



## 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](mailto: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.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

---

**THÈSE**

présentée en vue de l'obtention du titre de

**Docteur de l'Université de Lorraine  
en Chimie**

par

**Hatice GÖKCAN**

---

**Modélisation moléculaire de la réactivité de GABA-AT: de petits  
modèles représentatifs à la protéine complète, de la mécanique  
moléculaire à la chimie quantique, du statique au dynamique****Molecular modelling of GABA-AT reactivity: from small  
representative models to the full protein, from molecular mechanics  
to quantum chemistry, from static to dynamics**

---

Soutenue publiquement le 2 Septembre 2016  
à l'Université Istanbul Technique, Istanbul, Turquie**Membres du jury**

<b>Président :</b>	Prof. Mine Yurtsever	Istanbul Technical University, Istanbul, Turkey
<b>Rapporteurs :</b>	Prof. Safiye Erdem	Marmara University, Istanbul, Turkey
	Assoc. Prof. Pedro A. Fernandes	University of Porto, Porto, Portugal
<b>Examineurs :</b>	Prof. Pemra Doruker Turgut	Bogazici University, Istanbul, Turkey
	Prof. Xavier Assfeld	Université de Lorraine, Vandœuvre-lès-Nancy, France
<b>Directeurs :</b>	Prof. Gérald Monard	Université de Lorraine, Vandœuvre-lès-Nancy, France
	Assoc. Prof. F. Aylin Sungur	Istanbul Technical University, Istanbul, Turkey



## FOREWORD

I first would like to express my most sincere gratitude and appreciation to the jury members for taking their time to judge this work. Especially, I truly appreciate Prof. Erdem and Dr. Fernandes, for having reviewed this manuscript.

I also wish to express my gratitude to my advisors, Prof. Monard and Dr. Sungur, for their time, patience, and endless supports during the time that we shared together. I am grateful to be able to work with you who thought me so much.

I also want to thank to TUBITAK-Science Fellowships and Grant Programmes Department, and Embassy of France in Turkey for the Ph.D. financial support.

I had the opportunity to work in a wonderful environment of the TMS research team. I feel very lucky to be able to meet many talented researchers; Dr. Monari, Dr. Marazzi, Dr. Pastore, İlke, Antoine, Gülşah, Benjamin, Hugo, Wanlei, Marina, Thibaud, Lea, Shaurya, Yiğitcan, Marilyne... I wish to thank all of you for nice moments and the friendship that we shared. I also wish to thank to the members of the Computational Chemistry group in Informatics Institute; Mustafa, Erdem, and Melin. Thank you for your support and friendship.

I also feel very luck to have met with Evrim, Süha, and Derya in ITU. Thank you for being there when I needed you. I will never forget your kindness, and thoughtfulness. I am particularly grateful to Songül who is a wonderful friend in every way. Thank you for being by my side, and always giving me reasons to cheer.

Finally, I wish to extend my gratitude to my family. Thank you for sharing my dreams, for your devotion, support, and patience, for being the source of my strength. I cannot express how grateful I am for the incomparable love that you gave me, and how proud I am to have such a wonderful family.

September 2016

Hatice GÖKCAN





## TABLE OF CONTENTS

	<u>Page</u>
<b>FOREWORD.....</b>	<b>i</b>
<b>TABLE OF CONTENTS.....</b>	<b>iii</b>
<b>ABBREVIATIONS .....</b>	<b>vii</b>
<b>LIST OF TABLES .....</b>	<b>ix</b>
<b>LIST OF FIGURES .....</b>	<b>xi</b>
<b>SUMMARY .....</b>	<b>xxi</b>
<b>RÉSUMÉ .....</b>	<b>xxv</b>
<b>1. INTRODUCTION .....</b>	<b>1</b>
<b>2. METHODOLOGY .....</b>	<b>9</b>
2.1 Quantum Chemistry.....	9
2.1.1 The Schrödinger equation.....	9
2.1.2 The Born Oppenheimer approximation.....	11
2.1.3 Electron spin, Pauli principle and Slater determinant .....	11
2.1.4 Linear variation method and secular equation.....	13
2.1.5 The Hartree-Fock method.....	14
2.1.5.1 The Roothaan-Hall equation and self consistent field (SCF) method .....	15
2.1.6 Semiempirical methods .....	16
2.1.7 Density functional theory .....	19
2.1.7.1 Approximations to exchange-correlation energy functional .....	20
2.1.7.2 Hybrid functionals .....	22
2.1.8 Solvation effect.....	23
2.1.8.1 Polarized continuum model .....	25
2.1.8.2 Integral equation formalism.....	25
2.1.9 Potential energy surface.....	26
2.2 Molecular Mechanics .....	28
2.3 Molecular Dynamics .....	29
2.3.1 Integration of Newton's equations of motion .....	30
2.3.2 Periodic boundary conditions .....	32
2.3.3 Short and long range interactions .....	33
2.3.3.1 Ewald summation .....	34
2.3.4 Thermodynamic ensembles.....	35
2.3.5 Temperature coupling .....	36
2.4 Semiempirical Born-Oppenheimer Molecular Dynamics (SEBOMD).....	37
2.4.1 Simulations .....	37
2.4.2 Divide and conquer (D&C) .....	37
2.4.3 The <code>sebomd</code> module .....	40

<b>3. RESULTS .....</b>	<b>43</b>
3.1 Mechanistic Details of Transimination and HF Elimination Reactions .....	43
3.1.1 Computational details .....	44
3.1.2 Results and Discussion .....	45
3.1.2.1 Solution and bioactive conformations as zwitterions .....	45
3.1.2.2 Transimination reactions .....	46
3.1.2.3 Fluoride ion elimination .....	51
3.1.3 Concluding remarks .....	57
3.2 Molecular Dynamics Simulations of Apo, Holo and Inactivator Bound GABA-AT .....	59
3.2.1 Computational details .....	60
3.2.1.1 Preparation of samples .....	60
3.2.1.2 Force field modifications .....	62
3.2.1.3 Molecular dynamics simulation .....	63
3.2.1.4 Estimation of GABA-AT/PLP interaction free energies .....	63
3.2.1.5 Trajectory analysis .....	65
3.2.2 Results and Discussion .....	65
3.2.2.1 Preliminary study on GABA-AT .....	65
3.2.2.2 Protonation states in GABA-AT .....	67
3.2.2.3 PLP phosphate group .....	77
3.2.2.4 Carbonyl oxygen on the pyridine ring of PLP .....	78
3.2.2.5 PLP pyridine ring .....	81
3.2.2.6 Vigabatrin carboxyl group .....	81
3.2.2.7 Conserved Arg residue .....	83
3.2.2.8 The roles of active site residues on the stability of the enzyme ....	84
3.2.2.9 The effect of inactivator on the active site .....	86
3.2.2.10 Negative cooperativity .....	87
3.2.2.11 Remarks on the catalytic mechanism of GABA-AT .....	87
3.2.2.12 Solvent access to the active site .....	89
3.2.3 Concluding remarks .....	89
3.3 The Effect of Surrounding Residues on Inactivation Mechanism of GABA-AT with $\gamma$ -vinyl GABA .....	91
3.3.1 Computational details .....	93
3.3.2 Results and Discussion .....	94
3.3.2.1 Michael addition reaction with cluster models .....	95
3.3.2.2 Effect of Surrounding Residues on Geometries .....	102
3.3.3 Concluding remarks .....	105
3.4 Improvement of the Linear Scaling of the SEBOMD Simulations .....	107
3.4.1 Computational details .....	108
3.4.1.1 Water molecules .....	108
3.4.1.2 Protein .....	109
3.4.2 Results and Discussion .....	109
3.4.2.1 Implementation of eigenvalue solvers .....	110
3.4.2.2 External control .....	114
3.4.2.3 Preliminary study for profiling .....	115
3.4.2.4 Performance with LAPACK .....	118

3.4.2.5	Performance with SCALAPACK.....	124
3.4.2.6	Performance with Intel MKL threading .....	127
3.4.3	Concluding remarks.....	130
3.5	SEBOMD Simulations on C-N bond breaking reactions involving PLP complexes .....	133
3.5.1	Computational details .....	134
3.5.1.1	QM calculations.....	135
3.5.1.2	QM/MM calculations.....	135
3.5.1.3	SEBOMD simulations .....	136
3.5.2	Results and Discussion .....	136
3.5.2.1	Preliminary study on bond breaking reactions .....	137
3.5.2.2	SEBOMD simulations .....	139
3.5.3	Concluding remarks.....	140
<b>4.</b>	<b>CONCLUSIONS.....</b>	<b>145</b>
	<b>REFERENCES.....</b>	<b>149</b>
	<b>APPENDICES .....</b>	<b>169</b>



## ABBREVIATIONS

<b>AM1</b>	: Austin model 1
<b>ASC</b>	: Apparent surface charge
<b>B3LYP</b>	: Becke, three-parameter, Lee-Yang-Parr
<b>BLACS</b>	: Basic linear algebra communication subprograms
<b>CM1</b>	: Charge model 1
<b>DFT</b>	: Density functional theory
<b>DFTB</b>	: Density Functional based Tight Binding
<b>D&amp;C</b>	: Divide and conquer
<b>ESP</b>	: Electrostatic potential
<b>FFT</b>	: Fast Fourier transforms
<b>G09</b>	: Gaussian'09 program package
<b>GABA</b>	: $\gamma$ -aminobutyric acid
<b>GABA-AT</b>	: $\gamma$ -aminobutyric acid aminotransferase
<b>GAFF</b>	: General AMBER force field
<b>GB</b>	: Generalized Born
<b>GGA</b>	: Generalized gradient approximation
<b>GIST</b>	: Grid inhomogeneous solvation theory
<b>IEFPCM</b>	: Integral equation formalism variant of the PCM
<b>IRC</b>	: Intrinsic reaction coordinate
<b>LAPACK</b>	: Linear algebra package
<b>LDA</b>	: Local density approximation
<b>LSDA</b>	: Local spin density approximation
<b>MD</b>	: Molecular Dynamics
<b>MM</b>	: Molecular Mechanics
<b>MM-GB/SA</b>	: MM energies combined with GB and SA method
<b>MKL</b>	: Intel Math Kernel Library
<b>NBO</b>	: Natural bond orbital
<b>NPT</b>	: Isobaric-isothermal ensemble
<b>NVE</b>	: Microcanonical ensemble
<b>NVT</b>	: Canonical ensemble
<b>PBC</b>	: Periodic boundary conditions
<b>PCM</b>	: Polarized continuum model
<b>PDDG/PM3</b>	: Pairwise distance directed gaussian modification
<b>PES</b>	: Potential energy surface
<b>PLP</b>	: Pyridoxal-5'-phosphate
<b>PM3</b>	: Parametrized model number 3
<b>PME</b>	: Particle-mesh Ewald
<b>PPPM</b>	: Particle-Particle Particle-Mesh method
<b>QM</b>	: Quantum mechanical
<b>RMSd</b>	: Root-mean-square deviation
<b>SA</b>	: Surface area continuum solvation
<b>SAS</b>	: Solvent accessible surface

**SCALAPACK**: Scalable Linear algebra package  
**SCF** : Self consistent field  
**SCS** : Standard closed shell algorithm  
**SEBOMD** : Semiempirical Born-Oppenheimer Molecular Dynamics  
**SES** : Solvent excluded surface  
**SPME** : Smooth particle mesh Ewald  
**TS** : Transition state structure  
**UFF** : Universal force field  
**Vigabatrin** :  $\gamma$ -vinyl-GABA

## LIST OF TABLES

	<u>Page</u>
<b>Table 3.1</b> : Nomenclature of the Samples According to the Protonation States of the Asp298 and His190 Residues. ....	61
<b>Table 3.2</b> : Relative Free Energy of Interactions ( $\Delta\Delta G_{int}$ ) Between PLP and GABA-AT Generated from Holo-GABA-AT Trajectories, in kcal/mol.....	74
<b>Table 3.3</b> : Abbreviations of SEBOMD simulations and their explanations.....	111
<b>Table 3.4</b> : Newly implemented eigenvalue solvers. The number of routine represents the eigenvalue solver selection in SEBOMD simulations. ....	112
<b>Table 3.5</b> : The $C - N$ distances ( $\text{\AA}$ ) in transition state structures. The superscript $i$ indicates the quantum mechanical calculations with implicit water. The superscript $e$ indicates the QM/MM calculations with explicit MM water molecules. ....	138
<b>Table 3.6</b> : The activation free energy barriers (kcal/mol) of the located transition states and the relative free energies of the separated molecules. The reference structures for each reaction are the complexes in which the solute molecules are covalently bound. The superscript $s$ is used to denote the total relative free energies of the separated molecules. The superscript $i$ indicates the quantum mechanical calculations with implicit water. The superscript $e$ indicates the QM/MM calculations with explicit MM water molecules. ....	138
<b>Table 3.7</b> : The $C - N$ distances ( $\text{\AA}$ ) in transition state structures and the calculated activation free energy barriers ( $\Delta G^\ddagger$ , kcal/mol) in SEBOMD simulations of C-N bond breaking reactions <b>R01</b> , <b>R02</b> , <b>R03</b> , and <b>R04</b> . ....	139





## LIST OF FIGURES

	<u>Page</u>
<b>Figure R.1</b> : Représentation schématique des deux réactions successives qui apparaissent dans l'enzyme GABA-AT avec le substrat naturel GABA ou des analogues de GABA. ....	xxvii
<b>Figure R.2</b> : Sites actifs de GABA-AT. GABA-AT contient un cluster Fer-Soufre liant les deux monomères ensemble. Chaque site actif contient un PLP lié covalentement avec Lys239. ....	xxix
<b>Figure R.3</b> : Inactivation de GABA-AT par le $\gamma$ -vinyl-GABA (vigabatrine) via la formation d'un complexe enzyme-vigabatrine-PLP. ....	xxx
<b>Figure 1.1</b> : Schematic representation of two successive reactions occurring in PLP-dependent enzyme GABA-AT with the natural substrate GABA and the GABA analogues. ....	3
<b>Figure 1.2</b> : The active sites within GABA-AT in which an Iron-Sulfur cluster holds two monomers together. Each active site involves a PLP that is bound to Lys329. ....	3
<b>Figure 1.3</b> : GABA-AT inactivation with $\gamma$ -vinyl-GABA (vigabatrin) with the formation of enzyme-vigabatrin-PLP complex. ....	5
<b>Figure 2.1</b> : Solvent accessible surface (SAS) and solvent excluded surface (SES). SAS is represented with bold black line, while SES is represented with red line. ....	24
<b>Figure 2.2</b> : Minimum, maximum, and saddle points on a simple energy surface that is defined with $q_1$ and $q_2$ variables. ....	26
<b>Figure 2.3</b> : Potential energy in a molecular mechanics calculation is computed with the contributions from atomic motions. ....	29
<b>Figure 2.4</b> : A two dimensional representation of a simulation box with PBC. The unit cell is represented as the central box with thick black square. The interactions of particle $\chi$ with the images in the neighbour cells are depicted with red arrows. The movement of a particle from the central box, and the movement of its images in replicated boxes are represented with blue arrows. ....	33
<b>Figure 2.5</b> : Schematic representation of a SEBOMD simulation. ....	38
<b>Figure 2.6</b> : The division of a peptide system into subsystems with D&C Algorithm. ....	39
<b>Figure 2.7</b> : The subdivision of a subsystem into core and buffer regions with D&C Algorithm. ....	39
<b>Figure 2.8</b> : The SCF procedure with D&C algorithm to obtain the D&C energy of a single geometry conformation [85]. ....	41
<b>Figure 2.9</b> : A simple call scheme within a SEBOMD simulation using sander program. ....	42

<b>Figure 3.1</b> : Reaction pathway of (A) transamination with GABA and fluorine-substituted GABA analogues. (B) HF elimination from fluorine substituted GABA analogues via E1cb mechanism. ....	44
<b>Figure 3.2</b> : Newman projections of GABA conformations ( <b>00Ga-z</b> , <b>00Gb-z</b> and <b>00Gc-z</b> ), (R)-3-F-GABA ( <b>00Ra-z</b> , <b>00Rb-z</b> and <b>00Rc-z</b> ) and (S)-3-F-GABA ( <b>00Sa-z</b> , <b>00Sb-z</b> and <b>00Sc-z</b> ) in the zwitterionic form. Relative free energies (kcal/mol) are given in parentheses.....	47
<b>Figure 3.3</b> : B3LYP/6-31+G(d,p) free energy profiles of transamination reactions of GABA (pink), (R)-3-F-GABA (purple) and (S)-3-F-GABA (blue).....	48
<b>Figure 3.4</b> : Three-dimensional geometries of the first two steps of the transamination with the bioactive conformers of GABA ( <b>00TS01Gb</b> and <b>01TS02Gb_w</b> , respectively), (R)-3-F-GABA ( <b>00TS04Rb</b> and <b>04TS05Rb_w</b> , respectively) and (S)-3-F-GABA ( <b>00TS07Sb</b> and <b>07TS08Sb_w</b> , respectively). (A) Formation of substrate-PLP-methylamine complex. (B) 1,3 proton transfer between nitrogens of substrate and methylamine. The imaginary frequencies that define transition states are given in parentheses.....	49
<b>Figure 3.5</b> : Single point relative MP2 energies of Transamination Reactions of bioactive conformations and the binding step of solution conformations with MP2 level of theory at 6-311+G(d,p) basis set. Energies are given in kcal/mol.....	50
<b>Figure 3.6</b> : Three-dimensional geometries of transition state structures of the removal of methylamine with GABA ( <b>02TS03Gb</b> ), (R)-3-F-GABA ( <b>05TS06Rb</b> ) and (S)-3-F-GABA ( <b>08TS09Sb</b> ). The imaginary frequencies that define transition state are given in parentheses.....	52
<b>Figure 3.7</b> : B3LYP/6-31+G(d,p) free energy profile of (a). HF elimination from the (R)-3-F-GABA-PLP complex via E1cb (orange) and concerted (blue) mechanisms and (b). HF elimination from the (S)-3-F-GABA-PLP complex via E1cb mechanism with the first scenario (blue) and the second scenario (orange). ....	54
<b>Figure 3.8</b> : Three-dimensional geometries of HF elimination reactions of (R)-3-F-GABA-PLP complex. (A) HF elimination via stepwise E1cb mechanism with the assistance of methylamine. (B) HF elimination via concerted mechanism with the assistance of one water molecule. The imaginary frequencies that define transition states are given in parentheses. ....	55
<b>Figure 3.9</b> : Three-dimensional geometries of stationary structures of fluoride ion elimination reactions via E1cb mechanism of (S)-3-F-GABA-PLP-complex. (A) Abstraction of $\gamma$ proton and formation of structure 12post. (B) First scenario of fluoride ion elimination with the assistance of methylamine and water molecules. (C) Second scenario of fluoride ion elimination in the form of HF with the assistance of methylamine. The imaginary frequencies that define transition states are given in parentheses.....	56

<b>Figure 3.10:</b> Single point relative MP2 energies obtained with MP2 level of theory at 6-311+G(d,p) basis set. (a). Single point relative MP2 energies of HF elimination reactions of (R)-3-F-GABA-PLP complex. (b). Single point relative MP2 energies of HF elimination reactions of (S)-3-F-GABA-PLP complex. Energies are given in kcal/mol.....	58
<b>Figure 3.11:</b> Two dimensional LigPlot <sup>+</sup> representation of the interactions in the active site of vigabatrin bound GABA-AT.....	66
<b>Figure 3.12:</b> RMSd values of apo-GABA-AT samples <b>a.</b> $A^{-1,0}$ and $A^{-1,1}$ and <b>b.</b> $A^{0,0}$ and $A^{0,1}$ along the simulation.....	68
<b>Figure 3.13:</b> pKa of His190 versus pKa of Asp298 along the simulation of apo-GABA-AT samples <b>a.</b> $A^{-1,0}$ , <b>b.</b> $A^{-1,1}$ , <b>c.</b> $A^{0,0}$ and <b>d.</b> $A^{0,1}$ .....	69
<b>Figure 3.14:</b> RMSd values of holo-GABA-AT samples <b>a.</b> $H^{-1,0}$ and $H^{-1,1}$ and <b>b.</b> $H^{0,0}$ and $H^{0,1}$ along the simulation .....	69
<b>Figure 3.15:</b> Interaction between PLP pyridine ring and Asp298 in holo-GABA-AT. <b>a.</b> Distance variations between the side chain of Asp298 and pyridine ring of PLP for $H^{-1,0}$ (red) and $H^{-1,1}$ (blue). <b>b.</b> Typical three dimensional representations of the interaction between Asp298 and PLP in $H^{-1,0}$ (bottom) and $H^{-1,1}$ (top). .....	70
<b>Figure 3.16:</b> The deviation of the distance between the side chain H atom of neutral Asp298 and the N atom on the pyridine ring of PLP along the simulations of the holo-GABA-AT samples $H^{0,0}$ (red) and $H^{0,1}$ (blue). .....	71
<b>Figure 3.17:</b> The hydrogen bond network of Asp298 in holo-GABA-AT. Minimum distance probabilities between Asp298 and, respectively, His190, Asn140 and Val300 residues for <b>a.</b> $H^{-1,0}$ and <b>b.</b> $H^{-1,1}$ . <b>c.</b> Typical three dimensional representation of ionized Asp298, Asn140, neutral His190 and Val300 in $H^{-1,0}$ . <b>d.</b> Typical three dimensional representation of ionized Asp298, Asn140, ionized His190 and Val300 in $H^{-1,1}$ . .....	72
<b>Figure 3.18:</b> Minimum distance probabilities between Asp298 and, respectively, His190, Asn140 and Val300 in samples <b>a.</b> $H^{0,0}$ , and <b>b.</b> $H^{0,1}$ . .....	72
<b>Figure 3.19:</b> The interactions between the side chains of Asp298 and His190 in holo-GABA-AT samples. <b>a.</b> The variations of the minimum distances between $H_{\epsilon}$ of His190 and the carboxyl oxygens of Asp298 (red), and $H_{\delta}$ of Asp298 and $N_{\delta}$ of His190 (blue) in sample $H^{0,0}$ . <b>b.</b> The variations of the minimum distances between the side chain oxygen atoms of Asp298 and $H_{\epsilon}$ (red) and $H_{\delta}$ (green) of His190 in sample $H^{0,1}$ . <b>c.</b> The variations of the distance between the side chain oxygen atoms of Asp298 and $H_{\epsilon}$ (red) of His190 in sample $H^{-1,0}$ . <b>d.</b> The variations of the minimum distances between the side chain oxygen atoms of Asp298 and $H_{\epsilon}$ (red) and $H_{\delta}$ (green) of His190 in sample $H^{-1,1}$ . .....	73
<b>Figure 3.20:</b> Distributions of interaction free energies between PLP and GABA-AT computed for four Holo-GABA-AT trajectories: $H^{0,0}$ , $H^{0,1}$ , $H^{-1,0}$ , and $H^{-1,1}$ . .....	74

<b>Figure 3.21:</b> RMSd values of vigabatrin bound GABA-AT samples <b>a.</b> $V^{-1,0}$ and $V^{-1,1}$ and <b>b.</b> $V^{0,0}$ and $V^{0,1}$ along the simulation.....	75
<b>Figure 3.22:</b> Interactions of Asp298 in $V^{-1,1}$ and $V^{-1,0}$ . <b>a.</b> Distance variations between Asp298 side chain and the pyridine ring of PLP in $V^{-1,1}$ . <b>b.</b> Minimum distance probabilities between Asp298 and, respectively, His190, Asn140 and Val300 in $V^{-1,1}$ . <b>c.</b> The deviation of the minimum distance between the side chain O atoms of Asp298 and the H atom of N atom on the pyridine ring of PLP along the simulations of the vigabatrin bound GABA-AT sample $V^{-1,0}$ . <b>d.</b> Minimum distance probabilities between Asp298 and, respectively, His190, Asn140 and Val300 in sample $V^{-1,0}$ .....	76
<b>Figure 3.23:</b> Interactions of Asp298 in $V^{0,1}$ and $V^{0,0}$ . <b>a.</b> The deviation of the minimum distance between the side chain O atoms of Asp298 and the H atom of N atom on the pyridine ring of PLP along the simulations of the vigabatrin bound GABA-AT sample $V^{0,0}$ (red) and $V^{0,1}$ (blue). <b>b.</b> Minimum distance probabilities between Asp298 and, respectively, His190, Asn140 and Val300 in $V^{0,0}$ . <b>c.</b> Minimum distance probabilities between Asp298 and, respectively, His190, Asn140 and Val300 in sample $V^{0,1}$ .....	78
<b>Figure 3.24:</b> The interactions between the side chains of Asp298 and His190 in vigabatrin bound GABA-AT samples. <b>a.</b> The variations of the minimum distances between $H_\epsilon$ of His190 and the carboxyl oxygens of Asp298 (red), and $H_\delta$ of Asp298 and $N_\delta$ of His190 (blue) in sample $V^{0,0}$ . <b>b.</b> The variations of the minimum distances between the side chain oxygen atoms of Asp298 and $H_\epsilon$ (red) and $H_\delta$ (green) of His190 in sample $V^{0,1}$ . <b>c.</b> The variations of the distance between the side chain oxygen atoms of Asp298 and $H_\epsilon$ (red) of His190 in sample $V^{-1,0}$ . <b>d.</b> The variations of the minimum distances between the side chain oxygen atoms of Asp298 and $H_\epsilon$ (red) and $H_\delta$ (green) of His190 in sample $V^{-1,1}$ .....	79
<b>Figure 3.25:</b> Interactions of the PLP phosphate group in $H^{-1,1}$ and $V^{-1,1}$ . Minimum distance probabilities between the phosphate group oxygen atoms and, respectively, Gly136, Ser137 and Thr353* in <b>a.</b> holo-GABA-AT $H^{-1,1}$ and <b>b.</b> vigabatrin bound GABA-AT $V^{-1,1}$ . Typical three dimensional representation of the phosphate group oxygen atoms, Thr353* residue and the surrounding water molecules <b>c.</b> in $H^{-1,1}$ and <b>d.</b> in $V^{-1,1}$ .....	80
<b>Figure 3.26:</b> Interactions between the carbonyl oxygen atom on the pyridine ring and the Gln301 side chain. <b>a.</b> Variations of the minimum distances (left) and probability distributions (right) between the hydrogen atoms on the side chain of Gln301 and carbonyl oxygen on pyridine ring in $H^{-1,1}$ and $V^{-1,1}$ . <b>b.</b> Typical three dimensional representation of the interaction between PLP pyridine ring and the side chain of Gln301 in $H^{-1,1}$ and $V^{-1,1}$ . ....	80

<b>Figure 3.27:</b> Interaction between the pyridine ring of PLP and the phenyl ring of Phe189. <b>a.</b> Definitions of the distance $d_1$ , of the angle $\alpha$ , and of the dihedrals $\varphi$ , $\chi$ . <b>b.</b> Typical three dimensional representation of the interaction between the pyridine ring of PLP and phenyl ring of Phe189. <b>c.</b> Variation of the interaction in holo-GABA-AT ( $H^{-1,1}$ ). <b>d.</b> Variation of the interaction in vigabatrin bound GABA-AT ( $V^{-1,1}$ ).....	82
<b>Figure 3.28:</b> Interaction between the carboxyl group of vigabatrin and guanidine group of Arg192 in vigabatrin bound GABA-AT. <b>a.</b> Variations of the minimum distances between the oxygens of the carboxyl group of vigabatrin and the nitrogen atoms of the guanidine group along the simulations. <b>b.</b> Typical three dimensional representation of the interaction of vigabatrin with water molecules and Arg192.....	82
<b>Figure 3.29:</b> The variations of the minimum distances between the side chain hydrogen atoms of Arg192 and backbone oxygen atom of Tyr348* in sample $H^{-1,1}$ .....	83
<b>Figure 3.30:</b> Interaction between the side chains of Arg445 and Glu270 in holo-GABA-AT and vigabatrin bound GABA-AT. <b>a.</b> Variations of the minimum distances between the heavy atoms on the side chains along the simulations. <b>b.</b> Typical three dimensional representation of the salt bridge that is formed between Arg445 and Glu270.....	84
<b>Figure 3.31:</b> The computed B-factor values in simulated GABA-AT samples. <b>a.</b> Apo-GABA-AT samples $A^{0,0}$ and $A^{0,1}$ . <b>b.</b> Apo-GABA-AT samples $A^{-1,0}$ and $A^{-1,1}$ . <b>c.</b> Holo-GABA-AT samples $H^{0,0}$ and $H^{0,1}$ . <b>d.</b> Holo-GABA-AT samples $H^{-1,0}$ and $H^{-1,1}$ . <b>e.</b> Vigabatrin bound GABA-AT samples $V^{0,0}$ and $V^{0,1}$ . <b>f.</b> Vigabatrin bound GABA-AT samples $V^{-1,0}$ and $V^{-1,1}$ . The secondary structures are represented with different colors; the $\beta$ -sheets are colored with yellow and the $\alpha$ -helices are colored with red.....	85
<b>Figure 3.32:</b> The solvent access to the active site. The results obtained with GIST calculations <b>a.</b> on Monomer A of sample $H^{-1,1}$ . <b>b.</b> on Monomer B of sample $H^{-1,1}$ . <b>c.</b> on Monomer A of sample $V^{-1,1}$ . <b>d.</b> on Monomer B of sample $V^{-1,1}$ . The grid increments along each coordinate axis is 40x40x40, grid spacing is 0.5Å and center of the grid is chosen as imino nitrogen atom. ....	90
<b>Figure 3.33:</b> Two dimensional representations of model clusters of increasing size. Different protonation states of Asp298 and PLP in M4 and M4' are given on the right hand side in colored squares. ....	94
<b>Figure 3.34:</b> The Michael addition reaction in GABA-AT with $\gamma$ -vinyl GABA. <b>(a)</b> Two dimensional representation of reaction. <b>(b)</b> Three dimensional representation of Michael addition reaction with model <b>M1</b> . Frozen atoms are shown with black angle headed arrows. ....	96
<b>Figure 3.35:</b> Activation energy barriers of model clusters. The gas phase energies (red) are calculated with 6-311G(2d,2p) basis set while the solvent energies with $\epsilon = 4$ (blue) and $\epsilon = 80$ (green) are calculated with 6-31G(d,p) basis set.....	97

<b>Figure 3.36:</b> The Michael addition reaction with model cluster <b>M2</b> . Three dimensional representation of (a) phosphate group interactions, and (b) transition state center. Frozen atoms are shown with black angle headed arrows.....	98
<b>Figure 3.37:</b> The Michael addition reaction with model clusters with the involvement of Asp298. Three dimensional representation of the interaction between Asp298 side chain and pyridine ring N atom of PLP in (a) transition state of <b>M3</b> . (b) transition state of <b>M3'</b> , (c) reactant of <b>M3</b> and (d) reactant of <b>M3'</b> . Frozen atoms are shown with black angle headed arrows.....	101
<b>Figure 3.38:</b> The Michael addition reaction with model cluster <b>M4</b> . Three dimensional representation of the interaction between the pyridine ring N-Asp298-His190 triad in (a) reactant, and (b) transition state. Frozen atoms are shown with black angle headed arrows. ....	102
<b>Figure 3.39:</b> Geometrical parameters of six-membered MA transition state structures. (a) Two dimensional representation and labels of distances a-f, dihedrals ( $\alpha$ : H3-N2-H7-O8, $\beta$ : H6-C5-H9-O8, $\theta$ : C1-N2-C4-C5, $\eta$ : C14-C15-C5-C4, $\nu$ : C13-C14-C15-N16) and angles ( $\omega$ , $\phi$ , $\chi$ , $\delta$ , $\lambda$ , $\rho$ ). (b) The distribution of angles. (c) The distribution of distances. (d) The distribution of dihedrals. ....	103
<b>Figure 3.40:</b> The three dimensional representations of the fitted models <b>M1</b> (C atoms colored by cyan) and <b>M2</b> (C atoms colored by pink). (a) The alteration of the dihedral N16-C15-C5-C4 (red arrow) and its value (colored with red). (b) The dihedral of C5-H9-O8-H10 in <b>M1</b> (blue) and in <b>M2</b> (red). The frozen atoms are shown with black arrows.....	104
<b>Figure 3.41:</b> The three dimensional representation of the model <b>M2</b> (C atoms colored by pink) and the model <b>M3</b> (C atoms colored by magenta). The backbone C atoms of Ly329, Gly136, Ser137, Asp298, Thr353*, and the N atom on PLP of <b>M3</b> is fitted to the corresponding atoms of <b>M2</b> .....	104
<b>Figure 3.42:</b> The three dimensional representation of the interaction between Arg192 and the carboxyl group of carboxyl group of $\gamma$ -vinyl-GABA in model <b>M3</b> (C atoms colored by magenta) and in model <b>M4</b> (C atoms colored by yellow). The backbone C atoms of Ly329, Gly136, Ser137, Asp298, Thr353*, and the N atom on PLP of <b>M4</b> is fitted to the corresponding atoms of <b>M4</b> . The frozen atoms are depicted with black arrows. ....	105
<b>Figure 3.43:</b> Parameters that would effect consumed time by a SEBOMD simulation.....	110

- Figure 3.44:** Formation of BLACS grids with local\_masters that are defined by nthread. **a.** NPROCS=8 with nthread=1, number of grids=8, and number of local\_masters=8. **b.** NPROCS=8 with nthread=2, number of grids=4, and number of local\_masters=4. **c.** NPROCS=8 with nthread=4, number of grids=2, and number of local\_masters=2. **d.** NPROCS=8 with nthread=8, number of grids=1, and number of local\_masters=1. local\_masters are represented with purple color, and each grid is represented within black rectangles. .... 113
- Figure 3.45:** Execution times of the SCF routine, other routines and gradient routine as the percentage of the total SEBOMD simulation time with **a.** Standard Closed Shell algorithm (m0). and **b.** D&C algorithm (m2). The percentages are colored versus the cumulative number of simulation steps. SCF (bottom) and gradient routines (top) are depicted with dense colors while the other routines are depicted with light colors..... 117
- Figure 3.46:** A representation of the call scheme and number of calls in SEBOMD simulations. The number of calls for 64 and 216 water molecules are obtained with the condensed phase (n1) simulations, using full diagonalization (f1) and disregarding the long range interactions (lr0). The results obtained using D&C algorithm are indicated with blue boxes, and the results obtained using SCS algorithm are shown with red boxes. The densub subroutine is called only when D&C algorithm is used and denful subroutine is called only when SCS algorithm is used..... 119
- Figure 3.47:** Conservaion of total energy ( $E_{tot}$ ) in terms of deviation from average energy along 500 steps of SEBOMD simulation of 64 water molecules with standard closed shell algorithm using implemented eigenvalue solvers. .... 120
- Figure 3.48:** Execution times of eigenvalue solvers in one SCF iteration performed in SEBOMD simulations of different sizes of systems using the standard closed shell algorithm. Each eigenvalue solver is represented with different color. Execution times are calculated for both the integrated LAPACK library (left) and the Intel MKL (right). The Gnu compiler is used to compile the SEBOMD code in the case of the integrated LAPACK library while the Intel compiler is used instead of the Gnu one in the case of the Intel MKL. .... 121
- Figure 3.49:** Execution times of eigenvalue solvers for one SCF iteration performed in SEBOMD simulations of different sizes of systems using the D&C algorithm. Each eigenvalue solver is represented by different color. Execution times are calculated for both the integrated LAPACK library (left) and the Intel MKL (right). The Gnu compiler is used to compile the SEBOMD code in the case of the integrated LAPACK library while the Intel compiler is used instead of the Gnu one in the case of the Intel MKL. .... 122



- Figure 3.50:** The execution time of the eigenvalue solvers per SCF iteration with the automatic eigenvalue solver selection option. In the scanning section (first eight iterations), different solvers are used for each iteration. The different eigenvalue solvers are represented by different colors of points and the fastest solver (*dsyevr*) is depicted with an arrow. .... 123
- Figure 3.51:** Elapsed times obtained with SEBOMD simulations with standard closed shell algorithm using different eigenvalue solvers from SCALAPACK. Different number of processors are represented with different colors. Elapsed time obtained with the serial simulation using *dsyev* in Amber15 is shown with dashed line..... 125
- Figure 3.52:** Elapsed times obtained from SEBOMD simulations of 216 water molecules using the D&C algorithm with **a.** buffer=6Å on 1 local\_master. **b.** buffer=6Å on 2 local\_masters. Eigenvalue solvers are depicted with "?" as first letter to indicate both LAPACK and SCALAPACK versions. In SCALAPACK versions of the solvers, "?" indicates "p". Reference elapsed time obtained with the *dsyev* solver in Amber15 is shown with dashed line while elapsed times with LAPACK versions of the solvers are shown with black columns. Number of used processors with SCALAPACK routines are depicted with red (1 processor), blue (2 processor), green (4 processors), and yellow (8 processors) columns. .... 126
- Figure 3.53:** Elapsed times obtained from SEBOMD simulations of 216 water molecules using the D&C algorithm with **a.** buffer=8Å on 1 local\_master. **b.** buffer=8Å on 2 local\_masters. Eigenvalue solvers are depicted with "?" as first letter to indicate both LAPACK and SCALAPACK versions. In SCALAPACK versions of the solvers, "?" indicates "p". Reference elapsed time obtained with the *dsyev* solver in Amber15 is shown with dashed line while elapsed times with LAPACK versions of the solvers are shown with black columns. Number of used processors with SCALAPACK routines are depicted with red (1 processor), blue (2 processor), green (4 processors), and yellow (8 processors) columns. .... 128
- Figure 3.54:** Elapsed times obtained with 10 steps of SEBOMD simulations of 216 water molecules with buff=8Å and using implemented solvers with Intel MKL and threads. Reference elapsed time obtained with the *dsyev* solver in Amber15 is shown with dashed line while elapsed times without threading are shown with black columns. Number of threads that are used are represented with red (1 thread), blue (2 threads), green (4 threads), and yellow (8 threads) columns. .... 129

- Figure 3.55:** Elapses times obtained with 10 steps of SEBOMD simulations of solvated box with protein using buff=8Å and threaded the dsyevd solver in Intel MKL. Elapsed times using 1 thread are represented with solid columns, while 2 threads are represented with pattern filled columns. Different number of nodes are represented with red (1 node), blue (2 nodes), green (4 nodes), and yellow (8 nodes) columns. .... 131
- Figure 3.56:** Two dimensional representations of the reactions that are chosen for SEBOMD simulations. (a) First step of external aldimine formation. (b) The C-N bond breaking within the  $\gamma$ -vinyl-GABA-PLP-methylamine complex. (c) The attack of methylamine on the PLP imine carbon. (d) The formation of PLP-methylamine complex within the proposed Enamine pathway. The reactants are denoted with superscript *R*, the separated products are denoted with superscript *S*, and the modelled bond breaking reactions are depicted with red arrows. .... 134
- Figure 3.57:** The PMF of C-N bond breaking reactions. (a)R01. (b)R02. (c)R03. (d)R04. The results obtained with SEBOMD simulations are depicted with red line, and the results obtained with QM/MM calculations are depicted with green dashed line. The complex between the products is depicted with a black arrow. Continued on next page. .... 141
- Figure 3.57:** – continued from previous page. The PMF of C-N bond breaking reactions. (a)R01. (b)R02. (c)R03. (d)R04. The results obtained with SEBOMD simulations are depicted with red line, and the results obtained with QM/MM calculations are depicted with green dashed line. The complex between the products is depicted with a black arrow. .... 142
- Figure 3.58:** Three dimensional representations of different conformers of vigabatrin, PLP, and methylamine complexes that are located with the SEBOMD simulations. a. Complex conformers that are formed between vigabatrin and PLP-methylamine imine in R01. b. Complex conformers that are formed between vigabatrin-PLP imine and methylamine in R02. .... 143



**MOLECULAR MODELLING OF GABA-AT REACTIVITY:  
FROM SMALL REPRESENTATIVE MODELS TO THE FULL PROTEIN,  
FROM MOLECULAR MECHANICS TO QUANTUM CHEMISTRY,  
FROM STATIC TO DYNAMICS**

**SUMMARY**

Understanding enzymes and their catalytic mechanisms is very important in order to develop more effective drugs having little to no side effects. The active site residues of an enzyme may only be responsible for the facilitation of the enzymatic reaction or may be responsible for holding the substrates in place or both. In order to decipher the catalytic behaviour of enzymes, different approaches such as QM, MM-MD, and QM/MM can be used and their results can be correlated.

The quantum chemical approach is often the only reliable way to model chemical transformations. Unfortunately, the numerical solution of the Schrödinger equation becomes increasingly expensive in terms of computational power when size increases. Numerical procedures in quantum chemistry require the calculation of numerous integrals and the resolution of (non)-linear equations, thus limiting the size of the systems that can be modelled. As of today, only linear scaling approximate quantum chemical methods are capable of combining both the necessary speed and accuracy to model large biomolecular systems with reasonable CPU cost.

SEBOMD (SemiEmpirical Born-Oppenheimer Molecular Dynamics) is a method which is based on the Born-Oppenheimer approximation and uses a QM semiempirical NDDO Hamiltonian to perform molecular dynamics. It uses the linear scaling Divide & Conquer (D&C) algorithm to perform MD simulations of molecular systems. Today, the current D&C implementation is not the most efficient to run on very large supercomputers due to the diagonalization process of the Fock matrix. Thus, the diagonalization procedure should be optimized to be able to run MD simulations of molecular systems containing several hundreds of atoms.

The main aim of this thesis is to get a deeper understanding of the mechanistic insights of the reactivity and of the dynamics of the pyridoxal 5-phosphate (PLP) dependent

enzyme  $\gamma$ -aminobutyric acid aminotransferase (GABA-AT). GABA-AT catalyzes the reaction in which GABA is degraded to succinic semialdehyde. Reduction of GABA concentration in the brain is at the origin of many neurological diseases, and since GABA cannot cross the blood brain barrier, it cannot be directly administered. As a consequence, the inactivation of GABA-AT is considered as a valid target for the development of antiepileptic drugs. Because GABA-AT resembles many other PLP-dependent enzymes, understanding it could be of importance for the broad community of biochemists and computational chemists who study such class of proteins. Our work has consisted of five stages to pursue the comprehension of GABA-AT, and by extension that of PLP dependent enzymes.

First, the reaction and the preferred binding mode of the natural substrate GABA has been elucidated with different isomers by means of model systems with Density Functional Theory (DFT). The stereoelectronic differences between the studied structures enabled us to discuss the possible particularities of the transamination and HF elimination reactions in GABA-AT.

Second, the dynamics and the behavior of the enzyme has been studied with MM-MD through the use of apoenzyme, holoenzyme and holoenzyme with an inactivator. Our results demonstrate that several active site residues play a key role, and their protonation states as well as that of PLP require special attention.

Third, the effect of the active site residues in the inactivation mechanism has been investigated with the modeling of clusters at the QM level involving key residues. The biggest cluster has been modelled with 165 atoms and the effect of the solvent has been taken into account using implicit solvent. The inclusion of the active site residues within the models enabled us to emphasize the effects of these residues on the reactivity of GABA-AT. The presence of similar residues in many PLP dependent enzymes enabled us to provide some general aspects for modelling these enzymes.

Fourth, new diagonalization routines for the SEBOMD approach implemented in the Amber suite of programs, have been incorporated using LAPACK and SCALAPACK libraries, tested and evaluated. Our results showed that the main restriction on the improvement of the D&C algorithm is the diagonalization of the Fock matrix which is not efficient on parallel computers. However, when large number of cores were used in

SEBOMD simulations of large systems, using LAPACK solvers with threads that are provided by the Intel MKL did improve the efficiency of the SEBOMD calculations.

Fifth, reaction free energies of PLP containing systems have been investigated with SEBOMD simulations. Each molecular system contains 747 atoms and their reactivity has been simulated using umbrella samplings and the newly implemented diagonalization routines. This work is the first example of the SEBOMD simulations on chemical transformations, and provides some general insights on the capabilities of the SEBOMD methodology in modelling reactions and in computing reaction free energies.



**MODÉLISATION MOLÉCULAIRE DE LA RÉACTIVITÉ DE GABA-AT:  
DE PETITS MODÈLES REPRÉSENTATIFS À LA PROTÉINE COMPLÈTE,  
DE LA MÉCANIQUE MOLÉCULAIRE À LA CHIMIE QUANTIQUE,  
DU STATIQUE AU DYNAMIQUE**

**RÉSUMÉ**

PLP, la forme oxydée et phosphorylée de la vitamine B6, est un cofacteur organique qui forme une imine avec le groupe  $\epsilon$ -aminé d'une lysine dans les enzymes dépendantes au PLP [1]. La formation d'une imine est appelée formation d'une base de Schiff [2]. Cette espèce d'intermédiaire apparaît dans de nombreuses réactions catalysées par les enzymes et mène à une large gamme de produits. Il existe un intérêt particulier pour les enzymes dépendantes au PLP à cause de la grande diversité des réactions qu'elles catalysent comme la décarboxylation, l'élimination  $\beta$ , la coupure aldole, la transimination ou la transamination [1,3,4].

Il a été suggéré que l'azote du cycle pyridine de PLP est protoné dans les enzymes dépendantes au PLP. Un tel état acide pourrait causer un puit à électron qui aurait pour conséquence de faciliter les réactions enzymatiques [3, 5–10]. L'état de protonation du cycle pyridine de PLP en solution aqueuse a été exploré en utilisant de la RMN  $^{15}\text{N}$  sur des pyridines de PLP marqués au  $^{15}\text{N}$  en solution aqueuse à température ambiante [11]. Dans cette étude, Sharif et al. a reporté une valeur de pKa pour le cycle pyrdine de 5.8 [11] ce qui indique un cycle pyridine dans PLP qui serait largement non protoné dans les conditions physiologiques. Dans un autre travail de Sharif et al., des études à l'état liquide en  $^{15}\text{N}$  ont été conduites sur l'Aspartate Aminotransferase (AspAT) et sur des composés modèles [12]. Il a été alors reporté que les déplacements chimiques  $^{15}\text{N}$  pour le cycle pyridine de PLP sont consistants avec la présence d'un proton sur ce cycle. Une interaction de proton similaire a pu être reproduite sur un système modèle composé d'un complexe 1:2 entre une aldimine, représentative de Lys-PLP, et un acide aspartique dans un mélange de solvant d'une constante diélectrique  $\epsilon_r \approx 30$ . De plus, Limbach et al. ont reporté dans leur



étude par RMN  $^{15}\text{N}$  sur AspAT marqué à l'azote dans l'eau qu'à pH 7.5, la distance  $N_{\text{pyridine}} - H$  est courte alors que la distance  $H - O^{\text{Asp222}}$  est longue, ce qui indique un caractère zwitterionique de la liaison hydrogène OHN avec une forme protonée pour le cycle pyridine dans des conditions physiologiques [13]. Il a été proposé aussi que AspAT pourrait fournir des interactions supplémentaires, tels que Asp222, His143 et 2 molécules d'eau conservées, pour obtenir le transfert de proton vers le cycle pyridine de PLP [13].

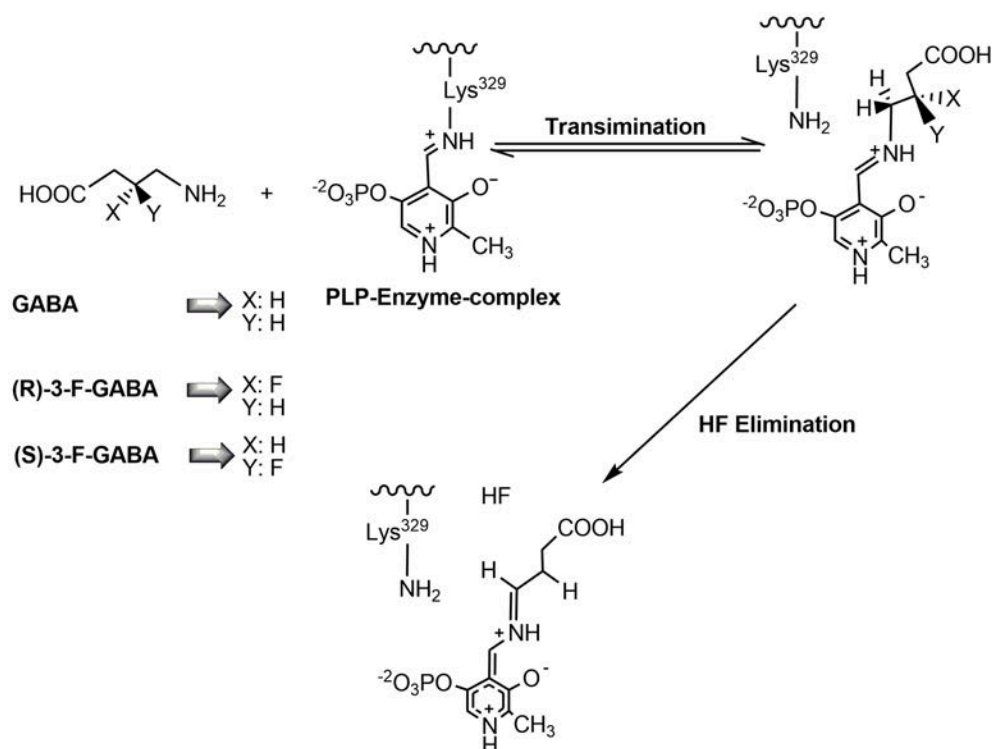
Le mécanisme catalytique des enzymes dépendantes au PLP est habituellement divisé en deux parties: une première et une seconde demi-réaction [14, 15]. La première demi-réaction implique la conversion de PLP en phosphate pyridoxamine (PMP) [14, 16]. La molécule de PMP dans le site actif est ensuite liée à l'enzyme à travers une seconde demi-réaction [14, 16, 17].

Dans les enzymes dépendantes au PLP, la quantité de recouvrement  $\sigma - \pi$  dans les cofacteurs dérivés des bases de Schiff détermine la spécificité des réactions enzymatiques engendrées, ce qui est en accord avec le modèle stéréochimique de Dunathans [8, 18–24]. Les bases de Schiff dans les enzymes dépendantes au PLP sont formées via une réaction de transimination qui implique une aldimine interne pour laquelle le PLP et le groupe  $\varepsilon$ -aminé du site active sont liés de manière covalente [1, 25]. Dureant la réaction, la liaison covalente entre PLP et l'enzyme est cassée et une nouvelle base de Schiff, nommée l'aldimine externe, est formé avec le substrat [1, 26, 27].

Après la formation de l'aldimine externe, les réactions chimiques possèdent une grande diversité de part la spécificité de l'enzyme mais aussi la nature du substrat. Dans le case des substrats fluorés, comme les fluoroamines  $\beta$ , les réactions de transimination sont suivies par l'élimination d'un ion fluor sous la forme HF (Figure R.1) [28–32]. Les études par chimie théorique sur les enzymes dépendantes au PLP sont principalement concentrées sur la formationm de l'aldimine externe [2–4, 25, 26, 33, 34]. Quelques études existent cependant sur les réactions d'élimination du HF dans les systèmes impliquant un cycle pyridine [33, 35].

L'acide  $\gamma$ -aminobutyrique (GABA) est un des neurotransmetteurs inhibiteurs du système nerveux central chez les mammifères. Il est dégradé en succinique semialdéhyde

via une réaction catalysée par la  $\gamma$ -aminobutyrique acide aminotransférase (GABA-AT, E.C. 2.6.1.19) [36]. GABA-AT, un membre de la sous-famille II des enzymes dépendantes au PLP de la famille  $\alpha$ , est un homo-dimère dans lequel toutes les sous-unités contiennent un site actif au PLP qui est lié de manière covalente sur une lysine 239 via une base de Schiff (Figure R.2).



**Figure R.1** : Représentation schématique des deux réactions successives qui apparaissent dans l'enzyme GABA-AT avec le substrat naturel GABA ou des analogues de GABA.

La réduction de concentration en GABA dans le cerveau entraîne des convulsions, ainsi que diverses maladies neurologiques comme l'épilepsie, la maladie de Parkinson, la chorée de Huntington, ou la maladie d'Alzheimer [29, 32]. L'augmentation de la concentration en GABA dans le cerveau a un effet anticonvulsant [37, 38]. Cependant, GABA ne peut pas traverser la barrière hémato-encéphalique, ainsi le contrôle direct du niveau de concentration en GABA ne peut pas être réalisée de manière effective [37, 39]. Par conséquent, plusieurs analogues de GABA ont été suggérés et testés pour contrôler le niveau de GABA dans le cerveau soit par inhibition, soit par inactivation de GABA-AT [39–42].

La séquence primaire de GABA-AT a été établie à partir de ADNc de cerveau de porc et de fragments peptidiques d'enzyme de foie de porc [43, 44]. La structure

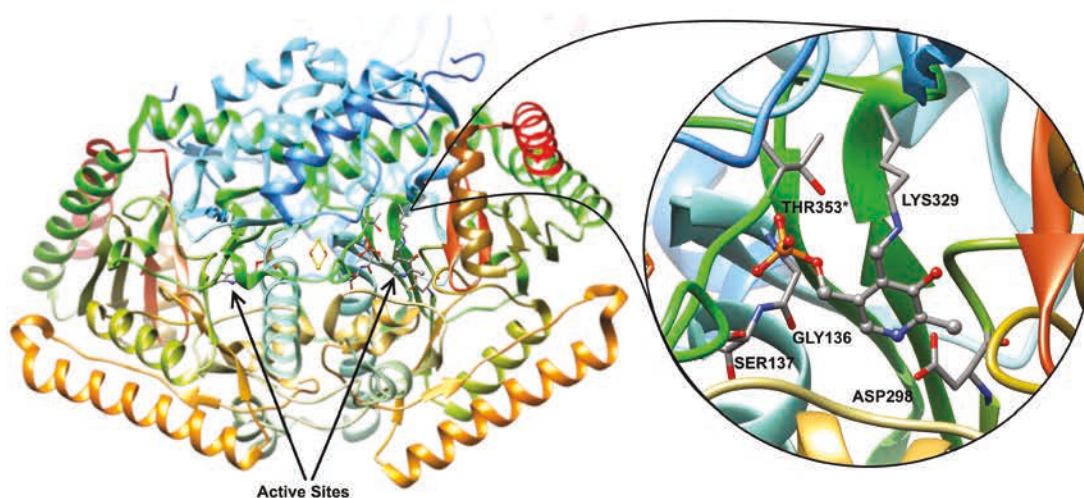
cristallographique de GABA-AT de foie de porc a été publié pour la première fois à 3 Å de résolution, pH 5.6 et à 293 K (code PDB 1GTX) [45]. La qualité de la structure native de GABA-AT a été améliorée et reportée à 2.3 Å de résolution, pH 5.7 et à 100K (code PDB 1OHV) [46]. Dans cette structure cristallographique, le cofacteur PLP est tenu en place non seulement par Gly136 et Ser137 mais aussi par le Thr353 de la seconde sous-unité via la donation de 5 liaisons hydrogènes vers le groupe phosphate [46]. Le cycle pyridine de PLP est pris en sandwich entre Phe189 et Val300 tandis que l'azote du cycle pyridine forme un pont salin avec Asp298 [46]. La mutation du résidu W101 qui interagit avec le cycle PLP de la phosphoserine aminotransferase (EhPSAT) d'*E. histolytica* en alanine montre la perte de l'activité enzymatique alors que la mutation W101F montre que la phenylalanine peut efficacement interagir avec le cycle de PLP [47]. De plus, dans l'Aspartate aminotransferase (AspAT), une autre enzyme dépendante au PLP, l'efficacité catalytique de l'enzyme est réduite respectivement d'environ 8600 fois et 20000 fois dans le cas des mutations D222A et D222N [48]. Egalement, il a été rapporté que ces mutations augmentent l'effet isotopique avec le deutérium, ce qui indique que la charge négative de Asp222 stabilise la protonation de l'azote du cycle PLP qui fonctionne alors comme un attracteur à électron durant la réaction enzymatique [48, 49]. D'un autre côté, la mutation de l'Asp279 qui interagit avec le cycle PLP en acide glutamique dans la 5-Aminolevulinate synthase (ALAS) révèle que le positionnement spatial de la chaîne latérale chargée négativement a aussi de l'importance dans l'activité de l'enzyme [50].

En plus de la structure native de GABA-AT, la structure cristallographique de GABA-AT inactivée par le  $\gamma$ -vinyl GABA (vigabatrine) a été reportée à 2.3Å de résolution, pH 5.7 et une température de 100K (code PDB 1OHV) [46]. Dans cette structure, le résidu Arg192 forme un pont salin courbé avec le groupe carboxylate de la vigabatrine [45, 46]. De plus, il a été observé que Arg445 est complètement conservé dans la famille  $\alpha$  des aminotransférases [51] avec sa charge positive qui est écranté par un pont salin avec Glu270 dans GABA-AT [45]. Il a été suggéré également que, pour obtenir un second point d'ancrage basique, les substrates dicarboxyliques auraient besoin de séparer la paire Glu270-Arg445 [45, 52].

Silverman et al. ont reporté quelques analogues fluorés de GABA qui sont conformationnellement restraints comme inactivateurs ou inhibiteurs possibles [28,

32]. Les études théoriques et expérimentales ont apportées quelques indices concernant la nature du mécanisme catalytique [28,33].

Dans un travail expérimental, Clift et al. ont proposé des énantiomères fluorés de GABA comme inhibiteurs potentiels des enzymes dépendantes au PLP comme GABA-AT [28] pour lesquels GABA (Figure R.1) est un substrat naturel. Ces analogues fluorés (3-F-GABA, Figure R.1) pourrait subir une élimination HF après la formation d'une aldimine externe. Les réactions d'élimination du HF pourrait alors procéder via des mécanismes d'élimination unimoléculaire par étape ou par une élimination bimoléculaire concerté [28].

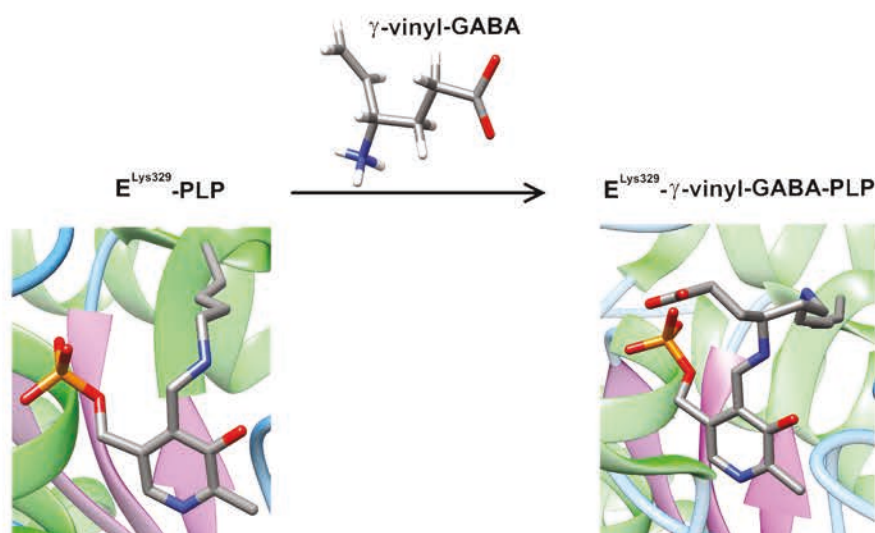


**Figure R.2 :** Sites actifs de GABA-AT. GABA-AT contient un cluster Fer-Soufre liant les deux monomères ensemble. Chaque site actif contient un PLP lié covalentement avec Lys239.

Vigabatrine, l'un des analogues naturelles du substrat GABA, a été reporté comme l'un des inactivateur les plus efficace pour GABA-AT avec un pouvoir élevé pour le traitement de l'épilepsie. Elle est utilisé dans plus de 60 pays [36, 46, 53]. Les mécanismes d'inactivation par la vigabatrine, qui se terminent par la formation d'un complexe enzyme-vigabatrine-PLP (Figure R.3), ont été étudiés à la fois expérimentalement [46, 54–57] et théoriquement [34], et le chemin d'addition de Michael a été indiqué comme le mécanisme préféré.

Les résultats expérimentaux et théoriques nous ont poussé à étudier les facteurs qui pourraient influencer la réactivité de GABA-AT:





**Figure R.3** : Inactivation de GABA-AT par le  $\gamma$ -vinyl-GABA (vigabatrine) via la formation d'un complexe enzyme-vigabatrine-PLP.

- Est-ce que le mode d'association préféré du substrat est une preuve de l'inhibition sélective ou est-ce que la suite de réactions successives est la cause principale de la sélectivité?
- Quels sont les effets des interactions électrostatiques et des différences stéréoelectroniques sur les réactions d'inhibition de GABA-AT?
- Quels sont les états de protonation des résidus du site actif, et quel est l'effet des états de protonation sur les interactions dans le site actif?
- Quel serait le comportement de l'enzyme lorsqu'un inactivateur est impliqué?
- Quels sont les effets des résidus entourant sur les paramètres géométriques des structures d'état de transition qui sont obtenues dans les études de modélisation?
- Quels résidus devraient être introduits dans des études de modélisation de manière à obtenir un modèle qui fournit une ressemblance optimale du site actif de GABA-AT?
- Est-ce que les effets dynamiques sont importants pour la réactivité?

Pour répondre à toutes ces questions, une approche pas à pas avec une complexité croissante a été choisie dans cette thèse.

D'abord, la réactivité dans GABA-AT a été évaluée avec des systèmes modèles, de petits modèles à de plus complexes, en passant par l'exploration des surfaces d'énergie

potentielle avec des outils de la mécanique quantique. Ensuite, la dynamique du système a été étudiée en mécanique moléculaire pour décrire la protéine entière. Enfin, la mécanique quantique et la dynamique moléculaire sont rassemblées pour apporter une plus grande compréhension de la réactivité des complexes au PLP.

Au total, le chapitre "Résultats" de cette thèse est divisé en cinq sections qui recouvrent notre travail sur la réactivité de GABA-AT. Premièrement, des petits modèles représentatifs ont été utilisés pour étudier deux réactions successives de GABA et de composés fluorés dérivés de GABA en mécanique quantique. Ensuite, des simulations de dynamique moléculaire ont été effectuées sur l'apoenzyme, l'holoenzyme, et la forme liée à la vigabatrine de GABA-AT pour évaluer les états de protonation et le rôle des résidus du site actif. Puis, en utilisant les informations fournies par les simulations de dynamique moléculaire, des calculs de chimie quantique ont été effectués sur des modèles de plus en plus grands en faisant intervenir les résidues entourant le site actif afin d'obtenir un modèle le meilleur possible du site actif de GABA-AT. Finalement, quelques réactions de cassure de liaisons C-N qui interviennent dans les mécanismes d'inactivation de GABA-AT ont été modélisées en utilisant une représentation explicitement quantique des molécules d'eau de solvant en utilisant la méthodologie SEBOMD. Avant ces simulations SEBOMD, de nouvelles routines de diagonalisation ont été implémentées pour l'approche SEBOMD dans le logiciel AMBER afin d'améliorer leur efficacité sur des ordinateurs parallèles.

### **Détails mécanistiques sur les réactions de transamination et d'élimination HF**

Les fluoroamines  $\beta$  sont couramment utilisés comme analogues de substrat pour déterminer les détails mécanistiques de réactions enzymatiques. Présence d'un atome de fluor donne lieu à des modifications dans le profil électronique et le pKa de molécules qui résultent en écarts mécanistiques. Les analogues de substrats basés sur un mécanisme substitués par fluor sont largement utilisés dans l'inactivation du pyridoxal 5'-phosphate (PLP) enzymes dépendantes. La présence de fluor atome modifie également la séquence de réactions qui ont lieu dans les enzymes PLP-dépendante où l'élimination de HF la réaction de la nation apparaît entre les réactions de transamination et d'inactivation [28]. En dépit de la quantité de les travaux sur fluoroamines  $\beta$ , l'effet des différences stéréoelectroniques sur le transamination et

l'élimination HF qui se déroulent dans les enzymes PLP-dépendante n'a pas encore été étudiés. Ici une densité étude de la théorie fonctionnelle est réalisée avec une enzyme PLP-dépendante GABA-AT pour comprendre l'effet de l'atome de fluor, de 3-amino-4-fluorobutanoïque d'acide (3-F-énantiomères GABA) sont choisis pour être étudiés en plus de l'acide 4-aminobutanoïque (GABA), qui est le substrat naturel pour aminotrans- d'acide  $\gamma$ -aminobutyrique ferase (GABA-AT). Les -fluoroamines  $\beta$  étudiés sont les inhibiteurs potentiels expérimentalement proposés d'enzyme PLP-dépendante du GABA-A.

Dans une étude expérimentale et l'amarrage combinée réalisée par Clift et al., L'élimination du HF et du Les réactions de transamination ont été étudiés sur la base des affinités de liaison et de l'efficacité relative des GABAA et ses analogues fluorés en présence d'une enzyme PLP-dépendante du GABA-AT [28]. Il a été déclaré cette solution et bioactifs conformères de GABA et ses analogues fluorés ne possèdent pas le même la conformation [28]. Ainsi, les calculs sont effectués à la fois pour la solution et conformations bioactives. Suivant le analyse conformationnelle des espèces zwitterioniques, la réaction de transamination est modélisé avec la solution et conformations bioactives pour discuter de l'efficacité de liaison des analogues du GABA. Ensuite, le HF élimination est modélisé avec les produits de la réaction de transamination pour étudier l'effet de la les interactions électrostatiques et les différences stéréoélectroniques sur le trajet de réaction. Le stéréoélectronique les différences entre les structures nous ont permis de discuter des dépendances possibles des réactions étudiées.

La dynamique des réactions enzymatiques dépendent non seulement des barrières d'énergie d'activation de la structures étatiques de transition, mais aussi sur les préférences conformationnelles et stéréoélectroniques du pondant réactions che correspondante, par ainsi que l'efficacité de liaison des substrats. Avant les mécanismes de réaction, la stabilité de la solution et bioactives conformères est étudiée et on a constaté que la con- bioactif agents des deux analogues du GABA substitués ne sont pas les structures mondiales minimales. Par conséquent, le rapport Population de conformère biologiquement actif de l'énantiomère (S) sera inférieure à celle de l'énantiomère (R). Les résultats indiquent que la faible efficacité de l'énantiomère (S) peuvent provenir de l'instabilité de la cor- répondant conformère biologiquement actif.

Les conformères bioactifs des fluor substitués analogues ne sont pas les structures minimales mondiales à cet égard, l'exigence des concentrations élevées de fluor substitué analogues à inhiber GABA-AT peuvent être expliquées.

Les rendements de liaison des différents énantiomères a également été étudiée pour de plus amples explications. le dynamique de l'étape de liaison, qui est la première étape dans la réaction de transimination, est analysée pour expliquer l'efficacité de liaison. On a constaté que le conformère R est plus commode pour la liaison. Le rapport de liaison calculée à partir des barrières énergétiques relatives des conformations bioactives (R) -3-F-GABA: (S) -3-F-GABA se trouve à 19,1: 1 qui est en harmonie avec le travail expérimental. Résultats soulignent l'importance de conformères réactives plutôt que analogue du substrat le plus peuplé dans les réactions d'inhibition impliquant PLP.

Dans les réactions de transimination, les différences stéréochimiques des analogues substitués par fluor affectent les barrières d'énergie d'activation par l'intermédiaire d'interactions électrostatiques. Ici, l'étape de détermination de vitesse est constaté que le transfert de protons 1,3 dans la réaction de transimination. Les différences entre les activa- barrières d'énergie tion des fluor substitués analogues du GABA ne sont pas très distinctif. Par conséquent, la mauvaise performance de l'énantiomère (S) ne peut pas être corrélé avec le profil d'énergie du transimination. Les réactions HF d'élimination des analogues de fluor-substitué sont également importants pour élucider la liaison de préférence de GABA-A, et d'analyser la raison de la faible performance de (S) -3-F-GABA, qui est présenté expérimentalement [28]. La première étape du mécanisme E1cb de l'énantiomère (R) (06TS10R) et (S) - énantiomère (09TS12S) ont des mêmes barrières d'énergie d'activation, où le carbanion intermédiaire de la (R) - énantiomère (10R) est plus stable. Pour R énantiomère, un état de transition sans barrières (10TS11R) est obtenu pour la dernière étape de l'élimination de HF. Deux scénarios différents sont pris en compte pour l'élimination de HF dans l'isomère (S) cas. On constate que l'état de transition de rotation (12postROT12ant) et l'état de transition implique des molécules d'eau que des navettes de protons (12postTS11) ont des énergies libres relatives similaires étant supérieure que la première étape d'élimination. La dernière étape du chemin est le caractère distinctif pour l'élimination de HF effi- les différences de carence pour les énantiomères. La libération d'ions



fluorure de la énantiomère (R) est plus rapide que la Énantiomère (S) supportant la vitesse d'élimination de HF inférieure de l'énantiomère (S), mais des résultats plus réalistes peut être obtenu en associant les acides aminés environnants du site actif.

**Simulations par dynamique moléculaire des formes apo, holo et inhibé par un inactivateur de GABA-AT** Le pyridoxal 5-phosphate (PLP) cofacteur est une molécule organique significative en chimie médicinale. On le trouve souvent lié de façon covalente à des résidus de lysine dans les protéines pour former des enzymes PLP-dépendantes. Un exemple de cette famille des enzymes dépendantes de PLP est le  $\gamma$ -aminobutyrique aminotransférase d'acide (GABA-AT) qui est responsable la dégradation du neurotransmetteur GABA. Son inhibition ou d'inactivation peut être utilisé pour empêcher la la réduction de la concentration en GABA dans le cerveau, ce qui est à l'origine de plusieurs troubles neurologiques.

Les résultats des études expérimentales [45–52], théoriques [33, 34] et RMN [11–13] nous amènent à étudier la protonation états des résidus de sites actifs dans l'enzyme dépendante du PLP GABA-A. Est-ce que l'anneau de pyridine de l'ours PLP un atome d'azote protoné et comment se comporter PLP dans le site actif? Quel est l'effet de la protonation états des résidus de sites actifs sur les interactions du site actif proposé? Faire ces résidus ont secondaire les rôles? Quel serait le comportement de l'enzyme quand un inactivateur est impliqué? Afin de répondre à ces questions, nous avons réalisé dynamique moléculaire (MD) des simulations. Apoenzyme, holoenzyme et vigabatrin Des échantillons de holoenzyme liés sont simulés pour évaluer les états de protonation et les rôles de l'actif Site résidus dans GABA-AT et, par extension, certains aspects généraux des enzymes dépendantes de PLP. Différent états de protonation ont été envisagées pour PLP et deux résidus de sites actifs clés: Asp298 et His190. Ensemble, 24 dynamique moléculaire indépendants trajectoires ont été simulées pour un total cumulatif de 2,88  $\mu$ s.

L'atome d'azote sur le cycle pyridine de PLP est suggéré d'être protonés dans le site actif de PLP dépen- enzymes dants. Cependant, les études de RMN indiquent qu'il ne peut pas être protoné dans des conditions physiologiques chez solution aqueuse. Nos simulations MD de apo, holo et vigabatrin liés échantillons GABA-AT jettent des lumières les états de protonation des résidus du site actif. Il est suggéré que Asp298

et His190 devraient être ionisée et chargée négativement Asp298 induit l'existence d'un proton sur l'atome d'azote du cycle pyridine du PLP. Par conséquent, comme on en trouve dans l'AAT par des études de RMN [12, 13], l'atome d'azote protoné du cycle pyridine de PLP dans GABA-AT devrait faciliter la première demi-réaction en provoquant un effet de puits d'électrons. Etant donné que le pKa de PLP est différente en solution et dans la protéine, il est diade Asp298 / His190 qui déplace le pKa PLP qui, à son tour, déclenche l'activation PLP pour la première demi-réaction. Ceci est le principal facteur qui détermine la réactivité du récepteur GABA-A.

Afin d'atteindre la seconde moitié réaction dans GABA-AT, PLP / PMP devrait être maintenu en place à l'actif site. Nos simulations d'échantillons GABA-AT MD présentent solide réseau de liaisons hydrogène entre la protéine et PLP. Le groupe phosphate est maintenu en place par le don d'au moins trois liaisons hydrogène tandis que le carbonyl oxygène du cycle de pyridine interagit avec Gln301. Phe181 forme un  $\pi - \pi$  interaction d'empilement avec l'anneau de pyridine et travaille en tant que gardien de la porte avec l'aide de Val300. Les interactions entre les résidus de sites actifs (par exemple, Asp298, His190, etc.) aident également à PLP tenue en place.

On a constaté que, dans presque 25% des protéines dépendantes de PLP disponibles dans la Protein Data Bank, un résidu Asp est orienté vers la bague de pyridine tandis qu'un résidu His est en contact étroit avec ce résidu Asp. Ces enzymes dépendantes PLP seraient afficher les propriétés de réactivité similaires comme dans GABA-AT. En revanche, l'absence d'un résidu Asp ou d'un résidu chargé positivement, qui facilite Asp ionisé pourrait provoquer différents problèmes de réactivité. Etant donné que la protonation de l'atome d'azote du cycle de pyridine est le principal facteur qui définit la réactivité de l'enzyme, l'interaction de cet atome avec doit être résolu dans ses environs Afin de définir la réactivité des enzymes dépendant du PLP.

On a constaté que les interactions de PLP dans l'échantillon holo-GABA-A ne sont pas altérées par l'implication de vigabatrin. les résultats de GIST montrent qu'une fois vigabatrin se lie à PLP, l'eau ne peut pas accéder au site actif autant que dans les échantillons holo-GABA-AT. Nos résultats confirment que le vigabatrin est un bon inhibiteur car il ne ne pas perturber les interactions principales de PLP et ne permet pas un autre substrat pour entrer dans le site actif.

**Effet des résidues environnants dans le mécanisme d'inactivation de GABA-AT par le  $\gamma$ -vinyle GABA** Afin d'obtenir des fonctionnalités plus fiables de la surface d'énergie potentielle, comme les barrières et la thermodynamique, une partie substantielle du site actif de l'enzyme est nécessaire pour être inclus dans le modèle. La protéine environnante peuvent affecter la description de la réaction de deux manières; l'effet stérique imposé par les environs résidus sur le modèle, et à long terme de polarisation qui peut affecter les énergies de réaction. stériques les effets sont modélisés par l'incorporation des résidus environnants à l'intérieur du modèle et en bloquant certaines Coordonnées atome où la troncature est effectuée. Le modèle de gagner en flexibilité avec l'augmentation de la taille et nous espérons obtenir des résultats plus précis que d'un petit et rigide modèle [58]. La polarisation de longue portée est modélisé par la technique du continuum polarisable qui fournit un milieu homogène de certaines constante diélectrique. L'effet de la solvation sur le noyau de la réaction diminue avec l'augmentation de de la taille du modèle [59, 60]. Si les effets stériques et à long terme de polarisation sont pris en compte, suffisamment grand modèle peut fournir mimique optimale en maintenant la réaction modélisé aussi proche que possible de la une expérimentale. Pour comprendre l'effet des résidus de sites actifs sur les réactions impliquant PLP, qui a été proposé d'avoir un atome de N protonée sur l'anneau de pyridine pour faciliter les réactions enzymatiques [3, 5–10], modèles de cluster quantiques ont été utilisés. Cette approche a été appliquée à beaucoup de classe d'enzymes [59–73] et a avéré être très utile pour étudier les réactions enzymatiques.

