On the role of protons in the reactivation of acetylcholinesterase : quantum and molecular mechanics studies

Thomas Driant

Université Pierre et Marie Curie

Ecole doctoral de Chimie Moléculaire de Paris Centre

Institut Parisien de Chimie Moléculaire/ Equipe MACO

On the Role of Protons in the Reactivation of Acetylcholinesterase: Quantum and Molecular Mechanics Studies

Présentée par :

Thomas Driant

Pour obtenir le grade de

Docteur de l'université Pierre et Marie Curie

Soutenance publique prévue le 22 septembre 2017

Devant un jury composé de :

Madame	Isabelle NAVIZET	Professeur	Rapporteur
Monsieur	Sam de VISSER	Professeur	Rapporteur
Madame	Shina Caroline Lynn KAMERLIN	Professeur	Examinateur
Monsieur	Florian NACHON	Ingénieur d'études et de fabrications HDR	Examinateur
Monsieur	Philippe KAROYAN	Professeur	Examinateur
Monsieur	Matthieu SOLLOGOUB	Professeur	Examinateur
Monsieur	Etienne DERAT	Maître de Conférences HDR	Directeur de thèse

Acknowledgments

My PhD Project was carried out at the Institut Parisien de Chimie Moléculaire, directed by Doctor Corinne Aubert. I worked as a member of the MACO team directed by Professor Louis Fensterbank.

Firstly, I would like to address sincere thanks to Professor Isabelle Navizet, Professor Sam de Visser, Professor Shina Caroline Lynn Kamerlin, Professor Philippe Karoyan, Docteur Florian Nachon, and Professeur Matthieu Sollogoub for accepting to review my thesis and be part of my Jury.

I am especially grateful to Doctor Etienne Derat for trusting me with this project. Looking back at my progress and maturation as a researcher I am proud of what I achieved and have you to thank for it. You have always been available for me to help me, guide me, teach me, and discuss the project. I am honoured by the confidence you put in me and I hope to have risen to the opportunity you have given me three years ago.

I am also grateful to Professor Shina Caroline Lynn Kamerlin for hosting me in her team in Uppsala University for three months. I thank you for the substantial contribution you have made to my scientific growth throughout our collaboration.

I also thank all the members of both the MACO team and the team of the Kamerlin lab, you are wonderful colleagues and I deeply enjoyed working with all of you.

Finally, I would like to thank my friends, my family and Anne for their love and support during those three years.

Table of Contents

ACKNOWLEDGMENTS 3

TABLE OF CONTENTS 4

GENERAL INTRODUCTION 11

I. Biological significance of acetylcholinesterase 11
I.1. Role as a mediator of signal transmission 11
I.2. Target of nerve agents 13

II. Scope of this work 15

CHAPTER 1: INTRODUCTION ON ACETYLCHOLINESTERASE: STRUCTURE, FUNCTION, AND MECHANISM 19

I. Structure of acetylcholinesterase 19
I.1. Active site 23
I.2. Enzyme channel of AChE 29

II. Activity of acetylcholinesterase 38
II.1. Regular catalytic activity 38
II.2. Inhibition by organophosphates 40
II.3. Reactivation of acetylcholinesterase 46

III. Molecular simulations of AChE 54
III.1. Regular catalytic activity 54
III.2. Inhibition by nerve agents 66
III.3. Reactivation by oximes 76

IV. Conclusions on acetylcholinesterase literature 84

CHAPTER 2: INTRODUCTION TO THE COMPUTATIONAL METHODS 87

I. Quantum mechanics 87
I.1. From molecular orbital method to Hartree-Fock theory 88
I.2. Density functional theory 93
I.3. Valence bond theory 97

II. Molecular mechanics 99
II.1. Classical force fields 99
II.2. Polarizable force fields 101
II.3. Molecular dynamics 103
II.4. Constant pH Molecular Dynamics 104

II.5. Empirical Valence Bond 105

III. QM/MM 87
III.1. Additive or Subtractive 109
III.2. Boundary scheme 111
III.3. Interactions between QM and MM 113
III.4. How to set up QM/MM simulations 114

IV. Conclusion 118

CHAPTER 3: TO DIVIDE AND CONQUER 121

I. Introduction 121

II. Minimal model 124

III. The role of Glu334 128

IV. The oxyanionic hole 131

V. The aromatic rings in the active site of AChE 136
V.1. Benzene 137
V.2. Phenol 144
V.3. Indole 150

VI. Discussion 156

VII. Conclusions 160

CHAPTER 4: QM/MM SIMULATIONS OF ACETYLCHOLINESTERASE REACTIVATION BY 2-PAM 163

I. Computational details 163

II. Destabilization of the frontier orbitals of the systems 169

III. Orientation of 2-PAM 173

IV. Study of the protonation state of Glu334 173

V. VX-inhibited AChE reactivation by 2-PAM 175
V.1. Reactivation of AChE with an unprotonated Glu202 176
V.2. Reactivation of AChE with a protonated Glu202 179
V.3. Reactivation with an unprotonated His447 and a protonated Glu202 181
V.4. QM model confirmation 182
V.5. Alternate approach of 2-PAM towards the phosphorus 184

VI. Deprotonation of 2-PAM in the active site of AChE 187
VI.1. Case of unprotonated His447 and unprotonated Glu202 187

VI.2. Case of unprotonated His447 and protonated Glu202 189
VI.3. Case of protonated His447 and unprotonated Glu202 189

VII. Discussion 190

VIII. Conclusion 195

CHAPTER 5: SIMULATIONS OF NON-PYRIDINIUM REACTIVATORS 199

I. Introduction 199

II. Truncated model QM study of the reactivation 200
II.1. 6M-2PA 201
II.2. 6M-3H-2PA 204
II.3. 6M-3M-2PA 207

III. QM/MM study of the reactivation of VX-inhibited AChE by non-pyridinium reactivators 210
III.1. 6M-2PA 210
III.2. 6M-3H-2PA 213

IV. QM/MM study of the deprotonation of non-pyridinium reactivators in the active site of VX-inhibited AChE 216
IV.1. 6M-2PA 216
IV.2. 6M-3H-2PA 217

V. Discussion 218

VI. Conclusion 223

CHAPTER 6: DIFFUSION OF PROTONS IN AND OUT OF THE ACTIVE SITE OF ACETYLCHOLINESTERASE 227

I. Introduction 227

II. Direct proton exchange between His447, Glu202, and the solvent 227
II.1. Proton exchange between His447 and Glu202 229
II.2. Proton exchange between the solvent and His447 230
II.3. Proton exchange between the solvent and Glu202 231

III. Molecular dynamics of acetylcholinesterase with polarizable force field 232

IV. Constant pH molecular dynamic simulations 235

V. A novel mechanism for proton diffusion in and out of the active site by QM/MM 241
V.1. VX-inhibited AChE with bis-protonated His447 244
V.2. Non-inhibited AChE with mono-protonated His447 246

VI. EVB simulations on proton diffusion in and out of the active site 249

VI.1. VX-inhibited AChE with a bis-protonated His447 256
VI.2. VX-inhibited AChE with a mono-protonated His447 257
VI.3. Non-inhibited AChE 258

VII. Discussion 259

VIII. Conclusion 264

GENERAL CONCLUSION AND PERSPECTIVES 267

APPENDIX 1: ENZYME KINETICS 273

APPENDIX 2: CHARMM PARAMETERS FOR SER203-VX ADDUCT AND 2-PAM 277

APPENDIX 3: AMOEBA PARAMETERS FOR SER203-VX ADDUCT 279

APPENDIX 4: GAFF PARAMETERS FOR SER203-VX ADDUCT 285

APPENDIX 5: Q-OPLSAA PARAMETERS FOR SER203-VX ADDUCT 287

APPENDIX 6: PYTHON SCRIPTS 290

General Introduction

General Introduction

I. Biological significance of acetylcholinesterase

I.1. Role as a mediator of signal transmission

Synapses are junctions between a neuron and another cell, either a neuron, or a different cell type. Their purpose is to allow for the transmission of a signal between two cells. In chemical synapses, the electric signal from the axon terminal of the presynaptic neuron is converted to a chemical signal.[1] A neurotransmitter is released into the synaptic cleft where it binds to a receptor on the postsynaptic neuron, which results in a response in that cell (an activation pathway of a new electrical signal) (see **Figure I-1**). For cholinergic synapses, that neurotransmitter is acetylcholine (ACh).

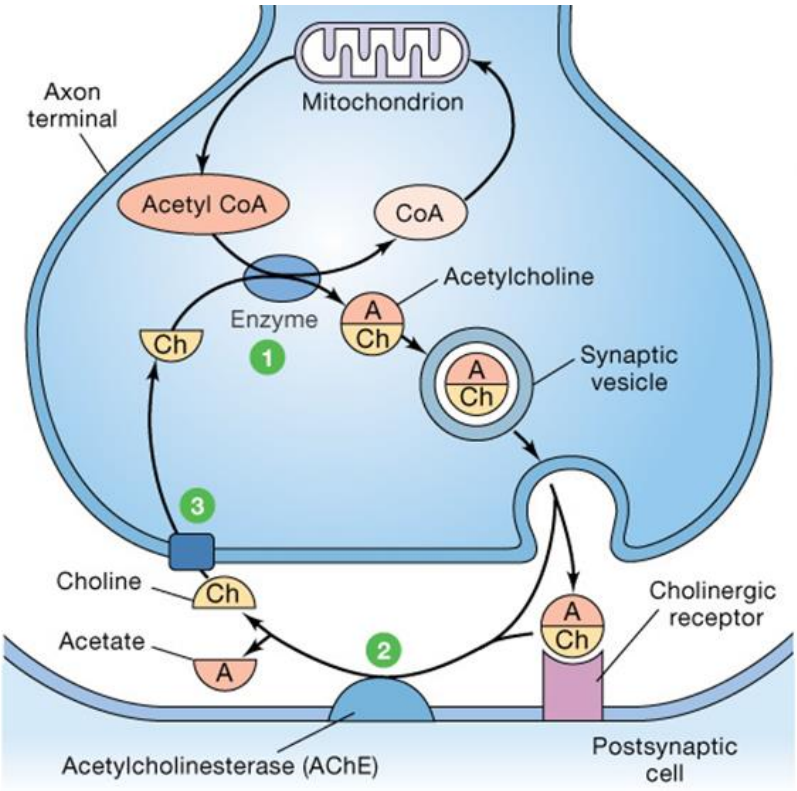
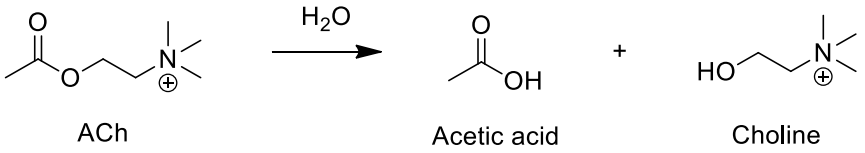


Figure I-1. Representation of a cholinergic synapse highlighting the role of acetylcholinesterase. Image found at ref [2]

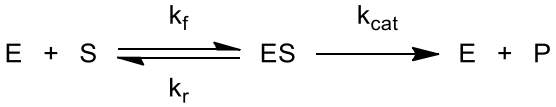
Every time ACh binds to a cholinergic receptor it produces a response in the postsynaptic cell. It is then rereleased into the synapse where it is free to activate a receptor once more. To prevent overactivation of the postsynaptic cell, ACh must then be removed from the synapse.[3] This removal is key to the precise temporal control of muscle contraction.

[1] R. L. Rapport, Nerve Endings: The Discovery of the Synapse, W.W. Norton, **2005**.
[2] "Acetylcholine synthesis and recycling.," can be found under http://faculty.pasadena.edu/dkwon/chap%208_files/images/image61.png, **n.d.**
[3] D. M. Quinn, *Chem. Rev.* **1987**, *87*, 955–979.

It is the role of acetylcholinesterase (AChE, Enzyme Code 3.1.1.7) which breaks ACh into and acetic acid and a choline (see **Scheme I-1**).[4] The catalysis of AChE follows Michaelis-Menten kinetics[5],[6] and as such catalysis proceeds according to **Scheme I-2** with a rate of product formation described in **equation I-I**.[7] Enzymatic kinetics are further expanded upon in **Appendix 1**.



Scheme I-1. Hydrolysis of ACh into an acetic acid and a choline.



Scheme I-2. Enzyme catalysis equation following the Michaelis-Menten model

$$v = \frac{d[P]}{dt} = \frac{V_{max}[S]}{K_M + [S]} = k_{cat}[E]_0 \frac{[S]}{K_M + [S]} \qquad \textbf{I-I}$$

$$k_s = \frac{k_{cat}}{K_M} \qquad \textbf{I-II}$$

The K_M of AChE, which is a descriptor of its affinity for the substrate, is 10^{-4} M towards ACh.[8] The lower the K_M, the greater the affinity of the enzyme for its substrate. The k_{cat} is a measure of the rate of the catalytic reaction independently from the concentration of reactants. The k_{cat} of AChE is 10^4 s^{-1}. Both measures are often combined into a catalytic efficiency constant, also called specificity constant (**Equation I-II**) which describes the efficiency of substrate to product conversion by an enzyme. For AChE, the specificity constant is 10^8 $M^{-1}.s^{-1}$. Those three constants are characteristic of an extremely efficient enzyme that reaches the theoretical diffusion limit for enzymatic reactions with specificity constants in the range of 10^8 to 10^{10}.[9],[10] For more detail see **Appendix 1**.

Once AChE releases the products of ACh hydrolysis the choline enters the axon terminal of the presynaptic neuron and is used to produce more ACh by acyl transfer from the acetyl CoA substrate. If the hydrolysis of ACh is too fast or too slow it has serious health consequences. The overexpression of AChE compared to the release of ACh in the synaptic cleft, present in Alzheimer's disease, leads to an incorrect transmission of the nervous signal

[4] T. L. Rosenberry, *Proc. Nat. Ac. Sc. USA.* **1975**, *72*, 3834–3838.
[5] L. Michaelis, M. L. Menten, *Biochem. Z.* **1913**, *49*, 333–369.
[6] K. A. Johnson, R. S. Goody, *Biochemistry* **2011**, *50*, 8264–8269.
[7] W. W. Chen, M. Niepel, P. K. Sorger, *Genes Dev.* **2010**, *24*, 1861–1875.
[8] H.-J. Nolte, T. L. Rosenberry, E. Neumann, *Biochemistry (Mosc.)* **1980**, *19*, 3705–3711.
[9] M. E. Stroppolo, M. Falconi, A. M. Caccuri, A. Desideri, *CMLS, Cell. Mol. Life Sci.* **2001**, *58*, 1451–1460.
[10] R. A. Alberty, G. G. Hammes, *J. Phys. Chem.* **1958**, *62*, 154–159.

and can result in a deterioration of cognitive and neuromuscular functions.[11] The diminution or loss of AChE function results in an overactivation of the synapses and the loss of the temporal control of muscle contraction.

I.2. Target of nerve agents

Several molecules and peptides are AChE inhibitors[12]. The fasciculin peptide (see **Figure I-2**) is a neurotoxin from mamba snakes (*dendroaspis*) that inhibits AChE by blocking access to its active site.[13] It binds to AChE by non-covalent interactions, usually associated with reversible inhibition. Smaller inhibitors, like donepezil, (see **Scheme I-3**) exhibit the same ability to inhibit AChE by blocking active site access and have been used therapeutically. [14] Carbamates such as carbofuran (see **Scheme I-3**) are covalent inhibitors of AChE used as pesticides.[15] As covalent inhibitors, they are often not spontaneously reversible.

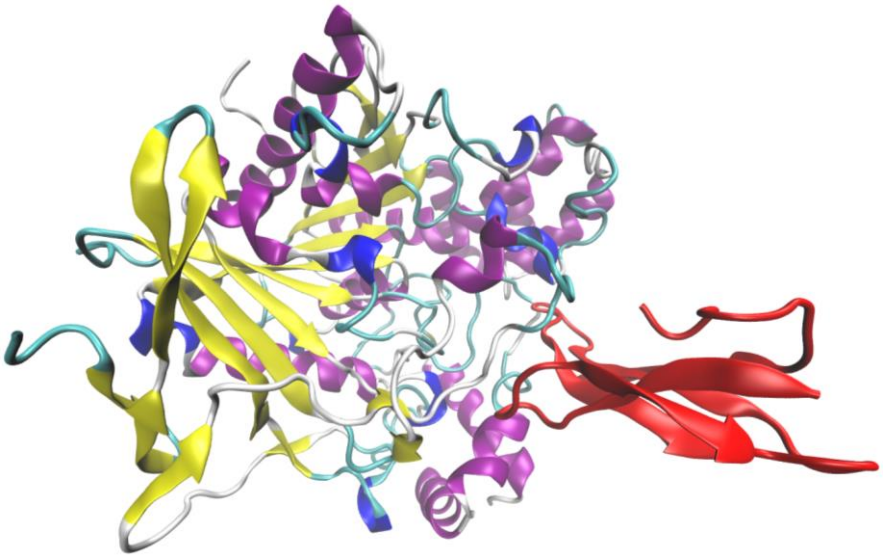
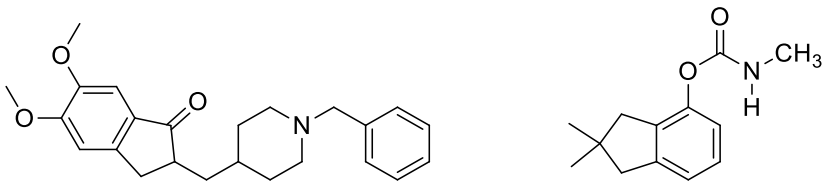


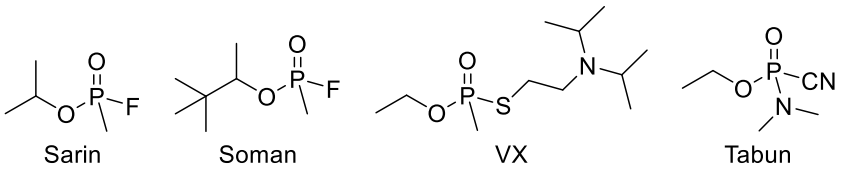
Figure I-2. 4EY8 PDB structure of AChE in complex with fasciculin.[16] Fasciculin is in red, AChE is coloured per secondary structure. Both are in ribbon representation.



Scheme I-3. Structure of donepezil and a sample carbamate

[11] P. A. Newhouse, A. Potter, E. D. Levin, *Drugs & Aging* **1997**, *11*, 206–228.
[12] M. Pohanka, *Biomed. Pap.* **2011**, *155*, 219–223.
[13] B. Rees, A. Bilwes, *Chem. Res. Toxicol.* **1993**, *6*, 385–406.
[14] A. Nordberg, A.-L. Svensson, others, *Drug Saf.* **1998**, *19*, 465–480.
[15] T. R. Fukuto, *Environ. Health Perspect.* **1990**, *87*, 245.
[16] J. Cheung, M. J. Rudolph, F. Burshteyn, M. S. Cassidy, E. N. Gary, J. Love, M. C. Franklin, J. J. Height, *J. Med. Chem.* **2012**, *55*, 10282–10286.

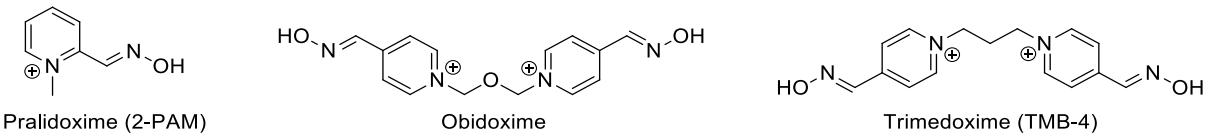
Many phosphate esters, also known as organophosphates (OP), are AChE inhibitors. This group of molecules has a high structural variety and many are commonly used as pesticides.[17] A number of organophosphates have also been used as chemical warfare agents and are known as such as neurotoxic agents or nerve agents.[18] The most important of those nerve agents are sarin, soman, VX, and tabun (see **Scheme I-4**).



Scheme I-4. Structure of common neurotoxic agents.

Those AChE inhibitors are covalent inhibitors that bind to the main catalytic residue of AChE's active site, preventing it to perform its function. The inhibition by organophosphates does not spontaneously revert.[19] When inhibited, AChE can no longer regulate the presence of ACh in the synaptic cleft which leads to the overactivation of receptors on the postsynaptic cell. This overactivation causes a large array of immediate symptoms from neurological symptoms like anxiety, confusion, failure to concentrate, convulsions, to respiratory paralysis and bradycardia.[17],[20] If one survives the immediate effects of OPs, he or she may develop organophosphate-induced delayed polyneuropathy. Polyneuropathy is a case of peripheral neuropathy affecting nerves on both sides of the body. This syndrome presents itself with a wide spectrum of symptoms of varying severity, such as numbness, ataxia (loss of limb coordination), parestesia (tingling, tickling, burning sensations), and muscle weakness; most of which remain permanent.

Fortunately, organophosphate poisoning can be treated. Three drugs are prescribed, atropine, benzodiazepine, and an oxime.[21] Atropine blocks the effects of the excess of ACh in cholinergic synapses. It acts in combination with a benzodiazepine (usually diazepam), a muscle relaxant in order to relieve the symptoms. Both those molecules treat the symptoms but do not restore AChE activity. This is the role of the oxime, usually pralidoxime, trimedoxime or obidoxime (see **Scheme I-5**)



Scheme I-5. Structure of common oxime reactivators.

[17] T. C. Marrs, *Pharmacol. Ther.* **1993**, *58*, 51–66.
[18] R. T. Delfino, T. S. Ribeiro, J. D. Figueroa-Villar, *J. Braz. Chem. Soc.* **2009**, *20*, 407–428.
[19] M. Trovaslet-Leroy, L. Musilova, F. Renault, X. Brazzolotto, J. Misik, L. Novotny, M.-T. Froment, E. Gillon, M. Loiodice, L. Verdier, et al., *Toxicol. Lett.* **2011**, *206*, 14–23.
[20] F. Worek, M. Koller, H. Thiermann, L. Szinicz, *Toxicology* **2005**, *214*, 182–189.
[21] M. Jokanović, M. Kosanović, *Environ. Toxicol. Pharmacol.* **2010**, *29*, 195–201.

Those compounds are known as reactivators; they remove the organophosphate from the active site by breaking its covalent bond with the enzyme. Thus, they restore AChE activity and proper regulation to ACh concentration in synapses. This treatment decreases the lethality of the OP inhibitor and the immediate symptoms. However known oximes have average activity at best and marked differences appear in the potency of reactivators depending on the nature of the inhibiting OP.[22] Tabun especially proves very resistant to reactivation.

II. Scope of this work

To guide the design of a new reactivator, active across most nerve agents, a better understanding of the active site of AChE must be achieved, especially its reactivity. The research presented in this manuscript was conducted using the tools of computational chemistry. The aim was to explore the key parameters during the reactivation process. The goal was also to clarify the enzymatic environment in AChE's active site, especially relative to protonation states.

[22] F. Worek, H. Thiermann, L. Szinicz, P. Eyer, *Biochem. Pharmacol.* **2004**, *68*, 2237–2248.

Chapter 1 Introduction on Acetylcholinesterase: Structure, Function, and Mechanism

Chapter 1: Introduction on Acetylcholinesterase: Structure, Function, and Mechanism

I. Structure of acetylcholinesterase

Enzymes have four structuration levels. The primary structure is the linear succession of amino acids bound by covalent amide bonds (*i.e.* peptide bonds). There are four types of AChE with different primary structures. They all share a common domain of 543 amino acids (for human and mouse AChE, 535 for *Tetronarce californica* AChE) and differ by their carboxyl-terminal region.[1] The inclusion of the C-terminal region brings the residue count to 583 and 575 respectively although it is rare to have it included in AChE crystal structures. Those are type H, type T, type R, and type S AChE. Type T exists in all vertebrates while the three other forms are not universally distributed among vertebrate species.[2] The secondary structure is the local folding of the primary structure into patterns such as α helixes or β strands. The tertiary structure is the folding of the secondary structure for instance, the formation of β sheets out of several β strands. It constitutes the three-dimensional structure of the protein. The quaternary structure is the assembly of one or several protein subunits to form the protein complex (sometimes called molecular form in the literature) encountered in vivo. An homomeric quaternary structure is a quaternary structure exclusively composed of identical subunits. When different subunits are involved the protein complex is heteromeric. The protein complex of AChE is not always a homomer of catalytic subunits, other elements can be part of the complex. Those elements can be polypeptides, such as collagen, or non peptidic, such a fatty acids, lipids or carbohydrates.[3] There are several protein complexes of AChE. All are described in details in two reviews by Taylor[4] or Massoulié *et al.*[5]. Those articles are briefly summarized in the following paragraph (see **Table I-1** and **Figure I-1**).

It should be noted that throughout this manuscript, the references are numbered starting from 1 at every chapter. Articles discussed in multiple chapters will be referenced again in every new chapter. The purpose was to avoid the reader to have to search references throughout the manuscript.
The numbering of figures, tables, schemes, and equations starts from 1 in every new section in each chapter. The section number is included in roman numerals in the name. Cross-chapter discussion of figures, tables, schemes, and equations is kept to a minimum.

[1] M. Schumacher, S. Camp, Y. Maulet, M. Newton, K. MacPhee-Quigley, S. S. Taylor, T. Friedmann, P. Taylor, *Nature* **1986**, *319*, 407–409.
[2] J. Massoulié, *Neurosignals* **2002**, *11*, 130–143.
[3] K. Méflah, S. Bernard, J. Massoulié, *Biochimie* **1984**, *66*, 59–69.
[4] P. Taylor, *J. Biol. Chem.* **1991**, *266*, 4025–4028.
[5] J. Massoulié, L. Pezzementi, S. Bon, E. Krejci, F.-M. Vallette, *Prog. Neurobiol.* **1993**, *41*, 31–91.

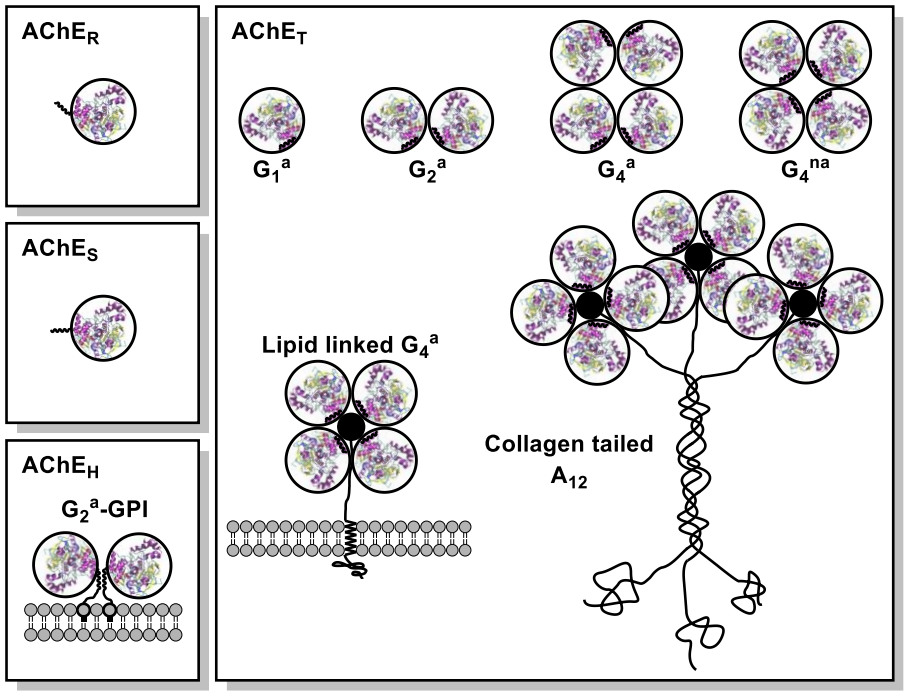


Figure I-1. Representation of the various protein complexes of AChE. (This figure is a revised version of the figure that inspired it and that can be found in reference [2])

AChE presents itself either in a heteromeric or a homomeric form. In the heteromeric form catalytic subunits are associated either to collagen-like subunits or to a lipid. The protein complexes with a collagen-like tail are called the asymmetric form because they are the only non-globular forms of AChE. The collagen tail is formed by the triple helical association of three collagenic subunits. Each of those subunits can bind one tetramer of catalytic AChE subunits, leading to complexes containing one (A_4), two (A_8) or three tetramers (A_{12}). For the heteromeric lipid linked G_4^a subunits a tetramer of catalytic subunits is bound to fatty acids by covalent bonds. Those fatty acids tether the tetramer of catalytic subunits to the surface of the cell. All the heteromeric forms of AChE have the type T catalytic subunit which is the most common. The homomeric forms of AChE are differentiated by the type of catalytic subunits, their quaternary organisation, and their solubility. The glycophospholipid linked G_2^a-GPI is the only protein complex to have type H catalytic subunits. Their C-terminal region binds to a glycophospholipid that anchors them to the cell membrane. The Amphiphilic G_1^a, G_2^a and G_4^a protein complexes are respectively mono-, di- and tetramers of type T subunits. They are easily solubilized in water but can also aggregate in the presence of detergent. The non-amphiphilic G_4^{na} protein complex is a fully water soluble tetramer of type T subunit free of any structural element. Both type S and R subunits produce water soluble monomers free of structural elements. A_4, A_8 and A_{12} protein complexes are the regular AChE expression in the synaptic clef of neuromuscular synapses.[6],[7] In central nervous system synapses, G_4^a type protein complexes are the regular expression of AChE.[8]

[6] I. Silman, J. L. Sussman, *Curr. Opin. Pharmacol.* **2005**, *5*, 293–302.

[7] E. Krejci, F. Coussen, N. Duval, J. M. Chatel, C. Legay, M. Puype, J. Vandekerckhove, J. Cartaud, S. Bon, J. Massoulié, *EMBO J.* **1991**, *10*, 1285–1293.

[8] A. L. Perrier, J. Massoulié, E. Krejci, *Neuron* **2002**, *33*, 275–285.

Table I-1. Main characteristics and composition of AChE *protein complexes*

Protein complex	Organisation of catalytic subunits	Structural elements	Homo/ Heteromeric	Fold type	Catalytic subunit type
A_4 A_8 and A_{12}	One to 3 tetramers	collagen	Heteromeric	Asymmetric	Type T
Lipid linked G_4^a	Tetramer	Fatty acid	Heteromeric	Globular	Type T
Glycophospholipid G_2^a-GPI	Dimer	Glycophospholipid	Homomeric	Globular	Type H
Amphiphilic G_1^a	Monomer	None	Homomeric	Globular	Type T
Amphiphilic G_2^a	Dimer	None	Homomeric	Globular	Type T
Amphiphilic G_4^a	Tetramer	None	Homomeric	Globular	Type T
Non Amphiphilic G_4^{na}	Tetramer	None	Homomeric	Globular	Type T
R	Monomer	None	Homomeric	Globular	Type R
S	Monomer	None	Homomeric	Globular	Type S

In the following sections only the catalytic subunit is discussed as it is the most relevant to the activity of the enzyme. No difference is made between the R, S, T and H type as they share the same common catalytic domain.[9] Differences found in the C-terminal region has only been shown to influence the complexation of AChE.

The primary structure of AChE was first deduced from its DNA sequence in 1986 by Taylor, MacPhee-Quigley and co-workers.[1] The acetylcholinesterase sequenced in this paper is from *Tetronarce californica*, an electric ray of the Pacific ocean (often called torpedo california or californica in the literature). This work was quickly followed by a study of the profile of the disulphide bonds of AChE by the same team.[10] It revealed that AChE has three intrachain disulfide bonds, between Cys69 and Cys96 (Cys67 and Cys94 in the cited literature), Cys257 and Cys272 (Cys254 and Cys265), and between Cys409 and Cys529 (Cys402 and Cys521). In *Tetronarce californica* two extra cysteine residues can be found. Cys231 is present in its reduced form although it is replaced by a glycine residue in human and mouse AChE. The final cysteine residue identified in the study is Cys580 (Cys572) present in the C-terminal region. This residue is involved in the only interchain disulfide bond of AChE to an identical AChE peptide.[11] Along with other AChE amino acid sequences it also allowed to reveal that AChE carries an overall negative charge ranging from -7 to -11.[12],[5] Finally, in 1991, Joel L. Sussman and Israel Silman published the first three-dimensional structure of

[9] J. Massoulié, S. Bon, N. Perrier, C. Falasca, *Chem. Biol. Interact.* **2005**, *157–158*, 3–14.
[10] K. MacPhee-Quigley, T. S. Vedvick, P. Taylor, S. S. Taylor, *J. Biol. Chem.* **1986**, *261*, 13565–13570.
[11] O. Lockridge, S. Adkins, B. N. L. Du, *J. Biol. Chem.* **1987**, *262*, 12945–12952.
[12] B. P. Doctor, T. C. Chapman, C. E. Christner, C. D. Deal, D. M. De La Hoz, M. K. Gentry, R. A. Ogert, R. S. Rush, K. K. Smyth, A. D. Wolfe, *FEBS Lett.* **1990**, *266*, 123–127.

acetylcholinesterase in *Science*.[13] The structure has been determined by X-ray analysis to a resolution of 2.8Å.

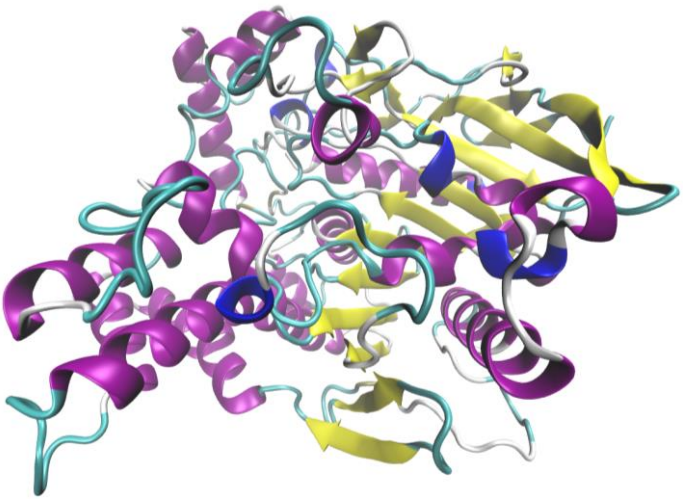


Figure I-2. 1EA5 AChE structure in ribbon representation with colouring per secondary structure.[14] α helixes are in purple, β strands in yellow, turns are in white, and the rest of the polypeptide in blue with

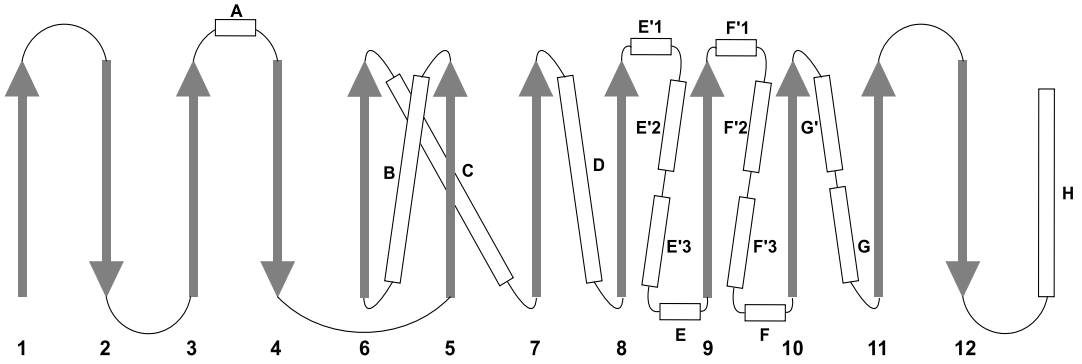


Figure I-3. Secondary structure figure showing the topology of AChE. β strands are represented as grey arrows, numbered per position in the central β sheet starting from the C-terminal. α helixes are represented as white rods, labelled per position in the same central β sheet. When several α helixes are positioned between β strands they share the same letter and are distinguished by the following number, except the last helix which only carries the letter. The labelling is made to fit the Figure in the 1991 Sussman paper the present figure is inspired from.

The X-ray data revealed the secondary and tertiary structure of AChE. Being composed of alternating α helixes and β strands, acetylcholinesterase classifies as an α/β protein fold type.[15] In the case of AChE, we have 12 β strands and 15 α helixes. All 12 β strands form a central β sheet with the first and last two β strands forming β hairpin loops (see **Figure I-2**). As

[13] J. L. Sussman, M. Harel, F. Frolow, C. Oefner, A. Goldman, L. Toker, I. Silman, *Science* **1991**, *253*, 872–879.

[14] H. Dvir, H. L. Jiang, D. M. Wong, M. Harel, M. Chetrit, X. C. He, G. Y. Jin, G. L. Yu, X. C. Tang, I. Silman, et al., *Biochemistry* **2002**, *41*, 10810–10818.

[15] a) M. Levitt, C. Chothia, *Nature* **1976**, *261*, 552–558. b) D. L. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman, J. Schrag, et al., *Protein Eng. Des. Sel.* **1992**, *5*, 197–211.

is common in such type of protein fold, the β strands are mostly parallel, with 7 consecutively parallel strands (5 to 11) and only 2, 4 and 12 breaking from the trend (See **Figure I-3**).

An additional observation was the concentration of potentially negatively charged residues such as glutamates and aspartates in the same hemisphere of the enzyme as the channel entrance. This observation stirred additional investigations that showed a strong dipolar moment for AChE.[16] In addition to the information collected on the general structure of AChE this first X-Ray structure allowed to have a precise geometry for the active site and revealed it to be buried at the bottom of a channel. Those specific substructures are detailed in the following sections.

The numbering used for residues in this manuscript is the numbering for mammalian AChE, commonly used for both human and mouse AChE structures. Starting from the original AChE structure by Sussman, many AChE crystals have been formed with *Tetronarce californica* AChE. Those structures have a slightly different numbering even though both enzymes, from mammals and from the electric ray, are highly similar in structure. In this work, the mammal numbering is used and when discussing the literature, the numbering is adjusted to avoid confusion.

I.1. Active site

The first model of the active site of AChE based on kinetic studies accounted for two subsites.[17] The "esteratic" subsite contains the catalytic residues needed for the breakdown of ACh into an acetic acid and a choline. Early kinetic studies showed that ACh has greater binding to AChE than a non-charged equivalent, 3,3-dimethylbutyl acetate.[18] This establishes the possible presence of a second subsite, an "anionic" subsite, with a choline binding role (See **Figure I-4**).

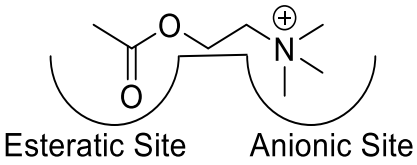


Figure I-4. Schematic representation of the active site of AChE.

A number of studies allowed to better define the active site of AChE into three subsites: a catalytic triad, and oxyanionic hole and a catalytic anionic subsite. All the residues in those three subsites have been identified using a combination of chemical marking, enzyme inhibition, selective mutations, and structural analysis (See **Figure I-5**). Those subsites, their residues, and the investigation methods used to uncover them is described in the following paragraphs.

[16] D. Porschke, C. Créminon, X. Cousin, C. Bon, J. Sussman, I. Silman, *Biophys. J.* **1996**, *70*, 1603–1608.

[17] D. Nachmansohn, I. B. Wilson, in *Advances in Enzymology and Related Areas of Molecular Biology* (Ed.: F.F. Nord), John Wiley & Sons, Inc., **1951**, pp. 259–339.

[18] D. H. Adams, V. P. Whittaker, *Biochim. Biophys. Acta* **1950**, *4*, 543–558.

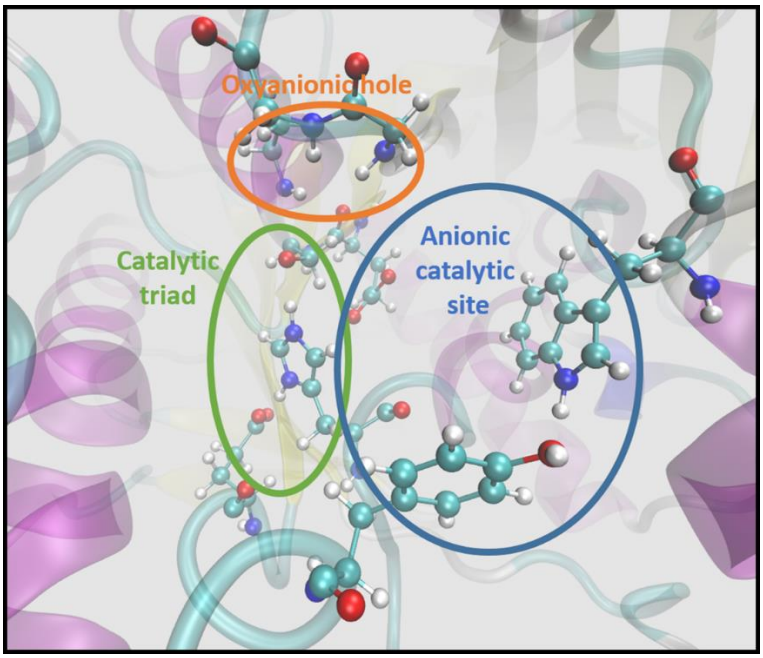


Figure I-5. Active site of AChE with its main residues highlighted in Ball&Stick representation and regular chemical element colouring (carbons in light lbue)

The main catalytic residue of the enzyme has long been established to be a serine. This residue, along with other putative catalytic residues, is present in the esteratic site. It was latter shown that the active site serine is Ser203 in the amino acid sequence of the enzyme (numbered 200 in the original article because it is the relevant numbering for *Tetronarce californica* AChE).[19] This was accomplished by marking the catalytic serine with a fluorophosphate.

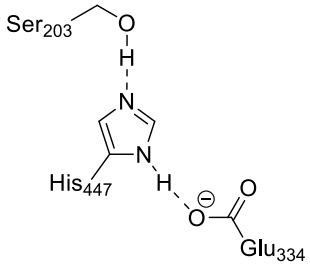
In 1974, the inhibition by ethoxyformic anhydride and subsequent reactivation by hydroxylamine suggested that a histidine was involved in the active site.[20] Although the involvement of the histidine could be merely structural, serine hydrolases are known to possess a catalytic histidine in their active site. Sequence comparison with other serine hydrolases and site directed mutagenesis have allowed to designate His447 as the catalytic histidine (numbered 440 in the cited literature as per *Tetronarce Californica* numbering).[21]

The comparison with other serine hydrolases suggests that AChE possesses a catalytic triad, composed of a serine, Ser203, a histidine, His447, and an aspartate or a glutamate. In X-ray crystallography, the amino acid placed is placed along the electron density map. The side chains are then positioned according to the electron density map. The chain tracing, that is the amino acid position in the tertiary structure, of the 1991 structure[13] placed Ser203, His447 and a glutamate, Glu334 close together. The side chain of the three residues form a planar array reminiscent of the catalytic triad of other serine proteases (See **Scheme I-1**).[22] The selective mutation of residues of the catalytic triad to alanine residues yields properly

[19] K. MacPhee-Quigley, P. Taylor, S. Taylor, *J. Biol. Chem.* **1985**, *260*, 12185–12189.
[20] R. Roskoski, *Biochemistry* **1974**, *13*, 5141–5144.
[21] G. Gibney, S. Camp, M. Dionne, K. MacPhee-Quigley, P. Taylor, *Proc. Nat. Ac. Sc. USA* **1990**, *87*, 7546–7550.
[22] T A Steitz, and R. G. Shulman, *Annu. Rev. Biophys. Bioeng.* **1982**, *11*, 419–444.

secreted and folded but yet inactive AChE.[23] This evidence cements the existence and role of the triad in AChE.

Such catalytic triad is heavily conserved among other serine protease enzymes such as *Geotrichum candidum* lipase[24], Bovine Pancreatic Cholesterol Esterase[25], Human Carboxylesterase 1[26] or p-nitrobenzyl esterase[27] with an aspartate replacing the glutamate in some cases. It belongs to a greater pattern of catalytic triads in which a nucleophile such as a serine or a threonine is activated by a base. This base is usually assisted, structurally and electrostatically, by an acidic residue like an aspartate of a glutamate.[28] In AChE, His447 acts as the base and deprotonates Ser203 in order to activate it. It is assisted in that role by Glu334, unprotonated and thus negatively charged, which stabilizes the charge of doubly protonated His447.[29]



Scheme I-1. Residues of the catalytic triad of AChE

Molecular dynamics have been used to show that while the hydrogen bond between Glu334 and His447 is very stable, the hydrogen bond between His447 and Ser203 is formed only part of the time.[30] This data is consistent with some crystallographic evidence of a possible secondary position for the histidine although in a different context than regular catalysis.[31] Proton NMR studies were also used to demonstrate the strength of the hydrogen bond between Glu334 and His447.[32]

The catalytic triad alone is not sufficient to explain the very high catalytic power of AChE. The existence of an oxyanionic hole was first alluded to by Sussman and his co-workers in their *Science* article in 1991.[13] It is composed of the NH moiety of the peptide bonds

[23] A. Shafferman, C. Kronman, Y. Flashner, M. Leitner, H. Grosfeld, A. Ordentlich, Y. Gozes, S. Cohen, N. Ariel, D. Barak, *J. Biol. Chem.* **1992**, *267*, 17640–17648.

[24] J. D. Schrag, Y. Li, S. Wu, M. Cygler, *Nature* **1991**, *351*, 761–764.

[25] J. C.-H. Chen, L. J. W. Miercke, J. Krucinski, J. R. Starr, G. Saenz, X. Wang, C. A. Spilburg, L. G. Lange, J. L. Ellsworth, R. M. Stroud, *Biochemistry (Mosc.)* **1998**, *37*, 5107–5117.

[26] V. A. de Souza, D. J. Scott, J. E. Nettleship, N. Rahman, M. H. Charlton, M. A. Walsh, R. J. Owens, *PLoS One* **2015**, *10*, e0143919.

[27] B. Spiller, A. Gershenson, F. H. Arnold, R. C. Stevens, *Proc. Nat. Ac. Sc. USA* **1999**, *96*, 12305–12310.

[28] G. Dodson, A. Wlodawer, *Trends Biochem. Sci.* **1998**, *23*, 347–352.

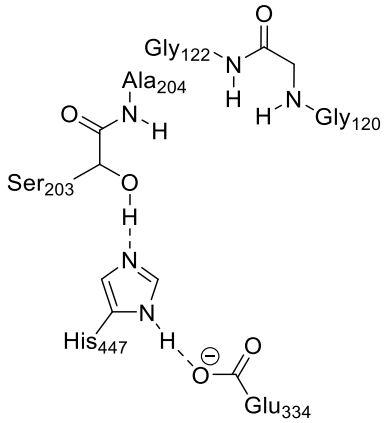
[29] A. V. Nemukhin, S. V. Lushchekina, A. V. Bochenkova, A. A. Golubeva, S. D. Varfolomeev, *J. Mol. Model.* **2008**, *14*, 409–416.

[30] J. Kua, Y. Zhang, J. A. McCammon, *J. Am. Chem. Soc.* **2002**, *124*, 8260–8267.

[31] A. Hörnberg, E. Artursson, R. Wärme, Y.-P. Pang, F. Ekström, *Biochem. Pharmacol.* **2010**, *79*, 507–515.

[32] M. A. Massiah, C. Viragh, P. M. Reddy, I. M. Kovach, J. Johnson, T. L. Rosenberry, A. S. Mildvan, *Biochemistry* **2001**, *40*, 5682–5690.

between Gly120 and Gly121, between Gly121 and Gly122 and between Ser203 and Ala204 (as shown in **Scheme I-2**). An oxyanion hole is a common sub-unit in serine protease enzymes.[22],[33] Its role is the stabilization of reaction intermediates. It improves the catalytic power of the enzyme (*i.e.* k_{cat}) by lowering the reaction barrier.



Scheme I-2. Catalytic triad and oxyanionic hole

The actual role of the oxyanionic site of AChE was first investigated by molecular modelling. In a 1993 *J. Am. Chem. Soc.* article, Quinn and co-workers used the Tripos force field to optimize the structure of the complex between AChE and acetylthiocholine (ACTh). [34] The interaction of the carbonyl with the oxyanionic hole that was observed suggested a role in the catalytic power of AChE. This putative oxyanionic hole was directly observed interacting with a transition state analogue in a crystal structure obtained by the team of Sussman in 1996 (See **Figure I-6**).[35] This crystal structure was produced by making an inhibitor, m-(N,N,N-trimethylammonio)-2,2,2-trifluoroacetophenone (TMTFA), react with AChE. The oxyanionic hole forms three donating hydrogen bonds with the inhibitor with the following hydrogen to oxygen lengths: 1.96Å, 2.12Å, and 2.64Å. The N-H-O angles are 158.64°, 142.14°, and 118.98° respectively. By analogy it is expected that the oxyanionic hole forms similar bonds with the tetrahedral intermediate during regular catalysis.

[33] F. K. Winkler, A. D'Arcy, W. Hunziker, *Nature* **1990**, *343*, 771–774.

[34] T. Selwood, S. R. Feaster, M. J. States, A. N. Pryor, D. M. Quinn, *J. Am. Chem. Soc.* **1993**, *115*, 10477–10482.

[35] M. Harel, D. M. Quinn, H. K. Nair, I. Silman, J. L. Sussman, *J. Am. Chem. Soc.* **1996**, *118*, 2340–2346.

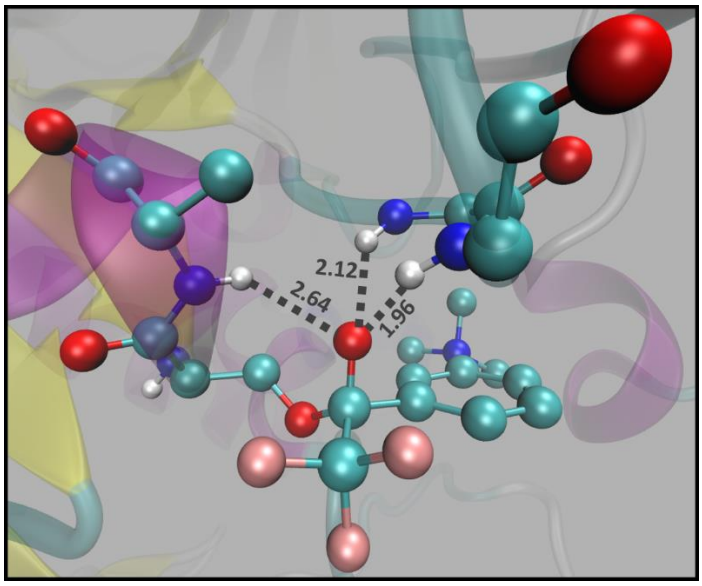


Figure I-6. 1AMN crystal structure. AChE with covalent inhibitor m-(N,N,N-trimethylammonio)-2,2,2-trifluoroacetophenone bound to the catalytic serine. The oxygen of the carbonyl is involved in three hydrogen bonds. Distances shown are in Ångström.

The Shafferman group used selective mutations to investigate the role of the oxyanionic hole. [36] Three mutant enzymes were produced with either Gly120, Gly121, or Gly122 selectively mutated to an alanine residue. A large variety of covalent and non-covalent inhibitors were used on those mutated enzymes, including transition state analogue TMTFA. The results show a drop in the reactivity of the mutated enzyme towards TMTFA, while the crystal structure shows that the mutations alters the structure of the oxyanion hole but maintains the overall geometry of the active site. It confirms the role of the oxyanion hole in AChE reactivity towards inhibitors. It is also one more evidence to its importance in the catalytic activity of AChE.

The catalytic triad and oxyanionic hole subsites are directly involved in AChE's activity. The third active site subsite, the catalytic anionic site (sometimes abbreviated to CAS) is involved in ligand binding. Molecular complementariness between AChE and inhibitors was used to show that this subsite is 5Å away from the esteratic subsite.[37] It binds quaternary ligands such as edrophonium[37], and N-methylacridinium[38]. At first, the anionic binding site was believed to be constituted of negatively charged residues, hence the name. This hypothesis was supported by kinetic studies at various ionic strengths that showed a concentration of negative charge near the binding site of AChE.[39] The comparison of the binding of many different ligand, including uncharged ones, coming from various studies later contradicted this data by indicating a largely uncharged and lipophilic subsite.[40] This

[36] A. Ordentlich, D. Barak, C. Kronman, N. Ariel, Y. Segall, B. Velan, A. Shafferman, *J. Biol. Chem.* **1998**, *273*, 19509–19517.
[37] I. B. Wilson, C. Quan, *Arch. Biochem. Biophys.* **1958**, *73*, 131–143.
[38] G. Mooser, D. S. Sigman, *Biochemistry* **1974**, *13*, 2299–2307.
[39] H.-J. Nolte, T. L. Rosenberry, E. Neumann, *Biochemistry* **1980**, *19*, 3705–3711.
[40] S. G. Cohen, E. Salih, M. Solomon, S. Howard, S. Bano Chishti, J. B. Cohen, *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1989**, *997*, 167–175.

controversy was addressed in a 1987 review by Daniel Quinn.[41] Evidence of at least one tryptophan residue in the vicinity of the binding site was provided by the quenching of tryptophan fluorescence bands by non-covalent inhibitor. [42] A tryptophan was also labelled p-dimethylaminobenzene diazonium fluoroborate in the binding site.[43] This data came with the evidence that quaternary nitrogen preferentially binds with aromatic groups through π interactions.[44] X-ray analysis was used to identify the residues involved in this catalytic anionic site. AChE crystals were soaked in solution of decamethonium, tacrine and edrophonium.[45] X-ray analysis of those crystals showed the ligands in close interaction with two residues, Trp86 and Tyr337 (Trp84 and Phe330 in the cited literature as per *Tetronarce Californica* numbering see **Figure I-7**). It is important to note that in *Tetronarce Californica* residue 330 is a phenylalanine, while both in human and mouse AChE the corresponding residue in the chain, residue 337, is a tyrosine.

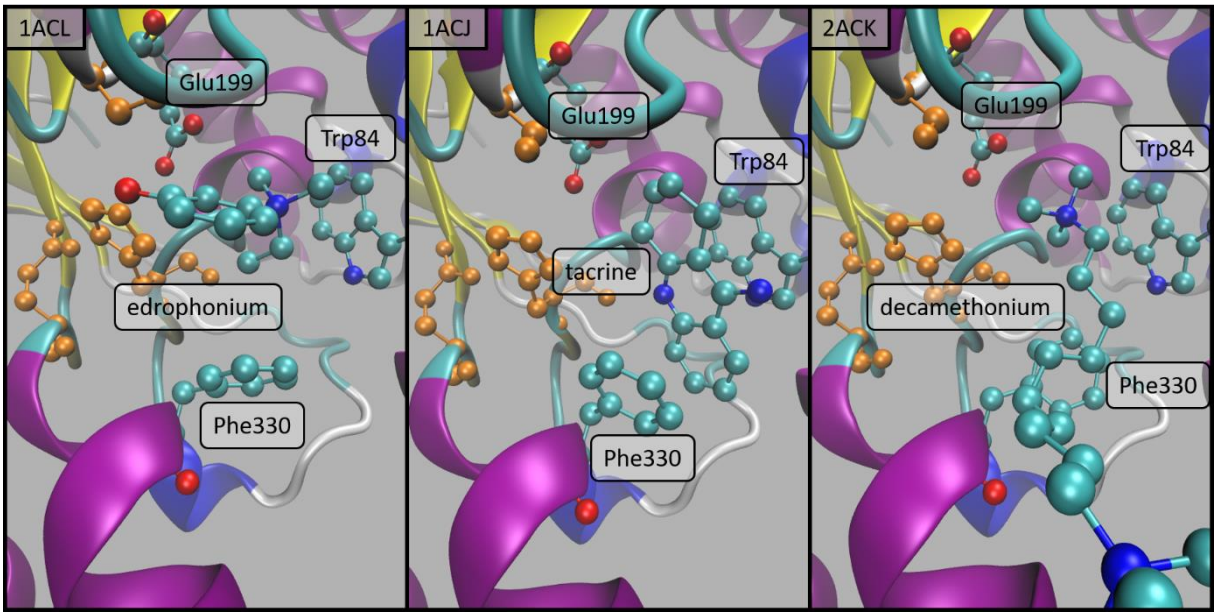


Figure I-7. Active site of AChE with catalytic triad (Ser200, His440, Glu327), Phe330, Glu199 and Trp84, and a ligand in Ball&Stick representation. The catalytic triad is coloured in orange for clarity. 1ACL: PDB structure with edrophonium docked in the CAS. 1ACJ: PDB structure with tacrine docked in the CAS. 2ACK: PDB structure with decamethonium docked in the CAS.

In structures 1ACL and 2ACK the ligand has a positively charged nitrogen. In both crystal structures the ammonium is observed close to Trp86. The third structure, 1ACJ, shows tacrine with its two aromatic cycles sandwiched between the indole ring of Trp84 (Trp86) and the phenyl ring of Phe330 (Tyr337). Structures 1ACL and 2ACK also show, although it is not discussed in the original paper, that Glu202 is in close proximity with the ammonium. A similar observation was made by Sussman and co-workers in the 1991 Science paper.[13] The

[41] D. M. Quinn, *Chem. Rev.* **1987**, *87*, 955–979.
[42] M. Shinitzky, Y. Dudai, I. Silman, *FEBS Lett.* **1973**, 30, 125–128.
[43] M. P. Goeldner, C. G. Hirth, *Proc. Nat. Ac. Sc. USA* **1980**, *77*, 6439–6442.
[44] D. A. Dougherty, D. A. Stauffer, *Science* **1990**, *250*, 1558–1560.
[45] M. Harel, I. Schalk, L. Ehret-Sabatier, F. Bouet, M. Goeldner, C. Hirth, P. H. Axelsen, I. Silman, J. L. Sussman, *Proc. Nat. Ac. Sc. USA* **1993**, *90*, 9031–9035.

ammonium of the choline moiety was found to be near Glu202 and Trp86. A mutation study showed that a Trp86Ala mutation heavily diminishes the affinity (660 fold increased K_M, K_M being inversely related to the affinity) and halves the reaction rate (the k_{cat}) of AChE towards a charged acetylcholine equivalent.[46] With uncharged analogues of ACh, the Trp86Ala mutation leaves the K_M unchanged but still diminishes the k_{cat}. The mutation of Tyr337 to an alanine has the same effect on both the charged and uncharged analogues of ACh, the affinity is not affected but the reaction rate diminishes. Several studies have been made on the effect of a mutation of Glu202.[47], [21] All show that an enzyme mutated at this position has a slightly higher K_M value and a lowered k_{cat}. Those results confirm that both Tyr337 and Glu202 have a role in substrate binding. They also both stabilize reaction intermediate and contribute to the high catalytic power of AChE. Trp86 contributes little to the catalytic power compared to its immense role in substrate binding. A mutation of this residue leads to a loss of affinity for ACh by an order of magnitude of 10^4. Only mutations of residues of the catalytic triad have such a dramatic effect on the activity of AChE. In the catalytic anionic site Trp86 has a dominant contribution to molecular recognition.[48] This predominant role of Trp86 leads to an increased importance of surrounding residues that could play a role in the positioning of its side chain. Tyr133, located behind Trp86 can be mutated to a phenylalanine with only marginal effect on affinity and the reaction rate. But a Tyr133Ala mutation represents a substantial affinity loss toward anionic substrates.[49] The combination of docking studies and molecular mechanics allowed to further detail the contributions of electrostatics and shape/size to the docking of the ammonium tail of ACh in AChE.[50] The cancellation of the partial charges on either of the three residues of the catalytic anionic site only marginally decrease the binding energy of the ammonium tail. The position and orientation of the tail, however, is modified and it leads to poorer docking of the acetyl head of ACh. Removing the charge on Glu202 decreases the docking energy and the proportion of fully docked (head and tail) ACh the most. The individual mutation of the residues of the catalytic anionic site to alanine residue decreases the docking energy to a much greater degree and disrupts the docking of both the ammonium tail and the acetyl head. This research indicates that the electrostatic is key, albeit not the main contribution, to the docking of the ligand in the catalytic anionic site.

I.2. Enzyme channel of AChE

One of the most noteworthy features of AChE is the fact that the active site is a buried one. It is located at the end of a 20Å deep channel lined with aromatic residues.[13] At the entrance of this channel is positioned a peripheral binding subsite called the peripheral anionic

[46] A. Ordentlich, D. Barak, C. Kronman, Y. Flashner, M. Leitner, Y. Segall, N. Ariel, S. Cohen, B. Velan, A. Shafferman, *J. Biol. Chem.* **1993**, *268*, 17083–17095.

[47] A. Shafferman, B. Velan, A. Ordentlich, C. Kronman, H. Grosfeld, M. Leitner, Y. Flashner, S. Cohen, D. Barak, N. Ariel, *EMBO J.* **1992**, *11*, 3561–3568.

[48] D. M. Quinn, S. R. Feaster, H. K. Nair, N. A. Baker, Z. Radić, P. Taylor, *J. Am. Chem. Soc.* **2000**, *122*, 2975–2980.

[49] A. Ordentlich, D. Barak, C. Kronman, N. Ariel, Y. Segall, B. Velan, A. Shafferman, *J. Biol. Chem.* **1995**, *270*, 2082–2091.

[50] J. Kua, Y. Zhang, A. C. Eslami, J. R. Butler, J. A. McCammon, *Protein Sci.* **2003**, *12*, 2675–2684.

site (sometimes referred as PAS in the literature). Finally, a backdoor into the active site of AChE has long been supposed to exist.

Early kinetic data introduced the possibility of a second anionic site.[51] This study showed that bisquaternary inhibitors like stilbamidine inhibit AChE more efficiently than nonquaternary ones. Indeed, the concentration needed to reach 50% inhibition of a solution of AChE is a thousand times lower for stilbamidine than for tetramethylammonium. It was hypothesized that the bisquaternary inhibitor would bind simultaneously to both binding sites. This idea was corroborated by several studies of bisquaternary inhibitors[52],[53] and eventually demonstrated X-Ray data showing decamethonium binding with both sites.[45] Several other inhibitors were used to gather data on that second binding site. Eserine was shown to bind to AChE, and thus inhibit it, without impeding the binding of ACh.[54] Atropine is another molecule which binds to AChE although contrary to eserine it does not inhibit it. Both react differently to a covalent inhibitor. While covalent inhibition on the triad does not prevent atropine binding it prevents the binding of eserine.[55], [56]

While the 1991 crystal structure was an important step into the identification of residues involved in the peripheral anionic site, the comparison of AChE to another similar enzyme gave the first tangible data. Butyrilcholinesterase (BuChE) is a cholinesterase enzyme that shows a lot of similarities with AChE. It does not share neither the propensity of AChE to be inhibited by large concentrations of substrate nor its peripheral anionic site. Sequence comparison of AChE with BuChE allowed for the identification of Trp286 (Trp279 in some studies), a residue close to the entrance of the channel, absent from BuChE structure.[57], [47] The mutation of Trp286 to Alanine decreases inhibition by peripheral anionic site specific ligands like propidium.[58] It does not affect catalytic anionic site ligands like edrophonium. The other components of the peripheral anionic site were identified by mutations studies to be Tyr124, Tyr72 and to a lesser degree Glu285.[59] This research showed that individual mutation would only slightly decrease the binding of propidium but multiple mutations of those residues would have a much larger effect, up to 100-fold. It was hypothesized that the role of Glu285, positioned behind Trp286 and not on the surface of AChE could be to polarize the π system of the tryptophan. A crystal structure obtained in 2003 allows to visualize clearly the interaction of the peripheral active site with a ligand (see **Figure I-8**).

[51] F. Bergmann, I. B. Wilson, D. Nachmansohn, *Biochim. Biophys. Acta* **1950**, *6*, 217–224.
[52] J.-P. Changeux, *Mol Pharmacol* **1966**, *2*, 369–392.
[53] P. Taylor, S. Lappi, *Biochemistry* **1975**, *14*, 1989–1997.
[54] J.-P. Changeux, W. Leuzinger, M. Huchet, *FEBS Lett.* **1968**, *2*, 77–80.
[55] G. Kato, J. Yung, M. Ihnat, *Biochem. Biophys. Res. Commun.* **1970**, *40*, 15–21.
[56] G. Kato, E. Tan, J. Yung, *Nature* **1972**, *236*, 185–185.
[57] J. L. Sussman, I. Silman, *Curr. Opin. Struct. Biol.* **1992**, *2*, 721–729.
[58] M. Harel, J. L. Sussman, E. Krejci, S. Bon, P. Chanal, J. Massoulié, I. Silman, *Proc. Nat. Ac. Sc. USA* **1992**, *89*, 10827–10831.
[59] D. Barak, C. Kronman, A. Ordentlich, N. Ariel, A. Bromberg, D. Marcus, A. Lazar, B. Velan, A. Shafferman, *J. Biol. Chem.* **1994**, *269*, 6296–6305.

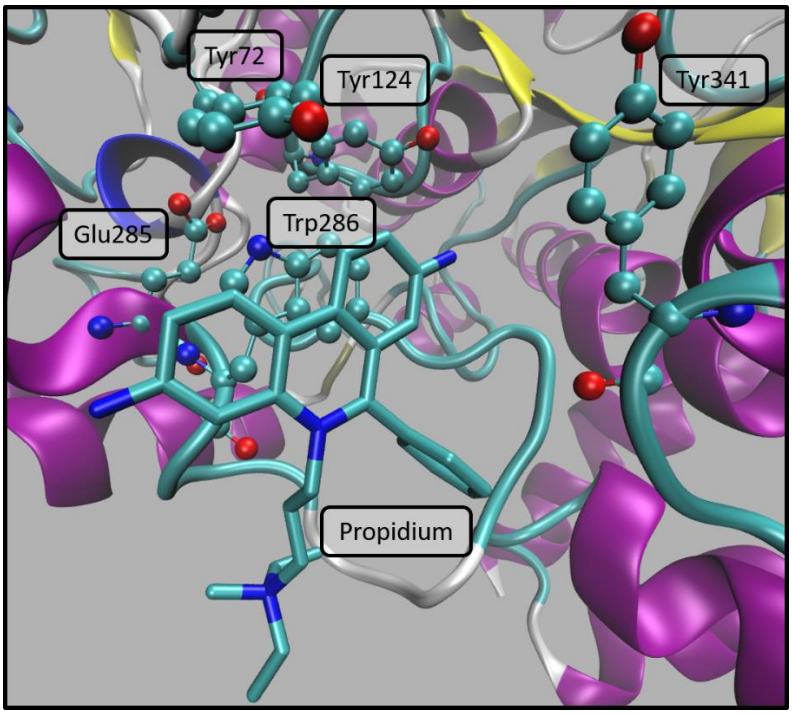


Figure I-8. Peripheral anionic site of AChE with propidium docked in. Tyr72, Try124, Glu285, Trp286 and Tyr 341 ar labelled and in Ball&Stick representation. Propidium is also labelled but in Licorice representation. 1N5R crystal structure.

Molecular simulations as well as docking studies show that while Tyr72, Tyr124 and Tyr341 form an aromatic cage of peripheral active site ligand, the π- π or cation- π interaction with Trp86 is the driving force for the binding of ligands on the PAS.[60]

The observation of the 20Å deep channel in 1991 showed multiple aromatic residues. The role of three phenylalanines present in the channel was then quickly investigated.[46], [58] The data showed that Phe295 (Phe288), Phe297 (Phe290), and Phe338 (Phe331) stabilize covalent and non-covalent complexes in the active site of AChE (See **Figure I-9**). In addition to this role, Phe295 and Phe297 (but not Phe338) are responsible for the specificity of AChE toward ACh. The mutation of those two residues to less bulky amino acids increases the activity of AChE towards bulkier substrates like butyrylcholine. Finally, double Phe295Ala Phe338Ala mutants have a severely decreased catalytic activity.[61] X-Ray data show that when both are mutated to smaller less structurally constraining residues, His447 is destabilized. Those two residues have thus a structural role to maintain the position of the crucial element of the catalytic triad. This structural role is was also highlighted in several additional studies and is not specific to those residues but rather to any residue in proximity with the triad and

[60] a) A. Cavalli, G. Bottegoni, C. Raco, M. De Vivo, M. Recanatini, *J. Med. Chem.* **2004**, *47*, 3991–3999. b) D. Branduardi, F. L. Gervasio, A. Cavalli, M. Recanatini, M. Parrinello, *J. Am. Chem. Soc.* **2005**, *127*, 9147–9155.

[61] D. Barak, D. Kaplan, A. Ordentlich, N. Ariel, B. Velan, A. Shafferman, *Biochemistry* **2002**, *41*, 8245–8252.

especially His447 which has been shown to be a very mobile residue.[49],[62],[63] Those three aromatic residues of the channel are often designated in the literature as the acyl pocket.

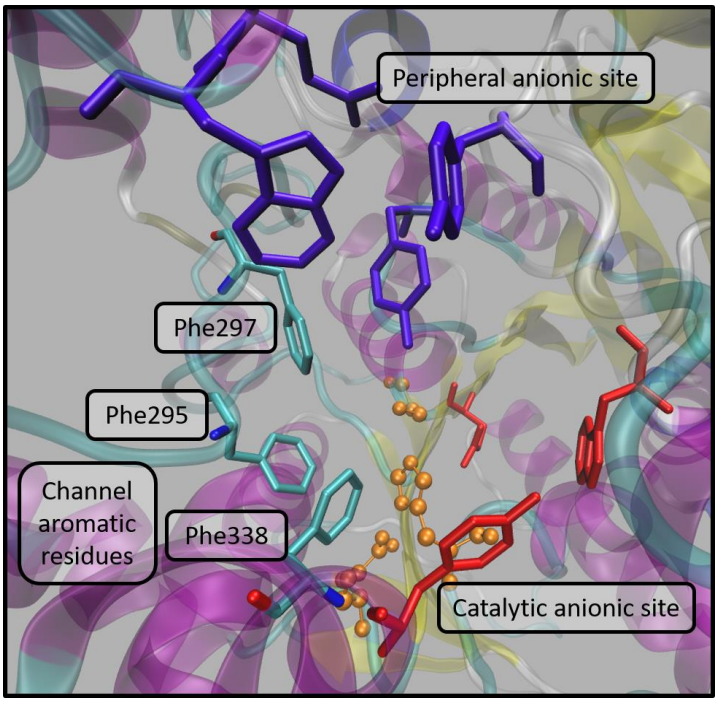


Figure I-9. Enzyme channel of AChE (3DL7 PDB structure) in ribbon representation with secondary structure colouring. The catalytic triad is represented in Ball&Stick representation in orange. The catalytic and peripheral anionic site are both represented in licorice representation in violet and red respectively. Aromatic residues of the channel, Phe295, Phe297 and Phe338 are in licorice representation and labelled.

A possible allosteric mechanism has long been postulated to exist in AChE. In the study of the peripheral anionic site it appeared that mutation of Trp86 in the catalytic anionic site would increase resistance to propidium inhibition in the peripheral active site without affecting binding of propidium.[46] That particular aspect of propidium inhibition can also be reproduced without mutating Trp86 but by a mutation of Asp74.[59] This aspartate is located at the entrance of AChE's channel opposite to the location of the peripheral anionic site. The hypothesis emerges throughout several studies that Asp74 and Tyr341 link the peripheral anionic site to the catalytic anionic site and are necessary for the allosteric effect observed when ligands are bound to the PAS.[47], [64] It was also hypothesized that this putative allosteric effect could also be explained by a steric blockade of the channel by ligands bound to the peripheral active site.[65] Brownian dynamics simulation also suggested that this aspartate, Asp74, could be involved in transferring ligands bound to the peripheral anionic site to the

[62] A. Shafferman, D. Barak, D. Kaplan, A. Ordentlich, C. Kronman, B. Velan, *Chem. Biol. Interact.* **2005**, *157–158*, 123–131.
[63] F. Ekström, C. Akfur, A.-K. Tunemalm, S. Lundberg, *Biochemistry* **2006**, *45*, 74–81.
[64] a) Z. Radić, P. D. Kirchhoff, D. M. Quinn, J. A. McCammon, P. Taylor, *J. Biol. Chem.* **1997**, *272*, 23265–23277. b)G. V. D. Ferrari, W. D. Mallender, N. C. Inestrosa, T. L. Rosenberry, *J. Biol. Chem.* **2001**, *276*, 23282–23287.
[65] T. Szegletes, W. D. Mallender, T. L. Rosenberry, *Biochemistry* **1998**, *37*, 4206–4216.

enzyme channel instead of participating in binding at the PAS. [66] It has been hypothesized that the allosteric effect involving Asp74 and Trp86 could be linked to a larger motion of the secondary structure containing both residues.[67] This structure is the Cys69-Cys96 cyclized Ω loop. This loop walls AChE's channel (See **Figure I-10**) and participates not only in the gating of the channel, but also in the opening of two alternate routes to the active site of AChE, the side door and the backdoor.

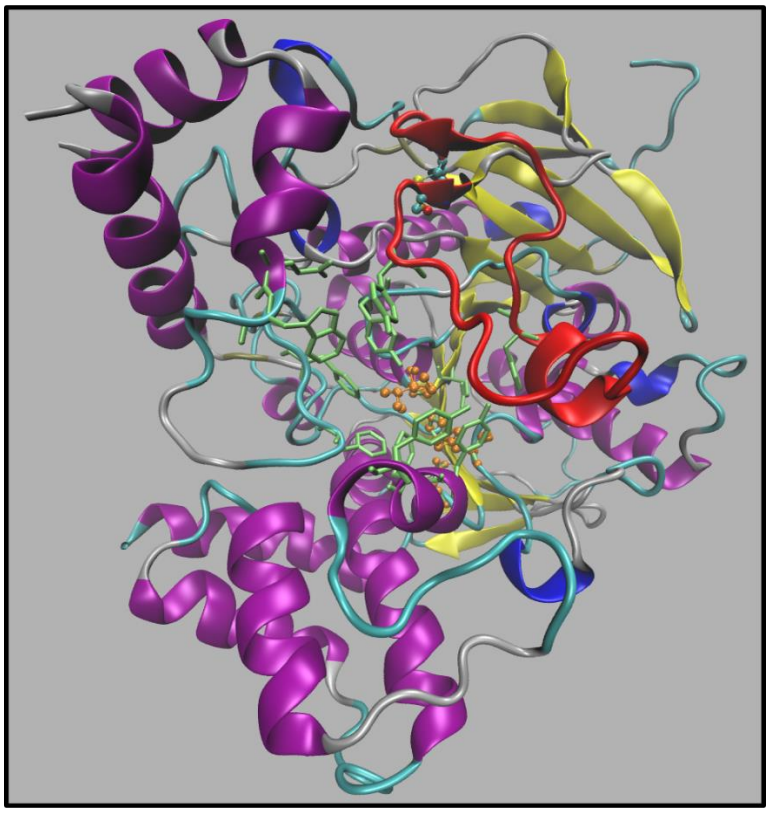


Figure I-10. AChE in ribbon representation and secondary structure colouring. Main residues of the channel (PAS, CAS, catalytic triad) are in licorice representation coloured in lime. The Cys69-Cys96 Ω loop is represented in red with both cysteines in Ball&Stick representation.

Several studies were made into the Ω loop structure and its role in AChE's activity. The individual mutations of residues of the loop involved in possible hydrogen bonds to the rest of the enzyme or conserved across cholinesterases have little effect on the catalytic activity. Only the deletion of five residues, residues 78 to 82, produces a displaced Trp86 and only slightly decreases the affinity of AChE for ACh by a magnitude of one.[68] In another research Gly80 and Val431 (*Tetronarce californica* numbering) were both mutated to cysteines to form

[66] S. Tara, A. H. Elcock, P. D. Kirchhoff, J. M. Briggs, Z. Radic, P. Taylor, J. A. McCammon, *Biopolymers* **1998**, *46*, 465–474.

[67] D. Barak, A. Ordentlich, A. Bromberg, C. Kronman, D. Marcus, A. Lazar, N. Ariel, B. Velan, A. Shafferman, *Biochemistry* **1995**, *34*, 15444–15452.

[68] B. Velan, D. Barak, N. Ariel, M. Leitner, T. Bino, A. Ordentlich, A. Shafferman, *FEBS Lett.* **1996**, *395*, 22–28.

a disulfide bond to restrict motions of the loop without significant change to the affinity and activity of AChE.[69] The hypothesized role of the Ω loop is the control of the size of the channel.

In 1997 fluctuations in the channel opening size were observed in molecular dynamic trajectories.[70] Opening and closing motions involving Tyr341 and Tyr124 are shown to accompany ligand clearance from the channel.[71] Further research by the same group showed that five residues were potentially involved in gating, Tyr124, Tyr341, Tyr338, Phe297 and Phe337.[72] In the same study the impact of the gating motions on ligand docking is evaluated. For a 2.4 Å probe, representing an acetylcholine sized ligand, the binding rate is divided by two, from $1.7x10^9$ $M^{-1}s^{-1}$ to $0.7x10^9$ $M^{-1}s^{-1}$, even though the channel is opened only 2.4% of the time in the MD trajectory. This is due to the high frequency of opening motions. For a 2.8Å probe, the binding rate is divided by three orders of magnitude to $0.7x10^6$ $M^{-1}s^{-1}$. This phenomenon is called dynamic selectivity.[73] in addition to those five residues involved in the gating, molecular dynamic studies have shown that the Ω loop could govern the opening of the channel by random non-concerted motions.[74]

The discovery of the buried nature of AChE's active site came with the question of ligand clearance. Electrostatic field calculations were used to evaluate the electrostatic properties of AChE and showed that electric field vectors pointed towards Glu202, Trp86 and Glu450 leading to the hypothesis made by the authors that Trp86 and Met85 could undergo conformational change to allow for the opening of a backdoor.[75] A subsequent examination by yet the same team using molecular dynamics showed transient opening of the backdoor that involved Trp86, Met85, Val132, Gly448 and Tyr449 (see **Figure I-11**).[76] The notion that the backdoor could have a significant role in AChE catalysis was refuted by two mutation studies. One showed that a saline bridge between Glu452 and Val132Lys that would block the backdoor had little effect of AChE's k_{cat} and K_M.[77] The second tested further mutations of Val132 to an arginine or even a bulky tryptophan that resulted in no change of AChE affinity or activity.[69] A multiple copy sampling study showed that the backdoor would be only be large enough for very small molecules like methane, ammonium or water. Those molecules could also fit in side door opening with the main channel entrance being the only opening large enough for larger substrates of AChE like ACh.[78] Molecular dynamic investigations of

[69] C. Faerman, D. Ripoll, S. Bon, Y. Le Feuvre, N. Morel, J. Massoulié, J. L. Sussman, I. Silman, *FEBS Lett.* **1996**, *386*, 65–71.
[70] S. T. Wlodek, T. W. Clark, L. R. Scott, J. A. McCammon, *J. Am. Chem. Soc.* **1997**, *119*, 9513–9522.
[71] S. Tara, A. H. Elcock, P. D. Kirchhoff, J. M. Briggs, Z. Radic, P. Taylor, J. A. McCammon, *Biopolymers* **1998**, *46*, 465–474.
[72] H.-X. Zhou, S. T. Wlodek, J. A. McCammon, *Proc. Nat. Ac. Sc. USA* **1998**, *95*, 9280–9283.
[73] T. Shen, K. Tai, R. H. Henchman, J. A. McCammon, *Acc. Chem. Res.* **2002**, *35*, 332–340.
[74] a) J. Shi, K. Tai, J. A. McCammon, P. Taylor, D. A. Johnson, *J. Biol. Chem.* **2003**, *278*, 30905–30911. b) A. E. Boyd, C. S. Dunlop, L. Wong, Z. Radić, P. Taylor, D. A. Johnson, *J. Biol. Chem.* **2004**, *279*, 26612–26618.
[75] D. R. Ripoll, C. H. Faerman, P. H. Axelsen, I. Silman, J. L. Sussman, *Proc. Nat. Ac. Sc. USA* **1993**, *90*, 5128–5132.
[76] M. K. Gilson, T. P. Straatsma, J. A. McCammon, D. R. Ripoll, C. H. Faerman, P. H. Axelsen, I. Silman, J. L. Sussman, *Science* **1994**, *263*, 1276–1278.
[77] C. Kronman, A. Ordentlich, D. Barak, B. Velan, A. Shafferman, *J. Biol. Chem.* **1994**, *269*, 27819–27822.
[78] Daniel Van Belle, L. De Maria, G. Iurcu, S. J. Wodak, *J. Mol. Biol.* **2000**, *298*, 705–726.

the backdoor led to the discovery of a side door, present in the wall of the channel near its entrance and composed entirely of residues from the Ω loop. [79] Val73, Asp74, Pro78, Gly79, Thr83, Glu84, Asn87, and Pro88 have all been supposedly involved to a greater or lesser extend in several openings in the Ω loop wall. Those openings are large enough for water molecules and small ligands (see **Figure I-11**).[80] Those articles also mention the backdoor. In addition to the already mentioned residues, molecular dynamic simulations show Trp86, Met85, Val132, Gly448 and Tyr449; Glu202 and Ile451 to be involved in backdoor opening motions.

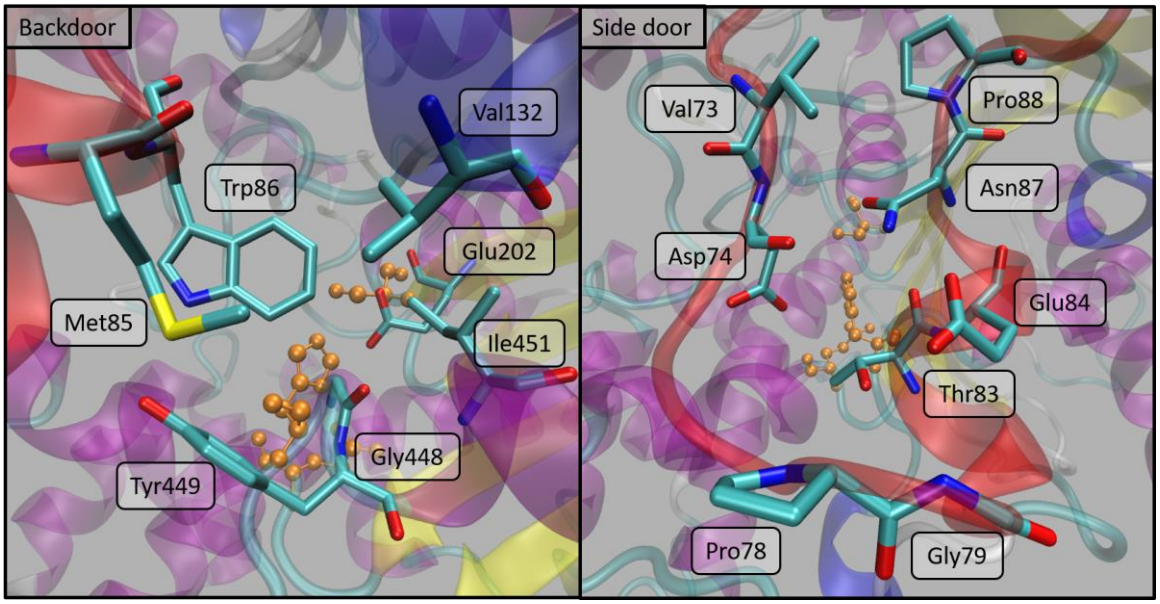


Figure I-11. Alternate routes to the active site with AChE in ribbon representation and secondary structure colouring. The omega loop is in red and the catalytic triad in CPK representation in orange. Relevant residues to both alternate routes are labelled.

As more simulations were conducted, the opening of the backdoor appears to be a very rare and short lived event in the case of free AChE but to increase in frequency when AChE is bound by fasciculin, huperzine A and tacrine.[81],[73] Those simulations are validated by two experimental evidences. First when AChE is covalently inhibited by 8-(*cis*-2,6-dimethylmorpholino)octylcarbamyoleseroline, the leaving group, eseroline, is not found in the active site of AChE in the X-ray structure.[82] This is an indication of the existence of a backdoor because this inhibitor completely obstructs the channel preventing leaving group clearance through the main channel entrance. Secondly enzymes bound by fasciculin, which

[79] a) S. T. Wlodek, T. W. Clark, L. R. Scott, J. A. McCammon, *J. Am. Chem. Soc.* **1997**, *119*, 9513–9522. b) S. Tara, V. Helms, T. P. Straatsma, J. A. McCammon, *Biopolymers* **1999**, *50*, 347–359

[80] a) S. Tara, V. Helms, T. P. Straatsma, J. A. McCammon, *Biopolymers* **1999**, *50*, 347–359. b) K. Tai, T. Shen, U. Börjesson, M. Philippopoulos, J. A. McCammon, *Biophys. J.* **2001**, *81*, 715–724.

[81] a) S. Tara, T. P. Straatsma, J. A. McCammon, *Biopolymers* **1999**, *50*, 35–43. b) K. Tai, T. Shen, R. H. Henchman, Y. Bourne, P. Marchot, J. A. McCammon, *J. Am. Chem. Soc.* **2002**, *124*, 6153–6161. c) J. M. Bui, J. A. McCammon, *Chem. Biol. Interact.* **2005**, *157*–158, 357–359.

[82] C. Bartolucci, E. Perola, L. Cellai, M. Brufani, D. Lamba, *Biochemistry* **1999**, *38*, 5714–5719.

completely obstructs the channel entrance, retain some catalytic activity.[83] Under the hypothesis that this residual catalytic activity could be linked to the backdoor serving as an alternate entrance, mutations of Trp86 (Trp83 in the article as per *Drosophila melanogaster* AChE numbering) to alanine or glutamate were attempted.[84] They produce an enzyme impervious to inhibition by ligands that close the main channel entrance. The protonation states used for molecular dynamic simulations of AChE are important factor to consider. As was shown in a 2010 article, the Ω loop is very mobile when Glu202 and Glu450 are simulated in their basic unprotonated form leading to a wider channel and frequent openings in the side door.[85] Oppositely when both those residues are protonated the mobility of the loop is reduced and the channel is smaller. Kinetic crystallography of an ACh equivalent showed that arsenocholine, the choline analogue used, could leave the active site either by the main channel entrance or by the backdoor.[86] A 2011 X-ray structure of an AChE dimer shows a rotated Tyr449 allowing for a large opening of the backdoor (see **Figure I-12**) in the presence of non-covalent PAS bound inhibitor aflatoxin.[87] This opening is large enough for polyethylene glycol to be at the backdoor in the crystal structure.

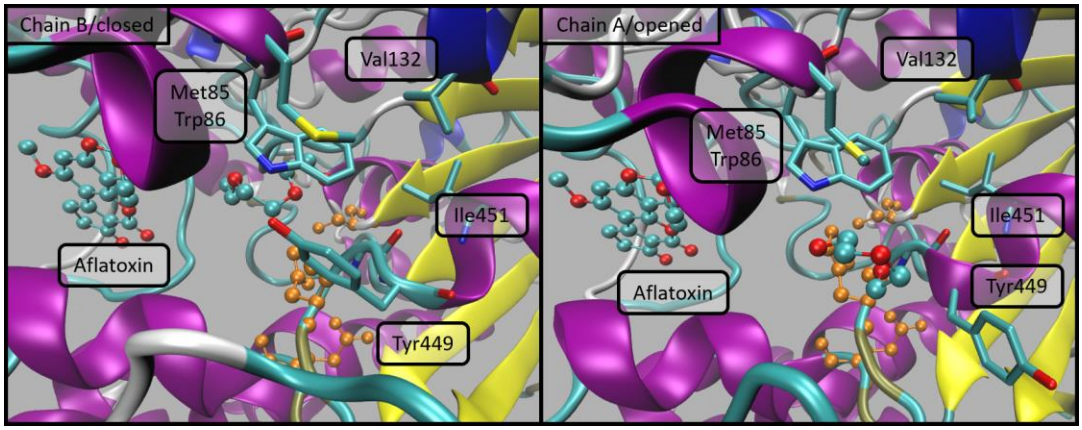
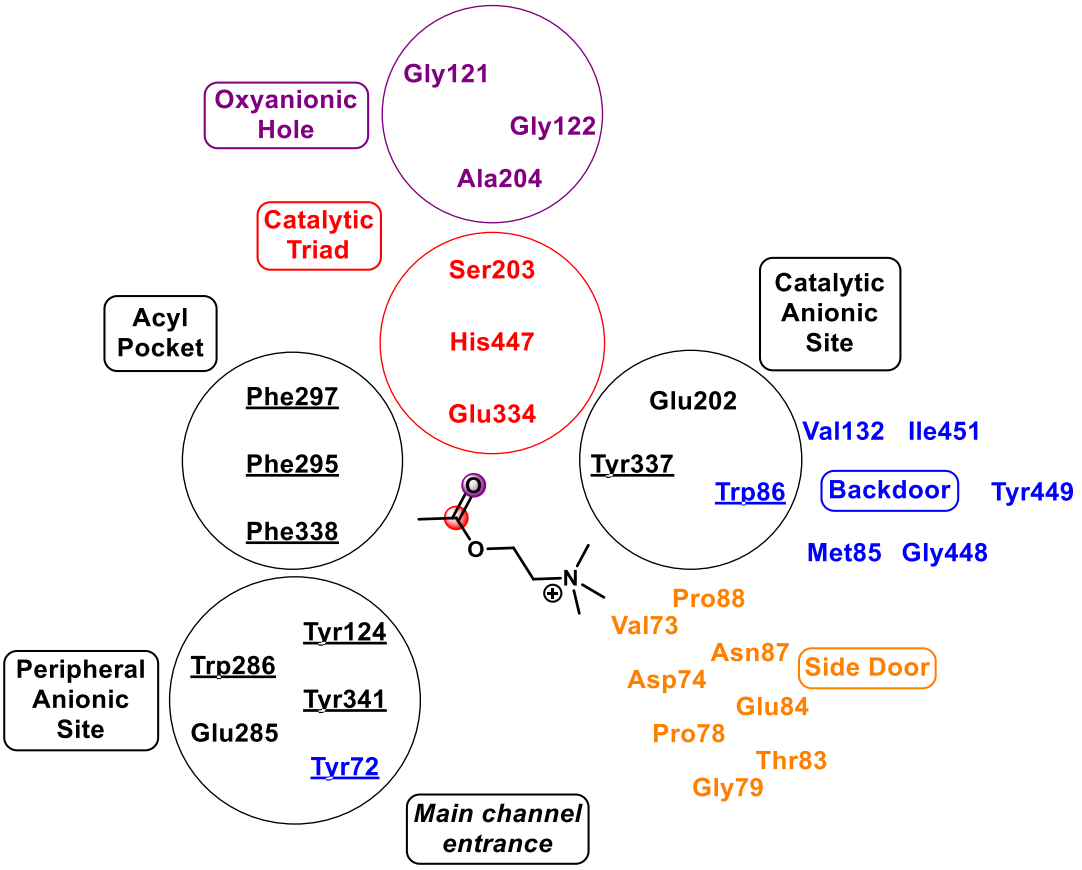


Figure I-12. 2XI4 PDB structure. Both in ribbon representation and secondary structure colouring. The catalytic triad is in Ball&Stick representation in orange, the backdoor residues are in licorice representation. Polyethylene glycol is at the backdoor in Ball&Stick representation like Aflatoxin in the enzyme channel.

From the previous sections on the active site and on the enzyme channel, it is evident that while only a handful of residues are crucial to the catalytic power of AChE, a larger number of residues are involved to a greater or lesser extend in the activity of this enzyme as parts of various substructures with different roles. Those residues and the structures and substructures they belong to are summarized in **Scheme I-3** bellow.

[83] a) Z. Radić, D. M. Quinn, D. C. Vellom, S. Camp, P. Taylor, *J. Biol. Chem.* **1995**, *270*, 20391–20399. b) T. L. Rosenberry, C.-R. Rabl, E. Neumann, *Biochemistry* **1996**, *35*, 685–690.
[84] F. Nachon, J. Stojan, D. Fournier, *FEBS J.* **2008**, *275*, 2659–2664.
[85] J. Wiesner, Z. Kříž, K. Kuča, D. Jun, J. Koča, *J. Biomol. Struct. Dyn.* **2010**, *28*, 393–403.
[86] J.-P. Colletier, A. Royant, A. Specht, B. Sanson, F. Nachon, P. Masson, G. Zaccai, J. L. Sussman, M. Goeldner, I. Silman, et al., *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2007**, *63*, 1115–1128.
[87] B. Sanson, J.-P. Colletier, Y. Xu, P. T. Lang, H. Jiang, I. Silman, J. L. Sussman, M. Weik, *Protein Sci.* **2011**, *20*, 1114–1118.



Scheme I-3. AChE residues grouped by subunit. Aromatic residues of the channel are underlined. Residues involved in the backdoor and in the side door are in blue and orange respectively. In red and purple are the residues of the catalytic triad and the oxyanionic hole respectively. Residues surrounded by a circle belong to a stable and recognised substructure.

This scheme allows to visualize the continuous string of aromatic residues that link the channel entrance and its nearby peripheral active site with the bottom of the channel. Another way to represent AChE in a simplified way is shown in **Figure I-13**. This figure is a reproduction of **Figure I-3** with the addition of the main residues and structures discussed above.

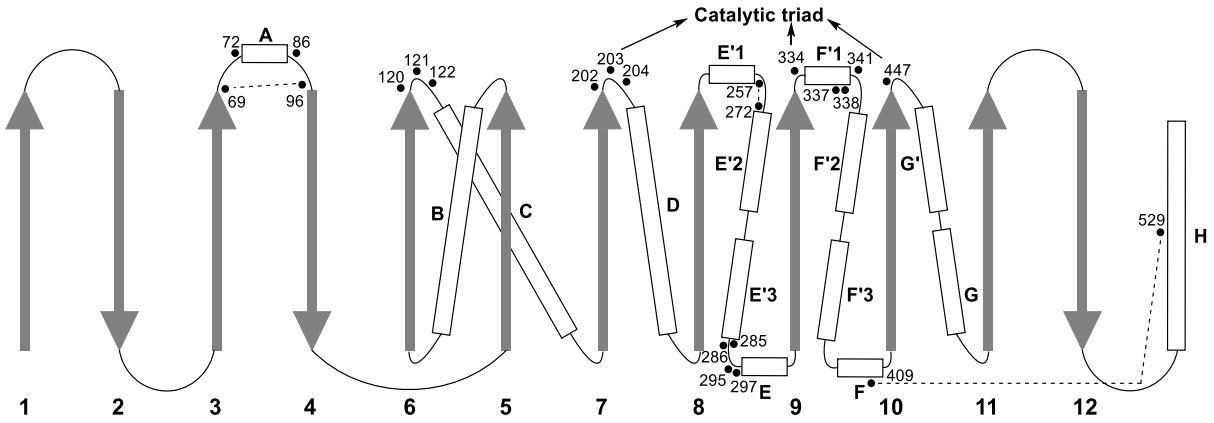


Figure I-13. Secondary structure figure showing the topology of AChE. β strands are represented as grey arrows, numbered per position in the central β sheet starting from the C-terminal. α helixes are represented as white rods, labelled per position in the same central β sheet. When several α helixes are positioned between β strands they share the same letter and are distinguished by the following number, except the last helix which only carries the letter. Important AChE residues are located as dots and disulfide bonds are shown as dashed lines.

Now that the main structure of the enzyme and the various substructures have been described, we have all the elements to describe AChE's catalytic activity.

II. Activity of acetylcholinesterase

II.1. Regular catalytic activity

Acetylcholinesterase catalyses the hydrolysis of acetylcholine into an acetate and a choline. This process is decomposed in two separate reactions, the acylation of AChE and the follow up deacylation that restores AChE into its original state. This was first hypothesized in 1950[88] and then proven in 1984 using tritium marked ACh.[89] This reaction is illustrated in **Figure II-1**.

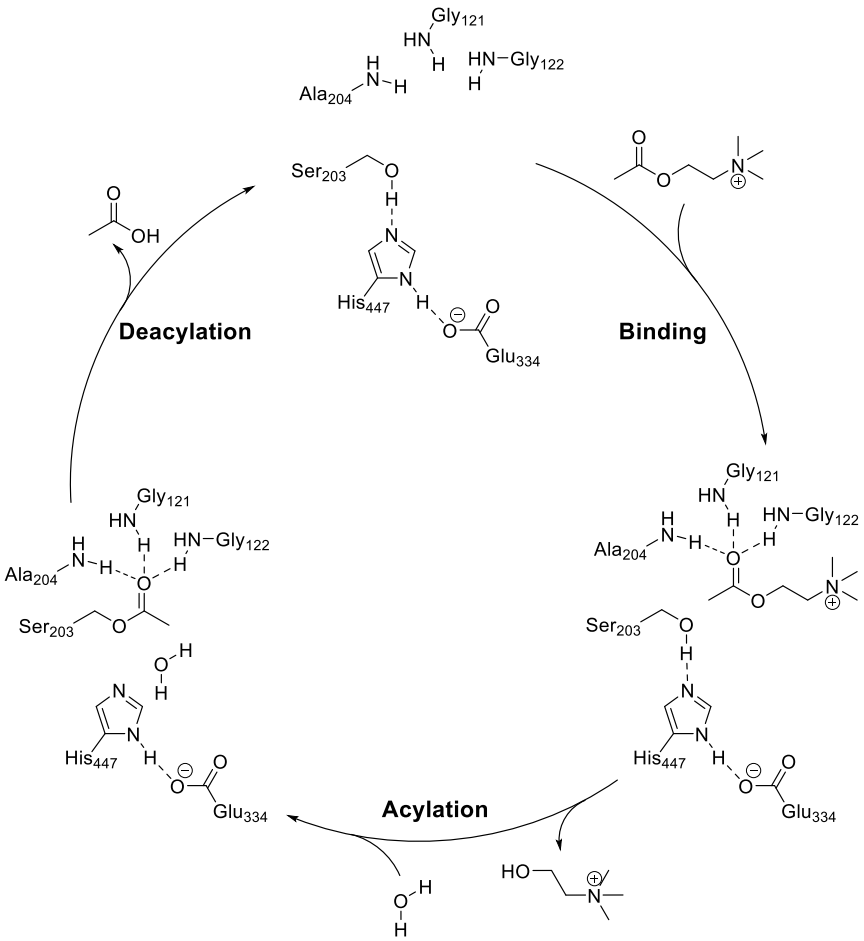


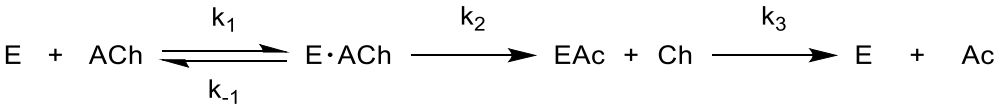
Figure II-1. Catalytic cycle of ACh hydrolysis by AChE

Acetylcholine enters the active site of AChE and reacts with the nucleophilic side chain of Ser203. The nucleophilicity of Ser203 is increased by its deprotonation by His447. His447 is assisted by the carboxylate side chain of Glu334. Glu334 was shown to fulfil that role by stabilizing doubly protonated His447 by electrostatic interactions and not a proton transfer

[88] I. B. Wilson, F. Bergmann, D. Nachmansohn, *J. Biol. Chem.* **1950**, *186*, 781–790.
[89] H. C. Froede, I. B. Wilson, *J. Biol. Chem.* **1984**, *259*, 11010–11013.

from the histidine to the glutamate as demonstrated by Arieh Warshel in 1989.[90] The introduction of this article summarizes very well the heated debate on this specific issue.

