Investigation of the deamidation reaction in the enzyme triosephosphate isomerase by means of computational chemistry tools

Ilke Ugur

To cite this version:

Ilke Ugur. Investigation of the deamidation reaction in the enzyme triosephosphate isomerase by means of computational chemistry tools. Other. Université de Lorraine, 2014. English. NNT : 2014LORR0020 . tel-01750678

HAL Id: tel-01750678
https://hal.univ-lorraine.fr/tel-01750678
Submitted on 29 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ
DE LORRAINE

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

Faculté des Sciences et Technologies

Collégium Sciences et Technologies

Ecole Doctorale Lorraine de Chimie et Physique Moléculaires (SESAMES)

THÈSE

présentée pour l'obtention du grade de
Docteur de l'Université Lorraine et
Docteur de Boğaziçi University, Istanbul (Turquie)
en Chimie Informatique et Théorique

par **Ilke UGUR**

Etude de la réaction de déamidation dans l'enzyme triosephosphate isomérase au moyen d'outils de calculs en chimie

Investigation of the deamidation reaction in the enzyme triosephosphate isomerase by means of computational chemistry tools

Soutenance publique le 27 Fevrier 2014 à Boğaçizi University (Istanbul, Turquie)

Membres du jury :

President:	Pr.	P. Doruker	Boğaziçi University, Turquie
Rapporteurs:	Pr.	S. Antonczak	Université Nice Sophia Antipolis, Nice
	Pr.	A. Dejaegere	IGBMC, Strasbourg
Examinateurs:	Pr.	V. Aviyente	Boğaziçi University, Turquie (co-directeur de thèse)
	Pr.	G. Monard	Université de Lorraine (co-directeur de thèse)
	Dr.	B. Balta	Istanbul Technical University, Turquie

Unité Mixte de Recherche CNRS-UL 7565 SRSMC
Faculté des Sciences et Techniques, BP 70239 - 54506 Vandœuvre-lès-Nancy Cedex

To the best chemistry teacher I know,

my mother,

Gönül Gökmeydan

ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to my advisor Prof. Viktorya Aviyente. She inspired me with her wisdom when I was an undergraduate student and from then, she leaded me with her invaluable advises through all the steps of my academical study. I am also grateful to her for motivating me to have a joint PhD thesis, that gave me the chance to work with my second advisor, Prof. Gerald Monard. I am appreciative of him for leading me with his profound intelligence during my PhD studies. His extensive guidance, and most importantly his unique critical thinking, contributed not only towards my scientific background but also towards my intellectual development. I would also like to thank him for his help and comprehension during my adaptation process to the new French culture.

I would also like to thank the members of the examining committee, Prof. Serge Antonczak, Assoc. Prof. Bülent Balta, Prof. Annick Dejaegere, and Prof. Pemra Doruker for their critical comments and advices.

I would like to express how lucky I find myself to have two brilliant office colleagues in both institutions; Burcu Dedeoglu and Antoine Marion. I am grateful for both of them for all of the scientific discussions we had, without which it would be impossible to complete this thesis. I would also like to thank them for their endless support, motivation, and sharing the most joyful times of my life. Special thanks to Antoine one more time, for his invaluable contributions to the main chapters of this thesis.

I am indebted to all the members of the Bogazici University Chemistry Department, especially to Prof. Naz Zeynep Atay, Eliza Kalvo and Hülya Metiner for supporting and motivating me throughout my education. My dearest friends, Sesil Agopcan, Melis Cağdas, and all the members of Aviyente group, whom were also always with me to share their knowledge and support. I would like to thank the former members of our group, especially Şaron Çatak and Nihan Çelebi Ölcüm for their wise advices. Special gratitude is extended to Sezgin Topçu for his friendship and unfading encouragement.

I am grateful of all of my dear friends; Gülsah Çifci, Cem Öztürk, Pınar Cakır, Gokhan Çaylı, and Hikmet Karayel, who made my research assistant duty not only easier, but also more enjoyable. Dear Hikmet, your memory continues to live on with everything that you have taught us all...

I also would like to express my gratitude to the members of SRSMC, each of them contributes to the unique environment of our laboratory. Xavier Assfeld, Manuel Ruiz Lopez, Antonio Monari, Frencasca Ingrosso, Thibaut Very, Daniel Bonhery, Marina Kasimova, and Thibaud Etiene, were always there for me for anything that I might need. It was a great chance for me to be inspired by them and enjoy our little town together.

My debts are owed to my mother Gönül Gökmeydan and my sister Simge Uğur for their everlasting encouragement, support and love. A special thanks to my mother, the sunshine of my life, for putting the love of chemistry in my heart when I was a little child, and inspiring me throughout my whole life with her incredible strength.

This research has been supported by TUBITAK Joint PhD Program (2214-B), TUBITAK-CNRS project (Project No. TUB-108T870), Joint PhD scholarship from French Embassy, Ankara. I acknowledge the support from Bogazici University Foundation (BUVAK). Part of this work was granted access to the HPC resources of CINES (Centre Informatique National de l'Enseignement Superieur) under an HPC-EUROPA2 allocation, made by GENCI (Grand Equipement National de Calcul Intensif).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

A	Helmoltz free energy
C	Carbon atom
E	Electronic energy
H	Hamiltonian operator
k	Force constant
N	Nitrogen atom
N	Total number of electrons in the system
$\mathbf{p}_i$	Momentum vector
q_A	Atomic charge
R^2	Pearson correlation coefficient
$\mathbf{r}_i$	Position vector
S	Sulfur atom
U	Bias potential
$v(0)$	Velocity vector
Z_A	The charge on nucleus for an atom A
Å	Angstrom
ΔG_{elec}	Electrostatic free energy
ΔG_{sol}	Solvation free energy
$E_H[\rho(r)]$	Electron-electron Coulombic energy
$E_{KE}[\rho(r)]$	Kinetic energy of the non-interacting electrons
ϵ	Dielectric constant
m_i	Mass of a particle
$\phi(\mathrm{r})$	Electrostatic potential
$\mathrm{p}K_a$	Acid dissociation constant
Ψ	Wave function
ρ	Charge density

LIST OF ACRONYMS/ABBREVIATIONS

Ala	alanine
AM1	Austin method 1
AMBER	Assisted Model Building with Energy Refinement
Arg	arginine
Asn	asparagine
Asp	aspartate
ATP	adenosine triphosphate
B3LYP	Becke-3-Lee-Yang-Parr
BLYP	Becke-Lee-Yang-Parr
CAP	3-chloroacetol phosphate
CG	Side chain carbon
CHelpG	Charges from Electrostatic Potentials using a Grid based method
CNDO	Complete neglect of differential overlap
CPCM	Conductor-like screening model
Cys	cysteine
DFT	Density Functional Theory
DHAP	dihydroxyacetone phosphate
ESP	Electrostatic Potential
GAP	D-glyceraldeyde 3-phosphate
Glu	glutamate
His	histidine
INDO	Intermediate neglect of differential overlap
iso-Asp	iso-aspartate
LCAO	Linear combination of atomic orbitals
Lys	lysine
MAD	Mean Average Deviation
MD	Molecular Dynamics
Met	methionionine

MK	Merz-Kollman model
MM	Molecular Mechanics
MNDO	Modified neglect of differential overlap
MO	Molecular Orbitals
NBO	Natural bond orbital
NDDO	Neglect of differential overlap
NPA	Natural population analysis
OD1	Side chain oxygen
O	Oxygen atom
PCM	Polarizable Continuum Method
PDB	Protein Data Bank
PM3	Parameterization method 3
PMF	Potential of mean force
QM/MM	Quantum Mechanics/Molecular Mechanics
QM	Quantum Mechanics
RESP	restrained electrostatic potential
SCF	Self consistent field
Ser	serine
Thr	threonine
TPI	Triosephosphate isomerase
TS	Transition state
Tyr	tyrosine
VDW	Van der Waals
WHAM	Weighted histogram analysis method

1. INTRODUCTION

1.1. Survey on Asparagine Deamidation

Deamidation of asparagine is the post translational modification of asparagine (Asn), resulting a mixture of aspartate (Asp) and iso-aspartate (iso-Asp) residues. Either free asparagines, or internal asparagines (i.e. bearing a sequence of XxxAsnYyy) might undergo deamidation reaction in peptides and proteins. The enzymes asparaginases are shown to catalyze deamidation of free asparagines. No evidence has been found for a natural enzyme which catalyzes deamidation of internal asparagine, revealing that, this reaction occur spontaneously under physiological conditions. This non-enzymatic chemical modification is also detected *in vitro* peptides and proteins resembling *in vivo* features.

Reaction mechanism of non-enzymatic deamidation is pH dependent. Under strong acidic conditions (pH $\leq$ 5), deamidation is catalyzed by acids, generating aspartate only. Suggested mechanism for acid catalyzed deamidation [1] is initiated by protonation of side chain NH_2 by the acid in the media (Figure 1.1). Equilibrium between conjugate base and water yields hydroxide ion which then attacks to the carbonyl of the asparagine side chain. That yields a tetrahedral intermediate bearing a good leaving group: NH_3. Finally, aspartate is formed by the release of NH_3. Deprotonation of side chain generates aspartic acid.

Under neutral and basic conditions (pH $\geq$ 6), the reaction is base catalyzed and the products are a mixture of aspartate, and iso-aspartate. Production of iso-aspartate requires cleavage of peptide backbone (Figure 1.2). This fact indicates that the base catalyzed mechanism is principally different than the acid catalyzed one which involves only side chain reaction. Experimentally isolated cyclic imide intermediate (succinimide) [2] justified the backbone contribution in basic and neutral media (Figure 1.2). Suggested mechanism of deamidation [3] under basic and neutral conditions is initiated by deprotonation of backbone amide of the neighboring residue (i.e. residue

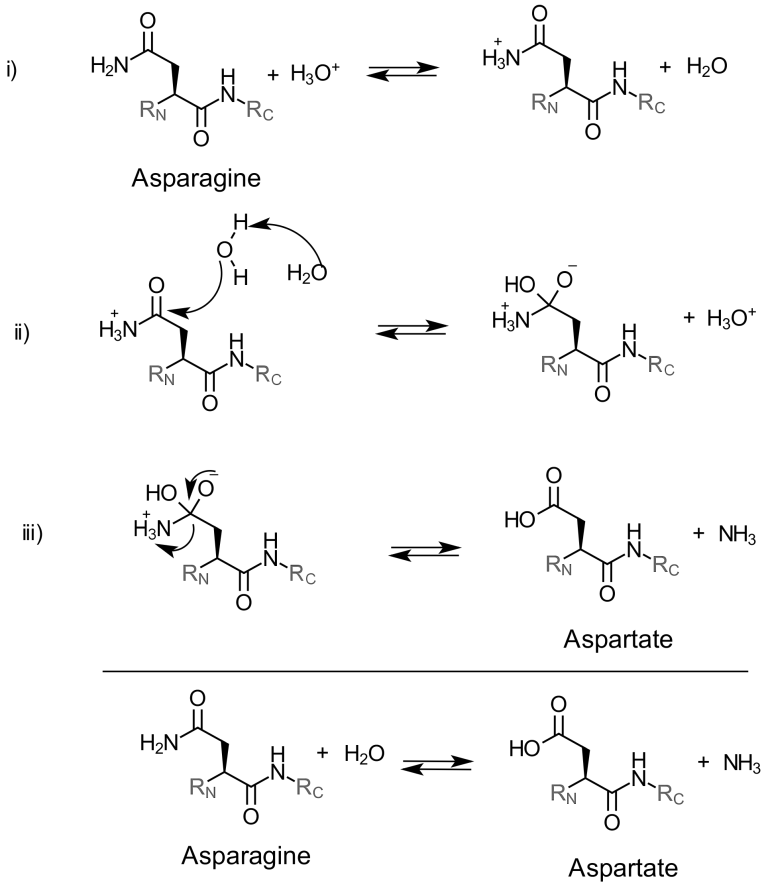


Figure 1.1. Suggested mechanism of asparagine deamidation under acidic conditions (pH ≤ 5) [1].

located C-terminal end of asparagine), in an equilibrium with the base (or water for $6 \leq \text{pH} \leq 8$) in the media (Figure 1.2). Nucleophilic attack of activated backbone nitrogen of neighboring residue to asparagine side chain generates cyclic tetrahedral intermediate. Protonation of amine functional with the excess proton in the media and release of NH_3 produces succinimide (Figure 1.2). Although there is no evidence of the formation of cyclic tetrahedral intermediate, the dielectric dependence of deamidation at pH 7.4, indicates an ionic transition state [4].

Hydrolysis of the succinimide intermediate by the attack of water to either carbonyl functionals yields iso-aspartate or aspartate (Figure 1.2). Capasso et. al. showed that hydrolysis of succinimide is relatively faster with respect to cyclization step, and defined succinimide formation as the rate determining step of the overall reaction [3].

The same group also suggested that the features of reaction mechanism of deamidation in small peptides are transferable to protein studies [3].

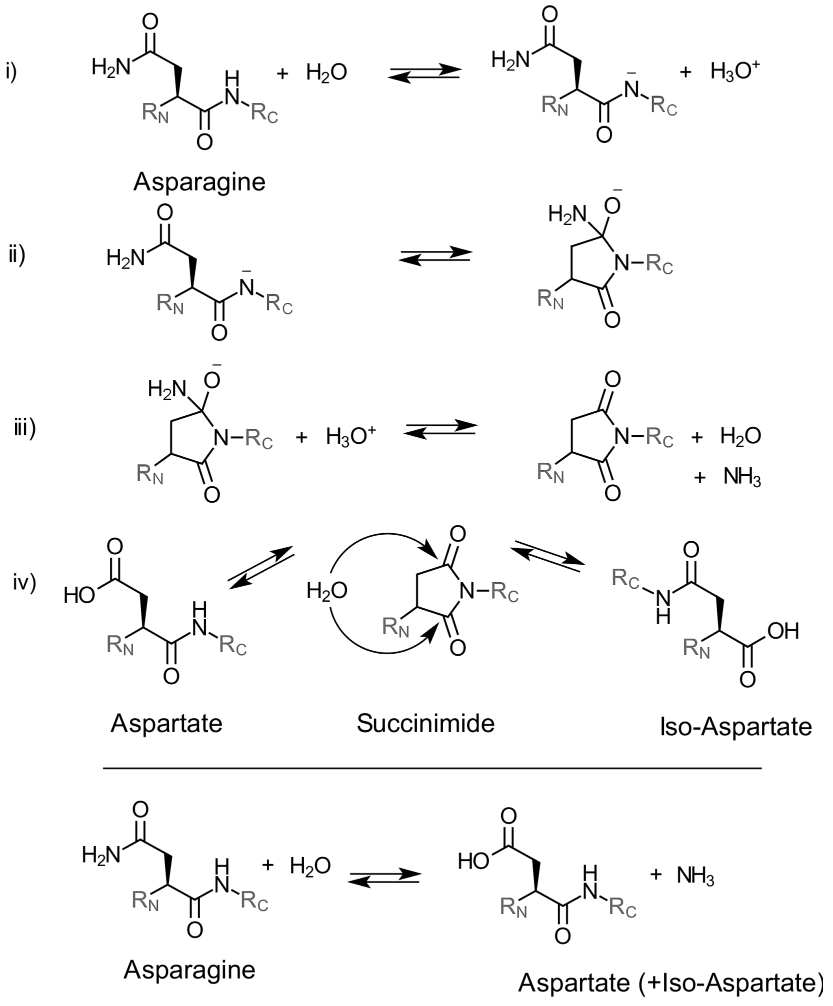


Figure 1.2. Suggested mechanism of asparagine deamidation under neutral or basic conditions (pH $\geq$ 6) [3]. $H_2O \leftrightarrow H_3O^+$ equilibrium was shown with respect to neutral media only. $H_2O \leftrightarrow H_3O^+$ should be replaced by $OH^- \leftrightarrow H_2O$ for basic media. Side chain equilibrium of aspartate was eliminated for simplicity.

As the succinimide formation is the rate determining step of the overall reaction, the nature of the neighboring residue (on the C-terminal side) becomes more of an issue. Extensive studies performed on deamidation rate explored that, the rate of deamidation is dominantly controlled by this neighboring residue (generally labeled as $n+1$ residue, as a reference to Asn, n) [5,6]. The most rapid deamidation was observed when $n+1$ residue is glycine. The facilitating effect of glycine on deamidation with respect to other residues was explained by several factors:

(i) Steric hindrance: It has been assumed that size of functionals on α carbon of $n+1$ residue controls the rate of deamidation, considering the up to 10 fold change of rate of deamidation when glycine is replaced by larger residues. However, further studies showed that degradation rates are not linearly correlated with the size of the functionals [4, 7].

(ii) Backbone amide acidity: Experimental studies relating deamidation rate with hydrogen exchange rate constants of backbone amide hydrogen of $n + 1$, created a new perspective to the facilitating effect of glycine [8]. The assumption here was that the rate of deamidation is directly correlated to acidity of backbone amide hydrogen of $n + 1$. This assumption fairly explained the impact of $n + 1$ residue for several non-glycine residues. However, relative rate of the peptides with glycine as a neighbor to asparagine was found to be faster than the hydrogen exchange rate estimations.

(iii) Conformational flexibility: The conformational dependence of backbone amide acidity was shown computationally [8]. The ease of reaction in glycine bearing peptides was explained by the wider conformational space that can be spanned by glycine compared to non-glycine residues.

All these results explaining the facilitating effect of glycine as the $n + 1$ residue were based on the experiments, and theoretical studies performed on small peptides. The effect of secondary and tertiary structure on deamidation was also examined and a diminishing trend was found for considerable amount of cases [9–12]. Half-time of deamidation was found to vary from a few hours to several years for the peptides and proteins bearing the same primary structure. Robinson et. al. developed a computational procedure [11–13] in order to predict the diminished rate of deamidation with respect to secondary, tertiary, and quaternary structure effects.

The extensive review on deamidation made by Robinson family [14], not only provided explanations on the sequence dependence of deamidation, but also advanced the discussions on the role of deamidation in biological systems. Deamidation was suggested to act as a biological molecular clock relying on the two facts: (i) being genetically controlled, (ii) taking place over a wide range of biologically relevant periods [14]. Although the validity of this hypothesis is still a matter of debate, several examples have been presented in the mentioned review [14], exploring time dependent degradation of protein caused by deamidation. Some outstanding examples of this hypothesis are: control of biological turnover rate in rabbit muscle aldolse, results indicating extensive amount of deamidated crystallin protein in cataract cells, and correlation between deamidation and Alzheimer.

In addition to time dependent degradation sourced by deamidation in living organisms, the chemical instability introduced by this reaction to *in vitro* cells causes several problems in pharmaceuticals. For instance, growth hormone-releasing factors were found to lose their biological activities because of being subject to deamidation during storage [15]. This major instability ascribed to deamidation was observed years before the recent developments in engineered protein drugs. Currently, controlling rate of deamidation is one of the major aspects of protein therapeutics optimization, since it is one of the most commonly occurred degradation in peptides [16, 17].

Although the role of deamidation in evolution is not clear yet, the continuous interest on deamidation in several multi disciplinary fields can be rationalized by the significant correlation with severe diseases and the undesirable effect on physico-chemical degradation of protein based drugs.

1.2. Essential Features of the Enzyme Triosephosphate Isomerase

Triosephosphate isomerase (TPI or TIM) has been a very well studied enzyme since 1962, following the discovery of the catalytic role of TPI on the energetically dis-favorable carbon-hydrogen bond cleavage [18]. The effect of several evolutionary improvements on enzyme efficiency is explored by Knowles et. al. by the studies on TPI activity [19, 20]. These studies not only classify TPI as a perfectly efficient enzyme, but also issue a statement on the power of enzyme catalysis in general. In addition, special elimination of side reactions, relating a number of severe diseases with the dysfunction of TPI, rapidly increased the interest on this enzyme [21]. Thus, more than 2000 publications are dedicated solely on TPI, and more than 150 crystal and NMR structures of this enzyme based on different organisms can be found in protein data bank (PDB).

TPI can be found in almost all of the living organisms ranging from least evolved bacteria to mammalian. It is functionalized in glycolysis pathway in which glucose is converted to pyruvate releasing free energy to form ATP (adenosine triphosphate). In glycolysis, TPI rapidly and reversibly catalyzes dihydroxyacetone phosphate (DHAP) to D-glyceraldeyde 3-phosphate (GAP) (Figure 1.3). This reversible conversion is accomplished using glutamate(Glu)167 as a catalytic base, and asparagine(Asn)11, lysine(Lys)13, histidine(His)95 to configure electronic stabilization.

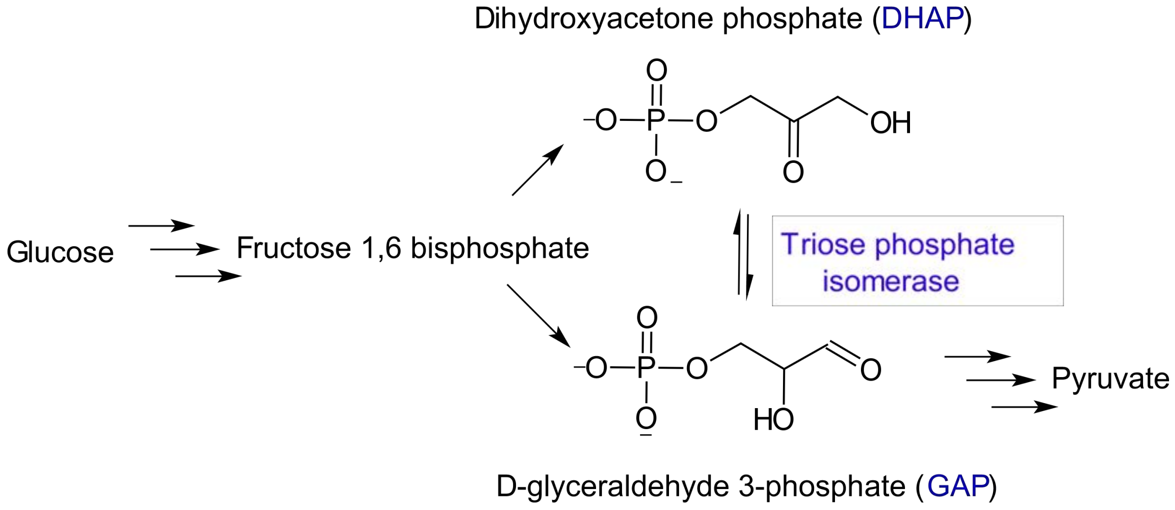


Figure 1.3. The reactions relevant to triosephosphate isomerase in glycolysis.

TPI is a homodimeric enzyme and the secondary structure of each monomer is constituted on a TIM-barrel topology (Figure 1.4). TIM-barrels are one of the most common protein folds in molecular biology and they were classified following the detection of the first crystal structure of TPI [22]. Triosephosphate isomerase is also abbreviated as TIM in literature, so TIM-barrels were actually named with respect to the enzyme that they were first detected [22]. Herein, and after, abbreviation "TPI" is used to refer the enzyme triosephosphate isomerase to avoid the confusion between other enzymes which also bears TIM-barrel structure.

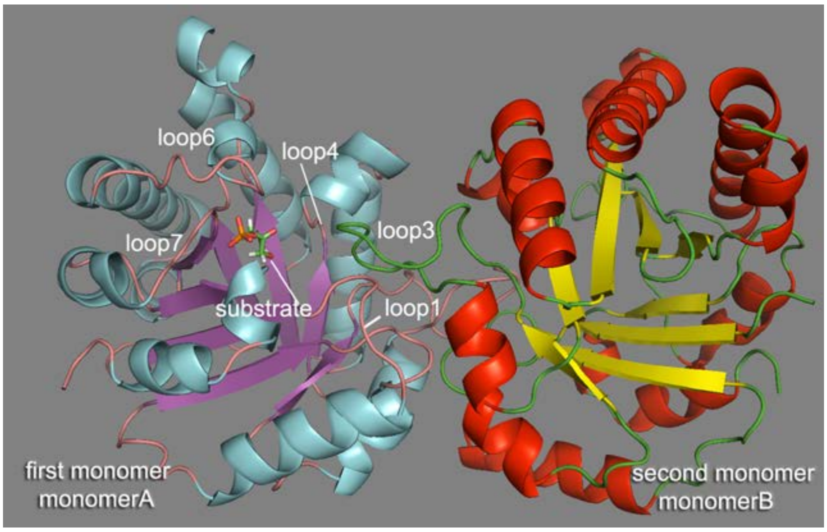


Figure 1.4. Structure of TPI dimer shown as cartoon model. Coloring is done with respect to secondary structure. Substrate (DHAP) is shown with stick model,located in monomerA.

TIM-barrels include an alternating pattern of α-helices and β-sheets combined by flexible loops [22]. This repeating pattern most commonly consist of 8 helices and 8 sheets resulting a toroid shape in which β-sheets are covered by α-helices. The loops, combining α-helices and β-sheets, are classified as stability and catalytic loops. As the name implies, catalytic loops provide the active site of the enzyme. In TPI, the active site residues are located on loop1 (Asn11, Lys13), loop4 (His95) and loop6 (Glu167) of each monomer of TPI (Figure 1.4). Although all the active site residues are located in one monomer, TPI is only active as a homodimer revealing the rigidifying effect of dimerization on the active site of each monomer [21].

Among the active site loops introduced above, substrate related conformational changes of loop6 forms one of most interesting feature of the catalytic activity of TPI. This loop, occupies two stable conformations: (i) closed form within a contact with the ligand, (ii) open form caused by 7 Å movement of the tip of loop6 with respect to the closed form. The most significant effects of the formation of a concerted closed conformation upon ligand binding is to stabilize the ligand and increase the interaction between the enzyme and the substrate with the aid of relatively small motion of loop7 (2 Å). The other influences are to protect the reaction site from bulk water and also to avoid the release of toxic intermediates which are formed during catalytic activity. General features of the active site loop in TPI is a subject of interest, and was investigated by different perspectives. [23–29]

1.3. Asparagine Deamidation in Triosephosphate Isomerase

In 1980, Gracy *et al.* detected asparagine deamidation in human TPI, on two distinct asparagines with the sequence of LysMetAsn(15)GlyArg and ValThrAsn(71)GlyAla [30]. Their findings explained previously observed heterogeneity of TPI in human cells [31, 32]. Further studies of this group showed that deamidation also occurs at Asn(15) and Asn(71) in other species of mammalian class [30].

Deamidation in mammalian TPI was shown to be an effective initiator of protein degradation [33, 34]. Gracy et. al. proposed that, corruption of intermonomeric interactions caused by deamidation might explain TPI degradation [35]. This proposal relies on the interrelated actions occurring with respect to crucial location of two distinct deamidation sites in TPI. These interrelated actions might be summarized as follows in accordance with the chronological order of findings of Gracy and coworkers.

Asn15 and Asn71 in mammalian TPI are located on loop1 and loop3 of each two monomers, returning four deamidation sites in total (Figure 1.5). Loop1 from monomer A and loop3 from monomer B (and viceversa) interdigidate in the interface of dimeric TPI, enabling Asn15 on monomerA (AsnA15) to interact with Asn71 on monomerB (AsnB71) (consequently AsnB15 to AsnA71, Figure 1.5).

Relative ratios of deamidation products (Asp71:Asp15) of TPI were found to be 2.54:1 and 1.73:1 for human and rabbit TPI respectively [30]. Relatively higher values of accumulated Asp71 with respect to Asp15 revealed the fact that Asn71 deamidates faster than Asn15. In addition to that, under alkaline conditions, in which mammalian

TPI deamidates rapidly, deamidation was not observed in chicken TPI and yeast TPI which bear Asn15, however Asn71 is replaced by Lys and Ser, respectively [36]. Further analysis showed that Asn15 deamidation in chicken TPI only observed following the oxidation of Cys126 [36]. This fact exposed that in mammalian TPI, deamidation of Asn71 and Asn15 might occur in a sequential order, in a way that Asn15 deamidation is triggered by Asp71 [34, 36].

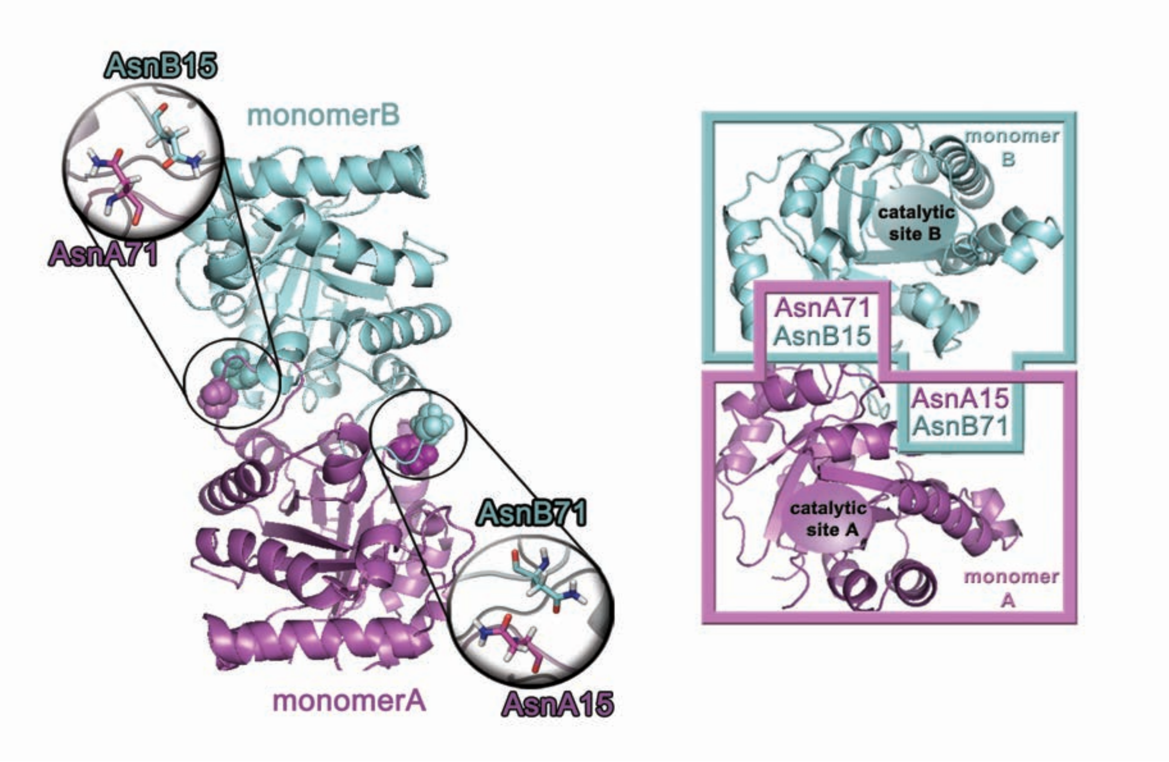


Figure 1.5. Locations of asparagine residues susceptible to deamidation in TPI. Protein structure is shown with respect to secondary structure. Asparagine residues are shown by ball and stick models.

Gracy *et al.* took their experiments a stage further to gain insights on the correlation between these two residues. One of these experiments was performed on a hybrid TPI which was formed by combining one monomer of rabbit TPI and yeast TPI [37]. Interestingly, in this hybrid TPI, Asn15 in the rabbit monomer, which lacks interactions with Asn71, did not undergo to deamidation. Hence, combining these results, it was concluded that deamidation of Asn71 is a prerequisite for Asn15 deamidation. However, predicted deamidation coefficients based on computational calculations performed by Robinson et. al. (as introduced in Section 1.2) was found as 2.27:1 for Asp71:Asp15 in human TPI, which is consistent with experimental results (2.54:1) [14]. These calculations are done considering only the protein environment of the residues in crystal structures. Strong accordance between theoretical predictions and experimentally found relative ratios concluded that these two deamidations would be independent, and the effect of Asp71 on Asn15 deamidation should be further analyzed considering the fact that the hybridization experiments would be incomplete [14].

The locations of Asn71 and Asn15 in TPI not only keep these two residues interacted, but also enable them to be affected by the catalytic site (Figure 1.5). Asn15 placed on loop1 which bears one of the catalytic residues. Asn71 is $\sim$ 25 Å away from the active site of the same monomer on which it is located. On the other hand, the distance between Asn71 of one monomer (A or B) and the active site of the other monomer (B or A) is $\sim$ 15 Å (Figure 1.5). These close arrangements of Asn15/71 and the active site increased the interest on the effect of catalytic activity on asparagine deamidation. Sun *et al.* [38] claimed that, binding of substrate analogue (3-chloroacetol phosphate-CAP) enhances deamidation of Asn71 based on their *in vitro* experiments. They observed conformational changes in the CAP-bound active site resembling to those detected during catalysis (i.e. motion of loop6, as introduced in Section 1.2).

Combination of all these experimental observations, Gracy and co workers concluded that the engaged features of asparagine deamidation in mammalian TPI serves as a regulator in protein turnover (i.e. the balance between protein synthesis and protein degradation). This hypothesis, commonly referred as *wear out* or *wear and tear mechanism*, is constituted as follows:

(i) As the catalytic activity increases, deamidation of Asn71 is accelerated,

(ii) then deamidation of Asn15 is triggered by Asp71 which is in close contact with Asn15,

(iii) deamidation of these two residues introduces strong destabilizations to the enzyme and enhances degradation (Figure 1.6).

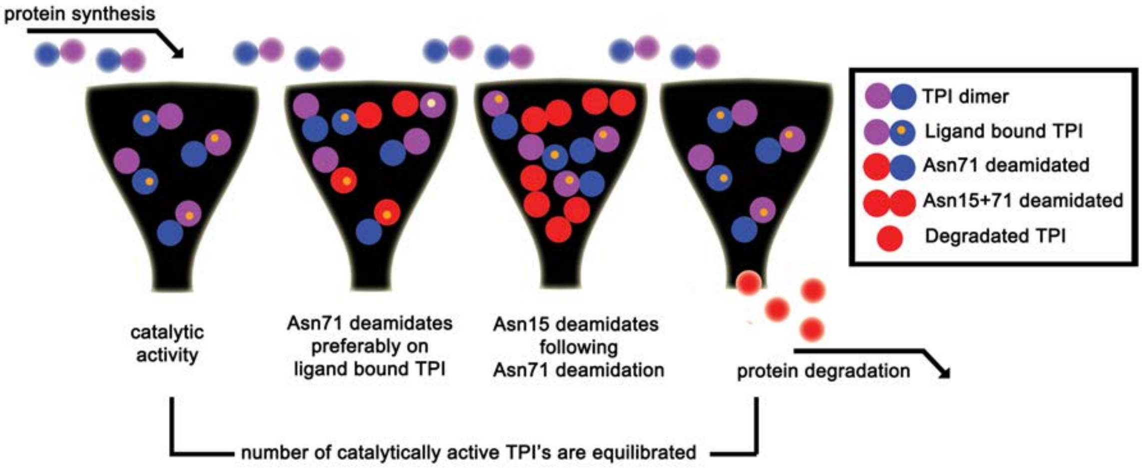


Figure 1.6. Simplified schematic representation of the role of asparagine deamidation on TPI protein turnover as proposed by Gracy and coworkers [14].

The main motive of proposing mammalian TPI as a protein turnover regulator, relies on the observed anomaly in older cells. In old cells, both Asn71 and Asn15 deamidate in the influence of catalytic activity. However, the enzyme does not degrade since degradation is known to be slower in old cells. That causes accumulation of heterogeneous TPI's in the cells, consequently occurrence of some diseases as cataract and Alzheimer. The same kind of regulation by TPI was also observed in chicken. Although Asn71 deamidation is not observed in chicken TPI, Asn15 deamidates following cysteine(126) oxidation that causes accumulation of TPI in older cells [39].

A number of other crucial diseases were found to be related to the role of mammalian TPI in protein turnover. A genetic variant of TPI was shown to deamidate much more rapid, that results a shift of protein turnover equilibrium to degradation [40]. Hemolytic anemia, myopathy, and mental retardation were reported examples of this case [14].

In this point, it is crucial to point out that deamidation of Asn71 (ThrAsnGlyAla) in TPI is much more slower than the peptide bearing the same primary structure. The single deamidataion half-life of TPI in pH 7.0, at 37°C in 0.05 M phosphate and triethanolamine buffers are 37.8, and 21.7 days respectively [14]. Pentapeptide GlyThrAsnGlyGly has deamidation half life of 1.04 days in similar conditions [14] (pH=7.4, 37°C, in 0.15M Tris-HCl buffer.) Thus, if deamidation in mammalian TPI serves as a regulator for protein turnover as suggested, it should also be noted that, higher structure effects control the predominance of degradation over synthesis by slowing Asn71 deamidation compared to isolated peptides.

The remarkable studies performed by Gracy and coworkers, constituted the milestone of protein turnover control in mammalian TPI. On the other hand, all of the explanations revealed by this group created pertinent questions:

- What are the factors that hinders Asn deamidation in mammalian TPI compared to isolated peptides?
- What is the reason behind the slower rate of Asn15 deamidation? Do Asn71 and Asn15 deamidate following the similar mechanisms?
- Is Asp71 a prerequisite for Asn15 deamidation? Or it is just a matter of occupying different locations on TPI?
- What kind of conformational changes brought by substrate facilitates Asn71 deamidation? Do these changes cause inter or intra-monomer effects on deamidation?

All of these questions might be expanded by combining the mechanistic features of deamidation as introduced in Section 1.1. Undoubtedly, explanations on those facts will put our understanding of TPI deamidation, and protein turnover one stage further.

1.3.1. Relevance of pK_a in TPI deamidation

The base catalyzed mechanism of deamidation under neutral pH was previously introduced in Section 1.1. Yuksel et. al. examined the pK_a dependency of deamidation rates in mammalian TPI using several buffers [41]. The buffers were chosen as bicine (pK_a = 8.35), borate (pK_a = 9.25), ethanolamine (pK_a = 9.44), glycine (pK_a = 9.60), Hepese (pK_a = 7.55), phosphate (pK_a = 7.21), and TEA (pK_a = 7.73). The linear relation between pK_a of buffer and rate constants in Brønsted plot, explored the fact that deamidation in TPI follows a general base catalysis mechanism [41].

The base catalyzed mechanism as shown in Figure 1.2 involves the deprotonation of the residue adjacent to asparagine. As it was stated above, this residue is glycine for both of the deamidation sites (Gly16 and Gly72). Although Asn15 and Asn72 bear the same primary structure (i.e. both followed by glycine), the pK_a of adjacent glycine (Gly16 or Gly72) might shift from the reference aqueous value with respect to the locations of Asn15 and Asn71 in TPI. Therefore, gaining insights of relative pK_a's of these two deamidation sites, and pK_a shifts that might be introduced by ligand binding would explain the key features of Asn deamidation in TPI.

2. OBJECTIVE AND SCOPE

Extensive studies performed on *in vivo* deamidation revealed that this chemical modification might be considered as a genetically controlled molecular clock, rather than being an ambiguous chemical degradation. The studies performed specifically on Triosephophate isomerase (TPI), which were introduced in the previous Chapter, exposed that the engaged features of asparagine deamidation in this enzyme would be a regulator for protein turnover. These experimental studies have remarkably enhanced our knowledge on the possible role of deamidation in TPI. In the present work, our goal is to obtain further explanations on this phenomena by focusing on the biochemical characteristics of two distinct deamidation sites in this enzyme using various computational chemistry tools.

In the following chapter (Chapter 3, Theoretical Background), the fundamental principles of the performed computations are introduced. More detailed explanations on the methodologies are presented within the relevant Chapters.

The results of the present study are introduced in four Chapters. Chapter 4 and 5 aim to provide informations on comparative reactivities of two distinct deamidation sites in TPI using classical molecular dynamics (MD) tools. In the light of the informations obtained by our classical MD studies, the kinetical calculations were performed on TPI using hybrid quantum mechanical/molecular mechanical(QM/MM) - MD methods. The relevant results are given in Chapter 6. The results obtained from these studies pointed out the strong correlation between deamidation and pK_a of the inclusive residues. This conclusion directed our focus to build a computational protocol in order to obtain accurate pK_a predictions of proteins. Chapter 7 includes the details of the proposed protocol. The crucial and interconnected conclusions are highlighted in Chapter 8.

3. THEORETICAL BACKGROUND

This chapter aimed to describe the basic principles of the theoretical approaches which were made use of in the present study. These approaches are presented in four main parts: Molecular Mechanics (MM), Molecular Dynamics (MD), Quantum Mechanics (QM), and hybrid methods (QM/MM). Regarding the rich content of each approach, this chapter is meant to provide fundamental aspects of the theoretical background by either exemplifying through the most commonly used methods or the ones directly related to the focus of the present study.

3.1. Molecular Mechanics

Molecular mechanics (MM) is based on the classical description of the molecules. In this method (also referred as force field method), the energy of a molecular system is calculated as a function of nuclear positions, in which the molecules and bonds are treated as perfect spheres and springs, respectively. Molecular systems containing significant number of atoms can not be fully modeled with the quantum approaches, because of the insuppressible computational cost. This problem is disposed by the classical approach of molecular mechanics, thus it is repeatedly used to perform calculations on biological systems with thousands of atoms as employed in the present study.

3.1.1. The Force Field

In molecular mechanics, the total energy of the system is described by parameterized equations, which are referred as force fields. The fundamental components of a force field are bonded terms (bond stretching, angle bending, torsion) and non bonded interactions (electrostatic, van der Waals). These terms are associated with pre-defined parameters, in order to predict particular properties of molecules. The contribution of bond stretching (V_{bond}) and angle bending (V_{angle}) to the energy can be expressed using harmonic potential:

$$V_{bond} = \sum_{bond} \frac{1}{2} k_{bond} (l_0 - l)^2 \tag{3.1}$$

and

$$V_{angle} = \sum_{angle} \frac{1}{2} k_{angle} (\theta_0 - \theta)^2 \tag{3.2}$$

where k_{bond} and k_{angle} are specific force constants, l_0 and θ_0 are reference values for bond length and bond angle, respectively. On account of the strong forces between bonded atoms, deviation from reference values of length and angle requires substantial energy. So that, significant variations are not observed for bond and angle terms.

The largest contribution to total energy is caused by torsional terms and non bonded interactions. Torsional potential originated by rotations about chemical bonds is described by cosine series expansion:

$$V_{torsion} = \sum_{n=0}^{N} \frac{1}{2} V_n [1 + cos(n\omega - \gamma)] \tag{3.3}$$

where V_n describes the amplitude of the barrier in the potential energy surface of rotations about ω, over periodicity of n, with the minimum angle of γ. Out of plane bending motions are other contributions of force field and they are commonly defined using the improper torsion ($V_{improper}$) functional form:

$$V_{improper} = \frac{1}{2} k_{\omega} (1 - cos2\omega) \tag{3.4}$$

in which improper torsion is the angle between four atoms that are not bonded in a sequential order.

The coupling between all of these terms mentioned above is represented by the contribution of cross terms to the force field. With respect to necessities of the calculations, either cross terms between all contributions, or only a few of them can be added to the force field.

In molecular mechanics, the electrostatic properties of the molecules are described by considering each atom as a point charge. The point charges form the basis of the electrostatic potential of a force field, and the electrostatic contribution (V_{el}) are defined accordingly by the Coloumb's law:

$$V_{el} = \sum_i \sum_j \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{3.5}$$

Point charges of a molecule are placed in positions (r_{ij}) which would give the optimum representation of experimental or quantum mechanical electric moments. They can be calculated using numerous methods. The force field used in this study, AMBER, utilizes point charges derived from electrostatic potentials [42].

Van der Waals interaction energy of molecules is calculated as the sum of all the interactions considering their separation and also their relative orientation. The most widely used form to describe Van der Waals interactions is the Lennard-Jones function:

$$V_{vdw} = 4\varepsilon_{AB} \left[\left(\frac{\sigma_{AB}}{r} \right)^{12} - 2 \left(\frac{\sigma_{AB}}{r} \right)^{6} \right] \tag{3.6}$$

here, ε_{AB} represents well depth of the potential, and σ_{AB} is the collision diameter, which equals to the arithmetic mean of diameters of pure species (σ_{AA}, σ_{BB}).

MM provides accurate geometries corresponding to minimum potential energies, that allows one to make comparisons between different conformations of systems. However, the minimum of a potential energy surface found by MM is only a static ideal geometry of a system. In reality, the molecules are in continuous motion, thereby the interactions of atoms are dynamical. These facts determine most of the chemical/physical phenomena in the nature (e.g. the functionality of biomolecular systems). In addition, the local minima found by MM would differentiate at non-zero absolute temperatures due to the thermodynamic variables which belong to the system. This motion dependent conformational changes are simulated by molecular dynamics by taking into account the thermodynamical variables.

3.2. Molecular Dynamics

Molecular dynamics (MD) generates consecutive configurations of biological systems to access their dynamical properties with a statistical mechanical approach. Production of consequential geometries is achieved by integrating Newton's laws of motion.

The time evaluated motion of the system containing N particles, can be described by classical Hamiltonian ($H(\mathbf{p}_i(t), \mathbf{r}_i(t))$ which is the sum of kinetical and potential energies of each particle:

$$H(\mathbf{p}_i(t), \mathbf{r}_i(t)) = \sum_{i=1}^{N} \frac{1}{2m_i}\mathbf{p}_i(t)^2 + V(\mathbf{r}_i(t)) \tag{3.7}$$

with $\mathbf{p}_i$(t) is the momentum vector, and $V(\mathbf{r}_i(t))$ is the potential energy.

The partial derivation of $H(\mathbf{p}_i(t), \mathbf{r}_i(t))$ gives the equations of motions. Hence, the velocity of the particle (v_i) and the force acting on it (F) are obtained as:

$$\frac{\partial \mathbf{r}_i}{\partial t} = \frac{\partial H}{\partial \mathbf{p}_i} = \frac{\mathbf{p}_i}{m_i} = \mathbf{v}_i \tag{3.8}$$

$$\frac{\partial \mathbf{p}_i}{\partial t} = -\frac{\partial H}{\partial \mathbf{r}_i} = -\frac{\partial V}{\partial \mathbf{r}_i} = F \tag{3.9}$$

which yields Newton's second law:

$$\frac{\partial^2 \mathbf{r}_i}{\partial^2 t} = \frac{F}{m_i} \tag{3.10}$$

with the particle bearing mass m_i, moving along any coordinate ($\mathbf{r}_i$) in the influence of the external force F acting on a particle.

In MD, the variations of the positions and the velocities of the particles are tracked in trajectories which are obtained by Equation 3.10. The force acting on each particle changes with respect to the variations of the position of the particle. The total force at time t ($F(t)$), is the vector sum of the interactions between particles, and it is assumed to be constant within a small time step (∂t). There are numerous algorithms implemented in MD softwares that are used to integrate the equations of motion. Herein, *Velocity Verlet* algoritm [43] is exemplified to simply introduce the basic MD scheme. In this algorithm the position (r) and velocity (v) of a particle at time t + ∂t are defined as follows:

$$\mathbf{r}(t+\partial t) = \mathbf{r}(t) + \partial t\mathbf{v}(t) + \frac{1}{2}\partial t^2 m^{-1} F(t) \tag{3.11}$$

$$\mathbf{v}(t+\partial t) = \mathbf{v}(t) + \frac{1}{2}\partial t m^{-1}(F(t) + F(t+\partial t)) \tag{3.12}$$

Initial coordinates of the particles, $r(0)$, might be obtained from several sources (e.g. for biological systems, the coordinates are obtained from Xray or NMR structures). Initial velocities, v(0), of the atoms are assigned randomly from the Maxwell-Boltzmann distribution, and they are adjusted to have total zero momentum of the system. The forces at time = 0 (F(0)), are derived by the potential energy from the force field as introduced in Section 3.1.1 using Equation 3.9.

Bearing the initial coordinates ($r(0)$), velocities $v(0)$, and forces ($F(0)$) for all atom molecular dynamics cycle can simply be described as follows:

(i) Displacement of coordinates ($r(0{+}\partial t)$) with respect to initial coordinates within the time interval ∂t is evaluated by the Equation 3.11,

(ii) F(0+∂t) is derived using r(0+∂t), and again the Equation 3.9,

(iii) new velocities, v(0+∂t), are found using F(0) and F(0+∂t) in the Equation 3.12,

(iv) steps (i) to (iii) are repeated until the sufficient simulation time is reached.

The sufficiency of simulation time is an ongoing debate in molecular dynamics studies [44]. As the ergodic hypothesis states, all accessible states of the system will be reached in infinite time. Yet, the simulations are limited with finite trajectories. Thus, the counterbalance between computational costs and the sensitivity of the system to various convergence criteria should be well considered. The time step, ∂t, would also be chosen as large as possible in order to maintain longer simulation times. However, ∂t is also limited with the accuracy of the integration algorithm introduced above. More specific discussion concerning the simulation of the enzyme used in this study will be detailed in Results Section.

In molecular dynamics, the macroscopic properties of the systems are investigated using *Periodic Boundary Conditions* (PBC). In PBC, the simulation box is replicated periodically and the replicas are aligned in all directions in order to form an infinite lattice. Among these replicas, only the coordinates of the simulation box (also referred as unit cell, or center box) are recorded. When a particle moves and leaves the simulation box, its periodic image moves in the same direction and replaces it. Thus, the number of particles within the simulation box remains constant. In periodic boundary conditions, the Ewald sum technique is commonly used to efficiently threat interatomic electrostatic interactions [45].

In order to obtain thermodynamic properties of the systems, statistical ensembles are used. Ensembles are formed based on number of particles (N), volume (V), energy (E), temperature (T), pressure (P). The most common ensembles in molecular dynamics are:

- NVE ensemble (microcanonical ensemble); constant N, V, and E,
- NVT ensemble (canonical ensemble); constant N, V, and T,

- NTP ensemble (isothermal-isobaric); constant N, T, and P.

Among those, NVT ensemble is the most commonly used one for the simulations of biological systems because of its computational efficiency. While simulating NVT ensemble, the exchange of energy is accommodated using different thermostat models by adjusting the temperature to a desired value. The average kinetic energy of an NVT ($< E_k >_{NVT}$) ensemble is defined classically:

$$< E_k >_{NVT} = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 \tag{3.13}$$

which is also related to temperature with the classical equipartition theory:

$$< E_k >_{NVT} = \frac{3N-6}{2} k_\beta T \tag{3.14}$$

where k_β is the Boltzman constant, and T is temperature. If the velocities are scaled at each step as $v_{new} = \lambda v_i$, the associated temperature change;

$$\Delta T = T_{new} - T(t) \tag{3.15}$$

can be found using the correlation of v_i and T in Equations 3.13 and 3.14:

$$\Delta T = \frac{1}{2}\sum_{i=1}^{N} \frac{2}{3} \frac{m_i(\lambda v_i)^2}{N k_\beta} - \frac{1}{2}\sum_{i=1}^{N} \frac{2}{3} \frac{m_i v_i^2}{N k_\beta} \tag{3.16}$$

returning;

$$\Delta T = (\lambda^2 - 1) T(t) \tag{3.17}$$

Combining Equation 3.16 and 3.17 yields the definition of scale factor λ with respect to target temperature (T_{new}) and instantaneous temperature $T(t)$:

$$\lambda = \sqrt{T_{new}/T(t)} \tag{3.18}$$

Thus, multiplying the velocities in each step by the factor λ (Equation 3.18) allows one to obtain the desired temperature (T_{new}) using the temperature obtained from kinetic energy ($T(t)$) . This process is the simplest thermostat [46] in molecular dynamics. The basic problem in this kind of thermostat is that, major temperature differences may arise between solute and solvent. Herein, a more sophisticated thermostat, Anderson thermostat [47] is used.

Anderson thermostat is based on stochastic collision method, in which the system is coupled in a heat bath [47]. Rather than multiplying the velocity by a factor (λ), the velocity of a particle in a random time interval is reassigned using Maxwell-Boltzman distribution at $T(t)$. The regenerated velocity is shown to be satisfying the normal Gaussian distribution. In each interval, the whole system is simulated at constant energy, so that temperature differences within the system is not observed. This fact makes this method a proper thermostat in order to generate accurate canonical ensembles.

The ensembles are built regarding the significance of solvent environment in the system of interest. For biological systems, the solvent environment is represented by water models to mimic the natural conditions. In MD, the solvent environment is commonly treated using explicit water molecules. Among the wide range of water molecules in literature, the rigid TIP3P [48] model has been chosen for this study, in which the electrostatic interactions are produced using Coloumb's law and Lennard-Jones potential.

With the interspersion of the mentioned fundamental components of MD, the routine molecular dynamics process can be summarized as follows:

(i) Initial coordinates of the system are prepared (either with explicit solvent molecules or not).

(ii) A complete set of parameters are assigned defining the number and mass of atoms, connection between them, eventually the interaction potential of the particles (referred as topology of the system).

(iii) Energy of the system is minimized with molecular mechanics using periodic boundary conditions.
(iv) The system is heated to an appropriate temperature (generally 300K, representing the room temperature).
(v) Equilibration of the system is performed.
(vi) Molecular dynamical trajectories are generated using a suitable ensemble and consequently a proper thermostat, with the aid of Newtonian equations of motion integrated with an appropriate time step.
(vii) Statistical mechanics analysis are performed on the trajectories with respect to the objectives of the study.

The details of equilibration, heating, time step, cutoff for interatomic interactions etc. might vary with respect to needs of the systems. More detailed calculation schemes of molecular dynamics simulations performed in this study are introduced within the relevant sections.

Molecular dynamics procedure described here uses classical Hamiltonian as mentioned earlier. Classical assumptions limit the applicability of MM (and MM-MD) to examine the electronic properties of molecules (e.g. charge transfer, bond formation, bond breaking etc.). Electronic properties are investigated by quantum mechanical (QM) tools, that will be introduced in Section 3.3. MD can also be established on quantum Hamiltonian, however application of QM-MD on systems with considerable number of atoms requires tremendous amount of computational cost that cannot be overcome for most of the cases. In addition to its computational efficiency advantage, the adequacy of classical mechanics on examining numerous natural events makes MM-MD as one of the most widely used molecular modeling tool.

3.3. General Features of Quantum Mechanics (QM)

Quantum mechanics were developed to overcome the requirement for a different approach, rather than classical mechanics, in order to describe microscopic systems. The foundation of quantum mechanics relies on exploring approximate solutions to the Schrödinger equation to describe the electronic properties of the molecules. The Schrödinger equation

$$H\psi = E\psi \tag{3.19}$$

returns the system energy, E as an eigenvalue of the Hamiltonian operator, H. The typical form of the Hamiltonian operator consists of the kinetic and potential energy terms of the nuclei and the electrons:

$$H = \frac{-\hbar^2}{2m_e}\sum_i \nabla_i^2 + \frac{-\hbar^2}{2m_k}\sum_k \nabla_k^2 + V_{e-e} + V_{e-N} + V_{N-N} \tag{3.20}$$

where m_e is the mass of the electron, m_k is the mass of the nuclei, $\hbar$ is the Planck's constant divided by 2π, ∇ is the Laplacian operator. When the system is defined within Cartesian coordinates, the Laplacian has the form

$$\nabla_i^2 = \frac{d^2}{dx_i^2} + \frac{d^2}{dy_i^2} + \frac{d^2}{dz_i^2} \tag{3.21}$$

The potential energy terms; V_{e-e}, V_{e-N}, and V_{N-N} are the electron-electron repulsion, attraction of electron and nuclei, and nuclei-nuclei repulsion, respectively. All of these are applied classically.

Accurate solutions of the Schrödinger equation can be obtained for small systems, however it is extremely difficult in the case of many-particle systems. This complexity is grounded on the fact that the motions of all of the particles are considered as correlated in the Schrödinger equation. Some approximations should be applied to remedy the complexity of the wave function. Under physical conditions, the nuclei of the molecule moves much more slowly than the electrons, since the mass of the nuclei is much more larger than the electrons. Considering this fact, Born-Oppenheimer approximation suggests that electronic energies can be computed for fixed nuclear positions. That provides elimination of the attractive electron-nuclear potential energy term, taking the nuclear kinetic energy term independent of the electrons, and considering the nuclear-

nuclear potential energy term constant. So the Schrödinger equation with respect to this approximation is

$$(H_{el} + V_{N-N})\Psi_{el} = E_{el}\Psi_{el} \tag{3.22}$$

The eigenfunctions of Equation 3.22 are the molecular orbitals (MO) of the molecular system and they can be constructed as a linear combination of atomic orbitals (LCAO, [49]). Atomic orbitals are defined by basis sets, usually with Gaussian functions.

3.4. *ab initio* Methods

Another complexity of the polyelectronic systems is due to the fact that the molecular orbitals mentioned in Section 3.3 are not provided as an input when the Schrödinger equation is attempted to be solved. The electronic structure calculations are done to obtain this variable. The variational theorem provides the route for these calculations, which states that the eigenvalue (energy) of an approximate wavefunction will always be higher than the eigenvalue of the true wavefunction. So that, starting from an estimated wavefunction, one should seek for the minimum value of energy at which the first derivative of the electronic energy with respect to the molecular orbital is zero. To reach the point at minimum energy, an iterative process is applied which is known as the self-consistent field (SCF) theory. SCF takes Coulomb and exchange integrals into account and uses Hartree-Fock (HF) equation:

$$f_i x_i = \varepsilon_i x_i \tag{3.23}$$

where f_i is the Fock operator, x_i is the molecular orbital of the electron(i), and ε_i is the energy of the system. The SCF procedure can briefly be described with the following simplified version:

(i) An input for molecular geometry is prepared
(ii) An initial guess for molecular orbitals is obtained

(iii) Hartree-Fock secular equation is constructed and solved to have ε_i

(iv) ε_i is used to obtain a second set of solutions to x_i which will be used in the next iteration,

(v) Step (iii) is repeated until the minimum energy point is reached at which the results are self-consistent.

Requirement of computation of many integrals makes these calculations expensive for large systems. SCF theory, also known as Hartree-Fock method, is an *ab intio* method which does not include empirical parameters and electronic correlation resulting from the instantaneous interactions of electrons.

3.5. Semiemprical Methods

Semiemprical methods were developed to overcome the computational cost of calculating and manipulating integrals in *ab initio* methods. In these methods, the core electrons are summed into the nuclear core, considering that only the valence electrons are involved in chemical properties that is needed to be investigated. This yields the neglect of some integrals and their compensation by empirical parameters. The accuracy of these methods relies on the parametrization quality. Since 1965, several semiempirical methods were developed with different approximations on parametrization [50–59].

The first methods: the complete neglect of differential overlap (CNDO) [50], intermediate neglect of differential overlap (INDO) [51] and neglect of differential overlap (NDDO) [52] were introduced by Pople and coworkers. Although those methods are not used frequently in modern computational chemistry, they provided the fundamentals of semiempirical approaches.

In 1977, Dewar *et al.* developed modified neglect of differential overlap (MNDO [55]). Following this development, all modern semiempirical methods are parameterized based on MNDO. Austin Model 1 (AM1) [56], PM3 [57], PM6 [59] are examples of the most commonly used semiempirical methods in modern quantum chemistry studies. The discussions on the efficiency of these methods are introduced in the relevant sections.