Des études de NMR de N <sup>15</sup> étiquetés cycle pyridine PLP en solution aqueuse a proposé que le cycle pyridine de PLP est le plus souvent non protoné dans des conditions physiologiques [11]. En outre, dans une étude de modélisation quantique le mécanisme de réaction  $\omega$ -transaminase, l'optimisation de la géométrie finissent par le transfert du proton à partir de N atome du cycle pyridine du PLP à un groupe carboxyle d'Asp259 [65]. D'autre part, la RMN 15studies sur les complexes de aldenamine silylée en présence d'un modèle ASP222 Boc-Asp-OtBu [12] et N marqué AspAT [13] suggéré que l'atome d'azote sur le cycle pyridine est protoné en AspAT. la protonation de l'azote sur le cycle pyridine du PLP a été suggérée pour faciliter les réactions enzymatiques enzymes dépendantes PLP par un effet de puits d'électrons [3, 5–10].

Il a également été proposé que la protonation du cycle pyridine N atome peut être réalisé avec l'aide de Asp222 et His143 en AspAT [13]. Notre molécule des simulations de dynamique de GABA-AT a révélé l'importance des résidus Asp298 et His190 sur la protonation de l'atome d'azote sur le cycle pyridine. Dans cette étude, nous montrons que l'interaction entre PLP et Asp298 peut être atteinte ou conservée uniquement avec un cycle pyridine protoné, une charge négative Asp298 une chaîne latérale et un His190 chargé positivement.

Pour étudier l'effet du site actif et ses résidus entourant le mécanisme d'inactivation de GABA-A, une étude approfondie est mise en œuvre en liaison avec la précédente expérimentale et résultats théoriques. Quel est l'effet de l'état de l'anneau PLP pyridine N atome sur enzymatique de protonation réactions? Quels facteurs font respecter l'atome N du cycle de la pyridine être protoné à l'actif site? Quels sont les effets des résidus environnants sur les paramètres géométriques des structures de l'Etat de transition? Quels résidus doivent être impliqués dans les études de modélisation pour obtenir le modèle qui offre optimale mimique du site actif de la GABA-A? Afin de répondre à ces questions, la réaction d'addition de Michael qui a été rapporté comme mécanisme d'inactivation le plus probable de  $\gamma$ -vinyl-GABA avec le GABA-AT [34,46] est étudiée en utilisant l'approche de cluster quantique. Les effets des résidus environnants, la protonation de N atome sur le cycle pyridine et comment appliquer cette protonation a été évaluée avec l'inclusion de différents entourant les résidus.

Dans cette partie de la thèse, la réaction d'addition de Michael de GABA  $\gamma$ -vinyle avec GABA-AT a été modélisée avec différentes tailles de grappes pour trouver le modèle optimal dans la mimique du site actif. Les tailles de cluster ont été augmentées jusqu'à ce que l'effet de la polarité de l'environnement sur la réaction était minimale. l'inclusion des résidus de sites actifs dans les modèles nous a aidé à expliquer les effets de ces résidus sur la réactivité. La présence de résidus similaires dans de nombreuses enzymes dépendantes PLP nous a permis de fournir certains aspects généraux de la modélisation de ces enzymes.

Dans le plus petit groupe Lys329 était représenté comme méthylamine et le groupe phosphate a été modélisé comme un groupe méthyle. Les calculs ont fait remarquer que le modèle est en mesure de fournir les perceptions générales à propos de

paramètres et énergétique géométriques, mais pas assez pour imiter le site actif. On a observé que l'inclusion du groupe phosphate, non seulement augmente les barrières, mais a également augmenté l'effet de la polarité de l'environnement. Une diminution de la barrière d'énergie d'activation avec l'inclusion de Gly136, Ser137 et Thr353 \* a fait remarquer que le groupe de phosphate doit être soit représenté comme méthyle ou devrait être introduit avec ses résidus environnants dans la modélisation des études pour obtenir une meilleure description du site actif et les résultats énergétiques fiables. Même si l'effet de la polarité de l'environnement a été réduit avec l'inclusion de Arg192 et Asp298, les barrières d'énergie d'activation augmentent. On a observé que le proton sur l'atome d'azote du cycle pyridine se déplace spontanément de l'atome d'azote correspondant, et la chaîne latérale de Asp298 comme il a été indiqué dans une étude de modélisation précédente impliquant Asp et PLP [65]. Il a été proposé que le fort effet de puits en raison du proton sur l'atome d'azote du cycle pyridine facilite les réactions enzymatiques [3, 5–10]. Par conséquent, le mouvement spontané du proton peut être la cause principale de l'augmentation de l'activation barrière d'énergie.

Nous avons signalé dans notre dynamique moléculaire étude que His190 dans GABA-AT peut augmenter l'interaction entre Asp298 et l'anneau de pyridine du PLP, et que le His190 chargé positivement peut aider à Asp298 rester chargé négativement [13, 74]. Par conséquent, la chaîne latérale de la His190 a été incluse dans le modèle pour appliquer la proposée effet de puits d'électrons. Il a été observé que la fourniture d'une liaison hydrogène entre chargé négativement Asp298 chaîne latérale et chaîne latérale His190 chargée positivement aide le proton à rester sur l'atome N du anneau de pyridine comme il a été proposé. Il a été constaté que les deux barrières d'énergie d'activation et l'effet de la polarité de l'environnement a diminué notablement en gardant le proton correspondant à la pyridine anneau N atome, ce qui confirme l'effet proposé de puits d'électrons.

Les paramètres géométriques du centre de l'Etat de transition, à l'exception des angles, sont modifiés lorsque la taille du modèle augmente. Les altérations sont principalement dues à l'élimination des atomes gelés par l'inclusion de plus l'interaction des résidus. Il a été observé que les distances dans le centre de l'Etat de transition étaient affecté avec l'inclusion du résidu Thr353 \* qui a modifié la

position de la molécule d'eau aider. Il a également observé que, l'inclinaison de la bague de PLP, qui pourrait être due à des rotations au sein de la  $\gamma$ -vinyl- GABA peut également effectuer dièdres et les distances à l'intérieur du centre de l'état de transition. Les résidus du site actif rapportés permettent d'obtenir une mimique optimale du site actif de la PLP une enzyme dépendante du GABA-AT. Il convient de noter que le groupe phosphate de PLP doit être soit représenté que méthyle ou devrait être fourni des liaisons hydrogène dans les études de modélisation pour éviter l'énergie inexactes valeurs. La présence d'un résidu chargé positivement qui peut interagir avec les substrats carboxyliques réduire les barrières d'énergie d'activation. L'inclusion de Asp-His diade, qui existe dans de nombreux PLP dépendants enzymes, peuvent appliquer la Asp pour rester chargé négativement. La présence de cette diade peut influencer sur l'électricité propriétés tronic de l'anneau PLP de pyridine et peut remarquablement améliorer la mimique du site actif dans les études de modélisation des enzymes dépendant du PLP.

**Amélioration de la mise à l'échelle linéaire des simulations SEBOMD** L'étude précise et fiable de systèmes biologiques reste un défi pour la communauté chimie quantique (QC). Le Chem quantique approche ical est souvent le seul moyen fiable pour modéliser les transformations chimiques: liaison de décision / rupture, transfert d'électrons (mécanismes d'oxydo-réduction), la transition électronique entre différents niveaux d'énergie, ionisation, etc. Malheureusement, la solution numérique de l'équation de Schrödinger devient de plus en plus (et souvent prohibitif) coûteux en termes de puissance de calcul lorsque la taille du système augmente. Le pro-numérique procédures en QC exigent le calcul de nombreuses intégrales et la résolution des équations (non) -Linear, limitant ainsi la taille des systèmes qui peuvent être modélisés. En outre, la plupart des algorithmes QM, dans leur standard, conduisent à une mise à l'échelle non linéaire du temps CPU par rapport au nombre d'atomes d'être pris en considération.

Au cours des dernières années, de nouveaux algorithmes ont été conçus pour calculer l'énergie d'une très grande moléculaire les systèmes (jusqu'à quelques milliers d'atomes) en utilisant ce qu'on appelle les méthodes de mise à l'échelle linéaire pour lequel la CPU le temps nécessaire pour évaluer l'énergie d'un système est

linéairement dépendant de la taille du système [75–78]. Parallèlement, en raison des nombreux degrés de liberté qui représentent ces grands systèmes, Car-Parrinello et techniques Oppenheimer dynamique moléculaire ont été développées pour explorer leur énergétique, structurelle et les propriétés dynamiques [79].

Aujourd'hui, seule l'échelle linéaire des méthodes de chimie quantique approximatives comme semiempirique ou Densité Binding (DFTB) approches fonctionnelles à base Tight sont capables de combiner à la fois la vitesse nécessaire et la précision pour modéliser les grands systèmes biomoléculaires avec un coût CPU raisonnable, et pour permettre à long terme simulations de dynamique moléculaire au niveau QC (jusqu'à la période de la nanoseconde) [80, 81]. Le Chem théorique groupe tère de Nancy a récemment mis en œuvre semi-empirique de Born-Oppenheimer Molecular Dynamics (SEBOMD) capacités de la suite de logiciels de AMBER en ajoutant à la partie MD de AMBER semiempir- QC routines iques qui utilisent l'échelle linéaire Divide & Conquer algorithme (D & C) [82, 83]. Aujourd'hui, cela permet MD simulations de systèmes moléculaires contenant plusieurs centaines d'atomes pour quelques nanosecondes utilisant un sin- poste de travail gle. Cependant, alors que l'algorithme D & C est intrinsèquement parallèle et échelles bien jusqu'à quelques des dizaines de cœurs de processeur, il ne semble pas pour nous que cet algorithme est le mieux adapté pour une utilisation sur massivement parallèle ordinateurs contenant des milliers de cœurs [84, 85]. Ainsi, notre objectif est d'améliorer l'évolutivité du semiempirique Born-Oppenheimer approche dynamique moléculaire sur des ordinateurs parallèles.

Aujourd'hui, l'implémentation D & C actuelle est pas le plus efficace de faire fonctionner le très grand superordinateurs teurs ou sur les GPU. En utilisant les bibliothèques numériques disponibles qui évoluent bien sur ces nouvelle architecture matérielle, tels que LAPACK et / ou scalapack, nous serons en mesure de lutter contre les systèmes plus importants par des simulations SEBOMD. Au lieu des simulations de courant qui peuvent produire des trajectoires de quelques nanosecondes sur les systèmes confinement ment quelques centaines d'atomes (par exemple un peptide dans une boîte a des molécules d'eau), nous devrions être en mesure d'effectuer simulations sur des milliers d'atomes, ce qui conduit, pour la première fois, aux simulations de (petit) vraiment solvaté protéines en utilisant une description quantique complète au niveau atomique.

La diagonalisation de la matrice Fock dans la partie semi-empirique des simulations SEBOMD est temps la consommation et la version actuelle de SEBOMD ne peut pas évoluer bien sur des ordinateurs parallèles. Afin de améliorer la performance de l'approche SEBOMD, nouveaux solveurs eigenvalue de LAPACK & scalapack ont été mis en œuvre et évalués.

Des études de profilage préliminaires ont fait remarquer que, si pleine diagonalisation est appliquée en phase condensée systèmes sans interactions à longue portée, les routines SCF devient la partie dominante dans SEBOMD par du temps écoulé. Ainsi, les solveurs mises en œuvre ont été profilée en utilisant ces options méthodologiques afin d'observer la performance qui est en grande partie dépendante de la part de l'EFC.

Sept différents solveurs LAPACK de eigenvalue ont été mises en œuvre au total et la déviation du total l'énergie de l'énergie moyenne a indiqué que tous les résolveurs trouvé eigenpairs correcte et peut être utilisé pour simulations de dynamique moléculaire. Profilage des études utilisant des tailles de systèmes de plus en plus (64-216 eau molécules) ont été effectuées en utilisant soit des compilateurs Gnu ou Intel pour évaluer l'effet du contrôle externe sur la performance. Le temps par procédure de diagonalisation a été calculée avec l'horloge interne de sander. Il a été observé que le compilateur Gnu avec les solveurs LAPACK intégrés n'a pas fourni haute performance à la fois pour la coquille fermée standard et D & C algorithmes. En outre, le solveur par défaut dans SE- BOMD (la routine diag) était toujours le plus rapide diagonalizer avec ces options, et seulement solveur dsyevr pourrait fournir des performances similaires à la routine diag. Cependant, le rendement augmente lorsque la Intel MKL a été utilisé avec le compilateur Intel. Quand un algorithme d'enveloppe de type fermé a été appliqué, tout le solveurs mises en œuvre étaient plus rapides que la routine diag compilé avec l'option Intel MKL. Dans le cas de l'algorithme D & C, le solveur par défaut a été trouvé pour être plus rapide que les solveurs qui utilise le stockage emballé (Dspev, dspevd et dspevx). Pour la coquille fermée standard et les algorithmes D & C, le solveur dsyevr présentait de meilleures performances que les deux solveurs récemment et la routine diag. le les résultats obtenus ont fait remarquer que les performances de la procédure de diagonalisation a été sensiblement affectée par le contrôle externe. Ainsi, pour assurer la sélection de la plus rapide diagonalizer, une valeur propre automatique module de sélection de solveur a été mis en œuvre. Lorsque



ce module a été appelé, chaque diagonalizer a été testé respectivement au cours d'une période de balayage (les huit premières itérations SCF) et le solveur le plus rapide a été choisi et utilisées pour d'autres calculs.

Afin d'améliorer l'efficacité des SEBOMD sur des ordinateurs parallèles, solveurs de scalapack ont également été mis en œuvre en plus des solveurs LAPACK. Ces solveurs utilise principalement des algorithmes similaires à calculer eigenpairs que dans leurs versions LAPACK respectives. La mise en œuvre des solveurs scalapack nous a permis d'utiliser l'algorithme de shell fermé standard dans les simulations SEBOMD, pour la première fois, sur le parallèle les machines. On a observé que les solveurs de scalapack n'échelle bien et le meilleur rendement parmi les solveurs SCAPACK pourraient être obtenus avec pdsyevx sur 2 noyaux. Simulations avec l'algorithme D & C ont fait remarquer que, solveurs scalapack ne fournissaient pas une plus grande efficacité comme dans la norme fermée shell rithme rithme. En fait, le D & C algorithme a de moins bonnes performances que l'algorithme standard de coquille fermée lorsque il n'y avait qu'un local\_master. En outre, l'utilisation de différentes tailles de tampons a montré que la rendement strictement dépendait de la taille de la matrice Fock. Étant donné que les tailles de matrices Fock dans le D & C algorithme ne sont pas assez grand, on ne peut obtenir une plus grande efficacité avec des solutionneurs de scalapack qu'avec les solveurs LAPACK.

La meilleure performance des solveurs LAPACK contre leurs versions scalapack nous a conduit à se demander si la performance peut être améliorée en utilisant les versions filetées des diagonalizers. Afin d'étudier la efficacité avec solveurs eigenvalue filetées, simulations préliminaires de la boîte de 216 molécules d'eau a été effectuée en utilisant l'algorithme D & C sur 1 core avec un nombre différent de fils. On a observé que on peut améliorer la performance de la procédure de diagonalisation avec l'utilisation de fils, en particulier avec le dsyevd solveur. En utilisant des études préliminaires en tant que référence, simulations avec une boîte contenant une petite protéine (TrpCage, 20 résidus, 284 atomes) et 730 molécules d'eau ont été réalisées avec l'algorithme D & C. Des simulations ont été réalisées avec un nombre différent de noyaux totaux sur un nombre différent de noeuds en utilisant nombre différent de fils. le temps écoulé ont été comparés au moyen du nombre total de noyaux. Résultats a fait remarquer que, lorsque le nombre total de cœurs étaient pas assez grand, en utilisant seulement 1 fil

fourni une meilleure performance. Cependant, les performances des simulations avec 1 et 2 fils est devenu semblable à l'augmentation en nombre total de cœurs. Lorsque le nombre total de processeurs qui ont été utilisés a atteint 128 noyaux, simulations requis moins de temps avec plus de 1 fil.

### **Simulations SEBOMD de la cassure de liaisons C-N dans des complexes au**

**PLP** L'utilisation de petites modèles représentatifs ont fourni quelques indices sur les réactions se produisant dans GABA-AT [3, 33, 34, 86], alors que grands modèles [65] et calculs QM / MM [87–90] ont fourni des informations sur l'effet des résidus environnants sur la réactivité des enzymes dépendantes PLP. L'effet de la solvation explicite sur les stabilités de GABA conformères ont été étudiés au niveau de QM [19] et les stabilités des analogues du GABA ont été abordés avec QM / calculs MM [91].

Outre les divers calculs et QM QM / MM, les simulations SEBOMD peuvent également être effectuées pour étudier des réactions chimiques impliquant des complexes PLP car l'approche utilise un NDDO hamiltonien comme un champ de force de réaction. méthodologie SEBOMD ont été appliquées avec succès à des molécules en condensé phases, comprenant de l'eau liquide et de N-méthylacétamide dans l'eau [82, 83]. Cependant, les transformations chimiques telles comme formant liaison / rupture n'a pas encore été étudié. Les solveurs eigenvalue nouvellement mis en œuvre dans SEBOMD permet maintenant les simulations de milliers d'atomes, mais des simulations de protéines plus grandes ne peuvent toujours pas être effectuée en raison de la demande de calcul. Au lieu de cela, les structures des modèles et des molécules d'eau explicites peuvent être inclus dans les échantillons.

Afin d'étudier les transformations chimiques avec des simulations SEBOMD, quatre obligations différentes C-N formation / réactions de rupture sont étudiés en utilisant parapluie échantillonnage. Les réactions réversibles sont impliquées dans les mécanismes d'inactivation proposé de GABA-A avec  $\gamma$ -vinyl-GABA [46] et ont été étudiés théoriquement [34].

La première réaction en cause implique l'attaque nucléophile de l'atome N de  $\gamma$ -vinyl-GABA pour la L'atome de carbone du complexe PLP-méthylamine. Cette réaction est la première étape de l'aldimine externe proposée formation entre  $\gamma$ -vinyl-GABA et PLP. La deuxième réaction est la rupture de la liaison C-N dans

le  $\gamma$ -vinyl-GABA-PLP-méthylamine complexe. Cette réaction aboutit à un complexe  $\gamma$ -vinyl-GABA-PLP où la méthylamine est le groupe labile. La troisième réaction est la première étape de la pathway énamine proposée, qui comprend l'attaque de la méthylamine sur le carbone imine PLP. La dernière réaction en question se produit également dans la voie proposée énamine et implique la rupture de la liaison C-N qui produit complexes PLP-méthylamine. Pour plus de simplicité, toutes les réactions réversibles sont modélisés comme rupture de liaison. Les réactions, ce qui signifie que les produits sont des molécules séparées.

Le premiers exemple de simulations SEBOMD de transformations chimiques fournit des informations générales sur les capacités de la méthodologie SEBOMD pour suivre des coordonnées de réaction and pour fournir des estimations d'énergie libre relative.

Afin d'étudier les possibles modifications des surfaces d'énergie potentielle, différentes approches informatiques et différents hamiltoniens ont été utilisés pour modéliser les réactions avec les simulations SEBOMD. Il a été trouvé que l'hamiltonien PM3 surestiment les distances C-N dans presque toutes les structures d'état de transition par rapport à celles localisées avec les méthodes QM et QM/MM. Les barrières d'énergie libre d'activation ont été surestimées dans presque tous les cas avec l'hamiltonien PM3. De plus, cet hamiltonien PM3 amène à des réactions endothermiques à la fois par les calculs QM (RO3 et RO4) et QM/MM (RO2, RO3, et RO4).

Les capacités de SEBOMD à modéliser des transformations chimiques ont été étudiées sur des variations de distance C-N et des énergies libres relatives. Il a été observé que la méthodologie SEBOMD a été capable de localiser des structures d'état de transition avec des distances C-N équivalentes que celles obtenues par des méthodes QM et QM/MM. Les barrières d'énergie libre d'activation sont surestimées d'environ 3-10 kcal/mol dans les simulations SEBOMD par rapport au calcul QM utilisant l'hamiltonien PM3. Quand les calculs QM/MM sont pris en compte, il est observé que les simulations SEBOMD surestiment les barrières d'énergie libre d'activation de moins de 4 kcal/mol.

## 1. INTRODUCTION

PLP, the phosphorylated and the oxidized form of vitamin B6, is an organic cofactor which forms an imine with the  $\epsilon$ -amino group of a lysine residue of PLP-dependent enzymes [1]. The formation of imine can be referred as a Schiff base formation [2]. This kind of intermediate appears in many different enzymatically catalyzed reactions and leads to a large diversity of products. There is a particular interest about PLP-dependent enzymes due to the diversity of their catalyzed chemical reactions such as decarboxylation,  $\beta$ -elimination, aldol cleavage, transamination and transamination [1,3,4].

The nitrogen on the pyridine ring of PLP is suggested to be protonated in PLP dependent enzymes. Such acidic state is believed to cause an electron sink effect that would thereby facilitate the enzymatic reactions [3,5–10]. The protonation state of the pyridine ring of PLP in water was explored using  $^{15}\text{N}$  NMR of  $^{15}\text{N}$  labeled PLP pyridine ring in aqueous solution at room temperature [11]. In this study, Sharif et. al. reported a pKa value for the pyridine ring of 5.8 [11] which indicates that the pyridine ring of PLP is mainly unprotonated under physiological conditions. In another work of Sharif et al.,  $^{15}\text{N}$  liquid state NMR studies were performed on Aspartate Aminotransferase (AspAT) and on model compounds [12]. They reported that the  $^{15}\text{N}$  chemical shifts for the pyridine ring of PLP are consistent with the presence of a proton on the pyridine ring. A similar proton interaction could be reproduced on a model system made of a 1:2 complex between an aldimine, representative of PLP-Lys, and a capped aspartic acid in a solvent mixture of dielectric constant  $\epsilon_r \approx 30$ . Furthermore, Limbach et. al. reported in their  $^{15}\text{N}$  NMR study on  $^{15}\text{N}$  labeled AspAT in water at pH 7.5 that the shortened distance of  $\text{N}^{\text{pyridine}} - \text{H}$  and lengthened distance of  $\text{H} - \text{O}^{\text{Asp222}}$  indicates an increased zwitterionic character of the OHN hydrogen bond with a protonated N on the pyridine ring under physiological conditions [13]. They also proposed that AspAT should provide extra interactions, such as Asp222,

His143 and two conserved water molecules, in order to achieve proton transfer to the pyridine ring N [13].

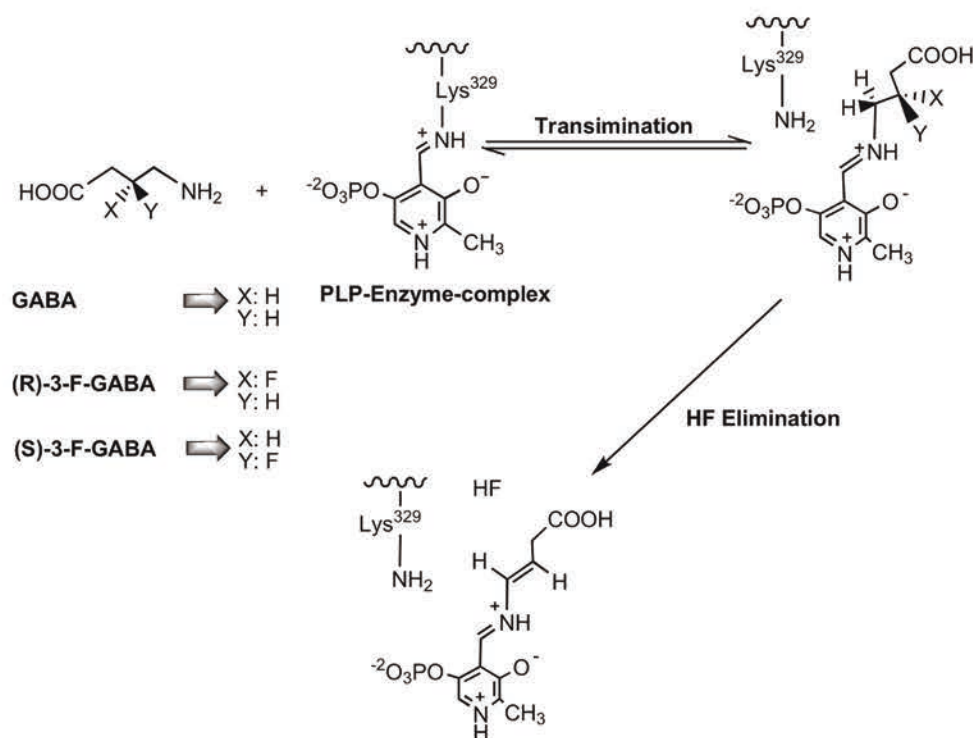
The catalytic mechanism of PLP dependent enzymes is usually divided into two parts: a first and a second half-reaction [14, 15]. The first half-reaction involves the conversion of PLP to pyridoxamine phosphate (PMP) [14, 16]. The PMP molecule in the active site is then bound to the enzyme again during the second half-reaction [14, 16, 17].

In PLP-dependent enzymes, the amount of  $\sigma - \pi$  overlaps in the cofactor-derived Schiff bases determines the specificity of the enzymatic reactions, which corroborates the Dunathans stereoelectronic model [8, 18–24]. The cofactor-derived Schiff bases in PLP-dependent enzymes are formed via the transimination reaction, which involves an internal aldimine where the PLP and  $\varepsilon$ -amino group of the active site residue is covalently bound [1, 25]. During the reaction, the covalent bond between PLP and the enzyme is broken and the new Schiff base, namely the external aldimine, is formed with the incoming substrate [1, 26, 27].

After the formation of the external aldimine, the chemical reactions have a wide range of diversity based on the specificity of the enzyme and also the nature of the substrate. In the case of fluorine-containing substrates such as  $\beta$ -fluoroamines, the transimination reactions are followed by the elimination of a fluoride ion in the form of HF (Figure 1.1) [28–32]. The computational studies related to PLP-dependent enzymes are mostly concentrated on the external aldimine formation [2–4, 25, 26, 33, 34]. Additionally, few studies exist on base-induced HF elimination reactions with systems that involve a pyridine ring [33, 35].

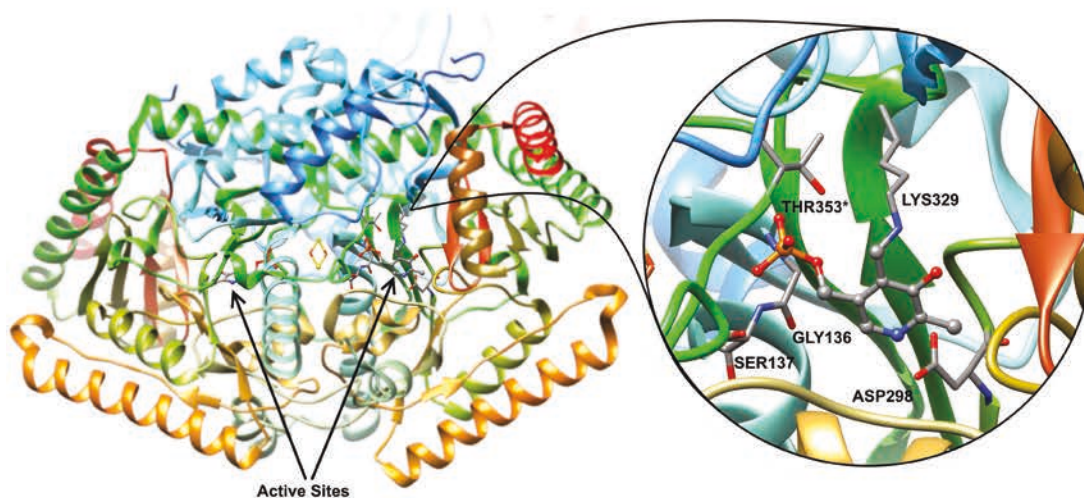
$\gamma$ -aminobutyric acid (GABA) is one of the inhibitory neurotransmitters in the mammalian central nervous system which is degraded to succinic semialdehyde via  $\gamma$ -aminobutyric acid aminotransferase (GABA-AT, E.C. 2.6.1.19) catalyzed reaction [36]. GABA-AT, a member of subfamily II of  $\alpha$ -family of pyridoxal 5-phosphate (PLP) dependent enzymes, is a homodimer in which all subunits contain an active site PLP that is covalently bound to Lys-329 via a Schiff base. Figure 1.2.

The reduction of GABA concentration in brain causes convulsion, besides several neurological diseases such as epilepsy, Parkinson's disease, Huntington's chorea, and



**Figure 1.1 :** Schematic representation of two successive reactions occurring in PLP-dependent enzyme GABA-AT with the natural substrate GABA and the GABA analogues.

Alzheimer's disease [29, 32]. Increasing the concentration of GABA in brain has an anticonvulsant effect in brain [37, 38]. Nevertheless, GABA cannot cross the blood-brain barrier, so the direct control of the level of the concentration of GABA cannot be performed effectively [37, 39]. Consequently, several GABA analogues have been suggested and tested to control the level of GABA in brain by either inhibition or inactivation of GABA-AT [39–42].



**Figure 1.2 :** The active sites within GABA-AT in which an Iron-Sulfur cluster holds two monomers together. Each active site involves a PLP that is bound to Lys329.

Deduction of the primary sequence of GABA-AT was established from pig brain cDNA and peptide fragments of the enzyme from pig liver [43,44]. The X-ray structure of GABA-AT from pig liver was first reported at 3Å resolution, pH 5.6 and at 293K (Protein Data Bank entry 1GTX) [45]. The quality of the X-ray structure of native GABA-AT has been improved and reported at 2.3Å resolution, pH 5.7 and at 100K (Protein Data Bank entry 1OHV) [46]. In this X-ray structure, the PLP cofactor is held in place not only by Gly136 and Ser137 but also by the Thr353 of the second subunit via the donation of five H bonds to the phosphate group [46]. The pyridine ring of PLP is sandwiched between Phe189 and Val300 while the nitrogen on the pyridine ring forms a salt bridge with Asp298 [46]. The mutation of the W101 residue that stacks with the ring of PLP in *E. histolytica* phosphoserine aminotransferase (EhPSAT) to alanine caused to the loss of the enzymatic activity while the W101F mutation showed that phenylalanine could efficiently stack with the ring of PLP [47]. Additionally, in Aspartate aminotransferase (AspAT), another PLP dependent enzyme, the catalytic efficiency of the enzyme was reduced by about 8600-fold and 20000-fold with the D222A and D222N mutations, respectively [48]. Furthermore, it was reported that these mutations increased the deuterium isotope effect which points out that the negatively charged Asp222 side chain stabilizes the protonated ring nitrogen of PLP which functions as an electron sink during the enzymatic reactions [48, 49]. On the other hand, the mutation of the Asp279 residue that interacts with the PLP ring to glutamic acid in 5-Aminolevulinate synthase (ALAS) revealed that the spatial positioning of the negatively charged side chain has also some importance in the activity of the enzyme [50].

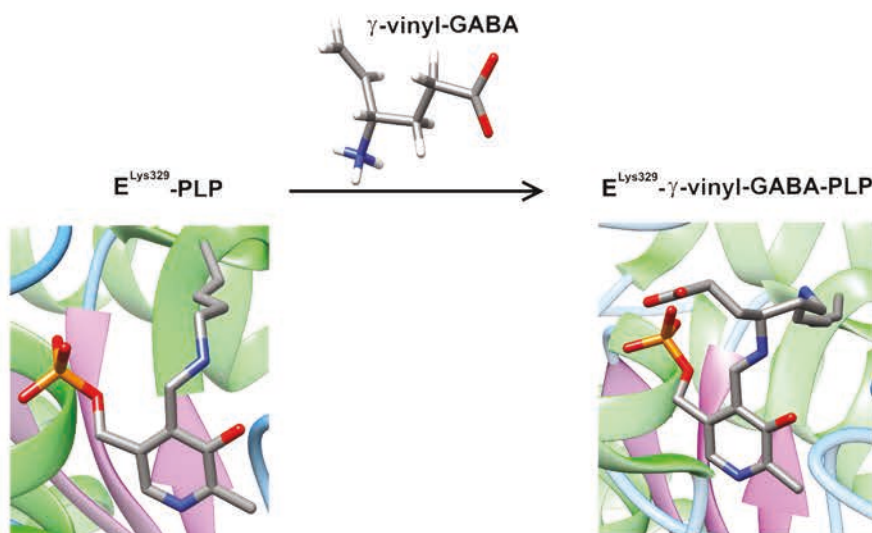
In addition to the native GABA-AT, the X-ray structure of the GABA-AT inactivated by  $\gamma$ -vinyl GABA (vigabatrin) was also reported at 2.3Å resolution, pH of 5.7 and temperature of 100K (Protein Data Bank entry 1OHV) [46]. In this X-ray structure, the Arg192 residue forms a bifurcated salt bridge with the carboxylate group of vigabatrin [45,46]. In addition to these conclusions from X-ray structures, it is also observed that Arg445 is fully conserved in the  $\alpha$ -family of aminotransferases [51] with its positive charge shielded by a salt bridge with Glu270 in GABA-AT [45]. It has been suggested that, to expose a second basic anchor point, the dicarboxylic substrates would require opening of the Glu270-Arg445 ion pair [45,52].



Silverman and co-workers have reported some fluorine containing conformationally restricted analogues of GABA as potential mechanism based inactivators and inhibitors [28,32]. Both experimental and theoretical studies have provided some clues regarding the nature of the catalytic mechanism [28,33].

In an experimental work, Clift et al. proposed fluorine-substituted GABA enantiomers as potential inhibitors of the PLP-dependent enzyme GABA-AT [28] where GABA (Figure 1.1) is the natural substrate for GABA-AT. The fluorine-containing analogues (3-F-GABA, Figure 1.1) could undergo HF elimination after the formation of the external aldimine. The HF elimination reactions may proceed either via stepwise elimination unimolecular conjugate base (E1cb) or concerted bimolecular elimination (E2) mechanisms [28].

Vigabatrin, one of the natural substrate analogues, was reported as the most effective mechanism-based inactivator of GABA-AT with a high potency for treatment of epilepsy and it is used in over 60 countries [36,46,53]. The inactivation mechanisms by vigabatrin, which ends up with an enzyme-vigabatrin-PLP complex (Figure 1.3), have been investigated both experimentally [46,54–57] and theoretically [34], and the Michael addition pathway is pointed out as the most preferred mechanism.



**Figure 1.3 :** GABA-AT inactivation with  $\gamma$ -vinyl-GABA (vigabatrin) with the formation of enzyme-vigabatrin-PLP complex.

The results from the experimental, theoretical and NMR studies lead us to investigate the factors that could effect the reactivity of GABA-AT:



- Can the preferred binding mode of the substrate be the evidence for the selective inhibition, or can the successive reactions be the main cause for the selectivity?
- What are the effect of the electrostatic interactions and the stereoelectronic differences on the inhibition reactions of GABA-AT?
- What are the protonation states of the active site residues, and what is the effect of the protonation states on the proposed active site interactions?
- What would be the behaviour of the enzyme when an inactivator is involved?
- What are the effects of the surrounding residues on the geometrical parameters of the transition state structures that are obtained in modelling studies?
- Which residues should be involved in modelling studies to get a model that provides optimum mimicry of the active site of GABA-AT?
- Are dynamical effects important for the reactivity?

To answer all these questions, a step-by-step approach with increased complexity has been chosen in this thesis. First the reactivity within GABA-AT will be assessed with model systems, from small ones to complexed ones and by exploring the potential energy surfaces with quantum mechanical tools. Then the dynamics of the system will be investigated using molecular mechanics to describe the full protein systems. Finally quantum mechanics and molecular dynamics will be gathered to deeper understand the reactivity of PLP complexes.

Overall, the "Results" chapter of this thesis is divided into five sections that cover our work on the reactivity of GABA-AT. First, small representative models have been used to study two successive reactions of GABA and fluorine containing GABA analogues with quantum mechanical calculations. Then, molecular dynamics simulations have been performed on apoenzyme, holoenzyme, and vigabatrin bound holoenzyme samples to evaluate the protonation states and the roles of the active site residues. Next, using the information provided by the molecular dynamics simulations, quantum mechanical calculations were performed with large representative models involving surrounding residues to obtain optimum mimicry of the active site of GABA-AT. Finally, the selected C-N bond breaking reactions that were proposed to be

involved in the inactivation mechanisms of GABA-AT were modelled using explicit water molecules with SEBOMD methodology. Prior to the SEBOMD simulations, new diagonalization routines have been implemented for the SEBOMD approach to improve its efficiency on parallel computers.



## 2. METHODOLOGY

### 2.1 Quantum Chemistry

#### 2.1.1 The Schrödinger equation

The evolution of the wave function ( $\Psi$ ) in time has been represented by Erwin Schrödinger in 1926 with a partial differential equation [92], which would be referred as the **time dependent Schrödinger equation**;

$$-\frac{\hbar}{i} \frac{\partial \Psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi(x,t)}{\partial x^2} + V(x) \Psi(x,t) \quad (2.1)$$

where  $\hbar$  is Planck's constant divided by  $2\pi$ ,  $m$  is mass of a particle and  $V$  is the potential field. The first step to solve this partial differential equation is expressing the wave function  $\Psi(x,t)$  as the product of two functions which will depend either only on the position ( $x$ ) or only on time ( $t$ );

$$\Psi(x,t) = \psi(x)\zeta(t) \quad (2.2)$$

Substitution of equation 2.2 to the equation 2.1 and reorganization will produce the following partial differential equation;

$$i\hbar \frac{1}{\zeta(t)} \frac{d\zeta(t)}{dt} = -\frac{\hbar^2}{2m} \frac{1}{\psi(x)} \frac{d^2\psi(x)}{dx^2} + V(x) \quad (2.3)$$

As it can be comprehended, the left hand side and the right hand side of the equation 2.3 depends on time and position respectively. Therefore, it can be separated into two parts; a time dependent part and a spatial dependent part which can be constructed by accepting either time or position as constant. The time dependent part can be expressed as

$$i\hbar \frac{d\zeta(t)}{dt} = E\zeta(t) \quad (2.4)$$

while the spatial dependent part can be represented by;

$$\left[ -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right] \psi(x) = E\psi(x) \quad (2.5)$$

In equation 2.5 the part in the parenthesis is the total energy expression in terms of position and momentum, and referred as **Hamiltonian operator** ( $\hat{H}$ ). If equation 2.5 is simplified with this operator, the time independent Schrödinger equation can be mentioned as:

$$\hat{H}\psi(x) = E\psi(x) \quad (2.6)$$

The total energy of a system, which is expressed with its Hamiltonian, is comprised of the kinetic and the potential energies ( $\hat{H} = T + V$ ). The kinetic energy part of the Hamiltonian ( $T$ ) is related to the linear momentum and can be expressed in three dimension as

$$T = -\frac{\hbar^2}{2m} \left[ \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right] = -\frac{\hbar^2}{2m} \nabla^2 \quad (2.7)$$

for a particle with mass  $m$ . In the case of an electron in the field of a nucleus, the potential energy term in Hamiltonian ( $V$ ) is the Coulomb potential that can be represented by:

$$V = -\frac{Ze^2}{4\pi\epsilon_0 r} \quad (2.8)$$

where  $Z$  is the atomic number of nucleus,  $\epsilon_0$  is the permittivity constant and  $r$  is the distance from electron to nucleus.

In the case of a  $n$  particle system, the kinetic energy is expressed as:

$$T = -\frac{\hbar^2}{2m_i} \sum_{i=1}^n \nabla_i^2 \quad (2.9)$$

and the potential energy term can be mentioned as:

$$V = \frac{1}{4\pi\epsilon_0 r} \left[ - \sum_i^{electrons} \sum_k^{nuclei} \left( \frac{Z_k e^2}{|\vec{R}_k - \vec{r}_i|} \right) + \sum_i^{electrons} \sum_{j < i}^{electrons} \left( \frac{e^2}{|\vec{r}_i - \vec{r}_j|} \right) + \sum_k^{nuclei} \sum_{l < k}^{nuclei} \left( \frac{Z_k Z_l e^2}{|\vec{R}_k - \vec{R}_l|} \right) \right] \quad (2.10)$$

In equation 2.10 the first term in parenthesis represents the electron-nuclei attraction, the second term in parenthesis accounts for the electron-electron repulsion, and the last term in parenthesis is the expression for the nuclei-nuclei repulsion.

When the number of the particles in a system is bigger than two, the Schrödinger equation cannot be solved analytically. In order to overcome this difficulty in many particle systems, it is fundamental to introduce approximations.

### 2.1.2 The Born Oppenheimer approximation

The Born-Oppenheimer approximation, which is the most common approximation for many quantum chemistry methods, relies on the idea of the great differences in masses between electrons and nuclei [93]. The nuclei are much slower than electrons and electrons respond to the displacements of the nuclei almost immediately. Thus, the approximation assumes the positions of the nuclei are fixed to solve Schrödinger equation. By doing so, the Schrödinger equation is divided into two parts; one part that is described with the electronic motion yielding the electronic wave function, and one part that is described with the nuclear motion. The electronic wave function depends only on the nuclear position but not their momenta. Therefore the total Hamiltonian is written as:

$$\begin{aligned}\hat{H}_{tot} &= T_n + V_{nn} + T_e + V_{ne} + V_{ee} \\ &= T_n + V_{nn} + \hat{H}_e\end{aligned}\tag{2.11}$$

where the kinetic energy of nuclei and electrons are represented by  $T_n$  and  $T_e$  respectively, and the potential energies of nuclear-electron attraction as  $V_{ne}$ , electron-electron repulsion as  $V_{ee}$  and nuclear-nuclear repulsion as  $V_{nn}$ .  $\hat{H}_e$  term is called the electronic Hamiltonian operator and depends only on the nuclear position.

### 2.1.3 Electron spin, Pauli principle and Slater determinant

The wave function that was described in section 2.1.1 determines the probability of an electron to be found at a given coordinate in three dimensional space. However, the question of placing electrons into orbitals is still not answered. Thus, a new concept should be introduced to answer this question, namely the **electron spin**.

In order to describe the spin of an electron, the magnitude of the total spin angular momentum should be defined first. Assuming that the square of the magnitude of the total spin angular momentum is defined by the linear and Hermitian operator  $\hat{S}^2$  and that the z component of the spin angular momentum is  $\hat{S}_z$ , the operator  $\hat{S}^2$  is expressed as:

$$\hat{S}^2 = \hat{S}_x^2 + \hat{S}_y^2 + \hat{S}_z^2\tag{2.12}$$

Using the commutation relations of operators, the eigenvalues of  $\hat{S}^2$  and  $\hat{S}_z$  can be defined respectively as;

$$s(s+1)\hbar^2 \quad s = 0, \frac{1}{2}, 1, \frac{3}{2}, \dots\tag{2.13}$$

$$m_s \hbar \quad m_s = -s, -s+1, \dots, s-1, s \quad (2.14)$$

The spin of a particle is defined by the quantum number  $s$  in equations 2.13 and 2.14.

In the case of electrons, the quantum number  $s$  can have a single value,  $s = +\frac{1}{2}$ , and  $m_s = +\frac{1}{2}$  or  $m_s = -\frac{1}{2}$ . Using the quantum number  $s$ , the total spin angular momentum of an electron can be given by:

$$\sqrt{\left[\frac{1}{2}\left(\frac{3}{2}\right)\hbar^2\right]} = \frac{1}{2}\sqrt{3}\hbar \quad (2.15)$$

With  $s = +\frac{1}{2}$ , the possible eigenvalues of  $\hat{S}_z$  of an electron can be found as  $\frac{1}{2}\hbar$  and  $-\frac{1}{2}\hbar$ . The eigenfunctions of the given eigenvalues are denoted simply as  $\alpha$  and  $\beta$ . These eigenfunctions are called spin eigenfunctions and are orthonormal.

The wave function ( $\psi(x, y, z)$ ) that was written as a function of the spatial coordinates can now be expressed with the spin variable ( $\psi(x, y, z, m_s)$ ). The Hamiltonian of a system does not involve the spin variable and has no effect on spin. Thus, one can simply express the new wave function of a one particle system as a product of spatial and spin parts;

$$\psi = \psi(x, y, z)g(m_s) \quad (2.16)$$

The value of  $g(m_s)$  can be either  $\alpha$  or  $\beta$  and depends on the value of the quantum number  $m_s$ . For an  $n$  particle system the wave function can be expressed as;

$$\psi = \psi(q_1, q_2, q_3, \dots, q_n) \quad (2.17)$$

where  $q_i$  stand for the spatial coordinates  $x_i, y_i, z_i$  and spin variable  $m_{s_i}$  of particle  $i$ . At this point, a new concept should be introduced; the **spin orbital**. A spin orbital ( $\chi$ ) is simply the product of a spatial orbital ( $\phi$ ) and the spin function ( $g$ ) in a given coordinate  $r$ :

$$\chi_i(r) = \phi(r)g(m_s) \quad (2.18)$$

If the Hamiltonian was separable, the many electron eigenfunctions ( $\psi$ ) could be expressed as product of one-electron eigenfunctions  $\phi_i$ . The newly expressed eigenfunction is called **Hartree product** and denoted as  $\Psi^{HP}$ . Using this formulation and combining it with equation 2.18, one can express the wave function of a system with  $n$  electrons in given coordinates  $x_i$  as a product of spin functions  $\chi_i$ :

$$\Psi^{HP}(x_1, x_2, x_3, \dots, x_n) = \chi_1(x_1)\chi_2(x_2)\chi_3(x_3)\dots\chi_n(x_n) \quad (2.19)$$

The most important phenomenon in here is that *no two electrons can occupy the same spin orbital* or else the wave function would vanish. In order to allow simultaneous exchange of electron space and spin coordinates, wave functions must be antisymmetric with respect to the interchange of any two electrons. This postulate is called the **Pauli's exclusion principle** [94]. The Hartree product given in equation 2.19 does not satisfy this postulate. In order to satisfy this principle, Slater pointed a way to express wave functions as a determinant in 1929 [95].

$$\Psi^{SD}(x_1, x_2, x_3, \dots, x_n) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \chi_1(x_1) & \chi_2(x_1) & \cdots & \chi_n(x_1) \\ \chi_1(x_2) & \chi_2(x_2) & \cdots & \chi_n(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(x_n) & \chi_2(x_n) & \cdots & \chi_n(x_n) \end{vmatrix} \quad (2.20)$$

The  $\Psi^{SD}$  term in equation 2.20 is called a **Slater determinant** of Hartree products.

#### 2.1.4 Linear variation method and secular equation

The linear variation method is based on the idea that a function can be expanded into a linear combination of other functions which span the same function space (i.e.  $e^{ikx} = \cos(kx) + i\sin(kx)$ ). This method along with orbital approximation allows a molecular orbital  $\phi$  to be expressed as a linear combination of  $n$  linearly independent atomic orbitals, which is referred as the **linear combination of atomic orbitals** (LCAO));

$$\phi(c_1, c_2, \dots, c_n) = c_1\phi_1 + c_2\phi_2 + \dots + c_n\phi_n \quad (2.21)$$

where  $\phi_i$  are the known atomic orbitals and  $c_i$  are a set of coefficients which determine the molecular orbital  $\phi$  that is correlated with the minimum electronic energy. At the minimum of the electronic energy, the partial derivative of the energy with respect to any  $c_i$  must be equal to zero. Thus, the following set of partial differential equations must be satisfied;

$$\frac{\partial E}{\partial c_1} = \frac{\partial E}{\partial c_2} = \dots = \frac{\partial E}{\partial c_n} = 0 \quad (2.22)$$

Construction of these  $n$  linear homogeneous equations and reorganizing them would give the following expressions;

$$\sum_{j=1}^n c_j \left[ \int \phi_i \hat{H} \phi_j dv - E \int \phi_i \phi_j dv \right] = 0 \quad i = 1, 2, \dots, n \quad (2.23)$$

$$\sum_{j=1}^n c_j [H_{ij} - \bar{E} S_{ij}] = 0$$



where  $H_{ij}$  is the Hamiltonian matrix element and  $S_{ij}$  is the overlap matrix element. A homogeneous system of linear equations can have a non-trivial solution if and only if its determinant vanishes.

$$\begin{vmatrix} H_{11} - ES_{11} & H_{13} - ES_{13} & \cdots & H_{1n} - ES_{1n} \\ H_{21} - ES_{21} & H_{23} - ES_{23} & \cdots & H_{2n} - ES_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ H_{n1} - ES_{n1} & H_{n3} - ES_{n3} & \cdots & H_{nn} - ES_{nn} \end{vmatrix} = 0 \quad (2.24)$$

Equation 2.24 is called the **secular equation** while the left hand side is called a **secular determinant**.

### 2.1.5 The Hartree-Fock method

The molecular Hartree-Fock (HF) wave function is written as a Slater determinant, and the HF molecular electronic energy ( $E_{HF}$ ) is given as;

$$E_{HF} = \langle \Psi | \hat{H}_{el} | \Psi \rangle + V_{NN} \quad (2.25)$$

where  $\hat{H}_{el}$  is the electronic Hamiltonian and  $V_{NN}$  is the internuclear repulsion term which is considered as constant. Since  $\Psi$  is normalized,  $E_{HF}$  can be expressed in terms of one electron core Hamiltonian ( $\hat{H}^{core}$ ), Coulomb integrals ( $J_{ij}$ ) and exchange integrals ( $K_{ij}$ );

$$\begin{aligned} E_{HF} &= \langle \Psi | \hat{H}_{el} | \Psi \rangle + V_{NN} \\ &= 2 \sum_{i=1}^{n/2} H_{ii}^{core} + \sum_{i=1}^{n/2} \sum_{j=1}^{n/2} (2J_{ij} - K_{ij}) + V_{NN} \end{aligned} \quad (2.26)$$

where

$$\begin{aligned} H_{ii}^{core} &\equiv \langle \chi_i(x_1) | \hat{H}^{core}(x_1) | \chi_i(x_1) \rangle \\ J_{ij} &\equiv \langle \chi_i(x_1) \chi_j(x_2) | \frac{1}{r_{12}} | \chi_i(x_1) \chi_j(x_2) \rangle \\ K_{ij} &\equiv \langle \chi_i(x_1) \chi_j(x_2) | \frac{1}{r_{12}} | \chi_j(x_1) \chi_i(x_2) \rangle \end{aligned} \quad (2.27)$$

From equations 2.26 and 2.27 one can derive the **Fock operator** ( $\hat{F}$ )

$$\hat{F}(x_1) = \hat{H}^{core}(x_1) + \sum_{j=1}^{n/2} (2\hat{J}_j(x_1) - \hat{K}_j(x_1)) \quad (2.28)$$

where  $\hat{J}$  and  $\hat{K}$  are the operators to solve Coulomb and exchange integrals in equation 2.27 respectively. Using the Fock operator, one can express the following one electron differential equation, which is simply called the **Hartree-Fock equations**;

$$\hat{F}(x_1) \chi_i = \epsilon_i \chi_i \quad (2.29)$$

In this equation  $\varepsilon_i$  denotes the orbital energy and is given by the following expression:

$$\varepsilon_i = H_{ii}^{core} + \sum_{j=1}^{n/2} (2J_{ij} + K_{ij}) \quad (2.30)$$

Since the operator  $\hat{F}$  depends on its own eigenfunctions, it is not possible to solve Hartree-Fock equations directly. Therefore, a methodology has been developed to solve HF equations iteratively.

#### 2.1.5.1 The Roothaan-Hall equation and self consistent field (SCF) method

In 1951 Roothaan [96] proposed to express the spatial orbitals as linear combinations of basis functions;

$$\psi_i = \sum_{\mu}^N c_{\mu i} \phi_{\mu} \quad (2.31)$$

and introduced the following expression;

$$\sum_{v=1}^N (F_{\mu v} - \varepsilon_i S_{\mu v}) c_{vi} = 0 \quad \mu = 1, 2, \dots, N \quad (2.32)$$

where

$$S_{\mu v} = \langle \phi_{\mu} | \phi_v \rangle \quad (2.33)$$

and

$$\begin{aligned} F_{\mu v} &= \langle \phi_{\mu} | \hat{F} | \phi_v \rangle \\ &= \left\langle \mu \left| -\frac{1}{2} \nabla^2 \right| v \right\rangle - \sum_k^{nuclei} Z_k \left\langle \mu \left| \frac{1}{r_k} \right| v \right\rangle \\ &\quad + \sum_{\lambda \sigma} P_{\lambda \sigma} \left[ (\mu v | \lambda \sigma) - \frac{1}{2} (\mu \lambda | v \sigma) \right] \end{aligned} \quad (2.34)$$

In equation 2.34, the terms in the form of  $(\mu \lambda | v \sigma)$  are the two electron integrals defined by:

$$(\mu \lambda | v \sigma) = \int \int \phi_{\mu}(x_1) \phi_{\lambda}(x_1) \frac{1}{2} \phi_{\lambda}(x_2) \phi_{\sigma}(x_2) dr(x_1) dr(x_2) \quad (2.35)$$

and  $P_{\lambda \sigma}$  are the elements of the **density matrix** (P) given by:

$$P_{\lambda \sigma} = 2 \sum_i^{occupied} c_{\lambda i} c_{\sigma i} \quad (2.36)$$

Hence equation 2.32 forms  $N$  set of linear homogeneous equation with  $S_{\mu v}$  as the overlap matrix elements, and  $c_{vi}$  as the unknown coefficients, it can be rewritten as;

$$FC = SC\varepsilon \quad (2.37)$$

which is generally called the **Roothaan-Hall equation** [96,97] in its matrix form. This equation can be solved iteratively with a procedure called the **Self-Consistent Field** (SCF). The general steps of the SCF procedure is as follows:

1. Calculate  $H_{\mu\nu}^{core}$ ,  $S_{\mu\nu}$ , and  $(\mu\nu | \lambda\sigma)$ .
2. Obtain a set of  $c_{ij}$  as initial guess and use them to calculate initial density matrix  $P^{(0)}$ .
3. Calculate  $F_{\mu\nu}^{(n)}$  from  $P^{(n-1)}$ . Use the  $S_{\mu\nu}$  and orthogonalization to obtain new orthonormal basis set. With this, the equation 2.37 transforms to  $F'C' = C'\epsilon$ ;
4. Use a diagonalization method to evaluate eigenvalue and eigenvectors of  $F_{\mu\nu}'^{(n)}$  which are  $\epsilon_i$  and  $c'_{ij}$  respectively.
5. Perform back transformation ( $C' \longrightarrow C$ ) to obtain a new set of  $c_{ij}$ .
6. Use new set of  $c_{ij}$  to construct a new density matrix  $P^{(n)}$ .
7. Compare  $P^{(n)}$  with previous  $P^{(n-1)}$ . Test for convergence;
  - If fails then begin the next iteration from step 3.
  - If succeeds then go on to perform other parts of the calculation.

### 2.1.6 Semiempirical methods

The term semiempirical can be defined simply as involving experimental data, assumptions or approximations with the physical theory [98, 99]. Semiempirical quantum mechanical methods are established upon the HF equations, but many of the integrals in calculations are either disregarded or parametrized based on experimental data to reduce the computational effort.

The first and the simplest semiempirical approach was the  $\pi$  electron method or Simple Hückel method (SHM) that generates molecular orbitals from a connectivity matrix [100]. In order to include all valence electrons, Hoffmann developed an extended version of  $\pi$  electron method in 1963, which would be called extended Hückel method (EHM) [101, 102]. However, both SHM and EHM methods involve only one electron integrals and are not iterative.

In 1953, Pariser, Parr, and Pople developed the first SCF type semiempirical method, Pariser-Parr-Pople (PPP), which was limited to  $\pi$  electrons [103, 104]. The first SCF type semiempirical method that is not limited to  $\pi$  electrons was reported by Pople et. al. and named as complete neglect of differential overlap (CNDO) [105, 106]. The limitations in CNDO arising due to the separation of states, led to the development of more sophisticated methods; such as neglect of diatomic differential overlap (NDDO) approximation [105], and intermediate neglect differential overlap (INDO) approximation [107]. In NDDO approximation, the overlap integrals  $S_{\mu\nu}$  are neglected if  $\mu \neq \nu$ ;

$$\varphi_\mu \varphi_\nu = 0 \quad \text{if} \quad \mu \neq \nu \quad (2.38)$$

where  $\varphi_\mu$  and  $\varphi_\nu$  are two different atomic orbitals. As a consequence, the overlap matrix  $S$  becomes unitary;

$$S_{\mu\nu} = \delta_{\mu\nu} \quad (2.39)$$

and the equation 2.32 becomes;

$$\sum_{\nu=1}^N (F_{\mu\nu} - \epsilon_i \delta_{\mu\nu}) c_{\nu i} = 0 \quad \mu = 1, 2, \dots, N \quad (2.40)$$

Additionally, the electron repulsion integrals,  $(\mu\lambda | \nu\sigma)$ , are neglected if two atomic orbitals are located on two different atoms, atom A and atom B;

$$\varphi_\mu \varphi_\lambda = 0 \quad \text{if} \quad \mu \in A, \text{ and } \lambda \in B \quad (2.41)$$

Using these approximations and assuming that  $\varphi_\mu$  and  $\varphi_\nu$  are located on atom A while  $\varphi_\lambda$  and  $\varphi_\sigma$  are located on atom B, the Fock matrix elements are defined as;

$$\begin{aligned} F_{\mu\mu} &= U_{\mu\mu} \sum_B V_{\mu\mu}^B + \sum_\nu^A P_{\nu\nu} \left[ (\mu\mu | \nu\nu) - \frac{1}{2}(\mu\nu | \nu\mu) \right] + \sum_B \sum_{\lambda\sigma}^B P_{\lambda\sigma} (\mu\mu | \lambda\sigma) \\ F_{\mu\nu} &= \sum_B V_{\mu\nu}^B + P_{\mu\nu} \left[ \frac{3}{2}(\mu\nu | \mu\nu) - \frac{1}{2}(\mu\mu | \nu\nu) \right] + \sum_B \sum_{\lambda\sigma}^B P_{\lambda\sigma} (\mu\nu | \lambda\sigma) \\ F_{\mu\lambda} &= H_{\mu\lambda}^c - \frac{1}{2} \sum_\nu^A \sum_\sigma^B P_{\mu\sigma} (\mu\nu | \lambda\sigma) \end{aligned} \quad (2.42)$$

where  $U_{\mu\mu}$  represents the atomic orbital ionization potential, the term  $V_{\mu\nu}^B$  stands for the electron attraction to the other nuclei, and  $H_{\mu\lambda}^c$  are the elements of one-electron core Hamiltonian.

There are many methods that have been developed by using the NDDO approximation. The first method that was proposed by Dewar et. al. in 1977 was called Modified

Neglect of Atomic Orbitals (MNDO) [108–110]. However, the repulsions between atoms that are separated by about their van der Waals distances are overestimated by MNDO method [98]. In order to overcome this problem, MNDO method was improved to Austin Method (AM1) by Dewar et. al. in 1985 [111]. Even though AM1 treated hydrogen bonding better than MNDO, the misrepresentation of the hydrogen bonds was still a problem [112]. A couple of years later Stewart reported another method and simply called it Parametrized model number 3 (PM3) [113–115]. In his method, Stewart used 18 parameters for each element (except H atom that has 11 parameters) while AM1 uses 10 to 19 parameters for each element [113].

After the development of AM1 and PM3, many versions have been reported in order to overcome the deficiencies of these models. In 1993 Dewar's group reported another semiempirical method under the name of semi ab initio method number 1 (SAM1) [116]. Another modified version of AM1, namely AM1/d, was first reported with the parameters for molybdenum in 2000 [117], which was followed by the parameters for magnesium in metalloenzymes [118], and for phosphoryl transfer reactions [119]. In 2002, two new semiempirical methods, PDDG/PM3 and PDDG/MNDO, have been developed with a Pairwise Distance Directed Gaussian modification [120]. It was reported that, the accuracy of the calculated heats of formation for compounds that contain C, H, N, and O atoms were improved relative to the previous NDDO methods [120]. In 2003 an extension of AM1 was reported under the name of AM1\*, in which the parameters for H, C, N, O and F atoms were unchanged but P, S and Cl atoms have been reparameterized [121]. Another modification to AM1, which is called Recife Method 1 (RM1) was reported by Rocha et. al in 2006 to obtain more accurate results from the semiempirical calculations of large biomolecules [122]. In 2007 Stewart reported PM6 by employing several modifications to core-core interactions, and these modifications allowed the parametrization of 70 elements [123]. In 2000, the Parametrizable Interaction Function (PIF) approach was introduced in order to correctly describe the water dimer and the intermolecular interactions in small water clusters [124]. Later on, the PIF parameters were extended to the molecules containing H, C, N, O and Cl atoms that interact with water and new parametrization was simply named PIF2 [125–127]. Marion et. al. reported an extension of PIF2 parameters, namely PIF3, and introduced hydrogen type specific parameters [128].

### 2.1.7 Density functional theory

The number of coordinates that is used to determine the wave function of a many electron system ( $3n$  spatial and  $n$  spin coordinates for a system containing  $n$  electrons) enforced the search for functions that has fewer variables. In 1964, it was proved that the ground state electronic energy ( $E_0$ ) of a molecule could be determined by knowing the molecule's ground state electron density ( $\rho$ ) [129]. In their theorem, Hohenberg and Kohn showed that  $E_0$  is a functional of  $\rho$ ;

$$E_0 = E_0[\rho] \quad (2.43)$$

The electronic Hamiltonian involves the potential energy arising from electron-nuclei attraction (see equation 2.10). In the Hohenberg-Kohn theorem, this potential energy is referred as an external potential acting on an electron. Using a similar notation as in equation 2.11, one can express the ground state electronic energy in terms of functionals;

$$E_0 = E_0[\rho] = T_e[\rho] + V_{ne}[\rho] + V_{ee}[\rho] \quad (2.44)$$

$$V_{ne}[\rho] = \left\langle \psi_0 \left| \sum_{i=1}^n v(r_i) \right| \psi_0 \right\rangle = \int \rho(r) v(r) dr \quad (2.45)$$

where  $v(r_i)$  is called external potential energy function. Even though  $V_{ne}$  is known, the functionals  $T_e$  and  $V_{ee}$  in equation 2.44 are unknown. Thus, using equation 2.44 is not practical to compute  $E_0$ . However, in 1965 Kohn and Sham constructed a method [130] to find  $E_0$  and  $\rho$ .

In their formulation, Kohn and Sham considered a Hamiltonian operator of a non-interacting system of electrons; a fictitious system in which the same external potential energy function acts on all non-interacting electrons. By doing so, the ground state wave function became the Slater determinant of the lowest energy spin-orbitals which are referred as **Kohn-Sham spin-orbitals** ( $\chi^{KS}$ );

$$\psi_0 = | \chi_1^{KS} \chi_2^{KS} \cdots \chi_n^{KS} |, \quad \chi_i^{KS} = \varphi_i^{KS}(r_i) g_i(m_s) \quad (2.46)$$

The spatial part of Kohn-Sham spin-orbital,  $\varphi_i^{KS}(r_i)$ , is the eigenfunction of one-electron Kohn-Sham Hamiltonian  $\hat{h}_i^{KS}$ ;

$$\hat{h}_i^{KS} \varphi_i^{KS}(r_i) = \varepsilon_i^{KS} \varphi_i^{KS}(r_i) \quad (2.47)$$

The term  $\varepsilon_i^{KS}$  in this equation is referred as the Kohn-Sham orbital energies.

In the Kohn-Sham method, the equation 2.44 is rearranged as;

$$E_0 = E_0[\rho] = \int \rho(r)v(r)dr + T_e^f[\rho] + \frac{1}{2} \int \int \frac{\rho(r_1)\rho(r_2)}{r_{12}}dr_1dr_2 + E_{XC}[\rho] \quad (2.48)$$

where  $T_e^f[\rho]$  is the kinetic energy of a fictitious system with non-interacting electrons, and  $E_{XC}[\rho]$  is called the **exchange-correlation energy functional** which is defined as;

$$E_{XC}[\rho] \equiv \Delta T[\rho] + \Delta V_{ee}[\rho] \quad (2.49)$$

The term  $\Delta T[\rho]$  in equation 2.49 is the difference between the true electronic kinetic energy of the molecule ( $T_e[\rho]$ ) and one of the fictitious system ( $T_e^f[\rho]$ );

$$\Delta T[\rho] \equiv T_e[\rho] - T_e^f[\rho] \quad (2.50)$$

and the term  $\Delta V_{ee}[\rho]$  is the difference between the Coulomb energy of the electrons in molecule and one of the fictitious system;

$$\begin{aligned} \Delta V_{ee}[\rho] &\equiv V_{ee}[\rho] - V_{ee}^f[\rho] \\ &\equiv V_{ee}[\rho] - \frac{1}{2} \int \int \frac{\rho(r_1)\rho(r_2)}{r_{12}}dr_1dr_2 \end{aligned} \quad (2.51)$$

Since it is not easy to evaluate  $E_{XC}[\rho]$  accurately, one can use approximations to exchange-correlation energy.

### 2.1.7.1 Approximations to exchange-correlation energy functional

It was shown by Hohenberg and Kohn that an accurate functional can be obtained if  $\rho$  varies slowly with position of electrons. Using a procedure called the **local density approximation** (LDA), they approximate  $E_{XC}[\rho]$  by:

$$E_{XC}^{LDA}[\rho] = E_X^{LDA}[\rho] + E_C^{LDA}[\rho] = \int \rho \epsilon_X^{LDA} \rho dr + \int \rho \epsilon_C^{LDA} \rho dr \quad (2.52)$$

where  $\epsilon_X^{LDA}$  and  $\epsilon_C^{LDA}$  are exchange and correlation energies per electron respectively.

The exchange parts of equation 2.52 are described as;

$$E_X^{LDA}[\rho] = C_x \int \rho^{4/3} dr \quad (2.53)$$

$$\epsilon_X^{LDA} = C_x \rho^{1/3} \quad (2.54)$$

where  $C_x = -\frac{3}{4} \left( \frac{3}{\pi} \right)^{1/3}$ .

Another approximation to  $E_{XC}$  allows different orbitals for electrons with different spins. This approximation is called the **local-spin density approximation** (LSDA) and expresses  $E_{XC}$  as;

$$E_{XC}^{LSDA} = E_{XC}^{LSDA}[\rho^\alpha, \rho^\beta] = E_X^{LSDA}[\rho^\alpha, \rho^\beta] + E_C^{LSDA}[\rho^\alpha, \rho^\beta] \quad (2.55)$$

where  $E_X[\rho^\alpha, \rho^\beta]$  is;

$$E_X[\rho^\alpha, \rho^\beta] = -2^{1/3} C_x \int [\rho_\alpha^{4/3} + \rho_\beta^{4/3}] dr \quad (2.56)$$

and  $\epsilon_X^{LSDA}$  is described as;

$$\epsilon_X^{LSDA} = -2^{1/3} C_x [\rho_\alpha^{1/3} + \rho_\beta^{1/3}] \quad (2.57)$$

In addition to LDA and LSDA, which are appropriate if  $\rho$  varies slowly with position, another approximation include gradients for the variation of  $\rho$  with position. This approximation is called the **generalized-gradient approximation** (GGA) which takes the gradients of  $\rho_\alpha$  and  $\rho_\beta$  into account. In the GGA,  $E_{XC}$  is expressed as;

$$E_{XC}^{GGA}[\rho_\alpha, \rho_\beta] = \int f(\rho_\alpha, \rho_\beta, \nabla \rho_\alpha, \nabla \rho_\beta) dr \quad (2.58)$$

where  $f$  is some function of  $\rho_\alpha$  and  $\rho_\beta$ , and their gradients. In the GGA,  $E_{XC}^{GGA}$  is usually divided into exchange and correlation parts and each part is modelled independently.

In 1988, Becke reported an exchange functional [131] using GGA as;

$$E_X^{B88} = E_X^{LSDA} - b \sum_{\sigma=\alpha,\beta} \int \frac{(\rho_\sigma)^{4/3} \xi_\sigma^2}{1 + 6b \xi_\sigma \sinh^{-1} \xi_\sigma} \quad (2.59)$$

where;

$$\xi_\sigma \equiv \frac{|\nabla \rho_\sigma|}{(\rho_\sigma)^{4/3}} \quad , \quad \sinh^{-1} = \ln \left[ x + (x^2 + 1)^{1/2} \right] \quad , \quad b = 0.0042 \quad (2.60)$$

Since exchange-correlation functionals can be split into exchange and correlation parts, any correlation functional can be used with any exchange functional. For instance, the calculations that are performed with the BLYP functional is a combination of B88 exchange functional (**BLYP**) and Lee-Yang-Parr correlation functional (LYP) [132, 133] (**BLYP**).



### 2.1.7.2 Hybrid functionals

According to the *adiabatic connection method* (ACM), one can connect a non-interacting reference system with the exact system. According to this method,  $E_{XC}$  can be expressed as;

$$E_{XC} = \int_0^1 \left\langle \Psi_\lambda \mid V_{XC}(\lambda) \mid \Psi_\lambda \right\rangle d\lambda \quad (2.61)$$

where  $V_{XC}$  is called **exchange-correlation potential** and defined as;

$$V_{XC}(r) = \frac{\partial E_{XC}[\rho]}{\partial \rho(r)} \quad (2.62)$$

If  $V_{XC}$  in equation 2.61 is linear in  $\lambda$ ,  $E_{XC}$  can be expressed as;

$$E_{XC} \cong \frac{1}{2} \left\langle \Psi_0 \mid V_{XC}(0) \mid \Psi_0 \right\rangle + \frac{1}{2} \left\langle \Psi_1 \mid V_{XC}(1) \mid \Psi_1 \right\rangle \quad (2.63)$$

When  $\lambda$  takes the value 0, it corresponds to a system with non-interacting electrons, so there is only exchange energy. Additionally, the exchange energy is exactly the same that is calculated with Hartree-Fock method when  $\lambda = 0$ . Hence,  $E_{XC}$  can have the form of;

$$E_{XC} = (1 - a)E_{XC}^{DFT} + aE_X^{HF} \quad (2.64)$$

where  $a$  is a constant. Using this concept, Becke established a three parameter functional expression [134];

$$E_{XC}^{B3PW91} = (1 - a)E_X^{LSDA} + aE_X^{HF} + b\Delta E_X^{B88} + E_C^{LSDA} + c\Delta E_C^{PW91} \quad (2.65)$$

where constants  $a = 0.20$ ,  $b = 0.72$ , and  $c = 0.81$ . Parameter  $\Delta E_X^{B88}$  is the gradient correction to LSDA for exchange [131], and parameter  $\Delta E_C^{PW91}$  is the gradient correction for correlation of Perdew and Wang [135–138]. The term  $E_C^{LSDA}$  stands for the correlation functional that has been reported by by Vosko, Wilk, and Nusair (VWN) [139]. In order to use LYP instead of PW91, Stephens et al. adapted exchange-correlation functional [131, 132, 134, 139, 140] that is given in equation 2.65 as;

$$E_{XC}^{B3LYP} = (1 - a)E_X^{LSDA} + aE_X^{HF} + b\Delta E_X^{B88} + (1 - c)E_C^{LSDA} + cE_C^{LYP} \quad (2.66)$$

where constants  $a, b$ , and  $c$  are same with  $E_{XC}^{B3PW91}$ . Functionals that are in the form of equations 2.65 and 2.66 are referred as **hybrid functionals** which are widely used in DFT calculations.

### 2.1.8 Solvation effect

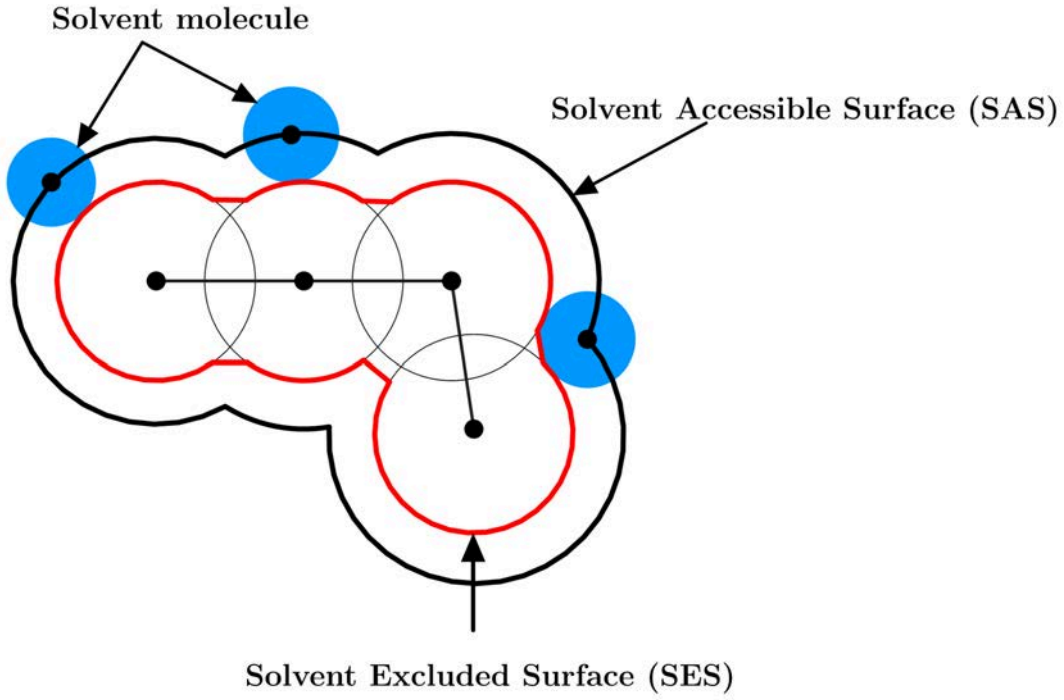
Theoretical modelling can be performed by treating the biomolecules as isolated and non-interacting species at low pressure. However, the accuracy of the calculations in vacuum is controversial since the biomolecules and their reactions are immersed mostly in water. In order to describe the physical nature of the biomolecules accurately, the effect of the solvent should be taken into account.

Methods to describe the effects of solvent are divided into two types; implicit and explicit solvent methods. The explicit solvent methods describe the individual solvent molecules. In implicit solvent methods, which is the main topic of this section, the solvent is treated as a continuum medium.

