were obtained also for anserine and ophidine [142, 192, 233]. Little kinetic data is available about other substrates.

So far, hCN1 remains the only known enzyme to enzymatically hydrolyze both carnosine and homocarnosine in blood and CSF [144, 233]. However, it is not completely clear why serum carnosinase is present in humans and other primates. It could be naively assumed that its role is to keep the serum carnosine concentration within certain homeostatic range. However, there are two reasons, why this is implausible. Firstly, many mammals with a large intake of carnosine or its analogs exhibit very low serum carnosinase activity in comparison with humans. Therefore, most probably hCN1 did not evolve to reduce (too) high carnosine concentrations in serum [30]. Moreover, studies on patients with carnosinase deficiency related diseases do not suggest that high carnosine concentrations would be directly causing mental retardation or any other clinical signs [50].

Finally a question arises, why would human body develop an enzyme that degrades a compound, showing protective properties in a number of disease, while by itself shows low toxicity? This has been called the human carnosinase paradox [97].

The only reasonable explanation for the existence of hCN1 so far is that it serves for the delivery of the hydrolysis products, such as β-alanine, L-histidine and GABA. hCN1 most probably gets in contact with its two substrates in the interstitium of certain tissues, where substrates are available through the release by cells. However, this explanation is still only hypothetical and further research is necessary in order to fully explain and understand the presence and the role of hCN1 in human serum [30].

3 Methods

3.1 Molecular dynamics

3.1.1 Statistical mechanics and molecular dynamics

By using statistical mechanics, we can connect information at microscopic level generated by molecular dynamics (MD) simulations (*e.g.* atomic positions, velocities) with experimental macroscopic quantities (*e.g.* pressure, temperature, energy) [7]. A system consisting of N particles can be defined through a set of particles' positions $\left(\vec{R} = \{\vec{R}_1, \dots, \vec{R}_N\}\right)$ and their corresponding momenta $\left(\vec{P} = \{\vec{P}_1, \dots, \vec{P}_N\}\right)$ and represented as a point in a $6N$- multidimensional space called phase space. For a particular point in phase space (a microscopic state of the system), named Γ, we can define the value of a property A as a function of Γ, $A(\Gamma)$. The experimentally observable value of A is defined as the ensemble average or thermodynamic average

$$A_{obs} = \langle A \rangle_{ens} = \int A\,(\Gamma) p(\Gamma) d\Gamma, \tag{3.01}$$

where $p(\Gamma)$ is the probability distribution function of collection of points Γ and $d\Gamma = d\vec{R}_1 \cdots d\vec{R}_N d\vec{P}_1 \cdots d\vec{P}_N$. The probability distribution function is dependent on the macroscopic parameters, which define the experimental conditions. If we consider a canonical ensemble, where the number of particles, N, the volume, V, and the temperature, T, of the system are constant, then we can write the probability distribution function as the Boltzmann distribution function

$$p_{NVT} = \frac{e^{-\frac{H(\Gamma)}{k_B T}}}{Z}, \tag{3.02}$$

where $H(\Gamma)$ is the classical Hamiltonian of the system defined as

$$H(\Gamma) = H\left(\{\vec{R}_I\}, \{\vec{P}_I\}\right) = \sum_{I=1}^{N} \frac{\vec{P}_I^{\,2}}{2M_I} + U\left(\{\vec{R}_I\}\right), \tag{3.03}$$

($\vec{R}_I$, $\vec{P}_I$ and M_I are the position, momentum and mass of the particle I), k_B is Boltzmann constant, and Z is the canonical partition function defined as

$$Z = \int d\Gamma p_{NVT}(\Gamma) = \int d\Gamma e^{-\frac{H(\Gamma)}{k_B T}}. \tag{3.04}$$

Since calculating Z is rather difficult in complex system like the biomolecular ones, we can obtain the ensemble average by an alternative strategy, namely by using MD simulations. Here, we consider the motion of Γ through the phase space as a function of time and with the initial point, Γ(0). By integrating the equation of motion from 0 to time τ we can obtain the evolution of Γ in a form of a trajectory consisting of a set of points Γ(t). Then, we can write the time average of A as

$$\langle A \rangle_\tau = \frac{1}{\tau} \int_0^\tau A[\Gamma(t)] dt. \tag{3.05}$$

The "ergodic hypothesis" connects $\langle A \rangle_\tau$ and $\langle A \rangle_{obs}$. For a system for which the ergodic hypothesis holds, if the system, is at equilibrium and is allowed to evolve infinitively long ($\tau \to \infty$), then it will pass through all the possible states. Therefore, we can write

$$\lim_{\tau \to \infty} \langle A(\Gamma) \rangle_\tau = \langle A(\Gamma) \rangle_\Gamma. \tag{3.06}$$

In practice, this means that the longer trajectory we generate, the closer we get to the ergodic condition and therefore the equality in (3.06) is better satisfied.

All molecular systems are assumed to be ergodic, which provides us with the theoretical justification for using MD simulations as tool to calculate ensemble averages of molecular systems [7].

3.1.2 Derivation of molecular dynamics equations

MD is an important and one of the most commonly used computational technique to study structure and function of biological (macro) molecules. The most accurate description of the time-dependent behavior of a molecular system can be obtained by solving its time-dependent Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t}\Psi\left(\{\vec{r}_i\},\{\vec{R}_I\};t\right) = H\Psi\left(\{\vec{r}_i\},\{\vec{R}_I\};t\right), \tag{3.07}$$

where $\{\vec{r}_i\}$ and $\{\vec{R}_I\}$ are the positions of the electrons and nuclei of the system, respectively, t is the time. Ψ represents the total wavefunction and H the Hamiltonian of the system usually defined as

$$\begin{aligned} H = &-\sum_I \frac{\hbar^2}{2M_I}\nabla_I^2 - \sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 + \frac{1}{4\pi\varepsilon_0}\sum_{i<j}\frac{e^2}{\left|\vec{r}_i-\vec{r}_j\right|} - \frac{1}{4\pi\varepsilon_0}\sum_{I,i}\frac{e^2 Z_I}{\left|\vec{R}_I-\vec{r}_i\right|} \\ &+ \frac{1}{4\pi\varepsilon_0}\sum_{I<J}\frac{e^2 Z_I Z_J}{\left|\vec{R}_I-\vec{R}_J\right|} \\ &= -\sum_I \frac{\hbar^2}{2M_I}\nabla_I^2 - \sum_I \frac{\hbar^2}{2m_e}\nabla_i^2 \\ &+ \frac{1}{4\pi\varepsilon_0}\sum_{i<j}\frac{e^2}{\left|\vec{r}_i-\vec{r}_j\right|} + V_{n-e}\left(\{\vec{r}_i\},\{\vec{R}_I\}\right) \\ &= -\sum_I \frac{\hbar^2}{2M_I}\nabla_I^2 + H_e\left(\{\vec{r}_i\},\{\vec{R}_I\}\right). \end{aligned} \tag{3.08}$$

In (3.08), the first term represents the kinetic energy of the nuclei, the second term the kinetic energy of the electrons and the next three terms the electron-electron, electron-nuclear and nuclear-nuclear Coulomb interactions, respectively. M_I and Z_I are the mass and atomic number of the I^{th} nucleus. m_e and $-e$ are the mass and charge of the electron and ε_0 is the vacuum permittivity. Finally, H_e is the electronic Hamiltonian.[h]

The total wavefunction, $\Psi\left(\{\vec{r}_i\},\{\vec{R}_I\};t\right)$, can be separated into the product consisting only of the electronic and the nuclear contributions (one-determinant approximation ansatz) [160, 241, 242].

---

[h] Note that this is the expression for the non-relativistic Hamiltonian, which is sufficient to get reasonably accurate description for the most of the biological systems.

$$\Psi\left(\{\vec{r}_i\},\{\vec{R}_I\};t\right)=\phi(\{\vec{r}_i\};t)\chi\left(\{\vec{R}_I\};t\right)e^{\frac{1}{\hbar}\int_{t_0}^{t}\tilde{E}_e(t')dt'}, \tag{3.09}$$

where the electronic $\left(\phi(\{\vec{r}_i\};t)\right)$ and nuclear $\left(\chi\left(\{\vec{R}_I\};t\right)\right)$wavefunctions are separately normalized to unity at every instant of time ($\langle\phi;t|\phi;t\rangle=1$ and $\langle\chi;t|\chi;t\rangle=1$). The phase factor in (3.09) can be rewritten considering that

$$\tilde{E}_e=\int\phi^*(\{\vec{r}_i\};t)\chi^*\left(\{\vec{R}_I\};t\right)H_e\phi(\{\vec{r}_i\};t)\chi\left(\{\vec{R}_I\};t\right)d\vec{r}d\vec{R}. \tag{3.10}$$

By inserting the last two equations, (3.09) and (3.10) into the Schrödinger equation (3.07), multiplying from left by $\langle\phi(\{\vec{r}_i\},t)|$ and $\langle\chi(\{\vec{r}_i\},t)|$, integrating over all the electronic and nuclear positions and applying the condition of energy conservation

$$\frac{d}{dt}E=\int\Psi^*\left(\{\vec{r}_i\},\{\vec{R}_I\};t\right)H\Psi\left(\{\vec{r}_i\},\{\vec{R}_I\};t\right)d\vec{r}d\vec{R}=0, \tag{3.11}$$

we get the following set of coupled time-dependent Schrödinger-like equations:

$$i\hbar\frac{\partial\phi}{\partial t}=-\sum_i\frac{\hbar}{2m_e}\nabla_i^2\phi+\left\{\int\chi^*\left(\{\vec{R}_I\};t\right)V_{n-e}\left(\{\vec{r}_i\},\{\vec{R}_I\}\right)\chi\left(\{\vec{R}_I\};t\right)d\vec{R}\right\}\phi, \tag{3.12}$$

$$i\hbar\frac{\partial\chi}{\partial t}=-\sum_i\frac{\hbar}{2M_I}\nabla_i^2\chi+\left\{\int\phi^*(\{\vec{r}_i\};t)\,H_e\left(\{\vec{r}_i\},\{\vec{R}_I\}\right)\phi(\{\vec{r}_i\};t)d\vec{r}\right\}\chi. \tag{3.13}$$

The electrons and the nuclei move in a time-dependent potential given by averaging over the other class of degrees of freedom.

In the next step, we want to approximate the quantum mechanical motion of the nuclei (much heavier than electrons) as a classical mechanical motion. To do this, we first rewrite nuclear wavefunction χ in polar representation

$$\chi\left(\{\vec{R}_I\},t\right)=A\left(\{\vec{R}_I\};t\right)e^{\frac{iS(\{\vec{R}_I\};t)}{\hbar}}, \tag{3.14}$$

where *A* is amplitude and *S* is phase. If we now insert this expression in equation (3.13) and separate real and imaginary parts we get

$$\frac{\partial S}{\partial t}=\sum_I\frac{1}{2M_I}(\nabla_I S)^2+\int\phi^*H_e\phi d\vec{r}=\hbar^2\sum_I\frac{1}{2M_I}\frac{\nabla_I^2A}{A}, \tag{3.15}$$

$$\frac{\partial A}{\partial t}+\sum_I\frac{1}{M_I}(\nabla_I A)(\nabla_I S)+\sum_I\frac{1}{2M_I}A(\nabla_I^2S)=0. \tag{3.16}$$

The latter equation can be rewritten as a continuity equation for the density probability $A^2 = |\chi|^2$ of the nuclei by multiplying from left by *2A*

$$\frac{\partial A^2}{\partial t} + \sum_I \frac{1}{M_I} \nabla_I (A^2 \nabla_I S) = 0. \tag{3.17}$$

However, more important for our purpose is the equation (3.15), which contains term that depends on $\hbar$. This contribution vanishes in the classical limit $\hbar \to 0$ and if we set $\nabla_I S$ as the momentum of nucleus I, $\vec{P}_I$, we get

$$\frac{\partial S}{\partial t} = \sum_I \frac{1}{2M_I} \vec{P}_I^2 + \int \phi^* H_e \phi d\vec{r} = 0. \tag{3.18}$$

This equation is isomorphic to the Hamilton-Jacobi equation [83, 160] of a classical mechanics system

$$\frac{\partial S}{\partial t} + H\left(\{\vec{R}_I\}, \{\vec{P}_I\}\right) = 0, \tag{3.19}$$

with the classical Hamiltonian *H* as

$$H\left(\{\vec{R}_I\}, \{\vec{P}_I\}\right) = T\left(\{\vec{P}_I\}\right) + V\left(\{\vec{R}_I\}\right) = \sum_I \frac{1}{2M_I} \vec{P}_I^2 + \int \phi^* H_e \phi d\vec{r}. \tag{3.20}$$

The corresponding Newtonian equation of motion is

$$\frac{d\vec{P}_I}{dt} = -\nabla_I \int \phi^* H_e \phi d\vec{r} \tag{3.21}$$

or

$$M_I \ddot{\vec{R}}_I = -\nabla_I \int \phi^* H_e \phi d\vec{r} = -\nabla_I V_e\left(\{\vec{R}_I\}\right). \tag{3.22}$$

Therefore, in this approximation the nuclei move according to classical mechanics in an effective potential V_e due to the electrons. This potential is a function of only the position of nuclei at a time *t* as a result of averaging H_e over the electronic degrees of freedom at a given instantaneous nuclear positions $\vec{R}_I(t)$.

In order to solve the equation for electronic wavefunction (3.12), which contains the full quantum nuclear wavefunction, in the classical limit $\hbar \to 0$ we can replace the nuclear density $\left|\chi\{\vec{R}_I\}\right|^2$ with a product of delta functions $\prod_I \delta\left[\vec{R}_I - \vec{R}_I(t)\right]$ centered at the instantaneous nuclear positions $\vec{R}_I(t)$. The expectation value of the position operator in (3.12) now becomes

$$\int \chi^*(\{\vec{R}_I\};t)\,\vec{R}_I\chi(\{\vec{R}_I\};t)d\vec{R} \rightarrow \vec{R}_I(t), \tag{3.23}$$

leading to a time-dependent wave equation for the electrons

$$\begin{aligned} i\hbar\frac{\partial\phi}{\partial t} &= -\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2\phi + V_{n-e}(\{\vec{r}_I\};\{\vec{R}_I(t)\})\phi \\ &= H_e(\{\vec{r}_I\};\{\vec{R}_I(t)\})\phi(\{\vec{r}_I\};\{\vec{R}_I\};t), \end{aligned} \tag{3.24}$$

which evolves self-consistently as the classical nuclei are propagated according to the equation (3.22).

Molecular dynamics simulations are branched out into two methodologies based on the treatment of the electronic part, *ab initio* (forces on nuclei are obtained from the electronic structure calculations) and classical MD (forces are derived from the predefined potential models by analytical gradient applications). Both methodologies are described in the following sections.

3.1.3 *Ab initio* molecular dynamics

3.1.3.1 Ehrenfest MD

This method is named after Ehrenfest who was the first to address the question of the derivation of Newtonian classical dynamics from the Schrödinger wave equation. It is relying on solving equations (3.22) and (3.24) simultaneously. The time-dependent electronic Schrödinger equation is solved "on-the-fly" as the nuclei are moving according to classical mechanics. The electronic wavefunction ϕ from the equation (3.24) can be expanded in terms of a basis of electronic states ψ_k

$$\phi(\{\vec{r}_i\},\{\vec{R}_I\};t) = \sum_{k=0}^{\infty} c_k(t)\,\psi_k(\{\vec{r}_i\},\{\vec{R}_I\}), \tag{3.25}$$

where $c_k(t)$ are complex time-dependent coefficients satisfying

$$\sum_k |c_k(t)|^2 = 1. \tag{3.26}$$

A possible choice for the basis functions ψ_k is the adiabatic basis obtained from solving the time-independent electronic Schrödinger equation

$$H_e(\{\vec{r}_i\},\{\vec{R}_I\})\psi_k(\{\vec{r}_i\},\{\vec{R}_I\}) = E_k(\{\vec{R}_I\})\psi_k(\{\vec{r}_i\},\{\vec{R}_I\}), \tag{3.27}$$

where $\vec{R}_I$ are the instantaneous nuclear positions at time t based on equation (3.22). Now we can write the equations of motion (3.22) and (3.24) as

$$M_I \ddot{\vec{R}}_I(t) = -\sum_k |c_k(t)|^2 \nabla_I E_k - \sum_{k,l} c_k^* c_l (E_k - E_l) \vec{d}_I^{kl} \tag{3.28}$$

$$i\hbar \dot{c}_k(t) = c_k(t) E_k - i\hbar \sum_{I,l} c_l(t) \dot{\vec{R}}_I \vec{d}_I^{kl}, \tag{3.29}$$

where the coupling terms are defined as

$$\vec{d}_I^{kl}\left(\{\vec{R}_I\}\right) = \int \psi_k^* \nabla_I \psi_l d\vec{r}. \tag{3.30}$$

In many cases, the electronic wavefunction (3.25) can be approximated with only one electronic state (usually the ground state ψ_0) bringing to the Ehrenfest equations of motion

$$M_I \ddot{\vec{R}}_I(t) = -\nabla_I \int \psi_0^* \nabla_I \psi_0 d\vec{r} \tag{3.31}$$

$$i\hbar \frac{\partial \psi_0}{\partial t} = H_e \psi_0. \tag{3.32}$$

It is important to mention that the wavefunction propagation is unitary, meaning that wavefunction preserves its norm and the set of orbitals used to build up the wavefunction will stay orthonormal [160].

3.1.3.2 Born-Oppenheimer MD

An alternative approach to *ab initio* MD is to solve the time-independent Schrödinger equation (3.27) for fixed nuclear positions at each step of the molecular dynamics of the nuclei. In this way, the time dependence of electronic structure is a only consequence of nuclear motion. The equations of motion in the so-called Born-Oppenheimer molecular dynamics can be written as

$$M_I \ddot{\vec{R}}_I(t) = -\nabla_I \min_{\psi_0} \langle \psi_0 | H_e | \psi_0 \rangle \tag{3.33}$$

$$H_e \psi_0 = E_0 \psi_0 \tag{3.34}$$

for the electronic ground state.

3.1.3.3 Car-Parrinello MD

The advantage of Ehrenfest MD lies in fact that we have to calculate the electronic wavefunction only at the beginning. However, the integration time step of equations of motion is dictated by the intrinsic dynamic of electrons, which is much faster that nuclear dynamics. On the other hand, Born-Oppenheimer MD has no electron dynamics and the relevant time scale is given by nuclear dynamics, resulting in the possibility to employ larger time steps to integrate equations of motion. However, this scheme requires that the electronics structure problem (3.34) be solved at every time step.

The *ab initio* scheme developed by Car and Parrinello [40] combines the advantages of the Ehrenfest and the Born-Oppenheimer MD. This scheme allows getting a minimized electronic wavefunction while the nuclei are propagated like in the Ehrenfest MD but with increased time step. This way, the computational expensive explicit electronic minimization at each time step is not needed. This so-called Car-Parrinello MD is based on the idea that the fast electronic motion can be adiabatically separated from the slower nuclear motion. A "classical" extended Lagrangian of the physical system can be written, which includes both electronic and nuclear degrees of freedom, where each electronic degree of freedom (orbital φ_i) is considered as a fictitious (classical) particle with a fictitious mass μ. If the fictitious mass of the electrons is small enough that the frequencies (energies) of the electronic motion is almost fully separated by the frequencies of the nuclear motion, the total electronic wavefunction (which is a function of the single orbitals) can follow the nuclear motion "adiabatically", i.e. remaining very close to the Born-Oppenheimer energy surface as done in the Born-Oppenheimer MD enforcing it by minimizing the wavefunction at each time step.

The extended Lagrangian postulated by Car and Parrinello reads

$$L_{CP} = \sum_I \frac{1}{2} M_I \dot{\vec{R}}_I^2 + \sum_i \frac{1}{2} \mu_i \langle \dot{\varphi}_i | \dot{\varphi}_i \rangle - \langle \psi_0 | H_e | \psi_0 \rangle + constraints, \qquad (3.35)$$

where the first two terms stand for kinetic energy of the nuclei and the fictitious electrons with the fictitious mass μ, respectively. The third term is the potential energy and the last one enforces the orthogonality of the single-electron orbitals φ_i, which are, in contrast with Ehrenfest MD, not conserved and have to be orthonormalized at every time step. The corresponding Newtonian equations of motion are obtained from the associated Euler-Lagrange equations

$$\frac{d}{dt}\frac{\partial L_{CP}}{\partial \dot{\vec{R}}_I} = \frac{\partial L_{CP}}{\partial \vec{R}_I} \tag{3.36}$$

$$\frac{d}{dt}\frac{\delta L_{CP}}{\delta \dot{\varphi}_i^*} = \frac{\delta L_{CP}}{\delta \varphi_i^*}, \tag{3.37}$$

which give

$$M_I \ddot{\vec{R}}_I(t) = -\frac{\partial}{\partial \vec{R}_I}\langle\psi_0|H_e|\psi_0\rangle + \frac{\partial}{\partial \vec{R}_I}\{constraints\} \tag{3.38}$$

$$\mu\ddot{\varphi}_i(t) = -\frac{\delta}{\delta\varphi_i^*}\langle\psi_0|H_e|\psi_0\rangle + \frac{\delta}{\delta\varphi_i^*}\{constraints\}. \tag{3.39}$$

While the nuclei move at physical temperature $\sim \sum_I M_I \dot{\vec{R}}_I^2$, this is not true for the electrons, which move at a fictitious temperature $\sim \sum_i \mu\langle\dot{\varphi}_i|\dot{\varphi}_i\rangle$. The fictitious mass is the key parameter in controlling the adiabatic separation of the nuclei from the electrons. The smaller μ, the closer is the instantaneous minimum energy $\langle\psi_0|H_e|\psi_0\rangle$ to the exact Born-Oppenheimer energy surface. On the other hand, smaller μ means smaller integration step. Therefore, a compromise between accuracy and efficiency in the simulations has to be found when choosing the value of μ.

3.1.3.4 Density Functional Theory

Ab initio molecular dynamics is not tied to any specific approach of describing electronic structure. However, its strength and weakness are connected with the strength and weakness of the chosen electronic structure method. Some of the electronic structure methods that are employed in combination with *ab initio* MD are the density functional theory (DFT), Hartree-Fock (HF), the full configuration interaction (full CI) and the semiempirical methods. In the work of this thesis, we mostly employed the DFT-based Car-Parrinello molecular dynamics approach. DFT is based on the idea that the ground state properties of a quantum system with *N* particles can be described starting from its charge density $\rho(\vec{r})$

$$\rho(\vec{r}) = N\int |\psi(\vec{r}_1, \dots, \vec{r}_N)|^2 d\vec{r}_1 \cdots d\vec{r}_N. \tag{3.40}$$

The two main advantages of this technique are that the density is an observable that can be easily measured and visualized and the dimensionality of the problem is reduced from 3*N* ($\psi(\vec{r}_1, \dots, \vec{r}_N)$) to 3, since the density is a function of a point in the space ($\rho(\vec{r})$).

DFT is founded on the two Hohenberg-Kohn theorems [100]:

a) The first Hohenberg-Kohn theorem states that the external potential, v_{n-e}, is a unique functional of the ground-state electron density $\rho(\vec{r})$.

$$v_{n-e}(\vec{r}) = -\frac{e}{4\pi\varepsilon_0}\sum_I \frac{Z_I}{\left|\vec{R}_I - \vec{r}\right|} + \frac{e^2}{4\pi\varepsilon_0}\sum_{I<J}\frac{Z_I Z_J}{\left|\vec{R}_I - \vec{R}_J\right|}. \tag{3.41}$$

Since the external potential determines the Hamiltonian of the system, full many-electron ground state is a unique functional of $\rho(\vec{r})$ and the total ground state energy is therefore written as a functional of $\rho(\vec{r})$.

$$\begin{aligned} E_0[\rho(\vec{r})] &= T[\rho(\vec{r})] + E_{e-e}[\rho(\vec{r})] + E_{n-e}[\rho(\vec{r})] \\ &= F_{H-K}[\rho(\vec{r})] + \int \rho(\vec{r})\, v_{n-e}(\vec{r}) d\vec{r}. \end{aligned} \tag{3.42}$$

$F_{H-K}[\rho(\vec{r})]$ is called the Hohenberg-Kohn functional and is independent of the external potential v_{n-e}.

b) The second Hohenberg-Kohn theorem states that for any trial electron density $\tilde{\rho}(\vec{r})$ we always have

$$E_0 \leq E_0[\tilde{\rho}(\vec{r})]. \tag{3.43}$$

This means that if we want to find the ground state energy E_0, we need to search for the minimum of the functional $E_0[\tilde{\rho}(\vec{r})]$.

Kohn-Sham equations. The Hohenberg-Kohn theorems provide the theoretical framework that proves the equivalence of a quantum description based on the electronic wavefunction and the one based on the electronic density. However, they do not give any "practical way" to rephrase the quantum problem in terms of the electronic density. This step was performed by Kohn and Sham some years after the original formulation of the Hohenberg-Kohn theorems.

Kohn and Sham assumed that for each physical electron system it is always possible to find a corresponding reference system of non-interacting electrons (Kohn-Sham electrons described by the orbitals φ_i^{KS}) that feels an effective potential (Kohn-Sham potential, v_{KS}) so that its ground state density is identical to that of the physical system (with interacting electrons) [130]. For the non-interacting system of electrons we know that the exact solution of the equations of motions in terms of φ_i^{KS} is the Slater determinant $\left|\left|\varphi_i^{KS}\right|\right|$. The Hohenberg-Kohn functional can therefore be written in terms of the non-interacting system orbitals:

$$F_{H-K}[\rho(\vec{r})] = -\frac{\hbar^2}{2m_e}\sum_i \int \varphi_i^{KS*}(\vec{r})\nabla^2\varphi_i^{KS}(\vec{r})d\vec{r} + \frac{e^2}{4\pi\varepsilon_0}\int \rho(\vec{r})\frac{1}{|\vec{r}-\vec{r}'|}\rho(\vec{r}')d\vec{r}d\vec{r}' + E_{xc}\,[\rho(\vec{r})], \tag{3.44}$$

where

$$\rho(\vec{r}) = \sum_i \left|\varphi_i^{KS}(\vec{r})\right|^2 \tag{3.45}$$

by definition, and $E_{xc}[\rho(\vec{r})]$ is called the exchange correlation energy term. $E_{xc}[\rho(\vec{r})]$ is the only unknown term and that has to be approximated.

The total electronic energy can then be rewritten as

$$\begin{aligned} E_{KS}[\rho(\vec{r})] &= F_{H-K}[\rho(\vec{r})] \\ &\quad + \int \rho(\vec{r})v_{n-e}(\vec{r})d\vec{r} = \\ &= -\frac{\hbar^2}{2m_e}\sum_i \int \varphi_i^{KS*}(\vec{r})\nabla^2\varphi_i^{KS}(\vec{r})d\vec{r} \\ &\quad + \frac{e^2}{4\pi\varepsilon_0}\int \rho(\vec{r})\frac{1}{|\vec{r}-\vec{r}'|}\rho(\vec{r}')d\vec{r}d\vec{r}' \\ &\quad + \int \rho(\vec{r})v_{n-e}(\vec{r})d\vec{r} + E_{xc}[\rho(\vec{r})]\,, \end{aligned} \tag{3.46}$$

Since (3.43) holds, we can apply the variational principle to the equation (3.46) together with the condition (3.45), obtaining a set of Kohn-Sham equations as functions of the Kohn-Sham orbitals $\{\varphi_i^{KS}\}$:

$$\left[-\frac{\hbar^2}{2m_e}\nabla^2 + v_{KS}(\rho)\right]\varphi_i^{KS}(\vec{r}) = \varepsilon_i\varphi_i^{KS}(\vec{r}), \tag{3.47}$$

where the Kohn-Sham potential is

$$v_{KS}(\rho) = v_{n-e}(\vec{r}) + \frac{1}{4\pi\varepsilon_0}\int \frac{1}{|\vec{r}-\vec{r}'|}\rho(\vec{r}')d\vec{r}' + v_{xc}(\rho) \tag{3.48}$$

and $v_{xc}(\rho)$ is the exchange-correlation potential defined as the functional derivative of E_{xc} with respect to $\rho(\vec{r})$

$$v_{xc}(\rho) = \frac{\delta E_{xc}[\rho(\vec{r})]}{\delta\rho(\vec{r})}. \tag{3.49}$$

If we can solve the electronic quantum problem in the Kohn-Sham DFT, we can rewrite the Car-Parrinello Lagrangian in terms of the Kohn-Sham orbitals as

$$L_{CP} = \sum_I \frac{1}{2} M_I \dot{\vec{R}}_I^2 + \sum_i \frac{1}{2}\mu \int \dot{\varphi}_i^{KS*}(\vec{r})\dot{\varphi}_i^{KS*}(\vec{r})d\vec{r} - E_{KS}[\rho(\vec{r})] + \sum_{i,j} \Lambda_{ij}\left[\int \dot{\varphi}_i^{KS*}(\vec{r})\dot{\varphi}_i^{KS*}(\vec{r})d\vec{r} - \delta_{ij}\right], \tag{3.50}$$

where Λ_{ij} are Lagrange multipliers.

Exchange-correlation functional. In the equation (3.46) the exchange-correlation energy term $E_{xc}[\rho(\vec{r})]$ is unknown. For any application of DFT, it is crucial to find a good approximation of it. Kohn and Sham approximated $E_{xc}[\rho(\vec{r})]$ through the "local density approximation" (LDA) [130, 187], which is widely used in physics. In the LDA, the exchange-correlation energy is assumed to be the same as the one of a homogenous electron gas with the same density at each point

$$E_{xc}^{LDA}[\rho(\vec{r})] = \int \rho(\vec{r})\varepsilon_{xc}^{hom}[\rho(\vec{r})]d\vec{r} = \int \rho(\vec{r})\{\varepsilon_x^{hom}[\rho(\vec{r})] + \varepsilon_c^{hom}[\rho(\vec{r})]\}d\vec{r}. \tag{3.51}$$

For the homogenous electron gas, the first integral, the "exchange" part can be exactly calculated:

$$\varepsilon_x^{hom}[\rho(\vec{r})] = -\frac{3}{4}\left(\frac{3\rho(\vec{r})}{\pi}\right)^{\frac{1}{3}}. \tag{3.52}$$

The second integral, the "correlation" part has only been determined analytically in the high and low density limit [41, 42] and by highly accurate quantum Monte-Carlo calculations [44]. Suitable analytical expressions to calculate ε_c^{hom} have been developed by Vosko, Wilk and Nusair [253] and by Perdew and Wang [194].

However, for chemical applications LDA is usually not sufficient and this led to the development of generalized-gradient approximations (GGAs) for E_{xc}, in which also the density gradient $\nabla\rho(\vec{r})$ is taken into account in the expression of E_{xc}:

$$E_{xc}^{GGA}[\rho(\vec{r})] = \int \rho(\vec{r})\varepsilon_{xc}[\rho(\vec{r}), |\nabla\rho(\vec{r})|]d\vec{r} = E_x^{GGA}[\rho(\vec{r})] + E_c^{GGA}[\rho(\vec{r})]. \tag{3.53}$$

One of the widely used GGA functionals is given by a combination of the Becke exchange functional and the Lee-Yang-Parr correlation functional (BLYP) [20, 139].

The Becke exchange functional is defined as

$$E_x^{GGA}[\rho(\vec{r})] = E_x^{LDA}[\rho(\vec{r})] - \beta \int \rho(\vec{r})^{\frac{4}{3}} \frac{x^3}{1 + 6\beta \sinh^{-1} x} d\vec{r}, \tag{3.54}$$

where

$$x = \frac{|\nabla\rho(\vec{r})|}{\rho(\vec{r})^{\frac{4}{3}}} \tag{3.55}$$

while β is determined by a fit on the data from an exact HF calculations and was fixed by Becke to 0.0042 au.

The Lee-Yang-Parr correlation functional is written as

$$\begin{aligned} E_c^{GGA}[\rho(\vec{r})] = -a \int \frac{1}{1 + \rho(\vec{r})^{\frac{1}{3}}} \Bigg(& \rho(\vec{r}) \\ & + b\rho(\vec{r})^{-\frac{2}{3}} \Bigg[C_F \rho(\vec{r})^{\frac{5}{3}} - 2t_w \\ & + \left(\frac{1}{9} t_w + \frac{1}{18} \nabla^2 \rho(\vec{r}) \right) \Bigg] e^{-\rho(\vec{r})^{\frac{1}{3}}} \Bigg) d\vec{r}, \end{aligned} \tag{3.56}$$

where

$$C_F = \frac{3}{10}(3\pi^2)^{\frac{2}{3}} \tag{3.57}$$

$$t_w = \frac{1}{8} \frac{|\nabla\rho(\vec{r})|^2}{\rho(\vec{r})} - \frac{1}{8} \nabla^2 \rho(\vec{r}) \tag{3.58}$$

with a = 0.04918, b = 0.132 derived from the Colle-Salvetti formula to calculate the correlation energies from HF second order density matrix.

Basis set. In the actual implementations of Kohn-Sham equations (3.47), the Kohn-Sham orbitals $\{\varphi_i(\vec{r})\}$ are expanded as linear combination of basis functions, f_l, with well-known behavior and which allow to perform mathematical operations on computers very quickly

$$\varphi_i(\vec{r}) = \sum_l c_{il} f_l\left(\vec{r}; \{\vec{R}_I\}\right), \tag{3.59}$$

where c_{il} are orbital expansion coefficients. In general the expression (3.59) holds exactly for any reasonable function in the limit of using a complete set of basis functions.

In this thesis, we used localized Gaussian-type basis functions for ligand parameterization at quantum level and plane wave basis set in the *ab initio* MD simulations. Both are described in following paragraphs.

Localized basis sets. The most widely used basis functions in quantum chemistry are the Slater-type basis functions (STOs)

$$f_{\vec{m}}^{S} = N_{\vec{m}}^{S} r_x^{m_x} r_y^{m_y} r_z^{m_z} e^{-[\zeta_{\vec{m}}|\vec{r}|]} \tag{3.60}$$

and the Gaussian-type basis functions (GTOs)

$$f_{\vec{m}}^{G} = N_{\vec{m}}^{G} r_x^{m_x} r_y^{m_y} r_z^{m_z} e^{-[\alpha_{\vec{m}}|\vec{r}|]}, \tag{3.61}$$

Where the integer triplet $\vec{m}$ is called the angular momentum channel and $N_{\vec{m}}, \zeta_{\vec{m}}$ and $\alpha_{\vec{m}}$ are constants that are typically kept fixed during an electronic structure calculation, so that only the coefficients c_{il} need to be optimized.

Plane wave basis set. The use of plane waves as a basis set is more convenient when periodic systems has to be described. Plane waves are defined as

$$f_G^{PW} = \frac{1}{\sqrt{V}} e^{i\vec{G}\vec{r}}, \tag{3.62}$$

where V is the volume of the periodic cell and $\vec{G}$ is the reciprocal lattice vector. The Kohn-Sham orbitals can be written in Bloch form [12]

$$\varphi_i(\vec{r}, \vec{k}) = e^{i\vec{k}\vec{r}} u_i(\vec{r}, \vec{k}), \tag{3.63}$$

where $\vec{k}$ is a wave vector in the first Brillouin zone of the reciprocal lattice and $u_i(\vec{r}, \vec{k})$ is a cell-periodic function

$$u_i(\vec{r}, \vec{k}) = u_i(\vec{r} + \vec{R}, \vec{k}). \tag{3.64}$$

The periodic functions can be expanded as a Fourier series

$$u_i(\vec{r}, \vec{k}) = \frac{1}{\sqrt{V}} \sum_{\vec{G}} c_i(\vec{G}, \vec{k}) e^{i\vec{G}\vec{r}} \tag{3.65}$$

and Kohn-Sham orbitals become

$$\varphi_i(\vec{r}, \vec{k}) = \frac{1}{\sqrt{V}} \sum_{\vec{G}} c_i(\vec{G}, \vec{k}) e^{i(\vec{G}+\vec{k})\vec{r}}. \tag{3.66}$$

Pseudopotentials. In order to minimize the size of the plane wave basis set, core electrons (electrons close to the nuclei) are considered as degrees of freedom but

simply replaced by pseudopotentials [92, 108, 199], in order to take into account their effects on the other (valence) electrons. This approximation is made on the well known observation that core electrons are not involved explicitly in the description of chemical reactions and are unaffected by the chemical environment. Therefore, core electrons effects are implicitly included in the nuclear potential, thus forming a pseudopotential, which takes into account joint effect of nucleus and its core electrons on the motion of the valence electrons. This way the total electronic wavefunction, describing only the valence electron and sometime referred as pseudo wavefunction $|\psi^{ps}\rangle$, is made much smoother close to the nuclear core region. As a consequence, the number of plane waves needed to expand this wavefunction can be considerably reduced.

The pseudo valence states can be found by solving the all-electron Schrödinger equation for an atom

$$\left[\hat{H} + \hat{V}_{nl}\right]|\psi^{ps}\rangle = E|\psi^{ps}\rangle, \tag{3.67}$$

where $\hat{H}$ is the all-electron Hamiltonian of an atom and $\hat{V}_{nl}$ is an energy-dependent non-local potential defined as

$$\hat{V}_{nl} = \sum_{i \in core} (E - E_i)|\chi_i\rangle\langle\chi_i|. \tag{3.68}$$

$|\chi_i\rangle$ and E_i are the single-electron eigenstates and eigenvalues of the core-electron states. Summing the repulsive $\hat{V}_{nl}$ and the strongly attractive Coulomb potential, $\hat{V}_{Cou}$, in $\hat{H}$ results in a much weaker pseudopotential operator $\hat{V}_{ps}$

$$\hat{V}_{ps} = \hat{V}_{Cou} + \sum_{i \in core} (E - E_i)|\chi_i\rangle\langle\chi_i| = \hat{V}_{Cou} + \hat{V}_{nl}. \tag{3.69}$$

Therefore, pseuodopotentials can be derived from *ab initio* calculations of isolated atoms by solving Kohn-Sham equations. Several procedures have been proposed for generating pseudopotentials to be used in combination with a plane waves basis set.. In our work, we used the "norm-conserving" pseudopotentials derived through the Martins-Troullier (MT) scheme [239], in which the following conditions have to be satisfied:

a) the lowest pseudo wavefunction does not contain any radial nodes;
b) the eigenvalues of pseudo and real wavefunctions are the same;
c) the radial pseudo wavefunction is equal to the radial all-electron wavefunction outside a cut-off radius r_c;

d) the integrated charge inside r_c is the same for both pseudo and real wavefunctions.

3.1.4 Force field-based molecular dynamics

In force field based MD, known also as classical MD, the interaction potential energy, U^{FF}, of the atoms of the system is approximated by simple analytical functions of the atomic coordinates (or more precisely, nuclear coordinates), instead of performing expensive calculations to solve the time-dependent Schrödinger equation as it is the case in *ab initio* MD. Therefore, in classical MD atoms are represented as points in the space and their internal structure, as well as the role of the electrons in it, is neglected.

Expressions for the potential energy U^{FF} contain sets of adjustable parameters that have to be fitted to data obtained from experiments or from high-level quantum calculations. Different types of force fields are available (*e.g.* AMBER [232], GROMOS [244], CHARMM [155]…) and they differ in both the functional forms and the parameter sets. In the work of this thesis the AMBER force field has been employed.

The potential energy for the AMBER force field reads

$$\begin{aligned} U^{FF}(r_1, \dots, r_N) = & \sum_{bonds} K_r\,(r - r_0)^2 + \sum_{angles} K_\theta\,(\theta - \theta_0)^2 \\ & + \sum_{torsions} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] \\ & + \sum_{i<j} \sigma_{ij}\left[\left(\frac{r_{0i,j}}{r_{ij}}\right)^{12} - \left(\frac{r_{0i,j}}{r_{ij}}\right)^{6}\right] + \sum_{i<j} \frac{q_i q_j}{\varepsilon r_{ij}} \end{aligned} \tag{3.70}$$

The first three terms in (3.70) represent bonded interactions between atoms, and the last two terms non-bonded ones. K_r, K_θ and V_n are bond, angle and torsion force constants. r and r_0 represent bond lengths and their equilibrium values, while θ and θ_0 are angles and their equilibrium values. n, ϕ and γ are the number of barriers, dihedral angle and phase, respectively. r_{ij} is the distance between atoms i and j, while $r_{0i,j}$ is their equilibrium distance and σ_{ij} is the well depth of a Lennard-Jones potential. q_i and q_j are the partial electric charges of atoms i and j, while ε is the dielectric constant.

As seen in the AMBER force field, Van der Waals interactions are usually described using the Lennard-Jones potential, which decreases with r_{ij}^{-6} at large distances and it is therefore considered a short-range potential. On the other hand Coulomb interactions decrease with r_{ij}^{-1} for large distances and therefore this potential is considered a long-range one.