Acetylcholinesterase activity toward AChE can be modelled by Michaelis-Menten kinetic under the steady state approximation.[91],[92] It is represented by the reaction equations in **Scheme II-1** with ACh as acetylcholine, Ac as acetic acid, Ch as choline, E as acetylcholinesterase (only in discussing kinetics as to not mistake ACh and AChE), E•ACh as the enzyme substrate complex and EAc as the acylated AChE.



Scheme II-1. ACh hydrolysis catalysis equation following the Michaelis-Menten model

In this equation, the first step is the formation of the enzyme substrate complex. k_1 is the rate constant of complex formation, k_{-1} is the rate constant of complex dissociation. The second step is the acylation step with rate constant k_2 and the third step is the deacylation with rate constant k_3. These constants can be used to define k_{cat} and K_M in the reaction rate v as a function of ACh concentration equation (**II-I**). Equations **II-II** and **II-III** are obtained with the King Altman method.[93]

$$v = \frac{k_{cat}[E]_0[S]}{K_M + [S]} \qquad \textbf{II-I}$$

$$k_{cat} = \frac{k_2 k_3}{k_3 + k_2} \qquad \textbf{II-II}$$

$$K_M = \frac{k_3}{k_3 + k_2} K'_M = \frac{k_3(k_{-1} + k_2)}{k_1(k_3 + k_2)} \qquad \textbf{II-III}$$

$$k_s = \frac{k_{cat}}{K_M} = \frac{k_1 k_2}{k_{-1} + k_2} \qquad \textbf{II-IV}$$

The K_M (defined by equation **II-III**) is defined as a product of K'_M, the dissociation constant for the enzyme substrate complex with the ratio of deacylation rate constant over the sum of acylation and deacylation rate constant. This formulation accounts for the accumulation of acylated AChE when k_2 is superior to k_3. Finally, the specificity constant is defined by equation **II-IV** as the product of k_{cat} by the inverse of K_M.

[90] A. Warshel, G. Naray-Szabo, F. Sussman, J. K. Hwang, *Biochemistry* **1989**, *28*, 3629–3637.

[91] T. L. Rosenberry, *Proc. Nat. Ac. Sc. USA* **1975**, *72*, 3834–3838.

[92] a) L. Michaelis, M. L. Menten, *Biochem. Z.* **1913**, *49*, 333–369.b) K. A. Johnson, R. S. Goody, *Biochemistry* **2011**, *50*, 8264–8269.

[93] E. L. King, C. Altman, *J. Phys. Chem.* **1956**, *60*, 1375–1378.

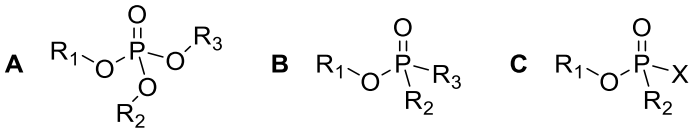
The k_{cat} of ACh hydrolysis is of the order of 10^4 s^{-1} with a K_M of the order of 10^{-4} M which gives a specificity constant k_s of 10^8 $M^{-1}.s^{-1}$.[39] This very high specificity constant puts AChE in the realm of diffusion limited enzymes.[94] Those enzymes catalyses the reaction with their substrates as fast or faster than the substrates reaches. Thus, they are only limited in their reaction rate by the diffusion of substrate to the enzyme. The debate on acetylcholinesterase's status as a diffusion limited enzyme steered the first efforts in simulating acetylcholinesterase. In their 1980 study, Rosenberry Nolte and Neumann hypothesize that AChE has an overall negative charge. This was supported by the sequencing of AChE's amino acid chain and the crystallisation of the enzyme.[10],[13] Both showed a large number of negatively charge residues, especially at the entrance of the enzymatic channel. Evaluation of the electrostatic potential[75] as well as electrooptical measurements[95] of AChE showed that it has a large dipole moment aligned with the channel. It was hypothesized by the authors of the study that this dipole would facilitate the approach and drive the cationic acetylcholine down the channel. Brownian dynamics simulations were used to validate that hypothesis.[96] Shafferman and co-worker refuted that hypothesis on the basis of mutation studies.[97] They demonstrated that the mutation of charged residues in the vicinity of the channel entrance to neutral ones would effectively nullify the dipole of AChE without decreasing it's catalytic effectiveness. In the introduction of the article presenting this work they argue that the catalytic effectiveness of AChE is slightly lower that what I would be if it was diffusion limited. Multiple Brownian dynamics and molecular dynamic studies from the group of J. Andrew MacCammon showed that placed in the hypothesis of a diffusion limited reaction, those simulations could reproduce experimental results including the mutation study of Shafferman and co-workers.[98] In an article published in 1998, Fuxreiter and Warshel make the point that the origin of the catalytic power of AChE is in it's reactivity.[99] They base their argument of the very high rate acceleration AChE provides to the hydrolysis of ACh. The reaction of ACh with an hydroxide in water has a rate constant of 2.2 $M^{-1}.s^{-1}$ which implied an acceleration of the rate by 8 orders of magnitude. This article will be analysed in detail along with other significant theoretical studies on AChE in a following section of this manuscript.

II.2. Inhibition by organophosphates

Organophosphorus compounds are organic (i.e. carbon containing) compounds with a phosphorus. This group is composed of a lot of very diverse compounds and can split into subgroups. Organophosphates, one of those subgroups, are organic esters derived from

[94] R. A. Alberty, G. G. Hammes, *J. Phys. Chem.* **1958**, *62*, 154–159.
[95] D. Porschke, C. Créminon, X. Cousin, C. Bon, J. Sussman, I. Silman, *Biophys. J.* **1996**, *70*, 1603–1608.
[96] R. C. Tan, T. N. Truong, J. A. McCammon, J. L. Sussman, *Biochemistry* **1993**, *32*, 401–403.
[97] A. Shafferman, A. Ordentlich, D. Barak, C. Kronman, R. Ber, T. Bino, N. Ariel, R. Osman, B. Velan, *EMBO J.* **1994**, *13*, 3448–3455.
[98] a) J. Antosiewicz, S. T. Wlodek, J. A. McCammon, M. K. Gilson, *Biochemistry* **1995**, *34*, 4211–4219. b) J. Antosiewicz, S. T. Wlodek, J. A. McCammon, *Biopolymers* **1996**, *39*, 85–94.c) H.-X. Zhou, J. M. Briggs, J. A. McCammon, *J. Am. Chem. Soc.* **1996**, *118*, 13069–13070. d) S. T. Wlodek, J. Antosiewicz, J. M. Briggs, *J. Am. Chem. Soc.* **1997**, *119*, 8159–8165.
[99] M. Fuxreiter, A. Warshel, *J. Am. Chem. Soc.* **1998**, *120*, 183–194.

phosphoric acid (which itself is not an organic compound). A strict definition would have organophosphates comprised of a phosphoryl group and three phosphate ester groups (see **Scheme II-2-A**). A broader definition would be that an organophosphate is comprised of a phosphoryl group and at least one phosphate ester (**Scheme II-2-B**). It could also be defined as a phosphine oxide with at least one phosphate ester substituent. Compounds belonging to this organophosphate subgroup are components of DNA, phospholipids, adenosine triphosphate and many other biologically relevant molecules. Out of all those biologically relevant molecules, many small organophosphate molecules have been identified as nerve agents.



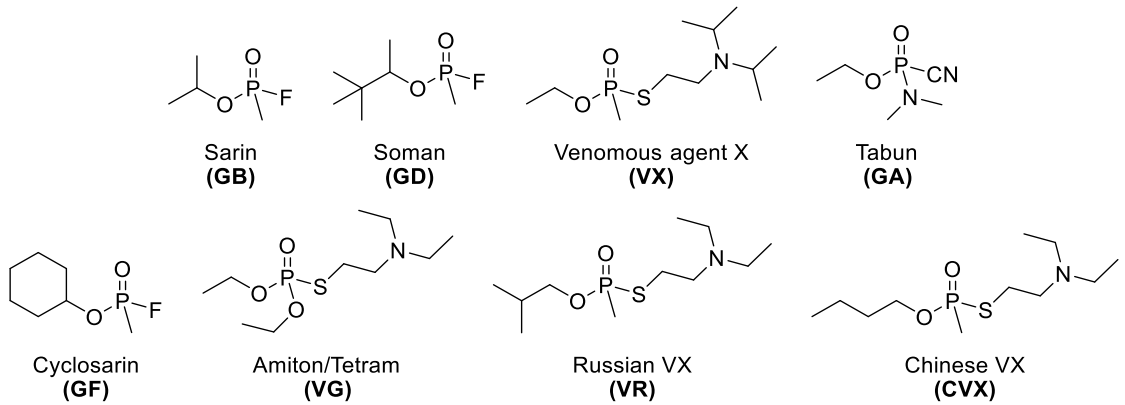
Scheme II-2. Generic structures for organophosphates (A and B) and nerve agents used as combat gases (C).

In the 1930s, the toxicity of some organophosphates (OP) on humans, and then on insects, was discovered. Soon after this discovery it was shown that the main target of those compounds is acetylcholinesterase.[100] This targeting of the nervous system lead to those compounds being designated as nerve agents. This inhibitory effect of nerve agents on AChE was exploited for use either as pesticides[101] or as combat gases[102]. The generic formula shown in **Scheme II-2-C** covers all known nerve agents used as combat gases. R_1 is an alkyl group, R_2 either an alkyl, an alkyl ester or in the case of tabun an amine; X is often a halogen, a cyanide or a thioester. The main nerve agents that have been used or designed as combat gases are in **Scheme II-3**. The most well-known and studied of those nerve agents are on the top row. It is important to note that although not all OPs are nerve agents it is very common in the literature to use the term organophosphate to designate nerve agents, especially considering that a large number of organophosphate have at least some inhibitory effect on AChE, be it in human AChE, vertebrate AChE, or invertebrate AChE.

[100] a) H. McCombie, B. C. Saunders, *Nature* **1946**, *157*, 287–289. b) J. F. Mackworth, E. C. Webb, *Biochem. J.* **1948**, *42*, 91–95.

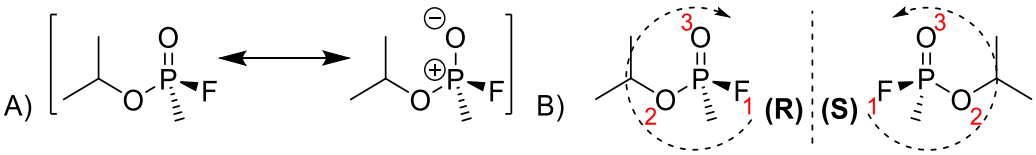
[101] J. E. Casida, G. B. Quistad, *J. Pestic. Sci.* **2004**, *29*, 81–86.

[102] C. MacIlwain, *Nature* **1993**, *363*, 3.



Scheme II-3. Structure of nerve agents used a combat gases. The common name is below the structure with the NATO identifier at the bottom.

A quick clarification of the absolute configuration of organophosphates around the phosphorus stereocenter is needed. While it is not true for all organophosphates, all nerve agents have a chiral phosphorus. A very common misconception is to treat the phosphoryl bond as a double bond. It has been shown though electron localization function, a quantum mechanical method used to analyse the electron density (ELF), that this mesomeric form is not the major representation of this bond which has a highly ionic character. This bond is thus better represented by a single bond with a single positive charge on the phosphorus and a negative charge on the oxygen (see **Scheme II-4-A**).[103] Considering the phosphoryl bond as a single bond shifts it priority behind that of the phosphate ester substituent and often reverses the absolute configuration (see **Scheme II-4-B**). Nevertheless, in the literature on AChE inhibition, aging, and reactivation, this bond is systematically represented by a double bond. In the rest of the manuscript it was decided to follow that custom.



Scheme II-4. A) mesomeric forms of sarin B) Sarin R and S enantiomers

[103] D. B. Chesnut, A. Savin, *J. Am. Chem. Soc.* **1999**, *121*, 2335–2336.

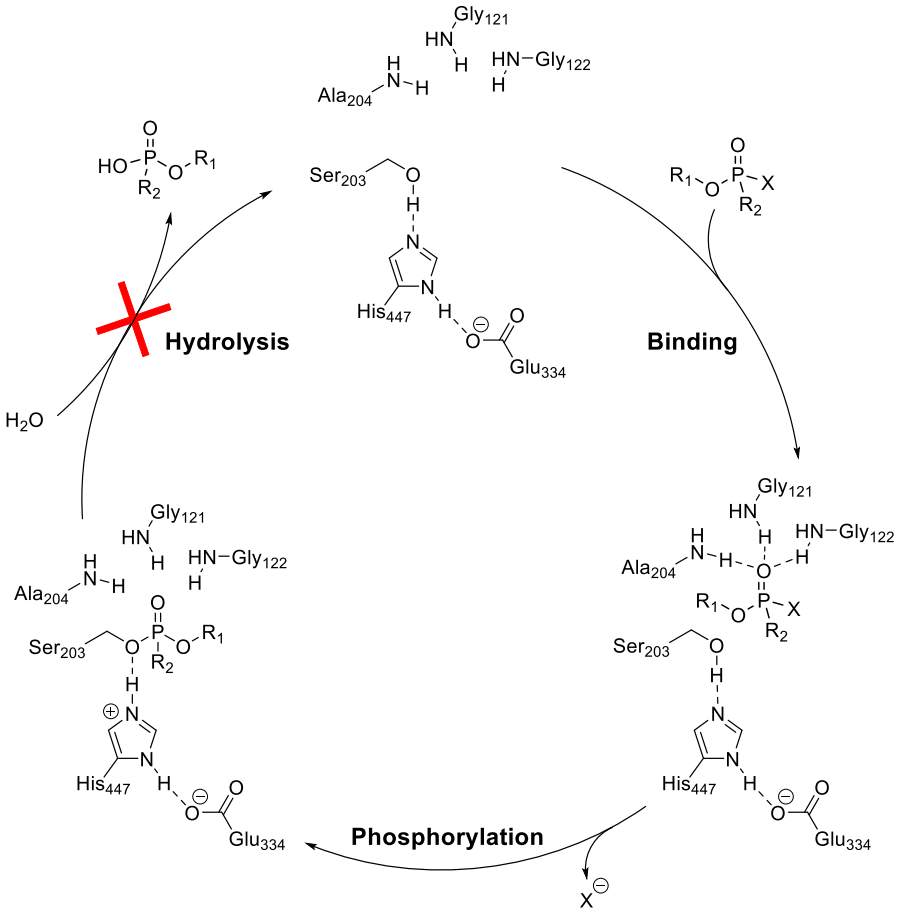


Figure II-2. Inhibition of AChE by a generic nerve agent.

Early kinetic studies suggested that the interaction between AChE and nerve agents would follow the same hydrolytic process than ACh hydrolysis by AChE (see **Figure II-2**).[104] The authors of that study suggested that the inhibition could be covalent at the esteratic site. That was later confirmed by the labelling of the catalytic serine with a nerve agent inhibitor.[19]

The inhibition of AChE is thus a two-step process. The nerve agents first reversibly bind to the active site of AChE after descending through the channel and then reacts with the catalytic serine, Ser203. This reaction is not reversible and produces the inhibited, and inactive, enzyme. It can be represented by the reaction equation of **Scheme II-5**. In this scheme, E is acetylcholine, PX the nerve agent, E•PX is the pre-reactive complex between AChE and the nerve agent, EP is the inhibited enzyme and X is the leaving group.

Scheme II-5. Reaction equation for the inhibition of AChE by a nerve agent.

Three constants are commonly used to describe the inhibition of AChE by a nerve agent. The first, the dissociation constant K_D, describes the binding of the nerve agent to the

[104] I. B. Wilson, S. Ginsburg, *Biochim. Biophys. Acta* **1955**, *18*, 168–170.

enzyme. It is defined in equation **II-V** as the ratio of the sum of rate constants of the steps breaking the bounded complex over the sum of rate constants of the steps forming the bonded complex. This definition of the K_D is preferred over the simple ration of k_{-1} over k_1 because k_2 is not significantly lower than k_1. The smaller the K_D the better the affinity between the substrate (or the inhibitor in that case) and the enzyme. The second is the phosphorylation constant k_2. Finally, it is very common to describe the whole process using bimolecular rate constant k_i as defined by equation **II-VI** as a ratio of the phosphorylation constant over the dissociation constant. It is also often called the inhibition rate constant in the literature and is convenient because of the ease with which this value can be obtained experimentally.

$$K_D = \frac{k_{-1} + k_2}{k_1} \qquad \textbf{II-V}$$

$$k_i = \frac{k_2}{K_D} = \frac{k_2 k_1}{k_{-1} + k_2} \qquad \textbf{II-VI}$$

Out of the four classical nerve agents, the most active one, characterized by the highest inhibition rate constant, is VX. The inhibition rate constant of tabun, the least active, is lower by an order of magnitude (see **Table II-1**).[105] All four nerve agents have inhibition rate constants at most one order of magnitude lower than the specificity constant of AChE towards ACh. It illustrates how potent those compounds are at AChE inhibition.

Because nerve agents are chiral, it was expected that the stereoisomers will have different toxicological properties because of the chiral environment of AChE. In 1988, H. P. Benschop and L. P. A. Jong studied the inhibition power of stereoisomers of VX, sarin, soman and tabun.[106] Their original work focuses on the optical properties of the enantiomers as associated absolute configurations had not been determined yet. Further studies showed the absolute configuration of all levorotary nerve agents to be S.[107], [108] The postulated role of chirality in the specificity of a nerve agent to AChE was confirmed as, for all nerve agents, the S enantiomers were shown to be the more active than the R enantiomers (see **Table II-1**). The inhibition constants for the racemic nerve agents are not the average of the inhibitions constants for the R and S enantiomers as the observed inhibition constant is driven by the faster reacting enantiomer. The rationalization of the regioselectivity of tabun has been provided by a previous QM/MM study in our group by Ophélie Kwasnieski.[109] Phe295 and

[105] F. Worek, H. Thiermann, L. Szinicz, P. Eyer, *Biochem. Pharmacol.* **2004**, *68*, 2237–2248.
[106] H. P. Benschop, L. P. A. De Jong, *Acc. Chem. Res.* **1988**, *21*, 368–374.
[107] N. Qian, I. M. Kovach, *FEBS Lett.* **1993**, *336*, 263–266.
[108] A. Ordentlich, D. Barak, C. Kronman, H. P. Benschop, L. P. A. De Jong, N. Ariel, R. Barak, Y. Segall, B. Velan, A. Shafferman, *Biochemistry* **1999**, *38*, 3055–3066.
[109] O. Kwasnieski, L. Verdier, M. Malacria, E. Derat, *J. Phys. Chem. B* **2009**, *113*, 10001–10007.

Phe297, the residues involved in the specificity of AChE towards ACh, have also been implicated in the specificity of AChE towards the S enantiomers of nerve agents.[110]

Table II-1. Inhibition constants of stereoisomers of the main nerve agents.

Nerve agent stereoisomer	k_i ($M^{-1}min^{-1}$)	Reference
(*S*)-VX	4×10^8	[106]
(*R*)-VX	2×10^8	[106]
(*RS*)-VX	1.2×10^8	[105]
(*S*)-Sarin	1.4×10^7	[106]
(*R*)-Sarin	$< 3 \times 10^3$	[106]
(*RS*)-Sarin	2.7×10^7	[105]
(*R, S*)-Soman	8×10^7	[108]
(*S, S*)-Soman	1.5×10^8	[108]
(*R, R*)-Soman	2×10^3	[108]
(*S, R*)-Soman	2×10^3	[108]
(*RS, RS*)-Soman	9.2×10^7	[105]
(*S*)-Tabun	2.8×10^7	[111]
(*R*)-Tabun	7.0×10^4	[111]
(*RS*)-Tabun	7.4×10^6	[105]

Once AChE is inhibited, barring any external interventions, there are two possible outcomes. The enzyme can, eventually, be spontaneously reactivated by a water molecule[112]; or the nerve agent can undergo a dealkylation of one of its substituents (see **Figure II-3**).

In the 1950's came the first observations that prolonged contact of AChE with organophosphate inhibitors would diminish the capacity of inhibited AChE to be reactivated.[113] Isotope marked organophosphates were used to identify the change that occurred in inhibited AChE.[114] It showed that while the phosphorus remained bonded to AChE, a substituent would be cleaved form the adduct. It was later showed that alkoxy substituents were subject to that reactivity and that the dealkylation would occur through cleavage of the C-O bond between the oxygen bound to the phosphorus and it alkyl chain.[115]

[110] a) A. Ordentlich, D. Barak, C. Kronman, N. Ariel, Y. Segall, B. Velan, A. Shafferman, *J. Biol. Chem.* **1996**, *271*, 11953–11962. b) A. Ordentlich, D. Barak, C. Kronman, H. P. Benschop, L. P. A. De Jong, N. Ariel, R. Barak, Y. Segall, B. Velan, A. Shafferman, *Biochemistry* **1999**, *38*, 3055–3066.

[111] O. Tenberken, H. Thiermann, F. Worek, G. Reiter, *Tox. Lett.* **2010**, *195*, 142–146.

[112] I. B. Wilson, *J. Biol. Chem.* **1951**, *190*, 111–117.

[113] F. Hobbiger, *Br. J. Pharmacol. Chemother.* **1955**, *10*, 356–362.

[114] F. Berends, C. H. Posthumus, I. V. D. Sluys, F. A. Deierkauf, *Biochim. Biophys. Acta* **1959**, *34*, 576–578.

[115] H. P. Benschop, J. H. Keijer, *Biochim. Biophys. Acta, Enzymol. Biol. Oxid.* **1966**, *128*, 586–588.

This process, also known as aging, produces an enzyme that cannot be reactivated, neither spontaneously nor by any known reactivator.

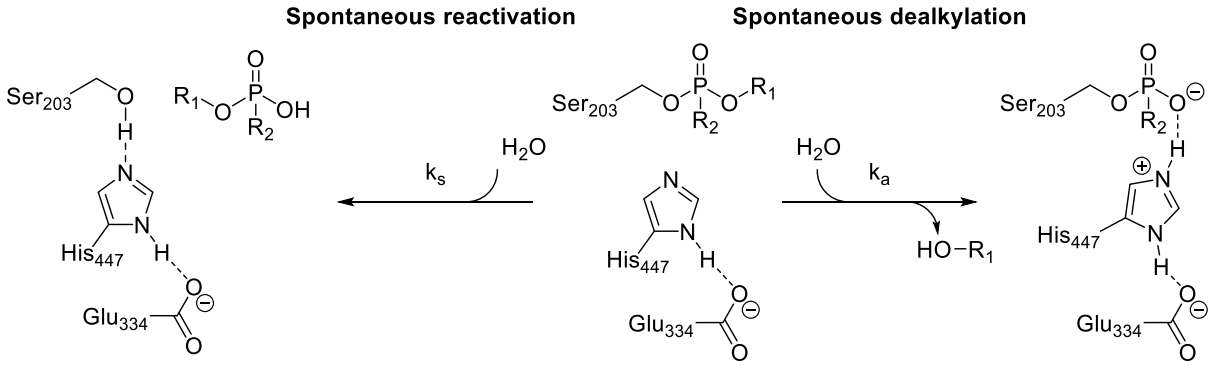


Figure II-3. Spontaneous outcomes of inhibition, dealkylation and reactivation

Both processes, spontaneous reactivation and dealkylation require a single water molecule to occur. The accessibility of the active site to water molecules from the solvent thus allows the kinetics of both reactions to be described using a single first order rate constant. This constant is k_a for the dealkylation and k_s for the spontaneous reactivation and both expressed in min^{-1} (or h^{-1} because of their extremely low rate constants). Franz Worek and his co-workers averaged several pre-existing values in a 2004 article for both processes and that data can be found in **Table II-2** bellow.[105]

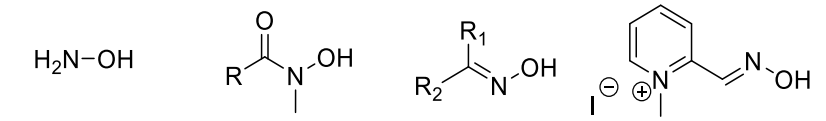
Table II-2. Spontaneous reactivation (k_s) and dealkylation (k_a) constants for all four main nerve agents

Nerve agent	k_s (min^{-1})	k_a (min^{-1})
VX	3.5×10^{-4}	3.2×10^{-4}
Sarin	Ø	3.8×10^{-3}
Soman	Ø	2.2×10^{-1}
Tabun	Ø	6.0×10^{-4}

The spontaneous reactivation was not observed for three of the four main nerve agents. Only VX can undergo spontaneous reactivation at an experimentally measurable rate and does so faster than dealkylation. We can notice than while dealkylation is a slow process with a rate constant indicating a few transformations per day at most, the dealkylation of soman-inhibited AChE is much faster. The dealkylation rate constant of soman is indicative of a few transformations per hour. Those constants mean that for sarin, soman, and tabun, the eventual outcome of nerve agent inhibition in the aging of the nerve agent adduct, permanently eliminating AChE activity. For VX, the pool of inhibited AChE will be split between aged and spontaneously reactivated AChE. The third possibility is the accelerated reactivation of AChE with a strong nucleophile to restore the catalytic activity of the enzyme.

II.3. Reactivation of acetylcholinesterase

For several years, nerve agents and organophosphate inhibitors in general were though to inhibit AChE irreversibly. Irwin B. Wilson first laid the principle of reactivation by a nucleophile on the assumption that the electrophilic phosphate ester is opened to a nucleophilic addition.[112] This first study showed reactivation activity for hydroxylamine. Five years later, after some success reactivating AChE with hydroxamic acids, I. B. Wilson used an oxime, 2-pyridinium aldoxime methiodide iodine (also known as pralidoxime, or 2-PAM, or even P2S when the counterion is a mesylate and not iodine), to reactivate diisopropyl fluorophosphate (DFP) inhibited AChE (see **Scheme II-6**).[116] The full IUPAC name of the classical oxime reactivators presented in **Scheme II-7** can be found at the bottom of the page.



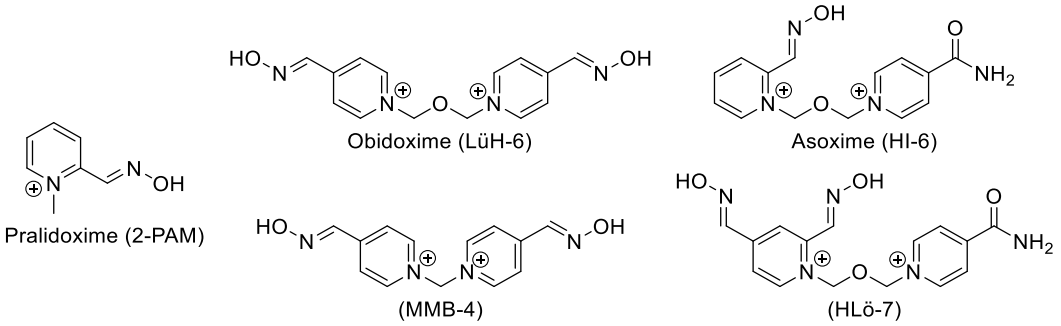
Scheme II-6. First known reactivators of OP inhibited AChE.

Oximes are imines in which the nitrogen bound substituent is a hydroxy group. As such, their IUPAC name is hydroxyimino. A distinction is made between aldoximes which have only one alkyl substituent and ketoximes which have two. The hydrogen of the oxime group is labile and several oximes are at equilibrium with their oximate form in aqueous solution. The potential of oximates as reactivators stems from the high nucleophilicity of the negatively charge oxygen due to the alpha effect. The alpha effect is the increased nucleophilicity observed in nucleophiles in which the nucleophilic atom is adjacent to an atom with a lone pair.[117] This effect is well known and studied but it's exact origin still has not been identified.[118] In oximates, the lone pair on the nitrogen atom increases the nucleophilicity of the negatively charged oxygen atom. 2-PAM was the first reactivator to see clinical use in nerve agent poisoning.[119] A notable feature of AChE is the pyridinium ring which is a staple of a generation of oxime reactivators. While the electron poor pyridinium diminishes the nucleophilicity of the oxime group, it was deemed a necessary sacrifice. The role of this quaternary nitrogen is to emulate the quaternary nitrogen of the choline moiety of acetylcholine. As such it's purpose is to contribute to the binding of 2-PAM in the active site of AChE. This first breach in the realm of AChE reactivation was quickly followed by the design of bis-pyridinium dioxime reactivators such as LüH6, also known as obidoxime, or MMB-4 (see

IUPAC names for reactivators are: 2-PAM: 2-[(hydroxyimino)methyl]-1-methylpyridin-1-ium LüH6: 1,1'-(oxy-bis-methylene)bis[4-(hydroxyimino)methyl]pyridinium MMB-4: 1,1'-methylene-bis[4-(hydroxyimino)methyl]pyridinium HLö7: 1-[[[4-(aminocarbonyl)pyridinio]methoxy]methyl]-2,4-bis-[(hydroxyimino)methyl] pyridinium HI-6: 1-[[[4-(aminocarbonyl)pyridinio]methoxy]methyl]-2-[(hydroxyimino)methyl]pyridinium

[116] I. B. Wilson, S. Ginsburg, *Biochim. Biophys. Acta* **1955**, *18*, 168–170.
[117] N. J. Fina, J. O. Edwards, *Int. J. Chem. Kinet.* **1973**, *5*, 1–26.
[118] W.-Y. Zhao, J. Yu, S.-J. Ren, X.-G. Wei, F.-Z. Qiu, P.-H. Li, H. Li, Y.-P. Zhou, C.-Z. Yin, A.-P. Chen, et al., *J. Comput. Chem.* **2015**, *36*, 844–852.
[119] Namba T, Hiraki K, *JAMA* **1958**, *166*, 1834–1839.

Scheme II-7).[120] The last classical reactivators are HI-6, also known as asoxime, and HLö7.[121] Both are bis-pyridinium reactivators with only one of those pyridinium substituted by one or more oxime functional groups. Those five oximes are some the most effective classical reactivator designs and have clinical use.



Scheme II-7. Classical oxime reactivators

The reactivations of AChE is another two-step process like the covalent inhibition of AChE (see **Scheme II-8**). The oxime first reversibly binds to the active site of AChE and then reacts on the phosphorus centre. This reaction restores the enzyme to its original uninhibited state while the oxime-nerve agent adduct can exit the active site of AChE. In **Scheme II-8** Ox is the oxime in the reactive oximate form, EP is the nerve agent inhibited enzyme, Ox•EP is the oxime phosphorylated enzyme complex E is the free enzyme and OxP is the oxime-nerve agent adduct.

Scheme II-8. Reaction equation for the reactivation of nerve agent inhibited AChE.

The parallel between inhibition and reactivation continues as the three constants commonly used to describe the reactivation are very similar to those used to describe inhibition. The dissociation constant K_{OX} is again described as the ratio of the sum of all step leading to or leading from the bonded complex between the inhibited enzyme and the oxime (see equation **II-VII**).[122] This constant is inversely correlated to the ability the oxime to bind the inhibited enzyme. The dephosphorylation constant k_2 is the rate constant of the dephosphorylation of the enzyme. The ratio of the dephosphorylation constant over the dissociation constant is the bimolecular rate constant of the reaction k_r that will be called reactivation rate constant throughout this work for consistency and clarity **II-VIII**. In many articles k_2 is called the reactivation constant and k_r is called the bimolecular rate constant.

[120] a) F. Hobbiger, P. W. Sadler, *Br. J. Pharmacol. Chemother.* **1959**, *14*, 192–201. b) E. Heilbronn, B. Tolagen, *Biochem. Pharmacol.* **1965**, *14*, 73–77.

[121] a) L. P. A. de Jong, G. Z. Wolring, *Biochem. Pharmacol.* **1980**, *29*, 2379–2387. b) L. P. A. de Jong, M. A. A. Verhagen, J. P. Langenberg, I. Hagedorn, M. Löffler, *Biochem. Pharmacol.* **1989**, *38*, 633–640.

[122] Z. Kovarik, N. Ciban, Z. Radić, V. Simeon-Rudolf, P. Taylor, *Biochem. Biophys. Res. Commun.* **2006**, *342*, 973–978.

$$K_{OX} = \frac{k_{-1} + k_2}{k_1} \qquad \textbf{II-VII}$$

$$k_r = \frac{k_2}{K_{OX}} = \frac{k_2 k_1}{k_{-1} + k_2} \qquad \textbf{II-VIII}$$

Table II-3. Dissociations (K_{OX}), reactivation (k_2) and bimolecular (k_r) rate constants for the five main reactivators towards the four most common nerve agents.

		Tabun	VX	Sarin	Soman
	K_{ox} (M)	706 x 10^{-6} [b]	28.1 x 10^{-6} [a]	27.6 x 10^{-6} [a]	Ø[d]
2-PAM	**k_2 (min^{-1})**	0.01[b]	0.215[a]	0.25[a]	Ø[d]
	k_r ($M^{-1}min^{-1}$)	1.4 x 10^1	9.9 x 10^3	9.1 x 10^3	Ø
	K_{ox} (M)	97.3 x 10^{-6} [b]	27.4 x 10^{-6} [a]	31.3 x 10^{-6} [a]	Ø[d]
LüH6	**k_2 (min^{-1})**	0.04[b]	0.893[a]	0.937[a]	Ø[d]
	k_r ($M^{-1}min^{-1}$)	4.1 x 10^2	3.3 x 10^4	3.0 x 10^4	Ø
	K_{ox} (M)	106.5 x 10^{-6} [b]	7.8 x 10^{-6} [a]	24.2 x 10^{-6} [a]	Ø[d]
HLö7	**k_2 (min^{-1})**	0.02[b]	0.49[a]	0.849[a]	Ø[d]
	k_r ($M^{-1}min^{-1}$)	1.9 x 10^2	6.3 x 10^4	3.5 x 10^4	Ø
	K_{ox} (M)	Ø[b]	11.5 x 10^{-6} [a]	50.1 x 10^{-6} [a]	Ø[d]
HI-6	**k_2 (min^{-1})**	Ø[b]	0.242[a]	0.677[a]	Ø[d]
	k_r ($M^{-1}min^{-1}$)	Ø	2.1 x 10^4	1.4 x 10^4	Ø
	K_{ox} (M)	2418 x 10^{-6} [e]	241.9 x 10^{-6} [e]	1544 x 10^{-6} [e]	Ø[d]
MMB-4	**k_2 (min^{-1})**	0.02[e]	1.56[e]	1.87[e]	Ø[d]
	k_r ($M^{-1}min^{-1}$)	8.3	6.5 x 10^3	1.2 x 10^3	Ø

a: from ref 123. b: from ref 105. c: from ref 124. d: from ref 125, shown to be unreactive due to the fast aging of soman. e: from ref 126.

For VX, the reaction rate with water is of the order of 10^{-4} min^{-1}. For the core pyridinium oxime, 2-PAM, the dephosphorylation rate constant is three orders of magnitude higher in the range of 10^{-1}. According to transition state theory (see equations **II-IX** and **II-X**), this is equivalent to a lowering of the energy barrier by 4.1 $kcal.mol^{-1}$.

[123] F. Worek, G. Reiter, P. Eyer, L. Szinicz, *Arch. Toxicol.* **2002**, *76*, 523–529.
[124] N. Aurbek, H. Thiermann, L. Szinicz, P. Eyer, F. Worek, *Toxicology* **2006**, *224*, 91–99.
[125] A. Bartling, F. Worek, L. Szinicz, H. Thiermann, *Toxicology* **2007**, *233*, 166–172.
[126] F. Worek, T. Wille, N. Aurbek, P. Eyer, H. Thiermann, *Tox. Appl. Pharmacol.* **2010**, *249*, 231–237.

$$k = Ae^{-\Delta E^{\ddagger}\beta} \qquad \textbf{II-IX}$$

$$\Delta\Delta E^{\ddagger} = ln\left(\frac{k_a}{k_b}\right) * k_B T \qquad \textbf{II-X}$$

The five main classical reactivators serve as references in the evaluation of new reactivators. The affinity and reactivity of those reactivators against tabun, VX, sarin and soman is listed in **Table II-3**. The first observation is that it was impossible for any reactivator to reactivate soman inhibited AChE. The dealkylation reaction is known to be extremely fast for soman and none of the five reactivators is able to reactivate aged soman. Another observation is the generally low reactivity of all reactivators toward tabun inhibited AChE. For all five reactivator tabun inhibited AChE is the nerve agent they have the least affinity for and the lowest reactivity towards. Regarding affinity, MMB-4 is the least affine of the reactivators across all nerve agents. When VX and sarin are concerned, HLö7 has the lowest K_{OX}, by a significant margin in the case of VX. The best affinity toward tabun inhibited AChE is with LüH6. When considering reactivity, MMB-4 has the highest dephosphorylation constant towards VX and Sarin. LüH6 has the highest activity towards tabun and very honourable dephosphorylation rate constants for VX and tabun. The lowest activity, against all nerve agents is 2-PAM. Overall, 2-PAM is not a good reactivator because of its low reactivity. Neither is MMB-4 because of its very low affinity. HLö7 would be a good reactivator except for its inability to reactivate tabun inhibited AChE. LüH6 is also a good reactivator, slightly inferior to HLö7 except for tabun where it is the best overall reactivator. HI-6 is a good reactivator although completely unable to reactivate tabun and inferior to HLö7 and LüH6 in every respect.

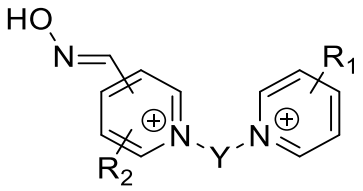
What this data shows is the complete inability of reactivators to reactivate two of the main combat never agents. It also shows that the reactivation constant towards the nerve agents that can be reactivated is in the range of 10^4, which remains low. What this data does not show is that for other nerve agents, specifically the numerous ones used as pesticides, the reactivation rate is not consistent.[105] Another lasting issue with the classical reactivators is their poor penetration of the blood-brain barrier to reactivate inhibited AChE in the brain.[127] The penetration of the oxime in the brain to reactivate inhibited AChE inhibited is key in the treatment of nerve agent poisoning. The reason is that part of the symptoms of nerve agent induced polyneuropathy are caused by a disruption of neuron to neuron synapses in the central nervous system.

A lot of research was dedicated to the design of new reactivators. The existing bipyridinium bis- or mono- oxime reactivators served as an inspiration for many potential new reactivators (see **Scheme II-9**).[127], [128] Those oxime designs vary in linker between the two

[127] J. Cabal, K. Kuča, J. Kassa, *Basic Clin. Physiol. Pharmacol.* **2004**, *95*, 81–86.

[128] a) M. Čalić, A. L. Vrdoljak, B. Radić, D. Jelić, D. Jun, K. Kuča, Z. Kovarik, *Toxicology* **2006**, *219*, 85–96. b) K.-A. Oh, G. Y. Yang, D. Jun, K. Kuca, Y.-S. Jung, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 4852–4855. c) K. Musilek, O. Holas, K. Kuca, D. Jun, V. Dohnal, M. Dolezal, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5673–5676.

pyridiniums (Y), in the position of the oxime, in the position and nature of an extra substituent on the first pyridinium (R_2), and in the position and nature of an extra substituent on the second pyridinium (R_1). While some of those oximes proved to be good reactivators, none had a significantly increased in activity against tabun or soman.



Scheme II-9. Generic oxime reactivator formula based on the bipyridinium bis- or mono- oxime

Several new reactivator designs were proposed outside of this rigid framework while keeping the oxime substituted pyridinium ring.[129] While some of them were able to reactivate AChE with reasonable efficiency, none showed exceptional ability to reactivate the enzyme.

A recent trend in reactivator design is to move away from the oxime substituted pyridinium.[130] This new class of reactivators showed, as expected, greater lipophilicity. It means that they have the potential to cross the blood-brain barrier with more ease than classical reactivators.[130a] The absence of the electron poor pyridinium also increases the reactivity of those oximes.[130b] The cost of that greater lipophilicity and nucleophilicity is a diminished affinity for AChE, confirming once again the role of the quaternary nitrogen in the binding phase of reactivation. Out of several studies of such non quaternary reactivators, one in particular emerged as a viable nerve agent reactivator (compound **1** in **Scheme II-10**).[131]

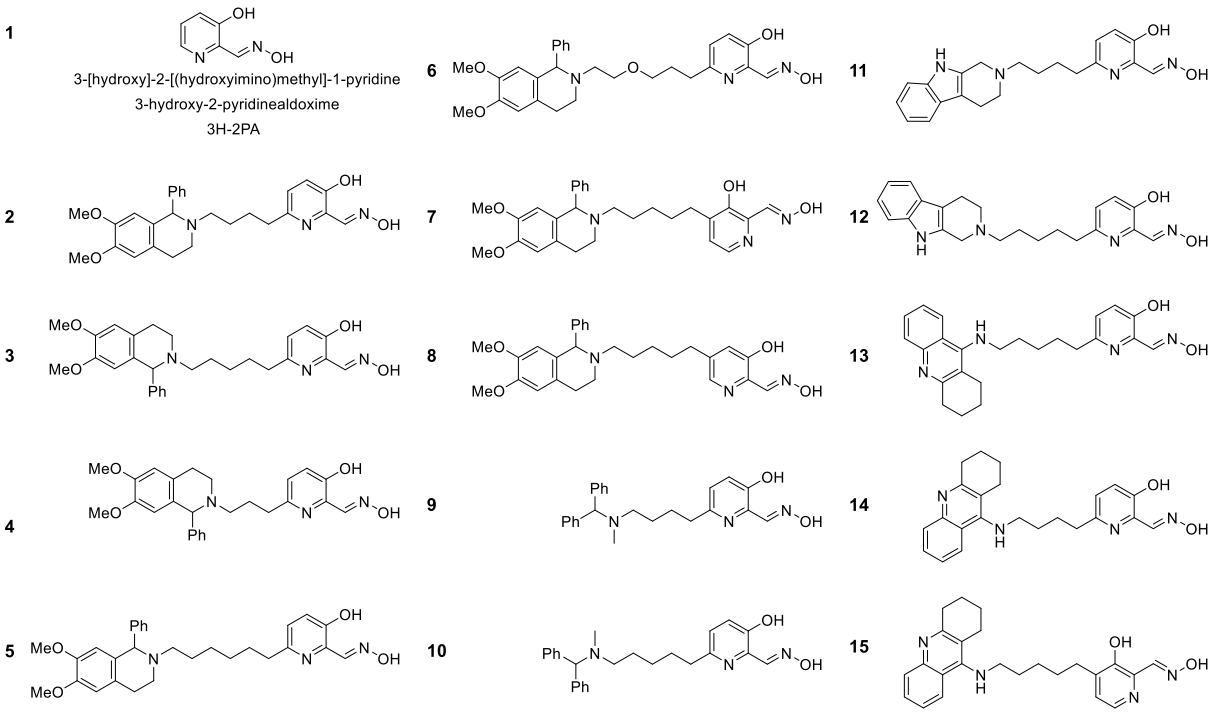
---

d) K. Kuca, D. Jun, J. Cabal, L. Musilova, *Basic Clin. Physiol. Pharmacol.* **2007**, *101*, 25–28. e) J. Kassa, J. Karasova, K. Musilek, K. Kuca, Y.-S. Jung, *Drug Chem. Toxicol.* **2008**, *31*, 371–381. f) J. Kassa, J. Karasova, K. Musilek, K. Kuca, *Toxicology* **2008**, *243*, 311–316. g) J. Acharya, A. K. Gupta, A. Mazumder, D. K. Dubey, *Toxicol. in Vitro* **2008**, *22*, 525–530. h) D. M. Maxwell, I. Koplovitz, F. Worek, R. E. Sweeney, *Toxicol. Appl. Pharmacol.* **2008**, *231*, 157–164. i) J. Kassa, D. Jun, J. Karasova, J. Bajgar, K. Kuca, *Chem. Biol. Interact.* **2008**, *175*, 425–427. j) Z. Kovarik, M. Čalić, G. Šinko, A. Bosak, S. Berend, A. L. Vrdoljak, B. Radić, *Chem. Biol. Interact.* **2008**, *175*, 173–179. k) J. Acharya, D. K. Dubey, S. K. Raza, *Toxicol. in Vitro* **2010**, *24*, 1797–1802. l) A. K. Sahu, B. Gupta, R. Sharma, Y. Signh, K. Musilek, K. Kuca, K. K. Ghosh, *Indian J. Chem. A* **2015**, *54A*, 40–45. m) M. Winter, T. Wille, K. Musilek, K. Kuca, H. Thiermann, F. Worek, *Toxicol. Lett.* **2016**, *244*, 136–142.

[129] a) K. Kuca, J. Picha, J. Cabal, F. Liska, *J. appl. Biomed.* **2004**, *2*, 51–56. b) R. Odžak, M. Čalić, T. Hrenar, I. Primožič, Z. Kovarik, *Toxicology* **2007**, *233*, 85–96. c) A. K. Valiveti, U. M. Bhalerao, J. Acharya, H. N. Karade, B. N. Acharya, G. Raviraju, A. K. Halve, M. P. Kaushik, *Bioorg. Med. Chem.* **2015**, *23*, 4899–4910. d) Pooja, S. Aggarwal, A. K. Tiwari, V. Kumar, R. Pratap, G. Singh, A. K. Mishra, *RSC Adv.* **2015**, *5*, 23471–23480.

[130] a) J. Kalisiak, E. C. Ralph, J. Zhang, J. R. Cashman, *J. Med. Chem.* **2011**, *54*, 3319–3330. b) Z. Radić, R. K. Sit, Z. Kovarik, S. Berend, E. Garcia, L. Zhang, G. Amitai, C. Green, B. Radić, V. V. Fokin, et al., *J. Biol. Chem.* **2012**, *287*, 11798–11809. c) S. F. da Cunha Xavier Soares, A. A. Vieira, R. T. Delfino, J. D. Figueroa-Villar, *Bioorg. Med. Chem.* **2013**, *21*, 5923–5930. d) S. F. McHardy, J. A. Bohmann, M. R. Corbett, B. Campos, M. W. Tidwell, P. M. Thompson, C. J. Bemben, T. A. Menchaca, T. E. Reeves, W. R. Cantrell Jr., et al., *Bioorg. Med. Chem. Lett.* **2014**, *24*, 1711–1714. e) Z. Wei, Y. Liu, Y. Wang, W. Li, X. Zhou, J. Zhao, C. Huang, X. Li, J. Liu, Z. Zheng, et al., *Toxicol. Lett.* **2016**, *246*, 1–6.

[131] L. Louise-Leriche, E. Păunescu, G. Saint-André, R. Baati, A. Romieu, A. Wagner, P.-Y. Renard, *Chem. Eur. J.* **2010**, *16*, 3510–3523.



Scheme II-10. Non-quaternary oxime reactivators build from the prototype 3H-2PA non-quaternary oxime.

The reactivity of nucleophiles with PhX, an analogue of VX with a phenyl instead of a methyl substituent, is used to evaluate the potential of new oxime reactivators. In two such studies,[132] an alcohol substituted 2-pyridinealdoxime (3H-2PA, compound **1** from **Scheme II-10**) was identified as a potential reactivator of nerve agent inhibited AChE. The absence of the positive charge of the pyridinium was postulated to cause a decreased affinity of 3H-2PA towards AChE compared to quaternary oximes. This was confirmed (see **Table II-4**) in a reactivation study. 3H-2PA has a good dephosphorylation constant, of the same order of magnitude than LüH6 and HLö-7, but has a much higher K_{OX} by at least 3 orders of magnitude.[133] The lack of affinity can be fixed by linking this reactivator to a good ligand of the peripheral active site of AChE. Several attempts were made to find a proper ligand to attach to the prototype with a linker. The results of this extended search can be found in a series of paper spanning three years.[133], [134], [135], [136], [137], [138] A few examples of those

[132] a) L. Louise-Leriche, E. Păunescu, G. Saint-André, R. Baati, A. Romieu, A. Wagner, P.-Y. Renard, *Chem. Eur. J.* **2010**, *16*, 3510–3523. b) G. Saint-André, M. Kliachyna, S. Kodepelly, L. Louise-Leriche, E. Gillon, P.-Y. Renard, F. Nachon, R. Baati, A. Wagner, *Tetrahedron* **2011**, *67*, 6352–6361.

[133] J. Renou, G. Mercey, T. Verdelet, E. Păunescu, E. Gillon, M. Arboléas, M. Loiodice, M. Kliachyna, R. Baati, F. Nachon, et al., *Chem. Biol. Interact.* **2013**, *203*, 81–84.

[134] G. Mercey, T. Verdelet, G. Saint-André, E. Gillon, A. Wagner, R. Baati, L. Jean, F. Nachon, P.-Y. Renard, *Chem. Commun.* **2011**, *47*, 5295–5297.

[135] G. Mercey, J. Renou, T. Verdelet, M. Kliachyna, R. Baati, E. Gillon, M. Arboléas, M. Loiodice, F. Nachon, L. Jean, et al., *J. Med. Chem.* **2012**, *55*, 10791–10795.

[136] J. Renou, M. Loiodice, M. Arboléas, R. Baati, L. Jean, F. Nachon, P.-Y. Renard, *Chem. Commun.* **2014**, *50*, 3947.

[137] M. Kliachyna, G. Santoni, V. Nussbaum, J. Renou, B. Sanson, J.-P. Colletier, M. Arboléas, M. Loiodice, M. Weik, L. Jean, et al., *E. J. Med. Chem.* **2014**, *78*, 455–467.

[138] J. Renou, J. Dias, G. Mercey, T. Verdelet, C. Rousseau, A.-J. Gastellier, M. Arboléas, M. Touvrey-Loiodice, R. Baati, L. Jean, et al., *RSC Adv.* **2016**, *6*, 17929–17940.

ligand/linkers pairs are shown in **Scheme II-10** and the corresponding reactivation constants are in **Table II-4.**

Table II-4. Dissociations (K_{OX}), reactivation (k_2) and biomelecular (k_r) rate constants for a class of new reactivators towards VX and tabun. K_{OX} in µM, k_2 in min^{-1} and k_r in $min^{-1}.mM^{-1}$. Reactivator numbering corresponding to the structures in **Scheme II-10.**

Reactivator	Reference	VX			Tabun		
		K_{OX}	k_2	k_r	K_{OX}	k_2	k_r
LüH6	105	27.4	0.893	32	97	0.04	0.4
HLö-7	105	7.8	0.49	63	106	0.02	0.2
1	133	30000	0.5	0.017	nd[c]	nd[c]	nd[c]
2	134	47	0.82	17	25	0.042	1.7
3	134	6	0.35	61	5	0.015	3.4
4	135	117	0.45	3.8	7	0.0034	0.5
5	135	38	0.29	7.6	14	0.021	1.5
6	135	nd[a]	nd[a]	3	4	0.006	1.5
7	135	113	0.28	2.4	75	0.08	1
8	135	nd[a]	nd[a]	0.2	nd[b]	nd[b]	nd[b]
9	133	nd[a]	nd[a]	2.6	nd[c]	nd[c]	nd[c]
10	133	26	0.075	2.9	nd[c]	nd[c]	nd[c]
11	136	885	5.6	6	100	0.06	0.6
12	136	125	1.1	9	180	0.06	0.3
13	137	41	0.56	13	30	0.0075	0.27
14	137	31	0.72	22	7.1	0.021	3
15	137	nd[b]	nd[b]	nd[b]	nd[c]	nd[c]	nd[c]

a: [reactivator] was too high compared to K_{OX}, the relation between observed rate and [reactivator] is purely linear and k_r and K_{OX} could not be determined. b: observable activity was too low. c: experiment was not carried out.

The first observation is that while the bimolecular rate constants for tabun remain low, some of the new reactivators present improvements over the classical ones. Namely compounds **2**, **3**, **4**, **5**, **6**, **7**, **11**, and **14**. This improvement comes from a greater affinity of those reactivators towards AChE. The greatest improvement is obtained when this good affinity is matched by a reactivity towards the tabun-AChE adduct equivalent to that of LüH6. The activity towards VX however is generally lower than HLö7. Only a single of those new oximes, oxime **3**, is as good a reactivator towards VX that HLö-7, despites its lower dephosphorylation

constant thanks to its lower K_D illustrating a very good affinity towards VX-inhibited AChE. The other reactivators show a generally diminished dephosphorylation constant compared to HLö7 with the two notable exceptions of compounds **2**, **11** and **12** whose affinity leave to be desired.

The data produced in the search for a good ligand/linker pair for 3-H-2-PA also gives indications on future designs. A proper linker needs to be in position 5 of the pyridine ring as illustrated by the difference in k_r between compounds **6**, **7** and **8** and between compounds **13** and **15**. It also needs to be connected to the pyridine ring by a five-atom chain as illustrated by the good k_r of compound **3** compared to compounds **2**, **4**, and **5**, of compound **10** compared to compound **9**, of compound **12** compared to compound **11** and of compound **14** compared to compound **13**.

In the work presented later in this manuscript, the dephosphorylation activity of 6M-3H-2PA, an analogue of 3H-2PA, the prototype of those new reactivators, has been compared to the prototype of classical reactivators, 2-PAM in **Chapter 5**.

The studies on 3-H-2-PA are not the only efforts on a new uncharged reactivator design as it is the subject of several studies from the group of Palmer Taylor in San Diego[139], José Daniel Figueroa-Villar in Rio de Janeiro[140], Stanton F. McHardy in San Antonio[141], and Song Li in Beijing[142].

III. Molecular simulations of AChE

Acetylcholinesterase, as a well-studied enzymatic system of high biological significance, has been the subject of a many theoretical studies, especially on its catalytic cycle. Those helped to lift the veil on the exact mechanism of AChE in its regular activity but also as an inhibitor target. However, many questions are yet unanswered, especially when it comes to the subject of AChE reactivation. The major publications in that domain are presented here with a special emphasis of reactivation as it is the main subject of the work presented in this manuscript.

III.1. Regular catalytic activity

The catalytic activity of AChE has been investigated experimentally more than 40 years before any theoretical mechanistic study. This work outlined a two-step catalysis in which the catalytic serine is first acylated by ACh with the elimination of its choline moiety and then deacylated by a water molecule. Both processes involved the addition of a nucleophile on the carbonyl of ACh and the departure of a leaving group. In both cases the addition and

[139] a) Z. Radić, R. K. Sit, Z. Kovarik, S. Berend, E. Garcia, L. Zhang, G. Amitai, C. Green, B. Radić, V. V. Fokin, *et al.*, *J. Biol. Chem.* **2012**, *287*, 11798–11809. b) R. K. Sit, Z. Radić, V. Gerardi, L. Zhang, E. Garcia, M. Katalinić, G. Amitai, Z. Kovarik, V. V. Fokin, K. B. Sharpless, *et al.*, *J. Biol. Chem.* **2011**, *286*, 19422–19430.

[140] S. F. da Cunha Xavier Soares, A. A. Vieira, R. T. Delfino, J. D. Figueroa-Villar, *Bioorg. Med. Chem.* **2013**, *21*, 5923–5930.

[141] S. F. McHardy, J. A. Bohmann, M. R. Corbett, B. Campos, M. W. Tidwell, P. M. Thompson, C. J. Bemben, T. A. Menchaca, T. E. Reeves, W. R. Cantrell Jr., et al., *Bioorg. Med. Chem. Lett.* **2014**, *24*, 1711–1714.

[142] Z. Wei, Y. Liu, Y. Wang, W. Li, X. Zhou, J. Zhao, C. Huang, X. Li, J. Liu, Z. Zheng, et al., *Toxicol. Lett*. **2016**, *246*, 1–6.

elimination can be concerted, going through an SN_2 mechanism, or stepwise, going through an addition-elimination mechanism. The two nucleophiles, Ser203 and water, are activated by a deprotonation that can occur before or concurrently to the nucleophilic addition. The possible mechanisms for acylation and deacylation are presented in **Figure III-1**, **Figure III-2**, **Figure III-3**, and **Figure III-4**. Those questions and other issues regarding the mechanism of AChE catalysis where addressed in several theoretical studies over the years. This section is a summary of that research.

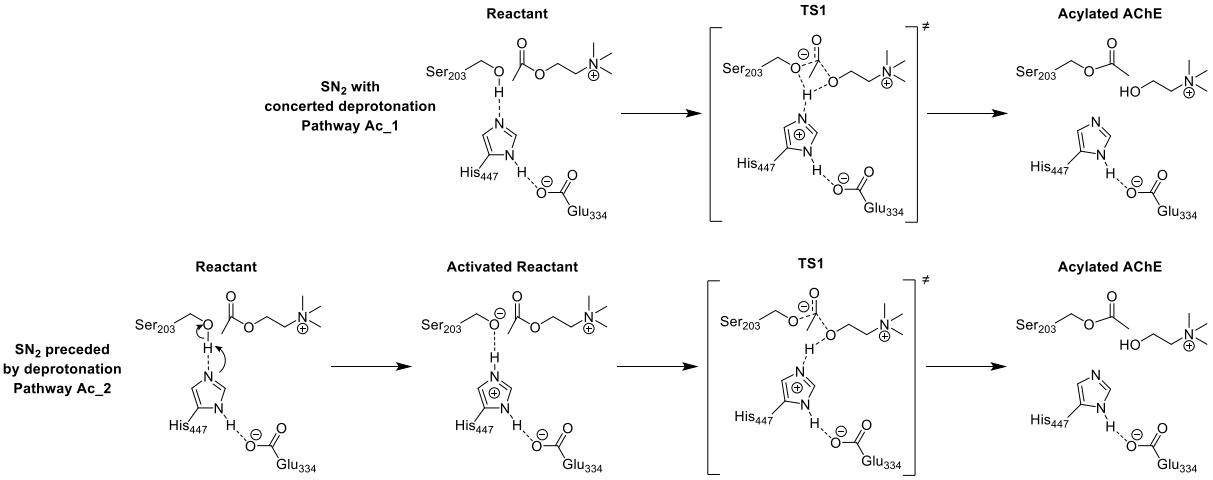


Figure III-1. Possible mechanisms for the acylation step of ACh hydrolysis by AChE in the hypothesis of a SN_2 mechanism

In the acylation, there is a leaving group that will be substituted by a nucleophile on a carbon centre. The leaving group is the choline moiety of acetylcholine and the nucleophile is the alcoolate side chain of Ser203. In the hypothesis of a SN_2 mechanism for the acylation step (**Figure III-1**), two pathways are available. The first is pathway **Ac_1** in which the nucleophile (Ser203) is activated concurrently to the nucleophilic addition on the carbonyl of acetylcholine. There is a single step between the reactant and TS1. In the structure of that transition state the proton of Ser203 is share between Ser203, the general base assisting in the deprotonation, His447, and the final proton acceptor, the choline moiety of ACh. The second pathway is pathway **Ac_2** in which the nucleophile is activated prior to the nucleophilic addition. This is a two-step process with the deprotonation of Ser203 followed by the nucleophilic substitution on the carbonyl of ACh. In this pathway, in the transition state TS1, the proton has already moved away from Ser203's alcohol and is between His447 and the choline moiety of ACh. In both pathways, the addition of the nucleophile (Ser203's oxygen) is concurrent to the elimination of the choline moiety.

In the hypothesis of an addition-elimination mechanism for the acylation step (**Figure III-2**), two pathways are again available. The first one is the succession of pathways **Ac_3** and **Ac_5**, the second is the succession, of pathways **Ac_4** and **Ac_5**. The addition step **Ac_3** is concurrent to the activation of the nucleophile by deprotonation. In the resulting TS, while the bend between the oxygen of Ser203 and the carbonyl shortens, the proton of Ser203 is being transferred to His447. The addition step of pathway **Ac_4** is preceded by the

deprotonation of the nucleophile. Both **Ac_3** and **Ac_4** pathways give the same product, the tetravalent intermediate in which the oxygen of the carbonyls forms a single bond with the carbon and carries a negative charge. The two possible addition-elimination mechanisms for the acylation share the same elimination step, **Ac_5**. In this step, the choline moiety is eliminated and the proton of His447 is transferred to the primary alcohol of choline.

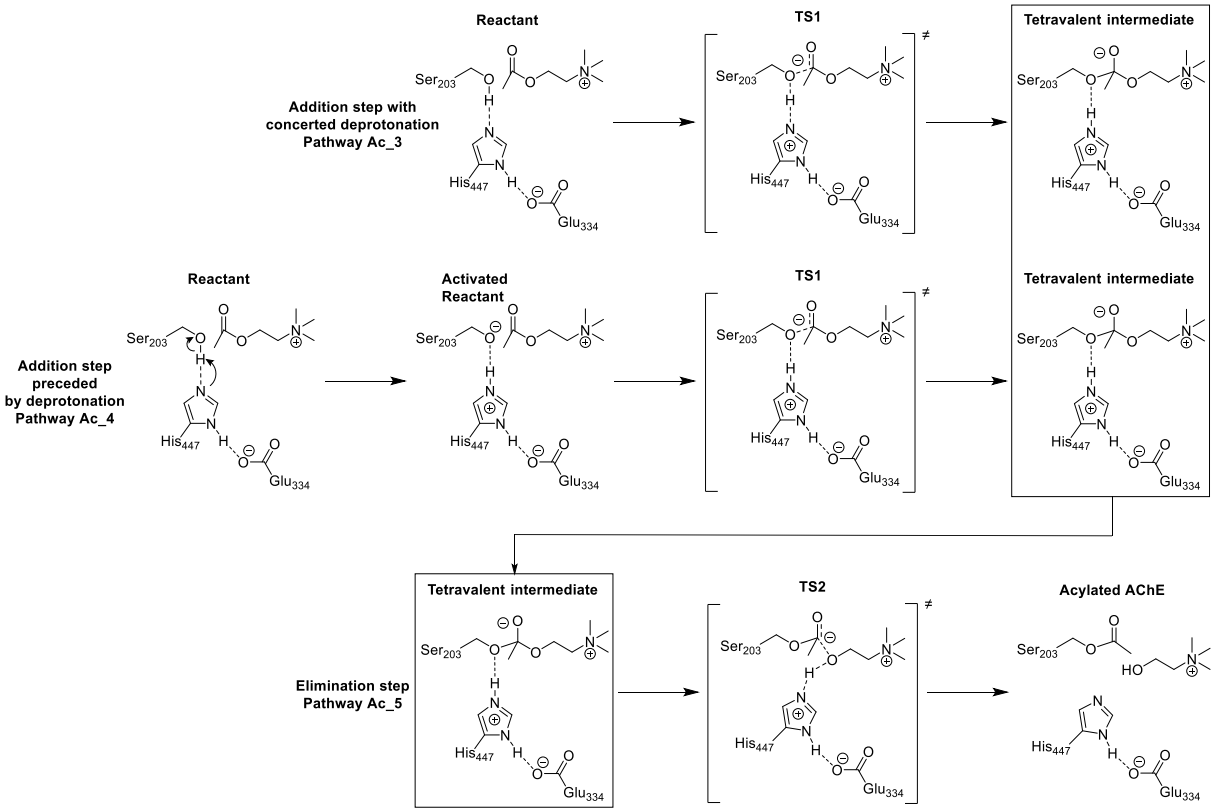


Figure III-2. Possible mechanisms for the acylation step of ACh hydrolysis by AChE in the hypothesis of an addition-elimination mechanism

The putative SN_2 mechanism for the deacylation step is almost identical to the putative SN_2 mechanism of the acylation (see **Figure III-3**). The main difference is a change in nucleophile and leaving group. In this case the nucleophile is a hydroxide and the leaving group is the AChE enzyme. There are two possible pathways, **Deac_1** and **Deac_2**. In the first, the proton of Ser203 is transferred while the nucleophilic substitution of the choline moiety by Ser203's side chain occurs. In the second one, **Deac_2**, the activation is stepwise and precedes the nucleophilic substitution on the carbonyl.

The putative addition-elimination mechanism of the deacylation is also quite similar to its acylation counterpart. Both possible pathways differ in their addition step and have a common elimination step. **Deac_3** involves a concerted deprotonation of the nucleophile and nucleophilic addition on the carbonyl. In **Deac_4** the deprotonation occurs first and is then followed by the nucleophilic addition. The **Deac_5** pathway is a very straightforward elimination of the enzyme from the carbonyl.

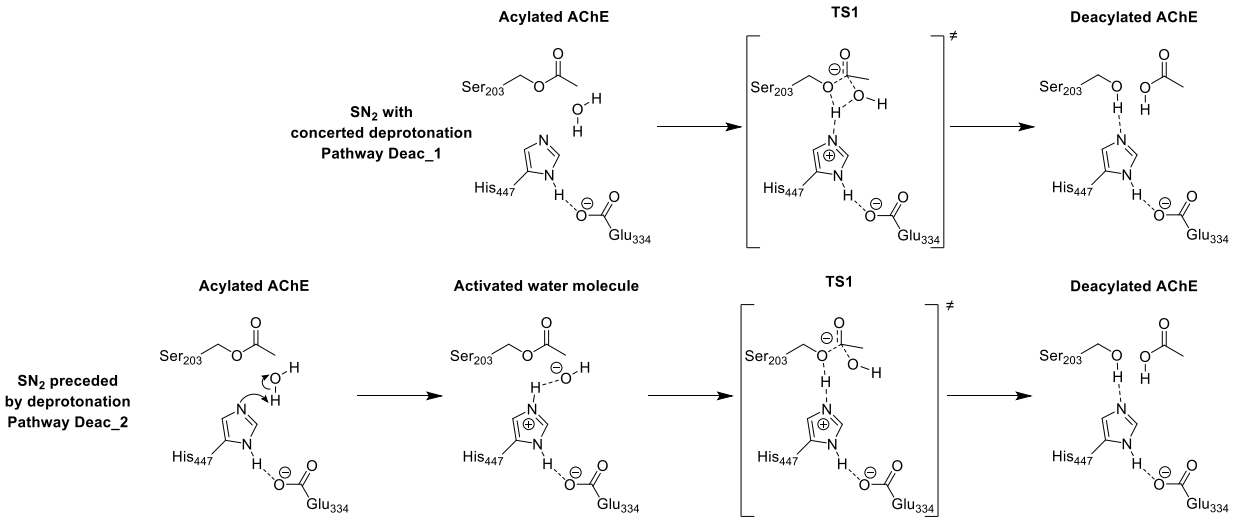


Figure III-3. Possible mechanisms for the deacylation step of ACh hydrolysis by AChE in the hypothesis of a SN_2 mechanism

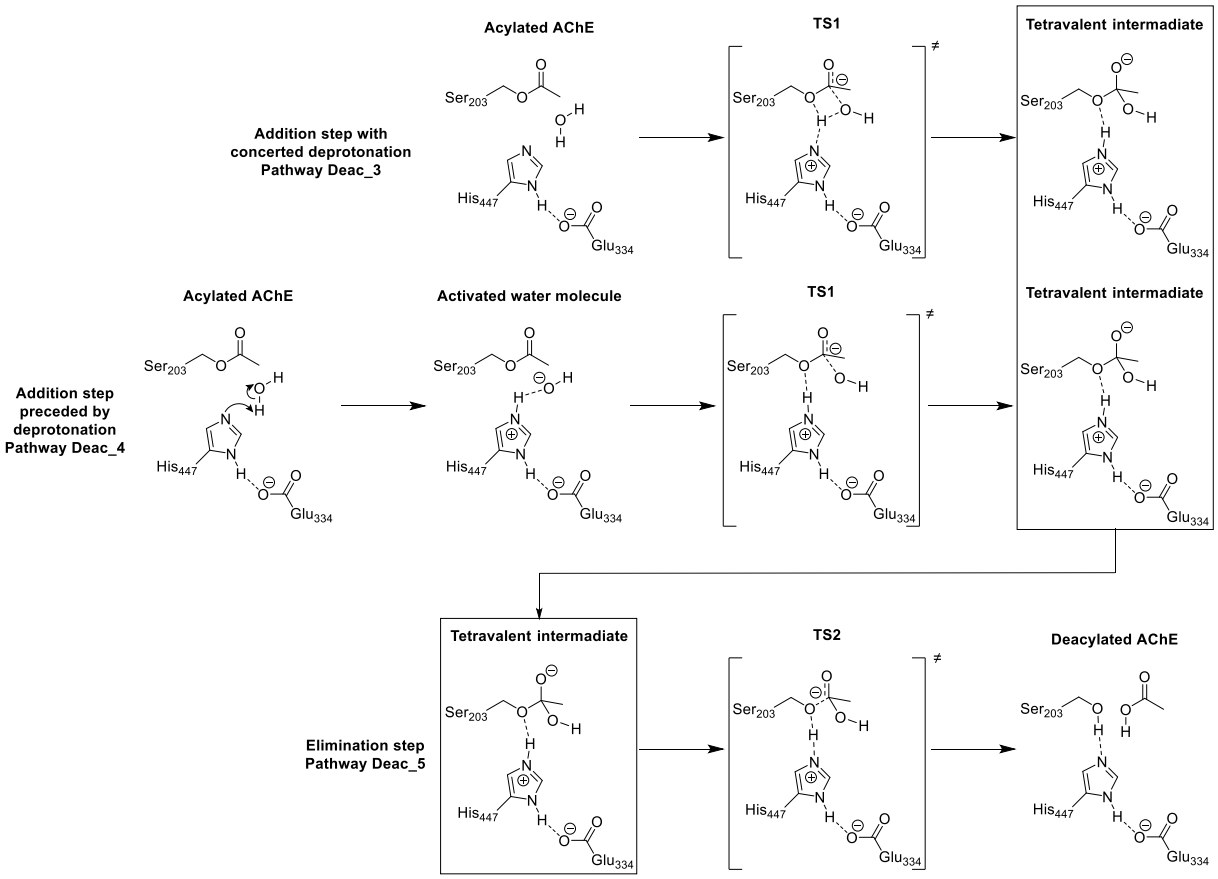


Figure III-4. Possible mechanisms for the deacylation step of ACh hydrolysis by AChE in the hypothesis of an addition-elimination mechanism

One of the first points addressed in simulations of the catalytic cycle of AChE is the mechanism of both acylation and deacylation steps. A very early quantum mechanical investigation of the mechanism of ACh catalysis by AChE is a set of semi-empirical QM/MM simulations.[143] The QM/MM method splits the systems into a subgroup treated with

[143] V. V. Vasilyev, *J. Mol. Struct. THEOCHEM* **1994**, *304*, 129–141.

quantum mechanics (the QM region), composed of the atoms directly involved in the reactivity; and a second larger classically treated subgroup (the MM region), whose influence needs to be taken into account but with no direct involvement in the reactivity. The atoms in the classically treated subgroup can be fully, only partially, or not fixed at all. This method allows to consider the influence of the apoenzyme beyond the atoms treated with quantum dynamics of the QM region without overinflating the calculation cost beyond what is possible. This article presents the optimisation of a stable tetravalent intermediate in the acylation step of ACh hydrolysis. The stability of this intermediate adds weight to the hypothesis of an addition-elimination mechanism for the acylation (see **Figure III-2**). This data was produced with the semi-empirical method PM3. This kind of method replaces the entirety of some part of the electron-interaction terms present at the Hartre Fock level by a parameter fixed to experiment. The limited accuracy of such method methods warrant a cautious use of this data, as well as the assumption of a stepwise deprotonation of Ser203. A truncated active site model was used with a semi-empirical methodology to simulate both the acylation and deacylation in two articles by Wan and co-workers. [144], [145] For both steps, stable tetravalent intermediates were optimized leading to the conclusion that an addition elimination mechanism is the most favourable. Once again one needs to keep in mind not only the limited accuracy of semi-empirical methods, but also the issues with a truncated model of the active site that only includes a few key residues and does not take the apo-enzyme into account. The issue is again raised in a 2002 article from the McCammon group.[146] In this article, they present the results of robust *ab initio* QM/MM simulations of the acylation step using Hartree Fock with singles points at a higher level. Those results, once again clearly point to an addition elimination mechanism with a stable tetravalent intermediate. The oxyanionic hole is identified in those simulation as having a critical role in the stabilization of the tetravalent intermediate and thus of the lowering of the energy barrier. This contribution of the oxyanionic hole and the literature dedicated to investigating it is discussed in a later paragraph

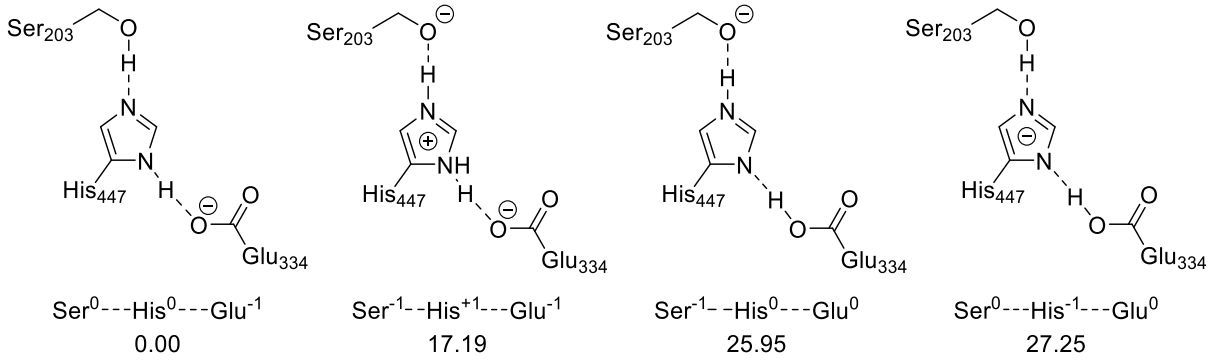
No evidence was ever provided for a SN_2 mechanism for the acylation nor the deacylation. The concurring evidence provided by different energy calculations methods and several models for active site simulations is compelling for an addition elimination mechanism for both steps (see **Figure III-2** and **Figure III-4**).

The second crucial mechanistic question on the catalytic cycle of AChE is the activation of the nucleophiles which can precede or be concurrent to the nucleophilic additions. The semi-empirical study of the catalytic cycle by Wang and co-worker also addressed this issue. In the first article the matter of stepwise versus concerted mechanism for the deprotonation of Ser203 by His447 in the acylation step is discussed.[144] The study concludes that the stepwise mechanism is less favourable than the concerted one because it is thermally prohibited to deprotonate Ser203 before the nucleophilic addition. The second article compares the stepwise and concerted mechanisms for the deprotonation of the water

[144] Q. Wang, H. Jiang, J. Chen, K. Chen, R. Ji, *Int. J. Quantum Chem.* **1998**, *70*, 515–525.
[145] Q. M. Wang, H. L. Jiang, K. X. Chen, R. Y. Ji, Y.-J. Ye, *Int. J. Quantum Chem.* **1999**, *74*, 315–325.
[146] Y. Zhang, J. Kua, J. A. McCammon, *J. Am. Chem. Soc.* **2002**, *124*, 10572–10577.

molecule in the deacylation step, in the addition-elimination mechanism hypothesis.[145] The study shows that a stepwise mechanism is thermally impossible due to the very high cost of the initial deprotonation of the water molecule. When this deprotonation occurs concurrently with the nucleophilic addition of the water molecule on the carbonyl of acylated AChE the energy barrier is lowered making this concerted pathway favourable. Both are subject to the same reservations as discussed above due to the semi-empirical and use of a truncated model system.

The catalytic triad is central in the activity of AChE and a well-studied substructure of AChE. Its protonation pattern is key to its function and the proton positioned "between" His447 and Glu334 has been the subject of several studies. The 1998 semi-empirical QM/MM study by V. V. Vasilyev focused in part on the protonation pattern of the catalytic triad.[143] Semi-empirical QM/MM was used to calculate the energy of the four possible protonation states (see **Scheme III-1**).



Scheme III-1. Possible protonation states for the catalytic triad of AChE with their respective energy with their respective energy calculated in PM3:OPLS. [143] energies in kcal.mol^{-1}.

The unprotonated Glu334 with mono protonated His447 is the preferred state for the catalytic triad of AChE. It was confirmed by further QM/MM calculations that showed that a proton transfer between His447 and Glu334 would destabilize the tetravalent intermediate in the acylation step.[146] The results are in conformity with the hypothesis derived from experimental data. X-Ray data on chymotrypsin showed that the glutamate or aspartate of the triad in likely to be deprotonated while the serine and the histidine would both be protonated (on a single protonation site for the histidine).[147] This was later confirmed by further X-Ray studies.[32]

[147] H. Tsukada, D. M. Blow, *J. Mol. Biol.* **1985**, *184*, 703–711.

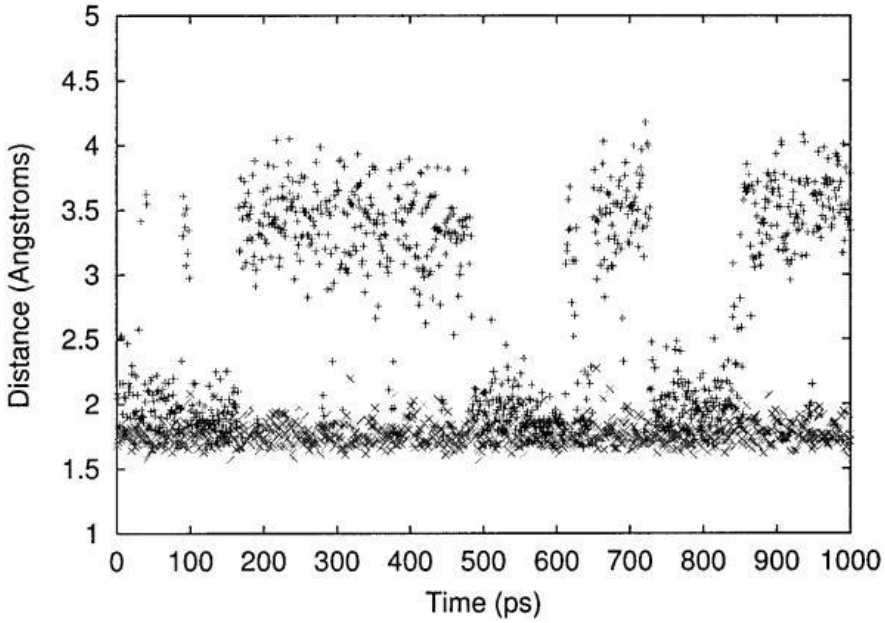


Figure III-5. Length of the two hydrogen bonds of the catalytic triad along a 1-ns trajectory. + refers to the Ser203-His447 hydrogen bond, x refers to the His447-Glu334 hydrogen bond. From ref [30]

To study the geometry of the catalytic triad, a simple static simulation is not always sufficient. The dynamics of the triad can be considered. The group of McCammon followed the length of the hydrogen bonds of the triad along classical molecular dynamics simulations of AChE.[30] This data is presented in **Figure III-5**. The hydrogen bond between His447 and Glu334 is very stable throughout the simulation while the hydrogen bond between Ser203 and His447 is much more unstable. It forms, breaks, and reforms several times in the 1000 ps trajectory illustrating the mobility of the triad. This mobility however is absent when AChE is complexed with ACh. In that case both hydrogen bonds have the same stability. This stability of the catalytic triad is promoted by nearby aromatic residues. As shown in another molecular dynamic study, Phe338 and Phe295 are trapping His447 in a position where the two hydrogen bonds are possible. The authors of the study postulate that this trapping is due to π- π and cation- π interactions.[61]

The postulated role of the oxyanionic hole in the reactivity of AChE is the stabilisation of transition states and reaction intermediates such as the tetravalent intermediate presented in **Figure III-2** and **Figure III-4**. Two sets of EVB studies were published in 1998 and 2000 to test this postulate. The EVB method can be summarized as a classical reactive molecular dynamic method and is detailed in the methods section. A reaction step whose energy in gas phase or aqueous solution is known is simulated in enzyme. It allows to study the influence of the enzyme, taking its dynamicity into account, on this reaction step. The first of those EVB studies focuses on the acylation[99] and the second on the deacylation.[148] Both were performed under the assumption that the mechanism of those steps is addition-elimination. They show that the enzyme, and more specifically by its oxyanionic hole, greatly stabilizes the tetravalent intermediate and transitions states leading to them.

[148] P. Vagedes, B. Rabenstein, J. Åqvist, J. Marelius, E.-W. Knapp, *J. Am. Chem. Soc.* **2000**, *122*, 12254–12262.

Classical molecular dynamics have also been used to study the role of the oxyanionic hole in the stabilization of this stable tetravalent intermediate and the transition state leading to it.[149] The structures of the substrate state, the transition state and the stable intermediate state were kept frozen while the rest of the enzyme, and the oxyanionic hole in particular, is free to move. This method is inherently flawed because the transition state cannot be represented in classical molecular mechanics in principle. In this instance it is a frozen geometry of a state with a lifetime of the order of the femtosecond kept frozen for several picoseconds. The length of the hydrogen bonds of the oxyanionic hole with the carbonyl oxygen of ACh was followed (see **Figure III-6**). What is shown is that in the reactant state, two of the hydrogen bonds of the oxyanionic hole are unstable. With the frozen transition state and the frozen stable intermediate, the hydrogen bonds are much more stable. The distance between the hydrogen and the oxygen of ACh's carbonyl rarely increases over 2.5 Å. The change in hydrogen bonding energy between the reactant state and the transition state (TS) show a 5.4 kcal.mol^{-1} stabilization in favour of the TS.

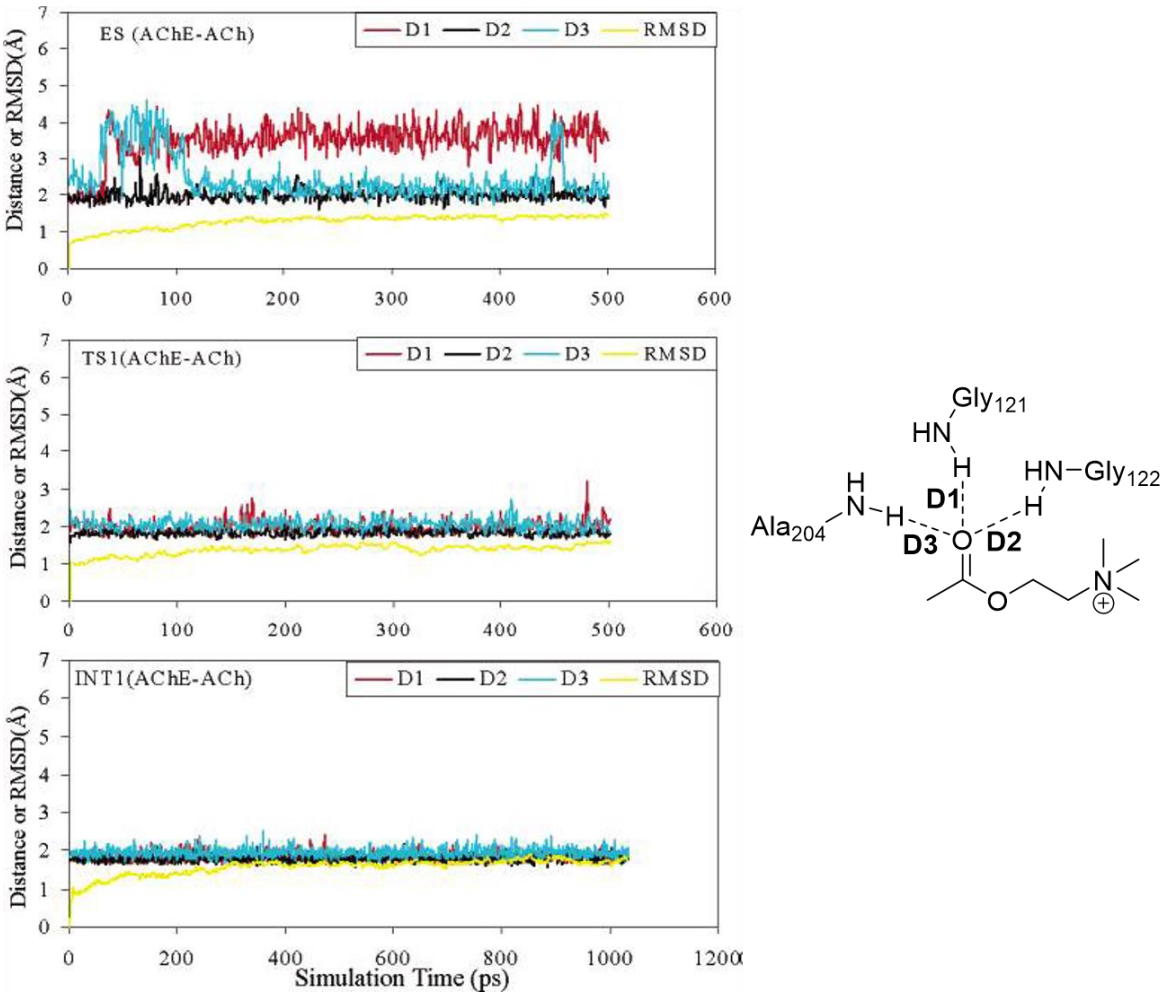


Figure III-6. Key distances between the carbonyl oxygen of ACh and the NH hydrogen of Gly121, Gly122, and Ala204 (D1, D2 and D3 respectively). The RMSD is the root-mean-square deviation of the backbone atoms of AChE from the initial structure. From ref [149]

[149] D. Gao, C.-G. Zhan, *J. Phys. Chem. B* **2005**, *109*, 23070–23076.

A glutamate in proximity of the catalytic triad, Glu202, has been identified as having a role in the binding of the choline moiety of ACh. The protonation state of that glutamate is a point of debate and the subject of some of the very first theoretical mechanistic studies of AChE's active site. EVB simulations of the acylation step showed a slightly lower energy barrier for unprotonated Glu202.[99] This is in line with previous results from classical simulations which indicated a much larger increase.[98d] The same type of EVB simulations on the deacylation step seemed to indicate that a protonated Glu202 would lower the energy barrier for deacylation.[150] The same article also presents Monte Carlo titration simulations that indicate that the acylated enzyme has a protonated Glu202. Two other articles presented results indicating that a protonated Glu202 would stabilize the acylation and deacylation more than a deprotonated Glu202.[151], [152]

The increase in computational power and the generalization of QM/MM and free energy methods derived from QM/MM allowed several research teams to obtain and publish partial or complete potential energy surfaces. Potential energy surfaces of the studies focused on acylation are presented in **Figure III-7**. **Figure III-8**, and **Figure III-10** present potential energy and free energy surfaces respectively of both the acylation and the deacylation. While they show many quantitative differences and even some qualitative ones, they converge on many mechanistic points. All agree on an addition-elimination mechanism for both the acylation and deacylation steps. There is also a consensus on the concerted activation of the nucleophile during the addition step by His447 for both nucleophiles (Ser203 for the acylation and H_2O for the deacylation). Finally, the kinetically limiting step is the nucleophilic addition during the acylation. It should be noted that for the studies that produced the potential energy surfaces presented in **Figure III-8** transition states for the elimination step of the acylation were not investigated. It was justified on the basis that the nucleophilic addition is the kinetically limiting step. The same assumption was made in those studies about the deacylation. The transition state for the elimination was not investigated because the addition was assumed to have a higher energy barrier. The key structures of the energy surfaces are shown in **Figure III-9**.

[150] P. Vagedes, B. Rabenstein, J. Åqvist, J. Marelius, E.-W. Knapp, *J. Am. Chem. Soc.* **2000**, *122*, 12254–12262.
[151] Y. Zhou, S. Wang, Y. Zhang, *J. Phys. Chem. B* **2010**, *114*, 8817–8825.
[152] S. V. Lushchekina, I. A. Kaliman, B. L. Grigorenko, A. V. Nemukhin, S. D. Varfolomeev, *Russ. Chem. Bull.* **2011**, *60*, 2196–2204.

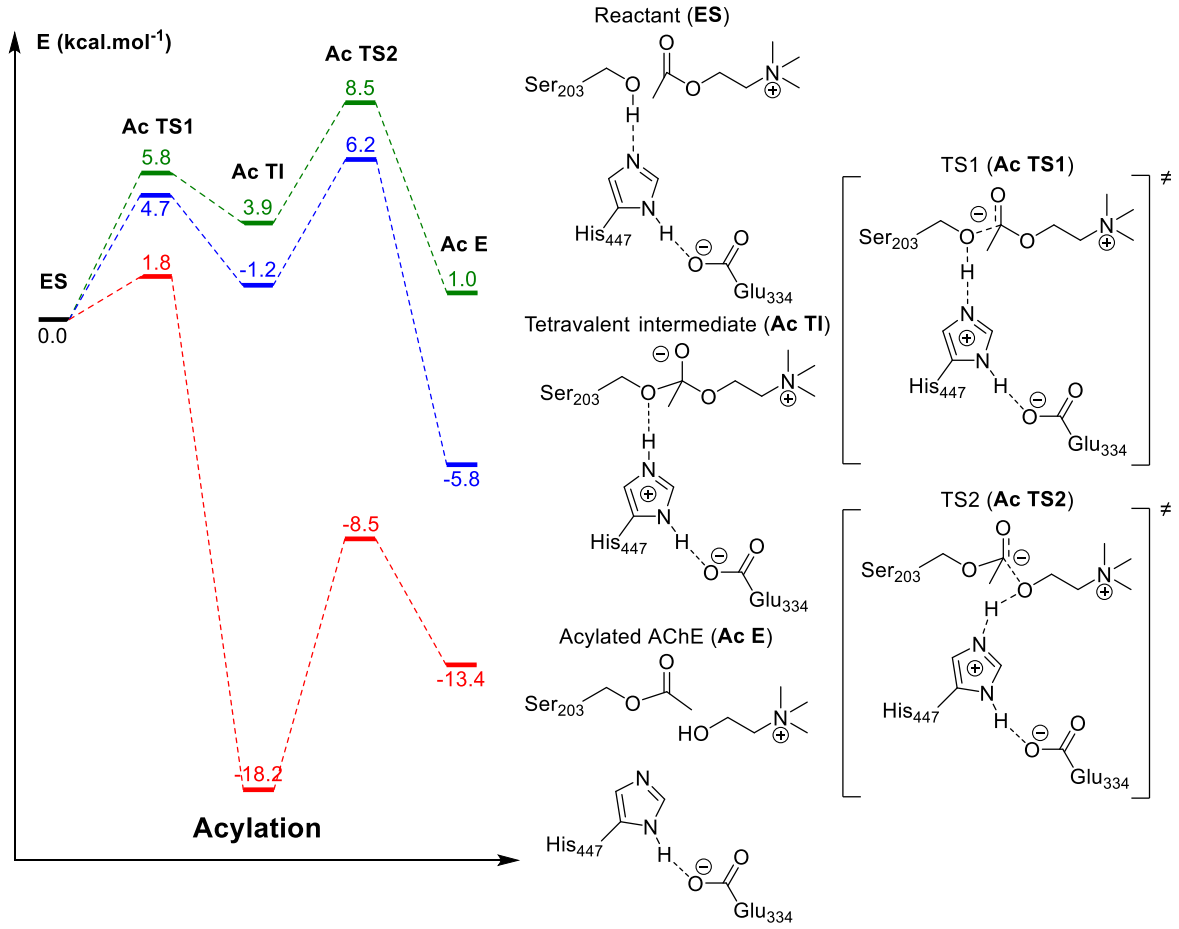


Figure III-7. Potential energy surface for the acylation step of ACh hydrolysis by AChE. The green surface comes from ref [153], the blue from ref [154] and the red from ref [152]. Key structures of the potential energy surface are to the side.

[153] J.-L. Fattebert, E. Y. Lau, B. J. Bennion, P. Huang, F. C. Lightstone, *J. Chem. Theory Comput.* **2015**, *11*, 5688–5695.

[154] A. V. Nemukhin, B. L. Grigorenko, D. I. Morozov, M. S. Kochetov, S. V. Lushchekina, S. D. Varfolomeev, *Chem. Biol. Interact.* **2013**, *203*, 51–56.

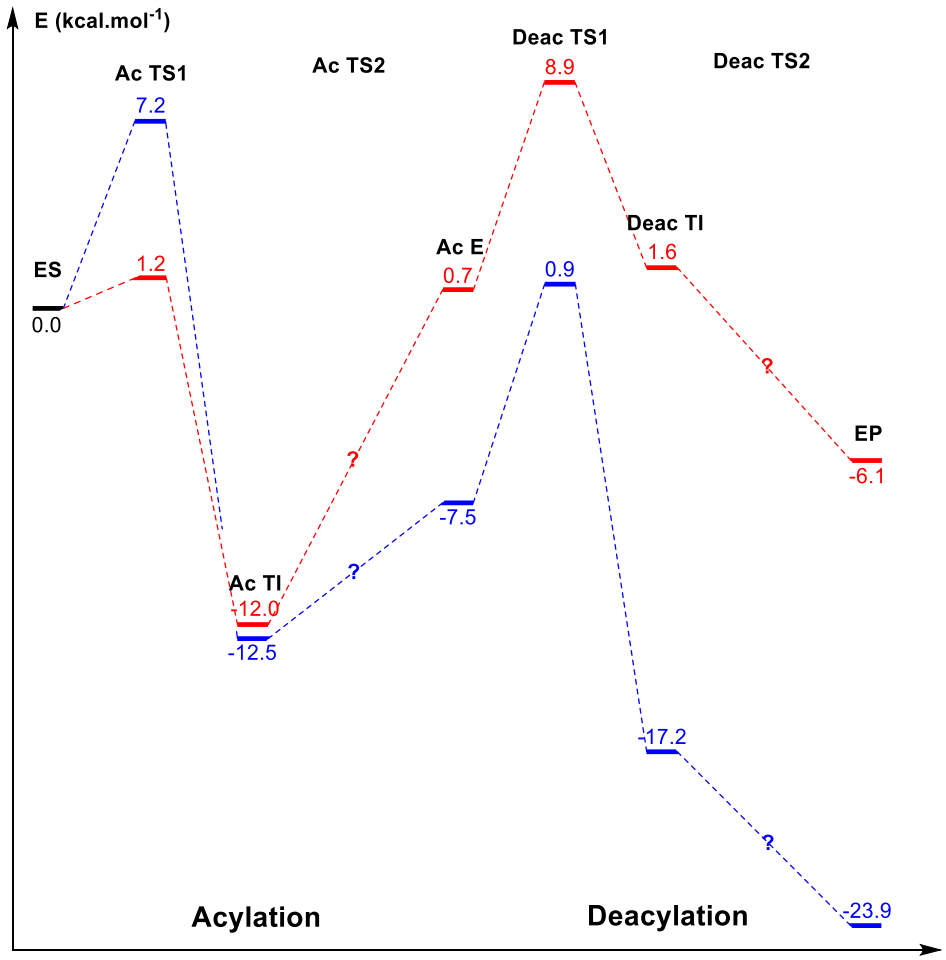


Figure III-8. Potential energy surface for both steps of ACh hydrolysis by AChE. The blue surface comes from ref [155] and the red from ref [156]. The key structures can be found in **Figure III-9**. Question marks indicate that the transition states were not searched for.

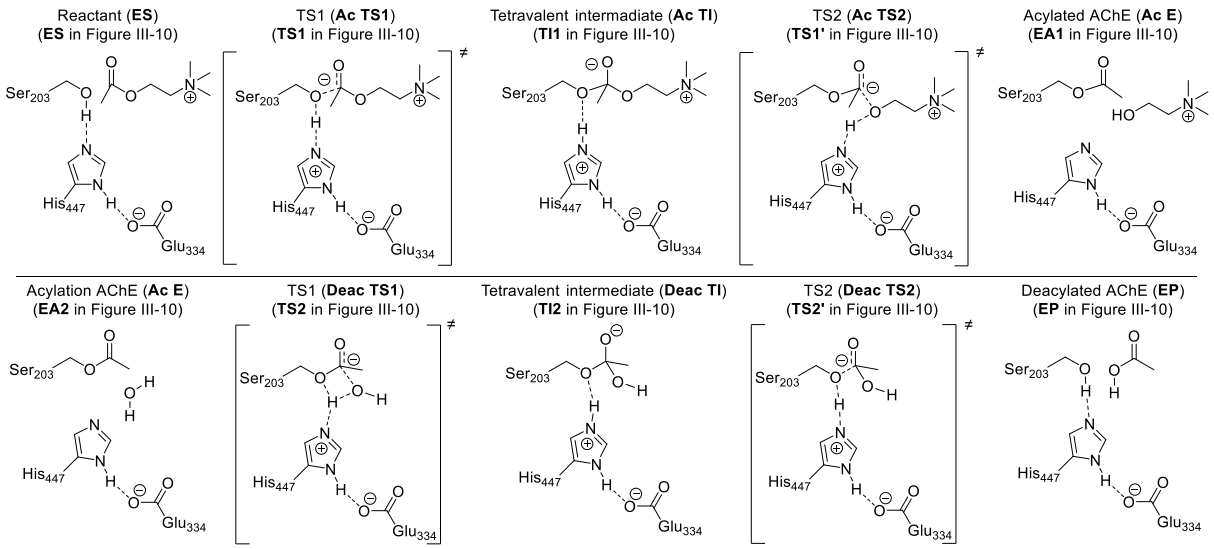


Figure III-9. Key structures of the potential energy surfaces displayed in **Figure III-8** and **Figure III-10**

[155] A. V. Nemukhin, S. V. Lushchekina, A. V. Bochenkova, A. A. Golubeva, S. D. Varfolomeev, *J Mol Model* **2008**, *14*, 409–416.

[156] O. Kwasnieski, Etude Théorique de La Réactivation de l'AChE Inhibée Par Le Tabun, UPMC, **2010**.