In implicit solvation techniques, the solute molecules are placed in a cavity in the continuum medium, and the interaction between the solute and the cavity is computed [141]. The calculations of the optimal size and the shape of the cavity have been the key steps in the development of the continuum models. The size of the cavity is related with the definition of the radii near the van der Waals (vdW) values [142]. The mostly used set of vdW radii values are Bondi [143, 144], Pauling [145], and universal force field (UFF) [146, 147]. At this point two different surfaces should be introduced; **solvent accessible surface** (SAS) and **solvent excluded surface** (SES) which are depicted in Figure 2.1. The shape and the size of the cavity defines these two surfaces. The SAS can be defined as the surface of the area around the solute, that can be reached by solvent sphere center. The second surface, SES, is the surface of the area around the solute that cannot be entered by the solvent.

After the creation of reliable SAS and SES surfaces with the definition of cavity, the energy between the solute and the solvent is calculated. The interaction energy that is calculated can be divided into four parts [141];

- Cavitation energy which depends on the solute size.
- Dispersion energy (caused by the solute-solvent dispersion forces).
- Reorganization energy (caused by the distribution of solvent-solvent dispersion forces).
- Electrostatic interaction energy between charges on solvent and solute.



**Figure 2.1** : Solvent accessible surface (SAS) and solvent excluded surface (SES). SAS is represented with bold black line, while SES is represented with red line.

The electrostatic interaction energy is the basic part that is used in the classification of the approaches in solvation methods. The calculation of this energy part uses a simplified version of Poisson equation [142] as a starting point;

$$\begin{aligned}\vec{\nabla}^2 V(\vec{r}) &= 4\pi\rho_M(\vec{r}) \quad \text{within } C \\ \epsilon\vec{\nabla}^2 V(\vec{r}) &= 0 \quad \text{outside } C\end{aligned}\tag{2.67}$$

where  $\rho_M$  is the charge distribution of the solute inside the cavity,  $C$  is the portion of the space that is occupied by the cavity, and  $\epsilon$  is the dielectric constant.  $V$  is the sum of the reaction potential created by the polarization of the dielectric medium ( $V_R$ ) and the electrostatic potential created by  $\rho_M$  ( $V_M$ ). The conditions on the surface of the cavity ( $\Gamma$ ) is defined as;

$$[V] = V_{in} - V_{out} = 0\tag{2.68}$$

$$[\partial V] = \left(\frac{\partial V}{\partial n}\right)_{in} - \epsilon\left(\frac{\partial V}{\partial n}\right)_{out} = 0\tag{2.69}$$

where the term  $\vec{n}$  is the vector perpendicular to the cavity surface. These equations are used to classify the developed approaches for solvation methods, however, only the **apparent surface charge** (ASC) methods [148] will be addressed. In all ASC methods, an apparent surface charge  $\sigma(s)$  is spread on the cavity surface ( $\Gamma$ ), and the

potential over the surface is defined as;

$$V_{\sigma}(\vec{r}) = \int_{\Gamma} \frac{\sigma(\vec{s})}{|\vec{r} - \vec{s}|} d^2s \quad (2.70)$$

### 2.1.8.1 Polarized continuum model

The polarized continuum model (PCM) reported in 1981 by Tomassi group [149] and have been implemented in various quantum mechanical calculation packages. The method has been modified [148], and contributed by many other groups [150–156]

In this method, solutes are described with their own cavity and ASC. The spherical surface is divided into small regions to localize and calculate the surface charges. The model allows the merging of two or more cavities and switching from one cavity to two cavities, which can be referred as association and dissociation [142]. Moreover, the model allows alteration of the ASC model into several ASCs [142]. In PCM, the polarization vector is expressed with the gradient of the total potential;

$$\vec{P}_i(\vec{r}) = -\frac{\epsilon_i - 1}{4\pi} \nabla V(\vec{r}) \quad (2.71)$$

where  $i$  defines the region number. The ASC distribution at the boundary of two regions is described with a unit vector that points from one region to another;

$$\sigma_{ij} = -(\vec{P}_j - \vec{P}_i) \cdot \vec{n}_{ij} \quad (2.72)$$

### 2.1.8.2 Integral equation formalism

The integral equation formalism (IEF), was developed in 1997 by Cancés and Mennucci [157–159]. While the previous continuum solvent approaches were limited to deal with standard isotropic liquids [157], the IEF approach allows the treatment of the media of different nature [157, 158].

The method reduces the dimensionality of the problem to the surface [157]. The IEF approach uses Green functions ( $G(x,y)$ ) to define the potential that is generated in position  $x$  by a unit point charge in position  $y$  [142]. It uses a single ASC procedure and does not need an additional finite element method step [157].

The method has been implemented in the Gaussian program package, and has become the default PCM approach [160].

### 2.1.9 Potential energy surface

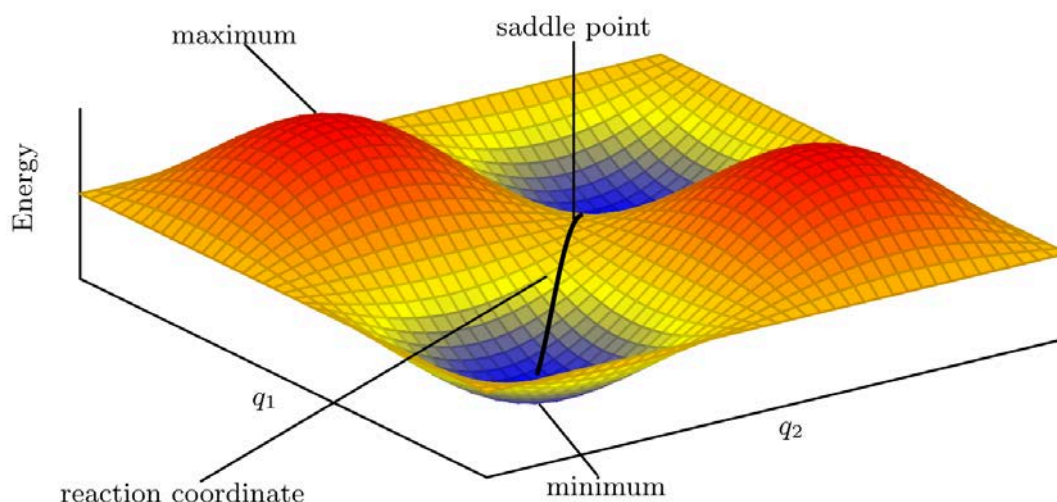
The potential energy of a molecule consisting of  $N$  atoms over all of the possible arrangements of these atoms creates a hypersurface with a dimension of  $3N + 1$ . This multi dimensional surface is called the **potential energy surface** (PES). A simple example of a three dimensional PES (is representative and does not correspond to a real molecular system) is depicted in Figure 2.2. All of the points on this surface are characterized with a certain molecular geometry, and computational chemistry programs are usually able to locate the geometries of the intermediates and the transition states that connect them.

The points on the PES that corresponds to the intermediates and the transition state that links them are called **stationary points**. On these points, the PES is flat;

$$\frac{\partial E}{\partial q_1} = \frac{\partial E}{\partial q_2} = \dots = 0 \quad (2.73)$$

where  $q_i$  are the variables that define the atomic arrangements and  $E = f(q_i)$ .

The points on the PES that corresponds to the intermediates with a finite lifetime are called **minima**. The path that connects these points and transition states is simply called the **reaction coordinate** or the **intrinsic reaction coordinate** (IRC) which is actually a two dimensional pathway from the multidimensional PES. The transition state, that is on this IRC, is the maximum along the pathway but minimum along all other directions. Due to the saddle-like shape of the surface on which the transition states are located, these points on PES are called **first order saddle points**. Since the



**Figure 2.2 :** Minimum, maximum, and saddle points on a simple energy surface that is defined with  $q_1$  and  $q_2$  variables.

PES is a hypersurface, the first derivatives of a PES with respect to the  $3N$  dimensions are manipulated with a **gradient** matrix;

$$\vec{g} = \begin{pmatrix} \left(\frac{\partial E}{\partial q_1}\right) \\ \left(\frac{\partial E}{\partial q_2}\right) \\ \vdots \\ \left(\frac{\partial E}{\partial q_N}\right) \end{pmatrix} \quad (2.74)$$

and the second derivatives are handled with the **Hessian** matrix;

$$\bar{H} = \begin{pmatrix} \frac{\partial^2 E}{\partial q_1 \partial q_1} & \frac{\partial^2 E}{\partial q_1 \partial q_2} & \cdots & \frac{\partial^2 E}{\partial q_1 \partial q_N} \\ \frac{\partial^2 E}{\partial q_2 \partial q_1} & \frac{\partial^2 E}{\partial q_2 \partial q_2} & \cdots & \frac{\partial^2 E}{\partial q_2 \partial q_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 E}{\partial q_N \partial q_1} & \frac{\partial^2 E}{\partial q_N \partial q_2} & \cdots & \frac{\partial^2 E}{\partial q_N \partial q_N} \end{pmatrix} \quad (2.75)$$

The minimum, the maximum, and the saddle points are defined where  $\vec{g}_i = 0$ , but not all such points are minima. If the gradient is zero, the minimum, the maximum, and the saddle points are defined as;

$$\text{Eigenvalues of } \bar{H} \quad \begin{cases} \text{are all positive} & \longrightarrow \text{Minimum} \\ \text{are all negative} & \longrightarrow \text{Maximum} \\ \text{are of both signs} & \longrightarrow \text{Saddle point} \end{cases} \quad (2.76)$$

In order to locate these points on a PES with respect to variables  $q_i$ , one need to perform a **geometry optimization**. The characterization of the minimum points is called **minimization**. Transition states (saddle points) are located via the **transition state optimizations**. In order to find the correct path that connects minima and transition states, **IRC calculations** are performed.

There are different algorithms that can be used to perform a geometry optimization. A Newton-Raphson algorithm can be performed, but the algorithm can cause large step sizes [161]. However, the step size can be controlled with the use of other techniques [162–165]. In the Gaussian09 quantum mechanical program package [160], the default algorithm for both minimizations and transition state optimizations is the Berny algorithm that uses GEDIIS [161]. In order to find correct IRC paths, Gaussian09 [160] performs calculations with an Hessian-based Predictor-Corrector

integrator algorithm [166–168], but different algorithms, such as the first-order Euler integration for predictor step, the damped velocity verlet integrator [169], and the local quadratic approximation [170, 171], are also provided.

## 2.2 Molecular Mechanics

Molecular mechanics (MM), unlike quantum mechanics, does not deal with an electronic Hamiltonian or a wave function. Instead of dealing with electrons, it treats atoms as single particles/balls, and the bonds between atoms are treated as springs. The method computes the energy of a molecular system using classical mechanics.

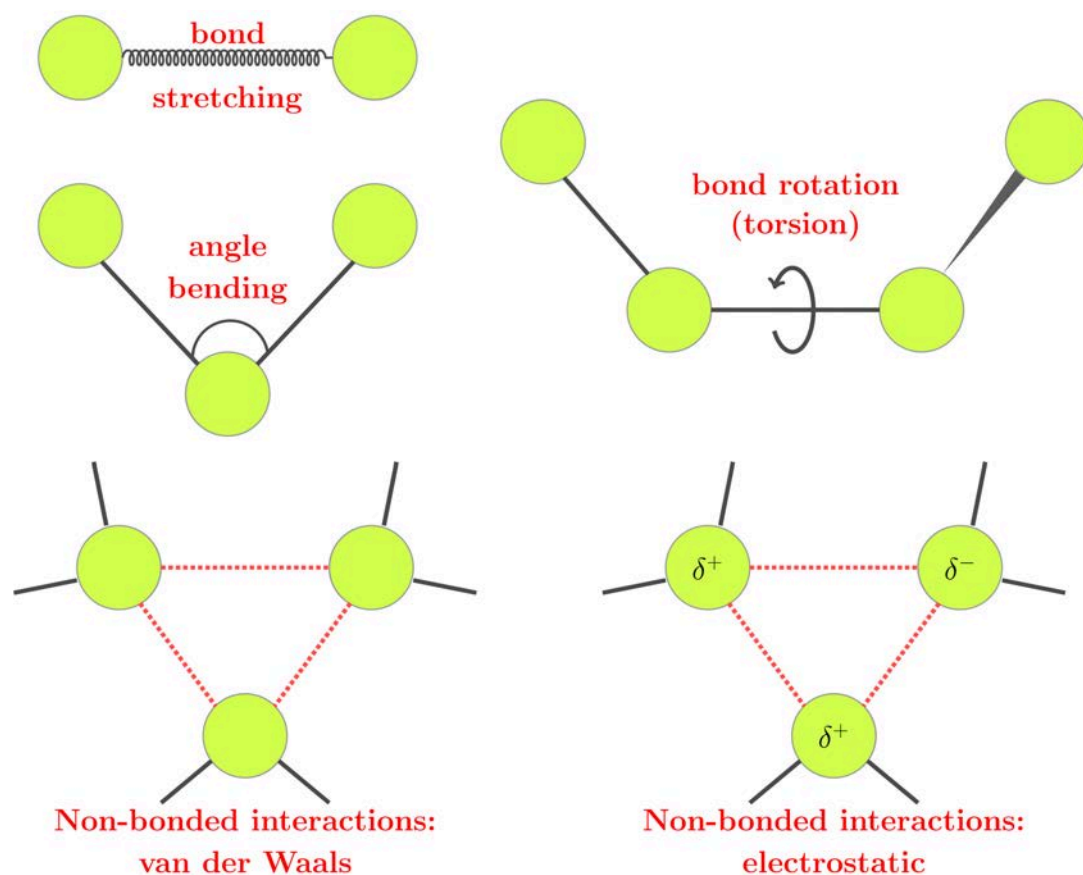
The potential energy in MM is the sum of the energies due to the bonded ( $V_{bonded}$ ) and non-bonded ( $V_{non-bonded}$ ) interactions. A representation of these interactions can be depicted in Figure 2.3. In this case, the energy due to the bonded interactions of a system consists of the addition of the bond stretching energy ( $V_{str}$ ), the energy due to the internal rotation about bonds (torsion) ( $V_{tors}$ ), and the energy due to the bond-angle bending ( $V_{bend}$ ). The energy due to the non-bonded interactions involves van der Waals ( $V_{vdW}$ ) and electrostatic ( $V_{elec}$ ) terms.

$$V = \sum_{bonds} V_{str} + \sum_{angles} V_{bend} + \sum_{dihedrals} V_{tors} + \sum_{pairs} V_{vdW} \sum_{pairs} V_{elec} \quad (2.77)$$

In equation 2.77, the vdW energy can be chosen as Lennard-Jones potential, and electrostatic interactions can be represented as Coulombic interactions.

The mathematical expression to calculate the potential energy is called a **force field**, and a force field can be parametrized using empirical data or data obtained with quantum mechanical calculations. There are various force fields that are parametrized in the literature [172–176]. One of them is the *Assisted Model Building with Energy Refinement* (AMBER) [177], which was developed for proteins, nucleic acids, and many related organic molecules in condensed phases and the potential energy is given by;

$$V = \sum_{bonds} K_{str}(r - r_0)^2 + \sum_{angles} K_{bend}(\theta - \theta_0)^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] \\ + \sum_{pairs} \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \sum_{pairs} \frac{q_i q_j}{\epsilon R_{ij}} \quad (2.78)$$



**Figure 2.3 :** Potential energy in a molecular mechanics calculation is computed with the contributions from atomic motions.

where  $K_{str}$ , and  $K_{bend}$  are the bond and the angle bending force constants respectively. The term  $r$  is the length of the bond when stretched and the term  $r_0$  is the equilibrium (natural) length of the bond. The terms  $\theta$  and  $\theta_0$  are the size of the angle when distorted and the equilibrium (natural) size of the angle respectively. The potential applied to dihedral angle  $\phi$  is represented with  $V_n$  where  $n$  is the multiplicity. The term  $\gamma$  is the phase angle which is usually  $0^\circ$  or  $180^\circ$ . The vdW energy is represented with 6-12 Lennard-Jones potential where  $R_{ij}$  represents the distance between the centers of the non-bonded atom  $i$  and atom  $j$ . The electrostatic interactions are represented with Coulombic interactions where  $q_i$  is the point charge of the atom  $i$  and  $\epsilon$  is the dielectric constant of the medium.

Since MM calculations are fast, they can be applied to large molecular systems unlike quantum mechanical calculations. However, electronic properties cannot be obtained with MM calculations since it ignores electrons.



## 2.3 Molecular Dynamics

Molecular dynamics (MD) is a technique for computing the evolution of the atomic positions of a system along time using the Newton's equations of motion with some initial conditions. MD is usually used to simulate a large system which leads to a large phase space. In order to obtain an representative average property of the system, all of the configurations within the phase space should be sampled during the simulations. By doing so, the time averages of an observable can be calculated, but it does not strictly ensures the experimental observable which is called the ensemble average. At this point the statistical mechanics provide the ergodic hypothesis, which states that the complete phase space can be explored with the simulations along an infinite time. Since the system will eventually pass through all the possible configurations, the time and ensemble averages will become equal to each other. Thus, the goal of a MD simulation is to generate enough representative configurations of a given system to obtain the equality between the time and the ensemble averages.

### 2.3.1 Integration of Newton's equations of motion

In a MD simulation, the trajectory of the particles (the evolution of atomic positions along time) in space is studied with the use of Newton's second law. In order to obtain the force ( $\vec{F}$ ) acting on a particle, the potential energy ( $U$ ) is differentiated with respect to the position ( $\vec{F} = \frac{\partial U}{\partial \vec{r}_i}$ ). Once the forces are calculated, Newton's equations of motion are integrated forward in time.

In models with intermolecular interactions, the force on each particle will change with the alteration of the position of the corresponding particle, and the position of any neighbour particle. Hence, the motions of all the particles will be coupled together and give rise to a many body problem. Under these circumstances, analytical solution of Newton's equations of motions will be not possible. Therefore, the equations are discretized and solved numerically using a finite difference method.

The most direct finite difference method for the integration is the **Verlet algorithm** [178]. This algorithm is based on truncated Taylor expansion of the coordinates of the particle. The Verlet algorithm uses the position and the acceleration of a particle at time  $t$ , and the position from the previous step ( $t - \Delta t$ ) to compute the new position at

next step  $(t + \Delta t)$ ;

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}\Delta t^2 a(t) + \dots \quad (2.79)$$

$$r(t - \Delta t) = r(t) - v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 - \dots \quad (2.80)$$

where  $a$  is the acceleration. The addition of the equation 2.79 and equation 2.80 yields;

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + a(t)\Delta t^2 \quad (2.81)$$

Even though the implementation of Verlet algorithm is straightforward, the possible loss of precision and the lack of velocity term in the equations makes it disadvantageous. The simplest algorithm that is more sophisticated than the Verlet scheme is **Leap-Frog algorithm** [179] which evaluates the velocities at half integer time step and then uses the velocities to compute the new positions.

$$r(t + \Delta t) = r(t) + v\left(t + \frac{1}{2}\Delta t\right)\Delta t \quad (2.82)$$

$$v\left(t + \frac{1}{2}\Delta t\right) = v\left(t - \frac{1}{2}\Delta t\right) + a(t)\Delta t \quad (2.83)$$

The new positions are computed based on the old position and velocities:

$$r(t + \Delta t) = r(t) + \Delta t v\left(t + \frac{1}{2}\Delta t\right) \quad (2.84)$$

Finally, the following expression is used to obtain velocities at time  $t$ :

$$v(t) = \frac{1}{2}\left[v\left(t + \frac{1}{2}\Delta t\right) + v\left(t - \frac{1}{2}\Delta t\right)\right] \quad (2.85)$$

Even though the loss of precision in the Verlet algorithm is prevented with Leap-Frog algorithm, the positions and velocities are not synchronized. Therefore, the calculation of kinetic energy contribution to the total energy is not performed at the same time as the positions are computed.

Another reformulation of Verlet algorithm that explicitly uses the velocity was developed in 1982 and is called the **velocity Verlet algorithm** [180]. In the first step of this algorithm, the positions at time  $t + \Delta t$  are computed using the velocities and accelerations at time  $t$ ;

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 \quad (2.86)$$

while the velocities at time  $t + \frac{1}{2}\Delta t$  are computed with the following expression;

$$v\left(t + \frac{1}{2}\Delta t\right) = v(t) + \frac{1}{2}a(t)\Delta t \quad (2.87)$$

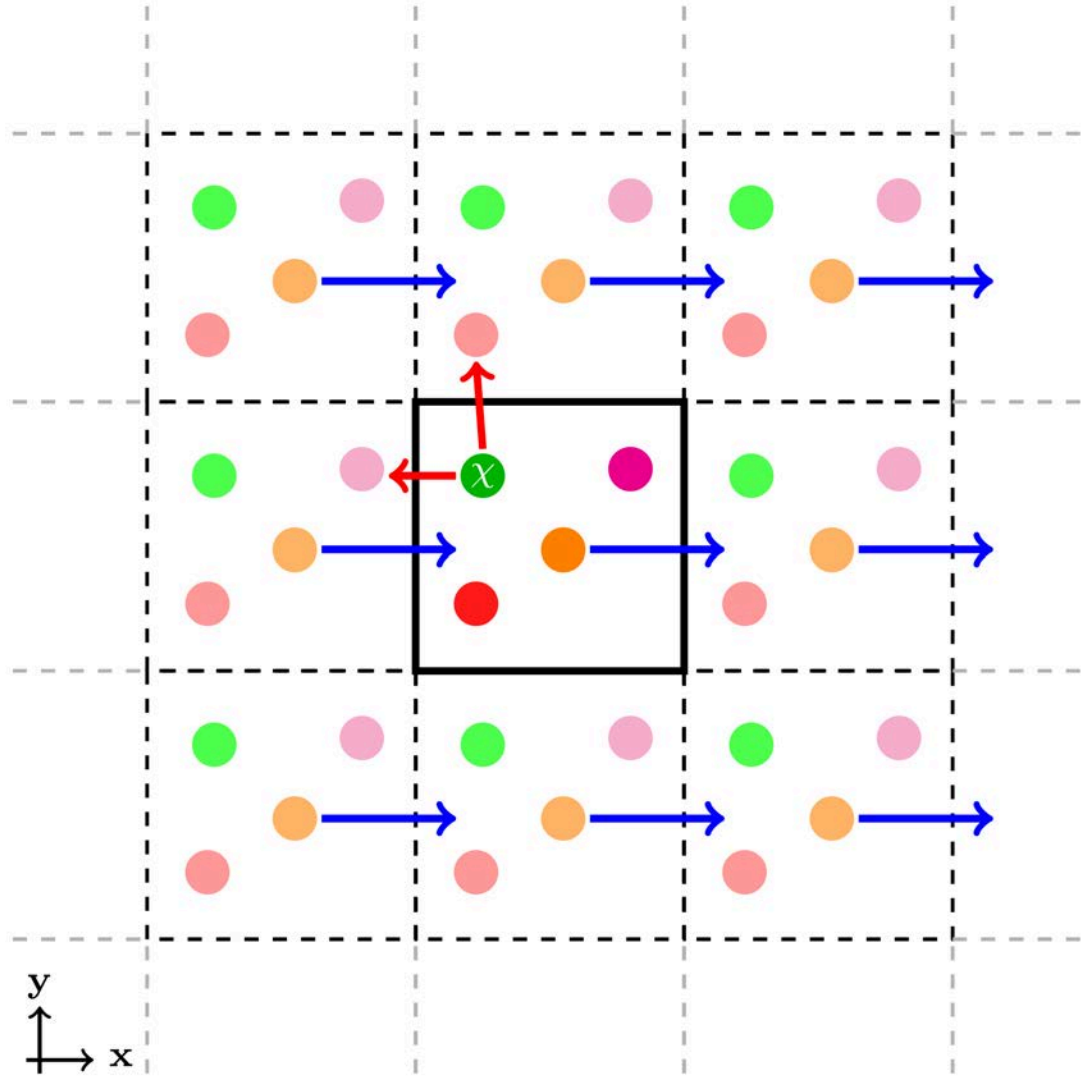
The force is determined from the newly computed positions, which gives the acceleration at time  $t + \Delta t$ , and finally the velocities at time  $t + \Delta t$  is computed with;

$$v(t + \Delta t) = v\left(t + \frac{1}{2}\Delta t\right) + \frac{1}{2}a(t + \Delta t)\Delta t \quad (2.88)$$

The most important point in the choice of the integrator is that it should conserve energy, permit long time step ( $\Delta t$ ), and be time reversible. The ideal integrator should be fast and demand minimum memory. However, the time that is consumed with the integrators are trivial relative to the other parts of the simulation (i.e. the calculation of the forces). The total time of a simulation mainly depends on the time that is consumed by the computation of the energy and the gradient. Therefore, the choice of the time step significantly effects the computational demand since in each iteration the energy and the gradient is calculated. If the time step is very small, the computational demand would considerably increase to simulate the same total time. Furthermore, the larger time steps to simulate the same total time can cause instability of the simulation (i.e. energy drift). The optimum time step should be about the smallest vibrational period in the system divided by two. In a typical system the fastest periodic motion belongs to the vibration of hydrogen atoms and the choice of time step should not exceed 1 fs. However, the time step can be increased with the use of other algorithms that allows to freeze the bonds involving hydrogen atoms, such as SHAKE [181, 182].

### 2.3.2 Periodic boundary conditions

The number of particles of a system in a MD simulation has a finite value that is far less than the Avogadro number. However, bulk systems can be considered as infinite at atomistic scale. In order to eliminate the border effects, a representative box of a model system can be used, considered as periodic, and replicated throughout the space to form an infinite lattice which is similar to a three-dimensional crystal. The replication of the box will create periodic copies of the particles that are called images. This procedure is called **periodic boundary conditions** (PBC) and a two dimensional representation of such a periodic system is depicted in Figure 2.4.



**Figure 2.4** : A two dimensional representation of a simulation box with PBC. The unit cell is represented as the central box with thick black square. The interactions of particle  $\chi$  with the images in the neighbour cells are depicted with red arrows. The movement of a particle from the central box, and the movement of its images in replicated boxes are represented with blue arrows.

If a particle within the central box moves during the simulation, its images within every periodic box moves with the same way. Therefore, if a particle leaves the central box, it will be replaced by its image (blue arrows in Figure 2.4). The PBC are applied using the **minimum image convention** that extend the system boundaries to the infinity. When applied, the minimum distance between two particles are calculated, and the particles on the edges of the box are allowed to interact with an image particle in the neighbour cell (red arrows in Figure 2.4).

### 2.3.3 Short and long range interactions

The size of the systems of interest increases with the increment of the computational power. However, computing all of the non-bonded pairwise potentials would drastically increase the computational effort of the MD simulations. Due to the rapid decay of the non-bonded pairwise potentials with large distances, the computation of all pair interactions can be avoided. There are several approaches that can be applied to reduce the computational effort, and respect to the long-range interactions, but only Ewald summation will be addressed in this section.

#### 2.3.3.1 Ewald summation

Let us assume a system containing  $N$  ions at locations  $r_1, r_2, \dots, r_N$  with point charges  $q_1, q_2, \dots, q_N$  in a cubic box with length of  $L$ . The total Coulomb interaction energy of such a system with  $N(N-1)/2$  ion pairs can be given by;

$$E_{coul} = \frac{1}{4\pi\epsilon_0} \sum_{(i,j)} \frac{q_i q_j}{|r_{ij}|} \quad (2.89)$$

where  $r_{ij} = r_i - r_j$ ,  $\epsilon_0$  is permittivity constant and the sum is performed over ion pairs  $(i, j)$ . If PBC are applied to this system, the ions in the central box will also interact with the particles in the replicated images. Describing a lattice vector  $\vec{n}$  that points from the central box to a periodic image that has the components of  $\vec{n}_x$ ,  $\vec{n}_y$ , and  $\vec{n}_z$  ( $\vec{n} = \pm\vec{n}_x\vec{L}, \pm\vec{n}_y\vec{L}, \pm\vec{n}_z\vec{L}$ ) where  $\vec{n}_x^2 + \vec{n}_y^2 + \vec{n}_z^2 \geq 1$ , the total Coulomb interaction energy evolves to;

$$E_{coul} = \frac{1}{4\pi\epsilon_0} \frac{1}{2} \sum_n \sum_i^N \sum_j^N{}' \frac{q_i q_j}{|r_{ij} + \vec{n}\vec{L}|} \quad (2.90)$$

where prime denotes the exclusion of the term  $j = i$  if  $\vec{n} = 0$ . However, the sum that is given in equation 2.90 converges very slowly and in fact it is conditionally convergent. In order to overcome this problem, Ewald introduced a method that evaluates the energy by transforming it into different summations [183], and expressed the Coulomb interaction energy in the following form;

$$\begin{aligned}
E_{coul} &= E^S + E^L - E^{self} \\
&= \frac{1}{4\pi\epsilon_0} \frac{1}{2} \sum_n \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{|r_i - r_j + nL|} \operatorname{erfc}\left(\frac{|r_i - r_j + nL|}{\sqrt{2}\sigma}\right) \\
&\quad + \frac{1}{2V\epsilon_0} \sum_{k \neq 0} \frac{e^{-\sigma^2 k^2/2}}{k^2} |\rho(k)|^2 \\
&\quad - \frac{1}{4\pi\epsilon_0} \frac{1}{\sqrt{2\pi}\sigma} \sum_{i=1}^N q_i^2
\end{aligned} \tag{2.91}$$

where  $\operatorname{erfc}$  is called complementary error function and  $\operatorname{erfc}(r) \equiv 1 - \frac{2}{\sqrt{\pi}} \int_0^r e^{-t^2} dt$ .

The Ewald summation becomes extremely demanding by means of computational power when the system size is large. Even if a fixed cutoff is chosen, Ewald summation is computationally intensive for large systems because the calculation of the  $E^L$  (reciprocal space part) scales as  $\mathcal{O}(N^2)$ . Since the bottleneck is caused by the reciprocal space part of the Ewald sum, several approaches have been reported to handle this part more efficiently. These techniques are called mesh based approaches in which the point charges are distributed to grid points. The mesh based techniques take advantage of Fast Fourier transforms (FFT) which are efficient on lattices. The earliest mesh based approach, Particle-Particle Particle-Mesh method (PPPM), was reported by Hockney and Eastwood in 1981 [184, 185]. Darden et. al. introduced particle-mesh Ewald (PME) approach after a decade and reported that it scales as  $\mathcal{O}(N \cdot \log(N))$  [186]. Two years later a smooth particle mesh Ewald (SPME) method was reported in which the PME method was reformulated with the use of B-spline interpolation [187].

### 2.3.4 Thermodynamic ensembles

An ensemble is a collection of the states of a system each of which have one or more common macroscopic property. It can be also defined as the probability distribution of the states of a system. The dynamic properties of a system is given with averages over an ensemble.

Even though different types of ensembles can be created depending on the thermodynamic properties, there are three major types of ensembles; **microcanonical ensemble**, **canonical ensemble**, and **isobaric-isothermal ensemble**.

- **Microcanonical ensemble (NVE):** This ensemble corresponds to an isolated system. In this ensemble a thermodynamic state is characterized by a fixed number of atoms ( $N$ ), a fixed volume ( $V$ ), and a fixed energy ( $E$ ).
- **Canonical Ensemble (NVT):** This ensemble corresponds to thermodynamic state that is characterized by a fixed number of atoms ( $N$ ), a fixed volume ( $V$ ), and a fixed temperature ( $T$ ). Each member of the ensemble represents the same closed, isothermal system.
- **Isobaric-Isothermal Ensemble (NPT):** The thermodynamic state of this ensemble is characterized by a fixed number of atoms ( $N$ ), a fixed pressure ( $P$ ), and a fixed temperature ( $T$ ).

### 2.3.5 Temperature coupling

Even though MD simulations rely on Newton's equations of motion and by default it samples microcanonical ensemble, a canonical ensemble is often desirable for compatibility with the experiments. In order to maintain the temperature constant, a basic MD scheme can be modified with the use of thermostat algorithms. These algorithms ensures a constant temperature along time by using the relationship between the kinetic energy ( $E_K$ ), and the instantaneous temperature ( $T$ );

$$E_K = \frac{N_{df}}{2} k_B T = \frac{1}{2} \sum_i^{N_{df}} \frac{p_i^2}{m_i} \quad (2.92)$$

where  $N_{df}$  is the number of degrees of freedom,  $k_B$  is the Boltzmann's constant,  $p_i$  is the linear momentum, and  $m_i$  is the mass of the particle  $i$ . The most direct approach is called **velocity rescaling** in which the velocities are rescaled at each step which will cause the kinetic energy to give the desired temperature. However, in this approach the correct energy fluctuations cannot be captured, thus it cannot generate the correct thermodynamic properties of canonical ensemble.

In 1980, Andersen introduced a method [188] in which the system is coupled with an imaginary heat bath to impose desired temperature. In this method, some stochastic impulsive forces act occasionally upon randomly selected particles, which cause collisions with the heat bath. The method designate the velocity of the randomly chosen particle using a Maxwell-Boltzman distribution at the desired temperature. The

collision time step of a randomly selected particle is defined with collision frequency  $\nu$ .

In 1984, Berendsen proposed a method that is similar to velocity rescaling [189]. However, instead of scaling the velocities at each time step, the method uses a time scale for the updating of the velocities. In this thermostat the coupling is weak and the scaling factor is defined as;

$$\gamma^2 = 1 + \frac{\Delta T}{\tau} \left( \frac{T_{i-1}}{T_{i+1}} - 1 \right) \quad (2.93)$$

where  $T_{i+1}$  and  $T_{i-1}$  are instantaneous temperature before and after the velocity scale,  $\Delta T = T_{i+1} - T_{i-1}$ , and  $\tau$  is a parameter.

## 2.4 Semiempirical Born-Oppenheimer Molecular Dynamics (SEBOMD)

SEBOMD (SemiEmpirical Born-Oppenheimer Molecular Dynamics) is a method which couples a semiempirical program DivCon99 [190] with a classical molecular dynamics software. It is based on Born-Oppenheimer approximation and uses a QM semiempirical NDDO Hamiltonian (AM1, PM3, etc.) as a force field to perform molecular dynamics [83].

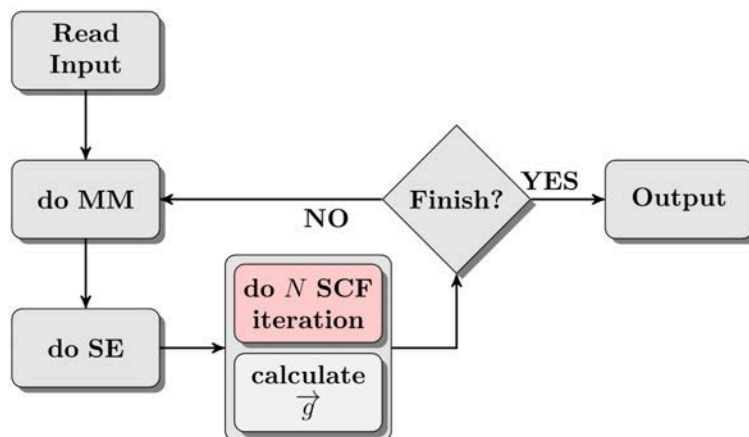
SEBOMD capabilities, with several changes from the original version of DivCon99, have been implemented in the AMBER suite of software [191] as the `sebomd` module that uses the linear scaling Divide & Conquer algorithm (D&C). Today, this permits MD simulations of molecular systems containing several hundreds of atoms for a few nanoseconds using a single workstation. SEBOMD has been successfully applied to the simulations of molecules in the condensed phase, including liquid water and N-methyl acetamide in water [82, 83].

### 2.4.1 Simulations

A SEBOMD simulation consists of two parts; a classical part and a semiempirical part. A schematic representation of a SEBOMD simulation is represented in Figure 2.5. The classical part of a SEBOMD simulation includes running the MD trajectory. MD runs can be performed using periodic boundary conditions, NVE, NVT or NPT thermodynamical ensembles, and long-range electrostatic interactions through Ewald or Particle Mesh Ewald summation. The semiempirical part of the SEBOMD method



consists in computing the energy and the gradient of a molecular system at the semiempirical quantum mechanical level of theory at each step of the MD trajectory. The most time consuming part of a SEBOMD simulation is the semiempirical part in which the SCF iterations are performed and the gradient is calculated.



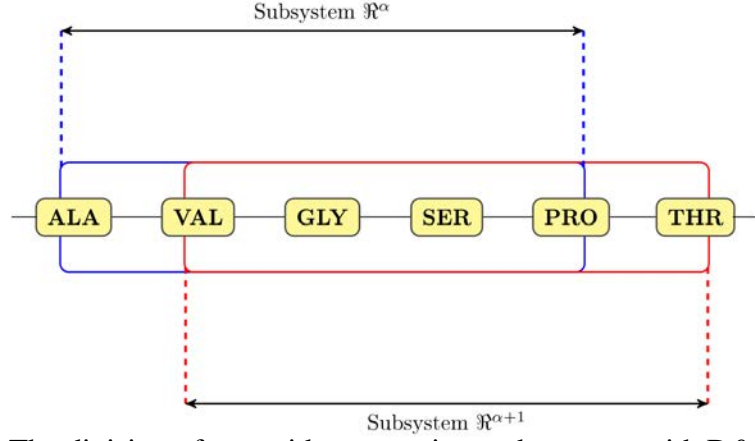
**Figure 2.5** : Schematic representation of a SEBOMD simulation.

#### 2.4.2 Divide and conquer (D&C)

A SCF cycle involves the construction and solution of the HF secular equation given in equation 2.29. However, the computational expense grows with the third power of the number of basis functions. The larger the system gets, the more CPU intensive is the solution of the diagonalization process. In order to overcome this bottleneck, Divide and Conquer (D&C) method divides the original system into a set of smaller subsystems ( $\mathfrak{R}^\alpha$ ,  $\alpha = 1, 2, \dots, N_{\text{subsystem}}$ ) that overlap with each other (see Figure 2.6). The D&C is based on the idea of the localization of LCAO-MO chemical bonding. The electronic information of each subsystem  $\mathfrak{R}^\alpha$  is computed separately, which are then combined to obtain the electronic description of the overall system. The Roothaan-Hall equations given in equation 2.37 becomes;

$$F^\alpha C^\alpha = C^\alpha E^\alpha \quad (2.94)$$

where the elements of  $F^\alpha$  are obtained from global Fock matrix elements. Since the number of basis functions will be decreased to  $N^\alpha$  ( $N^\alpha < N$ ) in the D&C algorithm, the computational expense will be lesser by a factor of  $(N^\alpha/N)^3$ . Thus, the diagonalization procedure of a set of subsystems would become less expensive relative to the diagonalization of the global Fock matrix if the system is large enough.



**Figure 2.6 :** The division of a peptide system into subsystems with D&C Algorithm.

In D&C algorithm, the most important part is the elimination of the truncation effects [84, 85]. These effects arise near the subsystem boundaries and can be addressed by applying the partitioning functions ( $D_{\mu\nu}^\alpha$ ) in order to weight the contributions of the subsystem to the global entries of the density matrix [85];

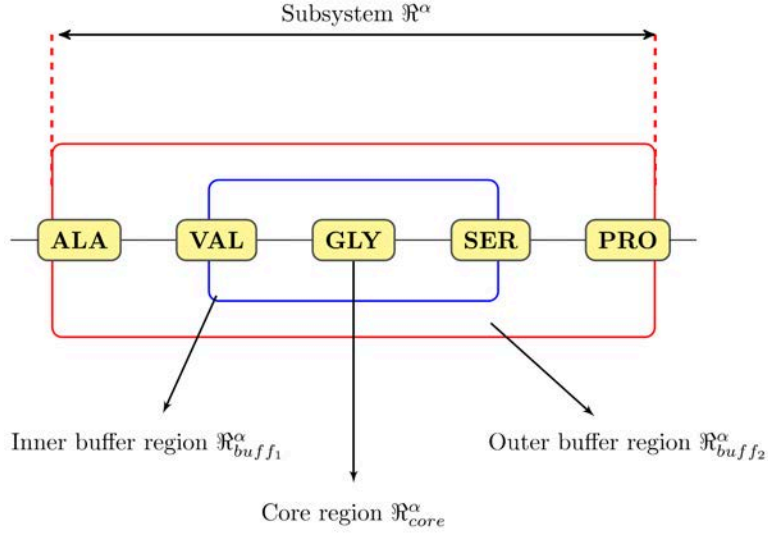
$$P_{\mu\nu} = \sum_{\alpha=1}^{N_{sub}} D_{\mu\nu}^\alpha P_{\mu\nu}^\alpha \quad (2.95)$$

The partitioning function given in equation 2.95, would be equal to zero if  $\chi_\mu$  or  $\chi_\nu$  is not in the subsystem, and if it constitutes an edge contribution which is most likely to be contaminated by truncation effects. Otherwise, the value of  $D_{\mu\nu}^\alpha$  would be equal to  $1/n_{\mu\nu}$  where  $n_{\mu\nu}$  is the total number of subsystems that overlaps. The contribution of the edges can be defined by the subdivision of the subsystems and are expressed with the definition of the buffer regions [85]. The subdivision of a subsystem into buffer and core regions is represented in Figure 2.7. The utilization of core and buffer regions expand the definition of partition function  $D_{\mu\nu}^\alpha$  to;

$$D_{\mu\nu}^\alpha = \begin{cases} 0, & \text{if } \chi_\mu \notin \mathcal{R}^\alpha \text{ or } \chi_\nu \notin \mathcal{R}^\alpha \\ 0, & \text{if } \chi_\mu \in \mathcal{R}_{buff_2}^\alpha \text{ or } \chi_\nu \in \mathcal{R}_{buff_2}^\alpha \\ 0, & \text{if } \chi_\mu \in \mathcal{R}_{buff_1}^\alpha \text{ and } \chi_\nu \in \mathcal{R}_{buff_2}^\alpha \\ 1/n_{\mu\nu}, & \text{otherwise} \end{cases} \quad (2.96)$$

Even though the partitioning scheme is useful to manage the quality of the information, the simple closed shell formula given in equation 2.36 is not applicable for the construction of  $P^\alpha$  [84, 85]. This issue is addressed by replacing the formula given in equation 2.36 with the equation proposed by Yang [192];

$$P_{\mu\nu}^\alpha = \sum_i^{occupied} n_i^\alpha (c_{i\mu}^\alpha)^* (c_{i\nu}^\alpha) \quad (2.97)$$



**Figure 2.7 :** The subdivision of a subsystem into core and buffer regions with D&C Algorithm.

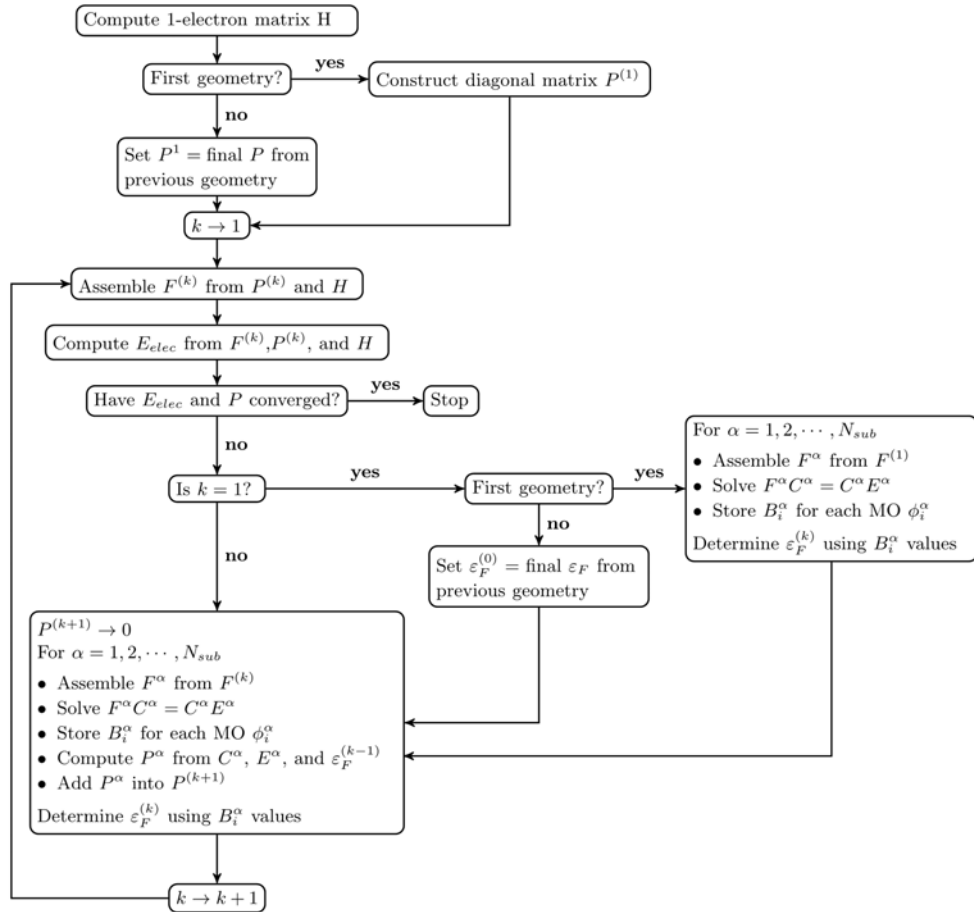
where the term  $n_i^\alpha$  is the occupation number that is defined by the MO energy  $\varepsilon_i^\alpha$  and a common Fermi energy  $\varepsilon_F$  [84,85];

$$n_i^\alpha = \frac{2}{1 + \exp[(\varepsilon_i^\alpha - \varepsilon_F)/kT]} \quad (2.98)$$

where  $k$  is the Boltzman constant and  $T$  is an absolute temperature. The value of the Fermi energy  $\varepsilon_F$  is adjusted iteratively so that the global density matrix gives the correct number of electrons where only the diagonal elements of density matrix are directly relevant. Thus, the total number of electrons are computed with;

$$n_{elec} = \sum_{\alpha=1}^{N_{sub\ occ.}} \sum_i n_i^\alpha \left[ \sum_{\mu=1}^N D_{\mu\mu}^\alpha |c_{i\mu}^\alpha|^2 \right] \equiv \sum_{\alpha=1}^{N_{sub\ occ.}} \sum_i n_i^\alpha B_i^\alpha \quad (2.99)$$

where the values of the term  $B_i^\alpha$  are calculated only once and stored for all subsystems. Using the conditions that are described above and the equations given in 2.95 - 2.99, an iterative SCF procedure is used to obtain the D&C energy which is depicted in Figure 2.8 [85]. The iterative procedure that is given in Figure 2.8 is applied to obtain the D&C energy for a single geometric conformation [85]. The initial density matrix ( $P^{(1)}$ ) is diagonal and this form is used only for the first conformation within the geometry optimization calculations. The density matrix of the subsystem  $\alpha$  ( $P^\alpha$ ) is constructed with the use of the adjusted Fermi energy from the previous cycle. The contribution of the subsystems are then added to the global density matrix ( $P^{(k+1)}$ ) that is going to be used in the next iteration.

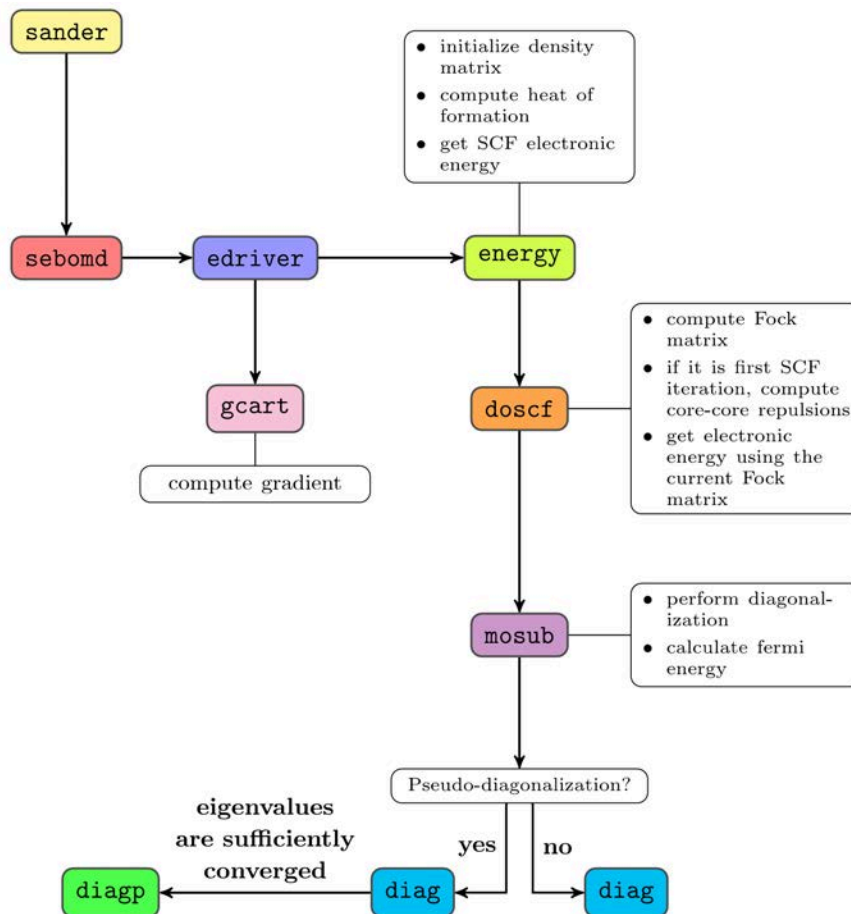


**Figure 2.8** : The SCF procedure with D&C algorithm to obtain the D&C energy of a single geometry conformation [85].

### 2.4.3 The `sebond` module

In AMBER, a SEBOMD simulation is performed with the `sander` MD program which directs the calculations by calling the `sebond`. A simple scheme of the calls to the modules and routines during a simulation is depicted in Figure 2.9. The `sebond` module consists many other subprograms and modules. In order to compute the Cartesian energy gradient, the `gcart` subroutine is used and the computation is performed analytically after the computation of the density matrix obtained from an SCF (Self-Consistent Field) procedure. During the SCF procedure, the Fock matrix  $F$  is diagonalized within `mosub` using the `diag` subroutine when a full SCF procedure is chosen. The `diag` subroutine calls other subroutines to perform tri-diagonalization on the Fock matrix and then to calculate both its eigenvectors and eigenvalues. On the other hand, when pseudo-diagonalization is turned on, the `diagp` subroutine is called to pseudo-diagonalize the Fock matrix. The `diagp` routine is only used when eigenvalues in SCF calculation have sufficiently converged. To do so, the

first steps of the SCF procedure make use of a diagonalizer solver (`diag`) then the pseudo-diagonalization is turned on.



**Figure 2.9 :** A simple call scheme within a SEBOMD simulation using `sander` program.

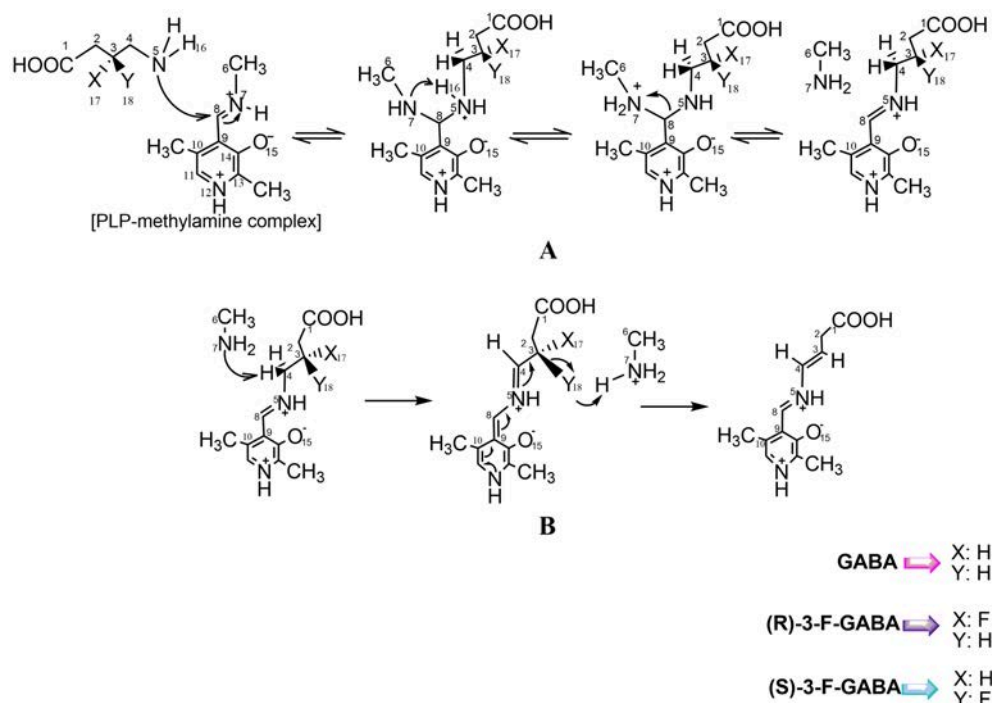
### 3. RESULTS

#### 3.1 Mechanistic Details of Transimination and HF Elimination Reactions

The  $\beta$ -fluoroamines are commonly used as substrate analogues to determine the mechanistic details of enzymatic reactions. Presence of fluorine atom gives rise to the alterations in the electronic profile and the pKa of molecules which results in mechanistic deviations. The fluorine-substituted mechanism-based substrate analogues are widely used in the inactivation of pyridoxal 5'-phosphate (PLP)-dependent enzymes. The presence of fluorine atom also alters the sequence of reactions taking place in PLP-dependent enzymes where the HF elimination reaction appears in between the transimination and inactivation reactions [28]. Despite the amount of the works on  $\beta$ -fluoroamines, the effect of stereoelectronic differences on the transimination and HF elimination reactions taking place in PLP-dependent enzymes has not been investigated yet. Herein a density functional theory study is conducted with a PLP-dependent enzyme GABA-AT to understand the effect of the fluorine atom, 4-amino-3-fluorobutanoic acid (3-F-GABA) enantiomers are chosen to be investigated besides 4-aminobutanoic acid (GABA), which is the natural substrate for  $\gamma$ -aminobutyric acid aminotransferase (GABA-AT). The investigated  $\beta$ -fluoroamines are the experimentally proposed potential inhibitors of PLP-dependent enzyme GABA-AT.

In a combined experimental and docking study conducted by Clift et al., the HF elimination and the transimination reactions were investigated based on the binding affinities and relative efficiencies of the GABA and its fluorinated analogues in the presence of PLP-dependent enzyme GABA-AT [28]. It was stated that solution and bioactive conformers of GABA and its fluorinated analogues did not possess the same conformation [28]. Thus, calculations are carried out both for solution and bioactive conformers. Following the conformational analysis of zwitterionic species, the transimination reaction is modelled with the solution and bioactive conformations

to discuss the binding efficiencies of the GABA analogues (Figure 3.1). Then the HF elimination is modelled with the products of the transimination reaction to investigate the effect of the electrostatic interactions and the stereoelectronic differences on the reaction path.



**Figure 3.1** : Reaction pathway of (A) transimination with GABA and fluorine-substituted GABA analogues. (B) HF elimination from fluorine substituted GABA analogues via E1cb mechanism.

### 3.1.1 Computational details

The presence of charged species along the reaction path makes it necessary to perform calculations in aqueous medium. Both the geometry optimizations and the frequency calculations in water environment are performed with the B3LYP [131, 132, 134, 139, 140] method and the 6-31+G(d,p) basis set. The study is conducted using Gaussian'09(G09) program package [160]. The solvation effect of water is considered by using the polarizable continuum model (PCM) of the Tomasi group as implemented in G09 [157, 158, 193, 194]. The energies that are reported in the text are the thermal free energies obtained with solvent optimization calculations using the G09 program package, unless otherwise stated. Moreover, natural bond orbital (NBO) [195] analysis is performed to figure out the importance of stereoelectronic effects on the stability of the structures along the reaction path and the charges reported in the text are based on these NBO calculations. The reactions of GABA and fluorine

substituted GABA analogues are studied with a model system where Lys329 residue is represented with a methylamine. This is a widely used and accepted representations in literature [33,34,196–198]. To reduce the charge delocalization on the pyridine moiety, the nitrogen of the pyridine ring of PLP is modelled as protonated [33, 199, 200]. On the other hand, the phosphate group of PLP is replaced with a methyl group which is another common representation [2, 4, 33, 199, 201]. The kinetics involved in the R:S binding ratio have been calculated according to the activated complex theory (Boltzman distribution) which describes the relative population at two different states

$$\frac{N_1}{N_2} = \frac{e^{-\Delta G_1/RT}}{e^{-\Delta G_2/RT}} \quad (3.1)$$

It has been reported that full geometry optimization calculations with higher level of quantum mechanical theories would not change the conformational properties where the energy values may change with the single point energy calculations [202–204]. Therefore, single point MP2 energy calculations with a larger basis set, 6-311+G(d,p), are performed for all species along the reaction coordinate [205]. The graphical representations were obtained using CYLview program [206]. A systematic labelling is applied for the species along the reaction coordinate. Structures that belong to the solution conformers are labelled with lower-case **a** where the bioactive conformers are depicted with lower-case **b**. In addition, the structures that are in their zwitterionic forms are labelled with "-z" and in the presence of water assistance **\_w** is used.

### 3.1.2 Results and Discussion

#### 3.1.2.1 Solution and bioactive conformations as zwitterions

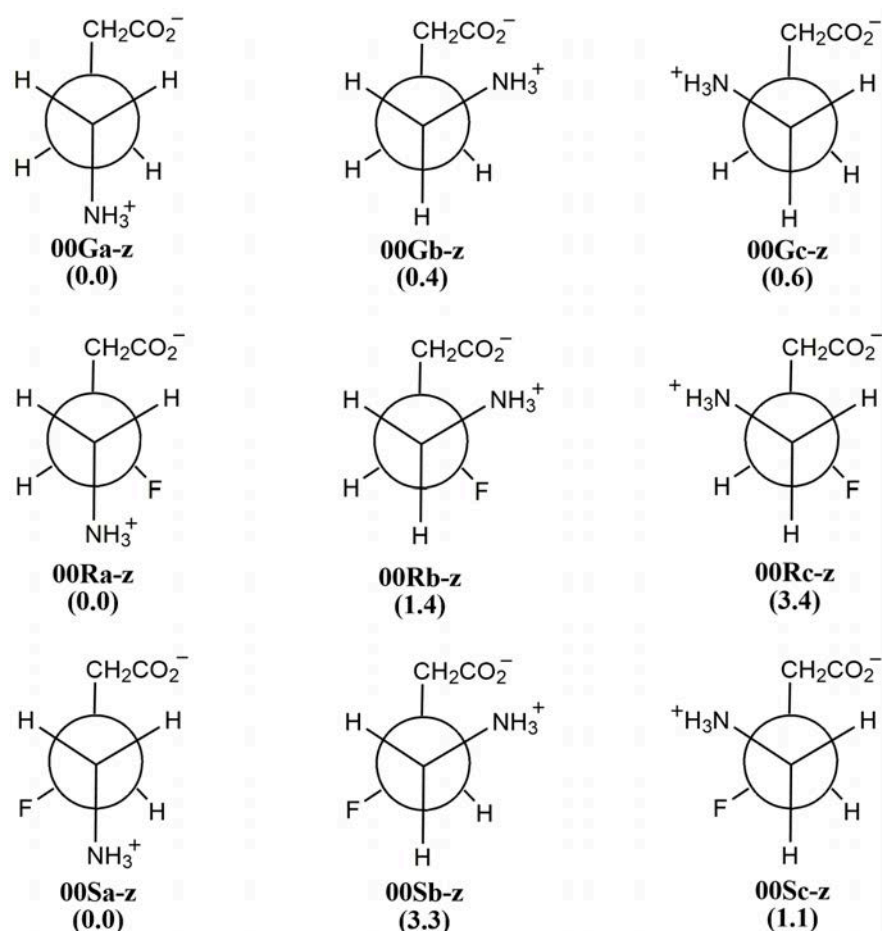
In  $\beta$ -fluoroamines, there is a stereoelectronic preference for the C-F and C-N bonds to align gauche in order to maximize the stabilizing overlap between the  $\sigma_{C-F}^*$  orbital and the vicinal  $\sigma_{C-H}$  orbital, which is known as stereoelectronic gauche effect [18, 30, 196, 207–209]. In an experimental study on the fluorinated substrate analogues of GABA-AT, a gauche alignment was proposed for the C-F and  $C-NH_3^+$  bonds for the binding conformation of (R)-3-F-GABA where an anti-alignment was suggested in the binding conformation of (S)-3-F-GABA [28]. The relative free energies of the conformers of GABA, (R)-3-F-GABA and (S)-3-F-GABA and the Newman projections of the corresponding conformations are given in Figure 3.2. It is



found that the stability of both gauche and anti-conformers of GABA (00Ga-z, 00Gb-z and 00Gc-z, Figure 3.2) are almost the same, which is consistent with previous works on GABA [19] and ethylimines [18]. Small energy differences between enantiomeric pairs of structures were not considered significant, and merely reflect slight differences in the torsion adopted by the  $CH_2COO$  group during optimization. Among the gauche conformations of (R)-3-F-GABA, solution conformer (00Ra-z) is lower in energy relative to the bioactive conformer (00Rb-z). The energy difference (1.4 kcal/mol) may stem from the stronger interaction between the imine hydrogen and the fluorine lone pairs in 00Ra-z, where similar results were reported before for  $\beta$ -fluoroamines and  $\beta$ -fluoroalcohols [18, 210–212]. The anti-conformer (00Rc-z) is found to be the least stable conformer. In case of S enantiomer, both the gauche conformers (00Sa-z and 00Sc-z) are found to be more stable than the anti-conformer (00Sb-z), which is known as the bioactive conformer. An energy difference on behalf of gauche preference is absolutely observed in all cases which is consistent with the stereoelectronic gauche effect [18, 19, 30, 196, 207–209]. However the bioactive conformers are not located as the global minimum structures for the fluorine substituted GABA analogues. Based on the relative free energies, the relative population of 00Rb-z will be higher than that of 00Sb-z, which may promote the inhibition efficiency of R. Experimental studies have also pointed out that high concentrations of substituted GABA analogues were required to inhibit the GABA transamination [28]. The energy differences found between conformations of the 3-F-GABA relative to the equally populated solution and bioactive conformers of GABA might be the possible reason for the required high concentrations during the inhibition. It is stated that substrates that are in zwitterionic form should undergo a deprotonation of the ammonium group before inactivation mechanisms take place [28]. The alterations in pKa values of substrate species may also influence the reactivity efficiencies as stated by O'Hagan et al. [208]. However, pKa calculations require a more sophisticated study which is not in the scope of this work.

### 3.1.2.2 Transamination reactions

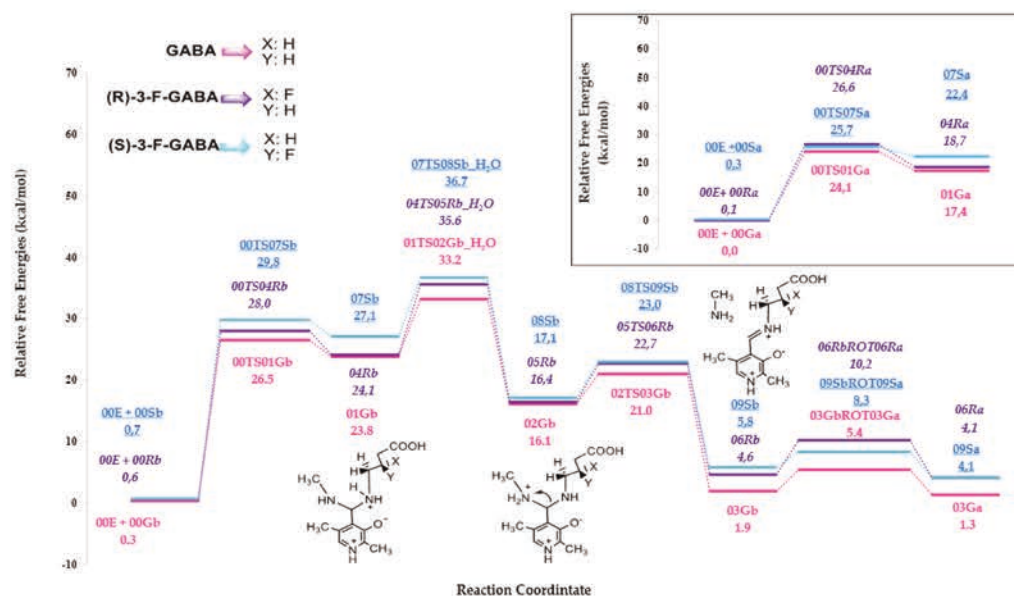
Transamination reaction proceeds through the following steps: nucleophilic attack of substrate amine to form geminal diamine; 1,3 proton transfer between the amino groups of substrate and methylamine; and external aldimine formation via the cleavage



**Figure 3.2 :** Newman projections of GABA conformations (**00Ga-z**, **00Gb-z** and **00Gc-z**), (R)-3-F-GABA (**00Ra-z**, **00Rb-z** and **00Rc-z**) and (S)-3-F-GABA (**00Sa-z**, **00Sb-z** and **00Sc-z**) in the zwitterionic form. Relative free energies (kcal/mol) are given in parentheses.

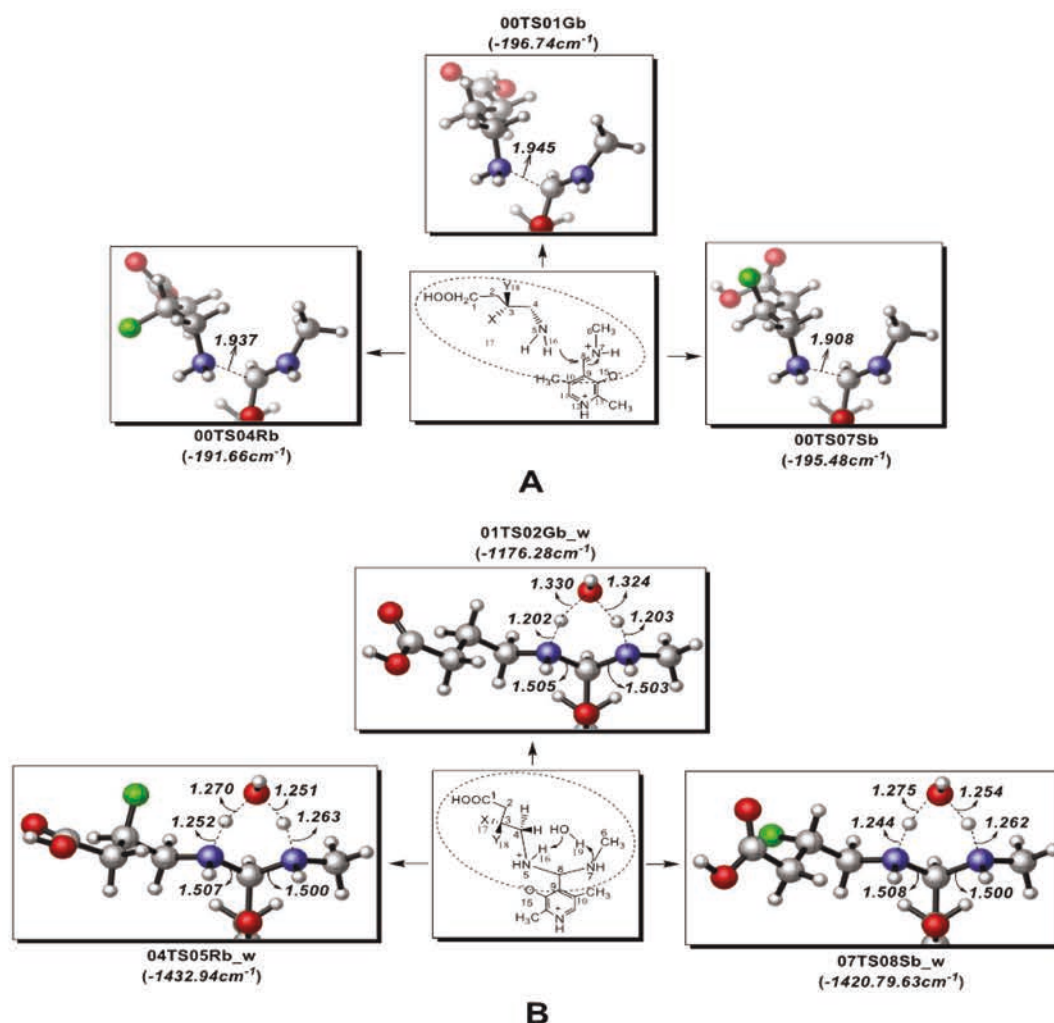
of the enzyme-PLP covalent bond. Energy profiles of transamination reactions of bioactive conformations of GABA, (R)-3-F-GABA and (S)-3-F-GABA are depicted in Figure 3.3 and are calibrated relative to the solution conformer of GABA (**00Ga**), which is the lowest-energy structure among the whole conformers. The energy profile of the binding step of solution conformers is also displayed in the box in Figure 3.3. It has been suggested that the substrates in the zwitterionic form undergoes a deprotonation at the beginning of the transamination reaction where Thr353 residue is proposed for the assistance of the process due to its proximity to the phosphate group of PLP in the active site [28]. The deprotonation before the binding lowers energy differences among the conformers relative to the corresponding zwitterionic forms (Figure 3.3). The nucleophilic attack of the amine group of substrate to C8 of the PLP-methylamine complex is known as the binding step, which regulates the enzymatic activity [28]. It is important to analyze the binding step energy barrier in

order to understand the dynamics of the substrate-binding efficiencies. The binding of the bioactive conformer of GABA (00Gb) to the GABA-AT is modelled with the transition state 00TS01Gb (Figure 3.4), having an energy barrier of 26.2 kcal/mol, which is higher than the solution conformer 00TS01Ga by 2.1 kcal/mol (Figure 3.3). The higher barrier value of 00TS01Gb can be attributed to the different nature of the steric crowding around the reacting center. The transition state yields 01Gb, which has a relative free energy of 23.8 kcal/mol (Figure 3.3). In our previous work on the fluorinated aromatic substrate, the binding step barrier is calculated to be 14.0 kcal/mol with N7-C8 distance being 2.133Å [33]. Although the steric hindrance created by the aromatic substituent is greater, the conformational preference of the carboxylic group in the previous transition state structure results in stronger hydrogen bonding interactions, thereby lowering the barrier height.



**Figure 3.3 :** B3LYP/6-31+G(d,p) free energy profiles of transimination reactions of GABA (pink), (R)-3-F-GABA (purple) and (S)-3-F-GABA (blue).

In the case of 3-F-GABA analogues, the binding step barriers for the bioactive conformers (00Rb and 00Sb) are calculated to be 27.4 kcal/mol, and 29.1 kcal/mol yielding 04Rb and 07Sb respectively (Figure 3.3). In both cases, the solution analogues have lower energy barriers. However, the order of energy barriers of the binding step reversed going from solution to bioactive conformers. The presence of a long range interaction between the hydrogen atom on the N5 and the fluorine atom (2.435Å) lowers the energy of the 00TS04Rb relative to 00TS07Sb as well as the charge-dipole interactions [213, 214]. In addition, the energetic stability of R enantiomer over S



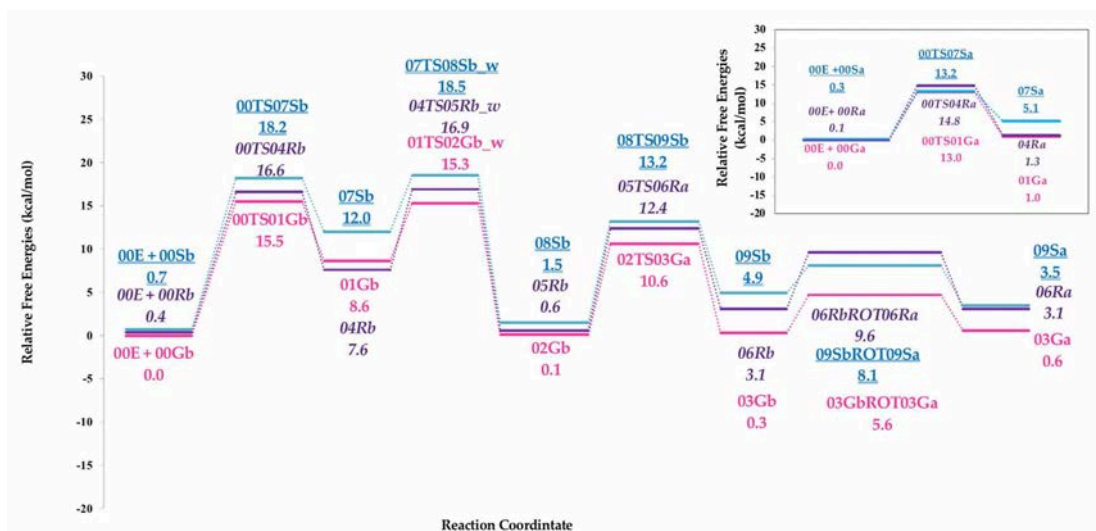
**Figure 3.4 :** Three-dimensional geometries of the first two steps of the transamination with the bioactive conformers of GABA (00TS01Gb and 01TS02Gb\_w, respectively), (R)-3-F-GABA (00TS04Rb and 04TS05Rb\_w, respectively) and (S)-3-F-GABA (00TS07Sb and 07TS08Sb\_w, respectively). (A) Formation of substrate-PLP-methylamine complex. (B) 1,3 proton transfer between nitrogens of substrate and methylamine. The imaginary frequencies that define transition states are given in parentheses.

enantiomer can also be explained by the stabilizing electrostatic interactions between the C-F bond and electropositive centers ( $F^{\delta-} - C_{\beta} - C_{\alpha} - N^{\delta+}$ ) [215,216].

Experiments indicated that the inhibition efficiency of the R-isomer is at least 10 times greater than the S-isomer [28]. Hence, the binding ratios are calculated for both of the solution and bioactive conformers of 3-F-GABA and compared with the experimental results. The binding ability of the S enantiomer of the solution conformer is found to be higher than that of the R at different levels of calculations which is inconsistent with the experimental results [28]. However the calculated binding efficiencies of bioactive conformers ( $((R) - 3 - F - GABA : (S) - 3 - F - GABA = 19 : 1)$  based on



B3LYP/6-31 + G(d,p) level reflect the experimental (R)-3-F-GABA:(S)-3-F-GABA binding efficiency ratio which has been proposed to be 10. The ratio obtained from MP2 single point energy calculations (9:1) (Figure 3.5) also supports the binding efficiency of R over S. Thus, succeeding calculations are conducted with the bioactive conformers.



**Figure 3.5 :** Single point relative MP2 energies of Transimination Reactions of bioactive conformations and the binding step of solution conformations with MP2 level of theory at 6-311+G(d,p) basis set. Energies are given in kcal/mol.