The time evolution of the system is calculated through the classical Newtonian equations of motion

$$\vec{F}_i = M_i \vec{a}_i, \tag{3.71}$$

where $\vec{F}_i$ is force acting on particle i, M_i is the mass of the particle i and $\vec{a}_i$ is the second derivative of the particle's position with respect to time t. By the force field one gets the values of the forces acting on the atoms:

$$\vec{F}_i = \frac{\partial U(\vec{R})}{\partial \vec{R}_i}. \tag{3.72}$$

The atomic trajectories in function of time are then calculating by integrating the equations (3.71).

Since (3.71) are second order ordinary differential equations, initial particles' positions and velocities have to be provided in order to integrate them. Velocities are usually randomly generated with the Maxwell-Boltzmann probability distribution corresponding to a given temperature T. At this point,

the forces acting on each particle can be calculated by using the force field and the atomic position of all other particles through eq. (3.72);

the set of Newton's equations of motion (3.71) can be numerically solved by one of the integration methods described in next section, and the positions and velocities at a time Δt are obtained.

These two steps are repeated, for the whole course of the simulation, with the only difference that the initial particles' positions and velocities are the updated coordinates and velocities from the previous cycle. This way the full MD trajectory is generated.

3.1.4.1 Integration of the equation of motions

In each cycle of the classical MD simulation, the Newton's equations of motion are integrated. The integration time step Δt is defined at the beginning of the simulation and is usually in the range of 1-2 fs. This is because the time step should be much smaller than the smallest time motions that are being reproduced in the MD simulation. Without any artificial constraints added to the system, the fastest motions are the oscillations of hydrogen atoms, which are typically of the order of 10 fs.

The simplest integration algorithm that results effective to integrate the Newtonian equation is the Verlet [247] algorithm, traced back to 1960s. In this thesis a variation

of the Verlet algorithm, the Leap-Frog [99] one was employed, which is also one of the nowadays most used algorithms in MD.

Verlet algorithm. In order to derive the expression for the Verlet algorithm, first we write the position of a particle at time $t + \Delta t$ and $t - \Delta t$ as a Taylor expansion around time *t*:

$$\vec{R}(t + \Delta t) = \vec{R}(t) + \frac{d\vec{R}(t)}{dt}\Delta t + \frac{1}{2!}\frac{d^2\vec{R}(t)}{dt^2}\Delta t^2 + \frac{1}{3!}\frac{d^3\vec{R}(t)}{dt^3}\Delta t^3 + O(\Delta t^4) \tag{3.73}$$

$$\vec{R}(t - \Delta t) = \vec{R}(t) - \frac{d\vec{R}(t)}{dt}\Delta t + \frac{1}{2!}\frac{d^2\vec{R}(t)}{dt^2}\Delta t^2 - \frac{1}{3!}\frac{d^3\vec{R}(t)}{dt^3}\Delta t^3 + O(\Delta t^4). \tag{3.74}$$

$\vec{R}(t - \Delta t)$ is the previous position of the particle, $\vec{R}(t)$ the current position and $\vec{R}(t + \Delta t)$ the new position. Since $\frac{d\vec{R}(t)}{dt} = \vec{v}(t)$ and $\frac{d^2\vec{R}(t)}{dt^2} = \vec{a}(t) = \frac{\vec{F}(t)}{2m}$ we can rewrite equations (3.73) and (3.74) as

$$\vec{R}(t + \Delta t) = \vec{R}(t) + \vec{v}\Delta t + \frac{1}{2!}\frac{\vec{F}(t)}{2m}\Delta t^2 + \frac{1}{3!}\frac{d^3\vec{R}(t)}{dt^3}\Delta t^3 + O(\Delta t^4) \tag{3.75}$$

$$\vec{R}(t - \Delta t) = \vec{R}(t) - \vec{v}\Delta t + \frac{1}{2!}\frac{\vec{F}(t)}{2m}\Delta t^2 - \frac{1}{3!}\frac{d^3\vec{R}(t)}{dt^3}\Delta t^3 + O(\Delta t^4). \tag{3.76}$$

Now, if we sum equations (3.75) and (3.76), and neglect higher order terms, we get

$$\vec{R}(t + \Delta t) \approx 2\vec{R}(t) - \vec{R}(t - \Delta t) + \frac{\vec{F}(t)}{2m}\Delta t^2. \tag{3.77}$$

This is the way how the Verlet algorithm updates the positions of the particles. Position updates are therefore independent from the velocities of the particles.

Leap-Frog algorithm. In contrast to Verlet's algorithm, Leap-Frog algorithm uses velocities at half-integer time steps, $\left(t - \frac{\Delta t}{2}\right)$ and $\left(t + \frac{\Delta t}{2}\right)$, to determine new particles' positions:

$$\vec{v}\left(t + \frac{\Delta t}{2}\right) = \vec{v}(t) + \frac{d\vec{v}(t)}{dt}\frac{\Delta t}{2} + \frac{1}{2!}\frac{d^2\vec{v}(t)}{dt^2}\frac{\Delta t^2}{4} + O(\Delta t^3) \tag{3.78}$$

$$\vec{v}\left(t - \frac{\Delta t}{2}\right) = \vec{v}(t) - \frac{d\vec{v}(t)}{dt}\frac{\Delta t}{2} + \frac{1}{2!}\frac{d^2\vec{v}(t)}{dt^2}\frac{\Delta t^2}{4} - O(\Delta t^3) \tag{3.79}$$

By subtracting (3.78) from (3.79) we get

$$\vec{v}\left(t+\frac{\Delta t}{2}\right)=\vec{v}\left(t-\frac{\Delta t}{2}\right)+\frac{\vec{F}(t)}{m}\Delta t+O(\Delta t^3) \tag{3.80}$$

and analogously for particles' positions

$$\vec{R}\left(t+\frac{\Delta t}{2}\right)=\vec{R}\left(t-\frac{\Delta t}{2}\right)+\vec{v}(t)\Delta t+O(\Delta t^3) \tag{3.81}$$

which, we can rewrite as

$$\vec{R}(t+\Delta t)=\vec{R}(t)+\vec{v}(t+\frac{\Delta t}{2})\Delta t+O(\Delta t^3) \tag{3.82}$$

Equation (3.82) and (3.80) provide the recipe for the Leap-Frog algorithm: Leap-Frog integration updates the positions and the velocities at interleaved time points, staggered in such a way that they 'leapfrog' over each other.

The Leap-Frog algorithm is algebraically equivalent to Verlet's algorithm, *i.e.* both algorithms provide identical trajectories. Moreover, both algorithms are numerically stable and time reversible. The last condition is crucial to enable the conservation of quantities such as energy and angular momentum. However, the Leap-Frog algorithm guarantees more accurate velocities, and consequently more accurate (kinetic) energy estimation of the dynamical system, and for this reason it is usually preferable.

3.1.4.2 Short-range interactions and neighbors list

The van der Waals term in equation (3.70) would require calculation of many pairwise interactions. However, since the range of interaction decreases very fast with the distance ($\sim 1/r^6$), it is computationally convenient to introduce a distance cutoff R_c, beyond which the interaction of two particles is not calculated and set to zero. The loss in accuracy is usually modest and it can become negligible with a correct choice of R_c and the use of the so-called long-range terms, which are constants added to the physical quantities, *a priori* calculated by considering the system distributed in the space with a uniform density.

To speed-up the calculation of the van der Waals term in (3.70), a list of particles within $R_c + \Delta_{list}$ (Δ_{list} is defined at the beginning of the simulation and is usually 1-2 Å) of each particle i is generated (neighbors lists). At each time step, a search for particles within R_c from particle i is performed from this list and only such particles are considered in the calculation of the short-range interactions contribution to F_i. Since particles moves and they can either leave or enter any sphere of radius $R_c + \Delta_{list}$ during the simulation, each neighbors list is updated after every certain number of steps (defined at the beginning of the simulation).

3.1.4.3 Periodic boundary conditions

Generally, in studying biological systems we do not consider our system as isolated particles (*i.e. in vacuo*), but rather in a (physiological) solution. However, due to computational limitations it is usually only possible to consider up to few hundred thousands of atoms in a single simulation. In these conditions, many particles would be at the edge of the system (on the boundary between system and vacuum), which would result in unrealistic surface effects.

To avoid the problem of the surface effects and to mimic the properties of bulk systems, special boundary conditions are introduced, in particular the periodic boundary conditions (PBC). When PBC are enforced, the simulation box represents a system surrounded by infinite number of replicas of itself. Still only N particles in the original simulation box are treated explicitly, namely when a particle exits the box on one side an image identical particle has to immediately appear on the other side.

PBC increase the number of interactions to be calculated. In the case of short-range interactions, *e.g.* van der Waals interactions, when a cutoff distance is introduced and the smallest side of the simulation box has a size larger than $2R_c$, it can be proved that each particle i can interact only with one of the infinite replicas of another particle j. Therefore, in MD simulation with PBC the so-called minimum image convention is applied to the particle searching algorithm for short-range interactions: only the closest image of the other $N-1$ particles is considered for such interactions.

On the other hand, long-range interactions cannot be effectively treated with cutoff approaches and all particle pairs has to be taken into consideration for a proper treatment of these interactions. If PBC are imposed on the system (formally simulating an infinite number of particles), correct calculation of long-range interactions require special techniques, such as the Ewald summation or the particle-mesh Ewald method described in the next section.

In addition, when defining the size of the simulation box, one must be careful that the macromolecule does not interact with its own periodic image, as this would create artifacts: ideally we would like to replicate the solvent molecules to mimic their bulk effects but we would not like to replicate the macromolecule(s) solute, unless to deal with very high (non-physiological) concentrations. For this reason, usually a length of the box vector larger than the sum of the length of macromolecule in that direction and twice the cutoff radius is employed, in order to minimize these unwanted interactions.

3.1.4.4 Long-range interactions

Pair-wise interactions are defined long-range when the force between the two particles at large distance falls to zero slower than r^{-d}, where d is the dimensionality of the system (in most of the common biological case $d = 3$). In this case we cannot use distance cutoff-based approximations and therefore interactions between all particles have to be taken into account. The Ewald summation (or its computationally faster variation particle-mesh Ewald method used in this thesis) is considered one of the best methods for treating long-range interactions in biological systems.

Ewald summation. This is a method to sum up electrostatic interactions in an infinite lattice, or, alternatively to compute electrostatic interactions under PBC [70]. In this cases the Coulomb potential can be written as

$$U_{Coulomb} = \frac{1}{2}\sum_{i,j=1}^{N}{\sum_{\vec{n}}}' \frac{q_i q_j}{\left|\vec{R}_{ij} + \vec{n}L\right|}\,, \tag{3.83}$$

where L is the box size, $\vec{n} = (n_1, n_2, n_3) = n_1 L\vec{x} + n_2 L\vec{y} + n_3 L\vec{z}$ is a lattice vector ($\vec{x}, \vec{y}, \vec{z}$ are unit vectors along the Cartesian axes), q_i and q_j are charges on particles i and j and $\sum_{\vec{n}}'$ means that for $n = 0$ the sum runs over $i \neq j$. $\left|\vec{R}_{ij} + \vec{n}L\right|$ is the distance between particle i in the original cell and the particle j in the image cell identified by vector $\vec{n}$. However, this sum converges slowly with n and therefore a very large number of images is required to achieve a reasonable estimation of the potential energy.

In the Ewald summation approach to the calculation of $U_{Coulomb}$, the charge distribution, $\rho(\vec{R})$, of the system is represented by an infinite set of point charges and every point charge is additionally surrounded with a Gaussian charge distribution of equal magnitude and opposite sign:

$$\rho_i^G(\vec{R}) = q_i \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} e^{-\alpha|\vec{R}_{ij} + \vec{n}L|^2}, \tag{3.84}$$

where α is the width of the Gaussian distribution (to be tuned in order to optimize the speed of convergence). This additional $\rho^G = \sum_i \rho_i^G(\vec{R})$ distribution acts like an ionic cloud, to screen the interaction between neighboring charges. But this screened charge distribution becomes short-ranged and the sum of interactions is now absolutely convergent. Since the total density of the system has to be only the physical $\rho(\vec{R})$, a third charge distribution identical to ρ^G but of opposite sign as to be added. However, this neutralizing distribution can be easily calculated in the reciprocal space where it is

absolutely convergent as well, and therefore both this term and the term associated to the density $\rho + \rho^G$ are convergent and suitable for numerical treatment.

Particle-mesh Ewald method (PME). The PME approach allows for a fast estimation of the electrostatic energy in PBC calculated according to the Ewald summation approach, by taking advantage of the fast Fourier transform (FFT) [56] in the calculation of discrete Fourier transforms. Here, the charges are assigned to a mesh using interpolation instead of directly summing wave vectors. Next, the grid is Fourier transformed using FFT algorithm and the reciprocal energy term is obtained by a single sum over the grid. The forces on each atom are obtained using the interpolation factors and the potential at the grid points is calculated by inverse transformations. Since PME scales as $Nlog(N)$ it is much faster than normal Ewald summation on medium and large systems.

3.1.5 Pressure and temperature control in MD simulations

It is important to perform MD simulations in the conditions that correspond to the experimental ones. This usually means that simulations should be performed keeping temperature and pressure constant for the entire run. Therefore, in order to mimic such experimental conditions, thermostat and barostat algorithms are used to control the temperature and pressure of the simulated system. In this thesis, equilibrium simulations were performed in NPT ensemble.

In the next sections, we introduce two different types of thermostats and barostats algorithms, the weak coupling and extended system methods, both used in this thesis but for different purposes.

3.1.5.1 Weak coupling method

Berendsen thermostat. In an experiment, the temperature is maintained constant by coupling the system to an external heat bath with constant temperature T_0. In the Berendsen approach [26], the temperature of the simulated system is corrected such that the deviation exponentially decays with some time constant τ_T:

$$\frac{dT(t)}{dt} = \frac{1}{\tau_T}[T_0 - T(t)], \tag{3.85}$$

where $T(t)$ is instantaneous temperature, T_0 is the target temperature. According to equipartition theorem, the kinetic energy of a system with $3N$ degrees of freedom is connected with the instantaneous temperature through the following expression

$$E_{kin} = \sum_I M_i \dot{\vec{R}}_i^2(t) = \frac{3}{2} N k_B T(t). \tag{3.86}$$

If the temperature is changed by δT at each (finite) interval of time δt, this results in a change of kinetic energy δE_{kin}

$$\delta E_{kin} = \frac{3}{2} N k_B \delta T(t) = \frac{3}{2} N k_B \frac{1}{\tau_T} [T_0 - T(t)]\, \delta t. \tag{3.87}$$

The idea in the Berendsen's approach is to achieve the kinetic energy change by rescaling the velocities of all the particles by a factor λ every δt, in order that

$$\delta E_{kin} = \sum_i M_i (\lambda^2 - 1) \dot{\vec{R}}_i^2(t). \tag{3.88}$$

Combining equations (3.86), (3.87) and (3.88) we get

$$\lambda = \sqrt{1 + \frac{\delta t}{\tau_T}\left(\frac{T_0}{T(t)} - 1\right)}. \tag{3.89}$$

The factor λ scales velocities every δt[i] in order to relax the temperature towards the target value T_0. The relaxation rate is controlled by the coupling constant τ_T.

Berendsen barostat. Pressure control is usually obtained by changing the dimensions of the simulation cell during the simulation. Similarly to the thermostat, in the Berendsen's approach to pressure control, the pressure of the simulated system is corrected such that the deviation exponentially decays with some time constant τ_P:

$$\frac{dP(t)}{dt} = \frac{1}{\tau_P} [P_0 - P(t)], \tag{3.90}$$

where $P(t)$ is instantaneous pressure and P_0 the target pressure. Rescaling the atomic positions by a scaling factor μ every a time interval δt, the new positions and consequently the volume of the box can be written as

$$x_{new} = \mu x_{old} \tag{3.91}$$

$$V_{new} = \mu^3 V_{old}. \tag{3.92}$$

[i] Please note that in general this is not the integration time step but much larger interval.

The pressure change δP after a volume change ΔV can be written as

$$\delta P(t) = -\frac{1}{\beta V}\Delta V, \tag{3.93}$$

where β is isothermal compressibility. Combining equations (3.90), (3.92) and (3.93) we get expression for μ

$$\mu = \sqrt[3]{1 - \frac{\beta\delta t}{\tau_P}\big(P_0 - P(t)\big)}. \tag{3.94}$$

3.1.5.2 Extended system method

This approach is based on the introduction of extra degrees of freedom, *S*, in the system, which can mimic the effect of the coupling with an heat/pressure bath. Therefore, in a Lagrangian formulation, a new extended Lagrangian has to be written in order to deal with the full (extended) dynamical system.

Nosé-Hoover thermostat. The extended Lagrangian in the Nosé approach [179] reads

$$L_{Nose} = \frac{1}{2}\sum_{i=1}^{N} M_i S^2 \dot{\vec{R}}_i^2 - U\big(\{\vec{R}_i\}\big) + \frac{Q}{2}\dot{S}^2 - g k_B T \ln S, \tag{3.95}$$

where *g* is the number of independent momentum degrees of freedom and *Q* is an effective mass associated with *S*, which should be chosen carefully along with systems. The conjugate momenta of $\vec{R}_i$ and *S* are

$$\vec{P}_i \equiv \frac{\partial L_{Nose}}{\partial \dot{\vec{R}}_i} = M_i S^2 \dot{\vec{R}}_i, \tag{3.96}$$

$$\vec{P}_S \equiv \frac{\partial L_{Nose}}{\partial \dot{S}} = Q\dot{S}. \tag{3.97}$$

The extended Hamiltonian for the extended system of *N* particles and the additional degree of freedom *S* can now be defined as

$$H_{Nose} = \sum_{i=1}^{N} \frac{\vec{P}_i^2}{2M_i S^2} + U\big(\{\vec{R}_i\}\big) + \frac{\vec{P}_S^2}{2Q} + g k_B T \ln S. \tag{3.98}$$

and the corresponding partition function of the microcanonical (NVE) ensemble can be written as

$$Z_{Nose} = \frac{1}{N!}\int \delta(E - H_{Nose}) dP_S dS d\vec{P} d\vec{R}. \tag{3.99}$$

By defining the scaled momentum $\vec{P}'$ as

$$\vec{P}' = \frac{\vec{P}}{S}. \tag{3.100}$$

The partition function simplifies as

$$Z_{Nose} \propto \frac{1}{N!}\int e^{\left[-\frac{3N+1}{k_B T g} H(\vec{P}', \vec{R})\right]} d\vec{P}' d\vec{R}, \tag{3.101}$$

where

$$H(\vec{P}', \vec{R}) = \sum_{i=1}^{N} \frac{\vec{P}_i'^2}{2M_i} + U(\{\vec{R}_i\}). \tag{3.102}$$

If $g = 3N + 1$, then Z_{Nose} corresponds to the partition function of a canonical (NVT) ensemble for the system with degrees of freedom $\{\vec{R}_i\}$ and $\{\vec{P}_i'\}$. $\vec{P}'$ can be interpreted as the "physical" momentum and $\vec{P}$ as the "virtual" one. If the physical variables are primed, they are related to the virtual (unprimed) ones as follows:

$$\vec{R}' = \vec{R}, \tag{3.103}$$

$$\vec{P}' = \frac{\vec{P}}{S}, \tag{3.104}$$

$$S' = S, \tag{3.105}$$

$$\Delta t' = \frac{\Delta t}{S}. \tag{3.106}$$

Now, we can obtain the equations of motion for the virtual variables by applying the standard Hamilton equations with the Hamiltonian (3.98):

$$\frac{d\vec{R}_i}{dt} = \frac{\partial H_{Nose}}{\partial \vec{P}_i} = \frac{\vec{P}_i}{M_i S^2}, \tag{3.107}$$

$$\frac{d\vec{P}_i}{dt} = -\frac{\partial H_{Nose}}{\partial \vec{R}_i} = -\frac{\partial U(\{\vec{R}_i\})}{\partial \vec{R}_i}, \tag{3.108}$$

$$\frac{dS}{dt} = \frac{\partial H_{Nose}}{\partial P_S} = \frac{P_S}{Q}, \tag{3.109}$$

$$\frac{dP_S}{dt} = -\frac{\partial H_{Nose}}{\partial S} = \left(\sum_i \frac{\vec{P}_i^2}{M_i S^2} - g k_B T\right) S^{-1}. \tag{3.110}$$

By using the eq. (3.103-3.106) the equations of motion for the physical variables become

$$\frac{d\vec{R}'_i}{dt'} = \frac{\vec{P}'_i}{M_i}, \tag{3.111}$$

$$\frac{d\vec{P}'_i}{dt'} = -\frac{\partial U(\{\vec{R}'_i\})}{\partial \vec{R}'_i} - \left(\frac{S' P'_S}{Q}\right) \vec{P}'_i, \tag{3.112}$$

$$\frac{1}{S}\frac{dS'}{dt'} = \frac{S' P'_S}{Q}, \tag{3.113}$$

$$\frac{d\left(\frac{S' P'_S}{Q}\right)}{dt'} = \left(\sum_i \frac{\vec{P}_i^2}{M_i} - g k_B T\right) Q^{-1}. \tag{3.114}$$

Hoover [102] simplified these equations by introducing the thermodynamic friction coefficient

$$\xi = \frac{S' P'_S}{Q}. \tag{3.115}$$

This way, after dropping the primes equations (3.111-3.114) can be finally written as

$$\dot{\vec{R}}_i = \frac{\vec{P}_i}{M_i}, \tag{3.116}$$

$$\dot{\vec{P}}_i = \frac{\partial U(\{\vec{R}_i\})}{\partial \vec{R}_i} - \xi \vec{P}_i, \tag{3.117}$$

$$\frac{\dot{S}}{S} = \frac{d \ln S}{dt} = \xi, \tag{3.118}$$

$$\dot{\xi} = \left(\sum_i \frac{\vec{P}_i^2}{M_i} - g k_B T\right) Q^{-1}. \tag{3.119}$$

That is the form usually implemented in the MD software packages.

Andersen barostat. In MD the pressure at a specific point in time can be calculated as

$$P(t) = \frac{2}{3V}(E_{kin} - \Xi), \tag{3.120}$$

where E_{kin} is the kinetic energy and Ξ is the inner virial tensor defined as

$$\Xi = -\frac{1}{2}\sum_{i<j} \vec{R}_{ij}\vec{F}_{ij}. \tag{3.121}$$

where $\vec{F}_{ij}$ is force between particles *i* and *j*. Therefore, during a simulation pressure can be controlled through a change in the inner virial tensor by rescaling the distances between particles $\vec{R}_{ij}$.

In the Andersen barostat method [9], the coordinates $\{\vec{R}_i\}$ are replaced by scaled coordinates $\{\vec{R}'_i\}$ defined as

$$\vec{R}'_i = \frac{\vec{R}_i}{V^{1/3}}. \tag{3.122}$$

Andersen wrote a new Lagrangian, where the volume *V* appears as an additional dynamical variable:

$$L_{AN} = \frac{1}{2}V^{\frac{2}{3}}\sum_i M_i\dot{\vec{R}}'^2_i - \sum_{i<j} U\left(V^{\frac{1}{3}}R'_{ij}\right) + \frac{1}{2}\mu\dot{V}^2 - p_0V, \tag{3.123}$$

where the first two terms are the Lagrangian of the unscaled system, the third term represents the kinetic energy for the motion of *V* and the fourth term is the potential energy associated with *V*. p_0 and μ are the target pressure and the pressure coupling constant, respectively.

Now, we can write the Hamiltonian and equations of motion for the scaled variables, and by mapping back to the unscaled variables we get the new equations of motion

$$\frac{d\vec{R}_i}{dt} = \frac{\vec{P}_i}{M_i} + \frac{1}{3}\vec{R}_i\frac{d\ln V}{dt} \tag{3.124}$$

$$\frac{d\vec{P}_i}{dt} = -\sum_{i<j}\vec{R}_{ij}U'\left(R_{ij}\right) - \frac{1}{3}\vec{P}_i\frac{d\ln V}{dt} \tag{3.125}$$

$$\mu\frac{d^2V}{dt^2} = p_0 + \left[\frac{2}{3}\sum_i\frac{\vec{P}_i^2}{2M_i} - \frac{1}{3}\sum_{i<j}\vec{R}_{ij}U'\left(R_{ij}\right)\right]V^{-1} \tag{3.126}$$

implemented in the MD codes.

Parrinello-Rahman barostat. Andersen method was extended by Parrinello and Rahman [188] to allow the simulation box to change its shape. The shape of the cell is

defined by the vectors $\vec{a}, \vec{b}$ and $\vec{c}$ and is completely arbitrary. The volume of the cell can be then written as

$$V = \det \boldsymbol{h} = \vec{a} \cdot \left(\vec{b} \times \vec{c}\right), \tag{3.127}$$

where **h** is a 3 matrix, whose columns are the three vectors $\vec{a}, \vec{b}$ and $\vec{c}$.

The position $\vec{R}_i$ of the particle i, can be written in terms of **h** and a column vector $\vec{S}_i = [\xi_i \, \eta_i \, \zeta_i]$ with $0 \leq \xi_i, \eta_i, \zeta_i \leq 1$

$$\vec{R}_i = \mathbf{h}\vec{S}_i = \xi_i \vec{a} + \eta_i \vec{b} + \zeta_i \vec{c}. \tag{3.128}$$

The squared distance between particles *i* and *j* can be therefore rewritten as

$$R_{ij}^2 = S_i^T \mathbf{G} S_j, \tag{3.129}$$

where **G** is the metric tensor

$$\mathbf{G} = \mathbf{h}^T \mathbf{h}. \tag{3.130}$$

Now the extended Lagrangian of Anderson can be cast as

$$L_{PS} = \frac{1}{2} \sum_i M_i \dot{S}_i^T \mathbf{G} \dot{S}_i - \sum_{i<j} U\left(R_{ij}\right) + \frac{1}{2} \mu \mathrm{Tr}\left(\dot{\mathbf{h}}^T \dot{\mathbf{h}}\right) - p_0 V \tag{3.131}$$

and the corresponding equations of motion are derived in a similar way to the isotropic case of Andersen.

Weak coupling methods are good in the equilibration phase when the system is far from equilibrium and the aim is to reach an equilibrated configuration as fast as possible. However, they do not provide a correct probability distribution. On the other hand, extended system methods are able to provide also correct probability distributions and therefore they are the most reliable ones for production run simulations, when the system is already at equilibrium and thermodynamic properties has to be predict.

In this thesis, we used Berendsen thermostat for pre-equilibrium simulations and Nosé-Hoover thermostat for equilibrium calculations. For pressure control, the Parrinello-Rahman barostat was applied in all the simulations.

3.1.6 Hybrid quantum mechanics/molecular mechanics (QM/MM) methods

To study chemical reactions or metal complexes in large biomolecular systems, a treatment of the entire system at quantum level would be computationally too expensive. However, the region where a quantum mechanical treatment is necessary usually consists of at most few hundreds of atoms. The rest of the system can usually be described at a lower level of theory, such as the classical one. In these cases, hybrid QM/MM approach [54, 212, 225] can be used. In this approach, the system is partitioned in two parts, one treated at quantum level (QM part) and the other treated at the force field level (MM part).

In our thesis we use the Car-Parrinello MD scheme for the quantum part, which was already described in the section 3.1.3.3 of this chapter.

More specifically, I used the QM/MM approach [134] developed by the group of Prof. Röthlisberger. This approach can be described through the extended Lagrangian:

$$\begin{aligned} L_{QM/MM} &= L_{CP} - E_{MM} - E_{QM/MM} \\ &= \frac{1}{2}\sum_I M_I \dot{\vec{R}}_I^2 \\ &+ \frac{1}{2}\mu \sum_i \int \dot{\varphi}_i^*(\vec{r})\dot{\varphi}_i(\vec{r})d\vec{r} - E_{KS}[\rho(\vec{r})] \\ &+ \sum_{ij} \Lambda_{ij}\left[\int \varphi_i^*(\vec{r})\varphi_i(\vec{r})d\vec{r} - \delta_{ij}\right] - E_{MM} - E_{QM/MM}, \end{aligned} \tag{3.132}$$

where E_{MM} is the force field-based energy function describing the MM part (here I used AMBER force field described in the equation (3.79)) and $E_{QM/MM}$ is the coupling term between the QM and MM parts.

The coupling term can be furthermore expanded as

$$E_{QM/MM} = E_{QM/MM}^{ele} + E_{QM/MM}^{vdw} + E_{QM/MM}^{bonded}, \tag{3.133}$$

where $E_{QM/MM}^{vdw}$ and $E_{QM/MM}^{bonded}$ describe the van der Waals and bonded interactions between QM and MM atoms and $E_{QM/MM}^{ele}$ stands for electrostatic coupling between the quantum charge distribution and the classical point charges:

$$E_{QM/MM}^{ele} = \sum_{i \in MM} q_i \int \rho(\vec{r}) v_i(|\vec{r} - \vec{r}_i|)d\vec{r}, \tag{3.134}$$

where q_i are the partial charges of the classical atoms at $\vec{r}_i$ and $v_i(|\vec{r} - \vec{r}_i|)$ is a modified Coulomb potential. In particular, this potential is modified in its short range

behavior, in order to avoid the so-called electron spill-out phenomenon, *i.e.* the unphysical accumulation of quantum charge density at the boundary of the QM box due to the presence of classical positive charges of the nearby MM region [134]

$$v_i = \frac{r_{ci}^n - r^n}{r_{ci}^{n+1} - r^{n+1}}, \tag{3.135}$$

where n is usually fixed to 4 and r_{ci} is the covalent radius of atom i. Therefore, $v_i(\vec{r})$ behaves as $\frac{1}{r}$ for large distances and goes smoothly to a constant value for small r.

Really, the calculation of the integrals in equation (3.134) is computationally expensive and therefore they are calculated only for MM atoms in the first shell, *i.e.* in the shell which lies within a cut-off radius (usually ~10 Å) from the QM/MM boundary. A second shell formed by MM atoms beyond the first shell but within a second cut-off radius (usually ~20 Å), is coupled to the QM part by the D-RESP (Dynamically Generated Potential Derived Charges) charge scheme [135]. In particular, D-RESP charges for the QM atoms are calculated by the wavefunction of the QM part. The contribution to the expression in (3.134) due to the MM atoms in the second shell is estimated by the electrostatic interactions between the partial charges of the MM atoms and the D-RESP charges of the QM part. Finally, interaction between the QM atoms and the rest of the MM atoms is explicitly kept into account by electrostatic interaction between multipole expansion of QM charge density and the classical point charges of the MM atoms.

Covalent bonds cut on the QM/MM boundary should always be non-polar (C-C) bonds. In order to saturate the dangling bonds and preserve the electronic structure of the QM part, I used an adapted monovalent carbon pseudopotential for the carbon atom in the C-C bond outside the QM part [252]. The bonded and van der Waals interactions between QM and MM regions are described by the AMBER force field.

3.2 Protein-ligand docking

Protein-ligand docking may be defined as an optimization problem, which would describe the “best-fit” orientation of a ligand that binds to a particular protein of interest. As a result of docking calculation, one or more best poses of the molecule inside the binding site and the corresponding binding affinities (or better score function values providing a measure of the strength of binding) are obtained.

3.2.1 Thermodynamics of non-covalent and reversible protein-ligand interactions

In docking, one usually takes into account only non-covalent protein-ligand interactions. Therefore, the process of ligand binding to the protein is a reversible process and can be sketched as an equilibrium reaction

$$P + L \rightleftharpoons PL, \tag{3.136}$$

where P is the protein and L is the ligand. The reaction rate constant for forming the complex PL is k_1 and k_{-1} is the reaction rate for the dissociation of the ligand from the protein. When the binding process is at equilibrium, we can define the binding constant, K_b

$$K_b = \frac{k_1}{k_{-1}} = \frac{[PL]}{[P][L]}, \tag{3.137}$$

where the square brackets represent molar concentrations. The binding constant is related to the free energy of binding through the equation

$$\Delta G = \Delta G^0 + RT \ln K_b. \tag{3.138}$$

At steady state $\Delta G = 0$ and therefore $\Delta G^0 = -RT \ln K_b$. Since in most cases we are interested in the value of the dissociation constant (called also inhibition constant) defined as

$$K_d = K_i = \frac{1}{K_b}, \tag{3.139}$$

we can now rewrite equation (3.138) at steady state condition as

$$\Delta G^0 = RT \ln K_i. \tag{3.140}$$

Free energy of binding consists of two contributions, enthalpic and entropic one. The enthalpic contribution includes the contribution from the hydrogen bonds, the ionic and polar interactions, the aromatic (stacking) interactions and the van der Waals interactions. The entropic contribution depends on the variation of the degrees of freedom for the entire system, which, as I will explain in the next section, docking algorithms model by rotatable bonds.

3.2.2 Docking calculations

Typical docking calculations consist of two major parts:

I) the search algorithm, which generates a large number of binding poses of the ligand inside the binding site;
II) the scoring function, which calculates a score (or binding affinity) for each particular pose.

While docking algorithms can be separated in two classes, rigid-body and flexible-body docking, only flexible-body docking is described here as it is the approach that usually yields to better results and it was the method chosen in this thesis.

In the flexible-body docking the flexibility of the ligand is taken into account through the torsional rotation around rotatable bonds, while the protein is still considered as (nearly) rigid. There are three main algorithms used for the flexible-body docking:

- Incremental construction algorithm
- Evolutionary algorithm
- Monte Carlo algorithm

The incremental construction algorithm [138] consists of three main steps. In the first step the ligand is broken into fragments and anchor group(s) (functional groups important for ligand binding) in the protein's binding site are selected. In the next step, the position of the first fragment is determined based on its affinity towards the anchor groups and *n* best poses are selected. Then, the second fragment is added to the fragment, which is already located in the binding site, and again *n* best poses are selected. This step is repeated until the entire ligand is reconstructed inside the binding site. For a schematic representation see also Fig. 3.1.

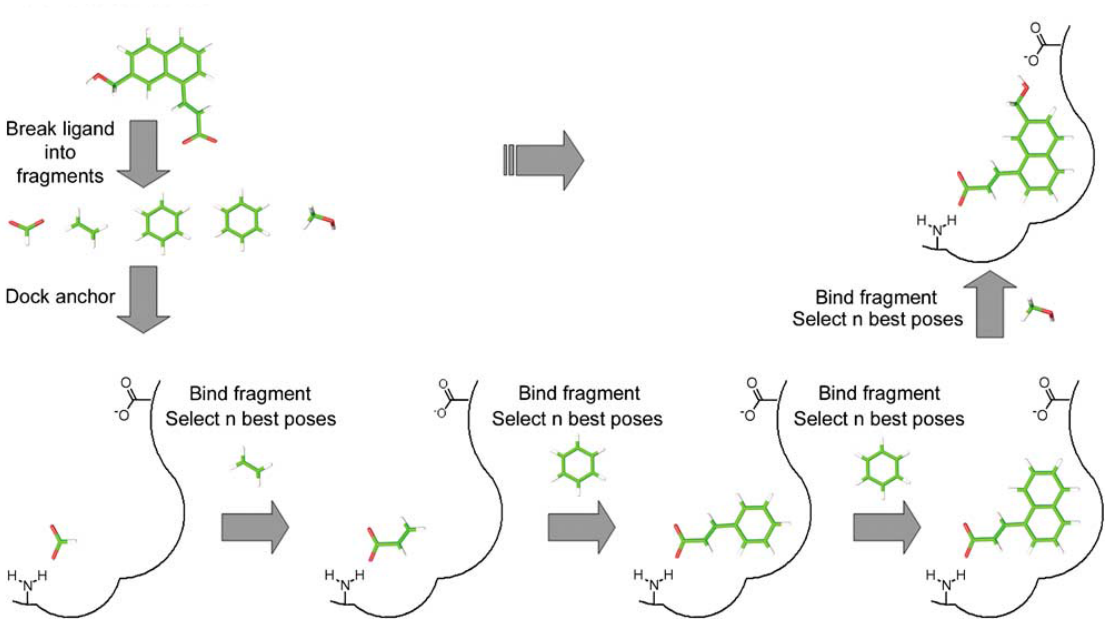


Figure 3.1: Schematic representation of incremental construction. The figure is taken from the reference [6].

The evolutionary algorithm (also known genetic algorithm) is based on ideas inspired by the Darwinian evolution [114]. Initial population of chromosomes (each chromosome represents a ligand's pose inside the binding site) is generated randomly. Pairs of high-scoring chromosomes (*i.e.* "parents") are combined to generate "children" (*e.g.* pose combined with torsions). Children are then randomly mutated by torsions, translation or rotation. Highest scorings of the new poses are then combined with the highest scoring of the original poses to make a new generation. This process is repeated for N generations or until no significant improvement is observed. For schematic representation see also Fig. 3.2.

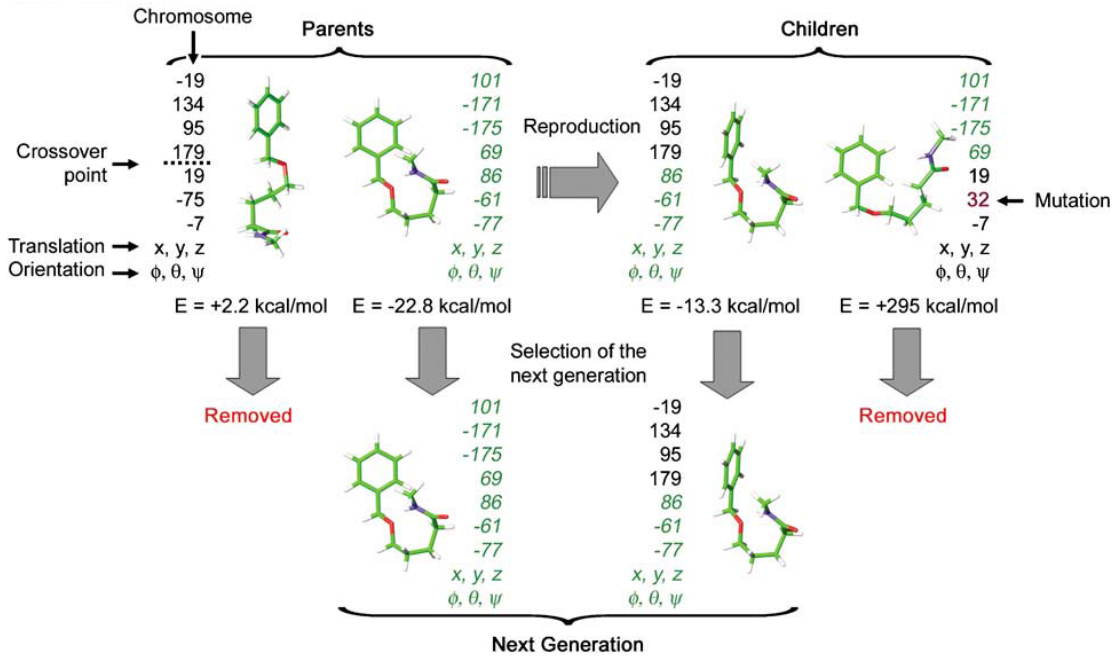


Figure 3.2: Schematic representation of genetic algorithm. The figure is taken from the reference [6].

The third algorithm used in the flexible-body docking is based on the Metropolis Monte Carlo method [166], in which the outcome of each trial depends only on the preceding trial and acceptance of the new trial depends on the Metropolis method. More specifically, in the first step, an initial conformation of the ligand is generated and evaluated in terms of free energy. Next, this orientation is subject to translational, rotational and torsional changes and again energetically evaluated. If the difference in free energies, ΔE, defined as

$$\Delta E = E_{new} - E_{old}, \tag{3.141}$$

is negative, then the new conformation is accepted. Otherwise its acceptance is determined using following equation

$$\mathrm{z} \leq e^{-\frac{\Delta E}{k_B T}}, \tag{3.142}$$

where z is a randomly generated number between 0 and 1 and the term on the right side of the disequation is called the Boltzmann factor. In the Boltzmann factor k_B is the Boltzmann constant and T is the temperature. If the randomly generated number is lower than the Boltzmann factor, then the new conformation is accepted. This process is repeated over N interactions. For schematic representation see also Fig. 3.3.