Most static QM/MM find the acylation to be exothermic but they only account for the enthalpic contribution to the free energy (both the red and blue curves in **Figure III-7** and the blue curve of **Figure III-8**. First-principle molecular dynamics simulations from Jean-Luc Fattebert with electrostatic embedding and an impressively large QM region finds the step to be very slightly endergonic.[153] Those simulations give free energy potential energy surfaces. Their results contradict the Born Oppenheimer molecular dynamic simulations of Zhou (See **Figure III-10**).[151] The energy from the calculations of Zhou is higher for the tetravalent intermediate and the TS leading to it, while the relative energy of the TS of the elimination is lower. The overall energy difference of the reaction is exergonic in the case of Zhou's simulations and endergonic in the case of Fattebert's simulations.

Fewer simulations have been reported on deacylation, maybe because the position of the reactive water molecule in the pre-reactional state is unclear.[156] Simulations to obtain a single potential energy surface for both steps prove to be even more challenging both from a technical and from a conceptual point of view. Nemukhin and co-workers do not provide an energy barrier for the second step of the acylation despite the endothermicity of the formation of the acylated enzyme (**Ac E** in **Figure III-7** and **Figure III-8**) on the basis of the postulated low energy of the barrier for this step. [155] This endothermicity might reflect the cost of positioning a water molecule for deacylation, a process difficult to simulate without extended sampling. This poses a technical problem that compelled the simulations of Zhou and co-workers to be split in two, the necessary water molecule being prepositioned prior to the deacylation simulations (See **Figure III-10**). [151]

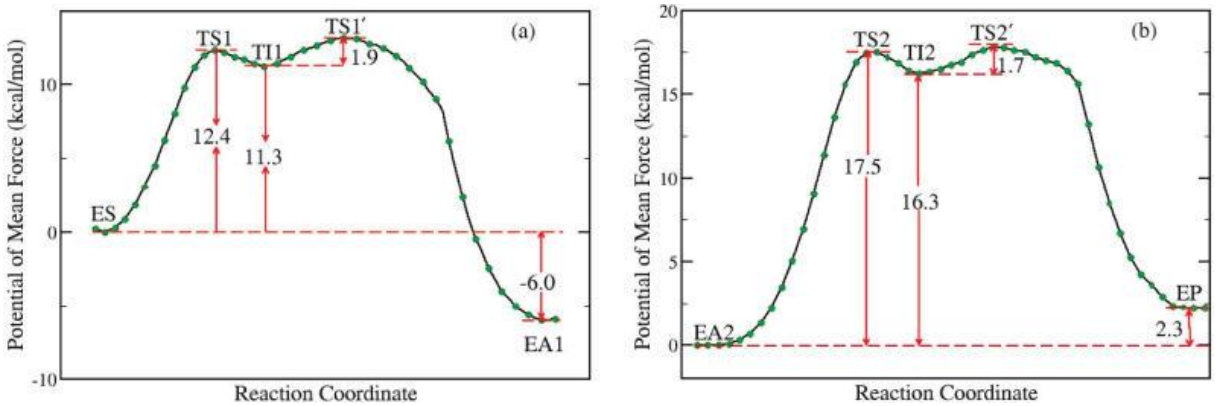


Figure III-10. Free energy profiles for the acylation step (a) and the deacylation step (b) from ref [151]. The key structures can be found in **Figure III-9**.

It has been postulated that Glu202 might have a role in the positioning of the water molecule, rationalizing its influence on the energy barrier of the deacylation.[155] The numerous possible avenues for a water molecule to enter the active site and access His447 muddles the issue. It makes the evaluation of the energy barrier of the positioning of the water molecule very difficult.

The energy difference of the deacylations step has been evaluated twice by static QM/MM and once by a free energy method (see **Figure III-8** and **Figure III-10b**). The QM/MM

simulations of Lushchekina and co-workers show a very large exothermicity while those reported in Ophélie Kwasnieski's thesis show a much more reasonable exothermicity. This difference has been traced back to a better choice of bonds crossing the QM/MM boundary. Similarly to the acylation, when the entropic factor is taken into account, like in Zhou's 2010 Born Oppenheimer molecular dynamic simulations, the step appears to be endergonic.[151] The similarities between the acylation and deacylation process are especially apparent when looking at **Figure III-10a** and **Figure III-10b** side by side. Not only do they go through the same mechanism and their nucleophile is activated at the same step, their respective free energy potential energy surface has an almost identical shape.

The overall picture drawn from the previous studies in the field is a two-step catalytic cycle (see **Figure III-11**). Both steps are going through a similar addition elimination mechanism in which the nucleophile is activated by being deprotonated concurrently to the nucleophilic addition. The nucleophiles of both steps are deprotonated by a histidine residue, His447, electrostatically assisted by the negatively charged Glu334 with whom His447 forms a very stable hydrogen bond. The tetravalent intermediates of both steps are stabilized by the oxyanionic hole which seems to be responsible to a large extend for the catalytic acceleration of the reaction. For both the acylation and the deacylation, the nucleophilic addition is the costlier step, with the one from the deacylation being the kinetically determinant step for the entire catalytic process.

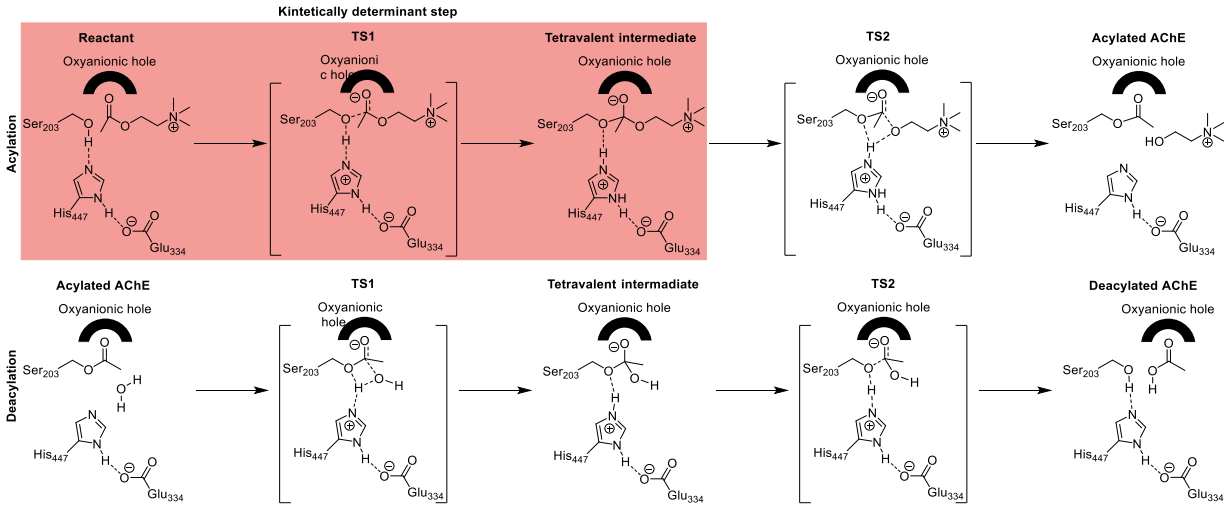


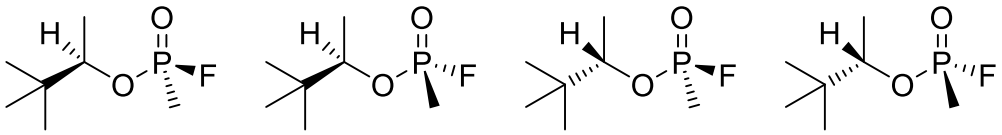
Figure III-11. Overall picture of the mechanism of the catalytic cycle of AChE.

Most of the non-catalytic activity of AChE shares main mechanistical traits with the regular catalytic activity, particularly the inhibition, which does, however, have its own particularities.

III.2. Inhibition by nerve agents

The description of the inhibition of AChE by organophosphate nerve agents and pesticides is challenging because of the large structural variety of the inhibitors. As the subject was tackled by various research teams with a wide range of methods of varying accuracy, not all of them studied the same organophosphates. The next section will attempt to establish the

common trends that unite all nerve agents, while still pointing out the standing out characters of some of the more peculiar ones like soman and its very specific aging mechanism.



Scheme III-2. Enantiomers of soman

The phosphorus centre of organophosphate nerve agents is chiral, as was detailed in the beginning of section II.2 of this manuscript. The enantiomers of individual nerve agents have varying reactivity towards AChE, usually favouring the S enantiomers. This enantioselectivity, often governed by nonbonding interactions between an enzyme and its substrate has been the subject of some of the first theoretical studies of the interaction of nerve agents with AChE. In 1993, Qian and Kovach performed molecular dynamic simulation to rationalize the increased reactivity of P_SC_R over P_SC_S (see **Scheme III-2**).[107]

The molecular dynamic simulations show that the P_SC_R enantiomer is associated with a rotation of His447 which prevents the formation of the favourable His447-Ser203 hydrogen bond. This favourable interaction stabilizes the P_SC_S complex more so than it does the P_SC_R. This hydrogen bond is also postulated to play a role in the reaction of AChE with organophosphates like the role it plays in the regular catalysis. Although this strictly classical study cannot be used to rationalize the mechanism, it can be postulated that if this His447-Ser203 hydrogen bond is relevant to the inhibition mechanism it could play a role into the specificity of AChE towards the P_SC_S enantiomer. This hypothesis will later be proven to be valid.

A few years later, Kovach was involved in another study that studied the interactions of P_RC_S and P_SC_S soman isomers with AChE.[157] The comparison of root of mean square deviation (RMSD) for all residues of AChE between P_RC_S soman-AChE complex and the P_SC_S soman-AChE complex provide a rationalization for the selectivity of AChE towards the P_SC_S isomer. The RMSD is on average higher for the P_RC_S soman-AChE complex, illustrating the greater displacement caused by this enantiomer, especially in key residues like His447. The energy necessary to move these side chains is linked to the increased energy cost associated with the interaction of the P_RC_S enantiomer over the P_SC_S enantiomer.

To positioning of VX in the active site of AChE was also studied by a combination of dynamic molecular mechanic simulations and static quantum mechanic calculations.[158] As was expected from the experimental data, (S)-VX docked more easily into the active site of AChE, as evidenced by the smaller RMSD of active site residues of the catalytic triad and the oxyanionic hole compared to (R)-VX. The ab-initio optimizations combined with the molecular dynamics allowed to establish a model of interaction of VX with AChE (see **Figure III-12**).

[157] A. Bencsura, I. Y. Enyedy, I. M. Kovach, *J. Am. Chem. Soc.* **1996**, *118*, 8531–8541.
[158] C. Albaret, S. Lacoutière, W. P. Ashman, D. Froment, P.-L. Fortier, *Proteins* **1997**, *28*, 543–555.

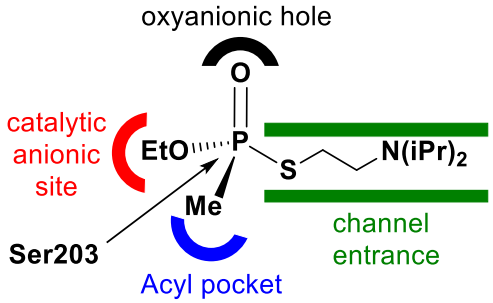


Figure III-12. interactions of (S)-VX with AChE.

In this model, the phosphoryl of VX fits into the oxyanionic hole, as did the phosphoryl of soman in the two previously mentioned molecular mechanics studies.[107], [157] This is hardly a surprise as the phosphoryl is an analogue of the carbonyl of ACh and its interactions with the oxyanionic hole are very stable in all molecular dynamics. The alkoxy substituent fits in the catalytic anionic site of AChE which is flexible and has more space than the acyl pocket to the opposite of the catalytic triad where the usually (for all four main nerve agents) smaller alkyl substituent is positioned. This orientation allows the leaving group, a bulky substituent in the case of VX, to be positioned in the channel pointed towards its entrance. In this position, it is ready for an in-line attack on the phosphorus by Ser203's side chain followed by its elimination. This model is likely to be applicable to other nerve agents and has been used to position nerve agent for simulation purposes in many studies. However, as was demonstrated by static QM/MM simulations, not all inhibitors favour in-line (anti-) attack of the oxygen on the phosphorus, some favour the syn nucleophilic addition.[109] This surprising result questions the applicability of the interaction mode shown in **Figure III-12** to all nerve agents. Static QM/MM calculations were used to obtain the potential energy surface for all four combinations of enantiomer and position of the leaving cyano group. For all four surfaces, the kinetically relevant step was determined to be the elimination of the cyano leaving group. The comparison of those four eliminations steps is presented in **Figure III-13**. The Syn-S and Anti-R are the two enantiomers that can produce the AChE-tabun adduct corresponding to the experimental X-Ray data.[159] What the data shows is a heavily favoured Syn positioning for the cyano leaving group as the step is exothermic only under the condition of a Syn approach. The expected in-line approach could still be the preferred one for other organophosphates such as VX which has a very bulky leaving group, but this research shows that the specificities of every nerve agent should not be overlooked.

[159] E. Carletti, H. Li, B. Li, F. Ekström, Y. Nicolet, M. Loiodice, E. Gillon, M. T. Froment, O. Lockridge, L. M. Schopfer, et al., *J. Am. Chem. Soc.* **2008**, *130*, 16011–16020.

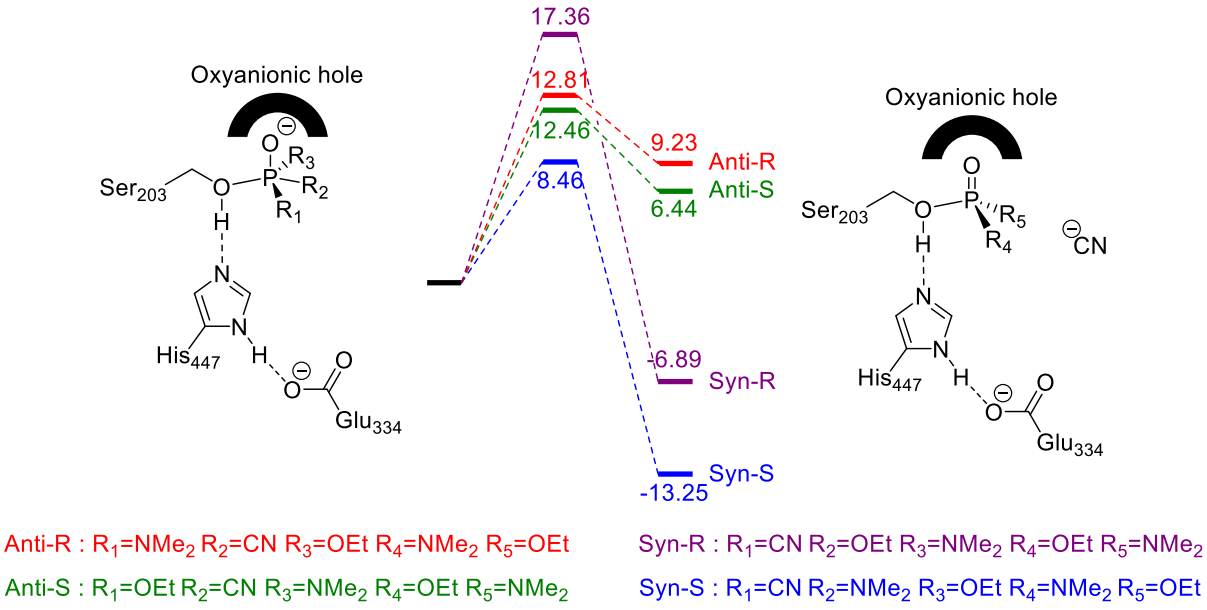


Figure III-13. Energy profiles for the elimination step of inhibition by soman depending on leaving group placement and enantioselectivity. Data in kcal.mol^{-1}. Drawn from the data in ref [109]

A significant number of studies performed on the reactivity of nerve agents, and similar compounds, towards AChE have been performed with static DFT calculations of truncated models of the active site without taking the apoenzyme into account.[160] In addition to the limitations inherent to DFT and to static simulations, the use of truncated model systems open the possibility for the reaction to reach states that would otherwise be unreachable in the enzyme, in particular when a poor choice of constraints is made in the geometry optimization. A particularly egregious example is a 2016 article from Rathnayake and Northrup in which the phosphoryl group of the organophosphate become protonated by His447 despite His447 being positioned opposite of the oxyanionic hole the phosphoryl is bound in.[160d] Despite those limitations and caveats, those studies have fuelled the hypothesis of an addition elimination mechanism for the covalent inhibition of AChE by Sarin and VX. The nucleophilic addition produces a stable intermediate with a trigonal bipyramidal geometry than can then undergo elimination of a leaving group. More recently static and dynamic QM/MM simulations have been performed and published that confirm the addition elimination mechanism as well as provide credible potential energy surfaces for the reaction.[161], [162], [163]

The kinetically determinant step for the inhibition of AChE by nerve agents is the step with the highest energy barrier. As we discussed earlier, the work of O. Kwasnieski showed the elimination step to be the more costly for tabun.[109] The QM/MM molecular dynamic

[160] a) J. Šečkutė, J. L. Menke, R. J. Emnett, E. V. Patterson, C. J. Cramer, *J. Org. Chem.* **2005**, *70*, 8649–8660. b) J. Wang, J. Gu, J. Leszczynski, *J. Phys. Chem. B* **2006**, *110*, 7567–7573. c) J. Wang, J. Gu, J. Leszczynski, *J. Phys. Chem. B* **2008**, *112*, 3485–3494. d) L. K. Rathnayake, S. H. Northrup, *Comput. Theo. Chem.* **2016**, *1088*, 9–23.

[161] J. M. Beck, C. M. Hadad, *Chem. Biol. Interact.* **2010**, *187*, 220–224.

[162] G. S. Sirin, Y. Zhang, *J. Phys. Chem. A* **2014**, *118*, 9132–9139.

[163] B. J. Bennion, S. G. Essiz, E. Y. Lau, J.-L. Fattebert, A. Emigh, F. C. Lightstone, *PLoS One* **2015**, *10*, e0121092.

study of the inhibition of AChE by soman (see free energy surface in **Figure III-14**) showed that both steps have low energy barriers which would point towards a diffusion limited inhibition process. Interestingly, the elimination was again showed to be more costly than the nucleophilic addition contrary to the catalytic cycle where the addition steps were more costly. The static QM/MM simulations of the phosphonylation of AChE by sarin however contradicted this trend. In their 2010 article, Beck and co-workers presented research indicating a much greater energy barrier for the nucleophilic addition.[161] While this difference could be attributed to the specificity of sarin it is also possibly a product of the methodology used in that study. Namely, the oxyanionic hole, which has been shown in every other study of inhibition by organophosphate to harbour the phosphoryl of the nerve agent, is not included in the QM region of the QM/MM model used. The role of the oxyanionic hole in stabilizing both the trigonal bipyramidal intermediate obtained after nucleophilic addition of Ser203 on the phosphorus, and the transition state leading to it, is well documented. The barrier for the nucleophilic addition could therefore be unrealistically increased by not including that structure in the QM region of the QM/MM model.

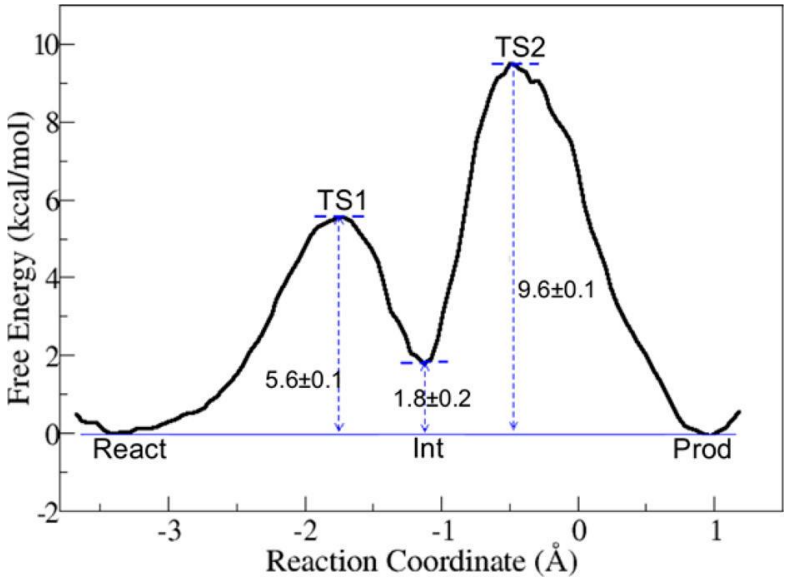


Figure III-14. Free energy surface for the phosphonylation of AChE by soman. The reaction coordinate is the linear combination of key reaction distances, see ref [162], from which this figure comes from.

Before addressing the role of some key residues in the inhibition process, a 2015 article published in *PLoS One* will be discussed.[163] To summarize the findings of that study, the AChE-soman adduct has a more stable active site region than uninhibited AChE. The long-term presence of a ligand on Ser203 and in the oxyanionic hole stabilizes both structures. This stabilization seems to have a ripple effect in the dynamics of the entire enzyme. It is making the active site a little bit less accessible from the channel but increases the mobility of the side chains of the backdoor. The tensing of the geometry of AChE as it is being inhibited by soman could also influence the protonation state of nearby Glu202. Unfortunately, neither this point nor the influence the protonation state of this residue has on inhibition was investigated. Most of the simulations of the inhibition of AChE were performed with an unprotonated Glu202[107],

[157], [109], [161], [163], with a notable exception.[162] The full static and dynamic QM/MM simulations of the inhibition have implicated Tyr124 in assisting the departure of the fluoride leaving group in sarin and soman.[161], [163], [162] An hydrogen bond is formed that stabilizes the fluoride by potentially over 10 kcal.mol^{-1}.

After the inhibition, before encountering a reactivator, the AChE-nerve agent adduct is not completely inert. The two possible natural outcomes, without the use of a reactivator, are the possible reactivation by a water molecule and the dealkylation of the organophosphate. The spontaneous reactivation of AChE by a water molecule has only been the subject of a single theoretical study of its mechanism.[164] The methodology employed was static QM/MM with DFT for the QM region. In this work, it is postulated that a water molecule can be deprotonated by an unprotonated His447 simultaneously to the nucleophilic addition of said water molecule on the phosphorus. Glu202 is unprotonated in those simulations and seem to only play a role of positioning the water molecule. The reaction goes through a reorganization of the reaction intermediate as can be seen in **Figure III-15**. This reorganisation, known as a Berry pseudorotation[165], changes the axis of the bypiramidal geometry of the phosphorus. From a pyramidal basis constituted of atoms P, O, OSer203, and OMe with an axis constituted of the NH_2, the phosphorus and the oxygen of the water molecule we go to a basis constituted of atoms P, O, OH and NH_2 with an axis made of OMe and OSer203. The bonds between the axial atoms are weaker and so this reorganization allows the bond between the phosphorus and the oxygen of the serine to be in the axis of the bipyramid and to be broken in the following elimination step.

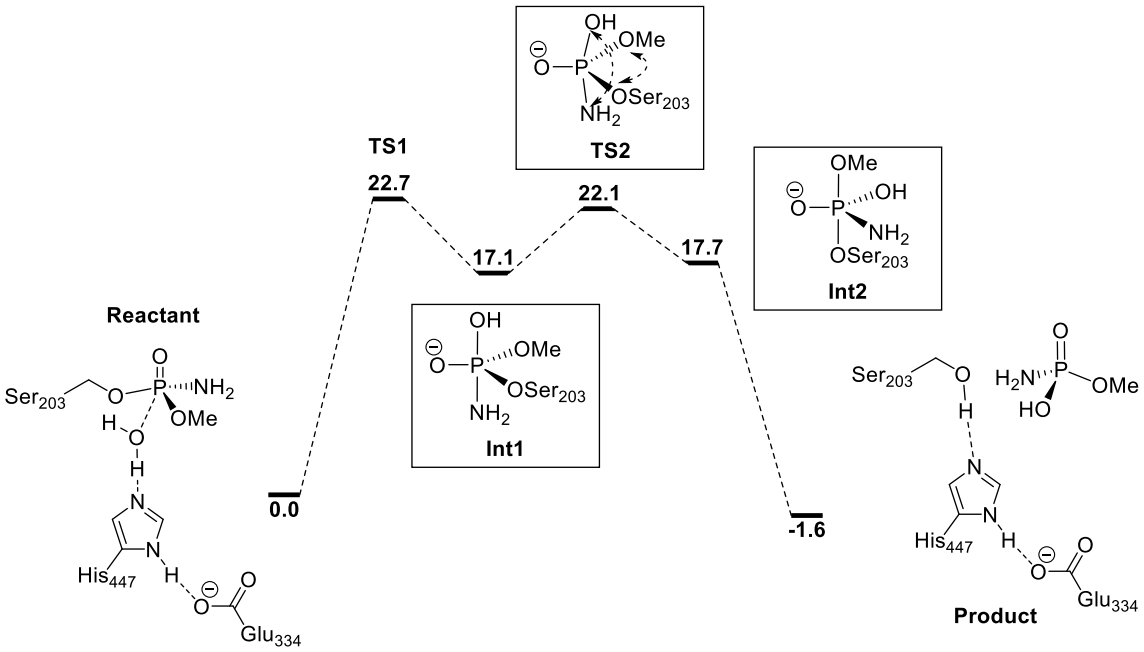


Figure III-15. Potential energy surface for the spontaneous reactivation by a water molecule of (-)-methamidophos inhibited AChE. Energies in kcal.mol^{-1}. Drawn from the data in ref [164]

[164] Y. Li, X. Sun, L. Du, Q. Zhang, W. Wang, *Computational and Theoretical Chemistry* **2012**, *980*, 108–114.
[165] R. S. Berry, *J. Chem. Phys.* **1960**, *32*, 933–938.

The spontaneous reactivation seems to go through an addition elimination mechanism and to have, as would be expected for such a slow process, a high energy barrier. The reaction of water molecule with a nerve agent-AChE adduct does not always yield the reactivated product, it often produces the dealkylated nerve agent adduct.[166]

The dealkylation of soman and tabun have been the subject of many theoretical and experimental studies. The focus on those two nerve agents is explained by the possibility for deamination opened by tabun and the speed of dealkylation in the case of soman, making both interesting study cases. Sarin and VX have had attracted less interest but can also be discussed by analogy with the previous two.

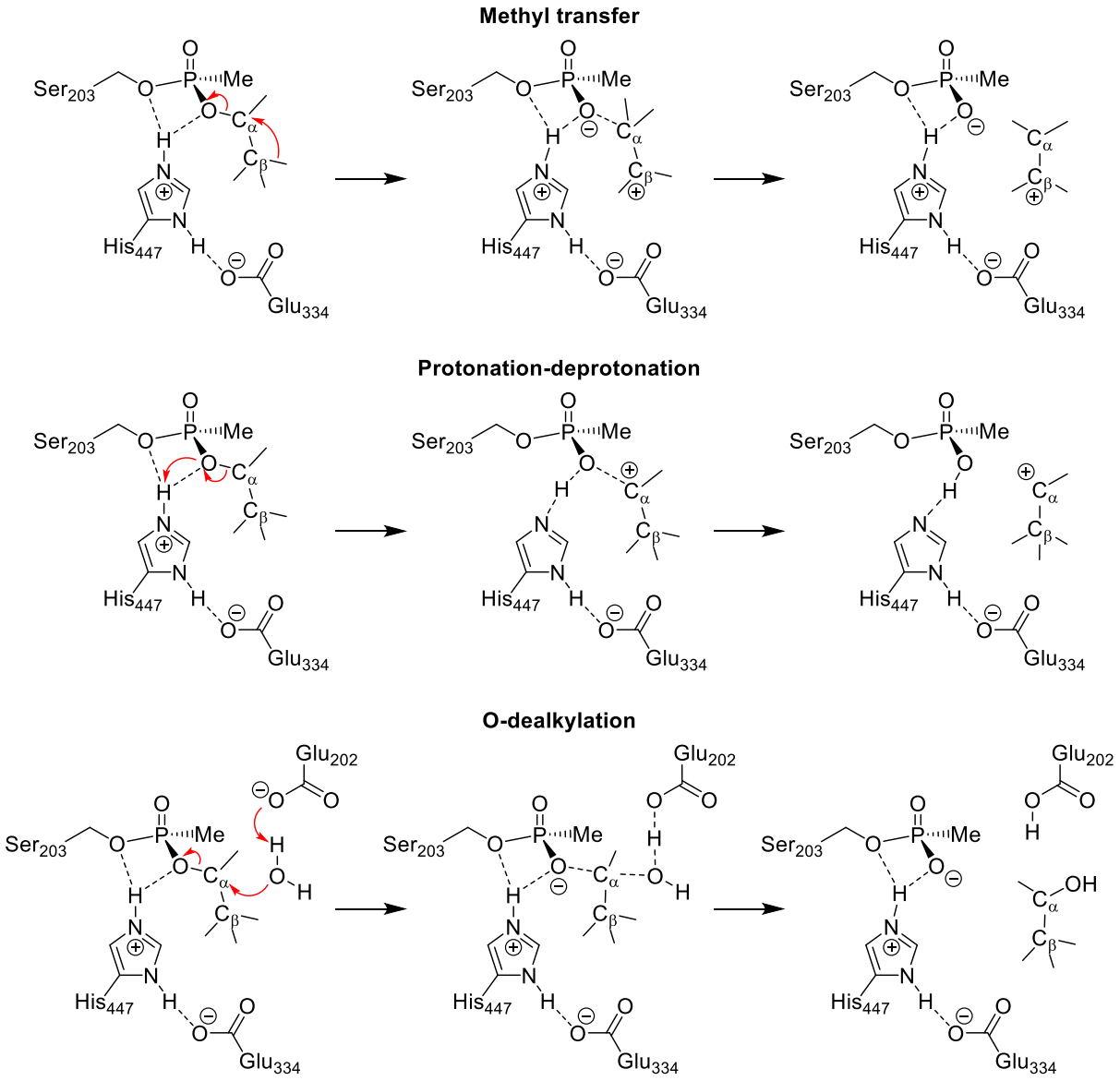


Figure III-16. Dealkylation mechanism hypothesis.

[166] H. O. Michel, B. E. Hackley, L. Berkowitz, G. List, E. B. Hackley, W. Gillilan, M. Pankau, *Arch. Biochem. Biophys.* **1967**, *121*, 29–34.

The dealkylation of soman is relatively fast, with a rate constant of the order of 10^{-1} transformations per minute. As such it has been the subject of much discussion in the literature. Experiments in inactivating His447[167] as well as pH dependence studies[168] point towards a major role of protonated His447 in the dealkylation. The mutation of Glu202 as well as the mutation of Glu450 have been shown to significantly decrease the rate of dealkylation, implicating those residues in the process as well.[169] Three mechanisms are commonly proposed for the dealkylation of soman (see **Figure III-16**).

The methyl transfer, or push pull, mechanism hypothesis has been proposed on the basis of experimental studies[170] and has been regarded as the most probable of the three propositions.[171] This mechanism is a [1,2] sigmatropic shift of a methyl.[172] The methyl on the β carbon of the pinacolyl migrates to the α carbon while the carbon-oxygen bond between the pinacolyl oxygen and the α carbon breaks. In this case a methyl group is displaced instead of a hydrogen because there are no hydrogen atoms available on the β carbon. The increased electron density on the pinacolyl oxygen would be stabilized by the positive charge of His447 while the carbenium on the β carbon would be stabilized by an unprotonated Glu202. This methyl migration mechanism is the only mechanism that can rationalize the products of dealkylation (see **Figure III-17**) observed with mass spectrometry.[173]

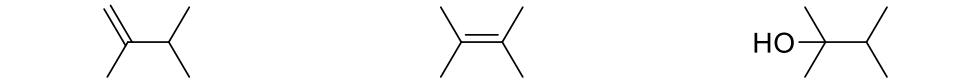


Figure III-17. Characterized products of soman dealkylation

The protonation deprotonation hypothesis suggests that the protonation of the pinacolyl oxygen by His447 would initiate the elimination of the alkyl chain with a carbenium on the α carbon. The third hypothesis is the O-dealkylation in which a water molecule is deprotonated to perform a nucleophilic addition on the α carbon, associated with the elimination of the pinacolyl oxygen. Ab initio QM/MM molecular dynamics simulations were performed by the group of Yingkai Zhang to compare those three possible mechanisms.[174] The protonation-deprotonation mechanism was quickly eliminated as very costly and producing an unstable product that would revert back to the reactants. More interestingly, the simulations showed that in the hypothesis of a protonated Glu202, the O-dealkylation pathway would be more favourable that the methyl transfer mechanism (see **Figure III-18**).

[167] G. Beauregard, J. Lum, B. D. Roufogalis, *Biochem. Pharmacol.* **1981**, *30*, 2915–2920.
[168] H. A. Berman, M. M. Decker, *J. Biol. Chem.* **1986**, *261*, 10646–10652.
[169] A. Ordentlich, C. Kronman, D. Barak, D. Stein, N. Ariel, D. Marcus, B. Velan, A. Shafferman, *FEBS Lett.* **1993**, *334*, 215–220.
[170] A. Shafferman, A. Ordentlich, D. Barak, D. Stein, N. Ariel, B. Velan, *Biochemical Journal* **1996**, *318*, 833–840.
[171] I. J. Enyedy, I. M. Kovach, A. Bencsura, *Biochem. J.* **2001**, *353*, 645–653.
[172] F. Bernardi, M. A. Robb, H. B. Schlegel, G. Tonachini, *J. Am. Chem. Soc.* **1984**, *106*, 1198–1202.
[173] C. Viragh, I. M. Kovach, L. Pannell, *Biochemistry* **1999**, *38*, 9557–9561.
[174] G. S. Sirin, Y. Zhou, L. Lior-Hoffmann, S. Wang, Y. Zhang, *J. Phys. Chem. B* **2012**, *116*, 12199–12207.

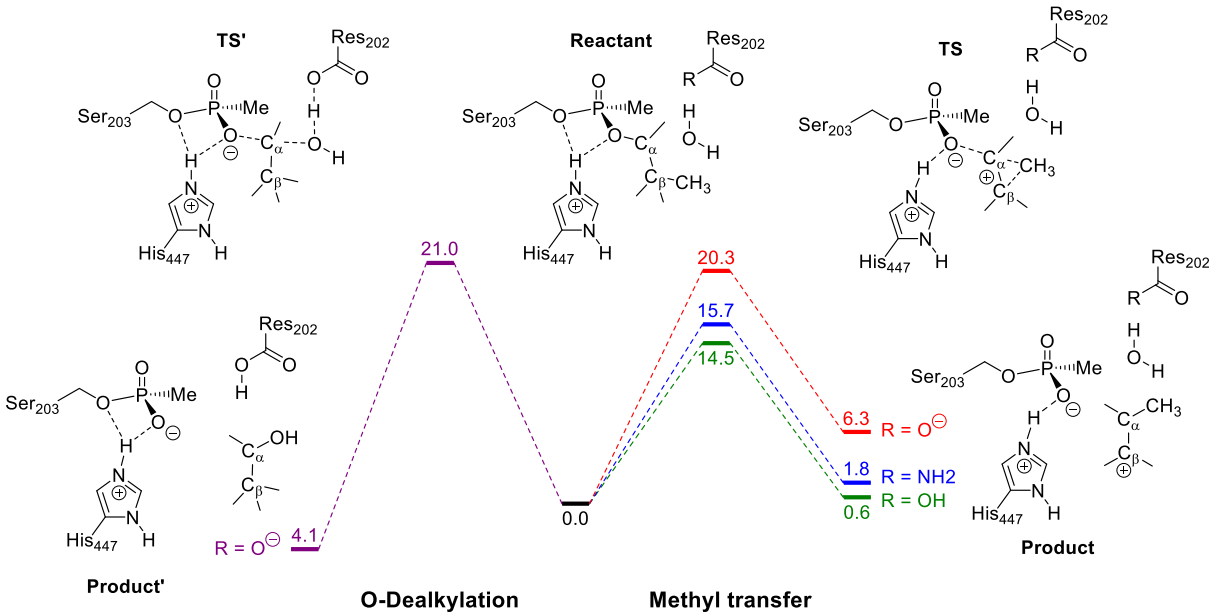


Figure III-18. Free energy surfaces for the O-dealkylation mechanism when Glu202 (noted Res202) is unprotonated, and the methyl transfer mechanism in the cases of protonated Glu202 (green), unprotonated Glu202 (red), and Gln202 (blue). Energies in kcal.mol^{-1}. Drawn from the data in ref [174]

The strength of this work is the comparison of the dealkylation in unprotonated Glu202, protonated Glu202 and Glu202Gln conditions. Among all hypothesis, the methyl transfer mechanism, in the case of a protonated Glu202, seems to be the more favourable pathway. This result is surprising as an unprotonated Glu202 was expected to assist the polarization of the carbon oxygen bond between the pinacolyl oxygen and the alkyl chain of the pinacolyl. The simulations showed no such role for Glu202. The expectation that an unprotonated Glu202 would stabilize the imidazolium in its polarizing role was also defeated. The important role of His447 in polarizing the carbon oxygen bond between the pinacolyl oxygen and the alkyl chain of the pinacolyl was further confirmed by a static DFT study on a truncated active site model in 2013.[175] The methyl transfer mechanism characterized in the work of Yingkai Zhang's group is a two-step process. The methyl transfer takes place concurrently to the cleavage of the C-O bond and is then followed by a hydration or a deprotonation step. This mechanism however has always been shown to be completely specific to soman-inhibited AChE.

All crystallographic studies have indicated that Sarin and VX as well as most of the O-alkyl substituted organophosphate age through the O-dealkylation mechanism.[176] The first notable exception was soman, the second is tabun. The results of mass spectrometry experiment showed that the dimethylamine group of tabun could be eliminated through P-N

[175] N. B. Chandar, B. Ganguly, *Chem. Biol. Interact.* **2013**, *204*, 185–190.

[176] a) C. B. Millard, G. Koellner, A. Ordentlich, A. Shafferman, I. Silman, J. L. Sussman, *J. Am. Chem. Soc.* **1999**, *121*, 9883–9884. b) A. Hörnberg, A.-K. Tunemalm, F. Ekström, *Biochemistry* **2007**, *46*, 4815–4825. c) F. Ekström, A. Hörnberg, E. Artursson, L.-G. Hammarström, G. Schneider, Y.-P. Pang, *PLoS One* **2009**, *4*, e5957.

bond scission instead of the alkyl chain of the alkoxy substituent.[177] This deamination of the nerve agent is specific to tabun. However, careful studies of crystal structures seemed to indicate that for tabun inhibited AChE, aging would go through the regular O-dealkylation pathway.[159] This pathway was modelled using a static QM/MM methodology.[178] A water molecule performed a SN_2 reaction on the α carbon of the ethoxy substituent while being deprotonated by Glu202 in a single step. The calculated energy barrier was of 19.9 kcal.mol^{-1}.

While there was a lack of structural evidence for the deamination in AChE, it was found to be possible for BuChE, which is very closely related to AChE. [179] Clever DFT simulations on a truncated model system of the active site that can be applied to both AChE and BuChE were performed on several possible pathways.[180] The reactivity of the R-tabun-ChE adduct (not necessarily obtained from R-tabun) with an hydroxide depends on the side of the tabun adduct that the hydroxide faces at the moment of the nucleophilic addition (see **Figure III-19**). Position A results in the nucleophilic addition of the hydroxide on the ethoxy to perform the O-dealkylation. Position B results in the nucleophilic addition of the hydroxide on the phosphorus and produces the standard deamination. Position C results in the nucleophilic addition of the hydroxide on the phosphorus and produces the deamination through rearrangement of the trigonal bipyramidal intermediate as suggested by Nachon and co-workers in 2010.[179] Position D was not investigated but is also a possibility. What calculations showed is that the deamination through rearrangement of the trigonal bipyramidal intermediate is the most favourable pathway. The determining energy barrier for this pathway is 12.1 kcal.mol^{-1}. The standard deamination is the least favourable pathways with a determining energy barrier of 21.1 kcal.mol^{-1}. The O-dealkylation has a determining energy barrier of 19.7 kcal.mol^{-1}.

[177] D. Barak, A. Ordentlich, D. Kaplan, R. Barak, D. Mizrahi, C. Kronman, Y. Segall, B. Velan, A. Shafferman, *Biochemistry* **2000**, *39*, 1156–1161.

[178] Y. Li, L. Du, Y. Hu, X. Sun, J. Hu, *Can. J. Chem.* **2012**, *90*, 376–383.

[179] F. Nachon, E. Carletti, F. Worek, P. Masson, *Chem. Biol. Interact.* **2010**, *187*, 44–48.

[180] M. K. Kesharwani, T. Bandyopadhyay, B. Ganguly, *Theor. Chem. Acc.* **2012**, *131*, 1175.

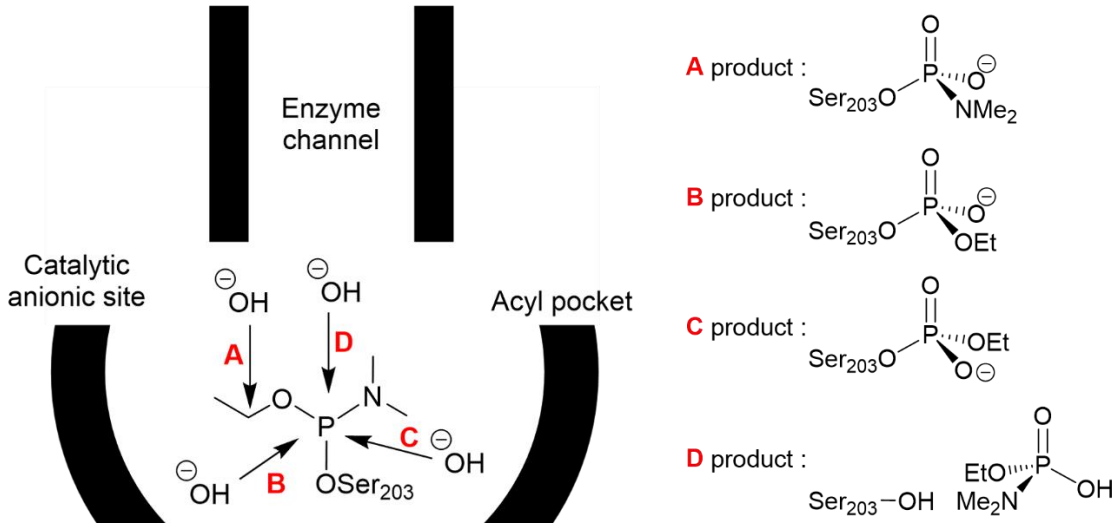


Figure III-19. Possibilities for a nucleophilic addition of a hydroxide on a (R)-tabun-ChE adduct. Top view with the phosphoryl group not shown

The logical conclusion would be that the R-tabun adduct goes through deamination and it does line up with the experimental evidence for BuChE, but not for AChE. It is however known in the literature that the acyl pocket of BuChE is much more open than that of AChE and so the position C might not be available for a water molecule to perform the deamination.[61] There is an effective dynamic control of dealkylation in tabun-inhibited AChE. The remaining lowest pathway would then be O-dealkylation as indicated by experimental data.

The common outcome of all aging mechanisms described here is the production of an oxygen anion of the nerve agent AChE adduct. This anion forms a very stable hydrogen bond with the imidazolium of His447 and prevents completely the reactivation of AChE by any means, including oximes.[181]

III.3. Reactivation by oximes

The fundamental goal of the study of the reactivation of AChE is the design of new, efficient, and universal reactivators. A perfect reactivator binds easily to AChE and goes down the channel of the enzyme to the catalytic site. It reacts easily on the phosphorus of any nerve agent, including tabun which is particularly hard to reactivate. The design of such a reactivator needs to be guided by a profound understanding of the reactivation process and its parameters.

The binding of the oxime to AChE and its progression along the channel to reach the active site of inhibited AChE has been the subject of multiple molecular dynamic simulations. Classical molecular mechanics are particularly adequate for this process which does not involve bond breaking or formation and is too long to be captured by static simulations. The group of Figueroa-Villar, in Brazil, is one of the most active in the field of AChE reactivation when it comes to simulations. One of their earliest studies was a set of molecular dynamic

[181] Y. Segall, D. Waysbort, D. Barak, N. Ariel, B. P. Doctor, J. Grunwald, Y. Ashani, *Biochemistry* **1993**, *32*, 13441–13450.

simulations of 2-PAM and two analogues of 2-PAM, DZP, a non-cationic analogue of 2-PAM, and the oximate form of DZP (DZP_{anion}) were placed at different points in the channel of AChE.[182] The main results are presented in **Figure III-20**.

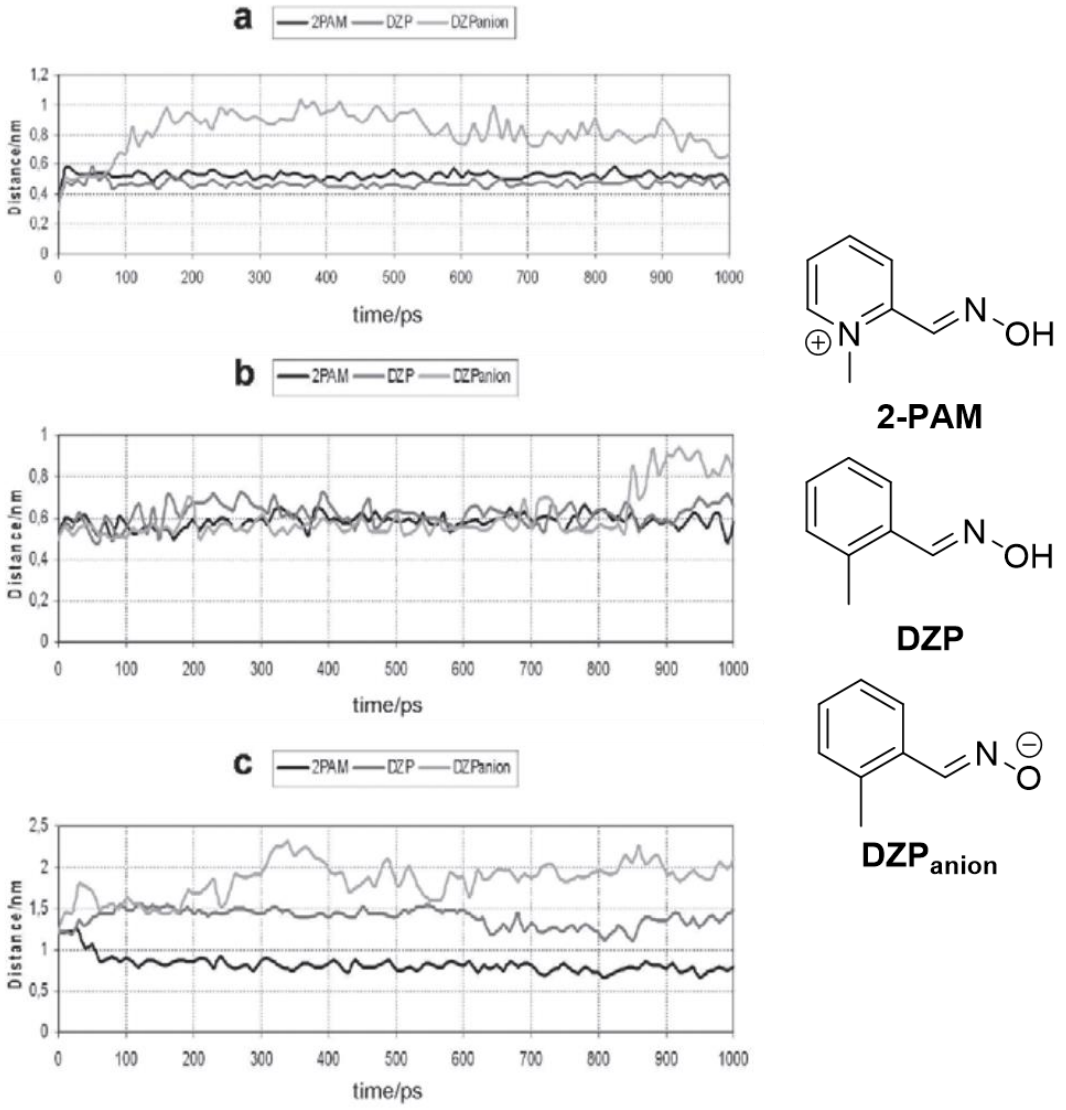


Figure III-20. Distance variation between the phosphorus atom of the nerve agent inhibitor and the oxygen of the oxime. 2-PAM is in black, DZP in dark gray and DZPanion in light gray. A: starting distance 3 Å, in the active site. B: starting distance 5 Å, in the middle of the channel. B: starting distance 8 Å, at the entrance of the active site. From ref [182]

Figure III-20 shows that when placed in the active site the oximate form of the non-cationic analogue of 2-PAM will move away from the nerve agent adduct (Ser-OP). When placed further away both oxime and oximate forms of DZP move away from the active site. In all three instances 2-PAM either holds in proximity to the nerve agent or moves closer. The article also shows the interactions of the reactivators when docked 5 Å away from the nerve agent. Only 2-PAM is seen successfully interacting with Trp86. This study reinforces the hypothesis that the positive charge of the pyridinium of 2-PAM is a good mimic of the

[182] A. da S. Gonçalves, T. C. C. França, A. Wilter, J. D. Figueroa-Villar, *J. Braz. Chem. Soc.* **2006**, *17*, 968–975.

ammonium of ACh for the binding of 2-PAM to the catalytic anionic site. The oximate form of DZP, DZP_{anion} struggles even more than DZP to bind within the active site or to progress through the channel. It could indicate that oximes are more likely to enter the active site in oxime as opposed to oximate form. It could have been more clearly demonstrated if this set of simulations also included the oximate form of 2-PAM and 2-pyridinealdoxime, without the methyl group on the pyridinium. In a pair of studies, steered molecular dynamics as well as docking simulations were used to compare four different reactivator prototypes and some designs based on those prototypes **Figure III-21**.[183]

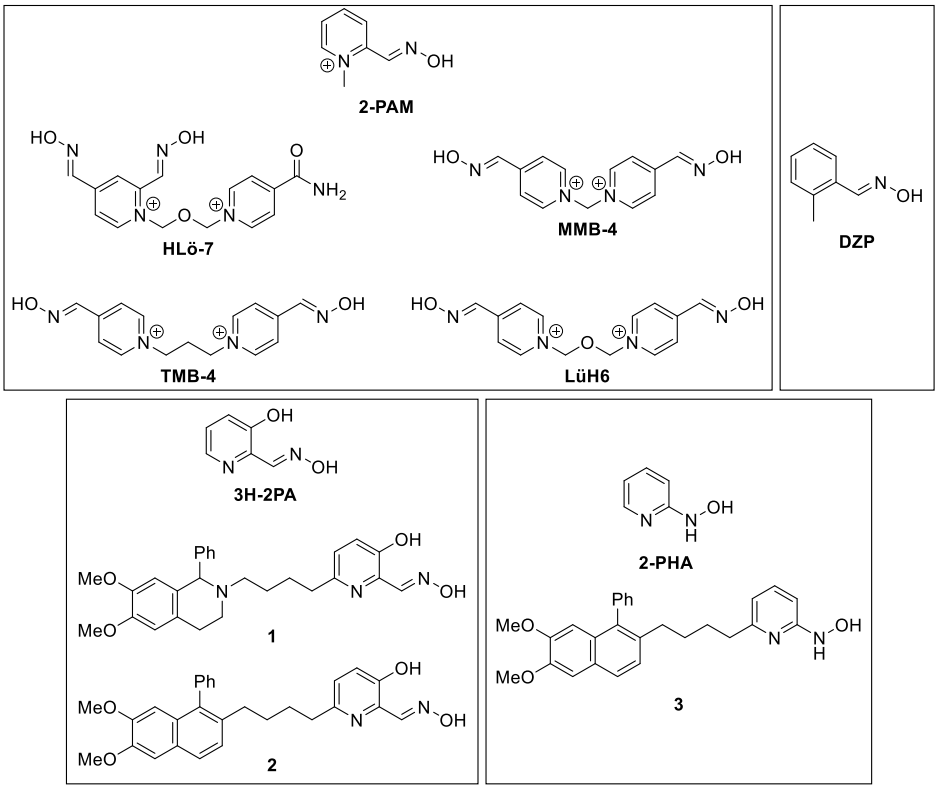


Figure III-21. Reactivators

The binding data for those reactivators can be found in **Table III-1**. Of the four reactivator prototypes 2-PAM has the best binding. Without the cationic charge of the pyridinium the binding of DZP, 3H-2PA, and 2PHA is greatly reduced which confirms the results of the 2006 study of Gonçalves and co-workers. The data in **Table III-1** however, is at odds with the experimental evidence. While MMB4 is expected to have worse binding than 2-PAM, LüH6 is expected to show more affinity towards tabun-inhibited AChE. This is not reflected in the binding energy as 2-PAM's is lower indicating better binding. The ligands used to improve the binding of the new oxime and hydroxylamine designs systematically decrease the binding energy

[183] a) R. Lo, B. Ganguly, *Mol. BioSyst.* **2014**, *10*, 2368–2383. b) R. Lo, N. B. Chandar, M. K. Kesharwani, A. Jain, B. Ganguly, *PLoS One* **2013**, *8*, e79591.

Table III-1. Force field calculated binding energies of reactivator-tabun inhibited AChE complexes. K_{OX} is the dissociation constant that describes the affinity of an oxime for AChE.

Reactivator	Binding energy (kcal.mol^{-1})	K_{OX} (M)
2-PAM	-27.6 [(b)]	706 x 10^{-6} [(c)]
TMB4	-21.9 [(a)]	1585 x 10^{-6} [(f)]
MMB4	-4.1 [(a)]	2418 x 10^{-6} [(d)]
LüH6	-18. [(a)]	97.3 x 10^{-6} [(c)]
Ortho-7	-38.1 [(b)]	Not found
DZP	-22.8 [(b)]	Not determined
3H-2PA	-10.5 [(b)]	Not determined
1	-30.9 [(b)]	25 x 10^{-6} [(e)]
2	-31.6 [(b)]	Not determined
2PHA	-14.4 [(a)]	Not determined
3	-19.9 [(a)]	Not determined

a: From ref [183a]. b: From ref [183b]. c: From ref [105]. d: From ref [126]. e: From ref [134]. f: From ref [128d]

The precise position of 2-PAM was further refined by molecular dynamic studies[184] as well as a little bit of docking.[185] What was found is that a good pre-reactive position for 2-PAM in the active site of AChE is with the pyridinium sandwiched between the aromatic rings of Try124 and Trp86. In this position, the oxime function forms a hydrogen bond with the phenol of Tyr337, locking it in place (see **Figure III-22**).

Steered molecular dynamic simulations of mutated AChE were employed to study more specifically the roles of Tyr337 and Phe338 in positioning the reactivator for the reactivation.[186], [187]The first such study, on tabun inhibited AChE, shows that when Tyr is mutated to an alanine the bipyridinium oxime K048 is allowed to rotate in the active site and the oxime functional group is moved away from the phosphorus atom.[186] The second study, still on tabun inhibited AChE but with the reactivator Ortho-7, indicated that the mutation of Tyr337 or both Tyr337 and Phe338 to alanine opens the channels and eases the approach of the reactivator. As residues of the acyl pocket and catalytic anionic site, they delimitate the size of the channel near the active site and therefore it is hard to believe that they have no influence on the binding of the reactivator. Their precise influence however is difficult to define. It is likely that while they do participate in the stabilization of the AChE-reactivator complex, their very bulky nature does impede the approach of the reactivator to some degree.

[184] A. da S. Gonçalves, T. C. C. França, J. D. Figueroa-Villar, P. G. Pascutti, *J. Braz. Chem. Soc.* **2011**, *22*, 155–165.
[185] K. S. Matos, D. T. Mancini, E. F. da Cunha, K. Kuča, T. C. França, T. C. Ramalho, *J. Braz. Chem. Soc.* **2011**, *22*, 1999–2004.
[186] N. B. Chandar, S. Ghosh, R. Lo, S. Banjo, B. Ganguly, *Chem. Biol. Interact.* **2015**, *242*, 299–306.
[187] R. Lo, N. B. Chandar, S. Ghosh, B. Ganguly, *Mol. BioSyst.* **2016**, *12*, 1224–1231.

Docking studies yields a docking energy that characterized the strength of the interaction of the enzyme with the docked substrate. For classical oximes in AChE, this docking energy agreed with the experimental values, with the lowest docking energies corresponding to reactivators with the lowest K_{OX}.[188] This good correlation between docking energies and experimental trends would tend to further validate the pre-reactive positions calculated from docking simulations.

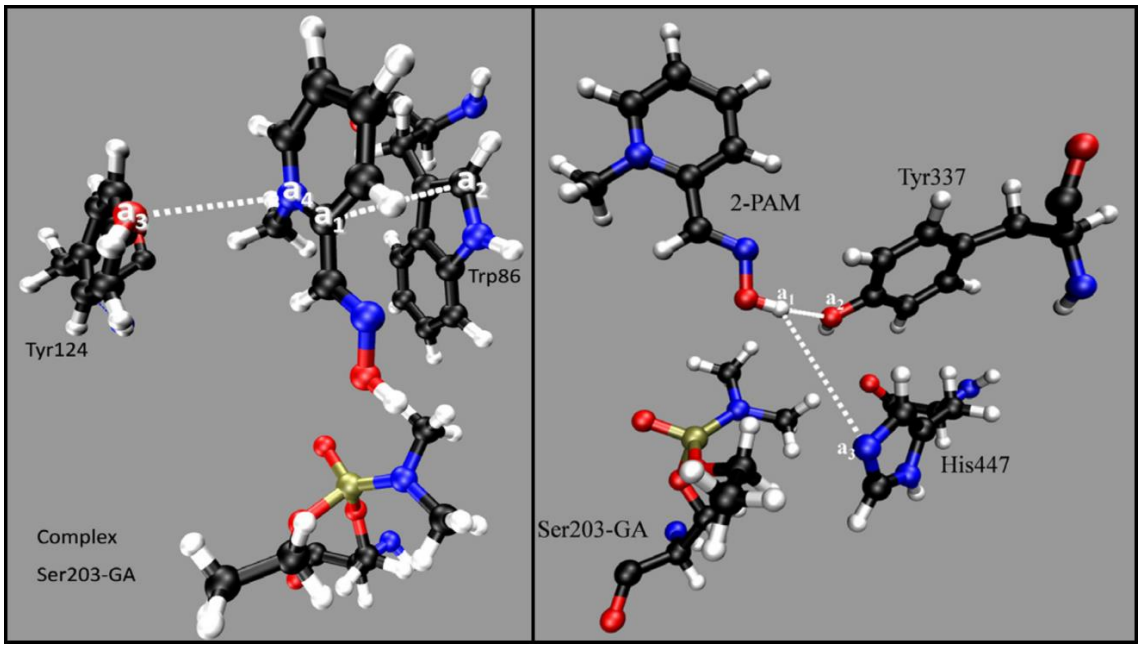


Figure III-22. Position of 2-PAM related to Tyr124, Trp86, the tabun-Ser203 adduct, Tyr337 and His447. From ref [184].

From the pre-reactive position, the reactivator is ready to react on the nerve agent inhibitor to separate it from Ser203's side chain. Known oxime reactivators are strongly nucleophilic in their oximate form and will perform a nucleophilic addition on the phosphorus which will drive the concerted or subsequent elimination of the enzyme from the phosphorus atom. This mechanism is very similar to the other reactive processes of AChE and the oxyanionic hole is expected to stabilize the bipyramidal trigonal intermediate of the addition elimination mechanism.[189] Several simulations of truncated model systems of reactivation, including only the reactivator and Ser203-nerve agent adduct, proceeded through an addition elimination mechanism as expected.[190], [183] A QM/MM study confirmed in 2012 the addition elimination mechanism for the reactivation of tabun inhibited AChE.[178] In this article, a full potential energy surface is presented. That potential energy surface has been reproduced in **Figure III-23**.

[188] K. S. Matos, E. F. F. da Cunha, A. da Silva Gonçalves, A. Wilter, K. Kuča, T. C. C. França, T. C. Ramalho, *J. Biomol. Struct. Dyn.* **2012**, *30*, 546–558.
[189] C. R. Hall, T. D. Inch, *Tetrahedron* **1980**, *36*, 2059–2095.
[190] J. Wang, J. Gu, J. Leszczynski, M. Feliks, W. A. Sokalski, *J. Phys. Chem. B* **2007**, *111*, 2404–2408.

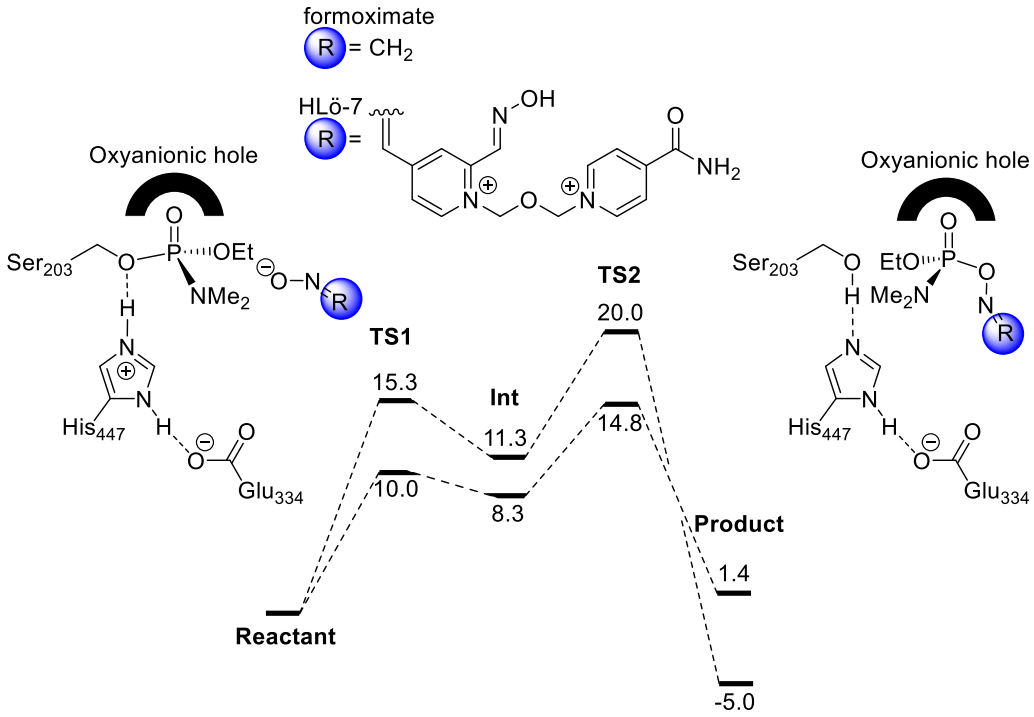


Figure III-23. Potential energy surfaces for the reactivation of tabun inhibited AChE with formoximate and HLö-7. Energies in kcal.mol^{-1}. From ref [178].

As can be seen on the potential energy surface, the rate determining step for both reactivators is the elimination. Even though the activation energy of the nucleophilic addition is higher, it has a smaller energetic span than the elimination which is thus rate determining, according to Kozuch and Shaik.[191] It contradicts the results obtained for the truncated model system simulations.[190], [185b] Two other sets of QM/MM simulations have been published that compare the reactivity of many classical reactivators.[185], [188] The differences in dephosphorylation energy barrier between classical oximes from those simulations were compared to the differences in dephosphorylation rate constant. They match very well which confirms the predictive power of the QM/MM methodology.

The truncated model system methodology was applied by Lo and co-workers to simulate the reactivation of tabun inhibited AChE by two classical oximes and two uncharged oximes.[183] The data from those two articles can be bound in **Table III-2** and the corresponding molecules in **Figure III-21**.

[191] S. Kozuch, S. Shaik, *Acc. Chem. Res.* **2011**, *44*, 101–110.

Table III-2. Energies of the reactants, products, intermediates, and transition states of the reactivation of tabun-inhibited AChE by oxime. Energies in kcal.mol^{-2}.

Reactivator	k_2	Reactant	TS1	Int	TS2	Product
2-PAM [b]	0.01 [c]	0.0	11.7	10.6	12.0	-3.8
TMB4 [a]	0.08 [e]	0.0	7.3	6.6	8.9	-9.5
MMB4 [a]	0.02 [d]	0.0	9.4	9.1	12.3	-4.4
LüH6 [a]	0.04 [c]	0.0	7.2	6.4	11.3	-6.8
Ortho-7 [b]	Not found	0.0	7.5	6.4	7.7	-7.7
DZP [b]	Not found	0.0	2.4	-4.4	-4.3	-22.4
3-H-2-PA [b]	Not found	0.0	2.6	-3.1	-2.8	-20.8
N-P-2-HA [a]	Not found	0.0	1.7	-7.4	-6.3	-25.1
3 [a]	Not found	0.0	6.1	-2.5	-2.3	-20.3

a: From ref [183a]. b: From ref [183b]. c: From ref [105]. d: From ref [126]. e: From ref [128d]

Beyond the confirmation of the addition elimination mechanism and of the addition as the rate determining step, this data shows a very clear divide between cationic oximes and oximes with a neutral prototypes. To the exception of compound **3** for all neutral oximes the energy barrier of the rate limiting addition step is below 3 kcal.mol^{-1}. The electron depleted pyridinium of cationic oximes leads to a reduced electron density on the nucleophile, diminishing its nucleophilicity. The increased nucleophilicity of oximes with a neutral π-system leads to this increased reactivity. This study also raises the question of hydroxylamine as an alternative to oxime based reactivators. It is based on two previous article which presented the comparison of several possible reactivators of AChE.[192] Several nucleophiles were evaluated (see **Figure III-24**). They evaluated the Gibbs free energy activation barrier of the rate determining addition step of reactivation of all six nucleophiles in a small model of the nerve agent adduct. The hydroxylamine anion, with a barrier of 2.3 kcal.mol^{-1} has a good reactivation potential compared to the formoximate.

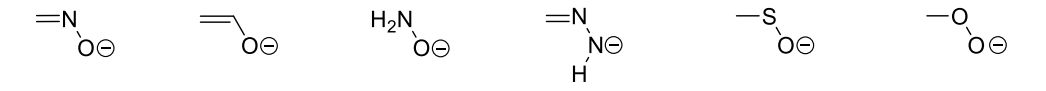


Figure III-24. Potential nucleophiles for AChE reactivation. Gibbs free energy activation barrier indicated in bold under the structure in kcal.mol^{-1}.

Amiss from most studies of reactivation is the question of the activation of the nucleophile. Oximates are Brønsted acids with pK_a ranging from 7 to 8 (**Table III-3**). As such they are expected to be protonated most of the time at physiological pH in the synaptic cleft,

[192] a) R. T. Delfino, J. D. Figueroa-Villar, *J. Phys. Chem. B* **2009**, *113*, 8402–8411. b) M. A. S. Khan, R. Lo, T. Bandyopadhyay, B. Ganguly, *J. Mol. Graphics Modell.* **2011**, *29*, 1039–1046.

around 7.2.[193] It is then logical to assume that oximes will enter the active site protonated and will be need to be deprotonated to their oximate form to react with the nerve agent adduct. Two articles published in 2014 and 2011 by Gonçalves and co-workers bring up the possible role of His447 in the deprotonation of the oxime reactivator 2-PAM.[184], [194] Their work also shows the spontaneous deprotonation of the methyl of 2-PAM by His447 after the elimination step of the reactivation. It raises serious questions about the quality of their simulations as this deprotonation of the methyl is not only very unlikely but also never evidenced in any other work of AChE reactivation, experimental or theoretical.

Table III-3. *p*K*a* of classical reactivators.

reactivator	2-PAM	LüH6	HLö7	HI-6	MMB4	TMB4	Ortho-7	3-H-2-PA
p*K*$_a$	7.7 [(a)]	7.9 [(a)]	7.0 [(a)]	7.3 [(a)]	Not found	8.2 [(a)]	Not found	8.2 [(b)]

a: From ref [105]. b: From ref [134].

Tabun is the greatest challenge of organophosphate reactivation. There is no known efficient reactivator for tabun-inhibited AChE. The main hypothesis for the limited reactivity of tabun-AChE adducts is that the electrons of the lone pair of the nitrogen are delocalized to the phosphorus atom rendering it less electrophilic. Ophélie Kwasnieski tested that hypothesis in her PhD thesis.[156] This work was realized under the assumption that a parallel could be established between the phosphorylation and the reactivation of tabun. The phosphorylation of Ser203 by several analogues of tabun was studied to compare the effect of substituents with different steric and electronic profiles (see **Figure III-25**). The assumption is that the phosphorylation and the reactivation are similar enough that the role of the N-dimethyl substituent will be the same in both processes.

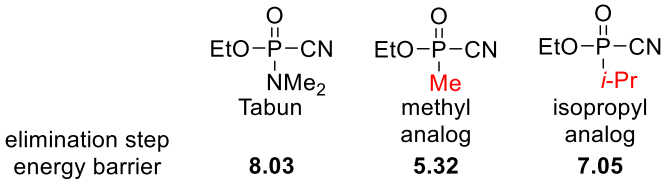


Figure III-25. Tabun and tabun analogues with matching energy barrier for the rate determining elimination step.

The substitution of the N-dimethyl substituent to a methyl greatly diminishes both steric strain and electronic delocalization. It also lowers the energy barrier confirming the influence of the N-dimethyl group on the reactivity of tabun. The substitution to iso-propyl, which provides a steric hindrance similar to that of the N-dimethyl group only very slightly diminishes the energy barrier. This indicates that most of the influence the N-dimethyl group plays on the activity of tabun is a steric influence. The electron delocalization plays only a

[193] R. Vroman, L. J. Klaassen, M. H. C. Howlett, V. Cenedese, J. Klooster, T. Sjoerdsma, M. Kamermans, *PLoS Biol.* **2014**, *12*, e1001864.

[194] A. da Silva Gonçalves, T. C. C. França, M. S. Caetano, T. C. Ramalho, *J. Biomol. Struct. Dyn.* **2014**, *32*, 301–307.

minimal role. Of course, this study on phosphorylation is only an indication and a similar study on dephosphorylation would be needed to confirm these data.

IV. Conclusions on acetylcholinesterase literature

Acetylcholinesterase is a crucial enzyme in the central and peripheral nervous systems of many species including our own. The first reason is its role in signal transmission at cholinergic synapses, but it also known to be the target of a deadly class of chemicals. Nerve agents covalently inhibit AChE which is often fatal. There is a treatment for nerve agent poisoning. It involves oxime compounds, strong nucleophiles that will react on the phosphorus adduct in AChE's active site to reactivate the enzyme. Their effectiveness however, is limited, especially against nerve agent tabun. The design of new potent and universal reactivators requires a thorough knowledge of AChE and its activity.

After decades of research, some facts have been established about the structure and activity of AChE. The main catalytic structure of AChE is a catalytic triad involving residues Ser203, His447 and Glu334. The regular mechanism of AChE catalysis in nucleophilic addition with the intermediate stabilized by an oxyanionic hole involving residues Gly120, Gly121, Gly122, and Ala204. This is a common feature of all reactive known processes taking place in AChE's active site from the catalysis to the reactivation, and including inhibition, aging and spontaneous reactivation by water. They share a common addition elimination mechanism with the intermediate stabilized by the oxyanionic hole. For both steps of regular catalysis and inhibition the main reacting nucleophile is activated by deprotonation by His447. It has been postulated that this deprotonation could also occurs in reactivation and aging, in some case with a substituted base, glutamate 202. Glu202 is a residue of the catalytic anionic binding site in very close proximity to the catalytic triad. It appears that this residue is generally protonated, contradicting the common expectation for glutamates, but some hypothesized reactive processes in AChE's active site involve an unprotonated Glu202. It precise role in the reactivation is still unknown.

Therefore, in our study of reactivation we chose to analyse the influence of the protonation state of this as well as other residues. On the realization that proton transfer steps and protonation states are a constant feature of the reactivity of AChE, we attempted, with the work presented here, to better understand the dynamic of proton transfer in the enzyme and the way it influences the reactivation of nerve agent inhibited AChE.

First a truncated model system of the active site was set up to single out the role of specific active site structures during the reactivation. The second part of the project, and fourth chapter of this thesis, saw the simulations of the in-situ deprotonation of classical reactivator 2-PAM and its subsequent reactivation of VX-inhibited AChE. In the fifth chapter, similar simulations of the non-cationic reactivators are presented. The final chapter is devoted to the question of protonation states a proton transfer between residues of AChE and with the solvent. In particular a, novel proton relay mechanism with the ability to maintain the protonation state of the active site of AChE is presented. The next chapter, however, is dedicated to the presentation of the simulation methods used throughout this project.

Chapter 2: Introduction to the Computational Methods

Chapter 2: Introduction to the Computational Methods

To introduce the field of computational chemistry seems, considering the tremendous role it has taken in modern chemistry and biochemistry, pointless. The precise methods used for the study of an enzyme and its reactivity however, do need to be introduced. The world of computational chemistry is divided between molecular mechanics and quantum mechanics but in order to achieve a quantum descriptions of reactivity in large systems, several methods bridge that gap. This section introduces EVB and QM/MM, two such methods, as well as the theoretical foundations of those methods.

I. Quantum mechanics

Quantum mechanics allows for the quantitative simulation of chemical phenomena. At its core, it allows for the calculation of the energy of a chemical system with a given geometry of known atomic nuclei and a given number of electrons. From this energy the lowest energy geometry, the electronic structure, the transition states, the excited states can all be calculated. The fundamental equation linking the energy of a system to the electronic wavefunction is the time-independent Schrödinger equation (equation **I-I**).[1]

$$\hat{H}\Psi = E\Psi \qquad \textbf{I-I}$$

In this equation, E is the energy of the system and Ψ is the electronic wavefunction. $\hat{\mathrm{H}}$ is the Hamiltonian operator defined as the sum of kinetic and potential operators (equation **I-II**), decomposed in electronic and nuclei terms (equation **I-III**).

$$\hat{H} = \hat{T} + \hat{V} \qquad \textbf{I-II}$$

$$\hat{H} = \widehat{T_e} + \widehat{T_n} + \widehat{V_{ne}} + \widehat{V_{ee}} + \widehat{V_{nn}} \qquad \textbf{I-III}$$

With $\widehat{T_e}$ as the kinetic energy operator for electrons, $\widehat{T_n}$ as the kinetic energy operator for nuclei, $\widehat{V_{ne}}$ as the operator for Coulombic attraction between nuclei and electrons, $\widehat{V_{nn}}$ as the operator for Coulombic repulsion between nuclei, and $\widehat{V_{ee}}$ as the operator for Coulombic repulsion between electrons. This decomposition is simplified because we are in a system where we can take the Born Oppenheimer approximation. It states that because of their very large mass compared to electron, nuclei move much slower. Thus, their kinetic energy operator $\widehat{T_n}$ can be canceled because they are considered immobile when considering the motion of electron and the Coulombic repulsion $\widehat{V_{nn}}$ between nuclei is considered constant. The Hamiltonian can be reformulated as the sum of an electronic Hamiltonian $\widehat{H_{el}}$ and the Coulombic repulsion $\widehat{V_{nn}}$ between nuclei (Equation **I-IV**). Those can then be used to separate

[1] E. Schrödinger, *Phys. Rev.* **1926**, *28*, 1049–1070.

the electronic energy term E_{el} from the Coulombic repulsion energy term V_{nn} in the time-independent Schrödinger equation (Equation **I-V**).

$$\widehat{H} = \widehat{T_e} + \widehat{V_{ne}} + \widehat{V_{ee}} + \widehat{V_{nn}} = \widehat{H_{el}} + \widehat{V_{nn}} \quad \textbf{I-IV}$$

$$\left(\widehat{H_{el}} + \widehat{V_{nn}}\right)|\Psi\rangle = (E_{el} + V_{nn})|\Psi\rangle \quad \textbf{I-V}$$

$$\widehat{H_{el}}|\Psi\rangle = E_{el}|\Psi\rangle \quad \textbf{I-VI}$$

Solving the electronic Hamiltonian solves the Schrödinger equation (equation **I-VI**). The electronic energy is the expectation value of the Hamiltonian (equation **I-VII**).

$$E_{el} = \frac{\langle\Psi|\widehat{H_{el}}|\Psi\rangle}{\langle\Psi|\Psi\rangle} \quad \textbf{I-VII}$$

The key question now is the formulation of the electronic wavefunction and electronic Hamiltonian. Three methods are commonly used in quantum chemistry to solve the electronic structure: the molecular orbital theory, the electronic band structure, and the valence bond theory. The first two are dominant in modern quantum chemistry and much of the formalism of quantum chemistry comes from the molecular orbital theory and the Hartree Fock method that arises from it.

All three methods rely on the variational principle. It states that the lowest energy E_0 attainable by the system is for the wavefunction of the ground state of the system Ψ_0. For any other wavefunction Ψ_i different to Ψ_0, E_i will be greater that E_0. The closer Ψ_i will be to Ψ_0, the lower E_i will be, with E_0 as a lower bound.

I.1. From molecular orbital method to Hartree-Fock theory

The molecular orbital theory is a method used to define the wavefunction. With the method, the wavefunction for a molecular orbital is equal the sum of weighted atomic orbitals (equation **I-VIII**).

$$\psi_{MO} = \sum_{i=1}^{N} c_i \psi_i \quad \textbf{I-VIII}$$

In this equation $\Psi_{MO}$is the wavefunction, ψ_i an atomic orbital, and c_i the weighting coefficient. The atomic orbital is the sole spatial component of a one electron function. To completely define an electron its spin needs to be taken into account. To describe the spin of an electron, two orthonormal functions are needed, $\alpha(\omega)$ for spin up and $\beta(\omega)$ for spin down. The spin orbital χ_i describe the one electron function with both spatial and spin components (equation **I-IX**)

$$\chi_i = \begin{cases} \psi_i \alpha(\omega) \\ or \\ \psi_i \beta(\omega) \end{cases} \quad \textbf{I-IX}$$

In Hartree Fock theory, a many-electron wavefunction is defined as a simple product of spin orbitals. As such, for electron i, with $h(i)$ the operator defining its kinetic and potential energy, χ_i is an eigenfunction of $h(i)$ with an eigenvalue of ε_i (equation **I-X**). The wavefunction in Hartree Fock theory is called a Hartree product (equation **I-XI**), and is the product of the spin orbital of each electron. The total electronic energy E_{el} is the sum of the eigenvalues (equation **I-XII**) and the electronic Hamiltonian operator $\widehat{H_{el}}$ is the sum of single electron operator $h(i)$s (equation **I-XIII**). Thoe can be used in a new formulation of the time-independent Schrödinger equation (equation **I-XIV**).

$$h(i)\chi_i = \varepsilon_i \chi_i \quad \textbf{I-X}$$

$$\Psi^{HP} = \prod_{i=1}^{N} \chi_i \quad \textbf{I-XI}$$

$$\widehat{H_{el}} = \sum_{i=1}^{N} h(i) \quad \textbf{I-XII}$$

$$E_{el} = \sum_{i=1}^{N} \varepsilon_i \quad \textbf{I-XIII}$$

$$\widehat{H_{el}} \Psi^{HP} = E_{el} \Psi^{HP} \quad \textbf{I-XIV}$$

To enforce the Pauli exclusion principle stating that two electrons cannot occupy the same spin orbital, wavefunction must abide by the antisymmetry principle. It states that a many-electron wavefunction must be antisymmetric with respect to the interchange of the coordinate x of any two electrons (equation **I-XV**).

$$\Psi^{HP}(x_1, \dots, x_i, \dots, x_j, \dots, x_N) = -\Psi^{HP}(x_1, \dots, x_j, \dots, x_i, \dots, x_N) \quad \textbf{I-XV}$$

The Hartree product does not satisfy this requirement so the wavefunction for a N-electron system must be rewritten as a Slater determinant (equation **I-XVI**).

$$\Psi(x_1, \dots, x_N) = (N!)^{-1/2} \begin{vmatrix} \chi_i(x_1) & \chi_j(x_1) & \dots & \chi_k(x_1) \\ \chi_i(x_2) & \chi_j(x_2) & \dots & \chi_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_i(x_N) & \chi_j(x_N) & \dots & \chi_k(x_N) \end{vmatrix} \quad \textbf{I-XVI}$$

The electronic Hamiltonian is still formulated as seen in equation **I-XVII**. It can be used to calculate the electornic energy E_{el} (equation **I-XVIII**). In this equation the integral and summation signs have been inversed to simplify the next step. The terms involving only a single electron i and those involving two electrons, i and j can be separated (equation **I-XIX**).

$$\widehat{H_{el}} = \widehat{T_e} + \widehat{V_{ne}} + \widehat{V_{ee}} = \frac{-\hbar^2}{2m_e}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\frac{Ze^2}{r_i} + \sum_{i=1}^{N}\sum_{j>i}\frac{e^2}{r_{ij}} \quad \textbf{I-XVII}$$

$$E_{el} = \int \Psi_i(i)\left\{\frac{-\hbar^2}{2m_e}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\frac{Ze^2}{r_i} + \sum_{i=1}^{N}\sum_{j\neq i}\frac{e^2}{r_{ij}}\right\}\Psi_i(i)d\tau_i \quad \textbf{I-XVIII}$$

$$\begin{aligned} E_{el} &= \sum_{i=1}^{N}\left\{\int \Psi_i(i)\left[\frac{-\hbar^2}{2m_e}\nabla_i^2 - \frac{Ze^2}{r_i}\right]\Psi_i(i)d\tau_i\right\} \\ &+ \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}\int\int \frac{\Psi_i^*(i)\Psi_j^*(j)e^2\Psi_i(i)\Psi_j(j) - \Psi_i^*(i)\Psi_i^*(j)e^2\Psi_j(i)\Psi_j(j)}{r_{ij}}d\tau_i d\tau_j \end{aligned} \quad \textbf{I-XIX}$$

$$E_{el} = 2\sum_{i=1}^{N/2}H_i + \sum_{i=1}^{N/2}\sum_{j=1}^{N/2}\left(2J_{ij} - K_{ij}\right) \quad \textbf{I-XX}$$

In equation **I-XX** the sums are over $N/2$ because the electrons are paired. H_i, K_{ij}, and J_{ij} are introduced. H_i is the core integral, defined as the sum of the potential energy of the electron in orbital Ψ_i and of the potential energy of the attraction between that electron and the nucleus (equation **I-XXI**). K_{ij} is also introduced (equation **I-XXII**) as the exchange operator, arising from the asymmetry of the wavefunction. J_{ij} finally, is introduced (equation **I-XXIII**) as the coulombian integral. It corresponds to the electron-electron repulsion energy from the electron of the *j*th orbital. The mean field approximation replaces the two-electron integral by a one electron integral corresponding to the mean interaction of the *i*th electron with all other electrons. The mono electronic Fock operator can then be defined as equation **I-XXIV**.

$$H_i = \int \Psi_i(i)\left[\frac{-\hbar^2}{2m_e}\nabla_i^2 - \frac{Ze^2}{r_i}\right]\Psi_i(i)d\tau_i \quad \textbf{I-XXI}$$

$$K_{ij} = \int\int \frac{\Psi_i^*(i)\Psi_j^*(j)e^2\Psi_j(i)\Psi_i(j)}{r_{ij}}d\tau_i d\tau_j \quad \textbf{I-XXII}$$

$$J_{ij} = \int\int \frac{\Psi_i^*(i)\Psi_i(i)e^2\Psi_j^*(j)}{r_{ij}} d\tau_i d\tau_j \approx \int \frac{e^2\Psi_j(i)}{r_{ij}} d\tau_j \quad \textbf{I-XXIII}$$

$$\widehat{F_\imath} = \widehat{H_\imath} + \sum_{i=1}^{n/2} 2\widehat{J_{\imath J}} - \widehat{K_{\imath J}} \quad \textbf{I-XXIV}$$

$$\widehat{F_\imath}\chi_i = \varepsilon_i\psi_i \quad \textbf{I-XXV}$$

Note that in equation **I-XXV** we use spatial orbitals as the case depicted is restricted Hartree Fock where both spin orbitals have the same spatial component. The spatial molecular orbitals are composed as a linear expansion of a finite set of K known basis functions ϕ_i (equation **I-XXVI**). We can then produce the Roothan equations (equation **I-XXVII**), composed of the C square matrix of expansion coefficients $C_{\mu i}$, the F Fock matrix, the S overlap matrix, and ε the diagonal matrix of orbital energies (equations **I-XXVIII**, **I-XXIX**, **I-XXX** and **I-XXXI** respectively).[2]

$$\psi_i = \sum_{\mu=1}^{K} C_{\mu i}\phi_\mu \quad \textbf{I-XXVI}$$

$$FC = SC\varepsilon \quad \textbf{I-XXVII}$$

$$C = \begin{vmatrix} C_{11} & C_{12} & \dots & C_{1K} \\ C_{21} & C_{22} & \dots & C_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ C_{K1} & C_{K2} & \dots & C_{KK} \end{vmatrix} \quad \textbf{I-XXVIII}$$

$$F = \begin{vmatrix} \int dr_1\phi_1^*(1)\hat{F}(1)\phi_1(1) & \dots & \int dr_1\phi_1^*(1)\hat{F}(1)\phi_K(K) \\ \vdots & \ddots & \vdots \\ \int dr_1\phi_K^*(K)\hat{F}(K)\phi_1(1) & \dots & \int dr_1\phi_K^*(K)\hat{F}(K)\phi_K(K) \end{vmatrix} \quad \textbf{I-XXIX}$$

$$S = \begin{vmatrix} \int dr_1\phi_1^*(1)\phi_1(1) & \dots & \int dr_1\phi_1^*(1)\phi_K(K) \\ \vdots & \ddots & \vdots \\ \int dr_1\phi_K^*(K)\phi_1(1) & \dots & \int dr_1\phi_K^*(K)\phi_K(K) \end{vmatrix} \quad \textbf{I-XXX}$$

[2] C. C. J. Roothaan, *Rev. Mod. Phys.* **1951**, *23*, 69–89.

$$\varepsilon = \begin{vmatrix} \chi_1 & 0 & \dots & 0 \\ 0 & \chi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \chi_K \end{vmatrix} \qquad \text{I-XXXI}$$

The Hartree Fock method, often called the Self Consistent Field (SCF) procedure consist of finding the eigenfunctions ψ_i of the Fock operator, which is initially defined from a guess set of eigenfunctions ψ_g. A summarized SCF cycle is represented in **Figure I-1**.

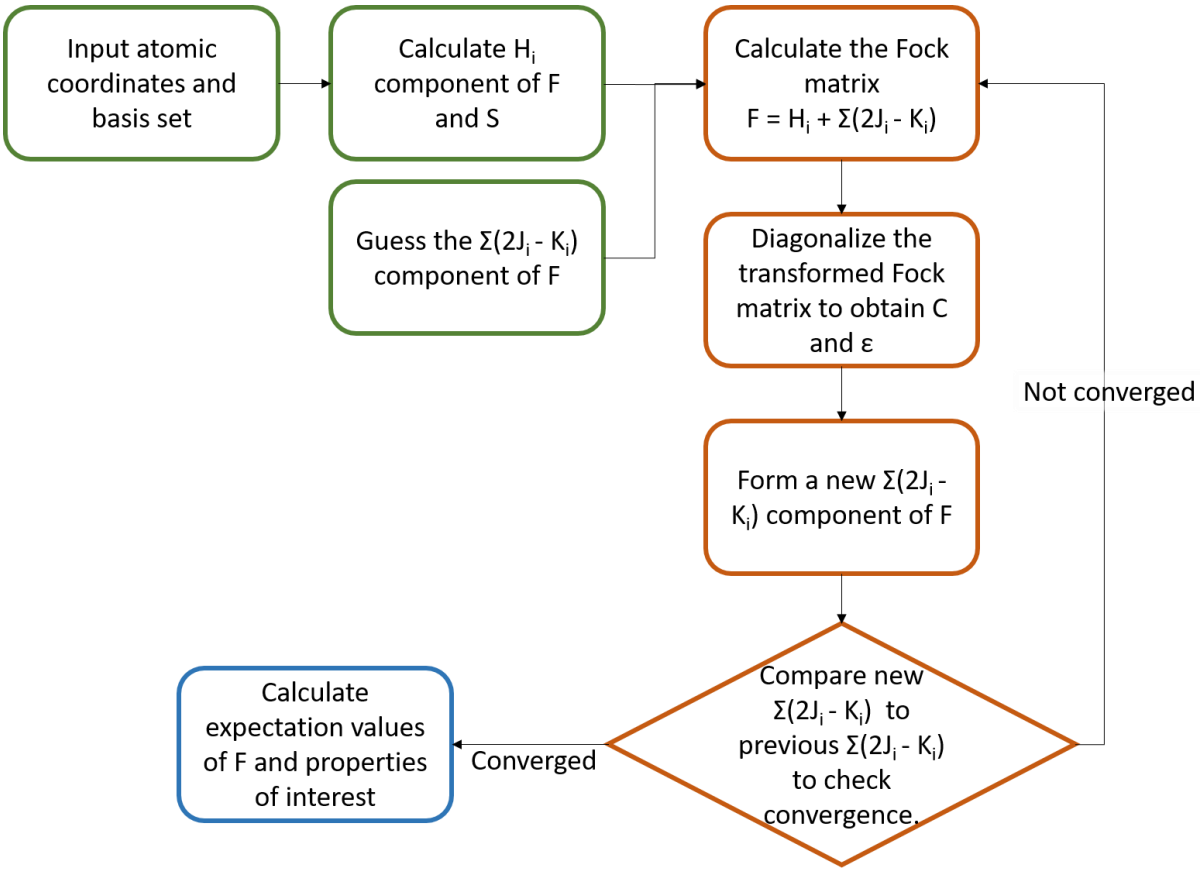


Figure I-1. Summary of the SCF procedure. Input steps are in green, iterative steps are in dark orange, and output is in blue.

The basis function of a basis set are Gaussian-type orbitals centred on the nucleus of an atom. Gaussian type orbitals instead of Salter type orbitals are most commonly used because the product of Gaussians on different two centres can be equated to a sum of Gaussian orbitals on a single centre. It allows for simpler and cheaper calculations despite the need for three Gaussian orbitals to describe a Slater orbital. Historically a single basis function would represent an atomic orbital but nowadays most basis sets use several functions to represent a single atomic orbital. For a chemical element, a basis set will contain a set of core orbitals, a set of valence orbitals and a set of diffuse orbitals. The valence orbitals can be better described when described with a greater number of functions. They can also be improved by adding function for the lowest unoccupied orbitals called polarization orbitals. Diffuse functions are extended Gaussian functions used to increase the size of the electron cloud.

This entire section has been written by taking inspiration from the explanations and representations used in the book Modern Quantum Chemistry by Szabo and Ostlund.[3]

As explained earlier, the Hartree Fock method uses an approximation of electron-electron interactions. The Hartree Fock energy will thus always be different from the absolute energy of the system. It lacks what has been called electronic correlation. Many methods have been based on the Hartree Fock wavefunction to establish a wavefunction that compensates for the errors of Hartree Fock. Those methods are called post-Hartree Fock and are not detailed here. The Density functional theory is another approach to the wavefunction that does include some level of electronic correlation.

I.2. Density functional theory

Thomas and Fermi devised what is considered the first description of electronic density as a cloud of electrons.[4] The electronic density is a function of a Cartesian vector and thus reduces a many body problem with 3N variables to a problem of 3 variables for the Cartesian coordinates of a vector. This electron density function can be obtained from the wavefunction as per equation **I-XXXII**.

$$\rho(\overrightarrow{r_1}) = N \int \dots \int |\Psi(\overrightarrow{x_1}, \overrightarrow{x_2}, \dots, \overrightarrow{x_N})|^2 \, d\overrightarrow{x_1} d\overrightarrow{x_2} \dots d\overrightarrow{x_N} \qquad \textbf{I-XXXII}$$

This function is an experimentally measurable observable. Thomas and Fermi proposed a formulation of the energy of a system only in terms electron density.

The first Hohenberg-Kohn theorem demonstrates that for any particle system of electrons interacting with an external potential $v_{ext}(\vec{r})$, *ie* the charge and position of nuclei, this potential is a unique functional of the electron density $\rho(\overrightarrow{r_1})$. What Hohenberg-Kohn demonstrates is that if two different $v_{ext}(\vec{r})$ potentials $v_1(\vec{r})$ and $v_2(\vec{r})$ can be used to obtain the same electronic density then they can only differ by a constant. Since the electron density flows from the wavefunction which can be obtained by appliying the external potential, which is a functional of the density, the electron density give access to all properties of the system.[5]

The second Hohenberg-Kohn theorem proves the validity of the variational principle when it comes to the relation between electron density and the energy. The minimal value of $E_{(v,N)}[\rho_0]$, for any N electrons and potential $v(\vec{r})$, is only obtained if $\rho_0(\vec{r})$ is the ground state density. The energy for any density that differs from the ground state is higher than the energy obtained for the ground state.

From those two theorems, the energy can be written not with the wavefunction and operators, but with functionals of the electronic density function (see equation **I-XXXIII**). From the second Hohenberg-Kohn the variational principle can be applied to this equation.

---

[3] A. Szabo, N. S. Ostlund, Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory, Courier Corporation, **1989**.

[4] a) L. H. Thomas, Mathematical *Proceedings of the Cambridge Philosophical Society* **1927**, *23*, 542–548. b) E. Fermi, *Rend. Accad. Naz. Lincei.* **1927**, *6*, 602–607.

[5] P. Hohenberg, W. Kohn, *Phys. Rev.* **1964**, *136*, B864–B871.

$$E[\rho(\vec{r})] = E_{Ne}[\rho(\vec{r})] + T[\rho(\vec{r})] + E_{ee}[\rho(\vec{r})] \quad \textbf{I-XXXIII}$$

$T[\rho(\vec{r})]$ is the kinetic energy, $E_{ee}[\rho(\vec{r})]$ is the electron-electron repulsion, and $E_{Ne}[\rho(\vec{r})]$ is the electron-nuclei attraction. The electron-electron repulsion and the kinetic energy can be united as the $F_{HK}[\rho(\vec{r})]$ functional (equation **I-XXXIV**). To know this functional is to have an exact solution of the Schrödinger equation. The electron-electron repulsion part can be broken down into its classical, $J[\rho(\vec{r})]$, and non-classical, $E_{ncl}[\rho(\vec{r})]$, components.

$$F_{HK}[\rho(\vec{r})] = T[\rho(\vec{r})] + E_{ee}[\rho(\vec{r})] \quad \textbf{I-XXXIV}$$

$$E_{ee}[\rho(\vec{r})] = J[\rho(\vec{r})] + E_{ncl}[\rho(\vec{r})] \quad \textbf{I-XXXV}$$

$$F_{HK}[\rho(\vec{r})] = T[\rho(\vec{r})] + J[\rho(\vec{r})] + E_{ncl}[\rho(\vec{r})] \quad \textbf{I-XXXVI}$$

The calculation of the exact kinetic energy of the systems proves to be very complex and thus Kohn and Sham elected to calculate the energy of a non-interacting systems with the density of the real interacting system (see equation **I-XXXVII**).[6] The non-classical component of the electron-electron interaction $E_{ncl}[\rho(\vec{r})]$ was rewritten into an exchange-correlation energy $E_{XC}[\rho(\vec{r})]$ (see equation **I-XXXVIII**).

$$T_S[\rho(\vec{r})] = \sum_i^N \left\langle \psi_i \middle| -\frac{1}{2}\nabla^2 \middle| \psi_i \right\rangle \quad \textbf{I-XXXVII}$$

$$F_{HK}[\rho(\vec{r})] = T_S[\rho(\vec{r})] + J[\rho(\vec{r})] + E_{XC}[\rho(\vec{r})] \quad \textbf{I-XXXVIII}$$

$$E_{XC}[\rho(\vec{r})] = (T[\rho(\vec{r})] - T_S[\rho(\vec{r})]) + (J[\rho(\vec{r})] - E_{ee}[\rho(\vec{r})]) \quad \textbf{I-XXXIX}$$

$$E[\rho(\vec{r})] = \sum_i^N \left\langle \psi_i \middle| -\frac{1}{2}\nabla^2 \middle| \psi_i \right\rangle - \left\langle \psi_i \middle| \sum_k^{nuclei} \frac{Z_k}{|r_i - r_k|} \middle| \psi_i \right\rangle + \sum_i^N \left\langle \psi_i \middle| \frac{1}{2} \int \frac{\rho(r')}{|r_i - r'|} dr' \middle| \psi_i \right\rangle + E_{XC}[\rho(\vec{r})] \quad \textbf{I-XL}$$

Equation **I-XL** describes the energy functional $E[\rho(\vec{r})]$ in four terms. The first is the kinetic energy of the system $T[\rho(\vec{r})]$. The second is the electron-nuclei attraction $E_{Ne}[\rho(\vec{r})]$. The third term is the classical part of the electron-electron interaction $J[\rho(\vec{r})]$. The final term is the exchange and correlation energy that describes both the non-classical terms of electron-electron interaction and the correction to kinetic energy. This equation is the key of energy calculation in DFT.

[6] W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133–A1138.