3.6. Density Functional Theory

Density Functional Theory (DFT) is based on Hohenberg-Kohn Existence Theorem [60] which proposes that with a known density of a system, one can form the Hamiltonian of the corresponding system and solve the Schrödinger equation to obtain energy. According to Hohenberg and Kohn, the potential due to the nuclei is the only external potential $V(r)$ affecting the electrons of the system and this potential is determined by the electron density $\rho(r)$:

$$\rho(r) = N\int \ldots \int |\Psi(r_1, r_2, \ldots r_n)|^2 dr_1 dr_2 \ldots dr_n \tag{3.24}$$

where r_i represents coordinates of the electrons. The real electron density is evaluated variationally as it is done in Molecular Orbital theory. According to Hohenberg-Kohn Variational Theorem [60], any electron density, other than the real one, will yield higher energy than the ground state energy. The ground state electronic energy is a function of electron density function, which implies the “functional” part of DFT and is represented as:

$$E[\rho(r)] = \int V(r)\rho(r)dr + F[\rho(r)] \tag{3.25}$$

where the first term describes the interactions of the electrons with the external potential, and $F[\rho(r)]$ is the sum of the kinetic energy of the electrons and interelectronic interactions. According to the Kohn and Sham theory [61] $F[\rho(r)]$ can be decomposed into:

$$F[\rho(r)] = E_{KE}[\rho(r)] + E_H[\rho(r)] + E_{XC}[\rho(r)] \tag{3.26}$$

where $E_{KE}[\rho(r)]$ is the kinetic energy of the non-interacting electrons, $E_H[\rho(r)]$ is the electron-electron Coulombic energy, and $E_{xc}[\rho(r)]$ is the exchange-correlation energy including both the effects of quantum mechanical exchange-correlation, as well as the correction for the classical self-interaction energy. $E_H[\rho(r)]$ is actually the classical electrostatic energy between two charge densities, therefore does not take into account the correlation between motions of electrons. It can be described as the summation of electron-nuclei interactions and electron-electron repulsion over all nucleus (M):

$$E_H[\rho(r)] = -\sum_{A=1}^{M} \int \frac{Z_A}{|r - R_A|} \rho(r) dr + \int \frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|} dr_1 dr_2 \tag{3.27}$$

and the kinetic energy of non-interacting electrons over all N electrons:

$$E_{KE}[\rho(r)] = \sum_{i=1}^{N} \int \Psi_i(r) - \frac{\nabla^2}{2} \Psi_i(r) dr \tag{3.28}$$

$E_{XC}[\rho(r)]$ term does not only contain the contributions of exchange and correlation, but also the correction between the real kinetic energy of the system and $E_{KE}[\rho(r)]$ term. It is rather more complicated than the other contributions of $F[\rho(r)]$ and can be computed using “local density approximation (LDA)” [62]. LDA is based on uniform electron gas model which assumes that the electron density is constant throughout all space. The electron correlation energy as a function of the density in the uniform electron gas ε_{XC} is expressed as:

$$E_{XC}[\rho(r)] = \int \rho(r) \varepsilon_{XC}(\rho(r)) dr \tag{3.29}$$

$E_{XC}[\rho(r)]$ is differentiated to yield the exchange correlation functional ($V_{XC}[\rho(r)]$):

$$V_{XC}[\rho(r)] = \frac{dE_{XC}[\rho(r)]}{d\rho} \tag{3.30}$$

consequently:

$$V_{XC}[\rho(r)] = \rho(r) \frac{d\varepsilon_{XC}(\rho(r))}{d\rho(r)} + \varepsilon_{XC}(\rho(r)) \tag{3.31}$$

Hence, the hamiltonian h_{KS} of the Kohn Sham equations

$$h_{KS}\phi_i = \varepsilon_i \phi_i \tag{3.32}$$

can be expressed as:

$$h_{KS} = -\frac{\nabla}{2} - \sum_{A=1}^{M} \frac{Z_A}{|r - R_A|} + \int \frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|} dr_1 dr_2 + V_{XC}[\rho(r)] \tag{3.33}$$

Thus, solving Kohn Sham equations only requires accurate interpretation for E_{XC}. E_{XC} is expressed analytically and generally treated separately as E_X and E_C. E_X and E_C are described by local or non-local (gradient corrected) methods. There are several density functional methods and the most popular E_X functionals are PW86, B88, PW91, PBE; and the E_C functionals are LYP, PW91, PBE, P86 [63]. The hybrid DFT methods add a component of the exact exchange energy calculated from Hartree-Fock theory. The exact exchange energy (${E_X}^{exact}$) is obtained from Kohn-Sham orbitals. The first hybrid DFT method was introduced by Becke, and named as B3LYP [64–66]:

$$E_{XC} = E_{XC}^{LSDA} + a_0(E_X^{exact} - E_X^{LSDA}) + a_x \Delta E_X^{GC} + a_c \Delta E_C^{GC} \tag{3.34}$$

with ${E_X}^{LSDA}$ is the exchange energy derived from local spin density approximation, ${E_X}^{GC}$ and ${E_C}^{GC}$ are gradient corrections for exchange and correlation, respectively. a_0, a_x, and a_c are empirical coefficients. The other hybrid functionals used in this study, i.e. O3LYP, PBE, and PBE0 [67–69], use the same method as B3LYP does, but different empirical coefficients. The rest, which were introduced by Truhlar and coworkers, i.e. M05, M05-2X, M06, M06-2X, M06L [70–73] uses different coefficients, but also the functionals depend on spin kinetic densities. Hybrid methods are proved to give more accurate results than their separable counterparts, i.e. B3LYP is more accurate that BLYP, etc. However, the accuracy and efficiency of hybrid functionals are subject to major discussions in computational chemistry. The details of this discussions will not be given in this section, thus in Chapter 7, the accuracy and efficiency of several DFT functionals is examined concerning the objective of this study.

3.6.1. Basis Sets

Basis sets are used to describe orbitals of a system mathematically, extensively based on Gaussian functions in modern DFT. The most commonly used basis sets are the split valence double zeta basis sets which were developed by Pople [74–81]. In these basis sets, the number of functions in the minimal basis is doubled for valence electrons. The inner shell is described by contracted and diffuse parts, therefore the functions are splitted. In order to obtain greater flexibility, split valence basis sets can be expanded by adding polarization functions, either on all of the heavy atoms or both heavy atoms and hydrogen atoms. Even after adding polarization functions, the space occupied by Gaussian functions might be insufficient to describe the systems where electrons are far from the core. To overcome this deficiency diffuse functions can be added to the basis set. The level of DFT functional and the basis sets should be arranged with respect to the properties and the behavior of the system modeled.

3.6.2. Atomic Charges Derived from Electron Density

There is no quantum mechanical operator to observe atomic charges of a molecular system, so that there is no unique definition of the quantum charges. In order to assign charges on atoms, population analysis methods screen the electron density between nuclei [82]:

$$q_A = Z_A - \int \rho_A(r) dr \tag{3.35}$$

where q_A is the atomic charge, and Z_A is the charge on nucleus for an atom A. The second term of the r.h.s of the Equation 3.35, integration over all electron density, is the total number of electrons in the system (N) and can be represented as:

$$N = \sum_{\mu}^{AO} (PS)_{\mu\mu} \tag{3.36}$$

with $P_{\mu\mu}$ and $S_{\mu\mu}$ are the electron density and the overlap population summed over all atomic orbitals AO. In Mulliken [83] population analysis, all of the electron density of an orbital is assigned to the atom with the corresponding atomic orbital. For each element of overlap population, the electron density is divided between atomic orbitals which are contribute to the corresponding overlap.

$$q_A = Z_A - \sum_{\mu \in A}^{AO} (PS)_{\mu\mu} \tag{3.37}$$

However, Mulliken charges are basis function dependent, since it is assumed that each function is centered on the nuclei. Also, the analysis might result with populations higher than 2 or even less than 0. Löwdin population analysis [84] has an adjustment to Mulliken approach in order to solve negative populations using nonorthogonal basis:

$$q_A = Z_A - \sum_{\mu \in A}^{AO} (S^{1/2} P S^{1/2})_{\mu\mu} \tag{3.38}$$

where the Löwdin atomic population is obtained by summing all the diagonal elements of the orbitals centered on a given atom. Natural population analysis (NPA) [85] gives another perspective than Mulliken and Löwdin charges. NPA is based on Natural Bond Orbitals (NBO) scheme which differentiate between lone pair electrons and the ones involved in covalent bonds corresponding to the Lewis structure of the molecule. The atomic orbitals are partitioned to core, valence and Rydberg orbitals. The partial charges are obtained by summing over Natural Atomic Orbitals (NAO) which are the optimal effective AO's.

3.6.3. Continuum Solvation Models

In quantum mechanics, continuum solvation methods are used to represent solvent environment of molecular systems in order to obtain more accurate and realistic results for reactions which occur in solution phase. Solvation free energy of a system (ΔG_{sol}) is the summation of the electrostatic component (ΔG_{elec}), van der Waals interaction between the solute and solvent (ΔG_{vdW}), and the energy required to form a cavity within solvent (ΔG_{cav}) [82].

$$\Delta G_{sol} = \Delta G_{elec} + \Delta G_{vdW} + \Delta G_{cav} \tag{3.39}$$

ΔG_{elec} represents the polarization of the solvent and is modeled as a uniform medium with dielectric constant (ε). When the solute is brought from gas phase to solvent environment, the dipole introduced by solute in solvent cavity induces an electric field within the cavity. The work required to create a charge distribution within the cavity can be defined as the interaction between the solute charge density $\rho(\mathrm{r})$ and electrostatic potential $\phi(\mathrm{r})$ at a given point r in space.

$$G = -\frac{1}{2}\int \rho(r)\phi(r)\partial r \tag{3.40}$$

Hence, ΔG_{elec} becomes:

$$\Delta G_{elec} = \int \Psi H \Psi d\tau - \int \Psi_0 H_0 \Psi_0 d\tau - \frac{1}{2}\int \rho(r)\phi(r)dr \tag{3.41}$$

with subscript 0 denoting the gas phase Hamiltonian. In polarizable continuum method (PCM) [86], which is the most widely used continuum solvation method, the cavity surface of the solute is prescribed by van der Waals interactions. The electrostatic potential ($\phi(\mathrm{r})$) is composed of the potential due to the solute ($\phi_\rho(\mathrm{r})$) and surface charges ($\phi_\sigma(\mathrm{r})$). Conductor-like screening model (COSMO) [87,88] is an other commonly used continuum method, in which the cavity is considered to be encapsulated in a conductor with an infinite dielectric constant.

3.7. Hybrid Methods : QM/MM

Modeling of reaction mechanisms is traditionally performed using small model systems due to the computational cost of the quantum mechanical tools. However, the environmental factors (i.e. interatomic interactions) has a significant effect on most of the chemical process. The most prevailing examples of this fact are the reactions

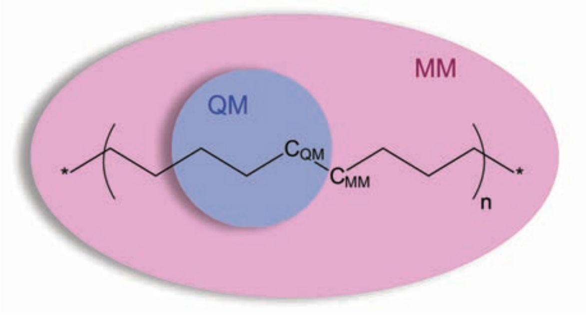


Figure 3.1. Schematic representation of partitioning the system to QM and MM subsystems.

catalyzed by enzymes under mild conditions. Apart from the enzymatic catalysis, in some cases, the reaction would have significantly different kinetics in an isolated form, compared to the model mimicking a real environment.

The interest on modeling chemical reactions in systems with a considerable amount of atoms, directed the theoreticians to improve sufficient methods that are not only constructed on quantum mechanical descriptors, but also computationally efficient. Herein, hybrid QM/MM tools are described that is one of the most widely used in the interest of modeling chemical reactions in systems with great number of systems [89–92].

The fundamental idea behind QM/MM method relies on the fact that the number of atoms directly involved in a chemical reaction is small compared to the rest of the system, and the rest only affects the reaction via non bonded interactions. Hence in QM/MM, the system is splitted into two parts: the reactive part with the small number of atoms are treated with QM, and the rest with MM as shown in Figure 3.1. This partition is accomplished by defining the Hamiltonian by the contribution of QM (H_{QM}), MM (H_{QM}), and the interaction between these two ($H_{QM/MM}$):

$$H = H_{QM} + H_{MM} + H_{QM/MM} \tag{3.42}$$

where H_{QM}, and H_{MM} has the same form as introduced in previous equations. The main components of $H_{QM/MM}$ term representing the electrostatic interaction of MM atoms with QM electrons, and repulsion between the MM and QM cores. In addition to electrostatic term, non-electrostatic interactions are described by a van der Waals term ($V^{vdW}_{QM/MM}$), yielding:

$$H_{QM/MM} = \sum_{k} \frac{q_M Z_k}{R_{kM}} - \sum_{iM} \frac{q_M}{r_{iM}} + V^{vdW}_{QM/MM} \tag{3.43}$$

with q_M is the point charge in MM part, Z_k is the core charge of QM atom k, R_{kM} is the distance between QM atom k and MM atom M, r_{iM} is the separation between QM electron and MM atoms. Considering the fact that, non electronic terms are defined classically, the partition of the QM/MM Hamiltonian can also be builded as follows:

$$H = H_{elec} + H_{nonelec} \tag{3.44}$$

where the non electronic terms are:

$$H_{nonelec} = H_{MM} + V^{vdW}_{QM/MM} + V^{nuclei}_{QM+QM/MM} \tag{3.45}$$

Thus, all the parts that require quantum treatment (H_{el} and $V^{nuclei}_{QM+QM/MM}$) are computed by quantum mechanical tools, whereas the rest (H_{MM} and $V^{vdW}_{QM/MM}$) are by molecular mechanics codes. QM calculations might be performed using *ab initio*, DFT, or semiempirical Hamitonians. The choice of Hamiltonian for QM part, as well as the force field for the MM part, is limited to computational efficiency and the accuracy of the method.

As it was stated earlier, the QM part of the system is defined as the reactive part and the rest is the MM part. Particularly in studies on enzyme reactivity or modeling reaction mechanisms within a protein, the reactive QM part is covalently bonded to the MM atoms as shown in Figure 3.1 ($C_{QM} - C_{MM}$). This covalent bond (or any others at the frontier between QM and MM parts) requires special treatment, because the MM atoms lack of valence electrons due to the definition of point charges. Among several methods in literature proposed for this special treatment, *link atom* method [93] has been chosen in this study. The idea in this method is to align a monovalent atom (generally, hydrogen) along the $C_{QM} - C_{MM}$ bond. This monovalent atom (generally refereed as *link atom*) interacts with the MM part as its QM counterparts do. *Link atom* can be managed to be either free to move, or fixed at 1 Å from C_{QM} [94].

The full minimization of the system might be performed on systems with small number of atoms using QM/MM Hamiltonian. In contrast, for biological systems with large number of atoms, other conformational sampling techniques should be applied like molecular dynamics (generally abbreviated as QM/MM-MD). The change of energy with respect to variations in conformational space -commonly referred as potential of mean force (PMF)- can be successfully generated by QM/MM-MD.

On the other hand, recalling the ergodic hypothesis, not all the components of a reaction coordinate can be sampled within a finite simulation time. Thus, some other techniques should be applied in order to obtain system's potential of mean force which includes components in higher energy levels. Umbrella sampling technique (Section 3.8) is one of the prosperous process that enables one to sample non accessible states within a limited simulation time.

3.8. Sampling Methods: Umbrella Sampling

The free energy of the system ($A(\xi)$) is related to the probability distribution ($P(\xi)$) of the components for the reaction coordinate:

$$A(\xi) = -k_\beta T \ln(P(\xi)) \tag{3.46}$$

For the states with the energy barrier higher than $k_\beta T$, free energy cannot be obtained because of the limitations of sampling within a finite simulation. In Umbrella Sampling technique, sampling around desired states is accomplished by adding a bias potential

(U) to the Hamiltonian:

$$H = H_0 + U \tag{3.47}$$

This biasing potential (U) is carried out on the system along the reaction coordinate, and it usually has the form of an harmonic potential:

$$U = \frac{1}{2}k(r - r_0)^2 \tag{3.48}$$

where k is the force constant, r is the parameter defining reaction coordinate and r_0 is the equilibrium value of the parameter. In order to accomplish sufficient sampling, the reaction coordinate is split to consecutive windows and the system is forced to be sampled within the restrained conformations in each window [95]. Among these windows, as the configurations deviate from equilibrium state, the value of biasing potential (U) increases quadratically.

The potential of mean force of the system (i.e the energy landscape between two states) can be retrieved by unbiasing and recombining the simulation windows. The unbiased energy in the window i is described as:

$$A_i(\xi) = -k_\beta T \ln(P'(\xi)) - U_i(\xi) + F_i \tag{3.49}$$

in which $P'(\xi)$ is the biased probability and F_i is the unknown free energy constant. The unbiasing and recombining procedure in order to obtain the complete PMF can be accomplished by using the weighted histogram analysis method (WHAM) [96]. In this method, the probability distribution's of each window $(P'(\xi))$ are calculated and modified with respect to the applied potential $(U_i(\xi))$. After obtaining the free energy of the each window separately, the curves are combined applying an iterative process.

The parameter defining reaction coordinate (ξ) can be chosen as distance, angle, dihedral, linear combination of distances etc. In order to obtain successful PMF from

Umbrella sampling technique, the distribution of the chosen reaction coordinate at window i should sufficiently overlap with the distribution at $i+1$ and $i-1$ windows. This adequate overlap is obtained by properly adjusting the force constant (k) of the harmonic potential and the window size. Large and small value of k respectively yields narrow and wide probability distribution within one window. Wide distribution shows that penalty applied to reaction coordinate is not enough to keep the system within desired limits. Narrow distribution, on the other hand, causes insufficient overlap. The number of frames within one window (simulation time) should also be long enough in order to obtain accurate statistics from each window. The choice of reaction coordinate, force constant, window size and simulation time is detailed in the relevant sections.

4. INITIATION OF THE REACTION OF DEAMIDATION IN TRIOSEPHOSPHATE ISOMERASE: INVESTIGATIONS BY MEANS OF MOLECULAR DYNAMICS SIMULATIONS*

4.1. Abstract

Deamidation of asparagine is the spontaneous degradation of this residue into aspartic acid. The kinetics of this slow reaction is mainly dependent on the nature of the adjacent amino acid that follows asparagine in a peptide or protein primary sequence. In the homodimer triosephosphate isomerase (TPI), there are two main deamidation sites per subunit: Asn15−Gly16 and Asn71−Gly72 for which deamidation dynamics are known to be interrelated. In this study, we investigate the initiation of the deamidation reaction in TPI by means of molecular dynamics. Simulations based on classical AMBER force field are performed in a 60 to 90 ns time scale for six distinct samples. Conformational changes, desolvation effects, and hydrogen bond networks are analyzed to interpret the experimental findings and previous quantum mechanical (QM) results. Results that are based on desolvation analysis clarify the assignments in the literature about the different behaviors of two deamidating sites in TPI. Conformational analysis supports findings suggested by QM studies: the most favorable reaction mechanism is the one that yields to succinimide intermediate via one or two step routes. The mechanism leading to the succinimide intermediate most likely involves the formation of a tetrahedral intermediate that is formed either directly from asparagine or via a side chain tautomer intermediate. In all cases, surrounding water molecules are present to assist the reaction.

4.2. Introduction

Deamidation is the conversion of the neutral amide side chain to the negatively charged carboxylate. Among the 20 neutral amino acids, asparagine (Asn) and glutamine (Gln) residues are known to undergo spontaneous nonenzymatic deamidation to form aspartic acid (Asp) and glutamic acid (Glu) residues under physiological conditions [2, 104–107]. This reaction is of significant biological interest because it causes time dependent changes in conformation and limits the lifetime of peptides and proteins. Deamidated proteins have been found in some aged and diseased tissues, such as human eye lens cataracts [108] and Alzheimer's plaques [109]. The timed processes of protein turnover, development, and aging have been suggested as possible roles for deamidation [110]. Robinson et al. have proposed the molecular clock hypothesis, which suggests that deamidation is a biological molecular timing mechanism that could be set to any desired time interval by genetic control [111]. Furthermore, they proposed that the instability of asparaginyl and glutaminyl residues is their primary biological function and that they serve as easily programmable molecular clocks [111]. Recent experiments [3, 3, 112–116] and computations [8, 117–121] have been in accord with this hypothesis and provided compelling evidence of its significance.

Some peptides and proteins, which are of special interest in the development of knowledge about deamidation, have been studied more thoroughly. In this study, mammalian triosephosphate isomerase (TPI) is chosen to elucidate the deamidation of asparagine by using molecular dynamical (MD) tools. This enzyme is biologically important and also has special features in order to investigate deamidation reactions. It is a homodimeric enzyme that can be found in almost all of the living organisms ranging from bacteria to mammalian. It rapidly and reversibly converts dihydroxyacetone phosphate (DHAP) into glyceraldehyde-3-phosphate (GAP) in the glycolytic pathway. The enzyme is only active in the dimeric form, which consists of identical monomers of 246 residues. The monomers are folded as TIM-barrel, which is built by eight α-helices on the outside, eight parallel β-strands on the inside, and eight loops referred to as loop1 to loop8 [122, 123].

It has been shown that mammalian TPI deamidation occurs preferentially at two distinct asparagines: MetAsn(15)Gly (on loop1) and ThrAsn(71)Gly (on loop3) [124]. A glycine residue next to asparagine is required for a deamidation reaction to occur.

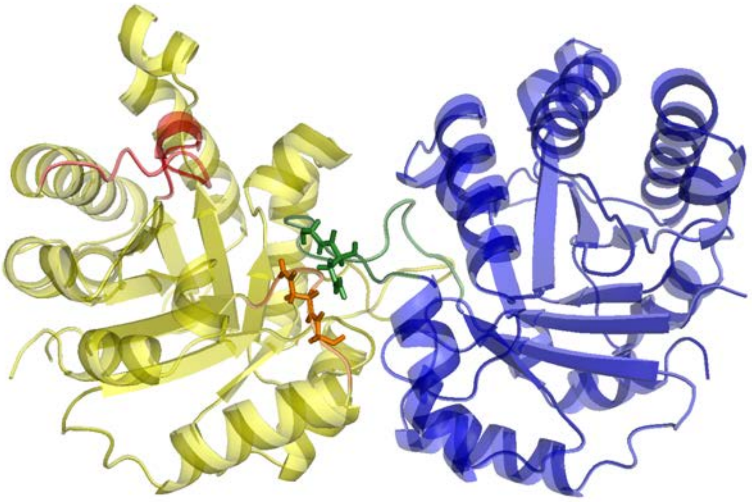


Figure 4.1. Cartoon representation of the TPI dimer. Two identical monomers A and B are represented by blue and yellow, respectively. LoopA6 is colored in red. LoopB1 and loopA3 are colored in orange and green, respectively. AsnGly residues are shown as sticks.

Gracy et al. have indicated that Asn71 deamidates earlier than Asn15, [124] so that the ThrAsn(71)Gly is named as primary deamidation site and that the MetAsn(15)Gly is the secondary deamidation site. They also suggest that the deamidation of Asn71 is a prerequisite for the deamidation of the juxtaposed Asn15 on the neighboring subunit [124]. However, the deamidation coefficient (CD) values of Asp71/Asp15 was computed as 162.27:1, which is in agreement with the results obtained by characterization of TPI deamidation products, 2.54:1 for human and 1.73:1 for rabbit TPI [124], so that these two deamidations can also be considered as independent events. [116]

On the basis of conclusions derived from several experiments, Robinson et al. reported that small peptides deamidate more rapidly than proteins and also that the secondary and tertiary structure of the enzyme has an essential impact on deamidation. [116] In the case of human TPI, the half-life of the deamidation is found as 21.7 days at pH 7, 37 °C. [124] For a pentapeptide that has the Val−Ser−Asn−Gly−Val sequence, the half-life is determined as 5.8 h under the same experimental conditions [125].

In each monomer of TPI, loop6 (residues 166 to 176) has a particular physiological function: this 11-residue loop region moves more than 7 Å and closes over the active site when the substrate binds and a hinge-type motion takes place. As the enzyme functions, a hinged lid over the active site closes, excluding water and preventing hydrolysis of triosephosphates to toxic products. Gracy et al. also suggest that, when the lid is closed, Asn71 in that monomer deamidates more readily [124] (Figure 4.1).

Capasso et al. have proposed that deamidation of relatively unrestrained Asn residues goes through a succinimide intermediate [3, 112, 126]. Quantum mechanical (QM) studies carried out by some of us have suggested that the deamidation at neutral pH can proceed through various mechanisms (see Figure 4.2 [117–119, 127–129]). Detailed discussions of these mechanisms will be provided in what follows. In this article, we explain the relationship between two deamidation sites, loop6 motion, and substrate binding using classical force field and molecular dynamics techniques. All-atom molecular dynamic simulations are performed in a 60 to 90 ns time scale for six distinct samples to interpret the experimental findings and the QM results. Deamidation of Asn residues is examined based on the configurational changes during the simulation. Considering the accelerating effect of solvation of Asn residues on deamidation, the differences of the desolvation of the residues and the hydrogen bonding networks are analyzed. The plausible reactive conformers suggested earlier in QM studies, including water assisted mechanisms, are also examined.

4.3. Computational Details

4.3.1. Preparation of the Samples

The crystal structure of the rabbit muscle TPI (PDB code 1R2R, resolution 1.5 Å [130]) was used as the starting structure of the simulations. This structure is apo and contains two identical dimers. The active site loop (loop6) in one of the four subunits of 1R2R is in the closed conformation. For simplicity, one of the dimers was deleted, and the dimer with one monomer that has the active site loop in the closed conformation was kept. The final structure of the TPI is a dimer in which the active site loop is open in one of the monomers and closed in the other. The monomer with

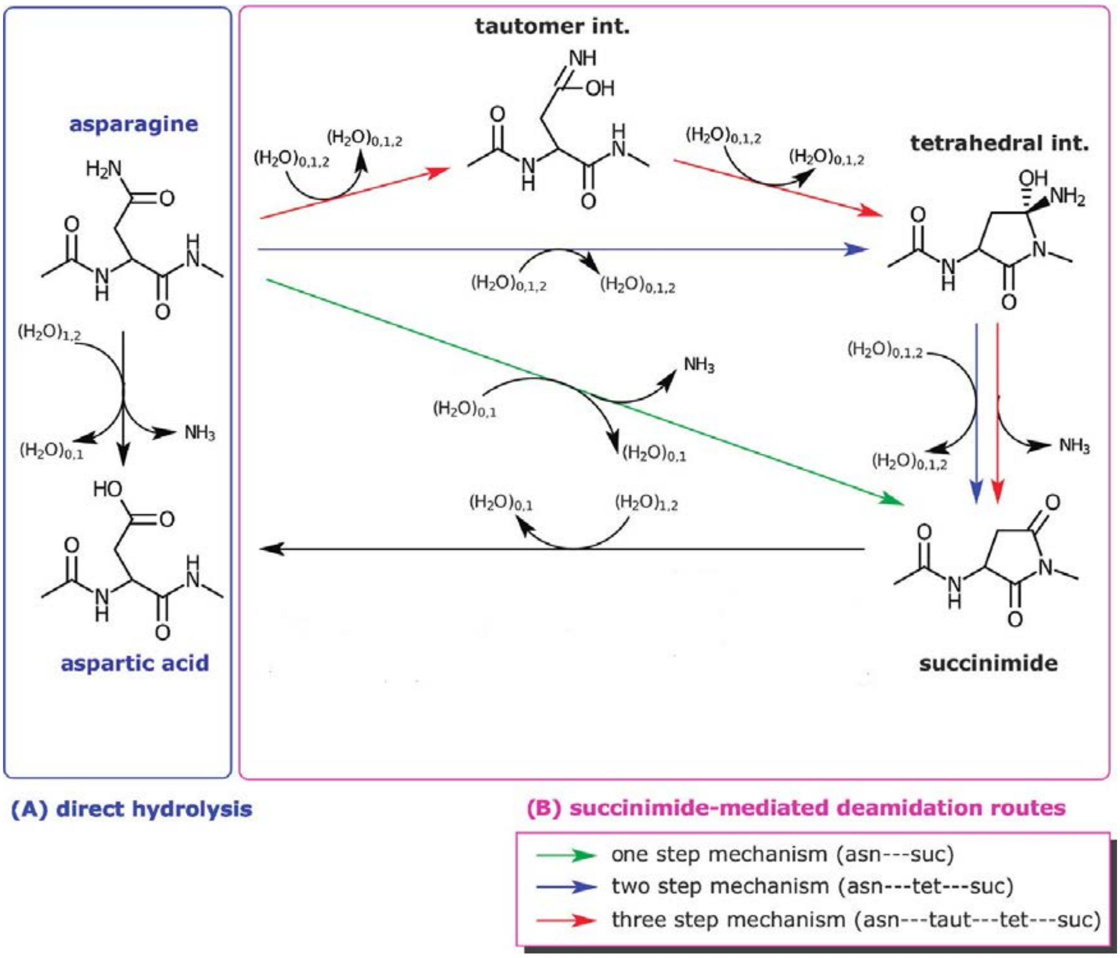


Figure 4.2. Summary of the previously suggested mechanisms for the conversion of asparagine to aspartic acid.

active site loop open is named as monomer A, and the one with the closed loop is monomer B. This nomenclature is used to label the loops and the residues regarding the corresponding monomer (loopA1 or loopB1, AsnA71 or AsnB71, etc.). Our basic concern was to understand the differences between the deamidations of Asn71 and Asn15 and to investigate the effect of motion of loop6 and substrate on deamidation. On the grounds of our concerns, six distinct samples, based on the 1R2R structure, were prepared. The nomenclature of the samples is as follows (Table 4.1): WOW: It is a wild type apo dimer TPI in which loopA6 is open and loopB6 is closed. This sample was prepared by keeping the exact structure of the crystal of the dimer of 1R2R. CC and OO: These samples were prepared to examine the effect of motion of the active site loop on deamidation. The closed structure of loopB6 is copied, and the loopA6, which has the open structure, was replaced with this copied closed loop. Both of the loop6's were kept closed at the beginning of the simulation (sample CC). The reverse of this procedure was performed to obtain the sample in which both loop6s were kept open at the beginning of the simulation (sample OO). CS: To understand the effect of substrate, one ligand- bound sample was prepared by using DHAP as the ligand. DHAP conformation inside the TPI active site was extracted from Yeast TPI structure (PDB code 1NF0, [131] resolution 1.2 Å) and incorporated into the active site of monomer B of 1R2R by active site structure fitting. According to a Needleman−Wunsch sequence alignment [132] comparison method, 1R2R and 1NF0 share 50% identical residues and have 67% sequence similarity. Fitting of the identical residues between both structures gives an all atom rms of 1.159 Å using the McLachlan algorithm as implemented in the ProFit software. 1R2R and 1NF0 respective active sites contain identical residues (i.e., Lys 12, His 95, Glu 165, Gly 232, and Gly 233) with very similar three-dimensional structure. The rms between both binding sites is evaluated at 0.232 Å by ProFit.35 AASP and BASP: Two mutant samples were prepared to investigate the effect of Asp71 on Asn15. Asn71 was mutated to Asp71 in monomer A (sample AASP) and in monomer B (BASP). The mutations were performed by simply patching Asp on Asn. One sample, with the substrate in the open loopA6, was also prepared and simulated up to 100 ns to observe the effect of the substrate. However, the substrate left the active site at the very beginning (4 ns) of the simulations, so this sample is discarded, and the results regarding this sample are not discussed in this article.

4.3.2. Molecular Dynamics Simulations.

The systems were hydrogenated, and the topology and coordinate files were built with the tLEAP module of AMBER 9 [133]. Molecular dynamics simulations were performed using the program AMBER with the ff03 force field parameters [134]. The protonation states of histidine residues were determined according to their polar environment. Samples WOW, CC, and OO were neutral. CS, AASP, and BASP had −2, −1, and −1 charges, respectively, and counter sodium ions were added to neutralize these systems. Waters from the crystal structure were deleted, and the systems were

Table 4.1. Nomenclature of the Samples.

	monomer A		monomer B	
name	**residue71**	**loop6**	**residue71**	**loop6**
WOW	N	open	N	closed
CC	N	closed	N	closed
OO	N	open	N	open
CS	N	open	N	open
AASP	N71D	open	N	open
BASP	N	open	N71D	open

solvated with explicit TIP3P water molecules [48]. The resulting systems are cubic, of edge length 93.13 Å, have an initial density of 1.0, and contain about 24 000 water molecules. This initial setting yields to molecular systems having a protein concentration in each cubic box of about half the 1R2R crystal concentration. The simulations were done using the SANDER module of AMBER 9. An NVT sample was used with Andersen temperature coupling [47]. Periodic boundary conditions were imposed on the systems. The particle mesh ewald summation technique [135] was used with a cutoff distance of 8 Å. A time step of 1.5 fs was used by the implementation of the SHAKE [136] algorithm for the bonds involving hydrogens.

The equilibration of the systems was performed in a four stage process. First, only the hydrogen atoms of the system were allowed to move during a 15 ps long molecular dynamics using a strong temperature coupling (i.e., velocities were randomly updated every 10 steps). This was to ensure a proper geometry of the protein hydrogen atoms that were added by the tLEAP program and to relax the hydrogen bond network between the water molecules and the protein. Second, the whole system was simulated for another 15 ps with the same strong temperature coupling, but here, all atoms were allowed to move. Third, a subsequent 100 K molecular dynamics was performed with looser temperature coupling (i.e., velocities were randomly assigned every 50 steps). Fourth, the thermostat temperature was increased to 300 K, and another 15 ps molecular dynamics was performed using the same previous temperature coupling. Finally, productions were performed at 300 K using standard Andersen thermostat as implemented in AMBER 9 (i.e., with atomic velocities updated every 1000 steps). The samples WOW and BASP were simulated up to 90 ns, and the remaining samples up to 60 ns at the same temperature.

4.3.3. Trajectory Analysis

The root-mean-square deviation (rmsd) results were calculated by using the ProFit software. Atomic fluctuation (AF) calculations and hydrogen bond analysis were performed by using the ptraj module of Amber 9. Generation of the plots was done using Gnuplot 4.2.

4.4. Results and Discussion

4.4.1. Stability of the Trajectories

To investigate the stability of the trajectories, the root-mean-square deviation of Cα atoms has been computed (see Figure 4.3 for WOW and Appendix A for other samples). For all samples, the reference structure corresponds to the initial structure of the molecular dynamics (e.g., the X-ray structure for WOW). The rmsd value increases moderately along the simulation to reach a rather high average value of ∼3−3.5 Å. By distinctly computing the rmsd of each monomer, it can be seen that each monomer does not unfold and keeps a low rmsd value along the trajectory at ∼1.5−2 Å. This indicates that the motions of the monomers with respect to each other are at the origins of the large backbone deviations in TPI. This behavior has been investigated earlier [24]. Considerable high values of the rmsd of the dimeric protein can be explained by those intramonomer motions. Our findings are compatible to what have been reported in other publications. Details of the investigations of those motions will not be discussed since they are out of context. Given that rmsd values of the distinct monomers or alpha helices and beta sheets stay within a 0.5−2 Å range, we consider our simulations as stable. The atomic fluctuations based on the Cα atoms with respect to the residues are calculated (Figure 4.4). For all samples, residues on helices and beta sheets of the TPI barrel fluctuates less than the residues on the loops. The secondary structure of the protein is conserved during the simulation. Excluding the terminal ends of the protein, most flexible residues are the hinges of the catalytic loops, loop1, 2, ..., loop8 (colored in yellow in Figure 4.4). Residues between 30−32 and 141−150, which are located on the stability loops of the TPI barrel, also have some flexibility. Loop6 and loop7, which are known as the most flexible loops of TPI, [21] have the highest fluctuation values for all of the samples. These results also show the reliability of the simulations.

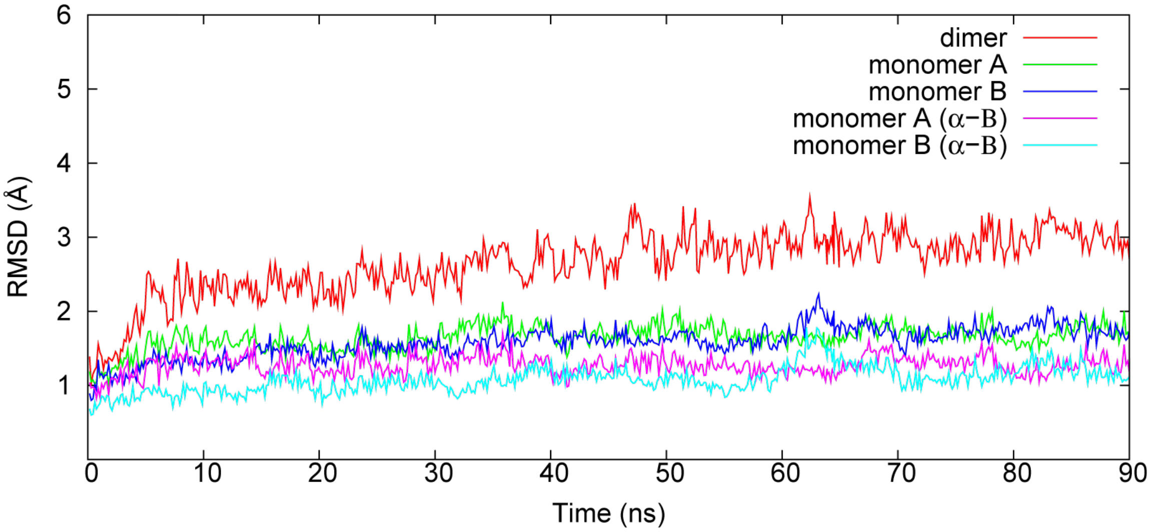


Figure 4.3. Variations of the rmsd (in Å) along the MD trajectory. Rmsd are computed for the complete dimer structure (red), monomer A only (green), monomer B only (blue), all α helices and β sheets of the subunit A (magenta), and all α helices and β sheets of the subunit B (cyan).

4.4.2. Solvent Accessibility.

4.4.2.1. Desolvation Effects on Residues Asn and Gly. Reaction paths leading to deamidation of Asn residue imply the deprotonation of the peptidic N−H group adjacent to Asn to form a succinimide ring. In the case of TPI, it corresponds to the N−H group of Gly16 and Gly72, respectively, bound to Asn15 and Asn71. The capability for the proton of this N−H group to leave glycine and to help the formation of the succinimide

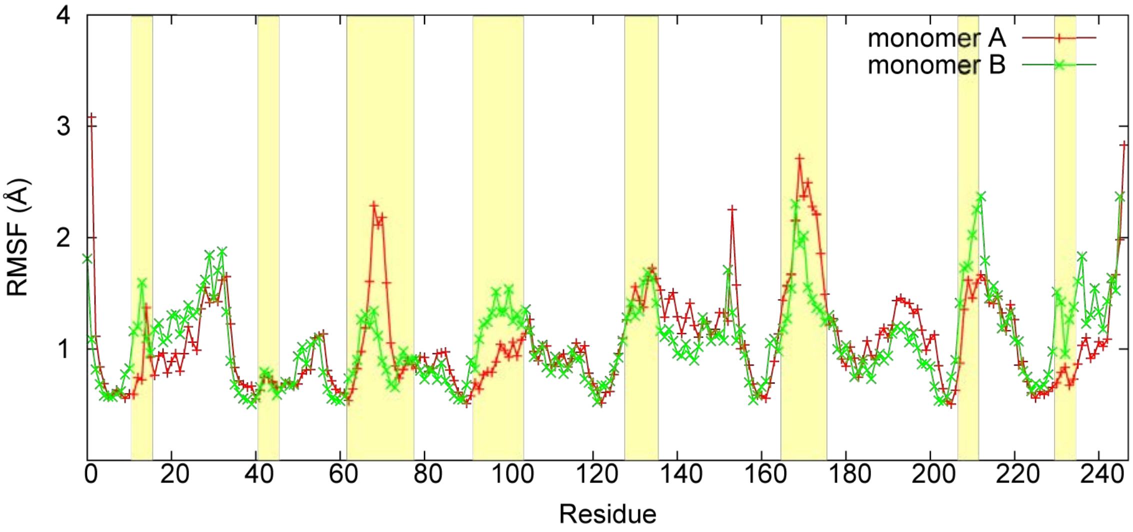


Figure 4.4. Atomic fluctuations (RMSF: root-mean-square fluctuations of Cα atoms) for the sample WOW. Yellow shading indicates the loops of the TIM barrel.

ring can be related to the pKa of the nitrogen involved in the peptide bond. There is no experimental measurement of such a pKa, but theoretical considerations can be brought on its variations with respect to local protein environment. Additionally, the peptidic side chain of Asn can also be recipient for protons, depending on the considered deamidation reaction mechanism. In the PROPKA program that has been developed to estimate pKas of ionizable protein side chains by Li et al., [137] the pKa of an ionizable residue is perturbed from its standard aqueous state by different factors: desolvation, hydrogen bonding, and charge−charge interactions [137]. In this study, pKa shift values due to desolvation of the Asn and Gly residues, located at the four deamidation sites, are calculated by 45using the equations proposed by Li et al. [137]. The calculations of pKa shifts of Asn and Gly are based on model residues, Asp and Arg respectively, and parameters are adopted as follows: (i) Asn71@CG, Asn15@CG, Gly72@N, and Gly16@N are chosen as center of ionizable groups; (ii) local radii (R_{local}) are determined as 4.5 Å and 5 Å ,respectively, in agreement with typical R_{local} used by Li et al. [137]; (iii) pKa shifts are calculated using the following equations: Local ($\Delta\mathrm{p}K_{\mathrm{LocalDes}}$):

$$\Delta\mathrm{p}K_{\mathrm{LocalDes}} = N_{\mathrm{local}}C_{\mathrm{local}} \tag{4.1}$$

where N_{local} is the number of non-hydrogen protein atoms within a distance R_{local} to the center point of the ionizable group, and C_{local} is a parameter representing the contribution of each neighbored atom to the local desolvation effect. The latter parameter is constant pKa shift per atom taken as 0.07, as suggested by Li et al. [137]. Global ($\Delta pK_{\text{GlobalDes}}$):

$$\Delta pK_{\text{GlobDes}} = (N_{15.5\text{Å}} - 400)C_{\text{global}} \tag{4.2}$$

where, $\text{N}_{15.5\text{Å}}$ is the number of non-hydrogen protein atoms within 15.5 Å of the center of the ionizable group, $\text{N}_{15.5\text{Å}} - 400$ is the excess of protein atoms, $C_{\text{global}} = 0.01$. Total (ΔpK_{Des}):

$$\Delta pK_{\text{Des}} = \Delta pK_{\text{GlobDes}} + \Delta pK_{\text{LocalDes}} \tag{4.3}$$

All the parameters used to compute the desolvation effect (i.e., C_{local}, Cglobal, R_{local}, and the 15.5 Å global radius) are chosen according to the discussions made by Li. et al. [137]. It is suggested that these constants give the best results to mimic desolvation effects on pKa deviations. In the original PROPKA formulation, only the non-hydrogen protein atoms outside the considered amino acid are counted. This is justified since only side chain ionizable groups are considered. In our case, since some ionizable groups are peptidic groups, we counted all non-hydrogen protein atoms within R_{local} or 15.5 Å of the center atoms and subtracted to N_{local} and $\text{N}_{15.5\text{Å}}$, respectively, a constant minimal number of non-hydrogen atoms. In our study, we subtracted 9 and 10 for Asn and Gly residues, respectively. Variations of $\text{N}_{15.5\text{Å}}$ along the simulations are reported in Figure 4.5.

The changes in the $\text{N}_{15.5\text{Å}}$ values of Asn@CG and Gly@N with respect to time are analyzed to monitor the variations of the global desolvation effects during the simulations. Residues having $\text{N}_{15.5\text{Å}}$ values greater than 400 are classified as buried residues based on the findings of Li et al. [137] while other residues are named surface residues. In the crystal structure Asn15@CG, Asn71@CG, Gly16@N, and Gly72@N are in the buried form with, respectively, $N_{15.5\text{Å}}$ values of 486, 428, 499, and 455 in subunit A and 506, 426, 499, and 441 in subunit B. In WOW, AsnA71@CG goes to the surface state at the very beginning of the simulation and goes back to the buried state after 70 ns of the simulation. AsnB71@CG keeps an almost constant state (i.e., $N_{15.5\text{Å}}$ between 300 and 400) during the whole trajectory. Gly16@N and Gly72@N follow the same trend as their neighboring residue with higher $N_{15.5\text{Å}}$ values in both of the monomers (Figure 4.5).

In other samples, both Asn15 side chains keep a buried state as in WOW. In the case of Asn71, trajectories show a change in state for Asn71 with, sometimes, some oscillations between buried and surface states. This is the case for AsnB71 in sample CC; AsnA71 and AsnB71 in samples CS, OO, and BASP; and AsnB71 in sample AASP.

Consequences of the global desolvation investigation show that loop3, which holds the primary deamidation site, can have two stable conformations: solvated and buried. The solvated conformation is different than the crystal structure. However, oscillations between those two conformations especially in WOW and BASP (which are simulated 30 ns longer than the rest) show that both of the conformations are accessible.

For all of the samples, secondary deamidation sites of both subunits always stay in buried conformations with $N_{15.5\text{Å}}$ values very close to 400 or higher than that. Gly@N has always higher values than Asn@CG, which indicates that Gly is more buried than Asn. It is observed that changes of $N_{15.5\text{Å}}$ value of Asn@CG and Gly@N are strongly correlated (Figure 4.6).

Following Li et al. equation for global desolvation, only residues having a $N_{15.5\text{Å}}$ larger than 400 atoms have their pKa shifted [137]. In our samples, $N_{15.5\text{Å}}$ is always lower for Asn than for its adjacent Gly (see Figure 4.6); this indicates that the N$-$H group in glycine residues (16 and 72 in both monomers) are usually more desolvated than the side chains of asparagines. In other terms, Gly is always more buried than Asn.

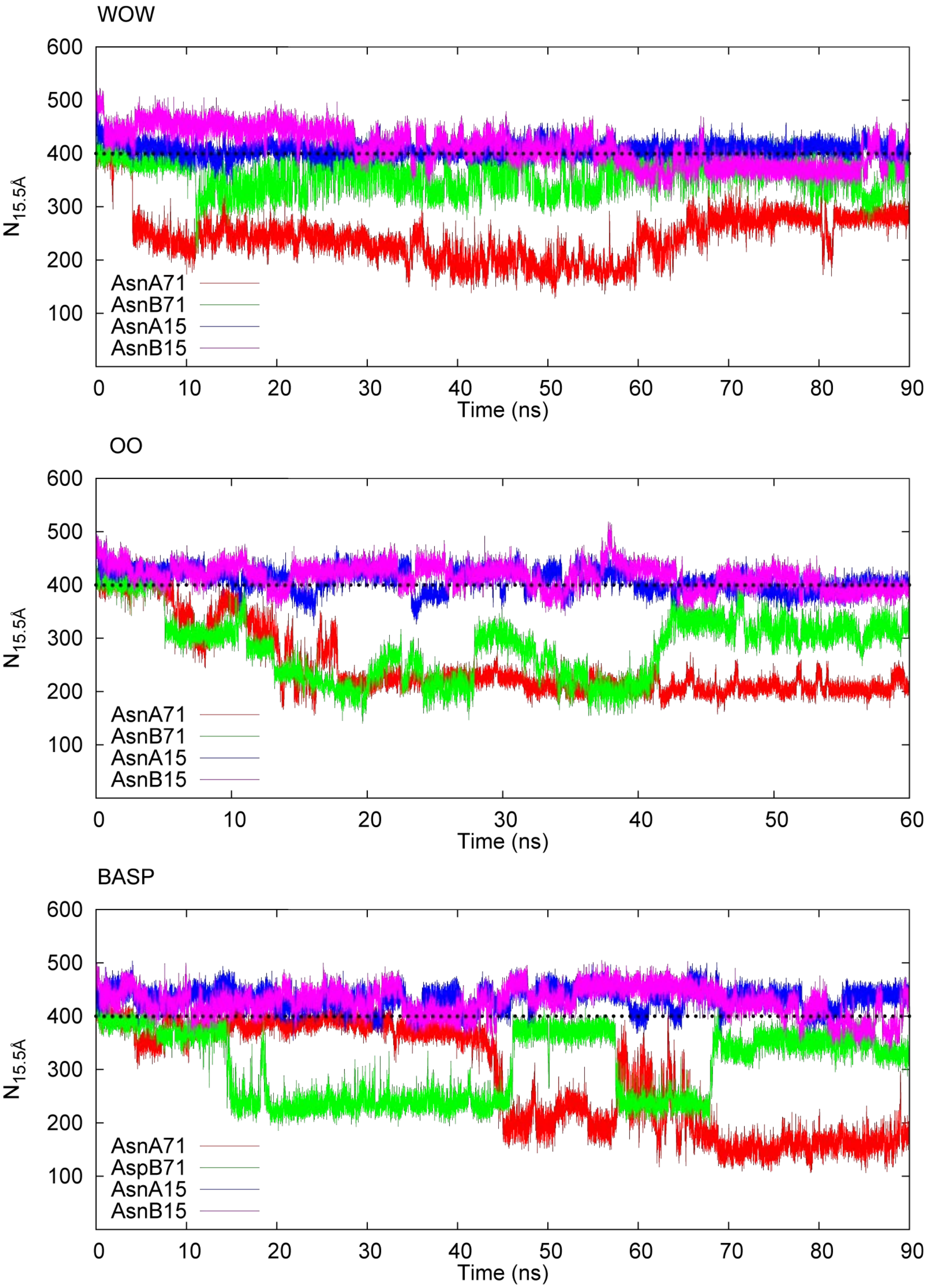


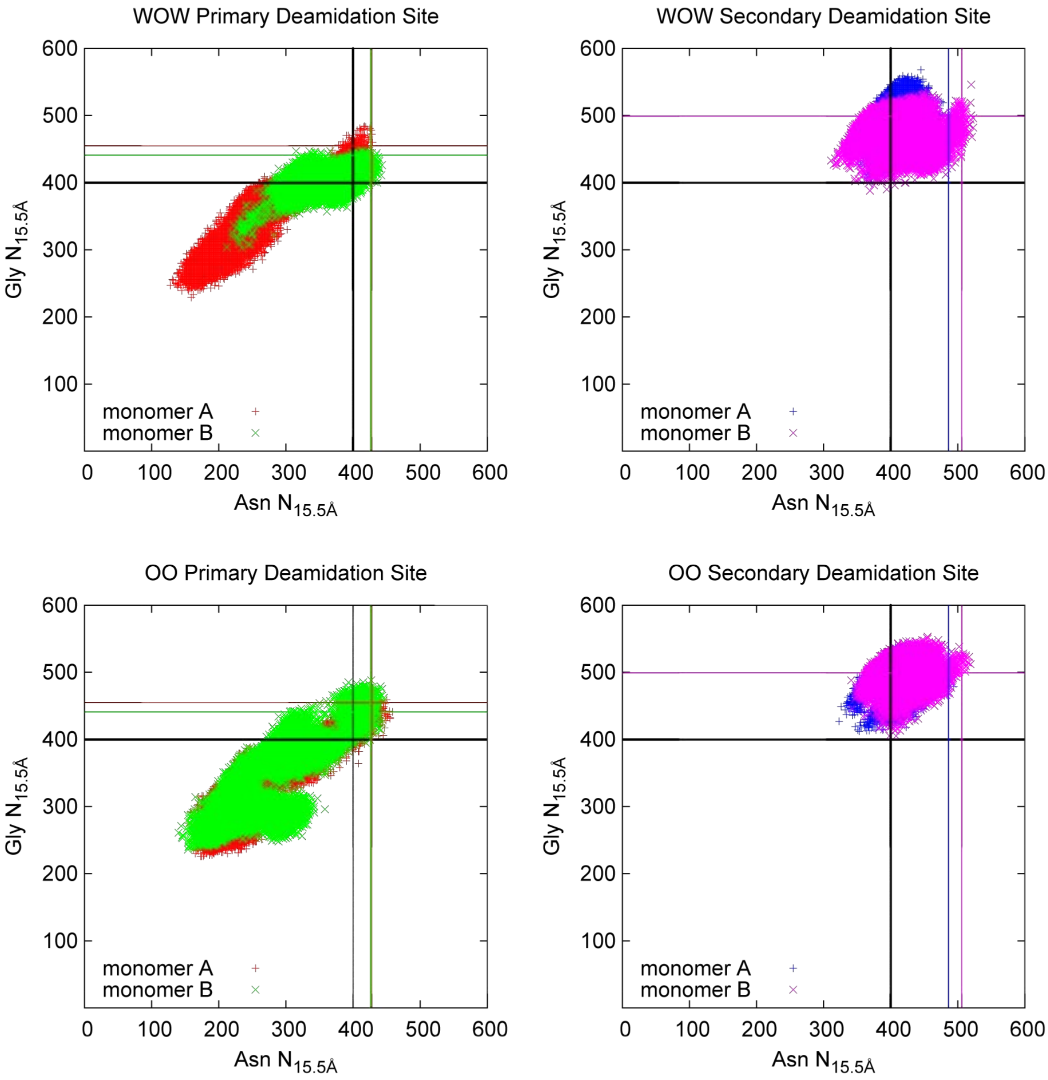
Figure 4.5. $N_{15.5Å}$ of Asn with respect to time (Asn 15A, 71A, 15B, and 71B are depicted in blue, red, magenta, and green, respectively).

When comparing Asn15Gly16 vs Asn71Gly72, it is found that Asn15Gly16 usually keeps a buried state along the trajectories, while Asn71Gly72 can be either buried or at the surface of TPI. This means that, in terms of desolvation effect, it will be more difficult to abstract the proton of the peptidic bond between Asn and Gly to form a succinimide intermediate in Asn15Gly16 than in Asn71Gly72 and that the pKa shift in Gly is higher than the one in Asn. Figure 4.7 4.8 and reports the distribution of the number of neighbors in the local vicinity (N_{local}) of Gly@H and Asn@OD1. We have not found any pattern linking local desolvation to samples' trajectories. Following Li et al. equation for local desolvation effect, it shows that local environment accounts in average for less than 1 pKa shift unit for all samples [137].

It has been found experimentally that Asn71 deamidation points to the deamidation of Asn15 [124]. Molecular dynamics simulations of Asn71 mutated structures (AASP and BASP) can produce some information on the influence of a deamidated residue on the second deamidation site.

Analysis of the correlation of $N_{15.5\text{Å}}$ between Asn and its adjacent Gly for AASP and BASP trajectories (see Appendix A) show that the presence of an aspartate residue in position 71 does not change the global desolvation state of the juxtaposed asparagine: the correlation between Asn15 and Gly16 $N_{15.5\text{Å}}$ numbers are not influenced by a N71D mutation, at least in the time scale of the present simulation.

4.4.2.2. Hydrogen Bond Analysis.. From QM static studies of deamidation on model systems done by Catak et al., [127–129] deamidation can occur when a hydrogen bond is formed between the side chain of Asn and the N−H group of the adjacent glycine. A possible descriptor of the reactant state for deamidation can be the formation of a hydrogen bond during the MD simulation between Asn@OD1 and Gly@H (defined hereafter as the hydrogen bond of interest). We report in Table 2 the percent occupancies, during the WOW trajectory, of hydrogen bonds involving Asn@OD1 as a hydrogen bond acceptor. Restrictions are made in Table 2 and, only the percent occupancies of hydrogen bonds involving Asn@OD1 as an acceptor that are higher or equal to the


WOW Primary Deamidation Site

WOW Secondary Deamidation Site

OO Primary Deamidation Site

OO Secondary Deamidation Site

Figure 4.6. $N_{15.5Å}$ of Asn vs Gly results. (Asn 15A, 71A, 15B, and 71B are depicted in blue, red, magenta, and green, respectively. $N_{15.5Å}$ of the crystal structure is shown with lines).

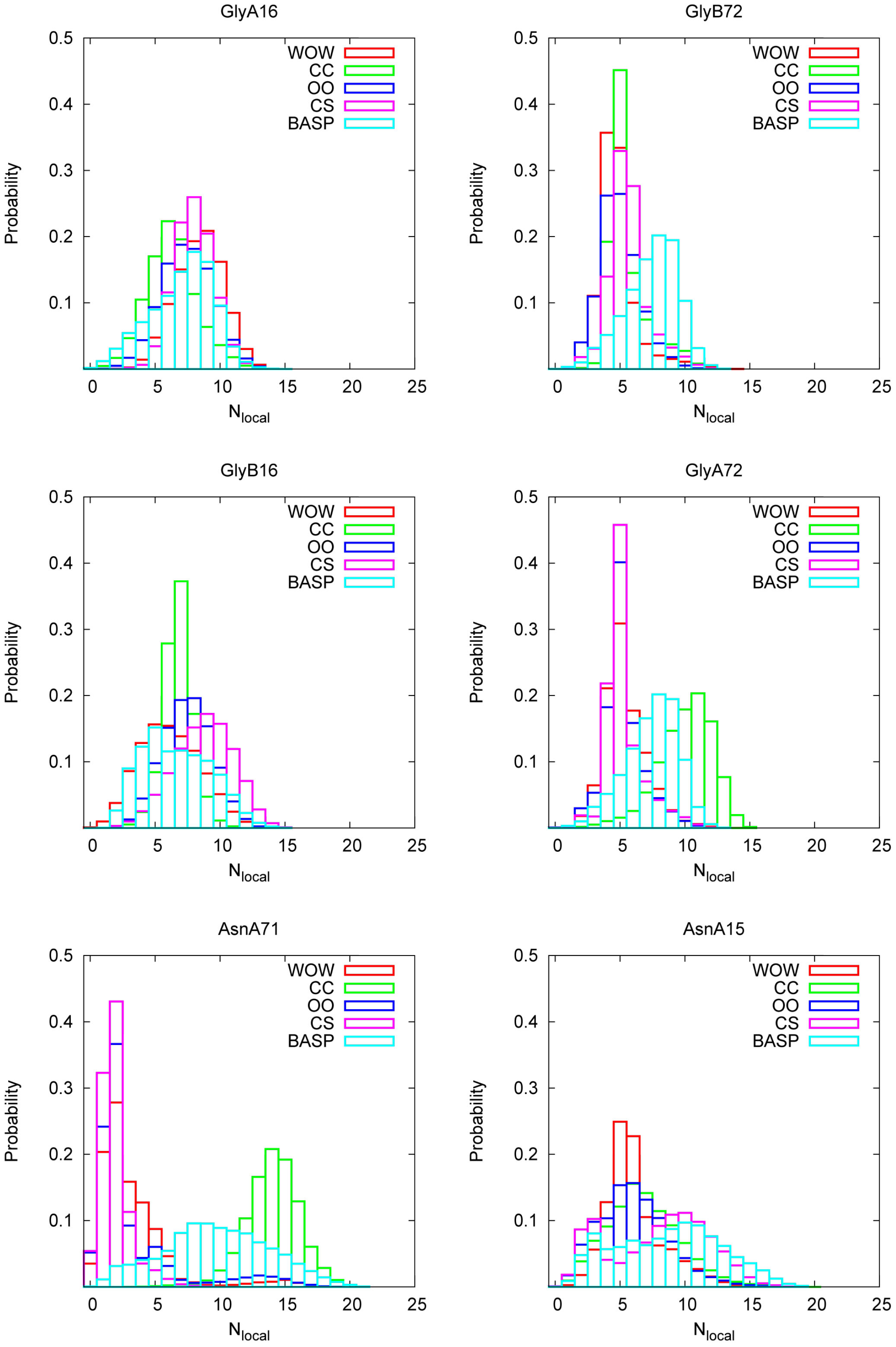


Figure 4.7. Probability of N_{local} values.