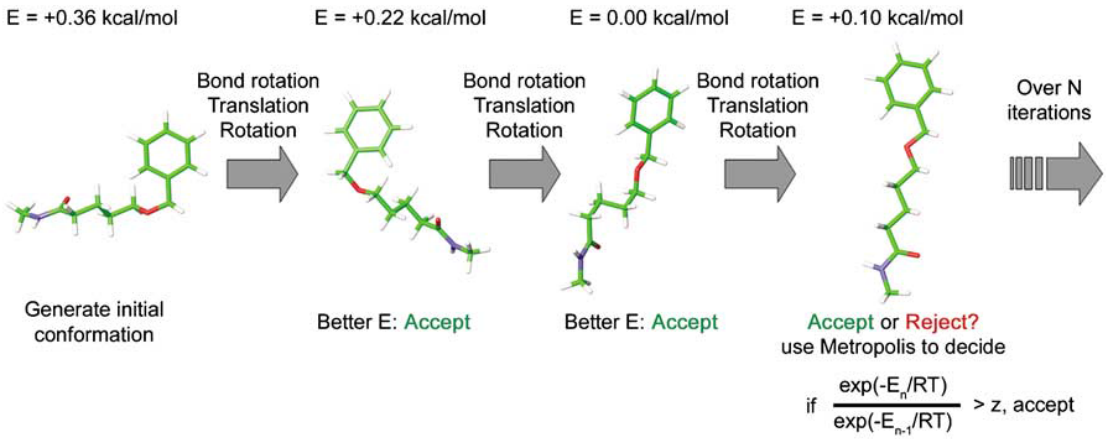


Figure 3.3: Schematic representation of Monte Carlo docking algorithm. The figure is taken from the reference [6].

While the incremental construction is an exhaustive but also computationally expensive method, the other two approaches are stochastic and computationally less demanding algorithms. In our work we used the evolutionary algorithm to determine the ligand orientation inside the binding site.

The second part of each docking calculation is the evaluation of the binding pose using a scoring function (SF). SF is a fast mathematical tool to estimate the strength of the interactions between the ligand and the protein. SFs can be classified into four main classes [151]:

1. Physics-based SF, where terms for calculation of ligand-protein interactions can be taken either from force field or from quantum mechanical calculations. Additionally, a (de)solvation energy term is usually calculated using the

Poisson-Boltzmann [82] or the Generalized Born [268] continuum solvent methods. The general functional form for physics-based SF reads

$$\Delta G_{binding} = \Delta E_{vdw} + \Delta E_{electrostatic} + \Delta E_{hbond} + \Delta G_{desolvation}, \quad (3.143)$$

where the first term represents the van der Waals interactions, the second term describes the electrostatic interactions, the third term takes into account hydrogen bonds and the last term is the desolvation energy.

2. Empirical SFs, which evaluate the fitness of ligand-protein interactions by summing up the rewarding (favoring ligand binding) and penalizing (disfavoring ligand binding) contributions of a number of individual terms, each representing an important energetic factor in binding. One example of empirical SF is ChemScore [67, 173] implemented in GOLD docking software [246], which was also used in some dockings in this thesis. ChemScore SF reads

$$ChemScore = S_{hbond} + S_{metal} + S_{lipophilic} + P_{rotor} + P_{strain} + P_{clash} + [P_{covalent} + P_{constraint}]. \quad (3.144)$$

Rewarding scores for hydrogen bonds, coordinated bonds and lipophilic contacts are denoted with S and penalties for frozen rotatable bonds, internal strain energy of the ligand and steric clashes are denoted with P. If covalent docking is required additional penalties are taken into account. Each term in the equation (3.144) does not contribute equally to the final value and therefore weight factors have to be applied. Normally, these factors are derived using the multivariate linear regression or the partial least-square analysis from the training set of ligand-protein complexes with known three-dimensional structures and binding affinity data.

3. Knowledge-based SF, where the basic principle is that occurrence frequency of a pairwise contact is assumed to be a measure of its energetic contribution to the ligand-protein binding. The summation of pairwise statistical potentials between ligand and protein is

$$A = \sum_{i}^{ligand} \sum_{j}^{protein} \omega_{ij}(r), \quad (3.145)$$

where $\omega_{ij}(r)$ is distance-dependent potential between atom pair *i-j* and it is derived from an inverse Boltzmann analysis

$$\omega_{ij}(r) = -k_B T \ln\left[g_{ij}(r)\right] = -k_B T \ln\left[\frac{\rho_{ij}(r)}{\rho_{ij}^{*}}\right]. \tag{3.146}$$

4. Descriptor based SFs introduces quantitative structure-activity relationship analysis into ligand-protein interaction. The idea behind descriptor based SFs is that properties of the ligand, the protein and their interaction patterns can be coded with certain descriptors, and from those descriptors machine-learning techniques can be applied to derive statistical models that compute ligand-protein binding scores. Also this SF needs a training set of ligand-protein complexes with known structures and binding data to derive the final models.

For the main part of the docking calculations in our work we used AutoDock scoring function [75], which is a physics-based SF and it is defined as

$$\begin{aligned}\Delta G = W_{vdW}\sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + W_{hbond}\sum_{i,j}E(t)\left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) \\ + W_{elec}\sum_{i,j}\frac{q_i q_j}{e(r_{ij})r_{ij}} + W_{tor}N_{tor} + W_{solv}\sum_{i,j}\left(S_i V_j + S_j V_i\right)e^{-\frac{r_{ij}^2}{2\sigma^2}}.\end{aligned} \tag{3.147}$$

The first term is the typical 6-12 potential for repulsion interactions (as in equation (3.70)) with parameters based on the AMBER force field. The second term is a directional hydrogen bond term based on 10-12 potential with parameters *C* and *D* assigned to give a maximal well depth of 5 kcal/mol at 1.9 Å for hydrogen bonds with oxygen and nitrogen, and a well depth of 1 kcal/mol at 2.5 Å for hydrogen bonds with sulfur; *E(t)* provides directionality based on the angle *t* from the ideal hydrogen bonding geometry. The third term is the screened Coulombic potential that described the electrostatics [165], where q_i are the atomic charges. The fourth term takes into account entropic contributions due to the loss of degrees of freedom upon binding. The last term is a desolvation potential based on the volume of atoms *V* that surround a given atom and shelter it from solvent, weighted by solvation parameter *S* and an exponential term with distance-weighting factor $\sigma = 3.5$ Å. Indexes *i* stand for ligand atoms and *j* for protein atoms. The weighting constant *W* have been optimized to calibrate the empirical free energy based on a set of experimentaly determined binding constants and r_{ij} is the distance between atoms *i* and *j*.

3.3 Ligand-based virtual screening

In this thesis, ligand-based virtual screening was used to screen a large library of compounds for those substrates that have structure similar to the already known substrates of the hCN1. Since we had the structure of only two substrates, which differ from each other only in the length of the alkyl chain, we used screened library of lead molecules (~4,500,000 compounds) obtained from ZINC database for an exact structural motive, which I believe it is important for enzyme inhibition.

4 Structural characterization of carnosine and homocarnosine

4.1 Introduction

Carnosine plays an important role in the biology of a human body. Its biochemical properties and their connection with possible therapeutic application of carnosine and its natural derivatives have already been covered in the second chapter. Here are presented results from my quantum chemical calculations and classical MD simulations aimed to better understand carnosine's and homocarnosine's structural and spectroscopic properties and their behavior in water environment.

Carnosine is highly water soluble dipeptide (1 g in 3.1 mL of water at 25°C) and has three ionizable groups, C- and N-terminus and imidazole ring with pK_a avalues of 2.76, 9.32 and 6.72 [251]. Based on the pK_a values of amino and carboxyl groups it is mainly found in its zwitter-ionic form at physiological conditions. Due to the pK_a value of the imidazole ring of 6.72, which is close to the physiological pH of 7.4, it can exist either in the neutral or ionized form. Indeed, IR and Raman studies of Torreggiani *et al*. [238] showed that at pH 7 both forms, ionized and non-ionized, are present, though the ratio is in favor of the non-ionized form. When in neutral form, proton can be found at either of the two imidazole nitrogen atoms, N_π or N_τ, resulting in tautomers

CT-I and **CT-II** (see Chart 4.1).[j] Experimental studies show the ratio between **CT-I** and **CT-II** is in range between 2:1 [76] and 3:1 [238].

Since the pK_a values of ionizable groups in homocarnosine have similar values to the one of carnosine (2.75, 9.88 and 6.79 in homocarnosine compared to 2.76, 9.32 and 6.72 in carnosine) [251], similar distribution of its tautomeric forms (**HT-I** and **HT-II**) is expected (see Chart 4.1).

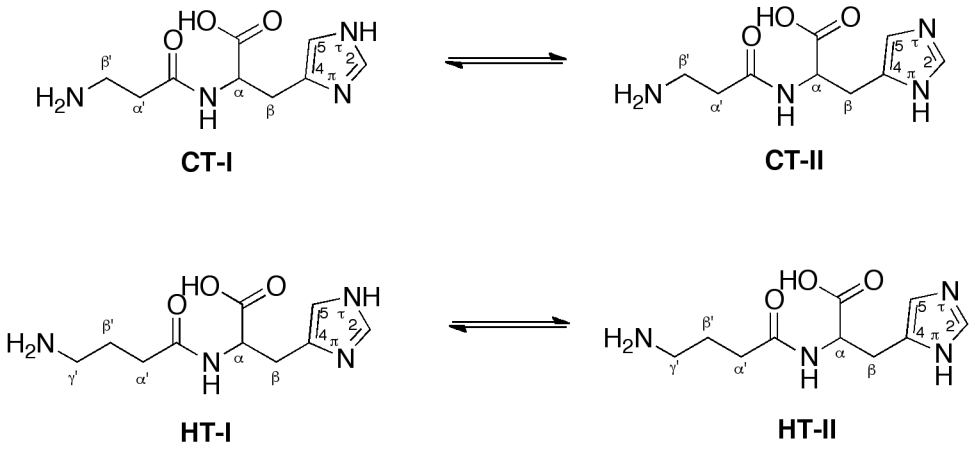


Chart 4.1: Tautomeric forms of carnosine (above) and homocarnosine (bellow). All tautomers are shown in their non-ionized form.

Using quantum chemical calculations and classical molecular dynamics I characterized carnosine and homocarnosine structural properties and compared them with available experimental data and previous computational studies. Moreover, I calculated ^{1}H and ^{13}C chemical shifts and compared them with experimental ones. Good agreement of all my results with readily available experimental and theoretical data confirmed the quality of my model.

4.2 Methods

First the four tautomers (**CT-I**, **CT-II**, **HT-I** and **HT-II** in Chart 4.1) were built using Molden [220] and underwent geometry optimization using B3LYP functional [21] and

[j] In some older literature N_π or N_τ were named also N1 and N3, however to avoid any confusion the former nomenclature is used here.

6-31G(d,p) basis set in aqueous solution. Water environment was treated implicitly [167] using IEF-PCM type of implicit solvent and with dielectric constant of water set at ε = 78.3553. The atomic radii were scaled by factor 1.1 from those of the UFF force field. Subsequent frequency calculations were used for free energies calculations as well as to show that the optimized structures correspond to local minima on the potential energy surface. The latter was confirmed by the absence of imaginary frequencies [236].

In the next step we performed μs-long classical MD calculations of **CT-I**, **CT-II**, **HT-I** and **HT-II** in aqueous solution at physiological pH and room temperature. All tautomer include one non-standard amino acid (β-alanine in carnosine and GABA in homocarnosine) which are not included in the basic classical force field libraries and therefore we parameterized them ourselves. The partial charges of all tautomers were derived from the *ab initio* calculations using the RESP algorithm [19]. The charges were fitted on the electrostatic potential grid calculated by CHelpG scheme [37] using Antechamber module (see Tab. A2.1 in Appendix 2) [255]. The parameters for β-alanine and GABA were obtained with the Antechamber module [255], using as input the bond lengths and bond angles data from the DFT calculations (see Tab. A2.2 and A2.3 in Appendix 2). Simulations were based on the AMBER99SB [104] force field with ILDN modification [150] and the TIP3P potential was used for water [116]. Each of the four tautomers was solvated in a truncated octahedron shaped box with the minimal distance of 12 Å between the solute and each boundary of the box. Periodic boundary conditions were applied. All simulations were performed at 298 K and 1 bar by coupling to Nosé-Hoover thermostat [102, 178] and Parrinello-Rahman barostat [189]. The temperature of 298 K was chosen because so that we could compare our results with experimentally obtained NMR data [264] as well as with the only so-far available MD calculations of carnosine in water [264]. The long range electrostatic interactions were taken into account using particle mesh Ewald method [56]. Time step in simulations was 1 fs. Each system was heated up to the final temperature of 298 K by using 20 steps of simulated annealing. In the first step, the temperature was increased from 0 to 10 K in 50 ps. In the following 8 steps, the temperature was increased for 10 K in each step of 25 ps. In the last 11 steps, the system was heated 20 K in each step of 25 ps.

The convergence of sampling in our simulations was checked using the cosine content of principal components (PCs), which is an indicator for sampling in MD simulations [94]. A cosine content close to 1 indicates bad sampling, while a value close to 0 suggests that the simulated time was long enough to sufficiently sample the system. The Hess analysis of both carnosine and homocarnosine MD simulations suggests that

the latter is the case in our simulations (see Tab. 4.1) Further independent simulations for both **CT-I** and **CT-II** were performed to test the robustness of our computational protocol.

Table 4.1: Cosine content of the first four principal components for **CT-I** and **CT-II** for the two independent simulations of each tautomer and for one simulation of **HT-I** and **HT-II**.

Principal component	Cosine content of principal component					
	CT-I		**CT-II**		**HT-I**	**HT-II**
1	$7.3*10^{-5}$	$1.5*10^{-5}$	0.0011	0.0019	0.0017	0.0007
2	0.0005	$4.6*10^{-5}$	$9.8*10^{-6}$	0.0020	$2.4*10^{-7}$	$6.7*10^{-5}$
3	0.0002	$1.4*10^{-5}$	0.0001	$4.8*10^{-6}$	$2.7*10^{-5}$	0.0003
4	$6.9*10^{-5}$	$2.6*10^{-6}$	$1.4*10^{-6}$	0.0008	$3.4*10^{-6}$	0.0006

The trajectories were clustered based on their RMSD value using the gromos method [57] implemented in the g_cluster analysis tool of the Gromacs software [96, 243]. This clustered the MD trajectory based on the RMSD values of the ligands' atoms. First, neighbors of each data point (*i. e.* snapshot) are defined based on a selected RMSD cutoff value (in my case 1.1 Å for carnosine and 1.3 Å for homocarnosine). Next, the data point with the largest number of neighbors is defined as the central structure of the first cluster, then this point and its neighbors defining this cluster are removed. The algorithm is finally iterated until all the data points have been assigned to a cluster. By definition, the central point of each cluster is the one with the most neighbors (within a cutoff distance). Therefore, this point is a representative of all its neighbors within the given cut-off distance.

To structurally characterize both compounds the following properties were calculated: (i) the square of the gyration radius defined as

$$\boldsymbol{R}_g^2 = \sum_i \left(\frac{\left(\boldsymbol{r}_i - \boldsymbol{r}_g\right)^2}{N} \right) \tag{4.1}$$

where $\boldsymbol{r}_i$ is the position of the atom i, $\boldsymbol{r}_g$ is position of the molecular center and N is number of atoms [264]. (ii) The free energies of representative structures of four highly populated clusters from **CT-I** and **CT-II** for carnosine (models **CT-Ia-d** and **CT-IIa-**

d) and from **HT-I** and **HT-II** for homocarnosine (models **HT-Ia-d** and **HT-IIa-d**) were calculated as single point energies using B3LYP functional [21] and 6-311+G(2d,p) basis set in implicit solvent [167]. This basis set was used because the protocol for the chemical shifts calculations [46], described in the next paragraph, requires the use of this basis set. So obtained free energy values were used in the calculation of Maxwell-Boltzmann statistics using the following equation

$$\frac{\langle N_X \rangle}{\langle N_0 \rangle} = \frac{e^{-\frac{E_X}{k_B T}}}{e^{-\frac{E_0}{k_B T}}} \tag{4.2}$$

where X is naming of the model, 0 stands for the lowest energy conformation, *i.e.* model **CT-Ia** for carnosine and **HT-Ib** for homocarnosine, $\langle N_X \rangle$ is the number of particles in the state X, E_X is the energy of corresponding conformation, k_B is Boltzmann constant and T is temperature (T = 298 K). (iii) ^{1}H and ^{13}C chemical shifts of each representative structure of the four highly populated clusters from all four tautomers were calculated. We followed the procedure of the reference [46]. First, structures of NMR experimental standards (tetramethylsilane (TMS) for ^{1}H and ^{13}C and sodium [2,2,3,3-2H4]-3-trimethylsilylpropanoate for ^{1}H in carnosine and sodium 2,2-dimethyl-2-silanpentane-5-sulfonate (DSS) for ^{1}H in homocarnosine) were optimized on the B3LYP/6-31G(d,p) level of theory. Then chemical shifts of the standards' structures along with representative structures **CT-I**, **CT-II**, **HT-I** and **HT-II** were calculated at the B3LYP/6-311+G(2d,p) level of theory. Next, the calculated chemical shift of each hydrogen and carbon atom in each representative structure was subtracted from the corresponding calculated chemical shift of standards. Finally, the so-obtained relative chemical shifts were weighted based on our Maxwell-Boltzmann statistics analysis and compared with experimental data [36, 76, 78, 85, 264].

The Gaussian09 software package [77] was used for *ab initio* calculations and Gromacs software package [96, 243] was used for classical MD simulations as well as for the analysis of the trajectories.

4.3 Results and discussion

4.3.1 *Ab initio* calculations

The calculated free energies corrected for the zero point energy obtained from the frequency calculations of each compound in implicit solvent show that **CT-II** and **HT-II** tautomers are more stable than **CT-I** and **HT-I** (difference in the free energy is 7.5

kcal/mol for both compounds). Namely, in **CT-II** and **HT-II** the proton is located on the N_π, which enables formation of an intramolecular hydrogen bond with the carboxylic group. On the other hand, in **CT-I** and **HT-I** the proton is bound to N_τ and the formation of an intramolecular hydrogen bond with the carboxylic group is sterically not possible.

4.3.2 Classical molecular dynamics simulations

1 μs-long classical MD simulations of **CT-I** and **CT-II** show that both tautomers mainly exist in two distinct conformations, defined in terms of the squared radius of gyration (R_g^2) [264]: the tautomer is in 'extended' conformation if $R_g^2 > 13.0$ Å^2 and in 'semi-folded' if 10.5 Å^2 < R_g^2 < 13.0 Å^2 (see Figure 4.1). While in the **CT-I** 'extended' conformation has predominant population, in the **CT-II** both conformations are equally populated.

Similarly, 'extended' and 'semi-folded' conformations can be defined for homocarnosine with values of $R_g^2 > 15.0$ Å^2 for the 'extended' conformation and 12.0 Å^2 < R_g^2 < 15.0 Å^2 for the 'semi-folded' one (Figure 4.1). For the **HT-I** the population of the two conformations is still shifted towards the 'extended' conformation, though not as strong as it is the case in carnosine. On the other hand in the **HT-II** both conformational states are more equally represented. This could be at least partially explained by the longer, and therefore also more flexible, alkyl chain of the GABA component of homocarnosine compared to the β-alanine in carnosine.

Next we clustered trajectories of each tautomer as described in the Methods part of this chapter and considered the representative structures of the four highest populated clusters for further calculations, namely **CT-Ia-d**, **CT-IIa-d**, **HT-Ia-d** and **HT-IIa-d**. They cover ~70% of the correspondent total conformational space. In **CT-I**, **CT-Ia** and **CT-Ib** (covering 81% of the conformational space), are 'extended', while **CT-Ic** and **CT-Id** have a 'semi-folded' conformation, showing a predominance of 'extended' conformations for this tautomer (Fig. 4.1). In **CT-II**, **CT-IIa** and **CT-IIc** (covering in total 54% of the correspondent conformational space) assume the 'extended' conformation, while the other two, **CT-IIb** and **CT-IId**, feature a 'semi-folded' [264] conformation. The populations of the two conformations in **CT-II** are comparable, similarly to what is reported in reference [264] (Fig. 4.1).

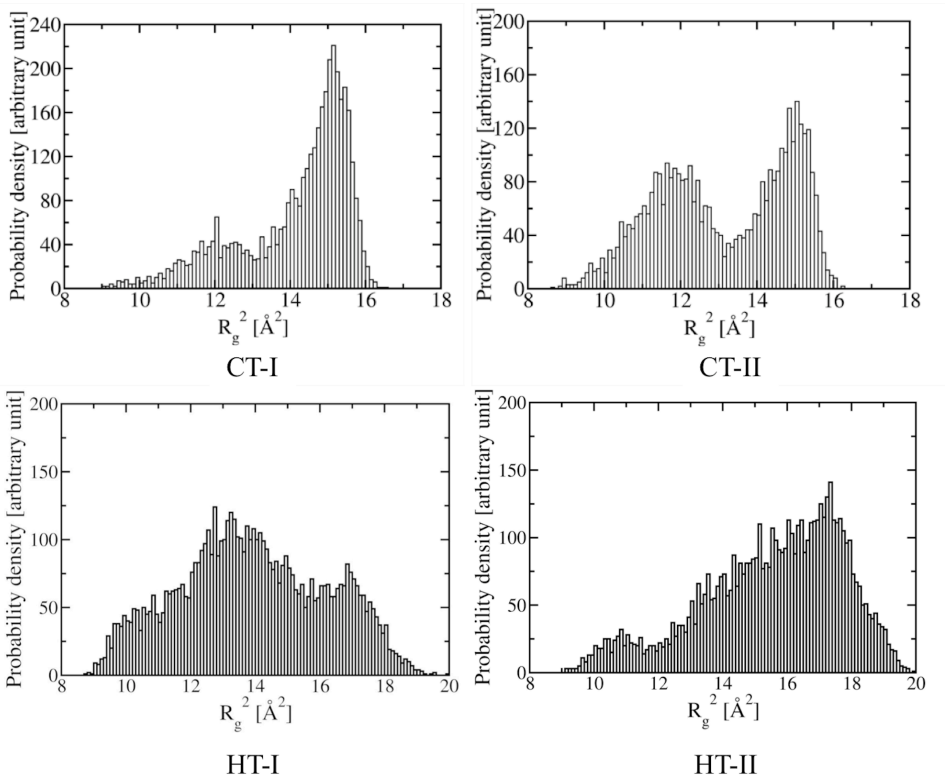


Figure 4.1: Probability distribution of R_g^2 in CT-I and CT-II (upper line) and HT-I and HT-II (lower line) in aqueous solution.

The 'extended' and 'semi-folded' representative conformations furthermore differ in the value of χ_1 dihedral angle of histidine side chain (defined as dihedral angle between N-C_α-C_β-C4). Namely, in all 'extended' conformation, the histidine side chain is in *trans* conformation (χ_1 is between -170° and -178°), while in 'semi-folded' representatives it can be either in *gauche*$^+$ (in **CT-Ic** χ_1 has value of 66°) or in *gauche*$^-$ (in **CT-IIb**, **CT-IId** and **CT-Id** χ_1 is between -52° an -64°) (Fig. 4.2). Each 'extended' and 'semi-folded' conformations can be further distinguished by χ_2 dihedral angle (defined as dihedral angle between C_α-C_β-C4-N_π), which defines the orientation of the imidazole ring of histidine in sense of its rotation in the axis of C_β-C4 bond. The 'extended', **CT-Ia** and **CT-IIb**, and the 'semi-folded', **CT-Id** and **CT-IIb**, have χ_2 values between 89° and 143°, while 'extended', **CT-Ib** and **CT-IIa**, and the 'semi-folded', **CT-Ic** and **CT-IId**, have χ_2 values between -60° and -125° (see Fig. 4.2).

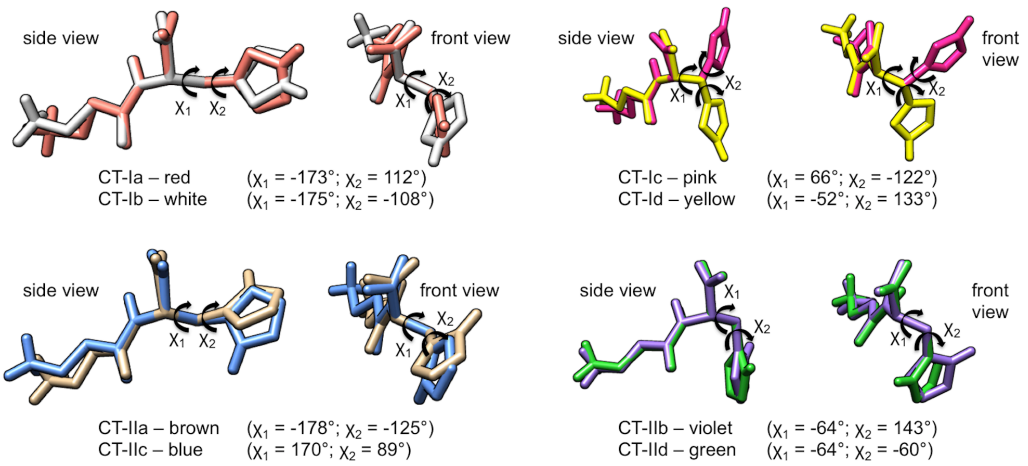


Figure 4.2: Representative structures of the largest populated clusters of carnosine tautomers **CT-I** and **CT-II** (see Chart 4.1). Each conformation is colored in its distinct color and it is shown from to different angles, as side view and as front view, with marked χ_1 and χ_2 dihedral angles. Only polar hydrogen atoms are shown due to clarity reasons.

Additionally, the RMSD, distance between terminal NH_3^+ and COO^-, accessible surface area, RMSF and radial distribution functions of **CT-II** compare well with previous classical MD study of this tautomer (see Fig. 4.1 and Fig. A2.1, A2.2 in Appendix 2) [264]. The RMSF analysis also shows that the **CT-II** is more flexible than the **CT-I** (Fig. A2.3 in Appendix 2). This is in agreement with the populations of 'extended' and 'semi-folded' conformations of each tautomer. Namely, in the **CT-I** the 'extended' conformation is the predominant one reflecting in the lower flexibility of imidazole ring compared to the **CT-II**, where the 'extended' and the 'semi-folded' conformations are more equally populated and therefore imidazole ring atoms have higher RMSF values (Fig. A2.3 in Appendix 2). The results obtained for both tautomers of carnosine were confirmed with further independent simulations of **CT-I** and **CT-II** (Tab. 4.1).

Further single point energy calculations of the representative structures of each cluster show that carnosine's and homocarnosine's most populared tautomers are **CT-I** and **HT-I**, respectively, according to the Maxwell-Boltzmann distribution analysis (see Tab. 4.2), which is in contrast with the results from our *ab initio* calculations. The difference between results from classical MD and DFT calculations can be explained, at least in part, by the different treatment of solvent environment. In classical MD simulations, ligand's hydrogen bond donor and acceptor atoms form hydrogen bonds with explicit solvent water molecules with around 80% probability for both protomers **CT-I** and **CT-II** (see also Tab. A2.4 in Appendix 2 for population of hydrogen bonds

between carnosine and water from MD simulations and Fig. A2.2 in Appendix 2 for radial distribution function of hydrogen bond acceptors and donors in carnosine with respect to water). On the other hand, in the DFT calculations, solvent is described implicitly and therefore explicit solvent-ligand hydrogen bonds are absent.

Table 4.2: Maxwell-Boltzmann statistics and relative free energies (in kcal/mol) for representative structures of highly populated clusters from MD simulations of carnosine's and homocarnosine's tautomers.[a]

Carnosine representative structure	$<N_X>/<N_{CT\text{-}Ia}>$[b]	Relative free energy with respect to **CT-Ia** in kcal/mol	Homocarnosine representative structure	$<N_X>/<N_{HT\text{-}Ib}>$[c]	Relative free energy with respect to **HT-Ib** in kcal/mol
CT-Ia	1	0	**HT-Ib**	1	0
CT-Ic	0.045	1.8	**HT-IIc**	0.001	4.3
CT-IIc	0.014	2.5	**HT-Ic**	$1.393*10^{-6}$	8.0
CT-IIa	0.012	2.6	**HT-Ia**	$3.895*10^{-8}$	10.1
CT-Id	$6.007*10^{-6}$	7.1	**HT-IIa**	$8.080*10^{-11}$	13.8
CT-IIb	$3.716*10^{-7}$	8.8	**HT-IIb**	$3.499*10^{-11}$	14.3
CT-Ib	$3.203*10^{-8}$	10.2	**HT-Id**	$1.417*10^{-13}$	17.5
CT-IId	$1.060*10^{-8}$	10.9	**HT-IId**	$9.503*10^{-16}$	20.5

[a] Please note that the clusters a-d are not the same for two ligands and their protonation states.

[b] All the structures are weighted against the structure with the lowest free energy – **CT-Ia**. Index X stands for the name of the model (**CT-Ia-d** and **CT-IIa-d**).

[c] All the structures are weighted against the structure with the lowest free energy – **HT-Ib**. Index X stands for the name of the model (**HT-Ia-d** and **HT-IIa-d**).

Finally we calculated ^{1}H and ^{13}C chemical shifts of all four representative structures of each tautomer and weighted them using Maxwell-Boltzmann statistics in order to verify the quality of our model by comparing them with available solution NMR data. The theoretical chemical shifts of carnosine and homocarnosine are generally in fair agreement with experimental values (Tab. 4.3-4.5) [36, 76, 78, 85, 264]. They differ from experiments by 27.0 ppm for ^{13}C (the biggest difference between experimental

and calculated ^{13}C chemical shifts was observed at C4, where experimental values also differ from each other for 16.9 ppm) and 1.8 ppm or less for ^{1}H for both carnosine and homocarnosine.[k]

Table 4.3: Calculated ^{13}C chemical shifts of both zwitterionic conformers of carnosine in comparison with experimental values.[a] For labeling of carbon atoms see Chart 4.1.

Carbon atom	Friedrich [76] [ppm] (T=294K)	Branham [36] [ppm] (T not reported)	Gaggelli [78] [ppm] (T=295K)	Calculated chemical shift[b] [ppm]	Assignments
CO_2^-	179.1	177.9	177.9	180.2 (182.0)	Carboxylate
C(O)NH	172.9	171.7	171.6	162.8 (164.7)	Amide carbonyl
C2	137.0	135.9	135.7	135.5 (137.4)	Imine carbon
C4	134.5	117.6	133.2	144.6 (146.5)	Ring carbon atom
C5	118.7	133.2	117.4	121.6 (123.5)	Ring methine carbon
C_α	56.5	55.3	55.2	67.9 (69.8)	α-His-methine
$C_{\beta'}$	37.2	32.5	32.3	47.9 (49.6)	β-Ala-methylene
$C_{\alpha'}$	33.7	36.1	35.8	47.0 (48.8)	α-Ala-methylene
C_β	30.2	29.0	29.0	39.7 (41.7)	β-His-methylene

[a] Friedrich[76] and Branham[36] used TMS as a reference compound, while Gaggelli[78] used sodium [2,2,3,3-2H4]-3-trimethylsilylpropanoate.

[b] Calculated chemical shifts are reported with respect to the TMS and sodium [2,2,3,3-2H4]-3-trimethylsilylpropanoate (in parentheses).

[k] Not-weighted ^{1}H and ^{13}C chemical shifts for carnosine and ^{1}H for homocarnosine and the statistical data of chemical shifts calculations are reported in Tab. A2.5-A2.8 in Appendix 2.

Table 4.4: Calculated and experimentally observed ^{1}H chemical shifts of carnosine.[a] For labeling of hydrogen atoms see Chart 4.1.

Hydrogen atom	Branham [36] [ppm] (T not reported)	Zhang [264] [ppm] (T=298K)	Friedrich [76] [ppm] (T=294K)	Gaggelli [78] [ppm] (T=295K)	Calculated chemical shifts[b] [ppm]	Assignments
H_2	7.66	7.60	7.38	6.93	7.72 (7.76)	Imine hydrogen
H_5	6.89	6.83	6.63	6.17	7.10 (7.14)	Ring methine
$H_{\beta R}$, $H_{\beta S}$	3.08; 2.92	3.01; 2.85[c]	2.65; 2.82	2.37; 2.21[c]	2.74; 3.38 (2.79; 3.42)	β-His-methylene
H_α	4.40	4.34	4.15	3.71	3.64 (3.68)	α-His-methine
NH	7.89	-	-	-	7.42 (7.46)	Amide N-H
$H_{\alpha'}(CH_2)$	3.16	2.53	2.35	2.44	2.64 (2.69)	α-Ala-methylene
$H_{\beta'}(CH_2)$	2.60	3.10	2.90	1.88	3.68 (3.72)	β-Ala-methylene

[a] Friedrich [76] and Branham [36] used TMS as a reference compound, while Gaggelli [78] used sodium [2,2,3,3-2H4]-3-trimethylsilylpropanoate.

[b] Calculated chemical shifts are reported with respect to the TMS and sodium [2,2,3,3-2H4]-3-trimethylsilylpropanoate (in parentheses).

[c] In the reference it is not mentioned which shift belongs to which hydrogen atom.

Table 4.5: Calculated and experimentally observed ^{1}H chemical shifts of homocarnosine. For labeling of hydrogen atoms see Chart 4.1.

Hydrogen atom	Govindaraju [85] [ppm] (T=310K)	Calculated chemical shifts [ppm]	Assignments
H_2	7.08	7.27	Imine hydrogen
H_5	8.08	6.88	Ring methine
$H_{\beta R}$, $H_{\beta S}$	3.00; 3.19[a]	2.82; 2.72	β-His-methylene
H_α	4.47	4.58	α-His-methine
$H_{\alpha'}$ (CH_2)	2.96	2.20	α-amino-butyryl-methylene
$H_{\beta'}$ (CH_2)	1.89	1.85	β-amino-butyryl-methylene
$H_{\gamma'}$ (CH_2)	2.37	3.17	γ-amino-butyryl-methylene
NH_3	6.40	4.65	amino-terminal group

[a] In the reference it is not mentioned which shift belongs to which hydrogen atom.

4.4 Conclusions

First I have performed μs-long classical MD of all four tautomers in aqueous solution and at physiological pH and identified the predominant tautomers of the two dipeptides in aqueous solution at physiological pH by μs-long MD simulations (**CT-I** and **HT-I** in Chart 4.1). Further I have characterized 'extended' and 'semi-folded' conformations of both dipeptides in terms of R_g^2 to show that energetically favorable conformation of both predominant tautomers is the 'extended' one. Calculations of additional properties, such as RMSD, solvent accessible surface, distance between terminal NH_3^+ and COO^-, RMSF and radial distribution functions compare well with previos computational studies of carnosine confirming the quality of our parameters for β-alanine and GABA. Finally, I evaluated our results by calculating ^{1}H and ^{13}C chemical shifts and comparing them with readily available solvent NMR data [36, 76, 78, 85, 264]. The fairly good agreement of those calculations proved the predictive power of my simulations.

5 Michaelis complexes of carnosine and homocarnosine with hCN1

5.1 Introduction

hCN1 is mainly present in human serum, where it degrades carnosine and therefore limits its possible therapeutic application. Moreover, it is highly specific enzyme, which showed activity only for carnosine and few other dipeptides (*i.e.* homocarnosine, anserine and ophidine) [233]. Since the other enzyme with known carnosine-degrading ability, hCN2, is not specific for carnosine, hCN1 was studied here.

Carnosine is considered hCN1's substrate and homocarnosine its competitive inhibitor as it is bound tighter inside the active site (K_m is reported to be 1.27 mM for carnosine and 0.2 mM for homocarnosine), though it is hydrolyzed at much slower rate than carnosine (with k_{cat} of 10.6 s^{-1} for carnosine and 0.2 s^{-1} for homocarnosine), resulting in higher catalytic efficiency of carnosine (hCN1's k_{cat}/K_m values for carnosine and homocarnosine are 8.6 and 1.1 $mM^{-1}s^{-1}$, respectively) [233]. Based on the structural differences between carnosine and homocarnosine and their binding inside hCN1 active site I try to explain here why one of the two ligands is considered as a substrate and the other one as a competitive inhibitor. Moreover, I try to clarify the role of particular residues in the hCN1 active site along with the effects of the enzyme dimerization on the course of the reaction and activity of the enzyme itself.

5.1.1 Structural characteristics of hCN1

The crystal structure of the apo hCN1 is known since 2006 (PDB ID 3DLJ; no article has been published so far). The crystal structure shows a homodimeric enzyme, with each monomer unit (monomers **M-I** and **M-II**) featuring a dizinc active site (Fig. 5.1a). Each monomer unit consists out of catalytic and dimerization domain [233]. Each monomeric unit has one active site, located in the catalytic domain. The main hallmarks of each active site are the two Zn^{2+} ions (Zn1 and Zn2 in Fig. 5.1b and c), coordinated by an acidic and a histidine residues, namely Zn1 is coordinated by one of the two carboxylate oxygens of Glu^{174} and the imidazole nitrogen of His^{452} and Zn2 is coordinated by imidazole nitrogen of His^{106} and one of the two carboxylate oxygens of Asp^{202}. The side chain of Asp^{139} is bridging both Zn^{2+} ions, coordinating to each of them with one carboxylic oxygen. The residues coordinating both Zn^{2+} ions (*i.e.* His^{106}, Asp^{139}, Glu^{174}, Asp^{202} and His^{452}) are forming first coordination shell. Another moiety bridging both Zn^{2+} ions was detected in the X-ray crystal structure but could not be identified.

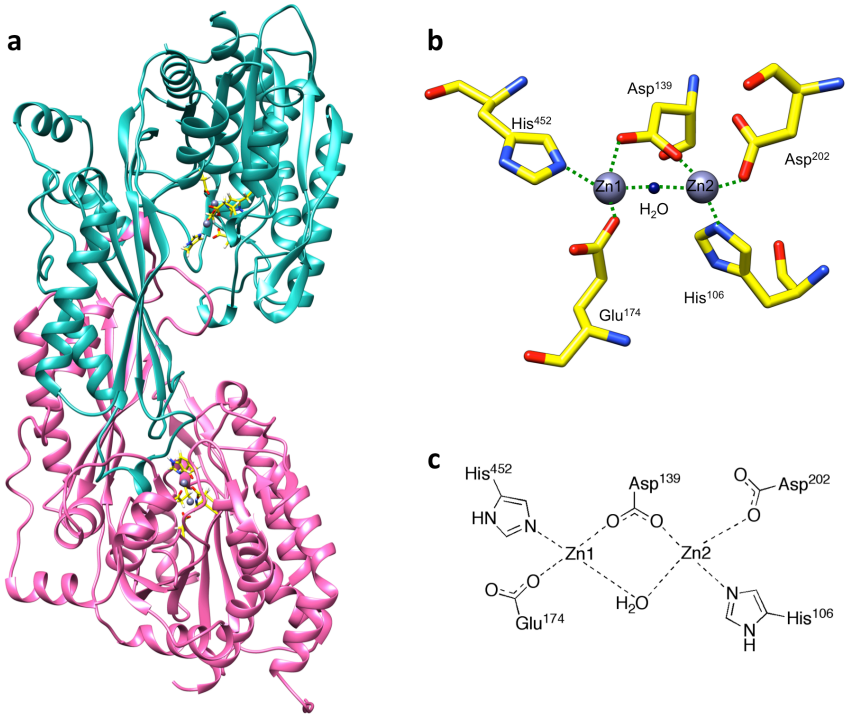


Figure 5.1: (a) hCN1 homodimer structure shown in ribbons representation. **M-I** and **M-II** are shown in pink and turquoise, while the active site residues are colored in yellow and Zn^{2+} ions in dark violet. (b) 3D representation of the hCN1 active site with only residues coordinating Zn^{2+} shown. Zn^{2+} ions are dark violet, oxygen atoms in red, nitrogen atoms in blue and bridging water molecule in dark blue. Coordination bonds are shown as dashed green lines. Hydrogen atoms are not shown due to clarity reasons. (c) 2D representation of the hCN1 active site. Coordination bonds are shown as dashed lines.

The other members of the M20 family for which the crystal structure is known are allantoate amidohydrolase from *E. Coli* (allC; PDB ID 1Z2L [5]), peptidase V from *Lactobacillus delbruecki* (PepV; PDB ID 1LFW [117]), carboxypeptidase G2 from *Pseudomonas sp*. (CPG2; PDB ID 1CG2 [213]) and peptidase T from *Salmonella typhimurium* (PepT; PDB ID 1FNO [88]). Comparison along them shows fully conserved structure of the active site (see Fig. 5.2a). The only difference is that Asp^{202} in hCN1 is substituted with histidine residue in allC. In all enzymes in the apo state the bridging species was determined to be water molecule [5, 213], except for the PepT, where no zinc bound water could be observed, possibly due to the low resolution of the data [88]. Additionally, other two metallopeptidases which feature identical active site to the M20 family, namely aminopeptidase from *Aeromonas proteolytica* (AAP; PDB ID 1AMP [47]) and aminopeptidase from *Streptomyces griseus* (SGAP; PDB ID 1CP7 [81]) also feature Zn^{2+}-bridging water molecule in their crystal structure (see Fig. 5.2b). Therefore it is reasonable to assume that the unknown species in the apoenzyme X-ray crystal structure of hCN1 is also a water molecule. Besides first shell residues coordinating both Zn^{2+} ions described above, also Glu^{173} from the second coordination shell, is preserved along all M20 family members as well as in AAP and SGAP (see Fig. 5.2).