This formulation of the energy allows to establish the Kohn Sham equations as equation **I-XLI** with h_i^{KS} defined in equation **I-XLII** and $V_{XC}\ [\rho(\overrightarrow{r_i})]$ defined in equation **I-XLIII**.

$$h_i^{KS}\psi_i = \varepsilon_i\psi_i \qquad \textbf{I-XLI}$$

$$h_i^{KS} = \left[-\frac{1}{2}\nabla^2 - \sum_A^M \frac{Z_A}{r_{iA}} + \int \frac{\rho(\vec{r_i})}{r_{ij}} d\vec{r_i} + V_{XC}[\rho(\vec{r_i})]\right] \qquad \textbf{I-XLII}$$

$$V_{XC}[\rho(\vec{r_i})] = \frac{\delta E_{XC}[\rho(\vec{r_i})]}{\delta\rho(\vec{r_i})} \qquad \textbf{I-XLIII}$$

The key of DFT is thus the formulation of the exchange and correlation functional $V_{XC}[\rho(\vec{r})]$. The classical approximation for the exchange and correlation functional is the local-density approximation (LDA). This approximation is based around the notion of a uniform electron cloud. As such, the exchange and correlation functional can be written as per equation **I-XLIV**. The exchange and correlation energy $\epsilon_{XC}\big(\rho(\vec{r})\big)$ is the exchange and correlation per particle (see equation **I-XLV**). It can be split into its two components with a simple formulation for exchange component (equation **I-XLVI**). This formulation comes from the uniform electron gas representation of Thomas Fermi. There no simple definition for the correlation component of the exchange and correlation energy as they vary from one functional to another.

$$E_{XC}^{LDA}[\rho(\vec{r})] = \int \rho(\vec{r})\epsilon_{XC}\big(\rho(\vec{r})\big)d\vec{r} \qquad \textbf{I-XLIV}$$

$$\epsilon_{XC}\big(\rho(\vec{r})\big) = \epsilon_X\big(\rho(\vec{r})\big) + \epsilon_C\big(\rho(\vec{r})\big) \qquad \textbf{I-XLV}$$

$$\epsilon_X\big(\rho(\vec{r})\big) = -\frac{3}{4}\left(\frac{3\rho(\vec{r})}{\pi}\right)^{1/3} \qquad \textbf{I-XLVI}$$

However, LDA's accuracy is too low for chemical applications. A more recent approximation is the generalized gradient approximation (GGA). It supplements the electron density with the gradient of the electron density to account for a non-homogeneous electron cloud. It is commonly formulated as per equation **I-XLVII**.

$$E_{XC}^{GGA}\big[\rho_\alpha(\vec{r})\rho_\beta(\vec{r})\big] = \int \rho(\vec{r})\epsilon_{XC}\big[\rho_\alpha(\vec{r}), \rho_\beta(\vec{r}), \nabla\rho_\alpha(\vec{r}), \nabla\rho_\beta(\vec{r})\big]d\vec{r} \qquad \textbf{I-XLVII}$$

There again, there is no simple formulation of the exchange and correlation functionals for GGA functionals. They have reached high enough accuracy to be usefully applied to chemical systems. The GGA approximations were improved to so-called meta-GGA functionals

that include the density of the kinetic energy, hybrid functionals that considers the occupancy of orbitals, and non-local functionals that take into account virtual orbitals (see **Figure I-2**).

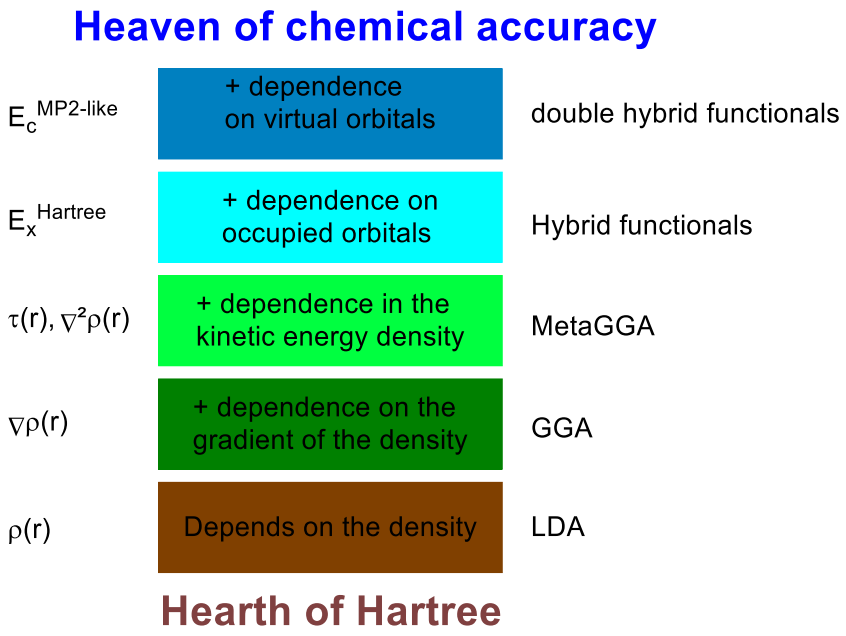


Figure I-2. Jacob's ladder of DFT exchange and correlation functionals established by John Perdew.

Regardless of the chosen functional, equation I-XLI is formulated similarly to Hartree Fock and thus a modified SCF procedure can be used which is detailed in **Figure I-3**.

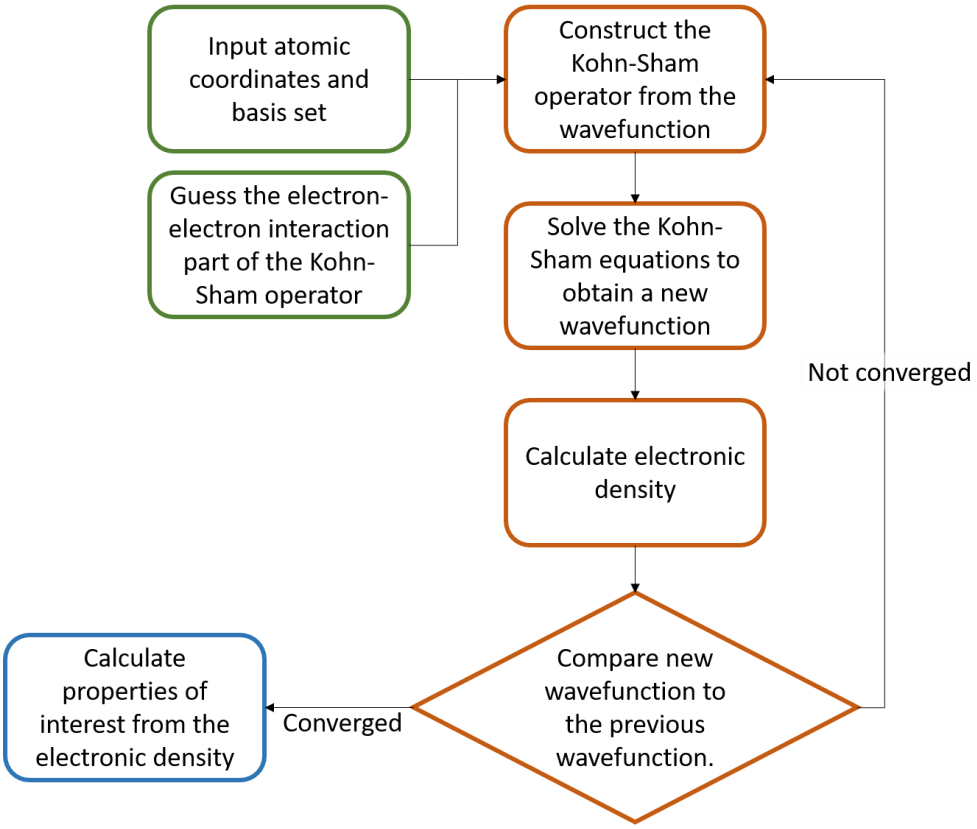


Figure I-3. Summary of the SCF procedure for DFT. Input steps are in green, iterative steps are in dark orange, and output is in blue.

The functional commonly used in the work presented in this project is B3LYP. It is a hybrid functional that combines LDA HF and GGA (see equation **I-XLVIII**). The exchange functional of B3LYP is based on a LDA exchange functional corrected by a GGA exchange functional from Becke[7] and Hartree Fock exchange functionals. The Hartree Fock and GGA corrections are weighted. The exchange functional is 72% GGA corrected, 20% Hartree Fock corrected, and 8% pure native LDA ($a_0 = 0.20$, $a_x = 0.72$). The correlation functional is 81% GGA functional from Lee Yang and Parr[8] and 19% pure native LDA ($\mathrm{a_c} = 0.81$).

$$E_{XC}^{B3LYP} = E_X^{LDA} + a_0(E_X^{HF} - E_X^{LDA}) + a_x(E_X^{GGA} - E_X^{LDA}) + E_C^{LDA} + a_c(E_C^{GGA} - E_C^{LDA}) \quad \textbf{I-XLVIII}$$

A third formulation exists to represent the wavefunction as a combination of available molecular resonance states for the system. It is called the valence bond theory.

I.3. Valence bond theory

In this paragraph, the basic principles of valence bond (VB) theory will be described. The theoretical backwall of valence bond theory will be presented now, as it is necessary in order to understand the empirical valence bond method that was used at the end of my thesis and which is discussed at the end of the next section.

This sub-section is not intended to be a *vade-mecum* per of VB and readers looking for something more detailed should read the review of Wu, Su, Shaik and Hiberty.[9] In valence bond theory, the wavefunction is built of valence bond structure (see equation **I-XLIX**). The system is described by a combination of structures that differ in the arrangement of the valence electrons. The electrons are localized on an atom/fragment.

$$\Psi^{VB} = \sum_{i=1}^{N} c_i \phi_i \quad \textbf{I-XLIX}$$

In this equation c_i is a weighting coefficient and ϕ_i a VB structure, or VB vector and N is the number of VB vectors. The VB vector ϕ_i is required to be a spin eigenfunction antisymmetric with respect to the permutation of electron indices. A VB vector is composed of a linear combination of atomic orbitals of the atom/fragment it will be localized on. The electrons of that atom/fragment occupy it. An analogy can be made between the VB vectors and the resonances structures of a chemical system. A VB wavefunction can be considered as composed of the full collection of resonance states (even ones that would seem chemically incorrect) of a system. The weighting coefficients c_i balance the importance a specific VB vector takes in the wavefunction of the system. A simple example is the dihydrogen molecule. The hydrogen molecule has four attainable VB structures. The most evident one is when both hydrogen atoms are covalently bonded. There are also the two limit ionic forms of this bond,

[7] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 1372–1377.
[8] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.
[9] W. Wu, P. Su, S. Shaik, P. C. Hiberty, *Chem. Rev.* **2011**, *111*, 7557–7593.

with both electrons either on hydrogen a or b. Finally, there is the possibility for both hydrogen to keep a single proton as a free radical. The composition of the VB structures can be written as the sum of the fragment orbital arrangements that correspond to this structure. For ϕ_1 for instance, we can have one spin up proton localized on hydrogen a and one spin down localized on hydrogen b as a first arrangement. The second arrangement is similar with inversed spin. For ϕ_1 both electrons are localized on hydrogen b, one with a spin up, the second with a spin down.

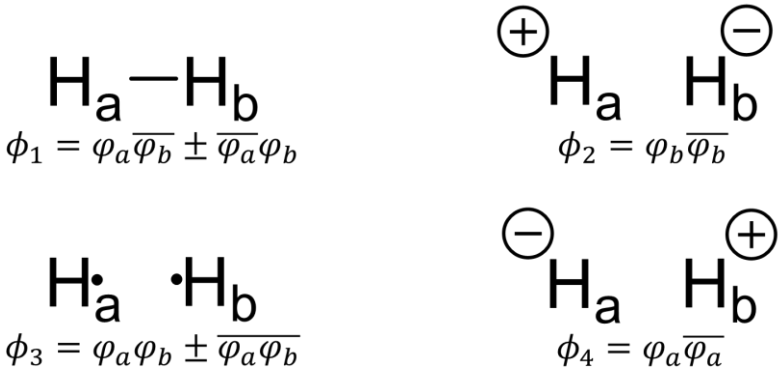


Figure I-4. VB vectors of dihydrogen with their composition in occupied fragment orbitals.

Now that we have a wavefunction, we need to use it to solve Schrödinger's equation. The variational principle and Schrödinger equation give equation **I-L**. This equation is often called the secular equation in the literature and written in the form taken by equation **I-L**.

$$\sum_{j=1}^{N}(H_{ij} - ES_{ij})c_j = 0 \qquad \textbf{I-L}$$

$$HC = ESC \qquad \textbf{I-LI}$$

In equation **I-LI**, H and S are the Hamiltonian and overlap matrix as defined by equations **I-LII** and **I-LIII**. This equation ties the Hamiltonian and the wave function to the energy.

$$H = \begin{vmatrix} \langle\phi_1|H|\phi_1\rangle & \langle\phi_1|H|\phi_2\rangle & \dots & \langle\phi_1|H|\phi_N\rangle \\ \langle\phi_2|H|\phi_1\rangle & \langle\phi_2|H|\phi_2\rangle & \dots & \langle\phi_2|H|\phi_N\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle\phi_N|H|\phi_1\rangle & \langle\phi_N|H|\phi_2\rangle & \dots & \langle\phi_N|H|\phi_N\rangle \end{vmatrix} \qquad \textbf{I-LII}$$

$$S = \begin{vmatrix} \langle\phi_1|\phi_1\rangle & \langle\phi_1|\phi_2\rangle & \dots & \langle\phi_1|\phi_N\rangle \\ \langle\phi_2|\phi_1\rangle & \langle\phi_2|\phi_2\rangle & \dots & \langle\phi_2|\phi_N\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle\phi_N|\phi_1\rangle & \langle\phi_N|\phi_2\rangle & \dots & \langle\phi_N|\phi_N\rangle \end{vmatrix} \qquad \textbf{I-LIII}$$

The diagonal elements H_{ii} of the Hamiltonian matrix represent the energy of the corresponding VB structure. One of the advantages of VB is that the weight of the different

VB structures can easily be calculated. Several definitions exist, the most commonly used is the Coulson-Chirgwin formula (equation **I-LIV**).

$$w_i = {c_i}^2 \sum_{j=1}^{N} c_i S_{ij} c_j \qquad \textbf{I-LIV}$$

The valence bond theory and methodology has been used with success in the study of [3 + 2] cycloaddition to show their radical character.[10] It has also been used in the study of hypervalent compounds such as XeF_2 to demonstrate that hypervalent bonding is charge-shift.[11], [12] It also constitutes part of the chemical basis of the EVB method that was used in this project and which is detailed at the end of the next section.

II. Molecular mechanics

Molecular mechanics are based around classical Newtonian mechanics. Atoms are represented as the etymology of the word indicates, indivisible particles. Molecules are treated as fixed entities, defined by their atoms and the spring-like bonds that connect them. All the terms that enter in the composition of the total energy of the system (see equation **II-I**) are calculated from set empirical parameters.

$$U_{total} = U_{bond} + U_{angle} + U_{torsion} + U_{vdW} + U_{ele}\ (+ U_{cross} + U_{impr})_{optional} \qquad \textbf{II-I}$$

This model does not account for electrons outside of their contribution to the overall charge of individual atoms. In classical force fields this charge is a set fixed parameter. Polarizable force field complete this set point charge with one or several environment-dependant multipoles. The total energy of the system is calculated from the Cartesian coordinates of the atoms, the topology of the molecules and atom type of every atom.

II.1. Classical force fields

The formulation of classical energy terms of classical force fields is detailed in the following section. Most common force fields (CHARMM, AMBER, OPLS) share the same formulation for those terms, otherwise it will be specified. The bond energy, also commonly called stretching as it is mechanically a spring, is formulated as per equation **II-II**.

$$U_{bond} = \sum_{bonds} k_b (r - r_0)^2 \qquad \textbf{II-II}$$

With r the length of the bond in Å, r_0 the length at equilibrium, and k_b the force constant (often called spring constant) in kcal.mol^{-1}. Å^{-2}. All three main force fields formulate

[10] B. Braida, C. Walter, B. Engels, P. C. Hiberty, *J. Am. Chem. Soc.* **2010**, *132*, 7631–7637.
[11] B. Braïda, P. C. Hiberty, *Nat. Chem.* **2013**, *5*, 417–422.
[12] S. Shaik, D. Danovich, W. Wu, P. C. Hiberty, *Nat. Chem.* **2009**, *1*, 443–449.

the stretching energy similarly. The angle energy is formulated very similarly as per equation **II-III**.

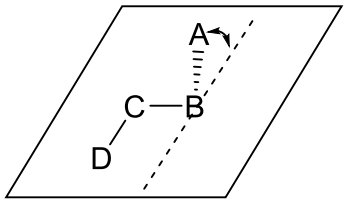
$$U_{angle} = \sum_{angles} k_\theta(\theta - \theta_0)^2 \qquad \textbf{II-III}$$

With θ the angle in °, θ_0 the angle at equilibrium, and k_θ the force constant in kcal.mol^{-1}.degrees^{-2}. There again all three force fields use the same formulation. CHARMM however, is the only force field to include an Urey-Bradley component to account for angle bending (equation **II-IV**).

$$U_{Urey-Bradley} = \sum_{distance} k_{ub}(S - S_0)^2 \qquad \textbf{II-IV}$$

In this equation, S is the distance between the two atoms in Å, $S_0$is the equilibrium distance and k_{ub} is the associated force constant in kcal.mol^{-1}. Å^{-2}.

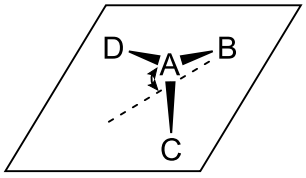
The torsion angles, also called dihedral angles, represent the angle a bond between atoms A and B makes with the plane of atoms B, C and D (see **Scheme II-1**). They are formulated according to equation **II-V**. in this equation, V_n is the torsion force constant, ϕ is the torsion angle, δ is the phase shift and n is the multiplicity of the function.



Scheme II-1. Torsion angle between atom A and the plane of atoms B, C and D.

$$U_{torsion} = \sum_{torsione} \sum_{n} V_n[1 + cos(n\phi - \delta)] \qquad \textbf{II-V}$$

Another divergence between CHARMM and other force fields is in the formulation used for improper angles. They represent out of plane deformations of planar atoms like sp^2 carbons (see **Scheme II-2**). CHARMM formulates it as the product of a force constant k_ω with the divergence of the angle ω from planarity (see equation **II-VI**). In Amber and OPLS, a Fourier term is used (see equation **II-VII**).



Scheme II-2. Improper angle between atom A and the plane of atoms B C and D.

$$U_{imp} = \sum_{impropers} \frac{k_\omega}{2}(\omega - \omega_0)^2 \qquad \textbf{II-VI}$$

$$U_{imp} = \sum_{impropers} \frac{k_\omega}{2}[1 + cos(2\omega)] \qquad \textbf{II-VII}$$

Those are the bonded interactions but the non-bonded interactions also enter the composition of force fields. Classical force fields have two terms for non-bonded interactions, a Lennard-Jones term, and an electrostatic term. The Lennard-Jones term, also called van der Waals term is formulated as per equation **II-VIII**.

$$U_{LJ} = \sum_{i,j} \varepsilon_{i,j} \left(\left(\frac{R^0_{i,j}}{R_{i,j}} \right)^{12} - 2 \left(\frac{R^0_{i,j}}{R_{i,j}} \right)^{6} \right) \qquad \textbf{II-VIII}$$

In this equation, $\varepsilon_{i,j}$ in the depth of the van Der Waals at distance $R_{i,j}$, with $R^0_{i,j}$ is the distance at which the Lennard-Jones potential is 0. Formally this term is the sum over all atom pairs of the difference between the attractive London dispersion energy (to the power of 6) and the Pauli repulsion (to the power of 12). The second term for non-bonded interaction is the electrostatic term (equation **II-IX**).

$$U_{coulomb} = \sum_{i,j} \frac{q_i q_j}{R_{i,j}} \qquad \textbf{II-IX}$$

Here q_i is the partial charge of atom i. This expression of the charge is the main difference between classical force fields and polarizable force fields.

II.2. Polarizable force fields

Polarizable force fields take into account the variation in electrostatics due to the interaction with the nearby charge of another atom. In quantum mechanics, the electron cloud of a fragment would respond to the presence of the electron cloud of another nearby fragment by being delocalized and polarized. This effect is accounted for in classical force fields for an average environment. However, in situations where the transferability to an average environment is not valid, classical force fields will fail to accurately represent this polarization. Polarizable force fields try to fix this oversight by including induced dipoles to regular force fields. Several models exist to account for that electronic property in a classical model. They will be briefly summarized over the following paragraphs.[13]

The first scheme involving polarization is the induced dipoles model. In this model, the system is a collection of charges the location of which varies from one force field to another

[13] Y. Shi, P. Ren, M. Schnieders, J.-P. Piquemal, in *Reviews in Computational Chemistry Volume 28* (Eds.: A.L. Parrill, K.B. Lipkowitz), John Wiley & Sons, Inc, **2015**, pp. 51–86.

but are usually located at atomic centres. The charge distribution at a given site i is the sum of a permanent charge and an induced charge.

$$M_i = M_i^0 + M_i^{ind} \quad \textbf{II-X}$$

In this equation M is the charge distribution, which can be a point charge or a point multipole. This point multiple can be expanded to as dipole, a quadrupole or even higher order moments. The induced charge distribution is defined by equation **II-XI**.

$$M_i^{ind} = \alpha_i \sum_{j \neq i} T_{ij}\left(M_j^0 + M_j^{ind}\right) \quad \textbf{II-XI}$$

In this equation, T_{ij} is the interaction operator and a function of the distance between sites i and j. It's exact formulation depends on the charge distribution scheme of a specific force field. α_i is the polarizability of site i. From the definition of the induced charge, two different possible expressions exist for the electrostatic energy.

$$U_{ele} = \frac{1}{2}\sum_1 \sum_{j \neq i} M_i^t T_{ij} M_j + \frac{1}{2}\sum_i \left(M_i^{ind}\right)^t \alpha_i^{-1} M_i^{ind} \quad \textbf{II-XII}$$

$$U_{ele} = \frac{1}{2}\sum_1 \sum_{j \neq i} (M_i^0)^t T_{ij} M_j^0 + \frac{1}{2}\sum_1 \sum_{j \neq i} \left(M_i^{ind}\right)^t T_{ij} M_j^0 \quad \textbf{II-XIII}$$

In both equations, the first term is the permanent electrostatic energy and the second term is the polarization energy.

The second scheme to include polarization is the Drude oscillator model. It attaches to every non-hydrogen atom a mobile point charge by a harmonic spring. This point charge will move in response of the electrostatic environment. The electrostatic energy is then expressed as the sum of pairwise interactions between atomic charges and the mobile point charge called Drude particle (equation **II-XIV**).

$$U_{ele} = \sum_{A<B}^{N} \frac{q_c(A) q_c(B)}{r_c(A) - r_c(B)} + \sum_{A<B}^{N,N_D} \frac{q_D(A) q_c(B)}{r_D(A) - r_c(B)} + \sum_{A<B}^{N_D} \frac{q_D(A) q_D(B)}{r_D(A) - r_D(B)} + \frac{1}{2}\sum_{A}^{N,N_D} k_D\left(r_D(A) - r_c(A)\right)^2 \quad \textbf{II-XIV}$$

In this equation N and N_D are the number of atoms and the number of Drude particles respectively. $q_c(A)$ is the charge of atom A and $q_D(A)$ is the charge of the Drude particle attached to atom A. $r_c(A)$ and $r_D(A)$ are the respective position of both. k_D is the force constant of the harmonic spring between the rude oscillator and the parent atom. The polarizability can then be defined as a function of the partial charge of the Drude particle and the force constant k_D (equation **II-XV**).

$$\alpha(A) = \frac{q_D^2(A)}{k_D} \qquad \textbf{II-XV}$$

The third scheme is the fluctuating charge model which is based on the charged equilibration method (CHEQ). The charge density is redistributed which equilibrates the chemical potential. In simpler terms, the electronegativity between two atoms forming a bond is compensated for by a change in charges until the electronegativity is the same on both atoms. Resulting energy for a system of M molecules each containing N_i atoms is found in equation **II-XVI**.

$$U_{ele}(R,Q) = \sum_{i=1}^{M}\sum_{\alpha=1}^{N}\chi_{i\alpha}Q_{i\alpha} + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{\alpha=1}^{N_i}\sum_{\beta=1}^{N_j}J_{i\alpha j\beta}Q_{i\alpha}Q_{j\beta} + \frac{1}{2}\sum_{i=1}^{MN}\sum_{j=1}^{MN}\frac{Q_iQ_j}{4\pi\epsilon_0 r_{ij}} + \sum_{i=1}^{M}\lambda_i\left(\sum_{\alpha=1}^{N_i}Q_{i\alpha} - Q_i^{Total}\right) \qquad \textbf{II-XVI}$$

In this equation $Q_{i\alpha}$ is the partial charge on atomic site α of molecule i. χ is the electronegativity of an atom, J is the hardness. λ is a Lagrangian multiplier which enforces the molecular charge of every molecule.

II.3. Molecular dynamics

In molecular dynamics, the objective is to study the evolution of a system over time. Molecular dynamics, due to the size of the studied systems and the time scale of the studied processes, generally precludes the use of quantum mechanics.[14] To simulate the evolution of a N-body system overtime, the algorithm must iteratively calculated forces on every particle, find the new position of said particle subject to those forces and then calculate new forces that will move the particle to a new position. The force applied to a particle A is the gradient of the potential energy of A with all other particles. It is also the product of the second order derivative of the position vector r_A of the particle with respect with time with the mass of the particle (see equation **II-XVII**).

$$F_A = -grad(U_A) = m_A\frac{d^2r_A}{dt^2} \qquad \textbf{II-XVII}$$

Of course, the second order derivative of the position vector with respect with time is the first order derivative of the velocity vector and is the acceleration vector of the particle.

$$\frac{d^2r_A}{dt^2} = \frac{dv_A}{dt} = a_A \qquad \textbf{II-XVIII}$$

[14] M. Karplus, J. A. McCammon, *Nat Struct Mol Biol* **2002**, *9*, 646–652.

A numerical approximation of the differential equation **II-XVII** can be obtained using a numerical integration method. The most common method used in molecular dynamics is the velocity Verlet algorithm, which is very closely related to the also common leapfrog algorithm.[15] The initial conditions for this algorithm is known particle positions $\vec{r}(t)$ and known particle velocities $\vec{v}(t)$. The first step of the velocity Verlet algorithm is the calculation of the velocities at time $t + ½\Delta t$ (equation **II-XIX**). Those velocities are then used to calculate the positions at time $t + \Delta t$ (equation **II-XX**). A new acceleration vector $\vec{a}(t + \Delta t)$ can be calculated from those new positions. Finally the velocities at time $t + \Delta t$ can be calculated (equation **II-XXI**).

$$\vec{v}\left(t + \frac{\Delta t}{2}\right) = \vec{v}(t) + \frac{\Delta t}{2}\vec{a}(t) \qquad \textbf{II-XIX}$$

$$\vec{x}(t + \Delta t) = \vec{x}(t) + \vec{v}\left(t + \frac{\Delta t}{2}\right)\Delta t \qquad \textbf{II-XX}$$

$$\vec{v}(t + \Delta t) = \vec{v}\left(t + \frac{\Delta t}{2}\right) + \frac{\Delta t}{2}\vec{a}(t + \Delta t) \qquad \textbf{II-XXI}$$

The leapfrog algorithm is very closely related to the velocity Verlet algorithm. The main difference is that velocities are calculated in full time step increments and, like in the velocity Verlet algorithm, as half of a time step difference of the positions.

Molecular dynamic simulations allow to simulate conformational changes that usually take place in a time scale of the order of the micro second. For large systems like biopolymers (DNA, proteins, bilipid layers) the simulation of those conformational changes is necessary to explore the conformational space of the system. Molecular dynamics allow to sample in this conformational space. Among the many additional methods derived from molecular dynamics, metadynamics are used to ensure that the full conformational space is sampled.[16]

Steered molecular dynamics is another type of simulations derived from simple molecular dynamics.[17] In steered molecular dynamics, an external force is applied to the system to provoke an event. A rare or otherwise very costly conformational change can be provoked and explored using this technique.

A third type of simulation based on the foundations of molecular dynamics are constant pH molecular dynamics.

II.4. Constant pH Molecular Dynamics

In classical molecular dynamics, because covalent bonds are set and cannot be either broken or formed, the protonation state of each residue is manually set at the beginning of the simulation. This approach ignores the effect of conformation on pK_a and protonation but

[15] W. C. Swope, H. C. Andersen, P. H. Berens, K. R. Wilson, *J. Chem. Phys.* **1982**, *76*, 637–649.
[16] A. Laio, M. Parrinello, *Proc. Nat. Ac. Sc. USA* **2002**, *99*, 12562–12566.
[17] B. Isralewitz, M. Gao, K. Schulten, *Curr. Opin. Struct. Biol.* **2001**, *11*, 224–230.

also misrepresents any residue with a pK_a close to the pH of the solvent. Because of the limits of force fields, titratable residue cannot be dynamically protonated and deprotonated to alleviate that issue. To solve this problem, calculate residue pK_a and explore the conformational effects of protonation changes, constant pH molecular dynamic (CpHMD) methods were proposed. Several method exist but here the method of Mongan, Case and McCammon[18], modified for explicit solvents by Swails, York and Roitberg[19], will be discussed. At regular time interval a protonation change for a titratable residue is proposed. If this protonation change produces a lower ΔG (equation **II-XXII**) than the reversal of the protonation to the previous state, it is accepted as a new protonation state. Else the simulation reverts to the previous state.

$$\Delta G = k_B T\left(pH - pK_{a,ref}\right) ln10 + \Delta G_{elec} - \Delta G_{elec,ref} \qquad \textbf{II-XXII}$$

In this equation pH is the set solvent pH, $pK_{a,ref}$ is the pH of a reference, ΔG_{elec} is the electrostatic component of the free energy for the titratable group in the protein and $\Delta G_{elec,ref}$ is the same electrostatic component of the free energy for the reference compound.

The CpHMD simulations directly yield the fraction of the time a specific residue was protonated at a specific pH. It also yields the conformational impact of a protonation state. Replicating CpHMD simulations a various pH values permits to establish the probability of a specific residue to be protonated at those different pH values. A titration curve can be produced from this data by fitting the results to the Hill equation using the least square method (see equation **II-XXIII**).[20]

$$pK_a = pH - nlog\left(\frac{[A^-]}{[AH]}\right) \qquad \textbf{II-XXIII}$$

In this equation, the pK_a of a specific residue, is the difference of the pH with the logarithm of the ratio of base over acid. The coefficient n is the Hill coefficient tied to the cooperativity of the protonation change. A coefficient of 1 indicates a protonation change independent of other residues. Deviations from 1 indicate that the protonation state change is cooperative.

II.5. Empirical Valence Bond

The empirical valence bond (EVB) method is a force field in which a specific reaction has been parametrized. It behaves like a reactive force field for this reaction and like a classical force field toward any other reactive process. It is empirical in the sense that the studied reaction in the studied environment must be compared to a reference reaction in a reference environment. The thermodynamic constants of this reference reaction in the reference

[18] J. Mongan, D. A. Case, J. A. McCammon, *J. Comput. Chem.* **2004**, *25*, 2038–2048.
[19] J. M. Swails, D. M. York, A. E. Roitberg, *J. Chem. Theory Comput.* **2014**, *10*, 1341–1352.
[20] a) J. L. Markley, *Biochemistry* **1973**, *12*, 2245–2250. b) J. L. Markley, *Acc. Chem. Res.* **1975**, *8*, 70–80.

environment need to be known either from theory or preferably from experiment. The two parameters, H_{ij}, and α_i^{gas} are fitted to reproduce thermodynamic constants of the reference. Those parameters are considered to be valid for the same reaction for any environment. For instance, once obtained, those two parameters can be used to simulate the reaction in any imaginable mutant of an enzyme.[21]

EVB shares the framework of VB as detailed in a previous section. Instead of VB vectors corresponding to a different electronic arrangement, VB states in EVB correspond to key energy minima for the studied reaction. Bluntly the EVB states correspond to the potential energy functions of the reactant state (state i), product state (state j), and stable intermediate states of the studied reaction. The system is described by a linear combination of the EVB states with the coupling between them described with reaction specific coupling terms. The resonance between the structures is thus taken into account and like a quantum mechanical method, the energy is obtained from the eigenvalues of the secular equation like it would in regular VB. The Hamiltonian matrix of equation **II-XXIV** can be constructed with the diagonal and off diagonal terms H_{ii} and H_{ij} as defined by equations **II-XXV** and **II-XXVI**.

$$H = \begin{vmatrix} H_{ii} & H_{ij} \\ H_{ji} & H_{jj} \end{vmatrix} \qquad \textbf{II-XXIV}$$

$$H_{ii} = U_{qq}^{i}(\boldsymbol{R}, \boldsymbol{Q}) + \alpha_{gas}^{i} = \varepsilon_i \qquad \textbf{II-XXV}$$

$$H_{ij} = Ae^{(-\mu(r-r_0))} \qquad \textbf{II-XXVI}$$

In equation **II-XXV**, which is defining diagonal elements of the matrix H, α_{gas}^{i} is the gas phase shift, a parameter fitted to a reference (the procedure for EVB calculations is detailed later). $U_{qq}^{i}(\boldsymbol{R}, \boldsymbol{Q})$ is the potential energy of atoms participating in the reaction with $\boldsymbol{R}$ representing the coordinates of those atoms and $\boldsymbol{Q}$ the charge. Additional potential energy terms can be added to account for solvent and protein environment. Finally ε_i is the energy of state H_{ii}. Equation **II-XXVI** shows how off-diagonal elements can be represented using exponential functions. A and μ are both parametrized constants fitted to experiments and the difference $r - r_0$ is the deviation from the equilibrium distance for relevant reacting atoms. Those diagonal elements H_{ij} are also parameters fixed to a reference as explained on the next page.

With the definition of all the matrix element of the H matrix the secular equation (equation **II-XXVII**) can be solved to obtain the lowest eigenvalue which is the ground state energy of the system. In EVB the system can often be defined with two states and thus the solution of the secular equation is easily obtained through equations **II-XXVIII** and **II-XXIX**.

$$HC = E_g C \qquad \textbf{II-XXVII}$$

[21] A. Shurki, E. Derat, A. Barrozo, S. C. Lynn Kamerlin, *Chem. Soc. Rev.* **2015**, *44*, 1037–1052.

$$H - E_g = 0 = \begin{vmatrix} \varepsilon_1 - E_g & H_{12} \\ H_{21} & \varepsilon_2 - E_g \end{vmatrix} \quad \textbf{II-XXVIII}$$

$$E_g = \frac{1}{2}\left[(\varepsilon_1 + \varepsilon_2) - \sqrt{(\varepsilon_1 - \varepsilon_2)^2 + 4H_{12}^2}\right] \quad \textbf{II-XXIX}$$

The previous sections detailed the process of calculating the groundstate EVB energy for a specific configuration of the system along the reaction path. To explore that reaction path from the reactant to the product while sampling the high energy states with molecular dynamics, several strategies can be used. In the EVB calculations presented in this manuscript, the coupling of EVB energy calculations with free energy perturbation (FEP) was used to explore those high energy states.

Free energy perturbation is based on the Zwanzig equation (equation **II-XXX**) which states that the free energy difference between two states can be calculated if the two states are both close and sufficient sampling has been performed on both states.

$$\Delta G_{A\to B} = -\beta^{-1} ln\langle exp\{-\beta(U_B - U_A)\}\rangle_A \quad \textbf{II-XXX}$$

There should be a lot of overlap between the two states. The mapping parameter λ is used and made to vary in small incremental steps (for a total of n step). The force field terms of a state are obtained by mixing the force field terms of the reactant and product states as per equation **II-XXXI.**

$$U_\lambda = (1-\lambda)U_1 + \lambda U_2 \quad (0 \le \lambda \le 1) \quad \textbf{II-XXXI}$$

Using the Zwanzig equation on steps k and $k+1$, we can obtain the energy difference between those steps. The sum of this energy differences over the n incremental steps yields the free energy difference between reactant state 1 and product state 2 (equation **II-XXXII**). For a given step k and a given set of coordinates r, the energy difference, called energy gap (x), between states 1 and 2 can be calculated (variable of equation **II-XXXIII**). This energy gap can be used as a reaction coordinate to obtain a free energy profile for the reaction (**Figure II-1**).

$$\Delta G_{\varepsilon_1\to\varepsilon_2} = \sum_{k=0}^{N-1} -\beta^{-1} ln\langle exp\{-\beta(U_{\lambda+1} - U_\lambda)\}\rangle_A \quad \textbf{II-XXXII}$$

$$\Delta G(x') = \Delta G_{\varepsilon_1\to\varepsilon_2} - \beta^{-1} ln\,\langle \delta(x - x') \times exp\left\{-\beta\left(E_g(x) - U_\lambda(x)\right)\right\}\rangle_{\varepsilon_m} \quad \textbf{II-XXXIII}$$

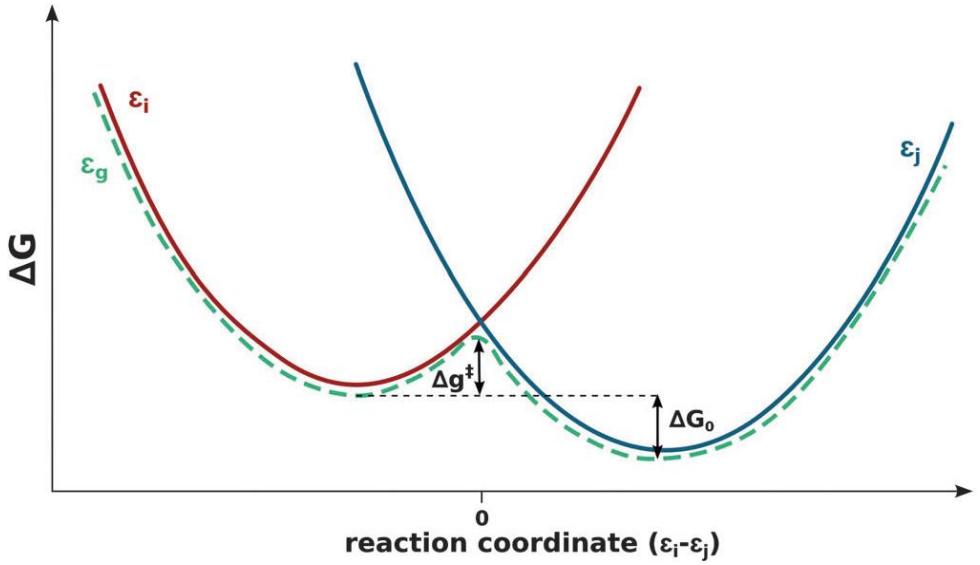


Figure II-1. EVB free energy profile as a combination of two states i and j.

As can be seen in **Figure II-1**, the reaction coordinate used to follow the reaction is not geometric or structural. The reaction coordinate is the energy difference between the two states of the reaction, the energy difference between the product and reactant state for a give conformation. It allows to represent all the conformational changes that accompany the reaction in a single reaction coordinate.

The EVB method has two main advantages. First it fully accounts for the dynamics of the environment of the reacting atoms. The second advantage is the assumption made in EVB that both the H_{ij} and the α_{gas}^{i} constants for a given reaction remain valid in any environment, be it gas phase, water solution or an enzyme. It allows to easily study the effect of mutation of protonation changes in an enzyme on the studied reaction.

Both the H_{ij} and the α_{gas}^{i} constants can be obtained by fitting an EVB calculation on an experimental value. When experimental ΔG^{*} and ΔG_0 values are available for a specific reaction in a known environment (gas phase, water solution, etc.), a corresponding EVB calculation is performed. Then, instead of using H_{ij} and α_{gas}^{i} to calculate theoretical ΔG^{*} and ΔG_0 values, the experimental ΔG^{*} and ΔG_0 are used to obtain H_{ij} and α_{gas}^{i}. Those can in turned be used to obtain ΔG^{*} and ΔG_0 from EVB simulation of that reaction catalysed by an enzyme.

The main caveat of EVB is that since it is a parametrized force field, it is not only limited by its quality, but it is also limited to the reaction and the mechanism for this reaction that is parametrized. A reaction (or reaction step for reactions that have one or more stable intermediates) is parametrized as a change from the set of parameters describing the reactant to the set of parameters describing the product. The reaction can only go through the described step. It means either an assumption must be made on the mechanism of the reaction or all possible mechanisms should be simulated and compared. EVB has been used at the very end of my project and the associated results are described in chapter 5 of the manuscript. In contrast to EVB, QM/MM relies on an electronic and unparametrized description of chemical mechanism.

III. QM/MM

QM/MM is a hybrid method combining quantum mechanics to molecular mechanics. It divides a large system into at least two regions.[22] The quantum mechanical (or QM) region which is electronically important and that needs a quantum treatment. This region contains the part of the system directly involved in the studied reaction or properties. The molecular mechanical region (or MM) that surrounds the QM region is a region that does not need to be treated at the electron level and where molecular mechanics are sufficient. This region encompasses everything not in the QM region. There are many applications to QM/MM, but it is a particularly pertinent method for the study of enzymatic reactions. Contrary to a truncated model[23] system (often called cluster model) the apoenzyme is simulated in full which natively includes long range interactions and allows for more significant conformational change to occur during the reaction. It also has the advantage to allow for "unexpected" chemical mechanism to occur. It is especially true when compared to EVB where a specific reaction must be parametrized.

There are two main schemes for the calculation of the QM/MM energy. The energy calculation can be additive, or subtractive. This issue is detailed in the next sub-section.

III.1. Additive or Subtractive

The subtractive scheme for QM/MM allows to perform QM/MM with very little modifications to standard quantum mechanics and molecular mechanics methods.[24] All regions as well as the inner regions encompassed in those regions are calculated independently as a whole at the chosen level of theory for that region. It means that, for a two-region system with a small inner QM region treated in DFT and the rest of the system in the MM region treated with a force field, two independent calculations are performed. The whole system is calculated with the force field, including both QM and MM regions, and the QM region is calculated with DFT.

[22] S. F. Sousa, A. J. M. Ribeiro, R. P. P. Neves, N. F. Brás, N. M. F. S. A. Cerqueira, P. A. Fernandes, M. J. Ramos, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2016**, DOI 10.1002/wcms.1281.

[23] B. Manta, K. E. Cassimjee, F. Himo, *ACS Omega* **2017**, *2*, 890–898.

[24] L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, et al., *Chem. Rev.* **2015**, *115*, 5678–5796.

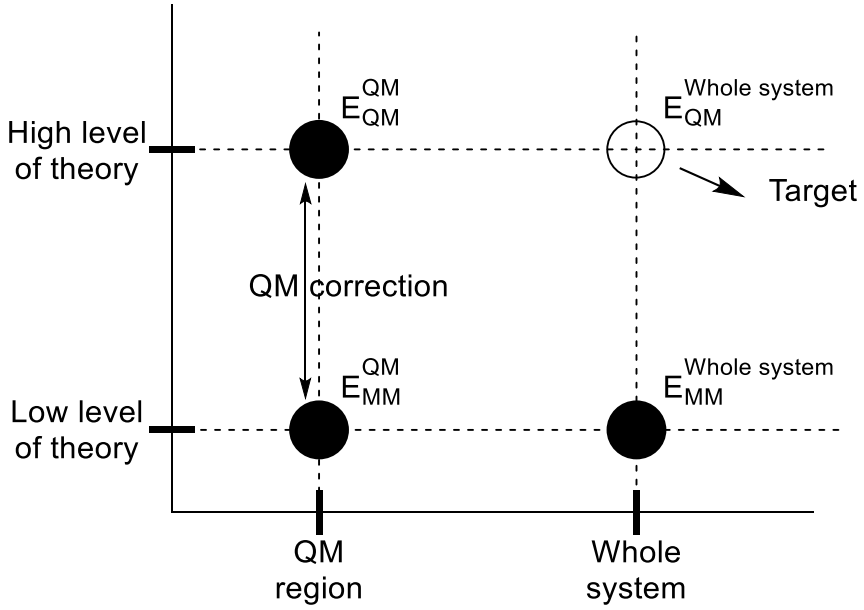


Figure III-1. Schematic partition of the whole system by the IMOMM method.

The energy MM of the QM region is then subtracted from the QM energy of the QM region to produce a QM correction (see **Figure III-1**). The main approximation of the subtractive scheme is that this QM correction of the QM region is transferable to the whole system. Thus, the total QM energy is approximated as the MM energy for the whole system plus that QM correction accounting for the QM treatment of the quantum region (see equation **III-I**). The subtractive scheme treats the coupling between the QM and MM regions implicitly. This is called the Integrated Molecular Orbital Molecular Mechanics method (IMOMM).[25]

$$E_{QM} \cong E_{QM/MM} = E_{MM}^{whole\ system} + \left(E_{QM}^{QM\ region} - E_{MM}^{QM\ region}\right) \qquad \textbf{III-I}$$

The subtractive scheme was initially designed for IMOMO[26], integrated MO + MO method. The principle was to split moderately large systems (in the range of 10^2 atoms) into a QM and a QM' region. The QM region would be treated with a high level method and the QM' region with a low level method.

This scheme allows to easily use more than two regions. If two QM regions and a MM region are used for instance, with a high cost inner region called the QM region and a low-cost outer region for mid-range interactions called the QM' region, equation **III-I** becomes equation **III-II**. This is part of the n-layered integrated MO and MM' method also called ONIOM.[27]

[25] F. Maseras, K. Morokuma, *J. Comput. Chem.* **1995**, *16*, 1170–1179.

[26] S. Humbel, S. Sieber, K. Morokuma, *J. Chem. Phys.* **1996**, *105*, 1959–1967.

[27] L. W. Chung, W. M. C. Sameera, R. Ramozzi, A. J. Page, M. Hatanaka, G. P. Petrova, T. V. Harris, X. Li, Z. Ke, F. Liu, et al., *Chem. Rev.* **2015**, *115*, 5678–5796.

$$E_{QM/MM} = E_{MM}^{whole\ system} + \left(E_{QM'}^{whole\ QM+QM'} - E_{MM}^{QM'region+QM\ region}\right) + \left(E_{QM}^{QM\ region} - E_{QM'}^{QM\ region}\right) \quad \textbf{III-II}$$

The main draw-back of subtractive QM/MM is that there is a disconnect between the inner region as calculated with MM method and the inner region as calculated with QM. While the error of the MM in the inner QM region is cancelled by the subtraction, the outer region is calculated with a flawed inner region leading to great difficulties with the geometry optimizations.

The additive QM/MM scheme calculates the energy of a specific region solely with the level of theory chosen for that region. In the case of a two-region system with a small inner QM region treated in DFT and the rest of the system in the MM region treated with a force field; the QM region is only calculated with DFT and the force field calculation does not include any atom from the QM region. A coupling term is added for the coupling between the two regions (see equation **III-III**).

$$E_{QM/MM} = E_{QM}^{QM\ region} + E_{MM}^{MM\ region} + E_{QM/MM\ coupling} \quad \textbf{III-III}$$

The disadvantage of the additive scheme is the added complexity of having to calculate a coupling term. This coupling term formally includes all interaction between the two regions, electrostatic, van der Waals.[28] Practically, depending on the boundary scheme and handling of interactions between the QM and MM regions, parts or even the entirety of the coupling term is integrated in the Hamiltonian of the QM region.[29]

In both schemes, when a covalent bond crosses the boundary between the QM and MM region must be handled by a specific boundary scheme.

III.2. Boundary scheme

An ideal system for QM/MM is one were there would be no covalent bond that crosses the boundary between QM and MM regions. Such systems, however, are extremely rare and often several covalent bonds cross the boundary between the MM and QM regions. There are two common strategies for dealing with the bonds that cross this boundary.

The first strategy is the link atom approach. When an atom in the QM region is bounded to an atom in the MM region this MM atom is replaced in the QM region by a "link atom".[30] This link atom (noted L in **Figure III-2**) is often a hydrogen atom but halogen atoms as well as methyl groups can also be used. The inconvenient of this approach is that the link atom will be geometrically very close to the atom it replaces in the MM region. To avoid artefacts due to the charge repulsion of the link atom with the original atom, the most

[28] H. M. Senn, W. Thiel, *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.

[29] P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, et al., *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 1–28.

[30] M. J. Field, P. A. Bash, M. Karplus, *J. Comput. Chem.* **1990**, *11*, 700–733.

commonly used methods is the charge shift scheme.[31] The charge on the original M_1 atom is deleted and shared evenly between the M_2 atoms connected to it (see **Figure III-2**). Point charges are added around those M_2 atoms to compensate for the lost dipole moment those M_2 atoms formed with the M_1 atom.

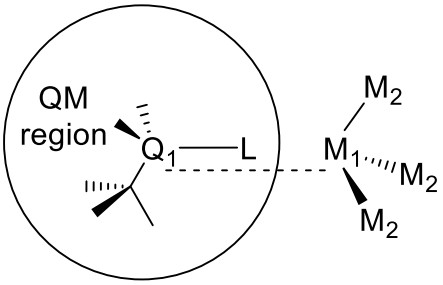


Figure III-2. Representation of the link atom approach

Several other means of dealing with the overpolarization due to the link atom at the boundary have been devised. The one electron integrals describing the link atom L can be deleted.[32] The point charges of all nearby atoms in the MM region can also be deleted.[33] Those two schemes deal with the overpolarization but the influence of the deleted charge or energy term on the system is lost. Finally, the point charge of atom M_1 can be replaced by a gaussian distribution of the charge that will retain long range electrostatic interactions but attenuate short range repulsion.[34]

The second common strategy for dealing with boundary crossing covalent bonds is the frozen orbital strategy. There are two implementations of this strategy. The frozen orbital can either be located on the Q_1 atom, it is the localized SCF scheme[35] (LSCF), or on the M_1 atom, it's the generalized hybrid orbital scheme[36] (GHO). In the LSCF scheme, a set of orbitals are placed on the Q_1 atom. One orbital, shown in gray on **Figure III-3a**, is directed towards M_1 and kept frozen. This frozen orbital is based on a calculation on a similar fragment. In the GHO scheme, a set or orbitals are placed of the M_1 atom. One orbital, shown in white on **Figure III-3b** and pointed towards Q_1, is active in the wavefunction of the QM region. The other three orbitals are frozen, populated with a fixed electron density calculated on a transferable model compound.

[31] A. H. de Vries, P. Sherwood, S. J. Collins, A. M. Rigby, M. Rigutto, G. J. Kramer, *J. Phys. Chem. B* **1999**, *103*, 6133–6141.
[32] D. Bakowies, W. Thiel, *J. Phys. Chem.* **1996**, *100*, 10580–10594.
[33] M. J. Harrison, N. A. Burton, I. H. Hillier, *J. Am. Chem. Soc.* **1997**, *119*, 12285–12291.
[34] D. Das, K. P. Eurenius, E. M. Billings, P. Sherwood, D. C. Chatfield, M. Hodošček, B. R. Brooks, *J. Chem. Phys.* **2002**, *117*, 10534–10547.
[35] V. Théry, D. Rinaldi, J.-L. Rivail, B. Maigret, G. G. Ferenczy, *J. Comput. Chem.* **1994**, *15*, 269–282.
[36] J. Gao, P. Amara, C. Alhambra, M. J. Field, *J. Phys. Chem. A* **1998**, *102*, 4714–4721.

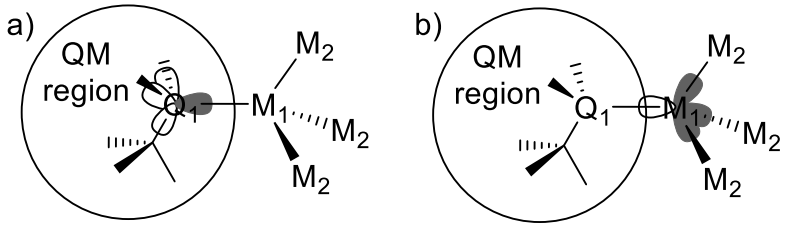


Figure III-3. Representation of the frozen orbital approach. a) LSCF scheme. b) GHO scheme. Grey orbitals are frozen, white are active

Those two strategies are not the only existing strategies to deal with a boundary crossing covalent bond but they are the most used.[37] The QM and MM region do not only interact through the boundary between QM and MM regions. The next sub-section treats the general electrostatic interactions between both regions.

III.3. Interactions between QM and MM

There are three classical possibilities for the treatment of the interactions between the QM and MM regions: mechanical embedding, electrostatic embedding, and polarized embedding.

Mechanical embedding is the simplest approach to the interactions between the two regions. The QM calculation is done as in gas phase. Structural constraints are imposed by the connections between the QM and MM regions. Besides those constraint, the MM region has no influence on the QM region. The influence of the QM region in the MM region is modelled as point charges and Lennard-Jones potentials as calculated with the normal force field. With mechanical embedding the effect of long and mid-range interactions of the apoenzyme on the reactive centre are lost. The point charges that represent the QM region in the MM calculation do not account for the changes occurring in the QM region over the course of the reaction. They only relay geometrical information on the QM region. The point charges can be updated along the potential energy surface of a reaction but it can produce discontinuities in the potential energy along the surface.

With the electrostatic embedding approach, the QM region is polarized by the MM region. The point charged of the atoms of the MM are included in the Hamiltonian as one electron integrals. This inclusion allows changes in the MM region to affect the reactivity in the QM region. The point charges of the MM region come from the parametrized partial charges from the force field. While they would fluctuate if they were treated in QM, the MM region is not involved in the reactivity and so the partial charges in this region are not expected to undergo massive changes. Like in the mechanical embedding scheme however, the QM region polarizes the MM region on the basis of parametrized force field charges. Those parametrized charges do not reflect the changes that occur in the QM region as the reaction follows its course.

[37] a) I. Antes, W. Thiel, *J. Phys. Chem. A* **1999**, *103*, 9290–9295. b) Y. Zhang, T.-S. Lee, W. Yang, *J. Chem. Phys.* **1998**, *110*, 46–54. c) A. Laio, J. Van de Vondele, U. Rothlisberger, *J. Chem. Phys.* **2002**, *116*, 6941–6947.

The last approach is the polarized embedding approach. In this approach both MM and QM regions polarize each other until they reach a consistent charge distribution. The polarized embedding requires the use of a polarizable force field able to respond to the external field of the QM region. This imposed choice of force field comes at a greater computational cost especially because to reach self-consistency of the charge distribution an iterative method must generally be used.

III.4. How to set up QM/MM simulations

This subsection will address general concerns when setting up QM/MM simulations. It focuses on the simulation of enzymatic systems as it was the subject of the work presented in this manuscript. The reader should go the review by Quesne, Borowski and de Visser for further details.[38]

To start QM/MM simulations on an enzymatic system the coordinates of the heavy-atoms of that enzymatic system need to be known. Such structures are available from X-ray crystallographic studies from many enzymes (see **Figure III-4A and B**). Very often, for a specific enzyme more than one structures are available, but not all those structures are of equal quality for QM/MM simulations. The first key element is the completeness of the X-ray structure. Often some residues, a cofactor, some side chains, or other elements of the enzymatic complex are missing because they have not been resolved or are absent from the crystallized enzyme. While some of this missing information can be manually filled in it requires a very good knowledge of the enzyme complex. The more complete the starting structure the better because modifying the structure by hand takes time and is error prone. The manual addition of elements like ligands or inhibitors that were absent from the crystallized structure cannot reproduce the conformational changes those ligands may have triggered in the structure of the enzymatic complex. The resolution of a structure is a second key element of choice. A higher resolution means a more reliable structure. Low resolutions can be associated with errors in atom position and atom type. Several other factors can enter in the choice of a starting structure like its prevalent in the pertinent literature. If a specific X-ray structure is consistently used in studies of a specific enzymatic system, using the same structure could facilitate the comparison with previous studies that used it.

[38] M. G. Quesne, T. Borowski, S. P. de Visser, *Chem. Eur. J.* **2016**, *22*, 2562–2581.

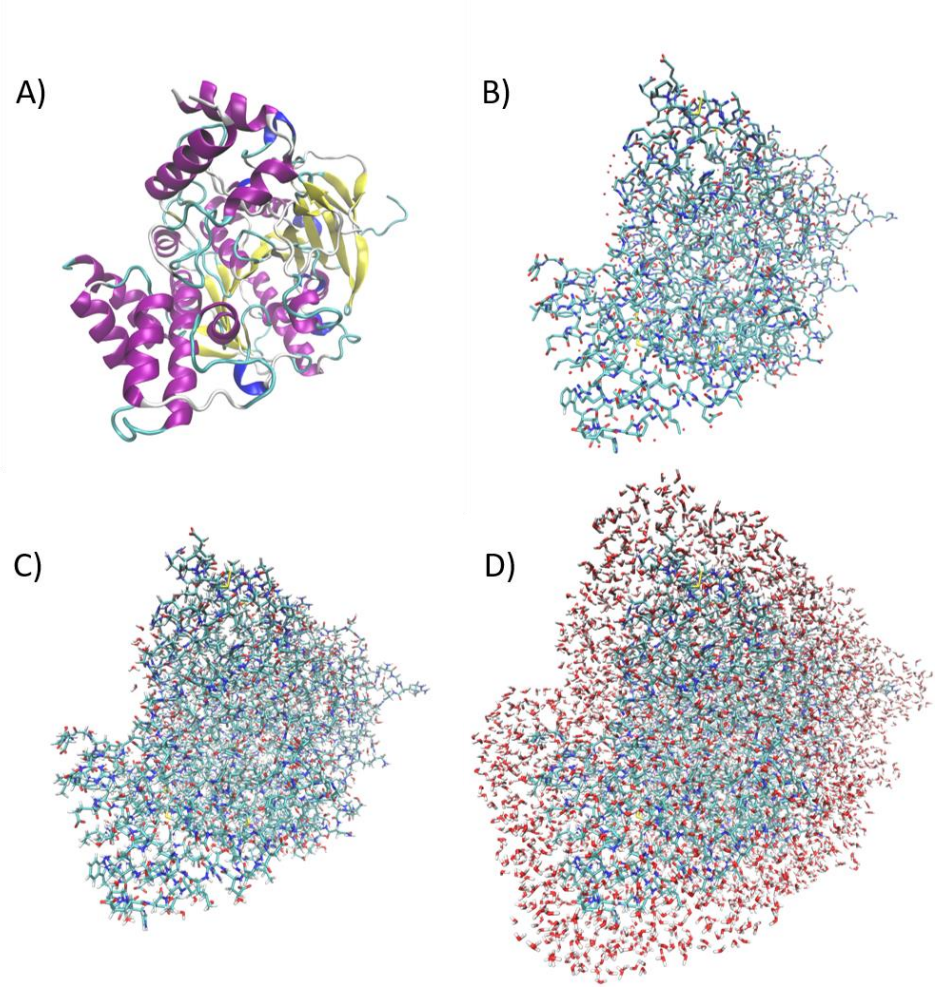


Figure III-4. Structures of AChE from the chain A of the PDB structure 3DL7. A) secondary structure of AChE in ribbon representation and secondary structure colouring. (crystal water molecules not apparent) B) heavy atoms of AChE. C) AChE structure including hydrogens. D) AChE solvated in a shell of water molecules.

After a good starting structure has been selected and modified to contain the proper amino acids, ligands, co-factors and inhibitors, further modification is often required. Enzymes are very reactive and as such the pre-reactive complex, has a short life span. This short life span makes the crystallization of the pre-reactive complex of the enzyme with its substrate extremely difficult and such a structure is rarely available. Several strategies exist to introduce substrate analogues or have unreactive enzyme-substrate complexes. Those complexes are easily modifiable into a reaction relevant structure, usually the pre-reactant complex. Those strategies have the advantage of producing enzyme structures whose conformations are adapted to the presence of the substrate. The other approach is to use docking software to place the substrate in an "empty" enzyme structure. The protocol of docking software establishes list of likely binding site from the structure of the substrate. They then place the substrate in those binding sites and evaluate the energy. The site that produces the lower is the most likely binding site for this substrate in the enzyme.

Now with a complete structure of the heavy atoms of the enzyme (see **Figure III-4B**) and a reaction relevant geometry for the substrate and active site, hydrogen atoms must be added to the structure of the enzyme (see **Figure III-4C**). Those hydrogens are from the original

X-ray structures as the exact location of every hydrogen atom cannot be resolved at the resolution range of protein crystal. Several software packages and online tools include protocols to add hydrogens according to force field topologies and structures. The use of such automatized protocols however, is not sufficient to add all hydrogen atoms to the structure of the enzyme. Some residues, with pK_as in the 5 to 9 range, have an ambiguous protonation pattern at physiological pH for the studied enzyme (usually around 7.2).[39] For those residues protocols exist to analyse the environment of a specific residue to evaluate the most plausible protonation state. CpHMD is also a method that can be used to evaluate the pK_a of specific residues in enzyme. The standard pK_a of amino acids in water solution is found in **Table III-1**.

Table III-1. pK_a of the side chain of residues with a labile proton in water solution.

Residue	Asp	Glu	His imidazolium/ imidazole	Cys	Tyr	Lys	Arg	His imidazole/ imidazolate
pK_a	3.9	4.1	6.0	8.1	10.1	10.7	12.5	>14

As can be seen in **Table III-1** cysteines and histidines are the most likely to have a reversal of their proton pattern. Histidines are special cases as there are two possible protonation sites on the imidazolium cycle. Thus, a histidine can be protonated on site δ, protonated on site ε, protonated on both ε and δ sites or completely unprotonated (see **Figure III-5**). While some software check the environment of histidines to find the best protonation site, it is best to visually check the environment of a histidine.

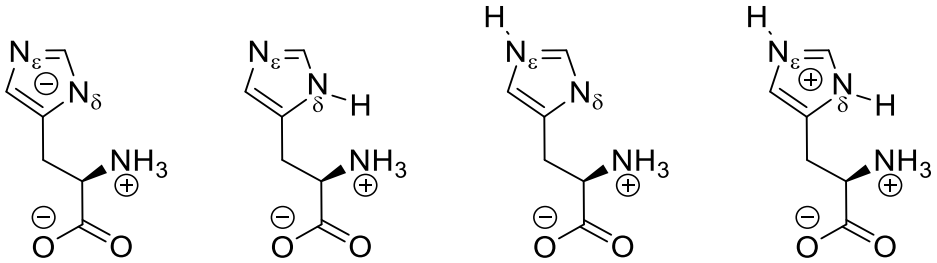


Figure III-5. Possible protonation states for the side chain of histidine residues

When all the hydrogens have been added, the now complete enzyme needs to be solvated (see **Figure III-4D**). The enzyme can be solvated in a water box or a water bubble centred on a specific residue or atom or the centre of mass of the enzyme. Another method is to encase the enzyme in a shell of water molecules up to a specified distance. Crystallization techniques usually remove most water molecules. The remaining water molecules are often structural water molecules with important roles in the function of the enzyme and should be kept.[40] Those water molecules do not represent all the water molecules naturally present in

[39] R. Vroman, L. J. Klaassen, M. H. C. Howlett, V. Cenedese, J. Klooster, T. Sjoerdsma, M. Kamermans, *PLoS Biol.* **2014**, *12*, e1001864.

[40] E. Meyer, *Protein Sci.* **1992**, *1*, 1543–1562.

the enzyme and specific protocols using molecular dynamics can be used to ensure that the water molecules around the enzyme properly penetrate the enzyme structure.

Through all the previous steps some energy minimization step or short dynamics might have been used, to minimize the position of water molecules for instance, but it should have been with a frozen backbone. To generate appropriate structures for QM/MM, snapshots should be drawn from molecular dynamics simulations at an appropriate temperature for the studied enzyme. Before running such molecular dynamics, the chemical systems must be heated from 0K to the desired temperature in increment of 50 or 25K over shorts periods of time. This is usually done with a frozen backbone. Once the system is at the desired temperature (usually 298K or 310K) the backbone is released and the system is equilibrated fully. When the total energy of the system is stabilized, snapshots can be drawn from the molecular dynamic. Ideally the same reaction or property is studied on several snapshots but the calculation cost will be a product of the number of snapshots and can quickly rise. Drawing snapshots after an extended molecular dynamic can also be ill advised as interesting structural features of a particular X-ray structure can be averaged out by a prolonged molecular dynamic, or if the force field is not relevant to treat a particular residue.

The final point to be raised is the selection of the atoms that will in the QM region. The QM region should obviously include at minimum all atoms involved in bond breaking or bond forming during the considered reaction. It requires a good knowledge of the enzyme and its reactivity to choose the QM region and no single protocol can be applied to choose an optimal QM region. Several guidelines however, can be followed to avoid common issues in the selection of the QM region. First the QM region should not be so small as to not include the first shell of atoms around the atoms involved in the reaction. Those atoms, including water molecules present in the active site, are important to distribute the electronic density and to describe the polarization in the active site of AChE. A large QM region will lead to more covalent bond crossing between the QM and MM regions. Those covalent bonds should be as apolar as possible with a strong σ character. Cutting through a very polar bond, like an amide bond, can over-polarize the QM region. In summary, when choosing the QM region, one should include enough atoms to properly describe the electronic environment of the reaction, while cutting through as few bonds as possible and making sure those bonds are simple apolar σ bonds. Those steps are the foundation of solid QM/MM simulations some examples of are presented along the next few paragraphs.

The first common use of QM/MM is for the refinement of the very same crystal structures necessary to perform QM/MM. In situations where an ambiguous structure was obtained from X-ray diffraction for the active site, QM/MM is now a refinement tool.[41] In some methods, raw crystallographic data can be used for geometry optimization to obtain refined structures.[42] The method was successfully used on human β alcohol dehydrogenase, the complex of yeast Hsp40 Ydj1 with its peptide substrate, and transition-state equivalent

[41] X. Li, S. A. Hayik, K. M. Merz Jr., *J. Inorg. Biochem.* **2010**, *104*, 512–522.
[42] U. Ryde, L. Olsen, K. Nilsson, *J. Comput. Chem.* **2002**, *23*, 1058–1070.

complexes of cytidine deaminase. But besides the betterment of structures, QM/MM can be used to evaluate important properties tied to the electronic structure.

QM/MM has been used in the study of excited states to evaluate spectroscopic properties.[43] One example is the study of the green fluorescent protein. The fluorescence of the chromophore of the enzyme was studied in the enzyme, in water solution and in gas phase. QM/MM with optimisation at the CASSCF level and energy evaluation at the CASPT2 level uncovered that the green fluorescent protein mimics a gas phase environment for the chromophore. The measured fluorescence in enzyme was closer to the measured florescence in gas phase with a single adjacent water molecule than it was to the measured fluorescence in water solution. QM/MM as a tool however, allows to go beyond enzymatic structure and electronic structure.

Finally, the most obvious use of QM/MM is the study of the enzymatic mechanism. An good example is the study of the role of a water molecule in the deprotonation of hydrogen peroxide in Horseradish peroxidase (HRP).[44] Previous QM/MM studies had found that in the reaction of HRP with hydrogen peroxide to form and iron oxo species the deprotonation of hydrogen peroxide was very large (near 20.0 kcal.mol^{-1}). Experimental kinetic studies however contradicted this and clearly indicated that the barrier was much lower around 1.6 kcal.mol^{-1}. The study by Derat, Shaik and co-workers showed that a water molecule would enter the active site and mediate the deprotonation of hydrogen peroxide. This mediation by a water molecule lowers the energy barrier to around 5 kcal.mol^{-1}.

IV. Conclusion

Theoretical chemistry is a great tool for the study of enzymatic systems. A perfect ideal method would reconcile the need for precision and accuracy in modelling chemical processes, the need to include fully the enzymatic system and its environment, and the need to take into account the dynamic dimension of the studied chemical phenomenon. Such method being unavailable, a compromise must be made, adapted to the subject at hand. A wide variety of methods exist, representing as many different compromises that can be made on these three axes. In this chapter, the principles behind the methods used throughout my PhD project have been presented, as well as how they are positioned on the three axes of chemical accuracy, systems size, and time scale.

[43] A. Sinicropi, T. Andruniow, N. Ferré, R. Basosi, M. Olivucci, *J. Am. Chem. Soc.* **2005**, *127*, 11534–11535.
[44] E. Derat, S. Shaik, C. Rovira, P. Vidossich, M. Alfonso-Prieto, *J. Am. Chem. Soc.* **2007**, *129*, 6346–6347.

Chapter 3: To Divide and Conquer

Chapter 3: To Divide and Conquer

I. Introduction

In this chapter, the study of the reactivation of VX-inhibited AChE by 2-PAM in a truncated model system is detailed (see **Figure I-1**). The goal of this study is to add residues and active site substructures one at a time to a simplified model of AChE's active site to rationalize the general role of those residues and structures in the reactivity. The minimal truncated model is composed of an analogue of the Ser203-VX adduct, $P(O)(OEt)_2CH_3$, the classical oxime 2-PAM and an imidazolium aromatic cycle below the oxygen of the ethoxy group opposition to 2-PAM which stands for the side chain of His447 (see **Figure I-1**). All other truncated model systems are built from this minimal model.

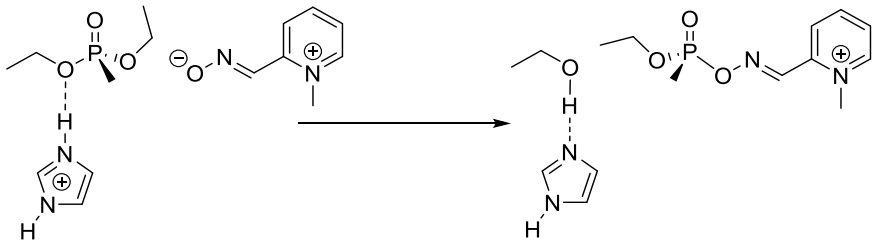


Figure I-1. Reactivation for the minimal truncated model of VX-inhibited AChE's active site with 2-PAM

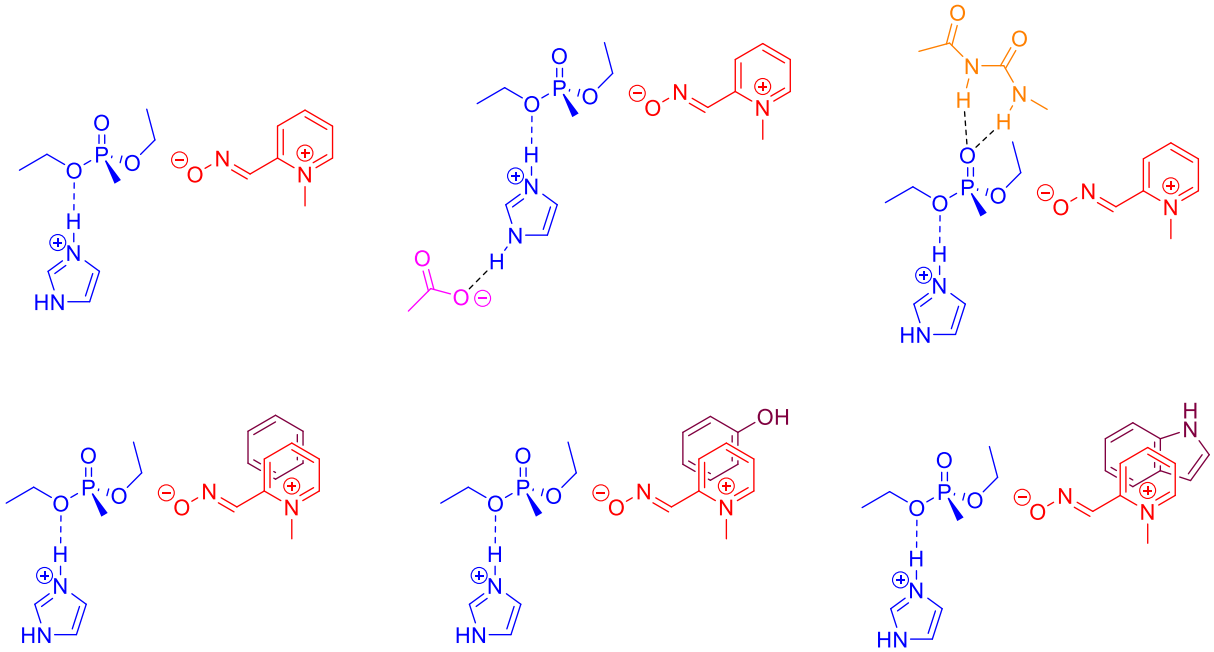


Figure I-2. Individual additions on the minimal model of VX-inhibited AChE's active site.

The minimal model includes analogues for two out of three members of the catalytic triad, Ser203 and His447(see **Figure I-2**). One of the first elements to be added to the truncated model was an analogue of Glu334, the third residue of the catalytic triad. An acetate was thus added near the second nitrogen of the imidazolium ring. Next a model of oxyanionic ring was added above the phosphoryl group of the VX model. Finally, a model with both

Glu334 and oxyanionic hole analogues was simulated. Another set of modifications was to add different aromatic rings analogues of tryptophan, phenylalanine, and tyrosine in position for π- π interactions and cation- π interactions near the pyridinium of 2-PAM. These combinations are individually shown on the second row of **Figure I-2**. All reactivator aromatic ring pairs have been simulated with the various active site models, from the minimal one to the complete one. All simulated combinations are listed in **Table I-1** as well as the section or subsection the results are presented in. These simulations will be discussed in section **VI**.

Table I-1. All simulated combinations of active site elements with the location of the relevant paragraphs.

	2-PAM	**2-PAM + benzene**	**2-PAM + phenol**	**2-PAM + indole**
VX-inhibited Ser203	II. Minimal model	V.1 Benzene	V.2 Phenol	V.3 Indole
VX-inhibited Ser203 + Glu334	III. The role of Glu334	V.1 Benzene	V.2 Phenol	V.3 Indole
VX-inhibited Ser203 + oxyanionic hole	IV. The oxyanionic hole	V.1 Benzene	V.2 Phenol	V.3 Indole
VX-inhibited Ser203 + Glu334 + oxyanionic hole	IV. The oxyanionic hole	V.1 Benzene	V.2 Phenol	V.3 Indole

As was discussed in the state of the art chapter, two mechanisms have been considered for the reactivation of nerve agent AChE (see **Figure I-3**). The first is the SN_2 mechanism with a single step and a single transition state. Both the distance between the phosphorus and the oxygen from the oxime and the distance between the phosphorus and the oxygen from Ser203 are expected to change simultaneously. The oxime will move closer to the phosphorus while the phosphorus moves away from the serine.

The second possibility is the addition-elimination mechanism. It is a two-step mechanism with two transition states surrounding a stable trigonal bipyramidal phosphorus intermediate. In this mechanism, the two reaction distances will change sequentially. First the distance between the phosphorus and the oxime will be reduced from the pre-reactive distance to a distance close to that of the covalent bond in the product. Then, the distance between the phosphorus and Ser203 will increase until the bond is broken.

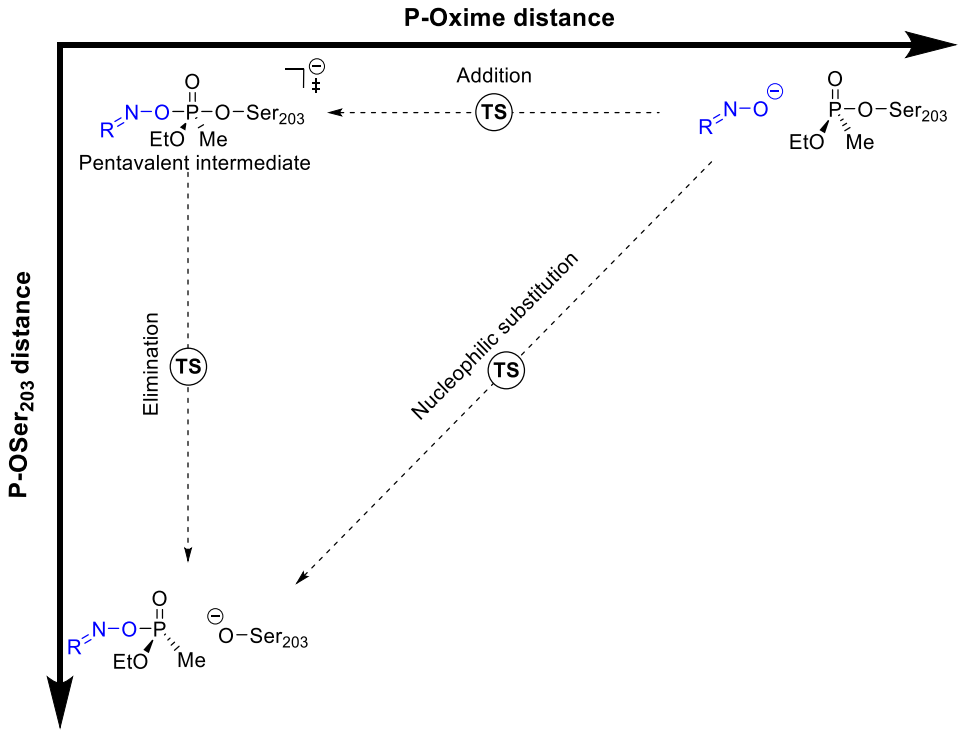


Figure I-3. Mechanistic rationalization for the reactivation of nerve agent inhibited AChE with key reaction coordinates.

The computational details for the simulations carried out in this chapter are given in the next two paragraphs. The simulations were performed in gas phase with the B3YLP functional,[1] the D3 dispersion correction,[2] and the def2-SV(P) basis set.[3] The calculation were performed using Turbomole V6.5.[4]

For every truncated model, the protocol was to optimise reactant and product states and to use a two-dimensional scan to obtain a geometry close to that of the TS. A two-dimensional scan was necessary after a monodimensional scan along the distance between the phosphorus and the oxygen of the oxime failed to locate a transition state. In the two-dimensional scan, the two reactive distances, between the phosphorus and the oxime and between the phosphorus and Ser203 were kept fixed and changed in incremental steps. Both were made to vary at from 1.5 Å to 2.5 Å in steps of 0.1 Å at least (in some specific cases the scans were extended or the step size was reduced. When relevant, it will be mentioned in the respective section). At every step, full geometry optimization was performed, except for the two restrained distances. On the two-dimensional potential energy surface obtained from the 2D scan the saddle point between the valley of the reactant and the valley of the product was located. The TS was then optimized using the geometry at the saddle point. Once the TS was localized two structures geometrically close to the product and reactant were identified and

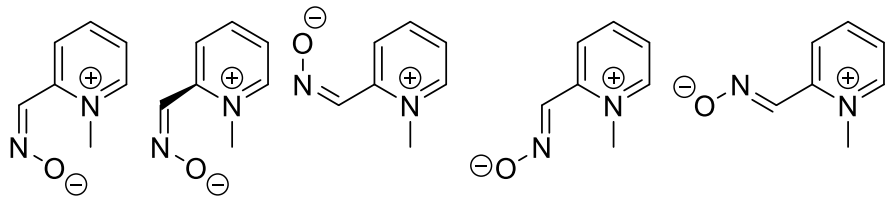
[1] a) A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652. b) C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.
[2] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
[3] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
[4] R. Ahlrichs, F. Furche, C. Hättig, W. Klopper, M. Sierka, F. Weigend, *TURBOMOLE*, TURBOMOLE GmbH, Development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, **2012**.

optimized with released constraints. Force calculations were used to confirm that true minima had been obtained for the reactant and product and that a first order saddle point for a single mode of vibration had been obtained for the transition state. A python script had to be written to make up for the lack of scanning tool in Turbomole and can be found in **Appendix 6**, Section **II.**

Throughout this study we have elected to do the simulations in gas phase. This choice was made in order to exclude another factor in the reactivity of the system and to ensure that any trend observed can be attributed unambiguously to the added or remove element to the model. In two cases, the simulations were repeated with the COSMO implicit solvation model, for the minimal truncated model and for the model with an acetate. These were performed to establish a comparison with previous truncated model studies. The first simulation to be detailed here is the simulation of the minimal truncated model.

II. Minimal model

Before the reactivation simulations, the possible geometries for 2-PAM must be investigated. Because of the double bond in the oxime functional group, 2-PAM can either be of *Z* or *E* configuration. Besides those stereo isomers, the rotation of the C-C bond between the oxime and the pyridinium give rise to two possible conformations *E* and *Z* depending on the N-C-C-C angle. The possible conformations are presented in **Figure II-1** with their relative energies compared to the one with the lowest energy. The energy was obtained after geometry optimization of the structure.



Isomer	*ZZ*	*ZZ_2*	*ZE*	*EZ*	*EE*
Energy (kcal.mol^{-1})	+11.79	+7.06	+3.45	+1.03	0.00

Figure II-1. Possible geometries for 2-PAM with their configuration and relative energy shown below.

As evidenced by the data, the E isomer of the C-N double bond in the oxime is systematically more favourable. Between the 2 possible conformations for E-2-PAM the difference in energy is very small. The E conformation, with the oxime functional group opposite to the pyridinium, is the most favourable one. This conformation is the one used in the simulations of VX-inhibited AChE reactivation in the truncated model. This data is coherent with previous investigations of this question.[5]

The minimal model has some marginal flexibility with the ethoxy substituents of the phosphorus but is mostly very rigid. It greatly simplified the simulations on this model. The

[5] C. S. Ewig, J. R. Van Wazer, *J. Mol. Struct.: THEOCHEM* **1988**, *168*, 235–246.

structures of the reactant, product and transition state can be found in **Figure II-2** and the two-dimensional potential energy surface in **Figure II-3**.

In the reactant, the distance between the phosphorus and Ser203 (P-Oser203) and between the phosphorus and the oxime (P-Ox) are 1.71 Å and 2.74 Å respectively (See **Figure II-2A**). This local minimum was obtained by removing the constraints on the structure with P-OSer203 1.70 Å and P-Ox 2.50 Å from the two-dimensional scan. The P-OSer203 and P-Ox of the product are 2.54 Å and 1.80 Å respectively (See **Figure II-2B**). This local minimum was obtained by removing the constraints on the structure with P-OSer203 2.50 Å and P-Ox 1.80 Å from the two-dimensional scan. Therefore, we can follow the minimum energy path on the potential energy surface. What can be seen is that to follow that minimum energy path both P-OSEr203 and P-Ox reaction coordinates must vary simultaneously. There is, along that energy path a single saddle point indicating a transition state. This transition state was optimized (See **Figure II-2C**). In the transition state, the P-Oser203 bond is 1.86 Å and the P-Ox bond is 2.11 Å. As indicated while the P-Oser203 has lengthened by 0.15 Å, the P-Ox bond has been shortened by 0.63 Å.

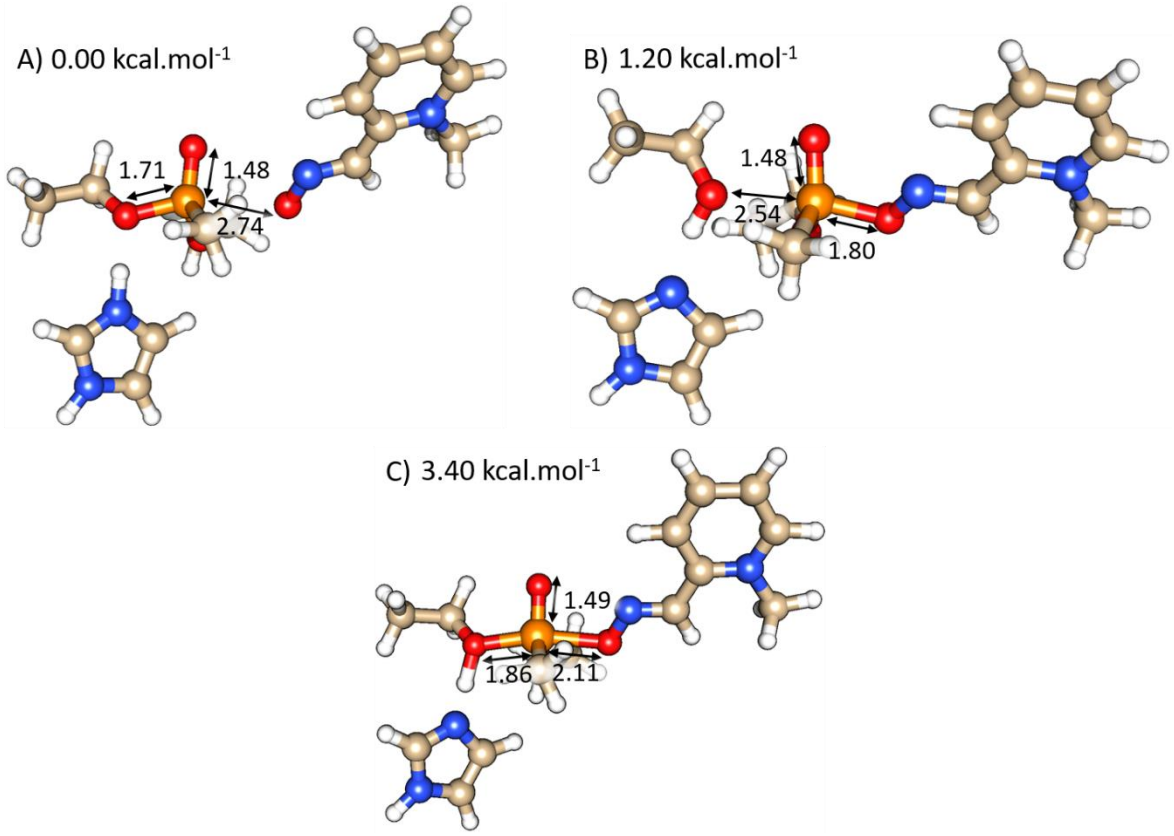


Figure II-2. Optimized structures of (A) the reactant, (B) the product, and (C) the transition state for the minimal model. Ball&Stick representation with regular chemical element colouring (carbons are in tan colouring).

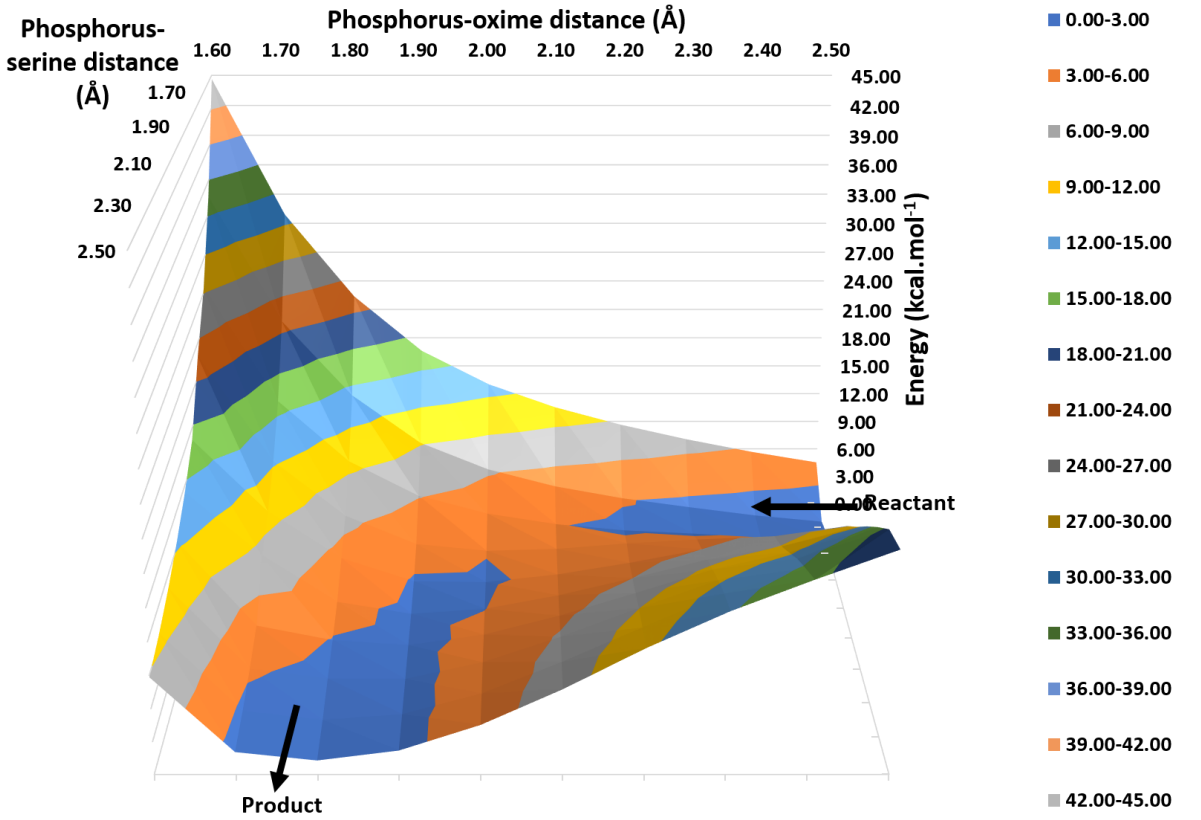


Figure II-3. Three-dimensional potential energy surface for the reactivation of the minimal model of VX-inhibited AChE by 2-PAM.

While not a reaction coordinate, the length of the phosphoryl bond is an interesting structural feature. It is supposed to be polarized in the intermediates structures and at the transition states involving nerve agents. In the reactant and product of the minimal model, the length of the phosphoryl bond is 1.48 Å. This distance only very slightly increases in the transition state structure to 1.49 Å. It will be interesting to compare this structural feature when more elements will be added to the model.

From this data, we can safely say that the reactivation process goes through a SN_2 mechanism. There is single transition state and on the potential energy surface both reactive distances vary simultaneously. The reactivation of VX-inhibited AChE in the minimal truncated model has an energy barrier of 3.40 kcal.mol^{-1} and is slightly endothermic, by 1.20 kcal.mol^{-1}.

Similar simulations were also conducted for the same system with the COSMO implicit solvent model and a dielectric constant of 80, for a water solvent.[6] The transition state was not localized in this set of simulations as the potential energy surface (**Figure II-5**) and reactant and product structures (**Figure II-4**) are sufficient to establish a comparison. Both the reactant and the product structures were obtained by releasing constraints on low energy reactant-like and product-like structures from the potential energy surface. In the reactant (**Figure II-4A**), the P-Ox and P-OSer are 2.76 Å and 1.68 Å respectively. The product (**Figure II-4B**) as a P-Ox of 1.71 Å and a P-OSer of 3.09 Å. These two structures can be connected by a minimum energy path. Following this minimum energy path it is evident that both reaction coordinates vary

[6] A. Klamt, G. Schüürmann, *J. Chem. Soc., Perkin Trans. 2* **1993**, 799–805.

very sequentially. First the P-Ox distance is reduced, then the P-OSer distance is increased. This profile is very different from the profile obtained in gas phase and much more evocative of an addition-elimination mechanism. The reaction is slightly endothermic with an energy difference between the reactant and the product of 0.73 kcal.mol^{-1}.

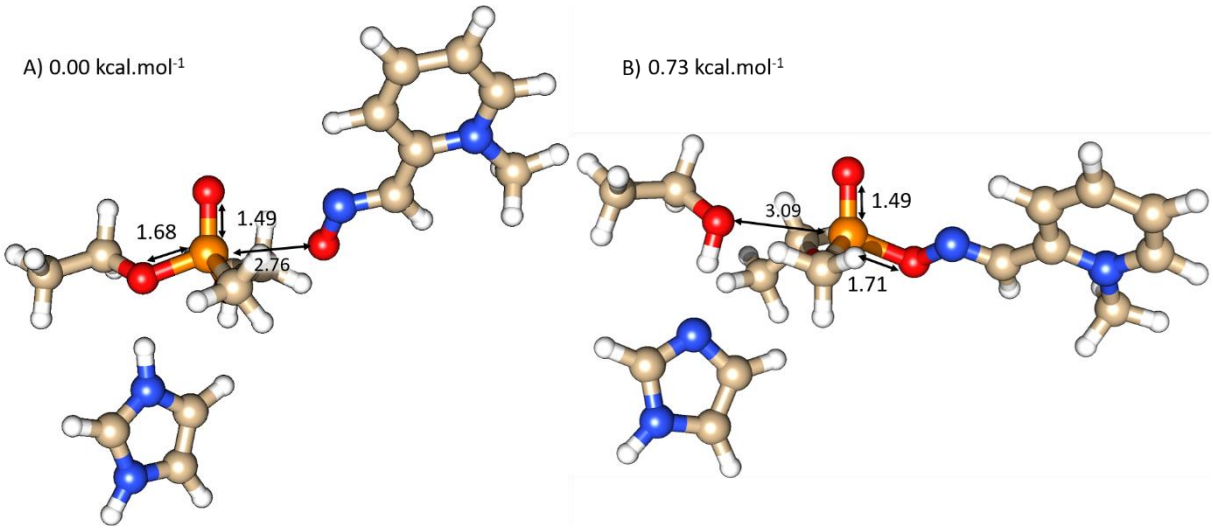


Figure II-4. Optimized structures of (A) the reactant and (B) the product for the minimal model in COSMO implicit solvation model.

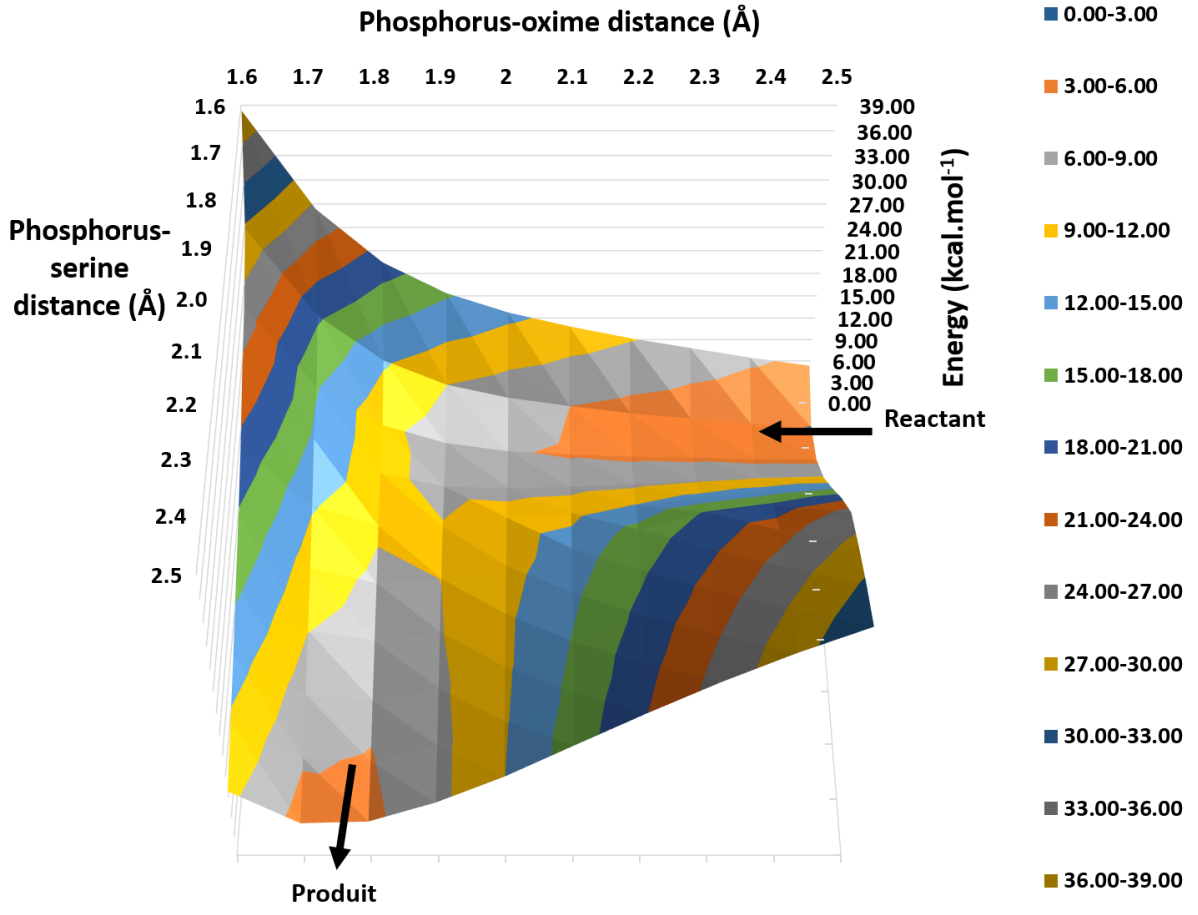


Figure II-5. Three-dimensional potential energy surface for the reactivation of the minimal model of VX-inhibited AChE by 2-PAM with the COSMO implicit solvent model

III. The role of Glu334

The role of Glu334 is studied by adding an acetate near the second protonated site of the imidazole standing for His447. The first step of the protocol is to optimize both reactant and products. The reactant structure that can be seen in **Figure III-1A**, has been obtained by lifting the constraints on both P-OSer203 and P-Ox bonds and optimizing the scan point where those distances where 1.70 Å and 2.50 Å respectively. In the reactant structure, the imidazolium cedes a proton to the acetate. In the optimized reactant structure, the oxime is much further away from the phosphorus than for the minimal model. It is at 3.81 Å from the phosphorus. The P-Oser203 bond has slightly shortened, compared to the minimal model, to 1.64 Å.

The optimization of the product seen in **Figure III-1B** proved to be very difficult. The 2-PAM-VX adduct had to be manually moved more than 7.0 Å away from the ethoxy group corresponding to the reactive serine. There again the imidazolium cedes a proton to acetate and a second proton to the ethoxy group representing Ser203. It forms a negatively charged imidazolate ring. In the optimized product structure, the P-OSer203 distance is 3.25 Å and the P-Ox distance is 1.75 Å. The energy of the product stands 48.18 kcal.mol^{-1} higher than that of the reactant indicating a very endothermic process. An extended two-dimensional scan along the two reaction coordinates was performed to attempt to locate a transition state (see **Figure III-2**).

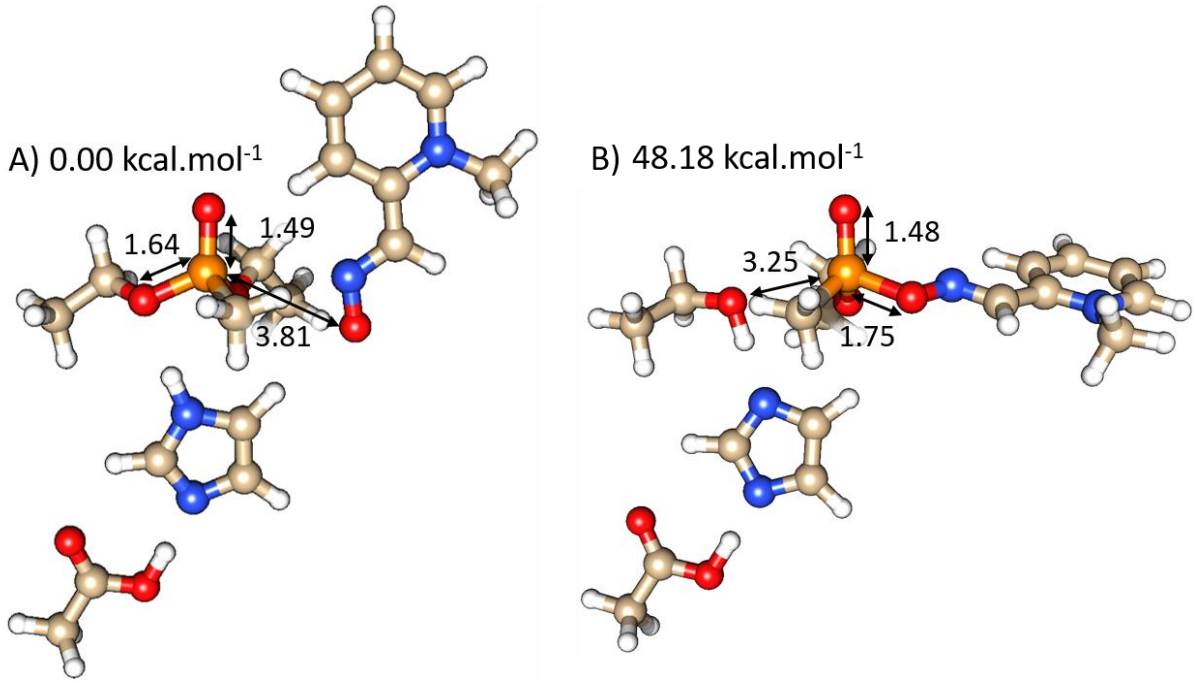


Figure III-1. Optimized structures of (A) the reactant and (B) the product for the model with acetate.