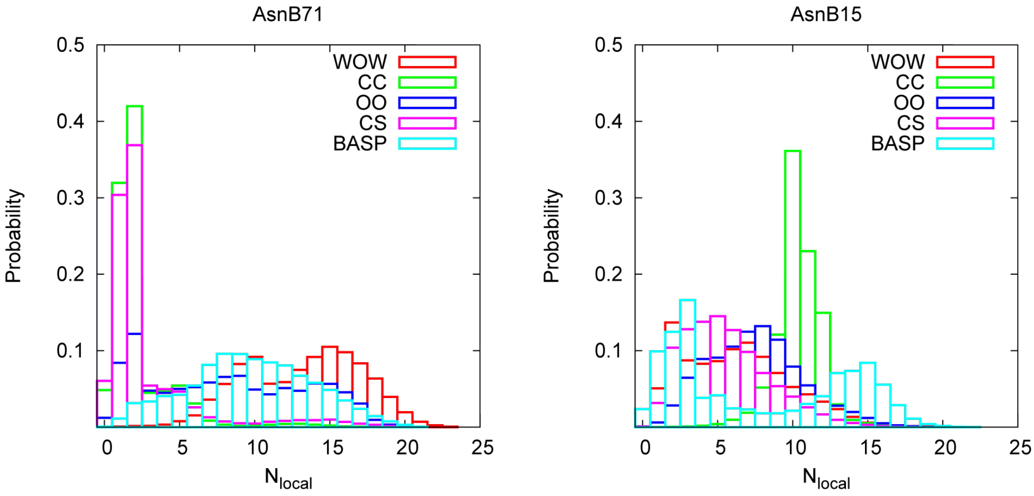


Figure 4.8. Probability of N_{local} values (cont.).

percent occupancy of the respective Asn@OD1···Gly@H hydrogen bond are reported.

During the WOW trajectory, direct hydrogen bonds between Asn@OD1 and Gly@H do not occur frequently. Percent occupancies are low (<10), and in the case of AsnB71@ OD1···GlyB72@H, nearly no hydrogen bond is present. In all cases but for AsnB71, no other hydrogen bond of interest is observed. In contrast, the dominant hydrogen bond interaction for AsnB71 involves AlaB73@H. The specific interaction occurs nearly half of the time during the WOW simulation.

When focusing on the hydrogen bond interactions of Gly@H, one can observe the preferential formation of hydrogen bond with oxygen from the backbone of TPI: GlyA@H interacts more with TrpA12@O, GlyB16@H with TrpB12@O, and GlyB72@ H with MetA14@O, respectively. Only GlyA72@H interacts with AsnA71@OD1, preferentially.

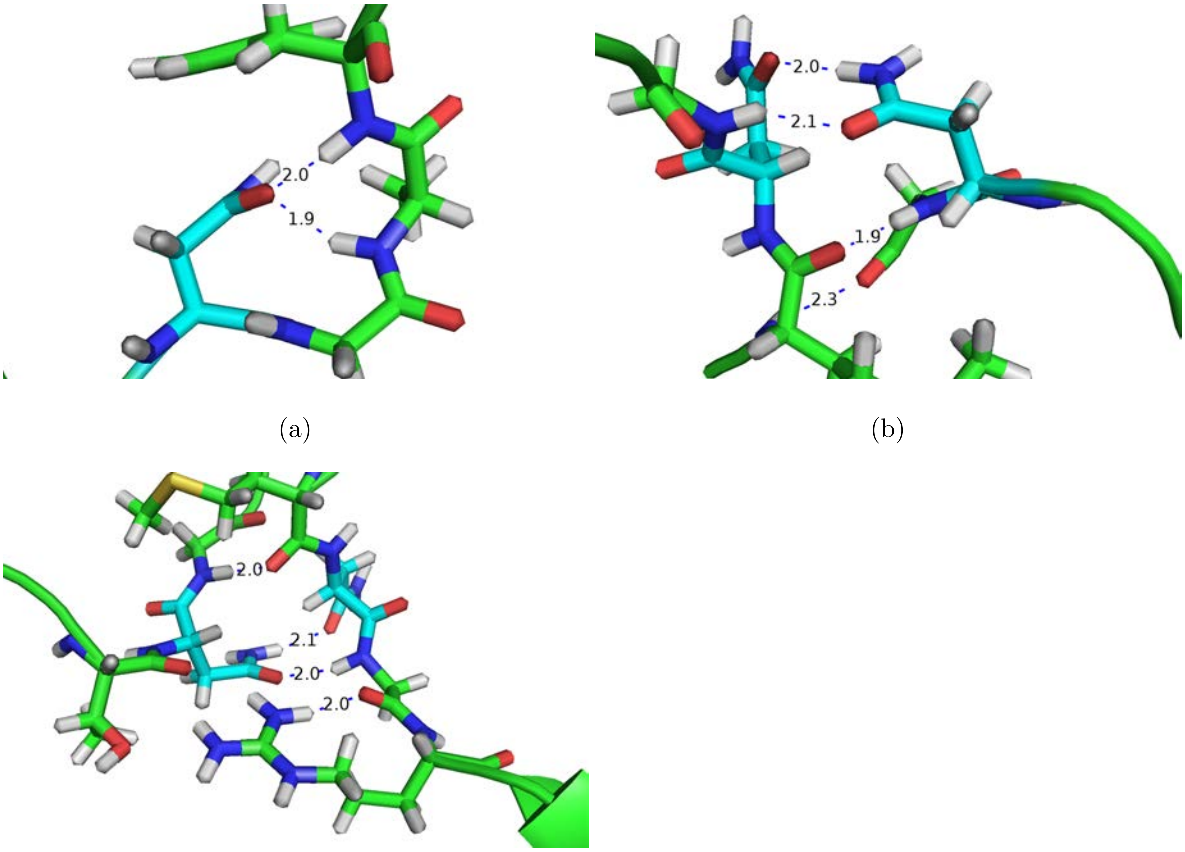
For all of the samples, the motion of loop3, containing Asn71, can be considered as the main factor that determines these interactions. When loop3 is interrelated with loop1, interactions between two monomers increase. In sample CC, loop3 of monomer A stays in the buried form. AsnA71@OD1 is bonded to GlyB16@H (Figure 4.9B), and also, AsnB15@OD1 is bonded to AsnA71@HD21 with significantly high percent

Table 4.2. Percent Occupancies of the Hydrogen Bonds of the Sample WOW.

<table>
<tr><th colspan="5">Asn OD1...Gly H interactions</th></tr>
<tr><th>acceptor</th><th>donor</th><th>% occupancy</th><th>distance(Å)</th><th>angle(deg)</th></tr>
<tr><td>AsnA15@OD1</td><td>GlyA16@H</td><td>8.21</td><td>2.933</td><td>44.73</td></tr>
<tr><td>AsnA71@OD1</td><td>GlyA72@H</td><td>6.42</td><td>3.040</td><td>44.87</td></tr>
<tr><td>AsnB15@OD1</td><td>GlyB16@H</td><td>5.79</td><td>2.958</td><td>46.50</td></tr>
<tr><th colspan="5">alternative interactions</th></tr>
<tr><th>acceptor</th><th>donor</th><th>% occupancy</th><th>distance(Å)</th><th>angle(deg)</th></tr>
<tr><td>AsnB71@OD1</td><td>AlaB73@H</td><td>47.78</td><td>3.013</td><td>29.20</td></tr>
<tr><td>TrpA12@O</td><td>GlyA16@H</td><td>20.94</td><td>3.260</td><td>45.40</td></tr>
<tr><td>TrpB12@O</td><td>GlyB16@H</td><td>15.70</td><td>3.244</td><td>39.42</td></tr>
<tr><td>MetA14@O</td><td>GlyB72@H</td><td>11.56</td><td>2.992</td><td>23.93</td></tr>
</table>

occupancies (76.19 and 67.83, respectively). These kinds of intramonomer interactions cannot be observed in the other samples in which loop3 goes to the solvated form. However, these interactions do not hinder the hydrogen bonds of interest. The occupancy of the hydrogen bond between AsnA71@OD1 and GlyA72@H is found as 10.28 and AsnB15@OD1 and GlyB16@H as 24.69 in the same sample. These percentages are found to be the highest among all of the samples for the computed occupancy. This fact can be explained as follows: when loop3 and loop1 are interrelated, the residues are less flexible so that the hydrogen bonds of interest are also more stable. From our simulations, we can distinguish three distinctive hydrogen bond acceptors for Gly@H within TPI: Thr70@O for the primary deamidation site, and Trp12@O and Met14@O for the second deamidation site. The former interaction occurs between the TPI monomers, while the latter interactions are intramonomeric.

4.4.2.3. Gly−H Interactions.. Analysis of hydrogen bonding patterns within TPI shows that hydrogen bonds between Asn side chain and its juxtaposed glycine backbone are not the preferential interaction during the MD simulations. Overall, simulations can



(a)

(b)

(c)

Figure 4.9. Stable hydrogen bonds ((a) WOW, AsnB71, 34 ns; (b) CC AsnA71 and AsnB15, 5 ns; (c) BASP AsnA71 and AsnB15, 34 ns).

be analyzed in terms of which kind of hydrogen bond Gly@H can make. Four states are possible for each TPI Gly@H: (i) hydrogen bond to the juxtaposed Asn@OD1 (Asn-bound state); (ii) hydrogen bound to another acceptor within TPI (protein-bound state); (iii) hydrogen bound to a water molecule (water-bound state); (iv) no hydrogen bond at all (free state). MD simulations have been analyzed for consecutive 1.5 ns windows, and percent probabilities for each four possible states have been collected. Evolutions of these probabilities during the WOW trajectory are reported in Figure 4.10 for each deamidation site Gly@H. The most significant information obtained from this analysis is that the free state is one of the most dominant states during the MD. One example for this fact is GlyA16. Figure 4.10 depicts that GlyA16 of WOW is desolvated after 12 ns and bonded to neither other residues nor to water up to the end of the simulation.

Figure 4.11A represents the environment around AsnA15 and GlyA16 for a snapshot at 34 ns of the WOW simulation. There, it can be seen that Gly@H points toward the inner part of the TPI but it does not make any hydrogen bond with TPI residues. The closest hydrogen acceptors are not correctly positioned in terms of distances and angles to form a stable hydrogen bond interaction. A similar free state is observed for the same snapshot for GlyB72 (Figure 4.11d): an interaction is formed with Met14@ SD that is too weak to be considered as a proton bound state.

Among all of the samples, GlyB16 of CC and GlyA72 of BASP are the only examples in which the probability of the protein- bound state is higher than the other states (see Appendix A). In CC, this state becomes more dominant after 20 ns. In BASP, the probability of having a protein-bond is high at the beginning of the simulation, and this interaction is lost after 42 ns.

The Asn-bound state cannot be considered as a dominant state in any of the samples.

Overall, analysis of the four possible states of Gly@H, as reported Figure 4.10, shows that Gly@H is mostly in a free state or in interaction with water molecules. To deamidate an Asn residue, Gly@H must not be involved in a stable interaction with a protein residue, other than the adjacent Asn@OD1. This is what we have found for all 4 Gly residues. This means that the proton of the glycine is mostly available during the time of the simulation to participitate to a deamidation reaction.

4.4.3. Near Attack Conformations (NAC)

In our previous studies [127–129], several models have been suggested to explain the mechanism of deamidation. In these studies, a small model was designed, and calculations have been carried out by using the B3LYP/6-31+G(d) methodology. The mechanisms depicted in Figure 4.2 can be classified as follows:

(i) Direct hydrolysis of asparagine to aspartate (Asn···Asp, black route in Figure

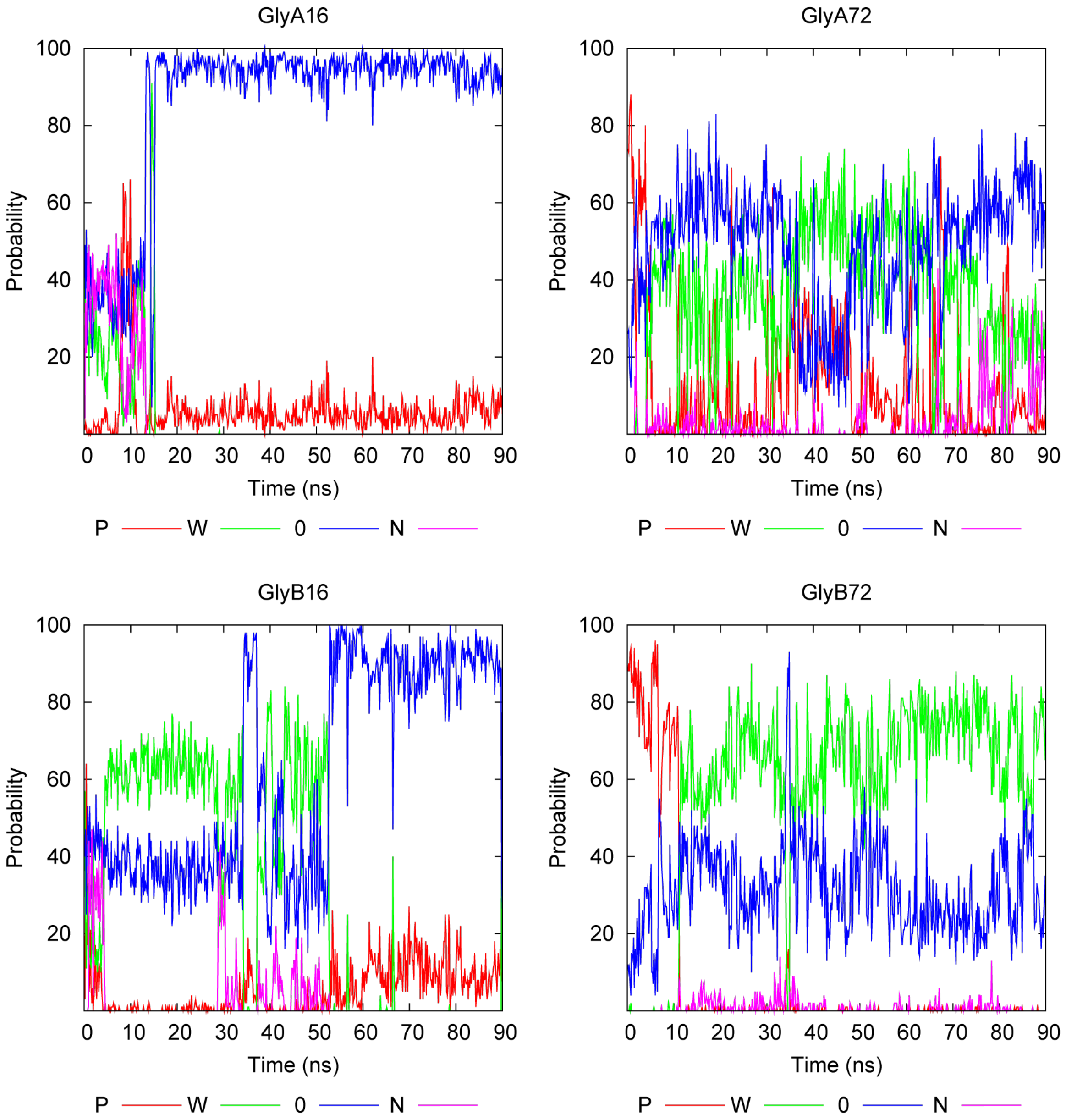


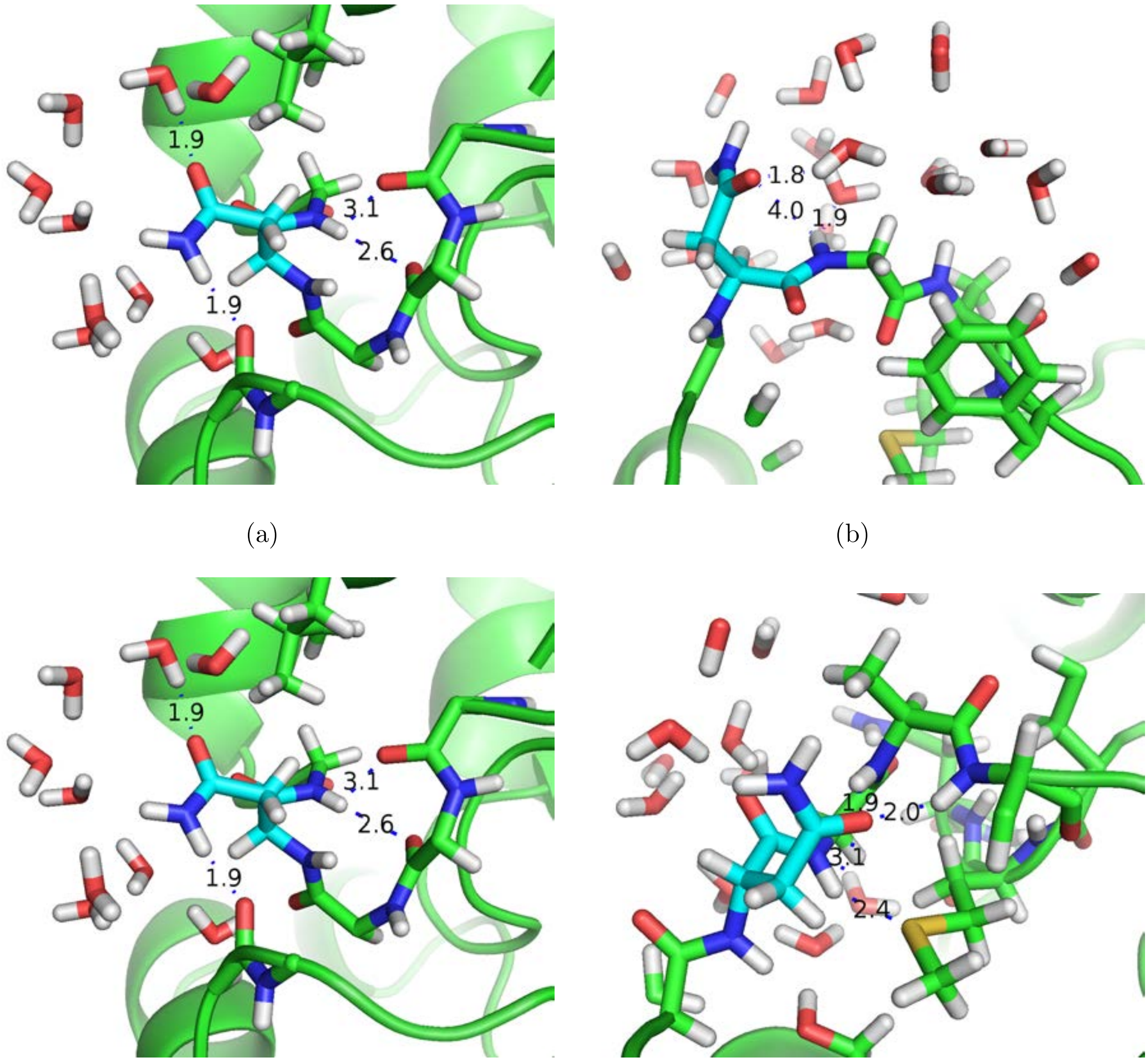
Figure 4.10. Probabilities of occupancy of Gly H of the sample WOW (red, green, blue, and magenta indicate protein-bound (P), water-bound (W), free (0), and Asn-bound (N) states, respectively).

4.2).

(ii) Succinimide-mediated deamidation of asparagine to aspartate (Asn···suc). This mechanism is suggested to proceed in three different ways.

  (a) One-step mechanism (Asn···suc, green route in Figure 4.2).

  (b) Two-step mechanism (Asn···tet···suc, blue route in Figure 4.2) via the for-



(a) (b)

(c) (d)

Figure 4.11. Gly H Interactions, WOW, at 34 ns. (a) GlyA16@H (distance between GlyA16@H and the closest H acceptor is 2.7 Å), (b) GlyA72@H, (c) GlyB16@H, and (d) GlyB72@H.

mation of a tetrahedral intermediate.

(c) Three-step mechanism where tautomerization of the Asn residue is followed by a tetrahedral intermediate (Asn···taut···tet···suc, red route in Figure 4.2).

The catalytic effect of water molecules has been taken into account in all of these mechanisms by introducing explicit water molecules (one or two) and the outcome of these studies can be summarized as follows:

Water assistance increases the rate of deamidation. The tautomerization route has the lowest barrier for the formation of the succinimide intermediate regardless of the number of water molecules that assist the reaction, including the waterless mechanism. Cyclization is the rate-determining step for succinimide formation in all water-assisted mechanisms.

For all mechanisms, intermediate and transition state structures have been determined.

One way to analyze our TPI simulations is to check if conformations of the deamidation sites geometrically closed to one of all possible reactant structures can be observed along the trajectories. From Figure 4.2 and References [127–129], one can extract near attack conformations (NAC) that are depicted in Figure 4.12. They correspond to the initial intermediates occurring for the four classes of all possible deamidation reactions (respectively, Asn···Asp, Asn···suc, Asn···tet···suc, and Asn···taut···tet···suc). In each class, reactions can be assisted by, respectively, 0, 1, or 2 water molecules, except for the direct hydrolysis mechanism Asn···Asp.

For comparison purposes, in addition to the TPI MD simulations mentioned above, an additional MD simulation has been carried out for a pentapeptide Gly−Gly−Asn−Gly−Gly in a water box environment during 100 ns, using a similar simulation protocol as for the TPI MD simulations.

All trajectories are analyzed in terms of geometrical criteria: intermolecular distances, angles, etc. (see Appendix A for details), and if occurring, near-attack conformers are identified and assigned to one of the corresponding reactant structures. Figure 4.14 reports the percent occurrences of favorable interactions during the different simulations.

Previous QM studies on model peptide have evaluated the energy barriers for such reactions [129]. For the nonassisted tetrahedral intermediate formation, the barrier is estimated at 49.7 kcal/mol, while the two water-assisted tautomerization routes have

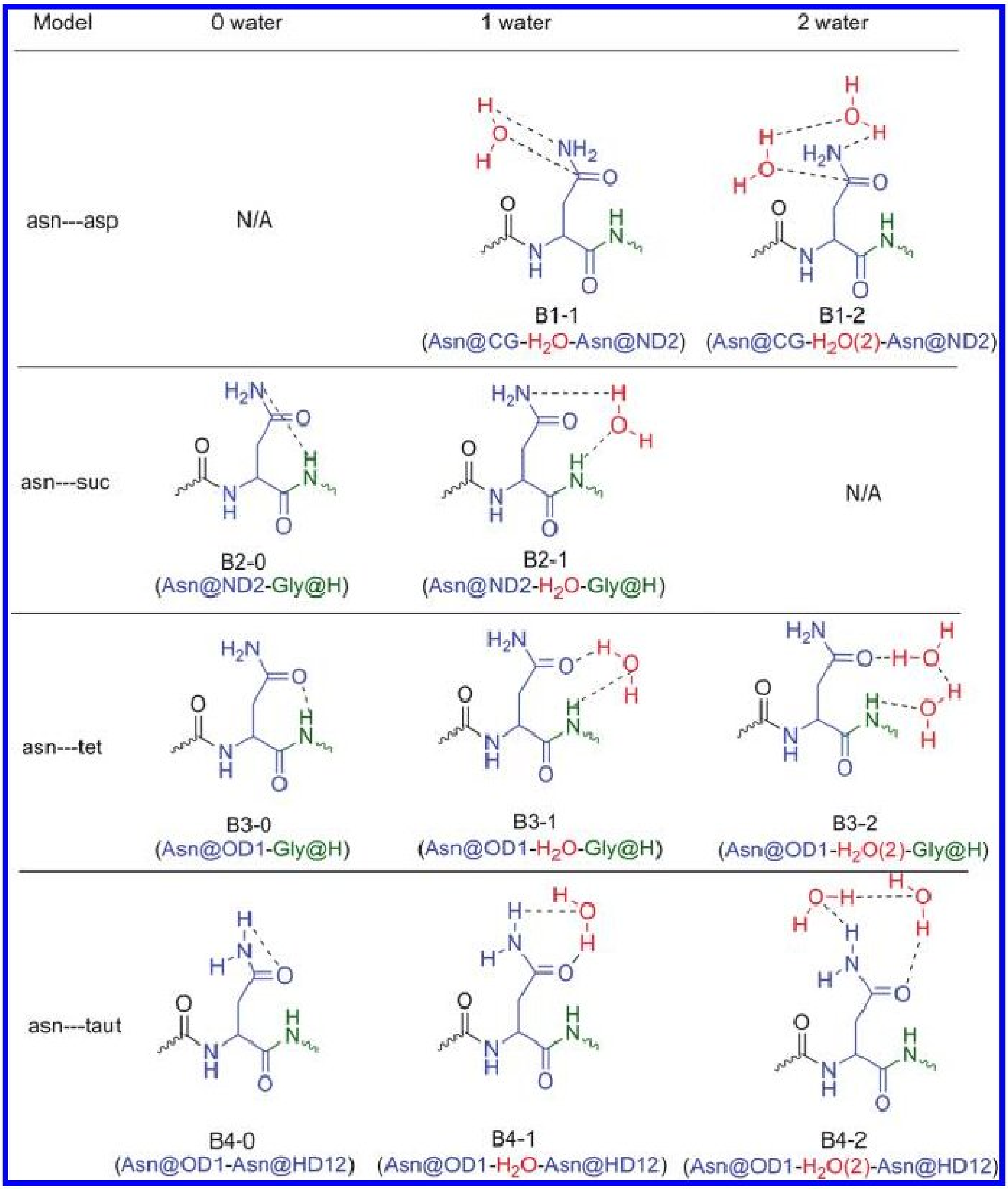


Figure 4.12. Representations of Models Suggested by QM Studies Asn, Gly, and Water Molecules Are Shown in Blue, Green, and Red, respectively.

two consecutive barriers: respectively, 18.1 kcal/mol to form the tautomer and 39.7 kcal/mol to continue up to the tetrahedral intermediates [129].

In all of the samples, there is nearly no occurrence for the near- attack conformation corresponding to the direct hydrolysis, and the direct formation of a succinimide intermediate. This supports the finding of the previous QM studies: not only do these reactions have higher barriers than the one corresponding to the water-assisted tetrahedral intermediate formation or the water-assisted Asn side chain tautomerization, respectively, but the conformation corresponding to geometries close to the initial reactant states of these difficult reactions do not occur frequently. As a consequence, in TPI, direct hydrolysis and direct succinimide formation are two model deamidation mechanisms that can be ruled out.

In addition, all TPI simulations reveal that NACs occur mostly for two reaction mechanisms: the direct formation of a tetrahedral intermediate (B3−0) and the Asn side chain tautomerization assisted by two water molecules. This suggests that the experimentally observed succinimide intermediate is formed after the formation of a metastable tetrahedral intermediate that can be obtained either directly or through a prior water-assisted tautomerization route (see Figure 4.2).

Finally, Figure 4.14 can be analyzed in terms of differences between primary and secondary deamidation sites. We have not found any obvious pattern differentiating these two sites. At first glance, our results do not suggest that one site could initiate deamidation faster than another by forming more preferentially NACs, as compared to the other deamidation sites.

4.5. Conclusions

We have performed molecular dynamics simulations of six possible samples for triosephosphate isomerase: wild-type protein extracted from crystallographic structure (WOW); substrate-bound system (CS); modeled structures to represent closed and open states of TPI (respectively, CC and OO); mutated simulations where the

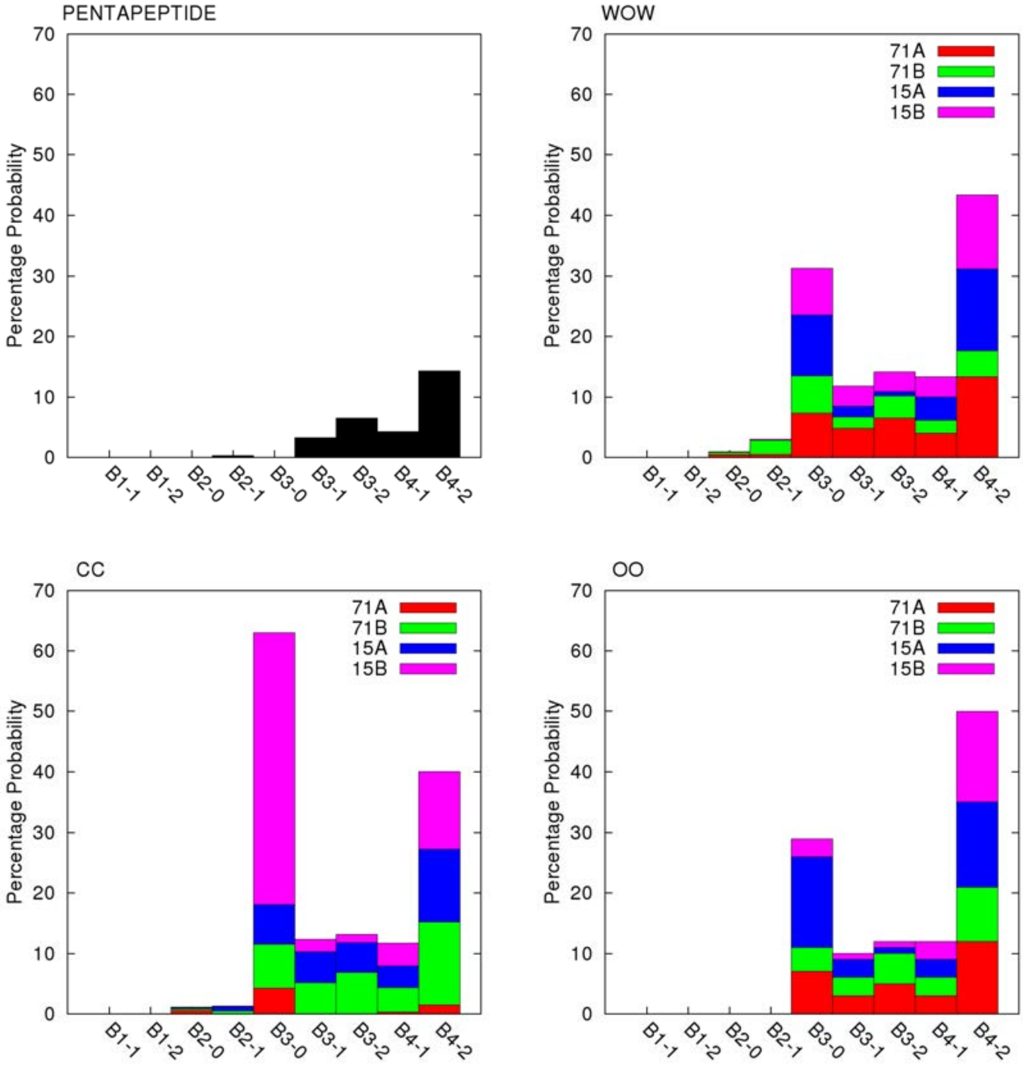


Figure 4.13. Percent occurrence of the interactions during the simulation. (Number of water molecules within the hydrogen bond network is shown as −0, −1, and −2 for the corresponding model.).

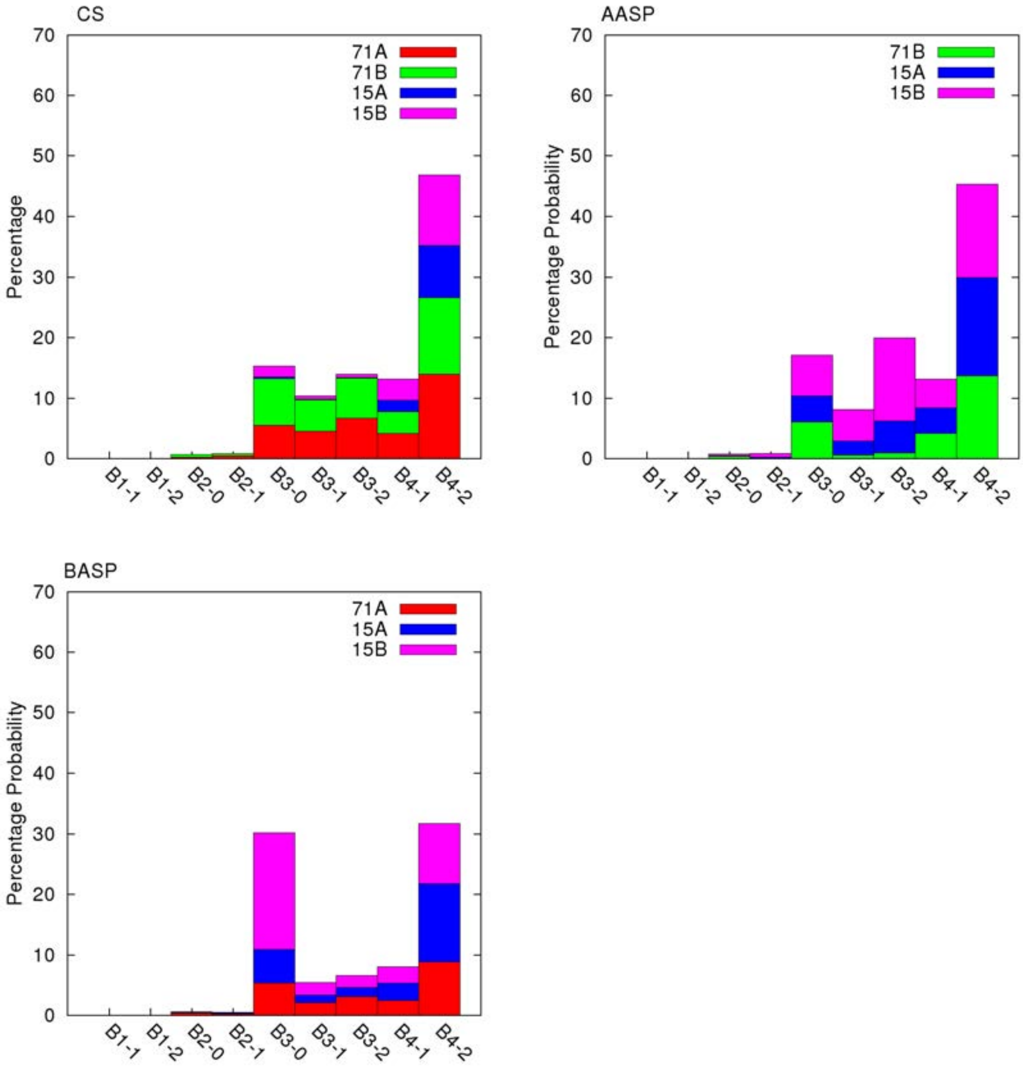


Figure 4.14. Percent occurrence of the interactions during the simulation. (Number of water molecules within the hydrogen bond network is shown as −0, −1, and −2 for the corresponding model.) (cont.).

asparagine of the primary deamidation site has been replaced by aspartate residue on each TPI monomer (respectively, AASP and BASP).

Analysis of the trajectories has been performed in terms of (i) global and local desolvation states for each deamidation site; 6298 (ii) hydrogen bond patterns; and (iii) near attack conformations. By using NAC analysis, we have found that only two initiations of the deamidation reaction seem possible in TPI: (i) a mechanism starting with the tautomerization of Asn side chain assisted by two water molecules; (ii) a mechanism involving the direct formation of the tetrahedral intermediate.

In each case, the first intermediate formed by the reaction will be a metastable intermediate according to the relative energies [127–129] reported by Catak et al. However, QM modeling have found that the tautomerization mechanism involved a lower TS barrier than the direct tetrahedral intermediate formation. Because NAC probabilities for the two mechanisms have been found similar, we can hypothesize that the kinetics involving the formation of the tautomer interaction will be faster than the one leading to the direct tetrahedral intermediate formation.

Comparison with a MD simulation of the Gly−Gly−Asn− Gly−Gly pentapeptide shows that NAC probabilities related to the tautomerization reaction as well as the tetrahedral inter- mediate formation are similar for TPI and the pentapeptide.

However, Gly@H solvent exposures are very different between the pentapeptide and the enzyme. This means that, after the tautomer or the tetrahedral intermediate has been formed, the NH group of the glycine in the pentapeptide will be more easily involved in the deamidation reaction to form either a tetrahedral intermediate or a succinimide intermediate. The metastable intermediate has more chance to go backward to Asn in TPI, whereas it will more likely go in the forward direction to form the succinimide intermediate in Gly−Gly−Asn−Gly−Gly. This should explain why experimentally TPI deamidates slower than the pentapeptide.

Experimental studies indicated that Asn71 deamidates faster than Asn15 [124].

NAC analyses show that Asn71Gly72 and Asn15Gly16 have almost the same probabilities to be found as a reactive conformer. However, NAC analyses have an interest only in the initiation of the deamidation reaction. To fully accomplish the deamidation, two main steps are required: (i) removal of Gly@H to obtain a succinimide intermediate and (ii) hydrolysis of succinimide intermediate to aspartate (Figure 4.2). Assistance of water is crucial in both steps because of particular reasons. In the first step, water facilities the reaction as suggested by QM studies. In the second one, the presence of water is mandatory as a reactant. When these facts are considered all together, it can be concluded that more solvated residues should undergo deamidation easier. Desolvation results show that Asn71Gly72 always has higher probability to be in water environment. So, we suggest that Asn71Gly72 will deamidate faster than Asn15Gly16 because of the differences in desolvation results, even though they both can accomplish the reactive conformers.

Another experimental finding is that deamidation of Asn71 is a prerequisite of the deamidation of Asn15 [124]. Analyses of the trajectories related to N71D mutation in TPI do not lead to clear findings. In the time scale of our simulation, we have not found strong effect of the presence of charged aspartate residue on the structural network around its juxtaposed Asn15. NAC analyses in all the trajectories also show that probabilities for both residues are almost the same.

One of the ultimate goals of the project was to analyze the relationship between the catalytic activity and the motion of the active site loop (loop6) with the deamidation. We have not found any clear evidence of the influence of the loop6 and the substrate at the time scale of the simulations. Although different behaviors are observed in each of the samples, these behaviors cannot be related to the motion of loop6.

5. WHY ASN71 DEAMIDATES FASTER THAN ASN15 IN THE ENZYME TRIOSE PHOSPHATE ISOMERASE? ANSWERS FROM μs MOLECULAR DYNAMICS SIMULATIONS

Deamidation is the uncatalyzed process by which asparagine or glutamine can be transformed into aspartic acid or glutamic acid, respectively. In its active homodimeric form, mammalian Triosephosphate Isomerase (TPI) contains two deamidation sites per monomer. Experimental evidence shows that the primary deamidation site (Asn71-Gly72) deamidates faster than the secondary deamidation site (Asn15-Gly16). To understand this phenomenon, we have performed GPU-enabled micro-second long molecular dynamics simulations of rabbit TPI as well as chicken TPI, which does not possess any primary deamidation site. Analysis of the simulations has been performed using independent global and local descriptors that can influence deamidation rates: desolvation effects, backbone acidity, and side chain conformations. Our findings show that all descriptors add up to favor the primary deamidation site over the secondary one in mammalian TPI: Asn71 deamidates faster because it is more solvent accessible, adjacent glycine NH backbone acidity is enhanced, and Asn side chain has a preferential near attack conformation. Our findings also shed light the effect of high order structure on deamidation.

5.1. Introduction

Deamidation of asparagine (Asn) is the reversible conversion of neutral amide side chain to negatively charged carboxylic acid, yielding aspartic acid residue (Asp). This conversion in the aminoacid sequence causes destabilization of peptide structures, which enhances chemical degradation of peptides and limits the lifetime of proteins [14]. The rate of deamidation of asparagine is mostly dependent on the primary structure of the peptides, and reaction with asparagine-glycine (Gly) sequence is known to be the most rapid conversion [105]. Under physiological conditions, deamidation occurs via succinimide intermediate, and this formation is the rate determining step of the overall reaction (Figure 5.1) [112]. As shown in the schematic representation of the suggested mechanism (Figure 5.1), formation of succinimide intermediate is initiated by deprotonation of backbone amide nitrogen of glycine (asn-gly(-)). Nucleophilic attack of backbone nitrogen to asparagine side chain produces a metastable cyclic tetrahedral intermediate (tet). This intermediate is protonated resulting in the release of one ammonia molecule. This step generates relatively stable succinimide (suc) intermediate which is converted to asparagine (asp-gly) in further hydrolysis. Because of the biological importance of deamidation reaction, there have been several studies in order to enlighten the details of the mentioned mechanism [8, 120, 121, 127, 129, 138–140]. All of these studies concluded that the reaction should start via removal of glycine backbone amide hydrogen. In addition, Catak et al. [129], emphasized the importance of solvation in deamidation, demonstrating that, water assistance enhances this phenomenon. This observation is in agreement with the experimental results showing that deamidation is prevented in low water concentration media [141].

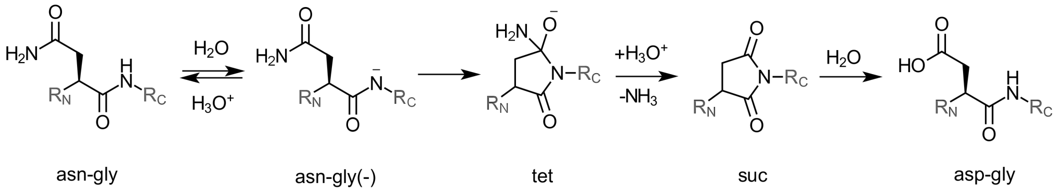


Figure 5.1. Suggested Deamidation Mechanism by Capasso and coworkers [126, 142].

Capasso et al. proved that the reaction mechanism of deamidation bears the same features in small peptides and proteins [3]. However, it is known that it occurs much more faster in peptides than proteins [14]. Besides, there is a wide range of deamidation half life's of several proteins which bear the same primary structure [12]. The main hypothesis on the effect of higher order sequence on deamidation focuses on two facts: (i) Decrease of acidity of glycine backbone amide would diminish deprotonation, (ii) Lose of flexibility, comparing to peptides, would prevent cyclization [12, 105]. In our previous study [143], in which the initiation of deamidation in triosephosphate isomerase (TPI) has been investigated by molecular dynamics, it has been suggested that global desolvation generated by secondary structure would affect the rate of deamidation. The effect of global desolvation on the rate of deamidation can be explained firstly by its relevance with acidity and flexibility. As the degree of desolvation increases, the acidity of backbone amide and flexibility of the aminoacid would decrease [137, 143]. Secondly, more desolvated residues would have less water in the first hydration shell that would diminish the rate of deamidation considering the enhancing effect of water in deamidation.

In the present study, deamidation of the enzyme triosephosphate isomerase (TPI) is elucidated. It has been proposed that deamidation in TPI regulates protein turnover, based on the fact that it is controlled by catalytic activity and it causes protein degradation [14]. Two distinct asparagines were found to be unstable under physiological conditions in each monomer of homodimeric TPI: Asn15 and Asn71 [14, 124]. It has been suggested that deamidation in TPI occurs in a sequential order in a way that deamidation of Asn71 triggers Asn15 deamidation [14]. This interrelated deamidation in TPI enables us to compare the significance of desolvation, backbone acidity, and flexibility on this reaction. The special features of deamidation in TPI can be summarized as follows:

(i) In mammalian TPI, deamidation occurs at MetAsn(15)GlyArg and ThrAsn(71)Gly Ala, which are located at the interdigitating loops (loop1 and loop3, respectively) of the two identical monomers of this homodimeric enzyme [124].

(ii) ThrAsn(71)GlyAla deamidates earlier than MetAsn(15)GlyArg, with the relative ratios of deamidated products as Asp(71):Asp(15) 2.54:1 for human, and 1.73:1 for rabbit TPI [124]. Asn71 and Asn15 are respectively depicted as primary (1°) and secondary (2°) deamidation sites in TPI.

(iii) Mammalian TPI is active only as a dimer. There are four deamidation sites in an active homodimeric TPI, and they are located in a way that Asn(71)Gly (1° deamidation site) of one monomer is juxtaposed to Asn(15)Gly (2° deamidation site) of the other monomer, and vice versa (Figure 5.2).

(iv) The rate of deamidation of Asn(71) is enhanced by the increased catalytic activity of mammalian TPI [124].

(v) Deamidation in TPI follows general base catalysis mechanism [41].

(vi) Chicken TPI bears Asn(15)Gly, but Asn(71) is replaced by Ser(71). This type of TPI does not undergo deamidation under mildly alkaline conditions, whereas mammalian TPI deamidates rapidly under this circumstances [144]. Based on this fact, and also on the experiments on hybrid TPI, it has been proposed that deamidated Asn(71) (Asp(71)) would be a prerequisite for Asn(15) deamidation [14].

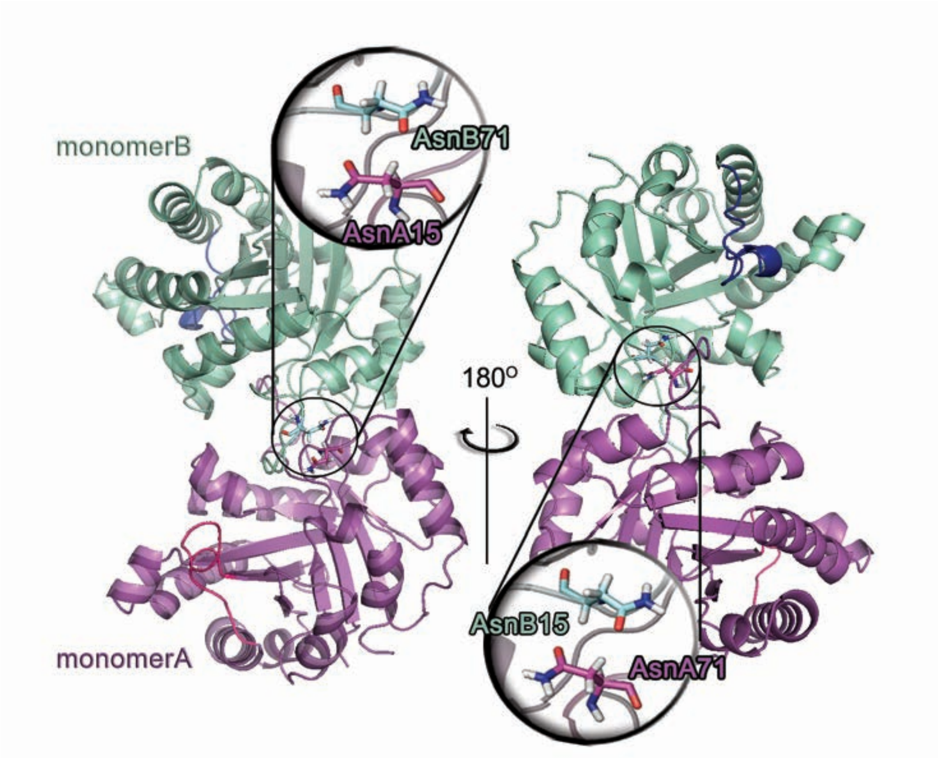


Figure 5.2. Locations of Asn(71) (1° deamidation site) and Asn(15) (2° deamidation site) on mammalian TPI crystal structure (1R2R, [130]).

Here, two distinct TPI, apo mammalian TPI and chicken TPI, were simulated by classical force field and molecular dynamics (MD) techniques in order to explain the experimental findings on TPI deamidation, and also to determine the predominance of the desolvation, acidity, and flexibility effects on deamidation. MD simulations were performed up to 1.5 μs using cuda-enabled graphics processing units (GPUs). To our knowledge, these simulations are the longest MD simulations applied on TPI. We discuss the impact of simulation time, while monitoring major changes on the analysis of a system with a considerable size. In addition, quantum mechanical (QM) calculations were performed to gain insights of backbone amide acidity of glycine.

5.2. Computational Details

5.2.1. Preparation of the Samples

Mammalian TPI structure was extracted from rabbit TPI crystal structure (PDB code 1R2R, resolution 1.5 Å [130]). This protein consists of two identical monomers, which are differentiated in the conformation of the active site. The monomers are labeled as monomer A and B, and this nomenclature is used in order to distinguish the amino acids located on either of these monomers (AsnA71, GlyB72, etc.). The crystal structure of the chicken TPI (PDB code 8TIM, resolution 2.5 Å) consists of two identical monomers as its mammalian counterpart. These systems, mammalian and chicken TPI are named as WOW, and CHI, respectively.

The systems were hydrogenated following the validation of protonation states of charged side chains with the PROPKA program [137, 145–147] and histidine residues were protonated according to their polar environment. tLEAP module of AMBER 12 [42] was used to build topology and coordinate files. The systems were solvated with explicit TIP3P [48] water molecules resulting in cubic boxes with the edge length of 93.13 Å. These systems contain about 24,000 water molecules (for a total of ∼80,000 atoms), with an initial density of 1.0 and have a protein concentration in each box of about half TPI crystal.

5.2.2. Molecular Dynamics Simulations

The simulations were performed using AMBER 12 [42] program package with pmemd module, and for the production runs the cuda-enabled graphics processing units (GPUs) version of pmemd was used [148,149]. All the simulations were performed using ff03 force field parameters [134]. Periodic boundary conditions were applied and an NVT ensemble was used with Andersen [47] temperature coupling. The particle mesh ewald (PME) [135] method was used to calculate long range electrostatic interactions within a cutoff distance of 8 Å. Only the bonds involving hydrogen atom were constrained with the SHAKE [136] algorithm. Time step was 2 fs for both equilibration and production. The equilibrations have been done in 5 stages: (i) To provide a proper geometry of hydrogen atoms, all the heavy atoms including water oxygen are restrained with a harmonic potential of 50 kcal/molÅ^2 for 100.0 ps, at 10K. (ii) The same potential was applied for another 100.0 ps, but removing the restraint on the water oxygens in order to ensure optimized positions of water with respect to protein environment. (iii) This last step was repeated by decreasing the harmonic potential to 5 kcal/molÅ^2 for 100.0 ps. (iv) The potential was removed for another 100.0 ps at 10K to equilibrate whole system. (v) The system was gradually heated to 300K for 2000.0 ps. The velocities were randomly updated every 10 steps for the equilibration stages 1 to 4 and every 100 for stage v. The production runs were performed for 1.5 μs with velocities updated every 1000 steps.

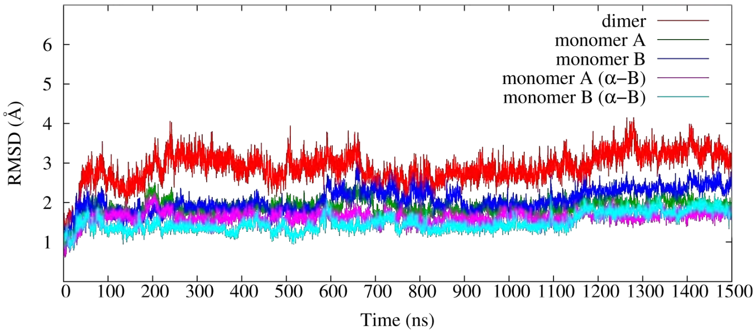
5.2.3. Quantum Mechanical Calculations

Geometry optimizations were performed with Gaussian09 software [150], using the density functional theory (DFT) at the B3LYP/6-31+G* level. Geometries of stationary points were optimized with constraints on two dihedrals of interest as detailed below. Polar environment for aqueous phase calculations were provided by Conductor-Like Polarizable Continuum Model (CPCM, $\varepsilon = 78.5$ [87, 88]). Energy values of gas and aqueous phase calculations include thermal enthalpy, and free energy corrections at 298.15 K and 1 atm.

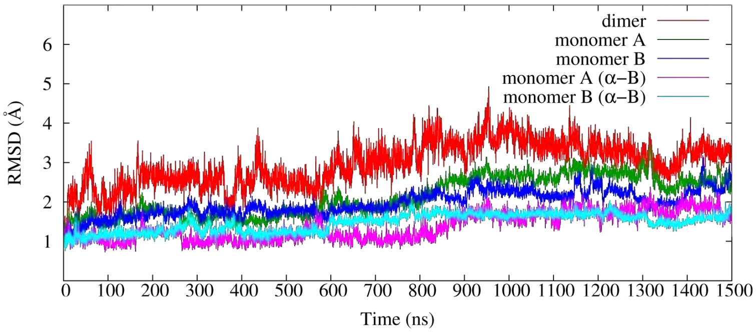
5.3. Results

5.3.1. Stability of the Simulations

Root mean square deviations (RMSD) of the Cα atom of (i) dimer, (ii) each monomer and (iii) α helices + β sheets of each monomer were monitored along the simulation with respect to reference crystal structure. Deviations are much more noticeable considering the RMSD of the dimer, compared to deviations on separated monomers for all of the samples (Figure 5.3). This result is in correlation with our previous molecular dynamics simulation analysis on six different TPI structures, which have a shorter time scale (60-90 ns) [143]. The motions of monomers with respect to each other in dimeric TPI enzyme, which was observed in other studies in literature [24], explains rather high RMSD values of the dimer. On the other hand, deviations of each monomer holds low RMSD value along the trajectory at 2.0 - 2.5 Å (Figure 5.3). Combining these remarks, we can conclude that our simulations are stable.



(a) WOW



(b) CHI

Figure 5.3. Variations of the rmsd (in Å) along the MD trajectory. Rmsd are computed for the complete dimer structure (red), monomer A only (green), monomer B only (blue), all α helices and β sheets of the subunit A (magenta), and all α helices and β sheets of the subunit B (cyan).

5.3.2. Solvent Accessibility

As it has been demonstrated previously, "global desolvation ($N_{15.5Å}$)" can be used as an efficient descriptor to investigate solvent accessibility of the aminoacids of interest

[143]. This descriptor was introduced by Li et al. [137] in order to predict the pK_a shift with respect to global desolvation of the aminoacids in protein environment, and it has been implemented in the previous versions of PROPKA program [137,145,147]. Global desolvation ($N_{15.5Å}$) is basically the measure of the non-hydrogenous protein atoms within a 15.5Å radius sphere centered on the ionizable atom of aminoacid of interest. The philosophy behind is that, the amount of protein atoms close to one aminoacid is inversely proportional to the amount of water molecules around, since the water molecules will be replaced by protein atoms or vice versa. Thus, as the number of protein atoms around a residue increases, it becomes less solvated than its free aminoacid form in aqueous solution. The degree of desolvation can be classified with the cutoff 400 (again as Li et al. suggest [137]), and the aminoacids having $N_{15.5Å}$ value greater than 400 are classified as buried while others are solvated.

Buried residues often hold higher pK_a 's compared to their solvated counterparts. As the degree of desolvation increases, the relative free energy of the negatively charged species rises, so that acid-base equilibrium shifts to neutral form. In this study, the global desolvation analysis was performed on Asn side chain and Gly backbone, first, to investigate the solvent exposure on Asn and Gly, which enhances deamidation reaction. Secondly, global desolvation analysis on Gly provides information on the pK_a shift of backbone amide: while the degree of burial increases, pK_a will be higher, and so the rate of deprotonation would be slower. Global desolvation analysis for both Asn side chain and Gly backbone was performed using the same protocol used in our previous study [143], and will not be detailed here.

$N_{15.5Å}$ of Asn side chain of all of the deamidation sites was monitored within simulation time (Figure 5.4), and Asn side chain desolvation with respect to Gly backbone desolvation were plotted for the sample WOW (Figure 5.5). In order to gain insights for populations of solvated residues along the simulation time, frames with Asn or Gly having $N_{15.5Å}$ values smaller than 400 were identified, and considered as solvated. Number of frames having solvated residues were summed up, and divided by the total number of frames in the simulation to obtain percent occurrence of "$N_{15.5Å}$ <400". (Table 5.1).

Side chain of Asn of 1° deamidation site on both monomers go to solvated state at the beginning of the simulation (Figure 5.4). They stay solvated along the simulation (Figure 5.5) resulting in percent occurrences of "$N_{15.5Å}$ <400" equals to 84.7 % and 98.0 % for monomer A and B, respectively (Table 5.1). Backbone of GlyA and GlyB of 1° deamidation site, have probability of 19.0 % and 96.2 % to be solvated.

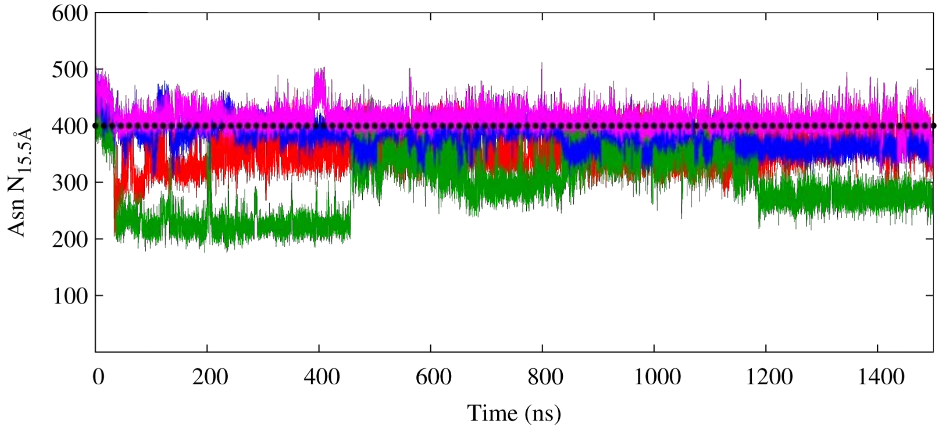
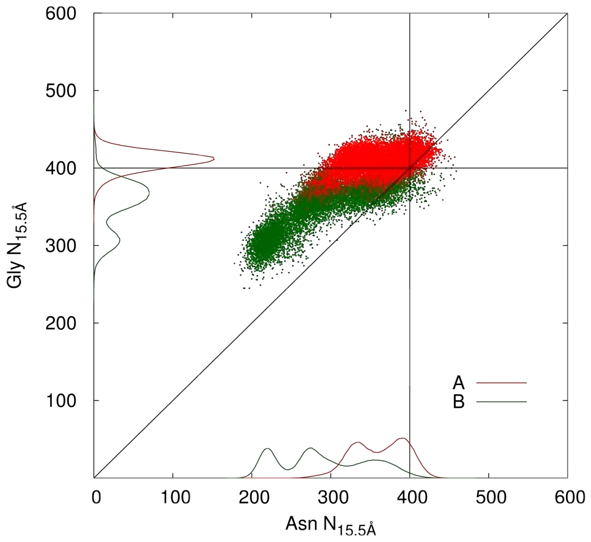


Figure 5.4. $N_{15.5Å}$ of Asn side chain with respect to time (Asn15A, 71A, 15B, and 71B are depicted in blue, red, magenta, and green, respectively).

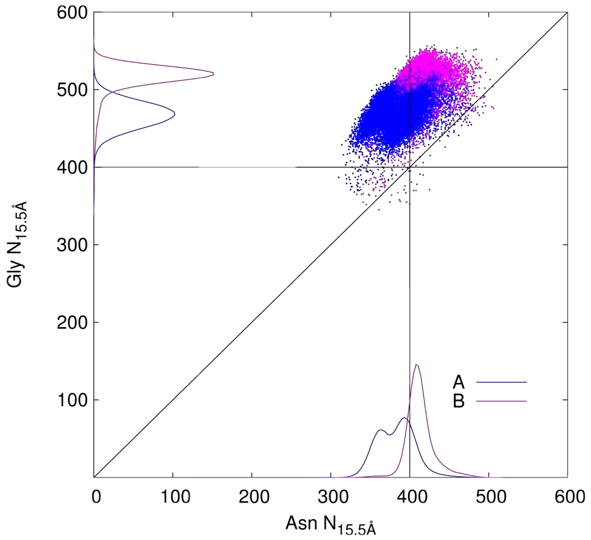
In 2° deamidation site, AsnA and AsnB are solvated 74.4 % and 15.3 % of the simulation. GlyA and GlyB of 2° deamidation site stay always buried along the simulation (Figure 5.5) with probability to be solvated almost zero (Table 5.1).

In chicken TPI (CHI), AsnA15 and AsnB15 have probabilities to be solvated about 58.1 % and 18.3 %. These values are inbetween the solvation results of Asn's of WOW in 1° and 2° deamidation sites. GlyA16 and GlyB16 in CHI can also be considered as buried since the solvation probabilities are around 10 %.