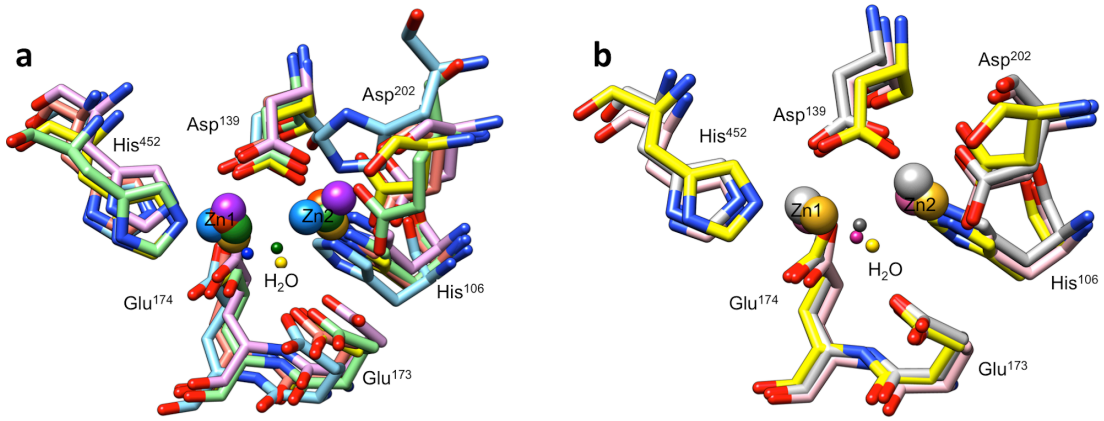


Figure 5.2: Superimposition of hCN1 active site (colored in yellow) with (a) available crystal structures of metallopeptidases from M20 family (allC colored in blue, PepV colored in violet, CPG2 colored in green and PepT colored in orange) and (b) AAP (colored in pink) and SGAP (colored in gray). Oxygen atoms are shown in red and nitrogen atoms in blue, while hydrogen atoms are not shown due to clarity reasons. Zinc ions are represented as big spheres and bridging water molecules as small sphere, both in the same color scheme as the corresponding enzyme. PepV and PepT are missing bridging water molecule because the crystal structure of the former has ligand present in the active site and in the later it was not detected due to the lower resolution of the crystallographic data. Oxygen atoms are shown in red and nitrogen atoms in blue, whereas hydrogen atoms are not shown due to clarity reasons.

hCN1 crystalizes and exists as a homodimer [110, 142, 233] as do many of the other members of the M20 family (*i.e.* allC [5], CPG2 [213], PepT [88] and human aminoacylase-1 (hAcy1) [149]). The homodimeric form of hCN1 is also the active one [110, 142, 233]. However, Peters *et al.* recently reported that hCN1 in human blood is actually present in both forms, though the activity of the monomeric form is considerably lower [197], suggesting that a possible rearrangement of the structure upon dimerization may affect the enzymatic function [149]. Equilibrium between both forms is age-dependent, where the monomeric form is predominant one in newborn children, while in the fully grown adults the dimeric form is the predominant one [197]. Similarly, it was for allC and hAcy1 that they are more active in their dimeric form [5, 149]. A mutational study on hAcy1 considered the mutation of the histidine residue, H206N, located on the second monomer, which is entangled in the active site of the first monomer (His^{206} in hAcy1 corresponds to His^{235} in hCN1). The mutation rendered the mutated enzyme significantly less active in comparison with the wildtype [149], leading to the suggestion that dimerization of the enzyme plays an important role in the enzymatic activity of the protein. No such mutational study has been performed for hCN1 yet, however based on the similarity of the enzymes from the same M20 family and previous findings that the dimer is hCN1's more active form, it may be plausible that the His^{235} is an important residue in the dipeptide hydrolysis.

5.1.2 Dipeptide hydrolysis mechanism in metallopeptidases

Reaction mechanism of hCN1 has not been determined yet. However, mutagenesis study on hCN1 has revealed some of the residues that could be possibly crucial for the reaction. Point muntations of hCN1 (H106A, D139A, E174A and E174A + D202A double mutation), all involving residues coordinating Zn^{2+} ions, led to the inactive enzyme [233]. The role of other residues in the hCN1 active site can only be assumed based on the studies of other metallopeptidases with the identical structure of the active site. Studies on CPG2 and hAcy1 show the importance of the preserved second shell glutamate residue, Glu^{424} in CPG2 and Glu^{147} in hAcy1 (corresponding to the Glu^{173} in hCN1), in the dipeptide hydrolysis across M20 family [128, 149]. Point mutations in both enzymes involving this glutamate residue resulted in highly reduced enzymatic activity or in completely inactive enzyme. In CPG2, the point mutation E424Q resulted in the 7-fold lower activity of the enzyme, while the E424A point mutation rendered the enzyme completely inactive. The enzyme retained some its enzymatic activity when a glutamate residue was substituted with a glutamine, because a glutamine can be involved in the same type of interactions as the protonated glutamate, while this is not the case for the alanine [128]. The same behavior was observed in the point mutations of Glu^{147} (E147A, E147Q and E147D) in hAcy1. When Glu^{147} was replaced with either

alanine or glutamine residue, the enzyme showed 1000-fold lower activity compared to the wild type. On the other hand the E147D mutation resulted in no enzymatic activity at all. This can be explained be the fact that in the E147D mutant the shorter side chain of the aspartate residue compared to the glutamate one results in location of the carboxylic group too far away to accept the proton from the bridging water molecule. Moreover, it also positions it closer to another aspartate residue (Asp^{348}), reducing the stability of the protein [149].

A DFT computational study by Navratil *et al*. tried to explain the mechanism of hydrolysis in M20 metallopeptidases on the basis of CPG2. Despite the fact that CPG2 is more active in its dimeric form [128], they performed their calculations on the model of active site based on the enzyme's PDB structure featuring only one monomeric unit. Even though the truncated model of the active site was used, the activation free energy was in good agreement with experimental values. However the role of the second monomer on the enzymatic activity cannot be evaluated from their calculations.

The only computational study orientated towards determining the structure of Michaelis complex between carnosine and hCN1 was performed prior to the determination of the hCN1's crystal structure. The hCN1 structure was therefore modeled based on the β-alanine synthetase. As a result some of the proposed residues coordinating both Zn^{2+} ions and the ligand are in fact far from the active site in the X-ray-determined structure. Moreover, only monomeric form of hCN1 was considered and Zn^{2+}-bridging moiety was excluded from the docking calculations [250].

At this point it is worth mentioning also the study by Holz *et al*., who proposed a general mechanism for dipeptide hydrolysis mechanism by metallopeptidase with co-catalytic active site, based on the X-ray crystallographic, kinetic, spectroscopic and thermodynamic data of AAP [101]. Despite the fact that AAP belongs to the M28 family of metallopeptidase enzymes, the active site is fully conserved in comparison with hCN1 and other M20 family enzymes. The main difference between AAP and hCN1 is that AAP is active as monomer. In AAP lid domain mimics the structure of a dimer by inserting the histidine residue in the active site [149]. Therefore despite this structural difference it is reasonable to assume that reaction follows the same pathway in both enzymes and also in both M20 and M28 families.

The main finding was that the second shell glutamate residue (Glu^{151} in AAP corresponding to Glu^{173} in hCN1) forms hydrogen bond with the bridging water molecule in the apo enzyme. Upon the presence of the substrate in the active site, it picks up the proton from the bridging water molecule and the so-formed hydroxide anion acts as nucleophilic agent attacking the peptide bond in the first step of reaction.

As mentioned before, the proton-accepting glutamate residue is conserved in the active site of all M20 family metallopeptidases with known structure. The protonated glutamate residue takes part also in the second step of the reaction, where it donates a proton to the peptide nitrogen, which causes the peptide bond to break [101]. For schematic course of the reaction proposed by Holz *et al*. [101] see Appendix 1.

5.2 Methods

5.2.1 hCN1 model structure

The apo hCN1 X-ray crystal structure (solved at 2.26 Å, PDB ID 3DLJ, paper not published yet) was used as a starting structure. It features a homodimeric enzyme, where each monomer contains an unknown residue, coordinating both Zn^{2+} ions. Based on the arguments discussed in the Introduction part of this chapter, we suggest that this residue is a water molecule in the apo state and a hydroxide anion in the holo state of hCN1 (Fig. 5.2 and Appendix 1).

M-I lacks residues 1-3, 77, 78, 208, 209, 437, 438 and 481 and **M-II** lacks residues 1-7, 77-79, 208, 209 413 and 481 (numbering as in the X-ray structure). They were added using the Modeller9.9 [217] code. The initial 26 residues (-25 to 0) belonging to the signal peptide[12] were not considered in our model, since this part is absent in the crystal structure. Moreover, the signal peptide normally does not play an active role in the activity of the enzyme. The histidine protonation states were determined using version 3.1 of H++ webserver [8, 84, 174] and are reported in the Table A2.9 in Appendix 2 along with other non-standard protonation states. The best initial models according to the Modeller's DOPE score [229] were optimized using loop refinement procedure [72] and re-ranked using DOPE score and evaluated using PROCHECK software [137]. The best structure of the dimeric hCN1 according to both scores was used. Our initial model of monomer hCN1 was obtained by simply removing **M-II** from the dimer.

The adducts of carnosine and homocarnosine with the monomer and the dimer were then constructed by molecular docking and refined with classical MD and QM/MM simulations. For studies of the dimeric complexes only active site on **M-I** was considered and the active site on **M-II** remained in the apo state.

5.2.2 Molecular docking

Initial Michaelis complexes of hCN1 in both monomeric and dimeric state with the most favorable tautomers of carnosine and homocarnosine (**CT-I** and **HT-I**), according to our simulations described in the previous chapter (and consistent with experimental data [76]), were determined using a two-step molecular docking calculations. Additionally we also docked other tautomeric forms of both ligands to the monomeric hCN1. The bellow described docking procedure is for docking to the **M-I** of hCN1, either in monomeric or in dimeric form.

(i) Step 1. In the first Zn^{2+} ions bridging residue in the refined enzyme structure was considered to be a water molecule. It is reasonable to assume that the position of the bridging water molecule is shifted during the ligand binding to the apoenzyme. Therefore the GOLD code [246] was used in this step as it is able to treat water molecules flexibly [245]. The complexes obtained by docking were ranked using Goldscore [114, 115] and rescored using ChemPLP scoring function [132]. Indeed, the complexes obtained in this first step of docking show an average shift of 1.6 ± 0.2 Å of the bridging water molecule with respect to the apoenzyme. The new position of water molecule was used in the neo-apoenzyme structure in further docking calculations.

(ii) Protonation states of hCN1 active site. It was shown that in the dipeptide hydrolysis by metallopeptidases with a hCN1-like catalytic active site the nucleophile attacking the peptide bond is actually the hydroxide anion [101]. Therefore water molecule has to donate one of its protons to a nearby residue. Based on a structural similarity analysis of hCN1 across other dizinc peptidases, for which structural information is available, described in the Introduction, the residue accepting the proton is expected to be Glu^{173} (Fig. 5.2). To verify this hypothesis, the residue preferentially picking up the proton from the water molecule using a model system was determined. The model system constituted by all the first shell Zn^{2+}-coordinating residues together with the second shell Glu^{173} and both Zn^{2+} ions in the hCN1 active site in the apo state and with the position of the bridging water molecule as obtained in the first step of docking. Next, modifications of the model system were built by varying the position of the proton, subtracted from the water molecule, among all the possible nearby proton acceptors (Glu^{173}, Glu^{174} and Asp^{202}). Finally, the model systems and its modifications were geometry minimized in the implicit solvent [167] and their potential energy was calculated. IEF-PCM model with atomic radii from UFF force field scaled by 1.1 was used. Different dielectric constant (ε) values were used as suggested in reference [148] In these calculations all atoms except hydrogen ones were frozen in order to maintain correct geometry of the active site. The calculations were carried out at the B3LYP[21]

level of theory, using the 6-31+G(d,p) basis set. My calculations confirm that indeed Glu^{173} picks the proton (named Glh^{173} hereafter), either at Oε2 or Oε1 oxygen atom of carboxylic group (resulting in tautomers hCN1**α** and hCN1**β**; Fig. 5.3). The differences in the potential energy by modifying the position of proton from Glh^{173} to Glu^{174} and Asp^{202} turned out to be rather large (15.8 kcal/mol or higher). Therefore it is reasonable to hypothesize (although it is not proven) that the entropic contribution would not modify such a trend.

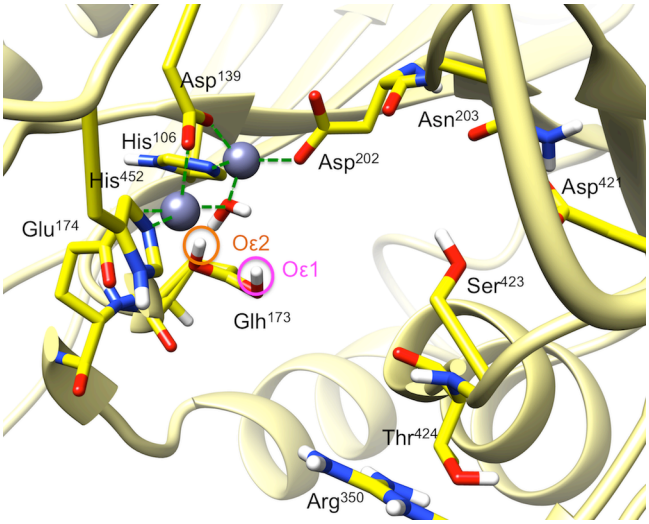


Figure 5.3: Tautomers **α** and **β** of hCN1 active site. Proton that changes position is encircled in orange for tautomer **α** (protonated on Oε2) and in pink for **β** (protonated on Oε1). Coordination bonds between Zn^{2+} ions and hCN1 active site residues are shown as dashed green lines. Oxygen atoms are shown in red, nitrogen atoms in blue and polar hydrogen atoms in white.

(iii) Step 2. Based on the results from previous step, we performed new set of docking calculations was performed with hydroxide anion as Zn^{2+} ions bridging moiety and starting from two different models of the monomer **M-I**, monomer hCN1**α** and monomer hCN1**β**. For this step AutoDock4.2 docking software [171], for which an improved force field for docking of small molecules to zinc metalloproteins has recently been developed [219], was used.

Additionally, following the approach for docking of small ligands to zinc metalloenzymes used by Park *et al*. [186], all atomic charges of enzyme and substrates were determined. Specifically, for the atomic charges of the hCN1 active site the same approach as it is used in Metal Center Parameter Builder (MCPB) [196] of AmberTools was used. Namely, the model system of the active site is created and its charges are determined using RESP methodology [19] on B3LYP/6-31G(d) level of theory. So calculated charges are then assigned to all atoms considered in the model,

except for the backbone heavy atoms (*i.e.* atoms forming peptide bonds), which are restrained to the values from AMBER94 force field [51]. My model systems consisted of the same residues as in the determination of the protonation states of hCN1 active site. The bridging residue was a hydroxide anion and Glu^{173} was considered in its non-ionized form. The protein residues were capped with acetyl and *N*-methylamine residues. The scheme used in MCPB adjusts the charge of capped residues to an integer value, thus allowing the formal charge of the cluster to disperse over the metal and the bound ligands [196]. Single point calculations were performed to preserve geometry of the active site within a computationally inexpensive procedure. For the rest of the enzyme atomic charges were taken from the AMBER94 force field [51]. Substrates' atomic charges were calculated using RESP method [19] on the structure of the substrates optimized with B3LYP functional [21] and 6-31G(d) basis set.

The docking calculations were performed using Lamarckian genetic algorithm [170] with 100 docking runs with the initial population of 150 individuals for each of the two substrates. The maximum number of generations and energy evaluations were set to 27,000 and 1×10^7. The enzyme and both, the Zn^{2+} ions and the hydroxide anion, were kept rigid during the calculations, while in the substrates all bonds were defined as rotatable with exception of the peptide one. The grid box had dimension of 51×51×51 points with spacing of 0.375 Å between them. The center of mass of the two Zn^{2+} ions was selected as the grid box center.

The representatives of the highest populated clusters, based on the RMSD value, of carnosine and homocarnosine Michaelis complexes with monomeric hCN1**α** and monomeric hCN1**β** from this second step of docking simulations were considered for the following calculations.

To determine whether monomeric hCN1**α** or monomeric hCN1**β** is more stable in the presence of substrate, I performed additional *ab initio* scanning with the same setup as described above. The difference was that here the active site model system, besides all the residues as in previously performed scanning, contained also the ligand.

It is possible that the protonation state of ligands' imidazole ring would change upon binding in the active site. Therefore, additional docking calculations to monomeric hCN1**α** were also performed with the other two protonation states of both ligands (where imidazole ring is protonated on N_π and with positively charged imidazole ring).

5.2.3 QM/MM simulations of Michaelis complexes

The complexes obtained with docking calculations of ligands in N_τ tautomeric forms to hCN1**α** were then subject to 40 ns long classical MD simulations, used as equilibration

simulations for following QM/MM simulations. Each system containing monomeric hCN1 was solvated using ~42,000 water molecules, with 135 Na^+ and 121 Cl^- ions ensuring neutrality and physiological conditions. Active site residues as well as ligand, Zn^{2+} and hydroxide ions were constrained in order to preserve geometries from docking. For the rest of the protein and ions we used AMBER99SB [104] force field with ILDN modification [150]. TIP3P potential was used for water. Simulations were performed with periodic boundary conditions at 298 K and 1 bar by coupling to Nosé-Hoover thermostat [102, 178] and Parrinello-Rahman barostat [189]. The temperature of 298 K was selected as the structural ensemble of the ligands was obtained by performing simulations at this temperature. Electrostatic interactions were taken into account using particle mesh Ewald method [56]. We applied the LINCS algorithm [95] to constrain all bond lengths in order to achieve time step of 2 fs. From the plot of RMSD value of the complexes we assumed that relaxation runs were long enough (see Fig. A2.4 in Appendix 2).

The last snapshots were then taken for further QM/MM simulations. Here, each system was partitioned into two parts: one treated at the quantum mechanical level (the QM part), and another part treated at the force field level (the MM part). The AMBER99SB [104] force field was used for the MM part. The QM region was identified by the proposed hCN1 active site (*i.e.* two Zn^{2+} ions, residues involved in the binding of these two ions – His^{106}, Asp^{139}, Glu^{174}, Asp^{202} and His^{452} and Glh^{173}, the substrate carnosine/homocarnosine, hydroxide anion needed for the nucleophilic attack on the peptide bond, and residues which can possibly form hydrogen bonds with the substrate (*i.e.* Asn^{203}, Arg^{350}, Ser^{423}, Thr^{424}, Asp^{421} and Glu^{451}). Open valencies were saturated by introducing a bond capping pseudopotential at the position of the first MM atom at the cutting border [252]. The time step for the dynamics was 0.12 fs, and the fictitious electronic mass was chosen to be 600 au. An energy cutoff of 80 Ry and the gradient corrected PBE exchange-correlation functional were used [193]. The core-valence electron interactions for the C, N, O, Zn and H atoms were described using non-conserving Martins and Trouiller pseudopotentials, corrected for a better description of van der Waals interactions [239, 252]. Integration of the nonlocal parts of the pseudopotential was obtained via the Kleinman-Bylander scheme [127] for all atoms except for Zn^{2+}, for which a Gauss-Hermite numerical integration scheme was used [59]. Electrostatic interactions between periodic images of the QM part were decoupled with the scheme of reference [159], while those of the classical part with periodic boundary conditions were treated by the particle-mesh method. A Nosé-Hoover thermostat [102, 178] was used to maintain a constant temperature of 298 K, as in reference [264] in order to be consistent with the temperature of the simulation in water. In total, 2 ps of QM/MM simulations were performed. Additionally, 1 ps

QM/MM simulation of **CT-I•**monomeric hCN1**α** was also performed at 303 K. This is the temperature at which K_m and k_{cat} were measured *in vitro* [233].

MD simulations and 2 ps QM/MM simulations of dimeric complexes were performed using the same setup as the monomeric ones. Both dimeric complexes were solvated using ~93,000 water molecules, with 295 Na^+ and 267 Cl^- ions ensuring neutrality and physiological conditions. The QM region in this case was the same as in monomeric complexes with additional His^{235} side chain from **M-II**, which can, according to our docking results, form hydrogen bond with the substrate.

The CPMD3.15.1 and GROMOS96 codes were used for the QM/MM calculations [231]. To evaluate the difference between monomeric and dimeric complexes in terms of protein structure the following properties were calculated: (i) the solvent accessible surface (SAS) was calculated using g_sas analysis tool of the Gromacs software [96, 243]. (ii) The hydrogen bonds were identified with the Hydrogen Bonds analysis tool of VMD visualization software, respectively.

5.3 Results

5.3.1 Carnosine• and homocarnosine•hCN1 complexes from docking calculations.

The structural determinants of apo hCN1 binding sites are very similar to those of the other dizinc peptidases for which structural information is available. These are allC, CPG2, PepT, PepV AAP and ASGP. The metal ions are coordinated by an acidic residue (Glu^{174} for "Zn1" and Asp^{202} for "Zn2", Fig. 5.1), along with a histidine (Fig. 5.1). An aspartate and a water molecule bridge the two metal ions in the apo state, while in the presence of the substrate, the bridging group is suggested to be a hydroxide anion and the Glu^{173} (which hydrogen bonds to the dizinc-bound hydroxide anion) to be protonated (if protonated named Glh^{173}; see also Fig. 5.3) [81]. Either of the two carboxylic oxygens can be protonated (Oε1 and Oε2 in Fig. 5.3, forming hCN1**β** and hCN1**α**, hereafter). *Ab initio* model calculations show that these two tautomers of hCN1 active site are far more stable than any other one (Tab. A2.10 and Tab. A2.11 in Appendix 2).

To obtain initial models of Michaelis complexes both tautomeric forms of each ligand as well as their positively charged counterpart were docked to the monomeric hCN1**α**. Aditionally, we performed dockings of **CT-I** and **HT-I** to the monomeric hCN1**β** and dimeric hCN1**α**.

*5.3.1.1 Docking of **CT-I** and **HT-I** to hCN1**α***

Docking calculations of **CT-I** to the monomeric hCN1**α** structure returned two different positions of **CT-I**. These differ in the orientation of *β*–alanine part. In one structure it is folded, while in the other one it is extended. However, both **CT-I**'s conformers form the same hydrogen bonds, namely imidazole ring of **CT-I** with backbone oxygen of Glu^{174}, peptide NH group with backbone oxygen of Thr^{424} and carboxylic group with peptide NH group of Thr^{424}. Additionally, **CT-I**'s carboxylic group forms a salt bridge with side chain of Arg^{350} and the peptide oxygen of **CT-I** forms a hydrogen bond with the bridging hydroxyl ion (see also Tab. 5.1 and Fig. 5.4). In the complex where **CT-I** is in extended conformation, the substrate's orientation seems more prone to the hydrolysis. Namely, when **CT-I** is in folded conformation, amino terminal group is orientated towards Glh^{173}, which plays an active role in the hydrolysis. This orientation of **CT-I**'s amino group could hamper the course of the reaction. On the other hand, **CT-I** in extended conformation does not feature any steric hindrances for the hydrolysis to take place and moreover, it could be additionally stabilized with hydrogen bonds between its amino group and nearby residues (Asp^{202}, Asn^{203} or Asp^{421}) [1] (see Fig. 5.4). Moreover, as reported above, the extended conformation is the most populated one in solution according to Maxwell-Boltzmann statistics. Therefore, we used structure of this complex for further calculations.

A similar interaction pattern to the above-discussed of **CT-I**, is also observed for **HT-I** docked into hCN1. The key difference between the two dipeptides is that **HT-I** forms an additional hydrogen bond between its amino group and side chain of Asn^{203}. Due to this additional interaction, **HT-I** is present inside the hCN1 active site only in the extended form (see Tab. 5.1 and Fig. 5.4).

[1] Please note that the latter interactions where not observed in the docking calculations, however some of them were present in the subsequent QM/MM simulations.

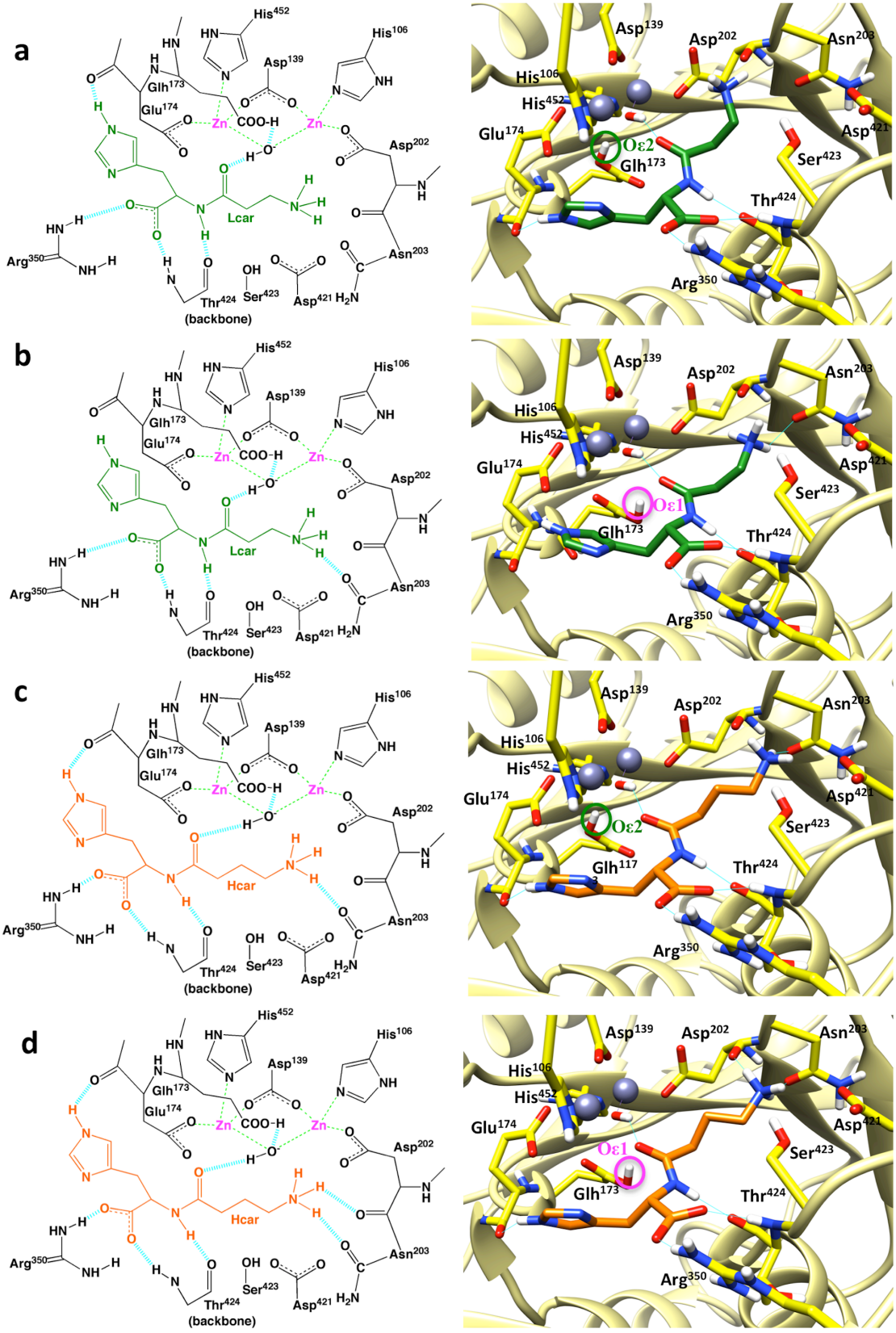

a
b
c
d
Lcar
Hcar
(backbone)
Zn
COO-H
His452
His106
Asp139
Asp202
Asn203
Asp421
Glh173
Glu174
Ser423
Thr424
Arg350
Oε2
Oε1

Figure 5.4: Highest populated clusters of monomeric hCN1 complexes (**CT-I•**hCN1**α** (a), **CT-I•**hCN1**β** (b), **HT-I•**hCN1**α** (c) and **HT-I•**hCN1**β** (d)) obtained by docking simulations (on Glh^{173}**α** proton is marked with green and on Glh^{173}**β** with pink circle). On the left side of the figure are schematic 2D representations of the complexes, while on the right side there are matching 3D figures. **CT-I** is shown in green and **HT-I** in orange. Hydrogen bonds are shown as dashed turquoise lines in 2D schemes and as full turquoise lines in the 3D representations. Coordination bonds between Zn^{2+} ions and hCN1 active site residues are shown on the 2D representations as dashed green lines. Zn^{2+} ions are shown as violet spheres in the 3D representation. Oxygen atoms are shown in red, nitrogen atoms in blue and polar hydrogen atoms in white.

*5.3.1.2 Docking of **CT-I** and **HT-I** to hCN1**β***

The docking results for the monomeric hCN1**β** show slightly different binding poses and interaction network, with respect to the monomeric hCN1**α**. Namely, **CT-I** is found only in extended conformation stabilized by a hydrogen bond between its amino group and Asn^{203} side chain, which is not present in hCN1**α**. On the other hand, the hydrogen bond between backbone oxygen of Glu^{174} and histidine ring of **CT-I** is not present anymore. For **HT-I**, we got the same set of interactions as in hCN1**α** with additional hydrogen bond between its amino group and backbone oxygen of Asp^{202} (see also Tab. 5.1 and Fig. 5.4).

5.3.1.3 Additional docking calculations of other ligand's tautomers

We performed also docking calculations with the other two possible protonation states of carnosine's and homocarnosine's imidazole ring, **CT-II** and **HT-II**, and with protonated imidazole ring (named **CT-III** and **HT-III**).

In the case of carnosine, the three protonation states result in similar binding poses. The difference is only in a hydrogen bond formed between the ligand's imidazole ring and hCN1: in the case of **CT-I** and **CT-III**, imidazole ring hydrogen bonds to Glu^{174} backbone oxygen, while in **CT-II** imidazole ring forms hydrogen bond with the side chain of Glh^{173} (see also Fig. 5.5 and Tab. 5.1). The RMSD of non-hydrogen atoms of **CT-II** and **CT-III** protonation states with respect to that of **CT-I** are 2.1 Å and 1.3 Å, respectively. The RMSD of the peptide bond in **CT-II** and **CT-III** relative to **CT-I** is 0.4 Å and 0.6 Å.

Also in the case of homocarnosine, the three ligand's protonation states result in similar binding poses. However, **HT-II** forms in total one hydrogen bond less with hCN1 with respect to the **HT-I** tautomer. Specifically, in **HT-II** imidazole ring forms

no hydrogen bonds with hCN1. On the other hand in **HT-III** the amino terminal group forms a hydrogen bond with Asp^{202}, instead of Asn^{203} as it is the case for **HT-I** (see also Fig. 5.5 and Tab. 5.1). Therefore, both **HT-II** and **HT-III** protonation states appeared less stabilized than **HT-I**. RMSD of non-hydrogen atoms of **HT-II** and **HT-III** protonation states with respect to **HT-I** is 1.1 Å and 1.4 Å, respectively, while the RMSD of the peptide bond in **HT-II** and **HT-III** relative to **HT-I** is 0.3 Å and 0.5 Å.

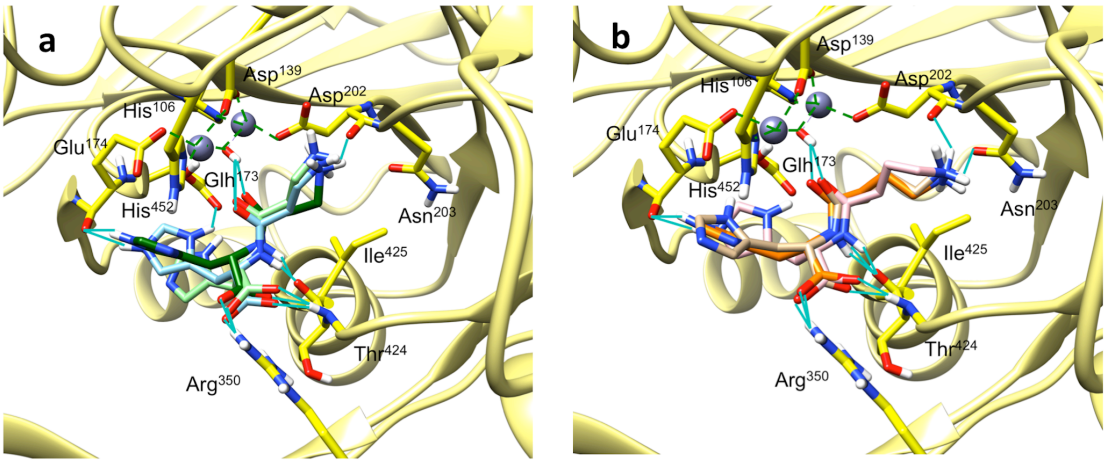


Figure 5.5: Highest populated clusters of monomeric hCN1**α** complexes obtained by docking simulations. a) binding poses of carnosine in all three protonation states (**CT-I** is shown in dark green, **CT-III** in light green and **CT-III** in light blue); b) binding poses of homocarnosine in all three protonation states (**HT-I** in shown in orange, **HT-II** in tan and **HT-III** in light pink). Hydrogen bonds are shown as full turquoise lines. The coordination bonds between Zn^{2+} ions and hCN1 active site residues are shown as dashed green lines. The Zn^{2+} ions are shown as violet spheres. Oxygen atoms are shown in red, nitrogen atoms in blue and polar hydrogen atoms in white.

It was shown that K_m for both anserine and ophidine has the same value as for carnosine. While anserine and ophidine differ from each other in the position of methyl group on the imidazole ring, on N_π in anserine and on N_τ in ophidine, carnosine can have a proton at either of the two imidazole nitrogen atoms. Based on the fact that K_m has similar value for all three compounds, one could assume that the location of either methyl group or proton in imidazole ring does not play a major role in substrate binding to hCN1. However no kinetic data is available to confirm whether this structural difference affects the reaction rate constant for the hydrolysis of the ligands. The former hypothesis was confirmed by our docking calculations of both tautomeric forms of each compound to the hCN1 and indeed no major difference in the ligands' orientation inside hCN1 were observed.

5.3.1.4 Docking to dimeric hCN1α

Each ligand, **CT-I** and **HT-I**, features two distinct docking binding poses, with population of 50% and 37% for carnosine (LBP1 and LBP2) and 68% and 16% for homocarnosine (HBP1, HBP2). The peptide bond in LBP2 and HBP2 has identical orientation to the one obtained from the monomer-based simulations (see Fig. 5.6 and Tab. 5.1). In the other binding pose (LBP1 and HBP1), which is by far that with the higher population, the peptide oxygen atom forms a hydrogen bond with His^{235} from the other monomer and therefore it does not coordinate the Zn^{2+} ion as in the monomer (see Fig. 5.6 and Tab. 5.1). Hence, LBP1 and HBP1 are most likely populated poses. Therefore, they were selected for further investigation using QM/MM simulations.

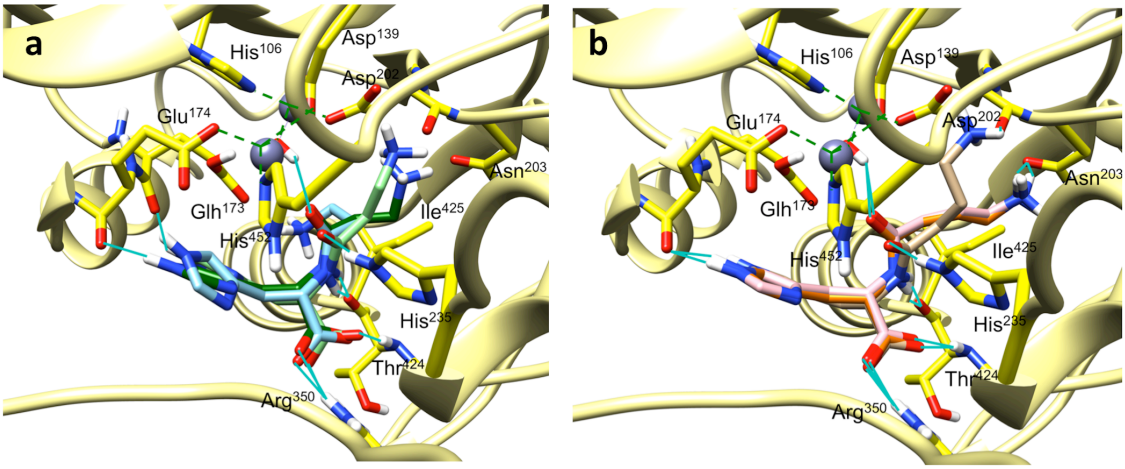


Figure 5.6: Dimeric hCN1 complexes in comparison with monomeric complexes obtained by docking simulations. a) binding poses of **CT-I** in hCN1 (**CT-I**•monomeric hCN1**α** is shown in dark green, LBP1 and LBP2 of **CT-I**•dimeric hCN1**α** are shown in light green and light blue, respectively); b) binding poses of **HT-I** in in hCN1 (**HT-I**•monomeric hCN1**α** is shown in orange, HBP1 and HBP2 of **HT-I**•dimeric hCN1**α** are shown in tan and light pink, respectively). Hydrogen bonds are shown as full turquoise lines. Coordination bonds between Zn^{2+} ions and hCN1 active site residues are shown as dashed green lines. Zn^{2+} ions are shown as violet spheres. Oxygen atoms are shown in red, nitrogen atoms in blue and polar hydrogen atoms in white.

Table 5.1: Hydrogen bond donors and acceptors inside the hCN1 active site from docking simulations. Asterisk indicates that the salt bridge is formed.

		Monomeric hCN1								Dimeric hCN1	
Hydrogen bond donor	Hydrogen bond acceptor	hCN1α						hCN1β		hCN1α[a]	
		CT-I	**CT-II**	**CT-III**	**HT-I**	**HT-II**	**HT-III**	**CT-I**	**HT-I**	**CT-I**	**HT-I**
Ligand histidine ring	Glu^{173} side chain		✔							✔	
Ligand histidine ring	Glu^{174} backbone oxygen	✔		✔	✔		✔		✔		✔
Ligand amino group	Asp^{202} backbone oxygen						✔		✔		✔
Ligand amino group	Asn^{203} side chain				✔	✔		✔	✔		
Arg^{350} side chain	Ligand carboxyl group	✔*	✔*	✔*	✔*	✔*	✔*	✔*	✔*	✔*	✔*
Ligand peptide NH group	Thr^{424} backbone oxygen	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
Thr^{424} peptide NH group	Ligand carboxyl group	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
Hydroxide anion	Ligand backbone oxygen	✔	✔	✔	✔	✔	✔	✔	✔		
His^{235} imidazole ring	Ligand peptide oxygen									✔	✔
Total number of hydrogen bonds		5	5	5	6	5	6	5	7	5	6

[a] Data is shown for LBP1 and HBP1.

5.3.2 The overall fold of hCN1

The crystal structure of hCN1 reveals no disulfide bridges between **M-I** and **M-II**, as all the cysteine residues of **M-I** are located far away from cysteine residues of **M-II**. The only interactions that keep both monomers together are hydrogen bonds and hydrophobic interactions. Moreover, a loop (residues 229 to 245) of the dimerization domain of each monomer is entangled in the catalytic domain of the other monomer. Especially, His^{235} residue, preserved across other M20 family members, seems to be the relevant residue for the catalysis. The role of this residue is further described in the QM/MM simulations section.

The solvent accessible surface **M-I** in the monomeric hCN1, calculated from the classical MD, is not too dissimilar to that of the **M-I** in the dimeric hCN1 (Tab. 5.2). However, the two structures differ in regions R**I** – R**V** (see Fig. 5.7a). In region R**I** (located between alpha helices 2 and 3), alpha helix 3 starts at residue 42 in the monomeric form and at residue 45 in the dimeric form and therefore also the whole loop region features different conformation (Fig. 5.7a). The turn in region R**II** (residues 77 – 79) is shifted towards the solvent of ~5 Å in the dimeric form compared to the monomeric one. The loop forming region R**III** (residues 228 – 245), between beta strand 11 and alpha helix 8, is entangled in the active site of the cognate momomer (and vice-versa) in the dimeric form. In the monomeric form this loop is solvent-exposed. Region R**IV** (residues 256 and 309) displays a difference in the position of the backbone atoms within the helical region. The most significant difference comes in the region R**V** (residues 328 – 340), which is entangled in the cognate monomeric unit in the case of dimeric hCN1. In the monomeric form, this region is stabilized by the salt bridge between Asp^{331} and Lys^{336} while this is sterically not possible in the case of the dimeric forms.