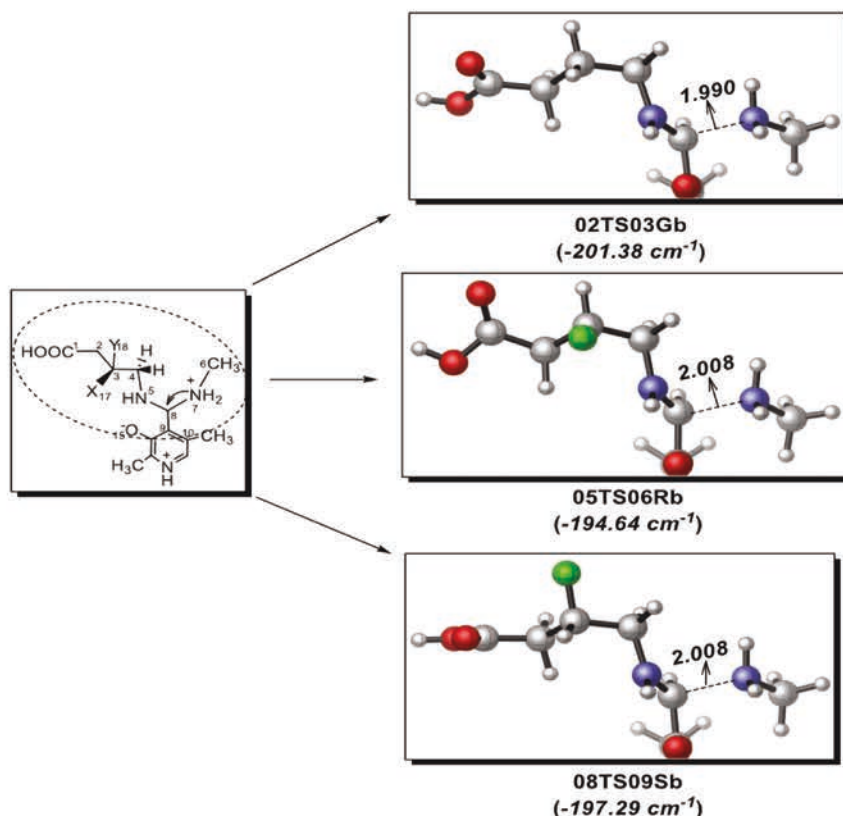
The free energy barrier of the 1,3 proton transfer step for the bioactive conformer of GABA (01TS02Gb\_w, Figure 3.4) is calculated as 9.5 kcal/mol (Figure 3.3). In the previous work, the barrier height of the 1,3 proton transfer step without water assistance was calculated as 21.8 kcal/mol, whereas it was found to be 17.2 kcal/mol with the water-assisted mechanism [33]. Herein, the energy barrier obtained for the GABA is lower than both barriers of the previous study [33]. The difference may stem from the conformational preference of the aromatic substrate, which creates electrostatic destabilization based on the nearby oxygen atoms around the reacting center. The fluorinated analogues, 04TS05Rb\_w and 07TS08Sb\_w (Figure 3.4), have free energy barriers of 11.5 kcal/mol and 9.6 kcal/mol, respectively (Figure 3.3), which are also lower than the previous results [33]. The NBO calculations pointed out that the interaction of the lone pair of C9 as donor with the lone pair of C10 as acceptor is stronger for the 04TS05Rb\_w than both for the 01TS02Gb\_w and 07TS08Sb\_w structures. Even though 04TS05Rb\_w has the highest stabilizing energy due to this interaction, the transition state structure is destabilized by the presence of the fluorine

atom next to the oxygen of the water molecule (3.420Å) which results in higher free energy value than 01TS02Gb\_w. The transition states of fluorine-substituted analogues yield the structures 05Rb and 08Sb with relative free energies of 16.4 and 17.1 kcal/mol, respectively. The charges on the reacting-site atoms have similar values in all transition state structures. Moreover, the MP2 single point energy calculations lowered the energy barrier values of the 1,3 proton transfer step about 3 kcal/mol, yet did not change the rate-determining step (Figure 3.5). The results indicate that the stereoelectronic differences affect the activation energy barrier.

The last step of the transamination reaction is the removal of the methylamine to form the substrate–PLP complex. The structure 02TS03Gb (Figure 3.6) has a free energy barrier of 4.9 kcal/mol (Figure 3.3). The transition states for R (05TS06Rb, Figure 3.6) and S (08TS09Sb, Figure 3.6) enantiomers have almost the same free energy barriers (6.3 kcal/mol and 5.9 kcal/mol respectively), which is compatible with previous works [33, 199]. The stability of 02TS03Gb is simply due to the interaction of the lone pair of the PLP ring carbon (C9) as acceptor and the lone pair of the reaction-center carbon atom (C8) as donor. These interactions are weaker in 05TS06Rb and 08TS09Sb. After the formation of structures 03Gb, 06Rb and 09Sb, rotations of 120° around the C3–C4 bond are required for further enzymatic reactions [28]. The free energy barrier for rotations are around 5 kcal/mol, yielding the structures 03Ga, 06Ra and 09Sa. The MP2 single point energy results drastically change the rotational energy barrier values to the point that the barriers are almost doubled or tripled (Figure 3.5). However, the mechanistic conclusion was not affected by the variations in the barrier heights.

### 3.1.2.3 Fluoride ion elimination

The fluorine-substituted substrate analogues are still contemporary issue not only due to the conformational dynamics but also due to their prominent efficiency on the enzymatic reactions. Even though the experimental [18, 26, 28, 217] and theoretical [33, 35, 218] studies have provided some clues regarding the fluorine-containing substrate analogues, the effect of the configurational changes of the substrates on the enzymatic reactions is still not explained. The alignment of C-F and C-N bonds would lead to different consequences for the HF elimination reaction, especially if



**Figure 3.6 :** Three-dimensional geometries of transition state structures of the removal of methylamine with GABA (02TS03Gb), (R)-3-F-GABA (05TS06Rb) and (S)-3-F-GABA(08TS09Sb). The imaginary frequencies that define transition state are given in parentheses.

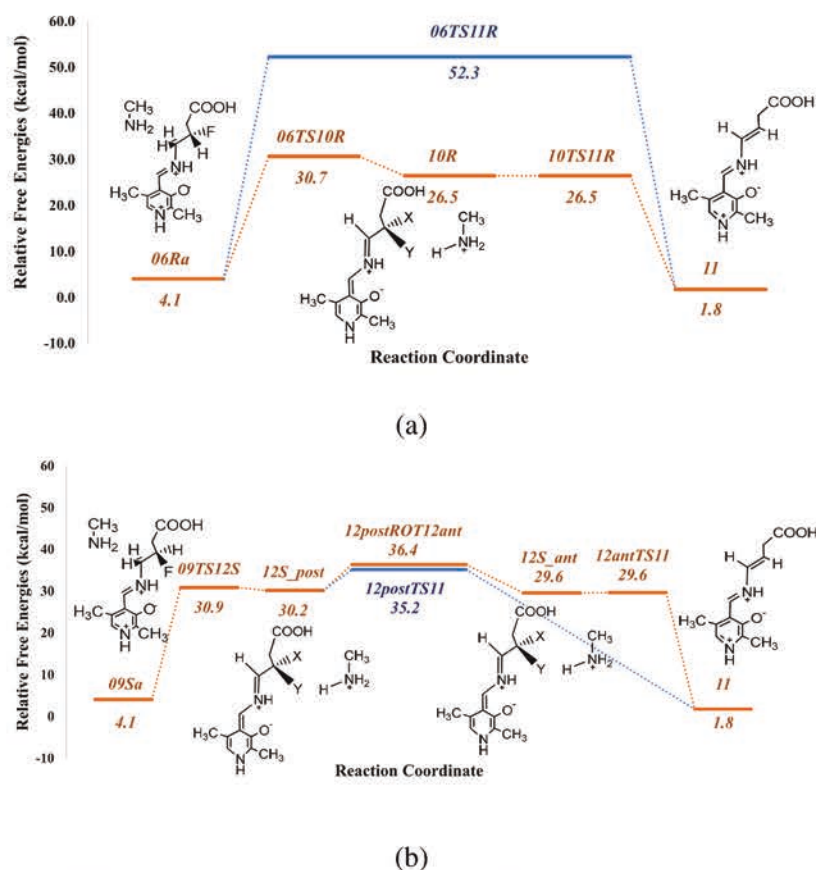
the mechanism is restricted by stereoelectronic circumstances [28, 35]. The effect of configurational and stereoelectronic differences on the HF elimination reaction can be explained computationally via model reactions of (R)-3-F-GABA and (S)-3-F-GABA.

It is known that the HF elimination mechanisms may proceed via either a stepwise E1cb or a concerted E2 mechanism. These elimination reactions are typical examples of the base-induced  $\beta$ -elimination reactions with the formation of a C-C double bond [202]. In both theoretical and experimental studies, it is proposed that the stepwise E1cb mechanism is generally assisted by a residue located near the active site, particularly a nearby Lys residue [28, 33, 42]. In a similar way, the concerted E2 elimination in enzymes, would likely be assisted either by water or by water and a nearby basic residue together [29, 33, 218]. Lys329 is reported as a candidate base for assistance of the E2 elimination reaction in GABA-AT [28, 42]. It is also noticeable that the protonated nitrogen on the pyridine ring of PLP may provide extra stabilization during the elimination reactions [33, 200, 219].

The first step in the E1cb mechanism is the abstraction of the  $\gamma$  proton at C4. The abstraction of the proton is then followed by HF elimination via proton transfer to the fluorine atom from the  $CH_3NH_3^+$  moiety. The energy profile of HF elimination from the(R)-3-F-GABA-PLP complex is given in Figure 3.7. The transition state structure (06TS10R, Figure 3.8) of the proton abstraction step in the E1cb mechanism is modelled with a free energy of 30.7 kcal/mol (Figure 3.7). The distance between the transferring proton and C4 is calculated as 1.517Å, while the distance between the transfer-ring proton and N7 is 1.232Å. The free energy barrier is calculated as 26.6 kcal/mol, while it was calculated as 7.8 kcal/mol in the previous work with a different substrate analogue which is consist of a six-membered ring [33]. This energy difference might stem from the higher charge delocalization ability of the aromatic ring bearing substrate, which would facilitate the proton abstraction in the previous work [33]. The proton abstraction from structure 06Ra yields the intermediate 10R (Figure 3.8), having a relative free energy of 26.5 kcal/mol (Figure 3.7) with a clockwise rotation  $93.9^\circ$  around the C4-N5 bond. The substrate tends to be planar to the molecular plane of PLP ( $\beta_{C8-N5-C4-C3} = 10.1^\circ$ ) in structure 10R, while it prefers to be perpendicular in 06Ra ( $\beta_{C8-N5-C4-C3} = 104.0^\circ$ ). It should be noted that the F atom is found to be anterior to the molecular plane of PLP in 10R, while it is posterior in 06a.

The proton transfer from  $CH_3NH_3^+$  moiety to the fluorine atom is rather problematic, and a possible transition state structure could not be modelled [33]. Presence of a shallow minimum due to the looseness of the C-F bond prevents modelling the transition state structure in a solution environment [33,35,218]. To our best knowledge, structure 10TS11R (Figure 3.8) is the first transition state structure that could be located for the HF elimination following E1cb mechanism. According to the solvent optimization calculations, the relative free energy of 10TS11R is calculated as 26.5 kcal/mol (Figure 3.7). Since structure 10R corresponds to a shallow minimum (Figure 3.7) owing to the looseness of the C-F bond(1.485Å, Figure 3.8), the transition state occurs barrierless. Product 11 has a relative free energy of 1.8 kcal/mol (Figure 3.7) and the dihedral angle  $\beta_{C8-N5-C4-C3}$  is calculated as  $25.8^\circ$ . The overall path of the(R)-3-F-GABA-PLP complex is found to be endothermic.

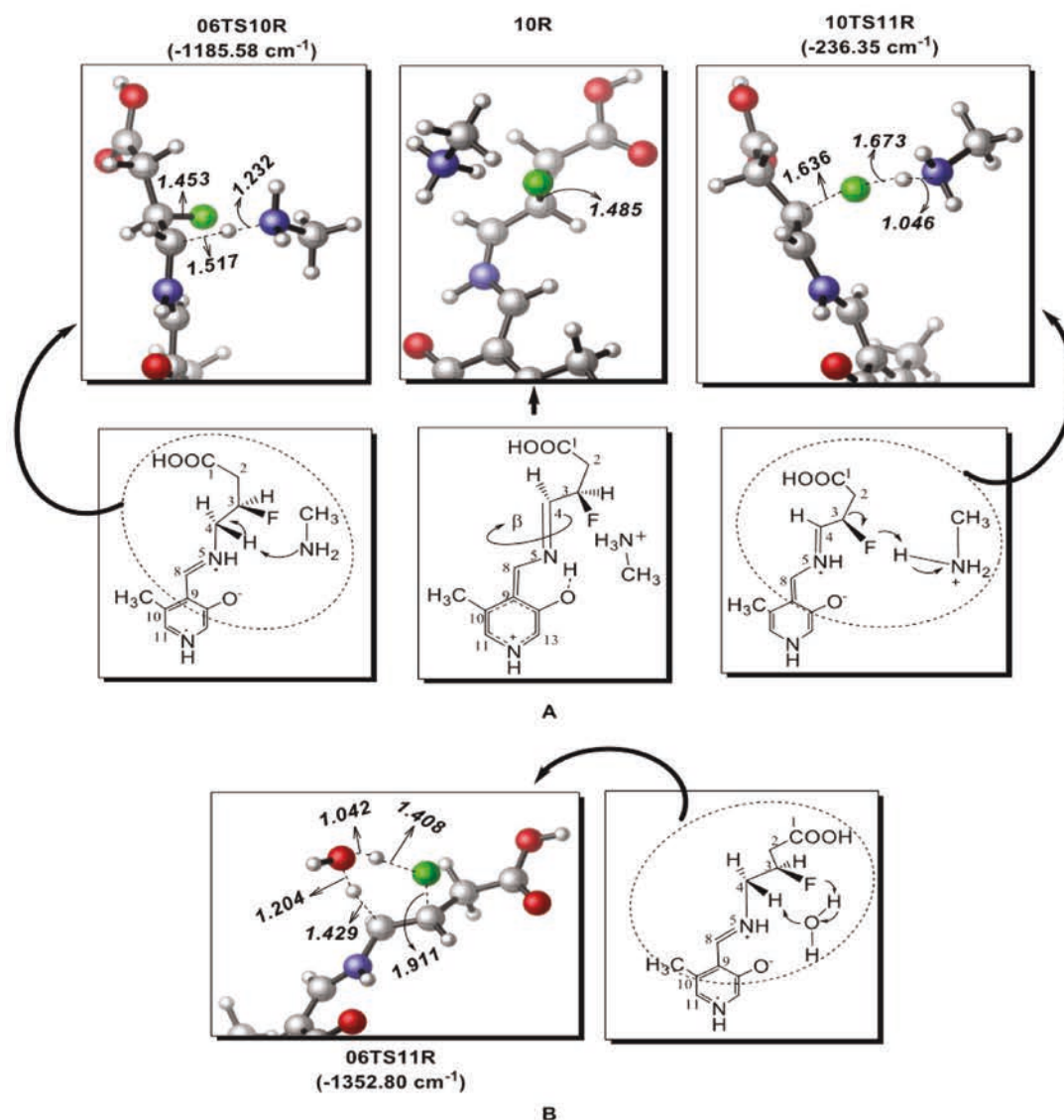




**Figure 3.7 :** B3LYP/6-31+G(d,p) free energy profile of (a). HF elimination from the (R)-3-F-GABA-PLP complex via E1cb (orange) and concerted (blue) mechanisms and (b). HF elimination from the (S)-3-F-GABA-PLP complex via E1cb mechanism with the first scenario (blue) and the second scenario (orange).

Clift et al. proposed that, because the (R)-3-F-GABA structure does not conform an anti-periplanar arrangement of the eliminating substituents it is not possible to achieve an E2 mechanism [28]. Therefore an alternative concerted mechanism is proposed for the HF elimination from (R)-3-F-GABA-PLP complex with the water assistance. During the course of the reaction an explicit water molecule is used as a proton shuttle for the  $\gamma$  proton abstraction. The free energy barrier for the formation of the **11** via the transition state **06TS11R** (Figure 3.8) is calculated as 48.2 kcal/mol in the concerted mechanism (Figure 3.7). The energetics of reaction paths reveals that there is a preference of the E1cb mechanism over the concerted mechanism in case of (R)-3-F-GABA-PLP complex.

In case of (S)-3-F-GABA, the E1cb mechanism is modelled in a similar way as in (R)-3-F-GABA. The energy profile of HF elimination reactions of (S)-3-F-GABA is given in Figure 3.7. The abstraction of the  $\gamma$  proton step, which is achieved with

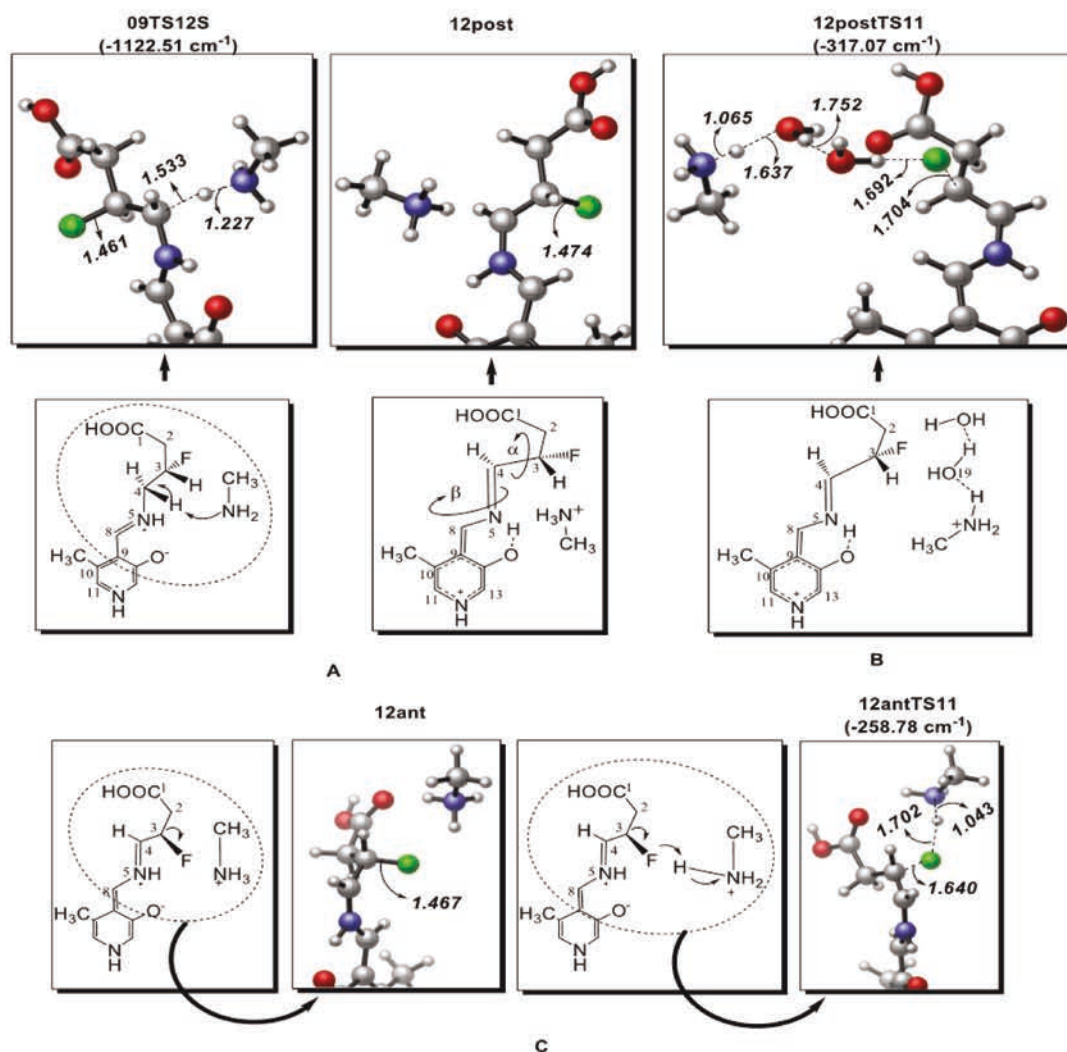


**Figure 3.8 :** Three-dimensional geometries of HF elimination reactions of (R)-3-F-GABA-PLP complex. (A) HF elimination via stepwise E1cb mechanism with the assistance of methylamine. (B) HF elimination via concerted mechanism with the assistance of one water molecule. The imaginary frequencies that define transition states are given in parentheses.

transition structure 09TS12S (Figure 3.9), has a relative free energy of 30.9 kcal/mol similar to the R enantiomer (06TS10R, Figure 3.9). Abstraction of the proton from 09Sa yields the structure 12post (Figure 3.9), in which the F atom is posterior to the molecular plane of PLP. It is observed that in the structure 12post, the carboxyl group of substrate is almost planar to the molecular plane of PLP and positioned in the opposite direction to the O atom of PLP as in 10R (Figure 3.8).

Two different scenarios can be proposed for the fluoride ion elimination since the fluorine atom and the protonated methylamine are in the opposite direction of the





**Figure 3.9 :** Three-dimensional geometries of stationary structures of fluoride ion elimination reactions via E1cb mechanism of (S)-3-F-GABA-PLP-complex. **(A)** Abstraction of  $\gamma$  proton and formation of structure 12post. **(B)** First scenario of fluoride ion elimination with the assistance of methylamine and water molecules. **(C)** Second scenario of fluoride ion elimination in the form of HF with the assistance of methylamine. The imaginary frequencies that define transition states are given in parentheses.

molecular plane of PLP. In the first scenario, the fluoride ion elimination occurs with the assistance of water molecules as proton shuttles. In the second scenario, in order to eliminate the fluoride ion in the form of HF, a rotation around the C3–C4 bond is required. The rotation allows the fluorine atom to be anterior to the molecular plane of PLP.

The transition state structure 12postTS11 (Figure 3.9) having a relative free energy of 35.2 kcal/mol is modelled for the first scenario (Figure 3.7). The distance between C3 and F is calculated as  $1.704 \text{ \AA}$ , while the distance between the transferring proton and

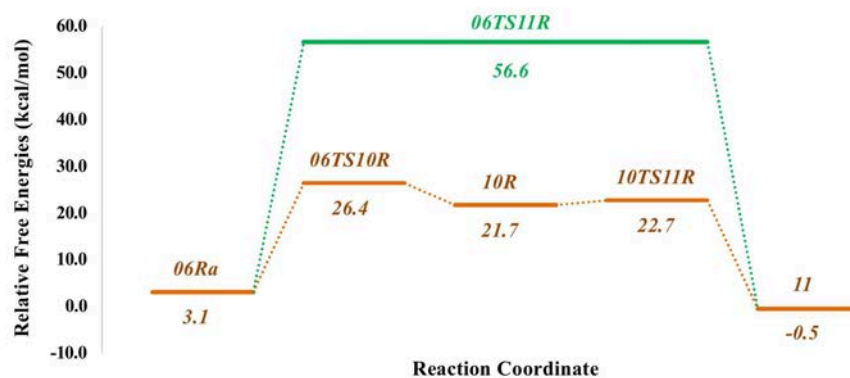
the O10 of water molecule is found to be 1.637Å. The dihedral angle  $\alpha_{N5-C4-C3-C2}$  in structure 12postTS11 shifted by  $\sim 10^0$  relative to 12post and then rotate back to its original position in structure 11 to maintain planarity of the carboxyl group relative to the PLP ring.

As for the second scenario, the proposed rotation around the C3–C4 bond is achieved with the transition state 12postROT12ant having a relative free energy of 36.4 kcal/mol (Figure 3.7) and yielding the intermediate structure 12ant (Figure 3.9). The dihedral angle  $\alpha_{N5-C4-C3-C2}$  is calculated as  $61.9^0$  for 12ant which was  $-178.0^0$  in the case of 12post. The proposed scenarios have almost equal probabilities to occur with a relative free energy difference of 1.2 kcal/mol. The results are also supported by the MP2 single point energy calculations (Figure 3.10). The stronger C3–F bond in structure 12ant(1.466Å) relative to the 10R (1.548Å) is expected to lower the relative energy. However, the strength of long range interactions predominates in structure 10R thereby generating 3.0 kcal/mol energy difference relative to structure 12ant. The last transition state structure 12antTS11 (Figure 3.9) of the second scenario is found to be similar to that of R analogue (10TS11R, Figure 3.8). The relative free energy of the transition state structure 12antTS11 is found to be 29.6 kcal/mol (Figure 3.7) where the transition state occurs barrierless as in 10TS11R. A relative free energy difference of 3.1 kcal/mol is calculated between the transition state structures of R and S analogues.

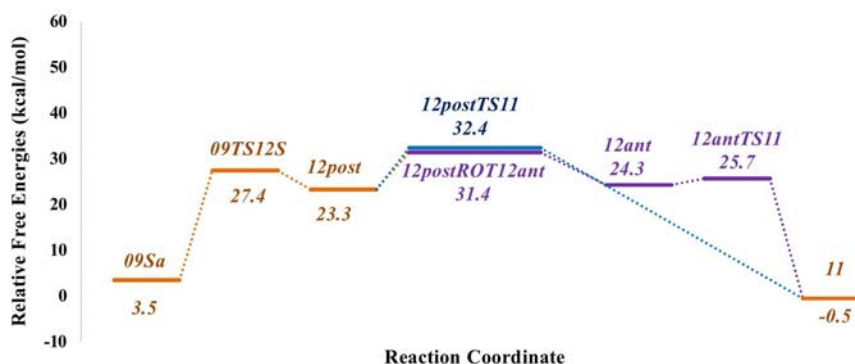
### 3.1.3 Concluding remarks

The mechanistic details of two successive reactions, transimination and HF elimination, are studied computationally with different conformers of GABA and its fluorine-substituted analogues. The stereoelectronic differences between the structures enabled us to discuss the possible dependencies of the studied reactions.

The dynamics of the enzymatic reactions depend not only upon the activation energy barriers of the transition state structures but also on the conformational and stereoelectronic preferences of the corresponding reactions, as well as the binding efficiencies of the substrates. Prior to the reaction mechanisms, the stability of the solution and bioactive conformers is investigated and it is found that the bioactive conformers of the two substituted GABA analogues are not the global minimum structures. Hence, the relative population of the bioactive conformer of



(a)



(b)

**Figure 3.10** : Single point relative MP2 energies obtained with MP2 level of theory at 6-311+G(d,p) basis set. **(a).** Single point relative MP2 energies of HF elimination reactions of (R)-3-F-GABA-PLP complex. **(b).** Single point relative MP2 energies of HF elimination reactions of (S)-3-F-GABA-PLP complex. Energies are given in kcal/mol.

the (S)-enantiomer will be less than that of the (R)-enantiomer. The results indicate that the poor efficiency of the (S)-enantiomer may stem from the instability of the corresponding bioactive conformer. The bioactive conformers of the fluorine substituted analogues are not the global minimum structures to this respect the requirement of the high concentrations of fluorine substituted analogues to inhibit GABA-AT can be explained.

The binding efficiencies of the different enantiomers was also investigated for further explanations. The dynamics of the binding step, which is the first step in the transamination reaction, is analysed to explain the binding efficiencies. It is found that the R conformer is more convenient for binding. The binding ratio calculated from the relative energy barriers of the bioactive conformers of (R)-3-F-GABA:(S)-3-F-GABA is found to be 19.1:1 which is in harmony with the experimental work. Results point

out the prominence of reactive conformers rather than the most populated substrate analogue in the inhibition reactions involving PLP.

In transimination reactions, the stereochemical differences of the fluorine-substituted analogues affect the activation energy barriers by means of electrostatic interactions. Herein, the rate-determining step is found to be the 1,3 proton transfer in the transimination reaction. The differences between the activation energy barriers of the fluorine substituted GABA analogues are not very distinctive. Therefore, the poor performance of the (S)-enantiomer cannot be correlated to the energy profile of the transimination. The HF elimination reactions of fluorine-substituted analogues are also important to elucidate the binding preference of GABA-AT and to analyze the reason for the poor performance of (S)-3-F-GABA, which is presented experimentally [28]. The firststep of the E1cb mechanism of the (R)-enantiomer (06TS10R) and(S)-enantiomer (09TS12S) have same activation energy barriers,where the carbanion intermediate of the (R)-enantiomer (10R) is more stable. For R enantiomer, a barrierless transition state(10TS11R) is obtained for the last step of the HF elimination. Two different scenarios are taken into account for the HF elimination in the (S)-isomer case. It is found that the rotational transition state (12postROT12ant) and the transition state involves water molecules as proton shuttles (12postTS11) have similar relative free energies being higher than the first step of elimination. The last step of the path is the distinctive for the HF elimination efficiency differences for the enantiomers. The fluoride ion release from the (R)-enantiomer is faster than the (S)-enantiomer supporting the lower HF elimination rate of the (S)-enantiomer.

### **3.2 Molecular Dynamics Simulations of Apo, Holo and Inactivator Bound GABA-AT**

The pyridoxal 5-phosphate (PLP) cofactor is a significant organic molecule in medicinal chemistry. It is often found covalently bound to lysine residues in proteins to form PLP-dependent enzymes. An example of this family of PLP dependent enzymes is  $\gamma$ -aminobutyric acid aminotransferase (GABA-AT) which is responsible for the degradation of the neurotransmitter GABA. Its inhibition or inactivation can be used to prevent the reduction of GABA concentration in brain which is the source of several neurological disorders.

The results from the experimental [45–52], theoretical [33, 34] and NMR [11–13] studies lead us to investigate the protonation states of the active site residues in PLP dependent enzyme GABA-AT. Does the pyridine ring of PLP bear a protonated nitrogen and how does PLP behave in the active site? What is the effect of the protonation states of the active site residues on the proposed active site interactions? Do these residues have secondary roles? What would be the behavior of the enzyme when an inactivator is involved? In order to answer these questions, we performed Molecular Dynamic (MD) simulations. Apoenzyme, holoenzyme and vigabatrin bound holoenzyme samples are simulated to evaluate the protonation states and the roles of the active site residues in GABA-AT. Different protonation states have been considered for PLP and two key active site residues: Asp298 and His190. Together, 24 independent molecular dynamics trajectories have been simulated for a cumulative total of 2.88  $\mu$ s.

### **3.2.1 Computational details**

#### **3.2.1.1 Preparation of samples**

Three different systems with different protonation schemes for each one has been built: the first one is GABA-AT apoenzyme (i.e., GABA-AT with a free Lys329), the second one is GABA-AT holoenzyme (i.e., GABA-AT with PLP covalently bound to Lys329), and the third one involves the  $\gamma$ -vinyl-GABA (vigabatrin) bound GABA-AT. The residue Asp298 was reported as an active site residue that interacts with the pyridine ring of PLP [45] via a hydrogen bond. In the literature, the nitrogen on the pyridine ring was depicted both as protonated or unprotonated [3, 33]. In case of a protonated nitrogen, the Asp298 residue should have a side chain with the charge of  $-1$ , while in the case of an unprotonated nitrogen on the pyridine ring a neutral Asp298 side chain should be considered. Therefore, because of the uncertainty on the position of the hydrogen in the hydrogen bond between the PLP pyridine ring and Asp298, two different protonation schemes have been considered: either a protonated pyridine ring with an unprotonated Asp or an unprotonated pyridine ring with a protonated Asp. In addition to the Asp residue, the analysis of hydrogen bond patterns in the X-ray structures might suggest two different protonation states for His190 which is hydrogen bonded to Asp298: either a neutral side chain or an ionized (positively charged) side

chain. In the following text, the samples of apoenzyme GABA-AT are named as A, the samples corresponding to GABA-AT holoenzyme are referred to H and vigabatrin bound GABA-AT samples are named V, respectively. The different protonation states of the residues in the samples are depicted with the charges of Asp298 and His190 residues respectively as superscripts; i.e. the GABA-AT apoenzyme sample that has a negatively charged Asp298 and a neutral His190 is named as  $A^{-1,0}$ . The nomenclature of all samples is given in Table 3.1.

**Table 3.1** : Nomenclature of the Samples According to the Protonation States of the Asp298 and His190 Residues.

System	Asp298	His190
Apoenzyme	$A^{-1}$	$A^{-1,1}$
		$A^{-1,0}$
	$A^0$	$A^{0,0}$
		$A^{0,1}$
Holoenzyme	$H^{-1}$	$H^{-1,1}$
		$H^{-1,0}$
	$H^0$	$H^{0,0}$
		$H^{0,1}$
Vigabatrin bound Holoenzyme	$V^{-1}$	$V^{-1,1}$
		$V^{-1,0}$
	$V^0$	$V^{0,0}$
		$V^{0,1}$

The samples were prepared from two different X-ray structures; 4-Aminobutyrate-aminotransferase from pig (Protein Data Bank entry 1OHV) and 4-Aminobutyrate-aminotransferase inactivated by  $\gamma$ -vinyl GABA (Protein Data Bank entry 1OHW). The apo-GABA-AT and holo-GABA-AT samples were prepared using the 1OHV PDB file while the vigabatrin bound GABA-AT samples were prepared using the 1OHW PDB file. Each PDB entry contains two homodimers in their unit cell, since the enzyme is active as a homodimer, only one of the dimer (chains A and B) of each Protein Data Bank entry was kept and subjected to simulations. The monomers are simply named as Monomer A and Monomer B. When considering the interactions between residues in one monomer, the residues that belong to the other



monomer are depicted with a star (i.e. Thr353\*) and the iron-sulfur cluster that binds two monomers is named as FES.

### 3.2.1.2 Force field modifications

The samples were hydrogenated, and the topology and coordinate files were built using the tLEAP module of AMBER 12 [220]. Amino acid residues have been modeled using the ff03 force field parameters [174]. The pKa values of the residues were computed with the PROPKA program [221, 222] and no amino acid was predicted with a shift towards a neutral pKa value. Therefore, all residues were protonated as in aqueous solution. The protonation states of histidine residues were determined according to the hydrogen bond pattern network surrounding each imidazole ring. Sodium counter ions were added to neutralize the samples. Waters from the crystal structure were deleted, and the samples were solvated with explicit TIP3P water molecules [223]. The simulation boxes are cubic, of edge length 111.8Å, have an initial density of 1.0, and contain about 40 000 water molecules.

In each monomer of holo-GABA-AT, a PLP ligand is covalently bound to Lys329. In vigabatrin-bound-GABA-AT, the PLP residues are covalently bound to the vigabatrin inhibitors which are themselves bound to the two Lys329. In all samples, Lys329 and the molecules that are covalently bound to their side chains have been modeled within AMBER as one whole residue using the Antechamber program [175]. The resulting non-standard residues have been modeled using the ff03 force field in order to be consistent with the force field used for the other amino acids. Atomic charges have been determined from SemiEmpirical Born-Oppenheimer Molecular Dynamics (SEBOMD) [82, 83] of the residues capped by acetyl and N-methyl groups: CM1 atomic charges [224], which have been developed to correctly reproduce molecular dipole moments of small molecules, were averaged from a 10ps SEBOMD simulation at 300K using the PM3 Hamiltonian.

In GABA-AT, the two monomers are linked together by an iron-sulfur cluster ( $Fe_2S_2$ ) bound by four cysteine residues. Based on ESP calculations of small  $Fe_2S_2$  bound to methylsulfur molecules, the iron atoms have been assigned an atomic charge of +0.7e, the sulfur atoms have been assigned an atomic charge of -0.45e, and the four cysteine residues (135, 137, 135\* and 137\*) have been modified in their thiolate form

so that their total charge is -0.125e each. Bond, angles and dihedrals angles between the iron-sulfur cluster and the four cysteine residues have been added using the GAFF force field [175].

### **3.2.1.3 Molecular dynamics simulation**

Molecular dynamics simulations were performed using the CUDA version of AMBER12's pmemd program [225,226]. Periodic boundary conditions were imposed on the systems. The Particle Mesh Ewald summation technique was used with the default 8Å cutoff [187]. A time step of 2.0fs was used by the implementation of the SHAKE algorithm for the bonds involving hydrogens [181]. The equilibration of the samples was performed in a five stage process [227]. First, after a short 1000 step minimization, a 100ps molecular dynamics was performed at 10K with a strong temperature coupling (i.e., velocities were randomly updated every 10 steps in the Andersen thermostat) and with a harmonic restraint of 50 kcal/mol/Å<sup>2</sup> on all non-hydrogen atoms to ensure a proper geometry of the hydrogen atoms that were added by the tLEAP program and to relax the hydrogen bond network between the water molecules and the protein. Second, a 100ps molecular dynamics simulation was performed in which only the protein heavy atoms were restrained. Third, another 100ps molecular dynamics simulation was performed in which the restrained atoms were identical as before but with a lighter 5 kcal/mol/Å<sup>2</sup> harmonic restraint. Fourth, a last 100ps molecular dynamics at 10K was performed with no restraint. Fifth, the thermostat temperature was linearly increased from 10K to 300K during a 2 ns simulation. Productions simulations were performed at 300K in the NVT ensemble using either the Andersen [188] or the Berendsen temperature coupling [189] for statistical comparison. A total of 24 different samples were simulated up to 120ns each (e.g., the three A, H, and V systems x 2 protonation states for Asp298/PLP x 2 protonation states for His190 x 2 different NVT thermostats). Since the results that are obtained from both thermostats are in agreement with each other, only the results obtained with Andersen temperature coupling are discussed below.

### **3.2.1.4 Estimation of GABA-AT/PLP interaction free energies**

From the Holo-GABA-AT trajectories, we have tried to evaluate the variations of interaction energies between PLP and GABA-AT with respect to the different

protonation states of His190 and Asp298. Because PLP is covalently bond to GABA-AT, it is difficult to decompose the potential energy of the fully solvated system computed at each step of the simulation into well-defined contributions. Therefore, in the spirit of the MM-GB/SA approach that is often used to evaluate binding free energies [228], we have designed the following approximate protocol to evaluate an approximate interaction free energies between GABA-AT and PLP that can be compared between different trajectories. First, the average interaction free energy between PLP and GABA-AT for a given Holo-GABA-AT trajectory is defined as:

$$\Delta G_{int} = G_H - (G_A + G_{PLP}) \quad (3.2)$$

where H and A refer to Holo-GABA-AT and Apo-GABA-AT respectively. Because we are interested in the variations of  $\Delta G_{int}$  between different Holo-GABA-AT trajectories, we define each G contribution of the right hand-side of equation 3.2 by:

$$G = \left\langle E_{MM} + G_{solv} - TS_{MM} \right\rangle \quad (3.3)$$

where  $E_{MM}$  is the total energy of the system, estimated at the molecular mechanics (MM) level,  $G_{solv}$  is the solvation free energy, and the last term is the absolute temperature multiplied by an entropy estimate. The bracket in equation 3.3 indicate an average over snapshots from the molecular dynamics trajectories.

In our computations, we have voluntary omitted the calculation of the entropy term because i) it is computationally very demanding either using a normal-mode approach or a quasi-harmonic approximation; ii) by assuming that the solute entropy should be the same for each trajectory, this term should cancel out when  $\Delta \Delta G_{int}$  between different simulations is computed, as well as the covalent term between PLP and the enzyme.

For each snapshot of a Holo-GABA-AT simulation,  $E_{MM} + G_{solv}$  is computed by removing all explicit water molecules and counter-ions and by using a Generalized Born approach to evaluate the solvation contribution [229]. Thus, the solvent contribution is averaged because of the use of an implicit solvent model.  $G_H$  is computed using the full Holo-GABA-AT system (i.e., GABA-AT with two covalently bound PLPs, one for each monomer),  $G_A$  is computed by removing all of the PLP atoms in one of the monomers of the Holo-GABA-AT snapshot, and  $G_{PLP}$  is computed from the coordinates of PLP extracted from this snapshot. In this approach,  $\Delta G_{int}$  does

not correctly correspond to a binding energy since the formation of the Schiff base between PLP and Lys329 yields the departure of a water molecule.

### **3.2.1.5 Trajectory analysis**

The root-mean-square deviation (rmsd), the atomic fluctuation (B-factor) calculations and the dihedral, distance and angle analysis were performed using the cpptraj module [230]. The pKa calculations on trajectories were performed using the PROPKA program [221, 222]. The two dimensional structure of the active site is investigated with the LigPlot+ software [231]. Probability distributions of the distance between the hydrogen donor atom and the heavy atom acceptor (O and N atom) were computed and any peak around 2Å was interpreted as a hydrogen bond interaction between the corresponding atoms.

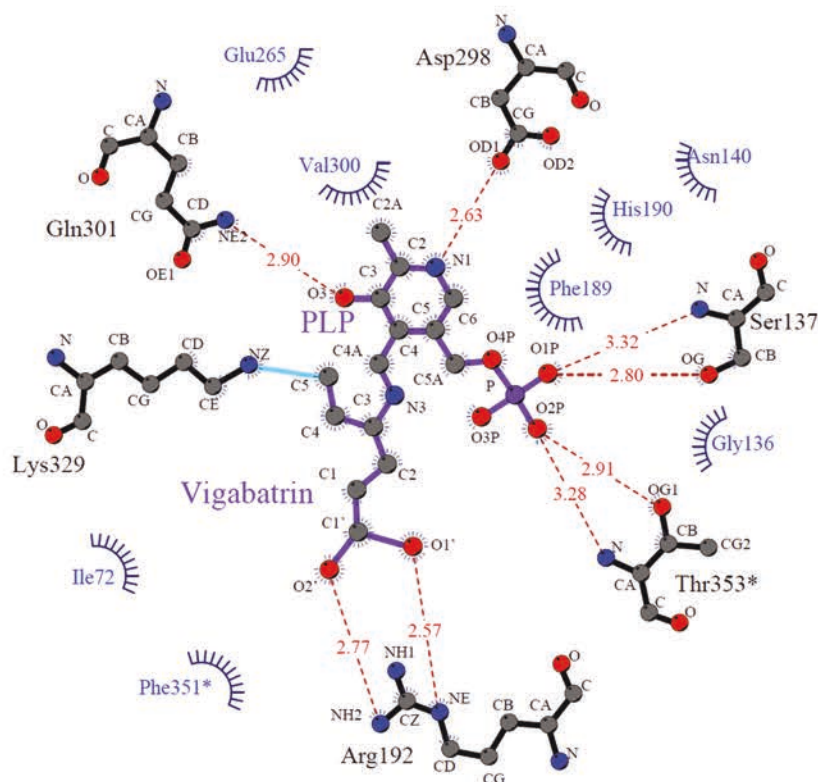
## **3.2.2 Results and Discussion**

### **3.2.2.1 Preliminary study on GABA-AT**

Experimental and theoretical studies have revealed that the pyridine ring of protein-bound PLP should be protonated to facilitate enzymatic reactions via an electron sink effect [5–8]. However the PLP molecule is mainly unprotonated in aqueous solution under physiological conditions [11]. Thus, GABA-AT should either bind low populated protonated PLP or protonate the pyridine ring during the Schiff base formation between PLP and the enzyme. Therefore, the pKa values of the residues in the active site, especially the ones that are in close proximity and can transfer a proton, are of great importance.

The crystallographic structure of GABA-AT is investigated to understand the active site residue interactions in GABA-AT before performing MD simulations with different protonation states. The active site residues in vigabatrin bound GABA-AT are drawn with LigPlot+ and are shown in Figure 3.11. Most of the active site residues are in close contact with either PLP or vigabatrin. The residue Phe351\* is proposed to narrow the active site and increases the hydrophobicity of the active site with Ile72 [45]. Glu270 is supposed to interact with the fully conserved Arg445 [45]. The iron-sulfur cluster, which holds the two monomer together, is bound to Cys135 and Cys138. Asn352\* is the neighbour residue of Thr353\* that is proposed to interact with the phosphate group

of PLP [45,52]. Glu265 is on the bottom of the coenzyme binding pocket and proposed to perform an intermediate function [52]. Phe189 has been suggested to sandwich PLP with Val300 [45], and Arg192 is proposed to interact with the carboxyl group of the substrates [45,52]. Among the active site residues that are close to PLP, Asp298 could be the candidate residue to protonate the nitrogen atom on the pyridine ring of PLP.



**Figure 3.11** : Two dimensional LigPlot<sup>+</sup> representation of the interactions in the active site of vigabatrin bound GABA-AT.

In order to investigate the protonation states of the active site residues in GABA-AT without PLP, the cofactor is excluded from the 1OHV X-Ray structure and a PROPKA calculation is performed on apo-GABA-AT. The residue that can transfer its proton to the pyridine ring N atom is Asp298 which should have therefore a neutral side chain to perform this action. While PROPKA indicates that Asp298 is 100% buried, the pKa's of Asp298 on both monomers are evaluated at about 2.8. Factors that favor a negative Asp298 are hydrogen bonds by Asn140, His190 and Val300, as well as charge-charge interactions with His190 and Lys329. Because His190 has the greatest contribution on the ionization state of the Asp298 residue, we have investigated the conservation of the His190 residue in the PLP dependent proteins that can be retrieved from the Protein Data Bank. In order to do so, from the Protein Data Bank (PDB), we have queried all

PDB entries that contain a PLP ligand. For all PLP residues in all the structures, we have:

1. Identify the PLP residue
2. Add a fictitious hydrogen atom bound at 1.0Å from the nitrogen atom of the PLP pyridine ring
3. Search for the closest residue to this hydrogen atom
4. If the closest residue is Asp and the distance is less than 2.3Å, then we consider to have a hydrogen bond between PLP and Asp and the structure is selected
5. If the structure is selected, we search if a His residue is hydrogen bonded to the Asp residue (with a distance < 3.3Å between one of the oxygen atoms of the carboxylic group and one of the nitrogen atom of the imidazole ring)

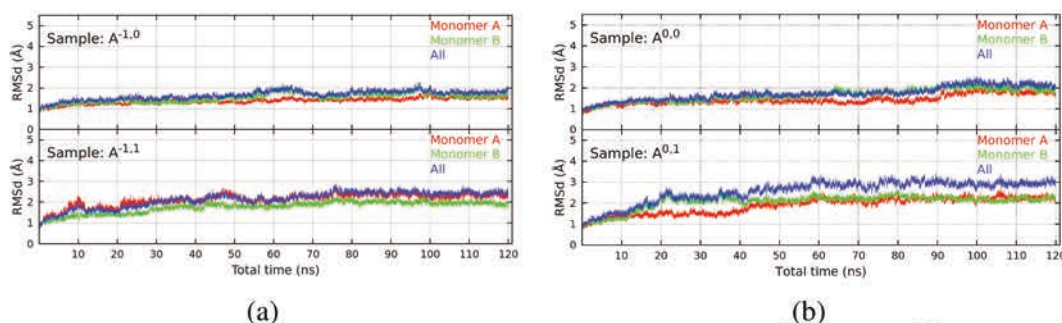
As of October 2014, we have found 714 PDB entries containing a PLP ligand. It is found that there is an Asp residue that can form a hydrogen bond with the nitrogen atom of the pyridine ring in 368 PDB entries among the found entries. From them, 171 PDB entries have a His residue that is in interaction with the corresponding Asp residue. This means that about 25% of PLP containing protein structures available in the PDB share a common pattern: PLP is hydrogen bonded to a Asp residue which is itself hydrogen bonded to a His residue. Since Asp is conserved in about 50% and His is conserved in about 25% of the PLP dependent proteins, the protonation states of these residues are important. The pKa values of His190 on both monomers of apo-GABA-AT are evaluated by PROPKA at around 5.8 where the crystallographic structure was reported at pH 5.7, so the calculated pKa of His190 is not enough to identify the protonation state of the imidazole ring. To identify the protonation states of the GABA-AT “triad” composed of PLP, Asp298 and His190, MD simulations with different ionized states have been performed.

### **3.2.2.2 Protonation states in GABA-AT**

**Apo-GABA-AT.** To elucidate the protonation states of His190 residues in the apoenzyme, MD simulations are first performed on apo-GABA-AT. The stability



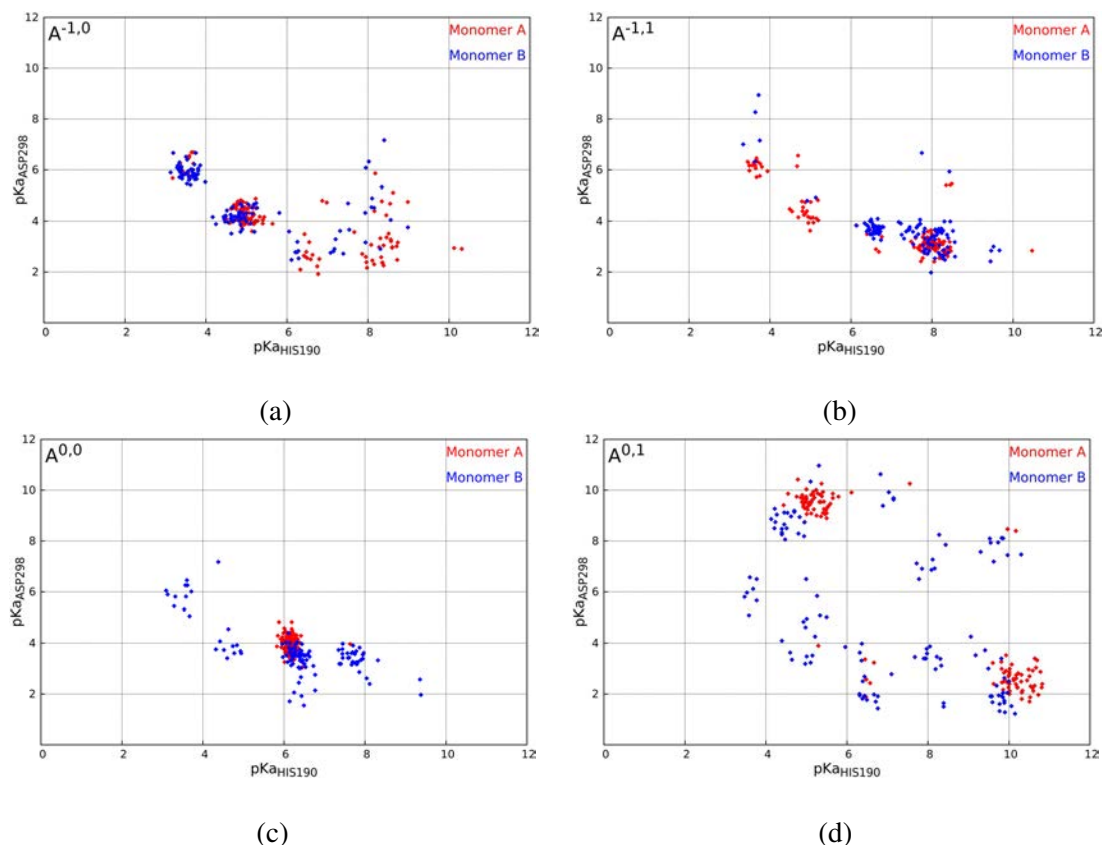
of the trajectories along the simulations are investigated by the calculations of the root-mean-square deviations (RMSd) of the backbone atoms (C,  $C_\alpha$  and N) in monomer A and monomer B, and in the dimer for all apo-GABA-AT samples. The reference structure corresponds to the initial structure of the corresponding system for all samples. The RMSd values of the samples with negatively charged Asp298 and neutral His190 are below  $2\text{\AA}$  ( $A^{-1,0}$ , Figure 3.12a). The presence of one proton between negatively charged Asp298 and positively charged His190 causes deviations up to more than  $2\text{\AA}$  ( $A^{-1,1}$ , Figure 3.12a). The RMSd results of the samples that have neutral Asp298 side chain (i.e.,  $A^{0,0}$  and  $A^{0,1}$ ) are all below  $2\text{\AA}$  and given in Figure 3.12b.



**Figure 3.12** : RMSd values of apo-GABA-AT samples **a.**  $A^{-1,0}$  and  $A^{-1,1}$  and **b.**  $A^{0,0}$  and  $A^{0,1}$  along the simulation.

PROPKA calculations were performed on 120 representative frames (i.e., one frame every 1ns of simulation) for each MD trajectory. The relationships between calculated pKa's of Asp298 and His190 on both monomers of the apo-GABA-AT samples ( $A^{-1,0}$ ,  $A^{-1,1}$ ,  $A^{0,0}$ , and  $A^{0,1}$ ) are depicted in Figure 3.13. In the case of sample  $A^{-1,0}$ , one would expect that the pKa of Asp298 would be low (i.e., corresponding to a negatively charged aspartate) while the pKa of His190 would be high (i.e., corresponding to a neutral imidazole ring). From Figure 3.13a, most frames correspond to a low pKa for His190 and a high pKa for Asp298. This is in contradiction with the protonation state of the simulation. In simulation of  $A^{0,0}$ , the pKa of Asp298 should be higher corresponding to a neutral aspartate, but found low. In case of sample  $A^{0,1}$  the pKa of Asp298 show different patterns on Monomer A and Monomer B which is also in contradiction with the protonation state of the simulation. In contrast, a low pKa prediction for Asp298 is obtained for the  $A^{-1,1}$  sample while His190 pKa is mostly predicted in the range 6-8, which is compatible with both a neutral or an ionized state for the imidazole ring. Overall, the RMSd stability and the pKa computation results

on the molecular dynamics of apo-GABA-AT support a negatively charged Asp298. While no clear conclusion can be drawn for the protonation state of His190, it would seem that PROPKA results as well as hydrogen bond patterns favor an ionized histidine residue.

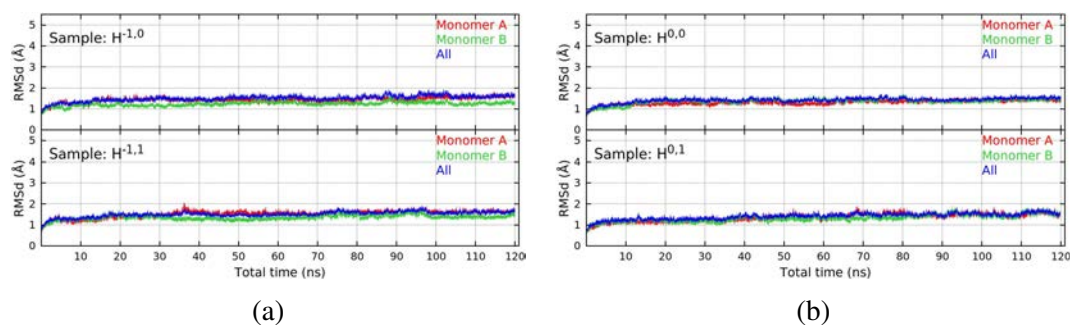


**Figure 3.13 :** pKa of His190 versus pKa of Asp298 along the simulation of apo-GABA-AT samples **a.**  $A^{-1,0}$ , **b.**  $A^{-1,1}$ , **c.**  $A^{0,0}$  and **d.**  $A^{0,1}$ .

**Holo-GABA-AT.** In order to assess the effect of the protonation states on the activity of the enzyme, the interaction between the side chain of Asp298 residue and the PLP pyridine ring is investigated for all holo-GABA-AT samples. The RMSd values that are obtained from the simulations of holo-GABA-AT samples are all computed below 2Å and the results are given in Figure 3.14.

The minimum distance between the oxygen atoms of the negatively charged Asp298 side chain and the hydrogen that is bound to the pyridine ring nitrogen is measured to investigate the possibility of having a hydrogen bond in the samples. The deviations of the minimum distances in  $H^{-1,0}$  and  $H^{-1,1}$  are depicted in Figure 3.15. It is observed that there is no interaction between Asp298 and PLP when the imidazole ring of the His190 is neutral (Figure 3.15a,  $H^{-1,0}$ , depicted in red). The disruptions

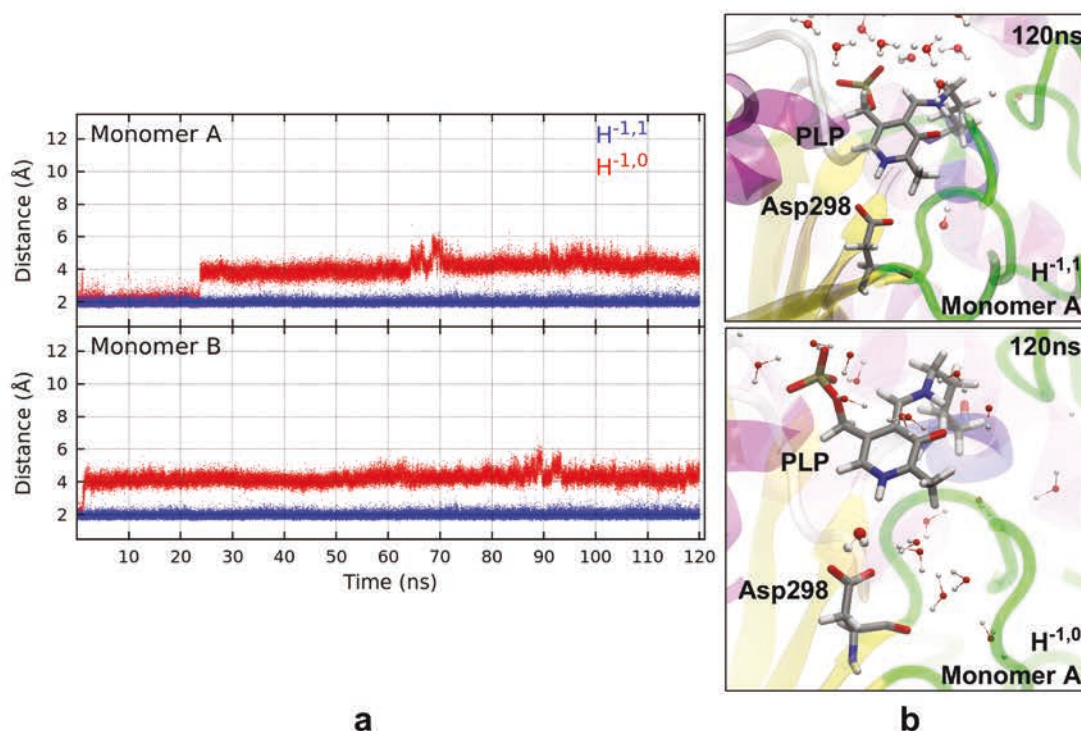




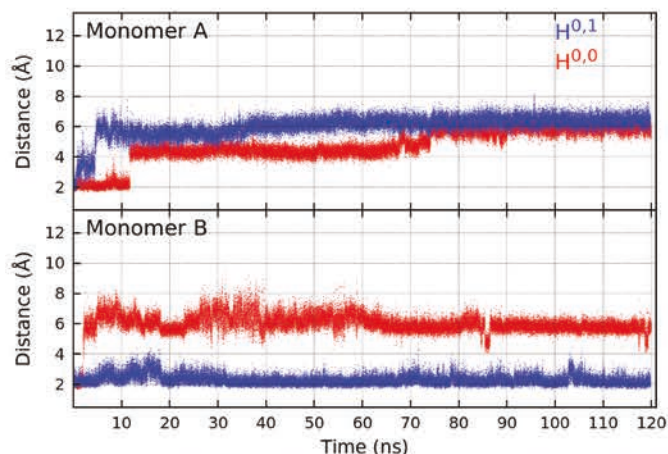
**Figure 3.14** : RMSd values of holo-GABA-AT samples **a.**  $H^{-1,0}$  and  $H^{-1,1}$  and **b.**  $H^{0,0}$  and  $H^{0,1}$  along the simulation

of the Asp298-PLP interaction is due to the movement of Asp298, and the proton on the nitrogen atom of the pyridine ring can form hydrogen bonds with the surrounding water molecules. On the other hand, when the imidazole ring is positively charged, Asp298 becomes less mobile and the interaction between Asp298 and the pyridine ring of PLP is stabilized (Figure 3.15a,  $H^{-1,1}$ , depicted in blue). The interaction between the neutral Asp298 and the pyridine ring in  $H^{0,0}$  and  $H^{0,1}$  has also been investigated. The distance variation between the nitrogen of the PLP pyridine ring and the hydrogen atom on the side chain of Asp298 is reported in (Figure 3.16). In both cases, it is found that the neutral Asp298 cannot form a stable hydrogen bond with the non-protonated N on the pyridine ring.

In addition to the interaction between the PLP pyridine ring and Asp298, there could be other interactions that would be affected by different protonation states of Asp298 and His190 side chains. Interactions of Asp298 with other fragments than the pyridine ring are also investigated to figure out if these are conserved or altered by the change in protonation states. Minimum distance probabilities between the side chain of Asp298 and, respectively, the imidazole ring of His190, the side chain of Asn140 and the backbone of Val300 residues are calculated for  $H^{-1,0}$  (Figure 3.17a) and  $H^{-1,1}$  (Figure 3.17b). The negatively charged Asp298 tends to have a hydrogen bond only with the hydrogen of the backbone of Val300 and it cannot interact with the imidazole ring of His190 or the side chain of Asn140 properly in  $H^{-1,0}$  (Figure 3.17a). On the other hand, the simulation of  $H^{-1,1}$  ends up with a stable hydrogen bond network with the involvement of the positively charged His190 side chain (Figure 3.17b). The side chain of Asp298 interacts with the  $H_\delta$  of His190 and can form hydrogen bonds with the H of the backbone of Val300 and the side chain of Asn140.

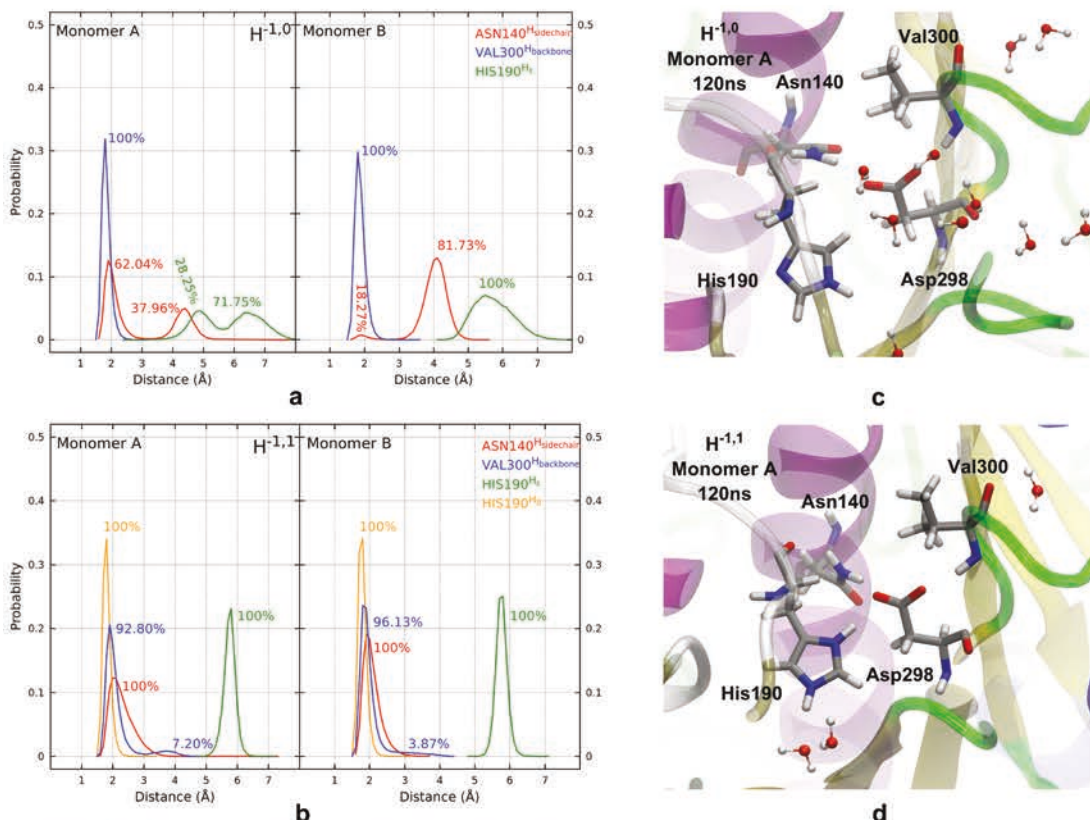


**Figure 3.15 :** Interaction between PLP pyridine ring and Asp298 in holo-GABA-AT. **a.** Distance variations between the side chain of Asp298 and pyridine ring of PLP for  $H^{-1,0}$  (red) and  $H^{-1,1}$  (blue). **b.** Typical three dimensional representations of the interaction between Asp298 and PLP in  $H^{-1,0}$  (bottom) and  $H^{-1,1}$  (top).



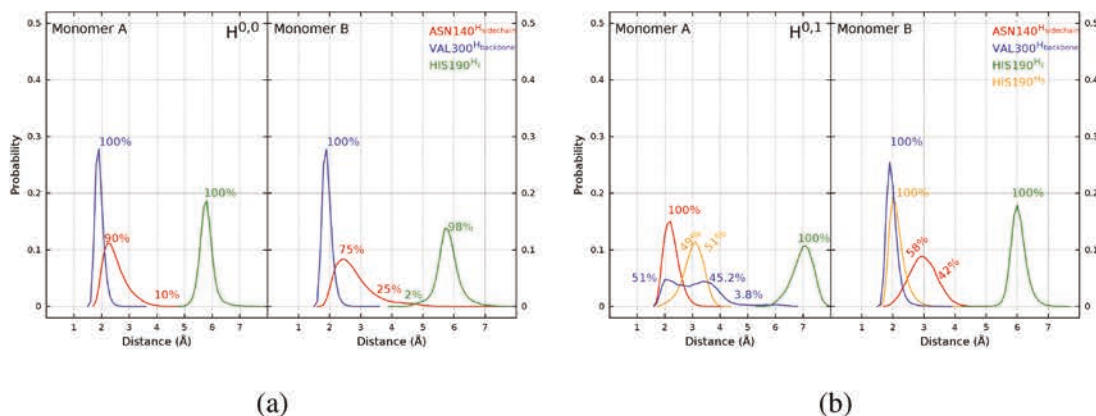
**Figure 3.16 :** The deviation of the distance between the side chain H atom of neutral Asp298 and the N atom on the pyridine ring of PLP along the simulations of the holo-GABA-AT samples  $H^{0,0}$  (red) and  $H^{0,1}$  (blue).

In addition to the holo-GABA-AT samples that contain negatively charged Asp298 side chain, the minimum distance probabilities between the side chain of Asp298 and, respectively, the imidazole ring of His190, the side chain of Asn140 and the backbone of Val300 residues are calculated also for the samples that have neutral Asp298 side chain ( $H^{0,0}$  and  $H^{0,1}$ ). The results are depicted in (Figure 3.18). When the imidazole ring of His190 is neutral ( $H^{0,0}$ ), the Asp298 can form hydrogen bonds with Asn140



**Figure 3.17 :** The hydrogen bond network of Asp298 in holo-GABA-AT. Minimum distance probabilities between Asp298 and, respectively, His190, Asn140 and Val300 residues for **a.**  $H^{-1,0}$  and **b.**  $H^{-1,1}$ . **c.** Typical three dimensional representation of ionized Asp298, Asn140, neutral His190 and Val300 in  $H^{-1,0}$ . **d.** Typical three dimensional representation of ionized Asp298, Asn140, ionized His190 and Val300 in  $H^{-1,1}$ .

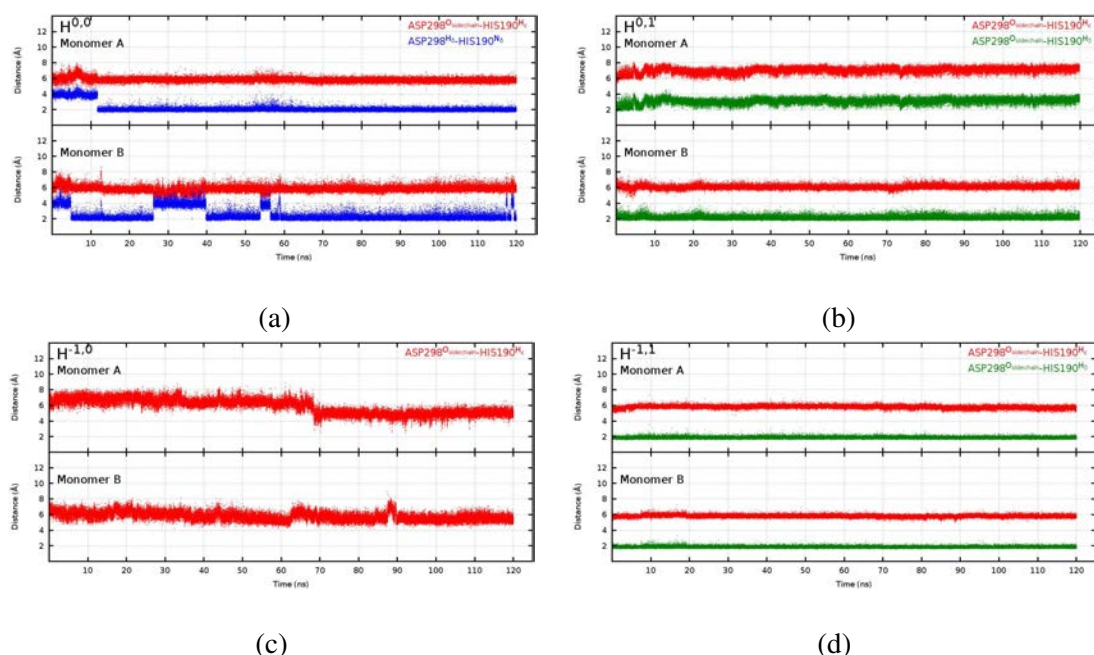
and Val300. Even though an extra hydrogen bond is possible with the presence of positively charged imidazole of His190 ( $H^{0,1}$ ), (see Figure 3.18b) the minimum distance probabilities of the hydrogen bonds of the Asp298 side chain decrease relative to the sample  $H^{-1,1}$ .



**Figure 3.18 :** Minimum distance probabilities between Asp298 and, respectively, His190, Asn140 and Val300 in samples **a.**  $H^{0,0}$ , and **b.**  $H^{0,1}$ .



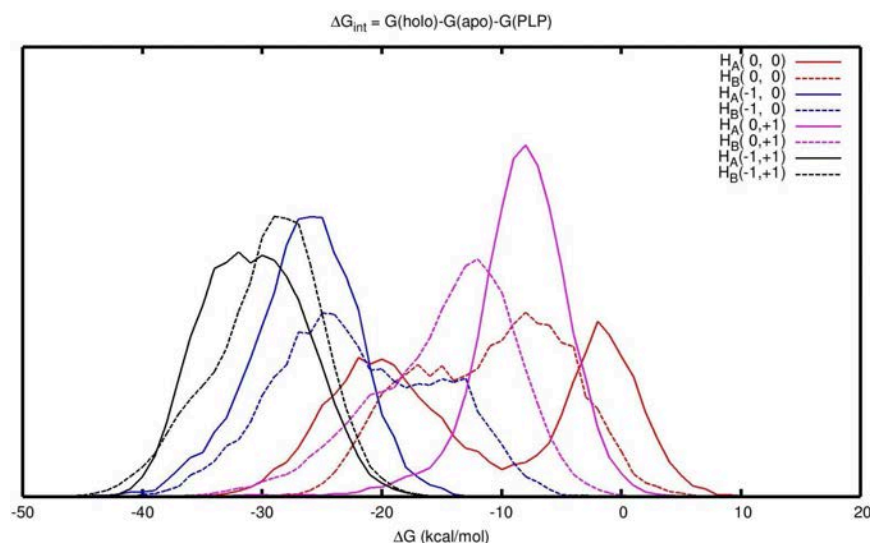
The computation of the variations of the distances between the side chain of His190 and the side chain of Asp298 on holo-GABA-AT samples yield to the presence of a stable hydrogen bond between the two residues only if both residues are ionized ( $H^{-1,1}$ ). The deviation of the computed distances are depicted in Figure 3.19. If both residues are neutral ( $H^{0,0}$ , Figure 3.19a) only one monomer can have stable interaction. When Asp298 is not ionized and His190 is ionized ( $H^{0,1}$ , Figure 3.19b), only one monomer have the interaction between these residues. In case of an ionized Asp298 and a neutral His190 ( $H^{-1,0}$ , Figure 3.19c), there is no interaction along the simulation.



**Figure 3.19 :** The interactions between the side chains of Asp298 and His190 in holo-GABA-AT samples. **a.** The variations of the minimum distances between  $H_{\epsilon}$  of His190 and the carboxyl oxygens of Asp298 (red), and  $H_{\delta}$  of Asp298 and  $N_{\delta}$  of His190 (blue) in sample  $H^{0,0}$ . **b.** The variations of the minimum distances between the side chain oxygen atoms of Asp298 and  $H_{\epsilon}$  (red) and  $H_{\delta}$  (green) of His190 in sample  $H^{0,1}$ . **c.** The variations of the distance between the side chain oxygen atoms of Asp298 and  $H_{\epsilon}$  (red) of His190 in sample  $H^{-1,0}$ . **d.** The variations of the minimum distances between the side chain oxygen atoms of Asp298 and  $H_{\epsilon}$  (red) and  $H_{\delta}$  (green) of His190 in sample  $H^{-1,1}$ .

Overall, the analysis of all holo-GABA-AT MD simulations shows that interactions within the active site of GABA-AT are stabilized by a positively charged His190 and a negatively charged Asp298. In terms of electrostatics, this means that GABA-AT active site is stabilized by charge-charge interactions where the negatively charged Asp298 is sandwiched between two positive fragments: the protonated His190 imidazole ring and the protonated PLP pyridine ring.

To further analyze the effect of the protonation states of Asp298 and His190 on the stability of the interactions between PLP and GABA-AT, the relative interaction free energies ( $\Delta G_{int}$ ) between the cofactor and the protein were computed from Holo-GABA-AT samples for both monomers. The calculated  $\Delta G_{int}$  values are depicted in Table 3.2. The distribution of the interaction free energies is reported at Figure 3.20. The lowest interaction free energy corresponds to the Monomer A of  $H^{-1,1}$ . The average interaction free energies of this system is then taken as a free energy reference. It is found that the minimum  $\Delta G_{int}$  value, which is chosen as reference, refers to the Monomer A of sample  $H^{-1,1}$ .  $\Delta \Delta G_{int}$  values that were obtained are lower in the samples with negatively charged Asp298 ( $H^{-1,1}$  and  $H^{-1,0}$ ) relative to the samples with neutral Asp298 ( $H^{0,1}$  and  $H^{0,0}$ ) by about 20 kcal/mol. Among the samples with negatively charged Asp298 ( $H^{-1,1}$  and  $H^{-1,0}$ ), positively charged His190 sample ( $H^{-1,1}$ ) has the lowest relative  $\Delta \Delta G_{int}$  values for both monomers. The calculated relative  $\Delta \Delta G_{int}$  values between PLP and GABA-AT confirm that the ionized states of His190 and Asp298 stabilize the protonated form of PLP in GABA-AT.