The differences of percent solvation of Asn in 2° deamidation site on monomer A (74.4 %) and B (15.3 %) is notable. The same kind of difference can be observed for Gly in 1° deamidation site with the solvation values: GlyA72 = 19.0 % and GlyB72 = 96.2 %. The question that arises from these results is, whether this large gap between the same residue on different monomers is a realistic picture indicating strongly different behavior of the two monomers, or it is just a matter of the way of defining



(a) 1° deamidation site (red=A, green=B)



(b) 2° deamidation site (blue=A, magenta=B)

Figure 5.5. $N_{15.5Å}$ of Asn side chain vs $N_{15.5Å}$ of Gly backbone. Red and green points represent 1° deamidation site of monomer A and B, respectively. Blue and magenta points represent 2° deamidation site of monomer A and B, respectively. The probability distribution of $N_{15.5Å}$ of each deamidation site is shown with corresponding colors.

the descriptor of solvation. From Figure 5.5, it can be seen that both GlyA72 (a, red, y-axis) and AsnB15 (b, magenta, x-axis) gather on the border of solvation limit (400). The solvation limit 400 has been chosen to describe the global desolvation of residues as it is proposed by Li et al. However, considering the observed specific characteristics of TPI here we suggest to use 420 as a limit to define solvated and desolvated frames. The updated solvation probabilities ("$N_{15.5\text{Å}}$ <420") of GlyA72 and AsnB15 are 77.6 % and 71.4 %, that are much more close to their counterparts in other monomers GlyB72 = 98.6 %, AsnA15 = 94.3 % (Table 5.1). Either choosing 400 or 420 as a limit of desolvation, the overall trend of solvation does not change, and can be summarized as follows:

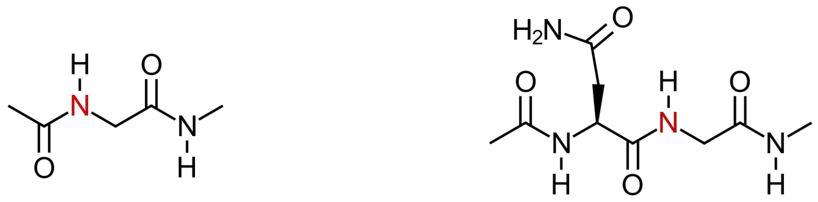
(i) Asn's in 1° deamidation site on both of the monomers are always solvated along the simulation.
(ii) Gly's in 1° deamidation site are also solvated, but having smaller probabilities compared to their Asn neighbors.
(iii) Asn's in 2° deamidation site are solvated but less than Asn's in 1° deamidation site.
(iv) Gly's in 2° deamidation site stay always buried.

Table 5.1. Percent occurrence of solvated conformations of Asn and Gly in 1° and 2° deamidation site (DS) of monomer A and B.

		1° DS		2° DS		1° DS		2° DS	
		Asn	Gly	Asn	Gly	Asn	Gly	Asn	Gly
WOW	A	84.7	19.0	74.4	0.4	97.8	77.6	94.3	1.0
	B	98.0	96.2	15.3	0.4	99.7	98.6	71.4	0.7
CHI	A			58.1	11.2			76.2	20.9
	B			18.3	10.3			33.2	23.6
		$N_{15.5\text{Å}}$ <400				**$N_{15.5\text{Å}}$ <420**			

5.3.3. Backbone Amide Acidity

5.3.3.1. Quantum Mechanical Calculations. Radkiewicz et al. [8], showed that acidity of Gly backbone amide hydrogen is highly correlated with peptide conformation by calculating the relative proton affinity of N-formyl-glycinamide as a function of Ramachandran plot (Φ and Ψ angles) (HF/6-31+G*//HF/3-21G). In the present study, a similar analysis was performed with a relatively larger model system (2-acetamido-N-methylacetamide, Figure 5.6, M1) using optimized structures with a higher basis set, DFT (B3LYP/6-31+G) (Figure 5.7 a). In addition to that, the change of Gibbs free energy in aqueous media (ΔG_{aq}) with respect to Ramachandran plot has been computed for the same model system using implicit water models (B3LYP/6-31+G (CPCM)) (Figure 5.7 b).



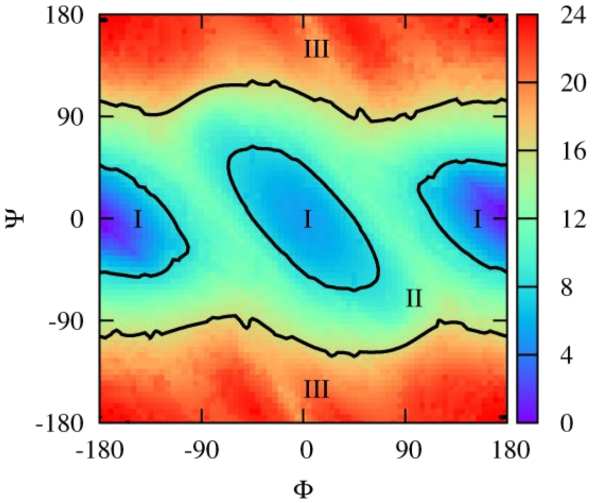
2-acetamido-N-methylacetamide
(M1)

2-acetamido-N'-(2-oxopropyl)malonamid
(M2)

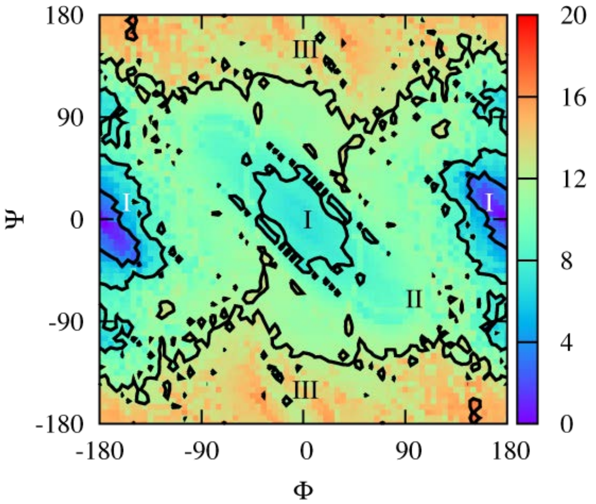
Figure 5.6. Model peptides for QM calculations without (M1) or with (M2) asparagine side chain. Backbone amide nitrogen is colored by red.

Our proton affinitiy results update findings of Radkiewicz et al. with some minor shifts. Either relative proton affinity (PA) of N-formyl-glycinamide (HF/6-31+G*//HF/3-21G) [8], or PA of 2-acetamido-N-methylacetamide (M1) (B3LYP/6-31+G //B3LYP/6-31+G) suggest that: (i) PA of Gly backbone amide deviates up to ∼24 kcal/mol with respect to changes in backbone conformation, (ii) Acidity decreases going from region I to III (Figure 5.7 a.), (iii) PA is mainly dependent on Ψ angle rather than Φ.

ΔG_{aq} deviations with respect to Ramachandran plot of M1, resembles the similar characteristics of gas phase proton affinitiy results (Figure 5.7 b.). The relatively most acidic conformations are observed in region I, while region III represents relatively more basic ones. $\Delta\Delta G_{aq}$ changes up to 17 kcal/mol on the Ψ axis, and 5 kcal/mol on the Φ axis indicating that acidity is more dependent to Ψ variations.



(a) ΔPA (kcal/mol)



(b) $\Delta\Delta G_{aq}$ (kcal/mol)

Figure 5.7. Correlation of acidity of amide backbone hydrogen of model M1 as a function of Ramachandran plot at B3LYP/6-31+G (CPCM). Contour lines divide the surfaces to three regions with an increment of 8 kcal/mol and 5.6 kcal/mol energy difference for ΔPA (kcal/mol) and $\Delta\Delta G_{aq}$. Acidity decreases going from blue to red.

Radkiewicz et al. [8] used the strong dependency of conformation and backbone amide acidity to explain the relative ease of deamidation in peptides which have Gly next to Asn, compared to other neighboring aminoacids [8]. They concluded that [8], the extra flexibility of glycine allows the system to occupy more acidic conformations, and consequently increases the rate of deamidation. Here we investigate if this dependency can also be used to explain the different rates of Asn15Gly and Asn71Gly in

TPI, since flexibility of glycine will be limited within a protein and might differentiate with respect to primary and secondary structure of the protein (i.e. Asn15Gly, and Asn71Gly are located on loop1 and loop3 in TPI, respectively).

Before proceeding on the analysis of TPI deamidation, the accuracy of PA and ΔG_{aq} analysis obtained by model M1 (Figure 5.6) should be validated. This model is limited to describe only Gly backbone. A similar analysis should be performed on a larger system which contains an asparagine residue in order to investigate side chain effect on acidity of glycine backbone amide (Model M2, Figure 5.6). Considering the computational cost of such an analysis, relative proton affinity and ΔG_{aq} of only 16 critical points (from most acidic to most basic) were calculated using the same QM level (Table 5.2). The best conformation of the asparagine side chain is adjusted in accordance with previous QM studies [127, 129] for the minimum structure only (Φ,Ψ = (-180.0°, +0.0°)). Rotations around Φ and Ψ angles are done on this minimum structure to constitute these 16 critical points.

Both ΔPA and $\Delta\Delta G_{aq}$ values of M2 are smaller than the values of M1 for each 16 points, concluding that the system is better stabilized by asparagine side chain interactions (Table 5.2).

The relative PA of M1 as a function of Ramachandran plot is highly symmetrical, because of the lack of a stereocenter in this molecule. The high symmetry of ΔPA of M1 is generated as a consequence of the isoenergetic points on Ramachandran plot. Among all the 16 critical points chosen to calculate with M2, the following points are isoenergetic on ΔPA of M1: Φ, Ψ = (+90.0°, -90.0°) $\leftrightarrow$ (-90.0°, +90.0°), (+90.0°, +0.0°) $\leftrightarrow$ (-90.0°, +0.0°), (+90.0°, +90.0°) $\leftrightarrow$ (-90.0°, -90.0°), and (+0.0°, -90.0°)$\leftrightarrow$ (+0.0°, +90.0°) (Table 5.2). For ΔPA of model M2, the difference of energetics for these points were found to be -0.3, 0.0, -0.2, and 0.6 kcal/mol, respectively. The small energy differences reveal the fact that isoenergetic points of PA of M1 is not affected by the asparagine side chain interactions.

For M2, the minimum energy points of ΔPA are found at Φ, Ψ = (-180.0°, +0.0°)

Table 5.2. ΔPA and $\Delta\Delta G_{aq}$ of model M1/M2 at 16 crucial points from Ramachandran plot (B3LYP/6-31+G). Energetics are given in kcal/mol. ΔPA of M1 is given relative to the minimum point of Figure 5.7. Aqueous phase calculations are done with CPCM.

	ΔPA			
$\Psi\backslash\Phi$	-180.0	-90.0	0.0	90.0
180.0	23.8/13.0	21.4/14.9	19.1/11.1	21.9/13.3
90.0	15.0/7.5	11.9/6.8	13.3/7.7	15.7/10.0
0.0	0.5/0.0	10.2/4.7	5.3/0.3	10.2/4.7
-90.0	14.6/7.9	15.7/10.2	13.3/7.1	11.9/6.5
	$\Delta\Delta G_{aq}$			
$\Psi\backslash\Phi$	-180.0	-90.0	0.0	90.0
180.0	15.0/9.2	14.2/12.9	12.8/9.1	14.2/11.8
90.0	6.0/6.2	9.9/7.4	10.9/8.7	11.4/8.9
0.0	0.0/0.0	9.9/7.2	7.1/3.9	9.9/7.4
-90.0	6.0/6.3	12.6/8.8	10.9/7.8	9.9/8.2

and Φ, Ψ = (+0.0°, +0.0°) (Table 5.2). These points correspond to the most acidic region (region I) of model M1. In addition, the four points with $\Psi = +180.0°$ correspond to the most basic region (region III) of model M1 and they are also found to be the highest energetic points for model M2. The rest of the points belongs to medium acidic region (region II) for M1, and the energetics of those points are in between the ones in region I and region III also for M2. All of these observations on difference between ΔPA of M1 and M2 are also valid for $\Delta\Delta G_{aq}$ results, concluding that using M1 as a model is acceptable to analyze the backbone amide acidity deviations with respect to conformations.

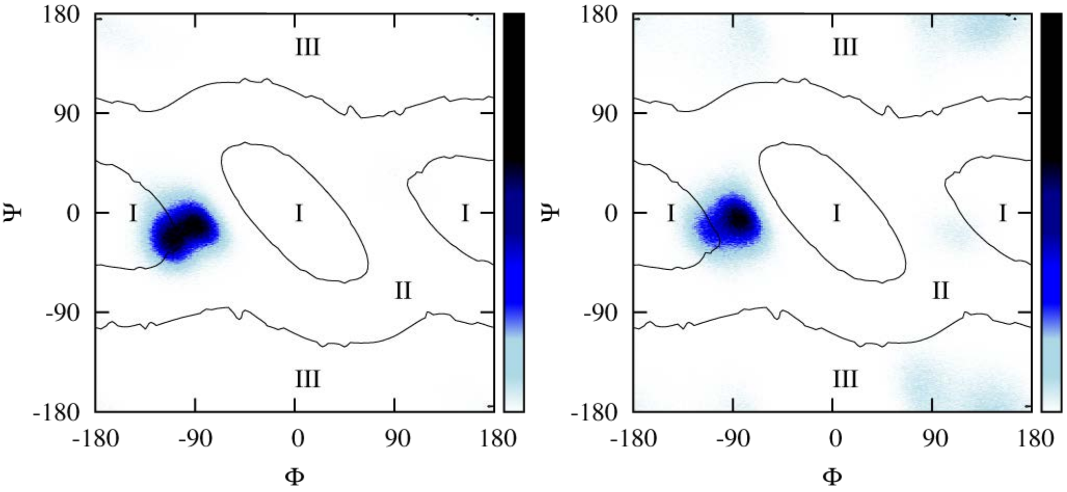
Although either of the models sufficiently describe the acidity dependence on backbone conformations, high values of $\Delta\Delta G_{aq}$ are noticeable. The highest $\Delta\Delta G_{aq}$ for M1 and M2 was found to be 16.9 kcal/mol and 12.9 kcal/mol, respectively. These values of $\Delta\Delta G_{aq}$ correspond to 12.4 units and 9.5 units of pK_a difference between the most acidic and most basic point on Ramachandran plot, for M1 and M2 respectively. This large gap between two conformational states stems from the poor stabilization of anions in implicit solvent. Since the anion is better stabilized in M2 with the influence of Asn side chain, $\Delta\Delta G_{aq}$ values are smaller in the case of M2, compared to M1. In order to gain insights on the quantitative pK_a deviations of glycine with respect to conformational changes, calculations should be performed on stabilized systems with the aid of explicit water molecules as suggested by Olivia et al. [151]. This kind of calculation is out of the scope of the present study, since our concern is to obtain a qualitative picture of relative acidity of 1° and 2° deamidation sites.

As a summary, the correlation between backbone amide acidity and conformational changes analysis for either M1 or M2 models for both ΔPA and $\Delta\Delta G_{aq}$ describe the most acidic conformations in region I, and the most basic ones in region III. These conclusions will be the leading point of our further analysis on differences of acidities of 1° and 2° deamidation sites of TPI.

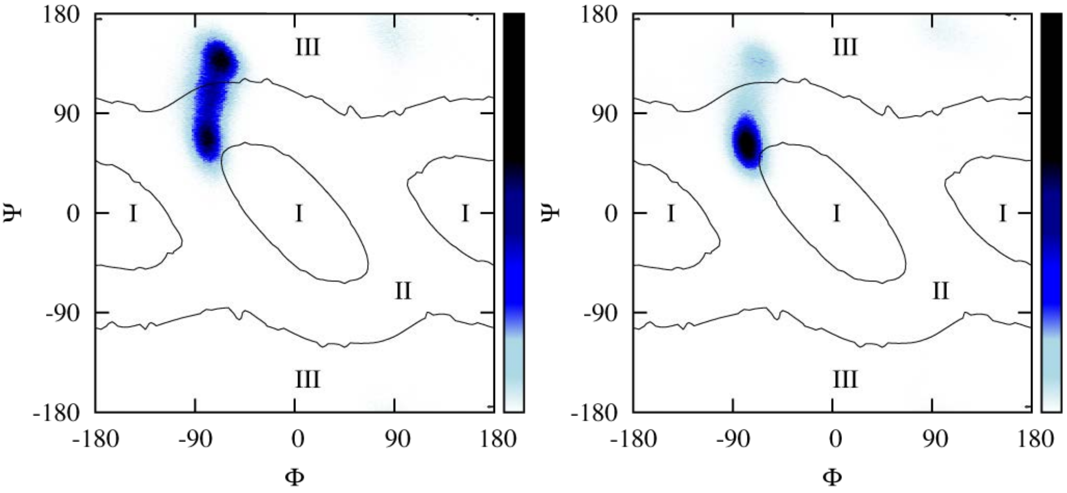
5.3.3.2. Deviations of Backbone Amide Acidity in TPI.

Ramachandran plots of the Gly backbone were plotted within the simulation time. In WOW, Gly of 1° deamidation site mostly populated inside the region I or on the border between region I and II (Figure 5.8), which corresponds to most acidic conformation of glycine with respect to ΔPA and $\Delta\Delta G_{aq}$ results (Figure 5.7). Whereas, Gly of 2° deamidation site occupies region III and some parts of region II that are closer to region III (Figure 5.8). These fields are less acidic than the regions occupied by Gly of 1° deamidation site. In the case of chicken TPI (CHI), GlyA16 and GlyB16 are mostly populated in region III which corresponds to least acidic part of the surface (Figure 5.9). Noticeably high amount of population is observed for GlyB15 in region II (Figure 5.9). This population is never observed in mammalian TPI 2° deamidation site.

The variations of conformations in the Ramachandran plot with respect to time for the 1° deamidation site is shown in Figure 5.10. Conformations at the border of

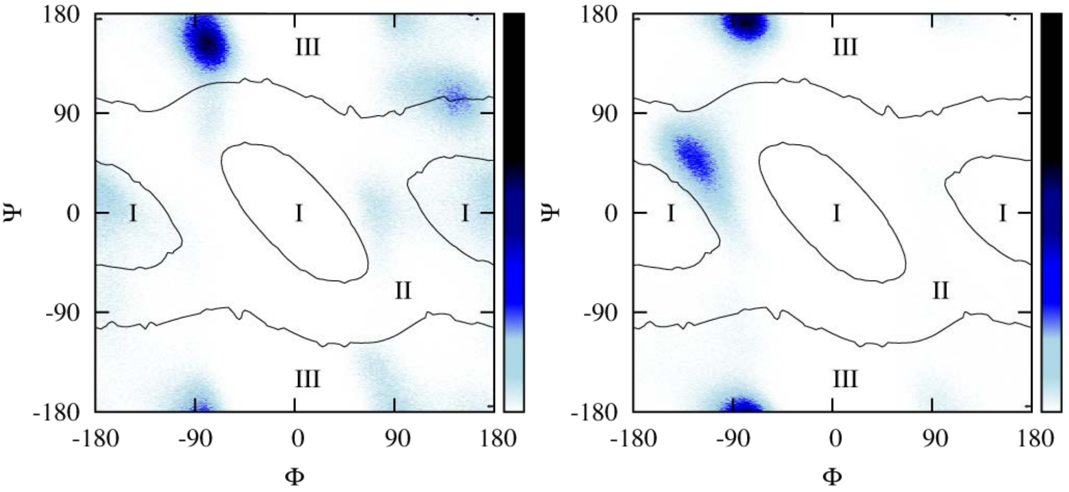


(a) GlyA72 (b) GlyB72



(c) GlyA16 (d) GlyB16

Figure 5.8. Probability distribution of Φ Ψ dihedrals of GlyA72 (a), GlyB72 (b), GlyA16 (c) and GlyB16 (d), sample WOW. Probability increases from white to black.



(a) GlyA16 (b) GlyB16

Figure 5.9. Probability distribution of Φ Ψ dihedrals of GlyA16 (a) and GlyB16 (b), sample CHI.Probability increases from white to black.

region I and II are classified as in region I for simplicity. The most acidic conformation I is observed after 32 ns for GlyA72, and this conformation is kept until the end of the simulation (1500 ns) (Figure 5.10). On the other hand, for GlyB72 , clustering on the most acidic conformation (region I) does not occur before 289 ns (Figure 5.10). This geometry is occupied for only 40 ns, then the system goes back to basic conformation for another 130 ns. The most probable conformation of this residue (region I) is filled starting from 462 ns to the end of simulation (Figure 5.10). Fluctuations are less significant for 2° deamidation site of WOW, since all GlyA16 and GlyB16 invade much more narrow area on the Ramachandran plot comparing to GlyA72 and GlyB72 (data not shown).

In order to gain insights on stabilizing factors of more acidic and more basic conformations, hydrogen bond analysis is performed in the sample WOW. Four sites are chosen to analyze hydrogen bond network which are related to the Ramachandran plot of Gly: H acceptors Gly(N-1)@O and Gly(N)@O; and H donors Gly(N)@H and Gly(N+1)@H. Gly(N-1) corresponds to Asn, and Gly(N+1) corresponds to Ala, Arg, for 1° and 2° deamidation site, respectively. The interactions with water is excluded and the schematic representation of frequently occurring H bonds between protein

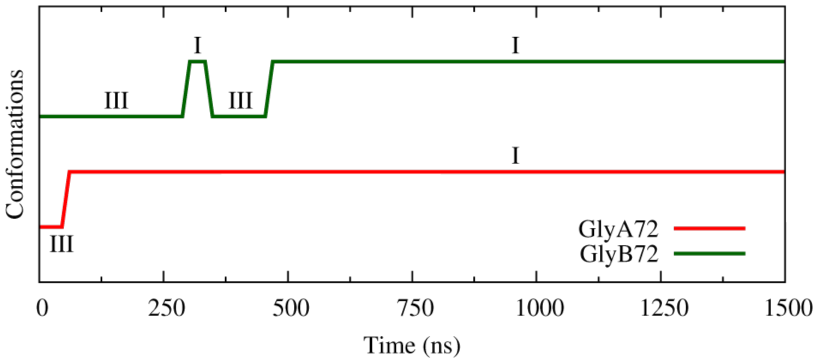


Figure 5.10. Changes of conformations in Ramachandran plot with respect to time, sample WOW.

atoms is shown in Figure 5.11.

For the 1° deamidation site, in both of the monomers, Asn71@O and Gly72@H are not involved in strong interactions, therefore they are free to be solvated by water (Figure 5.11). Gly72@O is occupied by the interactions with backbone hydrogens of the alternative monomer. Ala73@H interacts with the side chain oxygen of Asn71 of the same monomer more than half of the simulation. In the case of 2° deamidation site, all four atoms of interest are inhabited by strong protein interactions. Regarding the hydrogen bond between GlyB16@ and AsnA71@OD1, all of these interactions are intramonomeric. To compare 1° and 2° deamidation site, the main distinction is that, 2° deamidation site, for both A and B monomers, is engaged by strong protein interactions more significant than the 1° one (Figure 5.11). In addition, 2° deamidation site is rigidified by the intramonomeric interaction between Arg17@H with Asn16@O. This intramonomeric interaction, which involves two of the sides related to Ramachandran plot, might be the stabilizing effect of more basic conformations in 2° deamidation site of WOW.

5.3.4. Near Attack Conformations (NAC's)

One of the main hypothesis that explains diminishing rate of deamidation in proteins with respect to small peptides is that lose of flexibility would prohibit the

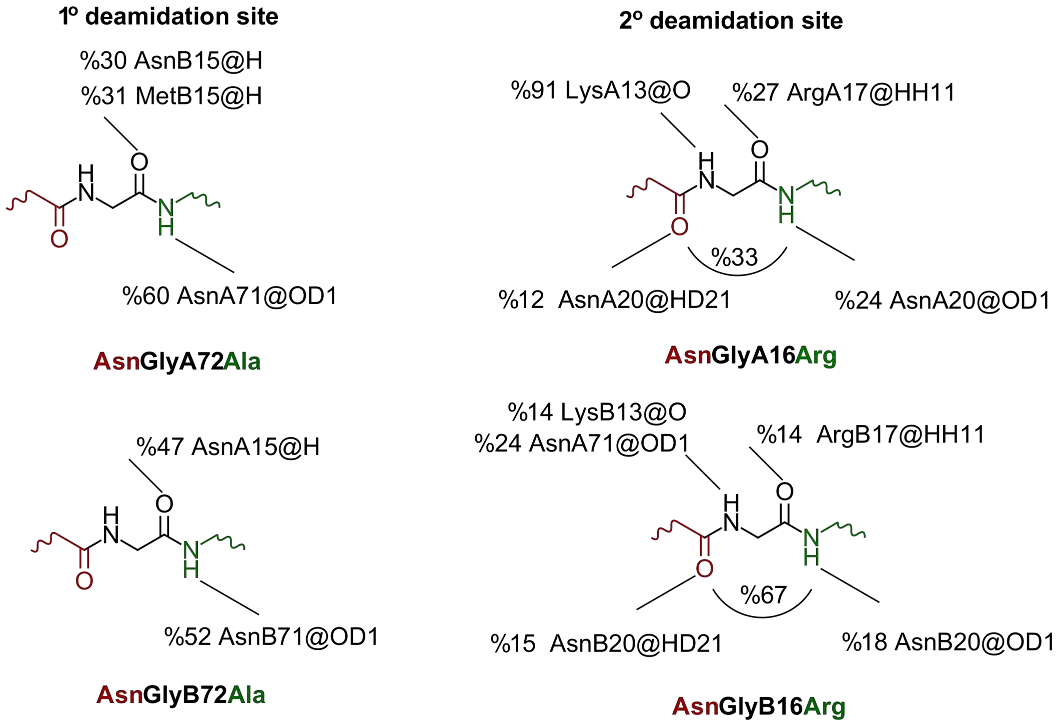


Figure 5.11. Hydrogen bond network of 1° and 2° deamidation sites, sample WOW. The percent occupancies of hydrogen bonds between Asn@O (or Gly@H, Gly@O, Ala/Arg@H) and protein atoms are given. Asn, Gly and Ala/Arg are shown as red, black and green colors, respectively.

required orientation of Asn and Gly to deamidate. To accomplish deamidation, peptide backbone should rotate freely and hold the Ψ dihedral of Asn (Ψ_{asn}) approximately 185° as shown in Figure 5.12 [14]. We have chosen to define the most favorable dihedral for tetrahedral formation in the range of $\Psi_{asn} \pm 75°$. In addition, Gly@N and Asn@CG distance should be small enough to produce succinimide ring via nucleophilic attack. Here, this critical distance was determined as 3.5 Å in accordance with previous QM studies [127,129]. The trajectories were analyzed, and the geometries which meet those two criteria were classified as near attack conformers (NAC), as generally referred in literature. The probabilities of finding NAC's within the whole simulation is tabulated in Table 5.3.

For mammalian TPI (WOW), 1° deamidation site holds the geometrical features susceptible to deamidation almost half or more of the simulation with the percent

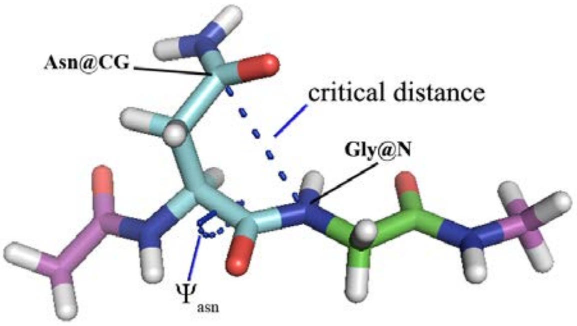


Figure 5.12. Necessary orientation for succinimide formation [14]. Asn, Gly and the rest of the peptide are colored with cyan, green and purple.

occurrence values of 67.2 % and 45.0 % in monomer A, and B, respectively. On the other hand, NAC's rarely occur in 2° deamidation site of this sample (% 1.5, and % 3.2 for monomer A, and B). For chicken TPI (CHI) NAC's have percent occurrence of values 7.6 %, and 35.1 % for monomers A, and B. Those values are higher than the percent occurrence of 2° deamidation site of mammalian TPI.

Table 5.3. Percent occurrence of Near Attack Conformers (Gly@N-Asn@CG distance $< 3.5\mathring{A}$, $+110.0° \leq \Psi_{asn} \leq +260.0°$).

	71A	**71B**	**15A**	**15B**
WOW	67.2	45.0	1.5	3.1
CHI			7.6	35.1

5.3.5. Comprehensive Results

Up to this point, differentiations in terms of global desolvation, acidity, and NAC analysis of WOW, and CHI samples were introduced. Here, the correspondence among these descriptors is evaluated, by analyzing if these events are observed within the same frame or different frames during the simulation time. Global desolvation ($N_{15.5\text{Å}}$) was defined as described in Section 5.3.2, and residues having $N_{15.5\text{Å}}$ values smaller than

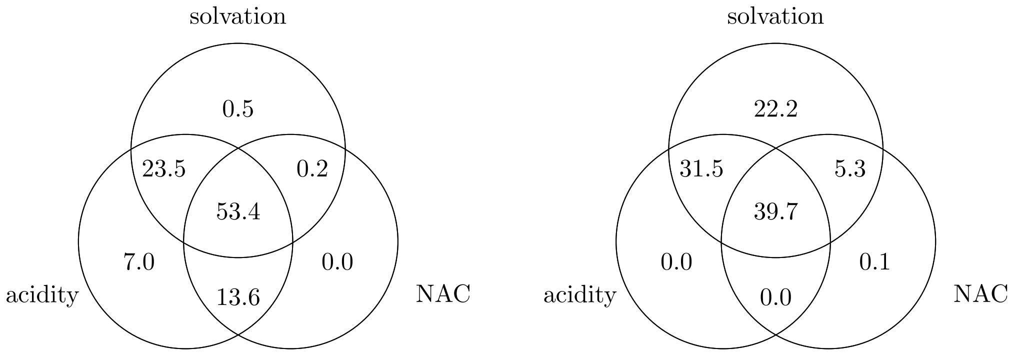
420 were considered as solvated. Due to the fact that the solvation percentage of Gly is always smaller than Asn adjacent to it (Table 5.1), "$N_{15.5\text{Å}}$ <420" of Gly considered as limiting solvation value of the corresponding AsnGly sequence.

For the acidity of the backbone, Ψ dihedral has been chosen as a descriptor, considering the fact that variations of either PA or ΔG_{aq} more dependent on Ψ than Φ as introduced in Section 5.3.3. To describe the relatively more acidic conformations, values of $\Psi \in$ [-60.0°, +60.0°] are extracted. This limit corresponds to a variety of acidity with a maximum $\sim$ 8 kcal/mol shift from the most acidic conformation in proton affinity (region I, Figure 5.7).

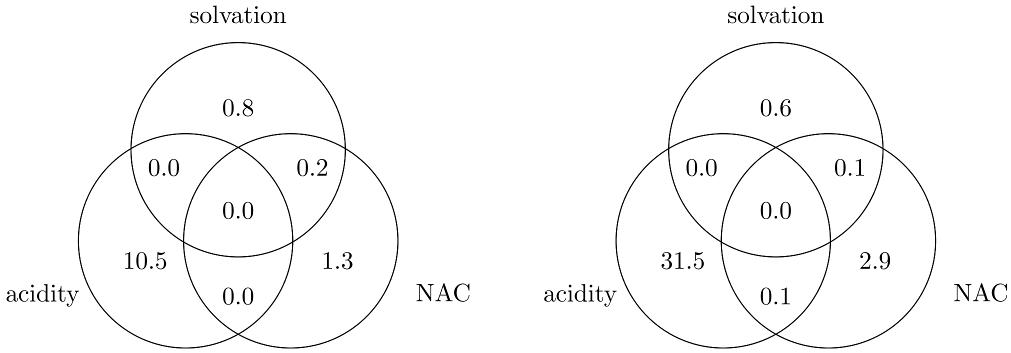
In each frame, we tested if Asn15Gly16 and Asn71Gly72 meet one or more of the criteria discussed above. For example, one frame, with Gly which has $N_{15.5\text{Å}} = 441$, $\Psi = 56.60^0$, $\Psi_{asn} = 160^0$,Gly@N-Asn@CG distance = 3.3 Å is classified as a system which is not solvated (441 > 420), acidic (-60.0° < 56.60° < +60.0°), and occupies near attack conformation (110.0° < 160.0° < 260.0° and 3.3 Å < 3.5 Å). Thus, for this frame, solvation, acidity, and NAC values are scored as 0, 1, and 1. The scores were summed along all the trajectories and divided by the total number of frames to find the percent occurrence of these criteria (Figure 5.13, 5.14).

For 1° deamidation site of mammalian TPI the probabilities of being both solvated, acidic, and resembling NAC (solvation $\cap$ acidity $\cap$ NAC) is 53.4 % and 39.7 % for monomer A and B, respectively (Figure 5.13). In the case of 2° deamidation site of WOW, there are no frames which met these criteria in the same frame, returning 0 probability in the intercept (Figure 5.13). The probability of being both solvated, acidic and NAC in CHI sample are different than 0 (Figure 5.13 1.4 % and 13.3 % for monomer A and B). However, those values are still smaller than the 1° deamidation site of WOW.

For 2° deamidation site of WOW, in which the solvation probabilities are almost 0, acidic conformations of GlyA and GlyB are observed as 10.5 % (A) and 31.6 % (B). 1° deamidation site A of WOW has also considerably high amount of acidic conformations

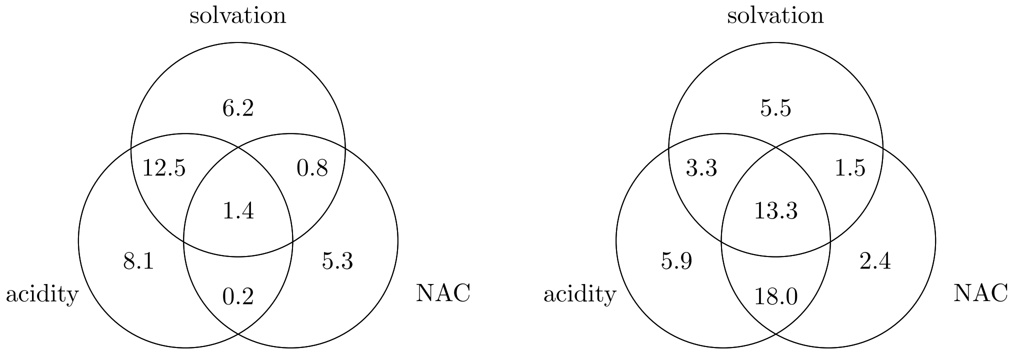


(a) 1° DS monomer A (b) 2° DS monomer A



(c) 1° DS monomer B (d) 2° DS monomer B

Figure 5.13. Sample WOW, Percent occurrences of solvation, glycine backbone amide acidity, and NAC (DS = deamidation site).

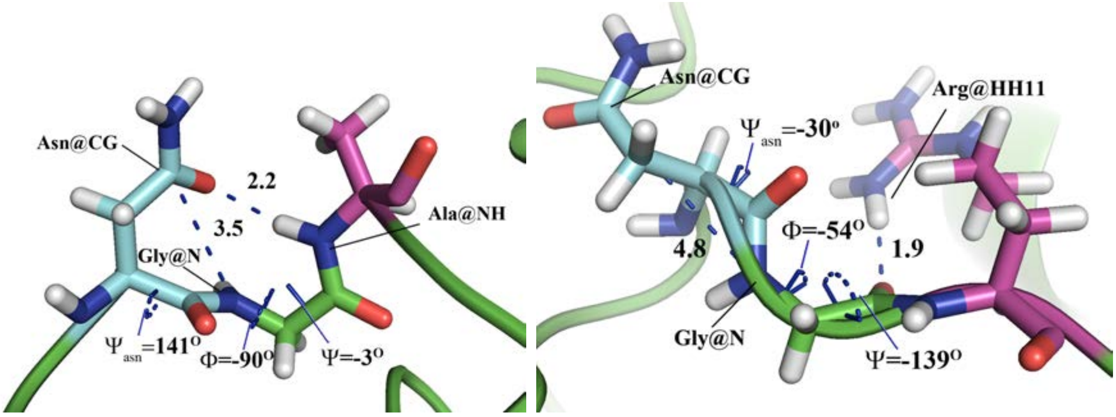


(a) monomer A (b) monomer B

Figure 5.14. Sample CHI, Percent occurrencies of solvation, glycine backbone amide acidity, and NAC.

in not solvated areas (total $20.6\ \% \rightarrow ((\text{acidity} \setminus (\text{solvation} \cap \text{NAC}) + ((\text{acidity} \cap \text{NAC}) \setminus \text{solvation}) = 7.0\ \% + 13.6\ \% = 20.6\ \%$). These results indicate that Gly might have the acidic conformations although it is buried inside the protein and its flexibility is limited by inter-molecular interactions.

Representative structures of 1° and 2° deamidation site of WOW including the most frequently observed dihedrals and hydrogen bond interactions are shown in Figure 5.15.



(a) 1° DS (b) 2° DS

Figure 5.15. Representative structures of 1° and 2° deamidation sites (DS) showing the most frequently observed interactions and dihedrals, sample WOW. Asn, Gly and the rest of the peptide are colored with cyan, green and purple, respectively.

5.4. Discussions

5.4.1. Why Asn71 deamidates faster than Asn15 in mammalian TPI?

In order to explain the relatively rapid reaction at 1° deamidation site in TPI; global desolvation, backbone amide acidity, and near attack conformer analysis were performed for Asn15Gly16 and Asn71Gly72. According to global desolvation analysis, 1° deamidation site is more solvated than 2° deamidation site (Figure 5.5), showing that this site would be exposed to water more frequently, compared to 2° deamidation site. This result would explain the faster reaction at 1° deamidation site, considering the enhancing effect of water on deamidation [129].

The more significant result obtained from global desolvation analysis is the great difference between desolvation of Gly16 and Gly72. Although Gly72 might be solvated more than half of the simulation, Gly16 is always buried (desolvated) (Figure 5.1 and Table 5.1). The greater value of desolvation on Gly16 would lead to higher pK_a in backbone amide. Higher pK_a of 2° deamidation site will cause the shift of equilibrium between asngly $\leftrightarrow$ asngly(-) in Figure 5.1 to reactant site (asngly). As a result of the decrease in the amount of activated Gly (asngly(-)), initiation of deamidation would be prohibited.

In addition to global desolvation, a more local view of glycine backbone amide acidity (Section 5.3.3) was also examined by its correlation with backbone conformations (Ramachandran plot). This analysis showed that 1° deamidation site has higher probability to be found in the most acidic regions (region I) of the Ramachandran plot compared to 2° deamidation site (Figure 5.8). Thus, Gly amide of 2° deamidation site would be less acidic, compared to 1° deamidation site. Hence, combining both global and local effects, deprotonation of Gly amide would require greater energy in the case of 2° deamidation site.

In deamidation reaction, deprotonation of Gly amide is followed by the formation of a cyclic tetrahedral intermediate (Figure 5.1). The energy required to form this cyclic intermediate would be diminished by the increased possibility of bearing conformers which resemble the transition state geometry (NAC). Percent occurrence of near attack conformers (NAC) were found to be around 50 % in the case of 1° deamidation site (Table 5.3). NAC's were almost never observed in 2° deamidation site, affirming that the formation of tetrahedral intermediate would be much more slower compared to 1° deamidation site.

Combining the remarks on glycine amide acidity and probability of having near attack conformers, it can be concluded that both deprotonation and cyclization steps require more energy in the case of Asn15 deamidation compared to Asn71.

Up to this point deamidation reaction was analyzed by discriminated steps (i.e. deprotonation, tetrahedral intermediate formation). In order to obtain successful reaction, the deamidation site should be solvated, Gly should be acidic enough, and at the same time geometries should resemble NAC's. In 1° deamidation site, these three descriptors were observed at the same time almost half of the frames of the simulation (Figure 5.13). In the case of 2° deamidation site, the probability to be found in these three states is zero (Figure 5.13). This result indicates that in the same conditions, almost half of the geometries of 1° deamidation site can initiate deamidation, whereas none of the geometries in 2° deamidation site are reactive.

5.4.2. Is Asp71 a prerequisite for Asn15 deamidation?

In the present study chicken TPI (CHI) was simulated to examine the behavior of Asn15Gly without the influence of Asn71Gly. In CHI, the probabilities of being reactive to initiate deamidation (being solvated, acidic and resembling NAC at the same frame) were found to be 1.4 % and 13.3 % for monomer A and B (Figure 5.14). Especially in the case of monomer B, there is a relatively high amount of reactive Asn15Gly compared to 2° deamidation site of WOW (which was found to be zero). If these two deamidations were independent in mammalian TPI, the amount of reactive conformations of Asn15Gly in WOW should be similar to that of CHI. Obtaining zero probability of reactive conformers in 2° deamidation site, different than the other deamidation sites of all samples, indicates that Asp71 might act as a prerequisite in order to increase the reactivity of Asn15 of mammalian TPI.

In our previous study [143], the correlation between the two deamidation sites were analyzed by simulating mutant TPI's (N71D) up to 60-90ns, and there has been no evidence found on the effect of Asp71 on Asn15. However, as introduced in Section 5.3.3 some crucial events related to deamidation can only be observed in longer simulation time. In order to have a better understanding of the effect of Asp71 on the reactivity of Asn15, here we suggest that the simulations with Asp71 mutant should be carried out within larger timescales.

5.4.3. Tertiary structure effect

The analysis of global desolvation along the simulation can shed a light on the effect of tertiary structure in different rates of deamidation. As it was already shown previously [143], the higher solvation values for the 1° deamidation site can be explained by the relative flexibility of loop3 on which Asn71Gly is located, compared to loop1 involving Asn15Gly. The smaller solvation percent of Asn15Gly of CHI compared to 1° deamidation site of WOW, is also in agreement with this conclusion (Table 5.1).

5.4.4. Primary structure effect

Most of the studies in the literature, related with the primary structure effect on deamidation, focus on the residues in the C-terminal ($n+1$) or N-terminal ($n-1$) position to Asn. In these studies, $n+1$ residues are found to be the most crucial factor on deamidation rates, in contrary the effect of $n-1$ residues are found to be negligible [14, 152]. Despite the observed changes in rates of pentapeptides bearing different $n+2$ residues [14], there are not enough studies that sufficiently explain its impact on deamidation. In mammalian TPI, AsnGly is followed by Ala and Arg in 1° and 2° deamidation sites, respectively. Hydrogen bond analysis of the present study indicates that both of the Asn are stabilized by their $n+2$ residues (Figure 5.11). The hydrogen bond between Ala backbone amide and Asn side chain in 1° deamidation site favors the formation of near attack conformations (Figure 5.15). On the other hand, the stable interaction between Asn and Arg backbone in the 2° deamidation site, limits the Φ/Ψ rotations, therefore the probability to be found in the relatively more acidic regions is diminished. These results illustrate that $n+2$ residue may favor or disfavor initiation of deamidation. Hence, the effect of $n+2$ residue should be included further formulations of deamidation rate prediction.

5.5. Conclusions

We have explored the different reactivity of two deamidation sites in the enzyme TPI using μs MD simulations. Three descriptors were introduced to analyze deamidation reaction: i) global desolvation, ii) backbone amide acidity, iii) near attack conformer analysis. These independent descriptors were efficiently used in order to gain insights on the sequential deamidation process in mammalian TPI. According to these analysis, initiation of deamidation would be facilitated at Asn71 because of its greater solvation and also the higher acidity of the adjacent glycine (Gly72) backbone amide. With respect to the results obtained from near attack conformer analysis, not only the initiation would be faster in the case of Asn71, but also the tetrahedral formation would require less energy. Combining these outcomes, it can be concluded that deamidation at Asn15 is hindered by both global and local factors.

Our findings also contribute to the understanding of the impact of peptide sequence on deamidation rate. The global desolvation analysis points the importance of tertiary structure on deamidation. The backbone acidity and the near attack conformation results highlight the possible role of residues $n+2$ to Asn. The relatively higher reactivity observed for Asn15 of chicken TPI signifies the possible effect of deamidated Asn71 (Asp71) on the deamidation of Asn15 in mammalian TPI. However, further analysis (e.g. μs MD simulations involving Asp71) should be performed in order to obtain a precise conclusion on this phenomenon.

6. COMPARISON OF THE REACTION KINETICS OF ASPARAGINE DIPEPTIDE AND TRIOSEPHOSPHATE ISOMERASE USING QM/MM TOOLS WITH UMBRELLA SAMPLING TECHNIQUE

The deamidation of asparagine (conversion to aspartate residue) in the enzyme triosephosphate isomerase (TPI) is proposed to have a significant impact on the protein turnover of this enzyme. There are two distinct asparagines in TPI that are known to undergo deamidation under physiological conditions: Asn15 and Asn71. Asn71 deamidates faster than Asn15 resulting a overall half life of 37.8 days *in vitro.* Under the same conditions, a pentapeptide bearing the same primary structure with the deamidation sites of TPI has a half life of 1.04 days. The different kinetics of the peptide and two deamidation sites in TPI are investigated using QM/MM tools with umbrella sampling technique. The crucial impact of the backbone amide acidity of the adjacent glycine on deamidation rate is highlighted. The deamidation in small peptide is favored firstly because of the higher reactivity of the asparagine, and also due to the stronger stability of the tetrahedral intermediate. In addition, the benchmark analysis over several semiempirical Hamiltonian's provides significant information on the efficiency of these methods modeling reaction mechanisms with umbrella sampling technique.

6.1. Introduction

The post translational modification of asparagine(Asn) to aspartate (Asp) under physiological conditions is designated as deamidation reaction. Non enzymatic deamidation of asparagine (Asn) is observed both *in vivo* and *in vitro* for peptides or proteins [14, 105, 152–154]. Deamidation of asparagine has been of special interest following the discoveries relating *in vivo* deamidation with cell-aging and severe human diseases [14]. Besides, scientific developments on engineered protein drugs raised the interest on deamidation of Asn, since it is known to be the most commonly observed chemical degradation in peptides, and directly affects the physicochemical stability of peptide base drugs [154].

Experimental studies performed by Capasso et. al. [3, 112, 155, 156] showed that reaction mechanism of deamidation is pH dependent. At physiological pH, succinimide intermediate was isolated and conversion of asparagine to succinimide was found to be the rate determining step of the overall reaction (Figure 6.1) [112]. The same group also showed that deamidation mechanism of a model peptide is transferable to protein studies [3]. The rate of deamidation was found to be dependent on the residues in the C-terminal position to Asn, and the fastest deamidation is observed when this residue is glycine (Gly). Numerous computational studies have been done to explain the experimental findings discussed above [8, 117, 120, 121, 127, 129, 151, 157]. Trout et. al. [121] calculated the pH dependent reaction pathways and introduced the overall rate constant of the reaction. They have proposed that, succinimide formation under physiological conditions (pH = 7.4) would occur in a stepwise mechanism which involves formation of a cyclic tetrahedral intermediate (tet, Figure 6.1). Their results also indicated the importance of acidity of glycine backbone amino group on deamidation rate [121]. The acidity of backbone amide is found to be strongly correlated with the variations of Gly backbone conformation [8]. Catak et. al. [127, 129], claimed that active water molecules enhance the rate of deamidation and also the rate of succinimide formation is determined by the energetics of tetrahedral formation (Figure 6.1, asn → tet). Recently, Olivia et. al. presented the influence of explicit waters on predicting backbone amide acidity [151]. The first dynamical study examining the kinetics of succinimide formation was done by Kaliman et. al. [157], in which they have modeled a fully solvated N-terminal asparagine by density functional theory methods (DFT) using metadynamics techniques. They have concluded that the rate determining step could be either deprotonation or cyclization (Figure 6.1, asn → asn^- and $\text{asn}^- \rightarrow$ tet).

In the present study, deamidation of the enzyme triosephosphate isomerase (TPI) is examined. This post translational modification in mammalian TPI has been proposed to serve as a regulator of protein turnover, and this proposal increased the interest on TPI deamidation. Deamidation takes place at the subunit interface of TPI,

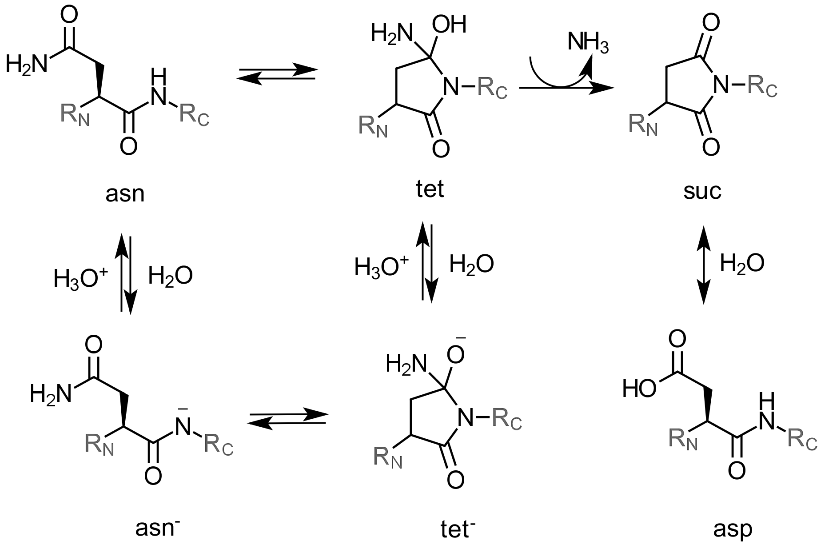


Figure 6.1. Deamidation of asparagine (asn) to aspartate (asp) via succicinimide (suc) intermediate under physiological conditions.

and there are two distinct asparagines (Asn15 and Asn71) known to deamidate in each unit of this homodimeric enzyme. Both of these asparagines that undergo deamidation in TPI has glycine in the C-terminal position to Asn. Following the experimental studies that show the relative ratios of deamidated products as Asp(71):Asp(15) 2.54:1 for human TPI, Asn71 and Asn15 are denoted as 1° and 2° deamidation sites, respectively [124]. The *in vitro* half life of the deamidation reaction in mammalian TPI was found as 37.8 days (in 0.05 M phosphate, at pH 7.0, at 37°C) [14]. Under the same conditions, the half life of a small peptide is found to be 1.04 days. The diminishing effect of the high-order structure on the deamidation rate has also been discovered for several other proteins [14]. A comparative study on the deamidation rate of asparagine dipeptide and TPI was done using molecular dynamics tools by our group [143]. In that study, it has been suggested that the different values of solvation would be an effective factor which decreases the deamidation rate in TPI compared to the asparagine dipeptide. The solvation results were also used to explain the relatively faster reaction at Asn71 compared to Asn15 [143].

A more detailed investigation has been done on the sequential order of deamidation in TPI using trajectories obtained from μs MD simulations [158]. In that study, the importance of global desolvation, backbone amide acidity, and near attack conformer formation on the initiation of deamidation reaction was explored. The relatively faster rate of deamidation at Asn71Gly, was explained by the more frequently occurrence of the solvated, acidic and reactive conformations compared to the other deamidation site (Asn15Gly).

The classical MD studies mentioned above focus on the differences on initiation of deamidation and they do not provide any kinetical information on this reaction. To the best of our knowledge, the only kinetical study on deamidation of TPI was performed by Konuklar et. al. [139] using QM/MM tools. They have shown that cyclization (asn → tet) and succinimide formation (tet → suc) are competing steps in TPI. However, the dynamical effects were not included in that study.

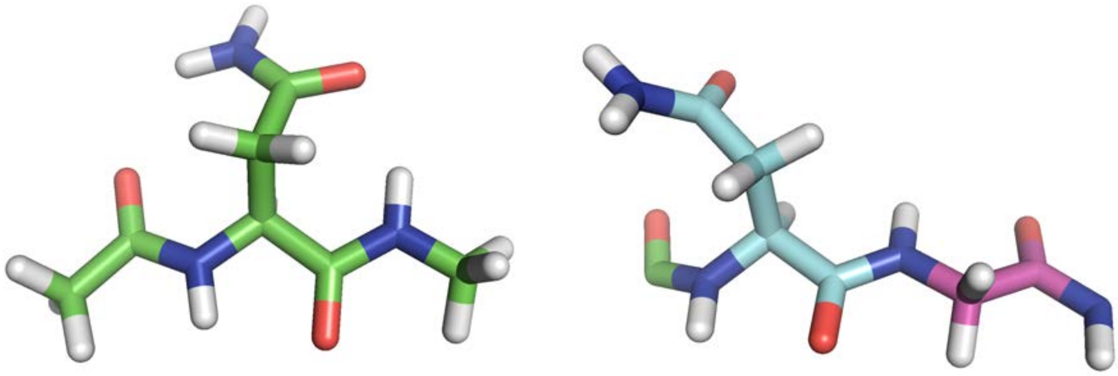
In the present study, our objective is to analyze the kinetics of this reaction for both deamidation sites of TPI using QM/MM-MD with umbrella sampling technique. In order to explain the slower rate in TPI with the effect of high-order structure, asparagine dipeptide is also modeled with the same protocol used for TPI. As it has been shown with extensive experimental and computational studies that succinimide formation is the rate determining step of the overall deamidation reaction, and also succicinimide formation is controlled by the rate of tetrahedral intermediate formation, the present work focuses on the tetrahedral formation only.

6.2. Computational Details

6.2.1. QM Calculations

QM calculations were performed for the asparagine dipeptide model which consists of acetamide-asparagine-Nmethylamide (Ace-Asn-Nme) (Figure 6.2 a). This model has been shown to be sufficient to study deamidation reaction in peptides [121,127,129]. Geometry optimizations and energy calculations were performed with Gaussian09 software [150] using density functional theory with B3LYP/6-31++G**. The polar environment was provided by Conductor-Like Polarizable Continuum Model (CPCM,

$\varepsilon = 78.5$ [87, 88]).



(a) asparagine dipeptide

(b) QM region of TPI

Figure 6.2. Representative figures for a.) asparagine dipeptide, b) QM region of TPI. Asparagine and glycine residues are colored with cyan and magenta in b.

6.2.2. Preparation of the Systems Prior to QM/MM-MD Calculations

The asparagine dipeptide was solvated with TIP3P [48] water molecules to result in a cubic system with edges of 55.8 Å, containing 5763 water molecules. Topology and coordinate files of this system were prepared with tleap module of AMBER 12 program package [42].

For TPI structures, three frames were extracted from molecular dynamics trajectories of our previous study [158]. The criteria to extract the structures is discussed in the relevant section. These structures consist of approximately 80,000 atoms. Due to the large dimensions of the system, stochastic boundary conditions were applied [159]. A sphere of radius 20 Å from the center of mass of Asn(15) and Asn(71) is chosen to form the reaction region, and atoms between 20 Å-26 Å form the buffer region. The rest of the system was excluded by replacing the exterior protein residues by N-methylamide and acetamide. A harmonic restraint with 20 kcal/molÅ^2 force constant was applied to atoms of both protein and water in the buffer region.

In order to obtain comparative results with the asparagine dipeptide, the water box of this system was also truncated with the same range defined for TPI (i.e. buffer region 20 Å-26 Å away from the center of mass of the solute).

6.2.3. Umbrella Sampling Calculations

6.2.3.1. Benchmark Studies Prior to QM/MM Calculations. Seabra et al. [160] the importance of the choice of QM and MM methods when the biological molecules are the focus of the study. Herein, several semiempirical methods were tested using umbrella sampling technique, in order to choose the most appropriate Hamiltonian to define QM part of our QM/MM calculations. The calculations were performed using only the asparagine dipeptide as a model and the polar environment was treated using modified generalized Born model implemented in AMBER 12 [42]. For the Hamiltonian's, AM1 [56], PM3 [57] and PM6 [59] were chosen as they are known to be the most commonly used semiempirical methods for QM/MM studies in the literature [161]. The modified versions of these methods; AM1/d [162,163], AM1-D* [164], RM1 [165], and PM6-D [164,166] were also tested. Density Functional Theory-based tight-binding (SCC-DFTB) method [167,168] was included to the benchmark.

6.2.3.2. Constructions of the Reaction Coordinates. The conversion of asparagine to tetrahedral intermediate with umbrella sampling technique was investigated with a two dimensional PMF. Deprotonation of N-H is prescribed by one reaction coordinate (RC1) which is defined as the antisymmetric combination of distance d1 and d2 (Figure 6.3). The distance between glycine backbone N and side chain C (d3 in Figure 6.3) was chosen as the second reaction coordinate (RC2).

6.2.3.3. QM/MM-MD Calculations. In the case of asparagine dipeptide, the solute and the water molecules are defined as the QM and MM parts of the system, respectively. For TPI, the QM part contains Asn and Gly residues of the corresponding deamidation site (e.g. for the 1° and 2° deamidation site Asn71Gly72, and Asn15Gly16 belongs to QM region, respectively). In addition to reactive Asn and Gly, four additional atoms from N-terminal and C-terminal residues of Asn and Gly were included to QM region to define the link atoms properly (i.e. to include the peptide bonds linking

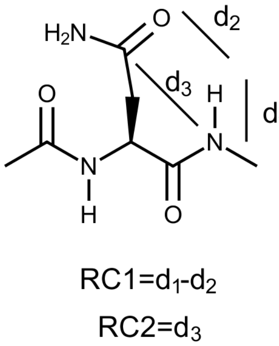


Figure 6.3. Definitions of the reaction coordinates which are used to model the formation of the tetrahedral intermediate (tet) from asparagine (asn)

the deamidation region to its neighboring aminoacids.). The resulting QM part of TPI consists of 27 reactive atoms including two link atoms (Figure 6.2b).

For both QM-MD and QM/MM-MD simulations were done using using AMBER 12 [42] program package with SANDER module. MM region of TPI is treated with ff03 force field parameters [134]. The QM part is treated by SCC-DFTB for the calculations of peptide in explicit waters and TPI.

The window size was defined as 0.10 Å, yielding a total number of 852 windows. In order to control reaction coordinates, a harmonic potential with a force constant of 300 kcal/molÅ was applied, which is centered to each window. Each window is simulated 15 ps with the time step of 0.5 fs. The results were analyzed using WHAM software to obtain potential of mean force (PMF) of the corresponding mechanisms.

6.3. Results

6.3.1. Reaction Mechanism

Prior to free energy calculations with umbrella sampling technique, we have performed a static QM calculation in order to characterize the reaction mechanism (i.e. stepwise/concerted, early/late transition state). This calculation also aimed to obtain an overview of the energetics of the mechanism. For this purpose, the conversion of the asparagine dipeptide to tetrahedral intermediate was modeled in continuum solvent model using DFT (B3LYP/6-31+G**//CPCM). Due to computational cost of the free energy calculations with umbrella sampling technique using DFT methods, the results obtained from this static calculation are used as a reference for further analysis on the free energy calculations performed with various of semiempirical Hamiltonian's. Although this approach does not give the most accurate comparison in terms of energetics, it provides significant information to distinguish between several semiempirical methods as discussed below.

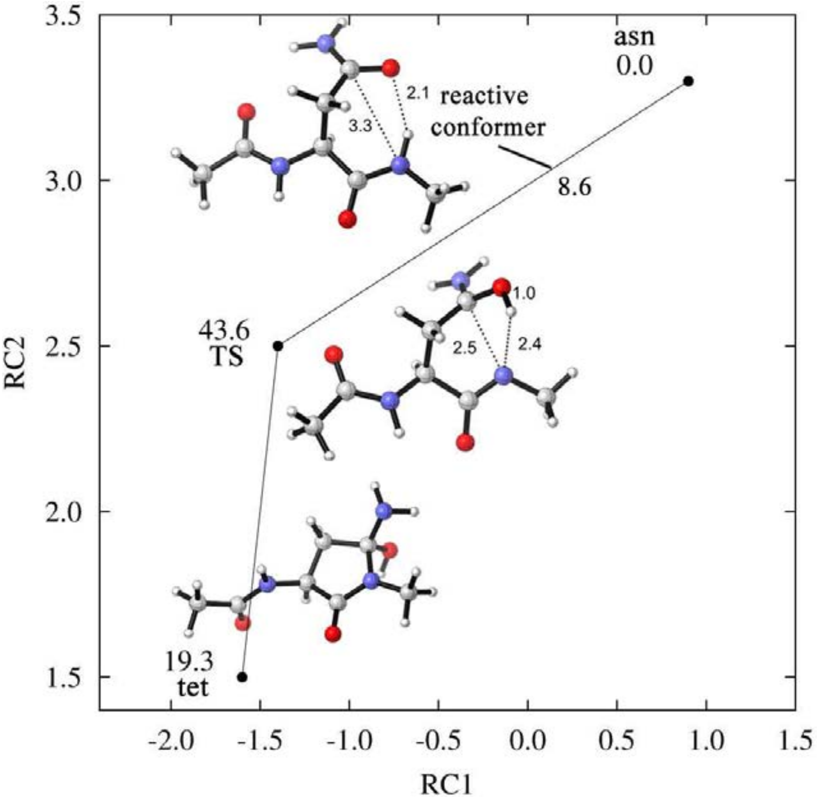


Figure 6.4. Potential energy profile of formation of tetrahedral (tet) intermediate from asparagine (asn) obtained form B3LYP/6-31+G**//CPCM calculations. Gibbs free energies are given in kcal/mol unit.