Table 5.2: The average values of solvent accessible surface of monomeric and dimeric (**M-I**) hCN1 from classical MD simulations.

Ligand	Monomeric hCN1	Dimeric hCN1
Carnosine	251 ± 3 nm^2	253.3 ± 3 nm^2
Homocarnosine	249 ± 3 nm^2	250 ± 2 nm^2

The distribution of intramolecular hydrogen bonds is different on passing from the monomeric to the dimeric forms. The average number of intramolecular hydrogen bonds is higher in the monomer than in the dimer (Fig. 5.7b). Parts of the hydrogen

bonds engaged in the formation of the dimer are indeed replaced by intramolecular hydrogen bond in the monomer. The number of intramolecular hCN1 hydrogen bonds is similar in the dimeric adducts of hCN1 with **CT-I** and **HT-I**. It is larger in the monomeric hCN1 adduct with **CT-I** relative to that of the **HT-I** adduct (see Fig. 5.7b).

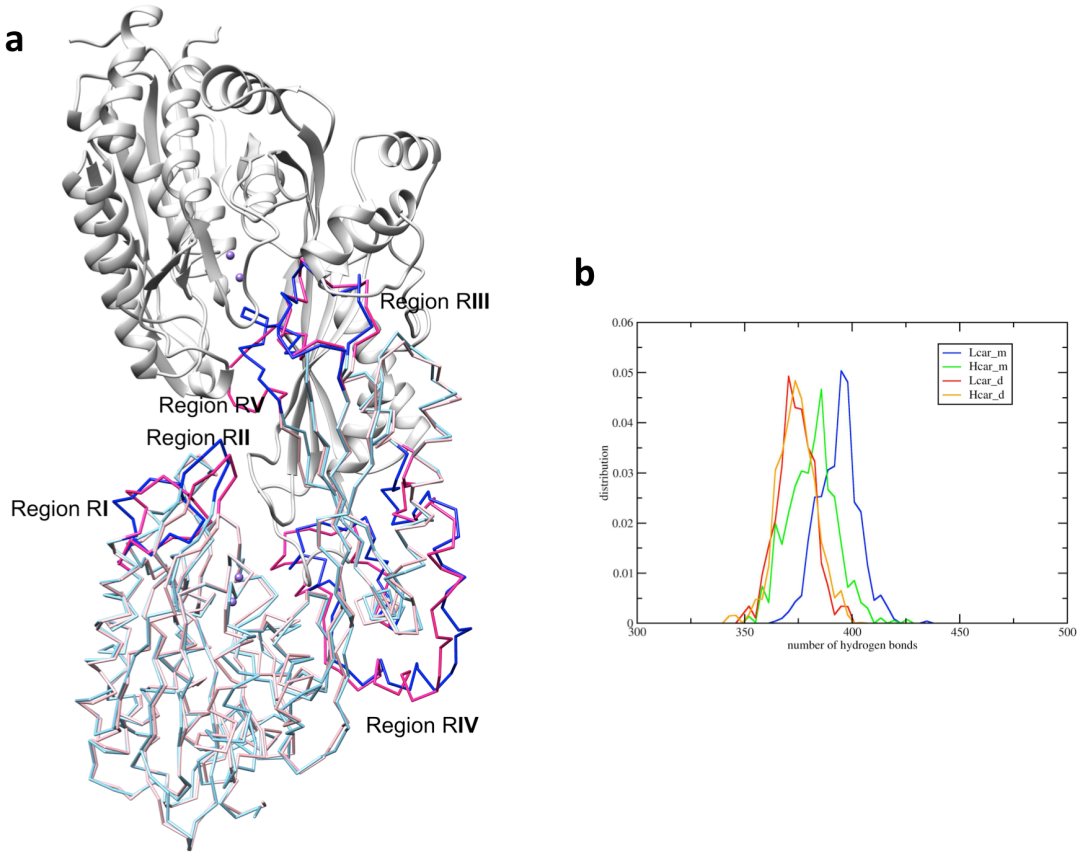


Figure 5.7: (a) Superimposition of monomeric hCN1 (light blue) with **M-I** of dimeric hCN1 (light pink) shown in chain trace representation. **M-II** is shown in light grey in ribbons representation and the two Zn(II) ions (Zn1 and Zn2) are shown as purple spheres. Regions R**I** – R**V** are shown in blue for monomeric hCN1 and in pink for **M-I**. (b) Distribution of intramolecular hydrogen bonds in the monomeric and **M-I** of dimeric complexes with ligands during 40 ns classical MD.

5.3.3 QM/MM simulations

Because of the presence of metal ions I used DFT-based QM/MM simulations to describe substrate/enzyme interactions as in references [54, 134]. Hence, the complexes underwent 2 ps long relaxation QM/MM simulations, preceded by 40 ns long classical MD-based equilibration (see Fig. A2.4 in Appendix 2).

Let us discuss first the monomer adducts. Both, **CT-I**•monomeric hCN1**α** and **HT-I**• monomeric hCN1**α** Michaelis complexes feature metal ions' coordination polyhedra similar to those of the apo enzyme. The metal-ligand atom distances are similar except those between Zn1 and the bridging aspartate (Asp^{139}; see Tab. A2.12 and A2.13 in Appendix 2), which is in both Michaelis complex longer than in the X-ray structure. The presence of weaker Zn1-Asp^{139} interaction may be consistent with observations of AAP reported in reference [47], where it is stated that, during the catalysis, the bridging aspartate residue can dissociate from either or both Zn^{2+} ions.[mn]

The substrates' histidine rings form a direct and a water-mediated hydrogen bond with Glu^{174} carbonyl group, for **CT-I** and **HT-I**, respectively (Fig. 5.8 and Tab. 5.3). The substrates' amino terminal hydrogen forms water-mediated and direct hydrogen bond with Asp^{202} backbone unit, for **CT-I** and **HT-I**, respectively. That of **CT-I** forms also a salt bridge with the side chain of this residue. The substrates' carboxylic group forms salt bridge with Arg^{350}. Both substrates' backbone amide groups form hydrogen bonds with Thr^{424} backbone unit. The latter interactions are more persistent in **HT-I** than **CT-I** (Tab. 5.3), though the calculated occurrences should be taken with care since they might strongly depend on the time scale of the simulations. The substrates' backbone carbonyl forms a water-mediated and direct hydrogen bond with Thr^{424} backbone, for **CT-I** and **HT-I**, respectively (Tab. 5.3). **HT-I**'s amino group interacts with Asp^{421} and Glu^{451} through water-mediated salt bridges for most of the simulation time, while the same interaction is completely absent in the case of **CT-I**. Same hydrophobic interactions are present for both ligands, namely between $(CH_2)_2$ chain in **CT-I** and $(CH_2)_3$ in **HT-I** and the side chain of Ile^{425}. Overall, **HT-I** appears to form more intermolecular interactions with the hCN1 than **CT-I** (Fig. 5.8 and Tab. 5.3). Therefore **HT-I** appears to be tightly bound inside the enzymatic active site, compared to **CT-I**.

In addition, while both ligands feature similar flexibility inside the active site, **CT-I**'s peptide bond appears to be more stable (see Fig. A2.5a and c in Appendix 2 for RMSF

---

[m] In addition, the distances between both zinc ions and the oxygen of the bridging moiety are slightly shorter in both Michaelis complexes (see Tab. A2.12 and A2.13 in Appendix 2). The changes in angle values of residues coordinating both zinc ions (reported in Tab. A2.14 and A2.15 in Appendix 2) are consequences of the changes in distances described above.

[n] The structure of the **CT-I**•monomeric hCN1**α** complex at 303 K was very similar to that obtained at 298 K (see Tab. A2.13 and A2.15 in Appendix 2).

of atoms of both ligands inside monomeric hCN1 active site), which could contribute to the **CT-I**'s higher propensity for hydrolysis.

The orientation of both substrates in **CT-I•**monomeric hCN1**β** and **HT-I•**monomeric hCN1**β** Michaelis complexes is less favorable for hydrolysis to occur (see Fig. A2.6, Tab. A2.16 and Section A2.1 in Appendix 2 for a discussion). Hence, our QM/MM calculations give evidence that the **CT-I•**monomeric hCN1**α** and **HT-I•**monomeric hCN1**α** are the most probable tautomers.

We further investigated this finding by *ab initio* calculations. In a model of the active site, we varied the position of the hydrogen atom among all the possible nearby proton acceptors (Glu^{173}, Glu^{174} and Asp^{202}). We calculated the relative stability of so-obtained tautomers at different dielectric constant values (same approach of reference [148], see Methods for details). Consistently with the QM/MM calculations, the monomeric hCN1**α** tautomers turn out to be more stable than the hCN1**β** ones by as much as ~13 and ~9 kcal/mol in **CT-I•**monomeric hCN1 and **HT-I•**monomeric hCN1, respectively (see Tab. A2.17 in Appendix 2). The other tautomers are even less stable than hCN1**β**.

We next turn our attention to the **CT-I•**hCN1 and **HT-I•**hCN1 complexes in their dimeric states. The MD and QM/MM structures, based on the most favorable docking binding pose show that the ligands' binding pose and most of the interactions described for the monomer between each ligand and hCN1 are indeed preserved (see Tab. 5.3 and Fig. 5.8). However, here **CT-I**'s and **HT-I**'s imidazole rings form, respectively, direct and water mediated hydrogen bond with the backbone carbonyl group of Glh^{173} (Fig. 5.8). In addition, the ligands' carbonyl group forms hydrogen bond with the side chain of Thr^{337} from the second monomer (Fig. 5.8). Finally, and most importantly, the ligands' carbonyl group forms one hydrogen bond with His^{235}, also from the second monomer (Fig. 5.8 and Tab. 5.3). This H-bond might play a role in the enzymatic hydrolysis through both the polarization of the substrate's C=O group and the conservation of a proper orientation for catalysis. The absence of this hydrogen bond in the monomeric complexes might indeed contribute to lower the enzymatic activity compared to the dimeric form. Investigations of the enzymatic reaction are required to further establish this aspect of hCN1 catalysis.

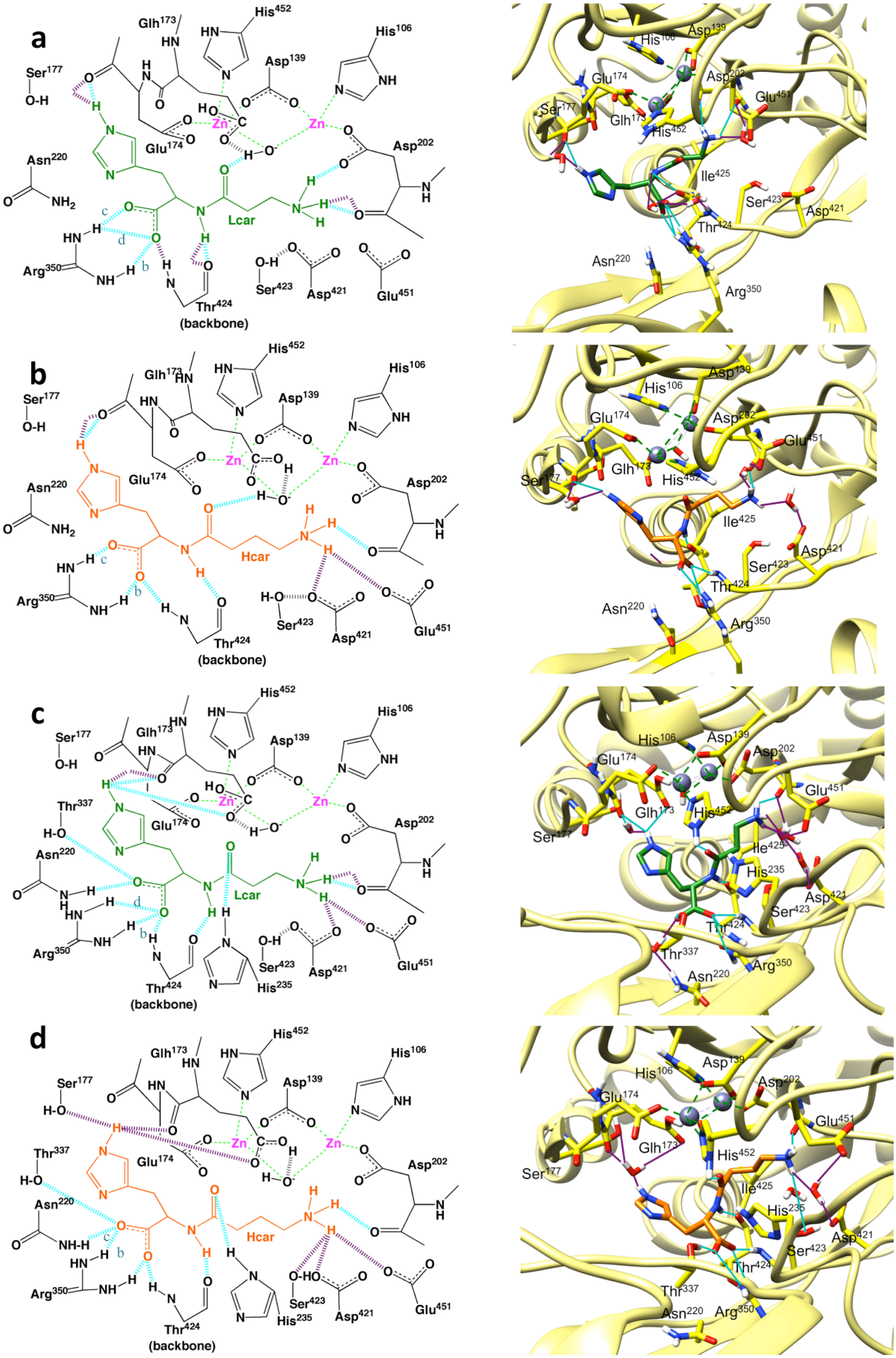

a
Glh173
His452
His106
Asp139
Ser177
O-H
Asp202
Glu174
Zn
Asn220
NH2
Lcar
Arg350
NH
Thr424
(backbone)
Ser423
Asp421
Glu451
Ile425
Thr424
b
Hcar
c
Thr337
H-O
d
His235

Figure 5.8: Structures of **CT-I•**monomeric hCN1**α** (a), **HT-I•**monomeric hCN1**α** (b), **CT-I•**dimeric hCN1**α** (c) and **HT-I•**dimeric hCN1**α** (d) as obtained by our QM/MM simulations. On the left side of each panel is schematic 2D representation of the Michaelis complex (bulk water molecules are not shown due to clarity reasons), while on the right side there is matching 3D figure (non-polar hydrogen atoms and bulk water molecules are not shown due to clarity). **CT-I** is shown in green and **HT-I** in orange. Hydrogen bonds between ligand and environment are shown as dashed turquoise lines in 2D schemes and as full turquoise lines in the 3D representations (for occupancy of hydrogen bonds between ligands and hCN1 see Tab. 5.3; hydrogen bond naming (a, b, c and d) between ligand's carboxylic group and side chain of Arg^{350} correspond to the one in Tab. 5.3). Water-mediated hydrogen bonds are shown as dashed purple lines in 2D schemes (due to the clarity, water molecules are omitted in these schemes) and as full purple lines in the 3D figures. Coordination bonds between Zn^{2+} ions and hCN1 active site residues are shown as dashed green lines on all figures. Zn^{2+} ions are shown as violet spheres in the 3D representation.

Overall, also in the dimeric hCN1 state, **HT-I** forms more extensive hydrogen bond network with respect to **CT-I** (see Tab. 5.3 and Fig. 5.8). This suggests that binding of **HT-I** to hCN1 is stronger than **CT-I**. This finding is consistent with the lower K_m value of **CT-I** compared to the **CT-I** [233]. However, as already mentioned, the hydrogen bond occurrence, especially that of the water-mediated hydrogen bonds, should be taken with care as they strongly depend on the time scale of the simulation.

Interestingly, the absence of the above discussed hydrogen bond between the ligand's carbonyl oxygen and the imidazole ring of His^{235} in the monomeric complexes might contribute to the experimentally observed lower enzymatic activity compared to the dimeric form [110, 142, 197].

Table 5.3: Probability of hydrogen bonds between ligands and the hCN1α active site from our 2 ps long QM/MM simulations. ^{WM} indicates water mediated hydrogen bond.

Hydrogen bond donor	Hydrogen bond acceptor	Carnosine• monomeric hCN1α	Homocarnosine• monomeric hCN1α	Carnosine• dimeric hCN1α	Homocarnosine• dimeric hCN1α
Ligand imidazole ring	Glh^{173} backbone oxygen	/	/	69% 4%^{WM}	71%^{WM}
Ligand imidazole ring	Glh^{173} carboxyl group	/	/	12%	1% 4%^{WM}
Ligand imidazole ring	Glu^{174} backbone oxygen	56% 18%^{WM}	63% 31%^{WM}	/	/
Ligand imidazole ring	Ser^{177} hydroxyl group	/	/	/	30%^{WM}
Ligand amino group	Asp^{202} backbone oxygen	4% 89%^{WM}	100%	26% 51%^{WM}	96%
Ligand amino group	Asp^{202} carboxyl group	100%	/	/	/
Ligand carboxyl group	Asn^{220} amino group	/	/	79%^{WM}	/
Arg^{350} side chain[a]	Ligands carboxyl group	a) / b) 49% c) 4% d) 40%	a) / b) 88% c) 97% d) /	a) 85% b) / c) 86% d) /	a) 100% b) / c) / d) 87%
Ligand amino group	Asp^{421} carboxyl group	/	83%^{WM}	63%^{WM}	76%^{WM}
Ligand amino group	Ser^{423} hydroxyl group	/	/	/	5%
Ligand peptide NH group	Thr^{424} backbone oxygen	3% 16%^{WM}	70%	100%	95%
Thr^{424} peptide NH group	Ligand carboxyl group	86%^{WM}	16%	7%	58%
Ligand amino group	Glu^{451} carboxyl group	/	69%^{WM}	87%^{WM}	8%^{WM}
Ligand backbone oxygen	His^{235} imidazole ring[b]	/	/	96%	75%
Ligand carboxylic group	Thr^{337} hydroxyl group[b]	/	/	96%	90%

[a] Between Arg^{350} and substrate's carboxylic group are 4 possible hydrogen bonds (marked with, a), b,) c) and d), see also Fig. 5.8); also salt bridge is formed here.

[b] Residues His^{235} and Thr^{337} are from **M-II**.

5.4 Discussion

Despite the biological relevance of hCN1 inhibition for a variety of human diseases [17, 58, 61, 98, 111, 123, 141, 206, 216], the current structural information for this enzyme is scarce. The enzyme exists as a monomer or a dimer *in vivo* [197], yet only the structure of the dimer is available (PDB ID 3DLJ). Structural information on the Michaelis complexes with the two dipeptides, carnosine and homocarnosine, selectively cleaved by hCN1, has so far relied on the elegant study by Vistoli *et al.* [250], which was however performed on the hCN1 monomer and before the X-ray structure of the enzyme had been reported.

Here I have predicted structures of Michaelis complexes formed between carnosine and homocarnosine with both monomeric and dimeric hCN1, by molecular simulations, based on the predominant tautomers of the two peptides in aqueous solution at physiological pH by μs-long MD simulations (**CT-I** and **HT-I** in Chart 4.1), which were docked into the monomeric and dimeric hCN1 using flexible docking simulations followed by MD simulations and finally QM/MM calculations.

5.4.1 Comparison between monomeric and dimeric hCN1

The pose of both ligands in the active site shares similarities between monomeric and dimeric hCN1. However, the ligands form two hydrogen bonds with the second monomer in the dimeric hCN1 (Fig. 5.8 and Tab. 5.3). Notably, these hydrogen bonds are absent in the monomeric complexes.

As a result, while in the monomer, the backbone oxygen is orientated towards the Zn^{2+} ion, while in the dimer it is oriented toward His^{235}. Such feature might play a role for the experimentally observed higher activity of the dimer hCN1 with respect the monomer [110, 142, 149, 250]. Indeed, one of those two additional hydrogen bonds might shift the electron density in the C=O bond away the oxygen, therefore making the carbonyl carbon more prone to the nucleophilic attack of the bridging hydroxide anion. This hypothesis might be consistent with mutagenesis experiments on the structurally similar hAcy1 homodimer enzyme [149]: the point mutation of His^{206} (corresponding to the His^{235} in hCN1) to Asn significantly reduces the efficiency of the enzyme [149]. In our structural prediciction such mutation might reduce polarizability of the C=O group, possibly affecting the enzymatic reactivity. Finally, we noticed that the number of intramolecular hydrogen bonds in the monomer is larger than that of the same subunit in the dimer.

5.4.2 Comparison between carnosine and homocarnosine binding to hCN1

Homocarnosine differs from carnosine only by the length of the alkyl chain of the non-histidine residue: it features three CH_2 groups instead of two. In both complexes with hCN1, monomeric and dimeric, homocarnosine forms a larger number of hydrogen bonds than carnosine. This might play a role for the larger affinity *in vitro* of homocarnosine relative to carnosine.

The proposed structures of our Michaelis complexes are radically different from the one previously suggested for hCN1 [250], prior the X-ray structure was deposited in the PDB. The latter model did not indeed include the bridging water molecule (either in its ionic or neutral state) and considered different residues to be involved in the coordination of the two Zn^{2+} ions and in the interaction with carnosine (Fig. A2.7 in Appendix 2).

5.5 Conclussions

Understanding the molecular basis of the increasing affinity on passing from carnosine to homocarnosine may help design hCN1's inhibitors retaining carnosine's beneficial effects but with far higher affinity for the target. These could turn out to be efficient therapeutic agents. Hence, based on the predicted structural determinants, it is tempting to propose some general considerations and principles for the rational design of selective monomeric and dimeric hCN1 inhibitors. The addition of a methylene group in homocarnosine can generate additional hydrophobic interactions in the **HT-I•**monomeric and dimeric hCN1 complex relative to those of carnosine and might therefore be an important structural reason that converts the preferred substrate carnosine into the worser substrate homocarnosine. This may provide a molecular basis for the smaller K_m for homocarnosine compared to that of carnosine. These hydrophobic interactions are formed mostly between ligand's alkyl chain and side chain of Ile^{425} (Fig. 5.8). Thus, the retention of a propyl (or longer) amine chain attached at the ligand carbonyl might help generating analogs of carnosine that fit well into hCN1, although without undergoing hydrolysis. These analogs might therefore bind and block the enzyme, inhibiting its hydrolytic function [60, 152].

Replacing the scissile amide bond with an electrophilic carbonyl element has been used to generate a stable but unproductive tetrahedral intermediate during amide hydrolysis in other enzymes that cleave amide bonds [164]. The geometry of the ligand's pose is not expected to change drastically. I propose therefore that the

substitution of the hydrolyzed amide function with bioisosteres of this group (including for example 5-membered heteroaromatic groups such as triazole or tretrazole rings or trifluoroethylamines [164]) or with a carbamate or urea group might be an alternative strategy to generate inhibitors.

In conclusion, I have proposed here structures of Michaelis complexes formed between carnosine and homocarnosine substrates with hCN1. Our observances from QM/MM simulations are consistent with the experimental evidence that homocarnosine has a strong competitive inhibitory effect on carnosine turnover mediated by hCN1 [197]. Namely, despite homocarnosine's Michaelis complex is better stabilized with respect to carnosine by stronger hydrogen bond network, carnosine's position inside active site is better orientated for hydrolysis reaction. These findings may provide the means for accurate modeling of non-hydrolysable carnosine derivatives with high affinity for their target.

6 Ligand screening for non-hydrolyzable carnosine derivatives

6.1 Introduction

As already mentioned in previous chapters, carnosine expresses many beneficial abilities such as heavy metal ion chelation, free radical and reactive oxygen species scavenging, and antioxidant activity to mention only few of them [61, 123, 206]. These properties are responsible for carnosine's ability to prevent many pathological conditions, from diabetes, ischemia damage and cancer to neurodegenerative disorders as shown in animal models [1, 13, 15, 43, 52, 64, 93, 140, 203, 207, 224, 240, 259, 265]. However, its possible therapeutic applications are severely hampered by human serum carnosinase activity that rapidly hydrolyses carnosine in human serum [233]. So far several studies have been performed aimed to design derivatives either by derivatisation of the terminal amino group, carboxyl group or ethylene chain of β-alanine [22-24, 27, 248, 249]. One of the proposed solutions is also the use of carnosine's stereoisomer, D-carnosine, which retains most of the beneficial abilities of L-carnosine and is not recognized by carnosinase enzymes [249]. However D-carnosine compound suffers from low bioavailability, which limits its possible use in therapeutics [23].

None of the so far designed carnosine's derivatives was based on the structure of the Michaelis complex between hCN1 and carnosine. To my best knowledge, a possible binding pose of both, carnosine and homocarnosine, inside the hCN1 active site based

on the enzyme's crystal structure, is proposed here for the first time [191]. In the previous chapters, I described and proposed, based on our results from QM/MM simulations of carnosine and homocarnosine in Michaelis complex with hCN1**α**, why carnosine is hCN1's substrate and homocarnosine is its competitive inhibitor. Briefly, homocarnosine forms stronger and more persistent network of hydrogen bonds with hCN1 than carnosine. Additionally, since the two compounds differ only in the length of the alkyl chain (carnosine has two CH_2 group while homocarnosine has three), homocarnosine can possibly form additional hydrophobic interactions with hCN1**α** in comparison with carnosine. Still, homocarnosine is also hydrolyzed by hCN1, though at a slower rate than carnosine, therefore limiting possible therapeutic application of also this compound. Substitution of homocarnosine's peptide group with a bioisoster (*e.g.* urea or carbamate group) would presumably make the ligand non-hydrolysable and possibly increase its chance for therapeutic application [164].

My goal in ligand screening was to find non-hydrolysable homocarnosine derivatives whose binding poses inside hCN1**α** resemble that of homocarnosine. I orientated ligand screening on compounds bearing propylamine chain attached to either carbamate or urea group, which are both non-hydrolysable bioisosters of the amide group.

6.2 Methods

For the ligand-based screening, we selected the 'Clean Leads' library from the ZINC database, which consists of 4,591,276 compounds. The library was screened against three structure motives, namely aminopropyl-urea, aminopropyl-carbamate and aminocarbamate (**T1-3** in Fig. 6.1). Additional condition for a positive hit was that the compound's only chiral carbon atom (if present) has *S*-orientation (as it is the case of Cα in carnosine and homocarnosine).

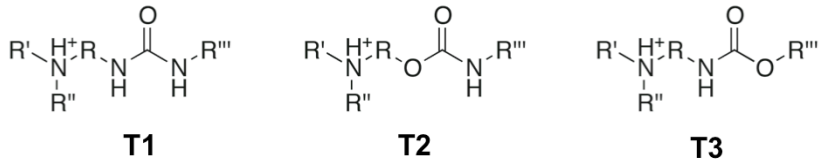


Figure 6.1: Structure motives used in the ligand-based screening. R is simple alkyl chain with the length of n methylene groups. R' and R'' can be either hydrogen atoms

or carbon atoms so that they form primary, secondary or tertiary amine. R''' can be a random functional group.

The selected compounds were then docked inside the monomeric hCN1**α** active site in order to further evaluate compounds obtained in screening. Docking calculations were performed using the same approach as we used in the second step of docking of carnosine and homocarnosine (see Chapter 5, Methods sections). The structures of the highest populated cluster of each ligand•monomeric hCN1**α** Michaelis complex were used for comparison with the **HT-I**•monomeric hCN1**α**.

Schrödinger software package was used for screening calculations and AutoDock4.2 [171] with the novel developed AutoDock force field for docking to zinc metalloproteins [219] was used in dockings.

6.3 Results and discussion

Six compounds from the 'Clean Leads' library satisfied both search criteria described above (**D1**-**6**; the compounds obtained by ligand-based screening are being tested now for their affinity and rate of hydrolysis by hCN1 in the group of Professor Enrico Rizzarelli at University of Catania and due to the confidentiality of the work I cannot disclose the structures of those compounds). Compounds **D1**-**5** are urea-based derivatives (based on **T1**), while **D6** is carbamate-based homocarnosine derivative (based on **T3**). Moreover, all compound have an amino group and compounds **D1**-**5** include also a carboxylic group. For structures of compounds **D1**-**6** see Fig. 6.2.

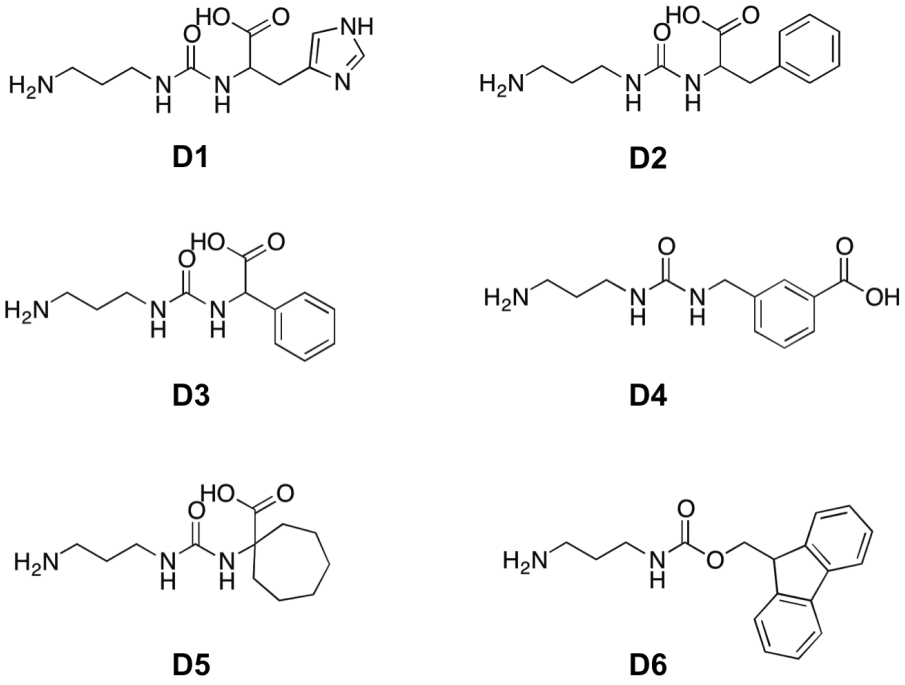


Figure 6.2: Structures of homocarnosine derivatives **D1**-**6** obtained from ligand-based screening.

In order to further evaluate compounds obtained in screening, I docked them in the hCN1**α** active site. The structures of the highest populated cluster of each ligand•monomeric hCN1**α** Michaelis complex obtained with docking calculations (name of the complex, **C1**-**6**, corresponds to the compound docked in the hCN1, **D1**-**6**) are presented here and compared with the previously obtained structure of **HT-I**•monomeric hCN1**α** Michaelis complex (see also Tab. 6.1 and Fig. 6.3).

Ligands in complexes **D1**-**4** and **D6** all have similar orientation compared to homocarnosine, while this is not the case in **D5**. In complexes **D1**-**5** the salt bridge between ligand and Arg^{350} is present as it is the case in **HT-I**•monomeric hCN1**α** complexes. Furthermore, the carboxylic group in complexes **C2, C4** and **C5** is stabilized with hydrogen bond with Asn^{220} side chain. On the other hand, the hydrogen bond between ligand's carboxylic group and peptide NH group of Thr^{424} observed in **HT-I**•monomeric hCN1**α** Michaelis complex is preserved only in **C3**. In **C6** these interactions are completely absent, since **D6** lacks a carboxylic group in that place.

Hydrogen bond between ligand's peptide amino group and Thr^{424} backbone oxygen, present in homocarnosine's Michaelis complex with hCN1**α**, is preserved in **C6.** In **C1-4** this interaction is strengthened, since ligands' urea group acts as a donor of two

hydrogen bonds to the Thr^{424} backbone oxygen. In **C5** this hydrogen bond is absent, due to the different orientation of the ligand inside the active site, namely carbonylic oxygen forms a hydrogen bond with Thr^{424} backbone NH group and is not orientated towards zinc ion as it is in **C1**-**4**,

Ligand's amino group in homocarnosine•monomeric hCN1**α** forms hydrogen bond with Asn^{203} side chain and this interaction is preserved in **C1** and **C3**, while in **C4** and **C5** form salt bridge with carboxylic groups of Asp^{421} and Glu^{451}, respectively. In **C2** and **C6** no hydrogen bonds between ligand and hCN1**α** could be observed. Hydrogen bonds between ligands and hCN1**α** are also reported in Tab. 6.1.

Hydrophobic interactions between propyl chain and Ile^{425} side chain observed in **HT-I**•monomeric hCN1**α** are mainly preserved in **C1**, **C3**, **C4** and **C6**, while in **C2** propyl chain has different orientation and forms hydrophobic interactions with Glu^{451}. No hydrophobic interactions between ligand's propyl chain and hCN1**α** are observed in **C5**.

Compounds **D1** and **D3** have the binding pose in hCN1 the most similar to the one of homocarnosine and also form similar network of interactions. On the other hand benzene ring of **D4** has different orientation compared to the imidazole ring of homocarnosine, though it forms more hydrogen bonds and hydrophobic interactions with hCN1**α** than homocarnosine. Finally, compound **D6** forms only one hydrogen bond with hCN1**α**, however it forms extensive network of hydrophobic interactions, which stabilize its position inside the active site. The other two compounds, **D2** and **D5**, seem to be less perspective, due to different orientation inside the active site and they appear to form fewer interactions with hCN1**α**.

Finally, I noticed that in all compounds obtained from screening calculations peptide group of homocarnosine is substituted by a longer urea or carbamate group. Therefore the aminoalkyl chain of these compounds is longer for one link in the chain in comparison with homocarnosine. Nevertheless, as it is seen from Fig. 6.3, this difference does not affect the orientation of compounds **D1**, **D3**, **D4** and **D6** inside the hCN1 active site. Even though the aminoalkyl chain of **D2** and **D5** has different orientation in the hCN1 active site than that of homocarnosine, which can be possibly explained by the longer aminoalkyl chain, the orientation of carbamate group of those two compounds is similar to the one of homocarnosine's peptide group.

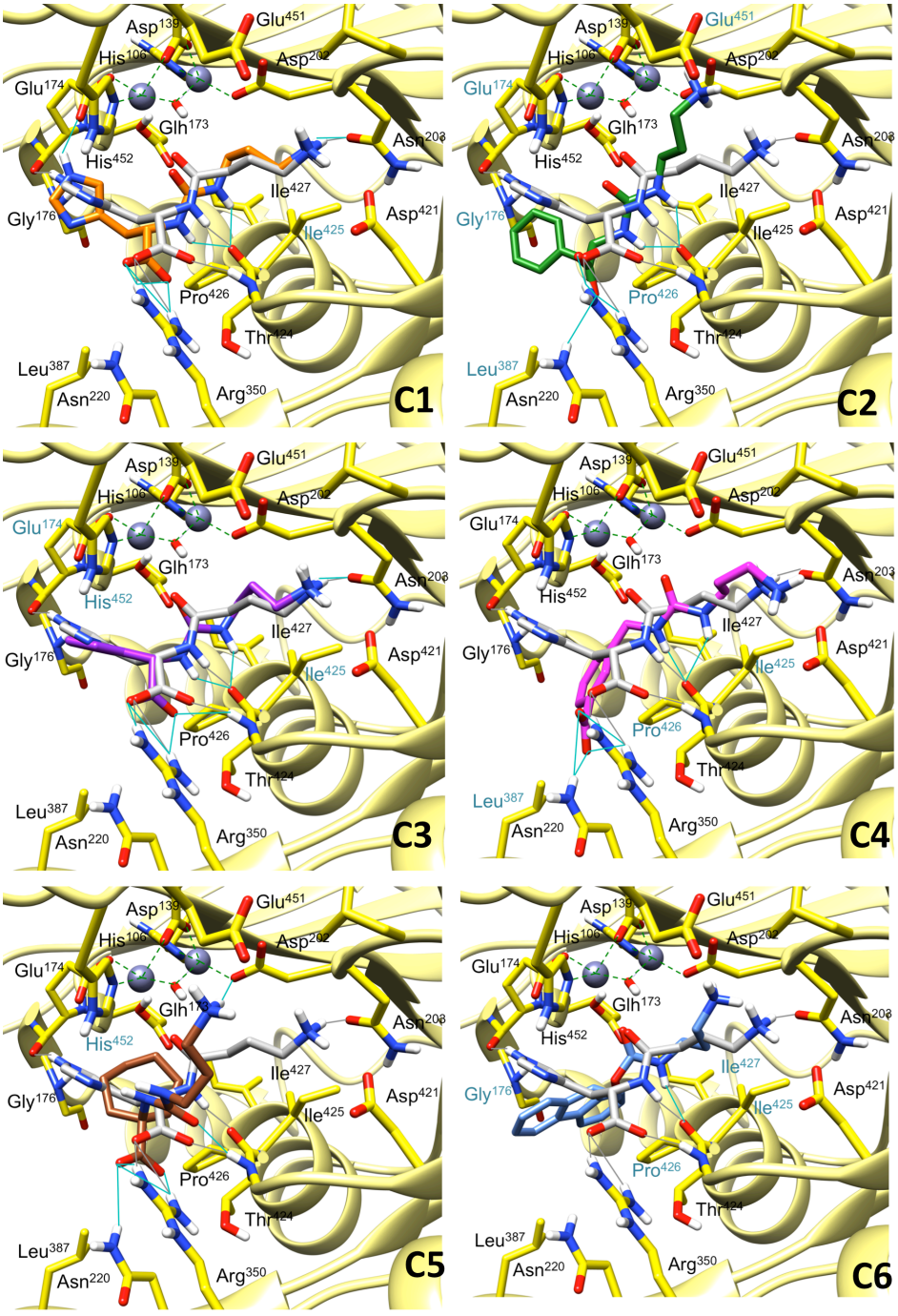


Figure 6.3: Structures of Michaelis complexes (**C1**-**6**) comparison with the structure of Hcar•monomeric hCN1 Michaelis complex. Hcar is always shown in grey, hCN1 in yellow, oxygen atoms in red, nitrogen atoms in blue, polar hydrogens in white and zinc ions as dark violet spheres. Non-polar hydrogens are not shown due to clarity. Hydrogen bonds between Hcar and hCN1 are shown as grey lines, while the cyan lines represent hydrogen bonds between compounds **D1**-**6** and hCN1. Names of residues forming hydrophobic interactions with the ligand are written in cyan. Dashed green lines represent bonds between zinc ions and coordinating residues.

Table 6.1: Hydrogen bond donors and acceptors between hCN1 active site and homocarnosine and its derivatives obtained from docking simulations. Asterisk indicates that the salt bridge is formed.

Hydrogen bond donor	Hydrogen bond acceptor	Homocarnosine• hCN1α	**D1•hCN1α (C1)**	**D2•hCN1α (C2)**	**D3•hCN1α (C3)**	**D4•hCN1α (C4)**	**D5•hCN1α (C5)**	**D6•hCN1α (C6)**
Ligand histidine ring	Glu^{174} backbone oxygen	✔						
Ligand histidine ring	Glu^{174} carboxylic group		✔					
Ligand amino group	Asn^{203} side chain	✔	✔		✔			
Asn^{220} side chain	Ligand carboxylic group			✔		✔✔[a]	✔	
Arg^{350} side chain	Ligand carboxylic group	✔*	✔*	✔*	✔*	✔*	✔*	
Ligand amino group	Asp^{421} carboxylic group					✔*		
Ligand peptide NH group		✔						
Ligand urea NH groups	Thr^{424} backbone oxygen		✔✔[b]	✔✔[b]	✔✔[b]	✔✔[b]		
Ligand carbamate NH group								✔
Thr^{424} peptide NH group	Ligand carboxylic group	✔			✔			
	Ligand urea CO group						✔	
Ligand amino group	Glu^{451} carboxylic group						✔*	
Total number of hydrogen bonds		5	5	4	5	6	4	1

[a] Each of the two carboxylic oxygen atoms forms one hydrogen bond with the Asn^{220} side chain.

[b] Each of the two urea NH groups forms one hydrogen bond with the Thr^{424} backbone oxygen.