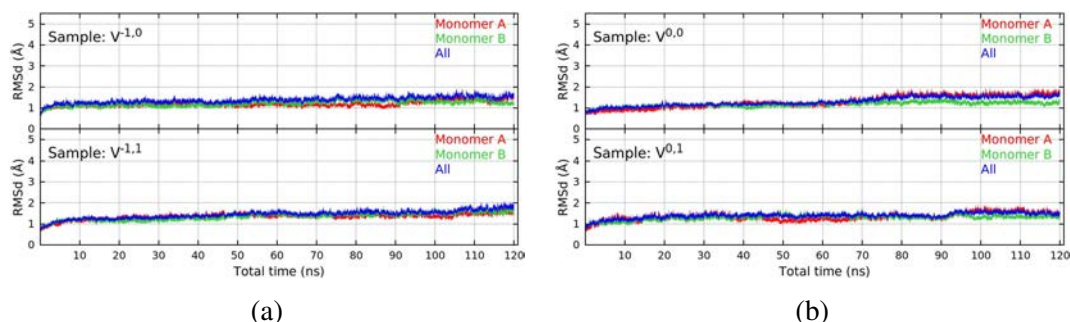
**Figure 3.20 :** Distributions of interaction free energies between PLP and GABA-AT computed for four Holo-GABA-AT trajectories:  $H^{0,0}$ ,  $H^{0,1}$ ,  $H^{-1,0}$ , and  $H^{-1,1}$ .

**Vigabatrin bound GABA-AT.** The simulations of holo-GABA-AT indicate that having a positively charged imidazole ring for His190 results in a less mobile Asp298 as well as stable interactions between Asp298 and the pyridine ring of PLP, the sidechains of His190 and of Asn140, and the backbone Val300. It is important to investigate if these interactions are conserved or altered in substrate bound GABA-AT,

**Table 3.2 :** Relative Free Energy of Interactions ( $\Delta\Delta G_{int}$ ) Between PLP and GABA-AT Generated from Holo-GABA-AT Trajectories, in kcal/mol.

Sample	Monomer A	Monomer B
$H^{-1,1}$	0.0	1.0
$H^{-1,0}$	4.1	8.7
$H^{0,1}$	22.3	15.9
$H^{0,0}$	18.7	19.7

and if there are alterations that may cause to it. Therefore, MD simulations with different protonation states for Asp298/PLP and His190 have been performed. The RMSd values that are obtained from the simulations of vigabatrin bound GABA-AT samples are all computed below  $2\text{\AA}$  and the results are given in Figure 3.21.



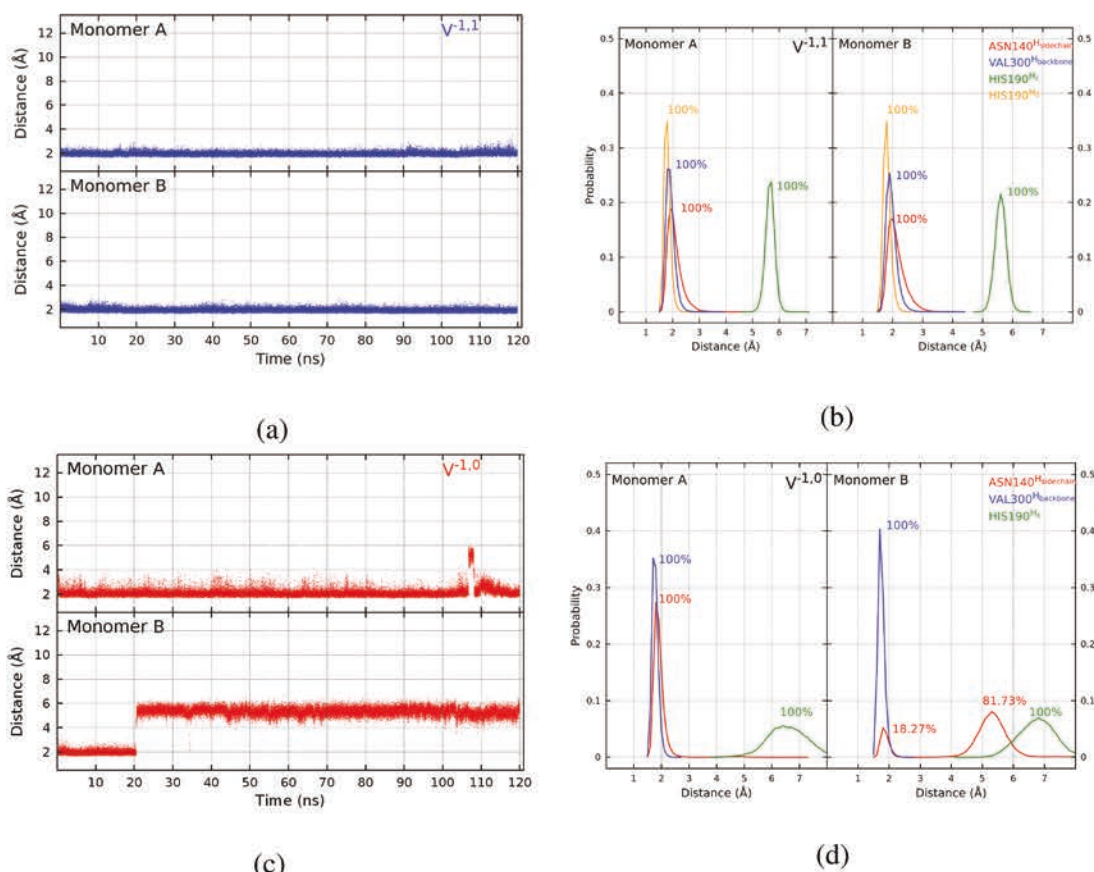
**Figure 3.21 :** RMSd values of vigabatrin bound GABA-AT samples **a.**  $V^{-1,0}$  and  $V^{-1,1}$  and **b.**  $V^{0,0}$  and  $V^{0,1}$  along the simulation.

The interaction between the Asp298 side chain and the pyridine ring of PLP is depicted in Figure 3.22a. The minimum distance between the oxygen atoms of the negatively charged Asp298 side chain and the hydrogen on the N atom of the pyridine ring is measured to investigate the possibility of having a hydrogen bond. It is observed that the interaction is conserved along the  $V^{-1,1}$  sample (Figure 3.22a) as well as in  $H^{-1,1}$  (Figure 3.22a). In addition, the hydrogen bond network that involves Asp298 is also investigated in vigabatrin bound GABA-AT. The minimum distance probability between the side chain of Asp298 and, respectively, the imidazole ring of His190, the side chain of Asn140 and the backbone of Val300 residues are calculated for  $V^{-1,1}$  (Figure 3.22b). It is found that the interaction between the imidazole ring of His190 and the side chain of Asp298 is not perturbed by the addition of vigabatrin. Furthermore, the probabilities of having an interaction between Asp298 and Asn140 as well as between Asp298 and Val300 are increased. The results that are obtained with

the samples of vigabatrin bound GABA-AT for other different protonation states are given in Figure 3.22c and Figure 3.22d. It was observed that there is an interaction between Asp298 side chain and pyridine ring of PLP when Asp298 is negatively charged but His190 is neutral ( $V^{-1,0}$ ). Nonetheless, the interaction is less stable in sample  $V^{-1,0}$  relative to  $V^{-1,1}$ . It was also found that the minimum distance probabilities between Asp298 and, respectively, His190, Asn140 and Val300 in sample  $V^{-1,0}$  also increase with the inclusion of vigabatrin relative to the holo-GABA-AT samples. However, the hydrogen bond between Asp298 and Asn140 side chains disrupts in one monomer (Figure 3.22d) with the presence of neutral imidazole. These results indicates that inclusion of vigabatrin may cause a steric effect within the active site to be able to have the aforementioned interactions regardless of the protonation state of the imidazole ring of His190. However, the positively charged side chain of His190 in  $V^{-1,1}$  enhances the hydrogen bond network of Asp298 relative to the neutral imidazole in  $V^{-1,0}$ .

In addition to the interactions of Asp298 within the samples  $V^{-1,1}$  and  $V^{-1,0}$ , the same interactions are also investigated for the samples that have neutral side chain of Asp298 ( $V^{0,1}$  and  $V^{0,0}$ ). The results are depicted in Figure 3.23. It was observed that the interaction between Asp298 and the pyridine ring of PLP becomes less stable (Figure 3.23a) relative to the  $V^{-1,1}$  and  $V^{-1,0}$  samples (Figure 3.22a and Figure 3.22c). The neutral Asp298 and His190 side chains in  $V^{0,0}$  disrupts the interaction between Asp298 side chain and PLP pyridine ring especially in one monomer (Figure 3.23a, red). When the minimum distance probabilities between Asp298 and, respectively, His190, Asn140 and Val300 are taken into account in  $V^{0,1}$  and  $V^{0,0}$  samples, it was observed that the probabilities decrease (Figure 3.23b and Figure 3.23c) relative to the ones that are obtained with negatively charged Asp298 side chain ( $V^{-1,1}$  and  $V^{-1,0}$ ). Overall, the obtained results indicates that the negatively charged Asp298 and positively charged His190 side chain provide more stabilization.

The computation of the variations of the distances between the side chain of His190 and the side chain of Asp298 on vigabatrin bound GABA-AT samples yield to the presence of a stable hydrogen bond between the two residues if and only if both residues are ionized ( $V^{-1,1}$ ). The deviation of the computed distances are depicted in Figure 3.24. When both residues are neutral ( $V^{0,0}$ , Figure 3.24a), or Asp298 is not



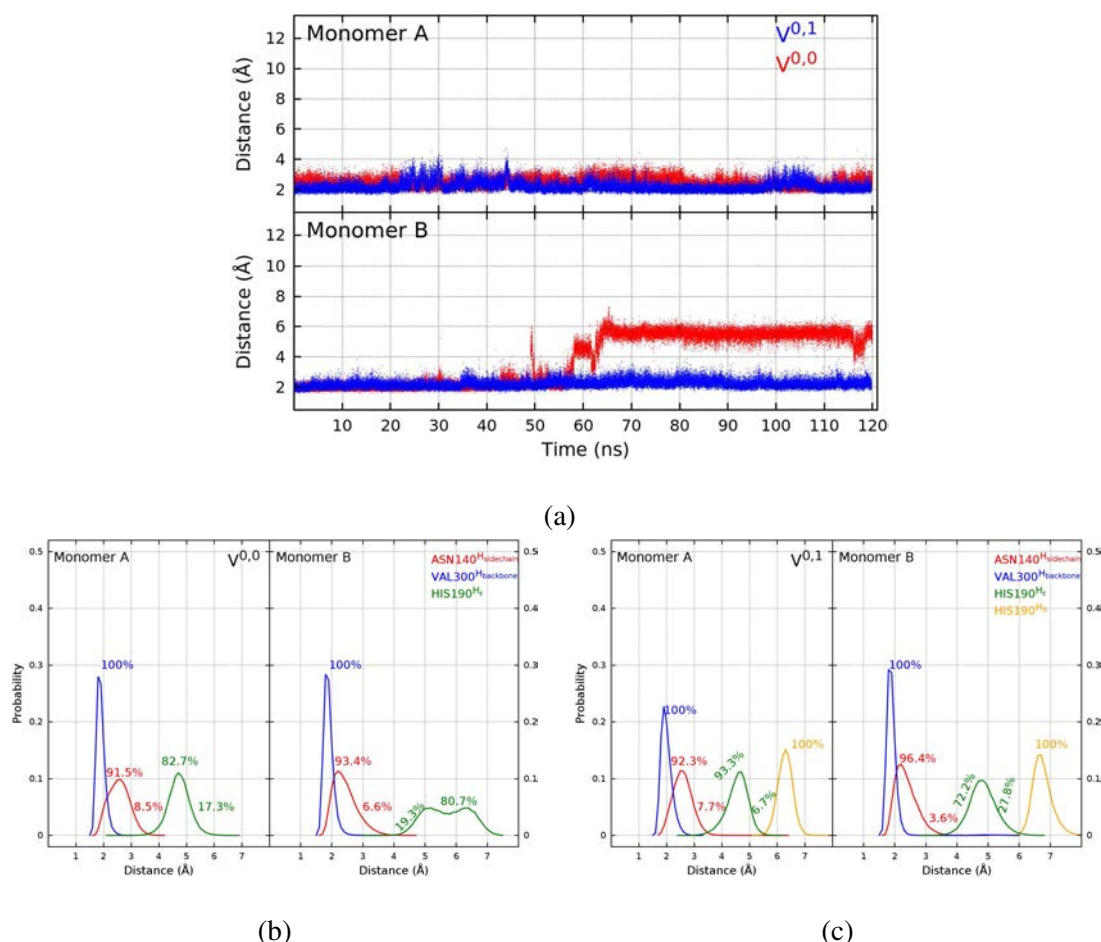
**Figure 3.22 :** Interactions of Asp298 in  $V^{-1,1}$  and  $V^{-1,0}$ . **a.** Distance variations between Asp298 side chain and the pyridine ring of PLP in  $V^{-1,1}$ . **b.** Minimum distance probabilities between Asp298 and, respectively, His190, Asn140 and Val300 in  $V^{-1,1}$ . **c.** The deviation of the minimum distance between the side chain O atoms of Asp298 and the H atom of N atom on the pyridine ring of PLP along the simulations of the vigabatrin bound GABA-AT sample  $V^{-1,0}$ . **d.** Minimum distance probabilities between Asp298 and, respectively, His190, Asn140 and Val300 in sample  $V^{-1,0}$ .

ionized and His190 is ionized ( $V^{0,1}$ , Figure 3.24b) and vice versa ( $V^{-1,0}$ , Figure 3.24c), there is no interaction along the simulations.

### 3.2.2.3 PLP phosphate group

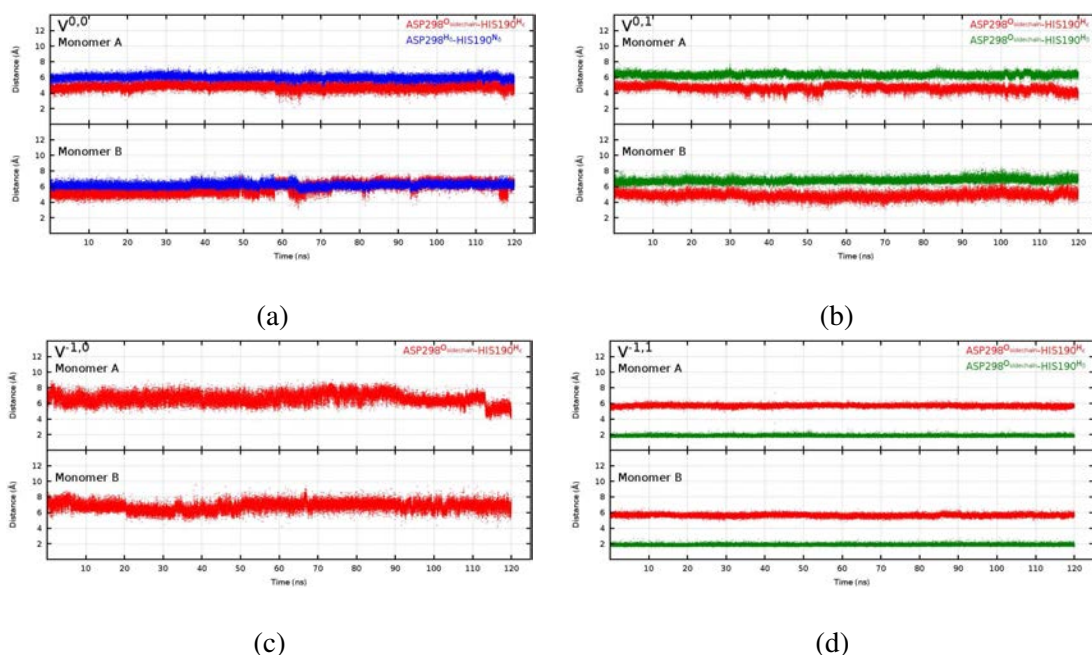
It has been suggested that the phosphate group of PLP is held in place via five hydrogen bonds that are donated by Gly136, Ser137 and Thr353\* [45, 52]. These interactions are investigated in holo and vigabatrin bound GABA-AT. The minimum distance probabilities between the oxygen atoms of the PLP phosphate group and Gly136, Ser137 and Thr353\*, respectively, are depicted in Figure 3.25. It is observed that the PLP phosphate group is held in place by at least three hydrogen bond from Gly136 and Ser137 in  $H^{-1,1}$  (Figure 3.25a). The most probable hydrogen bond is





**Figure 3.23 :** Interactions of Asp298 in  $V^{0,1}$  and  $V^{0,0}$ . **a.** The deviation of the minimum distance between the side chain O atoms of Asp298 and the H atom of N atom on the pyridine ring of PLP along the simulations of the vigabatrin bound GABA-AT sample  $V^{0,0}$  (red) and  $V^{0,1}$  (blue). **b.** Minimum distance probabilities between Asp298 and, respectively, His190, Asn140 and Val300 in  $V^{0,0}$ . **c.** Minimum distance probabilities between Asp298 and, respectively, His190, Asn140 and Val300 in sample  $V^{0,1}$ .

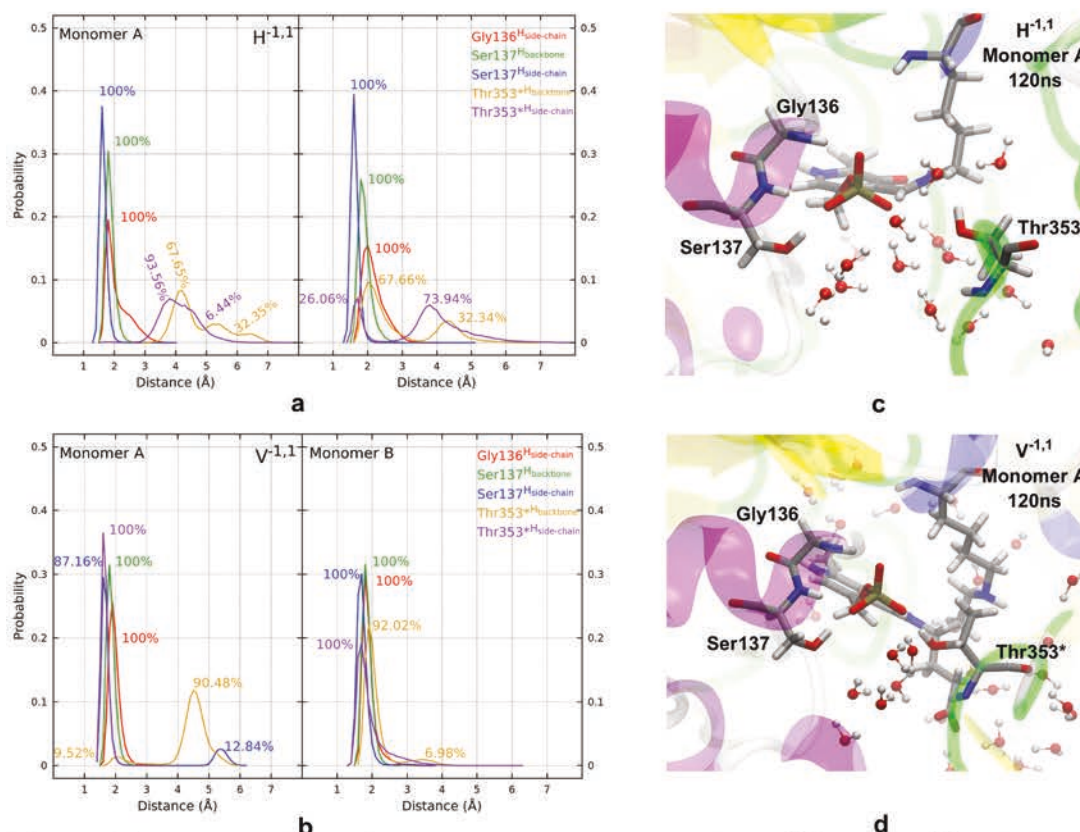
given by Ser137 in  $H^{-1,1}$ . In addition, the probability of hydrogen bond donation from Gly136 increases in both monomers when vigabatrin is involved (Figure 3.25b, sample  $V^{-1,1}$ ). The simulations point out that Thr353\* does not form a hydrogen bond with the PLP phosphate group in holo-GABA-AT (Figure 3.25a, sample  $H^{-1,1}$ ) and that water molecules can interact with the oxygen atoms of this phosphate group. However, when vigabatrin is bound to GABA-AT, there exists a hydrogen bond interaction between Thr353\* and the phosphate group. This indicates that an inter-monomeric interaction can be obtained via the involvement of the substrate (Figure 3.25b, sample  $V^{-1,1}$ ).



**Figure 3.24 :** The interactions between the side chains of Asp298 and His190 in vigabatrin bound GABA-AT samples. **a.** The variations of the minimum distances between  $H_{\epsilon}$  of His190 and the carboxyl oxygens of Asp298 (red), and  $H_{\delta}$  of Asp298 and  $N_{\delta}$  of His190 (blue) in sample  $V^{0,0}$ . **b.** The variations of the minimum distances between the side chain oxygen atoms of Asp298 and  $H_{\epsilon}$  (red) and  $H_{\delta}$  (green) of His190 in sample  $V^{0,1}$ . **c.** The variations of the distance between the side chain oxygen atoms of Asp298 and  $H_{\epsilon}$  (red) of His190 in sample  $V^{-1,0}$ . **d.** The variations of the minimum distances between the side chain oxygen atoms of Asp298 and  $H_{\epsilon}$  (red) and  $H_{\delta}$  (green) of His190 in sample  $V^{-1,1}$ .

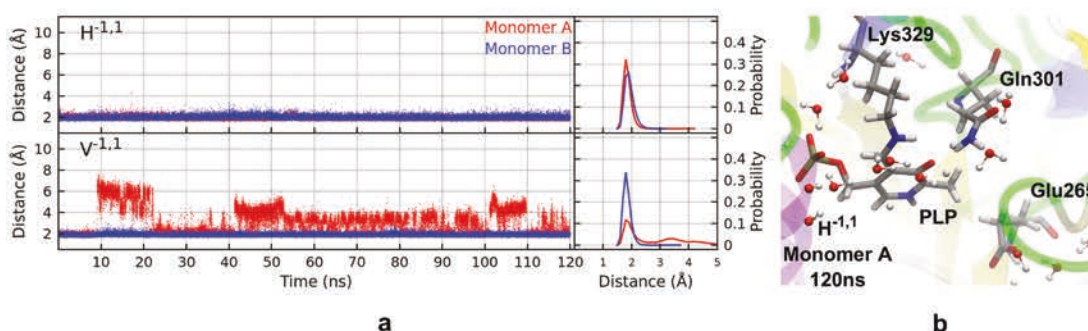
### 3.2.2.4 Carbonyl oxygen on the pyridine ring of PLP

The carbonyl oxygen on the pyridine ring of PLP can form a hydrogen bond with the surrounding residues in the active site. The amino group of Gln301 residue is close enough to donate hydrogen bonds to the corresponding oxygen atom in the crystallographic structure of GABA-AT. In order to investigate possible hydrogen bonds, the minimum distance between the hydrogen atoms on the amino group of the side chain of Gln301 and the carbonyl oxygen on the pyridine ring is calculated for samples  $H^{-1,1}$  and  $V^{-1,1}$ . The deviations in the minimum distances are depicted in Figure 3.26a. The results indicate that the amino group of Gln301 donates a hydrogen bond to the carbonyl oxygen in the holoenzyme  $H^{-1,1}$  and vigabatrin bound GABA-AT sample  $V^{-1,1}$ . However, it is observed that there are disruptions in monomer A of  $V^{-1,1}$ . The disruptions are due to the rotation of Gln301 side chain which ends up with an interaction between Gln301 and Glu265 side chains. Furthermore, the side chain of Gln301 can also be hydrogen bonded to the backbone oxygen atom of Ile267 when



**Figure 3.25 :** Interactions of the PLP phosphate group in  $H^{-1,1}$  and  $V^{-1,1}$ . Minimum distance probabilities between the phosphate group oxygen atoms and, respectively, Gly136, Ser137 and Thr353\* in **a.** holo-GABA-AT  $H^{-1,1}$  and **b.** vigabatrin bound GABA-AT  $V^{-1,1}$ . Typical three dimensional representation of the phosphate group oxygen atoms, Thr353\* residue and the surrounding water molecules **c.** in  $H^{-1,1}$  and **d.** in  $V^{-1,1}$ .

it does not interact either with the carbonyl oxygen on the pyridine ring or with the Glu265 side chain in  $V^{-1,1}$ . The flexibility of the Gln301 residue in vigabatrin bound GABA-AT may originate from the steric effects due to the involvement of vigabatrin.



**Figure 3.26 :** Interactions between the carbonyl oxygen atom on the pyridine ring and the Gln301 side chain. **a.** Variations of the minimum distances (left) and probability distributions (right) between the hydrogen atoms on the side chain of Gln301 and carbonyl oxygen on pyridine ring in  $H^{-1,1}$  and  $V^{-1,1}$ . **b.** Typical three dimensional representation of the interaction between PLP pyridine ring and the side chain of Gln301 in  $H^{-1,1}$  and  $V^{-1,1}$ .

It has been proposed that hydrogen bond donation from Gln301 would lower the pKa of the internal aldimine [52] and that Glu265, which is on the bottom of the coenzyme binding pocket, could have an intermediate function [52]. Our findings show here that, in vigabatrin bound GABA-AT, the Gln301 residue is more flexible than in holo-GABA-AT as resulting from a difference in interaction with Glu265.

### 3.2.2.5 PLP pyridine ring

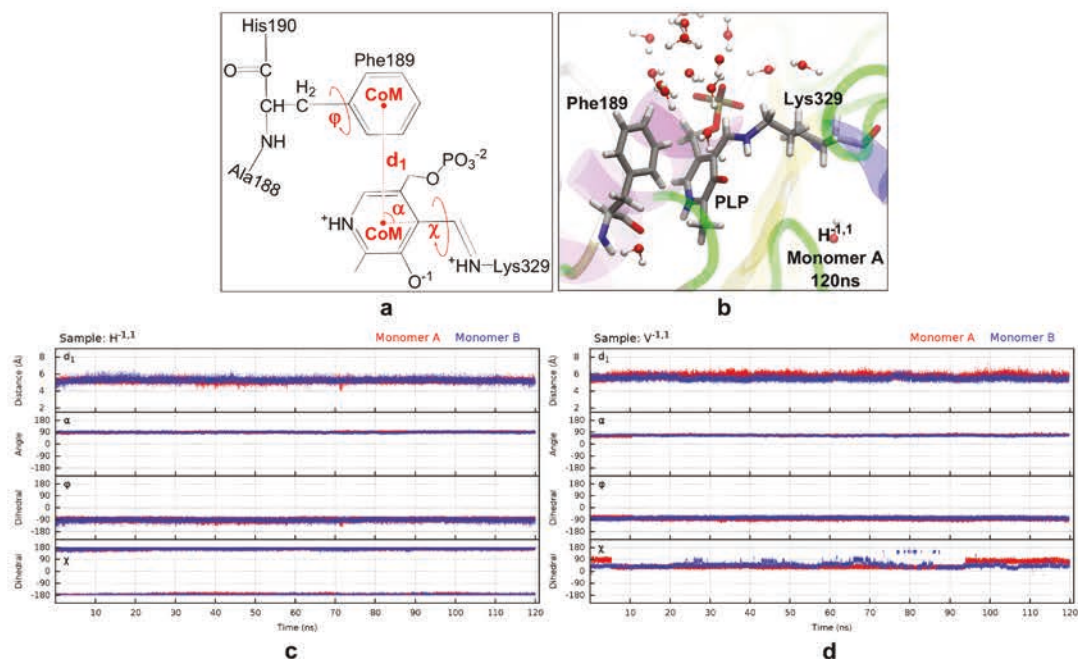
The six membered pyridine ring can be held in place by  $\pi - \pi$  stacking interaction with the phenyl ring of Phe189. The T-shape formed by the phenyl ring and the pyridine ring in the crystallographic structure could also have an effect on the activity of the enzyme. In order to investigate if the  $\pi - \pi$  stacking interaction is conserved along the simulations of  $H^{-1,1}$  and  $V^{-1,1}$ , the distance  $d_1$  between the center of masses of the six membered rings, the angle  $\alpha$  and the dihedral  $\chi$  in PLP, as well as the dihedral  $\varphi$  of Phe189 are calculated. The definitions of these parameters as well as their variations along our MD simulations are depicted in Figure 3.27. In holo-GABA-AT, the distance  $d_1$  is shorter than in vigabatrin bound GABA-AT. Furthermore, the distance  $d_1$  between the two rings is longer in Monomer A than in Monomer B for  $V^{-1,1}$ . The angle  $\alpha$  is smaller in  $V^{-1,1}$  than in  $H^{-1,1}$ . This is due to the alterations of the dihedral  $\chi$  of PLP pyridine ring. The phenyl ring is more flexible in holo-GABA-AT than in vigabatrin bound GABA-AT but there is no significant alterations of the dihedral  $\varphi$  in both samples.

It has been proposed that PLP is sandwiched between the Phe189 and Val300 residues to help PLP to be held in place in the active site [45]. The calculated distance, angle and dihedrals indicates that both the pyridine ring and Phe189 are not flexible in the active site. The tilt on the angle  $\alpha$  in  $V^{-1,1}$  stems from the involvement of vigabatrin which causes longer distance  $d_1$  and alterations on the dihedral  $\chi$ . However, the  $\pi - \pi$  stacking interaction between the pyridine ring and phenyl group of Phe189 is conserved in vigabatrin bound GABA-AT as well as in holo-GABA-AT.

### 3.2.2.6 Vigabatrin carboxyl group

The trajectories of vigabatrin bound GABA-AT are analyzed to investigate the interactions of the carboxyl group of vigabatrin with its surrounding residues. It is found that Arg192 guanidine group donates two hydrogen bonds to the carboxyl group

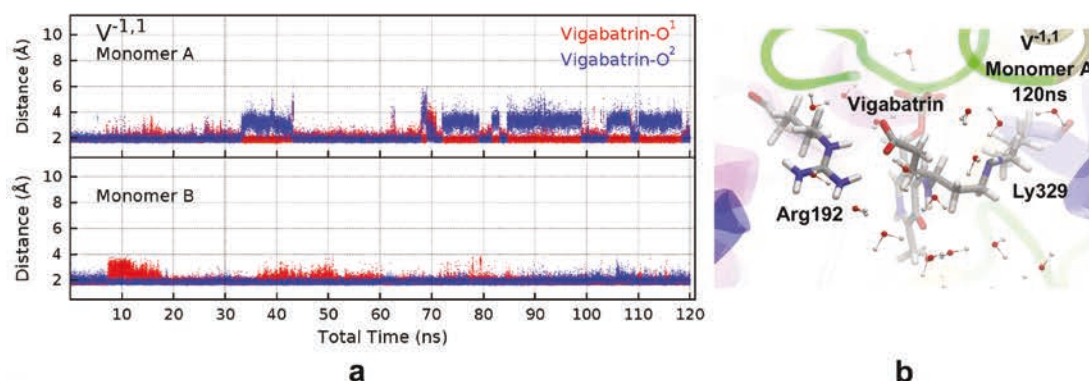




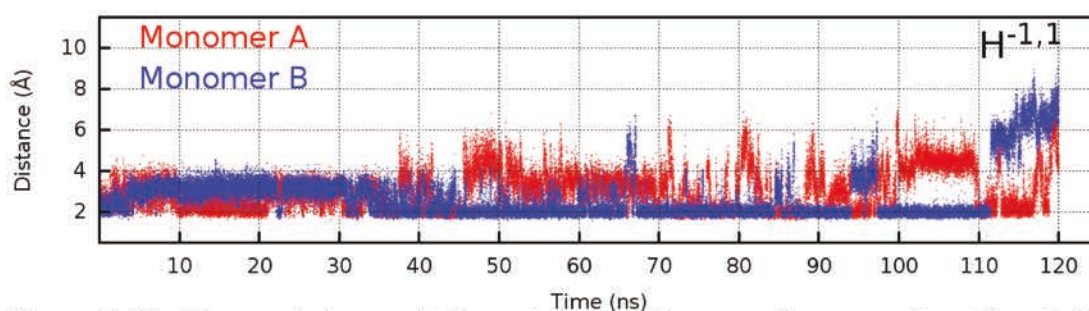
**Figure 3.27** : Interaction between the pyridine ring of PLP and the phenyl ring of Phe189. **a.** Definitions of the distance  $d_1$ , of the angle  $\alpha$ , and of the dihedrals  $\phi$ ,  $\chi$ . **b.** Typical three dimensional representation of the interaction between the pyridine ring of PLP and phenyl ring of Phe189. **c.** Variation of the interaction in holo-GABA-AT ( $H^{-1,1}$ ). **d.** Variation of the interaction in vibatrin bound GABA-AT ( $V^{-1,1}$ ).

oxygen atoms of vibatrin (Figure 3.28a). It is also observed that these hydrogen bonds can be disrupted along the simulations. However, the disruptions are due to water molecules that are in close vicinity of the carboxyl group of vibatrin. These water molecules form hydrogen bonds with the carboxyl group of vibatrin and the guanidine group of Arg192 when the interaction between these two groups is disrupted (Figure 3.28b). In addition, the carboxyl group of vibatrin and the side chain of Arg192 interact also with the side chain of Tyr348\* in  $V^{-1,1}$  due to the rotation of Tyr348\*.

It has been reported that the natural substrate GABA has a preferred binding mode both in experimental and theoretical works [28, 33]. The overall conformation of GABA within the active site is supposed to be controlled by the interaction between the carboxyl group of GABA and Arg192 [28]. Our results indicates that the conformation of the vibatrin in the active site can be controlled by the hydrogen bond with Arg192 in  $V^{-1,1}$  and the involvement of water molecules would yield more flexibility for the carboxyl group of vibatrin. In the case of  $H^{-1,1}$ , Arg192 guanidine group interacts with the surrounding water molecules and with the backbone of Tyr348\* (Figure 3.29).



**Figure 3.28 :** Interaction between the carboxyl group of vigabatrin and guanidine group of Arg192 in vigabatrin bound GABA-AT. **a.** Variations of the minimum distances between the oxygens of the carboxyl group of vigabatrin and the nitrogen atoms of the guanidine group along the simulations. **b.** Typical three dimensional representation of the interaction of vigabatrin with water molecules and Arg192.



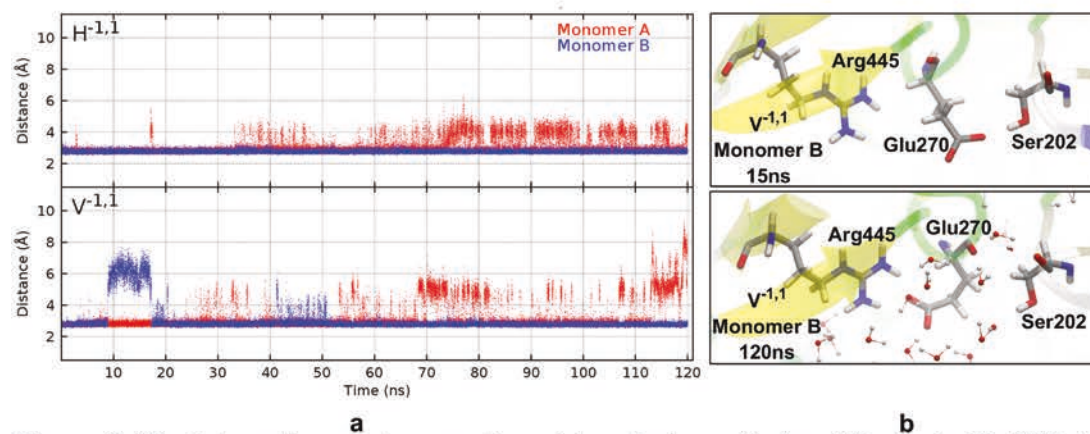
**Figure 3.29 :** The variations of the minimum distances between the side chain hydrogen atoms of Arg192 and backbone oxygen atom of Tyr348\* in sample  $H^{-1,1}$ .

### 3.2.2.7 Conserved Arg residue

Arg445 is a fully conserved residue in the  $\alpha$ -family of the aminotransferases [51] and it forms a salt bridge with the side chain of Glu270 in the crystallographic structures of GABA-AT. The minimum distances between the heavy atoms of the side chains of these residues (O and N atoms) are computed to observe if this salt bridge can be conserved along the simulations. The variations of the minimum distances are depicted in Figure 3.30. It is observed that the salt bridge can be disrupted on the Monomer A while it is conserved on the Monomer B of holo-GABA-AT. The same disruptions on Monomer A is also observed in the vigabatrin bound GABA-AT sample  $V^{-1,1}$ . However the interaction between these two residues is more flexible in the Monomer B of vigabatrin bound GABA-AT than in holo-GABA-AT. The disruption in the Monomer B of  $V^{-1,1}$  is due to the rotation of Glu270 residue which starts to interact with the side chain hydroxyl group of Ser202. Arg445 does not rotate towards



to the active site during the disruption and Glu270 rotates back to its original place to form the salt bridge again.



**Figure 3.30 :** Interaction between the side chains of Arg445 and Glu270 in holo-GABA-AT and vigabatrin bound GABA-AT. **a.** Variations of the minimum distances between the heavy atoms on the side chains along the simulations. **b.** Typical three dimensional representation of the salt bridge that is formed between Arg445 and Glu270.

It has been suggested that the positive charge of the Arg445 sidechain would be shielded by Glu270 [51] and that any approach of a dicarboxylic substrate would require the opening of this ion pair to allow substrate – Arg445 interaction [45]. Thus, a flexible interaction between these residues is to be expected. Since vigabatrin does not involve a secondary carboxyl group, Arg445 cannot rotate towards the active site and the disruption of the salt bridge is not stable along the entire simulation.

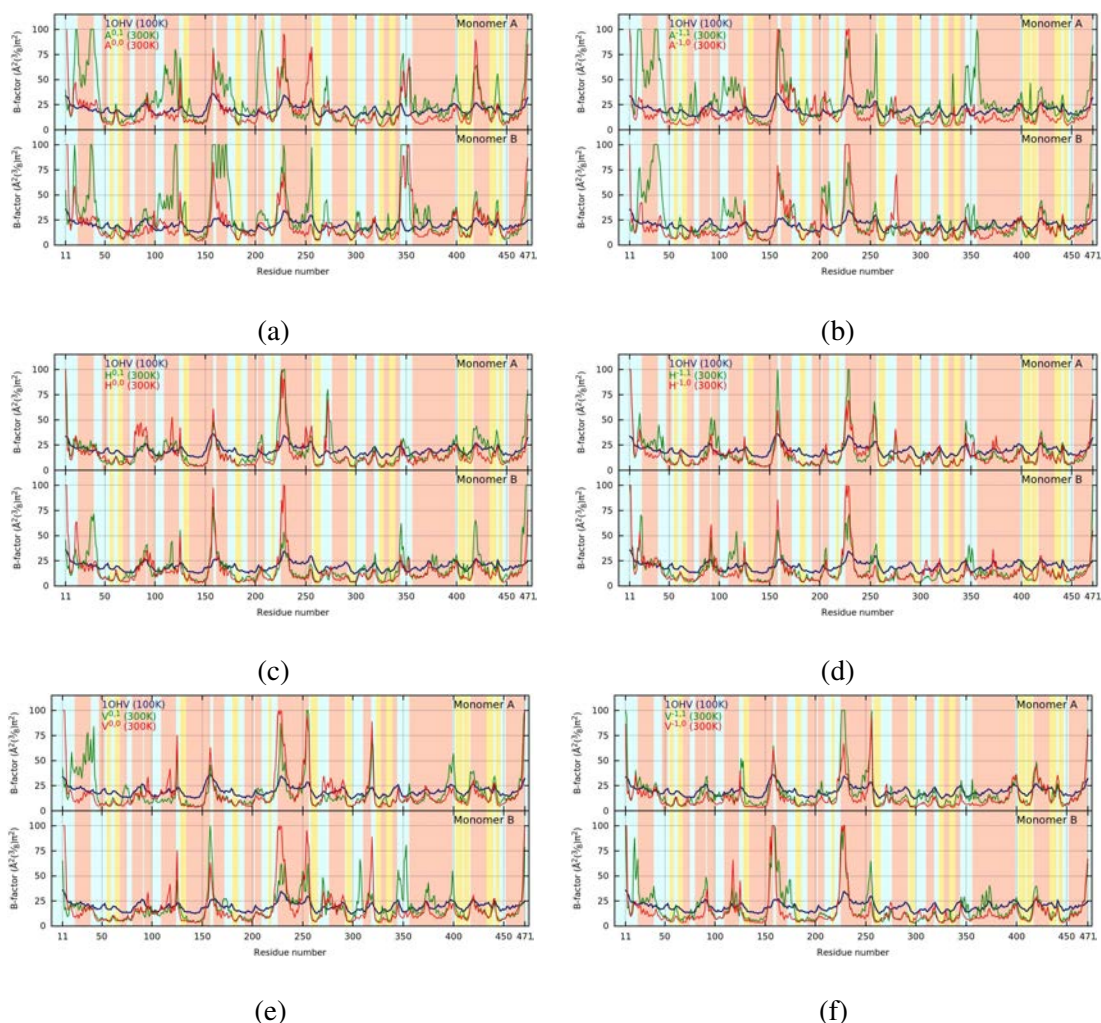
### 3.2.2.8 The roles of active site residues on the stability of the enzyme

It has been suggested that the nitrogen atom on the pyridine ring of PLP should be protonated in the enzyme [5–10]. Since the corresponding nitrogen cannot be protonated under physiological conditions [11], candidate residues that would help protonation are investigated. NMR studies on AspAT have revealed that this enzyme provides some extra interactions, such as with Asp222 and His143, to protonate the pyridine ring nitrogen [13]. In the case of GABA-AT, the active site residue Asp298 is the best candidate since the nitrogen atom on the pyridine ring faces towards the carboxyl group as in AspAT. PROPKA calculations reveal that the pKa of the buried Asp298 is evaluated at 2.8 with the greatest contribution to the ionization of Asp298 coming from His190. We have also found that Asp298 and His190 are conserved together in interaction in about 25% of all PLP dependent proteins available at the

Protein Data Bank. This indicates a possible joint effect on the protonation state of the pyridine ring nitrogen. Since the calculated pKa of the His190 residue is questionable, different GABA-AT samples are simulated with different protonation states. PROPKA calculations on the trajectories reveal that both the Asp298 and His190 residues should be ionized (with -1 and +1 charge, respectively) in apo-GABA-AT. The stability of the simulations are investigated by means of RMSd and B-factor calculations. The B-factors of the residues in apo-GABA-AT are higher than in both holo and vigabatrin bound GABA-AT (Figure 3.31). This result indicates that the inclusion of PLP stabilizes the samples via the interactions between the active site residues and the PLP cofactor.

It has been reported that mutations of Asp222 on AspAT caused reduced activity of the enzyme [48]. This potentially indicates the role of the interaction between Asp222 and PLP. In the case of GABA-AT, Asp298 is the conjugate residue for such a role. In order to reveal the effect of the protonation states of Asp298 and the PLP pyridine ring nitrogen in holo-GABA-AT, MD simulations are performed on different samples with different protonation states. Since the enzyme can provide extra interactions which would affect the protonation of the pyridine ring nitrogen as in AspAT [13], different protonation states for the His190 residue are also taken into account. It is observed that the pyridine ring and Asp298 can interact only if the side chains of Asp298 and His190 are ionized (sample  $H^{-1,1}$ ). The loss of the interaction in other samples would cause reduced activity of the holo-enzyme in a similar way as in AspAT. It was also suggested that the protonation of the nitrogen on the pyridine ring would cause an electron sink effect that can facilitate the enzymatic reactions [3, 5–9]. The results obtained from the MD simulations of holo-GABA-AT samples also indicate that the pyridine ring of PLP should be protonated in order to achieve Asp298-PLP interaction along the simulation. The role of the His190 is to preserve the interaction between Asp298 and PLP pyridine ring since the imidazole ring of His190 should be positively charged. It is noteworthy that the stability of the hydrogen bonding network of Asp298 is significantly affected when the ionized Asp298 is sandwiched by positively charged His190 and PLP pyridine ring. All these results point out that the secondary role of the positively charged His190 in the stabilization of the hydrogen bonds between Asp298 and the surrounding residues can also influence the enzymatic activity. In





**Figure 3.31** : The computed B-factor values in simulated GABA-AT samples. **a.** Apo-GABA-AT samples  $A^{0,0}$  and  $A^{0,1}$ . **b.** Apo-GABA-AT samples  $A^{-1,0}$  and  $A^{-1,1}$ . **c.** Holo-GABA-AT samples  $H^{0,0}$  and  $H^{0,1}$ . **d.** Holo-GABA-AT samples  $H^{-1,0}$  and  $H^{-1,1}$ . **e.** Vigabatratin bound GABA-AT samples  $V^{0,0}$  and  $V^{0,1}$ . **f.** Vigabatratin bound GABA-AT samples  $V^{-1,0}$  and  $V^{-1,1}$ . The secondary structures are represented with different colors; the  $\beta$ -sheets are colored with yellow and the  $\alpha$ -helices are colored with red.

addition to the interaction between PLP and Asp298, the activity of the enzyme can also be affected by the presence or absence of the  $\pi - \pi$  interaction between Phe189 and the PLP pyridine ring. The mutation of W101A in EhPSAT caused the loss of the enzymatic activity while the W101F mutation showed that Phe could efficiently stack with the pyridine ring [47]. Our MD simulations point out that the PLP pyridine ring is stacked by Phe189 regardless of the different protonation states. The role of the Phe189 residue is simply to hold PLP in place and it can affect the enzymatic activity indirectly since the stacking is always achieved.

### 3.2.2.9 The effect of inactivator on the active site

The inactivation of enzymes can be sometimes achieved by inhibitors that can either covalently modify the active site residues or produce covalently modified coenzymes. In the case of vigabatrin, the active site of the GABA-AT is covalently modified without the release of the cofactor PLP since the inactivator is also covalently bound to PLP. Our MD simulations on the vigabatrin bound GABA-AT sample  $V^{-1,1}$  point out that the interactions between the active site residues in holo-GABA-AT are conserved. However the probabilities of these interactions are found to be increased in most cases. The hydrogen bond network of Asp298 becomes more rigid when vigabatrin is involved. The loss of flexibility is simply due to the steric effect that is aroused by vigabatrin. This change in the steric effect results a more stable hydrogen bond network between Asp298, His190, Asn140 and Val300 than in holo-GABA-AT. The phosphate group oxygen atoms on PLP are held in place by at least three hydrogen bonds in holo-GABA-AT and Thr353\* does not donate any hydrogen bond to the group as suggested elsewhere [45, 52]. On the other hand, the inclusion of vigabatrin yields to a stronger network of hydrogen bonds between the phosphate group oxygen atoms and Thr353\* in one monomer. This indicates that an intermonomeric interaction is achieved. Other alterations are the interaction between the Gln301 side chain and the carbonyl oxygen of PLP, and a tilt of the angle between the center of mass of the phenyl ring of Phe189 and the PLP pyridine ring. However, these alterations do not destabilize the enzyme. It is also important to note that the interactions in one of the two monomers in vigabatrin bound GABA-AT are more flexible than in the other one. The intermonomeric interactions that are achieved by the inclusion of vigabatrin might be the cause of the different flexibilities among the two monomers.

### 3.2.2.10 Negative cooperativity

The flexibility differences between the two monomers of the GABA-AT samples are due to the asymmetry between the monomers. This asymmetry which yields to a negative cooperativity in GABA-AT has been reported before [8]. The dimer of the GABA-AT has two active sites but only one of them is functional [8, 17]. Thus, the differences between the two monomers that we obtain from our MD simulations could be simple representations of this negative cooperativity.

### 3.2.2.11 Remarks on the catalytic mechanism of GABA-AT

The catalytic mechanism of GABA-AT is usually divided into two parts as in Ornithine aminotransferase (OAT): a first and a second half-reaction [14, 15]. In the first half-reaction, the natural substrate GABA is converted to succinic semialdehyde and it involves the conversion of PLP to pyridoxamine phosphate (PMP) [14, 16]. The PMP molecule in the active site is then bound to the enzyme again during the second half-reaction [14, 16, 17]. The conversions of PLP to PMP in the first half-reaction and of PMP to PLP in the second half-reaction point out that holding the cofactor in the active site is necessary for catalytic activity. The interactions of PLP in the active site of holo and vigabatrin bound GABA-AT is investigated by means of MD simulations. The nitrogen on the pyridine ring of PLP interacts with the side chain of Asp298 while phosphate group of PLP is held in place via the donation of at least three hydrogen bonds from the neighboring active site residues. Gln301 donates hydrogen bond to the carbonyl oxygen of PLP and the pyridine ring is in  $\pi - \pi$  stacking interactions with the side chain of Phe189. The results that are obtained from MD simulations point out that even though one of the monomers of holo-GABA-AT and vigabatrin bound GABA-AT is more flexible than the other one, the interactions of PLP are always conserved, either directly or through water molecules that provide the necessary hydrogen bonds to hold the cofactor in the active site.

The second half reaction involves a succinic semialdehyde with two carboxyl groups. Many substrate analogues have been reported in the literature with a secondary carboxyl group [45]. In the case of OAT, which is very similar to GABA-AT, it has been suggested that there should be a second arginine residue to interact with this secondary carboxyl [232]. This second arginine would form a salt bridge with a glutamate residue when the substrate has only one carboxyl group [45, 232, 233]. In the case of GABA-AT, the second arginine is Arg445 which forms a salt bridge with Glu270. Our MD simulations on holo-GABA-AT and vigabatrin bound GABA-AT point out that this salt bridge is usually conserved along the simulations. Disruptions of the Arg445-Glu270 ion pair are found to be due to the rotation of the Glu270 residue. During disruption, the Glu270 residue is found to interact with Ser202 which was also considered in an active site model study on GABA-AT [52]. In the presence of a substrate with two carboxyl group, the disruption of the salt bridge via the rotation

of Glu270 would release Arg445 in a way that it can start to interact with the second carboxyl of the substrate. The role of Glu270 in GABA-AT was suggested to be a switch between the first and the second half-reactions as in OAT [234]. The results obtained from our MD simulations even without two carboxyl groups on the substrate provide information about the respective roles of Glu270 and Ser202 residues.

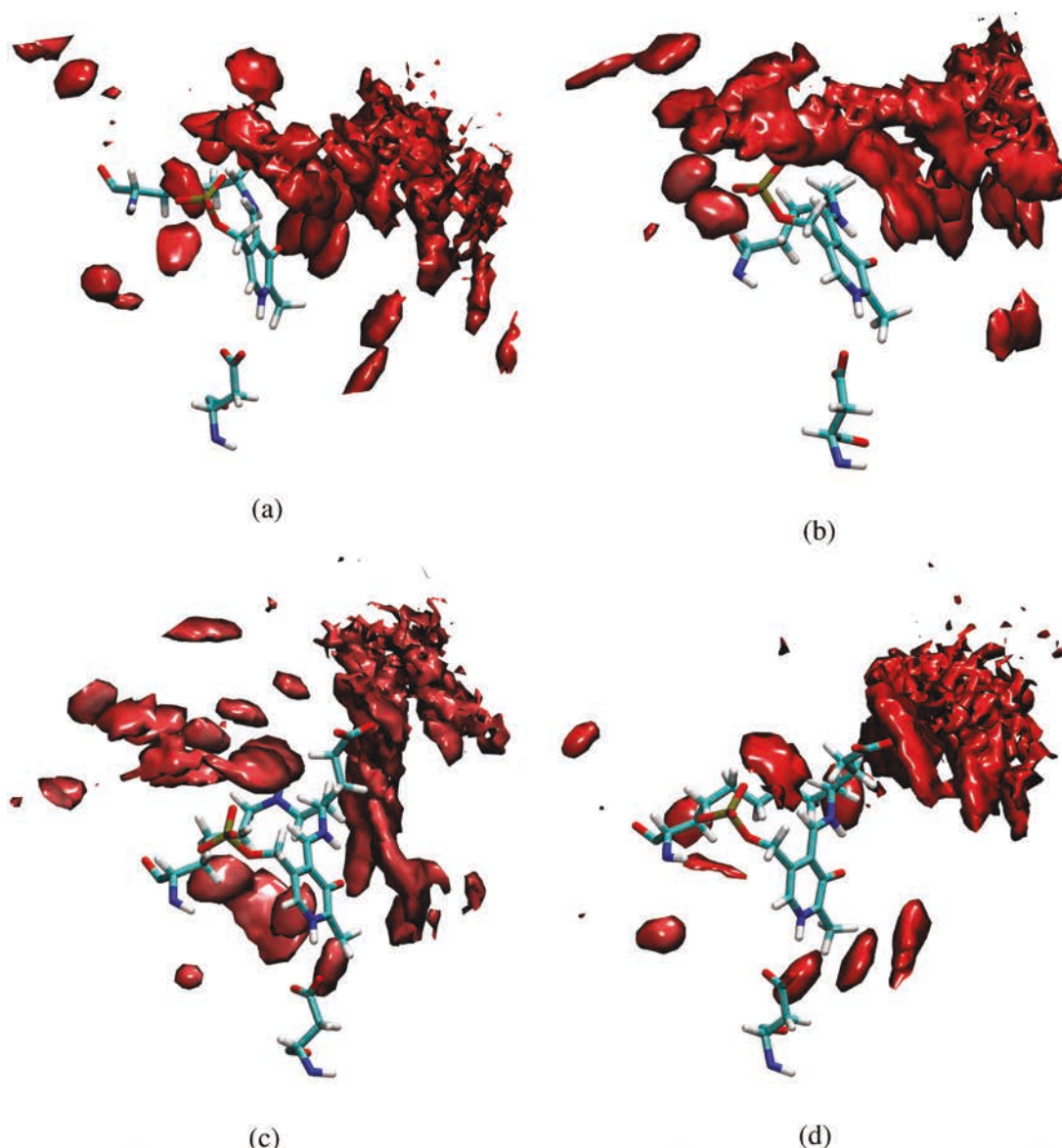
#### **3.2.2.12 Solvent access to the active site**

In order to investigate solvent accessibility to the active site, the solvent accessible area is also investigated by GIST [235, 236] calculations using cpptraj [230] and the results are depicted in Figure 3.32. Our results point out that the solvent can access the active site through a cavity in holo-GABA-AT. This cavity corresponds also to the region from which the substrates get into the active site of the enzyme. On the other hand, the inclusion of vigabatrin causes desolvation which in turn prevents the solvent to access the phosphate group of PLP. Hydrogen bond donations to the phosphate group oxygen atoms of PLP from the surrounding residues increase due to the desolvation in vigabatrin bound GABA-AT. This desolvation would also prevent any other substrate to get into the active site and ensures the inactivation of the protein. The inhibition of GABA-AT by vigabatrin starts with the formation of an external aldimine that is similar to the initial reaction of GABA in GABA-AT during the first half-reaction [55]. The simulations of  $V^{-1,1}$  point out that the conformation of vigabatrin in the active site can be controlled by Arg192. The role of Arg192 could be similar in the case of GABA as suggested previously [28].

#### **3.2.3 Concluding remarks**

Molecular dynamic simulations GABA-AT samples are studied in order to understand the main properties that define the activity of the enzyme. The simulations of apo, holo and vigabatrin bound GABA-AT samples enabled us to elucidate the ionization states and the roles of the active site residues in GABA-AT and, by extension, some general aspects of PLP dependent enzymes.

The nitrogen on the pyridine ring of PLP is suggested to be protonated in the active site of PLP dependent enzymes. However, NMR studies state that it cannot be protonated under physiological conditions in aqueous solution. Our MD simulations of apo, holo



**Figure 3.32 :** The solvent access to the active site. The results obtained with GIST calculations **a.** on Monomer A of sample  $H^{-1,1}$ , **b.** on Monomer B of sample  $H^{-1,1}$ , **c.** on Monomer A of sample  $V^{-1,1}$ , **d.** on Monomer B of sample  $V^{-1,1}$ . The grid increments along each coordinate axis is  $40 \times 40 \times 40$ , grid spacing is  $0.5 \text{ \AA}$  and center of the grid is chosen as imino nitrogen atom.

and vigabatrin bound GABA-AT samples shed lights on the protonation states of the active site residues. It is suggested that Asp298 and His190 should be ionized and that negatively charged Asp298 induces the existence of a proton on the nitrogen atom of the pyridine ring of PLP. Therefore, as found in AAT by NMR studies [12, 13], the protonated nitrogen atom of the pyridine ring of PLP in GABA-AT should facilitate the first half reaction by causing an electron sink effect [3, 5, 9]. Since the  $pK_a$  of PLP is different in solution and in the protein, it is the Asp298/His190 diad that shifts the

pKa of PLP which, in turn, triggers PLP activation for the first half-reaction. This is the main factor that defines the reactivity of GABA-AT.

In order to achieve the second half reaction in GABA-AT, PLP/PMP should be held in place in the active site. Our MD simulations of GABA-AT samples exhibit strong hydrogen bond network between the protein and PLP. The phosphate group is held in place by the donation of at least three hydrogen bonds while the carbonyl oxygen of the pyridine ring interacts with Gln301. Phe181 forms a  $\pi - \pi$  stacking interaction with the pyridine ring and works as a gate keeper with the assistance of Val300. The interactions between the active site residues (e.g., Asp298, His190, etc.) also help to hold PLP in place.

It is found that in almost 25% of the PLP dependent proteins available in the Protein Data Bank, an Asp residue is faced towards the pyridine ring while a His residue is in close contact with this Asp residue. These PLP dependent enzymes would display similar reactivity properties as in GABA-AT. In contrast, the lack of an Asp residue or of a positively charged residue that facilitates an ionized Asp might cause different reactivity issues. Since the protonation of the pyridine ring nitrogen atom is the main factor that defines the reactivity of the enzyme, the interaction of this atom with its surrounding should be resolved in order to define the reactivity of PLP dependent enzymes.

It is found that the interactions of PLP in holo-GABA-AT sample are not altered by the involvement of vigabatrin. GIST results show that once vigabatrin binds to PLP, water cannot access to the active site as much as in holo-GABA-AT samples. Our results confirm that vigabatrin is a good inhibitor since it does not disrupt the main interactions of PLP and it does not allow another substrate to get into the active site.

### **3.3 The Effect of Surrounding Residues on Inactivation Mechanism of GABA-AT with $\gamma$ -vinyl GABA**

In order to get more reliable features of the potential energy surface, such as barriers and thermodynamics, a substantial part of the enzyme active site is needed to be included in the model. The protein surrounding may affect the description of the reaction in two main ways; the steric effect imposed by the surrounding residues on

the model, and the long-range polarization that can affect the reaction energetics. The steric effects are modelled by the inclusion of the surrounding residues within the model and by blocking some atom coordinates where the truncation is made. The model will gain flexibility with the increase of the size and will hopefully produce better accurate results than a small and rigid model [58]. The long-range polarization is modelled with the polarizable continuum technique which provides a homogeneous medium of some dielectric constant. The effect of the solvation on the core of the reaction will decrease with the increase of the model size [59,60]. If the steric effects and long-range polarization are taken into account, a sufficiently large model can provide optimum mimicry by keeping the modelled reaction as close as possible to the experimental one. To understand the effect of the active site residues on the reactions involving PLP, which was proposed to have a protonated N atom on the pyridine ring to facilitate the enzymatic reactions [3,5–10], quantum cluster models have been used. This approach has been applied to many class of enzymes [59–73] and has proven to be quite useful to study enzymatic reactions.

NMR studies of  $^{15}\text{N}$  labeled PLP pyridine ring in aqueous solution proposed that the pyridine ring of PLP is mostly unprotonated under physiological conditions [11]. Besides, in a quantum modelling study on  $\omega$ -transaminase reaction mechanism, geometry optimizations end up with the transfer of the proton from N atom of the pyridine ring of PLP to carboxyl group of Asp259 [65]. On the other hand, the NMR studies on complexes of silylated aldenamine in the presence of an Asp222 model Boc-Asp-OrBu [12] and  $^{15}\text{N}$  labeled AspAT [13] suggested that the N atom on the pyridine ring is protonated in AspAT. The protonation of the nitrogen on the pyridine ring of PLP has been suggested to facilitate the enzymatic reactions in PLP dependent enzymes via an electron sink effect [3,5–10]. It was also proposed that the protonation of the pyridine ring N atom can be achieved with the assistance of Asp222 and His143 in AspAT [13]. Our Molecular dynamics simulations of GABA-AT has revealed the importance of the Asp298 and His190 residues on the protonation of the N atom on the pyridine ring (see Chapter 3.2). In this study, we show that the interaction between PLP and Asp298 can be achieved or conserved only with a protonated pyridine ring, a negatively charged Asp298 side chain and a positively charged His190 (see Chapter 3.2).



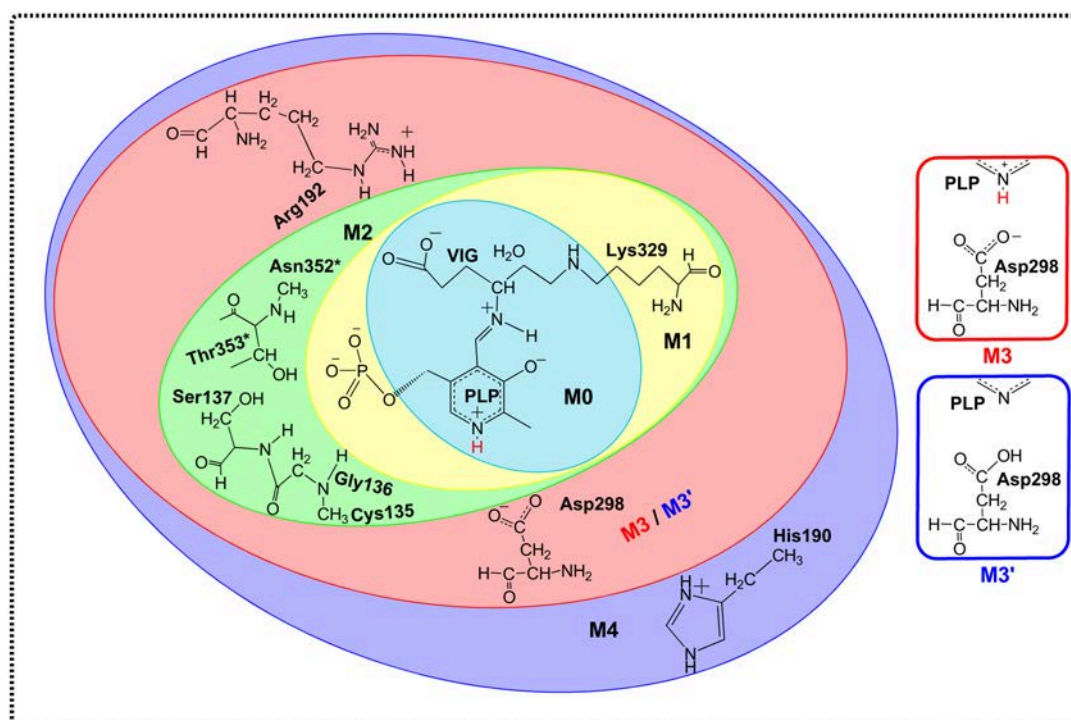
To investigate the effect of the active site and its surrounding residues on the inactivation mechanism of GABA-AT, a comprehensive study is carried out in conjunction with the previous experimental and theoretical results. What is the effect of the protonation state of the PLP pyridine ring N atom on enzymatic reactions? Which factors are enforcing the N atom of the the pyridine ring to be protonated in the active site? What are the effects of the surrounding residues on the geometrical parameters of the transition state structures? Which residues should be involved in modelling studies to get the model that provides optimum mimicry of the active site of GABA-AT? In order to answer these questions, Michael Addition reaction which was reported as the most probable inactivation mechanism of  $\gamma$ -vinyl-GABA with GABA-AT [34, 46] was investigated using quantum cluster approach. The effects of the surrounding residues, the protonation of N atom on the pyridine ring and how to enforce this protonation was evaluated with the inclusion of different surrounding residues.

### 3.3.1 Computational details

The calculations were performed using the hybrid DFT functional B3LYP [131, 132, 134, 139, 140] as implemented in the G09 [160] program package. Geometry optimizations were performed with the 6-31G(d,p) basis set and single point energy calculations were performed with the larger basis set 6-311G+(2d,2p) to get more reliable energies. Solvation effects were taken into account by single-point energy calculations at the same level of theory using the integral equation formalism variant of the polarizable continuum model (IEFPCM) [157, 158, 193, 194]. Although the active site of GABA-AT is known to contain water molecules and the modelled reaction is proposed to be assisted with a water molecule [34], solvents with a lower dielectric constant ( $\epsilon = 4$ ) are also commonly considered to be a good representation of the protein environment [58, 62, 237]. Therefore, a polarizable medium with two different dielectric constants,  $\epsilon = 80$  and  $\epsilon = 4$ , has been chosen to evaluate the effect of the environment. Frequency calculations were also carried out to identify the stationary points. Although frozen coordinates will rise few small imaginary frequencies, some atoms (mainly the backbone atoms of the residues) were kept frozen during geometry optimizations to restraint the atoms to their initial positions. Similar procedures has



been widely used in many applications. They are referred to a cluster approach [58–62, 237]. All of the energies that are reported in the text are electronic energies, unless otherwise stated.



**Figure 3.33 :** Two dimensional representations of model clusters of increasing size. Different protonation states of Asp298 and PLP in M4 and M4' are given on the right hand side in colored squares.

**Preparation of the models.** The initial geometries of the cluster models are extracted from the molecular dynamics trajectories at 600ps that were obtained with the MD simulations of  $\gamma$ -vinyl-GABA bound GABA-AT (see Chapter 3.2). In order to provide stabilization, the residues that are included in our cluster models were chosen in conjunction with the literature that reports interactions in the active site of PLP dependent enzymes [13, 28, 45, 46, 48, 52, 232] and our MD simulation results (see Chapter 3.2). Models of increasing size are shown in Figure 3.33. In order to simplify the models, some of the residues were truncated and only some part of the side chains were included during the calculations. As cluster size increases the backbone of the truncated residues was appended. The backbone atoms of Gly136 and Thr353\* were suggested to be in contact with the phosphate group oxygen atoms of PLP [45, 46, 52]. To include these atoms in our models, the adjacent residues were represented and

modelled with a methyl group (ie. Asn352\*, which forms a peptide bond with Thr353\* backbone N) and kept frozen during optimizations.

### 3.3.2 Results and Discussion

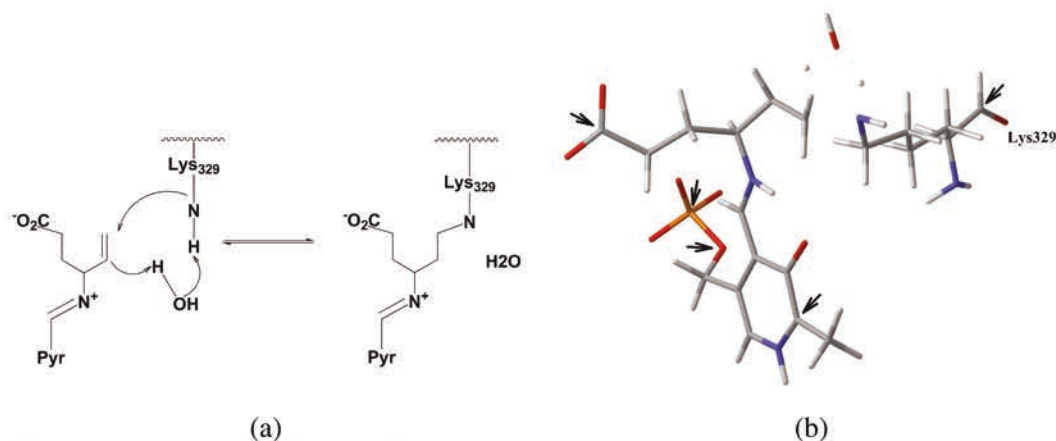
The mechanism based inactivators are designed on the basis of mechanistic details, hence different sizes of clusters were modeled to reveal the effect of the surrounding residues on the Michael addition reaction mechanism of  $\gamma$ -vinyl-GABA with GABA-AT. The Michael addition reaction, the second part of the inactivation mechanism of GABA-AT with  $\gamma$ -vinyl-GABA, takes place after the external aldimine formation [34, 46]. In order to explain the details of the inactivation mechanism of GABA-AT both the active site residues and the different protonation states were questioned with cluster models.

#### 3.3.2.1 Michael addition reaction with cluster models

The Michael addition reaction of  $\gamma$ -vinyl-GABA is a one-step reaction which is facilitated by the assistance of a water molecule [34,46]. The reaction takes place with the nucleophilic addition of the side chain N atom of Lys329 residue to the terminal C atom of the vinyl group of the substrate concomitantly with a 1,3 proton transfer (Figure 3.34).

**Truncated phosphate group and Lys329.** The first model, Model 0 (**M0**), is designed in conjunction with the structure in the previous computational study where the inactivation mechanism of GABA-AT was under investigation and the carboxyl group of  $\gamma$ -vinyl-GABA was protonated [34]. Herein,  $\gamma$ -vinyl-GABA is modelled in an ionic form different than the previous study. In previous computational studies, the phosphate group of PLP has been represented with a methyl group [2,4,33,34,199,200] while the Lys329 residue has been represented as methylamine [33,34,196–198]. The same representations are used for Lys329, and for the phosphate group of PLP in the smallest model. Single point energy calculations with a higher basis set on optimized structures led an activation energy barrier of 34.5 kcal/mol (Figure 3.35) while the activation free energy barrier with a neutral carboxyl group of  $\gamma$ -vinyl-GABA was reported as 40.7 kcal/mol in the previous study [34]. It was observed that the barrier was decreased when the solvent effect was taken into account (27.0 kcal/mol and 28.5

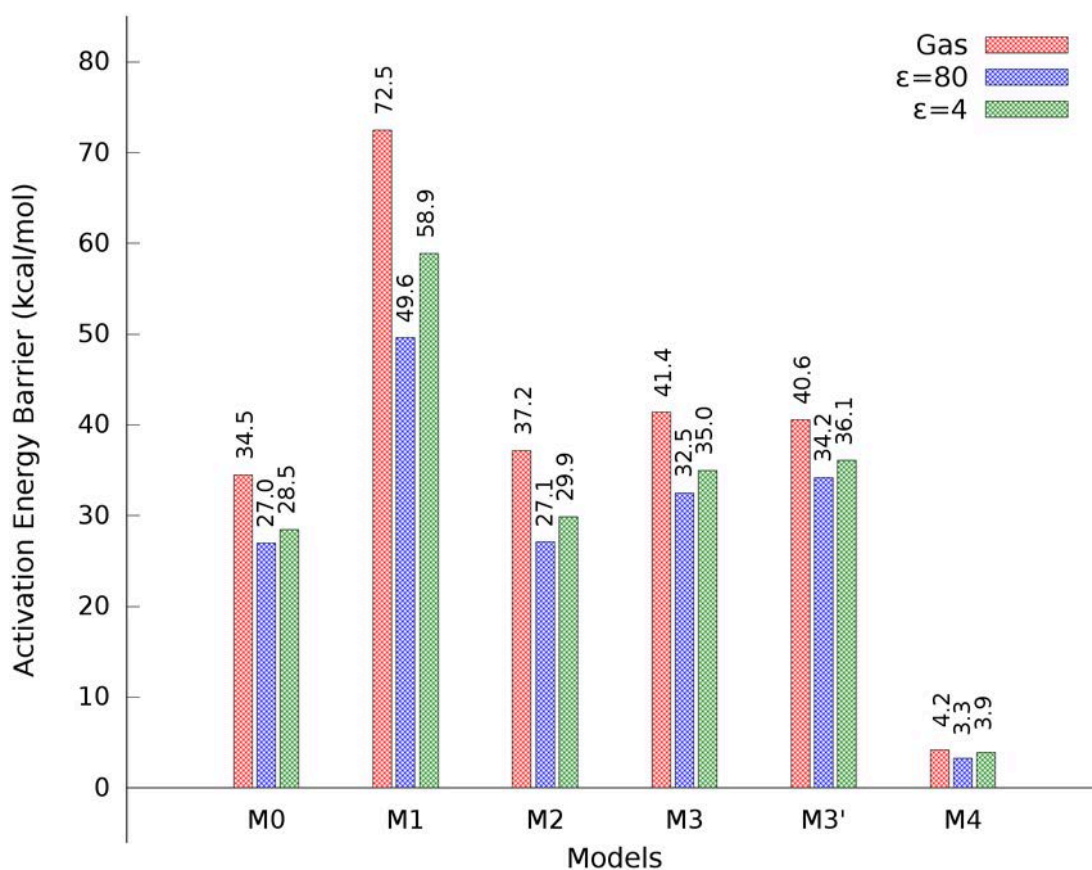
kcal/mol with  $\epsilon = 80$  and  $\epsilon = 4$  respectively). In the previous study, the energy gap between the gas phase and the solvent surfaces was reported at about 26 kcal/mol when  $\epsilon = 80$  [34]. However, in the presence of ionic carboxyl group this gap decreased to 7.5 kcal/mol. Nevertheless, the difference between the gas phase and solvent energy surfaces reveals that the model structure is affected with the polarity of the environment and is thus not large enough to mimic the active site. Thus, a larger model is required to obtain more accurate results.



**Figure 3.34 :** The Michael addition reaction in GABA-AT with  $\gamma$ -vinyl GABA. **(a)** Two dimensional representation of reaction. **(b)** Three dimensional representation of Michael addition reaction with model **M1**. Frozen atoms are shown with black angle headed arrows.

**Inclusion of phosphate group and Lys329 residue.** In order to obtain better representation of the active site residues, of the truncated Lys329 and the phosphate group of PLP were removed in Model 1 (**M1**) (Figure 3.34b). The absence of the surrounding residues and phosphate group may lead to disruptions on the geometry of the reactants and transition state. The phosphate group O and P atoms, the backbone C atom of Lys329 residue and also one of the C atoms on the pyridine ring were kept frozen during calculations to avoid possible disruptions (see Figure 3.34b). The activation energy barrier was found 72.5 kcal/mol with **M1** (Figure 3.35), which is about 40 kcal/mol higher than the barrier that was calculated for **M0**. It was observed that the phosphate group oxygen atoms form hydrogen bonds with the water molecule which accepts another hydrogen bond from the side chain of Lys329 in the reactant that was obtained with Intrinsic Reaction Coordinates (IRC) calculations. The presence of hydrogen bonds stabilizes the reactant, the absence of the corresponding interactions in transition state structure increases the activation energy barrier. In addition to the

high activation energy barrier with gas phase calculations, it was found that the barrier decreased drastically when the solvent effect was taken into account. The inclusion of the phosphate group, which was frozen due to the disruptions of the bonds, may be the main reason of the increase in the activation energy barriers and may hinder the decrease of the gap between the gas phase-solvent energy surfaces having difference dielectric constants. Addition of the negatively charged phosphate group and keeping it frozen during the geometry optimizations is not a convenient option for modelling reactions within the protein environment. In the absence of stabilizing interactions of the opposing residues, modelling charged side chains with methyl group provides better energy values.