According to the results obtained from QM calculations, conversion of asparagine to tetrahedral intermediate is a concerted mechanism involving a late transition state (Figure 6.4). Backbone nitrogen is simultaneously activated by deprotonation with the aid of the side chain oxygen. The five membered cyclic tetrahedral intermediate is formed via nucleophilic attack of glycine backbone nitrogen to asparagine side chain. Deprotonation occurs prior to the ring closure indicating the asynchronous feature of the transition state. The Gibbs free energy of activation ($\Delta G^{\#}$) and reaction (ΔG_{rxn}) are found to be 43.6 kcal/mol and 19.3 kcal/mol, respectively (Figure 6.4). These features are in accordance with the previous QM studies performed in gas phase [121, 127].

6.3.2. Benchmark Studies with Semiempirical Methods (QM-MD)

2D PMF of tetrahedral intermediate formation was calculated for the asparagine dipeptide with umbrella sampling technique, using the following semiempirical methods: AM1, AM1/d, AM1-D*, PM3, RM1, PM6, PM6-D, and SCC-DFTB. In order to obtain succesful comparison with DFT calculations, the calculations were done using continuum solvent model (modified generalized Born model). The reaction coordinates (RC1, RC2) were defined as introduced above (Figure 6.3). The reaction is modeled starting from a reactive conformer which has 3.3 Å distance between backbone N and side chain C (reactive conformer in Figure 6.4). This structure has the characteristic features of the minimum geometry found in gas phase calculations [127, 129].

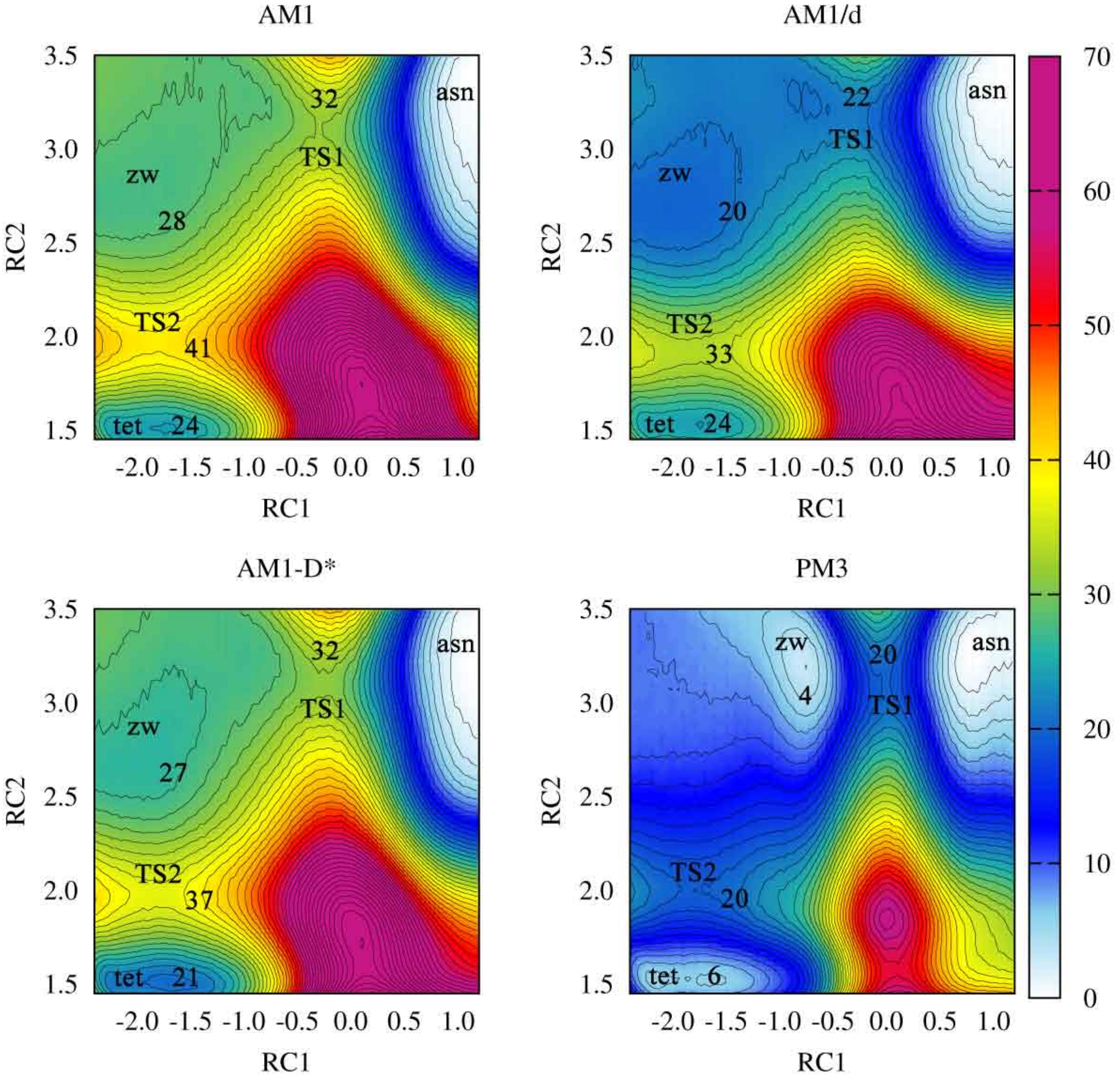
The tested semiempirical methods were evaluated by comparing the results obtained with DFT calculations (B3LYP/6-31+G** (CPCM)). The energetics of DFT results were recalculated taking the reactive conformation as a reference, which is the starting point of umbrella sampling calculations. The updated activation barrier and the reaction free energy was found as 35.0 kcal/mol and 10.7 kcal/mol, respectively.

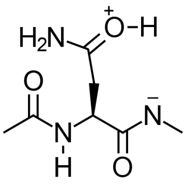
The energy surface of AM1, AM1/d, AM1-D* were found to be very similar to each other (Figure 6.5). These three methods predict a stepwise mechanism involving a metastable zwiterionic intermediate (zw, RC1, RC2 $\approx$ -2.0 Å, 3.0 Å) as represented in Figure 6.5. The formation of the zwiterionic intermediate is followed by cyclization which has higher barrier than the first step. The predicted energies for the barriers are not comparable with DFT since neither transition state for zwiterion formation, or this intermediate itself cannot be obtained by DFT methods. The product on the other hand, is approximately 10 kcal/mol less stabilized compared to DFT methods for all of these three AM1 based Hamiltonian's.

A similar kind of stepwise mechanism, which involves formation of zwiterionic intermediate prior to cyclization, is also predicted by PM3 (Figure 6.5). The stabilization of zwiterionic intermediate with PM3 is noteworthy. This intermediate is only 4 kcal/mol less stabilized than the reactant, and it is even more stable (2 kcal/mol) than the tetrahedral product (tet). Barriers of hydrogen abstraction and cyclization steps are isoenergetic. To the opposite of AM1 based methods, the relative energy of the tetrahedral product is smaller than the one predicted in DFT results.

In the case of PM6, PM6-D and RM1 in the range $2.5 \leq$ RC2 ≤ 3.5, as RC1 decreases, the energy increases gradually yielding a large, flat, high energy part in the PMF (Figure 6.6). For these three methods only one first order saddle point is determined revealing a concerted reaction mechanism. For PM6 and PM6-D, the TS is located at (RC1, RC2) $\approx$ (-1.7 Å, 2.3 Å). This TS has a barrier of $\sim$ 24 kcal/mol for both PM6 and PM6-D, and its structural features are similar to TS2 found in AM1 based methods and PM3. In the case of RM1, the TS is found at $\approx$ (-1.2 Å, 3.0 Å) with an energy barrier of $\sim$ 27 kcal/mol. Tetrahedral product stabilization is well predicted, with the relative free energies of $\sim$ 8 kcal/mol for all of these three methods (PM6, PM6-D, and RM1).

For SCC-DFTB, there are only two local minima detected, and those minimas correspond to reactant (asn) and product (tet) (Figure 6.6). The connection of those two minima and the saddle point constitutes a concerted reaction mechanism for the conversion of asparagine to tetrahedral intermediate. The first order saddle point at (RC1, RC2) = (-1.2Å, 2.6Å) bears the features of a late transition state as suggested in DFT calculations. In addition, for SCC-DFTB, the predicted activation barrier and the reaction energy is found to be $\sim$ 35 kcal/mol and $\sim$ 11 kcal/mol, respectively. These values are in a very good agreement with the energetics found by DFT (B3LYP/6-31++G**) calculations (35.0 kcal/mol and 10.7 kcal/mol, respectively).





zwiterionic intermediate (zw)

Figure 6.5. Free energy profile for the tetrahedral (tet) formation from asparagine (asn) obtained by QM-MD using the corresponding QM methods: AM1, AM1/d, AM1-D*, PM3. Energetic are given in kcal/mol (the counter level is 1.5 kcal/mol). Regions having free energy values greater that 70 kcal/mol use the color of 70 kcal/mol.

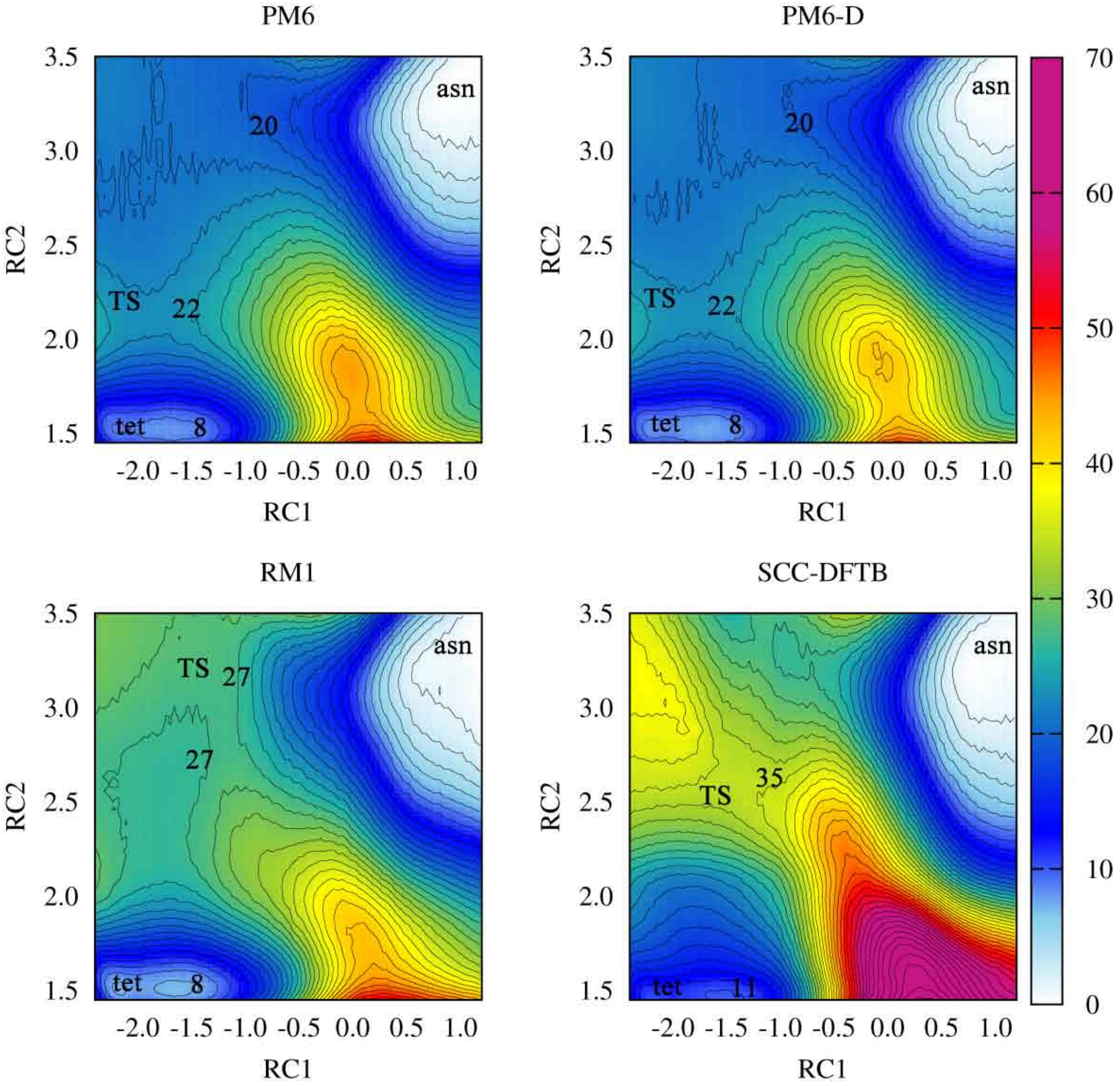


Figure 6.6. Free energy profile for the tetrahedral (tet) formation from asparagine (asn) obtained by QM-MD using the corresponding QM methods: PM6, PM6-D, RM1, SCC-DFTB. Energetic are given in kcal/mol (the counter level is 1.5 kcal/mol). Regions having free energy values greater that 70 kcal/mol use the color of 70 kcal/mol.

The methods which predict a stepwise mechanism (AM1, AM1/d, AM1D* and PM3) were discarded for further QM/MM calculations since neither zwiterionic intermediate, nor the transition state producing this intermediate were not observed in high level QM calculations. The PMF's obtained by the rest of the tested methods (PM6, PM6-D, RM1 and SCC-DFTB) indicate a concerted mechanism. Those PMF's are differentiated from each other by the energetics and by the location of the first order saddle point (i.e. geometrical feutures of the TS). In the case of PM6 (or PM6-D), the energetics of the activation barrier are $\sim$ 15 kcal/mol underestimated compared to high level QM calculations. For RM1, the TS is hardly detected because of the flat high energetic region around the TS. The geometrical feautures of the TS and the energetics predicted by SCC-DFTB is the most similar one to the TS of our high level QM calculations (B3LYP/6-31++G**). Besides, the relative energies and structures predicted by SCC-DFTB has been previously shown to be in a good agreement with B3LYP/6-31G* and MP2/6-31G* calculations. Thus, SCC-DFTB is chosen to perform further QM/MM calculations with asparagine dipeptide and TPI.

6.3.3. Asparagine Dipeptide versus TPI: a QM/MM-MD study

6.3.3.1. Asparagine Dipeptide. In order to treat the polar environment more efficiently, QM/MM-MD calculations were performed on the asparagine dipeptide model with explicit water models. The QM region consists of only the peptide, and the water molecules are in the MM region. QM calculations were done using SCC-DFTB Hamiltonian.

QM/MM calculations of the PMF of $asn \rightarrow tet$ step for asparagine dipeptide also gives a concerted mechanism with an energy barrier of $\sim$ 34.0 kcal/mol (Figure 6.7). The reactant (asn) is $\sim$ 12.0 kcal/mol more stable compared to the tetrahedral intermediate. Hence, the effect of explicit water molecules on energetics is negligibly small compared to continuum medium (Figure 6.5).

6.3.3.2. TPI. The deamidation of asparagine under physiological conditions is initiated by the deprotonation of backbone NH on glycine which is adjacent to asparagine. Thus the intrinsic acidity of this amide is a crucial factor that determines the different rates of deamidation. In our previous MD study [158], which focused on exploring the different reactivities of Asn15-Gly16 and Asn71-Gly72, the higher acidity of Gly72 amide (compared to Gly15) was found to be one of the main reasons which diminishes

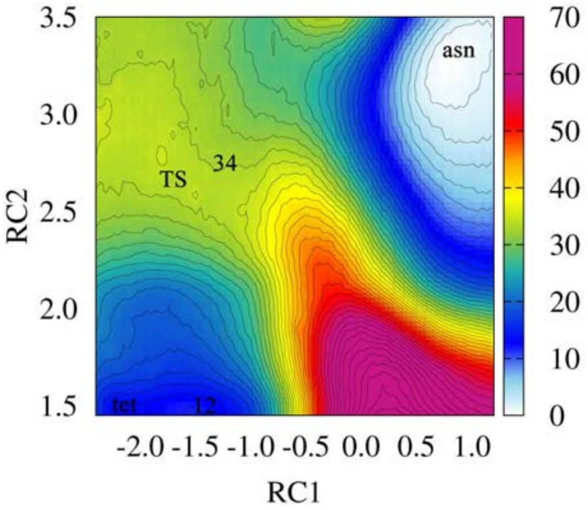


Figure 6.7. Free energy profile for the tetrahedral (tet) formation from asparagine (asn) obtained by SCC-DFTB/MM-MD calculations. Energetic are given in kcal/mol (the counter level is 1.5 kcal/mol). Regions having free energy values greater that 70 kcal/mol use the color of 70 kcal/mol.

the rate of deamidation at Asn15. The strong dependency of glycine amide acidity to backbone conformations, which was first shown by Radkiewicz et. al. [8], is used to accomplish the acidity comparison between Gly16 and Gly72. The method used and the results regarding this comparison obtained in our previous study can be summarized as follows:

- The deviations of Proton Affinity (PA) and differences in the change of Gibbs free energy in aqueous media (ΔG_{aq}) of the glycine amide were calculated with respect to changes in backbone conformations (Φ and Ψ dihedrals of the Ramachandran plot).
- The relative PA or ΔG_{aq} of this amide (in other way of saying, the relative acidity) was divided to three different regions of its Ramachandran plot considering the energy differences.
- Among those, the conformations of glycine bearing dihedral $\Psi \in$ [-60.0°, +60.0°] were found to be the most acidic ones.
- The Ramachandran plot of the glycine belonging two deamidation sites in TPI were also examined. When all of the frames of these 1.5 μs simulation is taken into consideration; Gly72 has the highest probability to be found in more acidic

region, whereas Gly16 always stays in less acidic regions.

In conclusion, it has been proposed that Asn71 should deamidate faster than Asn15, because Gly72 amide would be more easily deprotonated so that the reaction would initiate faster. In order to test this claim with kinetical studies, the statistics on Ramachandran plots obtained in our previous study is used to refine the possible frames which will be subject to umbrella sampling calculations. This approach enables us to combine the classical statistics with the kinetical results. The refinement is accomplished by extracting snapshots from the regions that Gly16 and Gly72 has the highest probability to be found in the Ramachandran plot according to our MD statistics. As mentioned above, these regions correspond to more acidic and less acidic regions for Asn71-Gly72 and Asn15-Gly16, respectively. These structures are labeled as C1-A (Asn71-Gly72) and C2 (Asn15-Gly16). For the QM/MM-MD calculations, the reactive parts of the samples (residues treated by QM) are defined as Asn71-Gly72 and Asn15-Gly16 for samples C1-A and C2, respectively. In order to make a more decent comparison on the glycine amide acidity, an additional snapshot is extracted from the Ramachandran plot of Asn71-Gly72 (C1-B). The geometrical features of this additional snapshot corresponds to a less acidic conformation of Gly72.

The formation of tetrahedral intermediate occurs via a concerted mechanism for all of the three samples (C1-A, C1-B, C2, Figure 6.8 a to c). In the case of 1° deamidation site, for the sample C1-A, the reaction proceeds through an energy barrier of $\sim$ 35 kcal/mol (Figure 6.8 a). The TS is located at (RC1, RC2) $\approx$ (-1.5 Å, 2.6 Å). The d1, d2 distances of RC1 were found to be 1.0 Å, and 2.5 Å indicating an asynchronous late transition state in which the proton transfer from backbone amide occurred prior to ring closure (RC2=d3=2.6 Å). The tetrahedral product is found at (RC1, RC2) $\approx$ (-2.1 Å, 2.5 Å) and it is $\sim$ 23 kcal/mol higher in energy than asn. For the less acidic conformation of this deamidation site (C1-B), the barrier increases to 41 kcal/mol which is located at (RC1, RC2) $\approx$ (-1.3 Å, 2.5 Å) (Figure 6.8 b). The reaction energy is found to be $\sim$ 26 kcal/mol which is similar to the reaction energy observed for the more acidic conformation (C1-A, $\sim$ 23 kcal/mol).

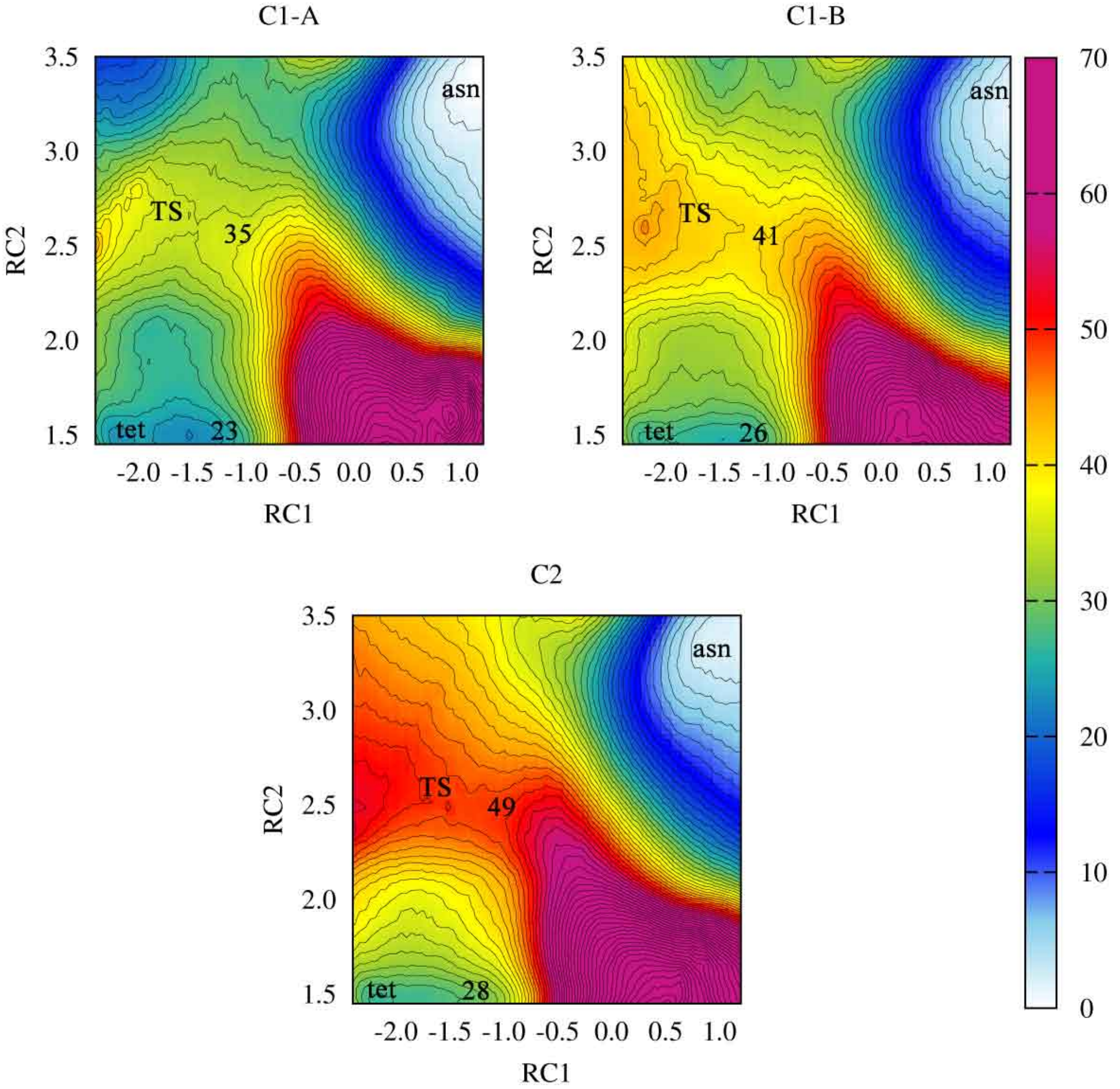


Figure 6.8. Free energy profile for the tetrahedral (tet) formation from asparagine (asn) in TPI for the samples C1-A, C1-B, and C2 (SCC-DFTB/MM-MD). Energetics are given in kcal/mol (the counter level is 1.5 kcal/mol). Regions having free energy values greater that 70 kcal/mol use the color of 70 kcal/mol.

In the case of 2° deamidation site (sample C2), the activation barrier is found to be ∼ 49 kcal/mol (Figure 6.8 c). The TS is located at (RC1, RC2) ≈ (-1.2 Å, 2.5 Å) with d1 and d2 equals to 1.0 Å and 2.2 Å, respectively. These distances indicate an asynchronous TS as it was observed in both of the samples from 1° deamidation site (C1-A and B). Although the TS is located approximately in the same region and shows similar geometrical features with either of the samples of the 1° deamidation site, the activation barrier is ∼ 14 kcal/mol and ∼ 8 kcal/mol higher than C1-A and

C1-B. The relative energy of the tetrahedral product is $\sim$ 28 kcal/mol, which is higher compared to C1-A ($\sim$ 23 kcal/mol) and C1-B ($\sim$ 26 kcal/mol). Keeping in mind that the activation barrier for C2 is higher than the rest of the samples, it can be concluded that the ratios between activation barriers and the relative energy of the intermediate are in accordance with Hammond's postulate [169].

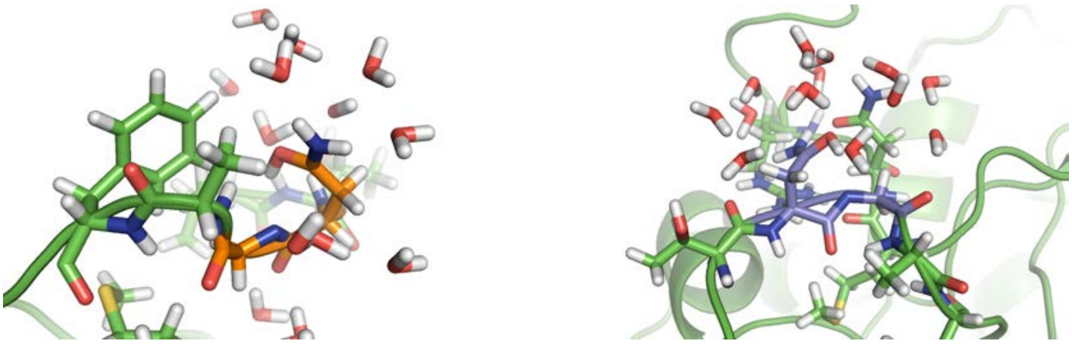
Comparing to asparagine dipeptide, the barrier of the more acidic conformation (C1-A) and the asparagine dipeptide are isoenergetic ($\sim$ 34 kcal/mol and $\sim$ 35 kcal/mol). On the other hand, the less acidic conformation of 1° deamidation site (C1-B) has 6 kcal/mol higher barrier compared to asparagine dipeptide. The tetrahedral intermediate is approximately 11 kcal/mol and 14 kcal/mol less stabilized compared to asparagine dipeptide for C1-A and C1-B, respectively. For 2° deamidation site, the barrier is 14 kcal/mol higher than the barrier found for the asparagine dipeptide, and the tetrahedral intermediate is approximately 16 kcal/mol less stabilized.

The representative frames of the transition states belonging three of the samples are shown in Figure 6.9 (a to c). In order to gain deeper aspects on the local environment of the TS, the residues within a distance of 5Å to Gly backbone nitrogen are highlighted with stick representations. In either of the cases (C1-A, C1-B, and C2) there are no residues observed that bring a strong attractive (or repulsive) interactions to the deamidation site (colored by orange, purple and magenta, respectively). In the case of 1° deamidation site, for both of the samples (C1-A and B) the reactive parts are surrounded by water molecules rather than the protein atoms. In contrast, sample C2 is found to have closer contacts with protein atoms than the water molecules indicating a more buried structure.

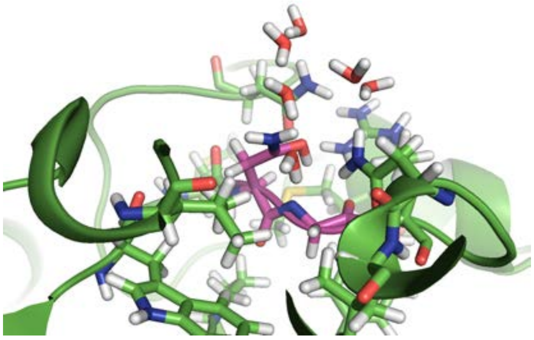
6.4. Discussions

6.4.1. Why is the half-life time of small Asn peptide smaller than TPI?

Under the same experimental conditions, the deamidation half-life times of a small peptide and of TPI are found to be 1.04 and 37.8 days, respectively. Previous



(a) C1-A
(b) C1-B



(c) C2

Figure 6.9. Representative structures of the transition states for samples C1-A (a) C1-B (b) and C2 (c). Asn-Gly residues are colored by orange (C1-A), purple (C1-B) and magenta (C2). Only the residues within a distance of 5Å to Gly backbone nitrogen are shown with stick representation, and the rest are shown with cartoon representation.

studies suggest that the slower rate of deamidation in proteins compared to their small counterparts might be caused by the less acidic character of the neighboring residue or by the lower flexibility in proteins [14, 120, 129]. The flexibility discussions involve the possible strong hydrogen bond network of the deamidation sites or being located on more rigid conformations (i.e. α helices). All of these proposals point out that deamidation rates of small peptides and proteins would be discriminated in the initiation step.

In this study, initial step of deamidation is modeled in the asparagine dipeptide in order to mimic a small peptide, consequently to compare its kinetics with TPI. Both Asn71 or Asn15 are located on flexible loops of TPI, this fact eliminates the possible effect of a more complex secondary structure (i.e. α helices). In addition, there has been no special interactions found that would hinder the formation of tetrahedral intermediate (Figure 6.9). Hence the comparison between asparagine dipeptide and the samples from TPI focus only on the relative acidity of the glycine backbone amide.

Sample C1-B or C2 have higher activation barrier (7 kcal/mol or 15 kcal/mol) compared to asparagine dipeptide (34 kcal/mol). Only the sample C1-A has an approximately isoenergetic barrier (35 kcal/mol) as asparagine dipeptide. The initial structures of C1-B and C2 are relatively less acidic compared to C1-A. Thus, if the geometrical features allow the TPI deamidation site to be acidic enough, the forward barrier (asn $\rightarrow$ tet) would be as small as the small peptide. These results are in agreement with the proposals which suggest that deamidation would be slower in proteins because of the diminished acidity of glycine amide.

On the other hand, more crucial difference between the kinetics of small peptide and TPI rise in the stabilization of the tetrahedral intermediate. This intermediate is more than $\sim$ 10 kcal/mol less stabilized for any of the samples of TPI compared to asparagine dipeptide. The tetrahedral intermediate is a five-membered heterocyclic compound. The relatively less stable intermediate in the case of TPI might be explained by the higher strain energy introduced by the long peptide chain in TPI. This result indicates that the barrier of reverse reaction (tet $\rightarrow$ asn) is much more smaller for TPI compared to the small peptide. Hence, the slower rate in TPI compared to its smaller counterpart does not only stem from the weaker reactivity (i.e. higher forward barrier), but also from poorer stability of the tetrahedral intermediate (i.e. lower backward barrier).

6.4.2. The importance of Gly backbone amide acidity on the rate of deamidation

The tetrahedral formation from asparagine at primary deamidation site of TPI was modeled using two samples; C1-A and C1-B. The main difference of these two samples is the relative acidity of glycine backbone amide, where C1-A and C1-B represent more and less acidic conformations, respectively. The activation barrier of the less acidic conformation (C1-B) is found to be 6 kcal/mol higher than the more acidic one (C1-A, Figure 6.9). In addition, the sample C2, in which tetrahedral formation is modeled at the secondary deamidation site, has 14 kcal/mol higher barrier than the sample C1-A. This sample also represents a less acidic structure of Gly16 with respect to the backbone conformations. Combining these facts, it can be concluded that, the acidity of the backbone amide has a crucial impact on the initiation of deamidation reaction as it was claimed previously.

Focusing on the primary deamidation site, the kinetic results also rise a query on the effective choice of initial geometries which are subject to umbrella sampling calculations. Most of the recent studies which were performed using umbrella sampling technique combined with QM/MM tools follow a short MD study and concentrate on only one snapshot. Our results show that the kinetics can differentiate up to 6 kcal/mol depending the intrinsic geometrical features of the samples.

6.4.3. Why does Asn71 deamidate faster than Asn15?

The barrier of tetrahedral formation is 14 kcal/mol and 8 kcal/mol smaller than Asn15 (C2) for the samples C1-A and C1-B, respectively. Considering the fact that the main difference between C1-A and C1-B is the relative acidity of the backbone amide, the great difference in the activation barrier between the sample C2 and the rest might stem from an additional reason.

As it was shown in previous molecular dynamics studies [143,158], Asn15 is more burried than Asn71 inside the protein due to their different locations in TPI (Asn15 and Asn72 is at loop1, and loop3, respectively). The greater solvation of the TS in the case of Asn71 can also be seen in the representative frames (Figure 6.9). In terms of kinetics of conversion of asparagine to tetrahedral intermediate, being in a more or less buried environment makes out an important factor considering the zwiterionic features of the TS, since the ionic species are less stabilized inside a nonpolar protein environment compared to aqueous media. This result shows that, the relatively faster reaction at Asn71 compared to Asn15 might be not only related to the relatively more acidic backbone amide, but also to the greater solvation that would stabilize the ionic TS.

6.5. Conclusions

In the present study the rate determining step (tetrahedral formation) of the overall deamidation reaction in the enzyme triosephosphate isomerase (TPI) was examined using umbrella sampling technique with QM/MM-MD tools. The same reaction was also modeled with asparagine dipeptide in order to compare the kinetics with the two different deamidation sites in TPI. This comparison lightened the impact of the high-order structure on deamidation rate. The most significant difference between the deamidation rates of TPI and the small peptide appears to be the stabilization of the tetrahedral intermediate. The tetrahedral intermediate is more than 10 kcal/mol less stabilized compared to asparagine dipeptide, possibly due to the stronger ring strain introduced by the peptide chain. To the best of our knowledge, these are the first kinetic results which relates the diminishing effect of the high-order structure with the stability of the intermediate in deamidation reaction.

The influence of glycine backbone amide acidity on the rate of tetrahedral formation was explored by performing calculations on three different samples of TPI. The activation barrier of the less acidic conformation is found be considerable higher compared to other samples. This result strengthened the proposal on the direct correlation between backbone amide acidity and the deamidation rate.

The higher barrier of the tetrahedral formation in the case of 2° deamidation site compared to both of the samples belonging to 1° one indicates that Asn71 deamidates faster than Asn15 due to its greater solvation and higher acidity of glycine backbone amide adjacent to it. These results are in a strong agreement with our previous results obtained by MD studies, exploring that the rate of deamidation of Asn15 is not only controlled by global effects (desolvation), but also the local ones (glycine amide acidity).

7. TOWARDS ACCURATE AND EFFICIENT AMINOACID pK_a PREDICTIONS USING CALCULATED ATOMIC CHARGES

7.1. Abstract

The acid dissociation constant (pK_a) of aminoacids play a crucial role in determining the chemical features of the proteins (i.e. high-order structure, reactivity, folding etc.). Although the experimental pK_a of aminoacids in aqueous solutions are well known, in protein environment the crucial shifts from the aqueous reference value are observed. The impact of these pK_a shifts on the chemical characteristics, increased the interest on predicting aminoacid pK_a's in protein. We have developed an efficient and accurate protocol to predict pK_a of a set of alcohols, phenols and thiols based on the linear regression between calculated atomic charges and experimental pK_a value. The protocol is tested with different environment approaches (gas phase, and PCM or CPCM models), with five distinct charge models (Mulliken, Löwdin, NPA, Mulliken, CHELPG), and with nine DFT-functional combined with sixteen different basis sets. As well as the high accuracy obtained from any of the methods and models, the best combination of DFT methods are found to be B3LYP/3-21G and M062X/6-311G gave the most precise predictions using CPCM model and NPA charges. The applicability of the suggested protocol is tested with tyrosine and cysteine aminoacids, and precise pK_a predictions are obtained. The stability of the aminoacid pK_a's with respect to geometrical changes also tested by MM-MD and DFT-MD calculations. Considering the strong accuracy and high computational efficiency, the pK_a prediction calculations using atomic charges indicate a promising method for predicting aminoacids pK_a in protein environment.

7.2. Introduction

Among the 20 standard amino acids that compose proteins, some of them are ionizable. Arginine (Arg), aspartic acid (Asp), glutamic acid (Glu), cysteine (Cys), histidine (His), lysine (Lys), and to some extent tyrosine (Tyr), serine (Ser), and threonine (Thr) all have a sidechain that can be protonated or unprotonated depending on acidic conditions. While the pK_a of these individual amino acids are well known in aqueous solutions [170], their pK_a can dramatically be changed in a protein environment. This change can be due to different factor: desolvation, hydrogen bonding, or charge-charge interactions [137, 145, 146, 171–173].

Desolvation effect is due to the nature of proteins. Many are globular polymers and some residues can be deeply buried inside the tertiary structure. In this case, buried amino acids are not in contact with solvent water anymore, and their pK_a is shifted to resemble more to their respective pK_a in non-polar medium (i.e., they tend to reach a neutral form) [174]. Many amino acids can be involved in hydrogen bonds either with their backbone or their sidechains. These hydrogen bonds between amino acids can also modify the amino acids sidechain pK_a' s. For example, Li and al. have shown that if an aspartic acid sidechain is involved in a hydrogen bond with a serine sidechain, its pK_a can be lowered by one pK_a unit [175] Finally, when two charged sidechains are in close contact (a.k.a., a salt-bridge), this stabilizing interaction shifts their relative pK_a' s to reflect this stable interaction: an acid becomes more acidic and a base becomes more basic [137].

Experimentally, amino acid pK_a in proteins are difficult to measure. Usually, pK_a values are assigned from titration curves obtained using NMR spectroscopy [176, 177]. In minor cases, other techniques like protein thermodynamics stability measurements can be used [174, 178] or UV spectroscopy [179]. Overall, the reported number of measured pK_a values [180] is somewhat modest compared to the number of tertiary structures available at the PDB [181].

Many theoretical approaches have been developed to estimate pK_a' s or protonation states in proteins. These methods can be divided in four main classes: i) QM based methods that use some kind of thermodynamical cycle [182–185]; ii) MM bases methods that use either alchemical modifications of ionizable residues to model (un)protonation or multiple protonation states [186–190]; iii) continuum solved based methods that use the dielectric difference between the interior of a protein and the outer solvent [191–194]; iv) knowledge based methods that use empirical parameters to estimate amino acids pK_a' s according to their respective local and global environment [137, 145, 146]. Accuracy measurement of the each model is usually expressed as an error between any set of experimental pK_a' s and their respective predicted pK_a' s. One of the problems that modellers face is a possible systematic discrepancy between the experimental values and the modelled values. In most cases, pK_a experiments are performed in aqueous solution while the model relies on 3D coordinates coming from the Protein Data Bank (PDB), which mostly contains structures obtained using X-ray or neutron diffractions. Therefore, the difference between a model pK_a and a set of experimentally measured pK_a' s can be due not only to the model but also to various other origins: variabilities of the experimental conditions within the experimental set of pK_a' s (i.e., temperature, ionic forces, presence of ligands, etc); X-ray resolutions of the proteins used in the model; packing effects in the crystallographic unit cell; sources (e.g., the organism for which a crystallographic structure exists can be different from the organism that was used for experimental pK_a measurements); etc. For some ionizable residues the number of reliable experimental data, today, can also be too small to provide a meaningful statistic.

For knowledge based methods, at least, having a trustworthy set of experimental pK_a' s accompanied with their respective protein coordinates is crucial. Ideally, the experimental pK_a measurement should be made directly on the crystallographic structures to ensure a possible direct comparison between the experiments and the model.

Among the recent works on estimating the pK_a of various organic compounds, the work by Zhang et al. is noteworthy [195, 196]. They have designed a methodology to predict pK_a values in aqueous solution from the computation of deprotonation energies.

For different acidic functional groups, they have performed linear regression fit to model pK_a' s as a function of the deprotonation energy ($E_{A^-} - E_{AH}$). They have obtained coefficients for five different acidic functional groups in the form $pKa = \alpha(E_{A^-} - E_{AH}) + \beta$ with very good correlation coefficient (R^2). They reported mean absolute deviations of ~0.4 pK_a unit and maximum error range of ±1.5 pK_a unit. Unfortunately, such approach is difficult to transpose to proteins since this methodology requires to compute the energies of both acidic (AH) and basic (A^-) states. In some proteins, steric interactions could lead to an incorrect, or nearly impossible, positioning of the proton in the AH form, especially when the basic form of an amino acid side chain is involved in a strong hydrogen bond or in a salt bridge.

To avoid any computation on the acidic form (AH), Roos et al. have suggested to use the atomic charge of the basic form (A^-) as a pK_a descriptor [197]. In the case of thiols, they have found a linear relationship between the NPA atomic charge on the sulfur atom of thiolates and the experimental pK_a of their related substituted thiols. The set of experimental pK_a was composed of (only) seven substituted thiols. Two of them had a pK_a around 3, while the other five had an experimental pK_a between 8 and 10. All thiolate forms were optimized at the B3LYP/6-31G* level using the Polarizable Continuum Model (PCM) to represent solvent effects. The linear regression fit was then applied to model the pK_a of cysteines in 3_{10}-helices and α-helices. Overall, this seminal work presented the potentiality of using atomic charges of a basic form to estimate pK_a as well as its possible application to problems of biological importance. From a set of acidic molecules (i.e, substituted thiols, alcohols or phenols), atomic charges on the corresponding anionic forms are computed to yield, after a linear regression fit, to an estimation of pK_a as a function of the atomic charges. In peptide or proteins, this linear function can then be used to estimate the pK_a of related amino acid side chains (i.e., cysteine, serine/threonine, or tyrosine side chains, respectively) after the computation of the atomic charge of the corresponding anionic forms.

In the present work, we extend the work of Roos et al. [197] to the case of substituted thiols and alcohols, including phenols. We limit our study to the linear regression step and explore how experimental pK_a' s can be related to the atomic charge of the basic forms RO^- (or RS^-) for alcohols and phenols (or thiols, respectively). Several questions are here addressed:

- can an atomic charge accurately reproduce experimental pK_a?
- which atomic charge model best fits experimental data?
- what is the influence of the basis set or DFT functional on the fitting?
- does solvent effect, using a continuum model approach, have to be included in the computation?
- what is the stability of the predicted pK_a with respect to the geometry of the molecules?

In the following sections, the methodology is presented; the linearity of the relationship between experimental pK_a and atomic charge is explored; the use of a continuum model to include averaged solvent effects is assessed; different computational protocols are compared; benchmarks using different DFT functionals and various basis sets are presented; and the stability of the present approach is analyzed. Finally, we present how accurately NDDO-based semiempirical hamiltonians can be used instead of DFT functionals to predict pK_a' s.

7.3. Experimental Database

The calculations were performed on a total 56 small organic molecules which consist of 14 phenols, 10 alcohols and 32 thiols. Phenols and alcohols were treated in one subset while thiols were in an other one. The subset containing phenols and alcohols will be simply referred as alcohols from now on. Among all of the molecules studied 19 alcohols and 25 thiols were used to construct the training set. 5 alcohols and 7 thiols were added to the test set which includes tyrosine and cysteine dipeptides.

Our fundamental criteria to choose the molecules in the subsets, is to be able to represent the widest range of experimental pK_a. Thus, in our subsets, the experimental pK_a of alcohols and thiols have a range of 3.71 to 15.92 and 5.24 to 11.22, respectively (Table 7.1 - 7.4). Within this widest range of experimental pK_a, the molecules bearing a strong electronegative atom (e.g. F, Br) were discarded. Accurate calculations on these kind of molecules can only be done using very large basis sets [198, 199]. This fact introduces restrictions to our scope of developing a computational protocol which would cover as much basis sets as possible. One molecule bearing F atoms was added only for the test set in order to check the validity of the proposed protocol (2,2,2-trifluoroethanethiol).

Most of the molecules in the subsets are not only small but also rather rigid molecules. The reason of avoiding flexible molecules is to overcome the risk of not obtaining the global minimum which would raise systematical errors in pK_a predictions as already discussed by Zhang et. al. [196]. For the few molecules which are relatively more flexible than the rest, conformational search was performed prior to the pK_a prediction calculations (the details are introduced in Computational Details Section).

7.4. Computational Details

7.4.1. Quantum Mechanical Calculations

All of the Quantum Mechanical (QM) calculations were done using Gaussian 09 [150] program package. Nine different density functionals (BLYP [222, 223], B3LYP [65, 222], OLYP [222, 224], O3LYP [222, 224, 225], PBEPBE [226], PBE1PBE [226], M06 [227, 228], M06L [228, 229], M062X [227, 228]), sixteen different basis sets were used. To interpret the aqueous solvent environment, the polarizable continuum model (PCM) and conductor model (CPCM) were used with the dielectric constant (ε) of 78.5. Five different types of atomic charge models were tested: Mulliken population analysis [83], Löwdin population analysis [84], Natural Population Analysis (NPA) [85] and two within the Electrostatic Potential (ESP) framework (Merz-Kollman model [230] and Charges from Electrostatic Potentials using a Grid based method (CHelpG) [231]). Unless otherwise stated, all the charge calculations were performed on the optimized geometries (after including or not the solvent effect).

Table 7.1. Alcohols Training Set: IUPAC nomenclature, molecule names and experimental pK_a values.

alcohols	name	pK_a
2,6-dinitrophenol	a01	3.71 [200]
2,4-dinitrophenol	a02	4.09 [200]
2,5-dinitrophenol	a03	5.21 [200]
2,5-dichlorophenol	a04	7.51 [201]
4-cyanophenol	a05	7.96 [202]
3-hydroxyquinoline	a06	8.06 [203]
3-methylsulfonylphenol	a07	8.40 [204]
5-hydroxyquinoline	a08	8.54 [203]
3-methoxyphenol	a09	9.65 [205]
phenol	a10	9.97 [200]
4-tert-butylphenol	a11	10.23 [206]
2,4,6-trimethylphenol	a12	10.87 [207]
2-methoxyethanol	a13	15.00 [208]
methanol	a14	15.20 [209]
phenyl-methanol	a15	15.44 [210]
ethanol	a16	15.50 [211]
2-propanol	a17	15.70 [209]
1-propanol	a18	15.87 [210]
1-butanol	a19	15.92 [210]

7.4.2. Molecular Dynamics Calculations

Aminoacids; tyrosine, cysteine and cysteine anions were represented in their dipeptide forms: the N and C terminal ends of the residues (tyrosine or cysteine) are substituted with N-methyl and acetamide, respectively. Parameters for phenol and phenolate were prepared using charges obtained from restrained electrostatic potential

Table 7.2. Thiols Training Set: IUPAC nomenclature, molecule names and experimental pK_a values.

thiols	name	pK_a
3-nitrobenzenethiol	t01	5.24 [212]
5-mercaptouracil	t02	5.30 [208]
4-acetylbenzenthiol	t03	5.33 [212]
3-chlorobenzenthiol	t04	5.78 [212]
4-chlorobenzenethiol	t05	6.14 [212]
benzenethiol	t06	6.61 [212]
2-methyl benzenethiol	t07	6.64 [212]
3-methyl benzenethiol	t08	6.66 [208]
4-methoxybenzenethiol	t09	6.78 [212]
4-methyl benzenethiol	t10	6.82 [212]
prop-1-ene-2-thiol	t12	7.86 [208]
ethyl-2-mercaptoacetate	t13	7.95 [213]
2-mercaptopropan-2-one	t14	7.99 [214]
N-2-Mercaptopropanoyl-glycine	t15	8.33 [215]
2,3-dimercapto-1-propanol	t16	8.62 [208]
2-ethoxyethanethiol	t17	9.38 [213]
phenylmethanethiol	t18	9.43 [213]
3-mercaptopropane-1,2-diol	t19	9.51 [213]
2-mercaptoethanol	t20	9.72 [213]
2-mercapto-2-methyl-1-propanol	t21	9.85 [216]
prop-2-ene-1-thiol	t22	9.96 [213]
methanethiol	t23	10.33 [216]
ethanethiol	t24	10.61 [217]
butane-1-thiol	t25	10.67 [213]
2-propanethiol	t26	10.86 [218]

Table 7.3. Alcohols Test Set: IUPAC nomenclature, molecule names and experimental pK_a values.

alcohols	name	pK_a
2,3,4,5,6-pentachlorophenol	a20	5.62 [219]
2,6-dichlorophenol	a21	6.79 [208]
tyrosine	a22	9.84 [220]
prop-2-yn-1-ol	a23	13.60 [211]
3-hydroxypropanenitrile	a24	14.03 [210]

Table 7.4. Thiols Test Set: IUPAC nomenclature, molecule names and experimental pK_a values.

thiols	name	pK_a
2,2,2-trifluoroethanethiol	t27	7.30 [221]
cysteine	t28	8.55 [220]
2-mercaptoacetic acid	t29	10.40 [208]
2-aminoethanethiol	t30	10.53 [208]
2,3-dimercaptopropan-1-ol	t31	10.57 [208]
2-methylpropane-2-thiol	t32	11.05 [218]
2-methylbutane-2-thiol	t33	11.22 [218]

(RESP [232]). The RESP charges were calculated using B3LYP/cc-pVTZ level of theory [233] on the geometries optimized at HF/6-31G**. Topology and coordinate files of these systems were prepared with tleap module of AMBER 12 program package [42]. Dipeptides were treated with ff03 forcefield [134]. GAFF forcefield of AMBER12 [42] was used for phenol, and phenolate. The polar environment was represented by modified generalized Born model [234]. Following minimization, the systems were heated up to 300 K using the Anderson coupling algorithm [47] during 50 ps with velocities updated every 1000 steps. Heated systems were equilibrated up to 100 ps. Using the same temperature algorithm, productions were performed for 150 ps.

Starting from classically equilibrated structures, molecular dynamic simulations were also performed using DFT potentials (DFT-MD). These calculations were accomplished using the external interface between AMBER 12 and Gaussian program packages [235]. B3LYP/3-21G was used for phenol and tyrosine, M062X/6-311G was used for cysteine. Choice of the DFT's will be detailed in the results and discussion section. DFT-MD simulations were performed using the protocol introduced for MM-MD calculations and productions were performed up to 50 ps.

7.5. Results and Discussions

We investigated the linear relationship between atomic charges and experimental pK_a for our two training sets of molecules representing various substituted thiols and alcohols. From the methodology suggested by Roos et. al. (*i.e.* computing the atomic charge of the basic form RS^- and RO^- respectively [197]) we found out that the choice of the computational framework plays an important role in the accuracy of the pK_a prediction. From the overall present study, we show that the combination of charge and solvent models giving the best results is NPA with CPCM for both thiols and alcohols sets. M062X/6-311G and B3LYP/3-21G are the DFT methods which are the most reliable for the present study of thiols and alcohols respectively. In what follows, we first show the linear relationship between experimental pK_a and atomic charges computed using the theoretical frameworks discussed above. Thus, using these results as a reference, we shall discuss the choice of charge model, DFT functional, basis-set and solvent model by changing one of these parameters while the others remain fixed to the best combination.

7.5.1. The Linearity of the Relationship Between Experimental pK_a and Atomic Charges

Anionic forms of all the molecules in two training sets (RO^- and RS^-) were optimized in aqueous phase (CPCM, $\varepsilon = 78.5$) using DFT (B3LYP/3-21G and M062X/6-

311G for alcohols and thiols, respectively). With the same solvent model, Natural Population Analysis were performed on these optimized geometries using the same DFT for the corresponding subsets ($O^-//O^-$ and $S^-//S^-$). The NPA charges of the atoms O^- and S^- were associated with the experimental pK_a's of each molecule in order to establish a standard linear fit. This method is referred below as method0. Figure 7.1 shows the relationship between experimental pK_a and computed NPA charge for the two training sets. A linear equation is obtained by a least-squares fit

$$pK_a = a * q(X^-) + b \tag{7.1}$$

where a and b are the fitted parameters, and $q(X^-)$ is the calculated charge on $q(O^-)$ and $q(S^-)$ for alcohols and thiols, respectively. The parameters a, b and the squared Pearson correlation coefficient (R^2) are also illustrated on Figure 7.1. The predicted pK_a's are computed using Equation 7.1 (i.e. by importing $q(X^-)$ of a given molecule into the parameterized equation).

For alcohols and thiols the R^2 values were found to be 0.995 and 0.986, respectively. For both of the training sets, no strong outlier molecules were observed. The maximum difference between theoretical and experimental pK_a among all the molecules, was found as 0.77 and 0.48 units for alcohols and thiols, respectively. These results indicate a highly strong correlation with the experimental pK_a and the charges on O^- and S^-. In order to analyze the influence of the charge model on the quality of the fit, the same protocol was applied with four other charge models.

7.5.2. Influence of the Charge Model

Fixing the DFT methods and the solvent model for both sets of molecules to: M062X/6-311G/CPCM for thiols and B3LYP/3-21G/CPCM for alcohols, alternative charge calculations were performed on the same optimized geometries that were used for NPA charge calculations. As it was done in the case of NPA, the charges on O^- and S^- of the anions were extracted to establish a linear fit with the experimental pK_a's. The results for alternative charge models (Mulliken, Löwdin, MK and CHelpG) are

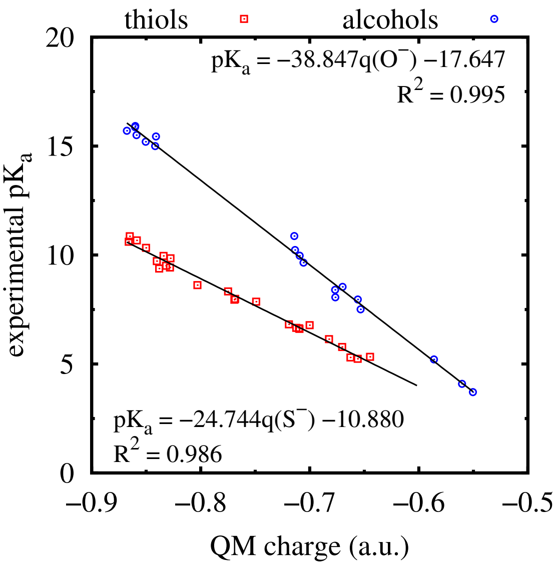
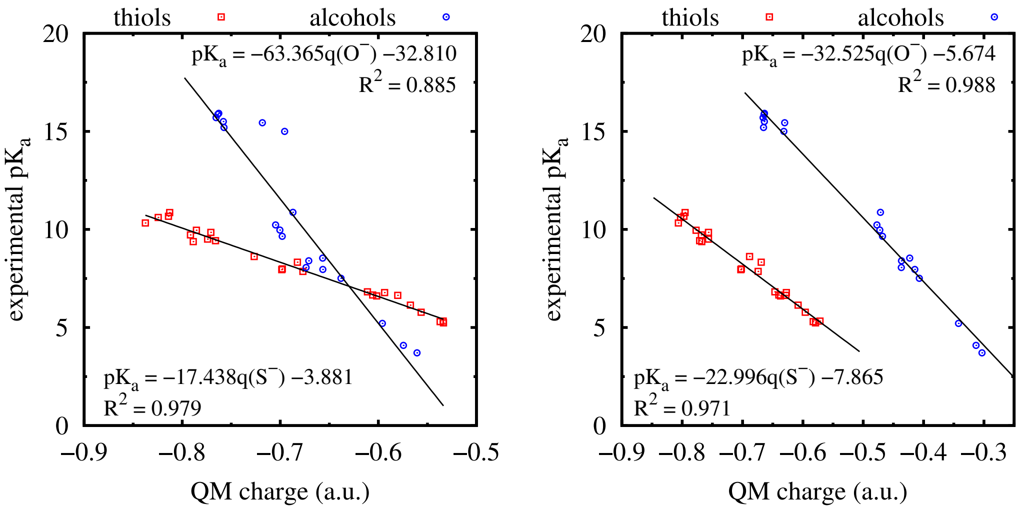


Figure 7.1. Linear regression between calculated NPA charges and experimental pK_a. Calculations were done using B3LYP/3-21G//CPCM and M062X/6-311//CPCM for alcohols and thiols, respectively.

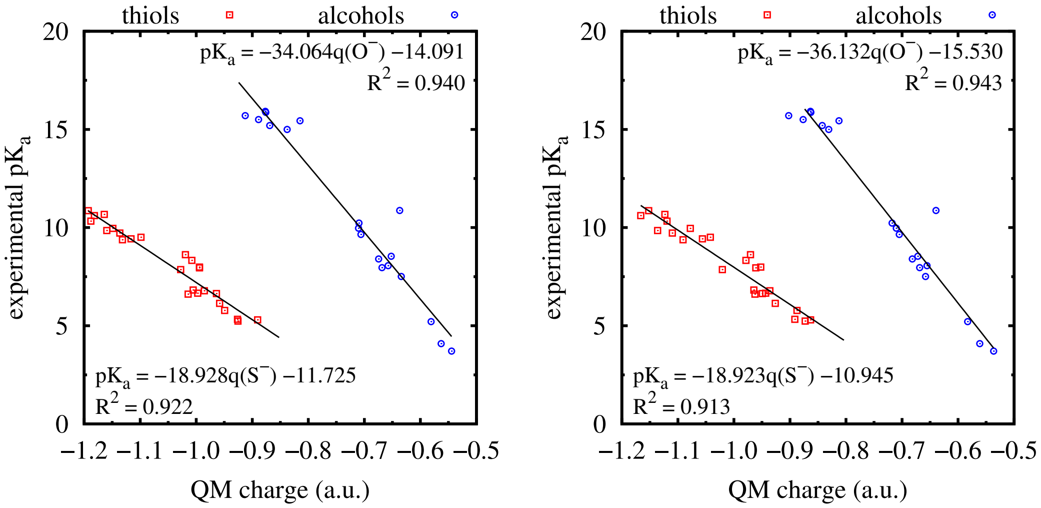
represented in Figure 7.2.

All of the tested charge models appear to have a strong correlation with the experimental pK_a, having $0.885 \leq R^2 \leq 0.988$. The smallest R^2 is found with Mulliken population analysis in alcohols subset (0.885, Figure 7.2 a). For this model, some molecules were found to be spreaded away from the fitting line, especially in the experimental pK_a range of 5 to 15. Using Equation 7.1, the theoretical pK_a of the strongest outliers were found to be 3.78 and 2.73 unit different than the experimental value (the molecules with the experimental value of 15.00 and 15.44, respectively). Again for the alcohols subset, better results were obtained using Löwdin charges with considerably high value of R^2 (0.988). The accurate predictions of Löwdin charges with alcohols have already been discussed in a computational study which focuses on pK_a predictions of substituted phenols [236]. The computed atomic charges for alcohols are similar using the MK and CHelpG models, meaning that the quality of the prediction is mainly leaded by the ESP framework. Although both of the methods have high R^2 values (MK = 0.940, and CHelpG = 0.943), noticeable amount of molecules deviates from the linear fit. The maximum deviation was observed for the molecule having



(a) Mulliken

(b) Löwdin



(c) MK

(d) CHELPG

Figure 7.2. Effect of charge model on the linear regression between calculated charges and experimental pK_a. Calculations were done with B3LYP/3-21G//CPCM and M062X/6-311//CPCM for alcohols and thiols, respectively.