6.4 Conclusions

Based on the ligand-based screening of the library with ~$4.5*10^6$ lead-like compound, I found six non-hydrolysable homocarnosine's derivatives (**D1**-**6**). Four of them (**D3**, **D4** and **D6**) seemed to be the most promising for further experimental evaluation to determine their affinity towards hCN1 and their propensity to be hydrolyzed by hCN1. They are being experimentally tested now in the group of Professor Rizzarelli from the University of Catania. The experimental evaluation of those compounds will validate my model described in the previous chapters and provide us with information on direction in which we should proceed in order to design novel non-hydrolysable carnosine derivatives and/or hCN1 inhibitors.

Since the **D3**, **D4** and **D6** compounds feature similar scaffold I am also performing additional structure-based screening based on the structure of **HT-I•**dimeric hCN1**α** Michaelis complex in order to find different type of scaffolds that could be used as a lead in subsequent design of novel hCN1 inhibitors.

7 Conclusions

The dipeptide carnosine shows many beneficial properties such as its ability to chelate heavy metals, to scavange free radical and ROS and to act as an antioxidant. It can also inhibit Aβ polymerization and α-synuclein oligomerization, thus acting as a neuroprotective agent. However, its potential therapeutic use is hindered by its rapid hydrolysis in tissues by the human tissue carnosinase and in the human bloodstream by the human serum carnosinase. The latter enzyme is a Zn-based enzyme and, intriguingly, displays different oligomerization states depending on the aging process. It mostly works as a monomer in children and as a dimer in adults. Here, I studied the role of human serum carnosinase dimerization for substrate binding. QM/MM simulations, complemented by molecular docking and classical molecular dynamics simulations, lead us to suggest that – in the dimeric complexes – one loop of the second monomer is entangled in the active site of the first monomer (and vice-versa, see Fig. 1.1). In the Michaelis complex, this loop features the His^{235} residue that forms one hydrogen bond with the substrate's peptide oxygen, thus affecting the polarization of the peptide bond. I suggest here that this interaction – obviously absent in the monomeric Michaelis complex – makes the dimer more prone to hydrolyze the substrate. This suggestion is consistent with the experimentally measured higher activity of the dimeric enzyme [233].

Carnosine derivatives could act either as human serum carnosinase inhibitors or as non-hydrolysable compounds, while retaining similar beneficial properties as carnosine. In an effort at designing new ligands retaining carnosine's beneficial properties and not being degraded in the serum, here I have used computational methods to identify the structural differences between the adducts of human serum carnosinase with carnosine and with the competitive inhibitor homocarnosine. My

calculations suggest that homocarnosine forms more hydrogen bonds inside the active site and is therefore bound stronger compared to carnosine. This might be consistent with the lower K_m of homocarnosine relative to carnosine [233].

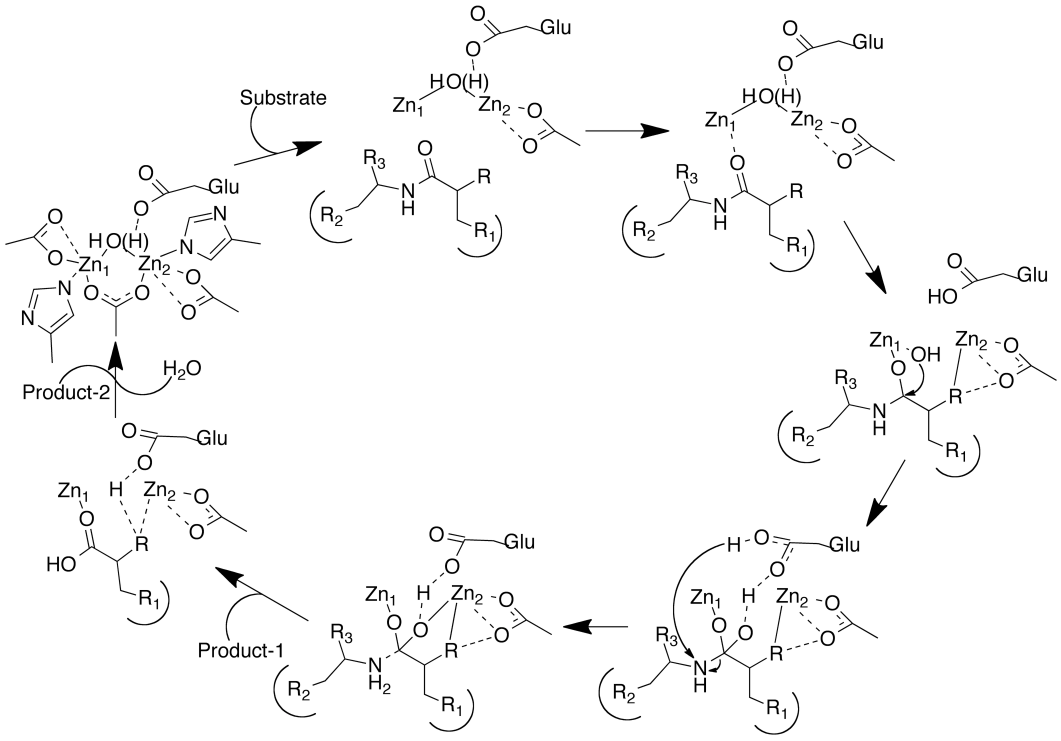
Next, I screened the libraries ("Clean Leads" library from ZINC database[o] and internal libraries) for compounds that would bind similarly inside the active site, as does homocarnosine. Since the only structural difference between carnosine and homocarnosine is an extra methylene group (Fig. 2.1), in these derivatives I selected either a carbamate or urea to replace the homocarnosine's amide group, in an effort to render the ligand non-hydrolysable while retaining its carnosine-like beneficial properties.

Most promising compounds from both screenings are now being tested *in vitro* for their affinity and activity against human serum carnosinase in the group of Professor Enrico Rizzarelli at University of Catania.

In conclusion, my thesis presents a molecular-level characterization of the Michaelis complex of carnosine and homocarnosine with human serum carnosinase in both monomeric and dimeric forms. My results, consistent with experiments, were used to screen potential candidates for therapeutic applications, including Alzheimer's and Parkinson's disease. These compounds are now being experimentally tested *in vitro*.

[o] http://zinc.docking.org/subsets/clean-leads

A1 Appendix 1 – Proposed mechanism of dipeptide hydrolysis



Scheme A1.1: Proposed general mechanism for the hydrolysis of a peptide catalyzed by a metallopeptidase with di-zinc co-catalytic active site based on the AAP as suggested by Holz *et al*. [101].

A2 Appendix 2 – Supporting information

QM/MM simulations of CT-I•hCN1β and HT-I•hCN1β

Michaelis complexes

CT-I•hCN1**β** and **HT-I**•hCN1**β** Michaelis complexes form the same hydrophobic interactions as in hCN1**α**, while few subtle differences for hydrogen bonds and salt bridges can be observed. Specifically, for **CT-I**, amino terminal group hydrogen bonds to backbone of Asp^{202} and the **CT-I**'s amide hydrogen bond with Thr^{424} backbone oxygen are more persistent than the same ones observed in hCN1**α** (see Tab. A2.16, Tab. 5.3 and Fig. A2.6). Further, the hydrogen bond between **CT-I**'s carboxyl group and Thr^{450} backbone unit is direct and not water-mediated as in hCN1**α** (despite being slightly less persistent). Finally, the **CT-I**'s amino terminal group salt bridge with Asp^{202}, observed in hCN1**α,** is now absent, while **CT-I**'s salt bridge with Arg^{350} side chain is more persistent.

In **HT-I**•hCN1**β**, Hcar's carbonylic oxygen forms a hydrogen bond with Glh^{173} side chain. **HT-I**'s amino terminal group forms hydrogen bonds with Asn^{203} and Ser^{423} side chains. These interactions are absent in hCN1**α** (see Tab. A2.16 and Tab. 5.3). **HT-I**'s amino terminal group forms also a water-mediated hydrogen bond with Asp^{202} backbone. In hCN1**α**, this hydrogen bond is direct. The hydrogen bond between **HT-I**'s amide and Thr^{424} backbone oxygen is less persistent than that in hCN1**α**. **HT-I**'s carboxyl group does not form a hydrogen bond with Thr^{424} backbone unit as in hCN1**α** (see Tab. A2.16 and Tab. 5.3). **HT-I**'s amino terminal group forms a persistent salt

bridge with Asp^{421}, while in hCN1**α** there is water-mediated hydrogen bond between those two residues. The **HT-I**'s salt bridge with Arg^{350} side chain is less persistent than in hCN1**α** (see Tab. A2.16 and Tab. 5.3).

HT-I forms one more hydrogen bond with hCN1 than **CT-I**. Therefore the flexibility of the ligand in the binding site is larger for **CT-I** than that of **HT-I** (see Fig. A2.5c).

As in hCN1**α**, the Zn^{2+} ions polyhedra are similar to those of the apo-enzyme (see Tab. A2.13 and A2.15). Distances between both Zn^{2+} ions and their coordinating residues are comparable to the values in the crystal structure of the apo enzyme (see Tab. A2.13 and A2.15). The major differences is the increased distance between Zn1 and carboxylic oxygen of Asp^{139} in the case of **HT-I•**hCN1**β**, while this is not the case in **CT-I•**hCN1**β** (see Tab. A2.13) and the decreasing distance between bridging moiety and both Zn^{2+} ions in both adducts, compared to the crystal structure (see Tab. A2.13).

Table A2.1: Partial atom charges of atoms in **CT-I** and **CT-II** together with legend of labeling of atoms in **CT-I** and **CT-II**. Atom types are gaff atom types [256].

CT-I			**CT-II**		
Atom type	Atom label	Charge	Aom type	Atom label	Charge
na	N_τ	-0.291	nc	N_τ	-0.557
nd	N_π	-0.583	na	N_π	-0.250
hn	H (on N_τ)	0.300	hn	H (on N_π)	0.266
cc	C_2	0.229	cd	C_2	0.232
h5	H (on C_2)	0.055	h5	H (on C_2)	0.057
cc	C_4	-0.266	cc	C_4	0.080
h4	H (on C_5)	0.194	h4	H (on C_5)	0.062
cd	C_γ	0.369	cd	C_γ	-0.030
c3	C_β	-0.122	c3	C_β	-0.072
hc	H (on C_β)	0.038	hc	H (on C_β)	0.029
c3	C_α	0.177	c3	C_α	0.327
c	C (carboxyl)	0.639	c	C (carboxyl)	0.605
o	O (carboxyl)	-0.650	o	O (carboxyl)	-0.624
h1	H (on C_α)	-0.024	h1	H (on C_α)	-0.039
n	N (amide)	-0.501	n	N (amide)	-0.628
hn	H (amide)	0.261	hn	H (amide)	0.281
c	C (carbonyl)	0.599	c	C (carbonyl)	0.660
o	O (carbonyl)	-0.553	o	O (carbonyl)	-0.563
c3	$C_{\alpha'}$	-0.271	c3	$C_{\alpha'}$	-0.272
hc	H (on $C_{\alpha'}$)	0.079	hc	H (on $C_{\alpha'}$)	0.088
c3	$C_{\beta'}$	0.233	c3	$C_{\beta'}$	0.173
hx	H (on $C_{\beta'}$)	0.041	hx	H (on $C_{\beta'}$)	0.060
n4	N (amine)	-0.441	n4	N (amine)	-0.427
hn	H (amine)	0.328	hn	H (amine)	0.328

Table A2.2: Bond lengths (in Å) for tautomers **CT-I** and **CT-II**. Atom types are gaff atom types [256] (see also Tab. A2.1).

CT-I		**CT-II**	
Bond	Bond length	Bond	Bond length
hn-n4	1.03	hn-n4	1.03
n4-c3	1.53	n4-c3	1.54
hx-c3	1.09	hx-c3	1.09
c3-c3	1.53	c3-c3	1.53
hc-c3	1.10	hc-c3	1.10
c3-c	1.58	c3-c	1.58
c-o	1.25	c-o	1.25
c-n	1.32	c-n	1.32
n-hn	1.06	n-hn	1.05
n-c3	1.46	n-c3	1.46
c3-h1	1.10	c3-h1	1.10
c3-cd	1.50	c3-cd	1.50
cd-cc	1.38	cd-cc	1.38
na-hn	1.01	na-hn	1.04
cc-h4	1.08	cc-h4	1.08
cc-na	1.37	cc-nc	1.38
nd-cc	1.32	nc-cd	1.32
cd-nd	1.39	cd-na	1.37
cc-h5	1.08	h5-cd	1.08

Table A2.3: Angle values (in °) for tautomers **CT-I** and **CT-II**. Atom types are gaff atom types [256] (see also Tab. A2.1).

CT-I		CT-II	
Angle	Angle value	Angle	Angle value
hn-n4-hn	107.9	hn-n4-hn	107.9
hn-n4-c3	111.0	hn-n4-c3	111.0
n4-c3-hx	106.8	n4-c3-hx	106.8
n4-c3-c3	111.3	n4-c3-c3	111.3
hx-c3-hx	110.1	hx-c3-hx	109.9
hx-c3-c3	110.8	hx-c3-c3	110.9
c3-c3-hc	109.9	c3-c3-hc	109.8
hc-c3-hc	107.6	hc-c3-hc	107.7
c3-c3-c	110.5	c3-c3-c	110.4
hc-c3-c	109.5	hc-c3-c	109.6
c3-c-o	116.1	c3-c-o	116.4
c3-c-n	113.5	c3-c-n	113.8
o-c-n	129.2	o-c-n	128.6
c-n-hn	126.8	c-n-hn	126.0
c-n-c3	128.0	c-n-c3	127.8
hn-n-c3	103.4	hn-n-c3	105.0
n-c3-c	102.5	n-c3-c	102.9
n-c3-h1	109.9	n-c3-h1	109.5
n-c3-c3	111.0	n-c3-c3	112.1
h1-c3-c3	109.1	h1-c3-c3	108.6
h1-c3-c	109.3	h1-c3- c	109.0
o-c-o	129.1	o-c-o	128.2
hc-c3-cd	108.5	hc-c3-cd	109.3
c3-cd-cc	129.6	c3-cd-cc	132.5
c3-c3-cd	116.0	c3-c3-cd	114.5
c3-cd-nd	120.2	c3-cd-na	122.4
cd-cc-na	105.3	cd-cc-nc	110.9
cd-cc-h4	129.4	cd-cc-h4	127.6
h4-cc-na	125.3	h4-cc-nc	121.5
cc-na-cc	107.4	cc-nc-cd	104.5
cd-nd-cc	105.5	cd-na-cd	107.1
na-cc-h5	122.4	nc-cd-h5	125.3
cc-na-hn	126.3	cd-na-hn	126.3
na-cc-nd	111.6	nc-cd-na	112.4
h5-cc-nd	126.0	h5-cd-na	122.3
nd-cd-cc	110.2	na-cd-cc	105.1

Table A2.4: Carnosine's hydrogen bond functionalities form hydrogen bonds with water molecules (similarly to what was found in reference [264]). The table shows the average and maximum water coordination numbers of hydrogen bond acceptors/donors for **CT-I** and **CT-II**. For both tautomers, all hydrogen bond acceptors and donors form hydrogen bond with water molecules with around 80% probability also consistent with their similar size of solvent accessible surface (see Fig. A2.1).

	CT-I		**CT-II**	
Hydrogen bond acceptor/donor	Average water coordination number	Maximum water coordination number	Average water coordination number	Maximum water coordination number
H_π	/	/	0.81 ± 0.33	1.07
H_τ	0.83 ± 0.32	1.05	/	/
H (amide)	0.79 ± 0.36	1.11	0.79 ± 0.36	1.11
H (primary amine)	0.85 ± 0.32	1.67	0.85 ± 0.32	1.56
H_α	0.78 ± 0.37	1.12	0.78 ± 0.37	1.10
H_2 (on C_2)	0.82 ± 0.36	1.19	0.82 ± 0.36	1.15
O (carboxyl)	0.85 ± 0.33	2.02	0.84 ± 0.33	1.89
O (carbonyl)	0.81 ± 0.34	1.37	0.81 ± 0.34	1.27
N_π	0.81 ± 0.34	1.08	0.79 ± 0.37	1.04
N_τ	0.80 ± 0.37	1.08	0.82 ± 0.33	1.24

Table A2.5: Not-weighted calculated ^{13}C chemical shifts of **CT-Ia-d** and **CT-IIa-d** conformers of carnosine. Chemical shifts were scaled with respect to the chemical shift of ^{13}C TMS.

Carbon atom	CT-Ia	CT-Ib	CT-Ic	CT-Id	CT-IIa	CT-IIb	CT-IIc	CT-IId
CO_2^-	180.1	180.3	180.4	182.3	176.3	176.2	188.3	182.1
C(O)NH	162.6	168.1	163.3	165.6	166.2	173.5	165.2	176.6
C_2	135.4	139.0	135.0	138.0	142.2	139.9	144.5	138.8
C_4	144.6	151.6	146.5	143.0	144.0	135.6	139.6	137.0
C_5	121.8	117.6	113.6	121.0	131.9	136.0	135.7	134.9
C_α	68.0	68.0	68.9	64.8	69.1	68.6	68.8	66.2
$C_{\beta'}$	48.4	35.7	57.0	39.0	40.8	36.3	44.7	50.0
$C_{\alpha'}$	47.3	37.5	4.0	45.9	46.8	39.8	43.3	40.2
C_β	39.7	45.3	35.4	39.0	37.0	37.8	33.3	35.5

Table A2.6: Not-weighted calculated ^{1}H chemical shifts of **CT-Ia-d** and **CT-IIa-d** conformers of carnosine. Chemical shifts were scaled with respect to the chemical shift of ^{1}H TMS.

Hydrogen atom	**CT-Ia**	**CT-Ib**	**CT-Ic**	**CT-Id**	**CT-IIa**	**CT-IIb**	**CT-IIc**	**CT-IId**
H_2	7.74	7.77	7.50	7.16	7.32	7.50	7.89	7.75
H_5	7.10	7.56	6.92	7.06	7.24	7.64	7.41	7.18
$H_{\beta R}$, $H_{\beta S}$	2.75; 3.40	2.83; 3.27	2.95; 3.41	2.51; 3.04	2.80; 3.29	2.94; 3.81	2.93; 2.90	3.01; 3.02
H_α	3.59	4.35	3.96	4.43	4.30	5.07	4.18	4.49
NH	7.39	7.59	9.51	7.48	6.67	8.01	8.31	6.34
$H_{\alpha'}$ (CH_2)	2.65	2.78	2.44	2.56	2.69	2.60	2.74	2.75
$H_{\beta'}$ (CH_2)	3.19	3.19	3.95	3.28	3.41	3.18	3.69	3.59

Table A2.7: Not-weighted calculated ^{1}H chemical shifts of **HT-Ia-d** and **HT-IIa-d** conformers of homocarnosine. Chemical shifts were scaled with respect to the chemical shift of ^{11}H DSS.

Hydrogen atom	**HT-IIa**	**HT-IIb**	**HT-IIc**	**HT-IId**	**HT-Ia**	**HT-Ib**	**HT-Ic**	**HT-Id**
H_2	7.25	8.42	8.06	7.72	8.06	7.27	7.43	7.45
H_5	7.12	7.17	7.33	7.21	7.50	6.88	6.91	7.08
$H_{\beta R}$, $H_{\beta S}$	3.02; 2.89	3.20; 3.03	3.58; 2.90	3.19; 2.85	3.53; 3.11	2.82; 2.71	3.31; 2.90	3.97; 2.95
H_α	5.15	4.96	4.58	4.12	4.03	4.58	3.93	4.23
$H_{\alpha'}$ (CH_2)	2.49	3.71	2.27	2.50	2.05	2.20	2.35	2.24
$H_{\beta'}$ (CH_2)	2.03	2.29	2.12	2.30	2.43	1.85	2.12	2.26
$H_{\gamma'}$ (CH_2)	3.28	3.45	3.60	3.76	3.82	3.17	3.74	3.24
NH_3	4.64	5.16	4.78	4.76	4.46	4.65	4.57	4.67

Table A2.8: Statistical data (in ppm) of calculated ^{1}H and ^{13}C chemical shifts (relative to the standard used). These include RMS error, mean error, mean absolute error, maximum error and standard deviation of absolute error.

	^{1}H			^{13}C	
	TMS	[2,2,3,3-2H4]-3-trimethylsilyl-propanoate	DSS	TMS	[2,2,3,3-2H4]-3-trimethylsilyl-propanoate
RMS error	0.52	1.01	0.91	11.3	11.2
Mean error	0.11	0.77	0.38	5.9	8.4
Mean absolute error	0.44	0.78	0.69	9.5	10.0
Maximum error	1.08	1.84	-1.75	27.0	17.3
Standard deviation of absolute error	0.27	0.71	0.63	6.4	5.4

Table A2.9: Protonation states of histidine residues and of Glu^{173}. The acidic amino acids other than Glu^{173} are negatively charged.

Residue number	Protonation state
106, 314, 324, 367, 452	HID
235, 244, 263, 388, 481	HIE
39, 288, 354, 436	HIP
173	GLH

Table A2.10: Relative free energies (in kcal/mol), calculated as described in the Part *(ii)* of "Molecular docking – Monomer" section of Methods part in the main text, of model system of apo hCN1 active site and its modifications (based on the hCN1 crystal structure), where different residues (*i.e.* $Glh^{173}\boldsymbol{\alpha}$ and $Glh^{173}\boldsymbol{\beta}$, Glu^{174}, Asp^{202}) are protonated and hydroxyl anion is bridging two Zn^{2+} ions. Relative free energies are calculated with respect to the model system, where water is bridging two Zn^{2+} ions and other residues are deprotonated.

Dielectric constant (ε)	$Glh^{173}\boldsymbol{\alpha}$	$Glh^{173}\boldsymbol{\beta}$	Glu^{174}	Asp^{202}
1	-13.0	-0.9	38.9	37.4
2	-4.4	3.9	34.1	31.9
4	0.6	6.4	30.5	28.3
8	3.4	7.5	28.2	26.3
16	4.8	8.1	26.9	25.2
32	5.5	8.4	26.3	24.6
64	5.9	8.5	25.9	24.3
80	6.0	8.5	25.8	24.3

Table A2.11: Relative free energies (in kcal/mol) of model system of apo hCN1 active site and its modifications (based on the hCN1 structure obtained after first step of docking), where different residues (*i.e*. $Glh^{173}\boldsymbol{\alpha}$ and $Glh^{173}\boldsymbol{\beta}$, Glu^{174}, Asp^{202}) are protonated and hydroxyl anion is bridging two Zn^{2+} ions. Relative free energies are calculated with respect to the model system, where water is bridging two Zn^{2+} ions and other residues are deprotonated.

Dielectric constant (ε)	$Glh^{173}\boldsymbol{\alpha}$	$Glh^{173}\boldsymbol{\beta}$	Glu^{174}	Asp^{202}
1	-29.2	-32.3	9.8	4.2
2	-22.9	-25.2	5.3	0.8
4	-19.1	-20.6	2.7	-1.1
8	-17.1	-18.1	1.2	-2.1
16	-16.0	-16.7	0.4	-2.6
32	-15.5	-16.0	0.0	-2.9
64	-15.2	-15.7	-0.2	-3.0
80	-15.1	-15.6	-0.3	-3.1

Table A2.12: Distances (in Å) between two zinc ions and their coordinating residues from my 2-ps long QM/MM simulations of carnosine and homocarnosine complexes with monomeric and dimeric hCN1α at 298 K and crystallographic data of apo-enzyme (PDB ID 3DLJ). Bond lengths from simulations are reported as average values.

Bond	Crystallographic data	Monomeric hCN1α		Dimeric hCN1α	
		CT-I•hCN1α	**HT-I•**hCN1α	**CT-I•**hCN1α	**HT-I•**hCN1α
Zn1-His^{452}	2.0	2.0 ± 0.1	2.0 ± 0.1	2.0 ± 0.1	2.0 ± 0.1
Zn1-Glu^{174}	2.1	2.0 ± 0.1	2.0 ± 0.1	2.1 ± 0.1	2.0 ± 0.1
Zn1-Asp^{139}	2.2	4.2 ± 0.5	4.1 ± 0.2	2.1 ± 0.1	2.1 ± 0.1
Zn1-OH^-	2.8	2.0 ± 0.1	2.0 ± 0.1	2.0 ± 0.1	2.1± 0.1
Zn2-Asp^{139}	2.0	2.1 ± 0.1	2.1 ± 0.1	2.0 ± 0.1	2.0 ± 0.1
Zn2-Asp^{202}	1.8	2.2 ± 0.1	2.1 ± 0.1	2.5 ± 0.5	2.1 ± 0.1
Zn2-His^{106}	2.0	2.2 ± 0.1	2.1 ± 0.1	2.1 ± 0.1	2.1 ± 0.1
Zn2-OH^-	2.8	2.2 ± 0.1	2.2 ± 0.1	2.1 ± 0.1	2.1 ± 0.1
Zn1-ligand's carbonyl oxygen	/	4.2 ± 0.3	4.0 ± 0.4	5.1 ± 0.3	4.5 ± 0.3
Zn1-Zn2	3.5	3.8 ± 0.1	3.5 ± 0.1	3.4 ± 0.1	3.3 ± 0.1

Table A2.13: Distances (in Å) between two zinc ions and their coordinating residues from my 1-ps long QM/MM simulations of carnosine complex with monomeric hCN1**α** at 298 K and 303 K, carnosine and homocarnosine complex with monomeric hCN1**β** at 298 K and crystallographic data of apo-enzyme (PDB ID 3DLJ). Bond lengths from simulations are reported as average values.

Bond	Crystallographic data	**CT-I•**hCN1**α** (T = 298 K)	**CT-I•**hCN1**α** (T = 303 K)	**CT-I•**hCN1**β** (T = 298 K)	**HT-I•** hCN1**β** (T = 298 K)
Zn1-His^{452}	2.0	2.0 ± 0.1	2.0 ± 0.1	2.1 ± 0.1	2.0 ± 0.1
Zn1-Glu^{174}	2.1	2.0 ± 0.1	2.0 ± 0.1	2.0 ± 0.1	2.1 ± 0.1
Zn1-Asp^{139}	2.2	3.8 ± 0.2	3.8 ± 0.2	2.1 ± 0.1	3.7 ± 0.3
Zn1-OH^-	2.8	2.0 ± 0.1	2.0 ± 0.1	2.0 ± 0.1	2.0 ± 0.1
Zn2-Asp^{139}	2.0	2.1 ± 0.1	2.1 ± 0.1	2.1 ± 0.1	2.0 ± 0.1
Zn2-Asp^{202}	1.8	2.2 ± 0.1	2.2 ± 0.1	2.1 ± 0.1	2.2 ± 0.1
Zn2-His^{106}	2.0	2.1 ± 0.1	2.1 ± 0.1	2.2 ± 0.1	2.2 ± 0.1
Zn2-OH^-	2.8	2.1 ± 0.1	2.1 ± 0.1	2.1 ± 0.1	2.1 ± 0.1
Zn1-ligand's carbonyl oxygen	/	4.1 ± 0.2	4.1 ± 0.2	5.5 ± 0.4	4.0 ± 0.3
Zn1-Zn2	3.5	3.8 ± 0.1	3.8 ± 0.1	3.3 ± 0.1	3.6 ± 0.1

Table A2.14: Angles (in °) between two zinc ions and their coordinating residues from my 2-ps long QM/MM simulations of carnosine and homocarnosine complexes with monomeric and dimeric hCN1**α** at 298 K and crystallographic data of apo-enzyme (PDB ID 3DLJ). Angle values from simulations are reported as average values.

Angle	Crystallographic data	Monomeric hCN1**α**		Dimeric hCN1**α**	
		CT-I•hCN1**α**	**HT-I•**hCN1**α**	**CT-I•**hCN1**α**	**HT-I•**hCN1**α**
His^{452}-Zn1-Glu^{174}	115	122 ± 6	129 ± 6	119 ± 7	120 ± 8
His^{452}-Zn1-Asp^{139}	83	80 ± 4	66 ± 5	98 ± 4	97 ± 5
His^{452}-Zn1-OH^-	115	124 ± 9	120 ± 5	121 ± 8	115 ± 6
Glu^{174}-Zn1-Asp^{139}	101	97 ± 12	95 ± 11	94 ± 8	101 ± 6
Glu^{174}-Zn1-OH^-	126	109 ± 7	103 ± 6	112 ± 6	110 ± 7
Asp^{139}-Zn1-OH^-	103	74 ± 8	85 ± 6	106 ± 6	111 ± 7
His^{106}-Zn2-Asp^{139}	106	91 ± 4	94 ± 5	114 ± 9	116 ± 6
His^{106}-Zn2-Asp^{202}	102	92 ± 5	101 ± 5	122 ± 11	114 ± 9
His^{106}-Zn2-OH^-	128	100 ± 7	89 ± 4	97 ± 4	98 ± 8
Asp^{139}-Zn2-Asp^{202}	127	139 ± 12	119 ± 7	111 ± 9	112 ± 7
Asp^{139}-Zn2-OH^-	101	102 ± 9	141 ± 9	101 ± 6	95 ± 7
Asp^{202}-Zn2-OH^-	96	117 ± 8	99 ± 8	105 ± 12	117 ± 9

Table A2.15: Angles (in °) between two zinc ions and their coordinating residues from my 1-ps long QM/MM simulations of carnosine complex with monomeric hCN1**α** at 298 K and 303 K, carnosine and homocarnosine complex with monomeric form of hCN1**β** at 298 K and crystallographic data of apo-enzyme (PDB ID 3DLJ). Bond lengths from simulations are reported as average values.

Angle	Crystallographic data	**CT-I•**hCN1**α** (T = 298 K)	**CT-I•**hCN1**α** (T =303 K)	**CT-I•**hCN1**β** (T = 298 K)	**HT-I•** hCN1**β** (T = 298 K)
His^{452}-Zn1-Glu^{174}	115	120 ± 6	119 ± 7	118 ± 6	111 ± 9
His^{452}-Zn1-Asp^{139}	83	78 ± 4	79 ± 5	100 ± 6	72 ± 5
His^{452}-Zn1-OH^-	115	129 ± 8	129 ± 9	110 ± 7	135 ± 15
Glu^{174}-Zn1-Asp^{139}	101	107 ± 7	108 ± 8	108 ± 6	93 ± 11
Glu^{174}-Zn1-OH^-	126	106 ± 7	107 ± 7	106 ± 7	104 ± 10
Asp^{139}-Zn1-OH^-	103	68 ± 6	69 ± 6	114 ± 4	79 ± 8
His^{106}-Zn2-Asp^{139}	106	91 ± 5	91 ± 5	92 ± 7	97 ± 5
His^{106}-Zn2-Asp^{202}	102	93 ± 5	93 ± 4	117 ± 4	97 ± 6
His^{106}-Zn2-OH^-	128	102 ± 7	102 ± 8	93 ± 5	94 ± 9
Asp^{139}-Zn2-Asp^{202}	127	134 ± 11	136 ± 10	93 ± 9	122 ± 7
Asp^{139}-Zn2-OH^-	101	105 ± 7	104 ± 8	99 ± 7	115 ± 5
Asp^{202}-Zn2-OH^-	96	118 ± 8	118 ± 7	135 ± 12	118 ± 10

Table A2.16: Probability of hydrogen bonds between the ligand and the hCN1**β** active site from 1-ps QM/MM simulations. ^{WM} indicates water mediated hydrogen bond.

Hydrogen bond donor	Hydrogen bond acceptor	**CT-I•**hCN1**β**	**HT-I•**hCN1**β**
Glh^{173} protonated carboxylic group	Ligand backbone oxygen	/	100%
Ligand histidine ring	Glu^{174} backbone oxygen	90%	43% 79%^{WM}
Ligand amino group	Asp^{202} backbone oxygen	91%	27% 68%^{WM}
Ligand amino group	Asp^{202} carboxylic group	/	/
Ligand amino group	Asn^{203} side chain	/	8%
Arg^{350} side chain*	Ligands carboxylic group	a) / b) 87% c) 55% d) 6%	a) 14% b) 28% c) 32% d) 12%
Ligand amino group	Asp^{421} carboxylic group	/	97%
Ligand amino group	Ser^{423} hydroxyl group	/	59%
Ligand peptide NH group	Thr^{424} backbone oxygen	46%	41%
Thr^{424} peptide NH group	Ligand carboxylic group	29%	/

* between Arg350 and carboxylic group are 4 possible hydrogen bonds (marked with, a), b,) c) and d), see also Fig. A2.6); also salt bridge is formed here.

Table A2.17: Relative free energies (in kcal/mol) of hCN1**α** and hCN1**β** of *holo* hCN1 active site (with **CT-I** and **HT-I** present inside the active site). Relative free energies were calculated with respect to the model system, where water is bridging two Zn^{2+} ions and Glu^{173} is deprotonated. The position of the hydrogen atoms was optimized. Single-point calculations were performed on the optimized structures. Values for the system with **HT-I** present inside the active site are in parentheses.

Dielectric constant (ε)	Glh^{173}**α**	Glh^{173}**β**
1	-24.8 (-25.2)	-4.3 (-16.6)
2	-18.9 (-19.3)	-1.6 (-8.0)
4	-15.3 (-16.0)	-0.1 (-3.2)
8	-13.4 (-13.8)	0.7 (-0.2)
16	-12.4 (-12.8)	1.1 (1.0)
32	-11.8 (-12.6)	1.3 (1.7)
64	-11.6 (-12.1)	1.4 (2.3)
80	-11.5 (-12.0)	1.4 (2.1)

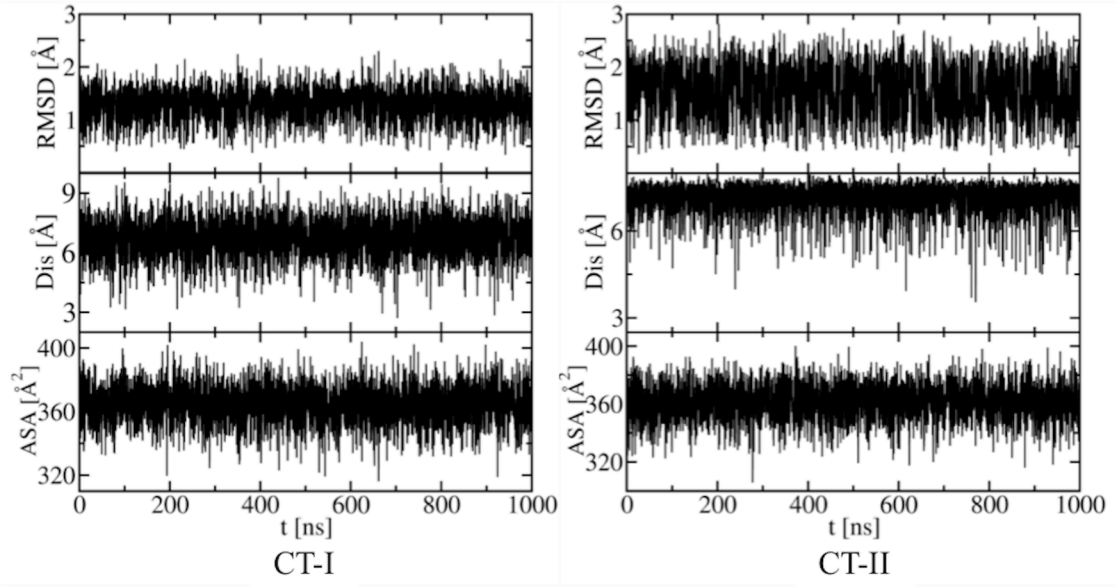


Figure A2.1: (a) RMSD, (b) distance between terminal NH_3^+ and COO^-, and (c) accessible surface area plotted as a function of simulated time for **CT-I** and **CT-II** tautomers in aqueous solution.

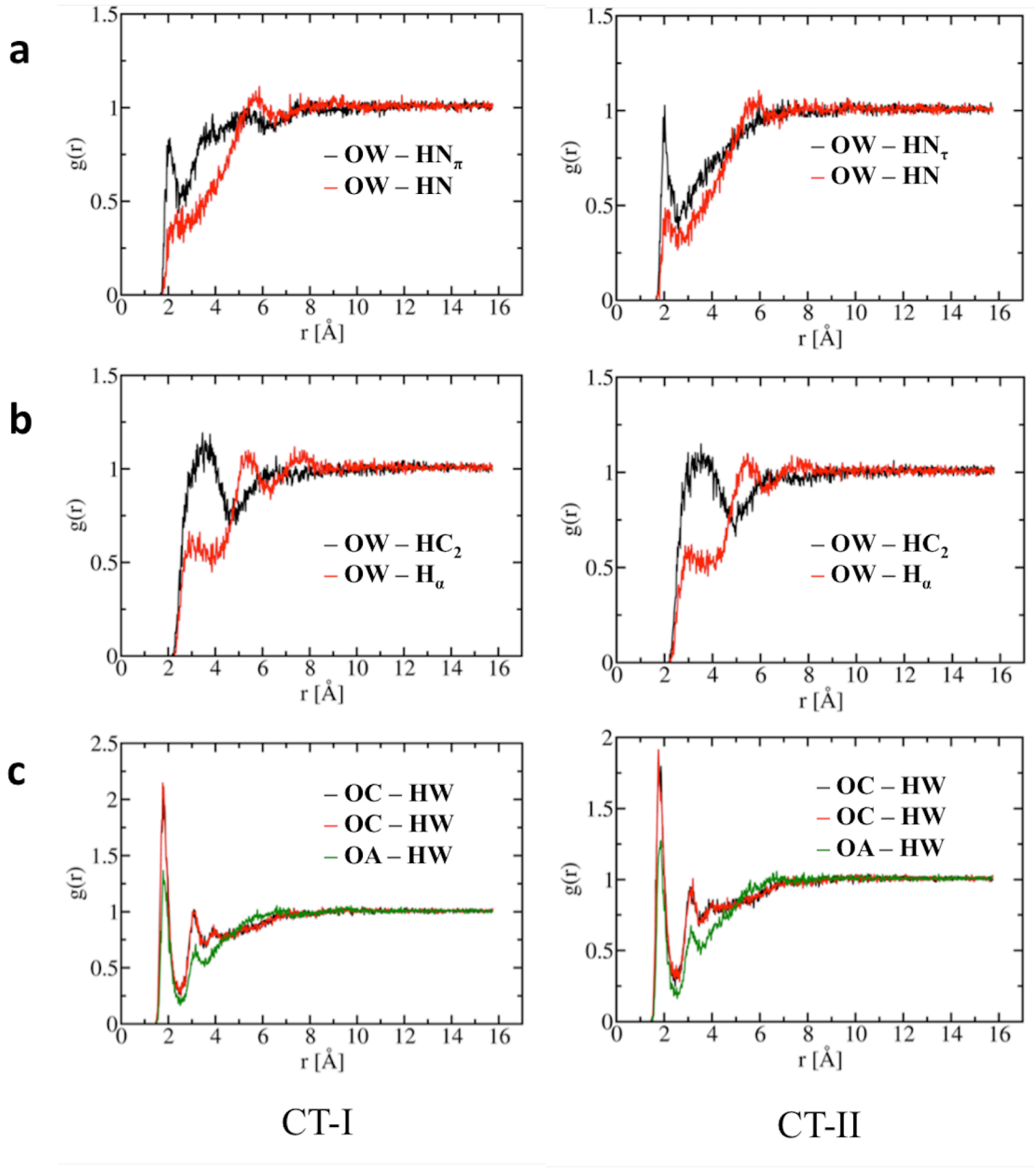


Figure A2.2: Radial distribution functions g(r) of **CT-I** and **CT-II** in water. (a) g(r) between the N-hydrogen atom in imidazole ring (HN_{π} and HN_{τ}) of carnosine and oxygen atom in water (OW) (black line); g(r) of amide hydrogen (HN) of carnosine and oxygen atom in water (OW) (red line). (b) g(r) of the C-hydrogen atom in imidazole ring (HC_2) of carnosine and oxygen atom of water (OW) (black line); g(r) of C_{α}-hydrogen of carnosine and oxygen atom in water (OW) (red line). (c) g(r) of both carboxylic oxygen atoms (OC) of carnosine and hydrogen atom in water (HW) (black and red lines); g(r) of amide oxygen atom (OA) of carnosine and hydrogen atom in water (HW) (green line). For the name of the atoms see Chart 4.1 in the Chapter 4.

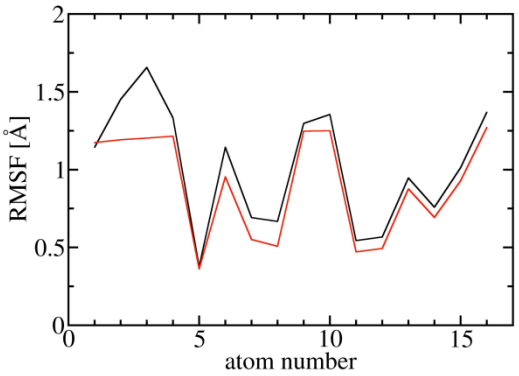


Figure A2.3: RMSF values of pairs of equivalent atoms (data shown only for C, N and O atoms) in **CT-I** (red line) and **CT-II** (black line) tautomers of carnosine during one of the two MD simulations reported here.