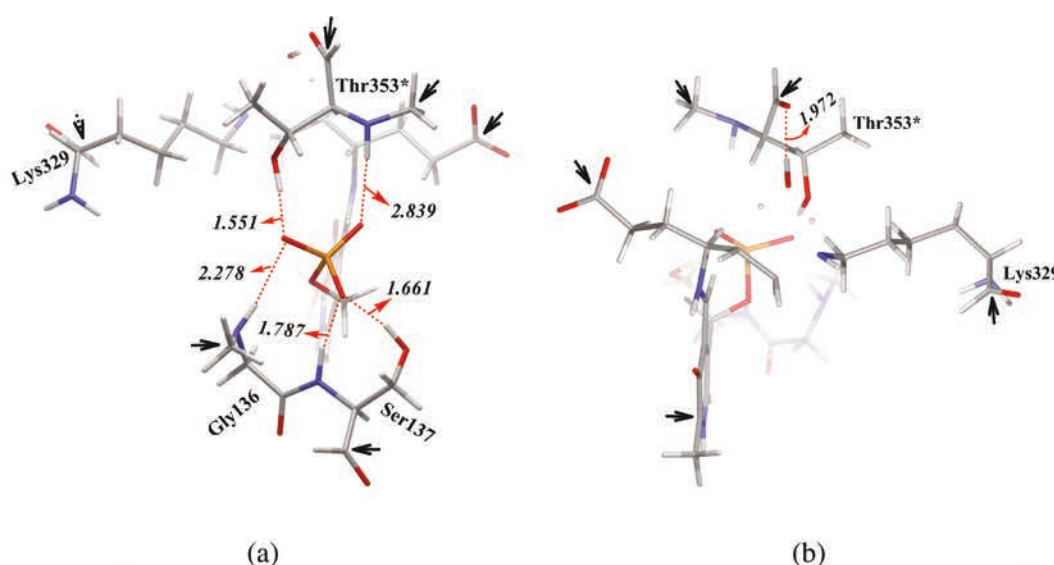


**Figure 3.35** : Activation energy barriers of model clusters. The gas phase energies (red) are calculated with 6-311G(2d,2p) basis set while the solvent energies with  $\epsilon = 4$  (blue) and  $\epsilon = 80$  (green) are calculated with 6-31G(d,p) basis set.

**Providing hydrogen bonds to phosphate group.** The energy gap between the potential energy surfaces in the previous model lead us to increase the size in Model 2 (**M2**, Figure 3.36). The active site residues Thr353\* of the other monomer, Gly136 and Ser137 were included in **M2** due to the close proximity to the phosphate group oxygen



atoms and their ability to form stabilizing hydrogen bonds [45, 46, 52](Figure 3.36). The backbone N atoms of Gly136 and Thr353\* are close enough to form hydrogen bonds with the phosphate group oxygens of PLP [45, 46, 52]. Therefore their adjacent residues, Cys135 and Asn352\* respectively, were modeled as methyl groups and their C atoms were kept frozen during the calculations. Since the interactions between the oxygens of the phosphate group and the surrounding residues are included in the model, the phosphate group was not kept frozen (Figure 3.36). The obtained results pointed out another H-bond which was formed between the Thr353\* backbone oxygen and the assisting water molecule (Figure 3.36b). The obtained barriers with gas phase calculations and calculations using different media with  $\epsilon = 80$  and  $\epsilon = 4$  are depicted in Figure 3.35. The presence of H-bond interactions between the phosphate group and the surrounding residues in **M2** caused a rapid decrease of the activation energy barrier in all media (Figure 3.35) relative to the previous model **M1**. It is apparent that when the negatively charged phosphate group is included in the model, including the surrounding residues to interact with it instead of keeping it frozen may provide more accurate results. However, the effect of the polarity of the environment on the reaction is still high to conclude optimum mimicry of the active site (Figure 3.35) even though the differences between the gas phase-solvent phase energy surfaces decreased according to the previous model **M1**.

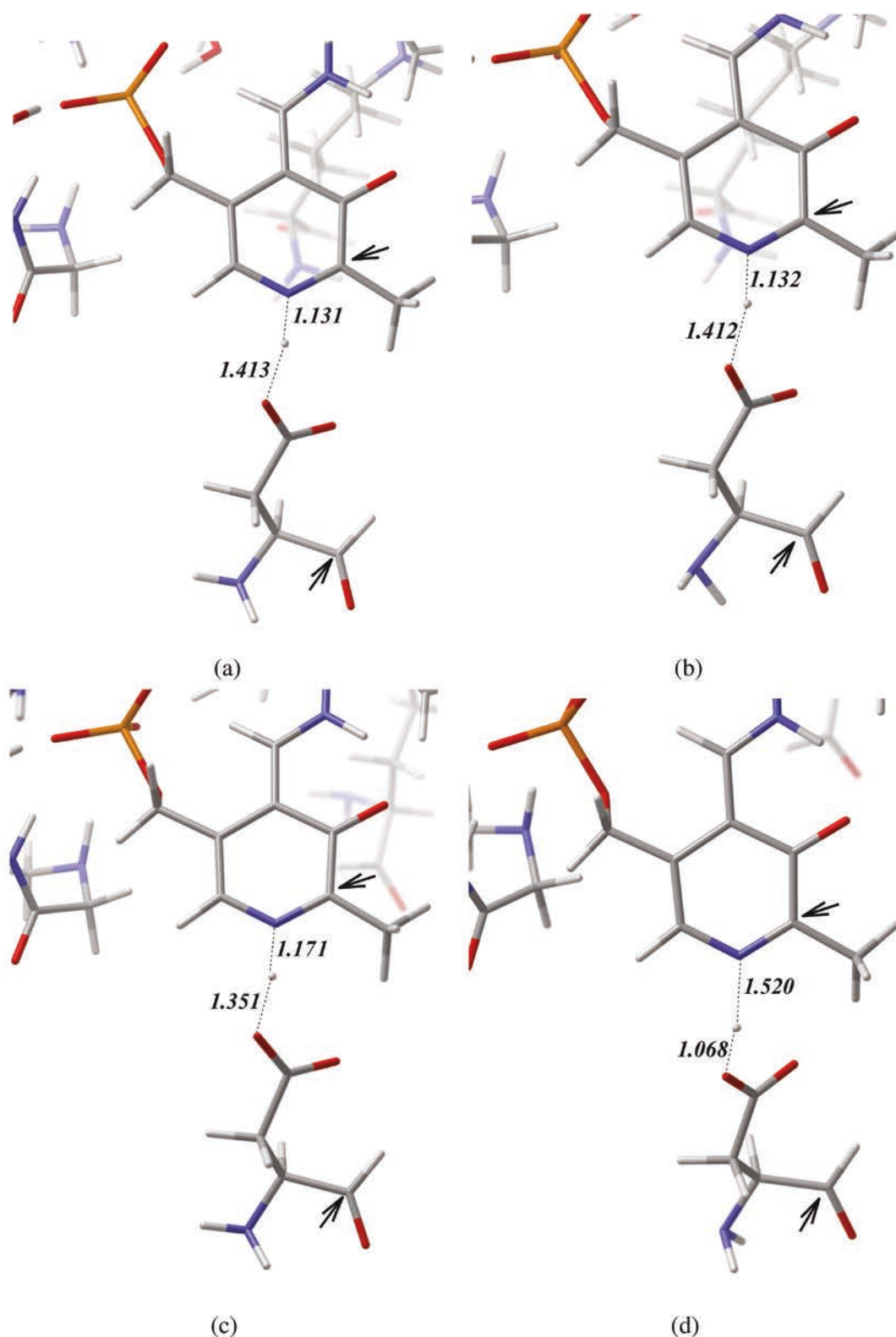


**Figure 3.36 :** The Michael addition reaction with model cluster **M2**. Three dimensional representation of **(a)** phosphate group interactions, and **(b)** transition state center. Frozen atoms are shown with black angle headed arrows.

**Inclusion of Arg192 and Asp298.** In the subsequent cluster models, the Michael addition reaction was modelled with the involvement of Arg192 and Asp298 in order to keep a total charge similar to the previous models (**M1** and **M2**). The carboxyl group of the substrate is suggested to interact with the guanidino group of Arg192 [45,46,52] and possible H-bonds may decrease the activation energy barriers. Furthermore, it was reported that Asp298 interacts with the pyridine ring of PLP [45,46]. It was proposed that the positively charged pyridine ring of PLP could cause an electron sink effect which facilitates the enzymatic reactions [3, 5–10] and that Asp298 could enforce the PLP ring nitrogen atom to be protonated in the enzyme [12, 13]. However, the pyridine ring nitrogen cannot be protonated under physiological conditions [11] yet there should be a proton between the pyridine ring N atom and side chain of Asp298 in order to form an interaction. Thus the proton should be either on the N atom of pyridine ring or on the side chain of Asp298. In order to question the effect of the position of the proton on the Michael addition reaction, different protonation states of PLP nitrogen was taken into account in Model 3 (**M3**) and Model 3' (**M3'**). In the case of a model with a non-protonated pyridine ring nitrogen, the proton should be shifted to the Asp298 side chain which leads to a neutral aspartate (**M3'**). In a quantum chemical study on the  $\omega$ -Transaminase enzyme from *Chromobacterium violaceum* (Cv-  $\omega$ TA) it was reported that the proton was located on the Asp259 residue since the proton moves spontaneously from the N atom of the pyridine ring and the side chain of Asp259 in geometry optimizations [65]. In **M3** and **M3'**, it was observed that the proton was located between the pyridine ring and Asp298 in the transition states (Figure 3.37). However, the reactants of **M3** and **M3'** which were obtained from IRC calculations show different characteristics by means of the positions of the aforementioned proton; it could be either bound to Asp298 side chain or close to the pyridine ring without binding to the N atom in the reactant. The activation energy barriers for the Michael addition reaction with these two models are depicted in Figure 3.35. The calculated gas phase activation energy barriers of the Michael addition reaction differs about 1 kcal/mol among the two models, while the single point PCM calculations yielded barriers that differ from each other about 1-2 kcal/mol. Even though the difference between the activation energy barriers of two models are not large enough to distinguish, the gap between the gas and solvent energy surfaces in **M3** found to be higher than in **M3'**. In addition to the gaps between surfaces, the

activation energy barrier of **M3** was calculated lower than that of **M3'** when solvent effects were taken into account with the two dielectric constants. It was found that the activation energy barriers increased relative to the previous model **M2**. The loss of the electron sink effect due to the proton which is not bound to the pyridine ring N atom in transition state structures (especially of **M2**) increases the barrier. It was reported that a negatively charged residue (mainly Asp) in PLP dependent enzymes can enforce the pyridine ring nitrogen to be protonated with the assistance of a positively charged residue [13]. Therefore, the loss of the electron sink effect can be prevented by the inclusion of another residue which could interact with the side chain of Asp298 and enforce the pyridine ring nitrogen to be protonated (see Chapter 3.2).

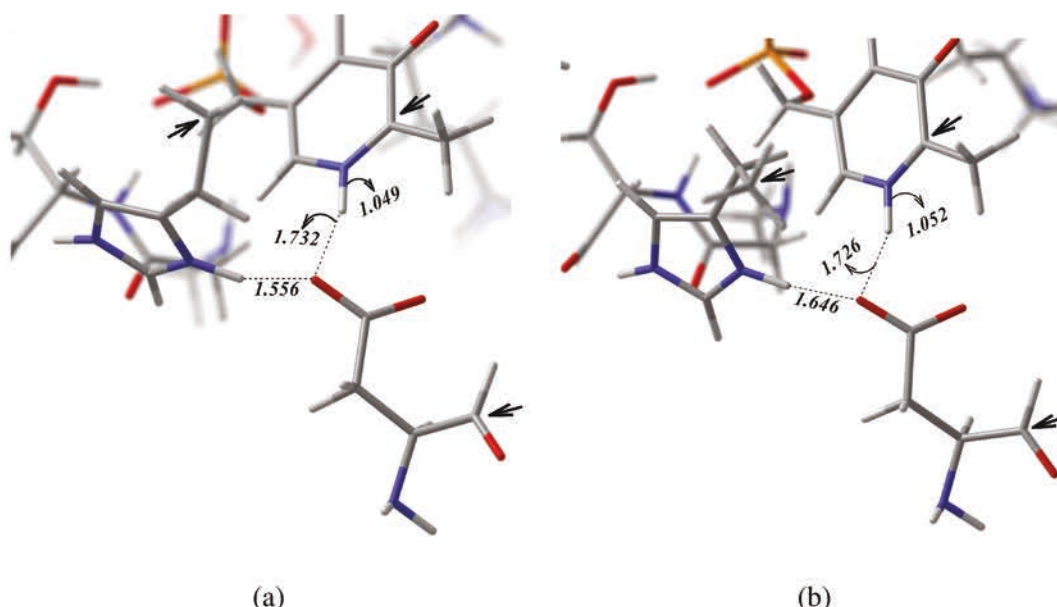
**Inclusion of His190 to form Asp-His diad.** The movement of the proton between the N atom of pyridine ring and the side chain of Asp298 may be prevented with a positively charged residue that can interact with Asp298 [13] and enforce it to stay ionized. In the crystallographic structure of GABA-AT, the only residue that can have such interaction with Asp298 is His190. Furthermore, in the molecular dynamics simulations study of GABA-AT, it was reported that the conservation of the interaction between Asp298 and pyridine ring of PLP can be achieved only with a positively charged His190 and a negatively charged Asp298 (see Chapter 3.2). It was computed that the relative interaction free energy between PLP and GABA-AT has the lowest value among other different protonation states (see Chapter 3.2). Thus, the Michael addition reaction was modelled with a negatively charged Asp298 and a positively charged side chain of His190 in Model 4 (**M4**) (Figure 3.38) to examine the capability of His190 to reduce the effect of the polarity of the environment on the potential energy surface. The activation energy barriers of **M4** is given in Figure 3.35. A decrease in the activation energy barriers is observed both for the gas phase and the solvent energy calculations. Furthermore, the difference between gas phase and solvent energy surfaces were found 0.9 kcal/mol with  $\epsilon = 80$  and 0.3 kcal/mol with  $\epsilon = 4$ . The energy gaps between the surfaces suggest that the Michael addition reaction with **M4** is not affected by the polarity of the environment as compared to the previous models. Thus, the obtained activation energy barriers are more accurate than the previous models since more flexibility was granted to the structure and the choice of dielectric constant became irrelevant [58, 60]. It was found that the Michael addition reaction



**Figure 3.37 :** The Michael addition reaction with model clusters with the involvement of Asp298. Three dimensional representation of the interaction between Asp298 side chain and pyridine ring N atom of PLP in (a) transition state of **M3**. (b) transition state of **M3'**, (c) reactant of **M3** and (d) reactant of **M3'**. Frozen atoms are shown with black angle headed arrows.



requires about 3-4 kcal/mol in **M4** which involves several surrounding residues. The inclusion of the positively charged His190 enabled Asp298 to stay negatively charged and thus enforced the PLP pyridine ring to be protonated. The enforced protonation states enabled the electron sink effect which then facilitated the enzymatic reaction by reducing the activation energy barrier.

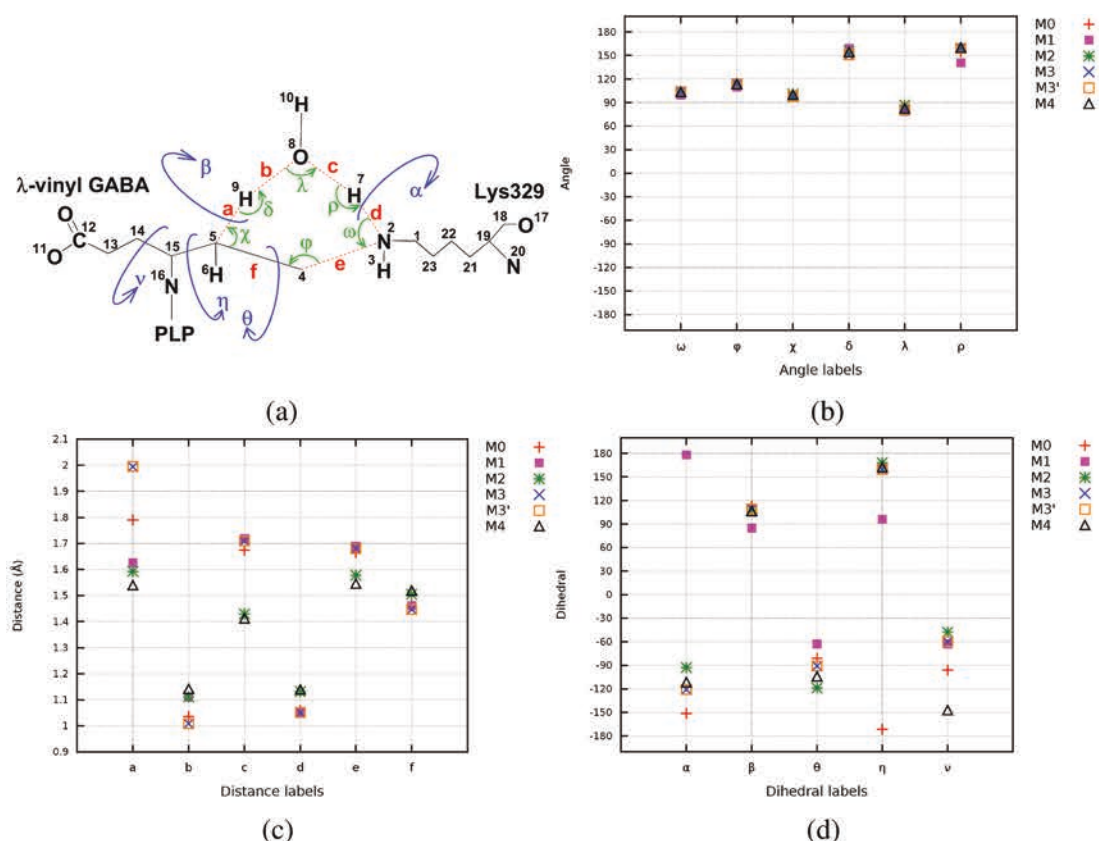


**Figure 3.38 :** The Michael addition reaction with model cluster **M4**. Three dimensional representation of the interaction between the pyridine ring N-Asp298-His190 triad in (a) reactant, and (b) transition state. Frozen atoms are shown with black angle headed arrows.

### 3.3.2.2 Effect of Surrounding Residues on Geometries

The geometrical properties of the reaction center were also investigated by means of distances, angles and dihedrals to understand the effect of the surrounding residues on the six-membered transition state. The calculated parameters within the six-membered transition state structure and their labels are depicted in Figure 3.39. It was observed that the angles within the transition state center were not altered by the increase of the size of the models (Figure 3.39a). The distances differ from each other about 0.1 Å within the models except the distances between the transferring proton H9 and C5 of  $\gamma$ -vinyl-GABA (a) and transferring proton H7 and O8 of water molecule (c) (Figure 3.39c).

The distances in **M0** and **M1** are similar. However, it was observed that the carboxyl group of  $\gamma$ -vinyl GABA and the methylamine in the smallest model **M0** have different



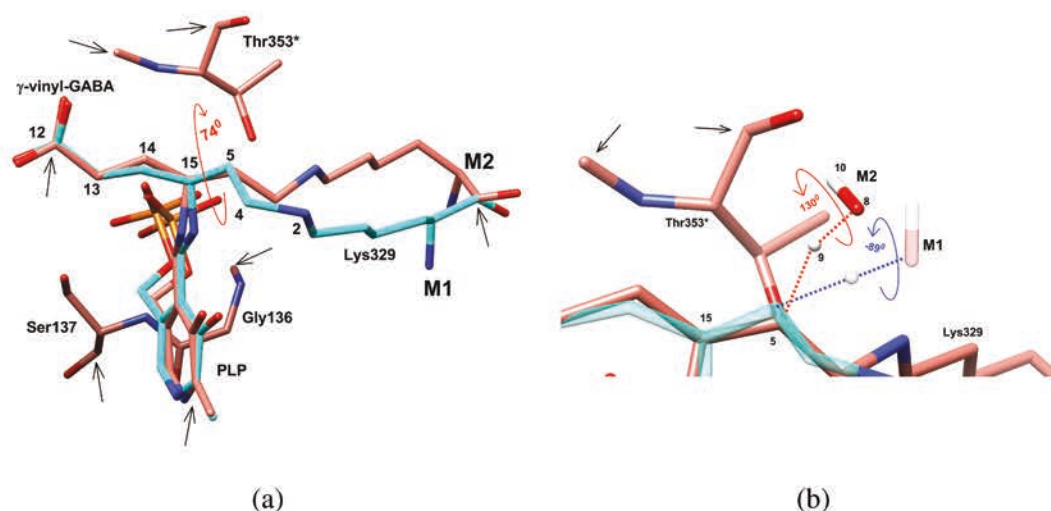
**Figure 3.39 :** Geometrical parameters of six-membered MA transition state structures. (a) Two dimensional representation and labels of distances a-f, dihedrals ( $\alpha$ : H3-N2-H7-O8,  $\beta$ : H6-C5-H9-O8,  $\theta$ : C1-N2-C4-C5,  $\eta$ : C14-C15-C5-C4,  $\nu$ : C13-C14-C15-N16) and angles ( $\omega, \phi, \chi, \delta, \lambda, \rho$ ). (b) The distribution of angles. (c) The distribution of distances. (d) The distribution of dihedrals.

orientations relative to the larger model **M1** due to the atoms that were kept frozen during geometry optimizations.

When the residues Gly136, Ser137 and Thr353\* were included in **M2**, it was observed that the water molecule rotated towards the backbone oxygen of Thr353\* and the distance **c** was shortened by about 0.3Å. The fitted three dimensional representations of **M1** and **M2** are depicted in Figure 3.40. The dihedral  $\eta$  and  $\theta$  are altered by about 70° and 90° respectively in **M2** relative to the **M1**. Furthermore, the dihedral  $\alpha$  diverged by about 90° in **M2** with the inclusion of Gly136, Ser137 and Thr353\*. The change of the distance **c** may stem from the differences in the aforementioned dihedrals.

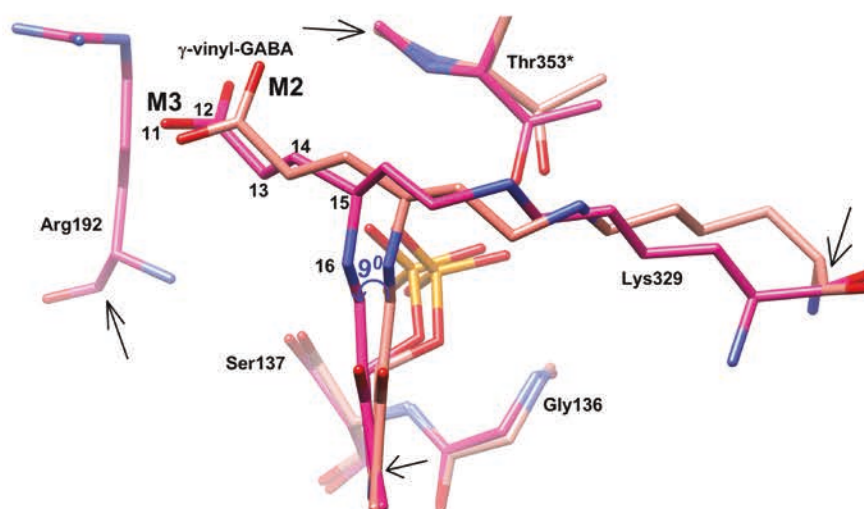
It was observed that with the inclusion of Arg192 and Asp298 in **M3**, the distance **a**, and **c** lengthen, and PLP is tilted by about 9°. The three dimensional representations of fitted **M2** and **M3** is depicted in Figure 3.41. The tilt in PLP might be caused by the removal of frozen carboxyl group of  $\gamma$ -vinyl-GABA with the inclusion of Arg192





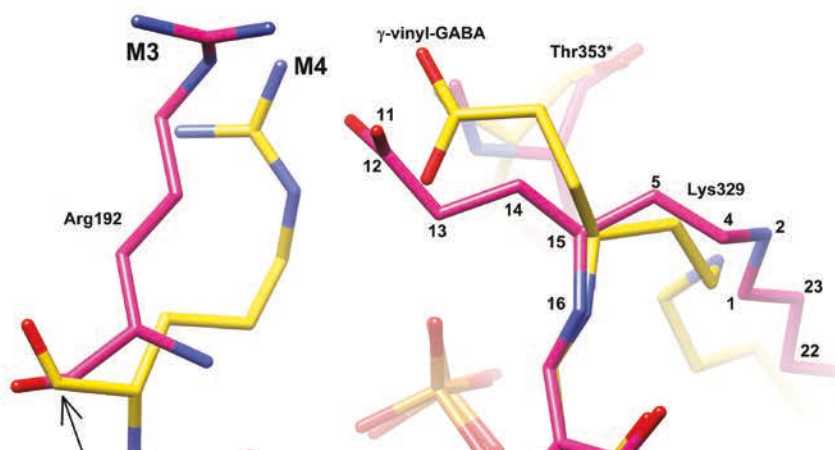
**Figure 3.40 :** The three dimensional representations of the fitted models **M1** (C atoms colored by cyan) and **M2** (C atoms colored by pink). **(a)** The alteration of the dihedral N16-C15-C5-C4 (red arrow) and its value (colored with red). **(b)** The dihedral of C5-H9-O8-H10 in **M1** (blue) and in **M2** (red). The frozen atoms are shown with black arrows.

in **M3**. The alteration of the distances **a**, and **c** may stem from the tilt of the PLP ring. With the inclusion of Arg192, the torsion of O11-C12-C13-C14 has changed by about  $60^\circ$  in **M3** relative to **M2**. It is also observed that both the dihedrals  $\alpha$  and  $\theta$  are altered by about  $30^\circ$  in **M3**. This may also be related to the tilt of the PLP ring. The same geometrical parameters were obtained in the geometry optimizations of the transition state structures with models **M3** and **M3'**.



**Figure 3.41 :** The three dimensional representation of the model **M2** (C atoms colored by pink) and the model **M3** (C atoms colored by magenta). The backbone C atoms of Ly329, Gly136, Ser137, Asp298, Thr353\*, and the N atom on PLP of **M3** is fitted to the corresponding atoms of **M2**.

The inclusion of the His190 side chain in **M4**, modified the distances **a**, and **c** as compared to **M2**. The dihedral  $\nu$  is changed by about  $100^\circ$  in **M4**. This may stem from the alterations of the internal torsions on Arg192. The three dimensional representation of the interaction between Arg192 and the carboxyl group of  $\gamma$ -vinyl-GABA in **M3** and **M4** is depicted in Figure 3.42. The dihedral C12-C13-C14-C15 was computed as  $-162^\circ$  in **M3**, a change from **M4** ( $-82^\circ$ ). In addition to the rotation of the carboxyl group of  $\gamma$ -vinyl-GABA, the internal torsions of Lys329 are altered. The dihedral C4-N2-C1-C23 in **M3** was computed as  $-166^\circ$  while it changed to  $-95^\circ$  in **M4**. Additionally the dihedral N2-C1-C23-C22 was measured at  $-73^\circ$  in **M4** from  $171^\circ$  in **M3**. These rotations, which may stem from the rotation of the  $\gamma$ -vinyl-GABA, might be the main reason of the changes in the distances **a** and **c**.



**Figure 3.42 :** The three dimensional representation of the interaction between Arg192 and the carboxyl group of carboxyl group of  $\gamma$ -vinyl-GABA in model **M3** (C atoms colored by magenta) and in model **M4** (C atoms colored by yellow). The backbone C atoms of Ly329, Gly136, Ser137, Asp298, Thr353\*, and the N atom on PLP of **M4** is fitted to the corresponding atoms of **M4**. The frozen atoms are depicted with black arrows.

### 3.3.3 Concluding remarks

In this part of the thesis, the Michael addition reaction of  $\gamma$ -vinyl GABA with GABA-AT was modeled with different sizes of clusters to find the optimum model in mimicry of the active site. The cluster sizes were increased until the effect of the polarity of the environment on the reaction was minimum. The inclusion of the active site residues within the models helped us to explain the effects of these residues on the reactivity. The presence of similar residues in many PLP dependent enzymes enabled us to provide some general aspects for modelling these enzymes.

In the smallest cluster Lys329 was represented as methylamine and the phosphate group was modeled as a methyl group. The calculations pointed out that the model is able to provide general perceptions about geometrical parameters and energetics but not enough to mimic the active site. It was observed that the inclusion of the phosphate group not only increased the barriers but also increased the effect of the polarity of the environment. A decrease in the activation energy barrier with the inclusion of Gly136, Ser137 and Thr353\* pointed out that the phosphate group should be either represented as methyl or should be introduced with its surrounding residues in modelling studies to obtain better description of the active site and reliable energy results. Even though the effect of the polarity of the environment was reduced with the inclusion of Arg192 and Asp298, the activation energy barriers increased. It was observed that the proton on the N atom of the pyridine ring moves spontaneously from the corresponding N atom and the side chain of Asp298 as it was stated in a previous modelling study involving Asp and PLP [65]. It was proposed that the electron sink effect due to the proton on the N atom of the pyridine ring facilitates the enzymatic reactions [3, 5–10]. Hence, the spontaneous movement of the proton might be the main cause of the increase in the activation energy barrier.

We reported in our molecular dynamics study that His190 in GABA-AT can increase the interaction between Asp298 and the pyridine ring of PLP, and that the positively charged His190 can help Asp298 to stay negatively charged [13, 74]. Therefore, the side chain of the His190 was included in the model to enforce the proposed electron sink effect. It was observed that providing a hydrogen bond between negatively charged Asp298 side chain and positively charged His190 side chain helps the proton to stay on the N atom of the pyridine ring as it was proposed. It was found that both the activation energy barriers and the effect of the polarity of the environment decreased remarkably by keeping the corresponding proton on the pyridine ring N atom, which confirms the proposed electron sink effect.

The geometrical parameters of the transition state center, except the angles, are modified when the model size increases. The alterations were mainly due to the removal of the frozen atoms by the inclusion of more interacting residues. It was observed that the distances within the transition state center were affected with the inclusion of the Thr353\* residue that altered the position of the assisting water

molecule. It was also observed that the tilt of the PLP ring, which might be due to the rotations within the  $\gamma$ -vinyl-GABA, may also effect the dihedrals and the distances within the transition state center.

### **3.4 Improvement of the Linear Scaling of the SEBOMD Simulations**

The accurate and reliable study of biological systems remains a challenge for the Quantum Chemistry (QC) community. The quantum chemical approach is often the only reliable way to model chemical transformations: bond making/breaking, electron transfer (redox mechanisms), electronic transition between different energy levels, ionization, etc. Unfortunately, the numerical solution of the Schrödinger equation becomes increasingly (and often prohibitively) expensive in terms of computational power when system size increases. The numerical procedures in QC require the calculation of numerous integrals and the resolution of (non)-linear equations, thus limiting the size of the systems that can be modelled. In addition, most of the QM algorithms, in their standard version, lead to a non-linear scaling of the CPU time with respect to the number of atoms to be considered.

In recent years, new algorithms have been designed to compute the energy of very large molecular systems (up to a few thousand of atoms) using so-called linear scaling methodologies for which the CPU time needed to evaluate the energy of one system is linearly dependent on the size of this system [75–78]. In parallel, due to the many degrees of freedom that represent such large systems, Car-Parrinello and Born-Oppenheimer Molecular Dynamics techniques have been developed to explore their energetical, structural and dynamical properties [79].

As of today, only linear scaling approximate quantum chemical methods like semiempirical or Density Functional based Tight Binding (DFTB) approaches are capable of combining both the necessary speed and accuracy to model large biomolecular systems with reasonable CPU cost, and to allow for long-run molecular dynamics simulations at the QC level (up to the nanosecond time frame) [80, 81]. The theoretical chemistry group of Nancy has recently implemented Semi-Empirical Born-Oppenheimer Molecular Dynamics (SEBOMD) capabilities in the AMBER suite of software by adding to the MD part of AMBER semiempirical QC routines that use the linear scaling Divide & Conquer algorithm (D&C) [82, 83]. Today, this permits

MD simulations of molecular systems containing several hundreds of atoms for a few nanoseconds using a single workstation. However, while the D&C algorithm is intrinsically parallel and scales nicely up to a few tens of CPU cores, it does not seem to us that this algorithm is the best suited for use on massively parallel computers containing thousands of cores [84, 85]. Thus, our target is to improve the scalability of the semiempirical Born-Oppenheimer Molecular Dynamics approach on parallel computers.

Today, the current D&C implementation is not the most efficient one to run on very large supercomputers or on GPUs. By using available numerical libraries that scale well on these new hardware architecture, such as LAPACK and/or SCALAPACK, we will be able to tackle larger systems by SEBOMD simulations. Instead of the current simulations that can produce trajectories of a few nanoseconds on systems containing a few hundreds of atoms (e.g. a peptide in a box a water molecules), we should be able to perform simulations on thousands of atoms, leading, for the first time, to the simulations of (small) truly solvated proteins using a complete quantum description at the atomic level.

### **3.4.1 Computational details**

Our reference systems for profiling studies are boxes of water molecules of different sizes (64  $H_2O$ , 128  $H_2O$ , and 216  $H_2O$ ) and a large solvated box that consists of a small protein (TrpCage, 20 residues, 284 atoms) and 730 water molecules. For preliminary study, only a box consisting of 216 water molecules is used since our objective is to find the best choice of methodology to profile diagonalization routines.

#### **3.4.1.1 Water molecules**

Our test case consists of running a 10 step molecular dynamics in the NVT ensemble at 300K using the Andersen temperature coupling [188]. A time step of 1.0 fs is used and no SHAKE algorithm [181] is applied.

For the QM part of the simulations, the PM3 Hamiltonian [124] is used with the PIF2 modification [125–127] for intermolecular core-core interactions. A polynomial interpolation of the density matrix elements is used from the last three steps as an initial (guess) density matrix for the current step. When the D&C algorithm is used,

the extent of the first buffer region from the core is fixed at 6.0 Å and no second buffer region from the core is used.

The number of simulation steps is increased from 1 to 10 and simulations are performed one by one for each chosen number of steps. In order to get more reliable results, all of the simulations are performed five times. Thus, 400 different simulations are performed in total for the preliminary study. In the case of the profiling studies of the diagonalization routines, 100 different simulations are performed in total for each eigenvalue solver, yielding 800 simulations in total. In order to get the consumed times per SCF iteration and gradient calculation, a linear regression is performed on the total times and the number of iterations or the number of simulation steps. All of the computations were carried out on a Dell PowerEdge T620 machine with two Intel Xeon CPU E5-2640 processors (two 6-cores at 2.50GHz with a total of 32Gbytes of memory) running under GNU/Linux.

#### **3.4.1.2 Protein**

The simulation set-up for the TrpCage protein system is similar to the water molecules with some slight differences. Instead of the Andersen thermostat, a Langevin temperature coupling is used and pressure regulation is performed using a Monte Carlo barostat. The collision frequency for the Langevin dynamics is chosen as  $1.0ps^{-1}$ .

For the QM energy calculation, instead of PM3 [124] with PIF2 modification [125–127], the AM1/d-PhoT semiempirical hamiltonian [119] is used. Additionally, an empirical force field correction is applied to all peptide bonds and a force constant for the peptidic correction of 5.9864 kcal/mol is chosen. The extent of the first buffer region from the core is fixed at 8.0 Å when the D&C algorithm is used.

Instead of increasing the number of simulation steps from 1 to 10 and performing simulations one by one for each chosen number of steps, the 10 steps of simulations are performed directly .

#### **3.4.2 Results and Discussion**

In order to understand the time consuming parts of the SEBOMD code, profiling studies must be performed. However, the profiling results may be affected mainly by four different parameters; the choice of the method, the choice of the



diagonalization scheme, the choice of using periodic boundary conditions and the choice of the computation of the long-range interactions. A schematic representation of these four parameters is given in Figure 3.43. First, the SEBOMD simulation can perform a standard SCF calculation or use a linear scaling D&C algorithm. Second, the diagonalization of the Fock matrix could be performed by using a pseudo-diagonalization procedure or by full diagonalization at each step of the SCF procedure. Third, periodic boundary conditions (PBC) can be applied or not. This corresponds to gas phase or condensed phase simulations. When applied, long-range interactions can be added either by using constant external charges (classical Particle Mesh Ewald summation) or by polarizing the wave functions using Mulliken charges (quantum Ewald summation). In order to find the best methodology for profiling eight different methodologies for SEBOMD simulations are conducted. The abbreviations that are used for the simulations are described in Table 3.3.



**Figure 3.43** : Parameters that would effect consumed time by a SEBOMD simulation.

### 3.4.2.1 Implementation of eigenvalue solvers

The most crucial point in implementation of eigenvalue solvers is that the solver should correspond to the problem that is at hand. Furthermore, the solver should deal with the

**Table 3.3** : Abbreviations of SEBOMD simulations and their explanations.

Method:	m0	Standard Closed Shell
	m2	Divide & Conquer
Diagonalization:	f0	Pseudo-diagonalization
	f1	Full-diagonalization
PBC:	n0	Gas
	n1	Condensed Phase
Longrange:	l0	Minimum Image
	l1	Classical PME Summation
	l2	Quantum Ewald Summation

same type of matrix that is in question (i.e. real or complex). Since a Fock matrix is a real symmetric matrix, eigenvalue solvers that are chosen should deal with the same type of matrices. In the *Linear Algebra PACKage* (LAPACK) there are eight different solvers that can be implemented in SEBOMD. These eigenvalue solvers and their properties are depicted in Table 3.4.

In addition to the implementation of new eigenvalue solvers, a new routine selection methodology is conducted. In order to improve the SEBOMD code, an automatic eigenvalue solver selection choice is implemented. The procedure is depicted in Algorithm 1. If the automatic selection option is used for diagonalization (`diag_routine=0`), the fastest eigenvalue solver is scanned during the first eight SCF iteration. After scanning, the fastest eigenvalue solver is chosen for further iterations.

**Algorithm 1** Procedure for automatic eigenvalue solver selection.

---

```
1: maxdiag = 8
2: iter = iteration_number
3: if first_call AND iter ≤ maxdiag then
4:   diag_routine = iter
5:   Perform diagonalization
6:   current_time = consumed_time
7: if fastest_time ≤ consumed_time then
8:   fastest_time = current_time
9:   fastest_routine = diag_routine
10: if iter > maxdiag then
11:   diag_routine = fastest_routine
```

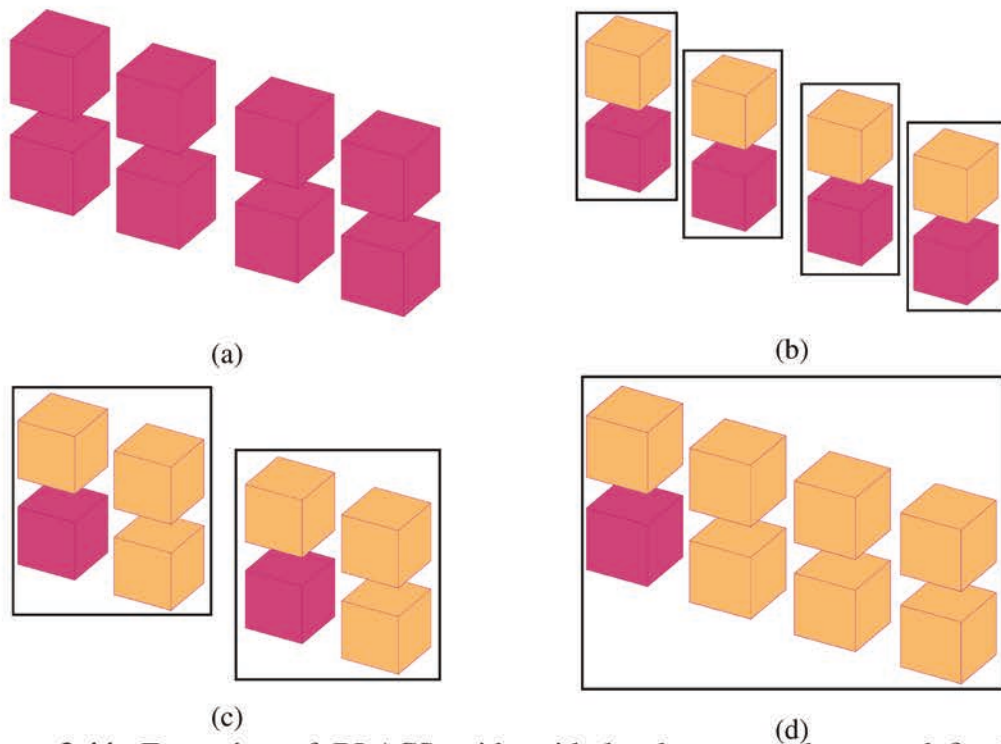
---

**Table 3.4 :** Newly implemented eigenvalue solvers. The number of routine represents the eigenvalue solver selection in SEBOMD simulations.

Routine Number	Eigenvalue Solver	Property
2	dsyev	Computes all eigenvalues and eigenvectors of a real symmetric matrix with QR iterations.
3	dsyevd	A divide and conquer algorithm is used to calculate eigenvectors.
4	dsyevr	Computes selected eigenvalues and eigenvectors of a real symmetric matrix. Uses Relatively Robust Representations to compute eigenspectrum Eigenvalues computed by the dqds algorithm, and eigenvectors are computed with Relatively Robust Representations.
5	dsyevx	Computes selected eigenvalues and, optionally, eigenvectors of a real symmetric matrix. Eigenvalues and eigenvectors can be selected by specifying either a range of values or a range of indices for the desired eigenvalues.
6	dspev	Computes all the eigenvalues and, optionally, eigenvectors of a real symmetric matrix in packed storage.
7	dspevd	Computes all the eigenvalues and, optionally, eigenvectors of a real symmetric matrix in packed storage. If eigenvectors are desired, it uses a divide and conquer algorithm.
8	dspevx	Computes selected eigenvalues and, optionally, eigenvectors of a real symmetric matrix in packed storage. Eigenvalues/vectors can be selected by specifying either a range of values or a range of indices for the desired eigenvalues.

The *SCAlable Linear Algebra PACKage* (SCALAPACK) has similar eigenvalue solvers than LAPACK. These solvers are simply named in a similar way as in their LAPACK version but with an extra letter to indicate parallelism; **pdsyev**, **pdsyevd**, **pdsyevr**, and **pdsyevx**. In order to implement SCALAPACK solvers, one should first create grids using *Basic Linear Algebra Communication Subprograms* (BLACS) that establish a linear algebra oriented message passing interface.

The implementation procedure of BLACS is depicted in Algorithm 2. The number of created grids are defined by the number of **local\_masters**, i.e., the processors that own the subsystems. For instance, if 16 processors are used with `local_master = 2`, the number of grids would be 2 with 8 processors in each grid if a D&C algorithm is used. A simple representation of the formation of BLACS grids is depicted in Figure 3.44. In the case of the **Standard Closed Shell** (SCS) algorithm only one grid is created since the number of subsystems is equal to one, and the number of `local_master`. During the implementation, the number of `local_master` is defined by the number of desired processors within the grid (Algorithm 2).



**Figure 3.44** : Formation of BLACS grids with `local_masters` that are defined by `nthread`. **a.** NPROCS=8 with `nthread`=1, number of grids=8, and number of `local_masters`=8. **b.** NPROCS=8 with `nthread`=2, number of grids=4, and number of `local_masters`=4. **c.** NPROCS=8 with `nthread`=4, number of grids=2, and number of `local_masters`=2. **d.** NPROCS=8 with `nthread`=8, number of grids=1, and number of `local_masters`=1. `local_masters` are represented with purple color, and each grid is represented within black rectangles.

After the creation of the grids, each `local_master` diagonalizes its own Fock matrix/matrices with the help of its corresponding workers (other processors within the same grid) using SCALAPACK eigenvalue solvers in parallel. The detailed procedure that is used in diagonalization of the Fock matrix with SCALAPACK solvers is given in Algorithm 3. The procedure generally consists of five steps. First, one should

---

**Algorithm 2** Creation of grids with BLACS.

---

```
1: define nthread      ► The number of processors in one grid
2: procedure CREATE SUB-GRIDS(setup_subgrid)
3:   subgrids = NPROCS/nthread      ► number of local_masters
4:   subnprocs = nthread
5:   define number of rows & columns (nprow&npcol) within the grids
6:   for k=1,subgrid do
7:     for i=1,nprow do
8:       for j=1,npcol do
9:         umap(k,i,j)= (k - 1) * subnprocs + (i - 1)npcol + (j - 1)
                                ► indicates the process-to-grid mapping
10:      call blacs_gridmap(subcontxt(k),umap(k,:,:),ldumap,nprow,npcol)
                                ► Maps available processes to grid
                                ► subcontxt defines the CONTEXT of the grid
                                ► ldumap is the leading dimension of the UMAP
11:      if subcontxt(k) ≥ 0 then
12:        mygrid = k
```

---

define the array descriptors; a vector that stores the information required to establish the mapping between an element and its corresponding process, and memory location. Second, Fock matrix should be sent to the worker processors. Third, every processor within the grid should set its corresponding local values, the parts that they are going to work on. Fourth, processors call SCALAPACK solver to diagonalize the Fock matrix. Last, since only the `local_master` own the subsystems, the results should be collected from workers for further computations.

### 3.4.2.2 External control

The configuration of the Amber installation can be performed with different compilers and flags which can effect the efficiency of SEBOMD. In most systems the **GNU** compiler works fine. If the **Intel** compiler is desired, the `intel` flag can be used instead of `gnu`.

An integrated LAPACK library can be utilized since it is already compiled and linked to *sander* in AmberTools. However, if the Fock matrix that is going to be diagonalized is large, the scaling of the diagonalization may not be as good as in small problems with the internal LAPACK solvers. Thus, one can give `-mkl` as a flag in order to use the eigenvalue solvers provided by the *Intel Math Kernel Library*



---

**Algorithm 3** Diagonalization procedure using SCALAPACK solvers

---

```
1: procedure ARRAY DESCRIPTORS(array_desc)
2:   call DESCINIT      ► Create descriptors
3:   call DGEBS2D       ► Broadcast send descriptors within the grid
4:   call DGEBR2D       ► Broadcast receive descriptors from
                        local_master
5: procedure SEND FOCK MATRIX(send_array)
6:   call IGEBS2D       ► Broadcast send the dimension of Fock matrix
7:   call IGEBR2D       ► Broadcast receive the dimension from
                        local_master
8:   call DGEBS2D       ► Broadcast send Fock matrix within the grid
9:   call DGEBR2D       ► Broadcast receive Fock matrix from
                        local_master
10:
11: procedure SET LOCAL VALUES(set_array)
12:   call PDELSET      ► Find local part of the matrix that correspond
13:
14: procedure DIAGONALIZATION(SCALAPACK)
15:   call PDSYEV*      ► Perform diagonalization with SCALAPACK solver
16:
17: procedure GATHER(gather_array)
18:   call NUMROC        ► Find local number of rows and columns
19:   call INDXL2G       ► Find global indices
20:   call DGSUM2D       ► Perform sum operation, similar to
                        MPI_Allreduce
```

---

(MKL). In this case, linking to the Intel MKL can be performed with additional flags. The advantage of using the Intel MKL is that the eigenvalue solvers are threaded and functions are thread-safe (safe execution by multiple threads). If the program execution environment has a shared memory architecture, using threaded eigenvalue solvers can improve efficiency. In order to use threaded versions of the solvers, `-mkl=parallel` should be depicted as a flag. The number of threads that are desired to be used during execution can be defined by the `OMP_NUM_THREADS` or `MKL_NUM_THREADS` environment variables.

### 3.4.2.3 Preliminary study for profiling

In order to get more accurate results from the profiling studies, the best combination of methodological options for SEBOMD simulation should be identified since elapsed time depends on these options. To identify time consuming parts of the simulations by means of percentages, four different methodological options are used (method, diagonalization, PBC, and longrange interaction, see Table 3.3). Sixteen different

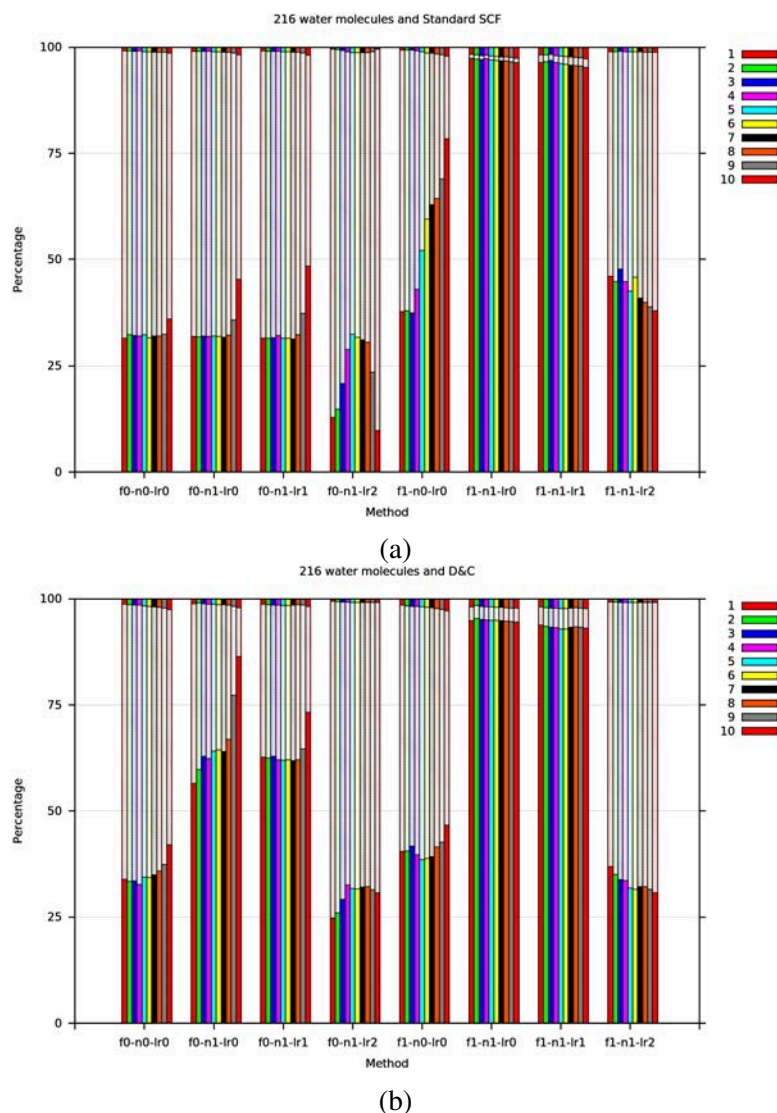
SEBOMD simulations of 216 water molecules are performed using either standard closed shell algorithm or D&C algorithm. Execution times by SCF, gradient calculation and other routines per simulation with increasing number of steps (1 to 10) are depicted in Figure 3.45.

Results obtained with standard closed shell algorithm indicates that the execution time by SCF routines are below than 50% when pseudo diagonalization (f0) is used (Figure 3.45a) and performing gas (n0) or condensed phase simulations (n1) does not have a significant effect unless full diagonalization (f1) is performed. Additionally, the fraction of execution times of these routines get lower when long-range interactions are added by polarizing the wave functions using Mulliken charges (quantum Ewald summation, lr2). SCF routines becomes the most dominant part of the simulations by means of consumed time when full diagonalization (f1) is performed with condensed phase (n1) and without long-range interactions (lr0).

In case of SEBOMD simulations using D&C algorithm, SCF routines consumes less than 50% with pseudo-diagonalization (f0) at gas phase (n0) and without long-range interactions (lr0) (Figure 3.45b). Performing simulations at condensed phase (n1) increases the percentage of consumed times of SCF routines in both diagonalization schemes (f0 and f1). Inclusion of long-range interactions with quantum Ewald summation (lr2) lowers the consumed times of the SCF routines by means of percentages as in standard closed shell algorithm. It was observed that, similarly to standard closed shell algorithm, SCF routines are the most dominant part of the simulations by means of consumed time when full diagonalization (f1) is performed with condensed phase (n1) and without long-range interactions (lr0).

As a summary, results obtained with preliminary profiling studies indicates that the SCF routines are the most dominant part of the SEBOMD simulations, using either standard closed shell algorithm or D&C algorithm, when full-diagonalization (f1) procedure applied in condensed phase (n1) without long-range interactions (lr0). In order to observe the effect of implemented eigenvalue solvers and used libraries more accurately, all of the profiling studies from herein are performed with these methodological options.





**Figure 3.45 :** Execution times of the SCF routine, other routines and gradient routine as the percentage of the total SEBOMD simulation time with **a.** Standard Closed Shell algorithm (m0). and **b.** D&C algorithm (m2). The percentages are colored versus the cumulative number of simulation steps. SCF (bottom) and gradient routines (top) are depicted with dense colors while the other routines are depicted with light colors.

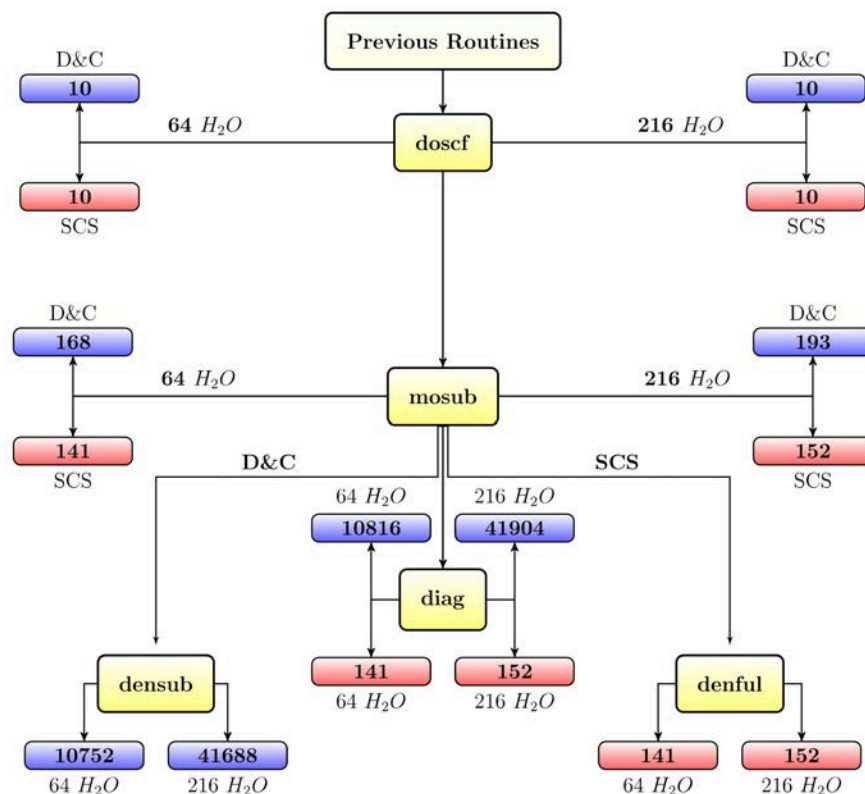
The critical point in profiling studies is that the number of calls to the subroutines in standard closed shell (SCS) and D&C algorithms are different from each other. Additionally, the number of calls to the subroutines alter with the size of the system. Therefore, the number of calls to some special subroutines are investigated for the smallest and the biggest system of water molecules (64 and 216) using both the SCS and D&C algorithms. The simulations are performed with 10 number of simulation steps at condensed phase (n1), using full diagonalization (f1) and disregarding the long range interactions (lr0). The obtained results are depicted in Figure 3.46, where  $\text{dos}_{\text{scf}}$  is interrelated with the number of steps in the simulation,  $\text{mosub}$  is correlated with the

number of SCF iterations, and `diag` is associated with the number of diagonalization. Since the number of simulation steps are same, the number of calls to the `doscfc` routine which computes the Fock matrix is same for all systems and algorithms. In case of 64 water molecules, the `mosub` subroutine is called for 141 and 168 times to perform SCF iterations when, respectively, SCS algorithm and D&C algorithm is used. When the system of 216 water molecules is simulated, the same subroutine is called for 152 and 193 times using SCS algorithm and D&C algorithm respectively. In simulations of 64 water molecules, the default subroutine to perform diagonalization in `seabond` module (`diag`) is called for 141 times with SCS algorithm while it is called for 10816 times with D&C algorithm. The obtained number of calls for `diag` is increased to 152 and 41904 using SCS algorithm and D&C algorithm respectively in simulations of 216 water molecule.

In order to built the whole density matrix, SCS algorithm uses the `denful` subroutine, while D&C algorithm uses the subroutine `densub`. The `denful` subroutine is called for 141 and 152 times in the simulations of 64 and 216 water molecules. In simulations using D&C algorithm, `densub` subroutine is called for 10752 and 41688 times with 64 and 216 water molecules respectively.

It is found that, even though the number of SCF iterations (`mosub`) is almost similar for the SCS algorithm and the D&C algorithm, the number of diagonalizations increases drastically when D&C (see Figure 3.46, `diag`, blue boxes) is used instead of SCS algorithm (see Figure 3.46, `diag`, red boxes). When the size of the system is taken into account, it is observed that the number of diagonalizations (`diag`) is almost similar for the smallest and the largest systems with SCS algorithm. However, a drastic increase is seen in the number of diagonalizations (`diag`) with the increase of the system size when D&C is applied. Additionally, the number of calls to the `denful` is related to the number of diagonalizations and the size of the system does not have a significant effect on it. However, the number of calls to the `densub` subroutine is effected severely with the size of the system. These results indicates that the profiling studies may be affected significantly with the number of calls which is related to the size of the system and the algorithm that is used (SCS or D&C). Since the number of calls may lead to less reliable results, one should consider the consumed time in one SCF iteration and also the effect of the system size. Therefore, the consumed time

per diagonalization step (one SCF iteration) is computed in the profiling studies of the implemented eigenvalue solvers. In case of the simulations using D&C algorithm, the consumed time per diagonalization step indicates the total time that is required for the diagonalization of all submatrices that belong to the subsystems.

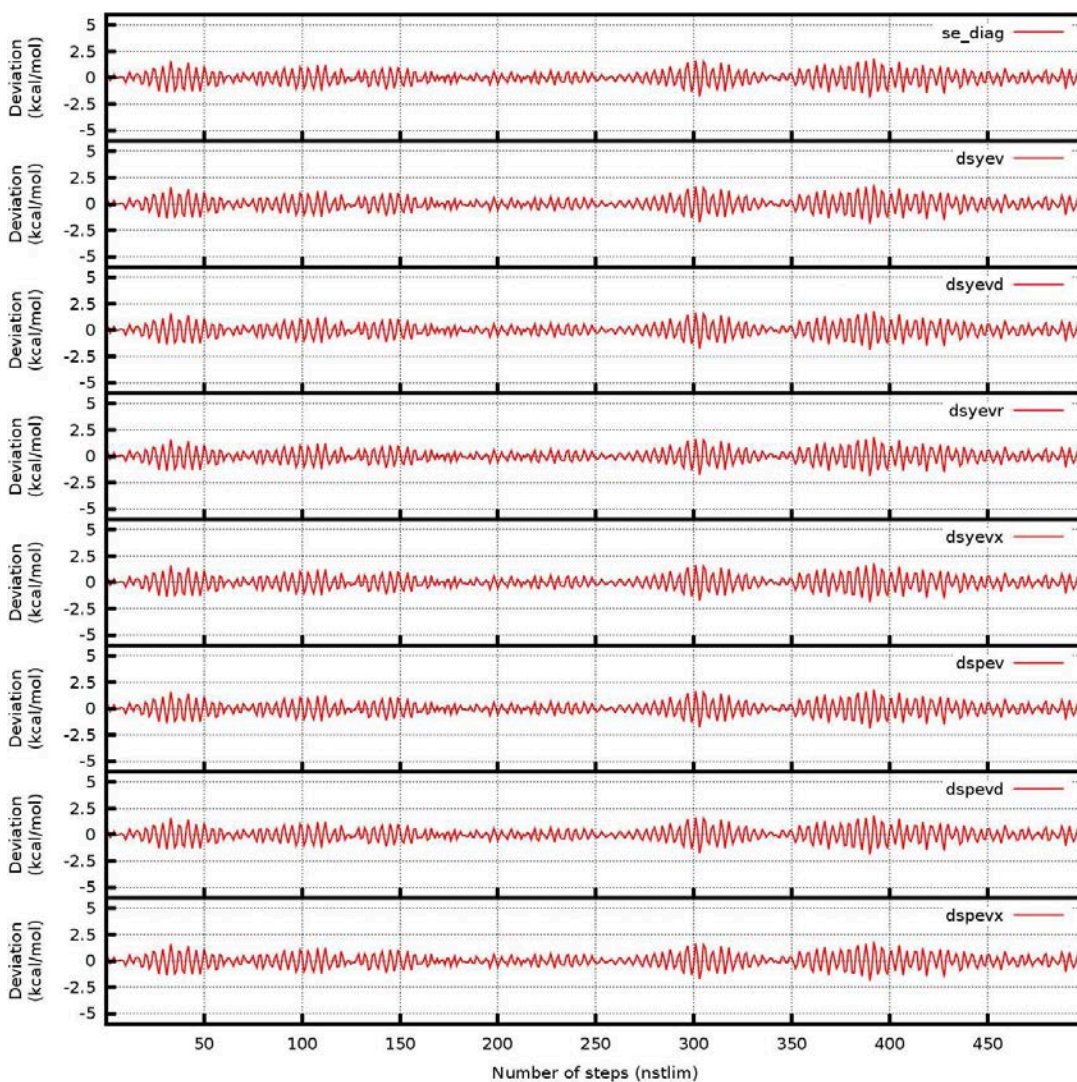


**Figure 3.46 :** A representation of the call scheme and number of calls in SEBOMD simulations. The number of calls for 64 and 216 water molecules are obtained with the condensed phase (n1) simulations, using full diagonalization (f1) and disregarding the long range interactions (lr0). The results obtained using D&C algorithm are indicated with blue boxes, and the results obtained using SCS algorithm are shown with red boxes. The densub subroutine is called only when D&C algorithm is used and denful subroutine is called only when SCS algorithm is used.

#### 3.4.2.4 Performance with LAPACK

The most time consuming part of the SEBOMD simulations (when long-range interactions are not calculated with a quantum Ewald summation) is the SCF procedure in which the Fock matrix is diagonalized. Diagonalization procedure in each SCF iteration is performed with a driver routine named diag (as diagp if pseudo-diagonalization is chosen) that uses a QR iteration algorithm. In addition to this solver, seven different eigenvalue solvers from LAPACK library (see Table 3.4) are implemented to improve the efficiency of SEBOMD simulations. The accuracies of the

implemented solvers are investigated by means of the conservation of the total energy along 500 steps of SEBOMD simulations of 64 water molecules in gas phase (n0) using a standard closed shell algorithm (m0). The deviations of the total energy from the average energies along the simulations that are performed with different eigenvalue solvers are depicted in Figure 3.47. It is observed that the deviation in total energy does not exceed 2.5 kcal/mol in all cases, indicating that all solvers compute the correct eigenvalue and eigenvector pairs.

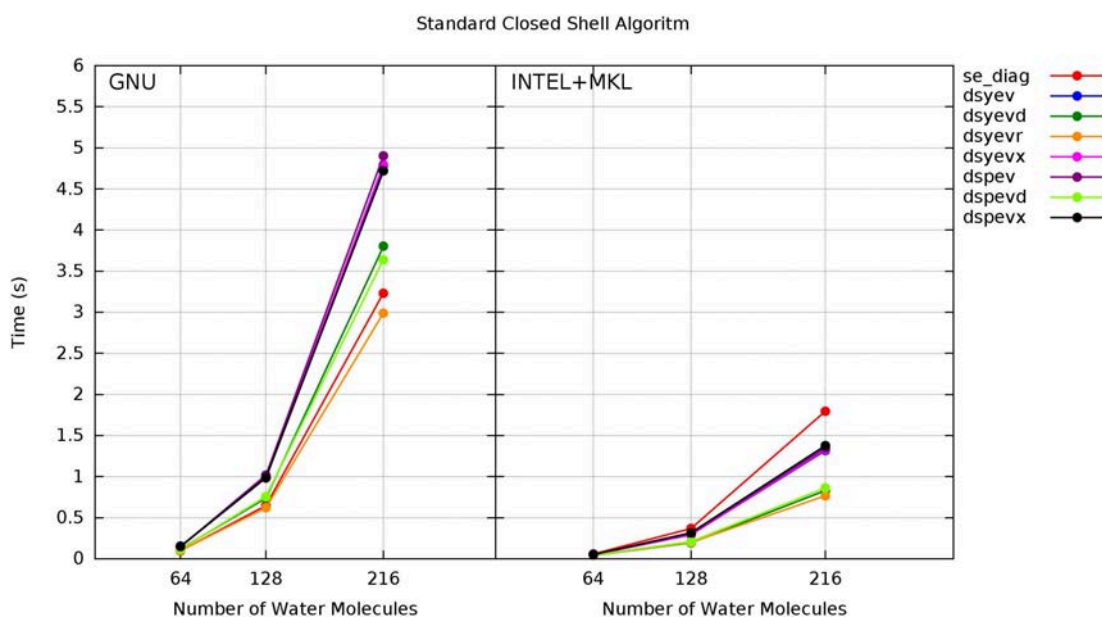


**Figure 3.47 :** Conservation of total energy ( $E_{tot}$ ) in terms of deviation from average energy along 500 steps of SEBOMD simulation of 64 water molecules with standard closed shell algorithm using implemented eigenvalue solvers.

Since eigenvalue solvers ensure the correct eigenvalue and eigenvector pairs, profiling studies are performed using three different boxes of water molecules (64, 128 and 216). SEBOMD simulations are performed with full-diagonalization in condensed

phase without long-range interactions. In order to observe the effect of different libraries from different vendors, Amber internal LAPACK libraries (compiled with gnu) or LAPACK libraries that are implemented in the Intel MKL are used. The consumed time per diagonalization step for each eigenvalue solver are calculated using the internal timer of `sander`.

The results obtained using the standard closed shell algorithm is depicted in Figure 3.48. It is observed that when the Amber internal LAPACK libraries are used, the default diagonalization routine `diag` is the fastest solver while the slowest one is `dsyev`. However, when Intel MKL is used, the `diag` solver becomes the slowest routine. Additionally, consumed time per diagonalization step for every solver is also reduced by using Intel compilers and MKL. The results indicates that the internal LAPACK library of Amber is not optimized enough to obtain high performance. It is found that, the fastest combination to diagonalize Fock matrix with standard closed shell algorithm is to use the `dsyevr` solver implemented in the Intel MKL and to compile the SEBOMD code with the Intel compiler.



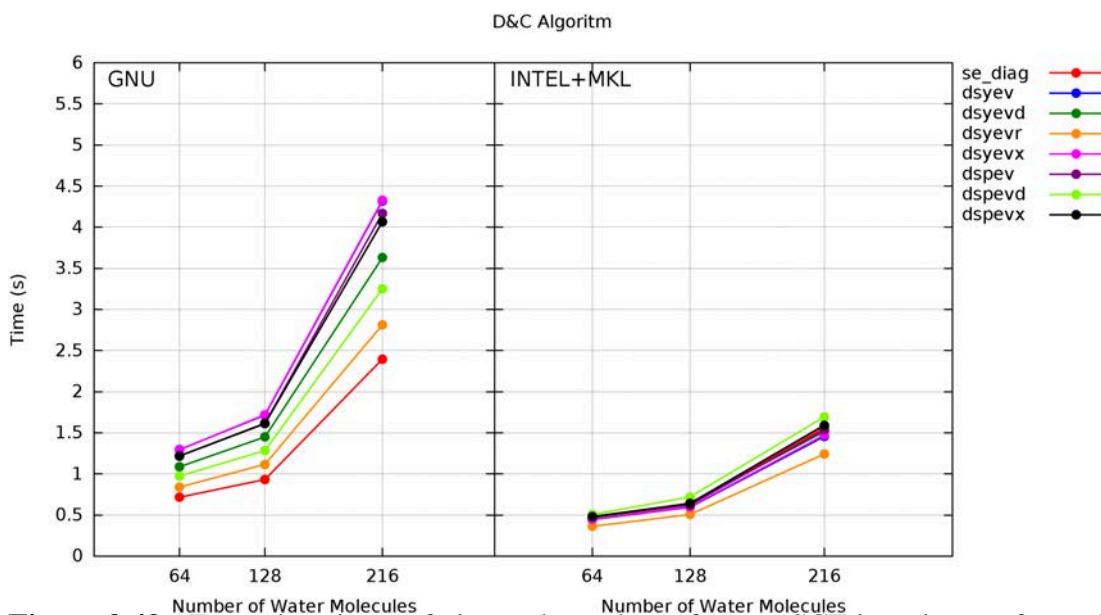
**Figure 3.48 :** Execution times of eigenvalue solvers in one SCF iteration performed in SEBOMD simulations of different sizes of systems using the standard closed shell algorithm. Each eigenvalue solver is represented with different color. Execution times are calculated for both the integrated LAPACK library (left) and the Intel MKL (right). The Gnu compiler is used to compile the SEBOMD code in the case of the integrated LAPACK library while the Intel compiler is used instead of the Gnu one in the case of the Intel MKL.

The diagonalization step is also profiled when the D&C algorithm is used and the results are depicted in Figure 3.49. It was observed that, if the Amber internal LAPACK library is used, the `diag` routine is still the fastest solver to diagonalize the Fock matrix. Similarly to the standard closed shell algorithm, the required time per diagonalization is reduced with the Intel MKL and the fastest solver is then `dsyevr`. The results indicates that the D&C algorithm compiled with the Intel MKL has better performance with large system size than with the Amber integrated LAPACK library.

In the SEBOMD simulations, the standard closed shell algorithm leads to better performance when the size of the system is small, however, the D&C algorithm provide better performance when the system size is large. This indicates that the performances of these two algorithms intersect at a point, namely the **crossover point**, when the performances are same with the same problem size. If the problem size exceeds this point, using the D&C algorithm instead of the standard closed shell algorithm would provide better performance. In case of the integrated LAPACK library (compiled with 'gnu' flag), the crossover point is found to be smaller than 216 water molecules. However, the crossover point is located beyond 216 water molecules when the Intel MKL is used. Additionally, the Intel MKL would allow larger systems to be diagonalized in lesser time with the D&C algorithm relatively to the integrated LAPACK library.

Profiling studies with different compilers and libraries, which pointed out the effect of external control on performance, led us to construct an algorithm to automatically choose the fastest eigenvalue solver (see Algorithm 1). This algorithm is implemented in the SEBOMD code and can be called with a diagonalization routine option. When a user give `diag_routine=0` as an input, the automatic eigenvalue selection is called and for the first eight SCF iterations all eigenvalue solvers that are implemented are scanned. After the first eight SCF iterations, the fastest solver is selected for further computations. The automatic solver selection is investigated with 1 step of SEBOMD simulation of 216 water molecules using standard closed shell algorithm and result is depicted in Figure 3.50. In the first SCF iteration the `diag` solver is used to diagonalize Fock matrix. In the following seven iterations `dsyev`, `dsyevd`, `dsyevr`, `dsyevx`, `dspev`, `dspevd`, and `dspevx` are called respectively. After the last eigenvalue solver is used (eight iteration) to diagonalize the Fock matrix, the





**Figure 3.49 :** Execution times of eigenvalue solvers for one SCF iteration performed in SEBOMD simulations of different sizes of systems using the D&C algorithm. Each eigenvalue solver is represented by different color. Execution times are calculated for both the integrated LAPACK library (left) and the Intel MKL (right). The Gnu compiler is used to compile the SEBOMD code in the case of the integrated LAPACK library while the Intel compiler is used instead of the Gnu one in the case of the Intel MKL.

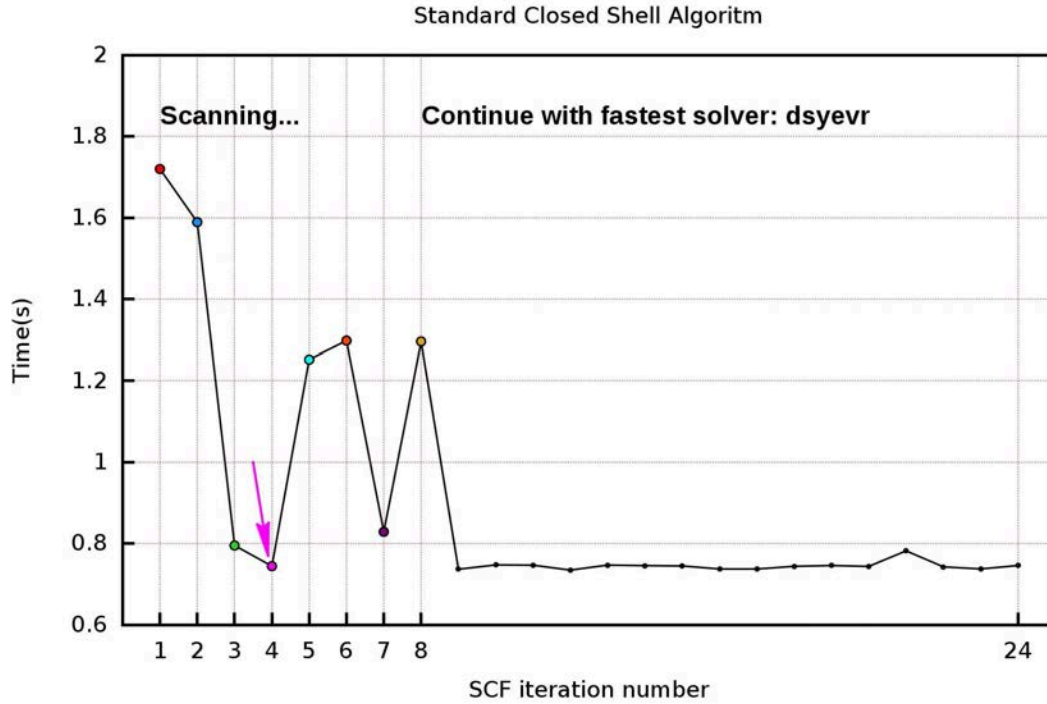
fastest routine, `dsyevr`, is chosen and the following iterations are performed with this solver.

### 3.4.2.5 Performance with SCALAPACK

In order to improve the scalability of SEBOMD simulations on parallel computers, the eigenvalue solvers that are in SCALAPACK (`pdsyev`, `pdsyevd`, `pdsyevr`, and `pdsyevx`) are implemented in the SEBOMD code. These SCALAPACK routines uses the same computation algorithms as in their LAPACK versions. In order to call these solvers, a linear algebra oriented message passing interface is established using BLACS routines. Implementation of the BLACS and SCALAPACK driver routines and of the calling procedures can be found in Algorithm 2 and Algorithm 3.