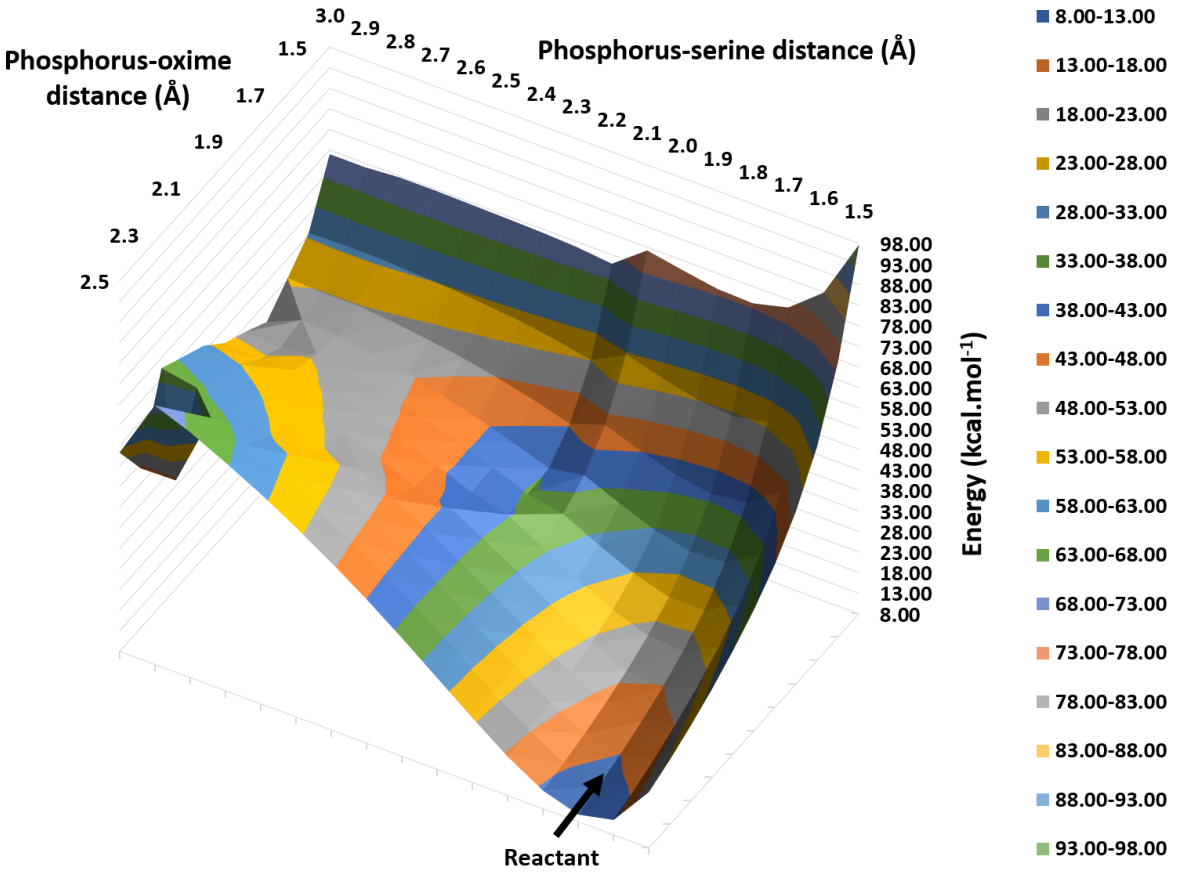


Figure III-2. Three-dimensional potential energy surface for the reactivation of the model with acetate of VX-inhibited AChE by 2-PAM.

From looking at the potential energy surface (**Figure III-2**) we see that reducing the distance between the oxime and the phosphorus or pulling Ser203 away from the VX adduct only increases the energy of the system. If the system is forced into a geometry with a P-Ox distance in the 1.8-1.7 Å range and a P-OSer203 in the 2.5-3.0 Å range the energy reaches a plateau and stagnates between 48 kcal.mol^{-1} and 53 kcal.mol^{-1}. Further extending the scan up to a P-OSer203 distance of 3.5 Å yields an unproductive geometry with 2-PAM sliding parallel to the imidazole to form a cation-π interaction. No transition state could be located on the potential energy surface. When optimized without constraints, any point from the scan returns to the reactant state.

From this data, what is apparent is that the addition of the acetate blocks the reactivation by deprotonating the imidazolium to an imidazole. The deprotonation of the imidazole into an imidazolate has a very large energetic cost that almost completely prevents the reactivation from taking place. As is, in this gas phase model, the reactivation is endothermic with an energy difference of 48.18 kcal.mol^{-1} and an energy barrier at least greater than 48.18 kcal.mol^{-1}.

A similar set of simulations was also conducted for the model with acetate with the COSMO implicit solvent model and a dielectric constant of 80, commonly used for a solvent as polar as water.[6]

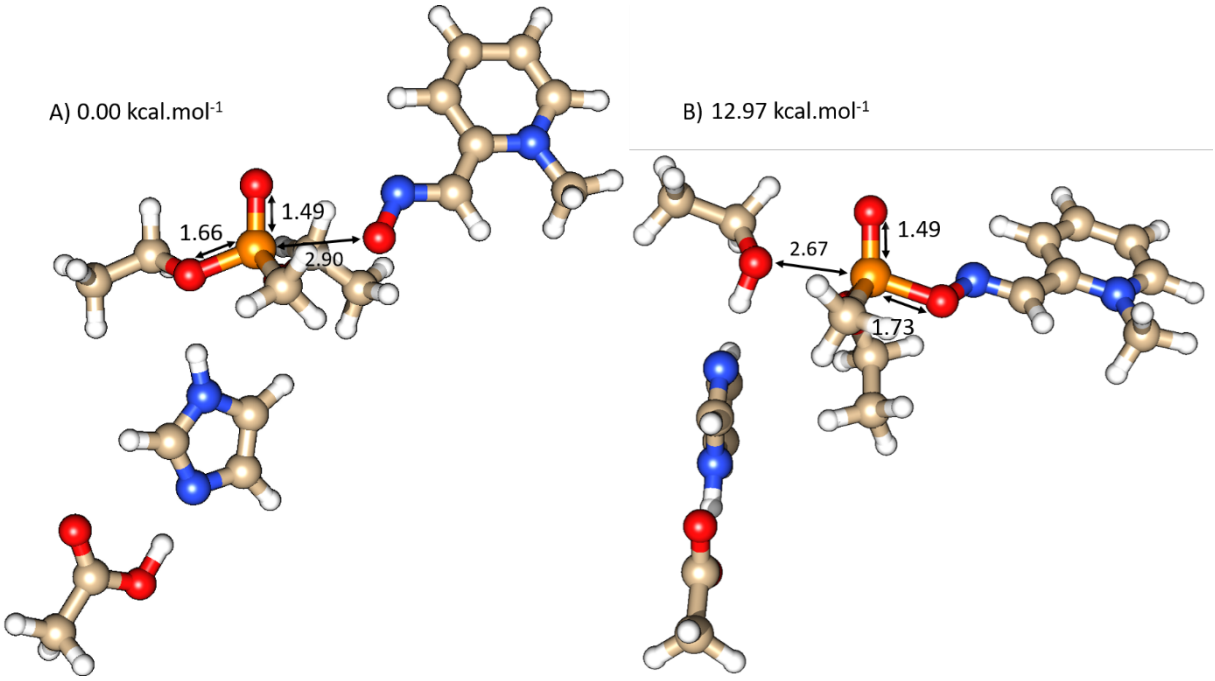


Figure III-3. Optimized structures of (A) the reactant and (B) the product for the model with acetate in COSMO implicit solvation model.

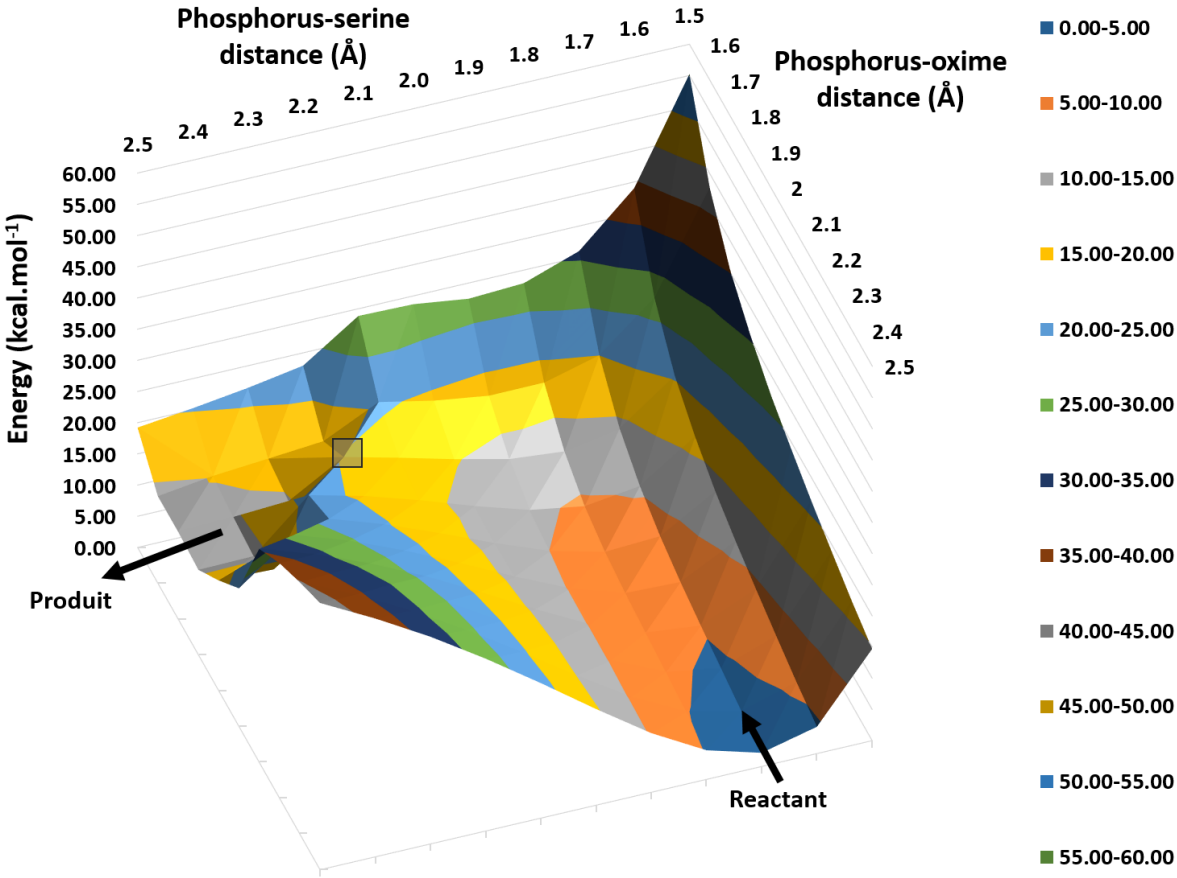
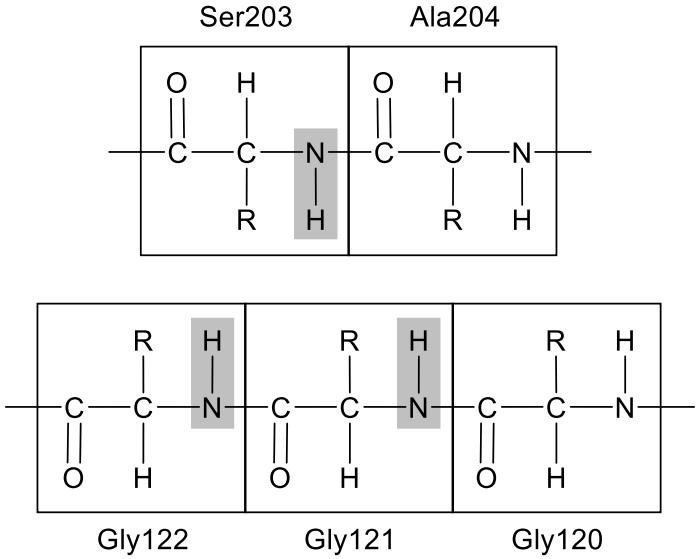


Figure III-4. Three-dimensional potential energy surface for the reactivation of the model with acetate of VX-inhibited AChE by 2-PAM with the COSMO implicit solvent model. The likely area of the transition state is materialized by a grey square.

No attempt was made to localize the transition state in this set of simulations. The potential energy surface (**Figure III-4**) and reactant and product structures (**Figure III-3**) are sufficient to establish a comparison with gas phase simulations. The structure of the reactant and the product were obtained by releasing constraints on low energy reactant-like and product-like structures from the potential energy surface. Contrary to the gas phase simulations of the model with acetate, an energy barrier clearly appears on the potential energy surface and the product was easily optimized. The reactant (**Figure III-3A**) has a P-Ox distance of 2.90 Å and a P-OSer distance of 1.66 Å. In the product (**Figure III-3B**) those distances are 1.73 Å and 2.67 Å respectively. Those two structures can be connected by a minimum energy path going through a single saddle point. Along this minimum energy path, the reaction distances vary sequentially, first the P-Ox is shortened, then the P-OSer is lengthened. This minimum energy path is evocative of an addition elimination mechanism. The reaction is highly endothermic with an energy difference between reactant and product of 12.97 kcal.mol^{-1}. The comparison between gas phase simulations and simulations performed with implicit solvent is done in the discussion section (section **VI**) of this chapter.

IV. The oxyanionic hole

The addition of the oxyanionic hole to the model system posed a challenge because of the mobility and flexibility of the components of this oxyanionic hole. The oxyanionic hole is composed of the three amino moieties of three peptide bonds involving five residues. The amides involved in the oxyanionic hole are between Gly120 and Gly121, Gly121 and Gly122, and Ser203 and Ala204 (see **Scheme IV-1**). Including exactly these residues would not provide a satisfactory model for the oxyanionic hole (see **Figure IV-1A**). It would leave three amide bonds cut in half with naked carbonyls and amine in the model. Instead a good model is to replace the naked carbonyls and amine by methyl groups. This produces the model of oxyanionic hole show in **Figure IV-1B**.



Scheme IV-1. Residues involved in the oxyanionic hole of AChE. The amino moieties involved in hydrogen bonding as part of the oxyanionic hole are in grey.

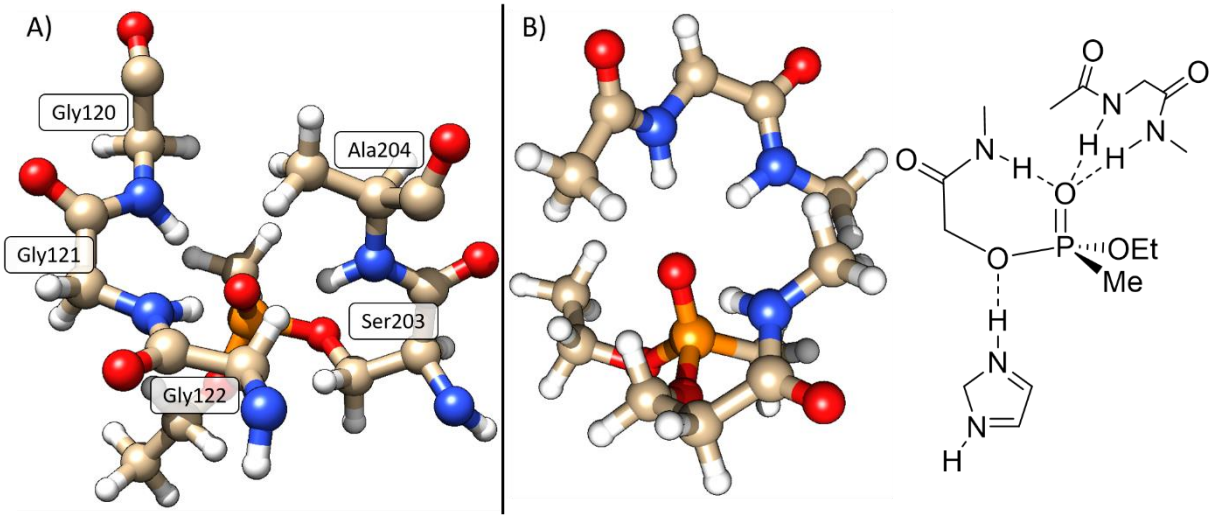


Figure IV-1. oxyanionic hole of AChE. A) residues of the oxyanionic hole taken directly from the enzyme. B) model of oxyanionic hole used

This model of oxyanionic hole proved to be too flexible and produced several unproductive conformations. The most common evolution of the system was the interaction of the amine moiety of the amide between Ser203 and Ala204 with the carbonyl moiety of the amide between Gly121 and Gly122 (see **Figure IV-2A**). This unproductive geometry sometimes evolves further by the unfolding of the amide between Ser203 and Ala204. In those conformations, the imidazolium would eventually move to interact with the carbonyl moiety of this amide rather than the phosphorus bond oxygen of Ser203. (see **Figure IV-2B**)

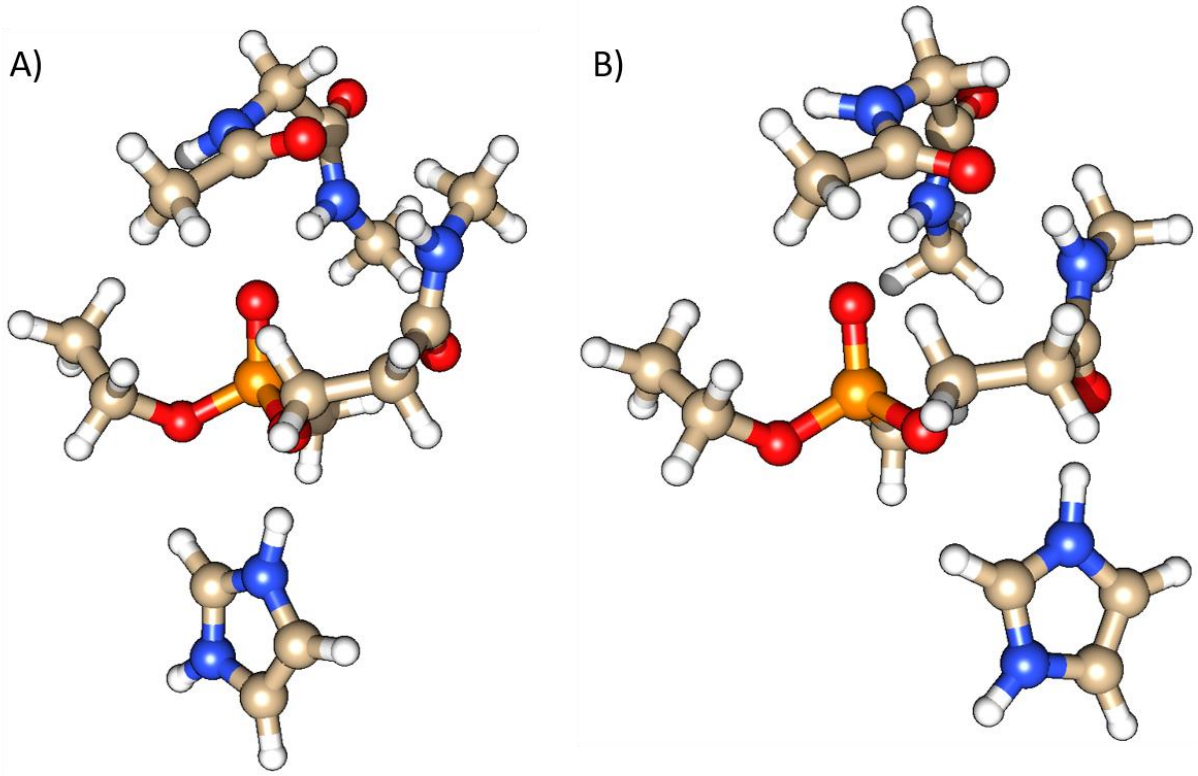


Figure IV-2. Unproductive geometries for oxyanion hole models.

To reduce the flexibility of the system and limit its instability, the amide bond connected to Ser203 was removed (**Figure IV-3**). While this new oxyanionic hole model is more different to the actual oxyanionic hole it maintains two hydrogen bonds with the VX adduct more consistently.

The reactant (**Figure IV-4A**) was optimized by releasing the constraints on the scan point with P-OSer203 and P-Ox at 1.70 Å and 2.70 Å respectively. The optimized reactant has a P-OSer203 bond of 1.68 Å and a P-Ox bond of 2.89 Å. The product (**Figure IV-4B**) was similarly optimized by releasing constraints on the scan point with P-OSer203 and P-Ox at 2.50 Å and 1.80 Å respectively. The optimized product has a P-OSer203 bond of 2.75 Å and a P-Ox bond of 1.74 Å. We can then follow on the potential energy surface the minimum energy path between the two points from which the reactant and product could be optimized (see **Figure IV-5**). Following the minimum energy path what we can see is that at first the only reaction distance to vary is the P-Ox bond, until it reaches the 2.10-2.00 Å range. During that approach of the oxime the P-OSer203 is quasi-constant. The P-Ox distance then varies very slowly and the P-OSer203 distance increases rapidly until the product is reached. There is a single saddle point along that minimum energy path. From this saddle point a transition state was optimized (see **Figure IV-4C**). The P-OSer203 bond is 1.74 Å and the P-O bond is 2.27 Å.

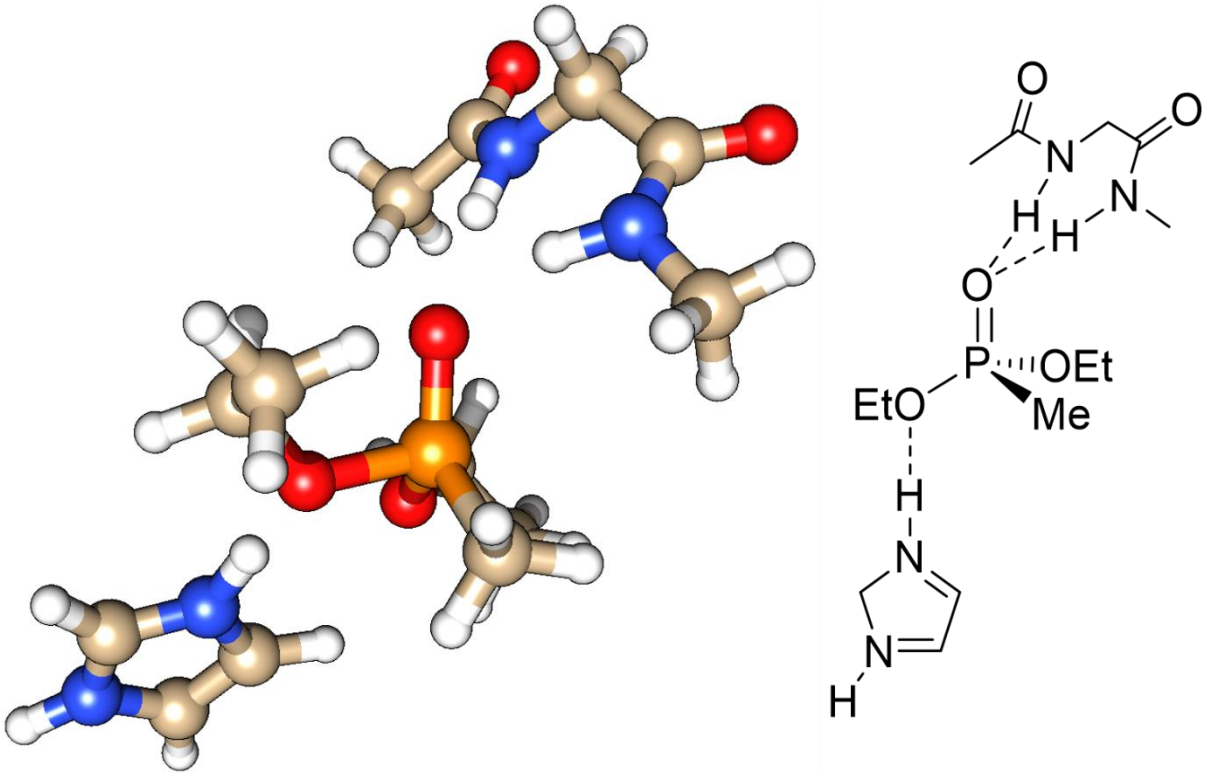


Figure IV-3. Selected oxyanionic hole model

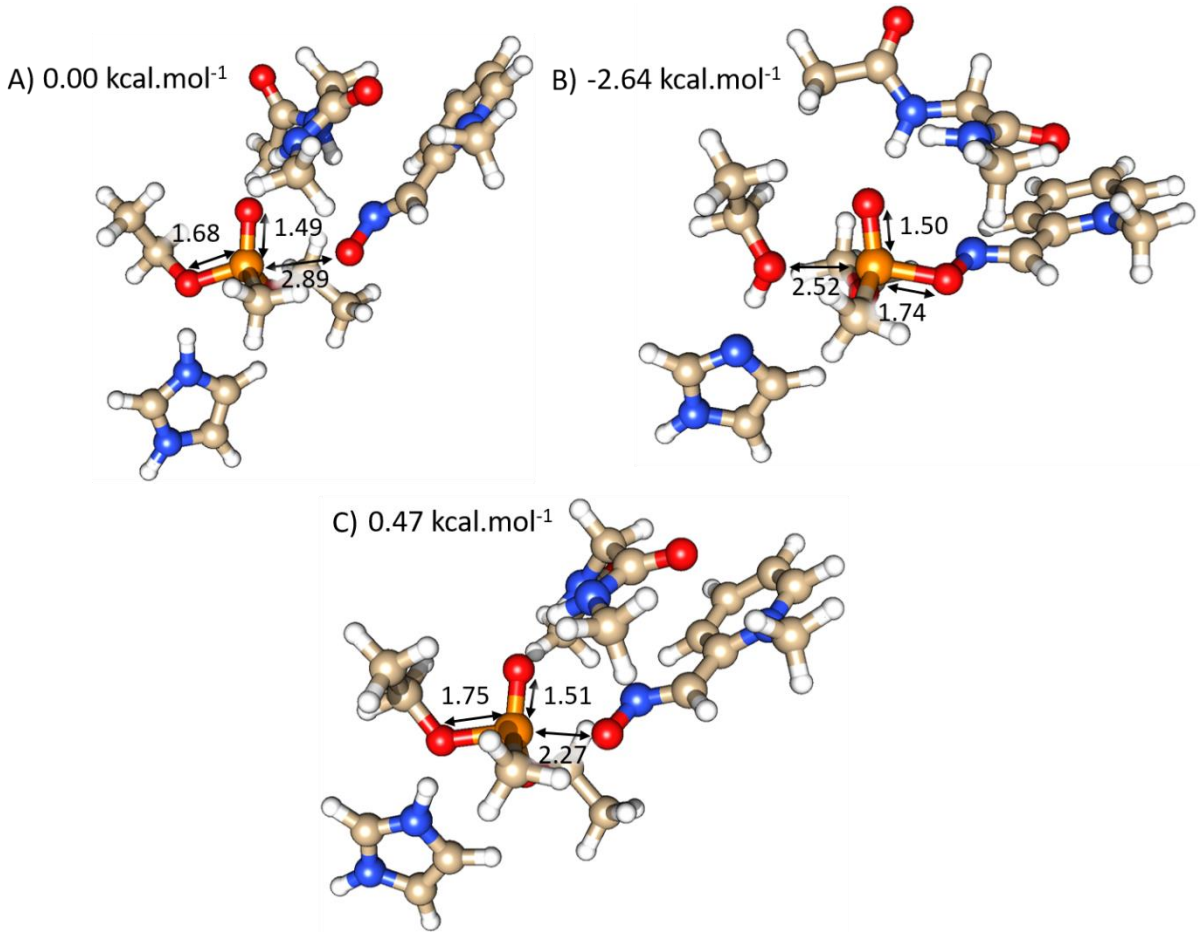


Figure IV-4. Optimized structures of (A) the reactant, (B) the product, and (C) the transition state for the model with oxyanionic hole.

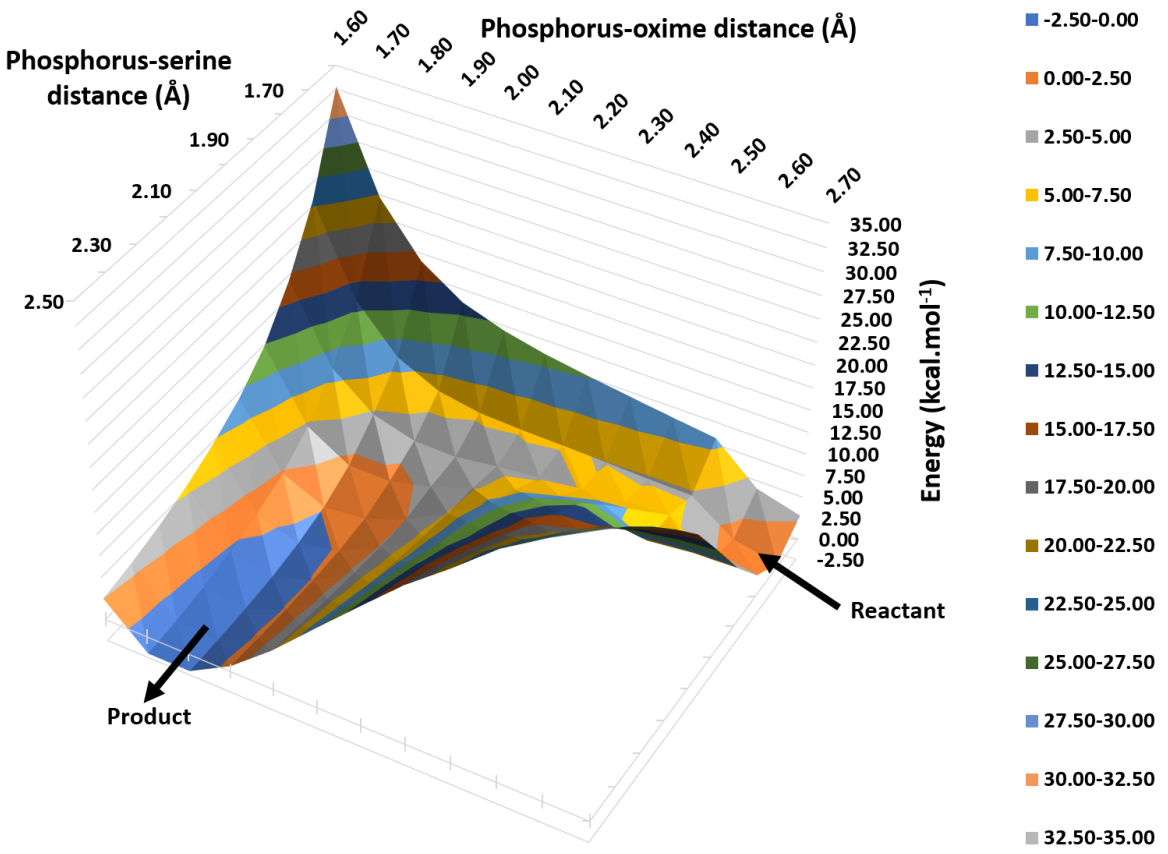


Figure IV-5. Three-dimensional potential energy surface for the reactivation of the model with oxyanionic hole of VX-inhibited AChE by 2-PAM.

When the acetate is added into the model however, once again the reactivation is not possible. The structure of a reactant was optimized (see **Figure IV-6**) but it was not possible to optimize a product structure. The reactant was optimized by releasing the constraints on the scan point with P-Oser 1.60 Å and P-Ox 2.50 Å. In the optimized structure, the reactivator moves away with a P-Ox bond of 3.04 Å. The other reactive distance doesn't diverge from expectation with a P-OSer at 1.65 Å.

The potential energy surface obtained from the two-dimensional scan (**Figure IV-7**) had a similar profile than the potential energy surface obtained with the acetate alone (see **Figure III-2**). Any attempt to progress towards a product by either shortening the P-Ox distance or increasing the P-OSer203 distance only results in an increase in energy. The potential energy surface does not allow to locate neither product or transition state.

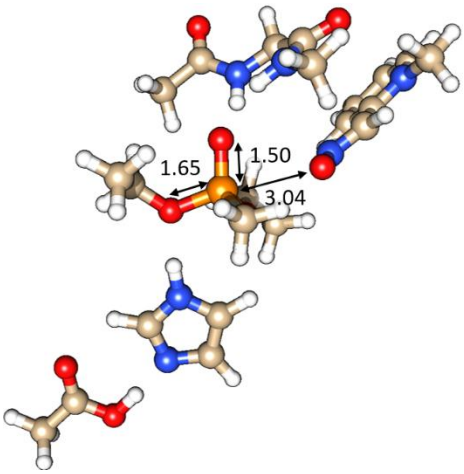


Figure IV-6. Optimized structure of the reactant for the model with acetate and oxyanionic hole.

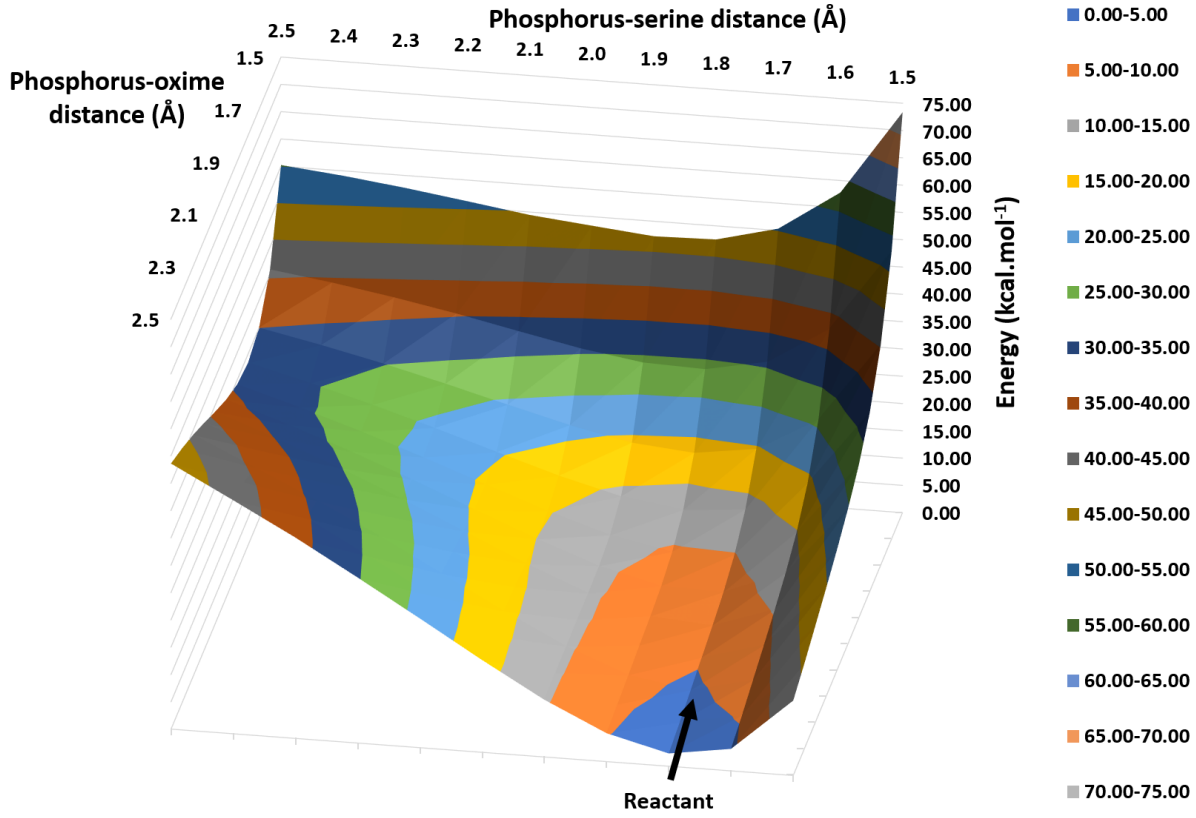


Figure IV-7. Three-dimensional potential energy surface for the reactivation of the model with acetate and oxyanionic hole of VX-inhibited AChE by 2-PAM.

The only thing that can be said on the energetics of the reactivation of the model with both the oxyanionic hole and the acetate is that the energy barrier is greater than 33 kcal.mol^{-1}. The acetate seems to completely block the reaction regardless of the influence of the oxyanionic hole.

V. The aromatic rings in the active site of AChE

The final addition to the truncated model of AChE's active site is an aromatic ring parallel to the pyridinium of 2-PAM to model the π-π and cation-π interactions aromatic

residues of the active site like Tyr337, Phe338, Phe295, Phe297, Tyr124, Tyr341, or Trp86. Three aromatic rings here used to model those residues, benzene, phenol, and indole (see **Figure V-1**). For those three aromatic rings, simulations were performed for truncated models with only the minimal model and the aromatic ring, with the acetate and the aromatic ring, with the oxyanionic hole and the aromatic ring, and with both the acetate and the oxyanionic hole, and the aromatic ring.

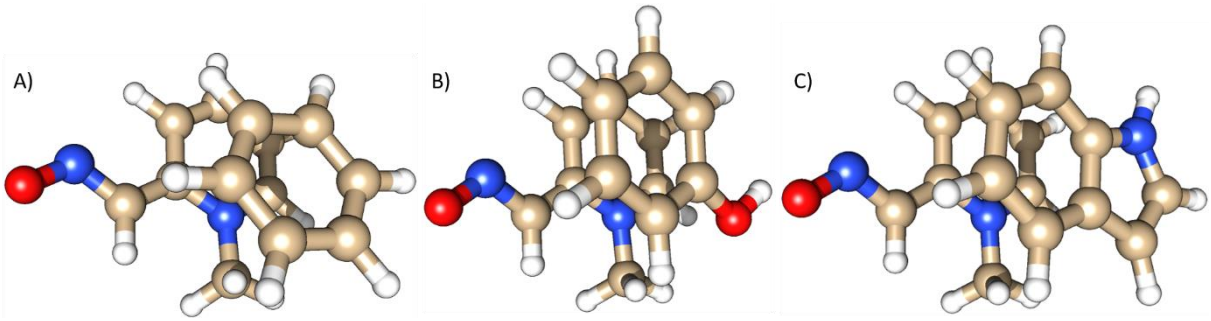


Figure V-1. aromatic rings used to model aromatic residues of the active site of AChE

In the next few paragraphs the optimized structures and potential energy surface for all twelve combinations are presented and briefly discussed. The global analysis of the results obtained for the various truncated models including aromatic rings is carried out at the end of this subsection. A global analysis of all truncated models is carried out in the next and final subsection.

V.1. Benzene

The first aromatic ring to be added to the model is benzene to model phenylalanine residues of AChE's channel. First benzene was added to the minimal model. Both reactant and product presented in **Figure V-2** were obtained by removing constraints on scan points on lower energy and geometrically closest to the desired local minima. The P-Ox and P-OSer203 distances in the reactant (**Figure V-2A**) are respectively 3.06 Å and 1.69 Å. The P-Ox and P-OSer203 distances in the product (**Figure V-2B**) are respectively 1.77 Å and 2.52 Å. We can follow the minimum energy path that connects the reactant to the product on the potential energy surface (**Figure V-3**). There is a single saddle point from which a transition state could be optimized (**Figure V-2C**). This transition state has P-Ox and P-OSer203 at 2.17 Å and 1.79 Å respectively. It is structurally closer to the reactant than to the product as we see that the P-OSer203 only increases by 0.1 Å while the P-Ox distance has been reduced by 0.89 Å. On the potential energy surface, the minimum energy path shows neither that the shortening of the P-Ox and lengthening of the P-Oser203 is sequential, nor that the two are concurrent. Overall, the reactivation in this model is exothermic with an energy difference of -1.54 kcal.mol^{-1} and a barrier of 4.07 kcal.mol^{-1}.

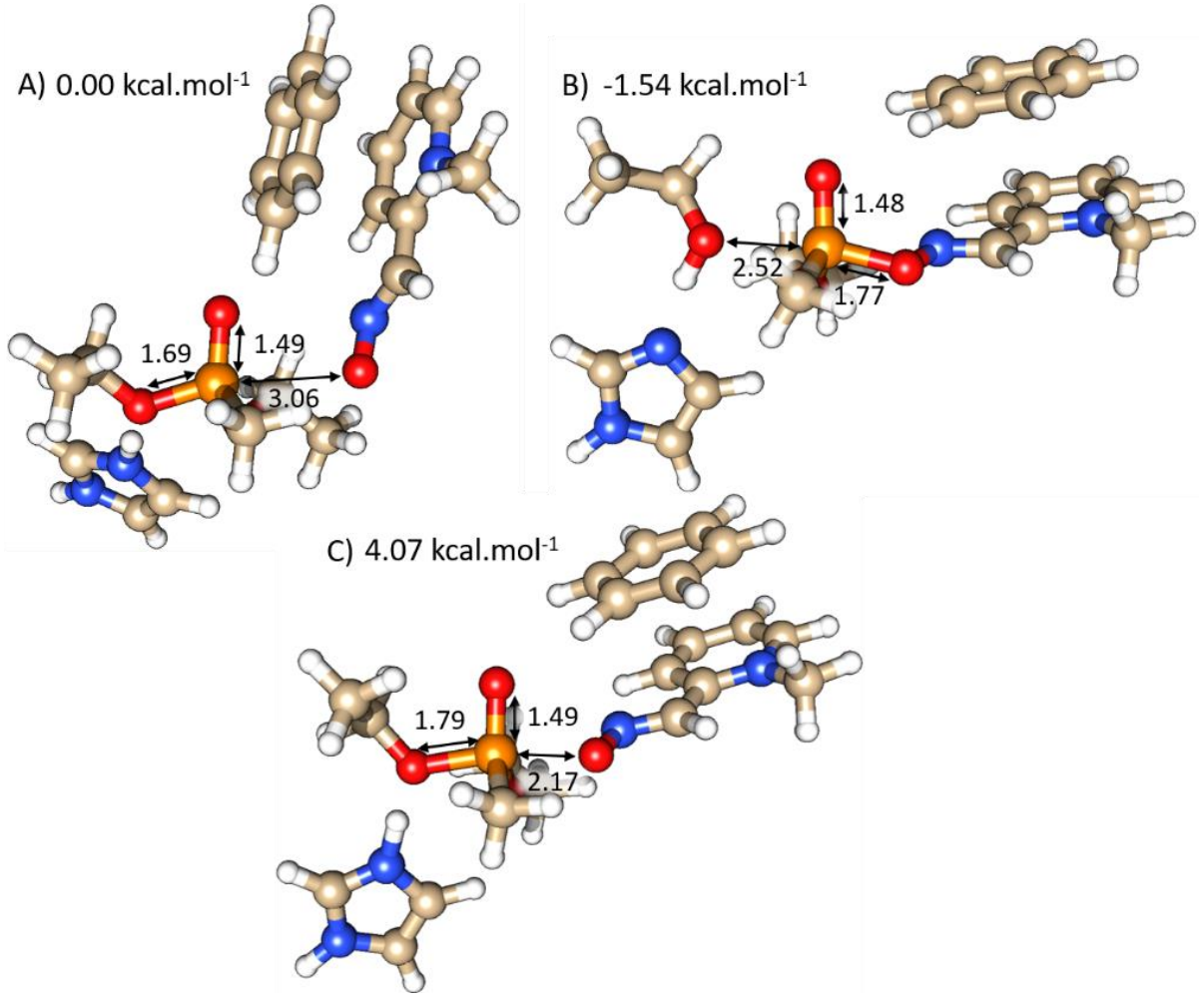


Figure V-2. Optimized structures of (A) the reactant, (B) the product, and (C) the transition state for the model with benzene.

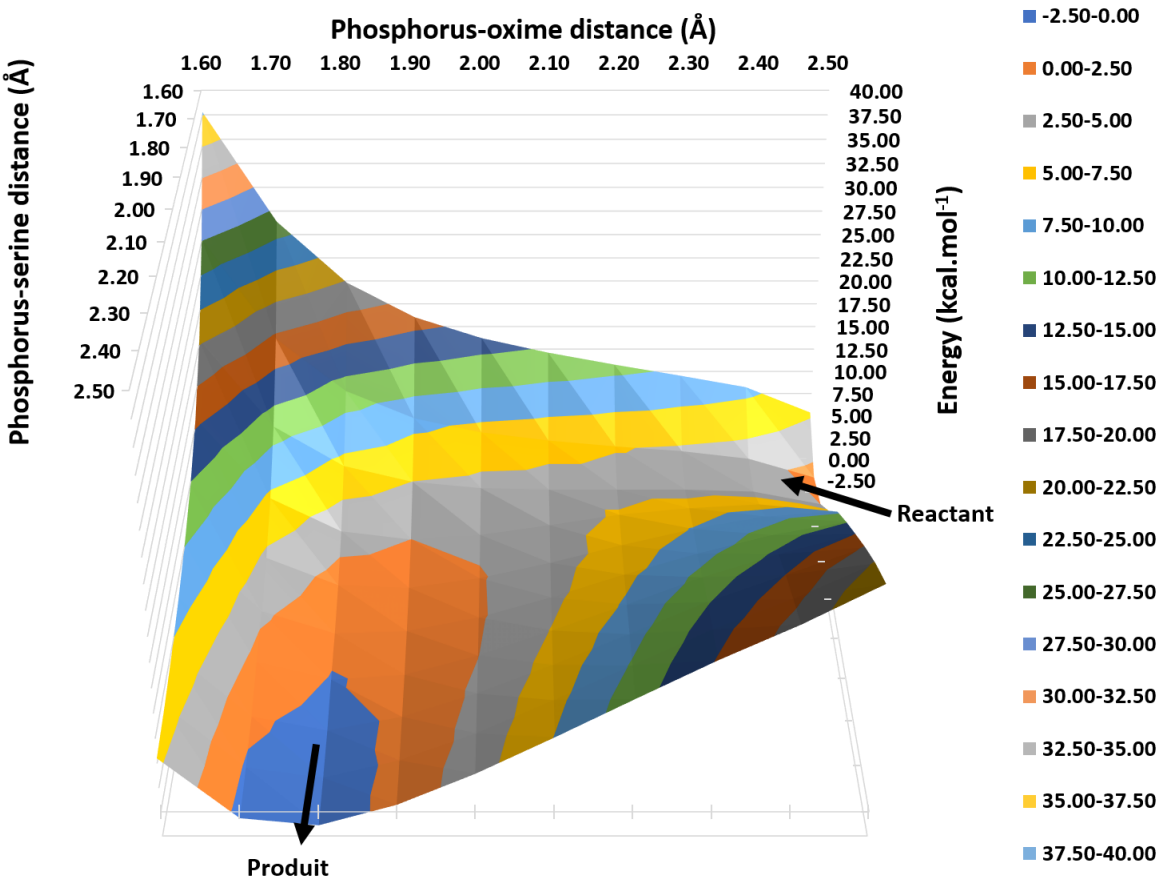


Figure V-3. Three-dimensional potential energy surface for the reactivation of the model with benzene of VX-inhibited AChE by 2-PAM.

In the next set of simulations, the oxyanionic hole analogue was added to truncated the model. Both reactant and product presented in **Figure V-4** were obtained by removing constraints on scan points on lower energy and geometrically closest to the desired local minima. In the reactant (**Figure V-4A**), P-Ox and P-OSer203 are respectively 2.57 Å and 1.70 Å in the product (**Figure V-4B**) those two distances are respectively 1.74 Å and 2.52 Å. We can follow the minimum energy path that connects the reactant to the product on the potential energy surface (**Figure V-5**). There is a single saddle point from which a transition state could be optimized (**Figure V-4C**). This transition state has P-Ox and P-OSer203 at 2.14 Å and 1.77 Å respectively. It is structurally closer to the reactant than to the product as we see that the P-OSer203 only increases by 0.07 Å while the P-Ox distance has been reduced by 0.43 Å. On the potential energy surface, the minimum energy path shows the shortening of the P-Ox and lengthening of the P-Oser203 is mostly sequential. the reactivation in this model is exothermic with an energy difference of -4.74 kcal.mol^{-1} and a barrier of 2.80 kcal.mol^{-1}.

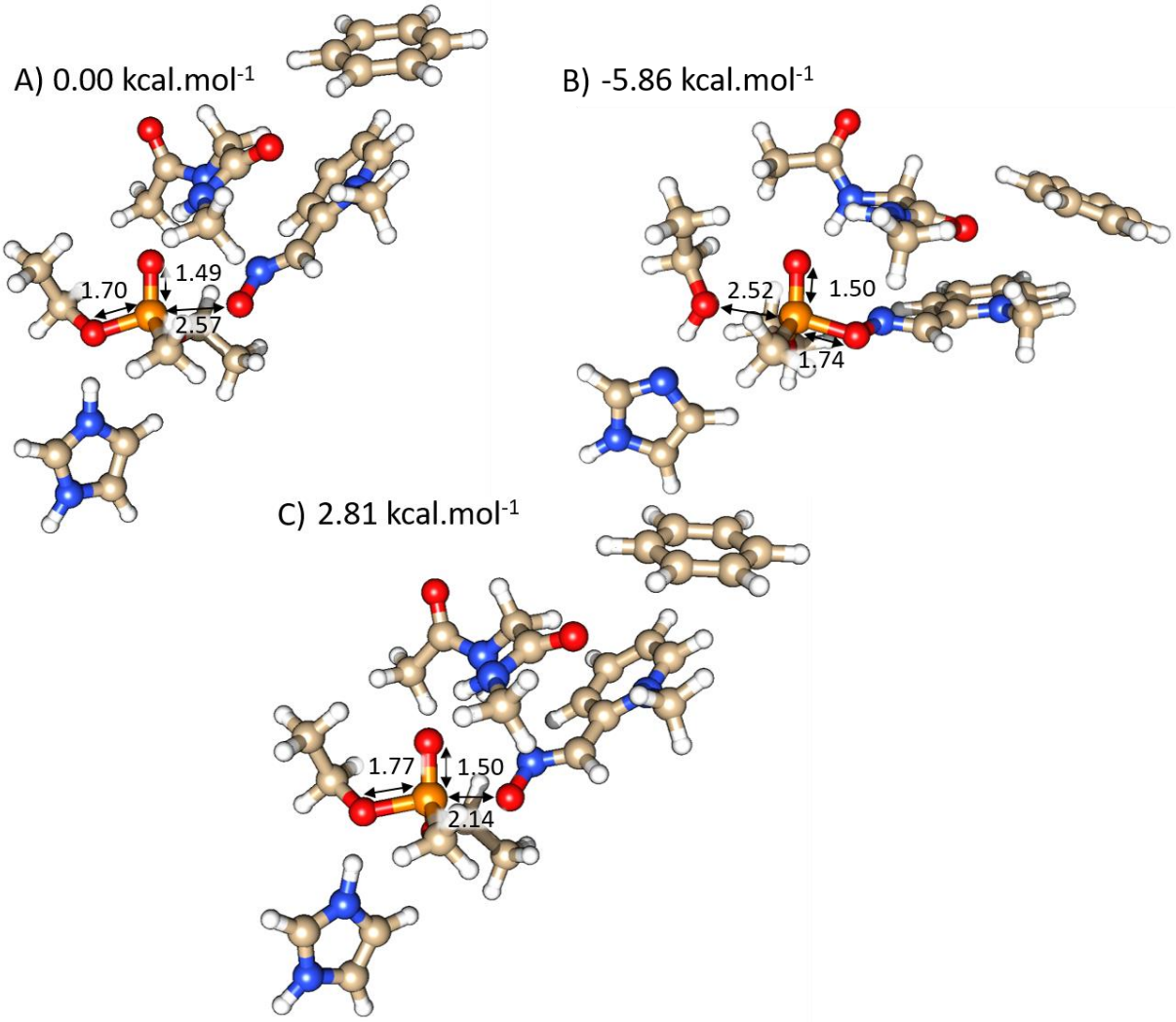


Figure V-4. Optimized structures of (A) the reactant, (B) the product, and (C) the transition state for the model with benzene and the oxyanionic hole.

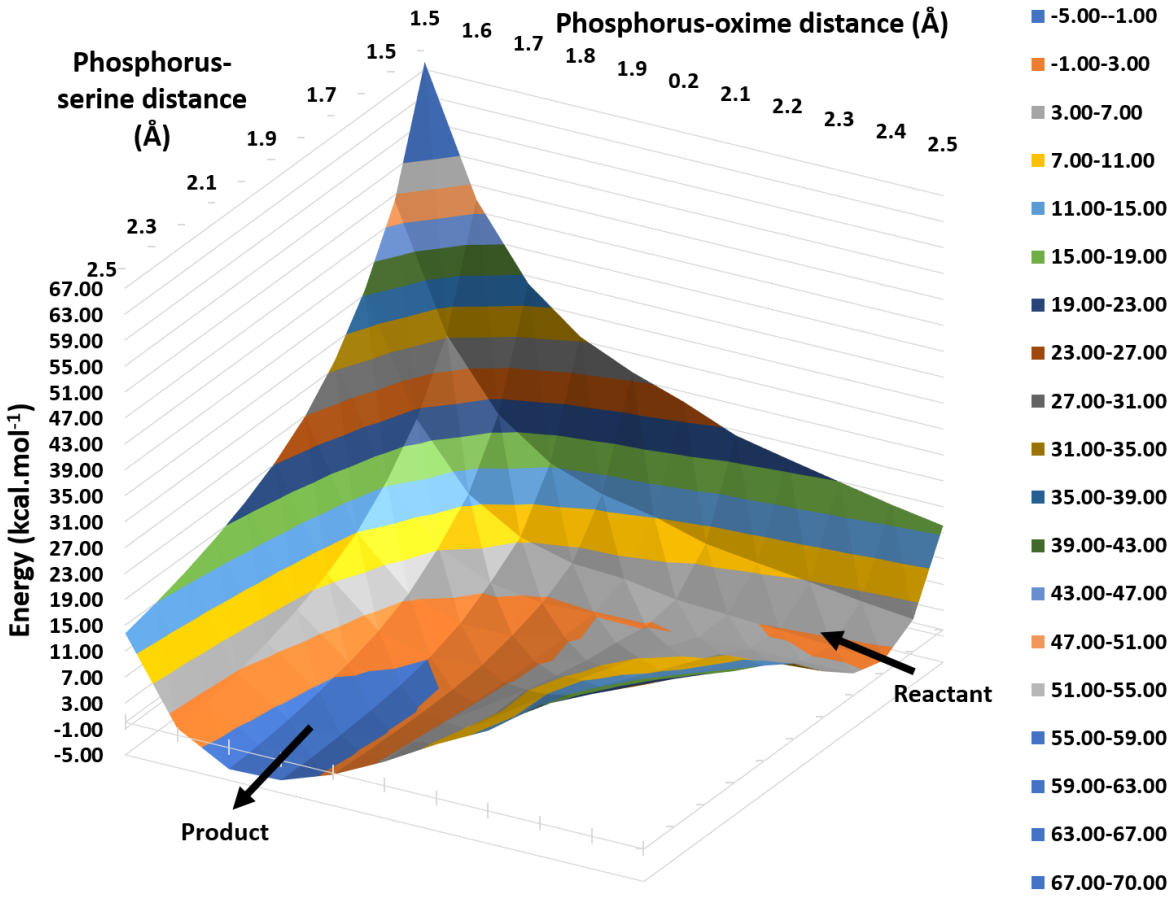


Figure V-5. Three-dimensional potential energy surface for the reactivation of the model with benzene and the oxyanionic hole of VX-inhibited AChE by 2-PAM.

Another possible addition to the truncated model is the acetate. With acetate in the model it was not possible to optimize a product for the reaction, only the reactant could be optimized by dropping constraints on a nearby scan point (**Figure V-6**). In this reactant, the P-Ox and P-OSer203 distances are 3.27 Å and 1.66 Å respectively. In the reactant structure the imidazolium has transferred a proton to the acetate and remains in imidazole form. As can be seen on the potential energy surface, any decrease in P-Ox or increase in P-OSer203 distances only increases the energy of the system (**Figure V-7**). Energetically, the only thing that can be said is that this reaction has a barrier at least higher than 37 kcal.mol^{-1}.

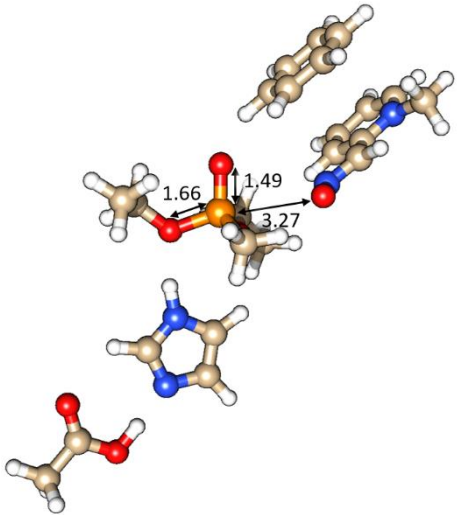


Figure V-6. Optimized structure of the reactant for the model with benzene and acetate.

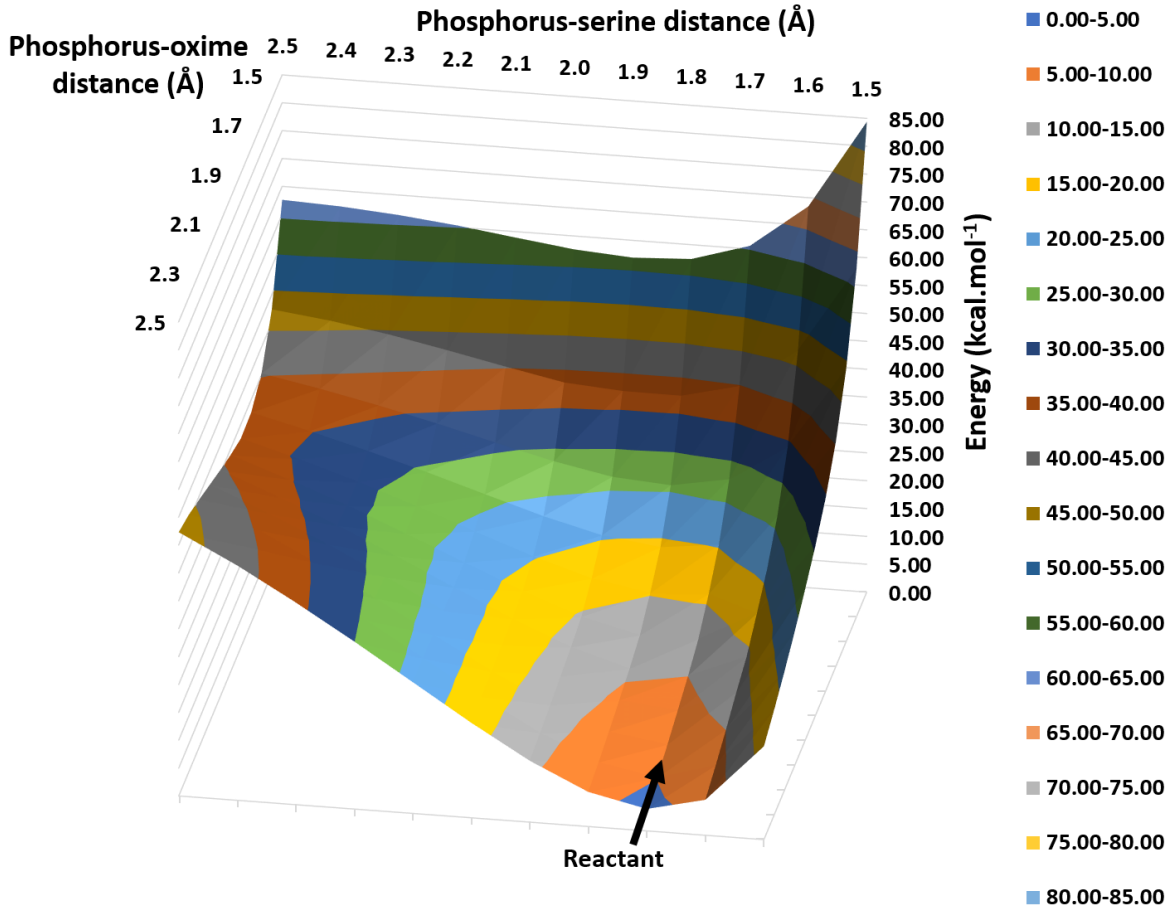


Figure V-7. Three-dimensional potential energy surface for the reactivation of the model with benzene and acetate of VX-inhibited AChE by 2-PAM.

In the complete model for benzene, the presence of the acetate still prevents the formation of a product by keeping the imidazole in its neutral form. It was not possible to obtain a structure for the product and only the reactant could be optimized (**Figure V-8**). Once again it was obtained by lifting constraints on a nearby scan point. In the reactant, the P-Ox and P-OSer203 distances are 1.65 Å and 3.02 Å respectively. On the potential energy surface

(**Figure V-9**) the reaction seems to be at an energetic dead-end with increases in energy with any change in reaction coordinate towards the product. Energetically, it is only possible to say that the energy barrier for the reaction is higher than 32 kcal.mol^{-1}.

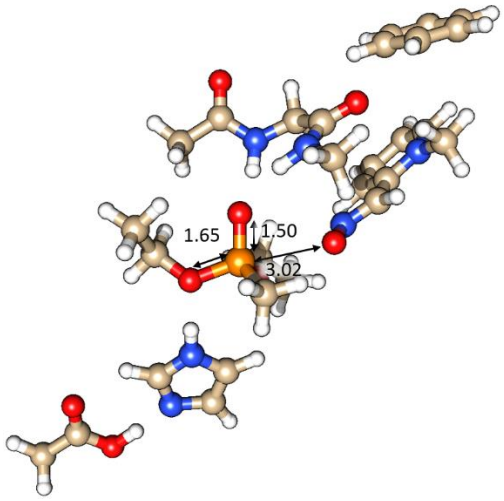


Figure V-8. Optimized structure of the reactant for the model with benzene, acetate, and the oxyanionic hole

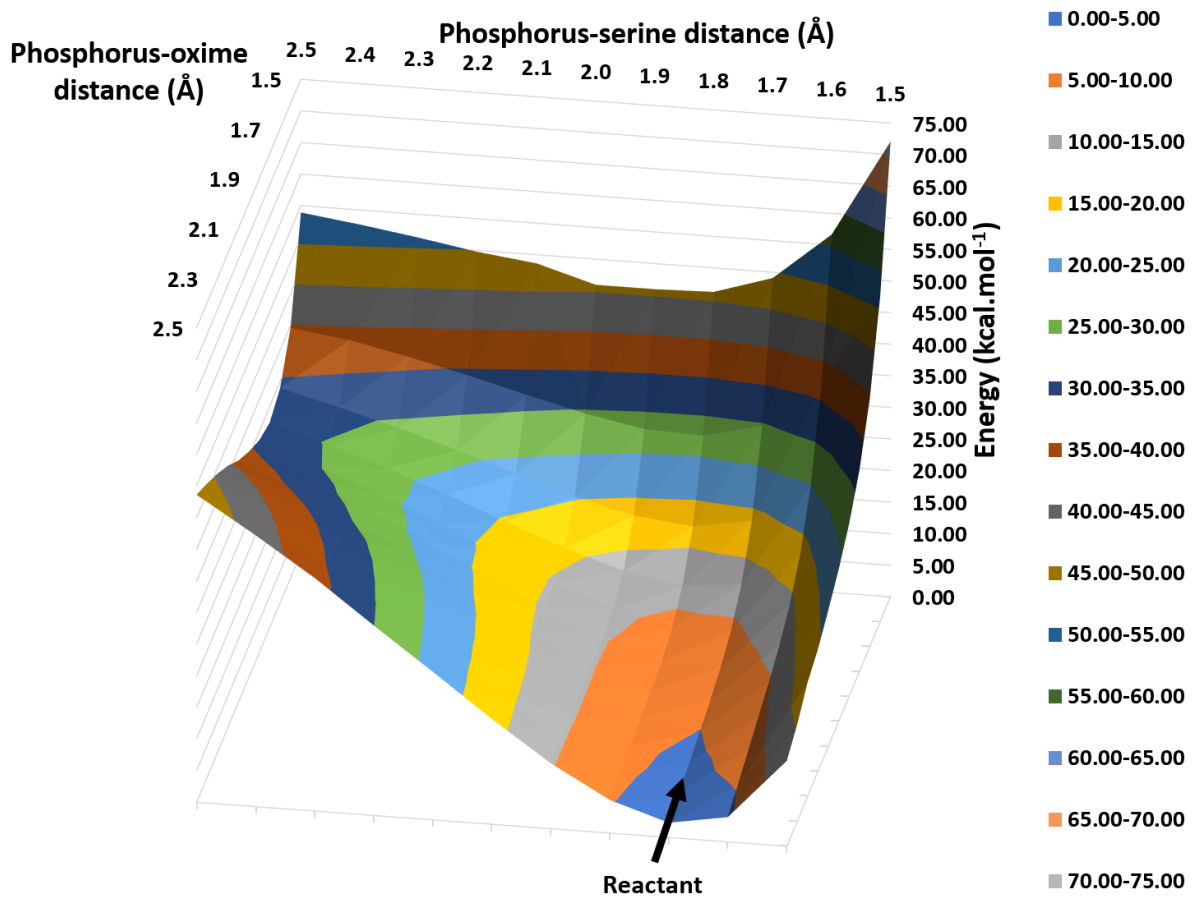


Figure V-9. Three-dimensional potential energy surface for the reactivation of the model with benzene, acetate, and the oxyanionic hole of VX-inhibited AChE by 2-PAM.

V.2. Phenol

The second aromatic ring whose influence on the reactivation has been studied is phenol. First it was added to the minimal model. Geometries for both the reactant and the product (**Figure V-10**) were obtained by releasing the constraints on scan points closest in energy to both local minima. In the reactant (**Figure V-10A**) the P-Ox and P-Oser203 distances are 2.75 Å and 1.71 Å respectively. In the product (**Figure V-10B**) those distances are 1.77 Å and 2.49 Å respectively. The minimum energy path that connect both structures can be followed on the potential energy surface (**Figure V-11**). Along this energy path there is a single saddle point that yielded the optimized transition state (**Figure V-10C**). It shows important structural similarities with the reactant with a P-OSer203 of 1.80 Å, only 0.09 Å longer. The main difference is a shortened P-Ox bond of 2.15 Å, 0.6 Å shorter than the reactant. The potential energy surface the variations of both reaction coordinates are neither concurrent nor completely sequential. The overall energetics of the reaction are an energy difference of -3.47 kcal.mol^{-1} and an energy barrier of 2.09 kcal.mol^{-1}.

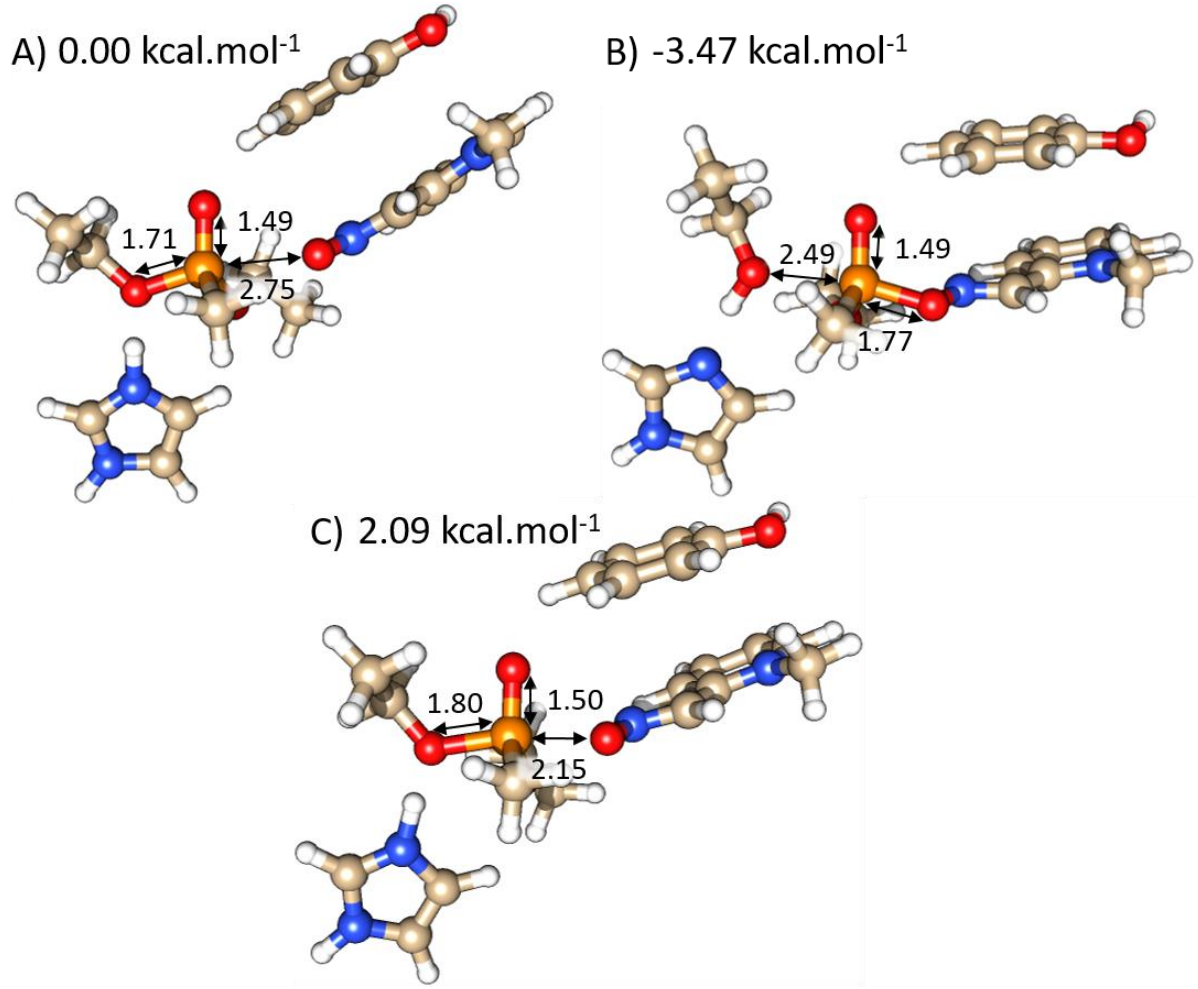


Figure V-10. Optimized structures of (A) the reactant, (B) the product, and (C) the transition state for the model with phenol.

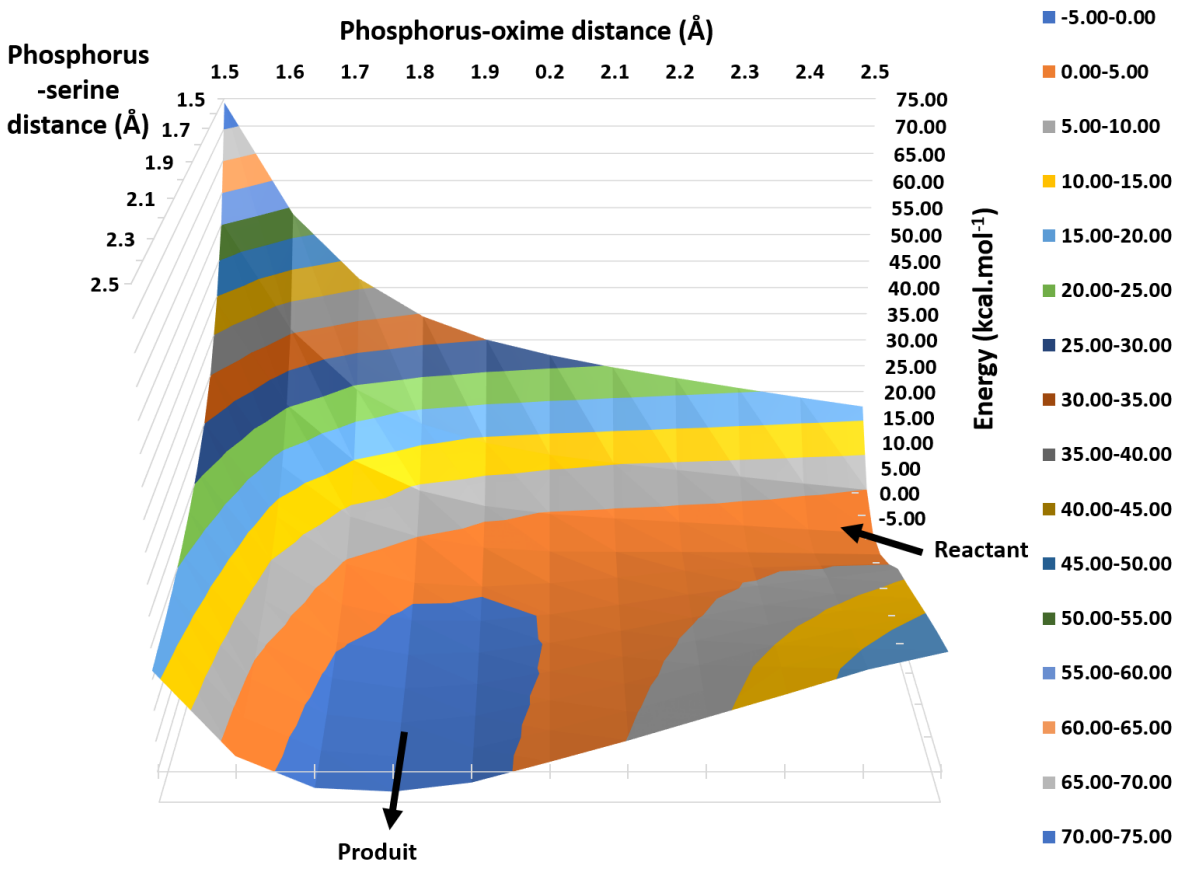


Figure V-11. Three-dimensional potential energy surface for the reactivation of the model with phenol of VX-inhibited AChE by 2-PAM.

The analogue of oxyanionic hole can also be introduced in the model alongside the phenol. The reactant and product structures presented in **Figure V-12** were obtained by optimizing selected scan points with released constraints. In the reactant (**Figure V-12A**), the P-Ox distance is 2.55 Å and the P-OSer203 distance is 1.70 Å. For the product (**Figure V-12B**), the P-Ox and P-OSer203 are 1.74 Å and 2.52 Å respectively. Between the reactant and the product, a minimum energy path can be traced on the potential energy surface (**Figure V-13**). A single saddle point has been detected on this minimum energy path. A nearby scan point was optimized into the transition state (**Figure V-12C**). The distances P-Ox and P-OSer203 are 2.39 Å and 1.73 Å respectively. This geometry is very reminiscent of the reactant with only a closer oxime by 0.16 Å. On the potential energy surface, the variation in P-OSer203 and P-Ox distances is almost completely sequential. There again the reaction is exothermic by 6.06 kcal.mol^{-1} with a reaction barrier of 1.36 kcal.mol^{-1}.

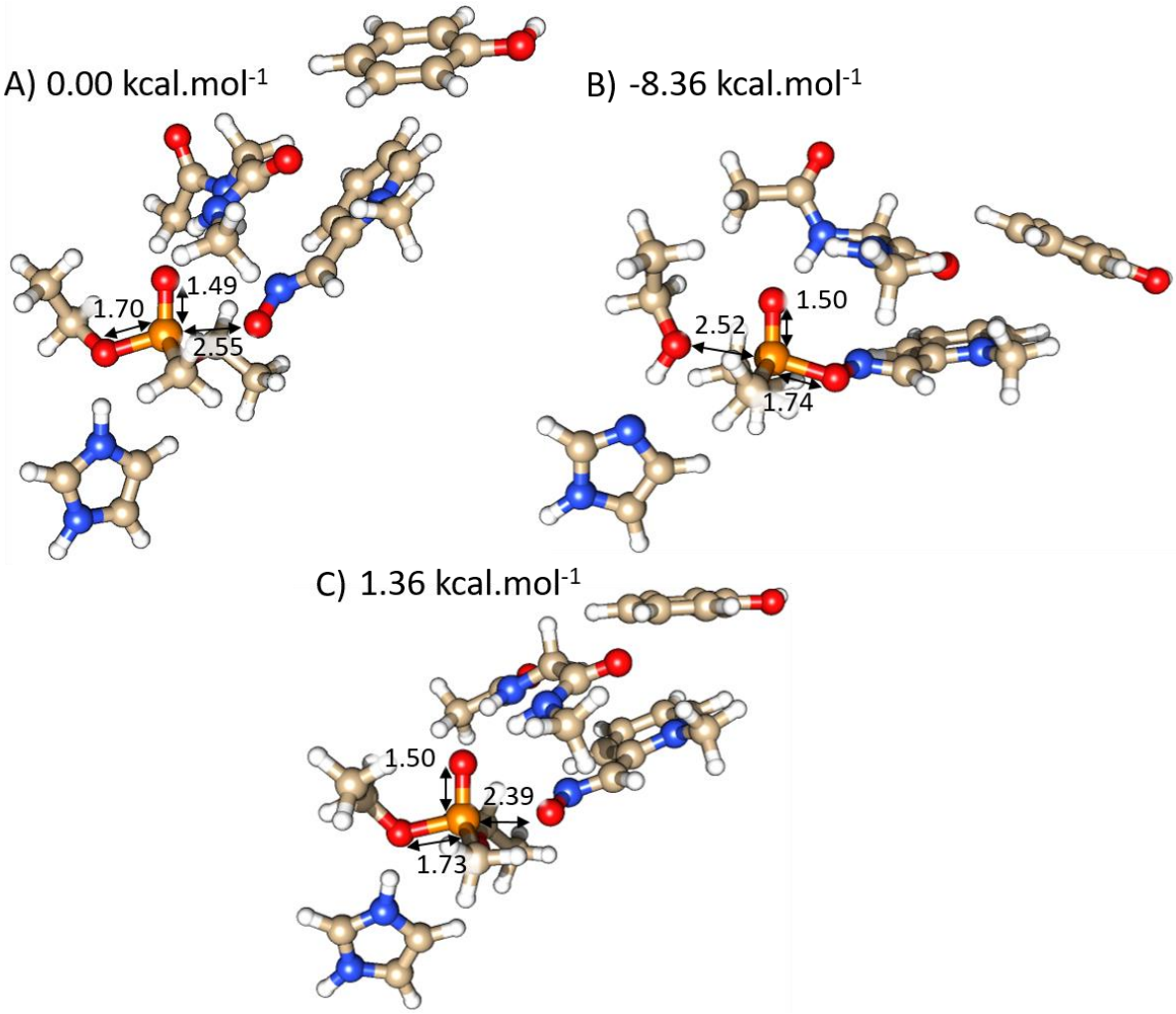


Figure V-12. Optimized structures of (A) the reactant, (B) the product, and (C) the transition state for the model with phenol and the oxyanionic hole.

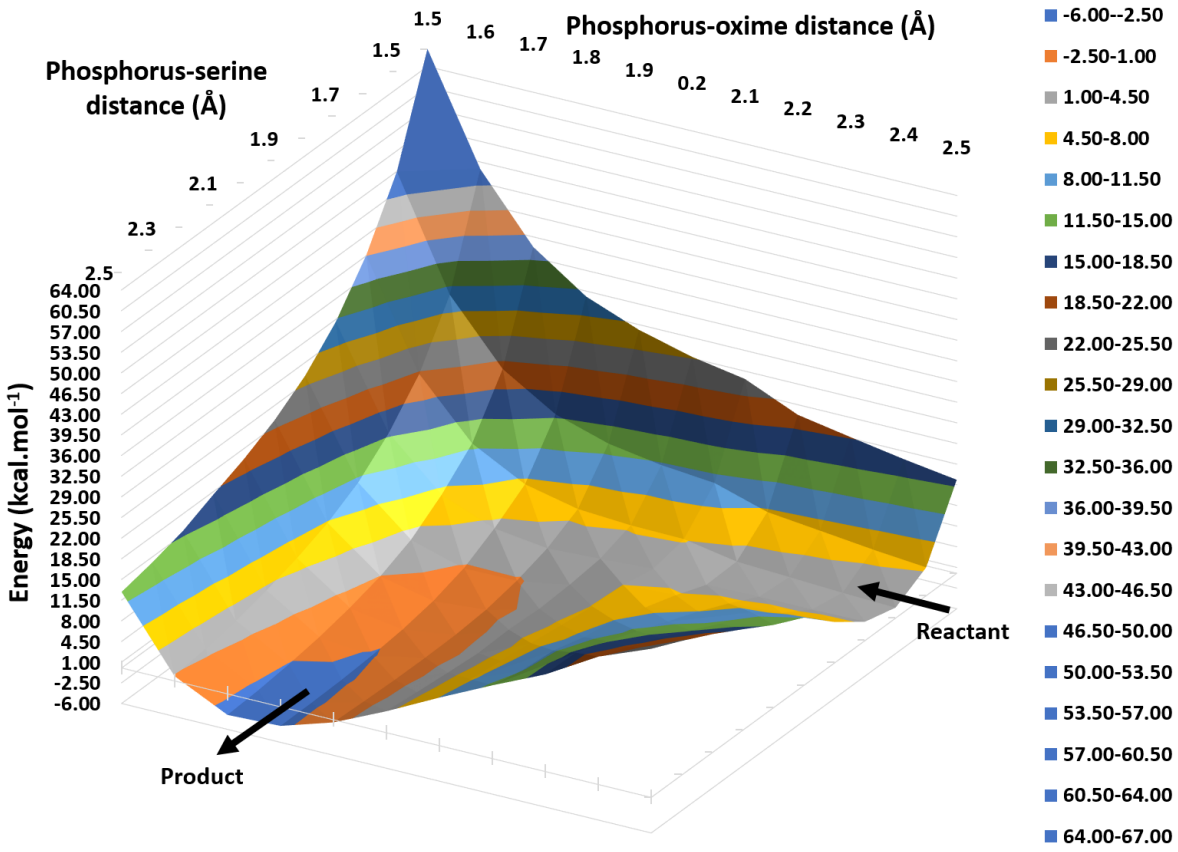


Figure V-13. Three-dimensional potential energy surface for the reactivation of the model with phenol and the oxyanionic hole of VX-inhibited AChE by 2-PAM.

Instead of the oxyanionic hole, the acetate can be added to the model. In this model with the acetate it was impossible to optimize a product for the reaction, only the reactant, by dropping the constraints on a nearby scan point (**Figure V-14**). In this reactant structure the imidazolium is deprotonated by the acetate. The P-Ox and P-OSer203 distances are 3.43 Å and 1.66 Å respectively. On the potential energy surface (**Figure V-15**), any change in the reactive distances towards the product increases the energy without. Energetically, the barrier for the reaction is at least higher than 36 kcal.mol^{-1}.

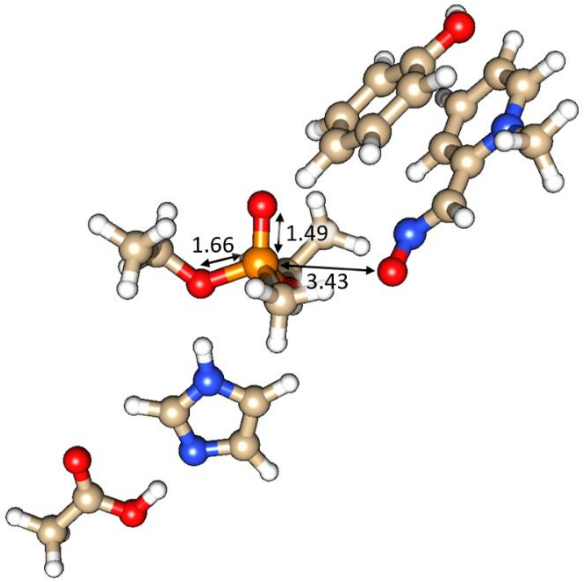


Figure V-14. Optimized structure of the reactant for the model with phenol and acetate.

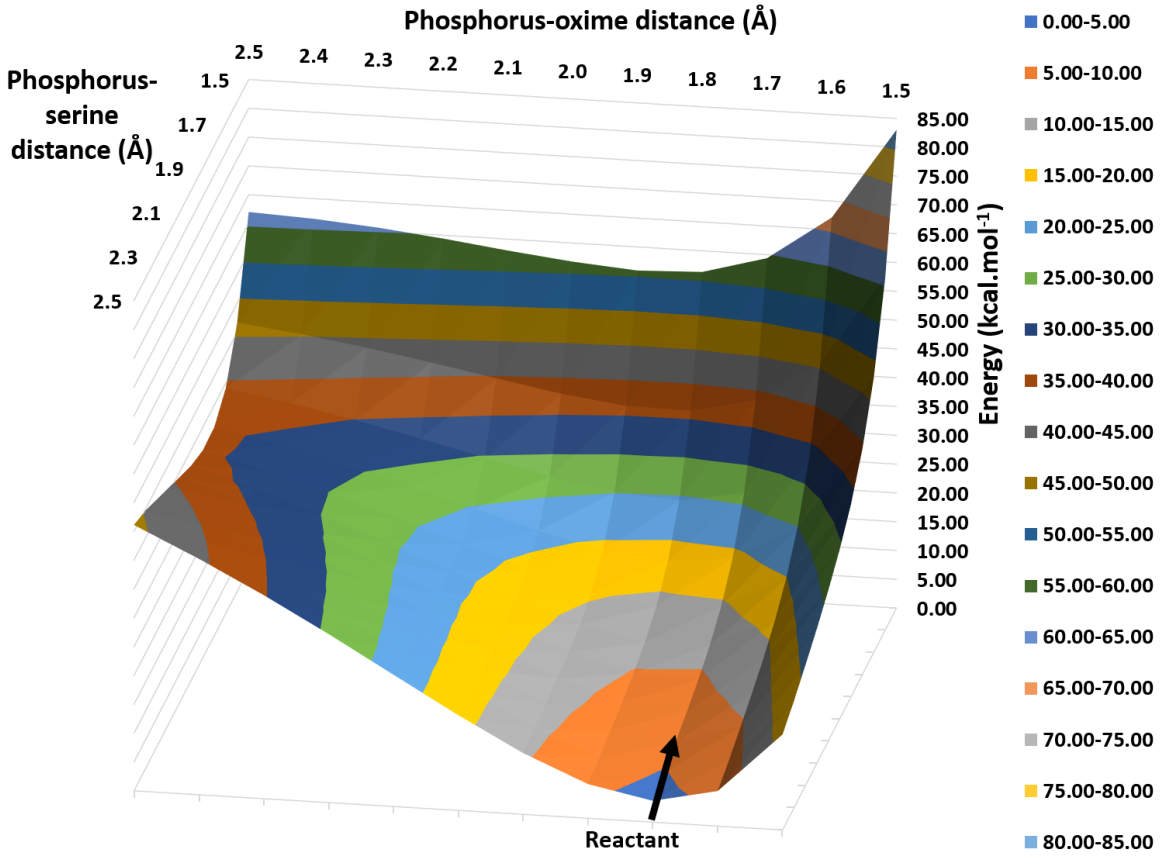


Figure V-15. Three-dimensional potential energy surface for the reactivation of the model with phenol and acetate of VX-inhibited AChE by 2-PAM.

The complete model for phenol includes both the acetate and the oxyanionic hole. The structure of a product could not be optimized in this model. The structure of the reactant (**Figure V-16**) comes from the optimization of a nearby scan point with released constraints. In this structure, the P-Ox distance is 3.04 Å and the P-OSer203 distance is 1.65 Å. The imidazole ring is in its neutral form. Any evolution of the reactive distances towards the

product increases steadily the energy as can be seen on the potential energy surface (**Figure V-17**). Energetically, the reaction has a barrier greater than 32 kcal.mol^{-1}.

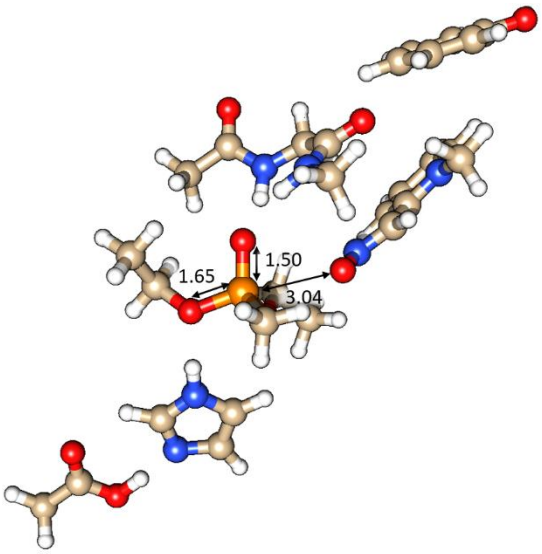


Figure V-16. Optimized structure of the reactant for the model with phenol, acetate, and the oxyanionic hole.

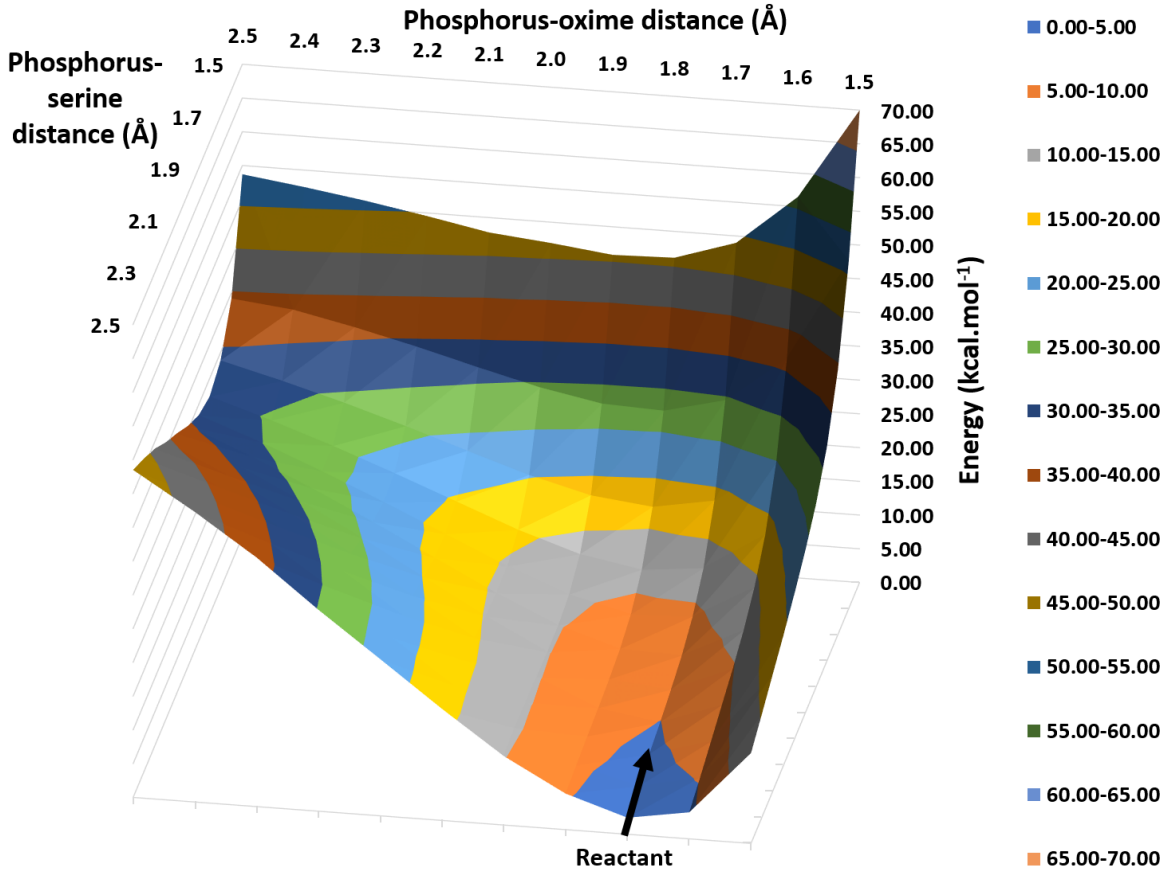


Figure V-17. Three-dimensional potential energy surface for the reactivation of the model with phenol, acetate, and the oxyanionic hole of VX-inhibited AChE by 2-PAM.

V.3. Indole

The third aromatic ring whose influence on the reactivation of VX-inhibited AChE by 2-PAM has been tested is the indole ring. First it was added to the minimal system. The optimized structures for this system are presented in **Figure V-18**. Both reactant and product were optimized by releasing constraints on relevant scan points. In the reactant (**Figure V-18A**), the P-Ox distance is 3.07 Å and the P-OSer203 distance is 1.69 Å. In the product (**Figure V-18B**), the P-Ox distance is 1.75 Å and the P-OSer203 distance is 2.63 Å. Those two structures can be connected by a minimum energy path on the potential energy surface along which there is a single saddle point. The localization of this saddle point allowed to optimize the structure of the transition state (**Figure V-18C**). In this transition state, the P-Ox is 2.07 Å and the P-OSer is 1.83 Å. On the potential energy surface (**Figure V-19**), the variations in reaction coordinate are neither sequential nor fully concurrent. The energetics of the reaction is a -4.03 kcal.mol^{-1} energy difference and a 4.25 kcal.mol^{-1} energy barrier.

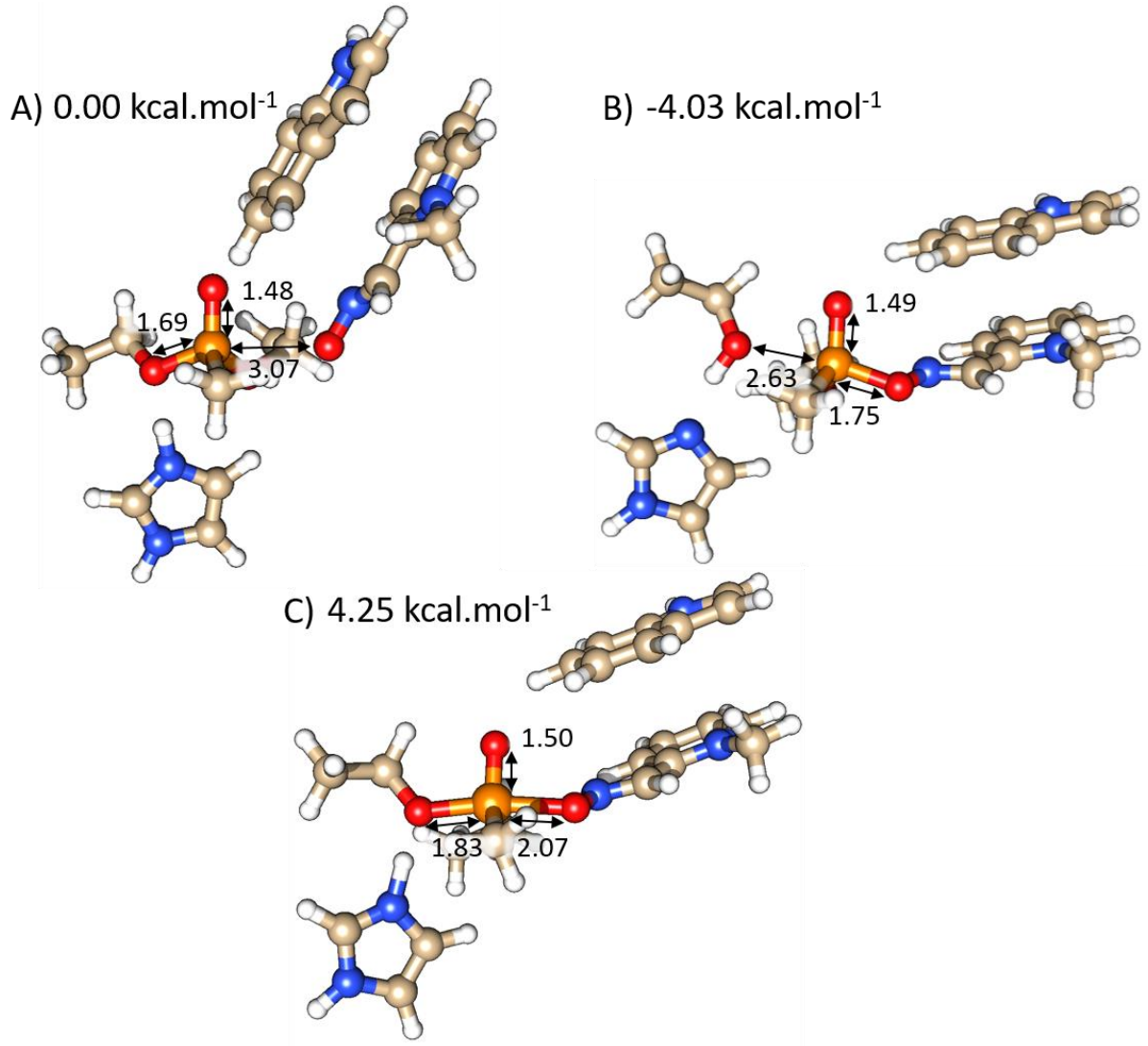


Figure V-18. Optimized structures of (A) the reactant, (B) the product, and (C) the transition state for the model with indole.