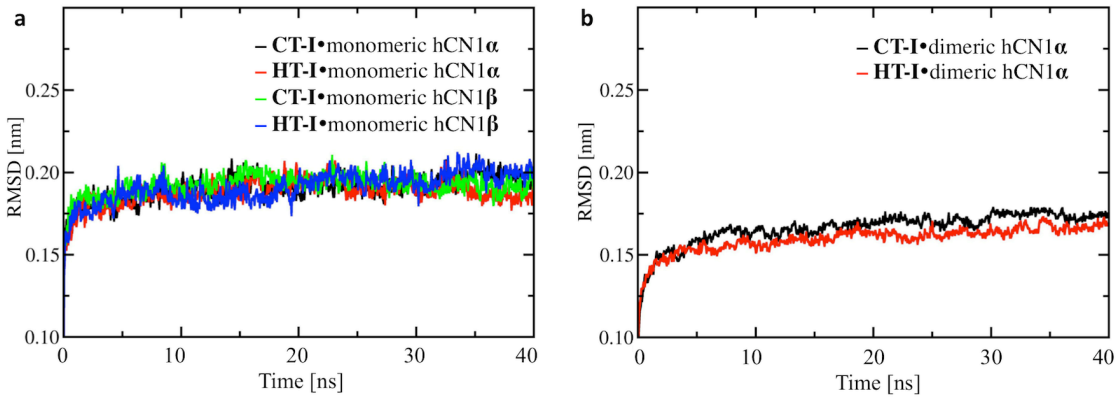


Figure A2.4: RMSD of hCN1 relative to the structure of the enzyme used in the second step of molecular docking during 40 ns classical MD. a) monomeric hCN1 complexes (**CT-I•**hCN1**α** in black, **CT-I•**hCN1**β** in green, **HT-I•**hCN1**α** in red and **CT-I•**hCN1**β** in blue). b) dimeric hCN1 complexes (**CT-I•**hCN1**α** in black and **HT-I•**hCN1**α** in red).

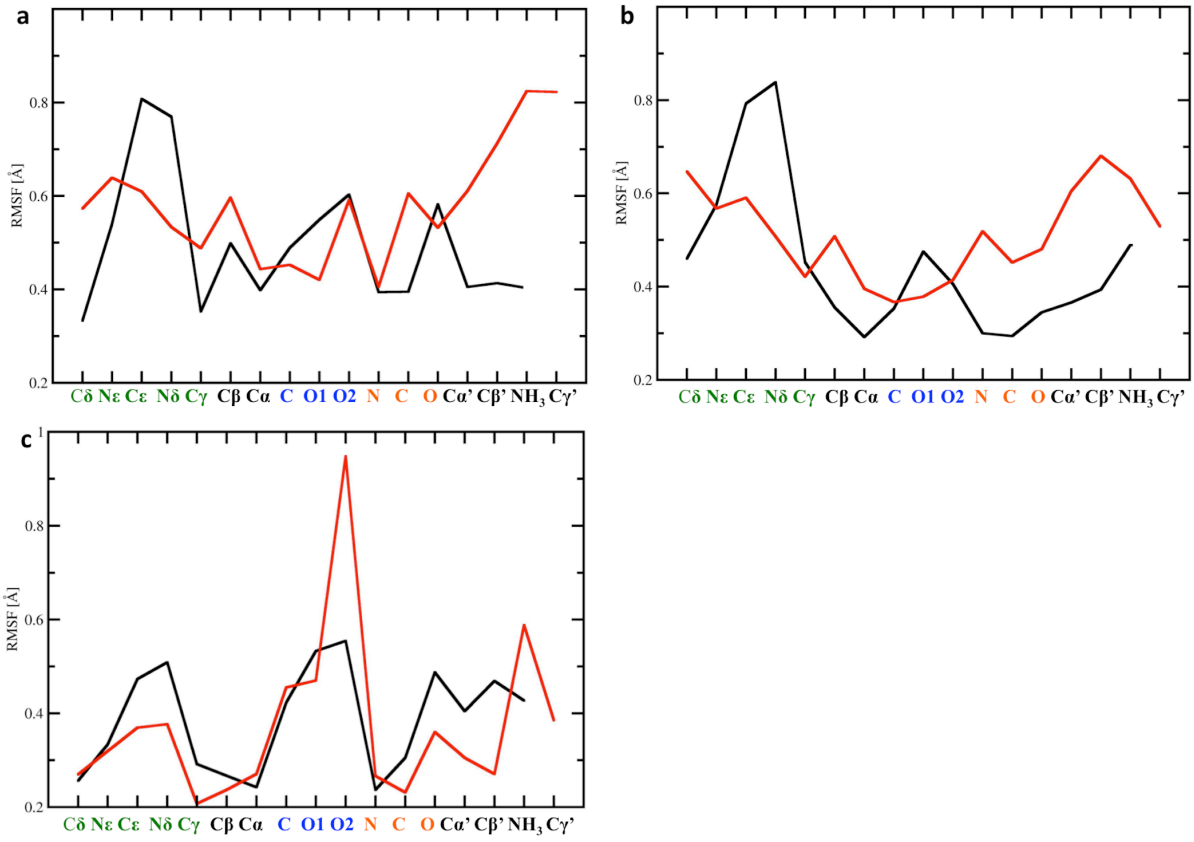


Figure A2.5: RMSF values for heavy atoms of both ligands inside a) monomeric hCN1**α** from 2 ps long QM/MM simulations; b) dimeric hCN1**α** from 2 ps long QM/MM simulations and c) monomeric hCN1**β** from 1 ps long QM/MM simulations. Carnsosine's RMSF is shown in black and homocarnosine's in red. Naming of the atoms is the same as in Chart 4.1. Green colored atom names represent atoms of histidine ring, blue ones represent carboxylic group and orange one amide bond.

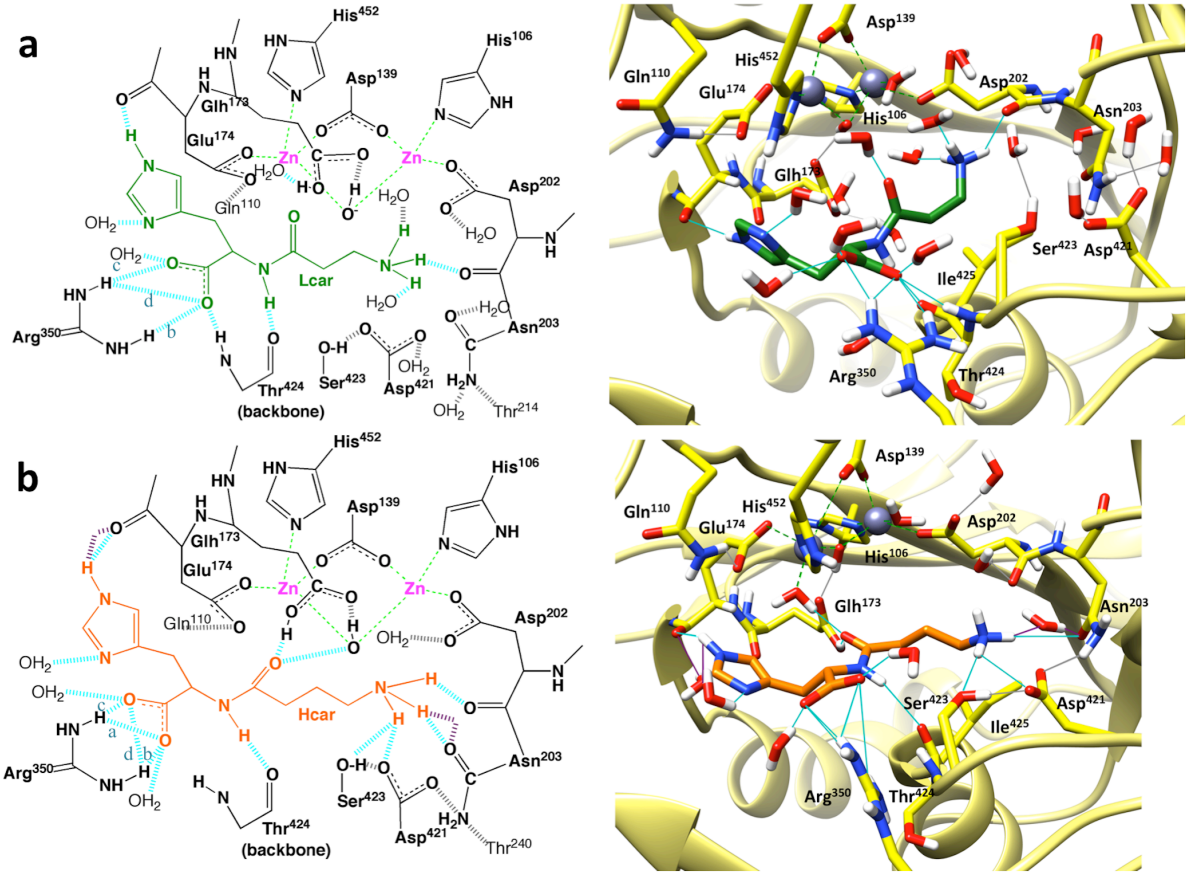


Figure A2.6: Structures of **CT-I**•monomeric hCN1**β** (a) and **HT-I**•monomeric hCN1**β** (b) Michaelis complexes obtained by one *ps*-long QM/MM simulations. On the left side of the figure are schematic 2D representations of the Michaelis complexes, while on the right side there are matching 3D figures (non-polar hydrogen atoms are not shown due to clarity). **CT-I** is shown in green and **HT-I** in orange. Hydrogen bonds between ligand and environment are shown as dashed turquoise lines in 2D schemes and as full turquoise lines in the 3D representations (for occupancy of hydrogen bonds between ligands and hCN1 see Tab. A2.16; hydrogen bond naming (a, b, c and d) between ligand's carboxylic group and side chain of Arg^{376} correspond to the one in Tab. A2.16), while those between hCN1 residues and water molecules are colored in grey. Water-mediated hydrogen bonds are shown as dashed purple lines in 2D schemes (due to the clarity, water molecules are omitted in these schemes) and as full purple lines in the 3D figures. Coordination bonds between Zn^{2+} ions and hCN1 active site residues are shown as dashed green lines on all figures. Zn^{2+} ions are shown as violet spheres in the 3D representation. Oxygen atoms are colored in red, nitrogen atoms in blue and polar hydrogen atoms in white. Non-polar hydrogen atoms are omitted due to clarity.

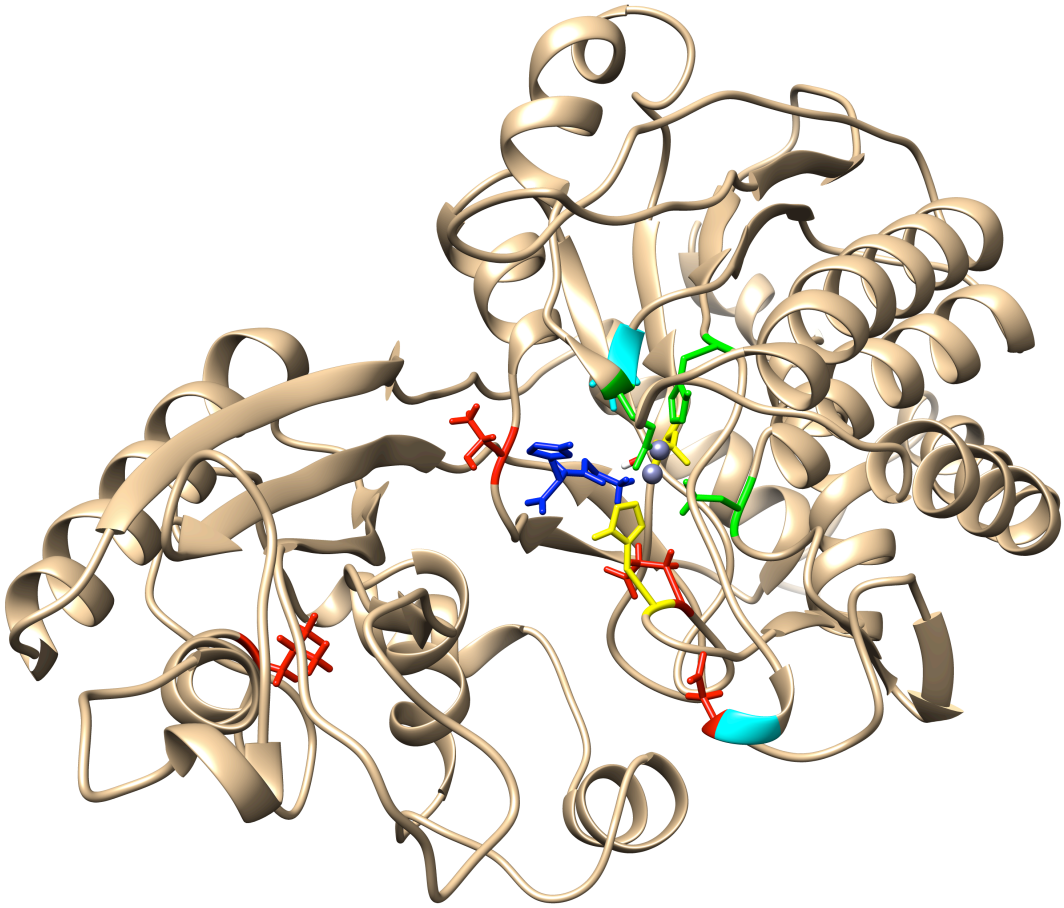


Figure A2.7: **CT-I** (colored in blue) docked in the monomeric hCN1**α** enzymatic cavity. Zn^{2+} ions are colored in pink. Residues coordinating Zn^{2+} according only to reference [250] are colored in cyan and the ones according only to our model are colored in yellow. Common residues in the two models coordinating Zn^{2+} are colored in green. Residues coordinating **CT-I** according to reference [250] are colored in red.

8 References

1. Abbasoğlu, L., et al., Beneficial effects of taurine and carnosine in experimental ischemia/reperfusion injury in testis. Pediatric Surgery International, 2012, **28**(11), 1125-1131.

2. Abe, H., Distribution of free L-histidine and related dipeptides in the muscle of fresh-water fishes. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry, 1983, **76**(1), 35-39.

3. Abe, H., et al., Role of histidine-related compounds to intracellular buffering in fish skeletal muscle. American Journal of Physiology - Regulatory, Integrative and Comparative Physiology, 1985, **249**(4), R449-R454.

4. Abou-Sleiman, P. M., Muqit, M. M. K., and Wood, N. W., Expanding insights of mitochondrial dysfunction in Parkinson's disease. Nat Rev Neurosci, 2006, **7**(3), 207-219.

5. Agarwal, R., Burley, S. K., and Swaminathan, S., Structural analysis of a ternary complex of allantoate amidohydrolase from Escherichia coli reveals its mechanics. Journal of Molecular Biology, 2007, **368**(2), 450-463.

6. Akdoğan, E. D., Computer-aided drug discovery (CADD), 2012: Kadir Has University, Istanbul.

7. Allen, M. P. and Tildesley, D. J., Computer simulation of liquids, 1989, USA: Oxford University Press.

8. Anandakrishnan, R., Aguilar, B., and Onufriev, A. V., H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. Nucleic Acids Research, 2012, **40**(W1), W537-W541.

9. Andersen, H. C., Molecular dynamics simulations at constant pressure and/or temperature. The Journal of Chemical Physics, 1980, **72**(4), 2384-2393.

10. Aristoy, M. C. and Toldrá, F., Concentration of free amino acids and dipeptides in porcine skeletal muscles with different oxidative patterns. Meat Science, 1998, **50**(3), 327-332.

11. Aristoy, M. C. and Toldrá, F., Histidine dipeptides HPLC-based test for the detection of mammalian origin proteins in feeds for ruminants. Meat Science, 2004, **67**(2), 211-217.

12. Ashcroft, N. W. and Mermin, N. D., Solid state physics 1976, Philadelphia: Saunders College Publishing.

13. Attanasio, F., et al., Protective effects of L- and D-carnosine on alpha-crystallin amyloid fibril formation: implications for cataract disease. Biochemistry, 2009, **48**(27), 6522-31.

14. Attanasio, F., et al., Carnosine Inhibits Aβ42 Aggregation by Perturbing the H-Bond Network in and around the Central Hydrophobic Cluster. ChemBioChem, 2013, **14**(5), 583-592.

15. Babizhayev, M. A., Antioxidant activity of L-carnosine, a natural histidine-containing dipeptide in crystalline lens. Biochim Biophys Acta, 1989, **1004**(3), 363-71.

16. Baran, E. J., Metal complexes of carnosine. Biochemistry (Moscow), 2000, **65**, 789-797.

17. Basun, H., et al., Metals and trace elements in plasma and cerebrospinal fluid in normal aging and Alzheimer's disease. Journal of Neural Transmission / Parkinson's Disease and Dementia Section, 1991, **3**(4), 231-258.

18. Bauchart, C., et al., Carnosine concentration of ingested meat affects carnosine net release into the portal vein of minipigs. The Journal of Nutrition, 2007, **137**(3), 589-593.

19. Bayly, C. I., et al., A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. The Journal of Physical Chemistry, 1993, **97**(40), 10269-10280.

20. Becke, A. D., Density-functional exchange-energy approximation with correct asymptotic behavior. Physical Review A, 1988, **38**(6), 3098-3100.

21. Becke, A. D., A new mixing of Hartree–Fock and local density−functional theories. Journal of Chemical Physics, 1993, **98**(2), 1372-1377.

22. Bellia, F., et al., New glycosidic derivatives of histidine-containing dipeptides with antioxidant properties and resistant to carnosinase activity. European Journal of Medicinal Chemistry, 2008, **43**(2), 373-380.

23. Bellia, F., et al., Neuroprotective features of carnosine in oxidative driven diseases. Molecular Aspects of Medicine, 2011, **32**(4–6), 258-266.

24. Bellia, F., Vecchio, G., and Rizzarelli, E., Carnosine derivatives: new multifunctional drug-like molecules. Amino Acids, 2012, **43**(1), 153-163.

25. Bellia, F., Vecchio, G., and Rizzarelli, E., Carnosinases, their substrates and diseases. Molecules, 2014, **19**(2), 2299.

26. Berendsen, H. J. C., et al., Molecular dynamics with coupling to an external bath. The Journal of Chemical Physics, 1984, **81**(8), 3684-3690.

27. Bertinaria, M., et al., Synthesis, physicochemical characterization, and biological activities of new carnosine derivatives stable in human serum as potential neuroprotective agents. Journal of Medicinal Chemistry, 2011, **54**(2), 611-621.

28. Blennow, K., de Leon, M. J., and Zetterberg, H., Alzheimer's disease. The Lancet, 2006, **368**(9533), 387-403.

29. Boldyrev, A. A., Carnosine: new concept for the function of an old molecule. Biochemistry (Moscow), 2012, **77**(4), 313-326.

30. Boldyrev, A. A., Aldini, G., and Derave, W., Physiology and pathophysiology of carnosine. Physiol Rev, 2013, **93**(4), 1803-45.

31. Boldyrev, A. A., et al., A comparative study of synthetic carnosine analogs as antioxidants. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry, 1989, **94**(2), 237-240.

32. Boldyrev, A. A., et al., The antioxidative properties of carnosine, a natural histidine containing dipeptide. Biochemistry international, 1987, **15**(6), 1105-1113.

33. Boldyrev, A. A., et al., Antioxidative properties of histidine-containing dipeptides from skeletal muscles of vertebrates. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry, 1988, **89**(2), 245-250.

34. Boldyrev, A. A., Gallant, S. C., and Sukhich, G. T., Carnosine, the protective, anti-aging peptide. Bioscience Reports, 1999, **19**(6), 581-587.

35. Botka, C. W., et al., Human proton/oligopeptide transporter (POT) genes: identification of putative human genes using bioinformatics. AAPS PharmSci, 2000, **2**(2), 76-97.

36. Branham, M. L., et al., Preparation, spectrochemical, and computational analysis of L-carnosine (2-[(3-aminopropanoyl)amino]-3-(1H-imidazol-5-yl)propanoic acid) and its ruthenium (II) coordination complexes in aqueous solution. Molecules, 2011, **16**(12), 10269-10291.

37. Breneman, C. M. and Wiberg, K. B., Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in

formamide conformational analysis. Journal of Computational Chemistry, 1990, **11**(3), 361-373.

38. Brown, C. E., et al., Copper(II) dimers in solution: evidence for motional averaging of coupling tensors without chemical dissociation. Journal of the Chemical Society, Dalton Transactions, 1982(12), 2371-2377.

39. Bush, A., et al., Rapid induction of Alzheimer A beta amyloid formation by zinc. Science, 1994, **265**(5177), 1464-1467.

40. Car, R. and Parrinello, M., Unified approach for molecular dynamics and density-functional theory. Physical Review Letters, 1985, **55**(22), 2471-2474.

41. Carr, W. J., Energy, specific heat, and magnetic properties of the low-density electron gas. Physical Review, 1961, **122**(5), 1437-1446.

42. Carr, W. J. and Maradudin, A. A., Ground-state energy of a high-density electron gas. Physical Review, 1964, **133**(2A), A371-A374.

43. Cartwright, S. P., Bill, R. M., and Hipkiss, A. R., L-carnosine affects the growth of Saccharomyces cerevisiae in a metabolism-dependent manner. PLoS ONE, 2012, **7**(9), e45006.

44. Ceperley, D. M. and Alder, B. J., Ground state of the electron gas by a stochastic method. Physical Review Letters, 1980, **45**(7), 566-569.

45. Chan, K. M., Decker, E. A., and Feustman, C., Endogenous skeletal muscle antioxidants. Critical Reviews in Food Science and Nutrition, 1994, **34**(4), 403-426.

46. Cheeseman, J. R., et al., A comparison of models for calculating nuclear magnetic resonance shielding tensors. The Journal of Chemical Physics, 1996, **104**(14), 5497-5509.

47. Chevrier, B., et al., Crystal structure of Aeromonas proteolytica aminopeptidase: a prototypical member of the co-catalytic zinc enzyme family. Structure, 1994, **2**(4), 283-291.

48. Christman, A. A., Factors affecting anserine and carnosine levels in skeletal muscles of various animals. International Journal of Biochemistry, 1976, **7**(9), 519-527.

49. Clements, A., et al., Aggregation and metal-binding properties of mutant forms of the amyloid Aβ peptide of Alzheimer's disease. Journal of Neurochemistry, 1996, **66**(2), 740-747.

50. Cohen, M., et al., Serum carnosinase deficiency: a non-disabling pehotype? Journal of Intellectual Disability Research, 1985, **29**(4), 383-389.

51. Cornell, W. D., et al., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. Journal of the American Chemical Society, 1995, **117**(19), 5179-5197.

52. Corona, C., et al., Effects of dietary supplementation of carnosine on mitochondrial dysfunction, amyloid pathology, and cognitive deficits in 3xTg-AD mice. PLoS ONE, 2011, **6**(3), e17971.

53. Crush, K. G., Carnosine and related substances in animal tissues. Comparative Biochemistry and Physiology, 1970, **34**(1), 3-30.

54. Dal Peraro, M., et al., Investigating biological systems using first principles Car–Parrinello molecular dynamics simulations. Current Opinion in Structural Biology, 2007, **17**(2), 149-156.

55. Danscher, G., et al., Increased amount of zinc in the hippocampus and amygdala of Alzheimer's diseased brains: a proton-induced X-ray emission spectroscopic analysis of cryostat sections from autopsy material. Journal of Neuroscience Methods, 1997, **76**(1), 53-59.

56. Darden, T., York, D., and Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. Journal of Chemical Physics, 1993, **98**(12), 10089-10092.

57. Daura, X., et al., Peptide folding: when simulation meets experiment. Angewandte Chemie International Edition, 1999, **38**(1-2), 236-240.

58. Davies, K. M., et al., Copper pathology in vulnerable brain regions in Parkinson's disease. Neurobiology of Aging, 2014, **35**(4), 858-866.

59. Davis, P. J. and Rabinowitz, P., Methods of numerical integration. 2nd ed., 1983, London: Academic Press Inc.

60. De Vivo, M., Bridging quantum mechanics and structure-based design. Frontiers in Bioscience (Landmark Edition), 2011, **16**, 1619-1633.

61. Decker, E. A., Livisay, S. A., and Zhou, S., A re-evaluation of the antioxidant activity of purified carnosine. Biochemistry (Moscow), 2000, **65**(7), 766-770.

62. Derave, W., et al., Muscle carnosine metabolism and β-alanine supplementation in relation to exercise and training. Sports Medicine, 2010, **40**(3), 247-263.

63. Dobbie, H. and Kermack, W. O., Complex-formation between polypeptides and metals. 2. The reaction between cupric ions and some dipeptides. Biochemical Journal, 1955, **59**(2), 246-257.

64. Dobrotvorskaya, I. S., et al., Characteristics of oxidative stress in experimental rat brain ischemia aggravated by homocysteic acid. Neurochemical Journal, 2011, **5**(1), 42-46.

65. Drozak, J., et al., Molecular identification of carnosine synthase as ATP-grasp domain-containing protein 1 (ATPGD1). J Biol Chem, 2010, **285**(13), 9346-56.

66. Dunnett, M., et al., Carnosine, anserine and taurine contents in individual fibres from the middle gluteal muscle of the camel. Research in Veterinary Science, 1997, **62**(3), 213-216.

67. Eldridge, M. D., et al., Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. Journal of Computer-Aided Molecular Design, 1997, **11**(5), 425-445.

68. Everaert, I., et al., Gene expression of carnosine-related enzymes and transporters in skeletal muscle. European Journal of Applied Physiology, 2013, **113**(5), 1169-1179.

69. Everaert, I., et al., Low plasma carnosinase activity promotes carnosinemia after carnosine ingestion in humans. American Journal of Physiology - Renal Physiology, 2012, **302**(12), F1537-F1544.

70. Ewald, P. P., Die Berechnung optischer und elektrostatischer Gitterpotentiale. Annalen der Physik, 1921, **369**(3), 253-287.

71. Farkas, E., Sovago, I., and Gergely, A., Studies on transition-metal-peptide complexes. Part 8. Parent and mixed-ligand complexes of histidine-containing dipeptides. Journal of the Chemical Society, Dalton Transactions, 1983(8), 1545-1551.

72. Fiser, A. and Šali, A., Modeller: generation and refinement of homology-based protein structure models, in Methods in Enzymology, Charles W. Carter, Jr. and Robert, M. S., Editors. 2003, Academic Press, 461-491.

73. Fitzpatrick, D. W. and Fisher, H., Carnosine, histidine, and wound healing. Surgery, 1982, **91**, 56-60.

74. Fonteh, N. A., et al., Free amino acid and dipeptide changes in the body fluids from Alzheimer's disease subjects. Amino Acids, 2007, **32**(2), 213-224.

75. Forli, S. and Olson, A. J., A force field with discrete displaceable waters and desolvation entropy for hydrated ligand docking. Journal of Medicinal Chemistry, 2012, **55**(2), 623-638.

76. Friedrich, J. O. and Wasylishen, R. E., A 1H and 13C nuclear magnetic resonance study of carnosine. Canadian Journal of Chemistry, 1986, **64**, 2132-2138.

77. Frisch, M. J., et al., Gaussian 09, Revision A.02, 2009: Wallingford CT.

78. Gaggelli, E. and Valensin, G., 1H and 13C NMR relaxation investigation of the calcium complex of β-alanyl-L-histidine (carnosine) in aqueous solution. Journal of the Chemical Society, Perkin Transaction 2, 1990(3), 401-406.

79. Gallant, S., Semyonova, M., and Yuneva, M., Carnosine as a potential anti-senescence drug. Biochemistry (Moscow), 2000, **65**, 866-868.

80. Gardner, M. L., et al., Intestinal absorption of the intact peptide carnosine in man, and comparison with intestinal permeability to lactulose. The Journal of Physiology, 1991, **439**(1), 411-422.

81. Gilboa, R., et al., Interactions of Streptomyces griseus aminopeptidase with a methionine product analogue: a structural study at 1.53 A resolution. Acta Crystallogr D Biol Crystallogr, 2000, **56**(Pt 5), 551-8.

82. Gilson, M. K., Given, J. A., and Head, M. S., A new class of models for computing receptor-ligand binding affinities. Cell Chemical Biology, 1997, **4**(2), 87-92.

83. Goldstein, H., Poole, C. P., and Safko, J. L., Classical mechanics, 2002, USA: Addison-Wesley.

84. Gordon, J. C., et al., H++: a server for estimating pKas and adding missing hydrogens to macromolecules. Nucleic Acids Research, 2005, **33**(suppl 2), W368-W371.

85. Govindaraju, V., Young, K., and Maudsley, A. A., Proton NMR chemical shifts and coupling constants for brain metabolites. NMR in Biomedicine, 2000, **13**(3), 129-153.

86. Gulewitsch, W. and Amiradžibi, S., Ueber das Carnosin, eine neue organische Base des Fleischextractes. Berichte der deutschen chemischen Gesellschaft, 1900, **33**(2), 1902-1903.

87. Haass, C., et al., Amyloid [beta]-peptide is produced by cultured cells during normal metabolism. Nature, 1992, **359**(6393), 322-325.

88. Hakansson, K. and Miller, C. G., Structure of peptidase T from Salmonella typhimurium. Eur J Biochem, 2002, **269**(2), 443-50.

89. Hanson, T. H. and Smith, E. L., Carnosinase: an enzyme of swine kidney. Journal of Biological Chemistry, 1949, **179**, 789-801.

90. Harding, J. W. and O'Fallon, J. V., The subcellular distribution of carnosine, carnosine synthetase, and carnosinase in mouse olfactory tissues. Brain Research, 1979, **173**(1), 99-108.

91. Harris, R. C., et al., Muscle buffering capacity and dipeptide content in the Thoroughbred horse, Greyhound dog and man. Comparative Biochemistry and Physiology Part A: Physiology, 1990, **97**(2), 249-251.

92. Heine, V., The pseudopotential concept, in Solid State Physics, Henry Ehrenreich, F. S. and David, T., Editors. 1970, Academic Press, 1-36.

93. Herculano, B., et al., β-alanyl-L-histidine rescues cognitive deficits caused by feeding a high fat diet in a transgenic mouse model of Alzheimer's disease. Journal of Alzheimer's Disease, 2013, **33**(4), 983-997.

94. Hess, B., Convergence of sampling in protein simulations. Physical Review E, 2002, **65**(3), 031910.

95. Hess, B., et al., LINCS: a linear constraint solver for molecular simulations. Journal of Computational Chemistry, 1997, **18**(12), 1463-1472.

96. Hess, B., et al., GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. Journal of Chemical Theory and Computation, 2008, **4**(3), 435-447.

97. Hipkiss, A. R., Glycation, ageing and carnosine: are carnivorous diets beneficial? Mechanisms of Ageing and Development, 2005, **126**(10), 1034-1039.

98. Hipkiss, A. R., Chapter 3 carnosine and its possible roles in nutrition and health, in Advances in Food and Nutrition Research, Steve, L. T., Editor 2009, Academic Press, 87-154.

99. Hockney, R. W., The potential calculation and some applications. Methods in Computational Physics, 1974, **9**, 135-211.

100. Hohenberg, P. and Kohn, W., Inhomogeneous electron gas. Physical Review, 1964, **136**(3B), B864-B871.

101. Holz, R. C., Bzymek, K. P., and Swierczek, S. I., Co-catalytic metallopeptidases as pharmaceutical targets. Current Opinion in Chemical Biology, 2003, **7**(2), 197-206.

102. Hoover, W. G., Canonical dynamics: equilibrium phase-space distributions. Physical Review A, 1985, **31**(3), 1695-1697.

103. Horinishi, H., Grillo, M., and Margolis, F. L., Purification and characterization of carnosine synthetase from mouse olfactory bulbs. Journal of Neurochemistry, 1978, **31**(4), 909-919.

104. Hornak, V., et al., Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins: Structure, Function, and Bioinformatics, 2006, **65**(3), 712-725.

105. Horning, M. S., Blakemore, L. J., and Trombley, P. Q., Endogenous mechanisms of neuroprotection: role of zinc, copper, and carnosine. Brain Research, 2000, **852**(1), 56-61.

106. House, E., et al., Aluminium, iron and copper in human brain tissues donated to the medical research council's cognitive function and ageing study. Metallomics, 2012, **4**(1), 56-65.

107. Hung, L. W., et al., Amyloid-β Peptide (Aβ) neurotoxicity is modulated by the rate of peptide aggregation: Aβ dimers and trimers correlate with neurotoxicity. The Journal of Neuroscience, 2008, **28**(46), 11950-11958.

108. Ihm, J., Total energy calculations in solid state physics. Reports on Progress in Physics, 1988, **51**(1), 105.

109. Imanari, M., et al., Accurate analysis of taurine, anserine, carnosine and free amino acids in a cattle muscle biopsy sample. Animal Science Journal, 2010, **81**(3), 369-376.

110. Jackson, M. C., Kucera, C. M., and Lenney, J. F., Purification and properties of human serum carnosinase. Clinica Chimica Acta, 1991, **196**(2), 193-205.

111. Jellinger, K. A., Chapter one - The relevance of metals in the pathophysiology of neurodegeneration, pathological considerations, in International Review of Neurobiology, Kailash, P. B. and Susanne, A. S., Editors. 2013, Academic Press, 1-47.

112. Jiménez-Jiménez, J. F., et al., Cerebrospinal fluid levels of transition metals in patients with Parkinson's disease. Journal of Neural Transmission, 1998, **105**(4), 497-505.

113. Jones, G., Smith, M., and Harris, R., Imidazole dipeptide content of dietary sources commonly consumed within the British diet. Proceedings of the Nutrition Society, 2011, **70**(OCE6), E363 (1 pages).

114. Jones, G., Willett, P., and Glen, R., A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. Journal of Computer-Aided Molecular Design, 1995, **9**(6), 532-549.

115. Jones, G., et al., Development and validation of a genetic algorithm for flexible docking. Journal of Molecular Biology, 1997, **267**(3), 727-748.

116. Jorgensen, W. L., et al., Comparison of simple potential functions for simulating liquid water. Journal of Chemical Physics, 1983, **79**(2), 926-935.

117. Jozic, D., et al., Crystal structure of the dinuclear zinc aminopeptidase PepV from Lactobacillus delbrueckii unravels its preference for dipeptides. Structure, 2002, **10**(8), 1097-106.

118. Jung Hoon, K. and Kyung Sik, K., Enhanced oligomerization of the a-synuclein mutant by the Cu,Zn-superoxide dismutase and hydrogen peroxide system. Mol. Cells, 2003, **15**(1), 87-93.

119. Junge, W. and McLaughlin, S., The role of fixed and mobile buffers in the kinetics of proton movement. Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1987, **890**(1), 1-5.

120. Just Kukurová, I., et al., Improved spectral resolution and high reliability of in vivo (1)H MRS at 7 T allow the characterization of the effect of acute exercise on carnosine in skeletal muscle. NMR in Biomedicine, 2016, **29**(1), 24-32.

121. Kalyankar, G. D. and Meister, A., Enzymatic synthesis of carnosine and related β-alanyl and γ-aminobutyryl peptides. Journal of Biological Chemistry, 1959, **234**(12), 3210-3218.

122. Kamal, M. A., et al., Influence of genetic knockout of Pept2 on the in vivo disposition of endogenous and exogenous carnosine in wild-type and Pept2 null mice. American Journal of Physiology - Regulatory, Integrative and Comparative Physiology, 2009, **296**(4), R986-R991.

123. Kang, J. H., et al., Carnosine and related dipeptides protect human ceruloplasmin against peroxyl radical-mediated modification. Molecules and Cells, 2002, **13**(3), 498-502.

124. Kayed, R., et al., Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. Science, 2003, **300**(5618), 486-489.

125. Kepp, K. P., Bioinorganic chemistry of Alzheimer's disease. Chemical Reviews, 2012, **112**(10), 5193-5239.

126. Kim, K. S., et al., The ceruloplasmin and hydrogen peroxide system induces α-synuclein aggregation in vitro. Biochimie, 2002, **84**(7), 625-631.

127. Kleinman, L. and Bylander, D., Efficacious form for model pseudopotentials. Physical Review Letters, 1982, **48**(20), 1425-1428.

128. Klusak, V., et al., Reaction mechanism of glutamate carboxypeptidase II revealed by mutagenesis, X-ray crystallography, and computational methods. Biochemistry, 2009, **48**(19), 4126-38.

129. Kohen, R., et al., Antioxidant activity of carnosine, homocarnosine, and anserine present in muscle and brain. Proceedings of the National Academy of Sciences, 1988, **85**(9), 3175-3179.

130. Kohn, W. and Sham, L. J., Self-consistent equations including exchange and correlation effects. Physical Review, 1965, **140**(4A), A1133-A1138.

131. Kontush, A., Amyloid-β: an antioxidant that becomes a pro-oxidant and critically contributes to Alzheimer's disease. Free Radical Biology and Medicine, 2001, **31**(9), 1120-1131.

132. Korb, O., Stützle, T., and Exner, T. E., Empirical scoring functions for advanced protein–ligand docking with PLANTS. Journal of Chemical Information and Modeling, 2009, **49**(1), 84-96.

133. Kozlowski, H., et al., Copper, iron, and zinc ions homeostasis and their role in neurodegenerative disorders (metal uptake, transport, distribution and regulation). Coordination Chemistry Reviews, 2009, **253**(21–22), 2665-2685.

134. Laio, A., VandeVondele, J., and Rothlisberger, U., A Hamiltonian electrostatic coupling scheme for hybrid Car–Parrinello molecular dynamics simulations. Journal of Chemical Physics, 2002, **116**(16), 6941-6947.

135. Laio, A., VandeVondele, J., and Röthlisberger, U., D-RESP: dynamically generated electrostatic potential derived charges from quantum mechanics/molecular mechanics simulations. Journal of Physical Chemistry B, 2002, **106**, 7300-7307.

136. Langston, J. W., Parkinson's disease: current and future challenges. NeuroToxicology, 2002, **23**(4–5), 443-450.

137. Laskowski, R. A., et al., PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography, 1993, **26**(2), 283-291.

138. Leach, A. R. and Kuntz, I. D., Conformational analysis of flexible ligands in macromolecular receptor sites. Journal of Computational Chemistry, 1992, **13**(6), 730-748.

139. Lee, C., Yang, W., and Parr, R. G., Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Physical Review B, 1988, **37**(2), 785-789.

140. Lee, Y.-t., et al., Histidine and carnosine delay diabetic deterioration in mice and protect human low density lipoprotein against oxidation and glycation. European Journal of Pharmacology, 2005, **513**(1–2), 145-150.

141. Lelie, H. L., et al., Copper and zinc metallation status of copper-zinc superoxide dismutase from amyotrophic lateral sclerosis transgenic mice. Journal of Biological Chemistry, 2011, **286**(4), 2795-2806.

142. Lenney, J. F., et al., Human serum carnosinase: characterization, distinction from cellular carnosinase, and activation by cadmium. Clinica Chimica Acta, 1982, **123**(3), 221-231.

143. Lenney, J. F., et al., Homocarnosinase: a hog kidney dipeptidase with a broader specificity than carnosinase. Archives of Biochemistry and Biophysics, 1977, **184**(1), 257-266.

144. Lenney, J. F., et al., Homocarnosinosis: lack of serum carnosinase is the defect probably responsible for elevated brain and CSF homocarnosine. Clinica Chimica Acta, 1983, **132**, 157-165.

145. Lenney, J. F., et al., Characterization of human tissue carnosinase. Biochemical Journal, 1985, **228**(3), 653-660.

146. Lenz, G. R. and Martell, A. E., Metal Complexes of Carnosine. Biochemistry, 1964, **3**(6), 750-753.

147. Lesné, S., et al., A specific amyloid-[beta] protein assembly in the brain impairs memory. Nature, 2006, **440**(7082), 352-357.

148. Liao, R.-Z., Yu, J.-G., and Himo, F., Quantum chemical modeling of enzymatic reactions: the case of decarboxylation. Journal of Chemical Theory and Computation, 2011, **7**(5), 1494-1501.

149. Lindner, H. A., et al., Essential roles of zinc ligation and enzyme dimerization for catalysis in the aminoacylase-1/M20 family. Journal of Biological Chemistry, 2003, **278**(45), 44496-44504.

150. Lindorff-Larsen, K., et al., Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins: Structure, Function, and Bioinformatics, 2010, **78**(8), 1950-1958.