Previous versions of Amber would allow only serial execution of SEBOMD simulations using standard closed shell algorithm. The implementation of SCALAPACK solvers allows this algorithm to be run on multi-cores. However, when standard closed shell algorithm is used to run SEBOMD simulations on multi-cores, there could be only one `local_master`. Since there is only one `local_master`,



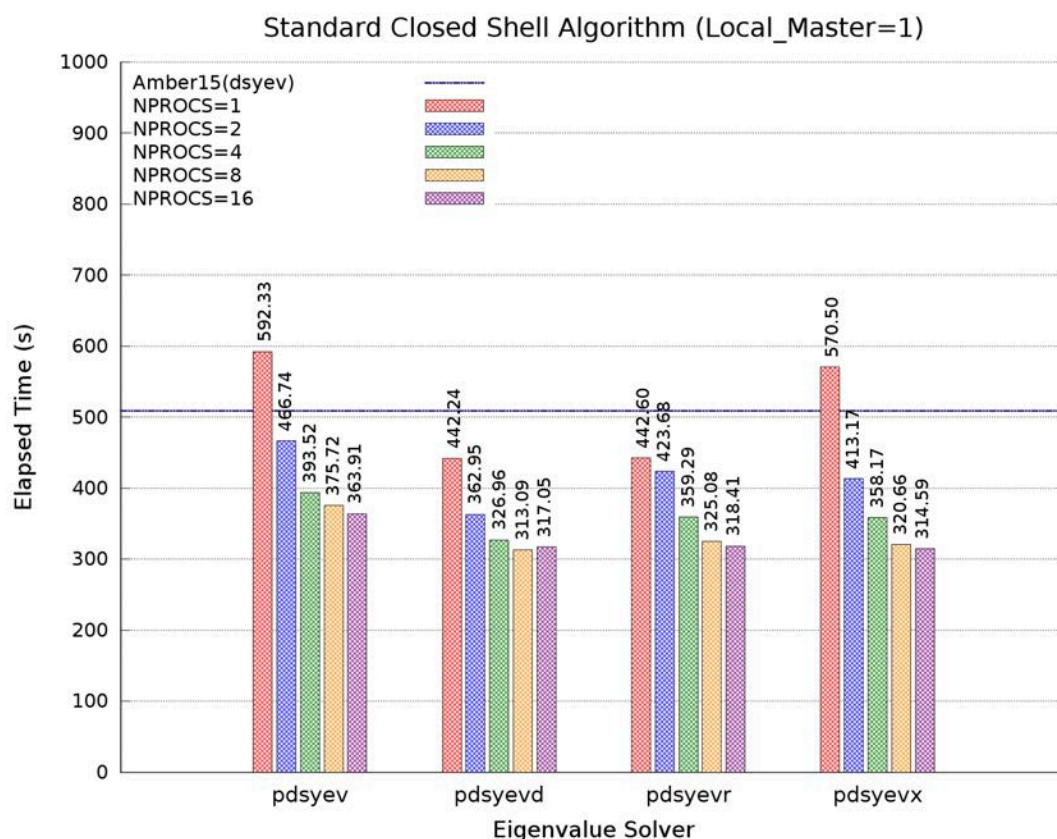


**Figure 3.50 :** The execution time of the eigenvalue solvers per SCF iteration with the automatic eigenvalue solver selection option. In the scanning section (first eight iterations), different solvers are used for each iteration. The different eigenvalue solvers are represented by different colors of points and the fastest solver (`dsyevr`) is depicted with an arrow.

all of the computations of SE part of the simulation are performed on single processor except the diagonalization procedure. In order to profile SCALAPACK routines, SEBOMD simulations of 216 water molecules are performed using this algorithm. The results that are obtained with SCALAPACK solvers are depicted in Figure 3.51. It was observed that the most scalable solver is `pdsyevx`. However, using SCALAPACK solvers with the standard closed shell algorithm does not provide high parallel efficiency ( $E(NPROCS)$ ) which is described as;

$$E(NPROCS) = \frac{Time_{serial}}{Time_{NPROCS} \times NPROCS} \quad (3.4)$$

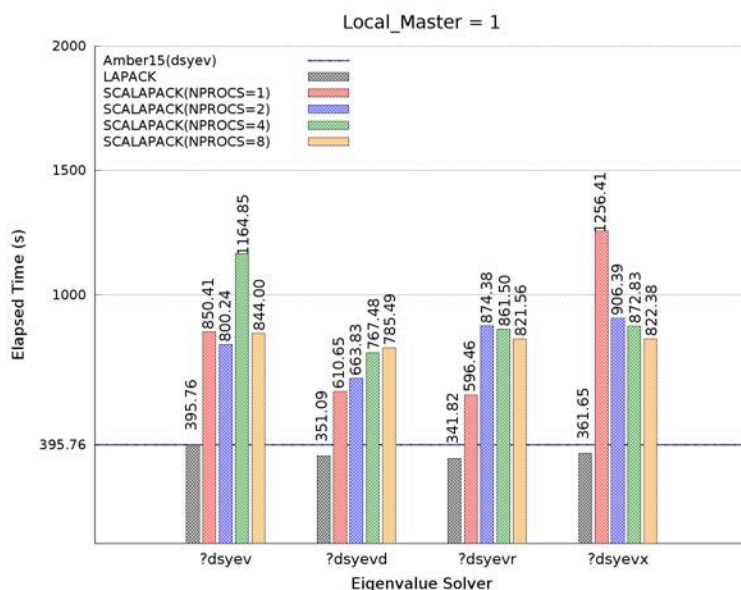
where  $Time_{serial}$  is the execution time with 1 processor,  $NPROCS$  is the number of processors, and  $Time_{NPROCS}$  is the execution time with  $NPROCS$  processors. The most efficient performance with standard closed shell algorithm can be obtained with `pdsyevx` on 2 cores (see Figure 3.51,  $E(2) = \frac{570.50s}{2 \times (413.17s)} = 0.69$ ). When a larger number of processors is used, the parallel efficiency is computed below 0.5 for all eigenvalue solvers.



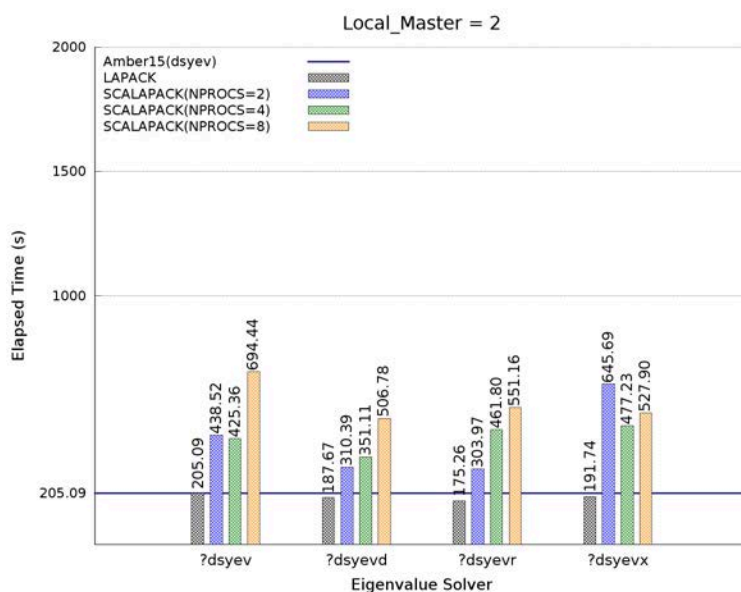
**Figure 3.51 :** Elapsed times obtained with SEBOMD simulations with standard closed shell algorithm using different eigenvalue solvers from SCALAPACK. Different number of processors are represented with different colors. Elapsed time obtained with the serial simulation using *dsyev* in Amber15 is shown with dashed line.

In the case of the D&C algorithm, the number of `local_masters` can be defined by the variable `nthread`, which is the total number of cores in a BLACS grid (see Algorithm 2). For instance, if `nthread=2` and the total number of processors (NPROCS) is equal to 8, there would be four `local_master` which are located on four BLACS grid (see Figure 3.44b). Results obtained using the D&C algorithm with a 6Å buffer and 216 water molecules are depicted in Figure 3.52. It is observed that the elapsed times obtained with SCALAPACK solvers are always larger than their LAPACK versions. Additionally, the elapsed times increase in most cases with the number of cores (i.e. `pdsyevd` in Figure 3.52a and Figure 3.52b). This indicates that the computation of small Fock matrices requires very short time relative to the required time for communication within the grid. Increasing the number of `local_masters` decreased the elapsed times (Figure 3.52b), yet LAPACK versions of the solvers are still faster than the SCALAPACK versions. Results that are obtained points out that using a small buffer region and performing diagonalization with SCALAPACK solvers

would not improve the performance of the D&C algorithm on parallel machines, since the Fock matrices are not large enough. Thus, a larger buffer region is chosen to increase the size of Fock matrices and to observe the effect of the size.



(a)



(b)

**Figure 3.52** : Elapsed times obtained from SEBOMD simulations of 216 water molecules using the D&C algorithm with **a.** buffer=6Å on 1 local\_master. **b.** buffer=6Å on 2 local\_masters. Eigenvalue solvers are depicted with "?" as first letter to indicate both LAPACK and SCALAPACK versions. In SCALAPACK versions of the solvers, "?" indicates "p". Reference elapsed time obtained with the dsyev solver in Amber15 is shown with dashed line while elapsed times with LAPACK versions of the solvers are shown with black columns. Number of used processors with SCALAPACK routines are depicted with red (1 processor), blue (2 processor), green (4 processors), and yellow (8 processors) columns.

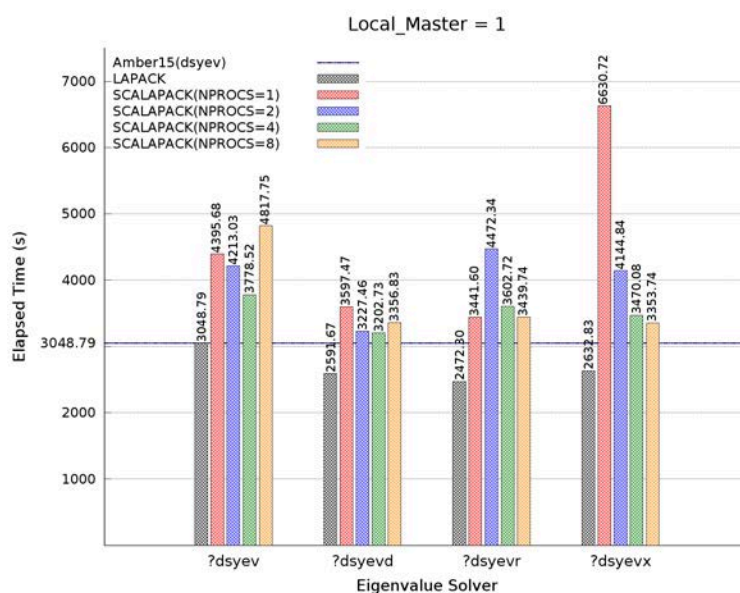
The larger buffer is chosen as 8Å, and the results that are obtained are given in Figure 3.53. The problem of the increment of elapsed times with the increment of cores is mainly solved by using large Fock matrices until eight cores with one `local_master` (Figure 3.53a). When eight cores are used, elapsed time increases except in `pdsyevx`. However, with two `local_master`, the increment of elapsed times is stopped (except with `dsyevr`). The main cause of different patterns obtained with different number of `local_masters` is due to the required communication time. Since the larger `local_master` number decreases the size of the grid (with the same number of cores and relative to less `local_master`), communication time would become lesser. Furthermore, the number of subproblems per master will decrease with larger `local_master` number.

Results point out that there is a critical balance between the size of Fock matrices, the number of subproblems (subsystems in SEBOMD) per `local_master`, and the size of the grids. It was observed that the most scalable solver is `pdsyevx`. However, LAPACK versions of the solvers are found to require less time than SCALAPACK solvers in all cases. Thus, using LAPACK solvers instead of SCALAPACK versions, would result in more efficient SEBOMD simulations.

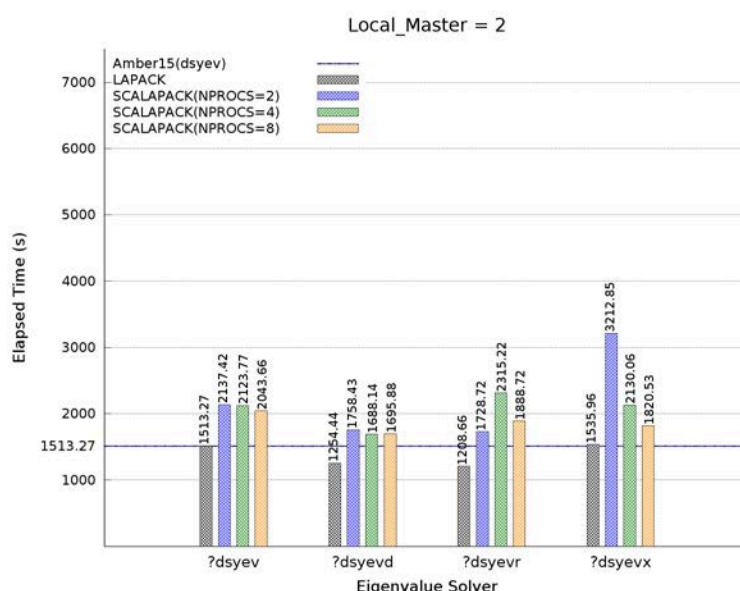
#### 3.4.2.6 Performance with Intel MKL threading

The SCALAPACK solvers could not improve the efficiency of SEBOMD simulations and using Gnu compilers with integrated LAPACK libraries cannot provide good performance. SEBOMD simulations can have best performance when Intel compilers and LAPACK solvers integrated in Intel MKL are used. The advantage of using the Intel MKL is that one can use threads which are thread-safe. Thus, threaded versions of LAPACK solvers have been investigated.

The first system in question was the simulation box containing 216 water molecules. In order to profile the threaded solvers, simulations are performed with the D&C algorithm using only one processor. This would allow to observe if using threaded versions of LAPACK solvers can improve performance. Results obtained using different number of threads are depicted in Figure 3.54. Elapsed times obtained with LAPACK solvers that do not use the threaded versions, and the reference elapsed time obtained with the `dsyev` solver that is implemented in Amber15 is also taken



(a)



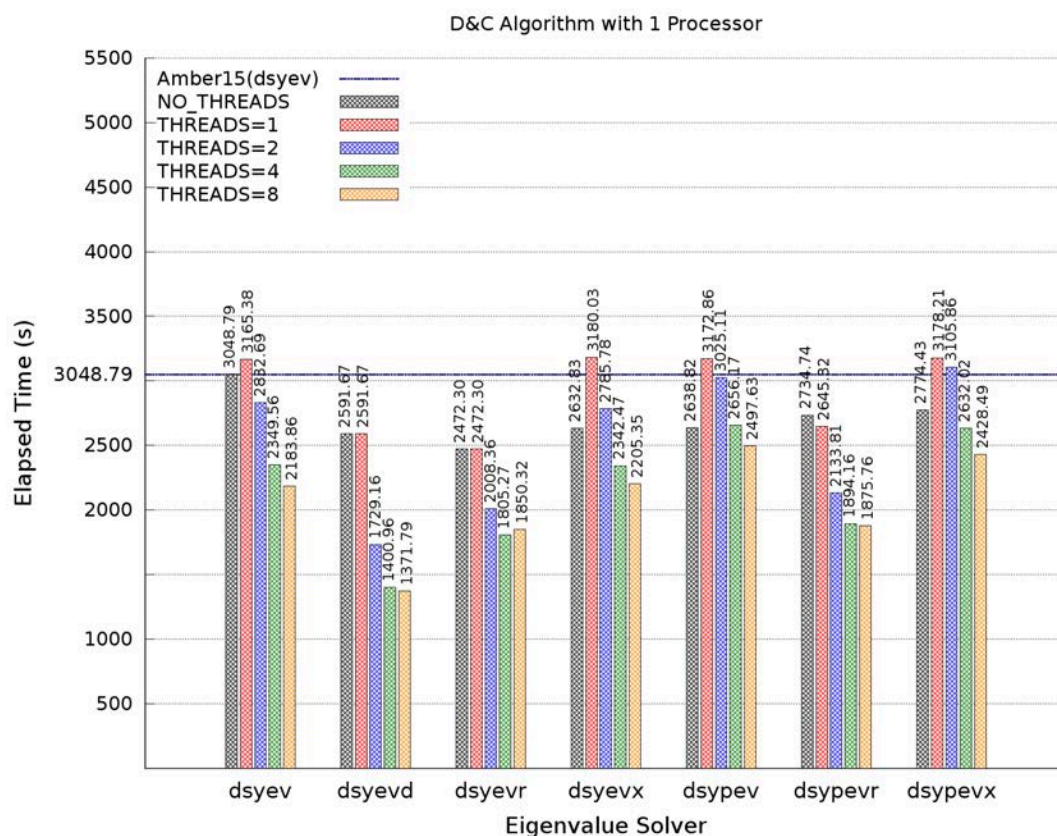
(b)

**Figure 3.53 :** Elapsed times obtained from SEBOMD simulations of 216 water molecules using the D&C algorithm with **a.** buffer=8Å on 1 local\_master. **b.** buffer=8Å on 2 local\_masters. Eigenvalue solvers are depicted with "?" as first letter to indicate both LAPACK and SCALAPACK versions. In SCALAPACK versions of the solvers, "?" indicates "p". Reference elapsed time obtained with the dsyev solver in Amber15 is shown with dashed line while elapsed times with LAPACK versions of the solvers are shown with black columns. Number of used processors with SCALAPACK routines are depicted with red (1 processor), blue (2 processor), green (4 processors), and yellow (8 processors) columns.

into consideration to inspect if the use of threads can improve performance of the current SEBOMD version. It was observed that, for all eigenvalue solvers, one can



perform simulations that requires less elapsed times using threads. In case of solvers `dsyev`, `dsyevx`, `dspev` and `dspevx`, threaded versions requires more time if only 1 thread is used. However, performance can be improved with the increase in the number of threads. In case of `dsyevx`, simulation ends with less elapsed time if threads are more than 2, while `dspev` and `dspevx` requires more than 4 threads to be faster than unthreaded versions. The threaded versions of `dsyevd`, `dsyevr` and `dspevr` are found to provide better performance than their unthreaded versions. The best performance obtained with eigenvalue solver `dsyevd`, which provides about 35% gain even with 2 threads. Since the best performance is obtained with eigenvalue solver `dsyevd`, further profiling studies with threaded versions are performed with this solver.



**Figure 3.54 :** Elapsed times obtained with 10 steps of SEBOMD simulations of 216 water molecules with  $\text{buff}=8\text{\AA}$  and using implemented solvers with Intel MKL and threads. Reference elapsed time obtained with the `dsyev` solver in Amber15 is shown with dashed line while elapsed times without threading are shown with black columns. Number of threads that are used are represented with red (1 thread), blue (2 threads), green (4 threads), and yellow (8 threads) columns.

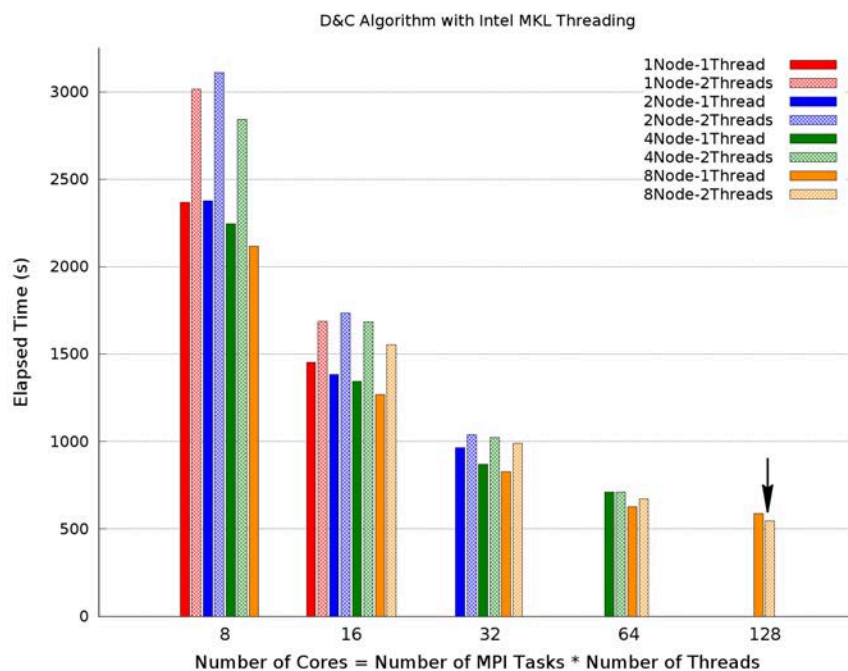
Simulations performed with 216 water molecules provide information on how to improve efficiency with LAPACK solvers. However, it is crucial to observe if efficiency of SEBOMD simulations can be improved with larger systems, such as a solvated box containing a small protein. Therefore, SEBOMD simulations with a simulation box that contains a small protein having 20 residues (TrpCage, 284 atom) and 730 water molecules are performed with the threaded version of the `dsyevd` eigenvalue solver. In all of the computations, D&C algorithm is used with `buffer=8Å`. Simulations are performed using different number of nodes where each node consists of 16 cores. Results obtained with the use of `dsyevd` on different number of nodes using 1 and 2 threads are depicted in Figure 3.55. Elapsed times are investigated in conjunction with the total number of cores (Total number of cores = Number of MPI tasks \* Number of threads). It is observed that the performances obtained using 1 thread are mostly better than the performances of the simulations using 2 threads. However, the elapsed times of the simulations using different number of threads become similar to each other with the increment of the total number of cores. When 64 cores are used in total, elapsed times obtained with 1 and 2 threads on 4 nodes becomes almost similar (i.e. 64 MPI tasks \* 1 thread vs 32 MPI tasks \* 2 threads). In case of 128 cores, using 2 threads instead of 1 thread provided better performance. Thus, if the total number of cores is larger than 64 cores, using at least 2 threads with `dsyevd` in simulations of large systems would require less time, and provide higher efficiency than using 1 thread or non-threaded version on the same number of cores.

### 3.4.3 Concluding remarks

The diagonalization of the Fock matrix in the semiempirical part of SEBOMD simulations is time consuming and the current version of SEBOMD cannot scale well on parallel computers. In order to improve performance of the SEBOMD approach, new eigenvalue solvers from LAPACK & SCALAPACK have been implemented and evaluated.

Preliminary profiling studies have pointed out that if full-diagonalization is applied in condensed phase systems without long-range interactions, SCF routines becomes the dominating part in SEBOMD by means of elapsed time. Thus, the implemented solvers





**Figure 3.55 :** Elapses times obtained with 10 steps of SEBOMD simulations of solvated box with protein using  $\text{buff}=8\text{\AA}$  and threaded the `dsyevd` solver in Intel MKL. Elapsed times using 1 thread are represented with solid columns, while 2 threads are represented with pattern filled columns. Different number of nodes are represented with red (1 node), blue (2 nodes), green (4 nodes), and yellow (8 nodes) columns.

have been profiled using these methodological options in order to observe performance that is mostly dependent on the SCF part.

Seven different LAPACK eigenvalue solvers have been implemented in total and deviation of the total energy from the average energy indicated that all solvers found the correct eigenpairs and could be used for molecular dynamics simulations. Profiling studies using increasing sizes of systems (from 64 to 216 water molecules) have been performed using either Gnu or Intel compilers to assess the effect of external control on the performance. The time per diagonalization procedure has been computed with the internal timer of `sander`. It was observed that the Gnu compiler with the integrated LAPACK solvers did not provide high performance both for standard closed shell and D&C algorithms. Moreover, the default solver in SEBOMD (the `diag` routine) was still the fastest diagonalizer with these options, and only solver `dsyevr` could provide similar performance to the `diag` routine. However, the performance increased when the Intel MKL was used with the Intel compiler. When a standard closed shell algorithm was applied, all of the implemented solvers were faster than the `diag` routine compiled with the Intel MKL option. In the case of

the D&C algorithm, the default solver was found to be faster than the solvers that uses packed storage (`dspev`, `dspevd`, and `dspevx`). For both the standard closed shell and the D&C algorithms, the solver `dsyevr` exhibited higher performance than both newly implemented solvers and the `diag` routine. The obtained results pointed out that the performance of the diagonalization procedure was substantially affected by external control. Thus, to ensure the selection of the fastest diagonalizer, an automatic eigenvalue solver selection module has been implemented. When this module was called, each diagonalizer was tested respectively during a scanning period (the first eight SCF iterations) and the fastest solver was chosen and used for further computations.

In order to improve efficiency of SEBOMD on parallel computers, solvers from SCALAPACK have also been implemented in addition to the LAPACK solvers. These solvers mainly uses similar algorithms to compute eigenpairs as in their respective LAPACK versions. Implementation of the SCALAPACK solvers allowed us to use standard closed shell algorithm in SEBOMD simulations, for the first time, on parallel machines. It was observed that SCALAPACK solvers did not scale well and the best efficiency among SCAPACK solvers could be obtained with `pdsyevx` on 2 cores. Simulations with the D&C algorithm pointed out that, SCALAPACK solvers did not provide higher efficiency as in standard closed shell algorithm. In fact, the D&C algorithm had worse performance than the standard closed shell algorithm when there was only one `local_master`. Additionally, the use of different sizes of buffers showed that the performance strictly depended on the size of the Fock matrix. Since the sizes of Fock matrices in the D&C algorithm are not large enough, one cannot obtain higher efficiency with SCALAPACK solvers than with the LAPACK solvers.

The higher performance of LAPACK solvers vs. their SCALAPACK versions led us to question if performance can be improved using the threaded versions of the diagonalizers. In order to investigate the efficiency with threaded eigenvalue solvers, preliminary simulations of the box of 216 water molecules has been performed using the D&C algorithm on 1 core with different number of threads. It was observed that one can improve the performance of diagonalization procedure with the use of threads, especially with the `dsyevd` solver. Using preliminary studies as a reference, simulations with a box containing a small protein (TrpCage, 20 residues, 284 atoms)

and 730 water molecules have been performed with the D&C algorithm. Simulations have been performed with different number of total cores on different number of nodes using different number of threads. Elapsed times were compared by means of total number of cores. Results pointed out that, when the total number of cores were not large enough, using only 1 thread provided better performance. However, performances of simulations with 1 and 2 threads became similar with the increase in total number of cores. When total number of processors that were used reached 128 cores, simulations required less time with more than 1 thread.

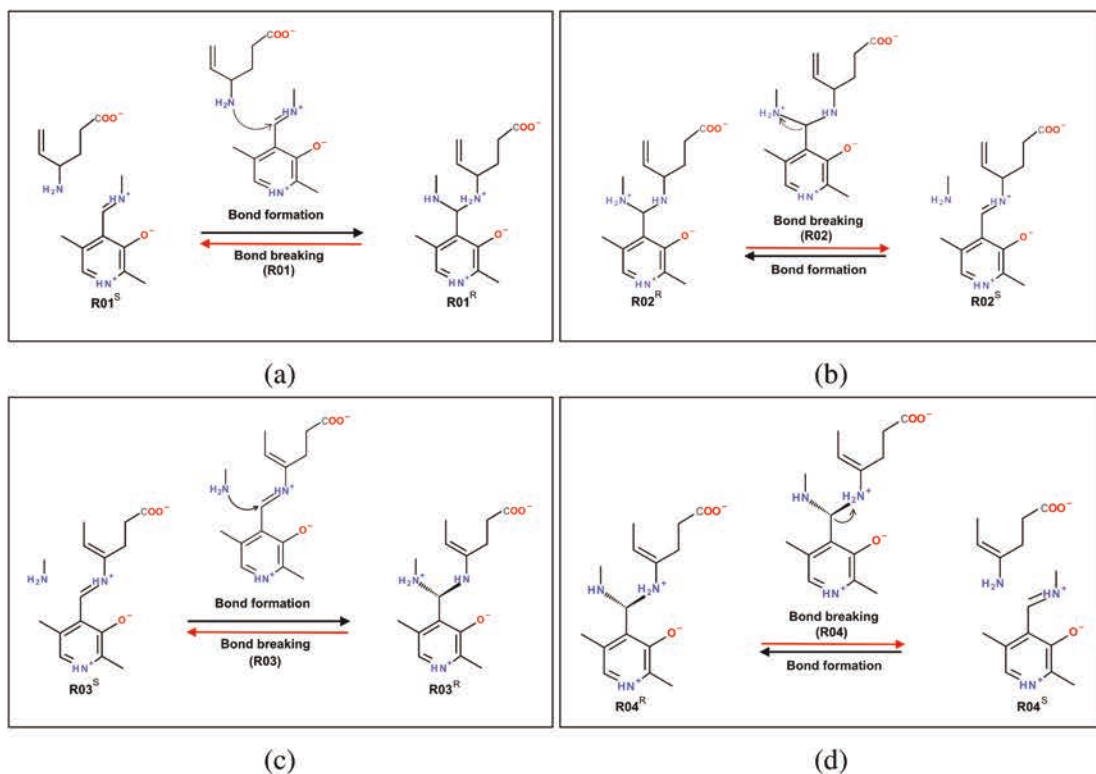
### **3.5 SEBOMD Simulations on C-N bond breaking reactions involving PLP complexes**

The use of small representative models provided some clues regarding the reactions occurring within GABA-AT [3, 33, 34, 86], while large models [65] and QM/MM calculations [87–90] provided information on the effect of the surrounding residues on the reactivity of PLP dependent enzymes. The effect of the explicit solvation on the stabilities of GABA conformers have been studied at QM level [19], and the stabilities of GABA analogues have been addressed with QM/MM calculations [91].

In addition to the various QM and QM/MM calculations, SEBOMD simulations can also be performed to investigate chemical reactions involving PLP complexes since the approach uses a NDDO Hamiltonian as a reactive force field. SEBOMD methodology have been successfully applied to molecules in condensed phase, including liquid water and N-methyl acetamide in water [82, 83]. However, chemical transformations such as bond forming/breaking have not been investigated yet. The newly implemented eigenvalue solvers in SEBOMD now enables the simulations of thousands of atoms, but simulations of larger proteins still cannot be performed due to the computational demand. Instead, model structures and explicit water molecules can be included within the samples.

In order to investigate chemical transformations with SEBOMD simulations, four different C-N bond formation/breaking reactions are studied using umbrella sampling. The reversible reactions are involved in the proposed inactivation mechanisms of GABA-AT with  $\gamma$ -vinyl-GABA [46] and have been studied theoretically [34]. The two dimensional representation of the investigated reactions are depicted in Figure 3.56.

The first reaction in question involves the nucleophilic attack of the N atom of  $\gamma$ -vinyl-GABA to the C atom of PLP-methylamine complex (Figure 3.56a). This reaction is the first step of the proposed external aldimine formation between  $\gamma$ -vinyl-GABA and PLP. The second reaction is the C-N bond breaking within the  $\gamma$ -vinyl-GABA-PLP-methylamine complex (Figure 3.56b). This reaction ends up with a  $\gamma$ -vinyl-GABA-PLP complex where the methylamine is the leaving group. The third reaction is the first step of the proposed enamine pathway which involves the attack of methylamine on the PLP imine carbon (Figure 3.56c). The last reaction in question also occurs in the proposed enamine pathway and involves the C-N bond breaking which produces PLP-methylamine complex (Figure 3.56d). For simplicity, all of the reversible reactions are modelled as bond breaking reactions, meaning that the products are the separated molecules.



**Figure 3.56** : Two dimensional representations of the reactions that are chosen for SEBOMD simulations. (a) First step of external aldimine formation. (b) The C-N bond breaking within the  $\gamma$ -vinyl-GABA-PLP-methylamine complex. (c) The attack of methylamine on the PLP imine carbon. (d) The formation of PLP-methylamine complex within the proposed Enamine pathway. The reactants are denoted with superscript *R*, the separated products are denoted with superscript *S*, and the modelled bond breaking reactions are depicted with red arrows.

### 3.5.1 Computational details

#### 3.5.1.1 QM calculations

Prior to the SEBOMD simulations, the reactions are modelled with the QM calculations. All of the QM calculations were performed with the Gaussian'09(G09) program package [160] and the solvation effect of water was considered using the integral equation formalism variant of the polarizable continuum model (IEFPCM) [157, 158, 193, 194]. Both geometry optimizations and frequency calculations in the water environment were performed with the B3LYP [131, 132, 134, 139, 140] method using the 6-31+G(d,p) basis set. Semiempirical AM1 [111], PDDG [120], and PM3 [113] Hamiltonians were also considered. Intrinsic Reaction Coordinate (IRC) calculations were performed for further characterization of the transition states.

#### 3.5.1.2 QM/MM calculations

The samples were prepared using the optimized geometries of the reactants that were obtained with the QM calculations. The molecules have been prepared using the Antechamber program [175] of AMBER 14 [238]. The samples were solvated with explicit TIP3P water molecules [223]. The simulation boxes were cubic, of edge length 37.252Å, had an initial density of 1.0, and contained about 1700 water molecules. The QM region includes all solute atoms (44 atoms in all samples), and the MM region corresponds to the water molecules.

Periodic boundary conditions were imposed on the systems. The Particle Mesh Ewald summation technique was used with the default 8Å cutoff [187]. Simulations were performed with the use of PM3 [113], and AM1/d [118, 119] Hamiltonians to describe the solute. The equilibrations of the samples were performed in a two stage process. First, after a short 1000 step minimization with a harmonic restraint of 50 kcal/mol/Å<sup>2</sup> on all non-hydrogen atoms, a 5ps molecular dynamics was performed using isotropic position scaling at 10K. Positional restraints were not applied during the simulations. A time step of 0.5fs was used by the implementation of the SHAKE algorithm for the bonds involving hydrogens [181] within the MM region. Second, the thermostat temperature was linearly increased from 10K to 300K during a 10ps simulation using the Andersen temperature coupling [188]. Production simulations were performed at 300K in the NPT ensemble using the Andersen temperature coupling [188] and a 8Å

cutoff distance was used for the QM region. The umbrella sampling protocol was applied to simulate the C-N bond breaking reactions. The C-N distance was increased from 1.4Å to 5.0Å by a step of 0.5Å. For each window, 15ps of simulation were performed which led to a total simulation time of 1095ps for each reaction. Extraction of the potential of mean force (PMF) was performed with the use of the weighted histogram analysis method (WHAM) [239].

### 3.5.1.3 SEBOMD simulations

The samples were solvated with explicit TIP3P water molecules [223]. The simulation boxes were cubic, of edge length 19.711Å, had an initial density of 1.0, and contained about 234 water molecules. Periodic boundary conditions were imposed on the systems. The Particle Mesh Ewald summation technique was used with a 6Å cutoff [187]. All of the simulations were performed at 300K using the Andersen temperature coupling [188]. The PM3 Hamiltonian [113] was used with the PIF3 modification [128]. Longrange interactions were added by polarizing the wave functions using Mulliken charges (quantum Ewald summation), and the D&C algorithm was used. The extent of the first buffer region from the core was fixed at 6Å. The equilibration of the samples were performed at constant volume using different  $\lambda$  values which enabled the computation of a mixed energy value between SEBOMD and full MM computations according to;

$$E_{pot} = \lambda E_{SEBOMD} + (1 - \lambda) E_{MM} \quad (3.5)$$

The  $\lambda$  value was increased from 0.0 to 1.0 by steps of 0.1 corresponding to 11 different 1ps SEBOMD simulations. Production simulations were performed in the NPT ensemble using isotropic position scaling, and the pressure was controlled with a Monte Carlo barostat. Umbrella sampling protocol was applied to simulate the C-N bond breaking reactions. The C-N distance was increased from 1.4Å to 5.0Å by steps of 0.1Å. For each window, 15ps of simulation were performed, leading to a total simulation time of 555ps for each reaction. Extraction of the PMF was performed with the use of the WHAM [239].

### 3.5.2 Results and Discussion

#### 3.5.2.1 Preliminary study on bond breaking reactions

SEBOMD uses a QM semiempirical NDDO Hamiltonian (AM1, PM3, etc.) to perform molecular dynamics [83]. In order to describe intermolecular interactions in water and solute-solvent interaction correctly, SEBOMD simulations were performed using the PM3 Hamiltonian [113] with the PIF3 modification [128]. However, it is important to investigate the effect of different computational approaches and Hamiltonians on the reactions to prospect the possible alterations on the potential energy surfaces. Therefore, the reactions were modelled by QM calculations using implicit water as solvent, and QM/MM calculations were performed with explicit water molecules that were located in the MM region. The C-N distances within the transition state centres, the activation free energy barriers of the bond breaking reactions, and the relative free energies of the separated products were benchmarked to study the variations on the reaction coordinates.

The C-N distances in the transition states that have been computed with QM and QM/MM calculations are given in Table 3.5. The transition states were located with a single negative eigenvalue in the QM calculations while they were located with a peak in the free energy computed by the WHAM in the case of QM/MM calculations. The calculated C-N distances with different computational approaches and Hamiltonians differ from each other up to 0.25Å. The QM calculations using DFT and AM1 both yielded similar C-N distances for every reaction, while PDDG (except for **R01**), and PM3 overestimated the same distance relative to DFT and AM1 counterparts. The QM/MM calculations using AM1d yielded almost similar C-N distances with the QM calculations using AM1. Additionally, the computed C-N distances in QM/MM calculations using PM3 were almost similar with its QM counterparts. We also tried to investigate the reactions using AM1 Hamiltonian with QM/MM calculations. However, a proton transfer observed within the reactants during the relaxation simulations for QM/MM calculations preventing the transition state structure modelling.

The activation free energy barriers and the relative free energies of the separated products computed with QM and QM/MM calculations are depicted in Table 3.6.



**Table 3.5 :** The  $C - N$  distances ( $\text{\AA}$ ) in transition state structures. The superscript  $i$  indicates the quantum mechanical calculations with implicit water. The superscript  $e$  indicates the QM/MM calculations with explicit MM water molecules.

	DFT <sup>i</sup>	PDDG <sup>i</sup>	AM1 <sup>i</sup>	PM3 <sup>i</sup>	AM1d <sup>e</sup>	PM3 <sup>e</sup>
<b>R01</b>	1.99	1.96	1.96	2.10	1.96	2.10
<b>R02</b>	1.90	2.04	1.98	2.15	2.01	2.12
<b>R03</b>	2.05	2.16	2.00	2.21	1.98	2.15
<b>R04</b>	1.96	2.06	1.96	2.15	1.94	2.10

It was observed that the computed activation free energy barriers were mainly overestimated in QM calculations with PM3 relative to both DFT, PDDG and AM1 except for **R01**. In addition to the QM calculations, PM3 Hamiltonian yielded highest activation free energy barriers for all reactions in QM/MM calculations. When the relative free energies of separated products were taken into account, it was observed that the bond breaking reactions are mainly exothermic with QM calculations. However, the use of a PM3 Hamiltonian yielded endothermic **R03** and **R04**. Furthermore, all of the reactions except **R01** were also found to be endothermic in the QM/MM calculations using PM3 Hamiltonian.

**Table 3.6 :** The activation free energy barriers (kcal/mol) of the located transition states and the relative free energies of the separated molecules. The reference structures for each reaction are the complexes in which the solute molecules are covalently bound. The superscript  $s$  is used to denote the total relative free energies of the separated molecules. The superscript  $i$  indicates the quantum mechanical calculations with implicit water. The superscript  $e$  indicates the QM/MM calculations with explicit MM water molecules.

	<b>R01</b>		<b>R02</b>		<b>R03</b>		<b>R04</b>	
<b>DFT<sup>i</sup></b>	4.9	−9.5 <sup>s</sup>	0.3	−13.0 <sup>s</sup>	7.8	−5.4 <sup>s</sup>	3.4	−5.5 <sup>s</sup>
<b>PDDG<sup>i</sup></b>	1.9	−12.8 <sup>s</sup>	4.0	−10.7 <sup>s</sup>	10.9	−0.8 <sup>s</sup>	5.8	−8.6 <sup>s</sup>
<b>AM1<sup>i</sup></b>	6.6	−10.0 <sup>s</sup>	5.0	−11.9 <sup>s</sup>	7.0	−10.5 <sup>s</sup>	6.7	−10.5 <sup>s</sup>
<b>PM3<sup>i</sup></b>	6.0	−9.5 <sup>s</sup>	9.0	−4.9 <sup>s</sup>	15.3	3.2 <sup>s</sup>	10.4	0.8 <sup>s</sup>
<b>AM1d<sup>e</sup></b>	9.0	−9.3 <sup>s</sup>	9.8	−4.6 <sup>s</sup>	10.6	−4.5 <sup>s</sup>	6.4	−14.9 <sup>s</sup>
<b>PM3<sup>e</sup></b>	12.6	−1.3 <sup>s</sup>	14.6	3.3 <sup>s</sup>	15.6	3.7 <sup>s</sup>	15.4	1.2 <sup>s</sup>

Overall, it was observed that the PM3 Hamiltonian yielded the largest C-N distances for the bond breaking reactions in both QM and QM/MM calculations. The highest activation free energy barriers were mainly computed with the PM3 Hamiltonian. Additionally, the QM calculations with PM3 yielded endothermic **R03** and **R04**, and the QM/MM calculations with PM3 yielded endothermic **R02**, **R03**, and **R04**.

### 3.5.2.2 SEBOMD simulations

The SEBOMD simulations were performed on selected the C-N bond breaking reactions (see Figure 3.56) using the PM3 [113] Hamiltonian with the PIF3 modification [128]. In order to examine the SEBOMD capabilities in modelling reactions and in computing reaction free energies, the variation of the *C – N* distances and the activation free energy barriers were investigated and compared to the QM and the QM/MM calculations using the same Hamiltonian.

The calculated *C – N* distances and the activation free energy barriers are given in Table 3.7. It is observed that, the *C – N* distances in transition state structures are very similar to the QM and QM/MM calculations (PM3<sup>i</sup>, and PM3<sup>e</sup> respectively, see Table 3.5). The computed distances in the SEBOMD simulations differed by up to 0.04Å from the PM3<sup>i</sup> and the PM3<sup>e</sup> calculations for all reactions.

**Table 3.7 :** The *C – N* distances (Å) in transition state structures and the calculated activation free energy barriers ( $\Delta G^\ddagger$ , kcal/mol) in SEBOMD simulations of C-N bond breaking reactions **R01**, **R02**, **R03**, and **R04**.

	C-N distance	$\Delta G^\ddagger$
<b>R01</b>	2.10	16.6
<b>R02</b>	2.18	18.4
<b>R03</b>	2.19	18.9
<b>R04</b>	2.11	15.4

It is observed that the SEBOMD simulations overestimated the activation free energy barriers about 3-10 kcal/mol relative to the QM counterparts (PM3<sup>i</sup>), and 0-4 kcal/mol relative to the QM/MM counterparts (PM3<sup>e</sup>). The activation free energy barrier for **R01** was overestimated by 10.6 kcal/mol relative to the PM3<sup>i</sup> calculations, and 4 kcal/mol relative to the PM3<sup>e</sup> calculations. The SEBOMD simulations overestimated

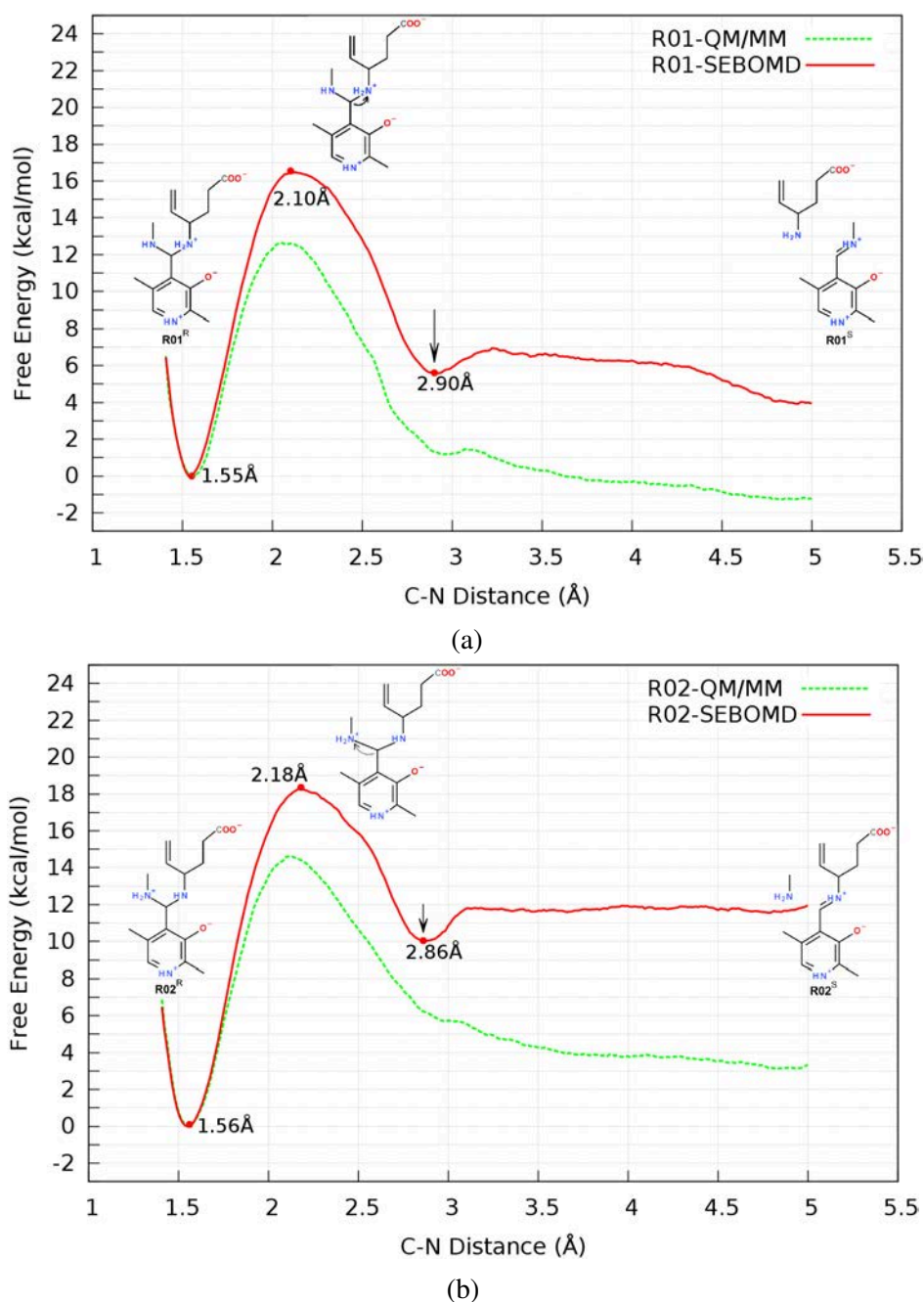
the barrier of **R02** about 9 kcal/mol relative to the the PM3<sup>i</sup> calculations, and 4 kcal/mol relative to the PM3<sup>e</sup> calculations. In case of **R03**, the barrier was overestimated about 3 kcal/mol relative to both of the PM3<sup>i</sup> and the PM3<sup>e</sup> calculations. The activation energy barrier of **R04** was overestimated by about 4 kcal/mol relative to the PM3<sup>i</sup> calculations, and was similar to the PM3<sup>e</sup> calculations.

The reaction coordinates were also investigated by means of PMF. The results are depicted in Figure 3.57. A complex formation was observed in the QM/MM calculations of the the first reaction (**R01**) where the corresponding C-N distance was about 2.90Å. However, SEBOMD simulations located a complex not only for the first reaction (**R01**) but also for the second reaction (**R02**). The distance between the corresponding C and N atoms in the complexes was computed about 2.9Å as in the QM/MM calculations but for both of the reactions. The three dimensional representations of some selected conformers of the complexes are depicted in Figure 3.58. In almost all of the complex conformers in **R01**, a water molecule was found to donate hydrogen bonds to the carboxyl group oxygen atom of vigabatrin, and also to the carbonyl oxygen of PLP (Figure 3.58a). In the case of **R02**, a water molecule that donates a hydrogen bond to the carboxyl group oxygen atom of vigabatrin (Figure 3.58b) was found. It was observed that, the same oxygen atom of the carboxyl group accepts another hydrogen bond from the imine nitrogen of the external aldimine (formed by PLP and vigabatrin).

### 3.5.3 Concluding remarks

The SEBOMD simulations have been performed using umbrella sampling on four different C-N bond breaking reactions that involve PLP complexes. The first example of the SEBOMD simulations on chemical transformations provide general insights on the capabilities of the SEBOMD methodology to follow reaction coordinates and to provide relative free energy estimates.

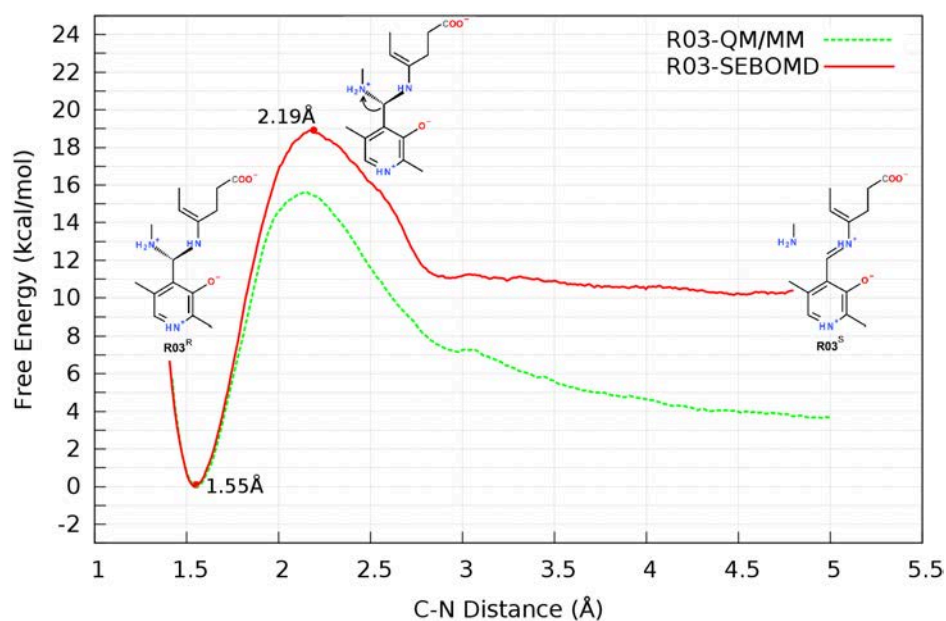
In order to investigate the possible alterations on the potential energy surfaces, different computational approaches and Hamiltonians have been used to model the reactions prior to the SEBOMD simulations. It was found that the PM3 Hamiltonian overestimated the C – N distances in almost all of the located transition state structures relative to the QM and the QM/MM counterparts. The activation free energy barriers



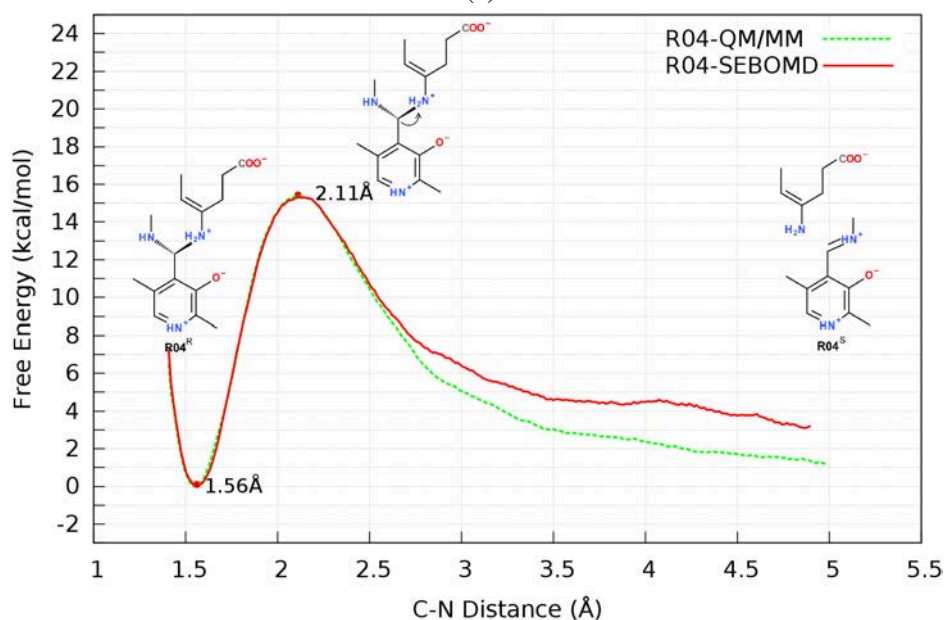
**Figure 3.57** : The PMF of C-N bond breaking reactions. (a)**R01**. (b)**R02**. (c)**R03**. (d)**R04**. The results obtained with SEBOMD simulations are depicted with red line, and the results obtained with QM/MM calculations are depicted with green dashed line. The complex between the products is depicted with a black arrow. Continued on next page.

were also overestimated almost in all cases with the PM3 Hamiltonian. Additionally, the PM3 Hamiltonian yielded endothermic reactions in both the QM (**R03**, and **R04**), and the QM/MM (**R02**,**R03**, and **R04**) calculations.

The SEBOMD capabilities in chemical transformations were investigated with the variation of the  $C - N$  distances and the relative free energies. It was observed



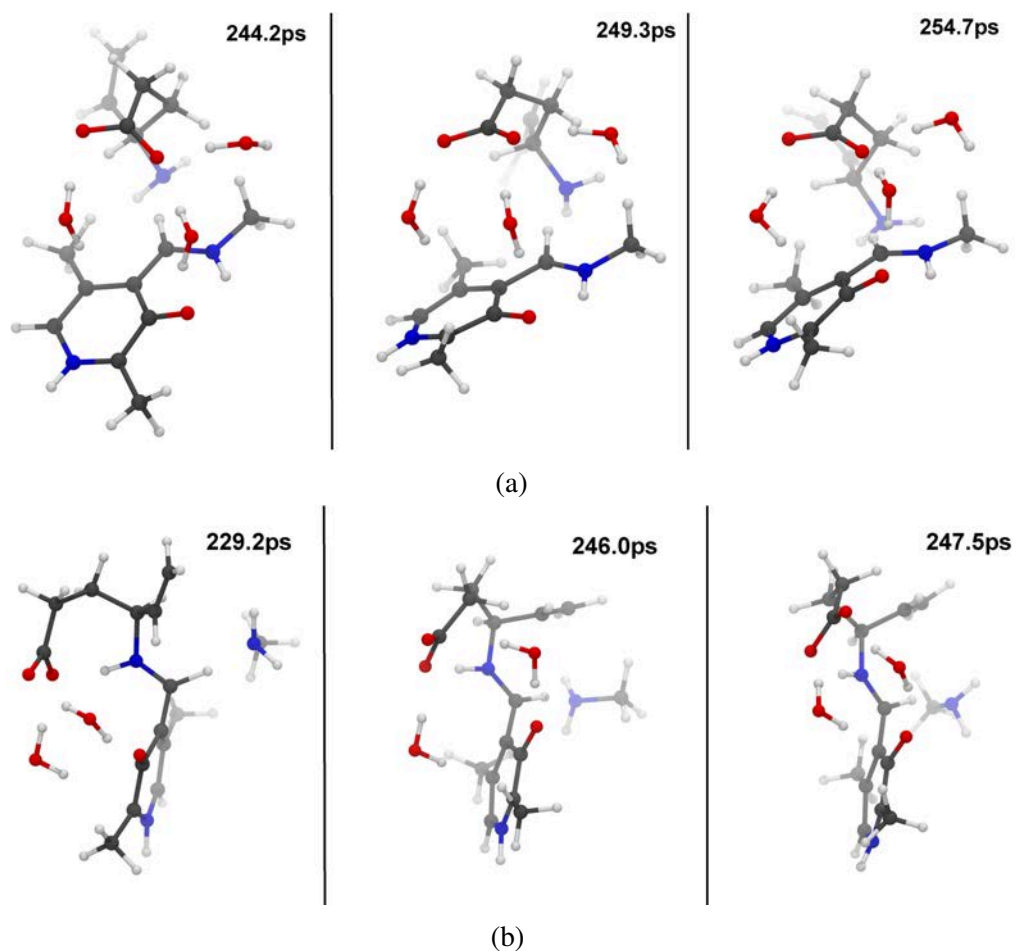
(c)



(d)

**Figure 3.57** : – continued from previous page. The PMF of C-N bond breaking reactions. (a)R01. (b)R02. (c)R03. (d)R04. The results obtained with SEBOMD simulations are depicted with red line, and the results obtained with QM/MM calculations are depicted with green dashed line. The complex between the products is depicted with a black arrow.

that the SEBOMD methodology was able to locate the transition state structures with the same  $C - N$  distances that were obtained in QM and QM/MM calculations. The activation free energy barriers were overestimated about 3-10 kcal/mol in the SEBOMD simulations relative to the QM calculations using PM3 Hamiltonian. When the QM/MM calculations were taken into account, it was observed that the SEBOMD overestimated the activation free energy barriers by less than 4 kcal/mol.



**Figure 3.58 :** Three dimensional representations of different conformers of vigabatrin, PLP, and methylamine complexes that are located with the SEBOMD simulations. **a.** Complex conformers that are formed between vigabatrin and PLP-methylamine imine in **R01**. **b.** Complex conformers that are formed between vigabatrin-PLP imine and methylamine in **R02**.





## 4. CONCLUSIONS

In this thesis, different approaches have been used to get a deeper understanding of the mechanistic insights of the reactivity and the dynamics of the PLP dependent enzyme, GABA-AT. Model structures have been used in quantum mechanical calculations to understand the mechanistic insights of different isomers and the effect of the fluorine in the inactivation of GABA-AT. Molecular dynamic simulations have been performed to reveal the roles of the active site residues. The effect of the active site residues on the reactivity has been studied with the use of quantum cluster models. Prior to the SEBOMD applications, the performance of the methodology has been improved with the implementation of new eigenvalue solvers. The first application of the SEBOMD approach on chemical transformations has been studied using these newly implemented routines. As a test case, C-N bond breaking reactions involving PLP complexes have been investigated.

The mechanistic details of transamination and HF elimination reactions, were studied computationally with different conformers of GABA and its fluorine-substituted analogues. Since the fluorine-substituted mechanism-based substrate analogues are widely used for the inactivation of PLP-dependent enzymes, our work provide some clues regarding the effect of the fluorine atom and the binding preference of GABA-AT with different isomers. The results indicated that the requirement of high concentrations of fluorine substituted analogues to inhibit GABA-AT and the poor efficiency of the (S)-3-F-GABA relative to (R)-3-F-GABA may stem from the instability of the bioactive conformers. It was observed that the poor performance of the (S)-3-F-GABA relative to (R)-3-F-GABA cannot be correlated to the energy profile of the transamination. However, the last step of the HF elimination path was found to be distinctive for the efficiency differences of the enantiomers.

In order to elucidate the ionization states and the roles of the active site residues in GABA-AT, molecular dynamic simulations of apo, holo and vigabatrin bound GABA-AT samples were studied. The presence of similar residues in many PLP

dependent enzymes enabled us to provide some general aspects of the active site residues of the same class of enzymes. The catalytic mechanism of PLP dependent enzymes is usually divided into a first and a second half-reaction. The activity of PLP dependent enzymes seems to be controlled by the amino acids present in the active site for both half reactions. Especially, the His190-Asp298 diad that is found in many PLP dependent enzymes generates an electric field which should greatly influence the electronic properties of PLP. Therefore, any QM or QM/MM study on the reactivity of any PLP dependent enzyme should at least include these two residues in the QM region.

The results obtained from the molecular dynamic simulations enabled us to provide some clues regarding the active site residues of the GABA-AT and their roles in general. In order to understand the effect of these active site residues on the reactions involving PLP, and to find an optimum model to mimic the active site, quantum cluster models were used. It was found that the phosphate group of PLP should be either represented as a methyl group or should be stabilized by neighbouring residues through hydrogen bonds in the modelling studies to avoid inaccurate energy values. The presence of a positively charged residue that can interact with the carboxylic substrates may reduce the activation energy barriers. The inclusion of the Asp-His diad, which exists in many PLP dependent enzymes, can enforce the Asp to stay negatively charged. The presence of this diad can remarkably improve the mimicry of the active site in the modelling studies of PLP dependent enzymes.

The increase of the computer power enabled the studies of large models using quantum mechanical descriptors. However, the numerical solution of the Schrödinger equation becomes increasingly expensive in terms of computational power when size increases. As of today, only linear scaling approximate quantum chemical methods are capable of combining both the necessary speed and accuracy to model large biomolecular systems with reasonable CPU cost. SEBOMD uses the linear scaling D&C algorithm which permits MD simulations of molecular systems containing several hundreds of atoms. However, the D&C implementation is not the most efficient to run on very large supercomputers due to the diagonalization process of the Fock matrix. In order to improve performance of the SEBOMD approach, new eigenvalue solvers from LAPACK & SCALAPACK have been implemented. Performance of the SEBOMD

simulations has been improved by implemented LAPACK solvers. The efficiency of the solvers strictly depended on the choice of the library and/or the compiler. SCALAPACK solvers could not provide higher efficiency than LAPACK solvers since the size of Fock matrix/matrices were not large enough. When a large number of cores was used (more than 64 cores) in SEBOMD simulations of large systems, using threads did lower the cost of the simulations. The automated selection of the diagonalizer has been implemented only for the serial version. Similar automation procedure can be implemented for the parallel computations by which the set of total number of cores, total number of threads, and solver can be scanned and optimized. As of today, a system that contains a small protein in water can be simulated with a simulation speed of 1.31 ps/day with the diag routine and 1.46 ps/day with the newly implemented dsyev solver on 128 cores. With the implementation of new eigenvalue solvers, the same system can be simulated on the same number of cores with a simulation speed of 1.59 ps/day with dsyevd. In order to overcome the bottleneck of the diagonalization, D&C algorithm should be replaced by another algorithm, such as the Density Matrix Minimization (DMM). DMM, in which the full density matrix is constructed, combines an electronic density minimization scheme with the use of sparse matrix-matrix multiplies to achieve linear scaling.

Thanks to the improved performance of the SEBOMD simulations, we present the first example of a chemical transformations studied through SEBOMD simulations. It provides general insights on the capabilities of the SEBOMD approach. Four different C-N bond breaking reactions that involve PLP complexes and 234 water molecules were studied using umbrella sampling techniques. It was observed that the SEBOMD simulations can be used to investigate chemical transformations. The approach is able to locate transition state structures. Additionally, it yields similar distances for the transition state center than what can be obtained with QM or QM/MM calculations. The activation free energy barriers are overestimated relative to the QM and the QM/MM calculations using the same Hamiltonian. This alteration in the potential energy surfaces may stem by the differences between implicit or explicit solvent modelling, and its description by means of QM or MM.



## REFERENCES

- [1] **Ondrechen, M.J., Briggs, J.M. and McCammon, J.A.** (2001). A Model for Enzyme-Substrate Interaction in Alanine Racemase, *Journal of the American Chemical Society*, 123(12), 2830–2834.
- [2] **Salvá, A., Donoso, J., Frau, J. and Muñoz, F.** (2002). Theoretical studies on Schiff base formation of vitamin b6 analogues, *Journal of Molecular Structure: {THEOCHEM}*, 577(2–3), 229 – 238.
- [3] **Casasnovas, R., Salvá, A., Frau, J., Donoso, J. and Muñoz, F.** (2009). Theoretical study on the distribution of atomic charges in the Schiff bases of 3-hydroxypyridine-4-aldehyde and alanine. The effect of the protonation state of the pyridine and imine nitrogen atoms, *Chemical Physics*, 355(2–3), 149 – 156.
- [4] **Salvá, A., Donoso, J., Frau, J. and Muñoz, F.** (2002). Theoretical studies on transimination of vitamin B6 analogs, *International Journal of Quantum Chemistry*, 89(1), 48–56.
- [5] **Griswold, W.R. and Toney, M.D.** (2011). Role of the Pyridine Nitrogen in Pyridoxal 5'-Phosphate Catalysis: Activity of Three Classes of PLP Enzymes Reconstituted with Deazapyridoxal 5'-Phosphate, *Journal of the American Chemical Society*, 133(37), 14823–14830.
- [6] **Jansonius, J.N.** (1998). Structure, evolution and action of vitamin B6-dependent enzymes, *Current Opinion in Structural Biology*, 8(6), 759 – 769.
- [7] **John, R.A.** (1995). Pyridoxal phosphate-dependent enzymes, *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, 1248(2), 81 – 96.
- [8] **Eliot, A.C. and Kirsch, J.F.** (2004). Pyridoxal Phosphate Enzymes: Mechanistic, Structural, and Evolutionary Considerations, *Annual Review of Biochemistry*, 73(1), 383–415.
- [9] **Banik, S.D. and Chandra, A.** (2014). A Hybrid QM/MM Simulation Study of Intramolecular Proton Transfer in the Pyridoxal 5'-Phosphate in the Active Site of Transaminase: Influence of Active Site Interaction on Proton Transfer, *The Journal of Physical Chemistry B*, 118(38), 11077–11089.
- [10] **Lin, Y.L., Gao, J., Rubinstein, A. and Major, D.T.** (2011). Molecular dynamics simulations of the intramolecular proton transfer and carbanion stabilization in the pyridoxal 5'-phosphate dependent enzymes L-dopa decarboxylase and alanine racemase, *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1814(11), 1438 – 1446.

- [11] **Sharif, S., Huot, M.C., Tolstoy, P.M., Toney, M.D., Jonsson, K.H.M. and Limbach, H.H.** (2007).  $^{15}\text{N}$  Nuclear Magnetic Resonance Studies of Acid-Base Properties of Pyridoxal-5'-Phosphate Aldimines in Aqueous Solution, *The Journal of Physical Chemistry B*, 111(15), 3869–3876.
- [12] **Sharif, S., Fogle, E., Toney, M.D., Denisov, G.S., Shenderovich, I.G., Buntkowsky, G., Tolstoy, P.M., Huot, M.C. and Limbach, H.H.** (2007). NMR Localization of Protons in Critical Enzyme Hydrogen Bonds, *Journal of the American Chemical Society*, 129(31), 9558–9559.
- [13] **Limbach, H.H., Chan-Huot, M., Sharif, S., Tolstoy, P.M., Shenderovich, I.G., Denisov, G.S. and Toney, M.D.** (2011). Critical hydrogen bonds and protonation states of pyridoxal 5'-phosphate revealed by NMR, *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1814(11), 1426 – 1437.
- [14] **Lee, H., Juncosa, J.I. and Silverman, R.B.** (2015). Ornithine Aminotransferase versus GABA Aminotransferase: Implications for the Design of New Anticancer Drugs, *Medicinal Research Reviews*, 35(2), 286–305.
- [15] **Shen, B.W., Hennig, M., Hohenester, E., Jansonius, J.N. and Schirmer, T.** (1998). Crystal structure of human recombinant ornithine aminotransferase, *Journal of Molecular Biology*, 277(1), 81 – 102.
- [16] **Cooper, A.J.,** (1985). Glutamate- $\gamma$ -aminobutyrate transaminase, Glutamate, Glutamine, Glutathione, and Related Compounds, volume 113 of *Methods in Enzymology*, Academic Press, pp.80 – 82.
- [17] **Churchich, J.E. and Moses, U.** (1981). 4-Aminobutyrate aminotransferase. The presence of nonequivalent binding sites., *Journal of Biological Chemistry*, 256(3), 1101–1104.
- [18] **Sparr, C., Salamanova, E., Schweizer, W.B., Senn, H.M. and Gilmour, R.** (2011). Theoretical and X-ray Crystallographic Evidence of a Fluorine-Imine Gauche Effect: An Addendum to Dunathan's Stereoelectronic Hypothesis, *Chemistry - A European Journal*, 17(32), 8850–8857.
- [19] **Crittenden, D.L., Chebib, M. and Jordan, M.J.T.** (2004). Stabilization of Zwitterions in Solution:  $\gamma$ -Aminobutyric Acid (GABA), *The Journal of Physical Chemistry A*, 108(1), 203–211.
- [20] **Vederas, J.C. and Floss, H.G.** (1980). Stereochemistry of pyridoxal phosphate catalyzed enzyme reactions, *Accounts of Chemical Research*, 13(12), 455–463.
- [21] **Hayashi, H., Wada, H., Yoshimura, T., Esaki, N. and Soda, K.** (1990). Recent Topics in Pyridoxal 5'-Phosphate Enzyme Studies, *Annual Review of Biochemistry*, 59(1), 87–110.
- [22] **Hayashi, H.** (1995). Pyridoxal Enzymes: Mechanistic Diversity and Uniformity, *The Journal of Biochemistry*, 118(3), 463–473.

- [23] **Bugg, T.** (2004). *Introduction to enzyme and coenzyme chemistry*, Wiley-Blackwell, 2 edition.
- [24] **Dunathan, H.C.** (1966). Conformation and reaction specificity in pyridoxal phosphate enzymes, *Proceedings of the National Academy of Sciences*, 55(4), 712–716.
- [25] **Rando, R.R. and Bangerter, F.W.** (1977). Reaction of the neurotoxin gabaculine with pyridoxal phosphate, *Journal of the American Chemical Society*, 99(15), 5141–5145.
- [26] **Silverman, R.B. and George, C.** (1988). Inactivation of  $\gamma$ -aminobutyric acid aminotransferase by (Z)-4-amino-2-fluorobut-2-enoic acid, *Biochemistry*, 27(9), 3285–3289.
- [27] **Fu, M. and Silverman, R.B.** (1999). Isolation and characterization of the product of inactivation of  $\gamma$ -aminobutyric acid aminotransferase by gabaculine, *Bioorganic & Medicinal Chemistry*, 7(8), 1581 – 1590.
- [28] **Clift, M.D., Ji, H., Deniau, G.P., O'Hagan, D., and Silverman, R.B.** (2007). Enantiomers of 4-Amino-3-fluorobutanoic Acid as Substrates for  $\gamma$ -Aminobutyric Acid Aminotransferase. Conformational Probes for GABA Binding, *Biochemistry*, 46(48), 13819–13828.
- [29] **Storici, P., Qiu, J., Schirmer, T. and Silverman, R.B.** (2004). Mechanistic Crystallography. Mechanism of Inactivation of  $\gamma$ -Aminobutyric Acid Aminotransferase by (1R,3S,4S)-3-Amino-4-fluorocyclopentane-1-carboxylic Acid As Elucidated by Crystallography, *Biochemistry*, 43(44), 14057–14063.
- [30] **Silverman, R.B. and Levy, M.A.** (1981). Mechanism of inactivation of  $\gamma$ -aminobutyric acid- $\alpha$ -ketoglutaric acid aminotransferase by 4-amino-5-halopentanoic acids, *Biochemistry*, 20(5), 1197–1203.
- [31] **Silverman, R.B. and Invergo, B.J.** (1986). Mechanism of inactivation of  $\gamma$ -aminobutyrate aminotransferase by 4-amino-5-fluoropentanoic acid. First example of an enamine mechanism for a  $\gamma$ -amino acid with a partition ratio of 0, *Biochemistry*, 25(22), 6817–6820.
- [32] **Wang, Z., Yuan, H., Nikolic, D., Breemen, R.B.V. and Silverman, R.B.** (2006). ( $\pm$ )-(1S,2R,5S)-5-Amino-2-fluorocyclohex-3-enecarboxylic Acid. A Potent GABA Aminotransferase Inactivator that Irreversibly Inhibits via an Elimination-Aromatization Pathway, *Biochemistry*, 45(48), 14513–14522.
- [33] **Gökcan, H. and Konuklar, F.A.S.** (2012). Theoretical Study on HF Elimination and Aromatization Mechanisms: A Case of Pyridoxal 5'-Phosphate-Dependent Enzyme, *The Journal of Organic Chemistry*, 77(13), 5533–5543.
- [34] **Durak, A.T., Gökcan, H. and Konuklar, F.A.S.** (2011). Theoretical studies on the inactivation mechanism of  $\gamma$ -aminobutyric acid aminotransferase, *Organic and Biomolecular Chemistry*, 9, 5162–5171.



- [35] **Alunni, S., Angelis, F.D., Ottavi, L., Papavasileiou, M., and Tarantelli, F.** (2005). Evidence of a Borderline Region between E1cb and E2 Elimination Reaction Mechanisms: A Combined Experimental and Theoretical Study of Systems Activated by the Pyridine Ring, *Journal of the American Chemical Society*, 127(43), 15151–15160.
- [36] **Choi, S., Storici, P., Schirmer, T. and Silverman, R.B.** (2002). Design of a Conformationally Restricted Analogue of the Antiepilepsy Drug Vigabatrin that Directs Its Mechanism of Inactivation of  $\gamma$ -Aminobutyric Acid Aminotransferase, *Journal of the American Chemical Society*, 124(8), 1620–1624.
- [37] **Fu, M. and Silverman, R.B.** (2004). Inactivation of  $\gamma$ -aminobutyric acid aminotransferase by (S)-4-amino-4,5-dihydro-2-furancarboxylic acid does not proceed by the expected aromatization mechanism, *Bioorganic & Medicinal Chemistry Letters*, 14(1), 203 – 206.
- [38] **Clift, M.D. and Silverman, R.B.** (2008). Synthesis and evaluation of novel aromatic substrates and competitive inhibitors of {GABA} aminotransferase, *Bioorganic & Medicinal Chemistry Letters*, 18(10), 3122 – 3125.
- [39] **Silverman, R.B., Durkee, S.C. and Invergo, B.J.** (1986). 4-Amino-2-(substituted methyl)-2-butenic acids: substrates and potent inhibitors of  $\gamma$ -aminobutyric acid aminotransferase, *Journal of Medicinal Chemistry*, 29(5), 764–770.
- [40] **Johnston, G.A.R., Curtis, D.R., Beart, P.M., Game, C.J.A., McCulloch, R.M. and Twitchin, B.** (1975). Cis- and trans-4-Aminocrotonic Acid as GABA Analogues of Restricted Conformation, *Journal of Neurochemistry*, 24(1), 157–160.
- [41] **Brehm, L., Hjeds, H. and Krosgaa, P.** (1972). Structure of Muscimol, a GABA Analog of Restricted Conformation, *Acta Chemica Scandinavica*, 26(3), 1298–1299.
- [42] **Bouclier, M., Jung, M.J. and Lippert, B.** (1979). Stereochemistry of Reactions Catalysed by Mammalian-Brain L-Glutamate 1-Carboxy-lyase and 4-Aminobutyrate: 2-Oxoglutarate Aminotransferase, *European Journal of Biochemistry*, 98(2), 363–368.
- [43] **Kwon, O., Park, J. and Churchich, J.** (1992). Brain 4-Aminobutyrate Aminotransferase - Isolation and Sequence of a cDNA-Encoding the Enzyme, *Journal of Biological Chemistry*, 267(11), 7215–7216.
- [44] **De Biase, D., Maras, B., Bossa, F., Barra, D. and John, R.A.** (1992). Protein structure of pig liver 4-aminobutyrate aminotransferase and comparison with a cDNA-deduced sequence, *European Journal of Biochemistry*, 208(2), 351–357.
- [45] **Storici, P., Capitani, G., Biase, D.D., Moser, M., John, R.A., Jansonius, J.N. and Schirmer, T.** (1999). Crystal Structure of GABA-Aminotransferase, a Target for Antiepileptic Drug Therapy, *Biochemistry*, 38(27), 8628–8634.