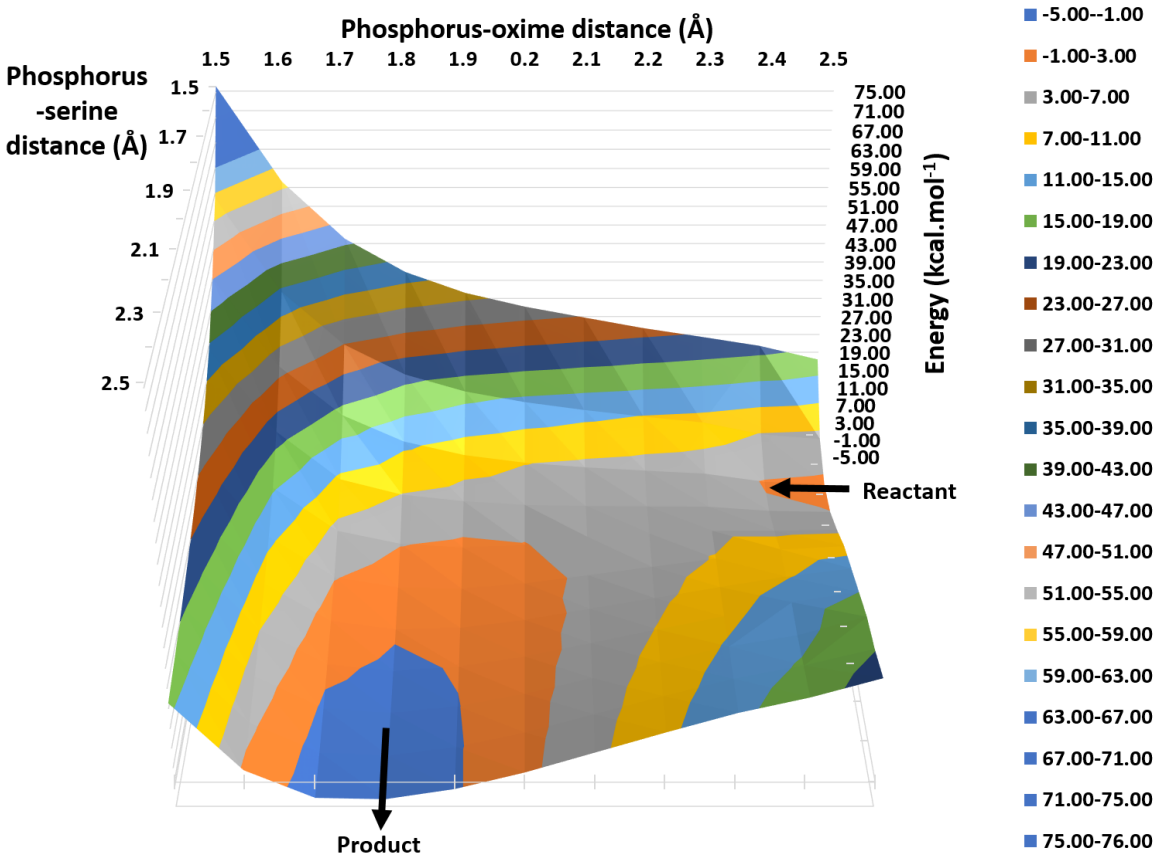


Figure V-19. Three-dimensional potential energy surface for the reactivation of the model with indole of VX-inhibited AChE by 2-PAM.

The model can be expanded by adding the oxyanionic hole. The reactant (**Figure V-20A**), obtained by releasing constraints on the scan point with P-Ox 2.50 Å and P-OSer203 1.70 Å, has a P-Ox distance of 2.62 Å and a P-OSer203 of 1.70 Å. The product (**Figure V-20B**), obtained by releasing constraints on the scan point with P-Ox 1.80 Å and P-OSer203 2.50 Å, has a P-Ox distance of 1.73 Å and a P-OSer203 of 2.50 Å. The reactant and product can be connected by a minimum energy path on the potential energy surface (**Figure V-21**). Following this minimum energy path a single saddle point corresponding to the optimized transition state (**Figure V-20C**) can be found. This transition state has a geometry with a P-Ox of 2.33 Å and a P-OSer203 of 1.74 Å. On the potential energy surface the variations in reaction coordinate distances are almost completely sequential. The reaction in this model is exothermic with an energy difference of -9.52 kcal.mol^{-1} and a reaction barrier of 0.34 kcal.mol^{-1}.

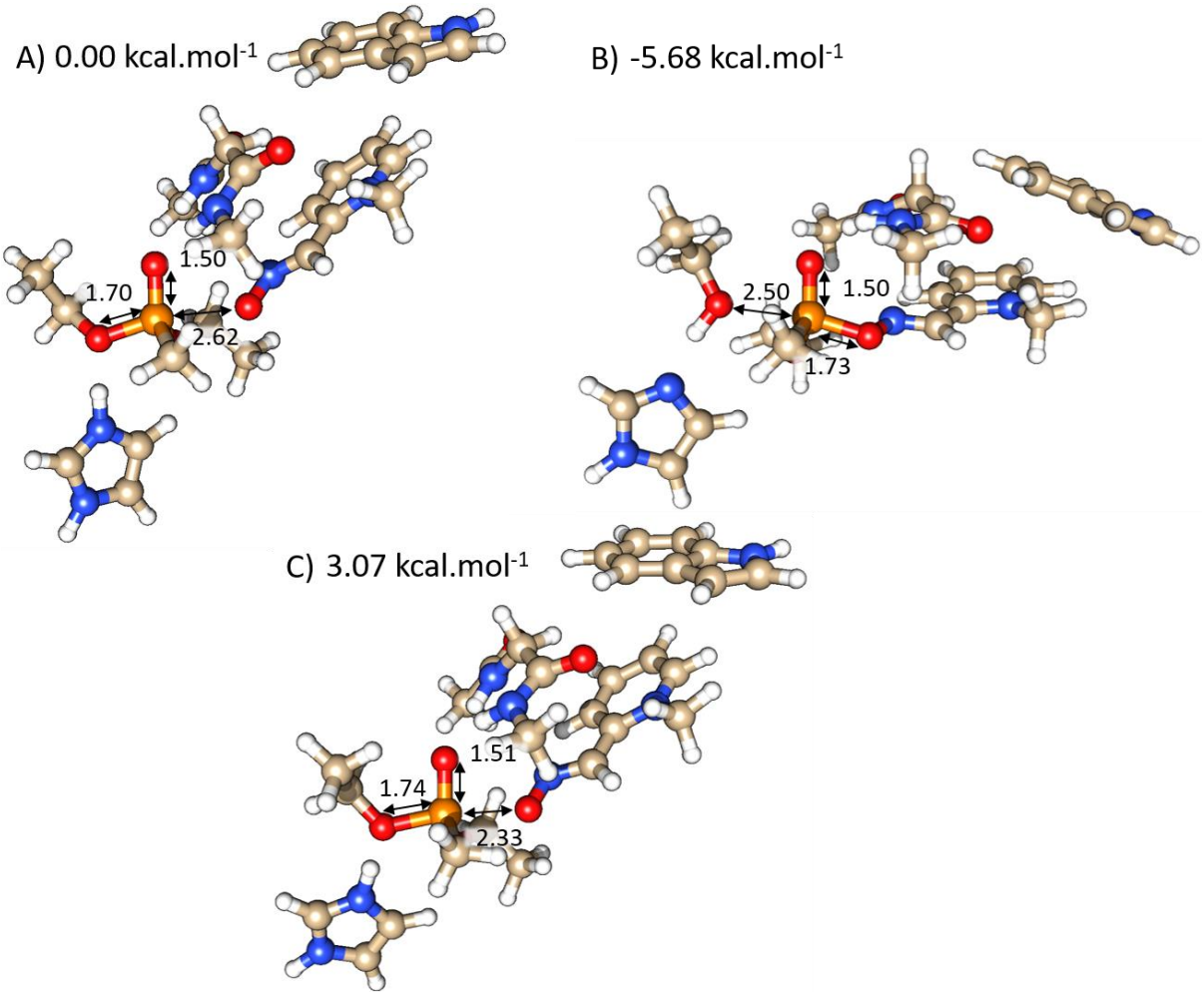


Figure V-20. Optimized structures of (A) the reactant, (B) the product, and (C) the transition state for the model with indole and the oxyanionic hole.

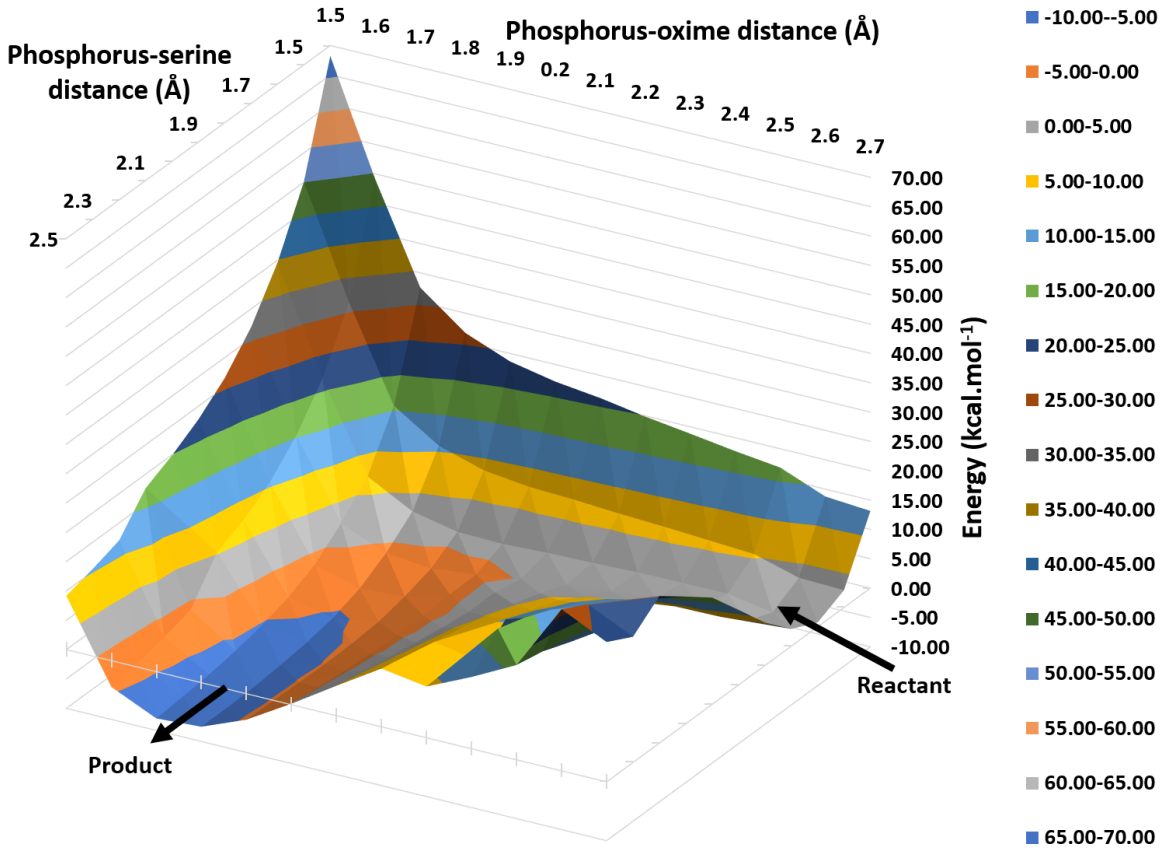


Figure V-21. Three-dimensional potential energy surface for the reactivation of the model with indole and the oxyanionic hole of VX-inhibited AChE by 2-PAM.

The model for indole can also be supplemented by adding an acetate. In the structure of the reactant, the imidazolium is deprotonated by the acetate (**Figure V-22**). In this structure, the P-Ox is 3.19 Å long and the P-OSer203 is 1.66 Å long. The product could not be optimized in this model. The potential energy surface (**Figure V-23**) shows an energetic dead-end for the reaction. The energy of the reaction barrier could not be precisely calculated but it's possible to say that it is at least greater than 34 kcal.mol^{-1}.

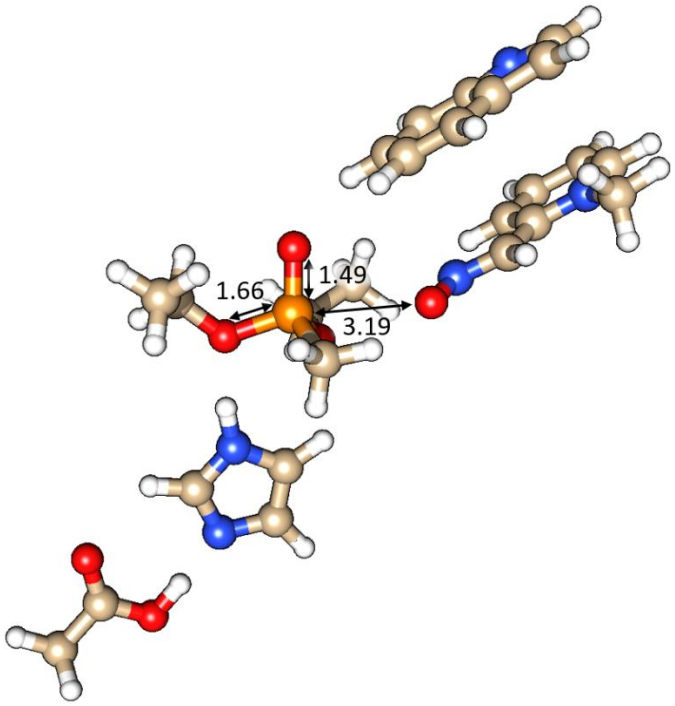


Figure V-22. Optimized structure of the reactant for the model with indole and acetate.

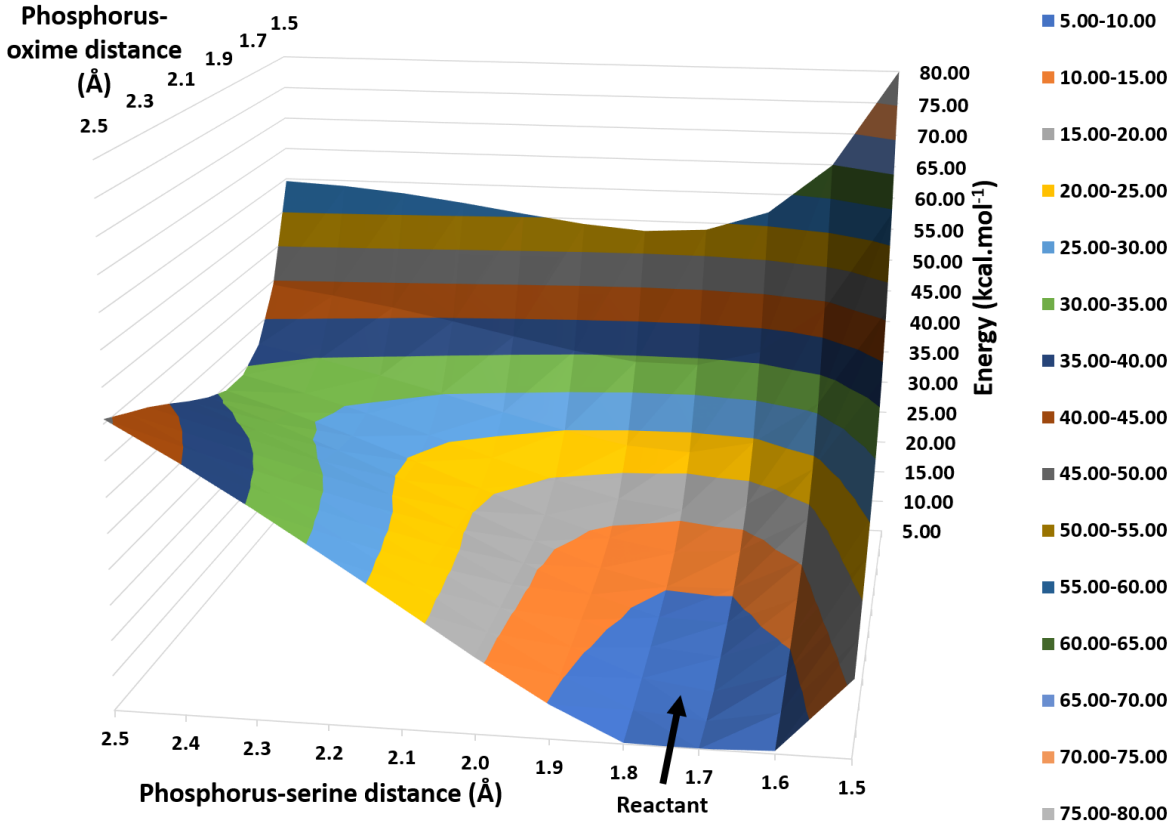


Figure V-23. Three-dimensional potential energy surface for the reactivation of the model with indole and acetate of VX-inhibited AChE by 2-PAM.

The complete model for indole includes both the acetate and the oxyanionic hole. Again, the imidazolium is deprotonated by the acetate in the structure of the reactant which is the only one that could be optimized (**Figure V-24**). In this structure, the P-Ox is 3.21 Å long

and the P-OSer203 is 1.65 Å long. The potential energy surface (**Figure V-25**) shows no product formation or energy barrier crossing and the only conclusion from an energetics standpoint is that the reaction barrier exceeds 31 kcal.mol^{-1}.

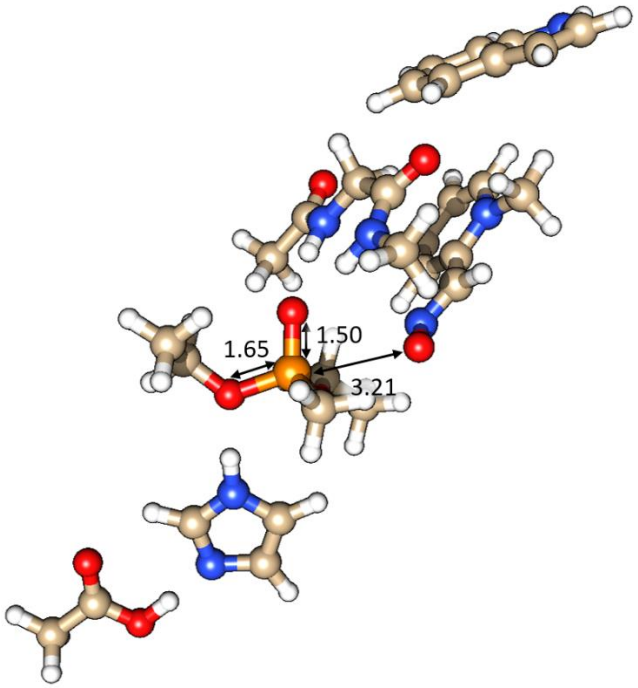


Figure V-24. Optimized structure of the reactant for the model with indole, acetate, and the oxyanionic hole.

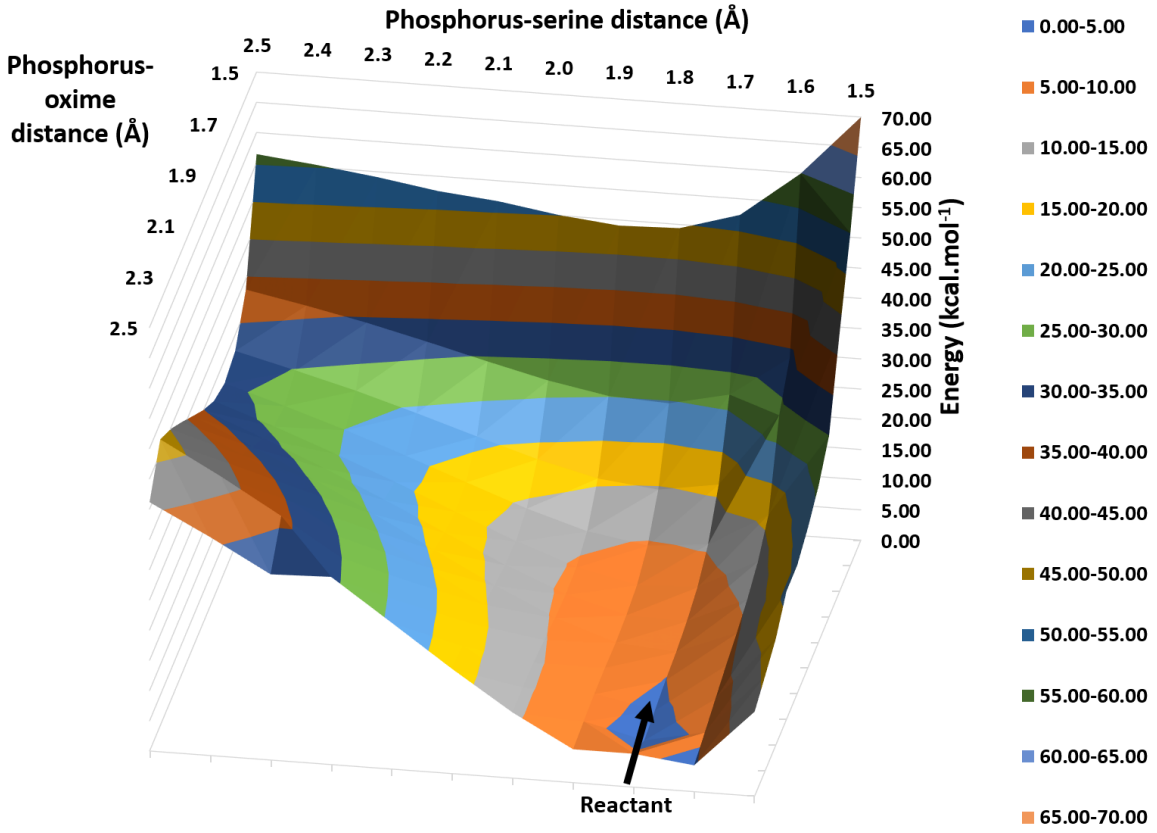


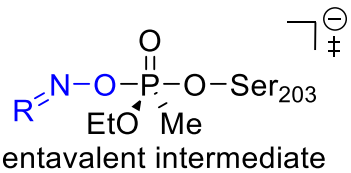
Figure V-25. Three-dimensional potential energy surface for the reactivation of the model with indole, acetate, and the oxyanionic hole of VX-inhibited AChE by 2-PAM.

Now that the results for the aromatic rings have been detailed, they will be analysed and contextualized with the other results in the next section.

VI. Discussion

Now that the results obtained for all 16 truncated models of AChE's active site have been presented, they can be discussed. The first element of note is that the inclusion of the acetate in the model systematically prevented the reaction. It was only possible to optimize a product structure for the minimal model but no transition state could be optimized. In those models, which mimics Glu334, the reactivation of VX-inhibited AChE with 2-PAM appears to be very endothermic with energy barriers at least superior to 30 kcal.mol^{-1}. This is surprising as imidazole has a pK_a of 7.0, higher than the pK_a of 4.8 of the acetate. This difference in pK_a should mean that the imidazolium should keep the proton over the acetate. However, classical pK_a values are measured in bulk solvent, and considering that the imidazolium forms a hydrogen bond with the VX-Ser203 adduct model, the acidity of the imidazolium might be increased over that of acetic acid. When the implicit solvent model COSMO was used for the minimal model with the acetate mimicking Glu334, the reaction was able to take place. A product was optimized despite the large endothermicity. It could explain the preventing of the reaction by the acetate in the gas phase as being due to a lack of compensation of the charge of the acetate. In the forthcoming QM/MM simulations, the acetate will be naturally compensated by a well described enzymatic environment.

For all the models that did not include an acetate the reaction could be simulated. In all eight cases where the acetate was not included, the reaction has a single transition state and no stable pentavalent intermediate (see **Scheme VI-1** and **Figure I-3**). Thus, in all those steps the mechanism is an SN_2. This was unexpected as all other investigations of AChE's reactivation using a similar truncated model methodology indicated an addition-elimination mechanism.[7] When the reactivation was modelled in the minimal model in water solvent implicit solvation using COSMO, the reactivation still has a SN_2 mechanism but with an addition elimination character. The other truncated model investigations of AChE's reactivation were made with tabun-inhibited AChE instead of VX-inhibited AChE. [7] The steric hindrance caused by the N-dimethyl substituent of tabun might explain the addition-elimination mechanism.



Pentavalent intermediate

Scheme VI-1. Pentavalent intermediate for the addition-elimination mechanism of VX-inhibited AChE reactivation by an oxime.

[7] a) R. Lo, B. Ganguly, *Mol. BioSyst.* **2014**, *10*, 2368–2383. b) R. Lo, N. B. Chandar, M. K. Kesharwani, A. Jain, B. Ganguly, *PLoS One* **2013**, *8*, e79591.c) J. Wang, J. Gu, J. Leszczynski, M. Feliks, W. A. Sokalski, *J. Phys. Chem. B* **2007**, *111*, 2404–2408.

In **Figure VI-1** the values for both key distances P-Ox and P-Oser203 of the reactants, products, and transition states are presented for the different truncated models. What is apparent is that the transition state is systematically closer in structure to the reactant than to the product. This early transition state is generally (to the exception of benzene) even earlier for models including the oxyanionic hole. The formation of the transition state from the reactant involves very little change in the P-OSer203 distance forming an almost straight line in **Figure VI-1**. In that respect, it evokes what could be the addition transition state for a hypothetic addition elimination mechanism.

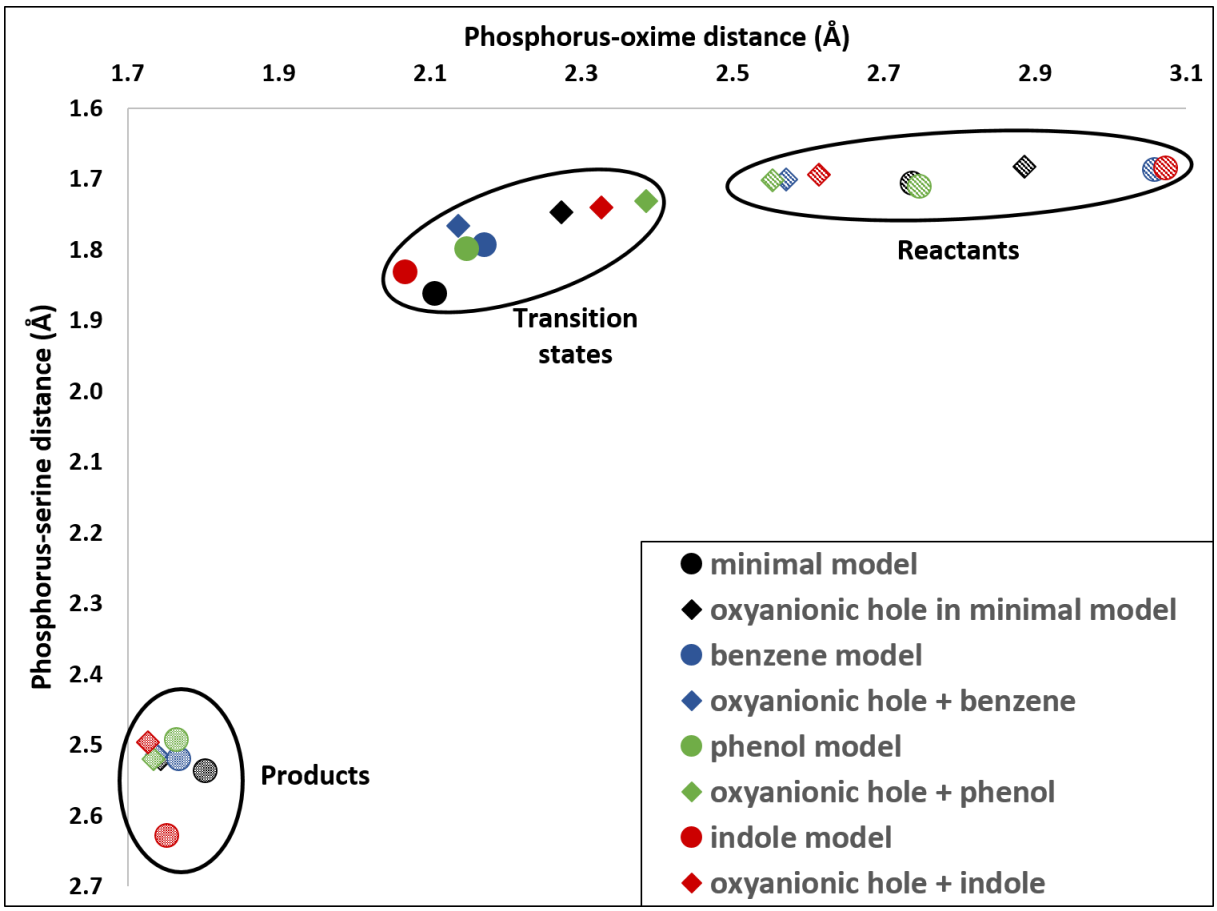


Figure VI-1. Reactive distances for all the optimized reactant product and transition state structures for all models that did not include acetate. Models including the oxyanionic hole are indicated by diamonds and models without the oxyanionic hole are indicated by circles. For reactants, the shapes are filled by bands, for the transition states the filling is plain, and for the products, the filling is dotted.

If the mechanism truly is an addition-elimination, a pentavalent intermediate should be found. It is not the case according to the calculations presented in this chapter. It could mean that the stabilization brought by the oxyanionic hole to the product state is greater than the stabilization it brings to the reactant state. It can be viewed as a case of a dissymmetric three state VB configuration mixing diagram where the products state rests lower in energy compared to the pentavalent state (**Figure VI-2**).[8]

[8] S. Shaik, A. Shurki, *Angew. Chem. Int. Ed.* **1999**, *38*, 586–625.

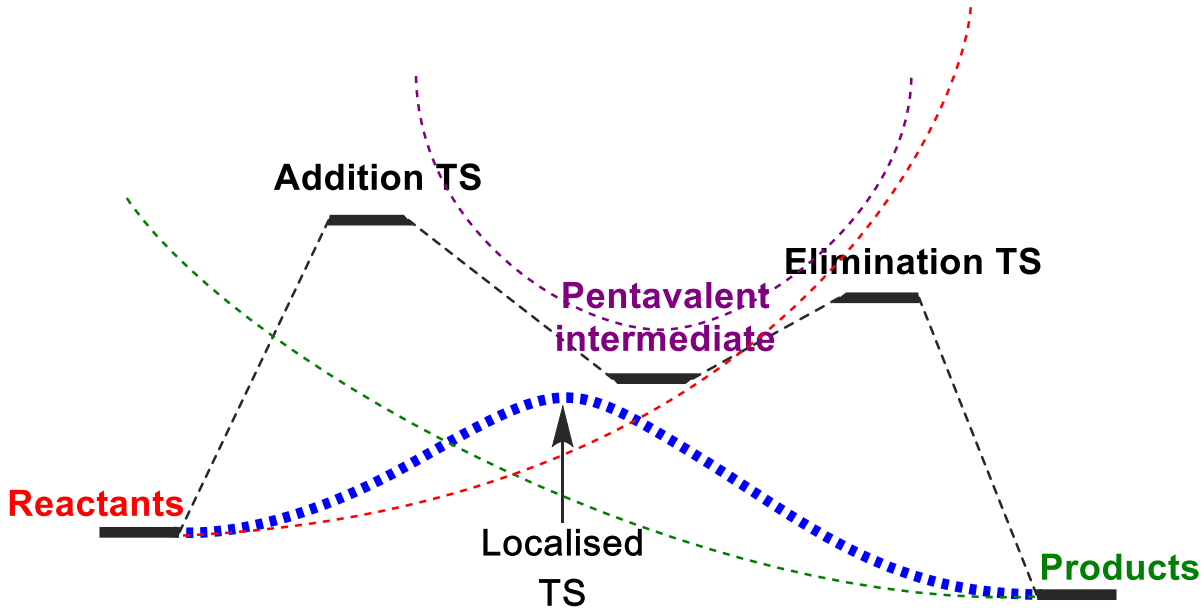


Figure VI-2. VB configuration mixing diagram from the reactivation.

The simulations confirm the role of the oxyanionic hole in the process of reactivation in the stabilisation of transition states.[9] As can be seen in **Table VI-1**, the inclusion of the oxyanionic hole systematically reduces the energy of the transition state compared to similar models without the oxyanionic hole. The hydrogen bonds the oxyanionic hole forms with the oxygen involved in the phosphoryl bond reduce the energetic cost of over-polarizing the phosphoryl bond. The energy difference of the reactivation is also decreased by the oxyanionic hole in all four cases. The oxyanionic hole appears to stabilize the product more than it does the reactant. In the models including the oxyanionic hole, the reactivation is always exothermic (see **Figure VI-3**).

Table VI-1. Changes in reaction barrier (ΔE*) and energy difference (ΔΔE) as a result of the inclusion of the oxyanionic hole in various models. All energies in kcal.mol^{-1}.

	ΔE* **(kcal.mol^{-1})**	**ΔΔE** **(kcal.mol^{-1})**
Minimal model	-2.93	-4.66
Model with benzene	-1.26	-4.32
Model with phenol	-0.73	-4.89
Model with indole	-1.18	-1.65

[9] a) N. Qian, I. M. Kovach, *FEBS Lett.* **1993**, *336*, 263–266. b) Y. Li, L. Du, Y. Hu, X. Sun, J. Hu, *Can. J. Chem.* **2012**, *90*, 376–383.

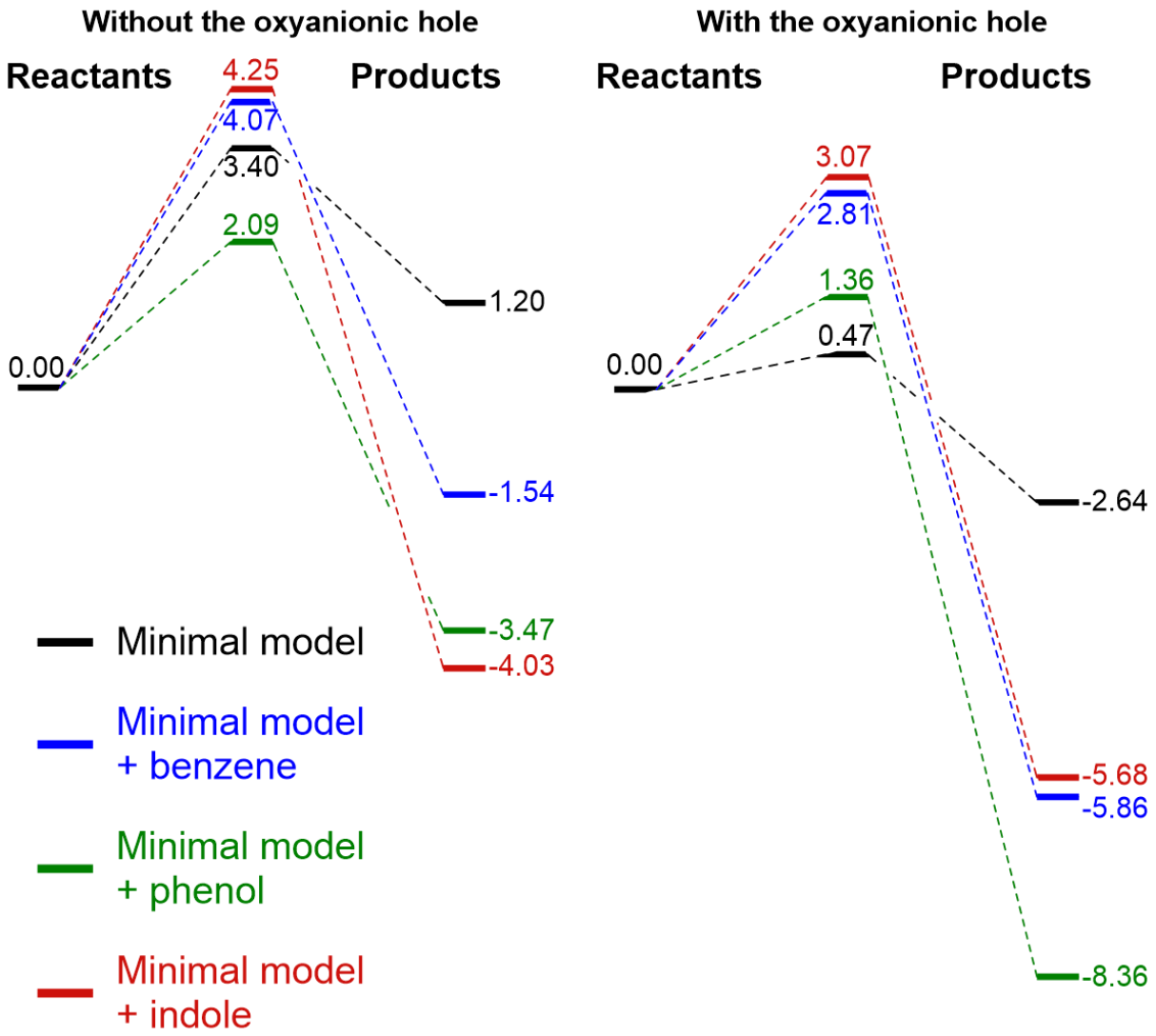


Figure VI-3. Reaction profiles for the reactivation in various truncated models. Energies were obtained with B3LYP/def2-SV(P) and are in kcal.mol^{-1}.

Figure VI-3 can be used as a basis for the discussion of the influence of the aromatic rings in the reactivation. Their inclusion in the model, with or without the oxyanionic hole systematically stabilizes the product more than the reactant. The increase in exothermic character of the reactivation when the oxyanionic hole is not identical for all aromatic rings. The exothermicity is increased by 4.89 kcal.mol^{-1} for phenol but only by 1.65 kcal.mol^{-1} for indole. This difference might be due to interactions between the aromatic rings and the oxyanionic hole that modify the way the oxyanionic hole interacts with the 2-PAM-VX and Ser203-VX adducts. The effect of aromatic rings on the energy barrier is inconsistent as benzene and indole increase the energy barrier while phenol decreases it. It could be concluded that phenol has a stabilizing interaction with 2-PAM but all activation energies for aromatic rings are within 3 kcal.mol^{-1} which falls in the limits of the truncated model methodology.

VII. Conclusions

The research presented in this chapter highlights the importance of the oxyanionic hole in stabilizing transition states of the reactivation. It also demonstrates that Glu334 on its own as a very negative impact on reactivation, completely preventing its occurrence. It is probable that in the enzyme, nearby residues mitigate the basicity of this glutamate to prevent in from blocking the reactivation because the reactivation process has been evidenced experimentally. Finally, those simulations show that the aromatic rings in AChE's active site seem to have a marginal influence in the reactivation process at best. This qualitative information will be valuable when tackling the full enzymatic system with QM/MM.

Chapter 4: QM/MM Simulations of Acetylcholinesterase Reactivation by 2-PAM

Chapter 4: QM/MM Simulations of Acetylcholinesterase Reactivation by 2-PAM

I. Computational details

To setup our QM/MM simulations of AChE it is necessary, as explained in Chapter 2, to choose an existing X-ray structure. The chosen structure is the PDB structure 3DL7, obtained and described by Carletti *et al*.[1] This structure has the advantage of being phosphorylated by aged tabun instead of being inhibitor free. It also has a good resolution of 2.5 Å. Finally, since this was structure used in the doctoral work of Ophélie Kwasnieski, it was initially thought that the new results could more easily be compared to those obtained in her *PhD* project if we kept using the same structure.[2]

The chosen PDB structure contains an AChE dimer with some ligands used to crystallize AChE. Those ligands are hexaethylene glycol, tetraethylene glycol, N-acetyl-D-glucosamine, and a chloride ion (see **Figure I-1A**). The ligands as well as one of the enzyme dimer were deleted from the structure to obtain the structure presented in **Figure I-1B**. This structure contains the enzyme chain A, the chloride ion and all 256 crystalline water molecules present in the X-ray data (**Figure I-2**).

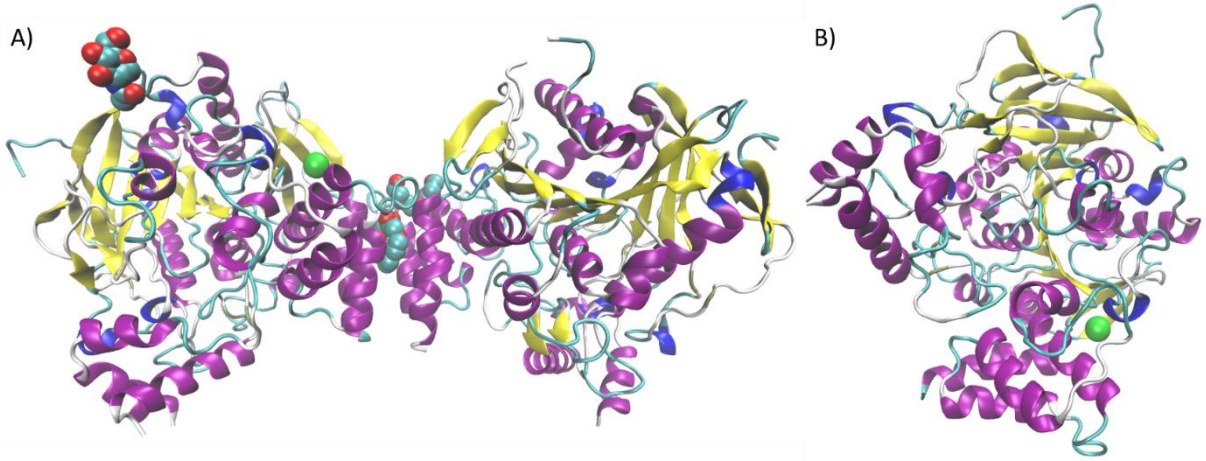


Figure I-1. A) PDB structure 3DL7 with both chains of the AChE dimer in ribbon representation and secondary structure colouring and the ligands in van der Waals radius representation. B) Chain A of 3DL7 with the chloride ion.

[1] E. Carletti, J.-P. Colletier, F. Dupeux, M. Trovaslet, P. Masson, F. Nachon, *J. Med. Chem.* **2010**, *53*, 4002–4008.

[2] O. Kwasnieski, Etude Théorique de La Réactivation de l'AChE Inhibée Par Le Tabun, UPMC, **2010**.

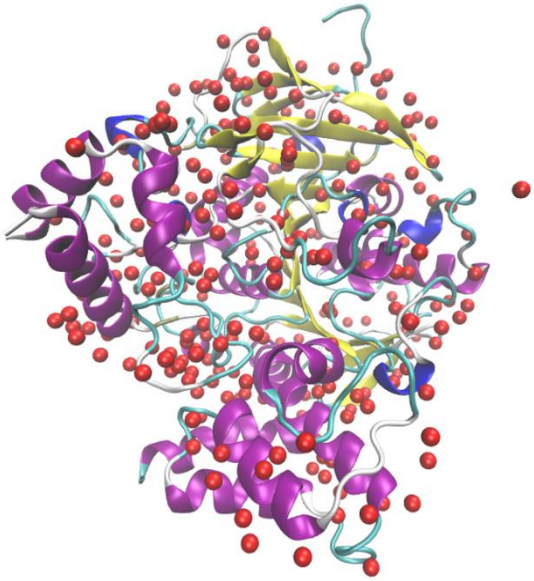


Figure I-2. Chain A of AChE structure 3DL7 with the crystal water molecules represented as red spheres.

In 3DL7 the X-ray data is inconclusive on the structure of Arg522, His447. The B chain of Arg522 was chosen over the A chain because the positive charge it carries appears to be better compensated by its environment (**Figure I-3**). Both A and B chains are exposed to the solvent but the chain B is located near a glutamine residue that create the potential for an hydrogen bond which is absent from chain A. For His447, the A chain was chosen because it is in a reactive position as opposed to the B chain where the hydrogen bond between His447 and Ser203 is broken, disrupting the triad (**Figure I-4**).

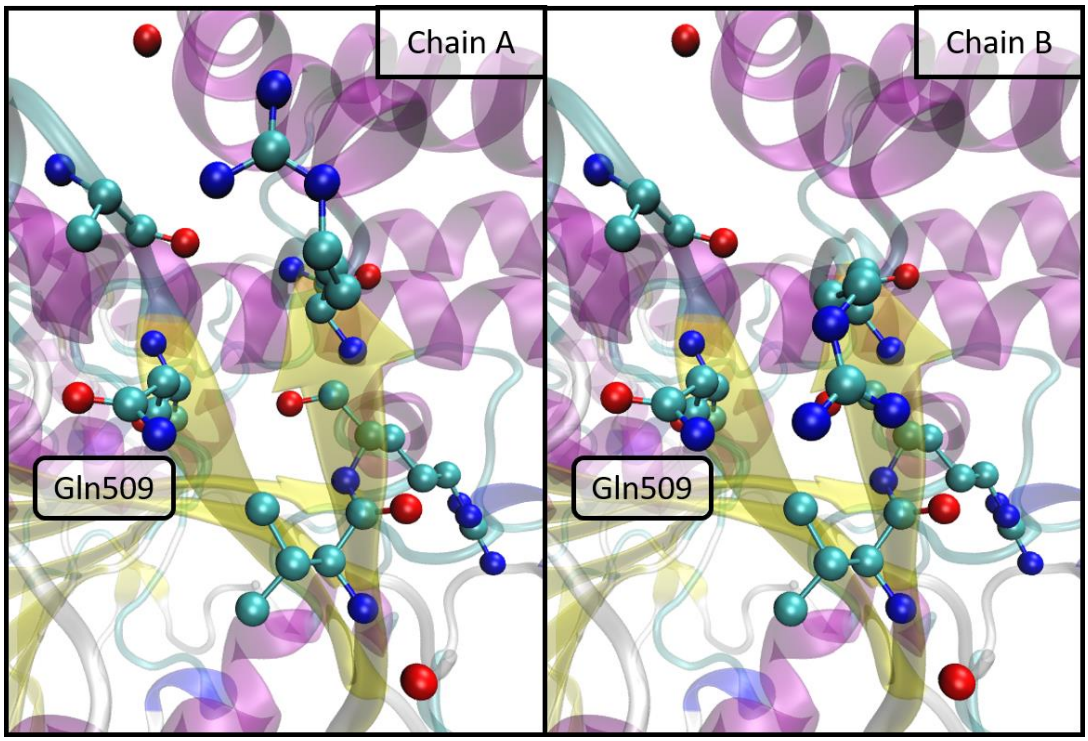


Figure I-3. A and B chains of Arg522 with some nearby residues highlighted in Ball&Stick representation and regular chemical element colouring (carbons in light blue)

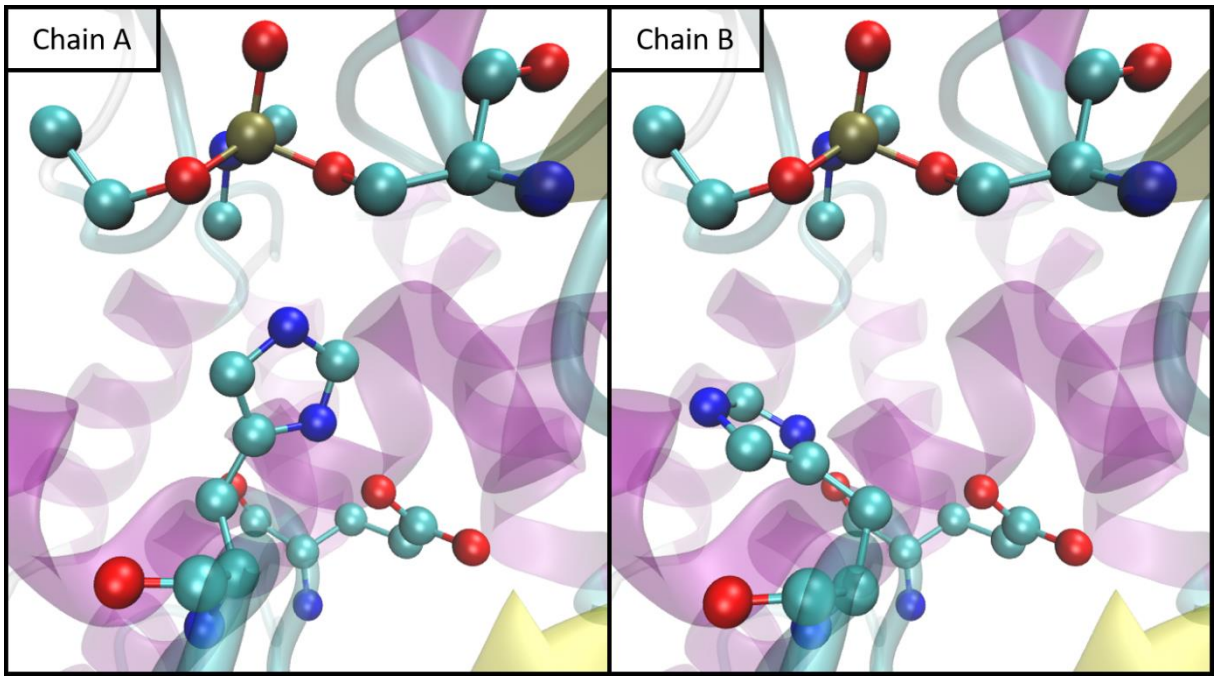


Figure I-4. Chains A and B of His447 in the catalytic triad.

The final heavy atom modifications were to replace the aged tabun adduct by VX. To do so, the N-dimethyl substituent of tabun is replaced by a methyl functional group.

After the heavy atoms modifications of the structure, the next step is the addition of hydrogens. To do so, the protonation state of various protonatable residues of the enzyme must be discussed before moving forward with the setup. The standard pK_a of protonatable residues does not account for the environment of the side chain of a specific residue which might influence its pK_a. To assess this local pK_a of a specific side chain, the software PROpK_a can be used.[3] In the case presented here, PROpK_a was used on the structure with all heavy atom modifications described earlier. All the pK_as calculated using PROpK_a are not presented here because for the sake of brevity as there is over a hundred protonatable residues in AChE. The most likely protonation state based on the calculated pK_a was used for most protonatable residues. The PROpK_a calculated value of a few important residues can be found in **Table I-1**.

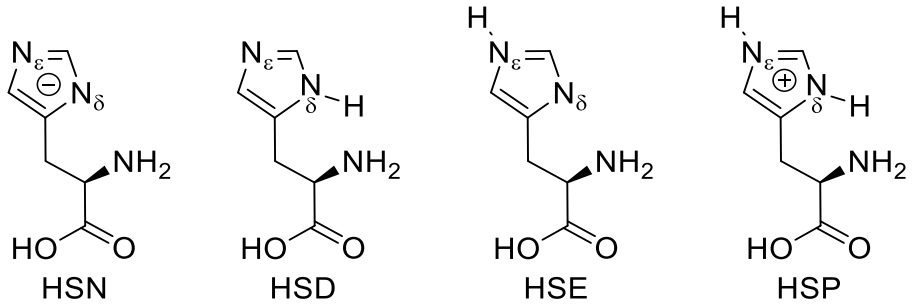
Table I-1. pK_a calculated using PROpK_a for a selection of important residues.

Residue	Glu202	Glu334	His447	Glu450	Glu452
PROpK_a calculated pK_a	10.10	6.42	7.12	6.87	7.27

An interesting feature of these residues whose protonation states will be discussed along the entire manuscript is their ability, except for Glu202, to exchange protons at

[3] a) M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, J. H. Jensen, *J. Chem. Theory Comput.* **2011**, *7*, 525–537. b) C. R. Søndergaard, M. H. M. Olsson, M. Rostkowski, J. H. Jensen, *J. Chem. Theory Comput.* **2011**, *7*, 2284–2295.

physiological pH of 7.2,[4] based on the calculated pK_a values. For the initial setup, Glu334 was kept unprotonated in line with the extensive research of its role in the catalytic triad.[5] Glu450 and Glu452 were kept unprotonated as it is more likely at pH 7.2. Finally, in this chapter the protonation state of His447 and Glu202 will be made to vary to study the role of proton exchange between these residues in the reactivation process. For histidines there is another level of complexity in that there two possible protonation states. Thus a histidine with a protonate imidazole ring can be protonated at site ε, or at site δ (see **Scheme I-1**). A histidine can also be protonated at both ε and δ sites and have an imidazolium side chain. The chosen protonation site for every histidine of AChE is indicated in **Table I-2**. The choice was based on a careful analysis of the environment of these histidines and especially the possibilities for hydrogen bonds with both sites.



Scheme I-1. Possible protonation states for the side chain of histidine residues

Table I-2. Chosen protonation state for histidine residues. Protonation states are designated as per the names defined in **Scheme I-1**.

Residue	His212	His223	His284	His287	His381	His387	His393	His405	His432	His447
Protonation state	HSE	HSE	HSD	HSE	HSD	HSE	HSE	HSD	HSE	HSP
PROpK_a calculated pK_a	2.60	5.77	6.48	5.97	3.50	6.76	6.39	2.14	6.08	7.12

Once these questions had been resolved, hydrogens were added to the rest of the enzyme structure using the tools built in the software CHARMM.[6] The chosen force field for this step as well as the molecular dynamics and the QM/MM simulations described in this chapter is CHARMM22.[7] A 24 Ångström thick shell of water molecules was added around the

[4] R. Vroman, L. J. Klaassen, M. H. C. Howlett, V. Cenedese, J. Klooster, T. Sjoerdsma, M. Kamermans, *PLoS Biol.* **2014**, *12*, e1001864.

[5] a) H. Tsukada, D. M. Blow, *J. Mol. Biol.* **1985**, *184*, 703–711. b) M. A. Massiah, C. Viragh, P. M. Reddy, I. M. Kovach, J. Johnson, T. L. Rosenberry, A. S. Mildvan, *Biochemistry* **2001**, *40*, 5682–5690.

[6] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, et al., *J. Comput. Chem.* **2009**, *30*, 1545–1614.

[7] a) A. D. Mackerell, M. Feig, C. L. Brooks, *J. Comput. Chem.* **2004**, *25*, 1400–1415. b) MacKerell A. D., D. Bashford, M. Bellott, Dunbrack R. L., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, et al., *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

system and in the channel of AChE using Chimera 1.10.1[8] to model a solvated enzyme. The overall charge of the enzyme being -8, the system, with the crystallized chloride ion, has an overall charge of -9. Nine water molecules were substituted with sodium ions using VMD 1.9[9] to have a system with a total neutral charge.

The next step is the minimization, heating, and equilibration of the system using NAMD 2.9.[10] First the full system, acetylcholinesterase, water molecules and ions is minimized with a frozen backbone. The system is then gradually heated to 310 K and finally a 5ns equilibration molecular dynamic is performed. During both the heating and the equilibration, only the inner 8 Å thick shell is mobile, the rest has been kept frozen to prevent the system from bursting. From this dynamic, a snapshot is extracted. The QM/MM simulations will be performed on this snapshot. The outer 16 Å of water molecules used to keep the molecular dynamics in a bubble are removed.

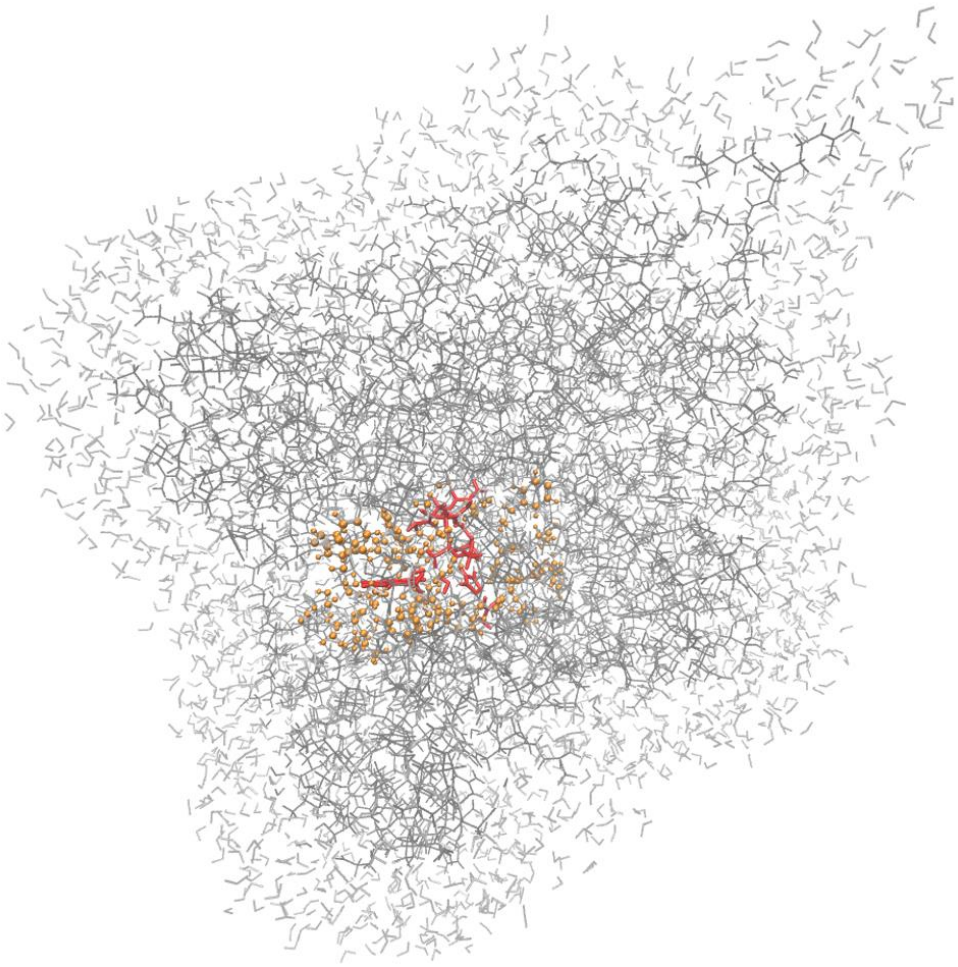


Figure I-5. AChE with QM region in red and in licorice representation, the optimized MM region in orange and ball and stick representation and the static MM region in wire representation and in grey.

[8] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin *J. Comput. Chem*. **2004**, *25*, 1605–1612.

[9] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33–38.

[10] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781–1802.

The QM/MM scheme used is additive QM/MM with three regions (see **Figure I-5**). The QM region, is composed of the catalytic triad, the reactivator ad the oxyanionic hole. The precise composition of this region has been changed at several points during the simulations described in this chapter. Thus, the precise composition is specified at the beginning of the relevant subsections. The MM region includes all other residues but a distinction is made between a fixed MM region and an optimized MM region.

The optimized MM region includes the QM region and most residues surrounding the QM region (**Figure I-6**). More specifically, residues of the enzymatic channel, catalytic anionic site, acyl pocket are included as well as all the nearby water molecules. A chain of residues and water molecules forming hydrogen bonds with Glu334, involving Glu202, Glu450, Glu452, Ser229, and Try428 are also included in the optimized MM region.

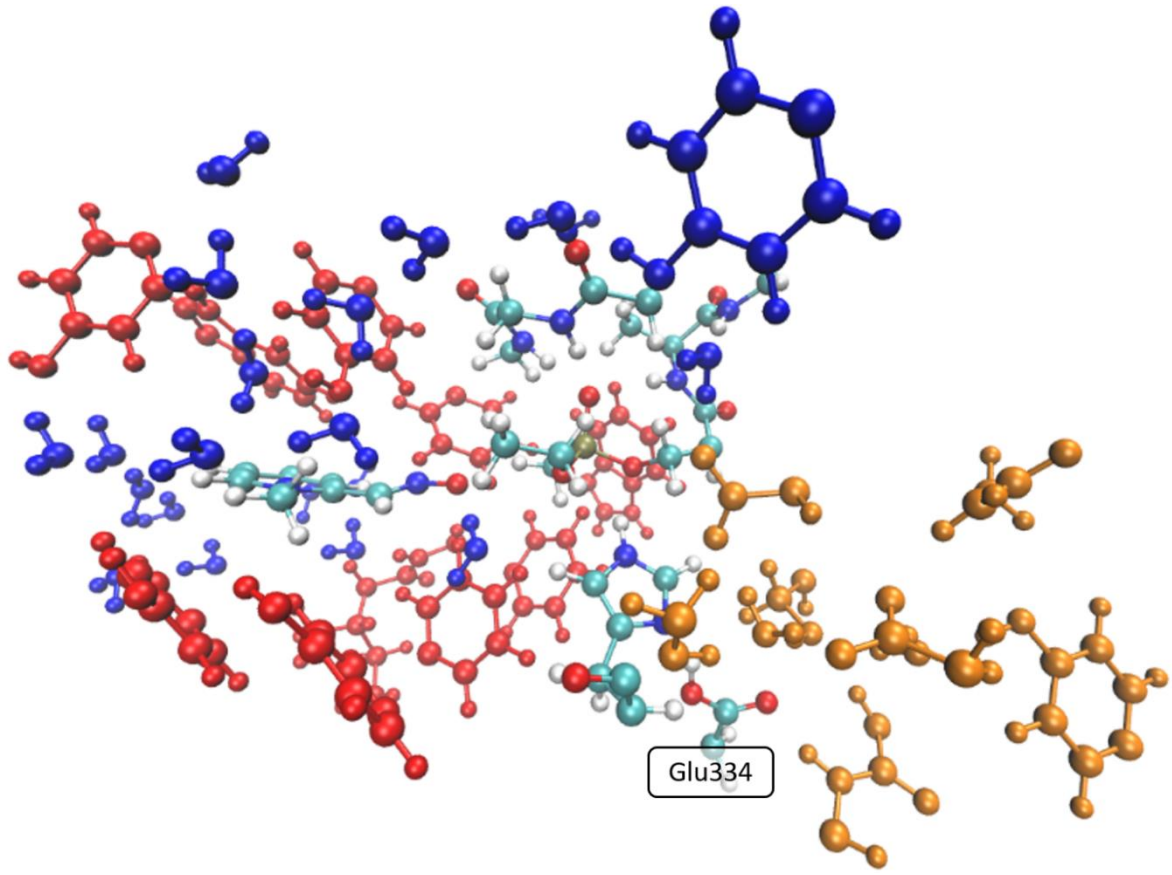


Figure I-6. Residues included in the optimized MM region. The residues of the standard QM region are coloured by chemical element. In red are the residues of the channel and acyl pocket. The water molecules as well as the tyrosine that structures the hydrogen bond network in AChE's active site are in blue. Finally, the residues in yellow are those that form a network of hydrogen bonds with Glu334.

The influence of the MM region on the QM region is taken into account with electrostatic embedding. The link atom with the charge shift scheme is used to handle the boundary between the QM and the MM regions. The geometries are optimized using the HDLCOpt module included in Chemshell 3.1b1.[11] The QM energetic data is provided by

[11] P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, et al., *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 1–28.

Turbomole V6.5[12] and calculated using B3LYP[1] with the D3 dispersion correction[13] and the def2-SV(P) basis set[14]. The MM data is provided by the DL_POLY_3 interface of Chemshell based on CHARMM22 parameters with modifications for the covalent adduct VX and 2-PAM (**Appendix 2**). A script was written to generate the list of atom index for the QM and MM region necessary to set up the Chemshell input. It can be found in (**Appendix 6**, Section **I**). In some important cases, single points were performed using Turbomole at the B3LYP-D3/def2-TZVP level and using ORCA[15] at the DLPNO-CCSD(T)/def2-TZVPP level.[16] In both cases single charges corresponding to the MM region are added to have consistent single points with electrostatic embedding.

The DLPNO-CCSD(T) method, for domain based local pair natural orbital coupled cluster method, is an approximation of the highly accurate CCSD(T) method. It was shown to be consistently within 1 kcal.mol^{-1} of standard CCSD(T) with the NormalPNO setting which is the one used in this work.[17] It is a particularly fast method compared to standard CCSD(T) with a very limited accuracy loss.[18]

II. Destabilization of the frontier orbitals of the systems

At the end of the first geometry optimization of the reactant state of the system, it was observed that the highest occupied orbitals where very high in energy (**Table II-1** row **1**). When compared to the truncated QM model calculations (**Table II-1** row **0**) the five highest occupied orbitals are less stable by 5.4 eV on average. The geometrical changes imposed by the enzyme cannot explain alone a high energy change. This was confirmed by doing a single point energy calculation of the QM region after removing the point charges that allow for the MM region to be accounted for in the QM energy calculation (**Table II-1** row **2**). The result of this single point calculation was that the energies of the five highest occupied orbitals are stabilized by 5.4 eV when removing the point charges. The role of the enzyme in destabilizing the orbitals is thus confirmed. The issue with such a destabilization of the orbitals is that it might indicate a flaw in the QM/MM setup. Thus, it warrants investigation.

Several hypotheses were made on the origin of this destabilization of the orbitals: an improper compensation of the charges from the ions in the water bubble around AChE, residues on the surface of AChE with a poorly compensated negative charge, inadequate description of the electron cloud of the phosphorus and the oxygen atoms bonded to it, a poorly designed QM region.

---

[12] a) R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, *Chem. Phys. Lett.* **1989**, *162*, 165–169. b) R. Ahlrichs, F. Furche, C. Hättig, W. Klopper, M. Sierka, F. Weigend, TURBOMOLE, TURBOMOLE GmbH, Development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, **2012**.

[13] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.

[14] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.

[15] F. Neese, *WIREs Comput Mol Sci* **2012**, 2, 73–78.

[16] a) C. Riplinger, F. Neese, *J. Chem. Phys.* **2013**, *138*, 034106. b) C. Riplinger, B. Sandhoefer, A. Hansen, F. Neese, *J. Chem. Phys.* **2013**, *139*, 134101

[17] D. G. Liakos, M. Sparta, M. K. Kesharwani, J. M. L. Martin, F. Neese, *J. Chem. Theory Comput.* **2015**, *11*, 1525–1539.

[18] N. Myllys, T. Olenius, T. Kurtén, H. Vehkamäki, I. Riipinen, J. Elm, *J. Phys. Chem. A* **2017**, *121*, 4812–4824.

Table II-1. Energy (in eV) of the five highest occupied orbitals for various single points of the reactant of AChE reactivation.

n°		HOMO - 4	HOMO - 3	HOMO - 2	HOMO - 1	HOMO
0	Truncated QM model with oxyanionic hole and glutamate	-6,4744	-6,3381	-6,3330	-5,7910	-5,2606
1	Regular QM/MM	-1,3294	-1,2467	-1,0868	0,0692	**0,5713**
2	QM region without point charges	-6,3847	-6,2400	-6,0283	-5,8825	-5,5573
3	Added diffuse functions on the phosphorus and all three oxygens bonded to it	-1,3264	-1,3057	-1,0601	0,0595	**0,5647**
4	Corrected QM/MM boundaries	-2,1157	-1,9005	-1,5951	-0,2277	**0,2465**
5	4 Na^+ deleted from the point charges	-1,0036	-0,7951	-0,4980	0,4420	0,9169
6	1 Na^+ and 1 Cl^- deleted from the point charges	-3,0092	-2,9717	-2,4745	-1,4723	**-1,0080**
7	1 Cl^- deleted from the point charges	-3,3150	-3,2484	-2,7749	-1,7449	-1,2828
8	All Na^+ deleted from the point charges	1,1601	1,2472	1,6001	2,3550	2,8505
9	All ions deleted from the point charged	0,3039	0,5298	0,9726	1,6804	2,1705
10	1 anionic surface residue deleted from the point charges	-3,5230	-3,4404	-2,9030	-1,9410	-1,4729
11	2 anionic surface residues deleted from the point charges	-3,9446	-3,8525	-3,2280	-2,4179	-1,9459
12	3 anionic surface residues deleted from the point charges	-4,3245	-4,2420	-3,7041	-2,7586	-2,2893
13	4 anionic surface residues deleted from the point charges	-4,7957	-4,7575	-4,2324	-3,2567	-2,7893
14	5 anionic surface residues deleted from the point charges	-5,9102	-5,6041	-5,5343	-4,4218	**-3,9539**
15	8 Na^+ ions moved near anionic surface residues to compensate the charges	-3,8615	-3,7588	-3,2345	-2,4478	**-1,9760**
16	+ 1 Cl^- ion and 1 Na^+ ion removed from the point charges	-3,8696	-3,7209	-3,2363	-2,5275	**-2,0480**
17	+ protonated Glu202	-5,4954	-5,3508	-4,8128	-3,9325	-3,**5244**

The first attempt to correct this destabilization of the high occupied orbitals was to add diffuses functions on the phosphorus and all three oxygen atoms bonded to it, the oxygen of Ser203, the oxygen of the phosphoryl, and the oxygen of the ethoxy substituent. The oxygen and nitrogen of the oxime received the same treatment. For those atoms, the sef2-SV(P) basis set was replaced by the def2-SVPD basis set. This change, as can be seen in **Table II-1** row **3** changed the energy of the orbitals very little, confirming once again that the destabilization of the orbitals was cause by elements outside of the QM region. In following rows, this change in basis set is already included in.

A common problem with QM/MM is a poorly handled boundary between QM and MM regions. To test the quality of the charge shift scheme for the covalent bond cuts that were made for our QM region, the systems, with the point charges was optimized with Turbomole instead of Chemshell. It revealed that the bonds that had been cut were too polar and the link atom hydrogens of a number of bonds were "ejected" far from the QM region, probably due to the electrostatic influence of the point charge of the atom they model. The two atoms that were ejected were the link atom standing for the nitrogen of the amide of Gly120 on the Cα and the hydrogen on the Cα carbon of Ala204 (see **Figure II-1A**). The Cα of Gly120 was removed from the QM region in order to cut in a C-C and not a C-N bond. Part of the backbone of Gly205 was added to move the cut to the C-C bond between the Cα of Glu205 and the C=O (see **Figure II-1B**). This change provided a marginal stabilization of the orbitals as can be seen in **Table II-1** row **5**. They were stabilized on average by 0.5 eV but the highest occupied orbital was only stabilized by 0.32 eV. This correction of the QM region is included in the following rows.

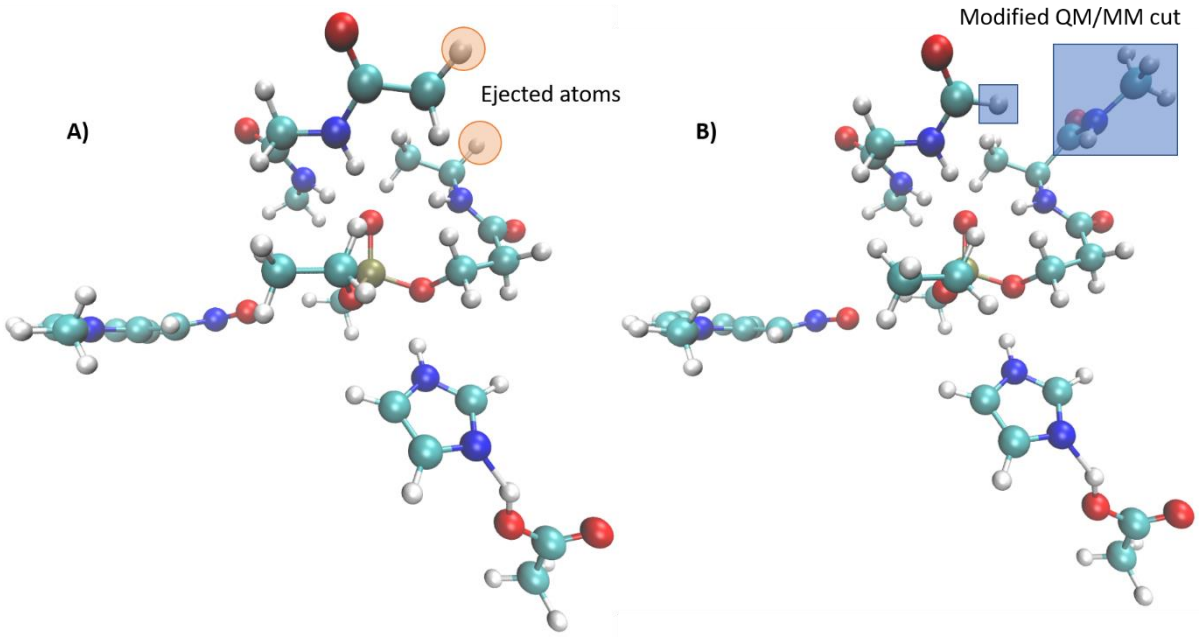


Figure II-1. A) original QM region. B) Modified QM region. Both in Ball&Stick representation and regular chemical element colouring (carbons in light blue).

In the third section of **Table II-1** (rows **5** to **9**), the effect of chloride and sodium ions on those orbitals as studied. To remove either half or all sodium cations completely

destabilizes the highest occupied orbitals by 0.7 eV and 2.6 eV on average respectively (rows **5** and **8**). The removal of a chloride ion is a positive influence as is produces stabilized orbitals by 1.3 eV on average (row **6**). When it is coupled with the removal of a single sodium cation to keep an overall chegre of 0, the stabilization, of 1.5 eV on average, is even greater (row **7**). If all the ions are removed, the orbitals are destabilized almost to the same degree than with the removal of the sodium cations (row **9**). This result seems to implicate long range negative charges in the destabilization of orbitals.

The stabilizing effect of the removal of the chloride ion inspired an investigation of other anions that could destabilize the system. To test the hypothesis that surface anionic residues such as glutamates and aspartates could be the cause of this destabilization, anionic surface residues were selectively removed from the point charges. The results are presented in the fourth section of **Table II-1** (rows **10** to **14**). What they show is that removing a single of those surface residues stabilizes the orbitals by 1.5 eV on average (row **10**) Every subsequent removal of an anionic residue from the point charges further bring further stabilization, up to 4.0 eV on average when five of those anionic residues are removed. In this case the HOMO of the systems goes from 0.25 eV to -3.95 eV, corresponding to a stabilization of 4.20 eV. *Those results indicate surface anionic residues to be poorly solvated and compensated*. To remedy to that problem, all eight sodium cations were selectively displaced closer to eight surface anionic residues to compensate their charge. The results of this change are presented in row **15** of **Table II-1**. While the stabilization this change brings is not as important as the removal of five anionic residues, the highest occupied orbitals are stabilized by 1.9 eV on average. It brings the HOMO of the system from 0.25 eV down to -1.98 eV. Those results show the important long and very-long range interactions can have on the active site, even when it is buried like AChE's. It raises the issue of the pertinence of a simple solvent bubble to simulate the environment of enzymes. AChE for instance is more often than not present as a dimer in synapses, tethered to the phospholipid bilayer and, in the case of inhibited AChE, in an environment overloaded with cationic ACh. To account for such an environment, would greatly increase the computational cost but also complexify the simulations of enzymatic systems.

Rows **16** and **17** of **Table II-1** summarize the efforts to improve the orbital destabilization problem. In row **16**, in addition to the displacement of the sodium cations, the chloride ion as well as a single sodium ion are removed from the system, resulting in a small stabilization of the HOMO and of the second highest occupied orbital from -1.98 to -2.05 and from -2.45 to -2.53 respectively. When catalytic anionic site residue Glu202 is protonated, in addition to all the previous changes to the QM/MM scheme, the energy of the orbitals is further decreased by 1.5 eV on average. It allows the energy of the HOMO to go down to -3.52 eV. This stabilization of the orbitals will be linked with the change in the capacity of AChE to be reactivated when the protonation state of Glu202 is changed which is discussed in section **V** of this chapter. Before discussing reactivation *per se*, a few minor points have been addressed.

III. Orientation of 2-PAM

To proceed with simulations of VX-inhibited AChE reactivation by 2-PAM, the question of the orientation of the reactivator needs to be resolved. Several orientations for 2-PAM in the pre-reactive position where considered. Of the many orientations that where attempted, only two orientations could be fully optimized (see **Figure III-1**). In both cases the plane of the pyridinium of 2-PAM is close to perpendicular to the plane constituted by the phosphoryl of VX with the oxygen of Ser203. In orientation A, the methyl of the methylpyridinium is on same side than the ethoxy substituent of VX. In orientation B, the methyl of the methylpyridinium is on same side than the methyl substituent of VX. Orientation A is more stable than orientation B by 3.59 kcal.mol^{-1} (B3LYP-D3/def2-SV(P):CHARMM). In position A, 2-PAM is in a pre-reactive position with a distance between the phosphorus and the oximate of 4.07 Å.

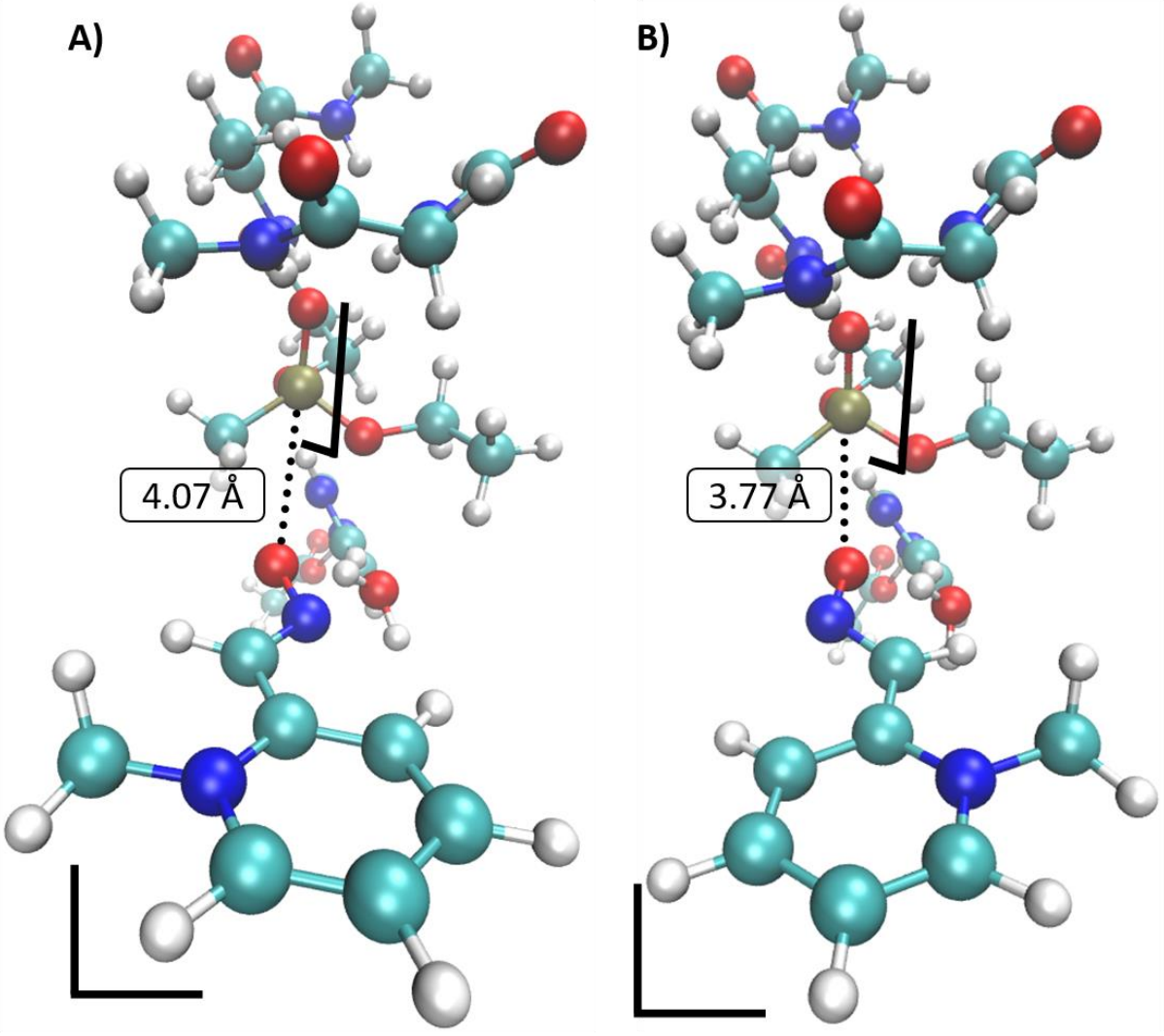


Figure III-1. Possible orientations optimized for 2-PAM in the active site of AChE.

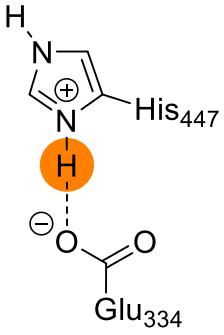
IV. Study of the protonation state of Glu334

One of the main information taken from the truncated model study presented in **Chapter 3** is the importance of the acidity of Glu334. When Glu334 is too basic and

deprotonates His447, the reactivation seems to be impossible. The residues involved in a hydrogen bonds network with Glu334 have been studied. The position of this shared proton between His447 and Glu334 depending on the calculation setup is reported in **Table IV-1**.

Table IV-1. Position of the proton between His447 and Glu334 depending on the QM region in QM/MM calculations or on the truncated QM model used.

Calculation level	system	Proton position
Truncated QM	Minimal truncated model + oxyanionic hole + acetate	Glu334
QM/MM	Minimum QM region (see **Figure II-1B** page **171**)	Glu334
QM/MM	Minimum QM region + Ser229	His447
Truncated QM	Minimal truncated model + oxyanionic hole + acetate + ethanol	Glu334
QM/MM	Minimum QM region (Ser229 rotated as to form a hydrogen bond with Glu334)	His447
QM/MM	Minimum QM region + Ser229, Glu450, Glu202 + 2 hydrogen bond forming water molecules	His447
QM/MM	Minimum QM region + Ser229, Tyr428 + one hydrogen bond forming water molecule	His447



Scheme IV-1. Hydrogen bond between His447 and Glu334

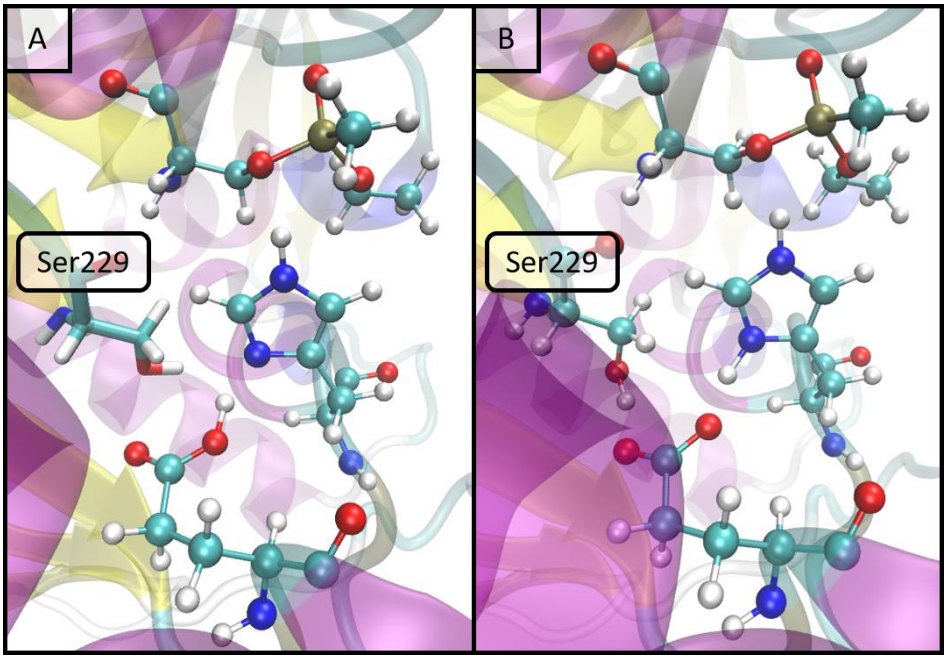


Figure IV-1. Active site of AChE with the catalytic triad in Ball&Stick representation A) Ser229 not included in the QM region in licorice representation B) Ser229 included in the QM region in Ball&Stick representation.

In the truncated QM model including the oxyanionic hole and the acetate, it was observed that the proton between the acetate and the imidazole is located on the acetate representing Glu334. The same observation (see **Figure IV-1A**) can be made in the QM/MM calculation with the minimal QM region presented in **Figure II-1B**. When Ser229 is included in the QM region, its geometry is rearranged to form a hydrogen bond with Glu334 (see **Figure IV-1B**). This serine is connected to several other residues through a network of hydrogen bonds (see **Figure IV-2**).

After this rearrangement, the proton on Glu334 is transferred to His447. To confirm the role of Ser229, an ethanol molecule was added to the truncated model system (**Chapter 3**) near the acetate to model Ser229. This addition does not change the protonation pattern and the proton remains on the acetate modelling Glu334. To further investigate this influence of Ser229 on reactivation the minimal QM region was used again but in the MM region Ser229 was given the optimized geometry it took when it was in the QM region. This new setup allows the proton to remain on His447 over Glu334. The protonation state of His447 and Glu334 was further observed in larger QM regions involving Ser229, Glu450, and Glu202 (**Figure IV-2A**), or Ser229 and Tyr428 (**Figure IV-2B**).

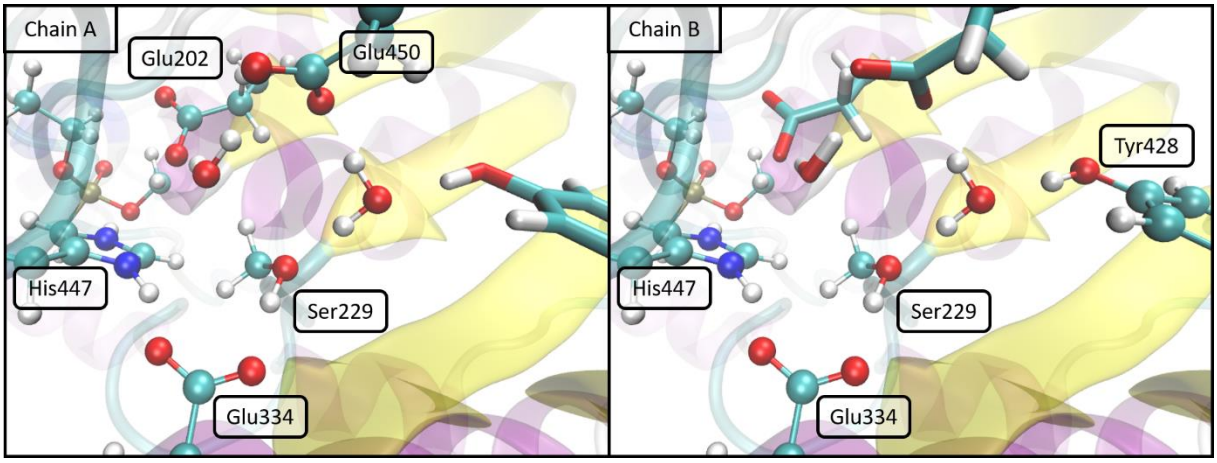


Figure IV-2. hydrogen bond network with Glu334. Residues included in que QM region are in Ball&Stick representation, residues not included are in licorice representation. A) Glu450, Glu202 and Ser229 is the QM region. B) Tyr428 and Ser229 in the QM region

These extended QM regions emphasize a hydrogen bond network ending with Glu334 and the catalytic triad. However, neither the first extended QM region nor the second change the protonation pattern and the proton remains on Glu334. Now that the experimentally determined[5] protonation state for Glu334 has been reached and understood, the reactivation simulations can be carried out.

V. VX-inhibited AChE reactivation by 2-PAM

The QM/MM simulations presented in this section are carried out as per the protocol described in section **I** of this chapter. The QM region used it the one presented in **Figure II-1B**. One of the goals of these simulations, besides the evaluation of the reactivation process, is to

study the effect the change of the protonation state of Glu202 ($Glu202^-$ AChE or $Glu202H^0$ AChE) on the reactivation process.

V.1. Reactivation of AChE with an unprotonated Glu202

The first set of simulations of the reactivation of VX-inhibited AChE by 2-PAM were carried out with an unprotonated Glu202. To localize the approximate transition state of this reaction a monodimensional energy scan along the phosphorus-oxime distance was perfomed. This distance is kept fixed during full geometry optimizations and made to change in small incremental steps. The results of this monodimensional scan did not allow to obtain reliable energy barriers. Thus, similarly to the truncated QM model, a bidimensional scan was performed with both phosphorus-oxime (P-Ox) and phosphorus-Ser203 (P-OSer203) distances controlled and scanned by 0.05 Å steps. From this scan, a three-dimensional potential energy surface was obtained (**Figure V-1**). On this potential energy surface, the high energy area combining a P-OSer203 superior to 2.05 Å and a P-Ox exceeding 2.05 Å was not fully scanned. This area was probed by several mono-dimensional scan and shown to be of very high energy and thus not worth the computational cost of a full two-dimensional scan.

Both the reactant and products were optimized by obtained by releasing constraints on scan points with reactant-like and product-like geometries respectively. In the reactant, called structure **1** (**Figure V-2**), there is a 1.65 Å covalent bond between the phosphorus and the oxygen of Ser203 while the P-Ox is 2.90 Å long. In the product, structure **2** (**Figure V-2**), the covalent bond is between the oxygen of the oxime and the phosphorus and is 1.75 Å long. The P-OSer203 has a length of 2.54 Å. In the reactant and the product the P-OSer203 and P-Ox distances fall outside of the range of the bidimensional scan. To verify that there is no interesting energetic feature between the area of the bidimensional scan and those two structures, mono-dimensional scan along either P-OSer203 or P-Ox distances were performed. Those scans confirmed that in this gap there nothing indicative of a saddle point or anything else than a flat potential energy surface. Mono-dimensional scans were also used to investigate the high energy area with a P-OSer203 superior to 2.05 Å and a P-Ox superior to 2.00 Å. This area was confirmed to be of very high energy and not worth the computational cost of investigating with a full bidimensional scan.

The minimum energy path can be followed between the reactant and the product. Along this minimum energy path there is a single saddle point indicating a transition state. Following this energy path both reaction vary sequentially. First the P-Ox is shortened to the range of 1.90 Å to 1.80 Å, while the P-OSer203 remains almost constant. There is no energy barrier in that part of the three-dimensional energy surface. Now that the P-Ox has been shorten to the same length range than the product state, the P-Oser203 bond increases. Along this increase of the P-OSer203 the minimum energy path goes through a saddle point. Structure **TS1-2** (**Figure V-2**) is the lowest energy point at this saddle point and the transition state of the reaction. In this structure, the P-OSer203 is 2.17 Å and the P-Ox is 1.82 Å, almost the same length than the product structure. This structure was localized by a finer two-dimensional scan around the area of the saddle point in steps of 0.01 Å.

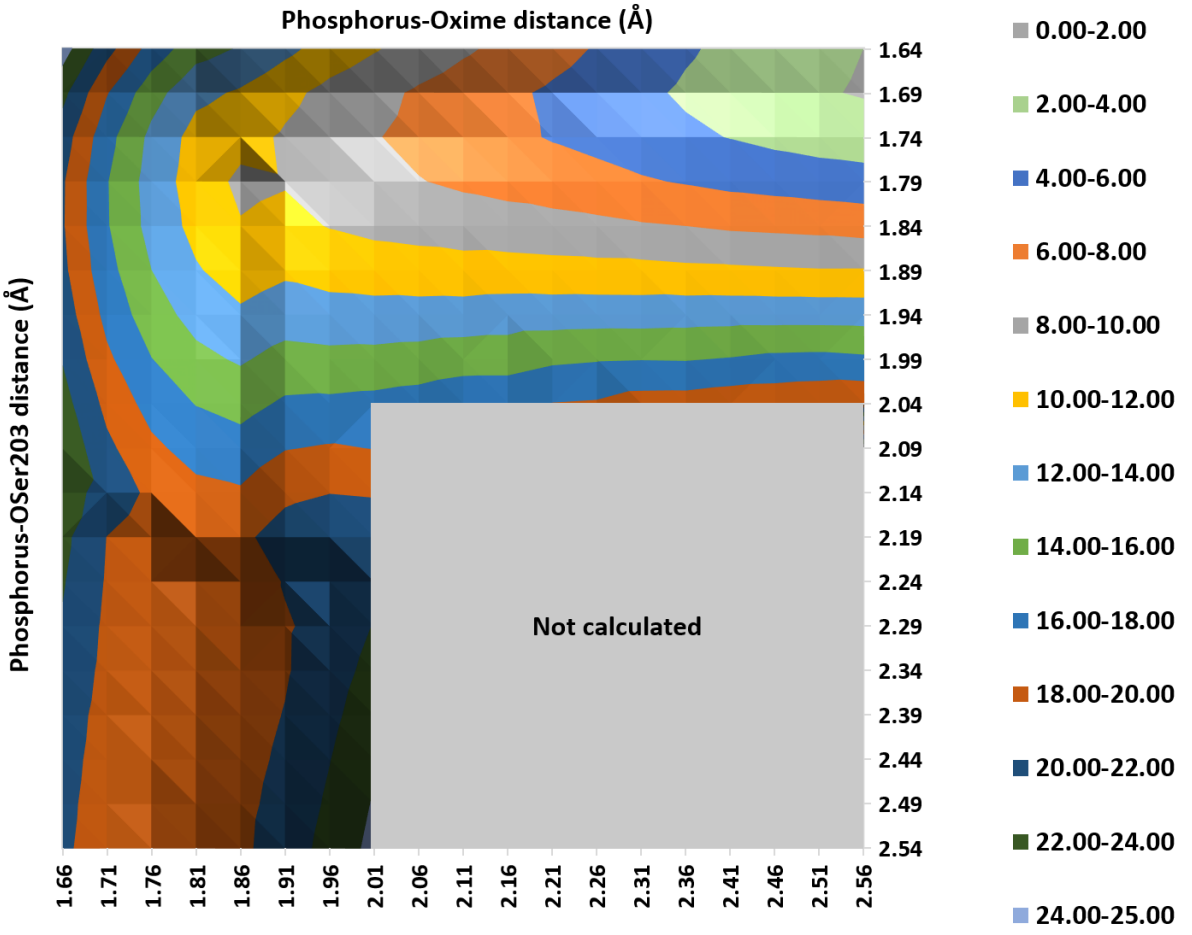


Figure V-1. Three-dimensional potential energy surface for the reactivation of VX-inhibited Glu202$^-$ AChE by 2-PAM.

A single-step reaction with a single transition state indicates a SN_2 mechanism. But a transition state of an SN_2 is characterized by roughly equal distances between leaving group and central atom and nucleophile and central atom which is not the case for structure **TS1-2**. This structure is much closer to the expected geometry for the transition state of the elimination step of an addition-elimination mechanism. The sequential evolution of both bond lengths does not reflect the simultaneity of bond breaking and formation expected in a SN_2 reaction. The reactivation of Glu202$^-$ AChE by 2-PAM thus seems to go through an addition-elimination-like mechanism. During this reactivation, there is no exchange of the proton between Glu334 and His447. This proton remains covalently bonded to His447 and maintains a hydrogen bond with Glu334.

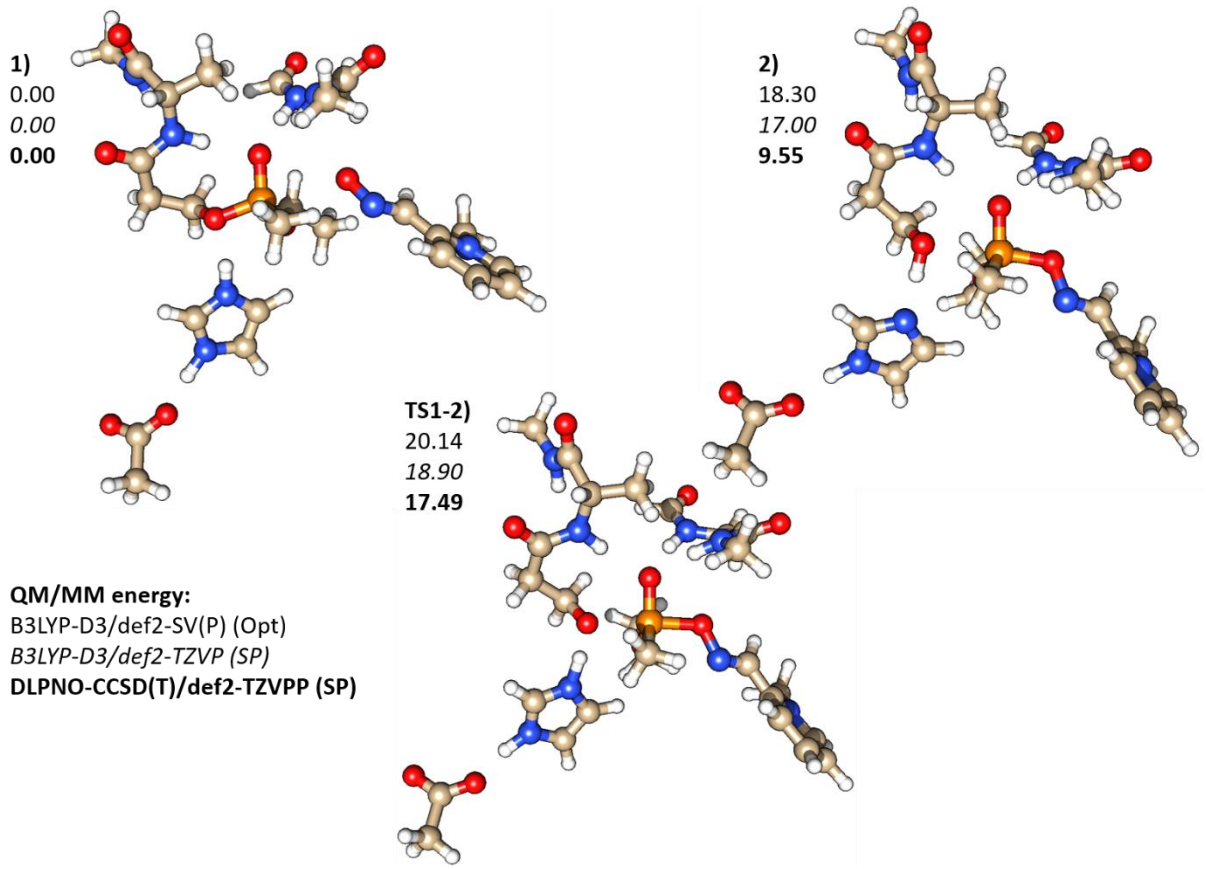


Figure V-2. Optimized structures of (1) the reactant, (2) the product, and (TS1-2) the transition state for the reactivation of VX-inhibited Glu202$^-$ AChE by 2-PAM. Ball&Stick representation with chemical element colouring (carbons in tan).

Due to the well documented underestimation of the activation barrier for SN_2 by B3LYP,[19] two single point calculations were conducted for every structure to confirm the quality of the B3LYP data. B3LYP-D3 single point energetics with a bigger basis set were also conducted. At the B3LYP-D3/def2-SV(P):CHARMM level used for geometry optimization, the reaction is endothermic with an energy barrier of 20.14 kcal.mol^{-1} and the product's (**2**) energy higher than that of the reactant (**1**) by 18.30 kcal.mol^{-1}. Single points with a larger basis set, def2-TZVP, gave similar energies with an energy barrier of 18.90 kcal.mol^{-1} and an endothermicity of 17.00 kcal.mol^{-1}. Finally, the DLPNO-CCSD(T)/def2-TZVPP single points gave a 17.49 kcal.mol^{-1} energy barrier and an endothermic reaction by 9.55 kcal.mol^{-1}. The DLPNO-CCSD(T) single points reproduce the trend of DFT calculations and therefore corroborate their reliability. The energy differences between the optimization energy and the two single point energies, especially between the B3LYP-D3 energies and DLPNO-CCSD(T) energies, are caused by differences in the potential energy surface at the different calculation level. The reactant and transition states at the B3LYP-D3/def2-SV(P) level are likely to have slight geometrical differences to the reactant and transition states at the DLPNO-CCSD(T)/def2-TZVPP level. This energy barrier is not too high for Glu202$^-$ AChE to overcome, but the endothermicity prevents reactivation in this protonation state situation by microreversibility.