151. Liu, J. and Wang, R., Classification of current scoring functions. Journal of Chemical Information and Modeling, 2015, **55**(3), 475-482.

152. Lodola, A. and De Vivo, M., Chapter 11 - The increasing role of QM/MM in drug discovery, in Advances in Protein Chemistry and Structural Biology, Christo, C. and Tatyana, K.-C., Editors. 2012, Academic Press, 337-362.

153. Lovell, M. A., A potential role for alterations of zinc and zinc transport proteins in the progression of Alzheimer's disease. Journal of Alzheimer's disease : JAD, 2009, **16**(3), 471-483.

154. Lovell, M. A., et al., Copper, iron and zinc in Alzheimer's disease senile plaques. Journal of the Neurological Sciences, 1998, **158**(1), 47-52.

155. MacKerell, A. D., et al., All-atom empirical potential for molecular modeling and dynamics studies of proteins. The Journal of Physical Chemistry B, 1998, **102**(18), 3586-3616.

156. Mandel, S., et al., Mechanism of neuroprotective action of the anti-Parkinson drug rasagiline and its derivatives. Brain Research Reviews, 2005, **48**(2), 379-387.

157. Mannion, A. F., et al., Carnosine and anserine concentrations in the quadriceps femoris muscle of healthy humans. European Journal of Applied Physiology and Occupational Physiology, 1992, **64**(1), 47-50.

158. Margolis, F. L., et al., Monoclonal antibodies to mammalian carnosine synthetase. Journal of Neurochemistry, 1987, **48**(2), 593-600.

159. Martyna, G. and Tuckerman, M., A reciprocal space based method for treating long range interactions in ab initio and force-field-based calculations in clusters. Journal of Chemical Physics, 1999, **110**(6), 2810-2821.

160. Marx, D. and Hutter, J., Ab initio molecular dynamics: theory and implementation, in Modern methods and algorithms of quantum chemistry, Grotendorst, J., Editor 2000, John von Neumann Institute for Computing: Jülich, 301-449.

161. Mattson, M. P., Pathways towards and away from Alzheimer's disease. Nature, 2004, **430**(7000), 631-639.

162. McFarland, G. A. and Holliday, R., Further evidence for the rejuvenating effects of the dipeptide L-carnosine on cultured human diploid fibroblasts. Experimental Gerontology, 1999, **34**(1), 35-45.

163. McKhann, G. M., et al., The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 2011, **7**(3), 263-269.

164. Meanwell, N. A., Synopsis of some recent tactical application of bioisosteres in drug design. Journal of Medicinal Chemistry, 2011, **54**(8), 2529-2591.

165. Mehler, E. L. and Solmajer, T., Electrostatic effects in proteins: comparison of dielectric and charge models. Protein Engineering, 1991, **4**(8), 903-910.

166. Metropolis, N., et al., Equation of state calculations by fast computing machines. The Journal of Chemical Physics, 1953, **21**(6), 1087-1092.

167. Miertuš, S., Scrocco, E., and Tomasi, J., Electrostatic interaction of a solute with a continuum. A direct utilizaion of ab initio molecular potentials for the prevision of solvent effects. Chemical Physics, 1981, **55**(1), 117-129.

168. Mineo, P., et al., Electrospray mass spectrometric studies of L-carnosine (β-alanyl-L-histidine) complexes with copper(II) or zinc ions in aqueous solution. Rapid Communications in Mass Spectrometry, 2002, **16**(7), 722-729.

169. Mora, L., Sentandreu, M. Á., and Toldrá, F., Contents of creatine, creatinine and carnosine in porcine muscles of different metabolic types. Meat Science, 2008, **79**(4), 709-715.

170. Morris, G. M., et al., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. Journal of Computational Chemistry, 1998, **19**(14), 1639-1662.

171. Morris, G. M., et al., AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. Journal of computational chemistry, 2009, **30**(16), 2785-2791.

172. Murphey, W. H., Patchen, L., and Lindmark, D. G., Carnosinase: a fluorometric assay and demonstration of two electrophoretic forms in human tissue extracts. Clinica Chimica Acta, 1972, **42**(2), 309-314.

173. Murray, C. W., Auton, T. R., and Eldridge, M. D., Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand-receptor binding affinities and the use of Bayesian regression to improve the quality of the model. Journal of Computer-Aided Molecular Design, 1998, **12**(5), 503-519.

174. Myers, J., et al., A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. Proteins: Structure, Function, and Bioinformatics, 2006, **63**(4), 928-938.

175. Myhre, O., et al., Metal dyshomeostasis and inflammation in Alzheimer's and Parkinson's Diseases: possible impact of environmental exposures. Oxidative Medicine and Cellular Longevity, 2013, **2013**, 19.

176. Nagai, K., et al., Action of carnosine and beta-alanine on wound healing. Surgery, 1986, **100**, 815-821.

177. Ng, R. H., et al., Metabolism of carnosine and homocarnosine in subcellular fractions and neuronal and glial cell-enriched fractions of rabbit brain. Journal of Neurochemistry, 1977, **28**(2), 449-452.

178. Nosé, S., A molecular dynamics method for simulations in the canonical ensemble. Molecular Physics, 1984, **52**(2), 255-268.

179. Nosé, S., A unified formulation of the constant temperature molecular dynamics methods. The Journal of Chemical Physics, 1984, **81**(1), 511-519.

180. Nunomura, A., et al., RNA oxidation in Alzheimer disease and related neurodegenerative disorders. Acta Neuropathologica, 2009, **118**(1), 151-166.

181. Olanow, C. W. and Tatton, W. G., Etiology and pathogenesis of Parkinson's disease. Annual Review of Neuroscience, 1999, **22**(1), 123-144.

182. Ono, K. and Yamada, M., Low-n oligomers as therapeutic targets of Alzheimer's disease. Journal of Neurochemistry, 2011, **117**(1), 19-28.

183. Otani, H., et al., Identification and characterization of a mouse dipeptidase that hydrolyzes L-carnosine. Journal of Biochemistry, 2005, **137**(2), 167-175.

184. Pan, J. W., et al., Intracellular pH in human skeletal muscle by 1H NMR. Proceedings of the National Academy of Sciences, 1988, **85**(21), 7836-7839.

185. Panayi, A. E., et al., Determination of cadmium and zinc in Alzheimer's brain tissue using Inductively Coupled Plasma Mass Spectrometry. Journal of the Neurological Sciences, 2002, **195**(1), 1-10.

186. Park, H., et al., A structure-based virtual screening approach toward the discovery of histone deacetylase inhibitors: identification of promising zinc-chelating groups. ChemMedChem, 2010, **5**(4), 591-597.

187. Parr, R. G. and Yang, W., Density-functional theory of atoms and molecules, 1989, US: Oxford University Press.

188. Parrinello, M. and Rahman, A., Crystal structure and pair potentials: a molecular-dynamics study. Physical Review Letters, 1980, **45**(14), 1196-1199.

189. Parrinello, M. and Rahman, A., Polymorphic transitions in single crystals: a new molecular dynamics method. Journal of Applied Physics, 1981, **52**(12), 7182-7190.

190. Pavlin, M., et al., The chemistry of neurodegeneration: kinetic data and their implications. Molecular Neurobiology, 2015, 1-16.

191. Pavlin, M., et al., Carnosine and homocarnosine degradation mechanisms by the human carnosinase enzyme CN1: insights from multiscale simulations. Biochemistry, 2016, **55**(19), 2772-2784.

192. Pegova, A., Abe, H., and Boldyrev, A., Hydrolysis of carnosine and related compounds by mammalian carnosinases. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 2000, **127**(4), 443-446.

193. Perdew, J., Burke, K., and Ernzerhof, M., Generalized gradient approximation made simple. Physical Review Letters, 1996, **77**(18), 3865-3868.

194. Perdew, J. P. and Wang, Y., Accurate and simple analytic representation of the electron-gas correlation energy. Physical Review B, 1992, **45**(23), 13244-13249.

195. Perry, T. L., Hansen, S., and Love, D. L., Serum-carnosinase deficiency in carnosinaemia. Lancet, 1968, **1**(7554), 1229-30.

196. Peters, M. B., et al., Structural survey of zinc-containing proteins and development of the Zinc AMBER Force Field (ZAFF). Journal of Chemical Theory and Computation, 2010, **6**(9), 2935-2947.

197. Peters, V., et al., Relevance of allosteric conformations and homocarnosine concentration on carnosinase activity. Amino Acids, 2010, **38**(5), 1607-1615.

198. Petroff, O. A. C., et al., Vigabatrin increases human brain homocarnosine and improves seizure control. Annals of Neurology, 1998, **44**(6), 948-952.

199. Pickett, W. E., Pseudopotential methods in condensed matter applications. Computer Physics Reports, 1989, **9**(3), 115-197.

200. Pimentel, C., et al., Oxidative stress in Alzheimer's and Parkinson's diseases: insights from the yeast Saccharomyces cerevisiae. Oxidative Medicine and Cellular Longevity, 2012, **2012**, 132146.

201. Poewe, W., Non-motor symptoms in Parkinson's disease. European Journal of Neurology, 2008, **15**, 14-20.

202. Premkumar, D. R. D., et al., Induction of heme oxygenase-1 mRNA and protein in neocortex and cerebral vessels in Alzheimer's disease. Journal of Neurochemistry, 1995, **65**(3), 1399-1402.

203. Preston, J. E., et al., Toxic effects of β-amyloid(25–35) on immortalised rat brain endothelial cell: protection by carnosine, homocarnosine and β-alanine. Neuroscience Letters, 1998, **242**(2), 105-108.

204. Prince, M., et al., World Alzheimer report 2015: the global impact of dementia: an analysis of prevalance, incidence, cost and trends, 2015, Alzheimer's Disease International (ADI): London.

205. Purchas, R. W., Triumf, E. C., and Egelandsdal, B., Quality characteristics and composition of the longissimus muscle in the short-loin from male and female farmed red deer in New Zealand. Meat Science, 2010, **86**(2), 505-510.

206. Quinn, P. J., Boldyrev, A. A., and Formazuyk, V. E., Carnosine: its properties, functions and potential therapeutic applications. Molecular Aspects of Medicine, 1992, **13**(5), 379-444.

207. Rajanikant, G. K., et al., Carnosine is neuroprotective against permanent focal cerebral ischemia in mice. Stroke, 2007, **38**(11), 3023-3031.

208. Rawlings, N. D., Barrett, A. J., and Finn, R., Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Research, 2016, **44**(D1), D343-D350.

209. Roberts, B. R., et al., The role of metallobiology and amyloid-β peptides in Alzheimer's disease. Journal of Neurochemistry, 2012, **120**, 149-166.

210. Rosenberg, A., The activation of carnosinase by divalent metal ions. Biochimica et Biophysica Acta, 1960, **45**, 297-316.

211. Rosenberg, A., Purification and some properties of carnosinase of swine kidney. Archives of Biochemistry and Biophysics, 1960, **88**(1), 83-93.

212. Röthlisberger, U. and Carloni, P., Simulations of enzymatics systems: perspectives from Car-Parrinello molecular dynamics simulations, in Theoretical and computational chemistry, Erikson, L. A., Politzer, P., and Maksić, Z., Editors. 2001, Elservier: Amsterdam, 215-251.

213. Rowsell, S., et al., Crystal structure of carboxypeptidase G2, a bacterial enzyme with applications in cancer therapy. Structure, 1997, **5**(3), 337-47.

214. Rubio-Aliaga, I., et al., Targeted disruption of the peptide transporter Pept2 gene in mice defines its physiological role in the kidney. Molecular and Cellular Biology, 2003, **23**(9), 3247-3252.

215. Rulon, L. L., et al., Serum zinc levels and Alzheimer's disease. Biological Trace Element Research, 2000, **75**(1), 79-85.

216. Sakae, K., et al., Effects of L-carnosine and its zinc complex (polaprezinc) on pressure ulcer healing. Nutrition in Clinical Practice, 2013, **28**(5), 609-616.

217. Šali, A. and Blundell, T. L., Comparative protein modelling by satisfaction of spatial restraints. Journal of Molecular Biology, 1993, **234**(3), 779-815.

218. Samudralwar, D. L., et al., Elemental imbalances in the olfactory pathway in Alzheimer's disease. Journal of the Neurological Sciences, 1995, **130**(2), 139-145.

219. Santos-Martins, D., et al., AutoDock4Zn: an improved AutoDock force field for small-molecule docking to zinc metalloproteins. Journal of Chemical Information and Modeling, 2014, **54**(8), 2371-2379.

220. Schaftenaar, G. and Noordik, J. H., Molden: a pre- and post-processing program for molecular and electronic structures. Journal of Computer-Aided Molecular Design, 2000, **14**(2), 123-134.

221. Schrag, M., et al., Effect of cerebral amyloid angiopathy on brain iron, copper, and zinc in Alzheimer's disease. Journal of Alzheimer's disease : JAD, 2011, **24**(1), 137-149.

222. Schrag, M., et al., Iron, zinc and copper in the Alzheimer's disease brain: a quantitative meta-analysis. Some insight on the influence of citation bias on scientific opinion. Progress in Neurobiology, 2011, **94**(3), 296-306.

223. Schröder, L., Schmitz, C. H. J., and Bachert, P., Carnosine as molecular probe for sensitive detection of Cu(II) ions using localized 1H NMR spectroscopy. Journal of Inorganic Biochemistry, 2008, **102**(2), 174-183.

224. Seidler, N. W., Yeargans, G. S., and Morgan, T. G., Carnosine disaggregates glycated alpha-crystallin: an in vitro study. Arch Biochem Biophys, 2004, **427**(1), 110-5.

225. Senn, H. M. and Thiel, W., QM/MM methods for biomolecular systems. Angewandte Chemie International Edition, 2009, **48**(7), 1198-1229.

226. Seubert, P., et al., Isolation and quantification of soluble Alzheimer's [beta]-peptide from biological fluids. Nature, 1992, **359**(6393), 325-327.

227. Sewell, D. A., et al., Estimation of the carnosine content of different fibre types in the middle gluteal muscle of the thoroughbred horse. The Journal of Physiology, 1992, **455**(1), 447-453.

228. Sharma, A. K., et al., Bifunctional compounds for controlling metal-mediated aggregation of the Aβ42 peptide. Journal of the American Chemical Society, 2012, **134**(15), 6625-6636.

229. Shen, M.-y. and Sali, A., Statistical potential for assessment and prediction of protein structures. Protein Science, 2006, **15**(11), 2507-2524.

230. Shoji, M., et al., Production of the Alzheimer amyloid beta protein by normal proteolytic processing. Science, 1992, **258**(5079), 126-129.

231. Shytle, R. D., et al., Oxidative stress of neural, hematopoietic, and stem cells: protection by natural compounds. Rejuvenation Research, 2007, **10**(2), 173-178.

232. Svozil, D., et al., Geometrical and electronic structure variability of the sugar–phosphate backbone in nucleic acids. The Journal of Physical Chemistry B, 2008, **112**(27), 8188-8197.

233. Teufel, M., et al., Sequence identification and characterization of human carnosinase and a closely related non-specific dipeptidase. Journal of Biological Chemistry, 2003, **278**(8), 6521-6531.

234. Teuscher, N. S., Keep, R. F., and Smith, D. E., PEPT2-mediated uptake of neuropeptides in rat choroid plexus. Pharmaceutical Research, 2001, **18**(6), 807-813.

235. Teuscher, N. S., et al., Carnosine uptake in rat choroid plexus primary cell cultures and choroid plexus whole tissue from PEPT2 null mice. Journal of Neurochemistry, 2004, **89**(2), 375-382.

236. Thomas, S., et al., Studies on adsorption of carnosine on silver nanoparticles by SERS. Chemical Physics Letters, 2010, **491**(1–3), 59-64.

237. Torreggiani, A., Bonora, S., and Fini, G., Raman and IR spectroscopic investigation of zinc(II)–carnosine complexes. Biopolymers, 2000, **57**(6), 352-364.

238. Torreggiani, A., Tamba, M., and Fini, G., Binding of copper(II) to carnosine: Raman and IR spectroscopic study. Biopolymers, 2000, **57**(3), 149-59.

239. Troullier, N. and Martins, J. L., Efficient pseudopotentials for plane-wave calculations. Physical Review B, 1991, **43**(3), 1993-2006.

240. Tsai, S.-J., et al., Antioxidative and anti-inflammatory protection from carnosine in the striatum of MPTP-treated mice. Journal of Agricultural and Food Chemistry, 2010, **58**(21), 11510-11516.

241. Tully, J. C., Molecular dynamics with electronic transitions. The Journal of Chemical Physics, 1990, **93**(2), 1061-1071.

242. Tully, J. C., Mixed quantum-classical dynamics: mean-field and surface-hoppin, in Classical and quantum dynamics in condensed phase simulations, Berne, B. J., Ciccotti, G., and Coker, D. F., Editors. 1998, World Scientific: Singapore.

243. Van Der Spoel, D., et al., GROMACS: fast, flexible, and free. Journal of Computational Chemistry, 2005, **26**(16), 1701-1718.

244. van Gunsteren, W. F., Biomolecular simulation: the GROMOS96 manual and user guide, 1996, Zürich: Hochschulverlag AG der ETH Zürich.

245. Verdonk, M. L., et al., Modeling water molecules in protein–ligand docking using GOLD. Journal of Medicinal Chemistry, 2005, **48**(20), 6504-6515.

246. Verdonk, M. L., et al., Improved protein–ligand docking using GOLD. Proteins: Structure, Function, and Bioinformatics, 2003, **52**(4), 609-623.

247. Verlet, L., Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. Physical Review, 1967, **159**(1), 98-103.

248. Vistoli, G., Carini, M., and Aldini, G., Transforming dietary peptides in promising lead compounds: the case of bioavailable carnosine analogs. Amino Acids, 2012, **43**(1), 111-126.

249. Vistoli, G., et al., Design, synthesis, and evaluation of carnosine derivatives as selective and efficient sequestering agents of cytotoxic reactive carbonyl species. ChemMedChem, 2009, **4**(6), 967-975.

250. Vistoli, G., et al., Homology modeling of human serum carnosinase, a potential dedicinal target, and MD simulations of its allosteric activation by citrate. Journal of Medicinal Chemistry, 2006, **49**(11), 3269-3277.

251. Vistoli, G., et al., Predicting the physicochemical profile of diastereoisomeric histidine-containing dipeptides by property space analysis. Chirality, 2012, **24**(7), 566-76.

252. von Lilienfeld, O. A., et al., Variational optimization of effective atom centered potentials for molecular properties. Journal of Chemical Physics, 2005, **122**(1), 14113.

253. Vosko, S. H., Wilk, L., and Nusair, M., Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. Canadian Journal of Physics, 1980, **58**(8), 1200-1211.

254. Wakabayashi, K., et al., The Lewy body in Parkinson's disease: molecules implicated in the formation and degradation of α-synuclein aggregates. Neuropathology, 2007, **27**(5), 494-506.

255. Wang, J., et al., Automatic atom type and bond type perception in molecular mechanical calculations. Journal of Molecular Graphics and Modelling, 2006, **25**(2), 247-260.

256. Wang, J., et al., Development and testing of a general amber force field. Journal of Computational Chemistry, 2004, **25**(9), 1157-1174.

257. Warburg, O., On the origin of cancer cells. Science, 1956, **123**(3191), 309-314.

258. Yamashita, T., et al., Cloning and functional expression of a brain peptide/histidine transporter. Journal of Biological Chemistry, 1997, **272**(15), 10205-10211.

259. Yan, H., et al., Effect of carnosine, aminoguanidine, and aspirin drops on the prevention of cataracts in diabetic rats. Mol Vis, 2008, **14**, 2282-91.

260. Yeum, K.-J., et al., Profiling histidine dipeptides in plasma and urine after ingesting beef, chicken or chicken broth in humans. Amino Acids, 2010, **38**(3), 847-858.

261. Yuneva, M. O., et al., Effect of carnosine on age-induced changes in senescence-accelerated mice. Journal of Anti-Aging Medicine, 1999, **2**(4), 337-342.

262. Zapp, J. A. and Wilson, D. W., Quantitative studies of carnosine and anserine in mammalian musce: II. The distribution of carnosine and anserine in various muscles of different species Journal of Biological Chemistry, 1938, **126**(1), 19-27.

263. Zatta, P., et al., Alzheimer's disease, metal ions and metal homeostatic therapy. Trends in Pharmacological Sciences, 2009, **30**(7), 346-355.

264. Zhang, R., Wang, D., and Wu, W., Conformations of carnosine in aqueous solutions by all-atom molecular dynamics simulations and 2D-NOSEY spectrum. Chinese Journal of Chemical Physics, 2013, **26**(1), 67-72.

265. Zhang, X., et al., Carnosine pretreatment protects against hypoxia–ischemia brain damage in the neonatal rat model. European Journal of Pharmacology, 2011, **667**(1–3), 202-207.

266. Zhou, S., et al., Identification of hydrazine in commercial preparations of carnosine and itsinfluence on carnosine's antioxidative properties. Analytical Biochemistry, 1998, **261**(1), 79-86.

267. Zoch, E. and Muller, H., Demonstration and determination of carnosinase activity of the human placenta. Enzymologia, 1971, **123**, 221-231.

268. Zou, X., Yaxiong, and Kuntz, I. D., Inclusion of solvation in ligand binding free energy calculations using the generalized-Born model. Journal of the American Chemical Society, 1999, **121**(35), 8033-8043.

List of figures

Figure 1.1: hCN1 homodimer structure shown in ribbons representation, where two monomers are shown in pink and grey, respectively. The active site residues are colored in yellow and Zn^{2+} ions in dark violet. Oxygen atoms are colored in red, nitrogen atoms in blue and bridging water molecule in dark blue. Coordination bonds are shown as dashed green lines. Non-polar hydrogen atoms are not shown due to clarity reasons. 3

Figure 2.1: Metabolic pathways of carnosine and its naturally occurring derivatives. The enzymatic reactions that are further discussed in the thesis are colored in green (synthesis) and in red (degradation). All compounds are shown in their non-ionized state. CN stands for carnosinase, CS for carnosine synthase, CMT for carnosine-*N*-methyltransferase and HDC for L-histidine decarboxylase. 5

Figure 3.1: Schematic representation of incremental construction. The figure is taken from the reference [6]. 48

Figure 3.2: Schematic representation of genetic algorithm. The figure is taken from the reference [6]. 48

Figure 3.3: Schematic representation of Monte Carlo docking algorithm. The figure is taken from the reference [6]..... 49

Chart 4.1: Tautomeric forms of carnosine (above) and homocarnosine (bellow). All tautomers are shown in their non-ionized form. 54

Figure 4.1: Probability distribution of R_g^2 in CT-I and CT-II (upper line) and HT-I and HT-II (lower line) in aqueous solution. 59

Figure 4.2: Representative structures of the largest populated clusters of carnosine tautomers **CT-I** and **CT-II** (see Chart 4.1). Each conformation is colored in its distinct color and it is shown from to different angles, as side view and as front view, with marked χ_1 and χ_2 dihedral angles. Only polar hydrogen atoms are shown due to clarity reasons. 60

Figure 5.1: (a) hCN1 homodimer structure shown in ribbons representation. **M-I** and **M-II** are shown in pink and turquoise, while the active site residues are colored in yellow and Zn^{2+} ions in dark violet. (b) 3D representation of the hCN1 active site with only residues coordinating Zn^{2+} shown. Zn^{2+} ions are dark violet, oxygen

atoms in red, nitrogen atoms in blue and bridging water molecule in dark blue. Coordination bonds are shown as dashed green lines. Hydrogen atoms are not shown due to clarity reasons. (c) 2D representation of the hCN1 active site. Coordination bonds are shown as dashed lines. 66

Figure 5.2: Superimposition of hCN1 active site (colored in yellow) with (a) available crystal structures of metallopeptidases from M20 family (allC colored in blue, PepV colored in violet, CPG2 colored in green and PepT colored in orange) and (b) AAP (colored in pink) and SGAP (colored in gray). Oxygen atoms are shown in red and nitrogen atoms in blue, while hydrogen atoms are not shown due to clarity reasons. Zinc ions are represented as big spheres and bridging water molecules as small sphere, both in the same color scheme as the corresponding enzyme. PepV and PepT are missing bridging water molecule because the crystal structure of the former has ligand present in the active site and in the later it was not detected due to the lower resolution of the crystallographic data. Oxygen atoms are shown in red and nitrogen atoms in blue, whereas hydrogen atoms are not shown due to clarity reasons. 67

Figure 5.3: Tautomers **α** and **β** of hCN1 active site. Proton that changes position is encircled in orange for tautomer **α** (protonated on Oε2) and in pink for **β** (protonated on Oε1). Coordination bonds between Zn^{2+} ions and hCN1 active site residues are shown as dashed green lines. Oxygen atoms are shown in red, nitrogen atoms in blue and polar hydrogen atoms in white.....72

Figure 5.4: Highest populated clusters of monomeric hCN1 complexes (**CT-I•**hCN1**α** (a), **CT-I•**hCN1**β** (b), **HT-I•**hCN1**α** (c) and **HT-I•**hCN1**β** (d)) obtained by docking simulations (on Glh^{173}**α** proton is marked with green and on Glh^{173}**β** with pink circle). On the left side of the figure are schematic 2D representations of the complexes, while on the right side there are matching 3D figures. **CT-I** is shown in green and **HT-I** in orange. Hydrogen bonds are shown as dashed turquoise lines in 2D schemes and as full turquoise lines in the 3D representations. Coordination bonds between Zn^{2+} ions and hCN1 active site residues are shown on the 2D representations as dashed green lines. Zn^{2+} ions are shown as violet spheres in the 3D representation. Oxygen atoms are shown in red, nitrogen atoms in blue and polar hydrogen atoms in white. 78

Figure 5.5: Highest populated clusters of monomeric hCN1**α** complexes obtained by docking simulations. a) binding poses of carnosine in all three protonation states (**CT-I** is shown in dark green, **CT-III** in light green and **CT-III** in light blue); b) binding poses of homocarnosine in all three protonation states (**HT-I** in shown in

orange, **HT-II** in tan and **HT-III** in light pink). Hydrogen bonds are shown as full turquoise lines. The coordination bonds between Zn^{2+} ions and hCN1 active site residues are shown as dashed green lines. The Zn^{2+} ions are shown as violet spheres. Oxygen atoms are shown in red, nitrogen atoms in blue and polar hydrogen atoms in white. 79

Figure 5.6: Dimeric hCN1 complexes in comparison with monomeric complexes obtained by docking simulations. a) binding poses of **CT-I** in hCN1 (**CT-I•**monomeric hCN1**α** is shown in dark green, LBP1 and LBP2 of **CT-I•**dimeric hCN1**α** are shown in light green and light blue, respectively); b) binding poses of **HT-I** in in hCN1 (**HT-I•**monomeric hCN1**α** is shown in orange, HBP1 and HBP2 of **HT-I•**dimeric hCN1**α** are shown in tan and light pink, respectively). Hydrogen bonds are shown as full turquoise lines. Coordination bonds between Zn^{2+} ions and hCN1 active site residues are shown as dashed green lines. Zn^{2+} ions are shown as violet spheres. Oxygen atoms are shown in red, nitrogen atoms in blue and polar hydrogen atoms in white. 80

Figure 5.7: (a) Superimposition of monomeric hCN1 (light blue) with **M-I** of dimeric hCN1 (light pink) shown in chain trace representation. **M-II** is shown in light grey in ribbons representation and the two Zn(II) ions (Zn1 and Zn2) are shown as purple spheres. Regions R**I** – R**V** are shown in blue for monomeric hCN1 and in pink for **M-I**. (b) Distribution of intramolecular hydrogen bonds in the monomeric and **M-I** of dimeric complexes with ligands during 40 ns classical MD. 83

Figure 5.8: Structures of **CT-I•**monomeric hCN1**α** (a), **HT-I•**monomeric hCN1**α** (b), **CT-I•**dimeric hCN1**α** (c) and **HT-I•**dimeric hCN1**α** (d) as obtained by our QM/MM simulations. On the left side of each panel is schematic 2D representation of the Michaelis complex (bulk water molecules are not shown due to clarity reasons), while on the right side there is matching 3D figure (non-polar hydrogen atoms and bulk water molecules are not shown due to clarity). **CT-I** is shown in green and **HT-I** in orange. Hydrogen bonds between ligand and environment are shown as dashed turquoise lines in 2D schemes and as full turquoise lines in the 3D representations (for occupancy of hydrogen bonds between ligands and hCN1 see Tab. 5.3; hydrogen bond naming (a, b, c and d) between ligand's carboxylic group and side chain of Arg^{350} correspond to the one in Tab. 5.3). Water-mediated hydrogen bonds are shown as dashed purple lines in 2D schemes (due to the clarity, water molecules are omitted in these schemes) and as full purple lines in the 3D figures. Coordination bonds between Zn^{2+} ions and hCN1 active site residues are shown as dashed green lines on all figures. Zn^{2+} ions are shown as violet spheres in the 3D representation. 87

Figure 6.1: Structure motives used in the ligand-based screening. R is simple alkyl chain with the length of n methylene groups. R' and R'' can be either hydrogen atoms or carbon atoms so that they form primary, secondary or tertiary amine. R''' can be a random functional group.....93

Figure 6.2: Structures of homocarnosine derivatives **D1**-**6** obtained from ligand-based screening.....95

Figure 6.3: Structures of Michaelis complexes (**C1**-**6**) comparison with the structure of Hcar•monomeric hCN1 Michaelis complex. Hcar is always shown in grey, hCN1 in yellow, oxygen atoms in red, nitrogen atoms in blue, polar hydrogens in white and zinc ions as dark violet spheres. Non-polar hydrogens are not shown due to clarity. Hydrogen bonds between Hcar and hCN1 are shown as grey lines, while the cyan lines represent hydrogen bonds between compounds **D1**-**6** and hCN1. Names of residues forming hydrophobic interactions with the ligand are written in cyan. Dashed green lines represent bonds between zinc ions and coordinating residues.....97

Scheme A1.1: Proposed general mechanism for the hydrolysis of a peptide catalyzed by a metallopeptidase with di-zinc co-catalytic active site based on the AAP as suggested by Holz *et al*. [101].....102

Figure A2.1: (a) RMSD, (b) distance between terminal NH_3^+ and COO^-, and (c) accessible surface area plotted as a function of simulated time for **CT-I** and **CT-II** tautomers in aqueous solution.....121

Figure A2.2: Radial distribution functions g(r) of **CT-I** and **CT-II** in water. (a) g(r) between the N-hydrogen atom in imidazole ring (HN_π and HN_τ) of carnosine and oxygen atom in water (OW) (black line); g(r) of amide hydrogen (HN) of carnosine and oxygen atom in water (OW) (red line). (b) g(r) of the C-hydrogen atom in imidazole ring (HC_2) of carnosine and oxygen atom of water (OW) (black line); g(r) of C_α-hydrogen of carnosine and oxygen atom in water (OW) (red line). (c) g(r) of both carboxylic oxygen atoms (OC) of carnosine and hydrogen atom in water (HW) (black and red lines); g(r) of amide oxygen atom (OA) of carnosine and hydrogen atom in water (HW) (green line). For the name of the atoms see Chart 4.1 in the Chapter 4.....122

Figure A2.3: RMSF values of pairs of equivalent atoms (data shown only for C, N and O atoms) in **CT-I** (red line) and **CT-II** (black line) tautomers of carnosine during one of the two MD simulations reported here.....123

Figure A2.4: RMSD of hCN1 relative to the structure of the enzyme used in the second step of molecular docking during 40 ns classical MD. a) monomeric hCN1 complexes (**CT-I•**hCN1**α** in black, **CT-I•**hCN1**β** in green, **HT-I•**hCN1**α** in red and **CT-I•**hCN1**β** in blue). b) dimeric hCN1 complexes (**CT-I•**hCN1**α** in black and **HT-I•**hCN1**α** in red). 123

Figure A2.5: RMSF values for heavy atoms of both ligands inside a) monomeric hCN1**α** from 2 ps long QM/MM simulations; b) dimeric hCN1**α** from 2 ps long QM/MM simulations and c) monomeric hCN1**β** from 1 ps long QM/MM simulations. Carnsosine's RMSF is shown in black and homocarnosine's in red. Naming of the atoms is the same as in Chart 4.1. Green colored atom names represent atoms of histidine ring, blue ones represent carboxylic group and orange one amide bond. 124

Figure A2.6: Structures of **CT-I•**monomeric hCN1**β** (a) and **HT-I•**monomeric hCN1**β** (b) Michaelis complexes obtained by one *ps*-long QM/MM simulations. On the left side of the figure are schematic 2D representations of the Michaelis complexes, while on the right side there are matching 3D figures (non-polar hydrogen atoms are not shown due to clarity). **CT-I** is shown in green and **HT-I** in orange. Hydrogen bonds between ligand and environment are shown as dashed turquoise lines in 2D schemes and as full turquoise lines in the 3D representations (for occupancy of hydrogen bonds between ligands and hCN1 see Tab. A2.16; hydrogen bond naming (a, b, c and d) between ligand's carboxylic group and side chain of Arg^{376} correspond to the one in Tab. A2.16), while those between hCN1 residues and water molecules are colored in grey. Water-mediated hydrogen bonds are shown as dashed purple lines in 2D schemes (due to the clarity, water molecules are omitted in these schemes) and as full purple lines in the 3D figures. Coordination bonds between Zn^{2+} ions and hCN1 active site residues are shown as dashed green lines on all figures. Zn^{2+} ions are shown as violet spheres in the 3D representation. Oxygen atoms are colored in red, nitrogen atoms in blue and polar hydrogen atoms in white. Non-polar hydrogen atoms are omitted due to clarity. 125

Figure A2.7: **CT-I** (colored in blue) docked in the monomeric hCN1**α** enzymatic cavity. Zn^{2+} ions are colored in pink. Residues coordinating Zn^{2+} according only to reference [250] are colored in cyan and the ones according only to our model are colored in yellow. Common residues in the two models coordinating Zn^{2+} are colored in green. Residues coordinating **CT-I** according to reference [250] are colored in red. 126

List of tables

Table 2.1: Differences between hCN1 and hCN2 as reported by Lenney *et al*. [142] and Teufel *et al*. [233]..... 13

Table 4.1: Cosine content of the first four principal components for **CT-I** and **CT-II** for the two independent simulations of each tautomer and for one simulation of **HT-I** and **HT-II**. 56

Table 4.2: Maxwell-Boltzmann statistics and relative free energies (in kcal/mol) for representative structures of highly populated clusters from MD simulations of carnosine's and homocarnosine's tautomers.[a] 61

Table 4.3: Calculated ^{13}C chemical shifts of both zwitterionic conformers of carnosine in comparison with experimental values.[a] For labeling of carbon atoms see Chart 4.1..... 62

Table 4.4: Calculated and experimentally observed ^{1}H chemical shifts of carnosine.[a] For labeling of hydrogen atoms see Chart 4.1. 63

Table 4.5: Calculated and experimentally observed ^{1}H chemical shifts of homocarnosine. For labeling of hydrogen atoms see Chart 4.1..... 64

Table 5.1: Hydrogen bond donors and acceptors inside the hCN1 active site from docking simulations. Asterisk indicates that the salt bridge is formed..... 81

Table 5.2: The average values of solvent accessible surface of monomeric and dimeric (**M-I**) hCN1 from classical MD simulations. 82

Table 5.3: Probability of hydrogen bonds between ligands and the hCN1**α** active site from our 2 ps long QM/MM simulations. ^{WM} indicates water mediated hydrogen bond..... 88

Table 6.1: Hydrogen bond donors and acceptors between hCN1 active site and homocarnosine and its derivatives obtained from docking simulations. Asterisk indicates that the salt bridge is formed. 98

Table A2.1: Partial atom charges of atoms in **CT-I** and **CT-II** together with legend of labeling of atoms in **CT-I** and **CT-II**. Atom types are gaff atom types [256]..... 105

Table A2.2: Bond lengths (in Å) for tautomers **CT-I** and **CT-II**. Atom types are gaff atom types [256] (see also Tab. A2.1).....106

Table A2.3: Angle values (in °) for tautomers **CT-I** and **CT-II**. Atom types are gaff atom types [256] (see also Tab. A2.1).....107

Table A2.4: Carnosine's hydrogen bond functionalities form hydrogen bonds with water molecules (similarly to what was found in reference [264]). The table shows the average and maximum water coordination numbers of hydrogen bond acceptors/donors for **CT-I** and **CT-II**. For both tautomers, all hydrogen bond acceptors and donors form hydrogen bond with water molecules with around 80% probability also consistent with their similar size of solvent accessible surface (see Fig. A2.1).....108

Table A2.5: Not-weighted calculated ^{13}C chemical shifts of **CT-Ia-d** and **CT-IIa-d** conformers of carnosine. Chemical shifts were scaled with respect to the chemical shift of ^{13}C TMS. 109

Table A2.6: Not-weighted calculated ^{1}H chemical shifts of **CT-Ia-d** and **CT-IIa-d** conformers of carnosine. Chemical shifts were scaled with respect to the chemical shift of ^{1}H TMS. 110

Table A2.7: Not-weighted calculated ^{1}H chemical shifts of **HT-Ia-d** and **HT-IIa-d** conformers of homocarnosine. Chemical shifts were scaled with respect to the chemical shift of ^{11}H DSS.....111

Table A2.8: Statistical data (in ppm) of calculated ^{1}H and ^{13}C chemical shifts (relative to the standard used). These include RMS error, mean error, mean absolute error, maximum error and standard deviation of absolute error. 112

Table A2.9: Protonation states of histidine residues and of Glu^{173}. The acidic amino acids other than Glu^{173} are negatively charged. 112

Table A2.10: Relative free energies (in kcal/mol), calculated as described in the Part *(ii)* of "Molecular docking – Monomer" section of Methods part in the main text, of model system of apo hCN1 active site and its modifications (based on the hCN1 crystal structure), where different residues (*i.e.* $Glh^{173}\boldsymbol{\alpha}$ and $Glh^{173}\boldsymbol{\beta}$, Glu^{174}, Asp^{202}) are protonated and hydroxyl anion is bridging two Zn^{2+} ions. Relative free energies are calculated with respect to the model system, where water is bridging two Zn^{2+} ions and other residues are deprotonated.....113

Table A2.11: Relative free energies (in kcal/mol) of model system of apo hCN1 active site and its modifications (based on the hCN1 structure obtained after first step of docking), where different residues (*i.e.* $Glh^{173}\boldsymbol{\alpha}$ and $Glh^{173}\boldsymbol{\beta}$, Glu^{174}, Asp^{202}) are protonated and hydroxyl anion is bridging two Zn^{2+} ions. Relative free energies are calculated with respect to the model system, where water is bridging two Zn^{2+} ions and other residues are deprotonated. 114

Table A2.12: Distances (in Å) between two zinc ions and their coordinating residues from my 2-ps long QM/MM simulations of carnosine and homocarnosine complexes with monomeric and dimeric hCN1**α** at 298 K and crystallographic data of apo-enzyme (PDB ID 3DLJ). Bond lengths from simulations are reported as average values..... 115

Table A2.13: Distances (in Å) between two zinc ions and their coordinating residues from my 1-ps long QM/MM simulations of carnosine complex with monomeric hCN1**α** at 298 K and 303 K, carnosine and homocarnosine complex with monomeric hCN1**β** at 298 K and crystallographic data of apo-enzyme (PDB ID 3DLJ). Bond lengths from simulations are reported as average values..... 116

Table A2.14: Angles (in °) between two zinc ions and their coordinating residues from my 2-ps long QM/MM simulations of carnosine and homocarnosine complexes with monomeric and dimeric hCN1**α** at 298 K and crystallographic data of apo-enzyme (PDB ID 3DLJ). Angle values from simulations are reported as average values. 117

Table A2.15: Angles (in °) between two zinc ions and their coordinating residues from my 1-ps long QM/MM simulations of carnosine complex with monomeric hCN1**α** at 298 K and 303 K, carnosine and homocarnosine complex with monomeric form of hCN1**β** at 298 K and crystallographic data of apo-enzyme (PDB ID 3DLJ). Bond lengths from simulations are reported as average values. 118

Table A2.16: Probability of hydrogen bonds between the ligand and the hCN1**β** active site from 1-ps QM/MM simulations. ^{WM} indicates water mediated hydrogen bond.119

Table A2.17: Relative free energies (in kcal/mol) of hCN1**α** and hCN1**β** of *holo* hCN1 active site (with **CT-I** and **HT-I** present inside the active site). Relative free energies were calculated with respect to the model system, where water is bridging two Zn^{2+} ions and Glu^{173} is deprotonated. The position of the hydrogen atoms was optimized. Single-point calculations were performed on the optimized structures. Values for the system with **HT-I** present inside the active site are in parentheses. 120