- [46] **Storici, P., De Biase, D., Bossa, F., Bruno, S., Mozzarelli, A., Peneff, C., Silverman, R.B. and Schirmer, T.** (2004). Structures of  $\gamma$ -Aminobutyric Acid (GABA) Aminotransferase, a Pyridoxal 5'-Phosphate, and [2Fe-2S] Cluster-containing Enzyme, Complexed with  $\gamma$ -Ethynyl-GABA and with the Antiepilepsy Drug Vigabatrin, *Journal of Biological Chemistry*, 279(1), 363–373.
- [47] **Mishra, V., Kumar, A., Ali, V., Nozaki, T., Zhang, K.Y.J. and Bhakuni, V.** (2011). Role of conserved active site tryptophan-101 in functional activity and stability of phosphoserine aminotransferase from an enteric human parasite, *Amino Acids*, 43(1), 483–491.
- [48] **Yano, T., Kuramitsu, S., Tanase, S., Morino, Y. and Kagamiyama, H.** (1992). Role of Asp222 in the catalytic mechanism of Escherichia coli aspartate aminotransferase: the amino acid residue which enhances the function of the enzyme-bound coenzyme pyridoxal 5'-phosphate, *Biochemistry*, 31(25), 5878–5887.
- [49] **Yano, T., Hinoue, Y., Chen, V.J., Metzler, D.E., Miyahara, I., Hirotsu, K. and Kagamiyama, H.** (1993). Role of an Active Site Residue Analyzed by Combination of Mutagenesis and Coenzyme Analog, *Journal of Molecular Biology*, 234(4), 1218 – 1229.
- [50] **Gong, J., Hunter, G.A. and Ferreira, G.C.** (1998). Aspartate-279 in Aminolevulinate Synthase Affects Enzyme Catalysis through Enhancing the Function of the Pyridoxal 5'-Phosphate Cofactor, *Biochemistry*, 37(10), 3509–3517.
- [51] **Mehta, P.K., Hale, T.I. and Christen, P.** (1993). Aminotransferases: demonstration of homology and division into evolutionary subgroups, *European Journal of Biochemistry*, 214(2), 549–561.
- [52] **Toney, M.D., Pascarella, S. and De Biase, D.** (1995). Active site model for  $\gamma$ -aminobutyrate aminotransferase explains substrate specificity and inhibitor reactivities, *Protein Science*, 4(11), 2366–2374.
- [53] **Qiu, J., Pingsterhaus, J.M. and Silverman, R.B.** (1999). Inhibition and Substrate Activity of Conformationally Rigid Vigabatrin Analogues with  $\gamma$ -Aminobutyric Acid Aminotransferase, *Journal of Medicinal Chemistry*, 42(22), 4725–4728.
- [54] **De Biase, D., Barra, D., Bossa, F., Pucci, P. and John, R.A.** (1991). Chemistry of the inactivation of 4-aminobutyrate aminotransferase by the antiepileptic drug vigabatrin., *Journal of Biological Chemistry*, 266(30), 20056–61.
- [55] **Nanavati, S.M. and Silverman, R.B.** (1991). Mechanisms of inactivation of  $\gamma$ -aminobutyric acid aminotransferase by the antiepilepsy drug  $\gamma$ -vinyl GABA (vigabatrin), *Journal of the American Chemical Society*, 113(24), 9341–9349.
- [56] **Metcalf, B.W.** (1979). Inhibitors of GABA metabolism, *Biochemical Pharmacology*, 28(11), 1705 – 1712.

- [57] **Lippert, B., Metcalf, B.W., Jung, M.J. and Casara, P.** (1977). 4-Amino-hex-5-enoic Acid, a Selective Catalytic Inhibitor of 4-Aminobutyric-Acid Aminotransferase in Mammalian Brain, *European Journal of Biochemistry*, 74(3), 441–445.
- [58] **Siegbahn, P.E. and Himo, F.** (2011). The quantum chemical cluster approach for modeling enzyme reactions, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(3), 323–336.
- [59] **Sevastik, R. and Himo, F.** (2007). Quantum chemical modeling of enzymatic reactions: The case of 4-oxalocrotonate tautomerase, *Bioorganic Chemistry*, 35(6), 444 – 457.
- [60] **Georgieva, P. and Himo, F.** (2010). Quantum chemical modeling of enzymatic reactions: The case of histone lysine methyltransferase, *Journal of Computational Chemistry*, 31(8), 1707–1714.
- [61] **Lovell, T., Himo, F., Han, W.G. and Noodleman, L.** (2003). Density functional methods applied to metalloenzymes, *Coordination Chemistry Reviews*, 238–239, 211 – 232, theoretical and Computational Chemistry.
- [62] **Georgieva, P. and Himo, F.** (2008). Density functional theory study of the reaction mechanism of the {DNA} repairing enzyme alkylguanine alkyltransferase, *Chemical Physics Letters*, 463(1–3), 214 – 218.
- [63] **Ertem, M.Z., Cramer, C.J., Himo, F. and Siegbahn, P.E.M.** (2012). N–O bond cleavage mechanism(s) in nitrous oxide reductase, *Journal of Biological Inorganic Chemistry*, 17(5), 687–698.
- [64] **Lind, M.E.S. and Himo, F.** (2014). Theoretical Study of Reaction Mechanism and Stereoselectivity of Arylmalonate Decarboxylase, *ACS Catalysis*, 4(11), 4153–4160.
- [65] **Cassimjee, K.E., Manta, B. and Himo, F.** (2015). A quantum chemical study of the  $\omega$ -transaminase reaction mechanism, *Organic and Biomolecular Chemistry*, 13, 8453–8464.
- [66] **Liao, R.Z. and Thiel, W.** (2013). On the Effect of Varying Constraints in the Quantum Mechanics Only Modeling of Enzymatic Reactions: The Case of Acetylene Hydratase, *The Journal of Physical Chemistry B*, 117(15), 3954–3961.
- [67] **Cui, F.C., Pan, X.L. and Liu, J.Y.** (2011). Reaction mechanism of isoflavone O-methyltransferase: A theoretical investigation, *Chemical Physics Letters*, 501(4–6), 502 – 507.
- [68] **Zhang, J., Sheng, X., Hou, Q. and Liu, Y.** (2014). Theoretical investigation on the dissociation of (R)-benzoin catalyzed by benzaldehyde lyase, *International Journal of Quantum Chemistry*, 114(6), 375–382.
- [69] **Sumner, S., Söderhjelm, P. and Ryde, U.** (2013). Effect of Geometry Optimizations on QM-Cluster and QM/MM Studies of Reaction Energies in Proteins, *Journal of Chemical Theory and Computation*, 9(9), 4205–4214.

- [70] **Liao, R.Z. and Thiel, W.** (2013). Convergence in the QM-only and QM/MM modeling of enzymatic reactions: A case study for acetylene hydratase, *Journal of Computational Chemistry*, 34(27), 2389–2397.
- [71] **Huang, W. and Gault, J.W.** (2012). Tautomerization in the UDP-Galactopyranose Mutase Mechanism: A DFT-Cluster and QM/MM Investigation, *The Journal of Physical Chemistry B*, 116(48), 14040–14050.
- [72] **Bruschi, M., Tiberti, M., Guerra, A. and Gioia, L.D.** (2014). Disclosure of Key Stereoelectronic Factors for Efficient H<sub>2</sub> Binding and Cleavage in the Active Site of [NiFe]-Hydrogenases, *Journal of the American Chemical Society*, 136(5), 1803–1814.
- [73] **Gleeson, D. and Gleeson, M.P.** (2015). Application of QM/MM and QM methods to investigate histone deacetylase 8, *Medicinal Chemical Communications*, 6, 477–485.
- [74] **Gökcan, H., Monard, G. and Sungur Konuklar, F.A.** (2016). Molecular dynamics simulations of apo, holo, and inactivator bound GABA-AT reveal the role of active site residues in PLP dependent enzymes, *Proteins: Structure, Function, and Bioinformatics*, 84(7), 875–891.
- [75] **Goedecker, S.** (1999). Linear scaling electronic structure methods, *Reviews of Modern Physics*, 71(4), 1085–1123.
- [76] **Goedecker, S. and Scuseria, G.** (2003). Linear scaling electronic structure methods in chemistry and physics, *Computing Science and Engineering*, 5, 14–21.
- [77] **Kussman, J., Beer, M. and Ochsenfeld, C.** (2013). Linear-scaling self-consistent field methods for large molecules, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 3(6), 614–636.
- [78] **Ochsenfeld, C., Kussmann, J. and Lambrecht, D.S.,** (2007). Linear-Scaling Methods in Quantum Chemistry, **K.B. Lipkowitz and T.R. Cundari,** editors, *Reviews in Computational Chemistry*, volume 23, chapter 1, John Wiley & Sons, Inc., Hoboken, NJ, USA, pp.1–82.
- [79] **Car, R. and Parrinello, M.** (1985). Unified Approach for Molecular Dynamics and Density-Functional Theory, *Physical Review Letters*, 55.
- [80] **Hutter, J. and Curioni, A.** (2005). Car-Parrinello molecular dynamics on massively parallel computers, *ChemPhysChem*, 6(9), 1788–1793.
- [81] **Elstner, M., Porezag, D., Jungnickel, G., Elsner, J., Haugk, M., Frauenheim, T., Suhai, S. and Seifert, G.** (1998). Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties, *Physical Review B*, 58(11), 7260–7268.
- [82] **Monard, G., Bernal-Uruchurtu, M.I., van der Vaart, A., Merz, K.M. and Ruiz-López, M.F.** (2005). Simulation of liquid water using semiempirical hamiltonians and the divide and conquer approach, *The Journal of Physical Chemistry A*, 109(15), 3425–3432.

- [83] **Ingrosso, F., Monard, G., Farag, M.H., Bastida, A. and Ruiz-López, M.F.** (2011). Importance of polarization and charge transfer effects to model the infrared spectra of peptides in solution, *Journal of Chemical Theory and Computation*, 7(6), 1840–1849.
- [84] **Dixon, S.L. and Merz, K.M.** (1996). Size Scaling, *The Journal of Chemical Physics*, 104(17), 6643–6649.
- [85] **Dixon, S.L. and Jr., K.M.M.** (1997). Fast, accurate semiempirical molecular orbital calculations for macromolecules, *The Journal of Chemical Physics*, 107(3), 879.
- [86] **Gökcan, H. and Konuklar, F.A.S.** (2014). Stereoelectronic explanations for the mechanistic details of transimination and {HF} elimination reactions, *Journal of Molecular Graphics and Modelling*, 51, 173 – 183.
- [87] **Zhang, J. and Liu, Y.** (2015). A QM/MM study of the catalytic mechanism of succinic semialdehyde dehydrogenase from *Synechococcus* sp. PCC 7002 and *Salmonella typhimurium*, *RSC Advances*, 5, 101672–101682.
- [88] **lin Lin, Y. and Gao, J.** (2011). Kinetic Isotope Effects of L-Dopa Decarboxylase, *Journal of the American Chemical Society*, 133(12), 4398–4403.
- [89] **Moya-García, A.A., Ruiz-Pernía, J., Martí, S., Sánchez-Jiménez, F. and Tuñón, I.** (2008). Analysis of the Decarboxylation Step in Mammalian Histidine Decarboxylase: A Computation Study, *Journal of Biological Chemistry*, 283(18), 12393–12401.
- [90] **Shoji, M., Hanaoka, K., Ujiie, Y., Tanaka, W., Kondo, D., Umeda, H., Kamoshida, Y., Kayanuma, M., Kamiya, K., Shiraishi, K., Machida, Y., Murakawa, T. and Hayashi, H.** (2014). A QM/MM Study of the l-Threonine Formation Reaction of Threonine Synthase: Implications into the Mechanism of the Reaction Specificity, *Journal of the American Chemical Society*, 136(12), 4525–4533.
- [91] **Cao, J., Bjornsson, R., Bühl, M., Thiel, W. and van Mourik, T.** (2012). Modelling Zwitterions in Solution: 3-Fluoro- $\gamma$ -aminobutyric Acid (3F-GABA), *Chemistry - A European Journal*, 18(1), 184–195.
- [92] **Schrödinger, E.** (1926). An undulatory theory of the mechanics of atoms and molecules, *Physical Review*, 28(6), 1049–1070.
- [93] **Born, M. and Oppenheimer, R.** (1927). Zur Quantentheorie der Molekeln, *Annalen der Physik*, 84(20), 457–484.
- [94] **Pauli, W.** (1925). Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren, *Zeitschrift für Physik*, 31(1), 765–783.
- [95] **Slater, J.C.** (1929). The Theory of Complex Spectra, *Physical Review*, 34, 1293–1322.

- [96] **Roothaan, C.C.J.** (1951). New developments in molecular orbital theory, *Reviews of Modern Physics*, 23(2), 69–89.
- [97] **Hall, G.G.** (1951). The Molecular Orbital Theory of Chemical Valency. VIII. A Method of Calculating Ionization Potentials, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 205(1083), 541–552.
- [98] **Lewars, E.G.**, (2011). Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics, chapter Semiempirical Calculations, Springer Netherlands, Dordrecht, 2 edition, pp.391–444.
- [99] **Thiel, W.**, (2005). Chapter 21 - Semiempirical quantum-chemical methods in computational chemistry, **C.E. Dykstra**, editor, Theory and Applications of Computational Chemistry, Elsevier, Amsterdam, pp.559 – 580.
- [100] **Hückel, E.** (1931). Quantentheoretische Beiträge zum Benzolproblem, *Zeitschrift für Physik*, 70(3), 204–286.
- [101] **Hoffmann, R.** (1963). An Extended Hückel Theory. I. Hydrocarbons, *The Journal of Chemical Physics*, 39(6), 1397.
- [102] **Hoffmann, R.** (1964). Extended Hückel Theory. II.  $\sigma$  Orbitals in the Azines, *The Journal of Chemical Physics*, 40(9), 2745.
- [103] **Pariser, R. and Parr, R.G.** (1953). A Semi-Empirical Theory of the Electronic Spectra and Electronic Structure of Complex Unsaturated Molecules. I., *The Journal of Chemical Physics*, 21(3), 466.
- [104] **Pople, J.A.** (1953). Electron interaction in unsaturated hydrocarbons, *Transactions of the Faraday Society*, 49, 1375–1385.
- [105] **Pople, J., Santry, D. and Segal, G.** (1965). Approximate Self-Consistent Molecular Orbital Theory. I. Invariant Procedures, *The Journal of Chemical Physics*, 43(10), S129–S135.
- [106] **Pople, J.A. and Segal, G.A.** (1965). Approximate Self-Consistent Molecular Orbital Theory. II. Calculations with Complete Neglect of Differential Overlap, *The Journal of Chemical Physics*, 43(10), S136–S151.
- [107] **Pople, J.A., Beveridge, D.L. and Dobosh, P.A.** (1967). Approximate Self Consistent Molecular Orbital Theory. V. Intermediate Neglect of Differential Overlap, *The Journal of Chemical Physics*, 47(6), 2026–2033.
- [108] **Dewar, M.J.S. and Thiel, W.** (1977). Ground states of molecules. 38. The MNDO method. Approximations and parameters, *Journal of the American Chemical Society*, 99(15), 4899–4907.
- [109] **Dewar, M.J.S. and Thiel, W.** (1977). Ground states of molecules. 39. MNDO results for molecules containing hydrogen, carbon, nitrogen, and oxygen, *Journal of the American Chemical Society*, 99(15), 4907–4917.

- [110] **Dewar, M.J.S. and McKee, M.L.** (1977). Ground states of molecules. 41. MNDO results for molecules containing boron, *Journal of the American Chemical Society*, 99(16), 5231–5241.
- [111] **Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. and Stewart, J.J.P.** (1985). Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model, *Journal of the American Chemical Society*, 107(13), 3902–3909.
- [112] **Dannenberg, J.J. and Evleth, E.M.** (1992). A critical examination of H-bonding interactions calculated using the AM1 molecular orbital method, *International Journal of Quantum Chemistry*, 44(5), 869–885.
- [113] **Stewart, J.J.P.** (1989). Optimization of parameters for semiempirical methods I. Method, *Journal of Computational Chemistry*, 10(2), 209–220.
- [114] **Stewart, J.J.P.** (1989). Optimization of parameters for semiempirical methods II. Applications, *Journal of Computational Chemistry*, 10(2), 221–264.
- [115] **Stewart, J.J.P.** (1991). Optimization of parameters for semiempirical methods. III Extension of PM3 to Be, Mg, Zn, Ga, Ge, As, Se, Cd, In, Sn, Sb, Te, Hg, Tl, Pb, and Bi, *Journal of Computational Chemistry*, 12(3), 320–341.
- [116] **Dewar, M.J., Jie, C. and Yu, J.** (1993). SAM1; The first of a new series of general purpose quantum mechanical molecular models, *Tetrahedron*, 49(23), 5003 – 5038.
- [117] **Voityuk, A.A. and Rösch, N.** (2000). AM1/d Parameters for Molybdenum, *The Journal of Physical Chemistry A*, 104(17), 4089–4094.
- [118] **Imhof, P., Noe, F., Fischer, S. and Smith, J.C.** (2006). AM1/d parameters for magnesium in metalloenzymes, *Journal of Chemical Theory and Computation*, 2(4), 1050–1056.
- [119] **Nam, K., Cui, Q., Gao, J. and York, D.M.** (2007). Specific Reaction Parametrization of the AM1/d Hamiltonian for Phosphoryl Transfer Reactions: H, O, and P Atoms, *Journal of Chemical Theory and Computation*, 3(2), 486–504.
- [120] **Repasky, M.P., Chandrasekhar, J. and Jorgensen, W.L.** (2002). PDDG/PM3 and PDDG/MNDO: Improved semiempirical methods, *Journal of Computational Chemistry*, 23(16), 1601–1622.
- [121] **Winget, P., Horn, A.H.C., Selçuki, C., Martin, B. and Clark, T.** (2003). AM1\* parameters for phosphorus, sulfur and chlorine., *Journal of Molecular Modeling*, 9(6), 408–14.
- [122] **Rocha, G.B., Freire, R.O., Simas, A.M. and Stewart, J.J.P.** (2006). RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I, *Journal of Computational Chemistry*, 27(10), 1101–1111.
- [123] **Stewart, J.J.P.** (2007). Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements, *Journal of Molecular Modeling*, 13(12), 1173–1213.



- [124] **Bernal-Uruchurtu, M., Martins-Costa, M., Millot, C. and Ruiz-López, M.F.** (2000). Improving Description of Hydrogen Bonds at the Semiempirical Level : Water - Water Interactions as Test Case, *Journal of Computational Chemistry*, 21(7), 572–581.
- [125] **Harb, W., Bernal-Uruchurtu, I.M. and Ruiz-López, F.M.** (2004). An improved semiempirical method for hydrated systems, *Theoretical Chemistry Accounts*, 112(4), 204–216.
- [126] **Arillo-Flores, O.I., Ruiz-López, M.F. and Bernal-Uruchurtu, M.I.** (2007). Can semi-empirical models describe HCl dissociation in water?, *Theoretical Chemistry Accounts*, 118(2), 425–435.
- [127] **Thiriot, E. and Monard, G.** (2009). Combining a genetic algorithm with a linear scaling semiempirical method for protein–ligand docking, *Journal of Molecular Structure: {THEOCHEM}*, 898(1–3), 31 – 41, theoretical treatment of large molecular systems.
- [128] **Marion, A., Monard, G., Ruiz-López, M.F. and Ingrosso, F.** (2014). Water interactions with hydrophobic groups: Assessment and recalibration of semiempirical molecular orbital methods, *The Journal of Chemical Physics*, 141(3).
- [129] **Hohenberg, P. and Kohn, W.** (1964). Inhomogeneous Electron Gas, *Physical Review*, 136, 864–871.
- [130] **Kohn, W. and Sham, L.J.** (1965). Self-Consistent Equations Including Exchange and Correlation Effects, *Physical Review*, 140, 1133–1138.
- [131] **Becke, A.D.** (1988). Density-functional exchange-energy approximation with correct asymptotic behavior, *Physical Review A*, 38, 3098–3100.
- [132] **Lee, C., Yang, W. and Parr, R.G.** (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, *Physical Review B*, 37, 785–789.
- [133] **Miehlich, B., Savin, A., Stoll, H. and Preuss, H.** (1989). Results obtained with the correlation energy density functionals of becke and Lee, Yang and Parr, *Chemical Physics Letters*, 157(3), 200 – 206.
- [134] **Becke, A.D.** (1993). Density-functional thermochemistry. III. The role of exact exchange, *The Journal of Chemical Physics*, 98(7), 5648.
- [135] **Perdew, J.P., Ziesche, P. and Eschrig, H.** (1991). *Electronic structure of solids'* 91, volume 11, Akademie Verlag, Berlin.
- [136] **Perdew, J.P., Chevary, J.A., Vosko, S.H., Jackson, K.A., Pederson, M.R., Singh, D.J. and Fiolhais, C.** (1992). Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation, *Physical Review B*, 46, 6671–6687.

- [137] **Perdew, J.P., Chevary, J.A., Vosko, S.H., Jackson, K.A., Pederson, M.R., Singh, D.J. and Fiolhais, C.** (1993). Erratum: Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation, *Physical Review B*, 48, 4978–4978.
- [138] **Perdew, J.P., Burke, K. and Wang, Y.** (1996). Generalized gradient approximation for the exchange-correlation hole of a many-electron system, *Physical Review B*, 54, 16533–16539.
- [139] **Vosko, S.H., Wilk, L. and Nusair, M.** (1980). Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis, *Canadian Journal of Physics*, 58(8), 1200–1211.
- [140] **Stephens, P.J., Devlin, F.J., Chabalowski, C.F. and Frisch, M.J.** (1994). Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields, *The Journal of Physical Chemistry*, 98(45), 11623–11627.
- [141] **Lewars, E.G.,** (2011). Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics, chapter Some "Special" Topics: Solvation, Singlet Diradicals, A Note on Heavy Atoms and Transition Metals, Springer Netherlands, Dordrecht, 2 edition, pp.521–560.
- [142] **Tomasi, J., Mennucci, B., and Cammi, R.** (2005). Quantum Mechanical Continuum Solvation Models, *Chemical Reviews*, 105(8), 2999–3094.
- [143] **Bondi, A.** (1964). van der Waals Volumes and Radii, *The Journal of Physical Chemistry*, 68(3), 441–451.
- [144] **Rowland, R.S., and Taylor, R.** (1996). Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van der Waals Radii, *The Journal of Physical Chemistry*, 100(18), 7384–7391.
- [145] **Pauling, L.** (1960). *The Nature of the Chemical Bond, An Introduction to Modern Structural Chemistry*, George Fisher Baker Non-Resident Lecture Series, Cornell University Press, Ithaca, United States.
- [146] **Rappe, A.K., Casewit, C.J., Colwell, K.S., III, W.A.G. and Skiff, W.M.** (1992). UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations, *Journal of the American Chemical Society*, 114(25), 10024–10035.
- [147] **Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R. and Pople, J.A.,** Gaussian 03, Revision D.01, Gaussian, Inc., Wallingford, CT, 2004.
- [148] **Cammi, R. and Tomasi, J.** (1995). Remarks on the use of the apparent surface charges (ASC) methods in solvation problems: Iterative versus matrix-inversion procedures and the renormalization of the apparent charges, *Journal of Computational Chemistry*, 16(12), 1449–1458.

- [149] **Miertuš, S., Scrocco, E. and Tomasi, J.** (1981). Electrostatic interaction of a solute with a continuum. A direct utilization of {AB} initial molecular potentials for the prediction of solvent effects, *Chemical Physics*, 55(1), 117 – 129.
- [150] **Luque, F.J., Bofill, J.M. and Orozco, M.** (1995). New strategies to incorporate the solvent polarization in self consistent reaction field and free energy perturbation simulations, *The Journal of Chemical Physics*, 103(23), 10183–10191.
- [151] **Colominas, C., Luque, F.J. and Orozco, M.** (1999). Monte Carlo–MST: New strategy for representation of solvent configurational space in solution, *Journal of Computational Chemistry*, 20(7), 665–678.
- [152] **Javier Luque, F., Curutchet, C., Munoz-Muriedas, J., Bidon-Chanal, A., Soteras, I., Morreale, A., Gelpi, J.L. and Orozco, M.** (2003). Continuum solvation models: Dissecting the free energy of solvation, *Physical Chemistry Chemical Physics*, 5, 3827–3836.
- [153] **Morreale, A., Gelpi, J.L., Luque, F.J. and Orozco, M.** (2003). Continuum and discrete calculation of fractional contributions to solvation free energy, *Journal of Computational Chemistry*, 24(13), 1610–1623.
- [154] **Curutchet, C., Cramer, C.J., Truhlar, D.G., Ruiz-López, M.F., Rinaldi, D., Orozco, M. and Luque, F.J.** (2003). Electrostatic component of solvation: Comparison of SCRF continuum models, *Journal of Computational Chemistry*, 24(3), 284–297.
- [155] **Basilevsky, M. and Chudinov, G.** (1991). Dynamics of charge transfer chemical reactions in a polar medium within the scope of the Born-Kirkwood-Onsager model, *Chemical Physics*, 157(3), 327 – 344.
- [156] **Orozco, M. and Luque, F.J.** (2000). Theoretical Methods for the Description of the Solvent Effect in Biomolecular Systems, *Chemical Reviews*, 100(11), 4187–4226.
- [157] **Cancès, E., Mennucci, B. and Tomasi, J.** (1997). A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics, *The Journal of Chemical Physics*, 107(8), 3032–3041.
- [158] **Mennucci, B., Cancès, E. and Tomasi, J.** (1997). Evaluation of Solvent Effects in Isotropic and Anisotropic Dielectrics and in Ionic Solutions with a Unified Integral Equation Method: Theoretical Bases, Computational Implementation, and Numerical Applications, *The Journal of Physical Chemistry B*, 101(49), 10506–10517.
- [159] **Cancès, E. and Mennucci, B.** (1998). New applications of integral equations methods for solvation continuum models: ionic solutions and liquid crystals, *Journal of Mathematical Chemistry*, 23(3), 309–326.

- [160] **Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R. and Fox, J.D.J.**, Gaussian 09 Revision A.1, gaussian Inc. Wallingford CT 2009.
- [161] **Li, X. and Frisch, M.J.** (2006). Energy-Represented Direct Inversion in the Iterative Subspace within a Hybrid Geometry Optimization Method, *Journal of Chemical Theory and Computation*, 2(3), 835–839.
- [162] **Schlegel, H.B.**, (1995). Geometry Optimization on Potential Energy Surfaces, **D.R. LYarkony**, editor, *Modern Electronic Structure Theory*, World Scientific, Singapore, pp.459 – 500.
- [163] **Schlegel, H.B.**, (1998). Geometry Optimization, **P.v.R. Schleyer, N.L. Allinger, P.A. Kollman, T. Clark, H.F. Schaefer, J. III Gasteiger and P.R. Schreiner**, editors, *Encyclopedia of Computational Chemistry*, volume 2, John Wiley & Sons, Inc., Chichester, U. K., pp.1136–1142.
- [164] **Banerjee, A., Adams, N., Simons, J. and Shepard, R.** (1985). Search for stationary points on surfaces, *The Journal of Physical Chemistry*, 89(1), 52–57.
- [165] **Simons, J. and Nichols, J.** (1990). Strategies for walking on potential energy surfaces using local quadratic approximations, *International Journal of Quantum Chemistry*, 38(S24), 263–276, 10.1002/qua.560382427.
- [166] **Hratchian, H.P. and Schlegel, H.B.** (2004). Accurate reaction paths using a Hessian based predictor–corrector integrator, *The Journal of Chemical Physics*, 120(21), 9918–9924.
- [167] **Hratchian, H.P. and Schlegel, H.B.**, (2005). Finding minima, transition states, and following reaction pathways on ab initio potential energy surfaces, **C.E. Dykstra, G. Frenking, K.S. Kim and G.E. Scuseria**, editors, *Theory and Applications of Computational Chemistry: The First 40 Years*, Elsevier, Amsterdam, p.195–249.
- [168] **Hratchian, H.P., and Schlegel, H.B.** (2005). Using Hessian Updating To Increase the Efficiency of a Hessian Based Predictor-Corrector Reaction Path Following Method, *Journal of Chemical Theory and Computation*, 1(1), 61–69.
- [169] **Hratchian, H.P., and Schlegel\*, H.B.** (2002). Following Reaction Pathways Using a Damped Classical Trajectory Algorithm, *The Journal of Physical Chemistry A*, 106(1), 165–169.
- [170] **Page, M. and McIver, J.W.** (1988). On evaluating the reaction path Hamiltonian, *The Journal of Chemical Physics*, 88(2), 922–935.
- [171] **Page, M., Doubleday, C. and McIver, J.W.** (1990). Following steepest descent reaction paths. The use of higher energy derivatives with abinitio electronic structure methods, *The Journal of Chemical Physics*, 93(8), 5634–5642.

- [172] **Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M.** (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations, *Journal of Computational Chemistry*, 4(2), 187–217.
- [173] **Jorgensen, W.L. and Tirado-Rives, J.** (1988). The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin, *Journal of the American Chemical Society*, 110(6), 1657–1666.
- [174] **Duan, Y., Wu, C., Chowdhury, S., Lee, M.C., Xiong, G., Zhang, W. and ... Kollman, P.** (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, *Journal of Computational Chemistry*, 24(16), 1999–2012.
- [175] **Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A. and Case, D.A.** (2004). Development and testing of a general amber force field, *Journal of Computational Chemistry*, 25(9), 1157–1174.
- [176] **Schmid, N., Eichenberger, A.P., Choutko, A., Riniker, S., Winger, M., Mark, A.E. and Gunsteren, W.F.** (2011). Definition and testing of the GROMOS force-field versions 54A7 and 54B7, *European Biophysics Journal*, 40(7), 843–856.
- [177] **Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M. and Kollman, .P.A.** (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules, *Journal of the American Chemical Society*, 117(19), 5179–5197.
- [178] **Verlet, L.** (1967). Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules, *Physical Review*, 159, 98–103.
- [179] **Hockney, R.W.** (1970). The Potential Calculation and Some Applications, *Methods in Computational Physics*, 9, 136–211.
- [180] **Swope, W.C., Andersen, H.C., Berens, P.H. and Wilson, K.R.** (1982). A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters, *The Journal of Chemical Physics*, 76(1), 637–649.
- [181] **Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J.** (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *Journal of Computational Physics*, 23(3), 327 – 341.
- [182] **Miyamoto, S. and Kollman, P.A.** (1992). Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models, *Journal of Computational Chemistry*, 13(8), 952–962.
- [183] **Ewald, P.P.** (1921). Die Berechnung optischer und elektrostatischer Gitterpotentiale, *Annalen der Physik*, 369(3), 253–287.

- [184] **Hockney, R.W. and Eastwood, J.W.** (1981). *Computer Simulation Using Particles*, McGraw-Hill, New York.
- [185] **Eastwood, J., Hockney, R. and Lawrence, D.** (1980). P3M3DP—The three-dimensional periodic particle-particle/ particle-mesh program, *Computer Physics Communications*, 19(2), 215 – 261.
- [186] **Darden, T., York, D. and Pedersen, L.** (1993). Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems, *The Journal of Chemical Physics*, 98(12), 10089–10092.
- [187] **Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H. and Pedersen, L.G.** (1995). A smooth particle mesh Ewald method, *The Journal of Chemical Physics*, 103(19), 8577–8593.
- [188] **Andersen, H.C.** (1980). Molecular dynamics simulations at constant pressure and/or temperature, *The Journal of Chemical Physics*, 72, 2384.
- [189] **Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R.** (1984). Molecular dynamics with coupling to an external bath, *The Journal of Chemical Physics*, 81(8), 3684–3690.
- [190] **Dixon, S.L., van der Vaart, A., Gogonea, V., Vincent, J.J., Brothers, E.N., Westerhoff, L.M. and Merz, J.K.M.**, (1999), DivCon99, The Pennsylvania State University.
- [191] **Case, D.A., Berryman, J.T., Betz, R.M., Cerutti, D.S., III, T.E.C., Darden, T.A. and Kollman, J.P.A.**, (2015), AMBER 2015, University of California, San Francisco.
- [192] **Yang, W.** (1991). Direct calculation of electron density in density-functional theory, *Physical Review Letters*, 66, 1438–1441.
- [193] **Cossi, M., Barone, V., Cammi, R. and Tomasi, J.** (1996). Ab initio study of solvated molecules: a new implementation of the polarizable continuum model, *Chemical Physics Letters*, 255(4–6), 327 – 335.
- [194] **Barone, V., Cossi, M. and Tomasi, J.** (1998). Geometry optimization of molecular structures in solution by the polarizable continuum model, *Journal of Computational Chemistry*, 19(4), 404–417.
- [195] **Reed, A.E., Curtiss, L.A. and Weinhold, F.** (1988). Intermolecular interactions from a natural bond orbital, donor-acceptor viewpoint, *Chemical Reviews*, 88(6), 899–926.
- [196] **Yamamoto, I., Deniau, G.P., Gavande, N., Chebib, M., Johnston, G.A.R. and O'Hagan, D.** (2011). Agonist responses of (R)- and (S)-3-fluoro- $\gamma$ -aminobutyric acids suggest an enantiomeric fold for GABA binding to GABAC receptors, *Chemical Communications*, 47, 7956–7958.
- [197] **Liao, R.Z., Ding, W.J., Yu, J.G., Fang, W.H. and Liu, R.Z.** (2007). Water-Assisted Transamination of Glycine and Formaldehyde, *The Journal of Physical Chemistry A*, 111(16), 3184–3190.

- [198] **Bach, R.D., Canepa, C. and Glukhovtsev, M.N.** (1999). Influence of Electrostatic Effects on Activation Barriers in Enzymatic Reactions: Pyridoxal 5'-Phosphate-Dependent Decarboxylation of  $\alpha$ -Amino Acids, *Journal of the American Chemical Society*, 121(28), 6542–6555.
- [199] **Cerqueira, N.M.F.S.A., Fernandes, P.A. and Ramos, M.J.** (2011). Computational Mechanistic Studies Addressed to the Transamination Reaction Present in All Pyridoxal 5'-Phosphate-Requiring Enzymes, *Journal of Chemical Theory and Computation*, 7(5), 1356–1368.
- [200] **Ortega-Castro, J., Adrover, M., Frau, J., Salvá, A., Donoso, J. and Muñoz, F.** (2010). DFT Studies on Schiff Base Formation of Vitamin B6 Analogues. Reaction between a Pyridoxamine-Analogue and Carbonyl Compounds, *The Journal of Physical Chemistry A*, 114, 4634–4640.
- [201] **Liao, R.Z., Ding, W.J., Yu, J.G., Fang, W.H. and Liu, R.Z.** (2008). Theoretical studies on pyridoxal 5'-phosphate-dependent transamination of  $\alpha$ -amino acids, *Journal of Computational Chemistry*, 29(12), 1919–1929.
- [202] **Das, G.** (2013). Zwitterionic conformers of pyrrolysine and their interactions with metal ions—a theoretical study, *Journal of Molecular Modeling*, 19(8), 2981–2991.
- [203] **Huang, Z. and Lin, Z.** (2005). Detailed Ab Initio Studies of the Conformers and Conformational Distributions of Gaseous Tryptophan, *The Journal of Physical Chemistry A*, 109(11), 2656–2659.
- [204] **qin Chen, Z., hua Zhang, C. and Xue, Y.** (2009). Theoretical Studies on the Thermodynamics and Kinetics of the N-Glycosidic Bond Cleavage in Deoxythymidine Glycol, *The Journal of Physical Chemistry B*, 113(30), 10409–10420.
- [205] **Frisch, M.J., Head-Gordon, M. and Pople, J.A.** (1990). A direct MP2 gradient method, *Chemical Physics Letters*, 166(3), 275 – 280.
- [206] **Legault, C.Y.**, (2009), CYLview, 1.0b, Université de Sherbrooke.
- [207] **Goodman, L., Gu, H., and Pophristic, V.** (2005). Gauche Effect in 1,2-Difluoroethane. Hyperconjugation, Bent Bonds, Steric Repulsion, *The Journal of Physical Chemistry A*, 109(6), 1223–1229.
- [208] **Deniau, G., Slawin, A.M.Z., Lebl, T., Chorki, F., Issberner, J.P., van Mourik, T., Heygate, J.M., Lambert, J.J., Etherington, L.A., Sillar, K.T. and O'Hagan, D.** (2007). Synthesis, Conformation and Biological Evaluation of the Enantiomers of 3-Fluoro- $\gamma$ -Aminobutyric Acid ((R)- and (S)-3F-GABA): An Analogue of the Neurotransmitter GABA, *ChemBioChem*, 8(18), 2265–2274.
- [209] **Humelnicu, I., Wurthwein, E.U. and Haufe, G.** (2012). The conformers of 3-fluoroalanine. A theoretical study, *Organic and Biomolecular Chemistry*, 10, 2084–2093.

- [210] **Briggs, C.R.S., Allen, M.J., O'Hagan, D., Tozer, D.J., Slawin, A.M.Z., Goeta, A.E. and Howard, J.A.K.** (2004). The observation of a large gauche preference when 2-fluoroethylamine and 2-fluoroethanol become protonated, *Organic and Biomolecular Chemistry*, 2, 732–740.
- [211] **Buissonneaud, D.Y., van Mourik, T. and O'Hagan, D.** (2010). A DFT study on the origin of the fluorine gauche effect in substituted fluoroethanes, *Tetrahedron*, 66(12), 2196 – 2202.
- [212] **Hyla-Kryspin, I., Grimme, S., Hruschka, S. and Haufe, G.** (2008). Conformational preferences and basicities of monofluorinated cyclopropyl amines in comparison to cyclopropylamine and 2-fluoroethylamine, *Organic and Biomolecular Chemistry*, 6, 4167–4175.
- [213] **Snyder, J.P., Chandrakumar, N.S., Sato, H., and Lankin, D.C.** (2000). The Unexpected Diaxial Orientation of cis-3,5-Difluoropiperidine in Water: A Potent CF - - - NH Charge-Dipole Effect, *Journal of the American Chemical Society*, 122(3), 544–545.
- [214] **Hunter, L.** (2010). The C–F bond as a conformational tool in organic and biological chemistry, *Beilstein Journal of Organic Chemistry*, 6, 38.
- [215] **Tanzer, E.M., Zimmer, L.E., Schweizer, W.B. and Gilmour, R.** (2012). Fluorinated Organocatalysts for the Enantioselective Epoxidation of Enals: Molecular Preorganisation by the Fluorine-Iminium Ion Gauche Effect, *Chemistry - A European Journal*, 18(36), 11334–11342.
- [216] **Gooseman, N., O'Hagan, D., Peach, M., Slawin, A., Tozer, D. and Young, R.** (2007). An Electrostatic Gauche Effect in  $\beta$ -Fluoro- and  $\beta$ -Hydroxy-N-ethylpyridinium Cations, *Angewandte Chemie International Edition*, 46(31), 5904–5908.
- [217] **Alunni, S., Laureti, V., Ottavi, L., and Ruzziconi, R.** (2003). Catalysis of the  $\beta$ -Elimination of HF from Isomeric 2-Fluoroethylpyridines and 1-Methyl-2-fluoroethylpyridinium Salts. Proton-Activating Factors and Methyl-Activating Factors as a Mechanistic Test To Distinguish between Concerted E2 and E1cb Irreversible Mechanisms, *The Journal of Organic Chemistry*, 68(3), 718–725.
- [218] **Angelis, F.D., Tarantelli, F. and Alunni, S.** (2006). Ab Initio Molecular Dynamics Simulations of Elimination Reactions in Water Solution: Exploring the Borderline Region between the E1cb and E2 Reaction Mechanisms, *The Journal of Physical Chemistry B*, 110(22), 11014–11019.
- [219] **Martell, A.E.**, (1984). Chemical and Biological Aspects of Vitamin B-6 Catalysis. Part A, **A.E. Evangelopoulos**, editor, Chemical and Biological Aspects of Vitamin B-6 Catalysis. Part A, Alan E. Liss Inc, New York, p. 63.
- [220] **Case, D., Darden, T., Cheatham, T., III, Simmerling, C., Wang, J., Duke, R. and Kollman, .P.**, (2012), AMBER 2012, University of California, San Francisco.



- [221] **Søndergaard, C.R., Olsson, M.H.M., Rostkowski, M. and Jensen, J.H.** (2011). Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values, *Journal of Chemical Theory and Computation*, 7(7), 2284–2295.
- [222] **Olsson, M.H.M., Søndergaard, C.R., Rostkowski, M. and Jensen, J.H.** (2011). PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions, *Journal of Chemical Theory and Computation*, 7(2), 525–537.
- [223] **Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L.** (1983). Comparison of simple potential functions for simulating liquid water, *The Journal of Chemical Physics*, 79(2), 926–935.
- [224] **Storer, J.W., Giesen, D.J., Cramer, C.J. and Truhlar, D.G.** (1995). Class IV charge models: A new semiempirical approach in quantum chemistry, *Journal of Computer-Aided Molecular Design*, 9(1), 87–110.
- [225] **Goetz, A.W., Williamson, M.J., Xu, D., Poole, D., Grand, S.L. and Walker, R.C.** (2012). Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born, *Journal of Chemical Theory and Computation*, 8(5), 1542–1555.
- [226] **Salomon-Ferrer, R., Goetz, A.W., Poole, D., Grand, S.L. and Walker, R.C.** (2013). Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald, *Journal of Chemical Theory and Computation*, 9(9), 3878–3888.
- [227] **Ugur, I., Marion, A., Aviyente, V. and Monard, G.** (2015). Why Does Asn71 Deamidate Faster Than Asn15 in the Enzyme Triosephosphate Isomerase? Answers from Microsecond Molecular Dynamics Simulation and QM/MM Free Energy Calculations, *Biochemistry*, 54(6), 1429–1439.
- [228] **Bill R. Miller, I., T. Dwight McGee, J., Swails, J.M., Homeyer, N., Gohlke, H. and Roitberg, A.E.** (2012). MMPBSA.py: An Efficient Program for End-State Free Energy Calculations, *Journal of Chemical Theory and Computation*, 8(9), 3314–3321.
- [229] **Onufriev, A., Bashford, D. and Case, D.A.** (2004). Exploring protein native states and large-scale conformational changes with a modified generalized born model, *Proteins: Structure, Function, and Bioinformatics*, 55(2), 383–394.
- [230] **Roe, D.R. and Thomas E. Cheatham, I.** (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data, *Journal of Chemical Theory and Computation*, 9(7), 3084–3095.
- [231] **Laskowski, R.A. and Swindells, M.B.** (2011). LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery, *Journal of Chemical Information and Modeling*, 51(10), 2778–2786.

- [232] **Storici, P., Capitani, G., Müller, R., Schirmer, T. and Jansonius, J.N.** (1999). Crystal structure of human ornithine aminotransferase complexed with the highly specific and potent inhibitor 5-fluoromethylornithine, *Journal of Molecular Biology*, 285(1), 297 – 309.
- [233] **Liu, W., Peterson, P.E., Carter, R.J., Zhou, X., Langston, J.A., Fisher, A.J. and Toney, M.D.** (2004). Crystal Structures of Unbound and Aminooxyacetate-Bound Escherichia coli  $\gamma$ -Aminobutyrate Aminotransferase, *Biochemistry*, 43(34), 10896–10905.
- [234] **Markova, M., Peneff, C., Hewlins, M.J.E., Schirmer, T. and John, R.A.** (2005). Determinants of Substrate Specificity in  $\omega$ -Aminotransferases, *Journal of Biological Chemistry*, 280(43), 36409–36416.
- [235] **Lazaridis, T.** (1998). Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory, *The Journal of Physical Chemistry B*, 102(18), 3531–3541.
- [236] **Nguyen, C.N., Kurtzman Young, T. and Gilson, M.K.** (2012). Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril, *The Journal of Chemical Physics*, 137(4), 044101.
- [237] **Siegbahn, P.E.M. and Himo, F.** (2009). Recent developments of the quantum chemical cluster approach for modeling enzyme reactions, *Journal of Biological Inorganic Chemistry*, 14(5), 643–651.
- [238] **Case, D.A., Berryman, J.T., Betz, R.M., Cerutti, D.S., III, T.E.C., Darden, T.A. and Kollman, J.P.A.**, (2014), AMBER 2014, University of California, San Francisco.
- [239] **Grossfield, A.**, WHAM: the weighted histogram analysis method, University of Rochester Medical Center, Rochester, NY, <http://membrane.urmc.rochester.edu/content/wham>.

## **APPENDICES**

### **APPENDIX A.1 : Force Field Modifications for Molecular Dynamics Simulations**



# APPENDIX A.1: Force Field Modifications for Molecular Dynamics Simulations

The AMBER files that have been used to prepare the samples with the tleap program, part of the AMBER package.

## Iron-Sulfur Cluster

### FES.gaff

This is the additional/replacement parameter set for Iron

MASS

FE 55.8

BOND

FE-SH 200.0 2.35

FE-S 200.0 2.35

ANGLE

SH-FE-SH 50.0 104.5

S-FE-S 50.0 104.5

S-FE-SH 50.0 104.5

CT-SH-FE 50.0 104.5

FE-S-FE 50.0 70.0

DIHE

FE-S-FE-S 1 0.0 0.0 0.

FE-S-FE-SH 1 0.0 0.0 0.

SH-FE-SH-CT 1 0.0 0.0 0.

CT-SH-FE-S 1 0.0 0.0 0.

CT-SH-FE-SH 1 0.0 0.0 0.

NONBON

FE 2.0000 0.4417

### FES.leapin

loadamberparams FES.gaff

set FES.800.S1 type S

set FES.800.S2 type S

set FES.800.FE1 type FE

set FES.800.FE2 type FE

set FES.800.S1 charge -0.45

set FES.800.S2 charge -0.45

set FES.800.FE1 charge 0.7

set FES.800.FE2 charge 0.7

bond FES.800.S1 FES.800.FE1

bond FES.800.S1 FES.800.FE2

bond FES.800.S2 FES.800.FE1

bond FES.800.S2 FES.800.FE2

## Cys Amino Acids (Bound to Iron-Sulfur Cluster)

### CYM.leapin

```
CYM = loadpdb CYM.pdb
set CYM.135.N charge -0.46300
set CYM.135.HN charge 0.25200
set CYM.135.CA charge 0.03500
set CYM.135.HA charge 0.04800
set CYM.135.CB charge -0.03300
set CYM.135.HB2 charge 0.07900
set CYM.135.HB3 charge 0.07900
set CYM.135.SG charge -0.23400
set CYM.135.C charge 0.61600
set CYM.135.O charge -0.50400
set CYM head CYM.135.N
set CYM tail CYM.135.C
check CYM
saveoff CYM substrat.lib
```

## Lys329-PLP Complex (The N on the pyridine ring is protonated)

### PLP-NH.gaff

This is the additional/replacement parameter set for Iron

MASS

BOND

ANGLE

HA-CA-N2	50.0	120.00	
CA-CA-N2	70.0	120.10	(CM-CA-N2)
CA-CT-H1	50.0	109.50	(CA-CT-HC)
CA-CT-OS	50.0	109.50	(CT-CT-OS)
CA-C -O	80.0	125.30	(CM-C -O)
C -CA-N2	70.0	120.10	(CM-CA-N2)
C -CA-CT	70.0	120.00	(CA-CA-CT)
CA-N2-CA	70.0	120.00	(CC-NA-CR)
CT-CA-N2	70.0	120.00	(CT-CC-NA)

DIHE

NONBON

### PLP-NH.leapin

loadamberparams PLP.gaff

PLP = loadpdb PLP.pdb

bondbydistance PLP

```
set PLP.2.N type N
set PLP.2.H type H
set PLP.2.C type C
set PLP.2.O type O
set PLP.2.CA type CT
set PLP.2.HA type H1
set PLP.2.CB type CT
set PLP.2.HB2 type HC
set PLP.2.HB3 type HC
set PLP.2.CG type CT
set PLP.2.HG2 type HC
set PLP.2.HG3 type HC
set PLP.2.CD type CT
set PLP.2.HD2 type HC
set PLP.2.HD3 type HC
set PLP.2.CE type CT
set PLP.2.HE2 type H1
set PLP.2.HE3 type H1
set PLP.2.NZ type N2
set PLP.2.HZ1 type H
set PLP.2.C4A type CA
set PLP.2.H4A type HA
set PLP.2.C4 type CA
set PLP.2.C3 type C
set PLP.2.O3 type O
set PLP.2.C2 type CA
set PLP.2.C2A type CT
set PLP.2.H2A1 type HC
```

```

set PLP.2.H2A2 type HC
set PLP.2.H2A3 type HC
set PLP.2.N1 type N2
set PLP.2.H1 type H
set PLP.2.C6 type CA
set PLP.2.H6 type HA
set PLP.2.C5 type CA
set PLP.2.C5A type CT
set PLP.2.H5A1 type H1
set PLP.2.H5A2 type H1
set PLP.2.O4P type OS
set PLP.2.P type P
set PLP.2.O1P type O2
set PLP.2.O2P type O2
set PLP.2.O3P type O2
set PLP.2.N charge -0.515
set PLP.2.H charge 0.340
set PLP.2.C charge 0.399
set PLP.2.O charge -0.443
set PLP.2.CA charge 0.007
set PLP.2.HA charge 0.128
set PLP.2.CB charge -0.103
set PLP.2.HB2 charge 0.064
set PLP.2.HB3 charge 0.064
set PLP.2.CG charge -0.113
set PLP.2.HG2 charge 0.076
set PLP.2.HG3 charge 0.076
set PLP.2.CD charge -0.120
set PLP.2.HD2 charge 0.068
set PLP.2.HD3 charge 0.068
set PLP.2.CE charge -0.019
set PLP.2.HE2 charge 0.090
set PLP.2.HE3 charge 0.090
set PLP.2.NZ charge -0.322
set PLP.2.HZ1 charge 0.391
set PLP.2.C4A charge 0.018
set PLP.2.H4A charge 0.190
set PLP.2.C4 charge -0.254
set PLP.2.C3 charge 0.469
set PLP.2.O3 charge -0.569
set PLP.2.C2 charge -0.400
set PLP.2.C2A charge -0.013
set PLP.2.H2A1 charge 0.038
set PLP.2.H2A2 charge 0.038
set PLP.2.H2A3 charge 0.038
set PLP.2.N1 charge 0.075
set PLP.2.H1 charge 0.288
set PLP.2.C6 charge -0.244
set PLP.2.H6 charge 0.155
set PLP.2.C5 charge -0.001
set PLP.2.C5A charge 0.216
set PLP.2.H5A1 charge 0.030
set PLP.2.H5A2 charge 0.030
set PLP.2.O4P charge -0.695
set PLP.2.P charge 1.990
set PLP.2.O1P charge -0.875
set PLP.2.O2P charge -0.875
set PLP.2.O3P charge -0.875
set PLP head PLP.2.N
set PLP tail PLP.2.C
check PLP
saveoff PLP substrat.lib

```

### Lys329-PLP Complex (The N on the pyridine ring is not protonated)

#### PLP-N.gaff

Remark line goes here

MASS

BOND

CM-NT 477.00 1.343 same as CR-NA, penalty score= 0.2

ANGLE

CM-NT-CT 70.000 120.000 G.M. (post) revision

HA-CM-NT	50.000	120.000	same as H5-CR-NA, penalty score=	1.2
CA-CM-NT	70.000	120.000	same as CC-CW-NA, penalty score=	1.5
CM-NT-H	50.000	120.000	same as CR-NA-H , penalty score=	1.1
C -CA-CM	63.000	120.000	same as C -CA-CA, penalty score=	4.9
CA-CA-CM	63.000	120.000	same as CA-CA-CB, penalty score=	4.2
CA-C -O	80.000	128.800	same as CB-C -O , penalty score=	3.0
C -CA-CT	70.000	120.000	G.M. (post) revision	
C -CA-NA	70.000	120.000	G.M. (post) revision	
CA-NA-CA	70.000	120.000	same as CC-NA-CR, penalty score=	7.4
CT-CA-NA	70.000	120.000	same as CT-CC-NA, penalty score=	4.5
H4-CA-NA	50.000	120.000	G.M. (post) revision	
CA-CA-NA	70.000	123.500	same as CB-CA-N2, penalty score=	6.8
CA-CT-H1	50.000	109.500	same as CA-CT-HC, penalty score=	0.1
CA-CT-OS	50.000	109.500	same as CM-CT-OS, penalty score=	4.9
DIHE				
HA-CM-NT-CT	1	1.500	180.000	2.000 same as X -CA-NA-X , penalty score=1080.0
CA-CM-NT-CT	1	1.500	180.000	2.000 same as X -CA-NA-X , penalty score=1080.0
HA-CM-NT-H	1	1.500	180.000	2.000 same as X -CA-NA-X , penalty score=1080.0
CA-CM-NT-H	1	1.500	180.000	2.000 same as X -CA-NA-X , penalty score=1080.0
IMPROPER				
CA-HA-CM-NT	1.1	180.0	2.0	Using the default value
C -CA-CA-CM	1.1	180.0	2.0	Using the default value
CA-CA-C -O	10.5	180.0	2.0	Using general improper torsional angle X-X-C-O, penalty score=20.0)
C -CT-CA-NA	1.1	180.0	2.0	Using the default value
CA-CA-NA-N	1.1	180.0	2.0	Using the default value
CA-H4-CA-NA	1.1	180.0	2.0	Using the default value
NONBON				

# PLP-N.prep

0 0 2

This is a remark line

molecule.res

PLP XYZ 0

CHANGE OMIT DU BEG

0.0000							
1	DUMM	DU	M	999.000	999.0	-999.0	.00000
2	DUMM	DU	M	999.000	-999.0	999.0	.00000
3	DUMM	DU	M	-999.000	999.0	999.0	.00000
4	N	N	M	-4.160000	1.199000	-1.237000	-0.546008
5	H	H	E	-4.985000	1.639000	-1.404000	0.356971
6	CA	CT	M	-4.216000	-0.251000	-0.980000	0.023103
7	HA	H1	E	-3.588000	-0.728000	-1.802000	0.116554
8	C	C	M	-5.640000	-0.859000	-1.095000	0.421096
9	O	O	E	-6.011000	-1.745000	-1.827000	-0.463174
10	CB	CT	3	-3.709000	-0.575000	0.419000	-0.112108
11	HB2	HC	E	-4.319000	0.046000	1.087000	0.063161
12	HB3	HC	E	-3.944000	-1.608000	0.664000	0.063161
13	CG	CT	3	-2.239000	-0.247000	0.652000	-0.128008
14	HG2	HC	E	-1.938000	0.767000	0.407000	0.088521
15	HG3	HC	E	-1.638000	-1.014000	0.039000	0.088521
16	CD	CT	3	-1.951000	-0.513000	2.038000	-0.106328
17	HD2	HC	E	-2.458000	0.229000	2.692000	0.034681
18	HD3	HC	E	-2.283000	-1.494000	2.344000	0.034681
19	CE	CT	3	-0.495000	-0.417000	2.501000	-0.026388
20	HE2	H1	E	0.064000	-1.291000	2.427000	0.065297
21	HE3	H1	E	-0.615000	-0.168000	3.622000	0.065297
22	NZ	NT	B	0.303000	0.677000	1.851000	-0.299819
23	HZ1	H	E	-0.104000	1.606000	1.740000	0.324666
24	C4A	CM	B	1.481000	0.444000	1.301000	-0.169733
25	H4A	HA	E	2.027000	-0.604000	1.312000	0.381932
26	C4	CA	B	2.197000	1.535000	0.621000	-0.313564
27	C3	C	B	1.580000	2.844000	0.621000	0.405118
28	O3	O	E	0.672000	3.190000	1.379000	-0.669643
29	C2	CA	B	2.230000	3.920000	-0.287000	-0.160315
30	C2A	CT	3	1.753000	5.323000	-0.306000	-0.012903
31	H2A1	HC	E	1.675000	5.780000	-1.274000	0.016576
32	H2A2	HC	E	0.708000	5.394000	0.047000	0.016576
33	H2A3	HC	E	2.393000	6.002000	0.337000	0.016576
34	N1	NA	S	3.213000	3.557000	-1.066000	-0.310101
36	C6	CA	B	3.883000	2.343000	-0.883000	-0.170134
37	H6	H4	E	4.716000	2.183000	-1.565000	0.103899
38	C5	CA	B	3.362000	1.323000	-0.067000	0.069579



39	C5A	CT	3	4.120000	0.045000	-0.038000	0.222106
40	H5A1	H1	E	5.262000	0.262000	-0.039000	0.010918
41	H5A2	H1	E	3.975000	-0.577000	0.845000	0.010918
42	O4P	OS	S	3.734000	-0.750000	-1.198000	-0.719859
43	P	P	3	3.296000	-2.282000	-0.643000	2.163796
44	O2P	O2	E	2.284000	-3.297000	-1.479000	-0.985207
45	O3P	O2	E	4.503000	-3.306000	-0.417000	-0.985207
46	O1P	O2	E	2.463000	-2.230000	0.693000	-0.985207

LOOP  
C5 C4  
IMPROPER  
-M CA N H  
CA +M C O  
C4 H4A C4A NZ  
C3 C5 C4 C4A  
C4 C2 C3 O3  
C3 C2A C2 N1  
C2 C6 N1 H1  
C5 H6 C6 N1  
C4 C6 C5 C5A  
DONE  
STOP

### Lys329-Vigabatrin-PLP Complex (The N on the pyridine ring is protonated)

#### VIG-NH.gaff

This is the additional/replacement parameter set for Iron

MASS

BOND

ANGLE

HA-CA-N2	50.0	120.00	
CA-CA-N2	70.0	120.10	(CM-CA-N2)
CA-CT-H1	50.0	109.50	(CA-CT-HC)
CA-CT-OS	50.0	109.50	(CT-CT-OS)
CA-C -O	80.0	125.30	(CM-C -O)
C -CA-N2	70.0	120.10	(CM-CA-N2)
C -CA-CT	70.0	120.00	(CA-CA-CT)
CA-N2-CA	70.0	120.00	(CC-NA-CR)
CT-CA-N2	70.0	120.00	(CT-CC-NA)

DIHE

NONBON

#### VIG-NH.leapin

loadamberparams VIG.gaff

VIG = loadpdb VIG.pdb

bondbydistance VIG

```

set VIG.2.N type N
set VIG.2.H type H
set VIG.2.C type C
set VIG.2.O type O
set VIG.2.CA type CT
set VIG.2.HA type H1
set VIG.2.CB type CT
set VIG.2.HB2 type HC
set VIG.2.HB3 type HC
set VIG.2.CG type CT
set VIG.2.HG2 type HC
set VIG.2.HG3 type HC
set VIG.2.CD type CT
set VIG.2.HD2 type HC
set VIG.2.HD3 type HC
set VIG.2.CE type CT
set VIG.2.HE2 type H1
set VIG.2.HE3 type H1
set VIG.2.NZ type NT
set VIG.2.HZ1 type H
set VIG.2.CI type CT
set VIG.2.HI2 type H1
set VIG.2.HI3 type H1
set VIG.2.CJ type CT
set VIG.2.HJ2 type HC

```

```

set VIG.2.HJ3 type HC
set VIG.2.CK type CT
set VIG.2.HK2 type H1
set VIG.2.CL type CT
set VIG.2.HL2 type HC
set VIG.2.HL3 type HC
set VIG.2.CM type CT
set VIG.2.HM2 type HC
set VIG.2.HM3 type HC
set VIG.2.CN type C
set VIG.2.ON1 type O2
set VIG.2.ON2 type O2
set VIG.2.NP type N2
set VIG.2.HP type H
set VIG.2.C4A type CA
set VIG.2.H4A type HA
set VIG.2.C4 type CA
set VIG.2.C3 type C
set VIG.2.O3 type O
set VIG.2.C2 type CA
set VIG.2.C2A type CT
set VIG.2.H2A1 type HC
set VIG.2.H2A2 type HC
set VIG.2.H2A3 type HC
set VIG.2.N1 type N2
set VIG.2.H1 type H
set VIG.2.C6 type CA
set VIG.2.H6 type HA
set VIG.2.C5 type CA
set VIG.2.C5A type CT
set VIG.2.H5A1 type H1
set VIG.2.H5A2 type H1
set VIG.2.O4P type OS
set VIG.2.P type P
set VIG.2.O1P type O2
set VIG.2.O2P type O2
set VIG.2.O3P type O2
set VIG.2.N charge -0.504
set VIG.2.H charge 0.352
set VIG.2.C charge 0.396
set VIG.2.O charge -0.441
set VIG.2.CA charge 0.014
set VIG.2.HA charge 0.107
set VIG.2.CB charge -0.115
set VIG.2.HB2 charge 0.085
set VIG.2.HB3 charge 0.085
set VIG.2.CG charge -0.097
set VIG.2.HG2 charge 0.051
set VIG.2.HG3 charge 0.051
set VIG.2.CD charge -0.123
set VIG.2.HD2 charge 0.084
set VIG.2.HD3 charge 0.084
set VIG.2.CE charge 0.003
set VIG.2.HE2 charge 0.043
set VIG.2.HE3 charge 0.043
set VIG.2.NZ charge -0.505
set VIG.2.HZ1 charge 0.315
set VIG.2.C1 charge -0.007
set VIG.2.HI2 charge 0.062
set VIG.2.HI3 charge 0.062
set VIG.2.CJ charge -0.118
set VIG.2.HJ2 charge 0.057
set VIG.2.HJ3 charge 0.057
set VIG.2.CK charge -0.010
set VIG.2.HK2 charge 0.117
set VIG.2.CL charge -0.087
set VIG.2.HL2 charge 0.069
set VIG.2.HL3 charge 0.069
set VIG.2.CM charge -0.151
set VIG.2.HM2 charge 0.038
set VIG.2.HM3 charge 0.038
set VIG.2.CN charge 0.513
set VIG.2.ON1 charge -0.685
set VIG.2.ON2 charge -0.658

```

```

set VIG.2.NP charge -0.311
set VIG.2.HP charge 0.376
set VIG.2.C4A charge 0.009
set VIG.2.H4A charge 0.248
set VIG.2.C4 charge -0.276
set VIG.2.C3 charge 0.489
set VIG.2.O3 charge -0.608
set VIG.2.C2 charge -0.503
set VIG.2.C2A charge 0.020
set VIG.2.H2A1 charge 0.019
set VIG.2.H2A2 charge 0.019
set VIG.2.H2A3 charge 0.019
set VIG.2.N1 charge 0.016
set VIG.2.H1 charge 0.270
set VIG.2.C6 charge -0.238
set VIG.2.H6 charge 0.133
set VIG.2.C5 charge -0.012
set VIG.2.C5A charge 0.205
set VIG.2.H5A1 charge 0.049
set VIG.2.H5A2 charge 0.049
set VIG.2.O4P charge -0.671
set VIG.2.P charge 1.843
set VIG.2.O1P charge -0.813
set VIG.2.O2P charge -0.813
set VIG.2.O3P charge -0.813
set VIG head VIG.2.N
set VIG tail VIG.2.C
check VIG
saveoff VIG substrat.lib

```

### Lys329-Vigabatrin-PLP Complex (The N on the pyridine ring is not protonated)

#### VIG-N.gaff

Remark line goes here

MASS

BOND

CM-NT 477.00 1.343 same as CR-NA, penalty score= 0.2

ANGLE

CM-NT-CT	70.000	120.000	G.M. (post) revision
HA-CM-NT	50.000	120.000	same as H5-CR-NA, penalty score= 1.2
CA-CM-NT	70.000	120.000	same as CC-CW-NA, penalty score= 1.5
CM-NT-H	50.000	120.000	same as CR-NA-H , penalty score= 1.1
C -CA-CM	63.000	120.000	same as C -CA-CA, penalty score= 4.9
CA-CA-CM	63.000	120.000	same as CA-CA-CB, penalty score= 4.2
CA-C -O	80.000	128.800	same as CB-C -O , penalty score= 3.0
C -CA-CT	70.000	120.000	G.M. (post) revision
C -CA-NA	70.000	120.000	G.M. (post) revision
CA-NA-CA	70.000	120.000	same as CC-NA-CR, penalty score= 7.4
CT-CA-NA	70.000	120.000	same as CT-CC-NA, penalty score= 4.5
H4-CA-NA	50.000	120.000	G.M. (post) revision
CA-CA-NA	70.000	123.500	same as CB-CA-N2, penalty score= 6.8
CA-CT-H1	50.000	109.500	same as CA-CT-HC, penalty score= 0.1
CA-CT-OS	50.000	109.500	same as CM-CT-OS, penalty score= 4.9

DIHE

HA-CM-NT-CT	1	1.500	180.000	2.000	same as X -CA-NA-X , penalty score=1080.0
CA-CM-NT-CT	1	1.500	180.000	2.000	same as X -CA-NA-X , penalty score=1080.0
HA-CM-NT-H	1	1.500	180.000	2.000	same as X -CA-NA-X , penalty score=1080.0
CA-CM-NT-H	1	1.500	180.000	2.000	same as X -CA-NA-X , penalty score=1080.0

IMPROPER

CT-O2-C -O2	10.5	180.0	2.0	Using general improper torsional angle X-O2-C-O2, penalty score=10.0)
CA-HA-CM-NT	1.1	180.0	2.0	Using the default value
C -CA-CA-CM	1.1	180.0	2.0	Using the default value
CA-CA-C -O	10.5	180.0	2.0	Using general improper torsional angle X-X-C-O, penalty score= 20.0)
C -CT-CA-NA	1.1	180.0	2.0	Using the default value
CA-H4-CA-NA	1.1	180.0	2.0	Using the default value

NONBON

# VIG-N.prep

```

0      0      2
This is a remark line
molecule.res
VIG XYZ 0
CHANGE OMIT DU BEG
0.0000
1 DUMM DU M 999.000 999.0 -999.0 .00000
2 DUMM DU M 999.000 -999.0 999.0 .00000
3 DUMM DU M -999.000 999.0 999.0 .00000
4 N N M -5.753000 0.848000 0.486000 -0.564059
5 H H E -6.681000 0.924000 0.852000 0.362093
6 CA CT M -5.797000 0.204000 -0.816000 0.021453
7 HA H1 E -5.423000 0.904000 -1.533000 0.113159
8 C C M -7.219000 -0.119000 -1.283000 0.425189
9 O O E -7.915000 0.759000 -1.777000 -0.471271
10 CB CT 3 -4.899000 -1.019000 -0.822000 -0.108953
11 HB2 HC E -5.187000 -1.602000 0.028000 0.049855
12 HB3 HC E -4.971000 -1.435000 -1.805000 0.049855
13 CG CT 3 -3.459000 -0.679000 -0.621000 -0.134414
14 HG2 HC E -3.327000 -0.509000 0.427000 0.076201
15 HG3 HC E -3.226000 0.095000 -1.323000 0.076201
16 CD CT 3 -2.586000 -1.877000 -0.984000 -0.083301
17 HD2 HC E -2.655000 -2.562000 -0.165000 0.041148
18 HD3 HC E -2.885000 -2.185000 -1.965000 0.041148
19 CE CT 3 -1.143000 -1.475000 -1.085000 -0.098696
20 HE2 H1 E -1.110000 -0.431000 -0.849000 0.116026
21 HE3 H1 E -0.599000 -2.175000 -0.487000 0.116026
22 NZ NT B -0.600000 -1.602000 -2.461000 -0.535773
23 HZ1 H E -0.113000 -0.755000 -2.677000 0.326364
24 CI CT 3 0.413000 -2.702000 -2.363000 -0.001531
25 HI2 H1 E 0.490000 -2.957000 -1.327000 0.044309
26 HI3 H1 E 0.129000 -3.446000 -3.077000 0.044309
27 CJ CT 3 1.783000 -2.187000 -2.789000 -0.138913
28 HJ2 HC E 1.603000 -1.576000 -3.649000 0.057424
29 HJ3 HC E 2.402000 -3.056000 -2.870000 0.057424
30 CK CT 3 2.486000 -1.279000 -1.785000 -0.038390
31 HK2 H1 E 2.168000 -0.269000 -1.937000 0.110071
32 CQ CT 3 4.001000 -1.367000 -1.958000 -0.098364
33 HQ2 HC E 4.434000 -1.073000 -1.025000 0.059265
34 HQ3 HC E 4.208000 -2.336000 -2.362000 0.059265
35 CM CT 3 4.464000 -0.352000 -2.994000 -0.166114
36 HM2 HC E 3.830000 0.504000 -2.897000 0.043959
37 HM3 HC E 4.525000 -0.868000 -3.930000 0.043959
38 CN C B 5.861000 0.093000 -2.645000 0.556961
39 ON1 O2 E 6.037000 1.164000 -2.058000 -0.721739
40 ON2 O2 E 6.804000 -0.631000 -2.947000 -0.721739
41 NP NT B 2.162000 -1.755000 -0.448000 -0.200063
42 HP H E 1.742000 -2.659000 -0.367000 0.292993
43 CA4 CM B 2.105000 -0.888000 0.575000 -0.205872
44 HA4 HA E 2.171000 0.157000 0.355000 0.396372
45 C4 CA B 1.951000 -1.302000 2.017000 -0.298490
46 C3 C B 1.720000 -2.639000 2.380000 0.410312
47 O3 O E 1.778000 -3.651000 1.439000 -0.657380
48 C2 CA B 1.532000 -2.950000 3.744000 -0.178773
49 CA2 CT 3 1.282000 -4.347000 4.223000 -0.008874
50 HA21 HC E 1.169000 -4.344000 5.287000 0.011798
51 HA22 HC E 0.390000 -4.725000 3.770000 0.011798
52 HA23 HC E 2.110000 -4.969000 3.954000 0.011798
53 N1 NA S 1.571000 -1.957000 4.693000 -0.323539
55 C6 CA B 1.797000 -0.649000 4.336000 -0.162488
56 H6 H4 E 1.837000 0.112000 5.088000 0.104171
57 C5 CA B 1.973000 -0.313000 3.001000 0.038163
58 CA5 CT 3 2.054000 1.156000 2.635000 0.208485
59 HA51 H1 E 2.047000 1.693000 3.560000 0.020329
60 HA52 H1 E 2.886000 1.249000 1.968000 0.020329
61 OP4 OS S 0.917000 1.611000 1.909000 -0.724030
62 P P 3 1.151000 2.873000 0.920000 2.181480
63 OP2 O2 E 2.019000 2.367000 -0.259000 -0.985643
64 OP3 O2 E -0.236000 3.387000 0.489000 -0.985643
65 OP1 O2 E 1.909000 3.944000 1.629000 -0.985643
LOOP
C5 C4

```

```

IMPROPER
-M   CA   N   H
CA   +M   C   O
CM   ON1  CN  ON2
C4   HA4  CA4  NP
C3   C5   C4  CA4
C4   C2   C3  O3
C3   CA2  C2  N1
C5   H6   C6  N1
C4   C6   C5  CA5
DONE
STOP

```



### Résumé :

La compréhension des enzymes et de leurs mécanismes catalytiques est d'une grande importance dans le développement de médicaments plus efficaces. Pour mieux appréhender ces phénomènes, différentes approches théoriques comme les méthodes QM, MM-MD et QM/MM, peuvent être utilisées.

L'objectif principal de cette thèse est d'obtenir une meilleure compréhension des mécanismes de réactivité et de la dynamique de l'enzyme GABA-AT ( $\gamma$ -aminobutyric acid aminotransferase), un modèle d'enzyme dépendante au phosphate pyridoxal (PLP).

Notre travail a consisté en 5 étapes vers une plus grande compréhension de GABA-AT. 1) la réaction et le mode d'attachement du substrat naturel GABA ont été étudiés pour différents isomères à l'aide de systèmes modèles et de la DFT. 2) l'enzyme a été simulée par dynamique moléculaire classique dans les cas de l'apoenzyme, l'holoenzyme et l'holoenzyme inactivée. Nos résultats montrent que plusieurs résidus du site actif jouent un rôle important et que leur état de protonation ainsi que celui du PLP sont cruciaux dans l'activité de GABA-AT. 3) l'influence des résidus du site actif sur la réactivité a été étudiée par la modélisation quantique de clusters moléculaires. Le plus gros cluster comprenait 165 atomes entouré d'un solvant implicite. 4) de nouvelles routines de diagonalisation pour SEBOMD ont été incorporées dans la suite AMBER à travers l'utilisation des bibliothèques LAPACK et SCALAPACK. Ces nouvelles routines ont été testées et leur efficacité a été évaluée. 5) des énergies libres de réaction ont été évaluées par dynamiques SEBOMD sur des intermédiaires réactionnels GABA-PLP.

### Mots-clés :

Mécanique quantique; Dynamique moléculaire; Méthodes semi-empirique; Croissance linéaire; Enzyme réactivité; GABA-AT.

---

### Abstract:

Understanding enzymes and their catalytic mechanisms is very important in order to develop more effective drugs having little to no side effects. In order to decipher the catalytic behavior of enzymes, different approaches such as QM, MM-MD, and QM/MM can be used and their results can be correlated.

The main aim of this thesis is to get a deeper understanding of the mechanistic insights of the reactivity and of the dynamics of the pyridoxal-5-phosphate (PLP) dependent enzyme  $\gamma$ -aminobutyric acid aminotransferase (GABA-AT). Because GABA-AT resembles many other PLP-dependent enzymes, understanding it could be of importance for the broad community of biochemists and computational chemists who study such class of proteins.

Our work has consisted of five stages to pursue the comprehension of GABA-AT. First, the reaction and the preferred binding mode of the natural substrate GABA has been elucidated with different isomers by means of model systems with DFT. Second, the dynamics and the behavior of the enzyme has been studied with MM-MD through the use of apoenzyme, holoenzyme and holoenzyme with an inactivator. Third, the effect of the active site residues in the inactivation mechanism has been investigated with the modelling of clusters at the QM level involving key residues. Fourth, new diagonalization routines for the SEBOMD (SemiEmpirical Born-Oppenheimer Molecular Dynamics) approach implemented in the Amber suite of programs, have been incorporated using LAPACK and SCALAPACK libraries, tested and evaluated to optimize the diagonalization procedure of the Fock matrix. Fifth, reaction free energies of PLP containing systems have been investigated with SEBOMD simulations.

### Keywords:

Quantum mechanics; molecular Dynamics; Semiempirical methods; Linear scaling methods; Enzyme reactivity; GABA-AT.

---

### Publication associée/Related publication:

H. Gökcan, and F. A. S. Konuklar, *Journal of Molecular Graphics and Modelling*, **2014**, 51, 173 - 183.

H. Gökcan, G. Monard, and F. A. S. Konuklar, *Proteins: Structure, Function, and Bioinformatics*, **2016**, 84(7), 875-891