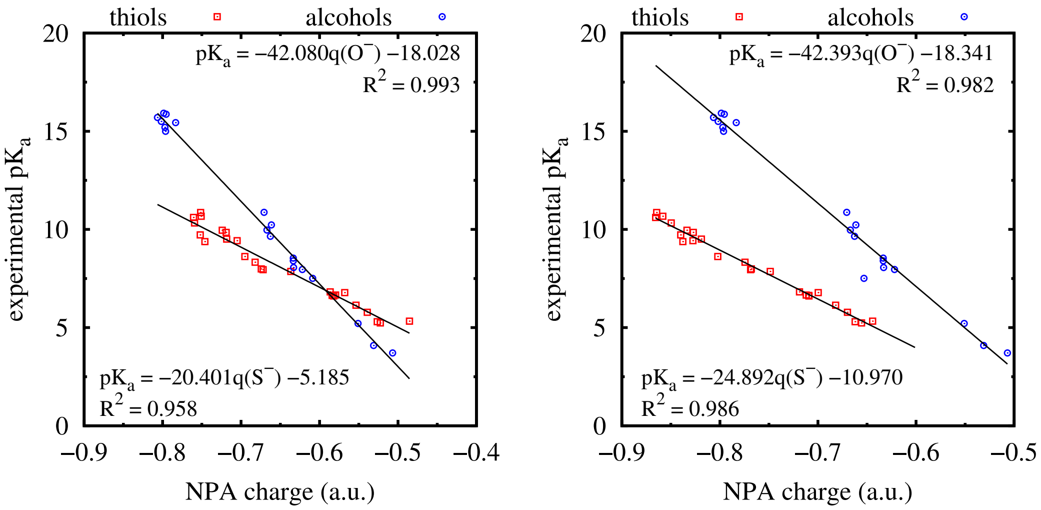
experimental pK_a = 10.87, and the difference with the predicted value was found to be 3.18, and 3.38 for the methods MK and CHelpG, respectively.

In the case of thiols, Mulliken, Löwdin and the two ESP methods (MK and CHelpG) all give quite high values of R^2 : 0.979, 0.971, 0.913 and 0.922 respectively. Interestingly, considering both the R^2 values and the spreading around the linear fit function, Mulliken and Löwdin methods give more adequate results compared to ESP methods.

To summarize, all of the tested charges generate acceptable linear fit with experimental pK_a. Nevertheless, the NPA methods appears to be the best when Figure 7.1 and 7.2 are compared.

7.5.3. Solvent Models

Zhang et. al. [196] proposed CPCM to represent solvent environment in their work focusing on pK_a prediction from deprotonation energies of the molecules. Roos et. al. used PCM for pK_a prediction calculations using QM charges. In addition to our CPCM calculations which were introduced up to this point, here the accuracy of PCM model was tested. Gas phase calculations were also performed considering its smaller computational cost compared to either of the solvent methods. Figure 7.3 presents the results of the linear regression of gas phase and PCM calculations using NPA charges and the DFT methods discussed in the previous sections.



(a) Gas phase (b) PCM

Figure 7.3. Solvent effect on the linear regression between NPA charges and experimental pK_a. Calculations were done with B3LYP/3-21G//CPCM and M062X/6-311//CPCM for alcohols and thiols, respectively.

Either gas phase or PCM calculations are as accurate as CPCM calculations with $R^2 \geq 0.958$ (Figure 7.3, 7.1). The lowest R^2 was observed for the gas phase calculations of thiols (0.958). The results are fairly accurate in the case of alcohols ($R^2 = 0.993$) in gas phase. Going from gas phase to PCM solvent model, the accuracy for thiols is increased ($R^2 = 0.986$). Interestingly, in the case of alcohols, PCM calculations yield less accurate results ($R^2 = 0.982$) compared to gas phase calculations. Comparing all of these three methods (gas phase, PCM, and CPCM) the best linearity is obtained when the CPCM solvent model is used.

7.5.4. DFT Functional and Basis-set Benchmarks

For a bench of DFT functionals and basis-sets, the same protocol was applied to our two training sets (optimization and NPA charge calculations on RO^- and RS^-, using CPCM solvent model). R^2, a and b values were obtained from the linear fit with experimental pK_a. In Figure 7.4, the Mean Absolute Deviation (MAD) of all the considered DFT functionals as a function of the selected basis-set are presented (box representations). For each combination of DFT functional and basis set, the difference between the experimental and predicted pK_a was calculated (ΔpK_a). The maximum value of this difference (MAX-ΔpK_a) is also shown in Figure 7.4 (black colored lines). The highest three ΔpK_a values and also the R^2's were tabulated in Appendix B and C for each DFT method.

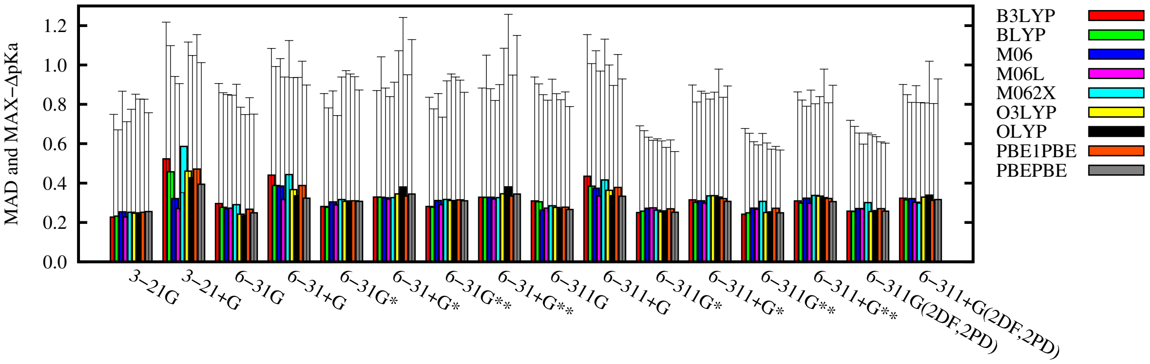
For both of the subsets, all of the DFT methods result a strong correlation between atomic charges and experimental pK_a with high R^2 and small MAD (or MAX-ΔpK_a) values. The R^2 was found to be $0.974 \leq R^2 \leq 0.995$ and $0.941 \leq R^2 \leq 0.986$ for alcohols and thiols, respectively (Appendix B and C). In the case of alcohols, MAD and ΔpK_a values are in a range indicating high accuracy ($0.227 \leq$ MAD ≤ 0.586, $0.560 \leq$ MAX-Δp$K_a \leq 1.217$, Figure 7.4). These descriptors are even smaller for thiols ($0.190 \leq$ MAD ≤ 0.350, $0.476 \leq$ MAX-Δp$K_a \leq 0.975$, Figure 7.4, b). For thiols, the quality of the results is very stable and is almost the same for all of the considered methods. Either for alcohols or thiols, increasing the basis set or adding difuse and/or polarization functions did not cause a significant improvement in accuracy. Moreover, in the case of alcohols, the accuracy is diminished when diffuse functions were added to the basis set, for any of the DFT functionals (e.g. 6-311+G* has higher MAD and MAX-ΔpK_a compared to 6-311+G*).

For alcohols, the smallest MAD's were found for the combination of all functionals with 3-21G basis set, and among all the tested methods B3LYP/3-21G give the most accurate result having MAD = 0.227. The MAX-ΔpK_a value for this subset is well predicted by 6-311G* and 6-311G** basis sets (with the smallest MAX-ΔpK_a of 0.560, PBEPBE/6-311G*). The highest R^2 (0.995) is equivalent for B3LYP/3-21G, M06L/3-21G and BLYP/6-311G**. To conclude, for alcohols, B3LYP/3-21G appears to be the most accurate method having the smallest MAD, highest R^2 and an acceptable value of MAX-ΔpK_a (0.749).

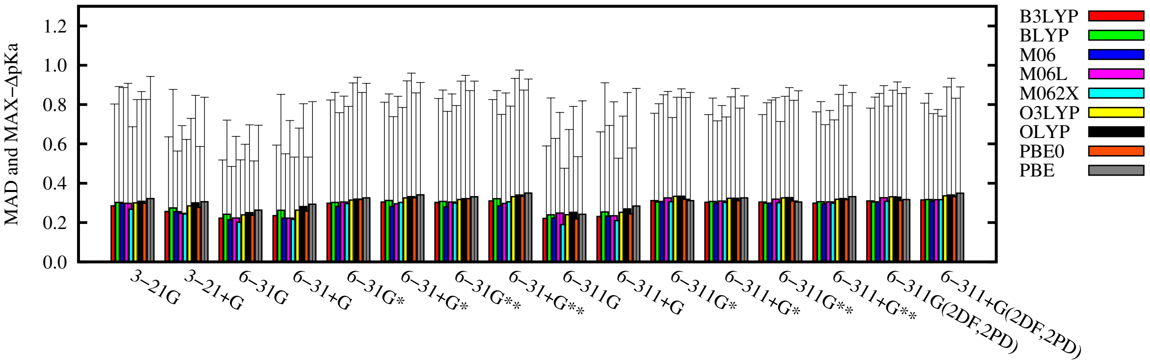
In the case of thiols, the smallest MAD and MAX-ΔpK_a, also the largest R^2 were found with M062X/6-311G (0.190, 0.476, and 0.986, respectively).

The methodology we propose in this paper aims at predicting accurate pK_a, not only for small organic molecules, but as a final goal, for aminoacids in proteins. Thus, both the accuracy of the prediction and the computational cost should be considered. Two representative molecules (1-butanol and 4-chlorobenzenethiol), which contain a number of atom representing the average of the corresponding subset, have been choosen and the MAD of all methods as a function of computational time have been plotted for these two molecules (Figure 7.5). The basis-sets were divided to four classes depending on bearing a diffuse or polarization function, both or none of them. As discussed above, for both of the training set, adding diffuse or/and polarization (red, blue, black points) functions did not the improve the accuracy. For alcohols, basis-sets containing only diffuse functions (red points) give systematically the worst results. The most accurate combination of DFT functionals and basis sets for alcohols and thiols (B3LYP/ 3-21G and M062X/6-311G, respectively) were also found to be one of the least time consuming methods.

In order to have an overview on the efficiency of the DFT functionals and basis

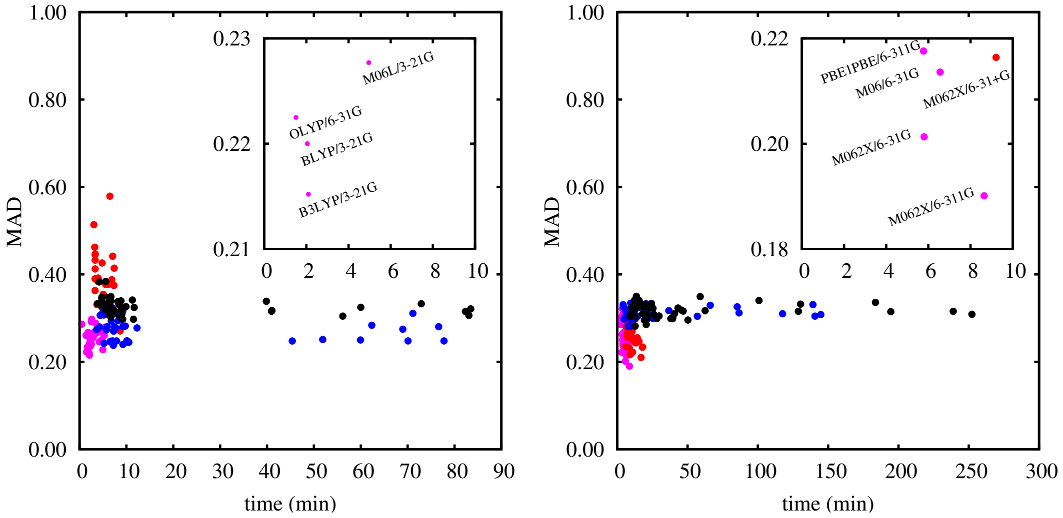


(a) alcohols



(b) thiols

Figure 7.4. MAD and MAX-ΔpK_a overall the DFT functionals and basis sets. Optimizations and NPA charge calculations were done using CPCM model.



(a) alcohols

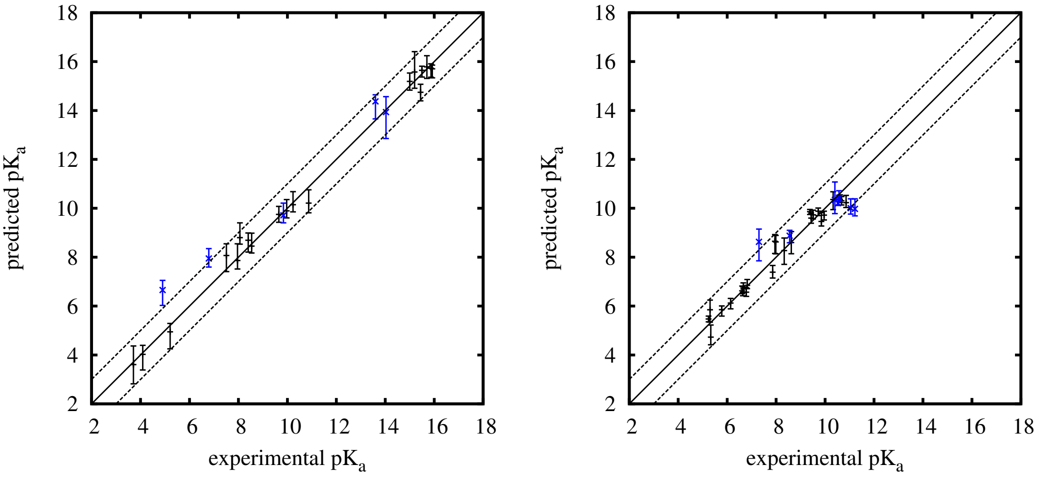
(b) thiols

Figure 7.5. MAD of pK_a for all of the DFT functionals and basis sets (solvation model CPCM, NPA charges). Red, blue, black and magenta colored points respectively represent diffuse, polarization, both diffuse and polarization, and none of them. The region of the smallest MAD and time values are shown in inset.

sets, the average predicted pK_a over all the methods was calculated. The minimum and maximum predicted pK_a among all the methods have been added to the predicted pK_a of each molecule as error bars. The predicted pK_a is plotted versus experimental values (Figure 7.6, black colored lines). For both alcohols and thiols, average, minimum and maximum value of the theoretical pK_a was found to be within the range of ± 1 unit compared to the experimental value.

The validity of the obtained protocol was tested using two small test sets which consist of 5 alcohols and 7 thiols. These test sets also contain tyrosine and cysteine dipeptides. Average, minumum and maximum values of theoretical pK_a were also calculated for these test sets (Figure 7.6, blue colored lines).

Accurate predictions were obtained for most the of the molecules in the test set. Moreover, the average pK_a of tyrosine and cysteine were calculated only -0.13 and +0.33 units different from the experimental pK_a. The maximum pK_a's were found to be 0.37 and 0.45 units more than the experimental pK_a's of tyrosine, and cysteine,



(a) alcohols (b) thiols

Figure 7.6. The predicted pK_a overall the DFT functionals and basis sets versus experimental pK_a (solvation model CPCM, NPA charges). Points show the average pK_a and the error bars denotes minimum and maximum predicted pK_a. Black and blue lines belong to the training and test set, respectively.

respectively. The minimum of the calculated pK_a overall the DFT functionals and basis sets is 0.45 unit smaller for tyrosine and 0.04 unit higher for cysteine compared to experimental data. For the best combination of DFT functionals and basis sets, the calculated pK_a's were found to be 0.05 and 0.39 units higher than the experimental one (B3LYP/3-21G and M062X/6-311G for tyrosine and cysteine, respectively). Thus, it can be concluded that the suggested protocol gives precise results for the predictions of aminoacid pK_a's.

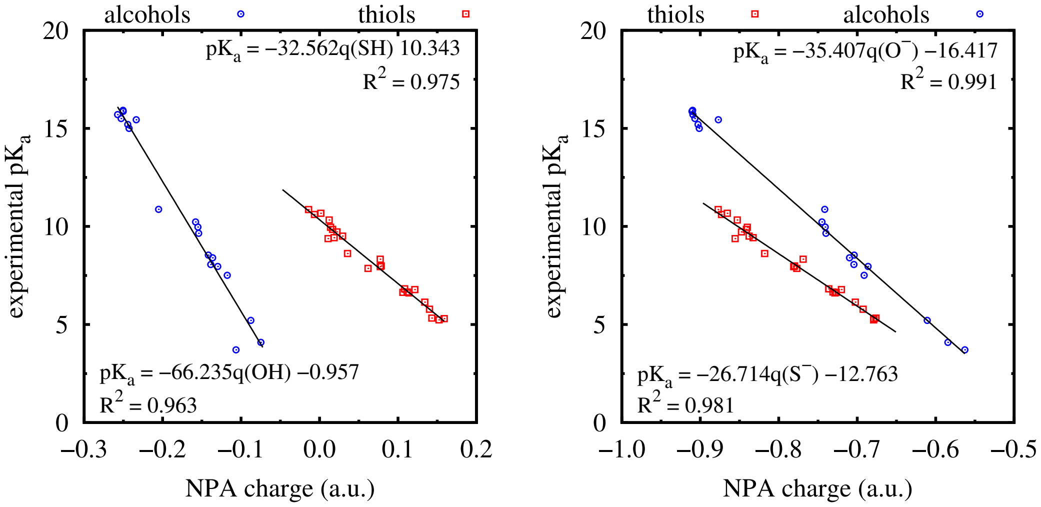
Few molecules in the test set have a deviation higher than 1 pK_a unit. For example, in the case of alcohols, pentachlorophenol and 2,6-dichlorophenol (with the experimental pK_a of 5.62 and 6.79) were found to be out of the ± 1 range limit. However, for the corresponding molecules the average value of the predicted pK_a was found to be very close to the minimum and maximum predicted values (error bars). This fact shows that all of the DFT methods result similar pK_a predictions. Thus, for these two molecules, we have reasonable doubt that the experimental pK_a could be inaccurate. For thiols, similar shift was observed for 2,2,2-trifluoroethanethiol (experimental pK_a equals to 7.30). To the opposite of the outliers in the alcohols subset, the error bars appear to be large, indicating that the pK_a predictions of this molecule is DFT method dependent. As it was discussed in Experimental Database Section, this protocol would not be sufficient enough to calculate the pK_a of molecules bearing strong electronegative atoms.

7.5.5. Geometries and Stability

In order to transfer the obtained pK_a prediction protocol to protein calculations, the different features of proteins in terms of ionizability should be considered. In the crystal structure of proteins, depending on the protein environment, ionizable aminoacids can be found either in their protonated or deprotonated forms. All the calculations introduced up to here were performed considering only the anionic forms of the molecules. Here, a similar protocol is tested for the molecules optimized in their neutral form. All of the calculations were performed with CPCM solvent, NPA charges and the DFT methods introduced above (B3LYP/3-21G and M062X/6-311G for tyrosine and cysteine, respectively). The charges are calculated in two ways:

- method1: For each optimized neutral alcohols (or thiols) the extracted atomic charge corresponds to the sum of the charge on the oxygen (or sulfur) and its bonded hydrogen. This method is referred below as X(0)//X(0) (X = O or S)
- method2: The atomic charges that enter the fitting procedure are defined as the charge of the oxygen (or sulfur) atom on its anionic form. This method is referred below as X($^-$)//X(0) (X = O or S), meaning that the atomic changes is extracted from an anion form after optimizing the corresponding neutral form.

In the case of alcohols, method1 gives a fit with R^2 equals to 0.963 (Figure 7.7 a). For the same method, R^2 was obtained as 0.975 for thiols (Figure 7.7 b). When the fit was performed on the charges obtained from single point calculations on the anions (method2), alcohols and thiols generated a fit with R^2 values 0.991 and 0.981, respectively. As a comparison, for both alcohols and thiols, the best fit was obtained by the protocol discussed in previous sections (method0) with R^2 values of 0.995 and



(a) method1 X(0)//X(0

(b) method2 X($^-$)//X(0)

Figure 7.7. The effect of geometries on pK_a predictions. X = O or S. B3LYP/3-21G and M062X/6-311 for alcohols and thiols, respectively. Solvent model CPCM, charge method NPA.

0.986, respectively. However, both method1 and method2 also give fairly good fit with the R^2 values higher than 0.98.

Another crucial point on pK_a predictions of proteins is the stability of the calculated pK_a's with respect to conformational changes. As a way to access the variability of pK_a prediction with respect to conformational changes, we have performed molecular dynamics (MD) simulations to provide multiple conformations and the optimum geometries. Short MM-MD simulations (150 ps) were performed for phenol, tyrosine and cysteine molecules using continuum solvent model. The NPA charges of X atom (which corresponds to O and S for alcohols and cysteine, respectively) were calculated in three ways in order to represent the methods introduced above:

(i) method0 (X($^-$)//X($^-$)): Deprotonated (anionic) forms of the molecules were simulated with MM-MD (X($^-$)). Single point NPA charge calculations were done on these anionic forms (X($^-$)).

(ii) method1 (X(0)//X(0)): MM-MD simulations on protonated (neutral) forms (X(0)) were followed by single point charge calculations on the same form (X(0)).

(iii) method1 (X($^-$)//X(0)): Hydrogen atoms were removed from the neutral geometries obtained from MM-MD calculations (X(0)). The single point charge calculations were done on these anionic forms (X($^-$)).

These protocols were performed on 50 frames extracted from simulations (between 100 and 150 ps) for each method and molecule. The predicted pK_a's were found using a and b values obtained from the fit with calculated charges and experimental pK_a for the corresponding method (Figure 7.1 and 7.7). All of the charge calculations were performed using CPCM with B3LYP/3-21G and M062X/6-311G methods for alcohols (phenol, phenolate, and tyrosine), and thiols (cysteine and cysteine anion), respectively. The fluctuations of calculated pK_a with respect to conformations were monitored by taking the experimental pK_a as a reference (Figure 7.8). The numerical average over all the frames were calculated (blue line).

In the case of phenolate (method0), the predicted pK_a's deviate within the range of -1 to 1 unit with respect to experimental pK_a (Figure 7.8 a). The average of these points yields a pK_a prediction which is very close to the experimental pK_a (0.04 unit). For the neutral form of this molecule (phenol), when the charges of both oxygen and hydrogen is considered (method1) the deviations were observed in a greater range (from -1.5 unit to 2 units, Figure 7.8 b). The predicted pK_a is 0.45 unit smaller than the experimental one. The best predictions were obtained with method2; the deviations were observed within 0.5 unit and the average pK_a was found to be 0.02 less than the experimental one (Figure 7.8 c). Overall, in the case of phenol, method0 and method2 can be considered as stable: pK_a prediction does not vary significantly with small geometrical deviations.

In the case of tyrosine, method0 was not studied since there are no available parameters for tyrosine anion within the AMBER predefined library of the parameters. For the other two methods, method1 results a wider spread in the calculated pK_a compared to method2, similar to the conlusions obtained in the case of phenol (Figure 7.8 d and e). The relatively wider fluctuations in the case of tyrosine indicate the effect of peptide backbone conformations on charge calculations. For method1, and 2; the average pK_a's are 0.74 unit and 0.41 unit smaller than the experimental one, which could be considered quite precise for the latter method.

In the case of cysteine, all of the three methods give relatively wider results compared to the two other molecules tested (Figure 7.8 f to h). Despite the large fluctuations, the average values are fairly close to the reference, which are -0.05, -0.20, and 0.07 unit different than the experimental pK_a for method0, method1 and method2.

Combining the results for these three molecules, method0 and method2 give highly accurate and stable results. Hence, no matter if the simulation is performed in neutral or anionic form, the single point charge calculations should be performed on the anionic form of the molecule.

In order to test the accuracy of the structures obtained from classical forcefields, the same kind of calculations were performed on the structures which were simulated using DFT based MD. For the DFT methods M062X/6-311G was used for cysteine, and B3LYP/3-21G for alcohols (phenol or tyrosine). Different from MM-MD calculations, DFT-MD simulations were performed in gas phase. All the three methods (method0-2) were tested, and the results are shown in Figure 7.9.

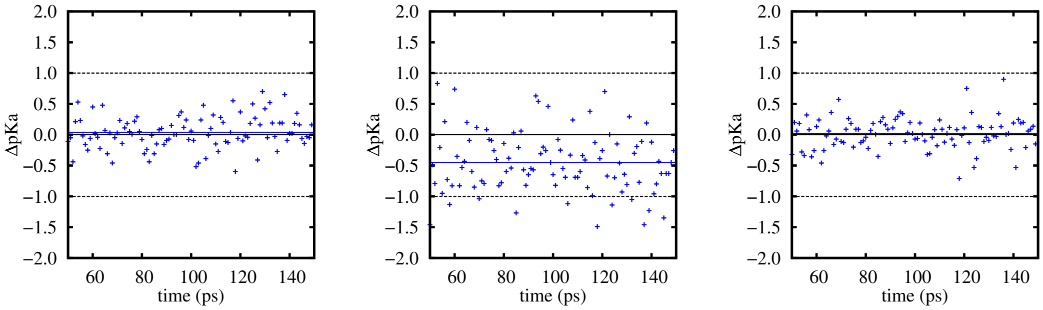
For phenol, the pK_a deviates between -1 to 0.5 units in method0. A wider range of deviations were observed for method1 (-1.5 to 2 units). Among all the three methods, method2 deviates least compared to experimental pK_a (-0.5 to 0.5). The average calculated pK_a's were found 0.38, 0.35 and 0.17 unit less than the experimental pK_a for method0, method1 and method2, respectively. Tyrosine behaves in a similar manner resulting 0.31 and 0.23 unit smaller predicted pK_a than the experimental one for method1 and method2 (Figure 7.9 d and e). For cysteine, all of the three methods deviates from the experimental reference within a considerably wider range (Figure 7.9 f to h). Correspondingly, the average calculated values were found to be imprecise: 0.65 unit less than the experimental pK_a for method0, and 0.79/0.71 higher than for method1/method2.

To conclude, for all of these three molecules, the narrowest range of deviations of the predicted pK_a's were obtained using method2 on the geometries obtained from classical forcefield calculations (MM-MD). The average pK_a of these structures also yield the closest prediction to the experimental value (-0.01, -0.41 and 0.07 unit deviation for phenol, tyrosine and cysteine). Calculating geometries with DFT forcefields did not improve the precision in most of the cases, in fact the accuracy is lost in the case of cysteine. At this point it should be noted that, DFT-MD calculations were performed in gas phase. The deviations of charges could be caused by the difference between gas and aqueous phase geometries. For either MM-MD or DFT-MD geometries, method1 yields the least accurate predictions for all of the three molecules.

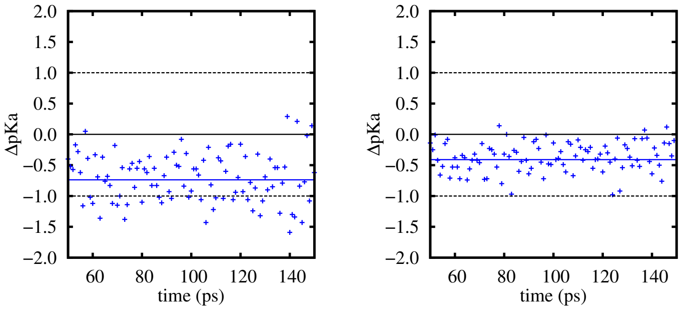
7.5.6. Conclusions

In the present study, we have suggested a protocol in order to obtain accurate and fast pK_a predictions of small alcohols and thiols. In addition, we have tested their applicability to predict aminoacid pK_a's. The suggested protocol is based on the linear regression of the the experimental pK_a with the atomic charges on the ionizable groups. Five charge models, two solvent models, gas phase calculations and several DFT methods (combination of nine DFT-functionals and sixteen basis sets) were tested, and all of these methods resulted with a strong linearity having R^2 higher than 0.8. Among those, NPA charge calculations performed on CPCM solvation model followed by optimizations in CPCM give the most accurate results. The best combination of DFT-functionals and basis sets are found to be B3LYP/3-21G and M062X/6-311G for alcohols and thiols, respectively. The energy optimization and NPA charge calculation procedures using these two most accurate DFT-methods (with CPCM) came up with an acceptable computational expense.

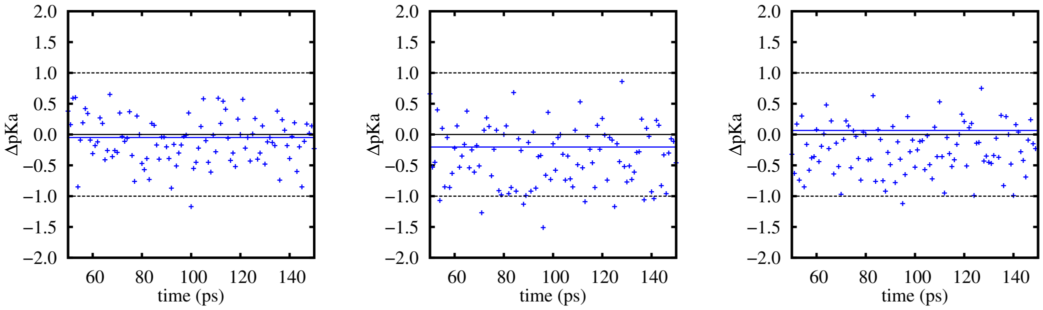
Among all the ionizable aminoacids, tyrosine and cysteine were chosen to be modeled for pK_a predictions. Using the best DFT combination, the difference between experimental and theoretical pK_a was found to be +0.05 and +0.39 unit for tyrosine, and cysteine, respectively. The average predicted pK_a over all the DFT-methods are -0.45 (tyrosine) and +0.04 (cysteine) unit different than the experimental value. MM-



(a) phenol method0 (b) phenol method1 (c) phenol method2

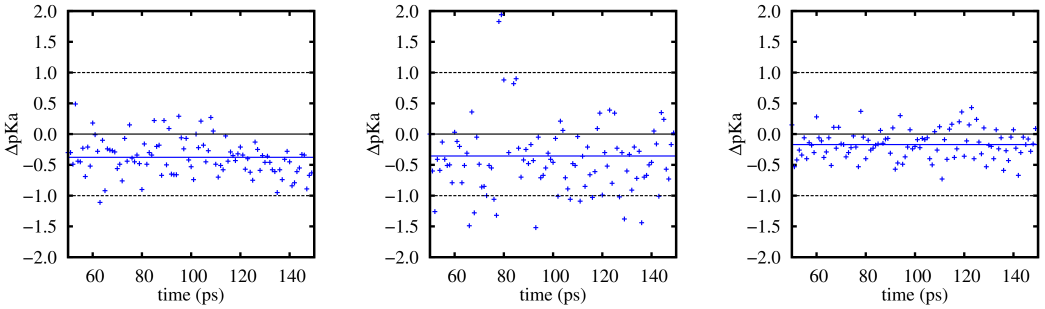


(d) tyrosine method1 (e) tyrosine method2

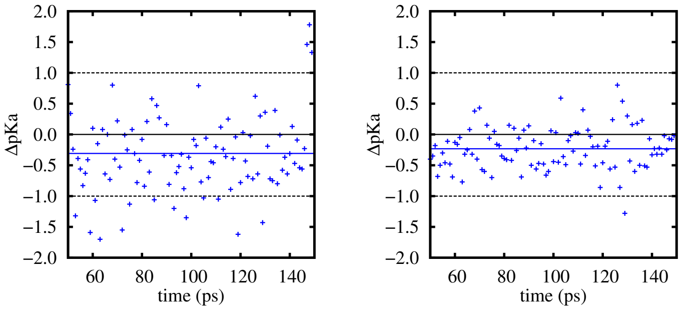


(f) cysteine method0 (g) cysteine method1 (h) cysteine method2

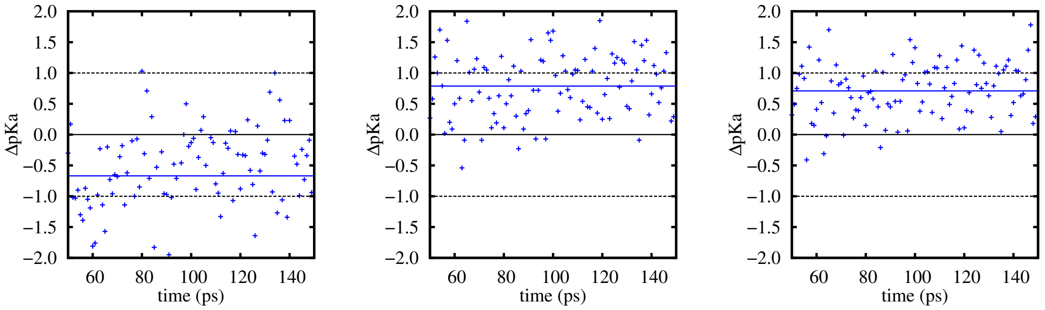
Figure 7.8. Deviations of predicted pK_a with respect to conformations (aqueous phase MM-MD, modified generalized Born model). B3LYP/3-21G and M062X/6-311G methods for alcohols (phenol and tyrosine), and cysteine were used for single point NPA charge calculations with QM (CPCM). The blue line shows the numerical average of the pK_a deviations.



(a) phenol method0 (b) phenol method1 (c) phenol method2



(d) tyrosine method1 (e) tyrosine method2



(f) cysteine method0 (g) cysteine method1 (h) cysteine method2

Figure 7.9. Deviations of predicted pK_a with respect to conformations (gas phase DFT-MD). B3LYP/3-21G and M062X/6-311G methods for alcohols (phenol and tyrosine) and cysteine were used for both DFT-MD and single point NPA charge calculations with QM (CPCM). The blue line shows the numerical average of the pK_a deviations.

MD calculations showed that the predicted pK_a's deviate within ± 1 unit with respect to different geometrical conformations. The most accurate predictions were obtained when the charge calculations were performed on the anionic species. The average theoretical pK_a over these different conformations yielded precise predictions (-0.41 and 0.07 for tyrosine and cysteine, respectively). Combining these results, the suggested protocol is found to be a promising method in order to predict pK_a of aminoacids, and the transferability to other ionizable aminoacids should be further analyzed.

8. GENERAL CONCLUSION

The findings of the present study provide explanations on the sequential deamidation in triosephosphate isomerase (TPI), the most plausible route of this reaction, and the impact of the high-order structure on deamidation.

Previous QM studies suggested several reaction mechanisms for deamidation reaction. Herein, it has been shown that for both pentapeptide (Gly-Gly-Asn-Gly-Gly) and the two deamidation sites of the TPI, there are only two plausible initial steps: tetrehadral intermediate formation, and tautomerization with two active water molecules.

According to global desolvation analysis, Asn15-Gly16 is found to be more buried (desolvated) in the enzyme, due to the less flexible nature of the loop on which this residue is located. Considering the enhancing effect of the active water molecules in deamidation reaction, the more desolvated residues, Asn15 in this case, would undergo a slower reaction. The comparison between the 1° (Asn71-Gly72) and 2° (Asn15-Gly16) deamidation sites showed the significance of global desolvation, and thereby the impact of the tertiary structure.

More buried residues are not only exposed to less water molecules. Indeed, they are known to have greater pK_a shifts. The more buried Gly16 in the 2° deamidation site of TPI would bear a higher pK_a, consequently it will be deprotonated slower compared to Gly72.

In order to obtain a more local perspective, the correlation between backbone amide conformations and the glycine amide acidity was analyzed. It is noteworthy that this analysis and the tool used for global desolvation are completely independent. However, both descriptors lead to the same conclusion: the Gly16 backbone amide would be less acidic compared to Gly72. This result strengthened the claim relating slower deamidation at Asn15 with the the less acidic nature of the adjacent Gly (Gly16). Moreover, this phenomenon is also proved by the kinetical studies. The activation barrier of Asn15-Gly16 is found to be approximately 14 kcal/mol higher compared to a relatively more acidic conformation of Asn71-Gly72.

The formation of tetrahedral intermediate does not only involve the deprotonation of the glycine backbone amide, but also requires the nucleophilic attack of this activated nitrogen to the Asn side chain. Compared to Asn15-Gly16, Asn71-Gly72 was found to have a higher probability to bear the geometrical features of a transition-state like structure (near attack conformations). This result indicates a thermodynamically less stable nature of the primary deamidation site, and consequently a more feasible reaction.

Asn71 deamidates faster than Asn15 due to several independent factors introduced above, however this residue deamidates much more slower compared to a small peptide. The classical molecular dynamical calculations performed on the pentapeptide showed that differences in solvation is one of the reasons that controls the rate of deamidation at Asn71. Although this deamidation site is found to be more solvated compared to 2° one, its solvation is conformation dependent, and not all the conformations are solvated as much as a free peptide. In addition to solvation criteria, according to our kinetics results, the pK_a shifts of glycine backbone amide is another factor that determines the rate of deamidation at TPI. The conclusions related to both solvation and acidity are in accordance with the previously suggested factors that would slow down the deamidation reaction in proteins. Another crucial outcome of the present study is the fact that tetrahedral intermediate will be approximately 10 kcal/mol less stabilized in TPI compared to a small peptide. Therefore the equilibrium between asparagine and the tetrahedral intermediate would shift more to reactant site in the case of TPI. Thus, for the very first time, the slower rate in TPI is found to be related to not only the reactivity of the deamidation sites but also to the stability of the intermediates.

Another ultimate goal of this study was to explain the impact of Asp71 (deamidated Asn71) on the deamidation of Asn15. Any clear evidence on this impact could not be detected when the mutated samples (i.e. Asn71 is mutated to Asp71) are compared with the wild type TPI. On the other hand, when Asn15 of chicken TPI, which is lack of Asn71, is compared to Asn15 of mammalian TPI; chicken TPI was found to be more reactive to initiate the reaction. Although the relation between Asn71 and Asn15 could not be clearly explained, the present results indicate that deamidation on these two sites cannot be classified as independent events, and requires further analysis.

In addition to the remarkable findings corresponding to deamidation in TPI, a computational protocol is suggested in order to obtain accurate and fast pK_a predictions of small alcohols and thiols. The suggested protocol was also found to give precise results for tyrosine and cysteine aminoacids. The strong accuracy and high computational efficiency make this protocol a promising candidate for predicting aminoacids' pK_a's in protein environment. This prediction would also be used to obtain more qualitative explanations on the relation between deamidation rate and the intrinsic pK_a of the proteins.

APPENDIX A: SUPPORTING INFORMATION OF THE MOLECULAR DYNAMIC SIMULATION ANALYSIS

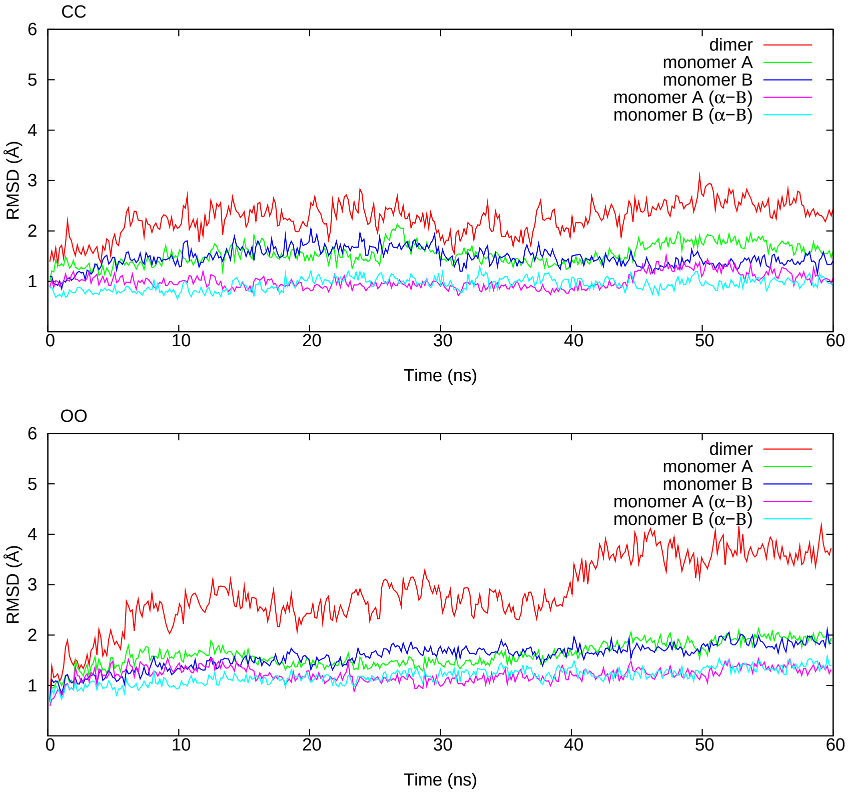


Figure A.1. Variations of the rmsd (in Å) along the MD trajectory (CC, OO and CS). Rmsd are computed for the complete dimer structure (red), monomer A only (green), monomer B only (blue), all α helices and β sheets of the subunit A (magenta), and all α helices and β sheets of the subunit B (cyan).

Table A.1. Interatomic distances for Near Attack Conformers (NAC).

	Donors	Acceptors	Cut off Distance (Å)
B1-1	WAT@O	Asn@CG	2.5
	Asn@ND2	WAT@H	2.5
B1-2	WAT@O	Asn@CG	2.5
	WAT@O	WAT@H	2.5
	Asn@ND2	WAT@H	2.5
B2-0	Asn@ND2	Gly@H	2.5
B2-1	WAT@O	Gly@H	2.5
	Asn@ND2	WAT@H	2.5
B3-0	Asn@OD1	Gly@H	2.5
B3-1	Asn@OD1	WAT@H	2.5
	WAT@O	Gly@H	2.5
B3-2	Asn@OD1	WAT@H	2.5
	WAT@O	WAT@H	2.5
	WAT@O	Gly@H	2.5
B4-0	Asn@OD1	Asn@HD12	2.5
B4-1	Asn@OD1	WAT@H	2.5
	WAT@O	Asn@HD12	2.5
B4-2	Asn@OD1	WAT@H	2.5
	WAT@O	WAT@H	2.5
	WAT@O	Asn@HD12	2.5

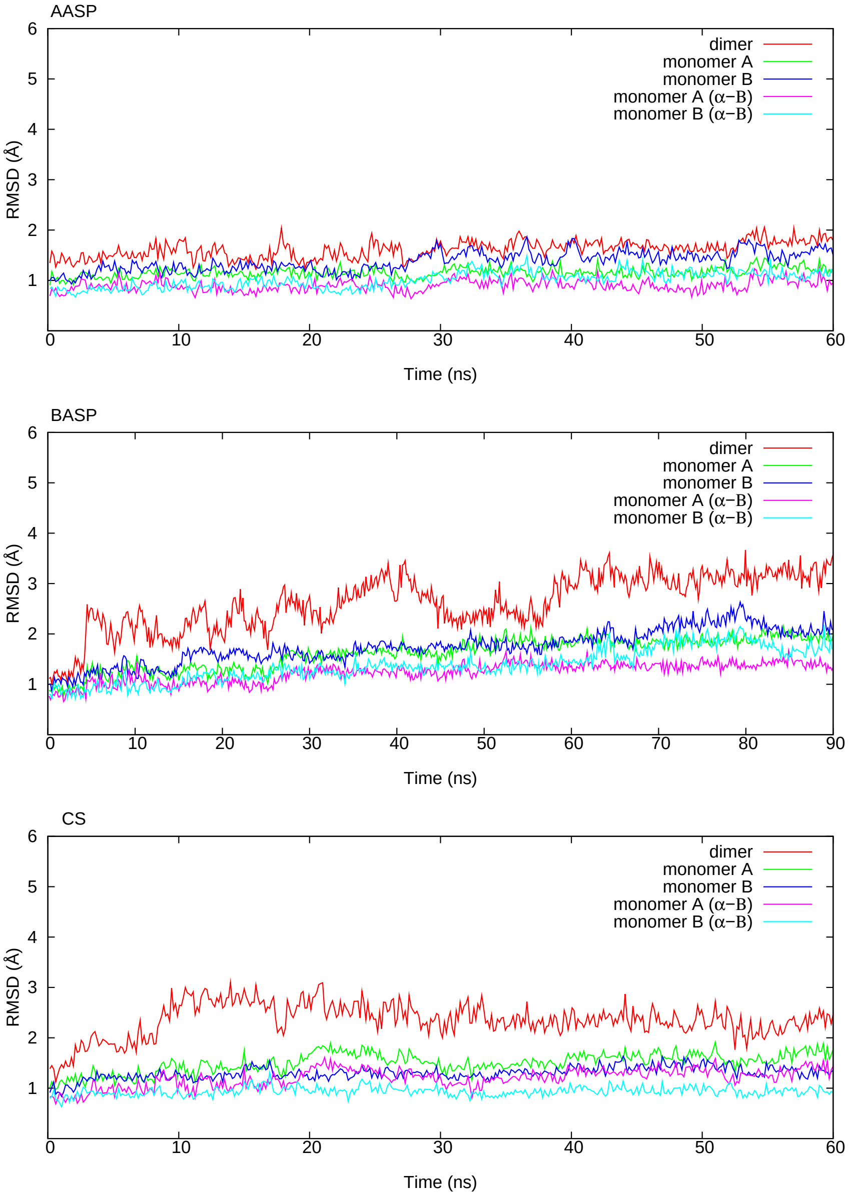


Figure A.2. Variations of the rmsd (in Å) along the MD trajectory (AASP, BASP). Rmsd are computed for the complete dimer structure (red), monomer A only (green), monomer B only (blue), all α helices and β sheets of the subunit A (magenta), and all α helices and β sheets of the subunit B (cyan).

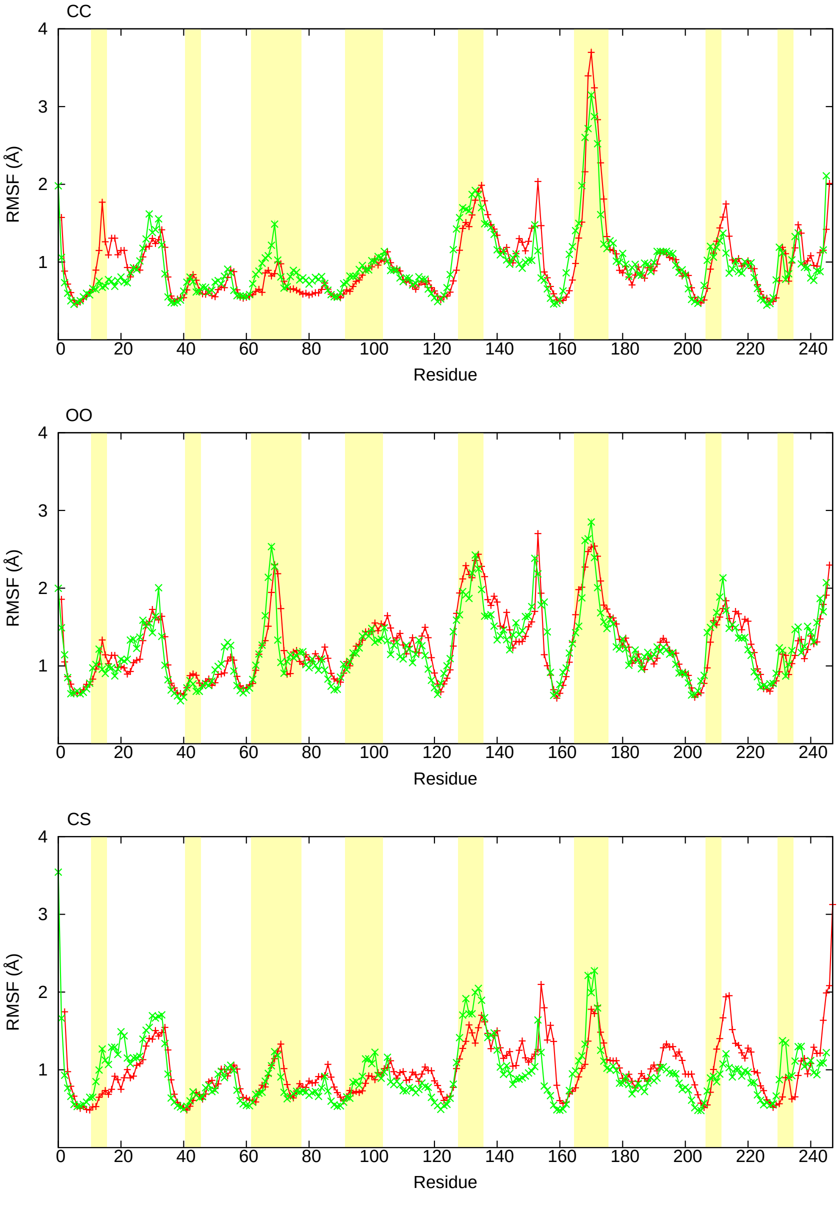


Figure A.3. Atomic fluctuations (RMSF: Root Mean Squared Fluctuations of C_α atoms) for the samples CC, OO, and CS. Yellow shading indicates the loops of the TIM barrel. Monomer A and monomer B are shown as red and green respectively.

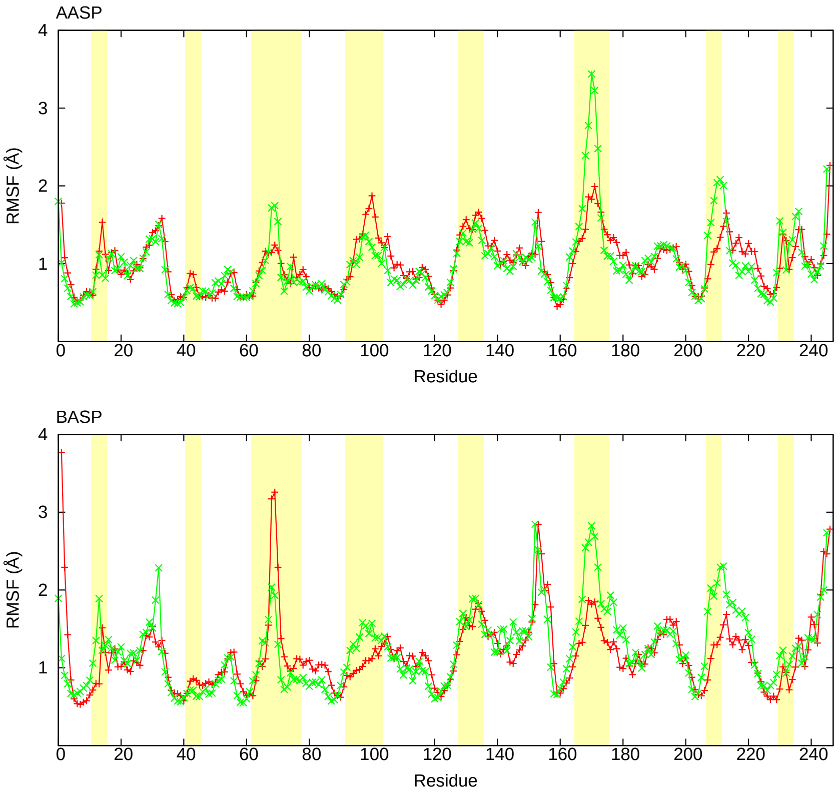


Figure A.4. Atomic fluctuations (RMSF: Root Mean Squared Fluctuations of C_α atoms) for the samples AASP, and BASP. Yellow shading indicates the loops of the TIM barrel. Monomer A and monomer B are shown as red and green respectively.

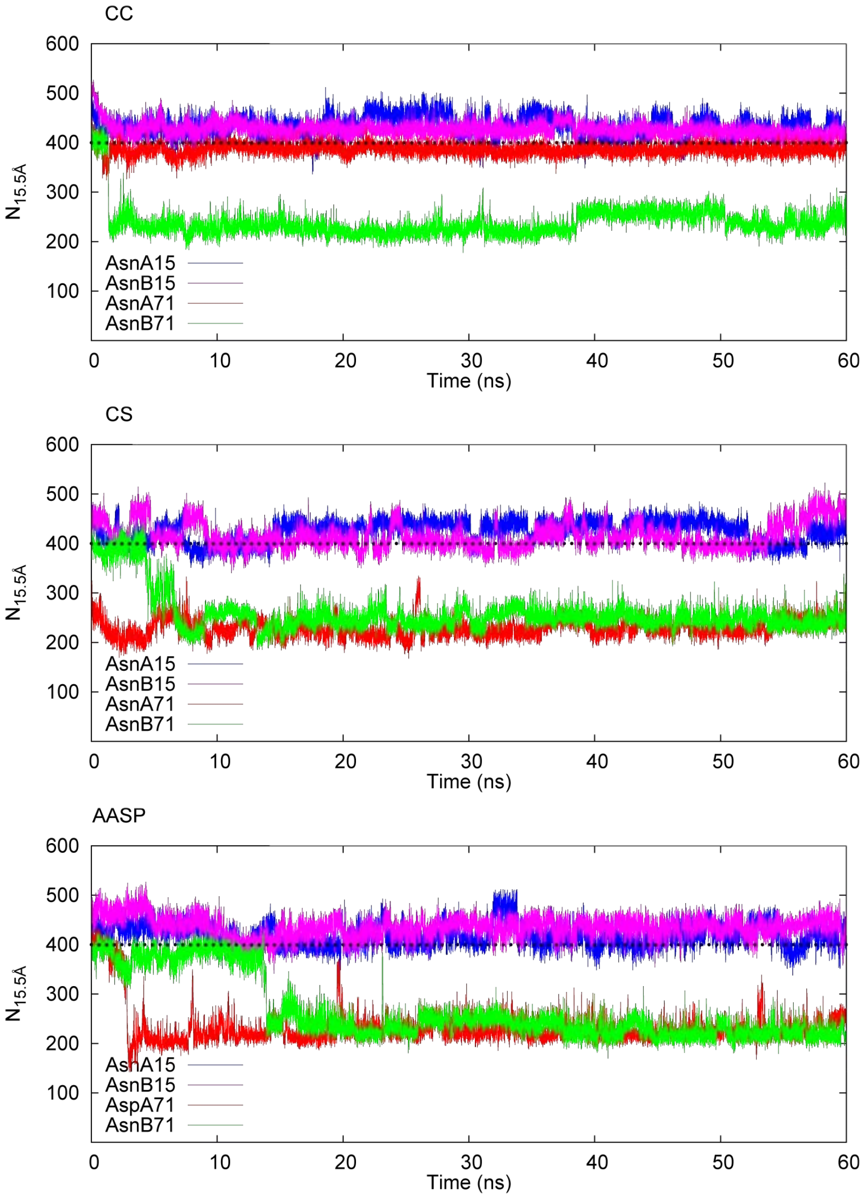


Figure A.5. $N_{15.5Å}$ of Asn with respect to time for the samples CC, CS, AASP. (Asn 15 A, 71 A, 15 B, 71B, are depicted in blue, red, magenta, green, respectively.).

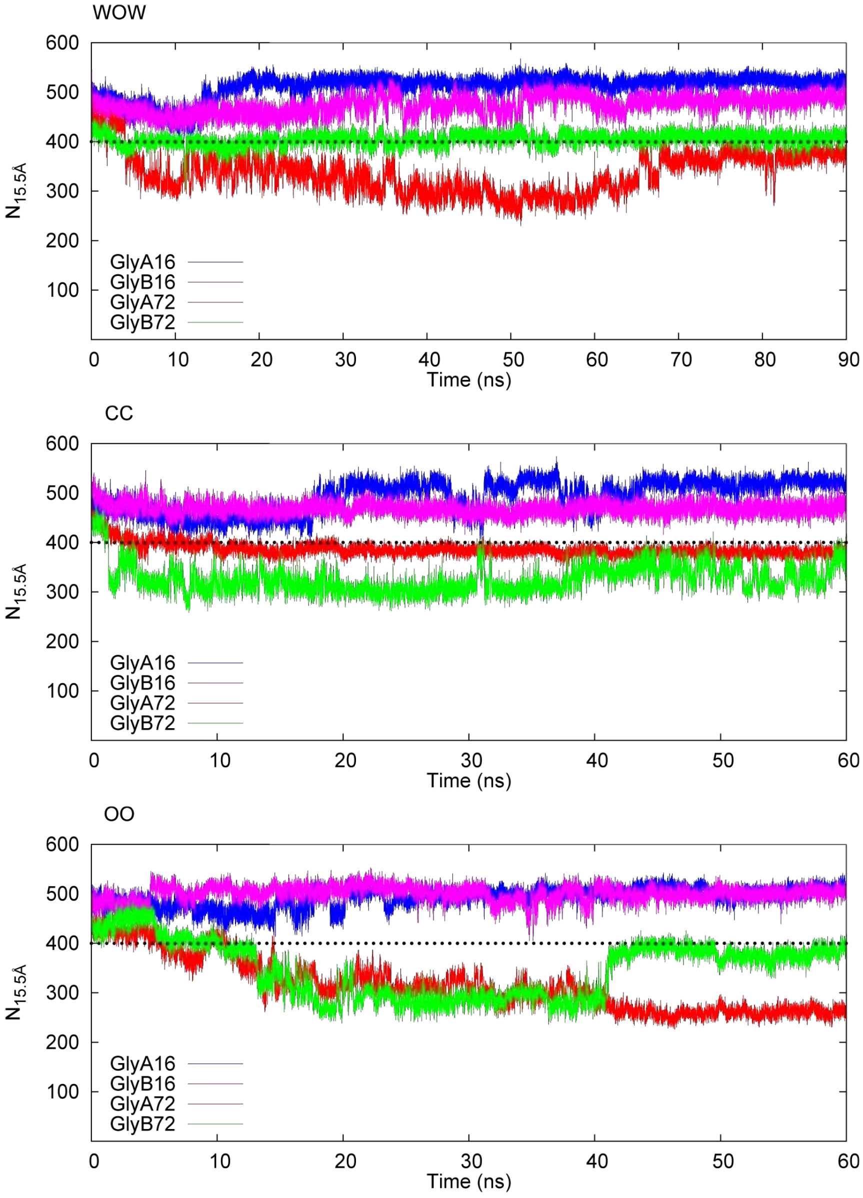


Figure A.6. $N_{15.5Å}$ of Gly with respect to time for WOW, CC, and OO. (Asn 15 A, 71 A, 15 B, 71 B, are depicted in blue, red, magenta, green, respectively.

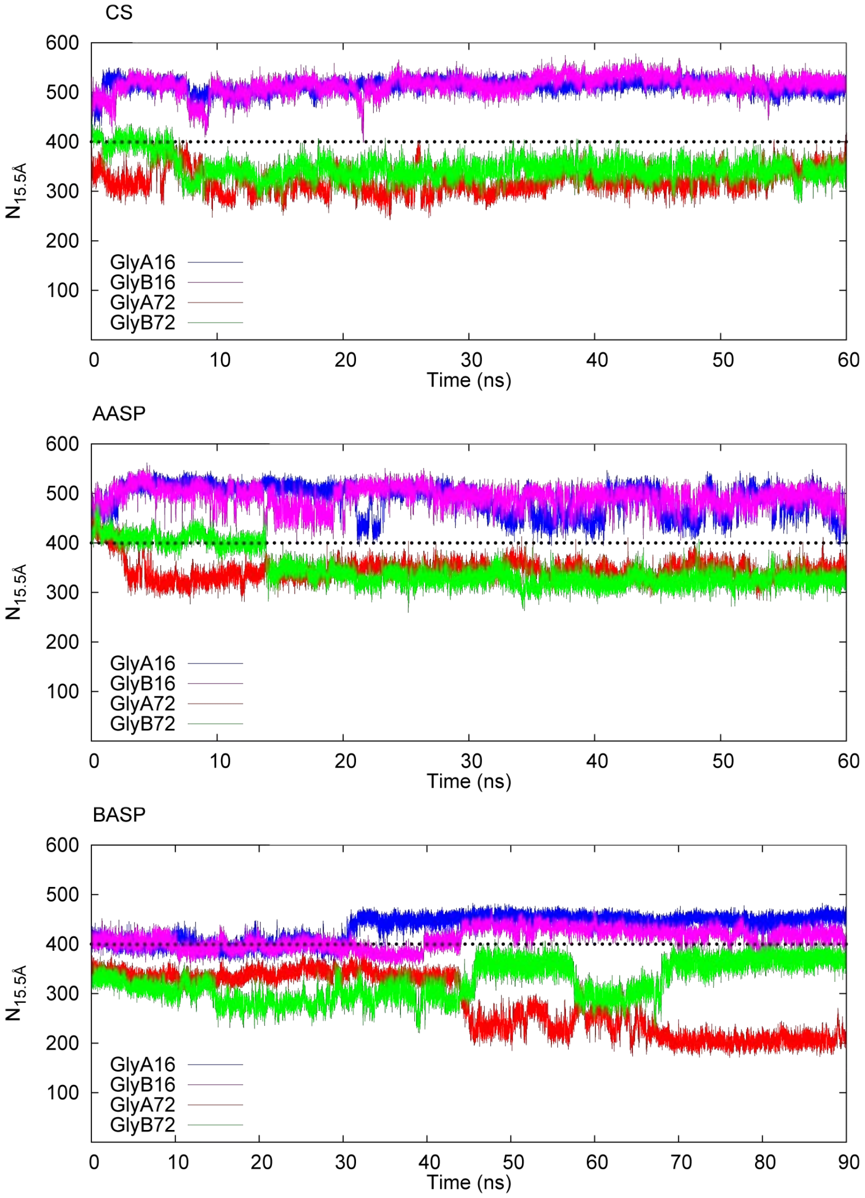


Figure A.7. $\mathrm{N}_{15.5\text{Å}}$ of Gly with respect to time for WOW, CC, and OO. (Asn 15 A, 71 A, 15 B, 71 B, are depicted in blue, red, magenta, green, respectively.