[19] O. V. Gritsenko, B. Ensing, P. R. T. Schipper, E. J. Baerends, *J. Phys. Chem. A* **2000**, *104*, 8558–8565.

V.2. Reactivation of AChE with a protonated Glu202

Because the reactivation of VX-inhibited AChE by 2-PAM when Glu202 is unprotonated seem not feasible in the light of the QM/MM results presented in the previous sub-section, a new protonation state should be envisaged for this residue. In the simulations presented in this sub-section, Glu202 is protonated (i.e. in its acid state)

Since a two-dimensional scan had to be resorted to for the unprotonated Glu202 case, a bidimensional scan was also used for protonated Glu202. Again, both P-Ox and P-OSer203 distances are controlled and scanned by 0.05 Å steps. This time 282 points were gathered, all optimized in B3LYP-D3 with def2-SV(P). All points of this two-dimensional scan yield an energy value letting three-dimensional potential energy surface to be plotted (**Figure V-3**). On this potential energy surface, the high energy area combining a P-OSer203 superior to 2.0 Å and a P-Ox exceeding 2.25 Å was not fully scanned. This area, similarly to the $Glu202^-$ scan, was probed by several mono-dimensional scan and shown to be of very high energy and thus not worth the computational cost of a full two-dimensional scan.

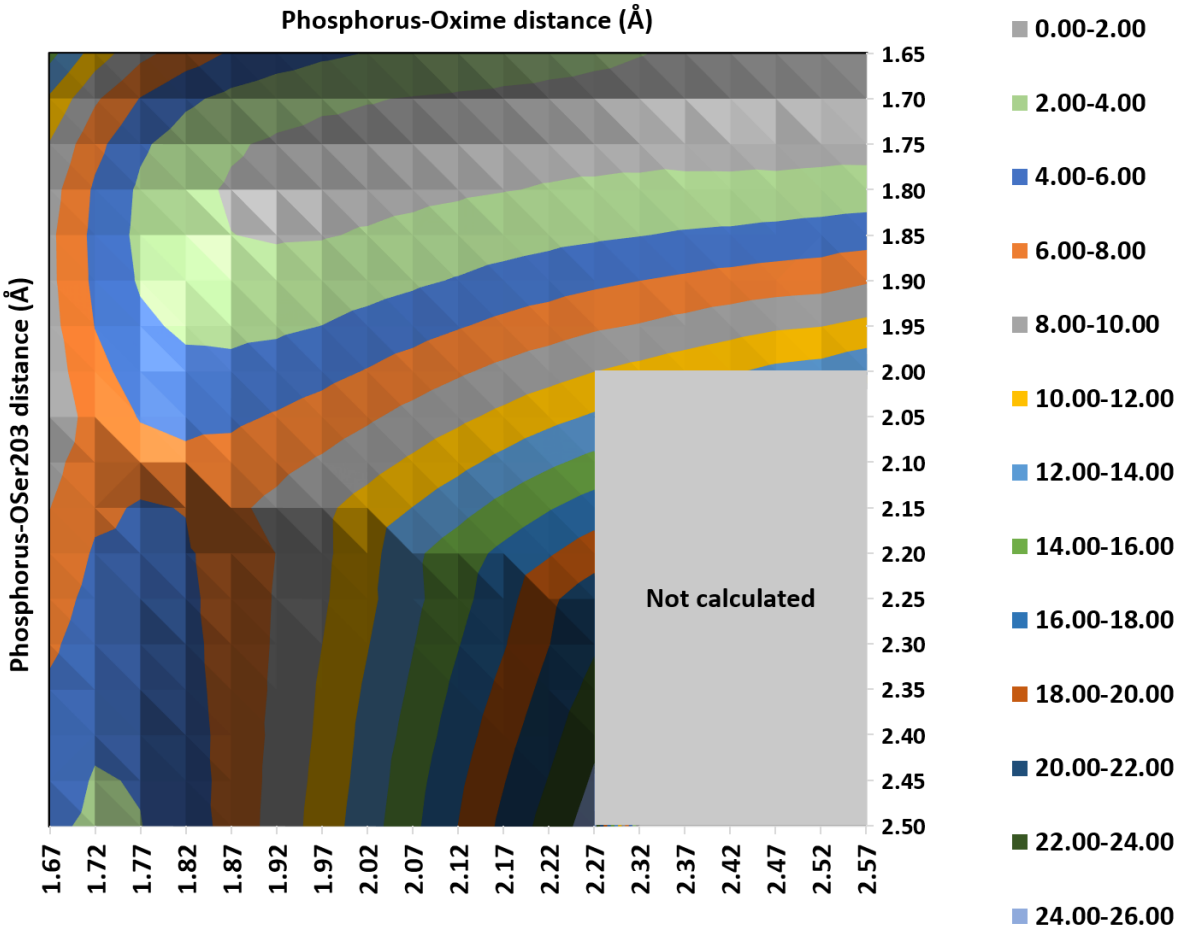


Figure V-3. Three-dimensional potential energy surface for the reactivation of VX-inhibited $Glu202H^0$ AChE by 2-PAM.

Both the reactant and products were optimized by obtained by releasing constraints on scan points with reactant-like and product-like geometries respectively. In the reactant,

called structure **1P** (**Figure V-4**), there is a 1.67 Å covalent bond between the phosphorus and the oxygen of Ser203 while the P-Ox is 2.75 Å long. In the product, structure **2P** (**Figure V-4**), the covalent bond is between the oxygen of the oxime and the phosphorus and is 1.72 Å long. The P-OSer203 has a length of 2.50 Å. To confirm the absence of energy variations between the edge of the potential energy surface an additional monodimensional scan with both P-OSer203 and P-Ox under constraints was realised. It confirmed the flatness of the potential energy surface in this section.

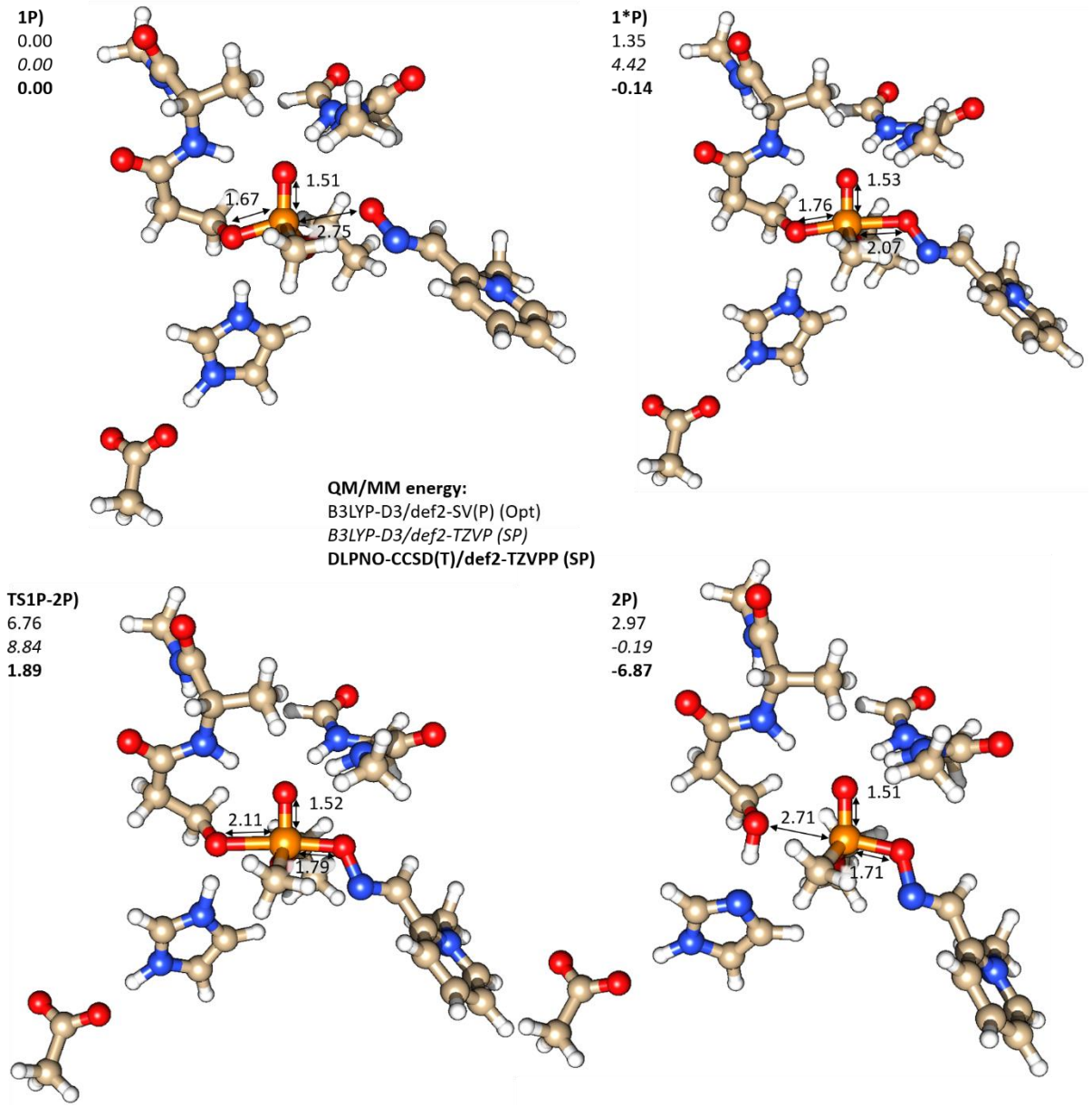


Figure V-4. Optimized structures of (1P) the reactant, (1*P) the stable intermediate, (2P) the product, and (TS1P-2P) the transition state for the reactivation of VX-inhibited Glu202H^0 AChE by 2-PAM.

The reactant and product can be connected on the potential energy surface by the minimum energy path. Following this minimum energy path, the two reaction distances P-Os and P-Oser203 vary sequentially. The shortening of the P-Ox to the range of 1.80-1.90 Å precedes the lengthening of the P-OSer203 that marks the departure of the leaving group, in

that case the whole enzyme, from the phosphorus. Along this energy path we see that the reduction of the P-Ox distance is associated with close to no energy change. After this reduction in the P-Ox distance, a stable reaction intermediate (**1*P**) was found and optimised. This reaction intermediate has a long P-Ox bond of 2.07 Å and a shorter P-OSer203 of 1.76 Å. No saddle point was located between this intermediate and the reactant due to a very flat potential energy surface in that section. When the P-Ox has been shortened to the range of 1.80-1.90 Å the minimum energy path shifts and the P-OSer203 increases. Along this increase of P-OSer203 along the minimum energy path the P-Ox remain almost constant and shifts form the range of 1.80-1.90 Å to the range of 1.70-1.80 Å. Following the minimum energy path as P-OSer203 is stretched, a saddle point is encountered. The transition state (structure **TS1P-2P** in **Figure V-4**) corresponding to this saddle point was localized by a finer scan on the saddle point area by 0.01 Å steps incremental variation of the P-Ox and P-OSer203. In this transition state, the P-Ox is 1.79 Å and the P-OSer203 is 2.11 Å. Interestingly it almost mirrors the structure of the pentavalent intermediate.

The potential energy surface is reminiscent of an addition-elimination mechanism where the two reactive distance vary sequentially and not simultaneously. The optimisation of a stable bipyramidal trigonal reaction intermediate further confirms the reactivation to likely have a two-step mechanism with a stable intermediate in between. We were unable to find the energy barrier for the first step between the reactant (**1P**) and the stable intermediate (**1*P**) due to the flatness of the surface. The stable intermediate is the pentavalent structure expected for an addition elimination mechanism on a phosphorus centre although distorted with a much longer phosphorus-oxime distance (2.07 Å) than the phosphorus-Ser203 distance (1.76 Å) where a more symmetrical intermediate could be expected. The **TS1P-2P** structure with its long phosphorus-Ser203 distance (2.11 Å) and the much shorter phosphorus-oxime distance (1.79 Å) is the transition state of the elimination step. Once again, there is no exchange of the proton between Glu334 and His447. This proton remains covalently bonded to His447 and maintains a hydrogen bonds with Glu334.

At the B3LYP-D3/def2-SV(P) level, used for geometry optimization, the reactivation of AChE by 2-PAM has an energy barrier of 6.76 kcal.mol^{-1} and is slightly endothermic with an energy difference of 2.97 kcal.mol^{-1}. The stable intermediate is only 1.35 kcal.mol^{-1} higher in energy than the product. Single points with the def2-TZVP basis set give an energy barrier of 8.84 kcal.mol^{-1} and an almost isoenergetic reaction with an energy difference of -0.19 kcal.mol^{-1}. The stable intermediate has an energy 4.42 kcal.mol^{-1} higher than that of the reactant. The DLPNO-CCSD(T) single points show the same trend with an energy gain of -6.87 kcal.mol^{-1} and a small reaction barrier of 1.89 kcal.mol^{-1}. The stable intermediate is -0.14 kcal.mol^{-1} more stable than the reactant. The reliability of the DFT calculations is once again confirmed. This case, contrary to the unprotonated Glu202 case, thus leads to reactivation. The energy barrier is small enough and the enthalpy is close to zero.

V.3. Reactivation with an unprotonated His447 and a protonated Glu202

The common consensus in the literature on the reactivation of VX-inhibited AChE by 2-PAM is that His447 must be protonated for the reactivation to take place. A quick set of simulations was performed on a system with an unprotonated His447 and a protonated Glu202 to check that assumption. The results are presented in **Figure V-5**. The product structure could not be optimized without constraints else it would revert to the reactant state. Structure **6** presented in **Figure V-5** has been obtained by keeping the distance between the phosphorus and the oxygen of Ser203 under constraint and is not a "true" product state.

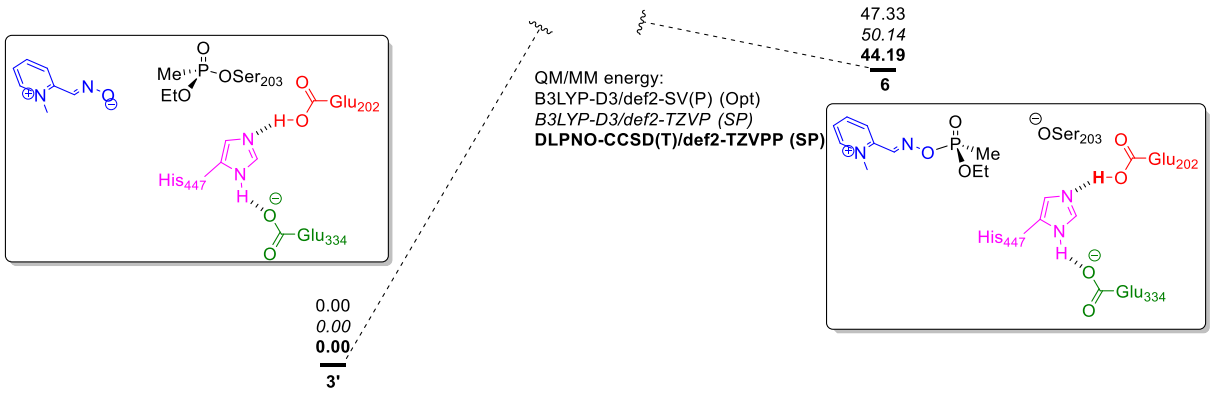


Figure V-5. Energy profile for the reactivation of VX-inhibited AChE by 2-PAM when His447 is not protonated and Glu202 is protonated. Energies in kcal.mol^{-1}.

The energy difference between the reactant (**3'**) and the product state (**6**) is 47.33 kcal.mol^{-1} at the B3LYP-D3/def2-SV(P) level. The single points also show a large endothermicity with 50.14 kcal.mol^{-1} at the B3LYP-D3/def2-TZVP level and 44.19 kcal.mol^{-1} at the DLPNO-CCSD(T)/def2-TZVPP level. There was no attempt to calculate the energy of the transition state as the energy difference between product and reactant is obviously too high for the reaction to occur. The only thing that can be said on that front is that the energy barrier is at least greater than the energy of the product.

V.4. QM model confirmation

The importance of the protonation state of Glu202 in the reactivation of VX-inhibited AChE can be further confirmed by returning to the truncated QM model methodology (see **Figure V-6**). A slightly modified truncated QM model was used with a geometry taken directly from the reactant structure in enzyme. In this truncated model, several atoms are kept fixed in order to prevent any unwanted interaction between the acetic acid modelling Glu202 and His447. When the acetic acid and the imidazole are not fixed, the imidazole flips to for a hydrogen bond with the acetate and ends up being non-reactive. Two set of simulations were conducted, with a protonated and an unprotonated Glu202 to check the influence this protonation state has in a simpler model. In both cases two-dimensional scan along the P-Ox and P-OSer203 distances were performed to compare the truncated QM model calculations with the QM/MM potential energy surfaces obtained with QM/MM. The method used was B3LYP-D3/def2-SV(P) as implemented within Turbomole V6.5.

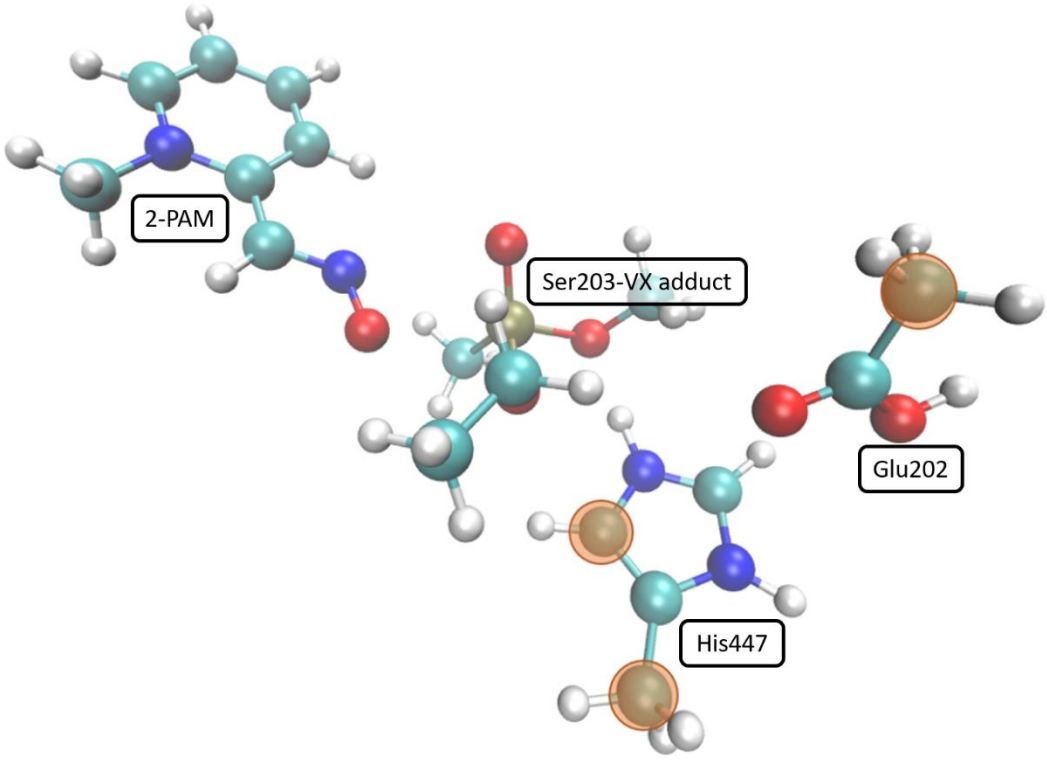


Figure V-6. Truncated QM model used with fixed atoms highlighted.

From the potential energy surface obtained with protonated and unprotonated Glu202 models, an energy profile was established (see **Figure V-7**). On this profile, it is evident that the observed trend in QM/MM is reproduced in the QM model. A protonated Glu202 diminishes the reaction barrier for the reactivation and increases the exothermicity. When Glu202 is unprotonated the energy barrier is very high at 24.01 kcal.mol^{-1} and the energy difference between reactant and product is 23.92 kcal.mol^{-1}. When Glu202 is protonated, the energy barrier is reduced by 17.05 kcal.mol^{-1} down to 6.96 kcal.mol^{-1}. The energy difference goes to 4.11 kcal.mol^{-1}.

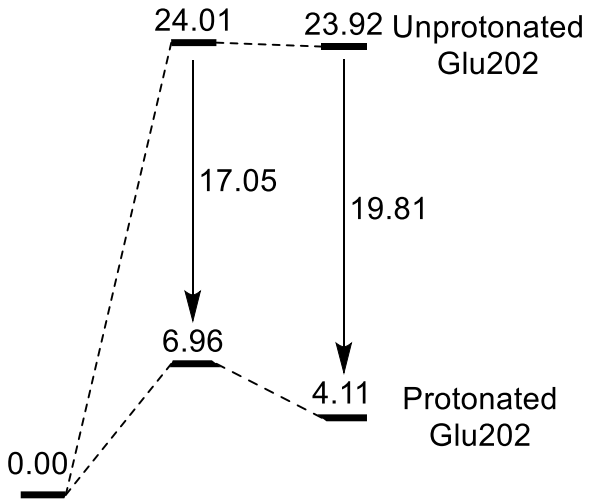


Figure V-7. Energy profile for the reactivation of VX-inhibited AChE by 2-PAM in the truncated QM model for the two possible protonation states of Glu202. Energies in kcal.mol^{-1}.

This stabilization confirms that the trends observed in QM/MM are in fact due to the influence of the protonation state of Glu202. It does not come from some secondary factor.

V.5. Alternate approach of 2-PAM towards the phosphorus

One of the remaining questions on the reactivation of AChE is the possibility of a different approach of 2-PAM towards the phosphorus than the classical approach with the oxime is trans position relative to the catalytic serine (see **Figure V-8A**). The alternate approach that could be envisaged is an approach with the oxime in trans position relative to the phosphoryl bond. In a view of AChE's active site with the oxyanionic hole on top and the phosphoryl bond completely vertical, the oxime would approach from underneath the phosphorus (see **Figure V-8B**). The pre-reactive position for 2-PAM trans from the phosphoryl is 9.42 kcal.mol^{-1} less stable trans the classical approach in B3LYP-D3/def2-SV(P). This large energy difference is confirmed by both the B3LYP-D3/def2-TZVP, and DLPNO-CCSD(T)/def2-TZVPP single points.

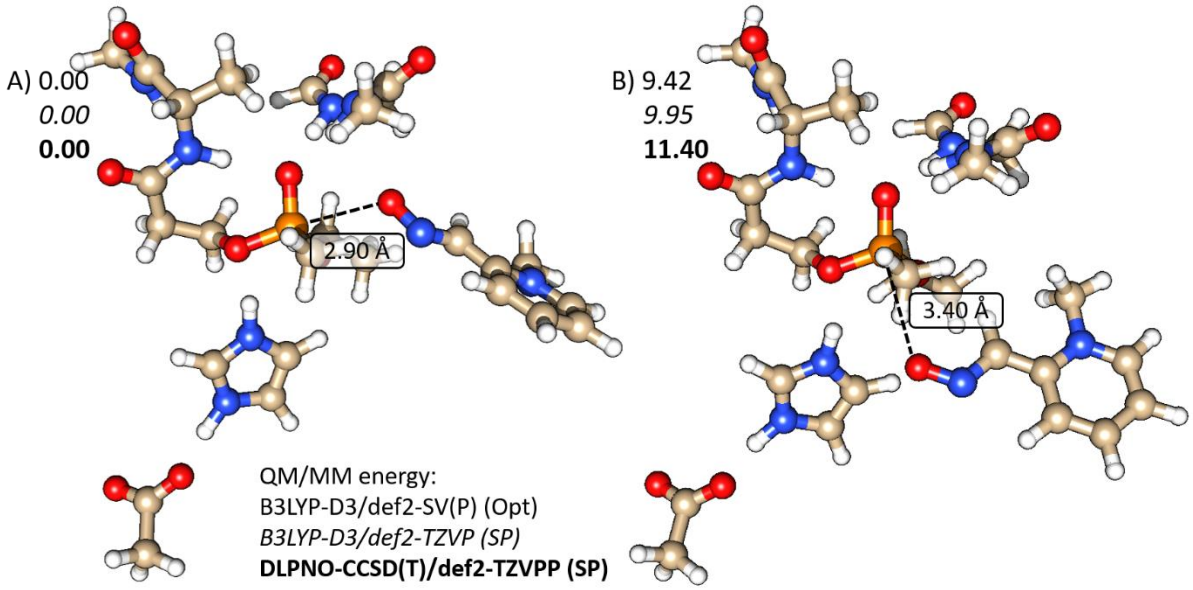


Figure V-8. Pre-reactive positions for 2-PAM in: A) trans from the serine position and B) trans from the phosphoryl position

To completely rule out the possibility of the alternate approach, the energy difference between the reactant and product with both approaches was established. In the subsequent discussion, the reactant for both approaches are used as the respective 0 for this approach. The structure of the product of the alternate approach is not known as some form of rearrangement is to be expected. To not bias the energy difference from this approach a mono-dimensional scan forcing 2-PAM to close the distance with the phosphorus what performed. Besides the distance constraint between the phosphorus and the oxygen of the oxime all atoms are free to move and the expected rearrangement does occur. As the oxime is brought closer to the phosphorus the oxime slides up and breaks linearity with the axis of

the phosphoryl bond. The methyl substituent of VX slides opposite to the catalytic serine simultaneously in what could be the result of a Berry pseudo-rotation (see **Figure V-9**).[20]

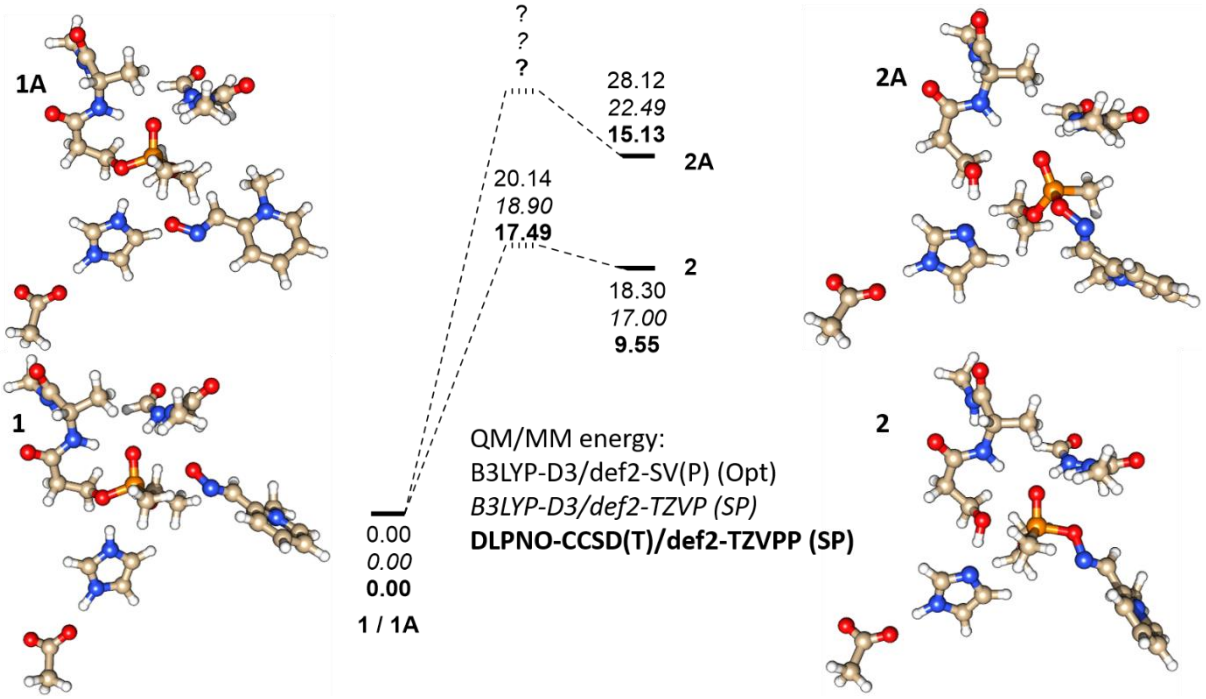


Figure V-9. Energy profile for the reactivation of VX-inhibited AChE by 2-PAM when Glu202 is unprotonated following two possible approaches for the reactivator: trans from the serine (**1**/**2**) and trans from the phosphoryl group (**1A**/**2A**)

The energy profile in **Figure V-9** contains the energetics of the reactivation of VX-inhibited AChE by 2-PAM through the classical approach in the case of unprotonated Glu202. This reactivation has an energy difference of 18.30 kcal.mol^{-1} in B3LYP-D3/def2-SV(P) (17.00 kcal.mol^{-1} in B3LYP-D3/def2-TZVP and 9.55 kcal.mol^{-1} in DLPNO-CCSD(T)/def2-TZVPP) and an energy barrier of 20.14 kcal.mol^{-1} in B3LYP-D3/def2-SV(P) (18.90 kcal.mol^{-1} in B3LYP-D3/def2-TZVP and 17.49 kcal.mol^{-1} in DLPNO-CCSD(T)/def2-TZVPP). When the reactivation occurs through the alternate approach the endothermicity is increased by 10.18 kcal.mol^{-1} in B3LYP-D3/def2-SV(P) (5.49 kcal.mol^{-1} in B3YLP-B3/def2-TZVP and 5.68 kcal.mol^{-1} in DLPNO-CCSD(T)/def2-TZVPP) to 28.12 kcal.mol^{-1} in B3LYP-D3/def2-SV(P) (22.49 kcal.mol^{-1} in B3LYP-D3/def2-TZVP and 15.13 kcal.mol^{-1} in DLPNO-CCSD(T)/def2-TZVPP). The endothermicity of the reactivation through the alternate approach exceeds the energy barrier for the reactivation through the classical approach. The energy barrier for the alternate approach should be at least as high in energy as the product for the alternate approach and is thus also higher than the barrier for the classical approach. Based on transition state theory, and using the endothermicity as a minimum boundary for the energy barrier, the alternate approach is expected to have a rate constant at least 420 times inferior to the classical approach. The choice was made not to spend tremendous computational resource and time to locate the transition state(s) for the reactivation from the alternate approach because the established

[20] R. S. Berry, *J. Chem. Phys.* **1960**, *32*, 933–938.

energy difference and energy barrier difference make this approach very unfavourable compared to the classical approach in the case of unprotonated Glu202. Calculations were also performed to check the feasibility of the alternate approach when Glu202 is protonated.

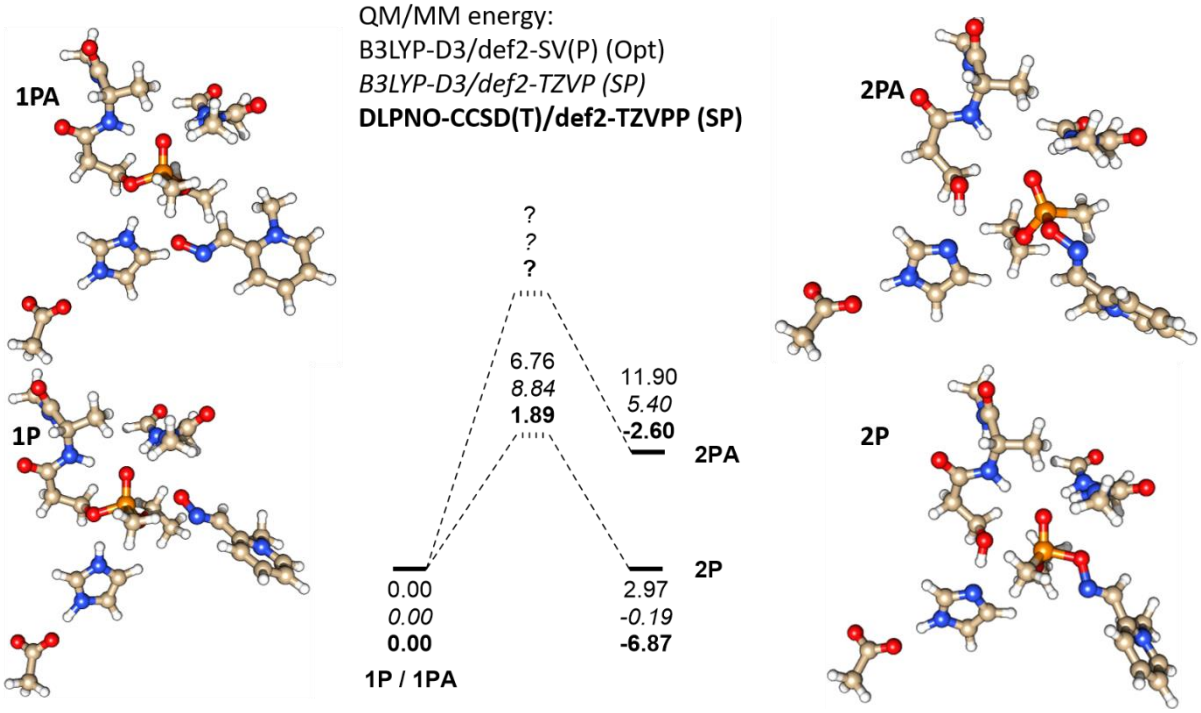


Figure V-10. Energy profile for the reactivation of VX-inhibited AChE by 2-PAM when Glu202 is protonated following two possible approaches for the reactivator: trans from the serine (**1P**/**2P**) and trans from the phosphoryl group (**1PA**/**2PA**)

The energy profile in **Figure V-10** presents the energetics of the react through both the classical and the alternate approach in the case of protonated Glu202.

The reactivation through the classical approach has an energy difference of 2.97 kcal.mol^{-1} in B3LYP-D3/def2-SV(P) (-0.19 kcal.mol^{-1} in B3LYP-D3/def2-TZVP and -6.87 kcal.mol^{-1} in DLPNO-CCSD(T)/def2-TZVPP) and an energy barrier of 4.42 kcal.mol^{-1} in B3LYP-D3/def2-SV(P) (8.84 kcal.mol^{-1} in B3LYP-D3/def2-TZVP and 1.89 kcal.mol^{-1} in DLPNO-CCSD(T)/def2-TZVPP).

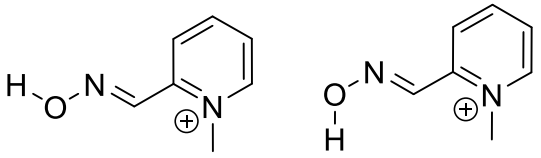
When the reactivation occurs through the alternate approach the endothermicity is increased by 8.93 kcal.mol^{-1} in B3LYP-D3/def2-SV(P) (5.59 kcal.mol^{-1} in B3LYP-D3/def2-TZVP and 4.27 kcal.mol^{-1} in DLPNO-CCSD(T)/def2-TZVPP) to 11.90 kcal.mol^{-1} in B3LYP-D3/def2-SV(P) (5.40 kcal.mol^{-1} in B3LYP-D3/def2-TZVP and -2.60 kcal.mol^{-1} in DLPNO-CCSD(T)/def2-TZVPP). The increase in endothermicity make the alternate approach much less favourable than the classical approach. Once again, the choice was made not to investigate the energy barrier for the reactivation through the alternate approach as there is compelling evidence that this approach is less favourable in the case of protonated Glu202 as well.

Now that the reactivation in both protonation states of Glu202 has been modelled in QM/MM, the question of the deprotonation of the reactivator in the active site remains to be studied for a full overview of the reactivation process.

VI. Deprotonation of 2-PAM in the active site of AChE

With a pK_a of 7.7, at physiological pH of 7.2, 76% of the reactivator is expected to be in its protonated (i.e. acidic) state. It means that an encounter between the VX-inhibited enzyme and the reactivator is much more likely to occur with 2-PAM in its protonated oxime form, whereas the reactive form is the oximate form. It has been envisaged in the literature that the oxime could be deprotonated in the active site of AChE but not seriously investigated computationally.[21] It was also hypothesized on the basis of X-Ray analysis of AChE crystals with reactivators in pre-reactive positions, that His447 could deprotonate the reactivator.[22] Assuming that only His447 is accessible to deprotonate 2-PAM with Glu202 assisting the deprotonation, three possibilities exist for the deprotonation of 2-PAM depending on the protonation state of Glu202 and His447: both Glu202 and His447 are unprotonated or either one is protonated and the other is not. If both residues are protonated, none of them is available for a possible deprotonation of 2-PAM.

The simulations presented in this section use the same QM/MM methodology used throughout but with the addition of residue Glu202 in the QM region. The transition states of all proton transfer steps were located though mono-dimensional scans with a single distance under constraint, between the proton and the atom that is supposed to receive it. This reaction distance was made to vary in 0.05 Å increments and when the area of the TS was located, the precise TS was obtained by a finer scan with 0.01 Å steps. The matter of the position of the proton on the oxime was addressed by comparing the energy of both positions. When the proton on the opposite side of the N-O bond compared to the carbon of the oxime, the total energy is 6 kcal.mol^{-1} compared to the Z position of the proton (see **Scheme VI-1**).



Scheme VI-1. Position of the proton of the oxime of 2-PAM

VI.1. Case of unprotonated His447 and unprotonated Glu202

The first set of simulations to be presented is the deprotonation of 2-PAM when both His447 and Glu202 are unprotonated. **Figure VI-1** summarizes the energy profile for this proton transfer step.

[21] a) A. da S. Gonçalves, T. C. C. França, J. D. Figueroa-Villar, P. G. Pascutti, *J. Braz. Chem. Soc.* **2011**, *22*, 155–165. b) A. da Silva Gonçalves, T. C. C. França, M. S. Caetano, T. C. Ramalho, *J. Biomol. Struct. Dyn.* **2014**, *32*, 301–307.

[22] a) F. Ekström, A. Hörnberg, E. Artursson, L.-G. Hammarström, G. Schneider, Y.-P. Pang, *PLoS One* **2009**, *4*, e5957. b) A. Hörnberg, E. Artursson, R. Wärme, Y.-P. Pang, F. Ekström, *Biochem. Pharmacol.* **2010**, *79*, 507–515. c) A. Allgardsson, L. Berg, C. Akfur, A. Hörnberg, F. Worek, A. Linusson, F. J. Ekström, *Proc. Nat. Ac. Sci. USA* **2016**, *113*, 5514–5519.

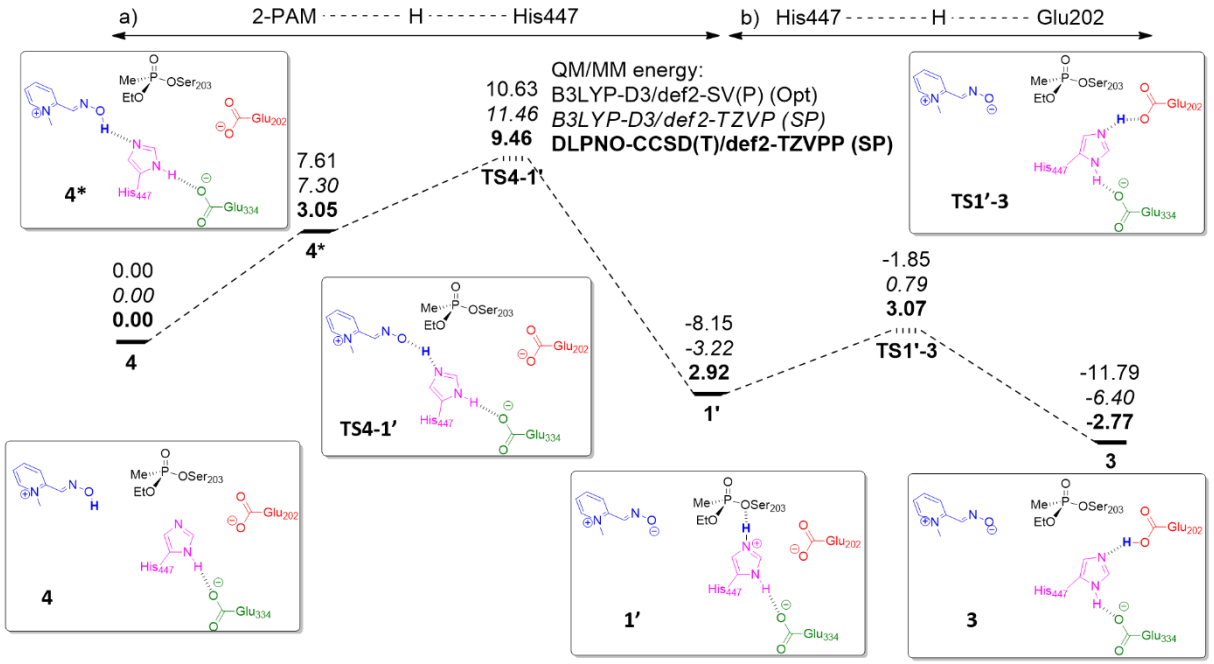


Figure VI-1. Energy profile for the proton transfer from 2-PAM to His447 (a) and from His447 to Glu202 (b). Energies in kcal.mol^{-1}.

The deprotonation of 2-PAM by His447 is a one-step exothermic reaction with an energy difference of -8.15 kcal.mol^{-1} in B3LYP-D3/def2-SV(P). The energy barrier is 10.63 kcal.mol^{-1} in B3LYP-D3/def2- SV(P). A stable intermediate (**4***) was optimized between the reactant (**4**) and the transition state (**TS4-1**). It has an energy of 7.61 kcal.mol^{-1} in B3LYP-D3/def2- SV(P), higher than the reactant. No energy barrier could be found between the reactant and this intermediate. When 2-PAM approaches His447 it forms a hydrogen bond between 2-PAM and His447 present in stable intermediate **4***. The intermediate is located in that energy plateau. The single point B3LYP-D3/def2-TZVP energies show a stable intermediate 7.30 kcal.mol^{-1} higher in energy compared to the reactant. The deprotonation has an energy barrier of 11.46 kcal.mol^{-1}, and an energy difference of -3.22 kcal.mol^{-1}. The DLPNO-CCSD(T)/def2-TZVPP energies show the deprotonation to have an energy difference of 2.92 kcal.mol^{-1} with a small energy barrier of 9.46 kcal.mol^{-1}. In that case the reaction is endothermic as opposed to exothermic because of geometry differences between the local minima from the potential energy surface at B3LYP-D3/def2-SV(P) and the potential energy surface at DLPNO-CCSD(T)/def2-TZVPP level. However, the DLPNO-CCSD(T)/def2-TZVPP single point does confirm the deprotonation to have a small energy difference and a small energetic barrier.

The deprotonation of 2-PAM by His447 is immediately followed by the deprotonation of His447 by Glu202. This reaction is also exothermic with an energy difference of -3.64 kcal.mol^{-1} at the B3LYP-D3/def2-SV(P) level between the reactant (**1'**) and the product (**3**). The energy barrier is only 6.30 kcal.mol^{-1} in B3LYP-D3/def2-SV(P) (**TS1'-3**). At the B3YLP-D3/ef2-TZVP level, the energy barrier is 4.01 kcal.mol^{-1} and the energy difference is 3.18 kcal.mol^{-1}. There again the DLPNO-CCSD(T) confirms the trends obtained with both basis sets in B3LYP-

D3. The single points show that the deprotonation of His447 as an energy difference of -5.69 kcal.mol^{-1} and has a very small energy barrier of 0.15 kcal.mol^{-1}.

VI.2. Case of unprotonated His447 and protonated Glu202

The second possibility for the deprotonation of 2-PAM, like for the reactivation, is a protonated Glu202. **Figure VI-2** summarizes the energy profile for this proton transfer step.

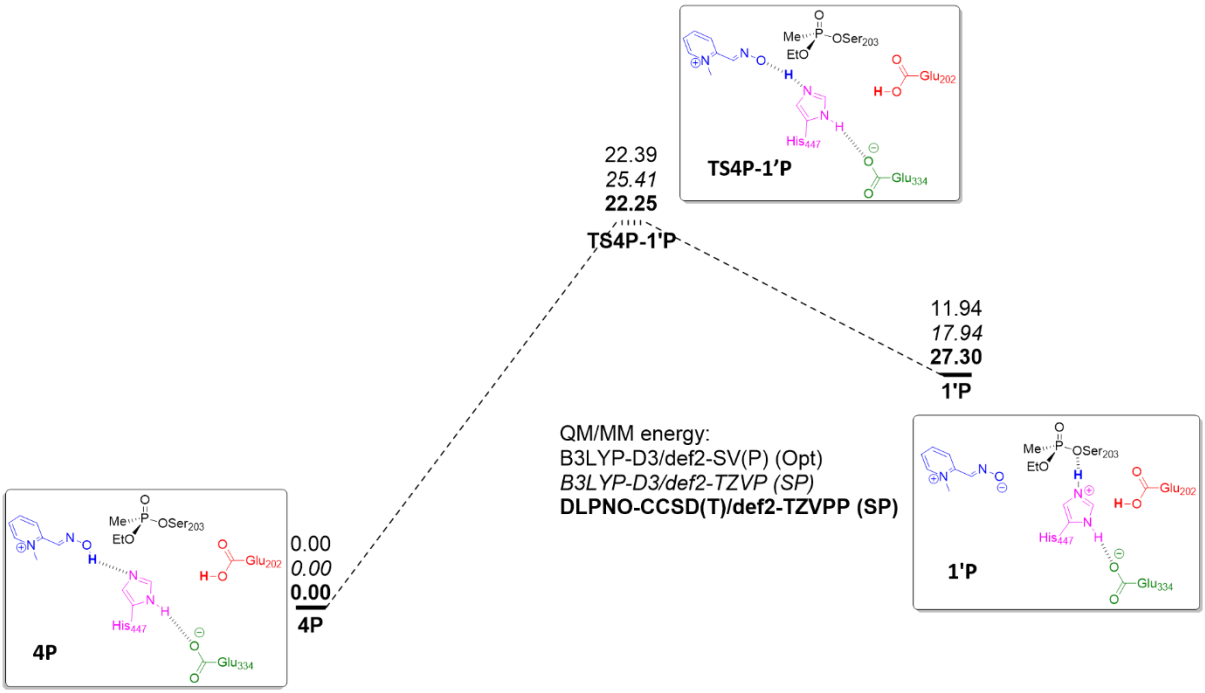


Figure VI-2. Energy profile for the proton transfer from 2-PAM to His447 when Glu202 is protonated. Energies in kcal.mol^{-1}.

The deprotonation of 2-PAM by His447 in the case of protonated Glu202 is endothermic with the energy difference between the reactant (**4P**) and the product (**1P**) of 11.94 kcal.mol^{-1} in B3LYP-D3/def2-SV(P). This deprotonation is a one-step reaction with a 22.39 kcal.mol^{-1} energy barrier in B3LYP-D3/def2-SV(P) (**TS4P-1'P**). At the B3LYP-D3/def2-TZVP level, the energy barrier is 25.41 kcal.mol^{-1} and the energy difference is 17.94 kcal.mol^{-1}. With DLPNO-CCSD(T)/def2-TZVPP the reaction is also endothermic with and energy difference of 27.30 kcal.mol^{-1} with a very high energy barrier of 22.25 kcal.mol^{-1}. Once again the difference in energy profiles from DLPNO-CCSD(T) come from geometrical differences between minima and saddle points of the potential energy surface corresponding to either calculation level. In any case, the DLPNO-CCSD(T) calculation does confirm that the energetic cost of 2-PAM deprotonation by His447 is much higher when Glu202 is protonated as opposed to unprotonated.

VI.3. Case of protonated His447 and unprotonated Glu202

The third possibility for the deprotonation of 2-PAM is the reactivator in oxime form entering the active site of VX-inhibited AChE with His447 protonated and Glu202 unprotonated. **Figure VI-3** summarizes the energy profile for this case.

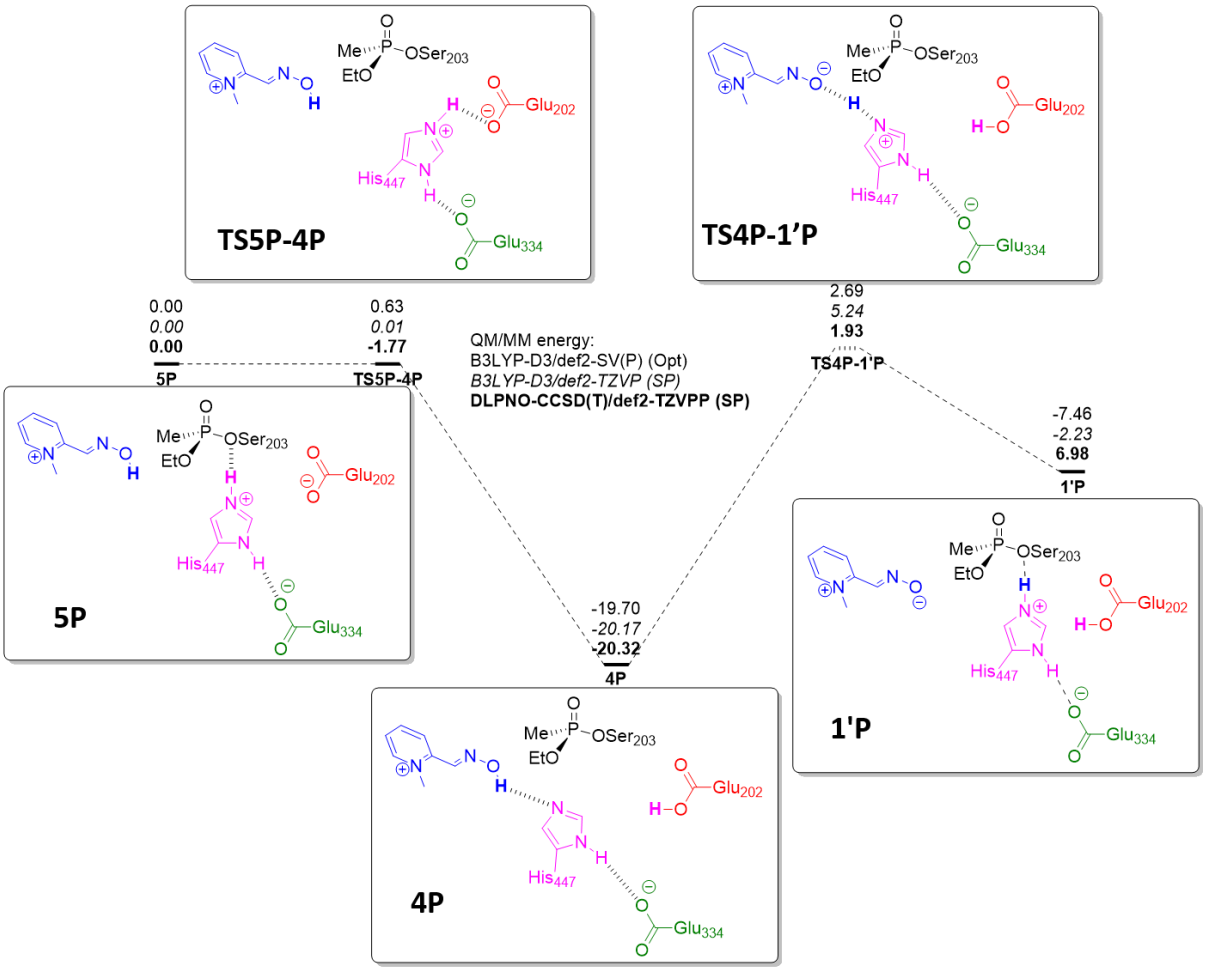


Figure VI-3. Energy profile for the proton transfer from His447 to Glu202 and from 2-PAM to His447. Energies in kcal.mol^{-1}.

For His447 to be able to deprotonate 2-PAM it first needs to have an available protonation site. Glu202, deprotonated at the beginning of this simulation, can deprotonate His447 to allow it to then active the oxime. This proton transfer step between His447 and Glu202 is highly exothermic with and energy difference of -19.70 kcal.mol^{-1} in B3LYP-D3/def2-SV(P) (-20.17 kcal.mol^{-1} in B3LYP-D3/def2-TZVP and -20.32 kcal.mol^{-1} in DLPNO-CCSD(T)/def2-TZVPP) and an energy barrier of 0.63 kcal.mol^{-1} in B3LYP-D3/def2-SV(P) (0.01 kcal.mol^{-1} in B3LYP-D3/def2-TZVP and -1.77 kcal.mol^{-1} in DLPNO-CCSD(T)/def2-TZVPP).

Once again, the DLPNO-CCSD(T) single point gives an energy profile indicative of a difference in reactant and saddle point geometries between the calculation levels. The DLPNO-CCSD(T), however, confirms the general trend with an extremely low energy barrier and a high exothermicity for the step. The following step is the deprotonation of 2-PAM by His447 with a protonated Glu202 as described in the previous subsection, subsection **VI.2**.

VII. Discussion

First and foremost, the trends obtained from QM/MM at the B3LYP-D3/def2-TZVP level are constantly in line with the trends obtained with high level ab-initio DLPNO-

CCSD(T)/def2-TZVPP. It asserts the pertinence of our DFT approach for the study of this system.

The data presented in this chapter are in partial agreement with previous QM/MM simulations of the reactivation of AChE on the nature of the reactivation mechanism.[23] While it was not possible to localize transition states for the addition step of the reactivation of VX-inhibited AChE, which has been localized in the literature by Li and co-workers, both this work and theirs point towards a kinetically determinant elimination step.[24] The addition-elimination mechanism is confirmed when Glu202 is protonated. However when Glu202 is unprotonated, which is also the case of the simulations of Li and co-workers, the mechanism could only be characterized as addition-elimination-like. This difference could be attributed to a higher energy barrier for the addition of the oxime on tabun, known to be sterically hindered,[25] compared to VX, used in our simulations. The effect of this stetric hindrance is potentially interesting to investigate further.

The truncated QM calculations presented in **Chapter 1** seemed to indicate that Glu334 could prevent the reactivation if its charge was not properly compensated. As the occurrence of reactivation is an undeniable experimental fact, it was concluded that either the role of Glu334 in the enzyme is different or, more likely, the enzymatic environment compensates the charge of this glutamate to keep it in carboxylate form and allow it to electrostatically stabilize His447 when it is protonated. A brief investigation of the protonation state of Glu334 depending on simulation conditions showed that this role in the enzyme is fulfilled by Ser229 and the network of hydrogen bonds connected to it. In order to connect Glu334 to this network of hydrogen bonds, Ser229 should rotate as to form and hydrogen bond with the glutamate. In all subsequent simulations, Glu334, as expected from experimental[5] and computational[26] evidence, remained unprotonated throughout, forming a very stable hydrogen bond with the δ protonation site of His447.

The protonation state of 2-PAM, Glu202 and His447 was shown to have an influence on the reactivation of AChE to be key parameters. As can be seen in the results presented over the previous sections, the protonation state of Glu202 has an effect on both the reactivation and the ability of 2-PAM to lose its proton in the active site of AChE. This effect of Glu202 is likely to be linked to the increased stabilization of high occupied molecular orbitals of the system when Glu202 is protonated. The protonation state of His447 has also been shown to be a key element of both the reactivation and the normal enzymatic activity of AChE. Finally, 2-PAM can only reactivate AChE when it is in its oximate form and thus must lose a proton at some point of the reactivation process. As the protonation state of those three elements has

[23] a) K. S. Matos, D. T. Mancini, E. F. da Cunha, K. Kuča, T. C. França, T. C. Ramalho, *J. Braz. Chem. Soc.* **2011**, *22*, 1999–2004. b) K. S. Matos, E. F. F. da Cunha, A. da Silva Gonalves, A. Wilter, K. Kuča, T. C. C. França, T. C. Ramalho, *J. Biomol. Struct. Dyn.* **2012**, *30*, 546–558.

[24] Y. Li, L. Du, Y. Hu, X. Sun, J. Hu, *Can. J. Chem.* **2012**, *90*, 376–383.

[25] O. Kwasnieski, Etude Théorique de La Réactivation de l'AChE Inhibée Par Le Tabun, UPMC, **2010**.

[26] a) J. Kua, Y. Zhang, J. A. McCammon, *J. Am. Chem. Soc.* **2002**, *124*, 8260–8267. b) V. V. Vasilyev, *J. Mol. Struct. THEOCHEM* **1994**, *304*, 129–141. c) Y. Zhang, J. Kua, J. A. McCammon, *J. Am. Chem. Soc.* **2002**, *124*, 10572–10577.

been shown to be crucial, their role in the reactivation process should be analysed as a whole. To that effect, all possible protonation state combinations for those three elements are enumerated in **Figure VII-1**.

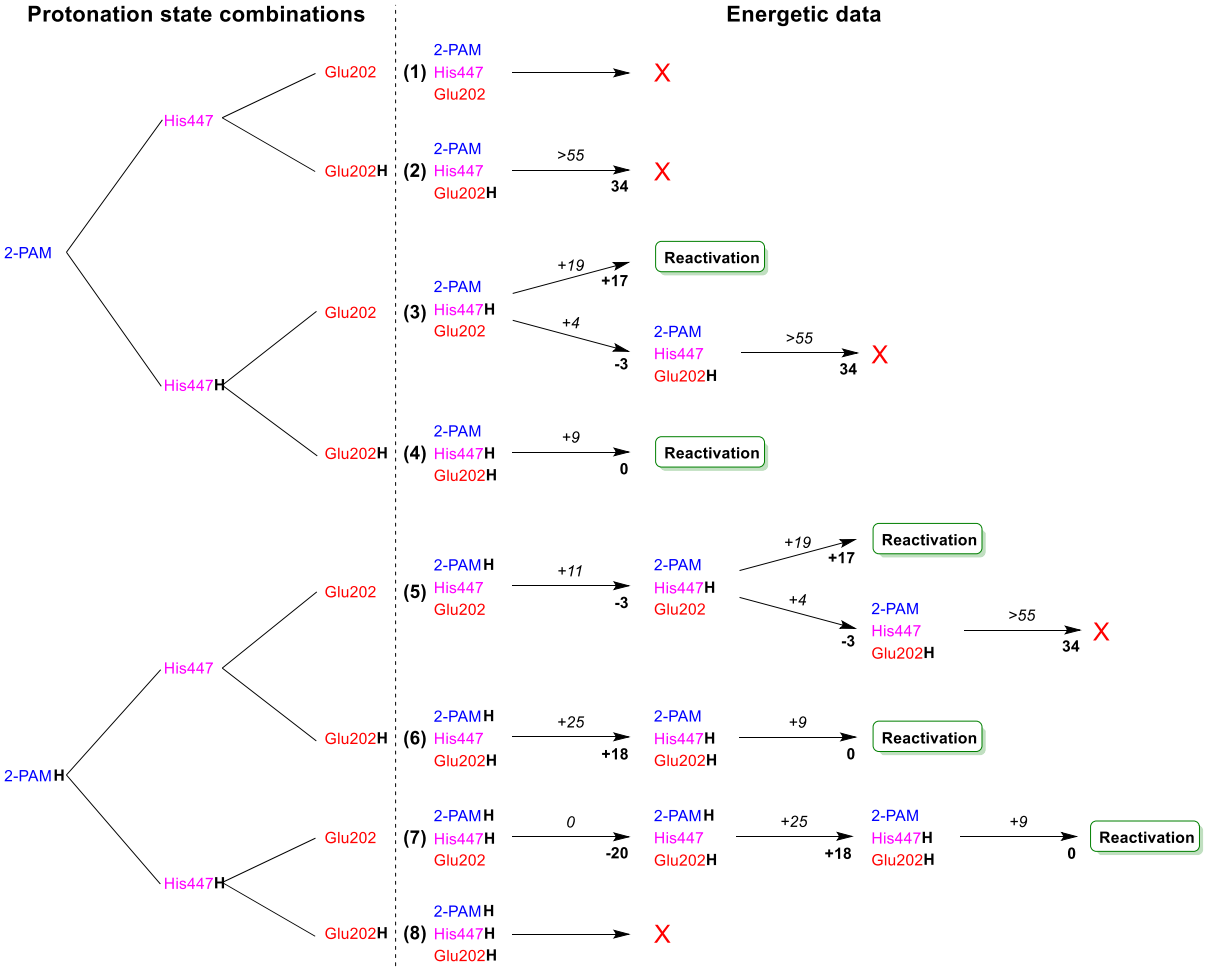


Figure VII-1. Tree of protonation states of 2-PAM, His447 and Glu202. The reactions are represented by an arrow with the reaction enthalpy at the bottom right and the energy barrier at the top in the middle. The energies are given at the B3LYP-D3/def2-TZVP:CHARMM level in kcal.mol^{-1} and rounded to the closest integer. At the end of every branch it is indicated whether reactivation occurs or not.

Figure VII-1 is presented like a tree of possibilities and it will be discussed branch by branch. Two branches have protonation states that completely exclude the possibility of reactivation. In **Branch 1** neither 2-PAM nor His447 nor Glu202 are protonated, which makes reactivation impossible because there is no proton in the system to protonate Ser203 after the elimination of the oxime-VX adduct. In **Branch 8**, all three are protonated. 2-PAM cannot lose its proton to His447 or Glu202 and be in its active oximate form. Thus, it is impossible for 2-PAM to reactivate VX-inhibited AChE in this Branch

When only Glu202 is protonated (**Branch 2**) reactivation is impossible. The high endothermicity of 50.14 kcal.mol^{-1} at the B3LYP-D3/def2-TZVP:CHARMM level and the energy barrier, which is equal or greater to 50.14 kcal.mol^{-1}, prevent it. It appears that without a proton on His447 for Ser203 to capture, the reactivation appears to be impossible.

For **Branch 3**, when only His447 carries a proton, two reactions are possible. The first is the reactivation of AChE by 2-PAM with the mechanism presented in subsection **V.1**. The second reaction is the proton transfer from His447 to Glu202 (see subsection **VI.1**). This second reaction gets the system to a state identical to the starting point of **Branch 2** and does not lead to reactivation. The first reaction is endothermic with a 17.00 kcal.mol^{-1} and a high barrier of 18.90 kcal.mol^{-1}; the second is exothermic with a 3.18 kcal.mol^{-1} energy gain and an energy barrier of 4.01 kcal.mol^{-1}. The exothermicity and low barrier of the second reaction makes the deprotonation of His447 by Glu202 much more favourable than the reactivation. After this deprotonation, the reaction is at an energetic dead-end and thus, **Branch 3** cannot lead to reactivation.

When both Glu202 and His447 are protonated (**Branch 4**) reactivation is possible (see subsection **V.2**). It is isoenergetic (-0.19 kcal.mol^{-1}) and has a low energy barrier of 8.84 kcal.mol^{-1}.

At the start of **Branch 5** only 2-PAM carries a proton. The first reaction than occurs is the proton transfer form 2-PAM to His447. This reaction is slightly exothermic (3.22 kcal.mol^{-1}) with a low energy barrier of 11.46 kcal.mol^{-1}. After the proton transfer the situation is in the same step as the start of **Branch 3** which, as it was already said, cannot lead to reactivation and falls into an energetic dead-end.

When both 2-PAM and Glu202 are protonated (**Branch 6**) reactivation can only occurs if 2-PAM loses its proton to His447. This deprotonation is endothermic (17.94 kcal.mol^{-1}) with a barrier of 25.44 (kcal.mol^{-1}) (see subsection **VI.2**). Once 2-PAM is deprotonated the situation is similar to the start of **Branch 4** that does lead to the reaction in one isoenergetic step with a small barrier. This branch can lead to reactivation but the system must go through a very expensive deprotonation step before reactivation is possible.

Finally **Branch 7** starts with 2-PAM and His447 protonated. Two steps are necessary for reactivation: The deprotonation of His447 by Glu202 and the deprotonation of 2-PAM by His447. The first step is an exothermic reaction (20.17 kcal.mol^{-1}) with a non-existent energy barrier (0.01 kcal.mol^{-1}) (see subsection **VI.3**). The second step is the deprotonation of 2-PAM by His447 as described in **Branch 6** which lead to reactivation.

Apart from **Branch 1** and **8** in **Figure VII-1**, which can be directly eliminated, 6 branches remain to discuss. Three other branches, **Branches 2**, **3**, and **5**, can be eliminated as, when there is a single proton on either 2-PAM, Glu202 and His447, it very favourably ends up on Glu202 where it cannot participate in the reactivation. Without a proton on His447 the reactivation has an unsurmountable endothermicity. Therefore, only 3 protonation states can lead to reactivation, those corresponding to **Branches 4**, **6** and **7**. The common feature of these branches is the fact that two protons are available in the active site. Once these protons end-up on His447 and Glu202, which is to say once 2-PAM is deprotonated, the reactivation barrier is relatively low (8.84 kcal.mol^{-1} at the B3LYP-D3/def2-TZVP level). With such a low barrier, it appears that AChE reactivation by oximes is mainly dependent on the protonation state of residues in the active site rather than closely related to the nucleophilicity of the oximes.

The ideal case for the reactivation of VX-inhibited AChE is the entrance in the enzyme channel of a reactivator in oximate form, which then binds into the active site where both His447 and Glu202 are protonated. Since oximes are basic compounds, this conjunction of protonation states has low probability at normal physiological pH in the synaptic cleft. Even in the unlikely cases of encounters between an inhibited AChE with a reactivator in oximate form, molecular dynamics simulations suggested that the descent down the channel of AChE is less favourable for oximes in their oximate form than in their oxime form.[27]

In this ideal case, the two necessary protons were already present in the active site. When the oxime enters the active site while being protonated, thus bringing one of the two necessary protons. The other proton can thus be located either on His447 or on Glu202. Our calculations indicate that, once the oxime is in the active site, a proton can easily be transferred from His447 to Glu202. What remains is the deprotonation of the oxime that will exacerbate its nucleophilicity and allow for reactivation. This step was found to be energetically costly (barrier of 25.44 kcal.mol^{-1} and reaction enthalpy of 17.94 kcal.mol^{-1}). Consequently, our computational study indicates that the main problem during the reactivation of inhibited AChE by oximes is not related to the intrinsic nucleophilic strength of the reactivator but rather to its ability to be deprotonated in the active site by the couple His447/Glu202. A strong nucleophilicity is commonly accompanied by a strong basicity. Finding a good oxime reactivator is thus balancing between these two contradictory effects for the reactivation process.

The role of Glu202 during the regular catalytic activity of AChE has already been investigated by experimental[28] and theoretical[29] means, particularly during the regular catalytic activity of AChE. It was shown by Monte-Carlo titration on non-inhibited AChE that this glutamic acid is protonated, for all configurations of Ser203. Further EVB calculations also reveal that the kinetics of the deacylation step is in line with experiments when this glutamic acid is in its protonated form. On the other hand, many roles were attributed to unprotonated Glu202. For example, it was speculated that charged Glu202 (unprotonated) is able to stabilize the binding of ACh by long-range electrostatic interactions.[30] Also, since His447 is in its protonated state when Ser203 is acetylated, it was conjectured by the same authors that the negative charge of Glu202 would stabilize the positive charge of the doubly protonated imidazole. Some aging mechanisms hypothezised that Glu202 could act as a general base for water molecules[24] or a transient proton acceptor[31]. Recently, Linusson and Ekström have studied by diffusion trap cryo-crystallography the dynamics of the sarin-inhibited AChE active site in conjunction with HI-6 reactivator.[22c] Notably, they found an interaction between

[27] A. da S. Gonçalves, T. C. C. França, A. Wilter, J. D. Figueroa-Villar, *J. Braz. Chem. Soc.* **2006**, *17*, 968–975.

[28] a) G. Gibney, S. Camp, M. Dionne, K. MacPhee-Quigley, P. Taylor, *Proc. Nat. Ac. Sc. USA* **1990**, *87*, 7546–7550. b) Z. Radic, G. Gibney, S. Kawamoto, K. MacPhee-Quigley, C. Bongiorno, P. Taylor, *Biochemistry* **1992**, *31*, 9760–9767.

[29] a) P. Vagedes, B. Rabenstein, J. Åqvist, J. Marelius, E.-W. Knapp, *J. Am. Chem. Soc.* **2000**, *122*, 12254–12262 b) Y. Zhou, S. Wang, Y. Zhang, *J. Phys. Chem. B* **2010**, *114*, 8817–8825. c) S. V. Lushchekina, I. A. Kaliman, B. L. Grigorenko, A. V. Nemukhin, S. D. Varfolomeev, *Russ. Chem. Bull.* **2011**, *60*, 2196–2204.

[30] S. T. Wlodek, J. Antosiewicz, J. M. Briggs, *J. Am. Chem. Soc.* **1997**, *119*, 8159–8165.

[31] C. Viragh, R. Akhmetshin, I. M. Kovach, C. Broomfield, *Biochemistry* **1997**, *36*, 8243–8252

Glu202 and the O-isopropyl moiety of sarin. They also find evidence that a proton transfer occurs close to the rate-limiting step, by deuteration experiments. These findings are in line with our computational analysis. The important role of Glu202 explains the severely decreased reactivation bimolecular rate constant when Glu202 is mutated to a glutamine.[32] We notice that this residue is highly conserved over the esterase superfamily[33] and we can thus speculate that it is more important than previously thought. Its exact role remains nevertheless elusive.

VIII. Conclusion

This computational study reveals for the first time the importance of Glu202 for the reactivation of VX-inhibited AChE. It shows that both glutamates in the vicinity of Ser203 have a role in reactivation. The possibility for 2-PAM to enter the active site in either its oxime or oximate form and to reactivate the enzyme is confirmed. The specific protonation states of the active site that allow for the reactivation to take place have been investigated. The reactivation reaction can only occur if both His447 and Glu202 are protonated, but those two residues do not need to be already protonated when 2-PAM enters the active site. The confirmation of those results using a dynamic method that includes the entropic contribution and that can follow the specific motions and role of water molecules of the active site would be a good addition to this work. This part of the PhD project has been published in *ChemBioChem*.[34] To pursue this line of study of AChE's reactivation, two avenues are open. New reactivators can be studied and evaluated with similar simulations to confront both their reactivation power and ability to be deprotonated in the active site to those of 2-PAM. The diffusion of protons in and out of the active site and the dynamics of proton exchanges should also be investigated. Both avenues have been investigated and the results are presented over the next two chapters.

[32] Y. Ashani, Z. Radic, I. Tsigelny, D. C. Vellom, N. A. Pickering, D. M. Quinn, B. P. Doctor, P. Taylor, *J. Biol. Chem.* **1995**, *270*, 6370–6380.

[33] a) J. C.-H. Chen, L. J. W. Miercke, J. Krucinski, J. R. Starr, G. Saenz, X. Wang, C. A. Spilburg, L. G. Lange, J. L. Ellsworth, R. M. Stroud, *Biochemistry (Mosc.)* **1998**, *37*, 5107–5117. b) V. A. de Souza, D. J. Scott, J. E. Nettleship, N. Rahman, M. H. Charlton, M. A. Walsh, R. J. Owens, *PLoS One* **2015**, *10*, e0143919. c) B. Spiller, A. Gershenson, F. H. Arnold, R. C. Stevens, *Proc. Nat. Ac. Sc. USA* **1999**, *96*, 12305–12310.

[34] T. Driant, F. Nachon, C. Ollivier, P.-Y. Renard, E. Derat, *ChemBioChem* **2017**, *18*, 666–675.

Chapter 5: Simulations of Non-Pyridinium Reactivators

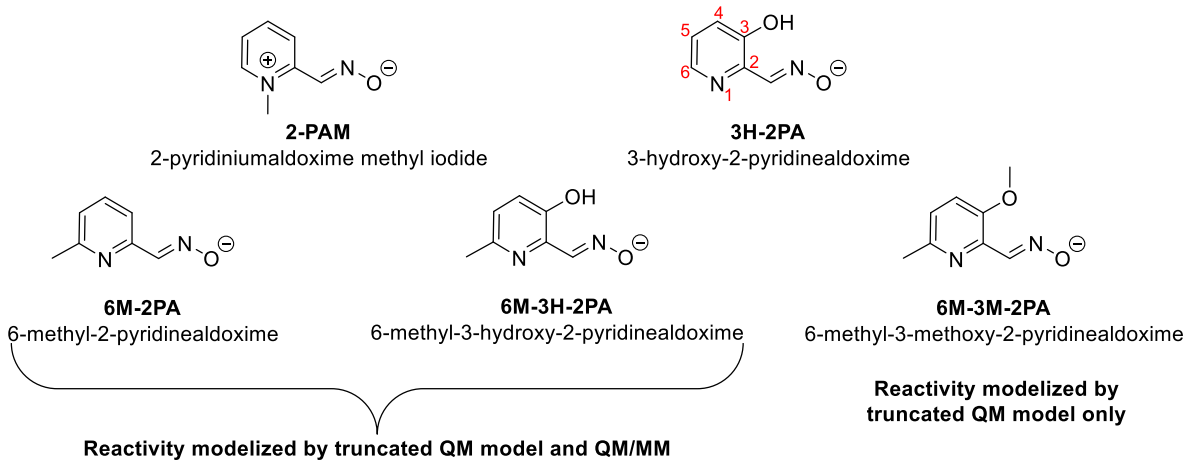
Chapter 5: Simulations of Non-Pyridinium Reactivators

I. Introduction

In the previous chapters, simulations of the deprotonation 2-PAM, the prototype of classical oxime reactivators, and the subsequent reactivation of VX-inhibited AChE by 2-PAM. In this chapter, a new type of reactivator prototypes, with an uncharged pyridine aromatic ring instead of the positively charged methyl pyridinium aromatic ring of 2-PAM (see **Scheme I-1**), is evaluated. In the manuscript, the methyl pyridinium is often called pyridinium for the sake concision.

Scheme I-1. Structure of pyridine and pyridinium aromatic rings.

The reactivator prototype 3-hydroxy-2-pyridinealdoxime (3H-2PA), as well as a number of reactivators derived from this prototype have been synthetized and experimentally investigated and it was shown that they have great reactivation potential but poor affinity.[1]



Scheme I-2. Non pyridinium reactivators

[1] a) L. Louise-Leriche, E. Păunescu, G. Saint-André, R. Baati, A. Romieu, A. Wagner, P.-Y. Renard, *Chem. Eur. J.* **2010**, *16*, 3510–3523. b) G. Saint-André, M. Kliachyna, S. Kodepelly, L. Louise-Leriche, E. Gillon, P.-Y. Renard, F. Nachon, R. Baati, A. Wagner, *Tetrahedron* **2011**, *67*, 6352–6361. c) J. Renou, G. Mercey, T. Verdelet, E. Păunescu, E. Gillon, M. Arboléas, M. Loiodice, M. Kliachyna, R. Baati, F. Nachon, et al., *Chem. Biol. Interact.* **2013**, *203*, 81–84. d) G. Mercey, T. Verdelet, G. Saint-André, E. Gillon, A. Wagner, R. Baati, L. Jean, F. Nachon, P.-Y. Renard, *Chem. Commun.* **2011**, *47*, 5295–5297. e) G. Mercey, J. Renou, T. Verdelet, M. Kliachyna, R. Baati, E. Gillon, M. Arboléas, M. Loiodice, F. Nachon, L. Jean, et al., *J. Med. Chem.* **2012**, *55*, 10791–10795. f) J. Renou, M. Loiodice, M. Arboléas, R. Baati, L. Jean, F. Nachon, P.-Y. Renard, *Chem. Comm.* **2014**, *50*, 3947. g) M. Kliachyna, G. Santoni, V. Nussbaum, J. Renou, B. Sanson, J.-P. Colletier, M. Arboléas, M. Loiodice, M. Weik, L. Jean, et al., *Eur. J. Med. Chem.* **2014**, *78*, 455–467. h) J. Renou, J. Dias, G. Mercey, T. Verdelet, C. Rousseau, A.-J. Gastellier, M. Arboléas, M. Touvrey-Loiodice, R. Baati, L. Jean, et al., *RSC Adv.* **2016**, *6*, 17929–17940.

In **Scheme I-2** the 2-PAM and 3H-2PA reactivator prototypes that have been experimentally evaluated are presented on the top row. In both those reactivators, the oxime is in position 2 on the aromatic ring. The positions are numbered for clarity of the structure of 3H-2PA. On the second row the reactivators that are evaluated in this chapter are presented. 6M-2PA is a neutral analogue of 2-PAM that retains an aromatic cycle with a nitrogen and a methyl substituent although displaced to an adjacent position on the pyridine ring. This methyl substituent also models the influence of the alkyl chain connecting this reactivator prototype to a ligand. This analog will allow to evaluate the advantages of a neutral pyridine cycle without the added hydroxy group. The other two reactivator prototype s, 6M-3H-2PA and 6M-3M-2PA, are evaluated to assess the benefits of an electron donating group on the reactivator prototype able to establish a hydrogen bond or not with the oxime reactivator.

This evaluation has been carried out with the same methods used for 2-PAM. Truncated QM model simulations were first employed to have a preliminary evaluation of the reactivation power of those three reactivator prototypes. QM/MM simulations have then been performed to evaluate to ability of these reactivators to be deprotonated and then to reactivate VX-inhibited AChE. Only 6M-2PA and 6M-3H-2PA have been evaluated using QM/MM due to a lack of time to perform the corresponding simulations for 6M-3M-2PA.

II. Truncated model QM study of the reactivation

The study of the reactivation with reactivators 6M-2PA, 6M-3H-2PA and 6M-3M-2PA in a truncated QM model was performed with similar models as the study on 2-PAM. The minimal model is composed of an organophosphate analogue of the Ser203-VX adduct, $P(O)(OEt)_2CH_3$, with an imidazolium aromatic cycle below the oxygen of the ethoxy group opposition to 2-PAM which stands for the side chain of His447 (see **Figure II-1**). To this minimal model, an equivalent of the oxyanionic hole are added. Then the model was further completed by an analogue of Glu334.

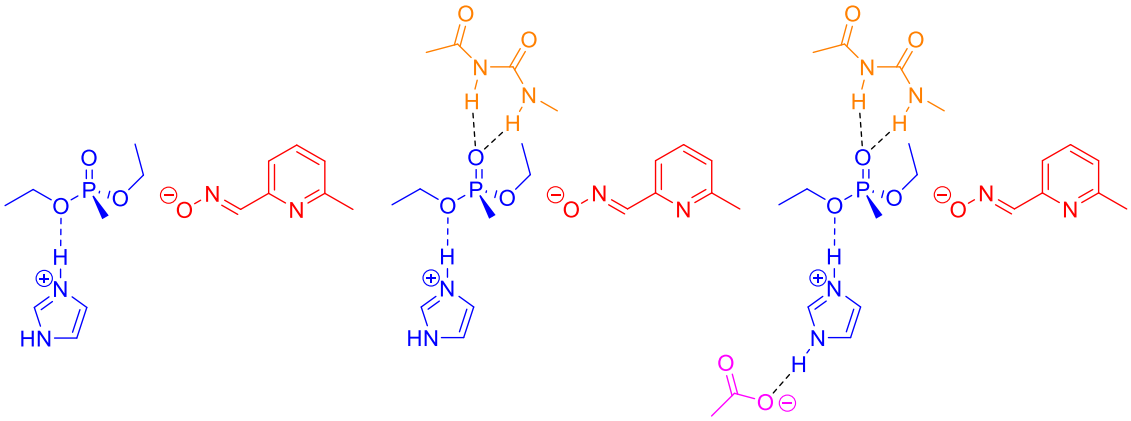
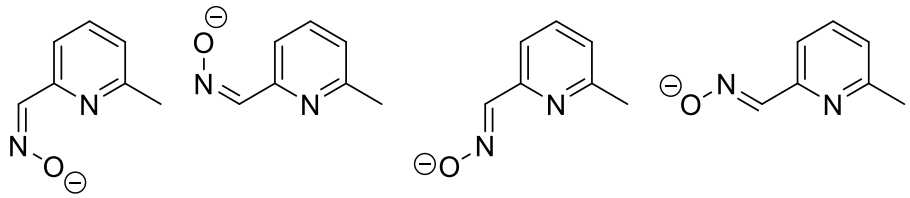


Figure II-1. Truncated models used to study the reactivation of a VX-inhibited AChE by non-pyridinium reactivators.

The simulations were performed in gas phase with the B3YLP functional,[2] the D3 dispersion correction,[3] and the def2-SV(P) basis set.[4] The calculations were performed using Turbomole V6.5.[5] The transition states for these reaction were not investigated as the optimization of the reactant structures proved very challenging and the optimization of the transition states promised to be as difficult.

II.1. 6M-2PA

For the reactivator 6M-2PA, a first step in the investigation of the most stable geometries. Like 2-PAM, because of the double bond in the oxime functional group, 6M-2PA can be in either *Z* or *E* configuration. The different rotamers possible for these stereoisomers are also explored.



Isomer	*ZZ*	*ZE*	*EZ*	*EE*
Energy (kcal.mol^{-1})	+12.14	+1.82	+5.53	0.00

Figure II-2. Possible geometries for 6M-2PA with their configuration and relative energy shown below.

The most stable conformation for 6M-2PA is, similarly to 2-PAM, the configuration with an *E* C-N double in the oxime and an *E* rotamer for the dihedral N-C-C-N along the C-C bond that connects the oxime to the pyridine ring. This configuration of 6M-2PA is the one used in the truncated QM model simulations presented in this section. The other configuration give rise to steric clash between either the oxygen and the pyridine ring or the two nitrogen from the oxime and the pyridine ring, or even both for the *ZZ* isomer.

First the reactivation power of 6M-2PA is investigated in the minimal model. Attempts were made to optimize the reactant but the reactivation spontaneously occurs and the product state is optimized (see **Figure II-3B**). This spontaneous reactivity could be due to a poor starting product for the optimization of the reactant or a low energy barrier and a high exothermicity for the reaction. A mono-dimensional scan with a constraint on the distance between the phosphorus and the oxygen of the oxime (P-Ox) was used to gradually move the reactivator away from the phosphorus which then reacts back on the serine. This scan allowed to successfully optimize a reactant structure (**Figure II-3A**).

[2] a) A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652. b) C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.
[3] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
[4] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
[5] R. Ahlrichs, F. Furche, C. Hättig, W. Klopper, M. Sierka, F. Weigend, *TURBOMOLE*, TURBOMOLE GmbH, Development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, **2012**.

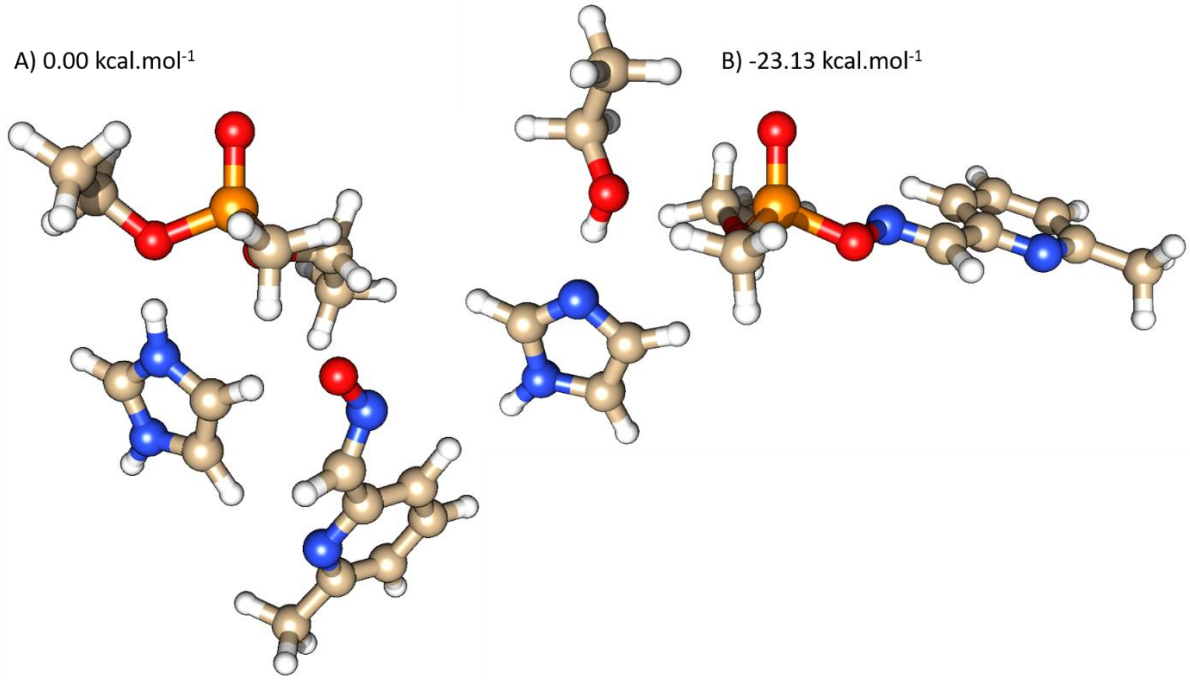


Figure II-3. Optimized structures of (A) the reactant and (B) the product for the reactivation by 6M-2PA in the minimal model.

The structure of the reactant shows an unrealistic position for the reactivator which slides parallel to the imidazole representing His447. In the enzyme, this position is not accessible to the reactivator. The increased stability due to the interactions between the imidazole and the pyridine of 6M-2PA, present in the reactant and absent from the product, adds a layer of complexity to the interpretation. The product is -23.13 kcal.mol^{-1} more stable than the reactant in this minimal model. The transition state was not investigated because the apparent spontaneous evolution of the reaction indicates that it is likely to be very small and thus difficult to locate. The difficulties encountered to optimize a product structure confirmed the sentiment that the search for a TS would be too time consuming to undertake. Throughout the study, no transition state was investigated.

The optimization of the reactant for the minimal model with the addition of the oxyanionic hole proved to be as difficult as in the minimal model. The optimization attempts for the reactant resulted in spontaneous reactivations and in the end the product structure was obtained (see **Figure II-4B**). A mono-dimensional scan along the P-Ox bond was performed to move the reactivator away from the phosphorus. From this scan, a reactant structure was successfully optimized (**Figure II-4A**).

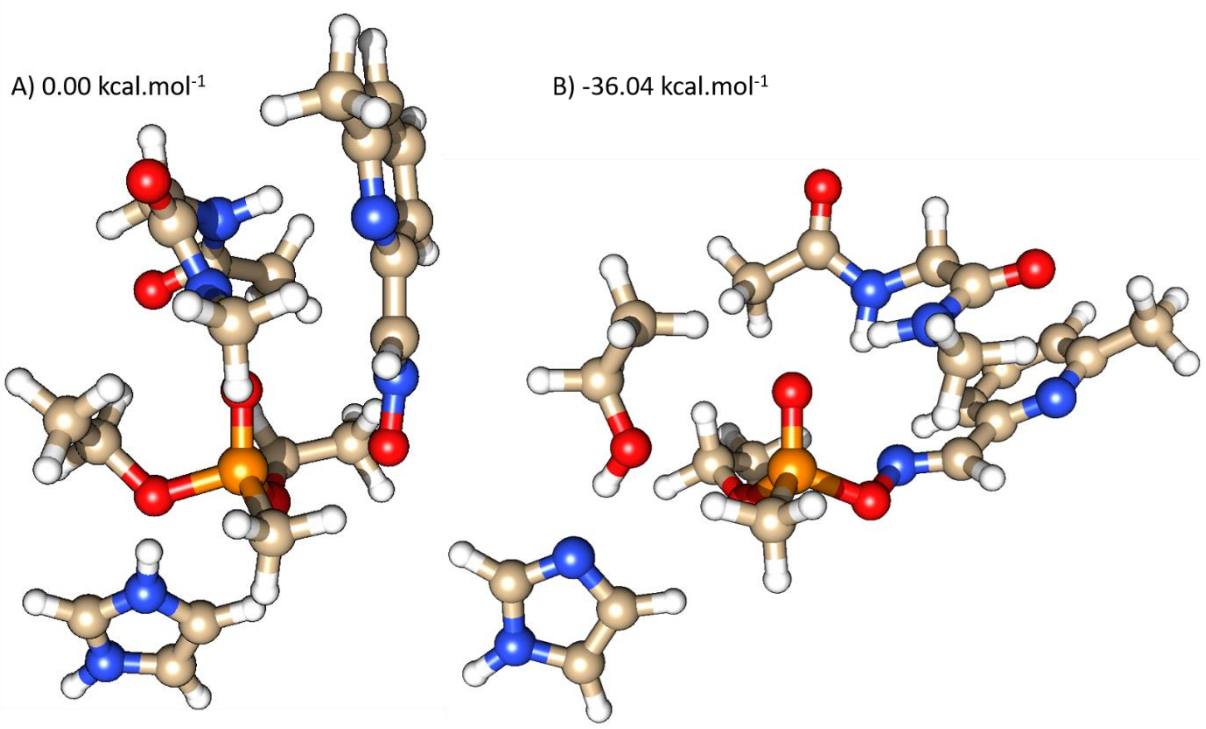


Figure II-4. Optimized structures of (A) the reactant and (B) the product for the reactivation by 6M-2PA in the model with the oxyanionic hole.

The reactant structure for the minimal model with the oxyanionic hole has an unrealistic position for the reactivator. 6M-2PA moves up and is positioned parallel to the oxyanionic hole. This position is not available for the reactivator in the enzyme. The added stabilization, absent from the structure of the product, prevents to attribute the energy difference between reactant and product exclusively to the reactivity. This energy difference is -36.04 kcal.mol^{-1}. Again, the transition state was not investigated as with all the transition states for all model for either of the three reactivator prototypes. This decision to not investigate the TS was made mainly due to some of the geometry of the reactant where the reactivator is in a position available in the truncated model but not available in the enzyme.

The addition of an acetate to the truncated model, to investigate the effect of Glu334 on the reactivation by 6M-2PA, completely changes the reactivity of the system. The attempt at optimizing a reactant directly evolve into a stable pentavalent intermediate (**Figure II-5C**). short scans on the P-Ox and P-OSer203 distances can be used to locate and then fully optimize reactant and product structures (**Figure II-5A** and **Figure II-5B**). The presence of a stable intermediate indicates that the mechanism of reactivation in this case is an addition-elimination.

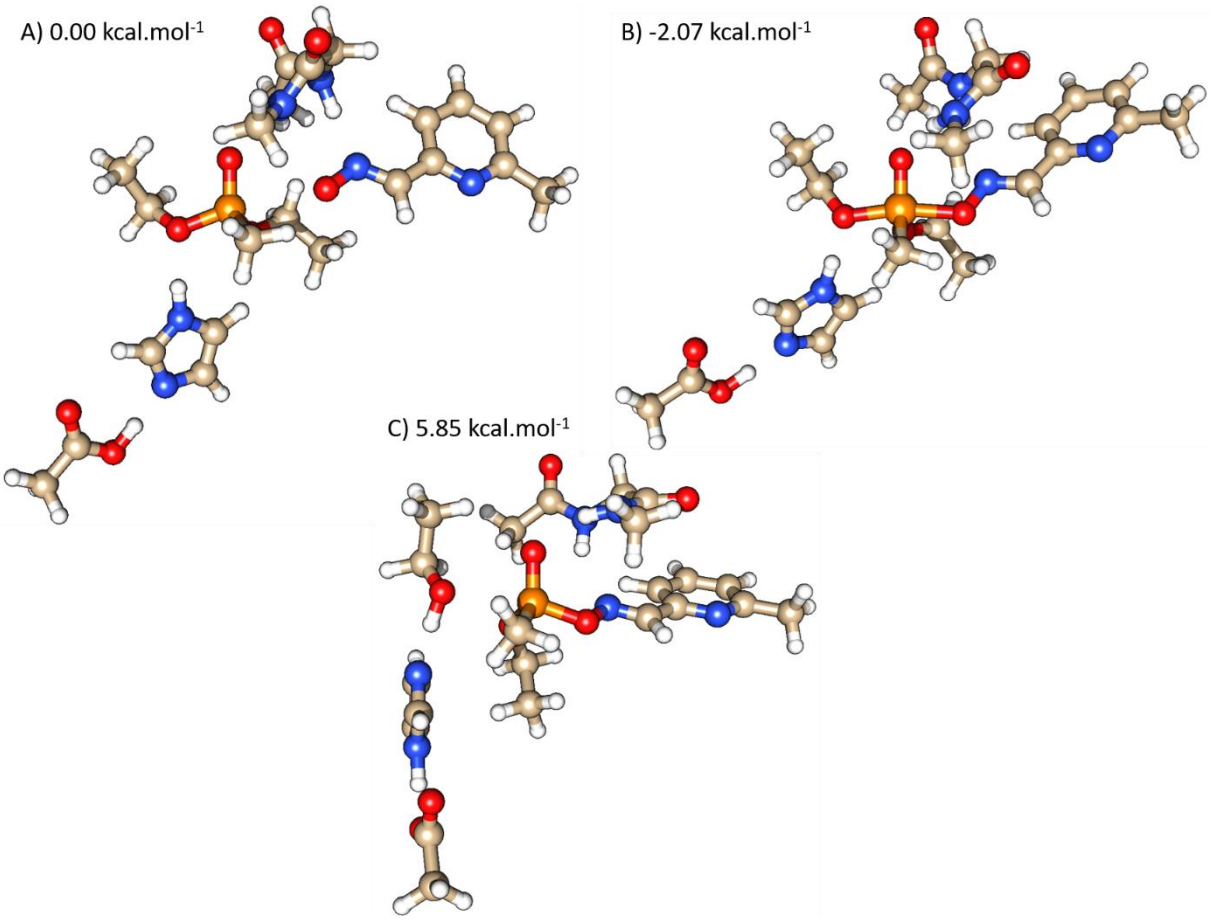
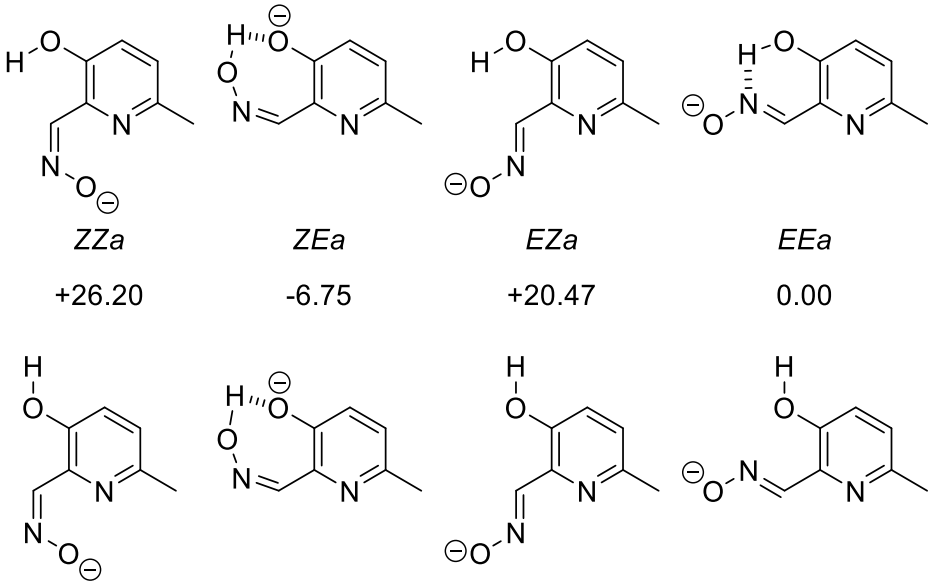


Figure II-5. Optimized structures of (A) the reactant, (B) the stable pentavalent intermediate, and (C) the product for the reactivation by 6M-2PA in the model with the oxyanionic hole and the acetate.

The pentavalent intermediate is formed in an exothermic nucleophilic addition step with an energy difference of -2.07 kcal.mol^{-1}. This reaction is then followed by the elimination step with an energy difference of 7.92 kcal.mol^{-1}. The overall energy difference of the reaction is 5.85 kcal.mol^{-1}. The added acetate stabilizes the reactant and completely negates the exothermicity of the process as previously found in **Chapter 3**.

II.2. 6M-3H-2PA

The most stable geometries for reactivator 6M-3H-2PA has been investigated. Like 6M-2PA, because of the double bond in the oxime functional group, 6M-3H-2PA can be in either *Z* or *E* configuration. The different possible rotamers for the C-C bond that connects the oxime and the pyridine aromatic ring are also explored. Finally, the rotation of the hydroxy group in position 3 on the pyridine aromatic ring is also evaluated (see **Figure II-6**). This rotation of the hydroxy group is noted *a* when the hydrogen is oriented towards the oxime and noted *b* when the hydrogen is oriented away from the oxime.



Isomer	*ZZa*	*ZEa*	*EZa*	*EEa*
Energy (kcal.mol^{-1})	+26.20	-6.75	+20.47	0.00

Isomer	*ZZb*	*ZEb*	*EZb*	*EEb*
Energy (kcal.mol^{-1})	+32.09	-6.70	+25.61	+26.01

Figure II-6. Possible geometries for 6M-3H-2PA with their configuration and relative energy shown below.

The most stable conformation of 6M-3H-2PA has the same *EEa* configuration as the most stable configuration of 6M-2PA. This conformation minimizes steric strain between the oxime and the pyridine ring of 6M-3H-2PA and between the oxime and the hydroxy group. The interactions between the hydroxy and the oxime are in fact optimized in that conformation where a short and stable hydrogen bond is formed between the hydroxy and the nitrogen of the oxime functional group. In both *ZEa* and *ZEb* conformations, the oxime deprotonated the hydroxy group producing near identical geometries. The *EEa* conformation is the one used throughout the study of reactivation by 6M-3H-2PA.

The reactivation power of 6M-3H-2PA was first investigated in the minimal model. The attempts to minimize the reactant failed and spontaneously evolved into the product state of the reaction. Even careful mono-dimensional scan could not produce a stable reactant but only aberrant states like the state shown in **Figure II-7A**. The deprotonation of the pyridine ring by the imidazole is not reported anywhere in the literature on the non-pyridinium reactivators of AChE. The product however was very easy to optimize as most starting geometries will optimize directly to the structure of the product (**Figure II-7B**). With only the product, not much can be said of this reaction energetically, only that it is highly exothermic and probably has a small energy barrier.