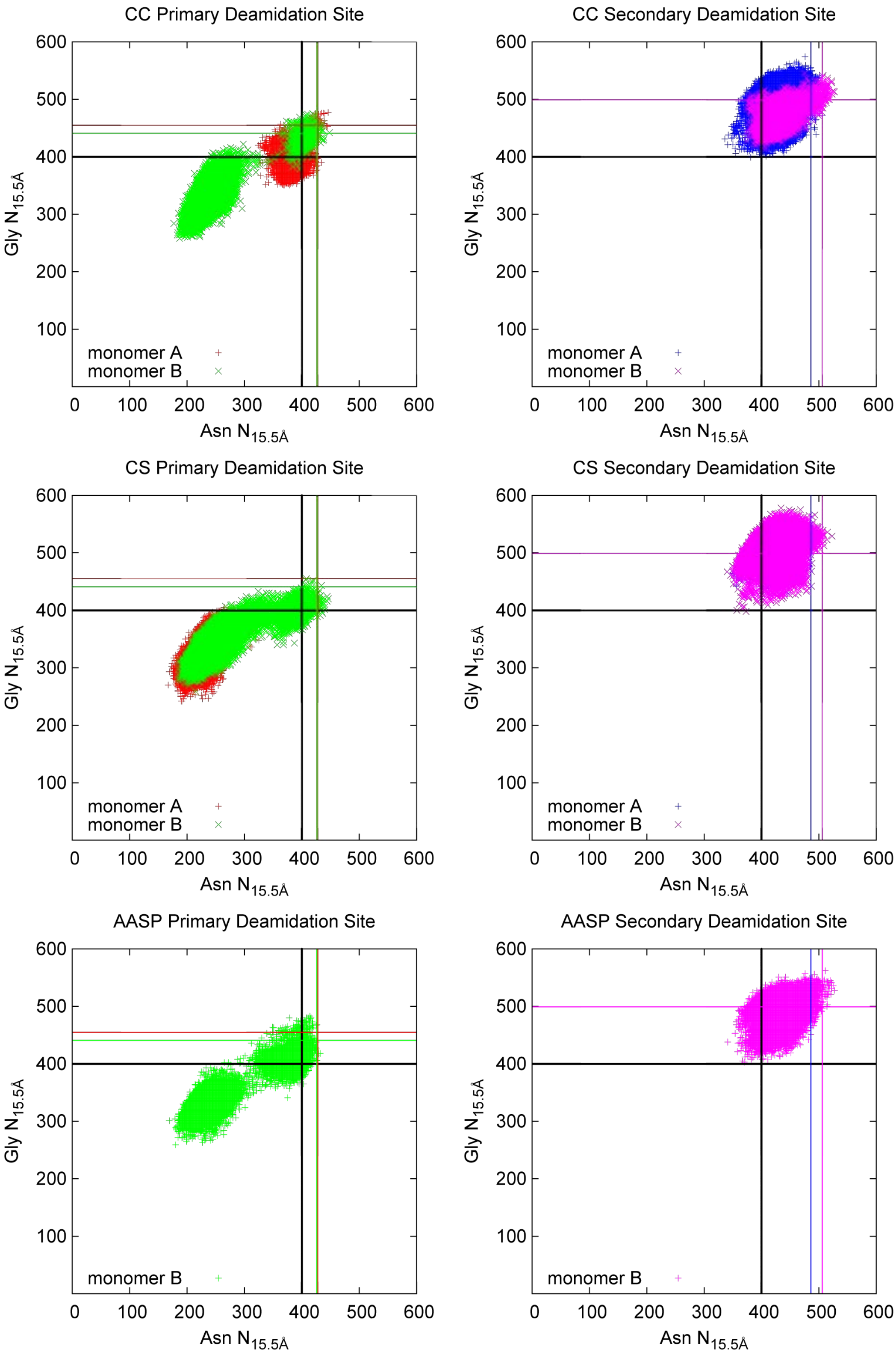


Figure A.8. $N_{15.5Å}$ of Asn vs Gly results for the samples CC, and CS. (Asn A15, A71, B15, B71, are depicted in blue, red, magenta, green, respectively. N15.5Å of the crystal structure is shown with lines.

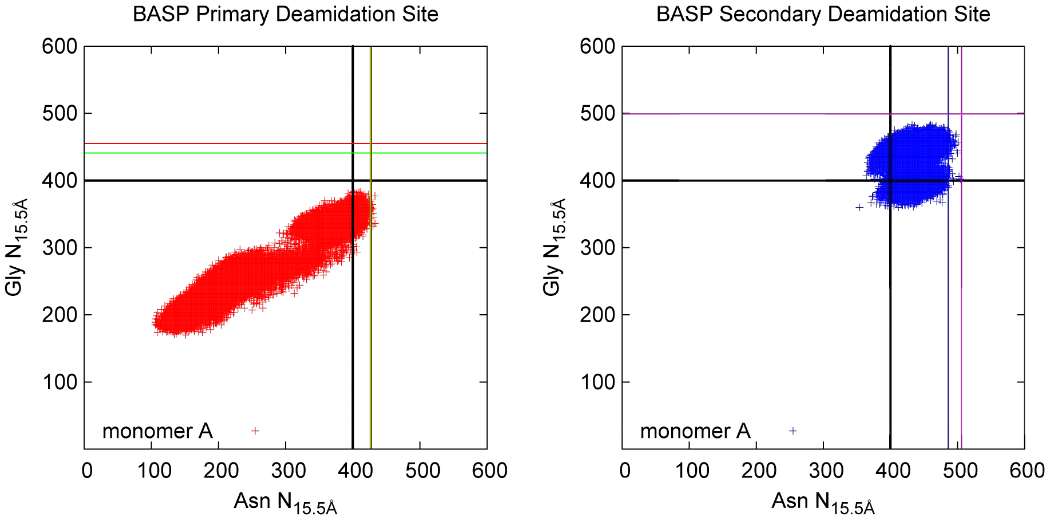


Figure A.9. $N_{15.5Å}$ of Asn vs Gly results for the samples AASP, and BASP. (Asn A15, A71, B15, B71, are depicted in blue, red, magenta, green, respectively. N15.5Å of the crystal structure is shown with lines.

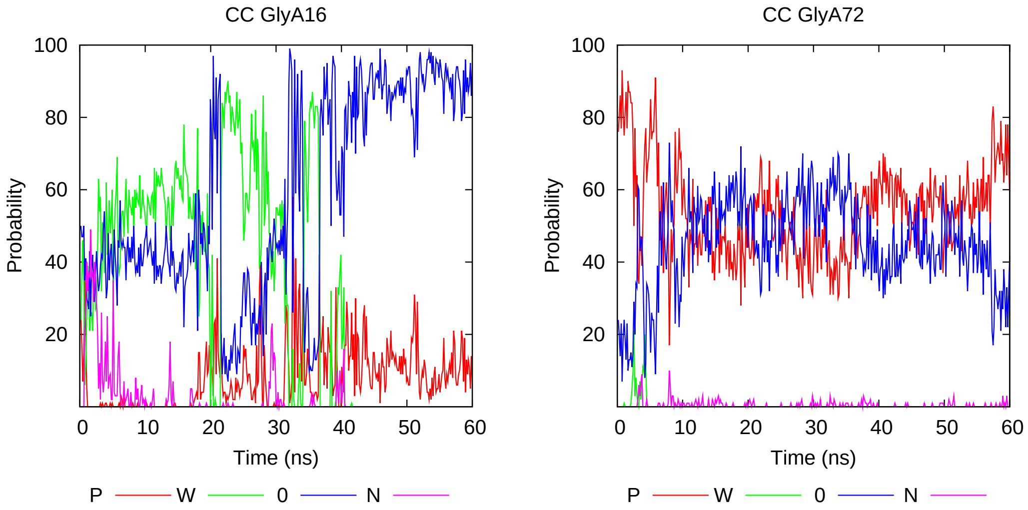


Figure A.10. Probabilities of occupancy of Gly H of the sample CC (red, green, blue, magenta indicate protein-bound (P), water-bound (W), free (0), Asn-bound (N) states respectively).

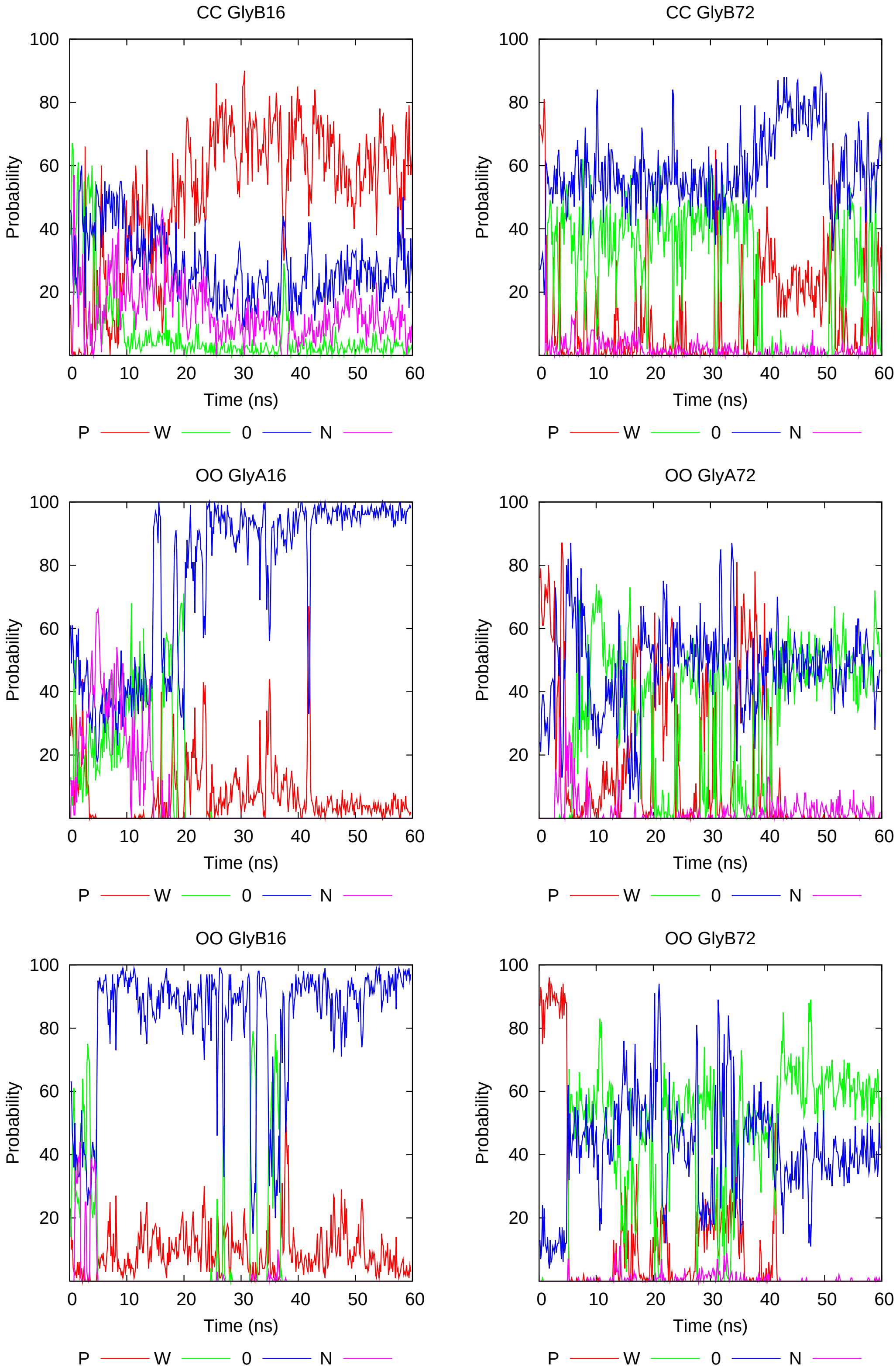


Figure A.11. Probabilities of occupancy of Gly H of the samples CC, OO (red, green, blue, magenta indicate protein-bound (P), water-bound (W), free (0), Asn-bound (N) states respectively).

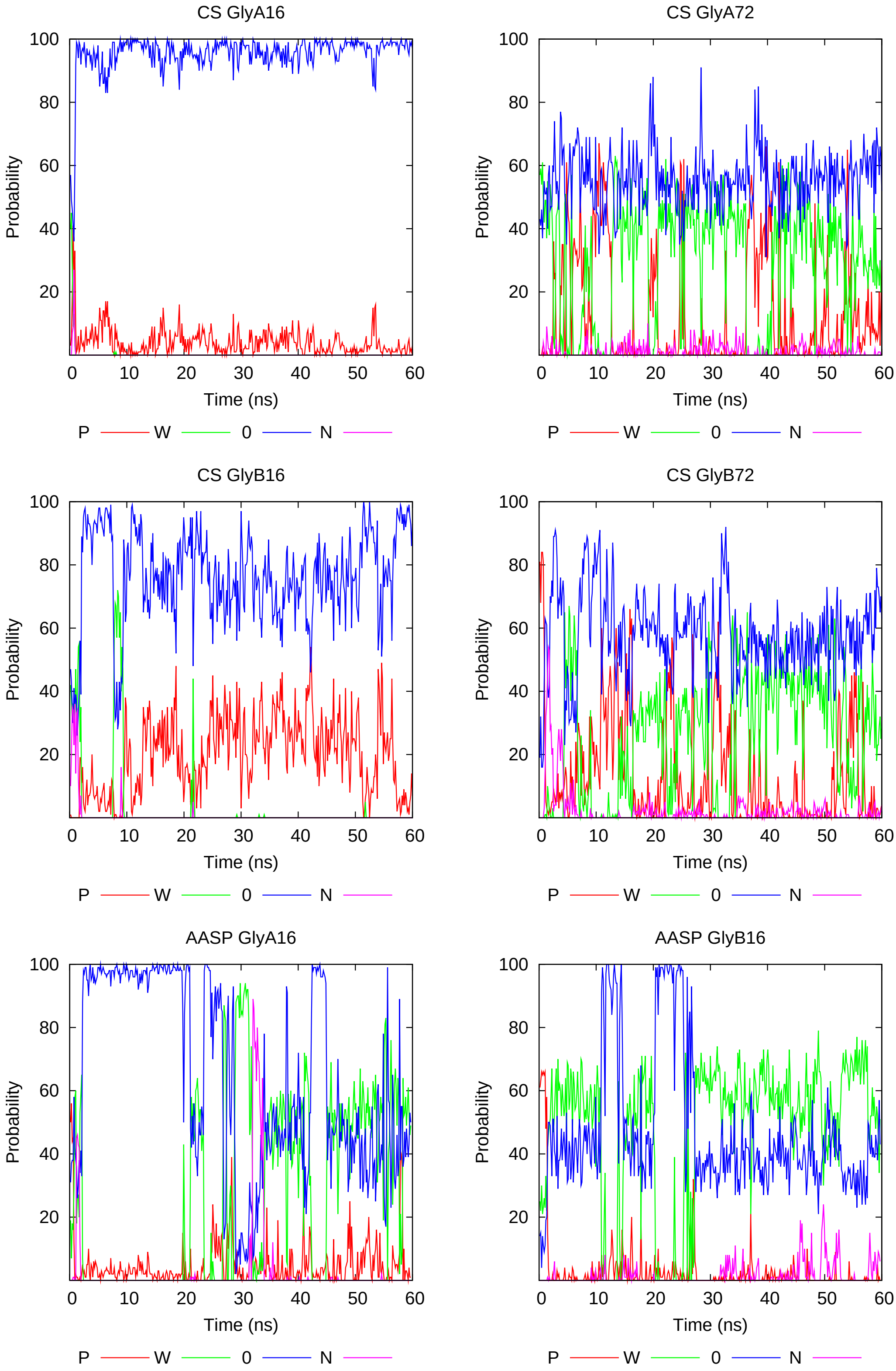


Figure A.12. Probabilities of occupancy of Gly H of the samples CS and AASP (red, green, blue, magenta indicate protein-bound (P), water-bound (W), free (0), Asn-bound (N) states respectively).

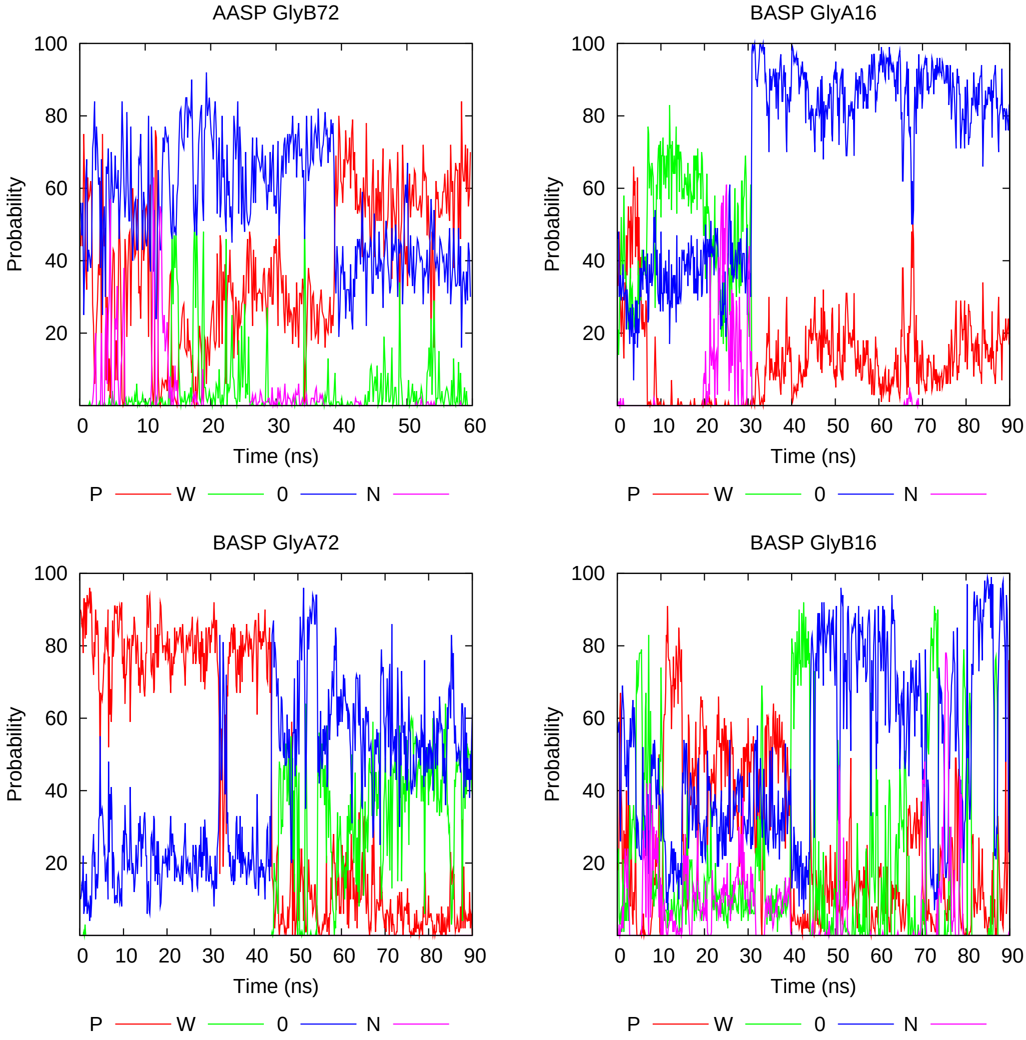


Figure A.13. Probabilities of occupancy of Gly H of the samples AASP and BASP. (red, green, blue, magenta indicate protein-bound (P), water-bound (W), free (0), Asn-bound (N) states respectively).

APPENDIX B: R^2, MAD AND MAX-ΔpK_a OVER ALL OF THE TESTED METHODS OF THE TRAINING SET (ALCOHOLS)

Table B.1. Alcohols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

		B3LYP	
	R^2	**MAD**	**MAX**
			a14 : 0.749
3-21G	0.995	0.227 ± 0.197	a08 : 0.588
			a25b : 0.440
			a08 : 1.217
3-21+G	0.979	0.523 ± 0.283	a21 : 0.973
			a06 : 0.867
			a08 : 0.906
6-31G	0.992	0.296 ± 0.220	a06 : 0.585
			a21 : 0.534
			a08 : 1.084
6-31+G	0.985	0.440 ± 0.260	a21 : 0.845
			a06 : 0.766
			a14 : 0.855
6-31G*	0.993	0.281 ± 0.220	a08 : 0.632
			a06 : 0.573
			a20 : 0.870
6-31+G*	0.988	0.329 ± 0.296	a14 : 0.838
			a08 : 0.796

Table B.2. Alcohols training set, R^2, MAD and MAX-$\Delta \mathrm{p}K_a$. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a14 : 0.836
6-31G**	0.993	0.281 ± 0.217	a08 : 0.623
			a06 : 0.567
			a20 : 0.883
6-31+G**	0.989	0.328 ± 0.294	a14 : 0.820
			a08 : 0.791
			a08 : 0.939
6-311G	0.992	0.309 ± 0.202	a06 : 0.530
			a01 : 0.514
			a08 : 1.154
6-311+G	0.985	0.435 ± 0.267	a21 : 0.833
			a04 : 0.811
			a08 : 0.691
6-311G*	0.994	0.250 ± 0.189	a06 : 0.566
			a14 : 0.502
			a08 : 0.898
6-311+G*	0.989	0.315 ± 0.285	a21 : 0.740
			a06 : 0.708
			a08 : 0.677
6-311G**	0.994	0.242 ± 0.187	a06 : 0.549
			a19 : 0.464
			a08 : 0.863
6-311+G**	0.990	0.309 ± 0.277	a21 : 0.746
			a06 : 0.697

Table B.3. Alcohols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a08 : 0.719
6-311G(2DF,2PD)	0.994	0.257 ± 0.196	a06 : 0.582
			a14 : 0.494
			a08 : 0.901
6-311+G(2DF,2PD)	0.989	0.323 ± 0.285	a21 : 0.748
			a06 : 0.705
		BLYP	
	R^2	**MAD**	**MAX**
			a14 : 0.670
3-21G	0.994	0.232 ± 0.201	a21 : 0.573
			a08 : 0.558
			a08 : 1.098
3-21+G	0.984	0.457 ± 0.267	a21 : 1.001
			a04 : 0.708
			a08 : 0.859
6-31G	0.993	0.277 ± 0.211	a21 : 0.653
			a14 : 0.430
			a08 : 0.992
6-31+G	0.987	0.388 ± 0.271	a21 : 0.896
			a20 : 0.759
			a14 : 0.782
6-31G*	0.993	0.277 ± 0.220	a21 : 0.681
			a08 : 0.614

Table B.4. Alcohols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a20 : 1.041
6-31+G*	0.988	0.328 ± 0.323	a21 : 0.855
			a14 : 0.854
			a14 : 0.777
6-31G**	0.993	0.278 ± 0.215	a21 : 0.677
			a08 : 0.603
			a20 : 1.050
6-31+G**	0.988	0.327 ± 0.318	a21 : 0.855
			a14 : 0.809
			a08 : 0.904
6-311G	0.993	0.305 ± 0.182	a21 : 0.541
			a01 : 0.464
			a08 : 1.007
6-311+G	0.988	0.384 ± 0.250	a21 : 0.853
			a04 : 0.639
			a08 : 0.666
6-311G*	0.994	0.257 ± 0.174	a21 : 0.500
			a06 : 0.477
			a20 : 0.812
6-311+G*	0.990	0.302 ± 0.282	a21 : 0.802
			a08 : 0.729
			a08 : 0.653
6-311G**	0.995	0.249 ± 0.173	a21 : 0.515
			a19 : 0.475

Table B.5. Alcohols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a20 : 0.822
6-311+G**	0.990	0.299 ± 0.281	a21 : 0.813
			a08 : 0.722
			a08 : 0.688
6-311G(2DF,2PD)	0.994	0.255 ± 0.188	a06 : 0.503
			a21 : 0.467
			a20 : 0.849
6-311+G(2DF,2PD)	0.990	0.315 ± 0.282	a21 : 0.815
			a08 : 0.731
		M06	
	R^2	**MAD**	**MAX**
			a14 : 0.867
3-21G	0.994	0.254 ± 0.196	a08 : 0.531
			a22 : 0.435
			a08 : 0.942
3-21+G	0.990	0.321 ± 0.274	a21 : 0.864
			a06 : 0.570
			a08 : 0.850
6-31G	0.993	0.270 ± 0.218	a21 : 0.576
			a14 : 0.565
			a08 : 1.032
6-31+G	0.987	0.385 ± 0.276	a21 : 0.863
			a06 : 0.721

Table B.6. Alcohols training set, R^2, MAD and MAX-$\Delta \mathrm{p}K_a$. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a14 : 0.868
6-31G*	0.992	0.304 ± 0.207	a08 : 0.625
			a21 : 0.558
			a14 : 0.883
6-31+G*	0.989	0.325 ± 0.278	a20 : 0.855
			a21 : 0.758
			a14 : 0.855
6-31G**	0.992	0.311 ± 0.198	a08 : 0.610
			a21 : 0.560
			a20 : 0.879
6-31+G**	0.989	0.328 ± 0.274	a14 : 0.861
			a21 : 0.760
			a08 : 0.849
6-311G	0.993	0.263 ± 0.210	a21 : 0.536
			a04 : 0.526
			a08 : 1.072
6-311+G	0.987	0.374 ± 0.282	a21 : 0.866
			a04 : 0.770
			a08 : 0.633
6-311G*	0.994	0.272 ± 0.186	a14 : 0.537
			a22 : 0.493
			a08 : 0.867
6-311+G*	0.990	0.310 ± 0.269	a21 : 0.802
			a14 : 0.708

Table B.7. Alcohols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a08 : 0.610
6-311G**	0.994	0.273 ± 0.181	a14 : 0.516
			a21 : 0.494
			a21 : 0.790
6-311+G**	0.991	0.323 ± 0.237	a08 : 0.785
			a14 : 0.730
			a08 : 0.654
6-311G(2DF,2PD)	0.994	0.271 ± 0.190	a14 : 0.544
			a21 : 0.503
			a21 : 0.810
6-311+G(2DF,2PD)	0.990	0.321 ± 0.250	a08 : 0.794
			a14 : 0.734
		M06L	
	R^2	**MAD**	**MAX**
			a14 : 0.712
3-21G	0.995	0.229 ± 0.196	a08 : 0.562
			a22 : 0.497
			a21 : 0.906
3-21+G	0.992	0.271 ± 0.263	a08 : 0.853
			a14 : 0.594
			a08 : 0.846
6-31G	0.993	0.274 ± 0.210	a21 : 0.612
			a06 : 0.486

Table B.8. Alcohols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a08 : 0.939
6-31+G	0.990	0.316 ± 0.270	a21 : 0.851
			a06 : 0.604
			a14 : 0.742
6-31G*	0.992	0.289 ± 0.220	a21 : 0.639
			a08 : 0.624
			a14 : 0.838
6-31+G*	0.990	0.319 ± 0.263	a21 : 0.806
			a20 : 0.731
			a14 : 0.735
6-31G**	0.992	0.290 ± 0.212	a21 : 0.641
			a08 : 0.605
			a14 : 0.820
6-31+G**	0.990	0.319 ± 0.264	a21 : 0.811
			a20 : 0.766
			a08 : 0.820
6-311G	0.993	0.272 ± 0.199	a21 : 0.630
			a06 : 0.468
			a08 : 0.969
6-311+G	0.989	0.334 ± 0.260	a21 : 0.899
			a06 : 0.629
			a08 : 0.617
6-311G*	0.993	0.274 ± 0.191	a21 : 0.566
			a06 : 0.553

Table B.9. Alcohols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a21 : 0.854
6-311+G*	0.991	0.297 ± 0.250	a08 : 0.728
			a14 : 0.713
			a08 : 0.594
6-311G**	0.994	0.267 ± 0.192	a21 : 0.593
			a22 : 0.546
			a21 : 0.873
6-311+G**	0.991	0.298 ± 0.245	a08 : 0.712
			a14 : 0.686
			a08 : 0.598
6-311G(2DF,2PD)	0.994	0.270 ± 0.193	a21 : 0.581
			a22 : 0.502
			a21 : 0.895
6-311+G(2DF,2PD)	0.991	0.303 ± 0.249	a08 : 0.717
			a14 : 0.701
		M062X	
	R^2	**MAD**	**MAX**
			a14 : 0.776
3-21G	0.994	0.252 ± 0.201	a08 : 0.583
			a20 : 0.492
			a08 : 1.351
3-21+G	0.974	0.586 ± 0.324	a06 : 1.042
			a21 : 1.024

Table B.10. Alcohols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a08 : 0.902
6-31G	0.992	0.290 ± 0.231	a06 : 0.700
			a14 : 0.509
			a08 : 1.124
6-31+G	0.984	0.443 ± 0.270	a06 : 0.830
			a21 : 0.814
			a14 : 0.939
6-31G*	0.992	0.316 ± 0.210	a06 : 0.639
			a08 : 0.571
			a14 : 0.912
6-31+G*	0.989	0.327 ± 0.281	a20 : 0.906
			a08 : 0.770
			a14 : 0.920
6-31G**	0.992	0.318 ± 0.207	a06 : 0.636
			a08 : 0.571
			a20 : 0.906
6-31+G**	0.989	0.327 ± 0.278	a14 : 0.892
			a08 : 0.770
			a08 : 0.928
6-311G	0.992	0.286 ± 0.226	a06 : 0.627
			a21 : 0.483
			a08 : 1.131
6-311+G	0.985	0.416 ± 0.294	a04 : 0.893
			a21 : 0.822

Table B.11. Alcohols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a14 : 0.625
6-311G*	0.994	0.262 ± 0.185	a08 : 0.613
			a06 : 0.521
			a14 : 0.827
6-311+G*	0.990	0.335 ± 0.236	a08 : 0.734
			a20 : 0.693
			a01 : 0.652
6-311G**	0.992	0.307 ± 0.190	a14 : 0.642
			a19 : 0.550
			a14 : 0.803
6-311+G**	0.990	0.337 ± 0.231	a08 : 0.727
			a20 : 0.695
			a14 : 0.655
6-311G(2DF,2PD)	0.993	0.301 ± 0.180	a08 : 0.581
			a19 : 0.541
			a08 : 0.809
6-311+G(2DF,2PD)	0.991	0.299 ± 0.262	a14 : 0.772
			a21 : 0.706
		O3LYP	
	R^2	**MAD**	**MAX**
			a14 : 0.851
3-21G	0.994	0.247 ± 0.196	a08 : 0.513
			a21 : 0.433

Table B.12. Alcohols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a08 : 1.116
3-21+G	0.983	0.461 ± 0.278	a21 : 0.944
			a06 : 0.819
			a08 : 0.785
6-31G	0.994	0.242 ± 0.222	a14 : 0.596
			a21 : 0.550
			a08 : 0.936
6-31+G	0.988	0.367 ± 0.273	a21 : 0.814
			a20 : 0.787
			a14 : 0.971
6-31G*	0.991	0.307 ± 0.226	a21 : 0.556
			a08 : 0.524
			a20 : 1.072
6-31+G*	0.987	0.345 ± 0.312	a14 : 0.989
			a21 : 0.738
			a14 : 0.954
6-31G**	0.991	0.311 ± 0.219	a21 : 0.553
			a08 : 0.516
			a20 : 1.085
6-31+G**	0.987	0.346 ± 0.310	a14 : 0.968
			a21 : 0.741
			a08 : 0.854
6-311G	0.993	0.276 ± 0.189	a06 : 0.507
			a21 : 0.464

Table B.13. Alcohols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a08 : 1.000
6-311+G	0.988	0.363 ± 0.256	a21 : 0.771
			a04 : 0.676
			a08 : 0.614
6-311G*	0.994	0.254 ± 0.186	a14 : 0.567
			a06 : 0.550
			a14 : 0.862
6-311+G*	0.989	0.336 ± 0.260	a20 : 0.836
			a21 : 0.679
			a08 : 0.604
6-311G**	0.994	0.250 ± 0.185	a06 : 0.536
			a14 : 0.532
			a20 : 0.839
6-311+G**	0.990	0.334 ± 0.258	a14 : 0.835
			a21 : 0.688
			a08 : 0.645
6-311G(2DF,2PD)	0.994	0.255 ± 0.196	a06 : 0.563
			a14 : 0.563
			a20 : 0.807
6-311+G(2DF,2PD)	0.989	0.329 ± 0.277	a14 : 0.793
			a08 : 0.758

Table B.14. Alcohols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

		OLYP	
	R^2	**MAD**	**MAX**
			a14 : 0.827
3-21G	0.994	0.249 ± 0.206	a21 : 0.593
			a08 : 0.482
			a08 : 1.048
3-21+G	0.985	0.427 ± 0.284	a21 : 0.976
			a06 : 0.743
			a08 : 0.748
6-31G	0.994	0.234 ± 0.220	a21 : 0.635
			a14 : 0.588
			a20 : 0.937
6-31+G	0.988	0.335 ± 0.301	a08 : 0.873
			a21 : 0.855
			a14 : 0.954
6-31G*	0.991	0.309 ± 0.229	a21 : 0.672
			a08 : 0.493
			a20 : 1.241
6-31+G*	0.985	0.380 ± 0.325	a14 : 1.039
			a21 : 0.794
			a14 : 0.939
6-31G**	0.991	0.311 ± 0.226	a21 : 0.670
			a08 : 0.486

Table B.15. Alcohols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a20 : 1.257
6-31+G**	0.985	0.381 ± 0.324	a14 : 1.019
			a21 : 0.798
			a08 : 0.823
6-311G	0.994	0.270 ± 0.179	a21 : 0.520
			a01 : 0.410
			a08 : 0.896
6-311+G	0.990	0.336 ± 0.256	a21 : 0.776
			a20 : 0.720
			a08 : 0.581
6-311G*	0.994	0.259 ± 0.175	a14 : 0.524
			a21 : 0.513
			a20 : 0.979
6-311+G*	0.989	0.329 ± 0.284	a14 : 0.862
			a21 : 0.724
			a08 : 0.572
6-311G**	0.994	0.255 ± 0.173	a21 : 0.529
			a14 : 0.491
			a20 : 0.979
6-311+G**	0.989	0.326 ± 0.282	a14 : 0.837
			a21 : 0.736

Table B.16. Alcohols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a08 : 0.636
6-311G(2DF,2PD)	0.994	0.261 ± 0.187	a14 : 0.536
			a06 : 0.520
			a20 : 1.019
6-311+G(2DF,2PD)	0.988	0.339 ± 0.286	a14 : 0.845
			a21 : 0.734
		PBE1PBE	
	R^2	**MAD**	**MAX**
			a14 : 0.826
3-21G	0.994	0.252 ± 0.207	a08 : 0.546
			a25b : 0.473
			a08 : 1.154
3-21+G	0.982	0.471 ± 0.283	a21 : 0.943
			a06 : 0.873
			a08 : 0.834
6-31G	0.993	0.267 ± 0.225	a06 : 0.638
			a14 : 0.562
			a08 : 1.019
6-31+G	0.987	0.388 ± 0.275	a21 : 0.820
			a06 : 0.772
			a14 : 0.940
6-31G*	0.992	0.310 ± 0.215	a06 : 0.574
			a08 : 0.561

Table B.17. Alcohols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a14 : 0.950
6-31+G*	0.988	0.334 ± 0.294	a20 : 0.939
			a08 : 0.716
			a14 : 0.923
6-31G**	0.992	0.315 ± 0.208	a06 : 0.569
			a08 : 0.555
			a20 : 0.949
6-31+G**	0.988	0.334 ± 0.291	a14 : 0.928
			a21 : 0.715
			a08 : 0.863
6-311G	0.993	0.278 ± 0.203	a06 : 0.599
			a21 : 0.471
			a08 : 1.053
6-311+G	0.987	0.378 ± 0.270	a21 : 0.815
			a04 : 0.732
			a14 : 0.619
6-311G*	0.994	0.269 ± 0.190	a08 : 0.561
			a06 : 0.527
			a14 : 0.836
6-311+G*	0.990	0.321 ± 0.252	a08 : 0.741
			a21 : 0.703
			a14 : 0.586
6-311G**	0.994	0.272 ± 0.186	a08 : 0.550
			a19 : 0.521

Table B.18. Alcohols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a14 : 0.808
6-311+G**	0.990	0.322 ± 0.247	a08 : 0.733
			a21 : 0.714
			a14 : 0.609
6-311G(2DF,2PD)	0.994	0.270 ± 0.185	a08 : 0.592
			a06 : 0.529
			a08 : 0.804
6-311+G(2DF,2PD)	0.990	0.312 ± 0.274	a14 : 0.782
			a21 : 0.755
		PBEPBE	
	R^2	**MAD**	**MAX**
			a14 : 0.757
3-21G	0.993	0.255 ± 0.217	a21 : 0.637
			a22 : 0.567
			a21 : 1.012
3-21+G	0.986	0.394 ± 0.282	a08 : 0.965
			a06 : 0.718
			a08 : 0.750
6-31G	0.994	0.249 ± 0.218	a21 : 0.685
			a14 : 0.529
			a21 : 0.898
6-31+G	0.989	0.323 ± 0.295	a20 : 0.889
			a08 : 0.842

Table B.19. Alcohols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a14 : 0.873
6-31G*	0.992	0.307 ± 0.218	a21 : 0.713
			a06 : 0.520
			a20 : 1.129
6-31+G*	0.987	0.344 ± 0.320	a14 : 0.942
			a21 : 0.867
			a14 : 0.861
6-31G**	0.992	0.310 ± 0.215	a21 : 0.715
			a06 : 0.515
			a20 : 1.150
6-31+G**	0.987	0.344 ± 0.321	a14 : 0.923
			a21 : 0.872
			a08 : 0.789
6-311G	0.994	0.266 ± 0.180	a21 : 0.574
			a06 : 0.431
			a08 : 0.929
6-311+G	0.989	0.333 ± 0.268	a21 : 0.873
			a20 : 0.641
			a08 : 0.560
6-311G*	0.994	0.252 ± 0.182	a21 : 0.548
			a06 : 0.538

Table B.20. Alcohols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a20 : 0.893
6-311+G*	0.990	0.307 ± 0.288	a21 : 0.827
			a14 : 0.808
			a21 : 0.568
6-311G**	0.994	0.249 ± 0.181	a08 : 0.549
			a06 : 0.522
			a20 : 0.897
6-311+G**	0.990	0.306 ± 0.287	a21 : 0.843
			a14 : 0.782
			a08 : 0.603
6-311G(2DF,2PD)	0.994	0.257 ± 0.186	a21 : 0.551
			a06 : 0.550
			a20 : 0.929
6-311+G(2DF,2PD)	0.989	0.316 ± 0.292	a21 : 0.849
			a14 : 0.787

APPENDIX C: R^2, MAD AND MAX-ΔpK_a OVER ALL OF THE TESTED METHODS OF THE TRAINING SET (THIOLS)

Table C.1. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model.

		B3LYP	
	R^2	**MAD**	**MAX**
			a14 : 0.753
3-21G	0.993	0.259 ± 0.215	a25b : 0.702
			a08 : 0.586
			a08 : 1.213
3-21+G	0.978	0.531 ± 0.295	a21 : 0.983
			a06 : 0.865
			a08 : 0.903
6-31G	0.991	0.321 ± 0.228	a25b : 0.652
			a06 : 0.584
			a08 : 1.081
6-31+G	0.984	0.459 ± 0.258	a21 : 0.855
			a06 : 0.764
			a14 : 0.859
6-31G*	0.991	0.312 ± 0.226	a08 : 0.629
			a06 : 0.571
			a20 : 0.858
6-31+G*	0.987	0.362 ± 0.289	a14 : 0.843
			a08 : 0.793

Table C.2. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a14 : 0.841
6-31G**	0.991	0.312 ± 0.224	a08 : 0.620
			a06 : 0.566
			a20 : 0.871
6-31+G**	0.988	0.360 ± 0.287	a14 : 0.824
			a08 : 0.788
			a08 : 0.936
6-311G	0.991	0.325 ± 0.223	a25b : 0.623
			a06 : 0.529
			a08 : 1.151
6-311+G	0.984	0.450 ± 0.269	a21 : 0.843
			a04 : 0.808
			a08 : 0.689
6-311G*	0.993	0.271 ± 0.212	a19 : 0.625
			a06 : 0.564
			a08 : 0.895
6-311+G*	0.988	0.346 ± 0.276	a21 : 0.750
			a06 : 0.706
			a08 : 0.675
6-311G**	0.993	0.265 ± 0.210	a19 : 0.652
			a06 : 0.547
			a08 : 0.860
6-311+G**	0.989	0.341 ± 0.270	a21 : 0.756
			a06 : 0.695

Table C.3. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a08 : 0.717
6-311G(2DF,2PD)	0.993	0.273 ± 0.223	a19 : 0.632
			a06 : 0.580
			a08 : 0.898
6-311+G(2DF,2PD)	0.988	0.353 ± 0.278	a21 : 0.758
			a06 : 0.703
		BLYP	
	R^2	**MAD**	**MAX**
			a14 : 0.675
3-21G	0.993	0.262 ± 0.221	a25b : 0.663
			a21 : 0.585
			a08 : 1.095
3-21+G	0.983	0.463 ± 0.281	a21 : 1.011
			a04 : 0.705
			a08 : 0.856
6-31G	0.992	0.304 ± 0.221	a21 : 0.664
			a25b : 0.633
			a08 : 0.989
6-31+G	0.986	0.410 ± 0.265	a21 : 0.906
			a20 : 0.747
			a14 : 0.786
6-31G*	0.991	0.308 ± 0.226	a21 : 0.692
			a08 : 0.612

Table C.4. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a20 : 1.028
6-31+G*	0.987	0.360 ± 0.315	a21 : 0.866
			a14 : 0.859
			a14 : 0.781
6-31G**	0.991	0.309 ± 0.222	a21 : 0.688
			a08 : 0.600
			a20 : 1.038
6-31+G**	0.987	0.359 ± 0.310	a21 : 0.865
			a14 : 0.814
			a08 : 0.901
6-311G	0.991	0.321 ± 0.205	a25b : 0.589
			a21 : 0.552
			a08 : 1.004
6-311+G	0.987	0.401 ± 0.252	a21 : 0.863
			a25b : 0.642
			a08 : 0.663
6-311G*	0.993	0.274 ± 0.199	a19 : 0.627
			a21 : 0.511
			a21 : 0.812
6-311+G*	0.989	0.334 ± 0.275	a20 : 0.799
			a08 : 0.726
			a19 : 0.663
6-311G**	0.993	0.266 ± 0.201	a08 : 0.650
			a21 : 0.526

Table C.5. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a21 : 0.823
6-311+G**	0.989	0.331 ± 0.274	a20 : 0.810
			a08 : 0.719
			a08 : 0.685
6-311G(2DF,2PD)	0.993	0.272 ± 0.209	a19 : 0.634
			a06 : 0.502
			a20 : 0.837
6-311+G(2DF,2PD)	0.988	0.347 ± 0.275	a21 : 0.825
			a08 : 0.728
		M06	
	R^2	**MAD**	**MAX**
			a14 : 0.871
3-21G	0.992	0.286 ± 0.215	a25b : 0.630
			a08 : 0.528
			a08 : 0.939
3-21+G	0.989	0.350 ± 0.266	a21 : 0.875
			a25b : 0.582
			a08 : 0.847
6-31G	0.992	0.300 ± 0.228	a25b : 0.624
			a21 : 0.588
			a08 : 1.029
6-31+G	0.986	0.414 ± 0.265	a21 : 0.873
			a06 : 0.718

Table C.6. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a14 : 0.872
6-31G*	0.991	0.336 ± 0.211	a08 : 0.623
			a21 : 0.569
			a14 : 0.888
6-31+G*	0.988	0.357 ± 0.274	a20 : 0.843
			a21 : 0.768
			a14 : 0.859
6-31G**	0.991	0.343 ± 0.202	a08 : 0.608
			a21 : 0.571
			a20 : 0.866
6-31+G**	0.988	0.360 ± 0.269	a14 : 0.866
			a21 : 0.771
			a08 : 0.847
6-311G	0.992	0.286 ± 0.228	a25b : 0.572
			a21 : 0.547
			a08 : 1.069
6-311+G	0.986	0.399 ± 0.273	a21 : 0.876
			a04 : 0.768
			a19 : 0.639
6-311G*	0.992	0.304 ± 0.194	a08 : 0.631
			a14 : 0.542

Table C.7. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a08 : 0.864
6-311+G*	0.989	0.341 ± 0.264	a21 : 0.812
			a14 : 0.713
			a19 : 0.666
6-311G**	0.992	0.306 ± 0.191	a08 : 0.607
			a14 : 0.520
			a21 : 0.800
6-311+G**	0.990	0.355 ± 0.229	a08 : 0.782
			a14 : 0.734
			a19 : 0.663
6-311G(2DF,2PD)	0.992	0.303 ± 0.200	a08 : 0.652
			a14 : 0.549
			a21 : 0.821
6-311+G(2DF,2PD)	0.989	0.352 ± 0.242	a08 : 0.791
			a14 : 0.738
		M062X	
	R^2	**MAD**	**MAX**
			a14 : 0.780
3-21G	0.993	0.284 ± 0.214	a25b : 0.694
			a08 : 0.581
			a08 : 1.347
3-21+G	0.973	0.596 ± 0.330	a06 : 1.040
			a21 : 1.034

Table C.8. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a08 : 0.899
6-31G	0.991	0.319 ± 0.237	a06 : 0.698
			a25b : 0.635
			a08 : 1.120
6-31+G	0.983	0.465 ± 0.263	a06 : 0.827
			a21 : 0.824
			a14 : 0.943
6-31G*	0.990	0.348 ± 0.214	a06 : 0.637
			a19 : 0.580
			a14 : 0.916
6-31+G*	0.988	0.360 ± 0.276	a20 : 0.893
			a08 : 0.767
			a14 : 0.924
6-31G**	0.990	0.350 ± 0.210	a06 : 0.635
			a19 : 0.593
			a14 : 0.896
6-31+G**	0.988	0.359 ± 0.273	a20 : 0.893
			a08 : 0.767
			a08 : 0.925
6-311G	0.991	0.304 ± 0.245	a06 : 0.625
			a25b : 0.599
			a08 : 1.127
6-311+G	0.984	0.438 ± 0.289	a04 : 0.890
			a21 : 0.833

Table C.9. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a19 : 0.682
6-311G*	0.993	0.294 ± 0.197	a14 : 0.630
			a08 : 0.611
			a14 : 0.831
6-311+G*	0.989	0.367 ± 0.229	a08 : 0.732
			a20 : 0.681
			a19 : 0.738
6-311G**	0.991	0.339 ± 0.197	a01 : 0.656
			a14 : 0.646
			a14 : 0.807
6-311+G**	0.989	0.368 ± 0.225	a08 : 0.725
			a20 : 0.682
			a19 : 0.729
6-311G(2DF,2PD)	0.991	0.333 ± 0.187	a14 : 0.659
			a08 : 0.579
			a08 : 0.806
6-311+G(2DF,2PD)	0.990	0.332 ± 0.257	a14 : 0.777
			a21 : 0.716
		M06L	
	R^2	**MAD**	**MAX**
			a14 : 0.716
3-21G	0.993	0.260 ± 0.214	a25b : 0.605
			a08 : 0.560

Table C.10. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a21 : 0.915
3-21+G	0.991	0.297 ± 0.262	a08 : 0.850
			a14 : 0.598
			a08 : 0.843
6-31G	0.992	0.305 ± 0.215	a21 : 0.623
			a25b : 0.596
			a08 : 0.936
6-31+G	0.989	0.345 ± 0.263	a21 : 0.862
			a06 : 0.602
			a14 : 0.747
6-31G*	0.991	0.320 ± 0.222	a21 : 0.650
			a08 : 0.622
			a14 : 0.842
6-31+G*	0.989	0.350 ± 0.258	a21 : 0.816
			a20 : 0.719
			a14 : 0.740
6-31G**	0.991	0.321 ± 0.216	a21 : 0.652
			a08 : 0.603
			a14 : 0.824
6-31+G**	0.989	0.350 ± 0.258	a21 : 0.822
			a20 : 0.754
			a08 : 0.818
6-311G	0.992	0.296 ± 0.214	a21 : 0.641
			a25b : 0.531

Table C.11. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a08 : 0.965
6-311+G	0.989	0.363 ± 0.249	a21 : 0.909
			a06 : 0.627
			a08 : 0.615
6-311G*	0.992	0.303 ± 0.199	a19 : 0.615
			a21 : 0.577
			a21 : 0.864
6-311+G*	0.990	0.328 ± 0.243	a08 : 0.725
			a14 : 0.717
			a19 : 0.643
6-311G**	0.992	0.299 ± 0.198	a21 : 0.604
			a08 : 0.592
			a21 : 0.883
6-311+G**	0.990	0.329 ± 0.239	a08 : 0.710
			a14 : 0.691
			a19 : 0.661
6-311G(2DF,2PD)	0.992	0.302 ± 0.197	a08 : 0.596
			a21 : 0.592
			a21 : 0.905
6-311+G(2DF,2PD)	0.990	0.334 ± 0.243	a08 : 0.714
			a14 : 0.705

Table C.12. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

		O3LYP	
	R^2	**MAD**	**MAX**
			a14 : 0.855
3-21G	0.993	0.278 ± 0.211	a25b : 0.621
			a08 : 0.511
			a08 : 1.112
3-21+G	0.982	0.476 ± 0.280	a21 : 0.954
			a06 : 0.817
			a08 : 0.782
6-31G	0.992	0.272 ± 0.233	a14 : 0.601
			a25b : 0.573
			a08 : 0.933
6-31+G	0.987	0.393 ± 0.267	a21 : 0.824
			a20 : 0.775
			a14 : 0.975
6-31G*	0.990	0.339 ± 0.229	a19 : 0.604
			a21 : 0.567
			a20 : 1.060
6-31+G*	0.986	0.378 ± 0.303	a14 : 0.993
			a21 : 0.749
			a14 : 0.958
6-31G**	0.990	0.344 ± 0.221	a19 : 0.617
			a21 : 0.564
			a20 : 1.072
6-31+G**	0.986	0.378 ± 0.301	a14 : 0.972
			a21 : 0.751

Table C.13. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a08 : 0.851
6-311G	0.992	0.292 ± 0.213	a25b : 0.562
			a19 : 0.517
			a08 : 0.996
6-311+G	0.987	0.384 ± 0.255	a21 : 0.781
			a04 : 0.674
			a19 : 0.637
6-311G*	0.993	0.275 ± 0.207	a08 : 0.611
			a14 : 0.572
			a14 : 0.867
6-311+G*	0.988	0.368 ± 0.249	a20 : 0.823
			a21 : 0.690
			a19 : 0.667
6-311G**	0.993	0.273 ± 0.204	a08 : 0.601
			a14 : 0.537
			a14 : 0.839
6-311+G**	0.988	0.367 ± 0.248	a20 : 0.827
			a21 : 0.699
			a08 : 0.643
6-311G(2DF,2PD)	0.993	0.273 ± 0.217	a19 : 0.641
			a14 : 0.568
			a14 : 0.798
6-311+G(2DF,2PD)	0.988	0.361 ± 0.268	a20 : 0.795
			a08 : 0.755

Table C.14. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

		OLYP	
	R^2	**MAD**	**MAX**
			a14 : 0.831
3-21G	0.993	0.280 ± 0.220	a21 : 0.605
			a25b : 0.583
			a08 : 1.044
3-21+G	0.984	0.445 ± 0.284	a21 : 0.986
			a06 : 0.740
			a08 : 0.745
6-31G	0.993	0.263 ± 0.234	a21 : 0.646
			a14 : 0.593
			a20 : 0.925
6-31+G	0.987	0.364 ± 0.293	a08 : 0.870
			a21 : 0.865
			a14 : 0.959
6-31G*	0.990	0.340 ± 0.232	a21 : 0.683
			a19 : 0.616
			a20 : 1.228
6-31+G*	0.984	0.413 ± 0.312	a14 : 1.043
			a21 : 0.804
			a14 : 0.944
6-31G**	0.990	0.343 ± 0.229	a21 : 0.682
			a19 : 0.633
			a20 : 1.244
6-31+G**	0.984	0.414 ± 0.311	a14 : 1.024
			a21 : 0.809

Table C.15. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a08 : 0.820
6-311G	0.993	0.286 ± 0.200	a21 : 0.530
			a19 : 0.528
			a08 : 0.893
6-311+G	0.989	0.356 ± 0.254	a21 : 0.786
			a20 : 0.709
			a19 : 0.636
6-311G*	0.993	0.277 ± 0.198	a08 : 0.579
			a14 : 0.529
			a20 : 0.967
6-311+G*	0.988	0.361 ± 0.270	a14 : 0.867
			a21 : 0.734
			a19 : 0.658
6-311G**	0.993	0.274 ± 0.195	a08 : 0.570
			a21 : 0.540
			a20 : 0.968
6-311+G**	0.988	0.358 ± 0.268	a14 : 0.842
			a21 : 0.746
			a19 : 0.638
6-311G(2DF,2PD)	0.993	0.277 ± 0.207	a08 : 0.634
			a14 : 0.541

Table C.16. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a20 : 1.007
6-311+G(2DF,2PD)	0.988	0.371 ± 0.271	a14 : 0.850
			a21 : 0.744
		PBE1PBE	
	R^2	**MAD**	**MAX**
			a14 : 0.831
3-21G	0.992	0.283 ± 0.222	a25b : 0.736
			a08 : 0.544
			a08 : 1.151
3-21+G	0.981	0.480 ± 0.293	a21 : 0.953
			a06 : 0.871
			a08 : 0.831
6-31G	0.992	0.296 ± 0.231	a06 : 0.637
			a25b : 0.609
			a08 : 1.016
6-31+G	0.986	0.416 ± 0.263	a21 : 0.831
			a06 : 0.770
			a14 : 0.944
6-31G*	0.990	0.342 ± 0.217	a06 : 0.573
			a08 : 0.559
			a14 : 0.954
6-31+G*	0.987	0.366 ± 0.286	a20 : 0.927
			a21 : 0.723

Table C.17. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a14 : 0.927
6-31G**	0.990	0.347 ± 0.210	a06 : 0.567
			a19 : 0.559
			a20 : 0.936
6-31+G**	0.987	0.367 ± 0.283	a14 : 0.933
			a21 : 0.726
			a08 : 0.860
6-311G	0.992	0.296 ± 0.225	a06 : 0.598
			a25b : 0.585
			a08 : 1.050
6-311+G	0.986	0.402 ± 0.262	a21 : 0.825
			a04 : 0.730
			a19 : 0.684
6-311G*	0.992	0.300 ± 0.199	a14 : 0.624
			a08 : 0.558
			a14 : 0.841
6-311+G*	0.989	0.352 ± 0.244	a08 : 0.738
			a21 : 0.714
			a19 : 0.709
6-311G**	0.992	0.304 ± 0.196	a14 : 0.591
			a08 : 0.547
			a14 : 0.812
6-311+G**	0.989	0.353 ± 0.239	a08 : 0.730
			a21 : 0.724

Table C.18. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a19 : 0.687
6-311G(2DF,2PD)	0.993	0.298 ± 0.196	a14 : 0.614
			a08 : 0.589
			a08 : 0.801
6-311+G(2DF,2PD)	0.989	0.344 ± 0.266	a14 : 0.786
			a21 : 0.765
		PBEPBE	
	R^2	**MAD**	**MAX**
			a14 : 0.761
3-21G	0.992	0.286 ± 0.233	a25b : 0.698
			a21 : 0.649
			a21 : 1.022
3-21+G	0.985	0.409 ± 0.283	a08 : 0.961
			a06 : 0.715
			a08 : 0.748
6-31G	0.992	0.279 ± 0.226	a21 : 0.696
			a25b : 0.579
			a21 : 0.908
6-31+G	0.988	0.349 ± 0.290	a20 : 0.878
			a08 : 0.839
			a14 : 0.877
6-31G*	0.990	0.338 ± 0.221	a21 : 0.724
			a19 : 0.528

Table C.19. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a20 : 1.116
6-31+G*	0.986	0.377 ± 0.309	a14 : 0.946
			a21 : 0.878
			a14 : 0.866
6-31G**	0.990	0.342 ± 0.218	a21 : 0.726
			a19 : 0.552
			a20 : 1.138
6-31+G**	0.986	0.377 ± 0.310	a14 : 0.927
			a21 : 0.882
			a08 : 0.786
6-311G	0.993	0.282 ± 0.207	a21 : 0.585
			a25b : 0.551
			a08 : 0.926
6-311+G	0.988	0.358 ± 0.263	a21 : 0.883
			a20 : 0.629
			a19 : 0.655
6-311G*	0.993	0.275 ± 0.202	a21 : 0.559
			a08 : 0.558
			a20 : 0.881
6-311+G*	0.989	0.340 ± 0.280	a21 : 0.837
			a14 : 0.812

Table C.20. Thiols training set, R^2, MAD and MAX-ΔpK_a. Optimizations and NPA charge calculations were done using CPCM model (cont.).

			a19 : 0.689
6-311G**	0.993	0.274 ± 0.200	a21 : 0.580
			a08 : 0.547
			a20 : 0.885
6-311+G**	0.989	0.338 ± 0.279	a21 : 0.854
			a14 : 0.786
			a19 : 0.660
6-311G(2DF,2PD)	0.993	0.275 ± 0.210	a08 : 0.600
			a21 : 0.562
			a20 : 0.917
6-311+G(2DF,2PD)	0.988	0.348 ± 0.284	a21 : 0.859
			a14 : 0.791

REFERENCES

1. Wright, H. T., “Nonenzymatic Deamidation of Asparaginyl and Glutaminyl Residues in Proteins”, *Critical Reviews in Biochemistry and Molecular Biology*, Vol. 26, No. 1, pp. 1–52, 1990.

2. Capasso, S. and S. Salvadori, “Effect of the Three Dimensional Structure on the Deamidation Reaction of Ribonuclease A”, *Journal of Peptide Research*, Vol. 54, pp. 377 – 382, 1999.

3. Robinson, N. E. and A. B. Robinson, *Molecular Clocks: Deamidation of Asparaginyl and Glutaminyl Residues in Peptides and Proteins.*, Althouse Press, Cave Junction, OR., 2004.

4. Capasso, S., L. Mazzarella, F. Sica and A. Zagari, “Deamidation via Cyclic Imide in Asparaginyl Peptides”, *Peptide Research*, Vol. 2, No. 2, pp. 195 – 200, 1989.

5. Capasso, S., L. Mazzarella, F. Sica, A. Zagari and S. Salvadori, “Kinetics and Mechanism of Succinimide Ring Formation in the Deamidation Process of Asparagine Residues”, *Journal of the Chemical Society, Perkin Transactions 2*, Vol. 2, pp. 679 – 682, 1993.

6. Aparicio, R., S. T. Ferreira and I. Polikarpov, “Closed Conformation of the Active Site Loop of Rabbit Muscle Triosephosphate Isomerase in the Absence of Substrate: Evidence of Conformational Heterogeneity”, *Journal of Molecular Biology*, Vol. 334, No. 5, pp. 1023 – 1041, 2003.

7. Geiger, T. C., “Deamidation, Isomerization, and Racemization at Asparaginyl and Aspartyl Residues in Peptides: Succinimide-linked Reactions That Contribute to Protein Degradation.”, *The Journal of Biological Chemistry*, Vol. 262, pp. 785 – 794, 1987.

8. Brennan, T. V. and S. Clarke, "Spontaneous Degradation of Polypeptides at Aspartyl and Asparaginyl Residues: Effects of the Solvent Dielectric.", *Protein Science*, Vol. 2, p. 331, 1993.

9. Tyler-Cross, R. and V. Schirch, "Effects of Amino Acid Sequence, Buffers, and Ionic Strength on the Rate and Mechanism of Deamidation of Asparagine Residues in Small Peptides", *The Journal of Biological Chemistry*, Vol. 266, pp. 22549 – 22556, 1991.

10. Robinson, N. E., Z. W. Robinson, B. R. Robinson, A. L. Robinson, J. A. Robinson, M. L. Robinson and A. B. Robinson, "Structure Dependent Nonenzymatic Deamidation of Glutaminyl and Asparaginyl Pentapeptides", *Journal of Peptide Research*, Vol. 63, pp. 426 – 436, 2004.

11. Stephenson, R. C. and S. Clarke, " Succinimide Formation from Aspartyl and Asparaginyl Peptides as a Model for the Spontaneous Degradation of Proteins.", *The Journal of Biological Chemistry*, Vol. 264, pp. 6164–6170, 1989.

12. Radkiewicz, J. L., H. Zipse, S. Clarke and K. N. Houk, "Accelerated Racemization of Aspartic Acid and Asparagine Residues via Succinimide Intermediates: an ab Initio Theoretical Exploration of Mechanism", *Journal of the American Chemical Society*, Vol. 118, No. 38, p. 9148–9155, 1996.

13. Kosky, A. A., U. O. Razzaq, M. J. Treuheit and D. N. Brems, "The Effects of Alpha-helix on the Stability of Asn Residues: Deamidation Rates in Peptides of Varying Helicity", *Protein Science*, Vol. 8, pp. 2519 – 2523, 1999.

14. Xie, M., Z. Shahrokh, M. Kadkhodayan, W. J. Henzel, M. F. Powell, R. T. Borchardt and R. L. Schowen, "Asparagine Deamidation in Recombinant Human Lymphotoxin: Hindrance by Three-dimensional Structures,", *Journal of Pharmaceutical Sciences*, Vol. 92, pp. 869 – 880, 2002.

15. Robinson, N. E. and A. B. Robinson, "Predicting Deamidation Rates of Human

Proteins from 3D Structures", *Proceedings of the National Academy of Sciences USA*, Vol. 98, p. 4367–4372, 2001.

16. Robinson, N. E., "Protein Deamidation", *Proceedings of the National Academy of Sciences USA*, Vol. 99, pp. 5283–5288, 2002.

17. Robinson, N. E. and A. B. Robinson, "Rediction of Primary Structure Deamidation Rates of Asparaginyl and Glutaminyl Peptides Through Steric and Catalytic Effects", *J. Peptide Res.*, Vol. 63, pp. 437 – 448, 2004.

18. Bongers, J., E. P. Heimer, T. Lambros, Y. E. Pan, R. M. Campbell and A. F. Felix, "Degradation of Aspartic Acid and Asparagine Residues in Human Growth Hormone-Releasing Factor", *International Journal of Peptide and Protein Research*, Vol. 39, pp. 364 – 374, 1992.

19. Liu, H., G. Caza-Bulseco, D. Faldu, C. Chumsae and J. Sun, "Heterogeneity of Monoclonal Antibodies.", *Journal of Pharmaceutical Sciences*, Vol. 97, No. 7, pp. 2426 – 2447, 2008.

20. Jenkins, N., "Modifications of Therapeutic Proteins: Challenges and Prospects", *Cytotechnology*, Vol. 53, No. 1 - 3, pp. 121 – 125, 2007.

21. Rose, I. A., "Mechanism of C–H Bond Cleavage in Aldolase and Isomerase Reactions", *Brookhaven Symposia in Biology*, Vol. 15, pp. 293 – 309, 1962.

22. Albery, W. J. and J. R. Knowles, "Efficiency and Evolution of Enzyme Catalysis", *Angewandte Chemie International Edition*, Vol. 16, pp. 285 – 293, 1977.

23. Cui, Q. and M. Karplus, "Catalysis and Specificity in Enzymes: A Study of Triosephosphate Isomerase and Comparison with Methyl Glyoxal Synthase", *Advances in Protein Chemistry*, Vol. 66, pp. 315 – 372, 2003.

24. Wierenga, R., E. Kapetaniou and R. Venkatesan, "Triosephosphate isomerase: a

highly evolved biocatalyst", *Cellular and Molecular Life Sciences*, Vol. 67, No. 23, pp. 3961–3982, 2010.

25. Banner, D. W., A. C. Bloomer, G. A. Petsko, D. C. Phillips, C. I. Pogson, I. A. Wilson, P. H. Corran, A. J. Furth, J. D. Milman, R. E. Offord, J. D. Priddle and S. G. Waley, "Structure of Chicken Muscle Triose Phosphate Isomerase Determined Crystallographically at 2.5 Angstrom Resolution Using Amino Acid Sequence Data", *Nature*, Vol. 255, pp. 609 – 614, 1975.

26. Joseph, D., G. A. Petsko and M. Karplus, "Anatomy of a Conformational Change: Hinged Lid Motion of the Triosephosphateisomerase Loop", *Science*, Vol. 249, pp. 1425 – 1428, 1990.

27. Cansu, S. and P. Doruker, "Dimerization Affects Collective Dynamics of Triosephosphate Isomerase", *Biochemistry*, Vol. 47, pp. 1358 – 1368, 2008.

28. Brown, F. K. and P. A. Kollman, "Molecular Dynamics Simulations of "Loop Closing" in the Enzyme Triose Phosphate Isomerase", *Journal of Molecular Biology*, Vol. 198, No. 3, pp. 533 – 546, 1987.

29. Xiang, J., J. Sun and N. S. Sampson, "The Importance of Hinge Sequence for Loop Function and Catalytic Activity in the Reaction Catalyzed by Triosephosphate Isomerase", *Journal of Molecular Biology*, Vol. 307, pp. 1103 – 1112, 2001.

30. Verlinde, C. L., C. J. Witmans, T. Pijning, K. H. Kalk, W. G. Hol, M. Callens and F. R. Opperdoes, "Structure of the Complex Between Trypanosomal Triosephosphate Isomerase and N-hydroxy-4-phosphono-butanamide: Binding at the Active Site Despite an Open Flexible Loop Conformation", *Protein Science*, Vol. 1, pp. 1578 – 1584, 1992.

31. Parthasarathy, S., H. Balaram, P. Balaram and M. R. Murthy, "Structures of Plasmodium Falciparum Triosephosphate Isomerase Complexed to Substrate Analogues: Observation of the Catalytic Loop in the Open Conformation in the

Ligand-bound State", *Acta Crystallographica Section D: Biological Crystallography*, Vol. 58, pp. 1992 – 2000, 2002.

32. Kurkcuoglu, O., R. L. Jernigan and P. Doruker, "Loop Motions of Triosephosphate Isomerase Observed with Elastic Networks", *Biochemistry*, Vol. 45, pp. 1173 – 1182, 2006.

33. Yuan, P. M., J. M. Talent and R. W. Gracy, "Molecular basis for the accumulation of acidic isozymes of triosephosphate isomerase on aging.", *Mechanisms of Aging and Development*, Vol. 17, pp. 151–162, 1981.

34. Skala, H., M. Vibert, A. Kahn and J. Dreyfus, "Phospho ADP Ribosylation of Human Glucose 6 Phosphate Dehydrogenase: Probable Mechanism of the Occurrence of Hyperanodic Forms", *Biochemical and Biophysical Research Communications*, Vol. 89, No. 3, pp. 988–996, 1979.

35. Kester, M. V., E. L. Jacobson and R. W. Gracy, "The Synthesis of a Labile Triosephosphate Isomerase Isozyme in Human Lymphoblasts and Fibroblasts", *Archives of Biochemistry and Biophysics*, Vol. 180, p. 562, 1977.

36. Lu, H. S., P. M. Yuan and R. W. Gracy, "Primary Structure of Human Triosephosphate Isomerase.", *The Journal of Biological Chemistry*, Vol. 259, pp. 11958–11968, 1984.

37. Tang, C., K. Ü. Yüksel, T. M. Jacobson and R. W. Gracy, "Isoforms of Chicken Triosephosphate Isomerase are Due to Specific Oxidation of Cysteine 126", *Archives of Biochemistry and Biophysics*, Vol. 283, No. 1, pp. 12–19, 1990.

38. Yüksel, K. U. and R. W. Gracy, *In A Study of Enzymes*, Vol. 2, CRC Press, Boca Raton, FL., 1991.

39. Zhang, Y., K. Yuksel and R. W. Gracy, "Terminal Marking of Avian Triosephosphate Isomerases by Deamidation and Oxidation", *Archives of Biochemistry and*

Biophysics, Vol. 317, No. 1, pp. 112–120, 1995.

40. Sun, A.-Q., K. U. Yuksel and R. W. Gracy, "Terminal Marking of Triosephosphate Isomerase: Consequences of Deamidation", *Archives of Biochemistry and Biophysics*, Vol. 322, No. 2, pp. 361–368, 1995.

41. Sun, A.-Q., K. Ü. Yüksel and R. W. Gracy, "Relationship Between the Catalytic Center and the Primary Degradation Site of Triosephosphate Isomerase: Effects of Active Site Modification and Deamidation", *Archives of Biochemistry and Biophysics*, Vol. 293, No. 2, pp. 382–390, 1992.

42. Gracy, R. W., J. M. Talent and A. I. Zvaigzne, "Molecular Wear and Tear Leads to Terminal Marking and the Unstable Isoforms of Aging", *Journal of Experimental Zoology*, Vol. 282, No. 1-2, pp. 18–27, 1998.

43. Eber, S., A. Pekrun, A. Bardosi, M. Gahr, W. Krietsch, J. Krüger, R. Matthei and W. Schröter, "Triosephosphate Isomerase Deficiency: Haemolytic Anaemia, Myopathy with Altered Mitochondria and Mental Retardation due to a New Variant with Accelerated Enzyme Catabolism and Diminished Specific Activity", *European Journal of Pediatrics*, Vol. 150, No. 11, pp. 761–766, 1991.

44. Yüksel, K. Ü. and R. W. Gracy, "In Vitro Deamidation of Human Triosephosphate Isomerase", *Archives of Biochemistry and Biophysics*, Vol. 248, No. 2, pp. 452–459, 1986.

45. Case, D., T. Darden, I. T.E. Cheatham, C. Simmerling, J. Wang, R. Duke, R. Luo, R. Walker, W. Zhang, K. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. Goetz, I. Kolossváry, K. Wong, F. Paesani, J. Vanicek, R. Wolf, J. Liu, S. B. X. Wu, T. Steinbrecher, H. Gohlke, Q. Cai, J. W. X. Ye, M.-J. Hsieh, G. Cui, D. Roe, D. Mathews, M. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P. Kollman, "AMBER 12", *University of California, San Fransisco*, 2012.

46. Verlet, L., “Computer Experiments on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules”, *Physical Review*, Vol. 159, pp. 98–103, 1967.

47. Genheden, S. and U. Ryde, “Will Molecular Dynamics Simulations of Proteins Ever Reach Equilibrium?”, *Physical Chemistry Chemical Physics*, Vol. 14, No. 24, pp. 8662–8677, 2012.

48. Cerutti, D. S. and D. A. Case, “Multi-Level Ewald: A Hybrid Multigrid/Fast Fourier Trans- form Approach to the Electrostatic Particle-Mesh Problem”, *Journal of Chemical Theory and Computation*, Vol. 6, pp. 443 – 458, 2010.

49. Woodcock, L.-V., “Isothermal Molecular Dynamics Calculations for Liquid Salts”, *Chemical Physics Letters*, Vol. 10, No. 3, pp. 257–261, 1971.

50. Andersen, H. C., “Molecular Dynamics Simulations at Constant Pressure and/or Temperature”, *Journal of Chemical Physics*, Vol. 72, pp. 2384 – 2394, 1980.

51. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, “Comparison of simple potential functions for simulating liquid water”, *The Journal of Chemical Physics*, Vol. 79, No. 2, pp. 926–935, 1983.

52. Clark, T., *A Handbook of Computational Chemistry*, John Willey&Sons, Canada,, 1985.

53. Pople, J. A. and G. A. Segal, “Approximate Self-Consistent Molecular Orbital Theory. II. Calculations with Complete Neglect of Differential Overlap”, *The Journal of Chemical Physics*, Vol. 43, No. 10, pp. S136–S151, 1965.

54. Pople, J. A., D. L. Beveridge and P. A. Dobosh, “Approximate Self-Consistent Molecular-Orbital Theory. V. Intermediate Neglect of Differential Overlap”, *The Journal of Chemical Physics*, Vol. 47, No. 6, pp. 2026–2033, 1967.

55. Pople, J. A. and D. L. Beveridge, *Approximate Molecular Orbital Theory*, McGraw-Hill, New York,, 1970.

56. Pople, J. A., D. P. Santry and G. A. Segal, “Approximate Self-Consistent Molecular Orbital Theory. I. Invariant Procedures”, *The Journal of Chemical Physics*, Vol. 43, No. 10, pp. S129–S135, 1965.

57. Bingham, R. C., M. J. Dewar and D. H. Lo, “Ground States of Molecules. XXV. MINDO/3. Improved Version of the MINDO Semiempirical SCF-MO Method”, *Journal of the American Chemical Society*, Vol. 97, No. 6, pp. 1285–1293, 1975.

58. Dewar, M. J. S. and W. Thiel, “Ground states of molecules. 38. The MNDO method. Approximations and parameters”, *Journal of the American Chemical Society*, Vol. 99, pp. 4899 – 4907, 1977.

59. Dewar, M. J., E. G. Zoebisch, E. F. Healy and J. J. Stewart, “Development and Use of Quantum Mechanical Molecular Models. 76. AM1: a New General Purpose Quantum Mechanical Molecular Model”, *Journal of the American Chemical Society*, Vol. 107, No. 13, pp. 3902–3909, 1985.

60. Stewart, J. J., “Optimization of Parameters for Semiempirical Methods I. Method”, *Journal of Computational Chemistry*, Vol. 10, No. 2, pp. 209–220, 1989.

61. Stewart, J. J., “MOPAC: a semiempirical molecular orbital program”, *Journal of computer-aided molecular design*, Vol. 4, No. 1, pp. 1–103, 1990.

62. Stewart, J. J. P., “Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements.”, *J. Mol. Mod.*, Vol. 13, pp. 1173 – 1213, 2007.

63. Hohenberg, P. and W. Kohn, “Inhomogeneous Electron Gas”, *Physical Review*, Vol. 136, pp. B864–B871, Nov 1964.

64. Gas, I. E., P. Hohenberg, W. Kohn, S.-C. E. I. Exchange and L. Sham, “Nobel Focus: Chemistry by Computer”, *Phys. Rev*, Vol. 136, p. B864, 1964.

65. Parr, R. G. and W. Yang, “Density-Functional Theory of Atoms and Molecules”, *Oxford: Oxford University Press.*, 1994.

66. Cramer, C. J., *Essentials of Computational Chemistry: Theories and Models*, John Wiley & Sons Ltd., USA, 2004.

67. Becke, A. D., “Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior”, *Physical Review A*, Vol. 38, No. 6, pp. 3098 – 3103, 1988.

68. Becke, A. D., “A New Mixing of Hartree–Fock and Local Density-Functional Theories”, *The Journal of Chemical Physics*, Vol. 98, pp. 1372 – 1377, 1993.

69. Becke, A. D., “Density-Functional Thermochemistry. III. The Role of Exact Exchange”, *The Journal of Chemical Physics*, Vol. 98, pp. 5648 – 5652, 1993.

70. Perdew, J. P., K. Burke and M. Ernzerhof, “Generalized Gradient Approximation Made Simple”, *Physical Review Letters*, Vol. 77, pp. 3865 – 3868, 1996.

71. Perdew, J. P., K. Burke and M. Ernzerhof, “Errata: Generalized Gradient Approximation Made Simple,”, *Physical Review Letters*, Vol. 78, p. 1396, 1997.

72. Cohen, A. J. and N. C. Handy, “Dynamic Correlation”, *Molecular Physics*, Vol. 99, pp. 607 – 615, 2001.

73. Zhao, Y., N. E. Schultz and D. G. Truhlar, “Exchange-Correlation Functional With Broad Accuracy for Metallic and Nonmetallic Compounds, Kinetics, and Noncovalent Interactions”, *The Journal of Chemical Physics*, Vol. 123, p. 161103, 2005.

74. Zhao, Y. and D. G. Truhlar, “Comparative DFT Study of Van der Waals Complexes: Rare-Gas Dimers, Alkaline-Earth Dimers, Zinc Dimer, and Zinc-Rare-Gas

Dimers", *The Journal of Physical Chemistry*, Vol. 110, pp. 5121 – 5129, 2006.

75. Zhao, Y. and D. G. Truhlar, "Density Functional for Spectroscopy: No Long-Range Self-Interaction Error, Good Performance for Rydberg and Charge-Transfer States, and Better Performance on Average than B3LYP for Ground States", *The Journal of Physical Chemistry A*, Vol. 110, pp. 13126 – 13130, 2006.

76. Zhao, Y. and D. G. Truhlar, "The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals", *Theoretical Chemistry Accounts*, Vol. 120, pp. 215 – 241, 2008.

77. Hehre, W. J., R. F. Stewart and J. A. Pople, "Self-Consistent Molecular Orbital Methods. 1. Use of Gaussian Expansions of Slater-Type Atomic Orbitals", *The Journal of Chemical Physics*, Vol. 51, pp. 2657 – 2664, 1969.

78. Collins, J. B., P. R. Schleyer, J. S. Binkley and J. A. Pople, "Self-Consistent Molecular Orbital Methods. 17. Geometries and Binding Energies of Second-Row Molecules. A Comparison of Three Basis Sets", *The Journal of Chemical Physics*, Vol. 64, pp. 5142 – 5151, 1976.

79. Binkley, J. S., J. A. Pople and W. J. Hehre, "Self-Consistent Molecular Orbital Methods. 21. Small Split-Valence Basis Sets for First-Row Elements", *Journal of the American Chemical Society*, Vol. 102, pp. 939 – 947, 1980.

80. Gordon, M. S., J. S. Binkley, J. A. Pople, W. J. Pietro and W. J. Hehre, "Self-Consistent Molecular Orbital Methods. 22. Small Split-Valence Basis Sets for Second-Row Elements", *Journal of the American Chemical Society*, Vol. 104, pp. 2797 – 2803, 1982.

81. Pietro, W. J., M. M. Francl, W. J. Hehre, D. J. Defrees, J. A. Pople and J. S. Binkley, "Self-Consistent Molecular Orbital Methods. 24. Supplemented Small Split-

Valence Basis-Sets for 2nd-Row Elements,", *Journal of the American Chemical Society*, Vol. 104, pp. 5039 – 5048, 1982.

82. Dobbs, K. D. and W. J. Hehre, "Molecular-Orbital Theory of the Properties of Inorganic and Organometallic Compounds. 4. Extended Basis-Sets for 3rd Row and 4th Row, Main-Group Elements", Vol. 7, pp. 359 – 378, 1986.

83. Ditchfield, R., W. J. Hehre and J. A. Pople, "Self-Consistent Molecular Orbital Methods. 9. Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules", *The Journal of Chemical Physics*, Vol. 54, p. 724, 1971.

84. Hehre, W. J., R. Ditchfield and J. A. Pople, "Self-Consistent Molecular Orbital Methods. 12. Further Extensions of Gaussian-type Basis Sets for Use in Molecular-Orbital Studies of Organic-Molecules", *The Journal of Chemical Physics*, Vol. 56, p. 2257, 1972.

85. Leach, A. R., *Molecular Modelling Principles and Applications*, Prentice Hall, England, 2001.

86. Mulliken, R. S., "Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I", *The Journal of Chemical Physics*, Vol. 23, No. 10, pp. 1833 – 1840, 1955.

87. Löwdin, P.-O., "On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals", *The Journal of Chemical Physics*, Vol. 18, pp. 365 – 375, 1950.

88. Reed, A. E., R. B. Weinstock and F. Weinhold, "Natural Population Analysis", *The Journal of Chemical Physics*, Vol. 83, pp. 735 – 746, 1985.

89. Tomasi, J., M. Benedetta and R. Cammi, "Quantum mechanical continuum solvation models", *Chemical Reviews*, Vol. 105, pp. 2999 – 3093, 2005.

90. Barone, V. and M. Cossi, “Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model”, *The Journal of Physical Chemistry A*, Vol. 102, pp. 1995 – 2001, 1998.

91. Cossi, M., N. Rega, G. Scalmani and V. Barone, “Energies, Structures, and Electronic Properties of Molecules in Solution with the C-PCM Solvation Model”, *Journal of Computational Chemistry*, Vol. 24, pp. 669 – 681, 2003.

92. Warshel, A. and M. Karplus, “Calculation of Ground and Excited State Potential Surfaces of Conjugated Molecules. I. Formulation and Parametrization”, *Journal of the American Chemical Society*, Vol. 94, No. 16, pp. 5612–5625, 1972.

93. Warshel, A. and M. Levitt, “Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme”, *Journal of Molecular Biology*, Vol. 103, No. 2, pp. 227–249, 1976.

94. Levitt, M. and A. Warshel, “Computer Simulation of Protein Folding”, *Nature*, Vol. 253, No. 5494, pp. 694–698, 1975.

95. Monard, G., X. Prat-Resina, A. González-Lafont and J. M. Lluch, “Determination of Enzymatic Reaction Pathways Using QM/MM Methods”, *International Journal of Quantum Chemistry*, Vol. 93, No. 3, pp. 229–244, 2003.

96. Field, M. J., P. A. Bash and M. Karplus, “A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations”, *Journal of Computational Chemistry*, Vol. 11, No. 6, pp. 700–733, 1990.

97. Reuter, N., A. Dejaegere, B. Maigret and M. Karplus, “Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches”, *The Journal of Physical Chemistry A*, Vol. 104, No. 8, pp. 1720–1735, 2000.

98. Mills, M. and I. Andricioaei, “An Experimentally Guided Umbrella Sampling Protocol for Biomolecules”, *The Journal of Chemical Physics*, Vol. 129, pp. 114101–

114111, 2008.

99. Kumar, S., J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, “Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method”, *Journal of Computational Chemistry*, Vol. 16, No. 11, pp. 1339–1350, 1995.

100. Pilar, F. L., *Elementary Quantum Chemistry*, Mc-Graw-Hil1, New York, 1990.

101. Hirst, D. M. A., *Computational Approach to Chemistry*, Blackwell Scientific Publications,Oxford, 1990.

102. Hehre, W. J., L. Radom, P. R. Schleyer and J. A, *Pople Ab Initio Molecular Orbital Theory*, John Willey and Sons, New York, 1986.

103. Lowe, J. P., *Quantum Chemistry*, Academic Press, 1993.

104. Szabo, A. and N. S. Ostlund, *Modern Quantum Chemistry Introduction to Advanced Electronic Structure Theory*, Dover, 1996.

105. Schlick, T., *Molecular Modeling and Simulation An Interdisciplinary Guide*, Springer, 2000.

106. Jiali, G. and T. M. A. (Editors), *Combined Quantum Mechanical and Molecular Mechanical Methods*, American Chemical Society, Washington, DC, 1998.

107. Robinson, A. and C. J. Rudd, “Deamidation of Glutaminyl and Asparaginyl Residues in Peptides and Proteins.”, *Current Topics in Cellular Regulation*, Vol. 8, pp. 247 – 295, 1974.

108. Robinson, A. B., J. W. Scotchler and J. H. McKerrow, “Rates of Nonenzymic Deamidation of Glutaminyl and Asparaginyl Residues in Pentapeptides”, *Journal of the American Chemical Society*, Vol. 95, No. 24, p. 8156–8159, 1973.

109. Bornstein, P. and G. Balian, “The Specific Nonenzymatic Cleavage of Bovine Ribonuclease with Hydroxylamine”, *Journal of Biological Chemistry*, Vol. 245, No. 18, pp. 4854–4856, 1970.

110. Meinwals, Y. C., E. R. Stimson and H. A. Scheraga, “Deamidation of the Asparaginyl-Glycyl Sequence”, *International Journal of Peptide and Protein Research*, Vol. 28, No. 1, pp. 79–84, 1986.

111. Takemoto, L., D. Boyle *et al.*, “Increased Deamidation of Asparagine During Human Senile Cataractogenesis”, *Molecular Vision*, Vol. 6, pp. 164–168, 2000.

112. Watanabe, A., K. Takio and Y. Ihara, “Deamidation and Isoaspartate Formation in Smeared Tau in Paired Helical Filaments Unusual Properties of the Microtubule-Binding Domain of Tau”, *Journal of Biological Chemistry*, Vol. 274, No. 11, pp. 7368–7378, 1999.

113. Weintraub, S. J. and S. R. Manson, “Asparagine Deamidation: a Regulatory Hourglass”, *Mechanisms of Ageing and Development*, Vol. 125, No. 4, pp. 255 – 257, 2004.

114. Robinson, N. E. and A. B. Robinson, “Molecular Clocks”, *Proceedings of the National Academy of Sciences*, Vol. 98, No. 3, pp. 944–949, 2001.

115. Capasso, S., L. Mazzarella, F. Sica, A. Zagari and S. Salvadori, “Kinetics and Mechanism of Succinimide Ring Formation in the Deamidation Process of Asparagine Residues”, *Journal of the Chemical Society, Perkin Transactions 2*, pp. 679–682, 1993.

116. Tam, J., M. Riemen and R. Merrifield, “Mechanisms of Aspartimide Formation: the Effects of Protecting Groups, Acid, Base, Temperature and Time.”, *Peptide Research*, Vol. 1, No. 1, p. 6, 1988.

117. Robinson, N., Z. Robinson, B. Robinson, A. Robinson, J. Robinson, M. Robin-

son and A. Robinson, “Structure-Dependent Nonenzymatic Deamidation of Glutaminyl and Asparaginyl Pentapeptides”, *The Journal of Peptide Research*, Vol. 63, No. 5, pp. 426–436, 2004.

118. Robinson, N. and A. Robinson, “Prediction of Primary Structure Deamidation Rates of Asparaginyl and Glutaminyl Peptides Through Steric and Catalytic Effects”, *The Journal of Peptide Research*, Vol. 63, No. 5, pp. 437–448, 2004.

119. Robinson, N. E. and A. B. Robinson, “Prediction of Protein Deamidation Rates from Primary and Three-Dimensional Structure”, *Proceedings of the National Academy of Sciences*, Vol. 98, No. 8, pp. 4367–4372, 2001.

120. Konuklar, F. A. S., V. Aviyente, T. Z. Sen and I. Bahar, “Modeling the Deamidation of Asparagine Residues via Succinimide Intermediates”, *Molecular Modeling Annual*, Vol. 7, No. 5, pp. 147–160, 2001.

121. Konuklar, F. A. S. and V. Aviyente, “Modelling the Hydrolysis of Succinimide: Formation of Aspartate and Reversible Isomerization of Aspartic Acid via Succinimide”, *Organic & Biomolecular Chemistry*, Vol. 1, No. 13, pp. 2290–2297, 2003.

122. Konuklar, F. A., V. Aviyente and M. F. Ruiz Lopez, “Theoretical Study on the Alkaline and Neutral Hydrolysis of Succinimide Derivatives in Deamidation Reactions”, *The Journal of Physical Chemistry A*, Vol. 106, No. 46, pp. 11205–11214, 2002.

123. Radkiewicz, J. L., H. Zipse, S. Clarke and K. N. Houk, “Neighboring side chain effects on asparaginyl and aspartyl degradation: an ab initio study of the relationship between peptide conformation and backbone NH acidity”, *Journal of the American Chemical Society*, Vol. 123, No. 15, pp. 3499–3506, 2001.

124. Peters, B. and L. Bernhardt, “Asparagine deamidation: pH-dependent mechanism from density functional theory”, *Biochemistry*, Vol. 45, No. 16, pp. 5384–5392,

2006.

125. Wierenga, R., “The TIM-barrel Fold: a Versatile Framework for Efficient Enzymes”, *FEBS Letters*, Vol. 492, No. 3, pp. 193–198, 2001.

126. Pujadas, G. and J. Palau, “Evolution of α-amylases: Architectural Features and Key Residues in the Stabilization of the (β/α) 8 Scaffold”, *Molecular Biology and Evolution*, Vol. 18, No. 1, pp. 38–54, 2001.

127. Yüksel, K. and R. W. Gracy, “In vitro deamidation of human triosephosphate isomerase”, *Archives of Biochemistry and Biophysics*, Vol. 248, No. 2, pp. 452 – 459, 1986.

128. Tyler-Cross, R. and V. Schirch, “Effects of Amino Acid Sequence, Buffers, and Ionic Strength on the Rate and Mechanism of Deamidation of Asparagine Residues in Small Peptides.”, *Journal of Biological Chemistry*, Vol. 266, No. 33, pp. 22549–22556, 1991.

129. Catak, S., G. Monard, V. Aviyente and M. F. Ruiz-López, “Reaction Mechanism of Deamidation of Asparaginyl Residues in Peptides: Effect of Solvent Molecules”, *The Journal of Physical Chemistry A*, Vol. 110, No. 27, pp. 8354–8365, Jul. 2006.

130. Catak, S., G. Monard, V. Aviyente and M. F. Ruiz-López, “Computational Study on Nonenzymatic Peptide Bond Cleavage at Asparagine and Aspartic Acid”, *The Journal of Physical Chemistry A*, Vol. 112, No. 37, pp. 8752–8761, 2008.

131. Catak, S., G. Monard, V. Aviyente and M. F. Ruiz-Lopez, “Deamidation of Asparagine Residues: Direct Hydrolysis versus Succinimide-Mediated Deamidation Mechanisms”, *The Journal of Physical Chemistry A*, Vol. 113, No. 6, pp. 1111–1120, Feb. 2009.

132. Jogl, G., S. Rozovsky, A. E. McDermott and L. Tong, “Optimal Alignment for Enzymatic Proton Transfer: Structure of the Michaelis Complex of Triosephos-

phate Isomerase at 1.2-Å Resolution", *Proceedings of the National Academy of Sciences*, Vol. 100, No. 1, pp. 50–55, 2003.

133. Needleman, S. B. and C. D. Wunsch, "A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins", *Journal of Molecular Biology*, Vol. 48, No. 3, pp. 443 – 453, 1970.

134. Case, D. A., T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, "The Amber Biomolecular Simulation Programs", *Journal of Computational Chemistry*, Vol. 26, No. 16, pp. 1668–1688, 2005.

135. Duan, Y., C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman, "A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations", *Journal of Computational Chemistry*, Vol. 24, No. 16, pp. 1999–2012, 2003.

136. Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, "A Smooth Particle Mesh Ewald Method", *The Journal of Chemical Physics*, Vol. 103, No. 19, pp. 8577–8593, 1995.

137. Ryckaert, J.-P., G. Ciccotti and H. J. Berendsen, "Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-alkanes", *Journal of Computational Physics*, Vol. 23, No. 3, pp. 327 – 341, 1977.

138. Li, H., A. D. Robertson and J. H. Jensen, "Very Fast Empirical Prediction and Rationalization of Protein pKa Values", *Proteins: Structure, Function, and Bioinformatics*, Vol. 61, No. 4, pp. 704–721, 2005.

139. Kosky, A. A., U. O. Razzaq, M. J. Treuheit and D. N. Brems, "The Effects of Alpha-helix on the Stability of Asn Residues: Deamidation Rates in Peptides of

Varying Helicity", *Protein Sciencence*, Vol. 8, No. 11, p. 2519–2523, 1999.

140. Konuklar, F. A. S., V. Aviyente, G. Monard and M. F. Ruiz Lopez, "Theoretical Approach to the Wear and Tear Mechanism in Triosephoshate Isomerase: A QM/MM Study", *The Journal of Physical Chemistry B*, Vol. 108, No. 12, pp. 3925–3934, 2004.

141. Takahashi, O., K. Kobayashi and A. Oda, "Modeling the Enolization of Succinimide Derivatives, a Key Step of Racemization of Aspartic Acid Residues: Importance of a Two-H2O Mechanism", *Chemistry & Biodiversity*, Vol. 7, No. 6, p. 1349–1356, 2010.

142. Garza-Ramos, G., M. Tuena de Gomez-Puyou, A. Gomez-Puyou, K. U. Yuksel and R. W. Gracy, "Deamidation of Triosephosphate Isomerase in Reverse Micelles: Effects of Water on Catalysis and Molecular Wear and Tear", *Biochemistry*, Vol. 33, No. 22, pp. 6960–6965, 1994.

143. Ugur, I., V. Aviyente and G. Monard, "Initiation of the Reaction of Deamidation in Triosephosphate Isomerase: Investigations by Means of Molecular Dynamics Simulations", *The Journal of Physical Chemistry B*, Vol. 116, No. 22, pp. 6288–6301, 2012.

144. Sun, A.-Q., K. U. Yüksel and R. W. Gracy, "Interactions Between the Catalytic Centers and Subunit Interface of Triosephosphate Isomerase Probed by Refolding, Active Site Modification, and Subunit Exchange.", *Journal of Biological Chemistry*, Vol. 267, No. 28, pp. 20168–20174, 1992.

145. Bas, D. C., D. M. Rogers and J. H. Jensen, "Very Fast Prediction and Rationalization of pKa Values for Protein–Ligand Complexes", *Proteins: Structure, Function, and Bioinformatics*, Vol. 73, No. 3, pp. 765–783, 2008.

146. Olsson, M. H. M., C. R. Søndergard, M. Rostkowski and J. H. Jensen, "PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical

pKa predictions", *Journal of Chemical Theory and Computation*, Vol. 7, No. 2, pp. 525 – 537, 2011.

147. Søndergaard, C. R., M. H. M. Olsson, M. Rostkowski and J. H. Jensen, "Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values", *Journal of Chemical Theory and Computation*, Vol. 7, No. 7, pp. 2284–2295, 2011.

148. Götz, A. W., M. J. Williamson, D. Xu, D. Poole, S. Le Grand and R. C. Walker, "Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born", *Journal of Chemical Theory and Computation*, Vol. 8, No. 5, pp. 1542–1555, 2012.

149. Salomon-Ferrer, R., A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, "Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald", *Journal of Chemical Theory and Computation*, Vol. 9, No. 9, pp. 3878–3888, 2013.

150. Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, "Gaussian 09 Revision B.01", Gaussian Inc. Wallingford CT, 2009.

151. Oliva, A., B. Henry and M. F. Ruiz-López, "Insights on Peptide Backbone N–H Acidity: Structure of Anions, Hydration Effects", *Chemical Physics Letters*, Vol. 561–562, No. 0, pp. 153 – 158, 2013.

152. Wakankar, A. A. and R. T. Borchardt, "Formulation Considerations for Proteins Susceptible to Asparagine Deamidation and Aspartate Isomerization", *Journal of Pharmaceutical Sciences*, Vol. 95, No. 11, pp. 2321–2336, 2006.

153. Robinson, N. E., "Protein deamidation", *Proceedings of the National Academy of Sciences*, Vol. 99, No. 8, p. 5283–5288, 2002.

154. Manning, M. C., D. K. Chou, B. M. Murphy, R. W. Payne and D. S. Katayama, "Stability of Protein Pharmaceuticals: an Update", *Pharmaceutical Research*, Vol. 27, No. 4, pp. 544–575, 2010.

155. Capasso, S., "Estimation of the Deamidation Rate of Asparagine Side Chains", *The Journal of Peptide Research*, Vol. 55, No. 3, p. 224–229, 2000.

156. Capasso, S. and P. Di Cerbo, "Kinetic and Thermodynamic Control of the Relative Yield of the Deamidation of Asparagine and Isomerization of Aspartic Acid Residues", *The Journal of Peptide Research*, Vol. 56, No. 6, p. 382–387, 2000.

157. Kaliman, I., A. Nemukhin and S. Varfolomeev, "Free Energy Barriers for the N-Terminal Asparagine to Succinimide Conversion: Quantum Molecular Dynamics Simulations for the Fully Solvated Model", *Journal of Chemical Theory and Computation*, Vol. 6, No. 1, pp. 184–189, 2010.

158. Ugur, I., A. Marion, V. Aviyente and G. Monard, "Why Asn71 deamidates faster than Asn15 in the Enzyme Triosephosphate Isomerase?: Answers from μs Molecular Dynamic Simulations", *The Journal of Computational Chemistry*, 2014, submitted.

159. Brooks, C. L. and M. Karplus, "Deformable Stochastic Boundaries in Molecular

Dynamics", *The Journal of Chemical Physics, 1983*, Vol. 79, No. 6312, 1983.

160. M. Seabra, G., R. C. Walker and A. E. Roitberg, "Are Current Semiempirical Methods Better Than Force Fields? A Study from the Thermodynamics Perspective", *The Journal of Physical Chemistry A*, Vol. 113, No. 43, pp. 11938–11948, 2009.

161. Senn, H. M. and W. Thiel, "QM/MM Studies of Enzymes", *Current Opinion in Chemical Biology*, Vol. 11, No. 2, pp. 182–187, 2007.

162. Imhof, P., F. Noé, S. Fischer and J. C. Smith, "AM1/d Parameters for Magnesium in Metalloenzymes", *Journal of Chemical Theory and Computation*, Vol. 2, No. 4, pp. 1050–1056, 2006.

163. Nam, K., Q. Cui, J. Gao and D. M. York, "Specific Reaction Parametrization of the AM1/d Hamiltonian for Phosphoryl Transfer Reactions: H, O, and P Atoms", *Journal of Chemical Theory and Computation*, Vol. 3, No. 2, pp. 486–504, 2007.

164. Korth, M., M. Pitonak, J. Rezac and P. Hobza, "A Transferable H-bonding Correction for Semiempirical Quantum-Chemical Methods", *Journal of Chemical Theory and Computation*, Vol. 6, No. 1, pp. 344–352, 2009.

165. Rocha, G. B., R. O. Freire, A. M. Simas and J. J. Stewart, "RM1: A Reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I", *Journal of Computational Chemistry*, Vol. 27, No. 10, pp. 1101–1111, 2006.

166. Jurečka, P., J. Černý, P. Hobza and D. R. Salahub, "Density Functional Theory Augmented with an Empirical Dispersion Term. Interaction Energies and Geometries of 80 Noncovalent Complexes Compared with ab initio Quantum Mechanics Calculations", *Journal of Computational Chemistry*, Vol. 28, No. 2, pp. 555–569, 2007.

167. Seabra, G. M., R. C. Walker, M. Elstner, D. A. Case and A. E. Roitberg, "Im-

plementation of the SCC-DFTB Method for Hybrid QM/MM Simulations within the Amber Molecular Dynamics Package.”, *The Journal of Physical Chemistry A*, Vol. 20, pp. 5655 – 5664, 2007.

168. Porezag, D., T. Frauenheim, T. Kohler, G. Seifert and R. Kaschner, “Construction of Tight-Binding-Like Potentials on the Basis of Density-Functional-Theory: Applications to Carbon.”, *Physical Review B*, Vol. 51, p. 12947, 1995.

169. Hammond, G. S., “A Correlation of Reaction Rates”, *Journal of the American Chemical Society*, Vol. 77, No. 2, pp. 334–338, 1955.

170. Thurlkill, R. L., G. R. Grimsley, J. M. Scholtz and C. N. Pace, “pK Values of the Ionizable Groups of Proteins”, *Protein science*, Vol. 15, pp. 1214–1218, 2006.

171. Jensen, J. H., H. Li, A. D. Robertson and P. A. Molina, “Prediction and Rationalization of Protein pKa Values Using QM and QM/MM Methods.”, *The Journal of Physical Chemistry. A*, Vol. 109, No. 30, pp. 6634–43, 2005.

172. Thurlkill, R. L., G. R. Grimsley, J. M. Scholtz and C. N. Pace, “Hydrogen Bonding Markedly Reduces the pK of Buried Carboxyl Groups in Proteins.”, *Journal of Molecular Biology*, Vol. 362, No. 3, pp. 594–604, 2006.

173. Laurents, D. V., B. M. Huyghues-Despointes, M. Bruix, R. L. Thurlkill, D. Schell, S. Newsom, G. R. Grimsley, K. L. Shaw, S. Treviño, M. Rico, J. M. Briggs, J. M. Antosiewicz, J. Scholtz and C. Pace, “Charge–Charge Interactions are Key Determinants of the pK Values of Ionizable Groups in Ribonuclease Sa (pI=3.5) and a Basic Variant (pI=10.2)”, *Journal of Molecular Biology*, Vol. 325, No. 5, pp. 1077–1092, 2003.

174. Isom, D. G., C. A. Castañeda, B. R. Cannon and B. García-moreno E., “Large Shifts in pK a Values of Lysine Residues Buried Inside a Protein”, *Proceedings of the National Academy of Sciences USA*, Vol. 108, No. 13, pp. 5260–6265, 2011.

175. Li, H., A. D. Robertson and J. H. Jensen, “The Determinants of Carboxyl pKa Values in Turkey Ovomucoid Third Domain.”, *Proteins*, Vol. 55, No. 3, pp. 689–704, 2004.

176. Quijada, J., G. López, R. Versace, L. Ramírez and M. L. Tasayco, “On the NMR analysis of pKa values in the unfolded state of proteins by extrapolation to zero denaturant.”, *Biophysical Chemistry*, Vol. 129, No. 2-3, pp. 242–50, 2007.

177. Baran, K. L., M. S. Chimenti, J. L. Schlessman, C. A. Fitch, K. J. Herbst and B. E. Garcia-Moreno, “Electrostatic Effects in a Network of Polar and Ionizable Groups in Staphylococcal Nuclease.”, *Journal of Molecular Biology*, Vol. 379, No. 5, pp. 1045–62, 2008.

178. Fitch, C. A., D. A. Karp, K. K. Lee, W. E. Stites, E. E. Lattman and E. B. García-Moreno, “Experimental pKa Values of Buried Residues: Analysis with Continuum Methods and Role of Water Penetration”, *Biophysical Journal*, Vol. 82, No. 6, pp. 3289–3304, 2002.

179. Nelson, K. J., D. Parsonage, A. Hall, P. A. Karplus and L. B. Poole, “Cysteine pK(a) Values for the Bacterial Peroxiredoxin AhpC.”, *Biochemistry*, Vol. 47, No. 48, pp. 12860–8, 2008.

180. Grimsley, G. R., J. M. Scholtz and C. N. Pace, “A Summary of the Measured pK Values of the Ionizable Groups in Folded Proteins.”, *Protein Science : a Publication of the Protein Society*, Vol. 18, No. 1, pp. 247–51, 2009.

181. Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, “The Protein Data Bank.”, *Nucleic Acids Research*, Vol. 28, No. 1, pp. 235–42, 2000.

182. Abul Kashem Liton, M., M. Idrish Ali and M. Tanvir Hossain, “Accurate pKa Calculations for Trimethylaminium Ion with a Variety of Basis Sets and Methods Combined with CPCM Continuum Solvation Methods”, *Computational and*

Theoretical Chemistry, Vol. 999, pp. 1–6, 2012.

183. Namazian, M., M. Zakery, M. R. Noorbala and M. L. Coote, "Accurate Calculation of the pKa of Trifluoroacetic Acid Using High-Level ab initio Calculations", *Chemical Physics Letters*, Vol. 451, No. 1-3, pp. 163–168, 2008.

184. Liptak, M. D., K. C. Gross, P. G. Seybold, S. Feldgus and G. C. Shields, "Absolute pK(a) Determinations for Substituted Phenols.", *Journal of the American Chemical Society*, Vol. 124, No. 22, pp. 6421–7, 2002.

185. Satchell, J. F. and B. J. Smith, "Calculation of Aqueous Dissociation Constants of 1,2,4-triazole and Tetrazole: A Comparison of Solvation Models.", *Physical Chemistry Chemical Physics*, Vol. 4, No. 18, pp. 4314–4318, 2002.

186. Baptista, A. M., P. J. Martel and S. B. Petersen, "Simulation of Protein Conformational Freedom as a Function of pH: Constant-pH Molecular Dynamics using Implicit Titration.", *Proteins*, Vol. 27, No. 4, pp. 523–44, 1997.

187. Wallace, J. a. and J. K. Shen, *Predicting pKa Values with Continuous Constant pH Molecular Dynamics.*, Vol. 466, Elsevier Inc., 1 edn., 2009.

188. Itoh, S. G., A. Damjanović and B. R. Brooks, "pH Replica-Exchange Method Based on Discrete Protonation States.", *Proteins*, Vol. 79, No. 12, pp. 3420–36, 2011.

189. Meng, Y., D. S. Dashti and A. E. Roitberg, "Computing Alchemical Free Energy Differences with Hamiltonian Replica Exchange Molecular Dynamics (H-REMD) Simulations.", *Journal of Chemical Theory and Computation*, Vol. 7, No. 9, pp. 2721–2727, 2011.

190. Swails, J. M. and A. E. Roitberg, "Enhancing Conformation and Protonation State Sampling of Hen Egg White Lysozyme Using pH Replica Exchange Molecular Dynamics", *Journal of Chemical Theory and Computation*, Vol. 8, No. 11,

pp. 4393–4404, 2012.

191. Bashford, D. and M. Karplus, “pKa’s of Ionizable Groups in Proteins: Atomic Detail from a Continuum Electrostatic Model.”, *Biochemistry*, Vol. 29, No. 44, pp. 10219–25, 1990.

192. Bashford, D., “An Object-Oriented Programming Suite for Electrostatic Effects in Biological Molecules an Experience Report on the MEAD Project”, *Scientific Computing in Object-Oriented Parallel Environment*, Vol. 1343, pp. 233–240, 1997.

193. Dillet, V., R. L. Van Etten and D. Bashford, “Stabilization of Charges and Protonation States in the Active Site of the Protein Tyrosine Phosphatases: A Computational Study”, *The Journal of Physical Chemistry B*, Vol. 104, No. 47, pp. 11321–11333, 2000.

194. Anandakrishnan, R., B. Aguilar and A. V. Onufriev, “H++ 3.0: Automating pK Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations.”, *Nucleic Acids Research*, Vol. 40, No. Web Server issue, pp. W537–41, 2012.

195. Zhang, S., J. Baker and P. Pulay, “A Reliable and Efficient First Principles-Based Method for Predicting pK(a) Values. 2. Organic Acids.”, *The Journal of Physical Chemistry. A*, Vol. 114, No. 1, pp. 432–42, 2010.

196. Zhang, S., J. Baker and P. Pulay, “A Reliable and Efficient First Principles-Based Method for Predicting p K a Values. 1. Methodology”, *The Journal of Physical Chemistry A*, Vol. 114, No. 1, p. 425–431, 2009.

197. Roos, G., S. Loverix and P. Geerlings, “Origin of the p K a Perturbation of N-Terminal Cysteine in α-and 310-Helices: A Computational DFT Study”, *The Journal of Physical Chemistry B*, Vol. 110, pp. 557–562, 2006.

198. Namazian, M., M. Zakery, M. R. Noorbala and M. L. Coote, "Accurate Calculation of the pKa of Trifluoroacetic Acid Using High-Level ab initio Calculations", *Chemical Physics Letters*, Vol. 451, No. 1–3, pp. 163 – 168, 2008.

199. Namazian, M. and M. L. Coote, "Gas-Phase Acidity, Bond Dissociation Energy and Enthalpy of Formation of Fluorine-Substituted Benzenes: A Theoretical Study", *Journal of Fluorine Chemistry*, Vol. 130, No. 7, pp. 621 – 628, 2009.

200. Hamann, S. D. and M. Linton, "Influence of Pressure on the Ionization of Substituted Phenols", *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, Vol. 70, p. 2239–2249, 1974.

201. Fischer, A., G. J. Leary, R. D. Topsom and J. Vaughan, "Ionic Dissociation of 4-substituted Phenols and 2, 6-dichloro-and 2, 6-dimethyl-phenols in Organic Solvents", *J. Chem. Soc. B*, p. 846–851, 1967.

202. Fickling, M. M., A. Fischer, B. R. Mann, J. Packer and J. Vaughan, "Hammett Substituent Constants for Electron-Withdrawing Substituents: Dissociation of Phenols, Anilinium Ions and Dimethylanilinium Ions", *Journal of the American Chemical Society*, Vol. 81, No. 16, p. 4226–4230, 1959.

203. Albert, A. and J. N. Phillips, "264. Ionization Constants of Heterocyclic Substances. Part II. Hydroxy-derivatives of Nitrogenous Six-Membered Ring-Compounds", *Journal of the Chemical Society (Resumed)*, p. 1294–1304, 1956.

204. De Maria, P., A. Fini and F. M. Hall, "Thermodynamic Acidity Constants of Ortho-Substituted Benzenethiols", *Journal of the Chemical Society, Perkin Transactions 2*, , No. 12, p. 1443–1445, 1974.

205. Bolton, P. D., F. M. Hall and I. H. Reece, "Effects of Substituents on the Thermodynamic Functions of Ionisation of Meta-Substituted Phenols", *J. Chem. Soc. B*, p. 709–712, 1967.

206. Palm, V. A., *Tables of Rate and Equilibrium Constants of Heterocyclic Organic Reactions*, Moscow, 1976.

207. Ko, H. C., W. F. O'Hara, T. Hu and L. G. Hepler, "Ionization of Substituted Phenols in Aqueous Solution", *Journal of the American Chemical Society*, Vol. 86, No. 6, p. 1003–1004, 1964.

208. Serjant E.P, B., Dempsey, *Ionisation Constants of Organic Acids in Aqueous Solution (IUPAC Chemistry Data Series No:23)*, Pregmon Press, 1979.

209. *Manual of ACD/Structure Elucidator*, MIT, 2013.

210. Takahashi, S., L. A. Cohen, H. K. Miller and E. G. Peake, "Calculation of the pKa Values of Alcohols From Sigma Constants and from the Carbonyl Frequencies of Their Esters", *The Journal of Organic Chemistry*, Vol. 36, No. 9, pp. 1205 – 1209, 1971.

211. Lide, D., *Handbook of Chemistry and Physics*, CRS Press, 87 edn., 2006-2007.

212. Maria, P., F. A. and F. M. Hall, "Thermodynamic Acid Dissociation Constants of Aromatic Thiols", *Journal of the Chemical Society, Perkin Transactions 2*, p. 1969–1971, 1973.

213. Kreevoy, M. M., E. T. Harper, R. E. Duvall, H. S. Wilgus III and L. T. Ditsch, "Inductive Effects on the Acid Dissociation Constants of Mercaptans", *Journal of the American Chemical Society*, Vol. 82, No. 18, p. 4899–4902, 1960.

214. Arnold, A. P. and A. J. Canty, "Methylmercury (II) Sulfhydryl Interactions. Potentiometric Determination of the Formation Constants for Complexation of Methylmercury (II) by Sulfhydryl Containing Amino Acids and Related Molecules, Including Glutathione", *Canadian Journal of Chemistry*, Vol. 61, No. 7, p. 1428–1434, 1983.

215. Pettit, L. D. and K. Powell, “IUPAC stability constants database”, *Chemistry International*, 2006.

216. Kreevoy, M. M., B. E. Eichinger, F. E. Stary, E. A. Katz and J. H. Sellstedt, “The Effect of Structure on Mercaptan Dissociation Constants”, *The Journal of Organic Chemistry*, Vol. 29, No. 6, p. 1641–1642, 1964.

217. Tsonopoulos, C., D. M. Coulson and L. B. Inman, “Ionization Constants of Water Pollutants”, *Journal of Chemical and Engineering Data*, Vol. 21, No. 2, p. 190–193, 1976.

218. Irving, R. J. and W. L. Nelander L., “Thermodynamics of the Ionization of Some Thiols in Aqueous Solution”, *Acta Chemica Scandinavica*, Vol. 18, p. 769–787, 1964.

219. Kortum, V. W. A. K., G., *Dissociation Constants of Organic Acids in Aqueous Solution*, London,Butterworths, 1961.

220. Thurlkill, R. L., G. R. Grimsley, J. M. Scholtz and C. N. Pace, “pK Values of the Ionizable Groups of Proteins”, *Protein Sciencence*, Vol. 15, No. 5, pp. 1214–1218, May 2006.

221. Hupe, D. J. and W. P. Jencks, “Nonlinear Structure-Reactivity Correlations. Acyl Transfer Between Sulfur and Oxygen Nucleophiles”, *Journal of the American Chemical Society*, Vol. 99, No. 2, pp. 451–464, 1977.

222. Lee, C., W. Yang and R. G. Parr, “Development of the Colle-Salvetti Correlation-Energy Formula Into a Functional of the Electron Density”, *Physical Review B*, Vol. 37, No. 2, p. 785, 1988.

223. Becke, A. D., “Density-functional Exchange-Energy Approximation with Correct Asymptotic Behavior”, *Physical Review A*, Vol. 38, pp. 3098–3100, 1988.

224. Handy, N. C. and A. J. Cohen, “Left-Right Correlation Energy”, *Molecular Physics*, Vol. 99, No. 5, pp. 403–412, 2001.

225. Hoe, W.-M., A. J. Cohen and N. C. Handy, “Assessment of a New Local Exchange Functional OPTX”, *Chemical Physics Letters*, Vol. 341, No. 3, pp. 319–328, 2001.

226. Perdew, J. P., K. Burke and M. Ernzerhof, “Generalized Gradient Approximation Made Simple”, *Physical Review Letters*, Vol. 77, pp. 3865–3868, 1996.

227. Zhao, Y. and D. G. Truhlar, “The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals”, *Theoretical Chemistry Accounts*, Vol. 120, No. 1-3, pp. 215–241, 2008.

228. Zhao, Y. and D. G. Truhlar, “Density Functionals with Broad Applicability in Chemistry”, *Accounts of Chemical Research*, Vol. 41, No. 2, pp. 157–167, 2008.

229. Zhao, Y. and D. G. Truhlar, “A New Local Density Functional for Main-Group Thermochemistry, Transition Metal Bonding, Thermochemical Kinetics, and Noncovalent Interactions”, *The Journal of Chemical Physics*, Vol. 125, p. 194101, 2006.

230. Singh, U. C. and P. A. Kollman, “An Approach to Computing Electrostatic Charges for Molecules”, *Journal of Computational Chemistry*, Vol. 5, No. 2, pp. 129–145, 1984.

231. Breneman, C. M. and K. B. Wiberg, “Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis”, *Journal of Computational Chemistry*, Vol. 11, No. 3, p. 361, 1990.

232. Bayly, C. I., P. Cieplak, W. Cornell and P. A. Kollman, “A Well-Behaved Elec-

trostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: the RESP Model", *The Journal of Physical Chemistry*, Vol. 97, No. 40, pp. 10269–10280, 1993.

233. Dunning Jr, T. H., "Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen", *The Journal of Chemical Physics*, Vol. 90, p. 1007, 1989.

234. Onufriev, A., D. A. Case and D. Bashford, "Effective Born Radii in the Generalized Born Approximation: The Importance of Being Perfect", *Journal of Computational Chemistry*, Vol. 23, pp. 1297 – 1304, 2002.

235. Götz, A. W., M. A. Clark and R. C. Walker, "An Extensible Interface for QM/MM Molecular Dynamics Simulations with AMBER", *Journal of Computational Chemistry*, Vol. 35, No. 2, pp. 95–108, 2014.

236. Vareková, R. S., S. Geidl, C.-M. Ionescu, O. Skrehota, M. Kudera, D. Sehnal, T. Bouchal, R. Abagyan, H. J. Huber and J. Koca, "Predicting pKa Values of Substituted Phenols from Atomic Charges: Comparison of Different Quantum Mechanical Methods and Charge Distribution Schemes", *Journal of Chemical Information and Modeling*, Vol. 51, p. 1795–1806, 2011.

Résumé (français): La déamidation est la modification post-traductionnelle de l'asparagine (Asn) et de la glutamine (Glu). Elle est communèment observée dans les peptides et les protéines. Il a été démontré que la déamidation limite la durée de vie de ces macromolécules. Dans ce travail, la déamidation de l'asparagine dans des petits peptides et dans l'enzyme triosephosphate isomérase a été modélisée. La déamidation dans la triosephosphate isomérase de mammifères a été observée sur deux sites distincts: Asn15 et Asn71. Asn71 a une vitesse de déamidation plus élevée que Asn15 et moins grande que pour un petit peptide. Il a été suggéré que la déamidation de Asn15 se produit sous l'influence de la déamidation de Asn71. Pour expliquer ces résultats expérimentaux, des simulations de dynamiques moléculaires classiques à l'échelle de la microseconde et des calculs d'énergie libre, de type *umbrella sampling*, à l'aide de méthodes combinées mécanique quantique/mécanique moléculaire ont été réalisés. Nous montrons que la déamidation séquentielle dans la triosephosphate isomérase est due à la fois à des effets locaux et globaux. Ces résultats apporte une nouvelle perspective sur l'impact de l'ordre structurel sur la vitesse de déamidation Nous avons également déterminé la voie la plus plausible de cette reaction ainsi que l'influence de la variation du pK_a, dans la chaîne principale, de la partie amide du résidu adjacent de l'asparagine sur la vitesse de déamidation. En regard de l'importance des variations de pK_a dans l'environnement protéique, nous avons élaboré un protocole informatique permettant d'évaluer de manière rapide et précise des pK_a. Ce protocole a été appliqué à des petites molécules organiques et nous avons montré qu'il était également applicable à des études relatives à la prédiction de pK_a dans les protéines.

Résumé (anglais): Deamidation is the posttranslational modification of asparagine (Asn) and glutamine (Glu) residues, which is observed in several proteins and peptides. It has been shown that deamidation limits the lifetime of these macromolecules. In this work, deamidation of asparagine in small peptides and in the enzyme triosephosphate isomerase has been modeled. Deamidation in mammalian triosephosphate isomerase has been observed at two distinct deamidation sites: Asn15 and Asn71. Asn71 deamidates faster than Asn15 and slower than a small peptide. It has been suggested that, deamidation at Asn15 occurs with the influence of deamidated Asn71. In order to explain these experimental findings, microsecond long classical molecular dynamics simulations and free energy calculations using quantum mechanics/molecular mechanics tools combined with umbrella sampling technique have been performed. The sequential deamidation in triosephosphate isomerase has been shown to be related with both global and local effects. These results bring a new perspective to the impact of the high-order structure on deamidation rate. The most plausible route of this reaction was also determined. The pK_a shift of backbone amide of the residue adjacent to asparagine has been found to be one of the most crucial factor determining the rate of deamidation. Considering the importance of pK_a shifts in protein environment, a computational protocol was suggested in order to obtain accurate and fast pK_a predictions. This protocol was applied to small organic molecules, and it has been shown to be applicable to studies concerning aminoacid pK_a predictions.

Mots-clés: déamidation ; asparagine ; triosephosphate isomerase; dynamique du moléculaire, calculs d'énergie libre, mécanique quantique/mécanique moléculaire, calculs du pK_a.