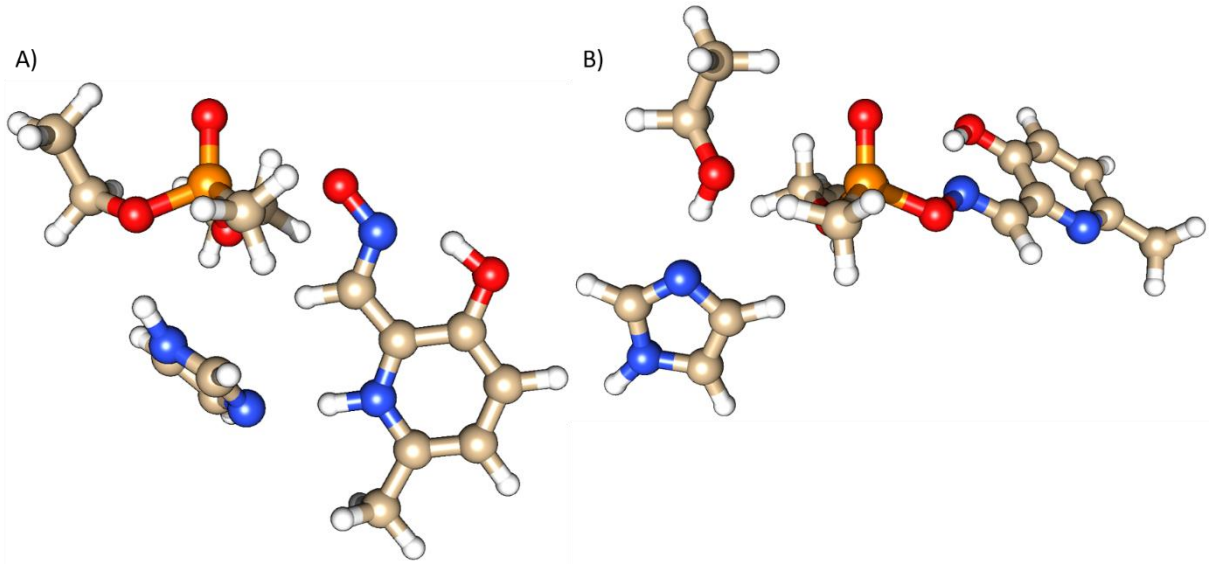


Figure II-7. Optimized structures of (A) the reactant and (B) the product for the reactivation by 6M-3H-2PA in the minimal model.

The model can be expanded by adding an oxyanionic hole analogue to the minimal model. This addition however, do not allow for the optimization of the reactant which spontaneously evolves towards to product (**Figure II-8B**). Like in the minimal model, attempts to scan the P-Ox distance to optimize the reactant resulted in aberrant structures (**Figure II-8A**). In this case, the oxyanionic hole flips to stabilize the deprotonation of the hydroxy group in favour of the nitrogen of the oxime functional group. The product again is easily optimized as most states of the system spontaneously evolve into the product (**Figure II-8B**). The product structure alone yields little energetic information but once again the reactivation seems to be very exothermic and have a low energy barrier.

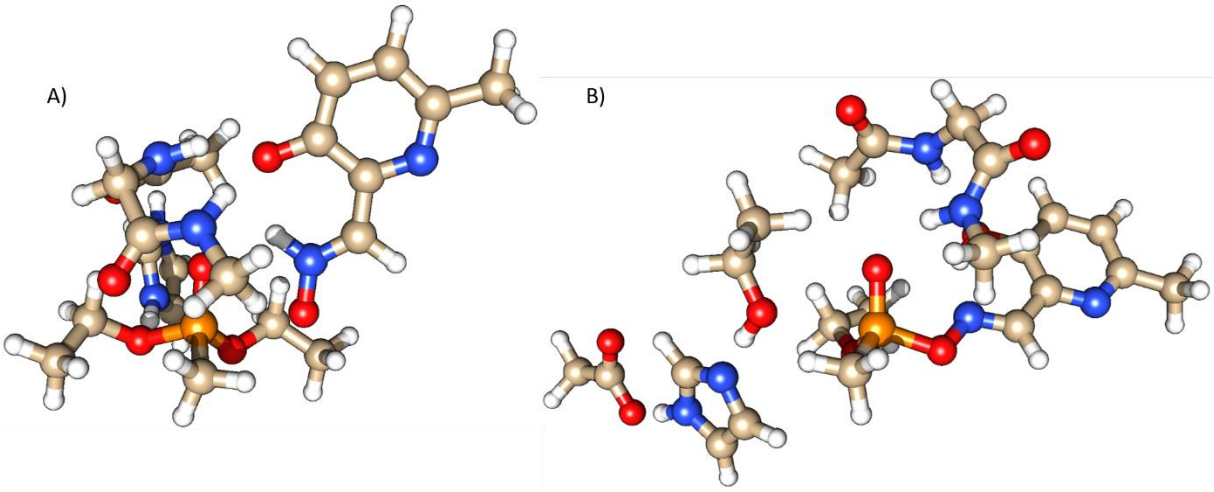


Figure II-8. Optimized structures of (A) the reactant and (B) the product for the reactivation by 6M-3H-2PA in the model with the oxyanionic hole.

The model is further expanded by adding an acetate near the second protonation site of the imidazole. The larger model allows the optimization of both reactant and product (**Figure II-9A** and **Figure II-9B**).

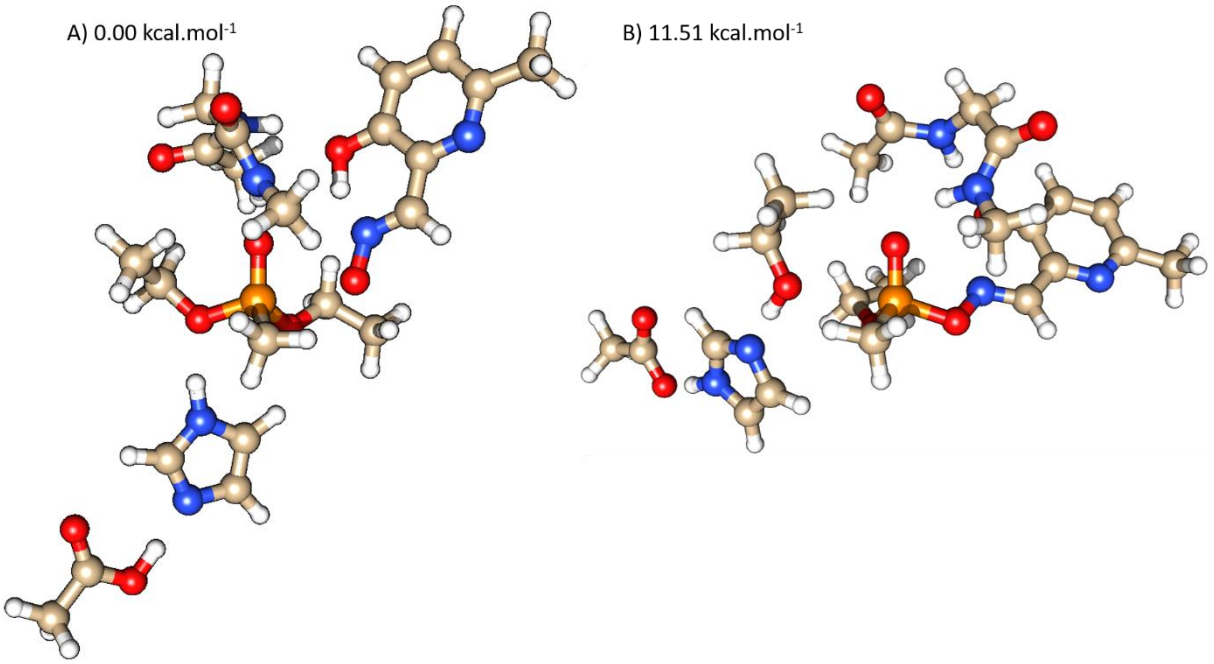


Figure II-9. Optimized structures of (A) the reactant and (B) the product for the reactivation by 6M-3H-2PA in the model with the oxyanionic hole and the acetate.

The energy difference between the reactant and the product in the model with both the oxyanionic hole and the acetate is 11.51 kcal.mol^{-1}. The third non-pyridinium reactor tested in this truncated QM model study was 6M-3M-2PA.

II.3. 6M-3M-2PA

6M-3M-2PA has a methoxy group instead of the hydroxy group in position 3 on the pyridine ring. Once more the possible conformations of this reactivator were explored. The explored conformations involve rotation of the C-N bond of the oxime, rotation of the C-C bond connecting the oxime to the pyridinium, and the rotations along the C-O bond of the methoxy group. The rotamers of the C-O bond of the methoxy group are *a*, *b*, and *c*. Rotamer *a* has the methoxy oriented towards the oxime, in the plane of the pyridine ring. Rotamer *b* has the methoxy oriented away from the oxime, in the plane of the pyridine ring. Rotamer *c* has the methoxy out of the plane of the pyridine ring away from the oxime. The results are presented in **Figure II-10**. For all four combinations of C-N stereoisomers and C-C rotamers, the most favourable position of the methoxy group is out of the plane of the pyridine ring (*ZZc*, *ZEc*, *EZc* and *EEc*). Among the conformations that do have an out of plane methoxy group, the most stable one is the *EEc* conformation. It minimizes steric clash with both the pyridine and the methoxy group. As such, this conformation of 6M-3M-2PA is the one used throughout this study.

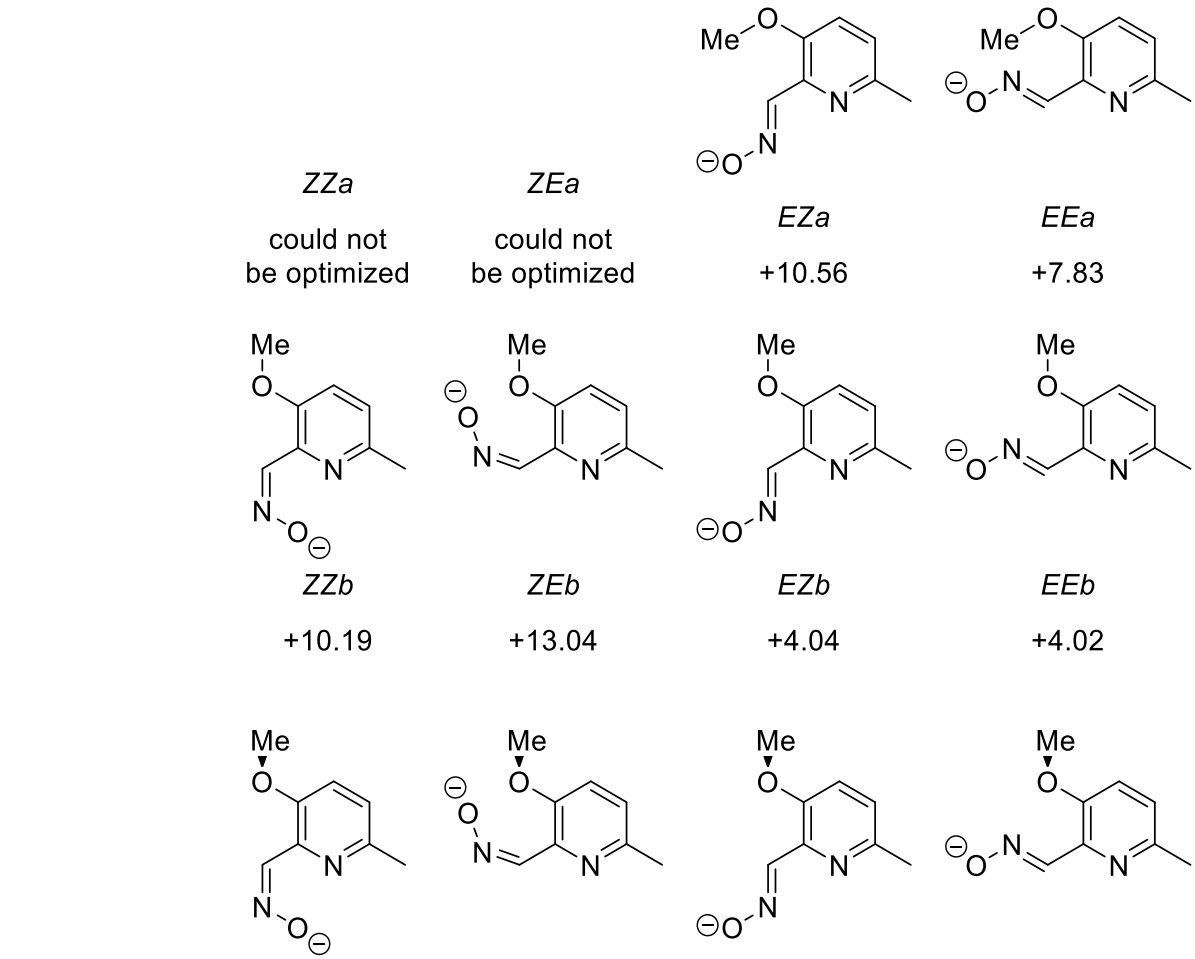


Figure II-10. Possible geometries for 6M-3M-2PA with their configuration and relative energy shown below.

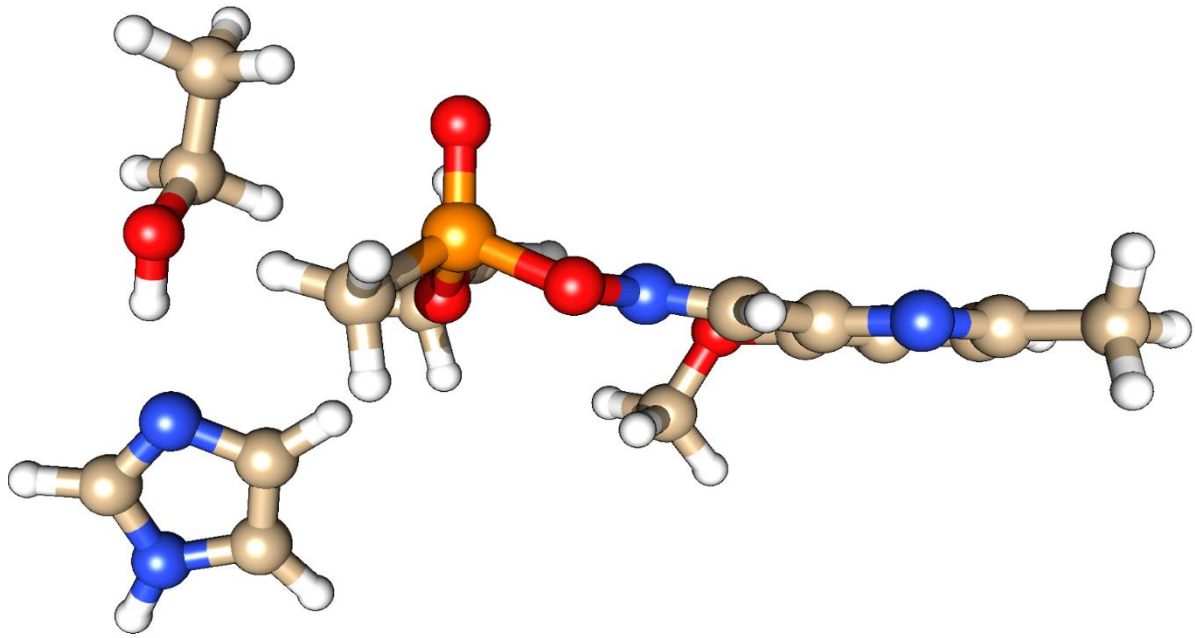


Figure II-11. Optimized structures of the product for the reactivation by 6M-3M-2PA in the minimal model

The reactivator 6M-3M-2PA is first evaluated in the minimal model. In this model, it was impossible to optimize any reactant structure. All attempts, even using cautious scan to generate high quality starting geometries, evolve into the product state show in **Figure II-11**. As such, not much can be said but that the reactivation with this reactivator in this model is assumed to have a high exothermicity and a low energy barrier.

A similar observation was made when the oxyanionic hole was added to the model. No reactant structure could be optimized by any mean as the system spontaneously evolve to the product state. This product structure is presented in **Figure II-12**. Once again, the only observation that can be made is that the reaction is likely to be exothermic and to have a low energy barrier.

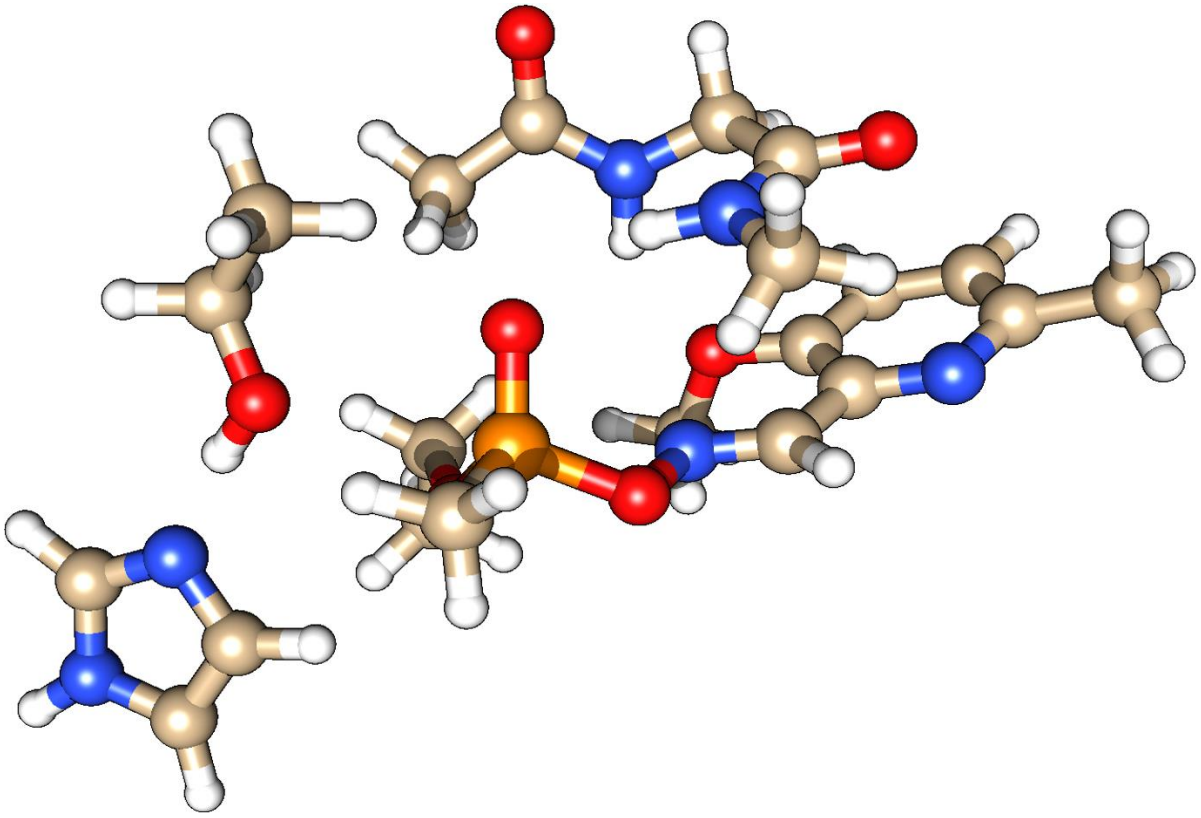


Figure II-12. Optimized structures of the product for the reactivation by 6M-3M-2PA in the model with the oxyanionic hole

When an acetate is added to the model the reactant, a stable pentavalent intermediate, and the product can be optimized. Those three structures are presented in **Figure II-13**. Energetically, the first step is an exothermic nucleophilic addition step with an energy difference of -2.74 kcal.mol^{-1}. The elimination step that follows has an energy difference of 7.99 kcal.mol^{-1}. The overall energy difference of the reaction is 5.25 kcal.mol^{-1}.

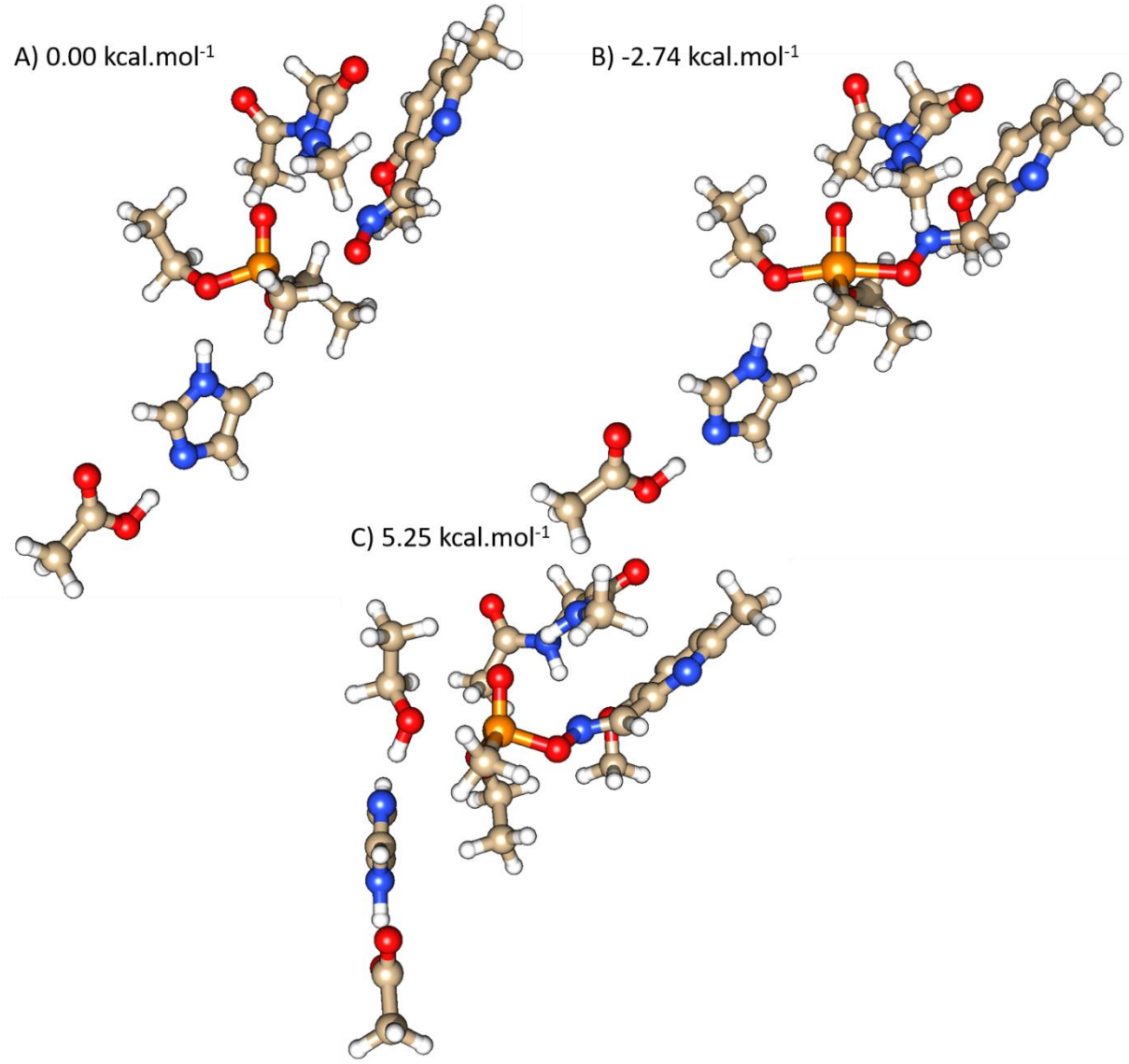


Figure II-13. Optimized structures of (A) the reactant, (B) the stable intermediate, and (C) the product for the reactivation by 6M-3M-2PA in the model with the oxyanionic hole and the acetate.

The evaluation of non-pyridinium reactivators with truncated QM models is flawed because the influence of the enzyme goes beyond the mere role of the catalytic residues and some steric strain. Those flaws are addressed by the use of a method that takes the full influence of the enzyme into account, QM/MM.

III. QM/MM study of the reactivation of VX-inhibited AChE by non-pyridinium reactivators

The QM/MM simulations of the reactivation of VX-inhibited AChE by non-pyridinium reactivators were conducted with similar methods and setup as presented in **Chapter 4**. A notable exception is the addition of a single sodium ion to compensate for a neutral non-positively charged aromatic ring in the reactivator prototype. The simulations were performed with a protonated Glu202 as the simulations performed in **Chapter 4** showed that the reactivation can only go through with a protonated Glu202.

III.1. 6M-2PA

First the reactivator 6M-2PA was modelled in the active site of AChE. since bidimensional scans were necessary to investigate the energy profile of the reactivation of VX-inhibited AChE by 2-PAM, a bidimensional scan was also used to investigate the energy profile of VX-inhibited reactivation by 6M-2PA. This bidimensional scan was performed along both phosphorus oxime (P-Ox) and phosphorus Ser203 (P-OSer203) distances. These two distances were made to vary in 0.05 Å steps and full optimization was performed at every point. The three-dimensional potential energy surface obtained from this bidimensional scan is presented in **Figure III-1**. In this scan, a high energy area with a P-Ox superior to 2.03 Å and a P-OSer203 superior to 2.07 Å was not fully investigated. This area was probed using mono-dimensional scans that confirmed that this area is very high in energy and thus not worth the investment of a full bidimensional scan.

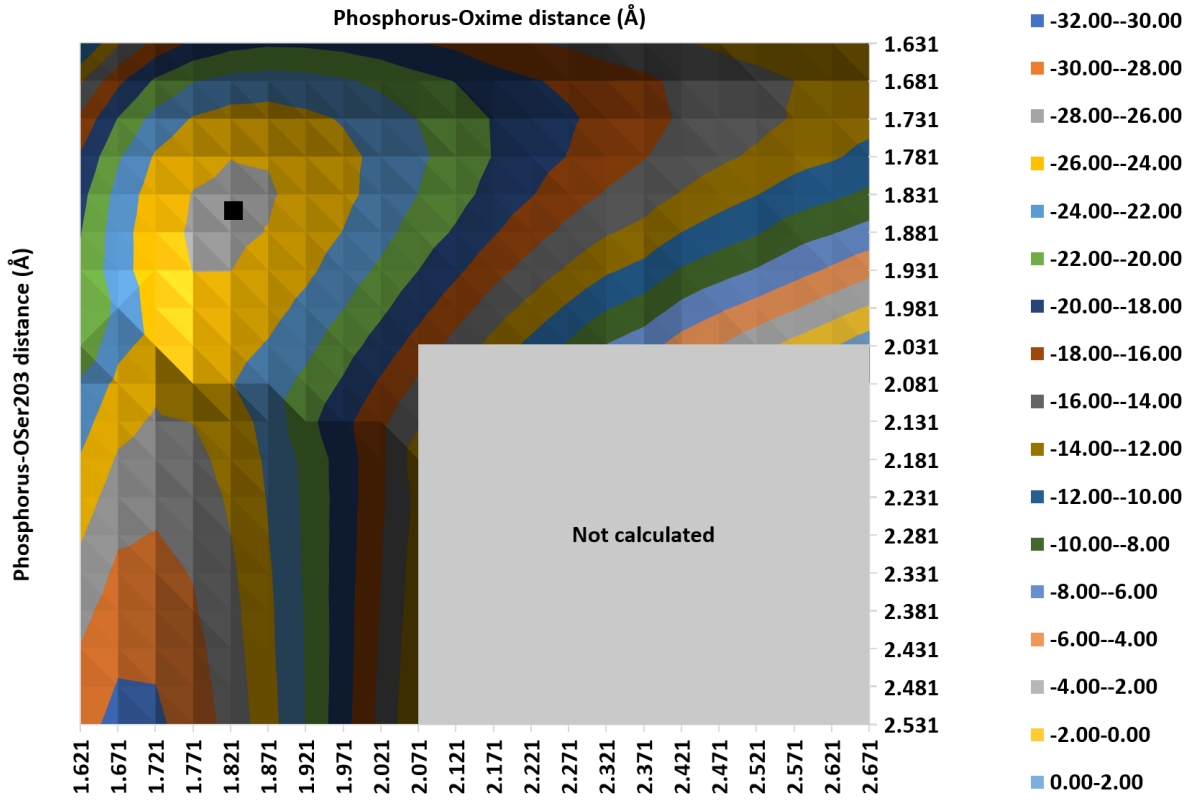


Figure III-1. Three-dimensional potential energy surface for the reactivation of VX-inhibited $Glu202^-$ AChE by 6M-2PA. The position of the optimized pentavalent on the potential energy surface is symbolized by a black square.

The reactant, product, and pentavalent structure were optimized by releasing constraints on select scan points with a reactant-like, product-like, pentavalent-like structure (see **Figure III-2**). The reactant structure, called structure **1_1**, has a P-Ox distance of 3.83 Å and a P-OSer203 distance of 1.66 Å. The pentavalent, called structure **1_2**, is almost symmetrical with a P-Ox of 1.82 Å and a P-OSer203 of 1.85 Å. The product finally, called structure **1_3**, has a P-Ox of 1.67 and a P-OSer203 of 2.83 Å. Those three structures, obtained from releasing scan points can be connected by a minimum energy path on the potential energy surface.

On the potential energy surface, the reactive distances P-Ox and P-OSer203 vary sequentially and not simultaneously. The reaction has two steps. First the P-Ox distance is reduced while the P-OSer203 remains roughly constant to connect the reactant and the pentavalent intermediate. This is the nucleophilic addition step. Then, following the minimum energy path, the P-Ox remains mostly constant while the P-OSer203 increases to form the product structure. This is the elimination step.

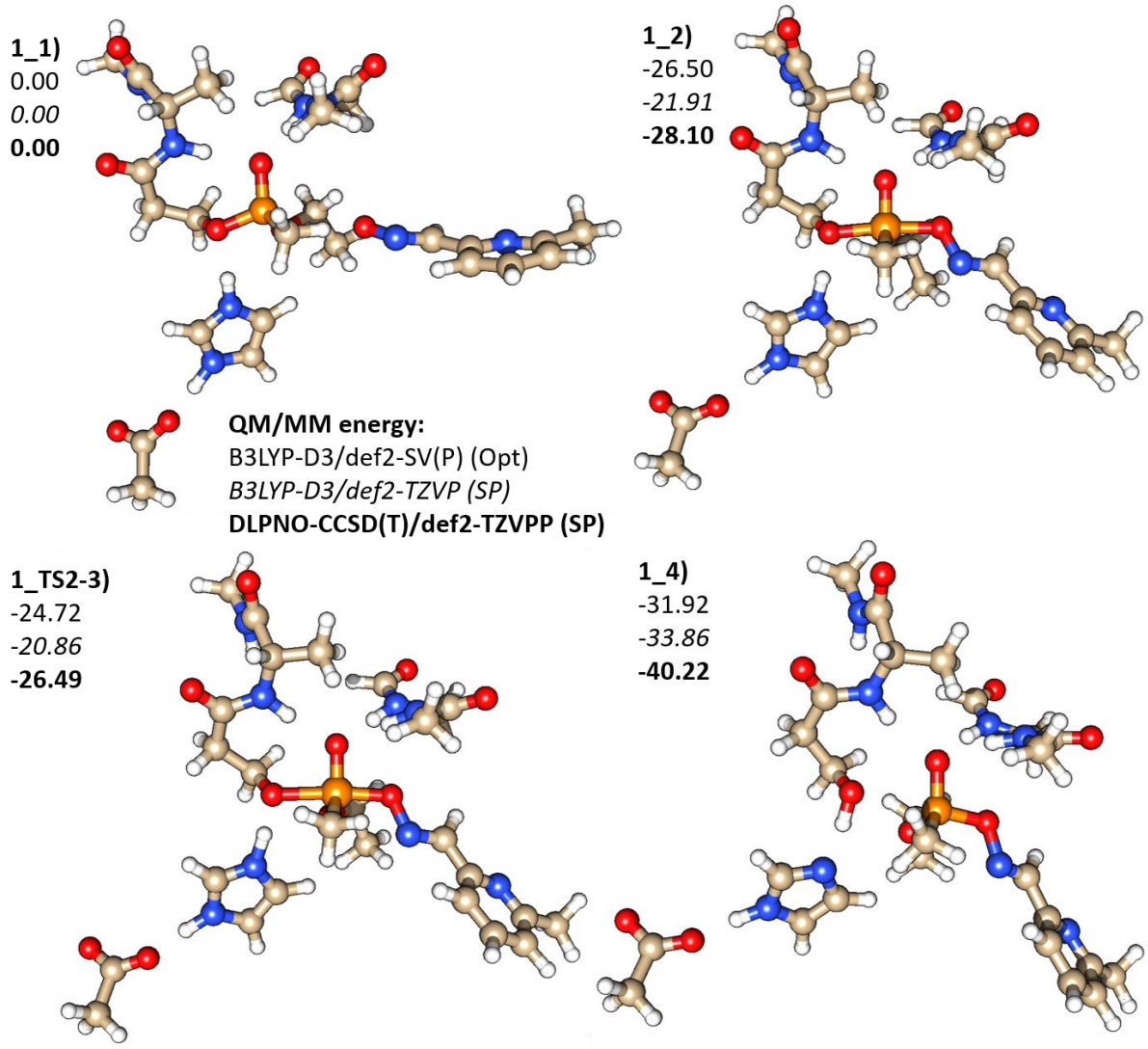


Figure III-2. Optimized structures of (1_1) the reactant, (1_2) the stable pentavalent intermediate, (1_4) the product, and (1_TS2-3) the transition state for the reactivation of VX-inhibited Glu202H^0 AChE by 6M-2PA.

Along that minimum energy path, a single saddle point could be observed between the pentavalent and the product. A finer scan with 0.01 Å steps was performed around the saddle point and allowed to locate the transition state **1_TS2-3** (see **Figure III-2**) for the elimination step. In this transition state, the P-Ox is 1.76 Å and the P-OSer203 is 2.07 Å. No transition state was located for the nucleophilic addition step, likely because this step is too exothermic. As the structure of the reactant has a P-Ox distance that falls outside of the boundary of the bidimensional scan, several monodimensional scan were used to probe the gap between the potential energy surface and structure **1_1**. Those scans showed no indication of a saddle

point in that gap. From the minimum energy path and the geometry of the pentavalent intermediate and the transition state, the reactivation of VX-inhibited AChE by 6M-2PA most definitely goes through an addition-elimination mechanism.

The reaction goes through a first nucleophilic addition step which is exothermic with an energy difference of -26.50 kcal.mol^{-1} at the B3LYP-D3/def2-SV(P) level that was used for geometry optimization. The energy barrier of the second step, the elimination, is 1.78 kcal.mol^{-1} and an exothermicity of -5.42 kcal.mol^{-1} at the B3LYP-D3/def2-SV(P) level. At this level the overall energy difference for the reaction is -31.92 kcal.mol^{-1}. Single points with the def2-TZVP basis set showed an exothermic first step with an energy difference of -21.91 kcal.mol^{-1}. With this basis set, the elimination has an energy barrier of 1.05 kcal.mol^{-1} and an exothermicity of -11.95 kcal.mol^{-1}. The overall energy difference is -33.86 kcal.mol^{-1}. At the DLPNO-CCSD(T)/def2-TZVPP level, the nucleophilic addition becomes exothermic by -28.10 kcal.mol^{-1} and the elimination step has an energy barrier and energy difference of 1.61 and -12.12 kcal.mol^{-1} respectively. The overall energy difference at the DLPNO-CCSD(T)/def2-TZVPP level is -40.22 kcal.mol^{-1}. The DLPNO-CCSD(T) calculations confirm the low barrier for the elimination as well as the exothermicity of both addition and elimination steps.

The reactivation of VX-inhibited AChE by 6M-2PA is a very favourable process due to the large exothermicity and the low energy barrier. The next subsection details how the addition of a hydroxide in position 3 of the pyridine ring affects the reactivation.

III.2. 6M-3H-2PA

Once again, to model the reactivation of VX-inhibited AChE by 6M-3H-2PA, a bidimensional scan was used to keep a consistent methodology across all the QM/MM simulations of the reactivation. This bidimensional scan scanned along the P-Ox and P-OSer203 distances in 0.05 Å steps with a full geometry optimization at every point. The three-dimensional potential energy surface obtained from this scan is presented in **Figure III-3**. A presumed high energy area was not investigated by this scan when the P-Ox exceeds 2.06 Å and the P-OSer203 exceeds 1.91 Å. Mono-dimensional scans were used to probe this area and confirm that in the range of distances for the P-Ox distance and the P-OSer203 distance the energy is very high. As such, this area is not worth the cost a full bidimensional scan.

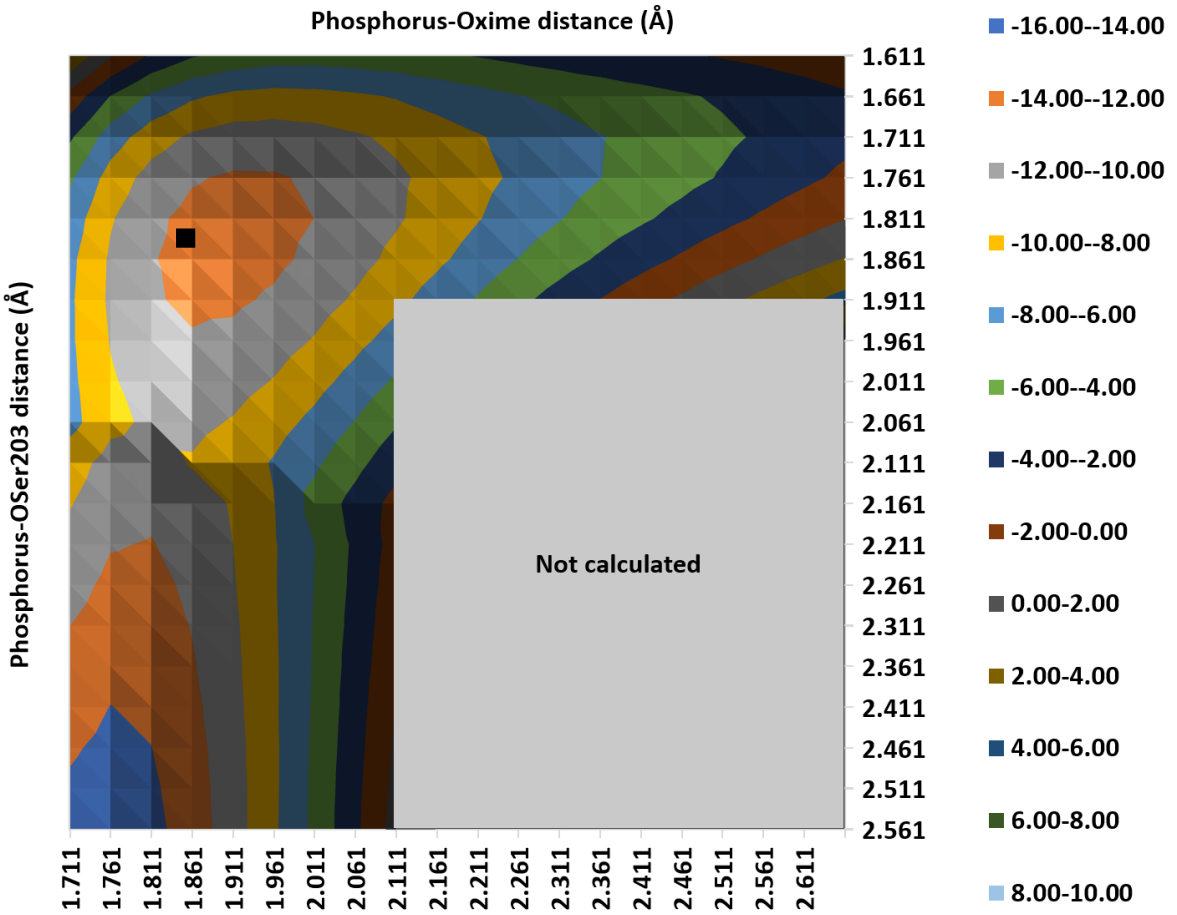


Figure III-3. Three-dimensional potential energy surface for the reactivation of VX-inhibited Glu202$^-$ AChE by 6M-3H-2PA. The position of the optimized pentavalent on the potential energy surface is symbolized by a black square.

The reactant, pentavalent of product structures were optimized from scan points with released constraints on P-Ox and P-OSer203 (see **Figure III-4**). The structure of the reactant, structure **2_1**, ha s a P-Ox distance of 3.21 Å and a P-OSer203 of 1.66 Å. In the pentavalent, structure **2_2**, the P-Ox is much shorter at 1.84 Å but the P-OSer203 has changed very little with a 1.83 Å distance. The product, structure **2_3**, has a P-Ox and a P-OSer203 of 1.68 Å and 2.87 Å respectively. Those three structures having been obtained from points of the bidimensional scan, they can be connected by the minimum energy path on the potential energy surface drawn from the results of the scan.

Following this potential energy surface, the reaction appears to have two steps. In the first step, the P-Ox distance is shortened while the P-OSer203 remain almost constant. In a second step the P-Ox now reduced to the length of a covalent P-O bond, remains roughly constant while the P-OSer203 increases until the product structure is reached. Those two steps are the nucleophilic addition and the elimination steps respectively.

Along the minimum energy path, a single saddle point can be found corresponding to the elimination step. In the area around this saddle point a finer scan was performed in 0.01 Å steps to localize the transition state **2_TS2-3** (see **Figure III-4**). This transition state has a P-Ox distance of 1.78 Å and a P-OSer203 distance of 2.10 Å. No transition state could be located

for the nucleophilic addition step, even as mono dimensional scans were used to probe the non-scanned area between the reactant and the area of the bidimensional scan. Despite the absence of a transition state for the nucleophilic addition, there is enough evidence to say that the reactivation of VX-inhibited AChE by 6M-3H-2PA goes through an addition-elimination mechanism.

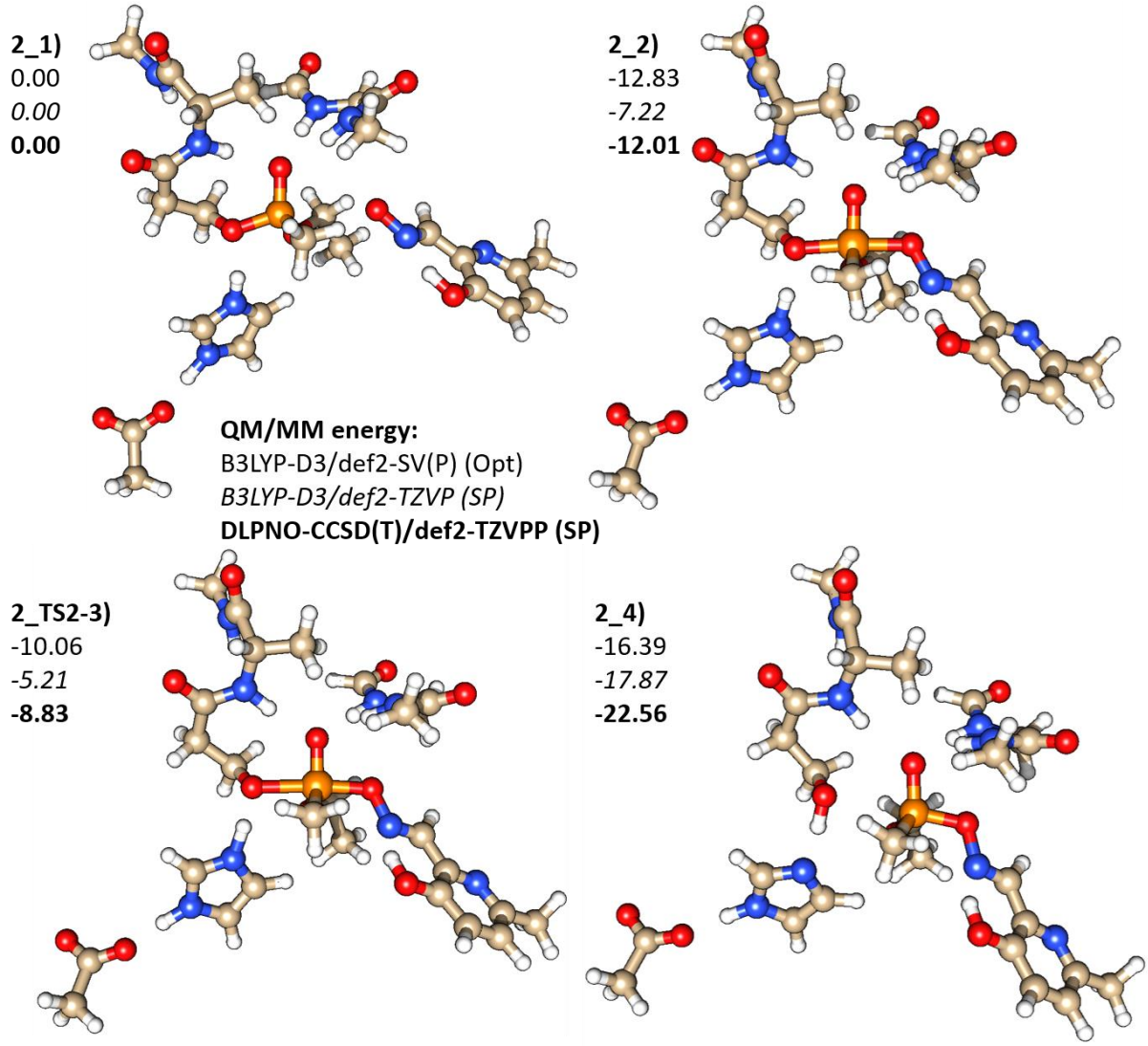


Figure III-4. Optimized structures of (2_1) the reactant, (2_2) the stable intermediate, (2_4) the product, and (2_TS2-3) the transition state for the reactivation of VX-inhibited Glu202H^0 AChE by 6M-3H-2PA.

The nucleophilic addition step of the reaction has, at the B3LYP-D3/def2-SV(P) level, an exothermicity of -12.83 kcal.mol^{-1}. For the elimination step, the energy barrier is 2.77 kcal.mol^{-1} and the energy difference is -3.56 kcal.mol^{-1} at the B3LYP-D3/def2-SV(P) level that was used for geometry optimization. The overall energy difference is -16.39 kcal.mol^{-1}. The def2-TZVP single points indicate an energy difference for the nucleophilic addition of -7.22 kcal.mol^{-1}. They show for the elimination an energy barrier of 2.01 kcal.mol^{-1} and an energy difference of -10.65 kcal.mol^{-1}. The overall exothermicity of the reaction is -17.87 kcal.mol^{-1}. At the DLPNO-CCSD(T)/def2-TZVPP level, the elimination as an energy difference of -12.01 kcal.mol^{-1}. The nucleophilic addition has an energy barrier of 3.18 kcal.mol^{-1} and an energy

difference of -10.55 kcal.mol^{-1} at this level. The overall energy difference is -22.56 kcal.mol^{-1}. The DLPNO-CCSD(T) values are in line with the B3LYP-D3 optimization and single point energies.

Although less exothermic (from -33.86 to -17.87 kcal.mol^{-1}) and with slightly higher energy barrier(from 1.05 to 1.73 kcal.mol^{-1}), the energetics of the reactivation of VX-inhibited AChE by 6M-3H-2PA show a very favourable reaction with a high exothermicity and low energy barrier. The reactivation was not the only parameter of those reactivators that was evaluated. The ability to be deprotonated inside the active site of AChE was also studied, since it is probably one of the key parameters for the reactivation as shown in **Chapter 4**.

IV. QM/MM study of the deprotonation of non-pyridinium reactivators in the active site of VX-inhibited AChE

As was shown in **Chapter 4**, the reactivation power of a reactivator is not the only measure of the efficiency of a reactivator. This efficiency is also governed by the ability of the reactivator to be deprotonated inside the active site of AChE by His447. The deprotonation of the reactivator is much more favourable in a deprotonated Glu202 but this leads to an energetic dead end (see section **V** in **Chapter 4**). The deprotonation simulations were thus conducted with a protonated Glu202, which has proven very energetically challenging for the reactivator 2-PAM. The simulations presented in this section rely on the same QM/MM methodology used throughout. The only change is the addition of residue Glu202 in the QM region as it was the case in **Chapter 4**. All the transition states were located though mono-dimensional scans with a single distance under constraint, between the proton and the receiving atom. This reaction distance was made to vary in 0.05 Å increments and when the area of the TS was located, the precise TS was obtained by a finer scan with 0.01 Å steps.

IV.1. 6M-2PA

The deprotonation of the reactivator 6M-2PA by His447 has been simulated with a protonated Glu202. Those simulations give the energy profile presented in **Figure IV-1**.

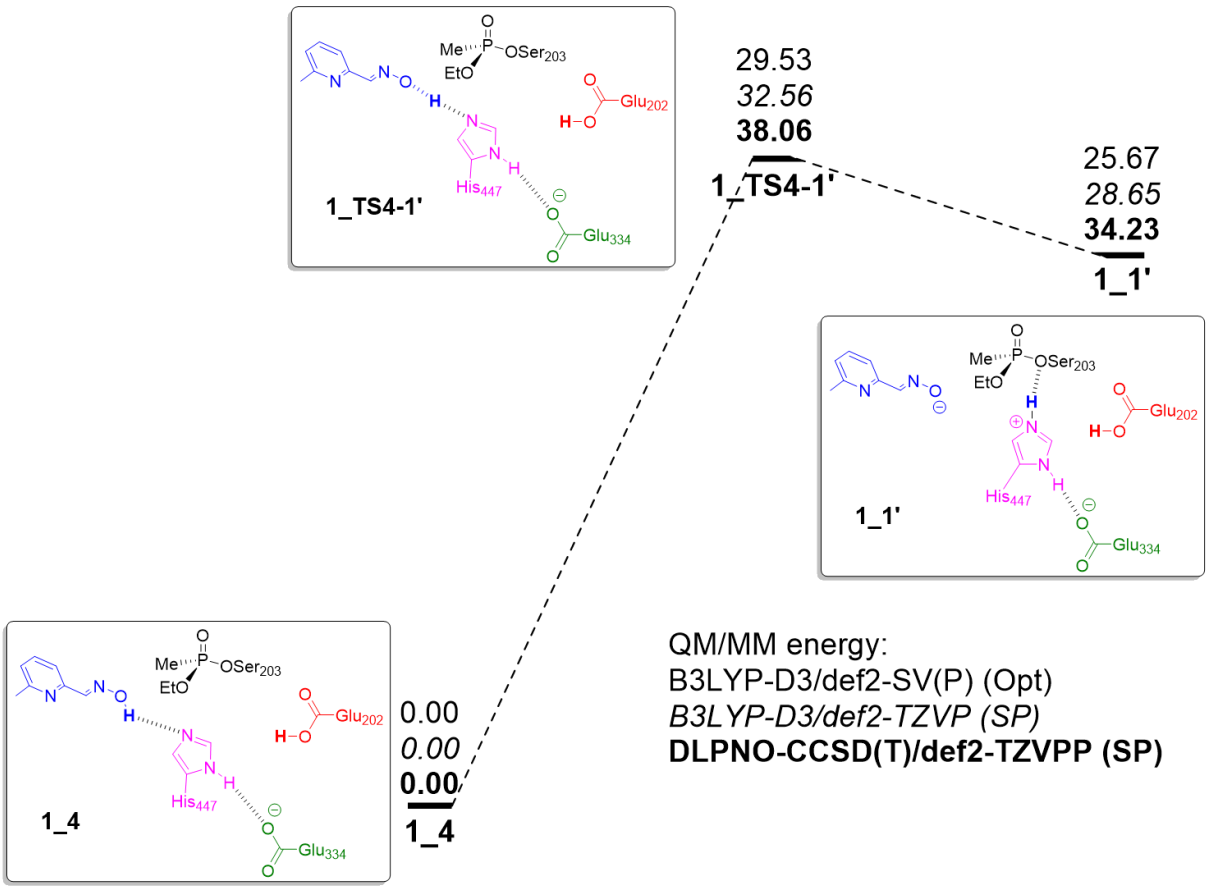


Figure IV-1. Energy profile for the proton transfer from 6M-2PA to His447 when Glu202 is protonated. Energies in kcal.mol^{-1}.

The deprotonation of 6M-2PA by His447 is a one-step endothermic reaction. The transition state **1_TS4-1'** could be localized to obtain the energy barrier of the process. At the B3LYP-D3/def2-SV(P) level that was used for geometry optimization this barrier is 29.53 kcal.mol^{-1} and the energy difference between the reactant **1_4** and the product **1_1'** is 25.67 kcal.mol^{-1}. At the B3LYP-D3/def2-TZVP level, the energy barrier is 32.56 kcal.mol^{-1} and the energy difference is 28.65 kcal.mol^{-1}. The DLPNO-CCSD(T)/def2-TZVPP single points gave an energy barrier and an energy difference of 38.06 kcal.mol^{-1} and 34.23 kcal.mol^{-1} respectively. These DLPNO-CCSD(T) values show some numerical differences with B3LYP-D3 obtained values, probably due to a difference in the potential energy surface corresponding to either method. Despite these numerical differences, the DLPNO-CCSD(T) completely confirms the trends outlined by the B3LYP-D3 energies.

IV.2. 6M-3H-2PA

Simulations have also been performed on the deprotonation of reactivator 6M-3H-2PA by His447 in the case of protonated Glu202. Those simulations are presented in **Figure IV-2**.

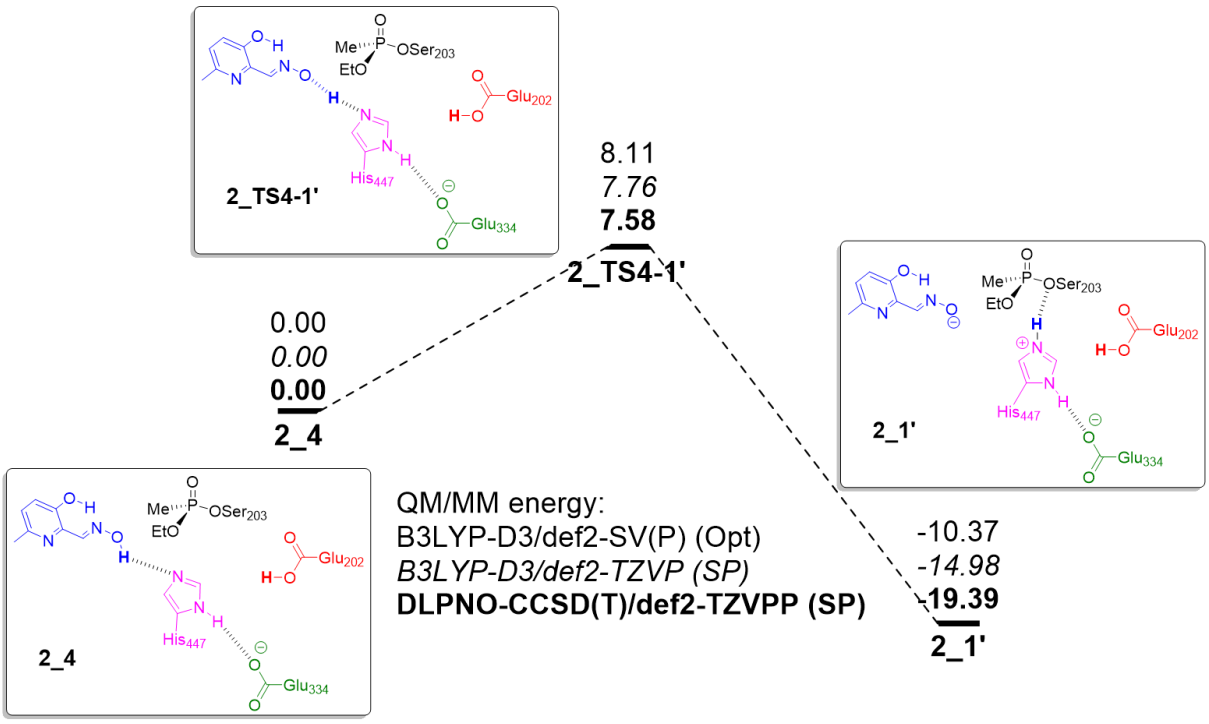


Figure IV-2. Energy profile for the proton transfer from 6M-3H-2PA to His447 when Glu202 is protonated. Energies in kcal.mol^{-1}.

This reaction is a one-step exothermic reaction. The localization of the transition state **2_TS4-1'** allowed for the evaluation of the energy barrier of the reaction. This energy barrier is 8.11 kcal.mol^{-1} at the B3LYP-D3/def2-SV(P) level used for geometry optimization. The energy difference is -10.37 kcal.mol^{-1} at this level. With def2-TZVP as a basis set, the energy barrier is 7.76 kcal.mol^{-1} and the energy difference is -14.98 kcal.mol^{-1}. The DLPNO-CCSD(T)/def2-TZVPP single points gave an energy barrier and an energy difference of 7.58 kcal.mol^{-1} and -19.39 kcal.mol^{-1} respectively. The DLPNO-CCSD(T) single points give an almost identical energy barrier as the B3LYP-D3/def2-TZVP single points but a higher exothermicity. This difference in exothermicity does not change the overall energy profile of the process which DLPNO-CCSD(T) does confirm.

V. Discussion

The first observation that can be made is the agreement between the DLPNO-CCSD(T)/def2-TZVPP calculations and the B3LYP-D3/def2-TZVP calculations. While there are some numerical differences between values obtained with both methods, the trends produced are systematically similar.

The preliminary evaluation of the three non-pyridinium reactivators 6M-2PA, 6M-3H-2PA, and 6M-3M-2PA in a truncated full QM model of the active site of VX-inhibited AChE showed that they have a greatly increased reactivity compared to 2-PAM. In the minimal model, the reactivation by 2-PAM had an energy difference of 1.20 kcal.mol^{-1}. In the same model, the reactivation by 6M-2PA is highly exothermic with an energy difference of -23.13 kcal.mol^{-1} (**Figure II-3**) and both 6M-3H-2PA, and 6M-3M-2PA show signs of a similar level of exothermicity, although the energetics for these reactivators could not be made available,

precisely because of the extend of the exothermicity. The addition of the oxyanionic hole to the model produces similar results. The energy difference for 6M-2PA in that model is -36.04 kcal.mol^{-1} (**Figure II-4**) while for 2-PAM the energy difference is only -2.64 kcal.mol^{-1}. Again, while precise energetics could not be obtained for the reactivation by 6M-3H-2PA and 6M-3M-2PA in this model, they show signs of high exothermicity. It should also be mentioned that as was previously observed in **Chapter 3**, the addition of the oxyanionic hole to the model apparently increases the exothermicity of the reactivation. For 6M-2PA this exothermicity increases by 12.91 kcal.mol^{-1}. Another evidence of the potency of non-pyridinium reactivators comes with the addition of the acetate to the model.

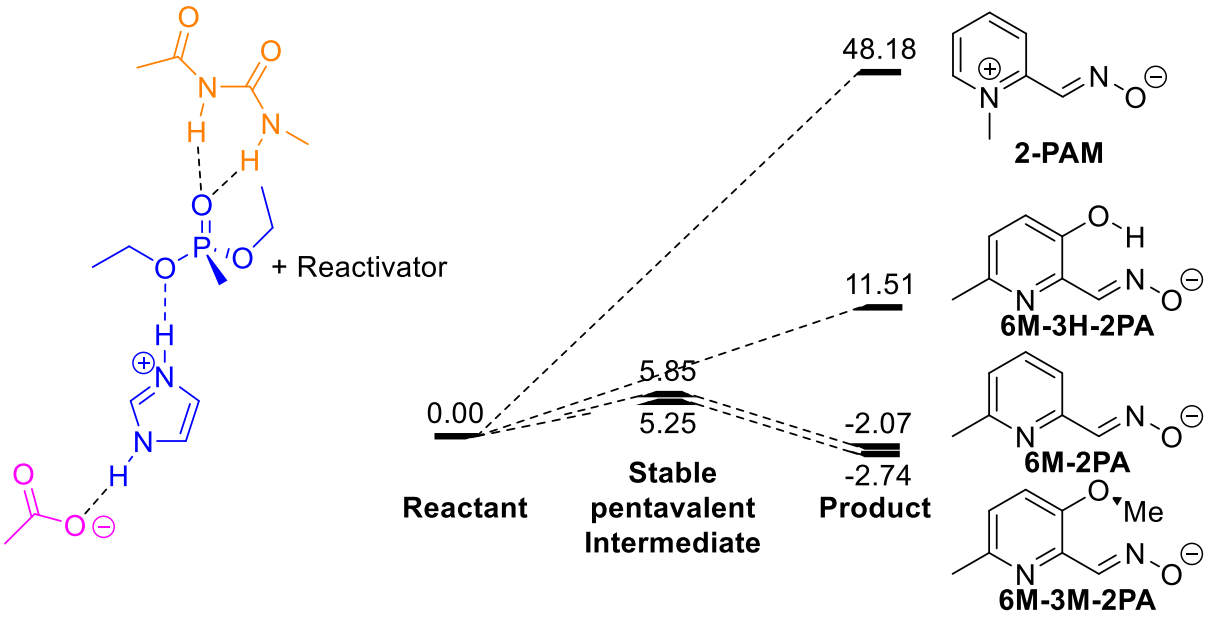


Figure V-1. Energy profile for the reactivation of a truncated model of the active site of VX-inhibited AChE. Energies in kcal.mol^{-1} obtained at the B3LYP-D3/def2-SV(P) level.

When the acetate that models Glu334's influence in the reactivation is added to the model, the reactivation by 2-PAM become very endothermic with an energy difference of 48.18 kcal.mol^{-1} at the B3LYP-D3/def2-SV(P) level (see **Figure V-1** and Section **III** of **Chapter 3**). With the non-pyridinium reactivator 6M-2PA, the addition of Glu334 reduces the energy difference by -50.25 kcal.mol^{-1}. The reactivation becomes exothermicity with an energy difference of -2.07 kcal.mol^{-1}. A stable pentavalent intermediate can also be optimized. The addition of a hydroxy group on position 3 of the pyridine aromatic cycle (6M-3H-2PA) increases the energy difference compared to 6M-2PA between the reactant and the product to 11.51 kcal.mol^{-1}. This effect is unexpected as the hydroxy group has a mesomeric donating effect that should increase the electron density on the pyridine ring and thus increase the electron density in the oxime and increase nucleophilicity. The hydrogen bond the hydroxide forms with the nitrogen of the oxime might also be responsible for a reduced electron density on the oxime and a decreased nucleophilicity. This is confirmed by the reactivation simulations in this truncated model by 6M-3M-2PA where the hydroxy is replaced with a methoxy group. With this reactivator, the energy difference of the reaction is -2.74 kcal.mol^{-1}, very close to the

value obtained with 6M-2PA. This indicates that the hydroxy group does reduce the nucleophilicity of the oxime. With 6M-3M-2PA, the pentavalent intermediate and the product are slightly more stable that with 6M-2PA by 0.60 and 0.67 kcal.mol^{-1} respectively, indicating a possible mesomeric donating effect of the methoxy.

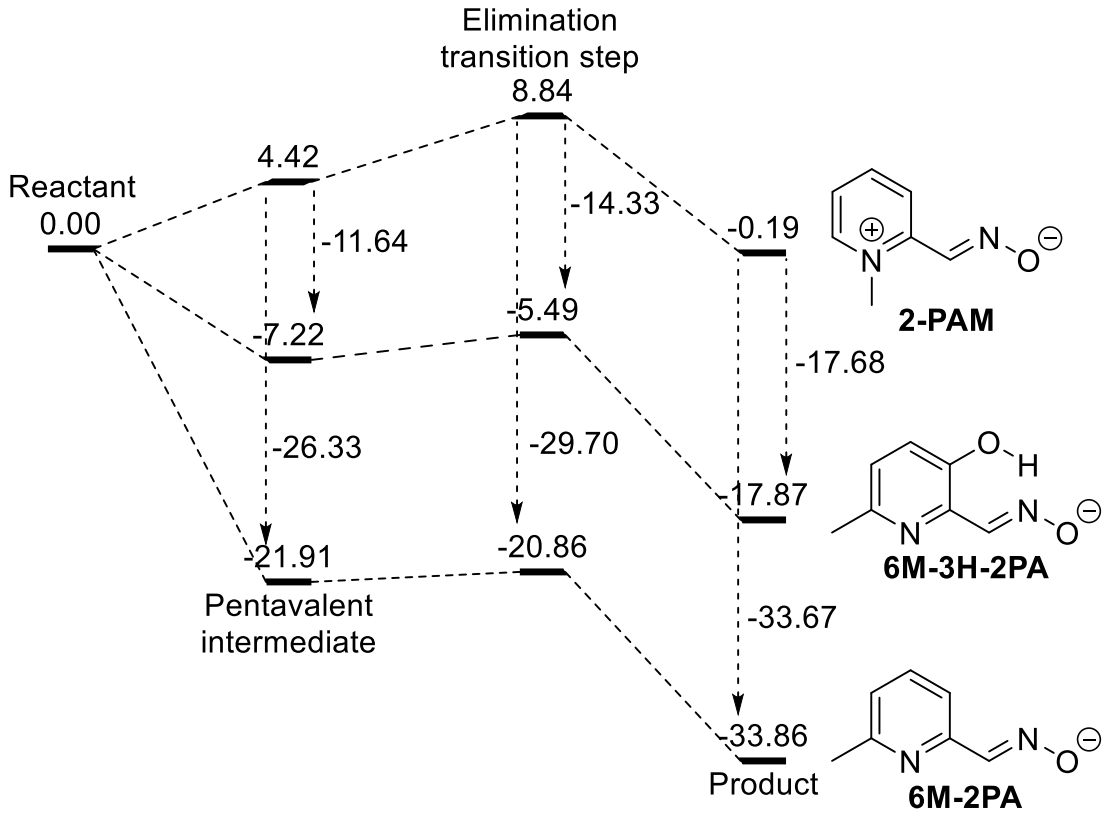


Figure V-2. QM/MM Energy profile for the reactivation of VX-inhibited AChE for three reactivators. Energies in kcal.mol^{-1} obtained at the B3LYP-D3/def2-TZVP:CHARMM level.

The QM/MM simulations of VX-inhibited AChE reactivation confirm the lesser nucleophilicity of 6M-3H-2PA compared to 6M-2PA. As can be seen in **Figure V-2**, both the addition and elimination steps are more exothermic for 6M-2PA than for 6M-3H-2PA by 14.69 and 15.99 kcal.mol^{-1} respectively. Both non-pyridinium reactivators have greater exothermicity for both steps than the pyridinium reactivator 2-PAM. The energy barrier for the elimination is 4.42 kcal.mol^{-1} with 2-PAM and is reduced to 1.73 and 1.05 kcal.mol^{-1} for 6M-3H-2PA and 6M-2PA respectively. Despite the diminished nucleophilicity of 6M-3H-2Pa due to the hydroxy group, both 6M-2PA and 6M-3H-2PA reactivators are more potent that 2-PAM. This result is surprising in the light of the research of the team of Pierre-Yves Renard at Rouen University comparing reactivator prototypes on the basis of their reactivity with PhX.[6] This study showed that 3-hydroxy-2-pyridinealdoxime (3H-2PA), almost identical to 6M-3H-2PA, is expected to have a slightly better reaction rate with PhX than 2-PAM. Both have a higher reaction rate than 6M-2PA by an order of magnitude which seems to contradict the

[6] L. Louise-Leriche, E. Păunescu, G. Saint-André, R. Baati, A. Romieu, A. Wagner, P.-Y. Renard, *Chem. Eur. J.* **2010**, *16*, 3510–3523.

QM/MM obtained energies. Experimental reactivation of VX-inhibited AChE by 2-PAM and 3-hydroxy-2-pyridinealdoxime (3H-2PA) does confirm the superior reactivity of 6M-3H-2PA, despite a lower affinity to AChE.[7]

Both the QM/MM simulations and the truncated QM model simulations with oxyanionic hole and acetate in the model agree on an addition-elimination mechanism for the reactivation of VX-inhibited AChE with non-pyridinium reactivators. This is in agreement with the previous computational evaluation of non-pyridinium reactivators in truncated QM,[8] and the general consensus on AChE reactivation by oximes.[9] The elimination step of the reactivation is the kinetically determinant step for both QM/MM and truncated QM simulations.

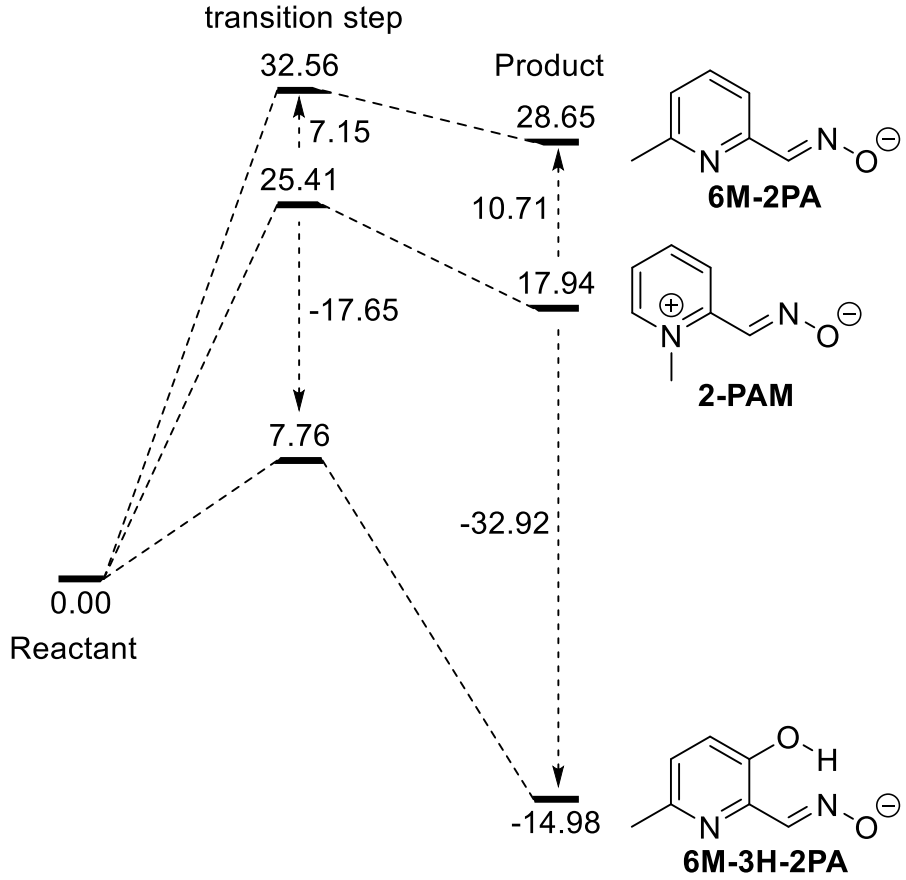


Figure V-3. QM/MM Energy profile for the deprotonation of reactivators in the active of VX-inhibited AChE. Energies in kcal.mol^{-1} obtained at the B3LYP-D3/def2-TZVP:CHARMM level.

The second parameter of VX-inhibited AChE reactivation is the ability for the reactivator to be deprotonated inside the active site of AChE. In this aspect, the reactivator

[7] J. Renou, G. Mercey, T. Verdelet, E. Păunescu, E. Gillon, M. Arboléas, M. Loiodice, M. Kliachyna, R. Baati, F. Nachon, et al., *Chem. Biol. Interact.* **2013**, *203*, 81–84.

[8] R. Lo, N. B. Chandar, M. K. Kesharwani, A. Jain, B. Ganguly, *PLoS One* **2013**, *8*, e79591.

[9] a) Y. Li, L. Du, Y. Hu, X. Sun, J. Hu, *Can. J. Chem.* **2012**, *90*, 376–383. b) K. S. Matos, D. T. Mancini, E. F. da Cunha, K. Kuča, T. C. França, T. C. Ramalho, *J. Braz. Chem. Soc.* **2011**, *22*, 1999–2004. c) K. S. Matos, E. F. F. da Cunha, A. da Silva Gonalves, A. Wilter, K. Kuča, T. C. C. França, T. C. Ramalho, *J. Biomol. Struct. Dyn.* **2012**, *30*, 546–558.

6M-2PA is very much worse than classical reactivator 2-PAM (see **Figure V-3**). The energy barrier for the deprotonation of 6M-2PA by His447 is 32.56 kcal.mol^{-1} with an energy difference of 28.65 kcal.mol^{-1}, respectively 7.15 and 10.71 kcal.mol^{-1} higher than 2-PAM. 6M-3H-2PA however, probably thanks to the stabilization of the high electron density on the oxime, is considerable more easily deprotonated by His447 than 2-PAM. The endothermic deprotonation of 2-PAM is exothermic in the case of 6M-3H-2PA by -14.98 kcal.mol^{-1} with a small energy barrier of 7.76 kcal.mol^{-1}. The inability to effectively deprotonate 6M-2PA demonstrated by this result can be used to rationalize the results obtained by the research team of Pierre-Yves Renard. Those experiments were performed in a phosphate buffer at a pH of 8.5 in which the higher basicity of 6M-2PA is a hindrance to reactivation.[6]

The reactivation of VX-inhibited AChE by 3-hydroxy-2-pyridinealdoxime (3H-2PA) has been reported by Renou *et al.* to have a reactivation rate constant 8 times greater than the reactivation by 2-PAM.[7] The energy barrier difference for the reactivation between 6M-3H-2PA, which is almost identical to 3H-2PA, and 2-PAM indicates, using transition state theory, that the reactivation rate constant for 6M-3H-2PA should be more than 60 times greater than for 2-PAM. When considering the energy barrier difference for the deprotonation between 6M-3H-2PA and 2-PAM, the reactivation rate constant should be more than 35 000 times greater for 6M-3H-2PA than for 2-PAM. If either the reactivation or the deprotonation steps are considered rate limiting, neither is in agreement with the experiment. This would however be an oversimplification of the issue. The experimental reactivation rate value is based on reactivation kinetics and the evolution of an observed bimolecular rate constant. This observed reaction rate combines all events that result in reactivation of VX-inhibited AChE. It includes the binding of the reactivator to AChE, its progression through the channel, estimated to be more difficult than for 2-PAM,[10], [8] the deprotonation of the oxime functional group followed by reactivation as modelled in this chapter, in the case of protonated Glu202. It also includes reactivation in other protonation states for the residues of the active and reactivation in the case of a reactivator entering the active site of AChE deprotonated. If one of those pathways constitutes a significant proportion of reactivation event in addition to the reactivation pathway modelled in this chapter, it could produce an experimental reactivation rate ratio in disagreement with the ratio calculated from the energy barriers of individual processes. It should be reminded that the static QM/MM simulations are performed at 0 K, while experiments are performed around 300 K.

[10] a) A. da S. Gonçalves, T. C. C. França, A. Wilter, J. D. Figueroa-Villar, *J. Braz. Chem. Soc.* **2006**, *17*, 968–975. b) R. Lo, B. Ganguly, *Mol. BioSyst.* **2014**, *10*, 2368–2383.

VI. Conclusion

The results presented in this chapter help rationalize the potency of the non-pyridinium reactivator designs compared to classical reactivator based on the 2-PAM prototype. It was shown that the charge on the pyridinium severely deactivates the oxime by reducing its nucleophilicity. Simply getting rid of the charge by removing the methyl group on the nitrogen of methyl pyridinium to have a pyridine aromatic ring is not sufficient to produce an efficient reactivator. As was shown with 6M-2PA, the increased nucleophilicity of a non-pyridinium reactivator increases the basicity of the oxime and thus increases the cost of oxime deprotonation in the active site of VX-inhibited AChE. The addition of a hydroxy group on the pyridine ring able to form hydrogen bond with the nitrogen of the oxime slightly decreases the nucleophilicity. This small reduction is the price to pay for the much-reduced cost of deprotonating the oxime. While not as potent as 6M-2PA in the reactivation reaction, 6M-3H-2PA is a more effective reactivator because of its ability to easily be deprotonated.

Chapter 6: Diffusion of Protons In and Out of the Active Site of Acetylcholinesterase

Chapter 6: Diffusion of Protons In and Out of the Active Site of Acetylcholinesterase

I. Introduction

In the previous chapters, the protonation state of active site residues was demonstrated to play a role in the reactivation process. The energetics of proton transfer steps, most important of which the deprotonation of the reactivator, have been shown to vary depending on the protonation state of active site residues. This observation prompts the question of the diffusion of protons in and out of the active site. To study this diffusion several approaches at different levels of theory were used. Computational details on these approaches will be given at the beginning of every section. First, the possibility of direct proton exchange between the solvent and the residues of the active site was studied with QM/MM. A molecular dynamic study of the diffusion of water molecules in the enzyme using polarizable force field AMOEBA was initiated. Although it is still underway and no result will be discussed the setup of these dynamics will be quickly discussed. During a stay at Uppsala University, in the context of a collaboration with Professor Lynn Kamerlin, CpHMD simulations were used to evaluate the protonation state of most AChE protonatable residues. Finally, a proton relay mechanism was uncovered and modelled using both QM/MM and EVB simulations.

II. Direct proton exchange between His447, Glu202, and the solvent

Direct proton transfer and between the solvent and protonatable residues of the active and between active site residues can occur before the reactivator enters the active site. These two issues were tackled using QM/MM. To perform these simulations a new setup must be employed without the reactivator in the active site of AChE. To avoid this new setup to be too different form the previous setup presented in **Chapter 4** and thus to allow more pertinent comparisons between results obtained with either setups, the new setup was prepared from the previous setup presented in **Chapter 4**.

The first step in preparing a new setup is to remove the reactivator from the setup from **Chapter 4**. In this setup, there is a 8 Ångström thick shell of water molecules around the enzyme. A 16 Ångström thick shell was added around the system to bring the water molecule shell around the enzyme to 24 Ångström thick using Chimera 1.10.1.[1] The system was then equilibrated for 50 ns, with the outer 16 Å thick water molecules shell kept frozen, using NAMD 2.9.[2] This frozen shell is used to prevent the system from bursting. From this dynamic, a snapshot is extracted. The QM/MM simulations will be performed on this snapshot. The outer 16 Å of water molecules used to keep the molecular dynamics in a bubble are removed.

[1] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.

[2] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J. Comput. Chem.* **2005**, 26, 1781–1802.

The composition of the QM region is almost identical to the QM region used for deprotonation simulations in **Chapters 4** and **5** except for the water molecules occupying the same space as the reactivator before them. The MM region is also almost identical to the one used in the QM/MM simulations presented in **Chapters 4** and **5** except for the water molecules occupying the active site which have moved during the equilibration step. Both regions are respectively presented in **Figure II-1A** and **Figure II-1B**.

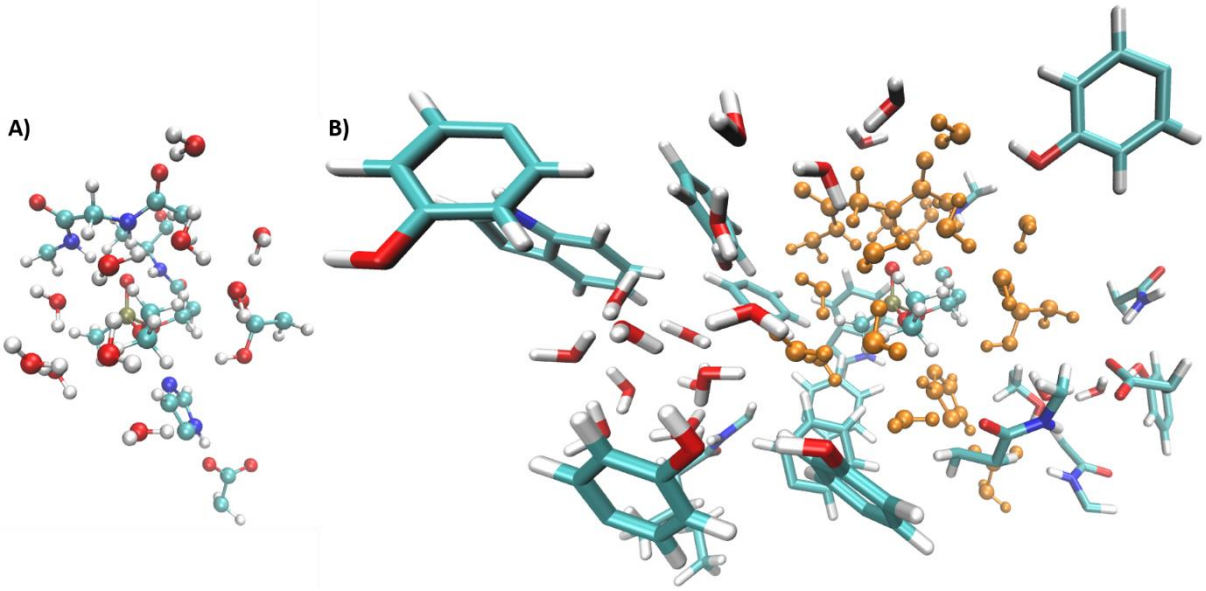


Figure II-1. (a) QM region, (B) MM region with the atoms of the QM region in orange and the Ser203-VX adduct in standard elemental colouring.

All other aspects of the QM/MM simulations are strictly identical with the simulations performed in **Chapters 4** and **5**. The electrostatic embedding scheme is used to account for the MM region in the QM calculations, the link atom with the charge shift scheme is used to handle the boundary between the QM and the MM regions. The geometries are optimized using the HDLCOpt module included in Chemshell 3.1b1.[3] The QM energetic data is provided by Turbomole V6.5[4] and calculated using B3LYP[1] with the D3 dispersion correction[5] and the def2-SV(P) basis set[6]. The MM data is provided by the DL_POLY_3 interface of Chemshell based on CHARMM22 parameters with modifications for the covalent adduct VX (**Appendix 2**). In some important cases, single points were performed using Turbomole at the B3LYP-D3/def2-TZVP level and using ORCA[7] at the DLPNO-CCSD(T)/def2-TZVPP level.[8]

[3] P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, et al., *J. Mol. Struct.: THEOCHEM* **2003**, 632, 1–28.

[4] a) R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, *Chem. Phys. Lett.* **1989**, 162, 165–169. b) R. Ahlrichs, F. Furche, C. Hättig, W. Klopper, M. Sierka, F. Weigend, TURBOMOLE, TURBOMOLE GmbH, Development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, **2012**.

[5] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, 132, 154104.

[6] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, 7, 3297.

[7] F. Neese, *WIREs Comput Mol Sci* **2012**, 2, 73–78.

[8] a) C. Riplinger, F. Neese, *J. Chem. Phys.* **2013**, 138, 034106. b) C. Riplinger, B. Sandhoefer, A. Hansen, F. Neese, *J. Chem. Phys.* **2013**, 139, 134101

II.1. Proton exchange between His447 and Glu202

In this new setup, the transfer of a proton between His447 and Glu202 has been modelled. The results of these calculations are presented in **Figure II-2**. The transition state was located though a mono-dimensional scan with a single distance under constraint, between the proton and the receiving atom. This reaction distance was made to vary in 0.05 Å increments and when the area of the TS was located, the precise TS was obtained by a finer scan with 0.01 Å steps.

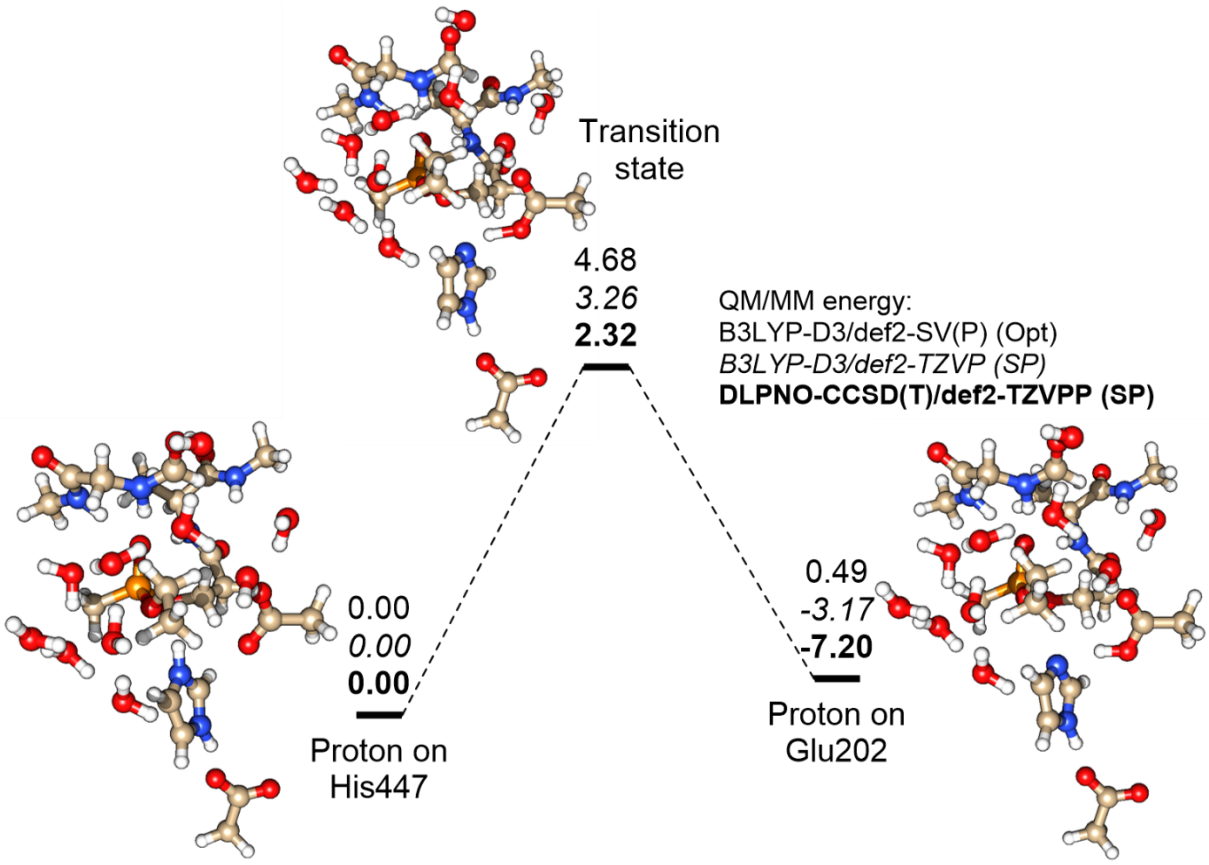


Figure II-2. Energy profile for the proton transfer from His447 to Glu202. Energies in kcal.mol^{-1}. The structures are places adjacent to the respective states.

The proton transfer from His447 to Glu202 has an energy barrier of 4.68 in B3LYP-D3/def2-SV(P) relative to the energy of the reactant state (with the proton on His447). The reaction is endothermic with an energy difference of 0.49 kcal.mol^{-1}. When the larger basis set def2-TZVP is used for single points however, the energy difference is reversed and protonated Glu202 is more stable than protonated His447 by 3.17 kcal.mol^{-1}. The energy barrier with this basis set is 3.26 kcal.mol^{-1}. The DLPNO-CCSD(T)/def2-TZVPP single points give an energy barrier of 2.32 kcal.mol^{-1} and an exothermicity of the proton transfer towards Glu202 of -7.20 kcal.mol^{-1}. The DLPNO-CCSD(T) energies are in agreement with the B3LYP-D3/def2-TZVP single points on the exothermicity of the transfer of a proton from His447 to Glu202. Like the B3LYP-D3/def2-TZVP energies they contradict the B3LYP-D3/def2-SV(P) energy on the energy difference.

II.2. Proton exchange between the solvent and His447

QM/MM Simulations were performed to evaluate the protonation and deprotonation of His447 by the solvent. Those simulations were performed with either a protonated or unprotonated Glu202 and the results are presented in **Figure II-3**. The transition states could not be located for any of the presented steps.

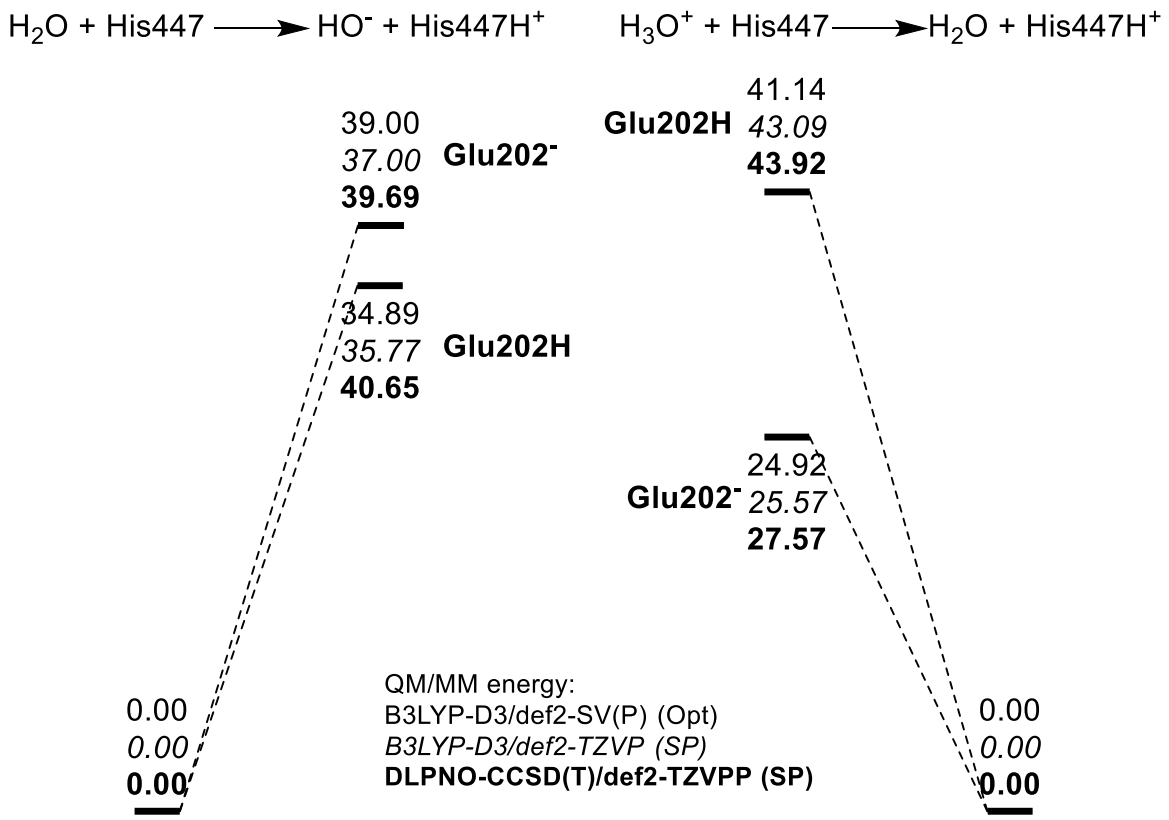


Figure II-3. Energy profile for the exchange of protons between His447 and the solvent in the two available protonation states of Glu202.

The deprotonation of a water molecule by His447 is a barrierless endothermic process. The energy difference is 39.00, 37.00 and 39.69 kcal.mol^{-1} in B3LYP-D3/def2-SV(P), B3LYP-D3/def2-TZVP, and DLPNO-CCSD(T)/def2-TZVPP respectively when Glu202 is unprotonated. When Glu202 is protonated, the reaction remains endothermic and barrierless. However, he energy difference slightly decreases to 34.89, 35.77 and 40.65 kcal.mol^{-1} in B3LYP-D3/def2-SV(P), B3LYP-D3/def2-TZVP, and DLPNO-CCSD(T)/def2-TZVPP respectively. The repulsion between the charge of an unprotonated Glu202 and the hydroxy ion produced by the deprotonation of a water molecule might explain why a protonated Glu202 slightly reduces the endothermicity of the reaction. If the reaction is considered from the other end, as a deprotonation of His447H$^+$ by a hydroxy ion, it is a very exothermic process in both protonation states of Glu202.

The deprotonation of a hydronium ion by His447 is a barrierless exothermic process. When Glu202 is protonated, its energy difference is 41.14, 43.09, and 43.92 kcal.mol^{-1} at the B3LYP-D3/def2-SV(P) level, at the B3LYP-D3/def2-TZVP level, and at the DLPNO-CCSD(T)/def2-

TZVPP level respectively. When Glu202 is unprotonated the exothermicity is much reduced to 25.92, 25.57, and 27.57 kcal.mol^{-1} at the B3LYP-D3/def2-SV(P) level, at the B3LYP-D3/def2-TZVP level, and at the DLPNO-CCSD(T)/def2-TZVPP level respectively. The second protonatable residue of the active site of VX-inhibited AChE was also subject of a similar study. If the perspective on the reaction is reversed, the deprotonation of His447H$^+$ by a water molecule is a very endothermic barrierless process in both protonation states of Glu202, with reduced endothermicity for unprotonated Glu202 but still impossible.

II.3. Proton exchange between the solvent and Glu202

The protonation and deprotonation of Glu202 was also evaluated using QM/MM simulations. These simulations have been performed with either a mono- (His447^0) or bis-protonated His447 (His447H$^+$) side chain. The results are presented in **Figure II-4**. For all the steps presented, no transition state could be located.

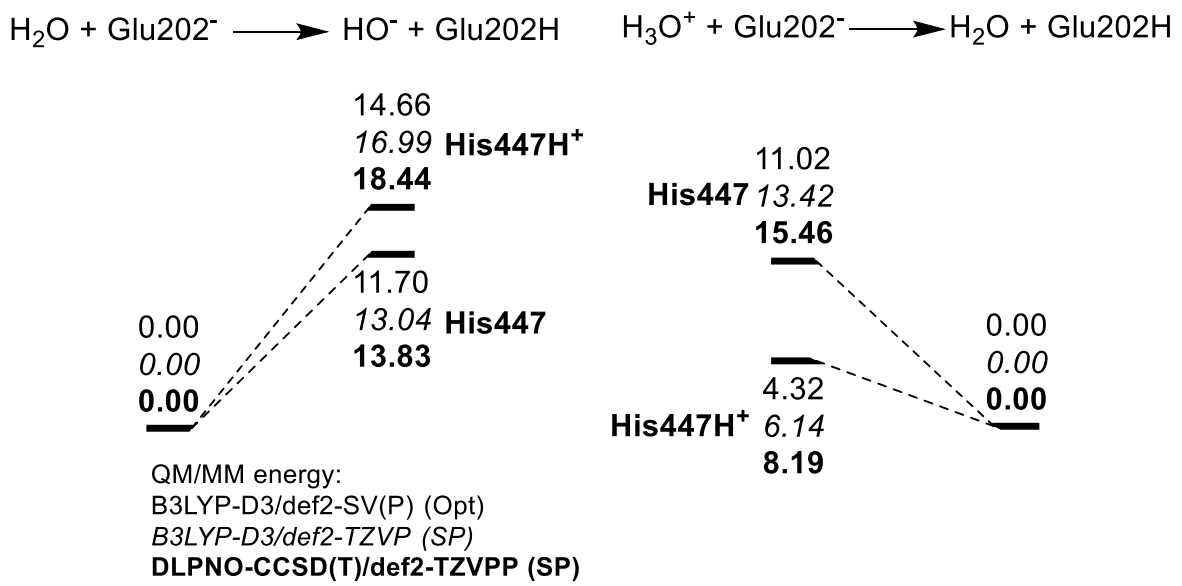


Figure II-4. Energy profile for the exchange of protons between His202 and the solvent in the two available protonation states of Glu447.

The transfer of a proton from a water molecule to an unprotonated Glu202 has been evaluated in both protonate states of His447. When His447 is bis-protonated and thus has an imidazolium side chain, the proton transfer from a water molecule to Glu202 is endothermic without energy barrier. The energy difference is 14,66, 16.99, and 18.44 kcal.mol^{-1} at the B3LYP-D3/def2-SV(P) level, the B3LYP-D3/def2-TZVP level, and the DLPNO-CCSD(T)/def2-TZVPP level respectively. When His447 is mono-protonated, with an imidazole side chain, the transfer of a proton from a water molecule to the unprotonated Glu202 has a reduced endothermicity and remains barrierless. The energy difference is 11.70, 13.04, and 13.83 kcal.mol^{-1} at the B3LYP-D3/def2-SV(P) level, the B3LYP-D3/def2-TZVP level, and the DLPNO-CCSD(T)/def2-TZVPP level respectively. This reduced endothermicity could be tied to a stabilized unprotonated Glu202 state by His447H$^+$ due to electrostatic interactions of the charges of the imidazolium with the carboxylate. The reverse reaction, the deprotonation of

protonated Glu202 by a hydroxy ion, is very exothermic in both protonation states of His447. The exothermicity is increased when His447 is bis-protonated.

The deprotonation of a hydronium ion by Glu202 in carboxylate form is exothermic and barrierless in both protonation states of His447. When His447 is bis-protonated, the proton transfer from the hydronium to Glu202 has an energy difference of 4.32, 6.14, and 8.19 kcal.mol^{-1} in B3LYP-D3/def2-SV(P), B3LYP-D3/def2-TZVP, and DLPNO-CCSD(T)/def2-TZVPP respectively. When His447 is mono-protonated, the proton transfer from the hydronium to Glu202 has an energy difference of 11.02, 13.42, and 15.46 kcal.mol^{-1} in B3LYP-D3/def2-SV(P), B3LYP-D3/def2-TZVP, and DLPNO-CCSD(T)/def2-TZVPP respectively. Once again the reduced exothermicity of the proton transfer from the hydronium to Glu202 when His447 is mono-protonated might be due to a stabilization of unprotonated Glu202 by electrostatic interactions of the carboxylate with the imidazolium. Considering the reverse reaction, the deprotonation of Glu202 by a water molecule is also possible. This process is highly endothermic with no energy barrier. This endothermicity is reduced for His447H$^+$.

One of the main information of this study of proton exchange between the solvent and residues of the active site is that the protonation or deprotonation of a water molecule by His447 or Glu202 has a high energetic cost. Proton exchange between these residues and water ions (hydronium and hydroxy ions) however is systematically spontaneous. It brings the question of the rapid diffusions of water molecules in an out of the active site as it would favour the entrance of a hydroxy of hydronium ion. The simulations presented in the next section were performed to address this point.

III. Molecular dynamics of acetylcholinesterase with polarizable force field

In order to study the movements of water molecules in the active site and through the channel of AChE, molecular dynamics simulations in polarizable force field have been performed. The intention was to obtain 25ns trajectories after equilibration for both four systems: VX-inhibited AChE with either protonated and unprotonated Glu202 and non-inhibited AChE with either protonated or unprotonated Glu202. Molecular dynamic simulations of AChE at various protonation states of Glu202 is not in itself a novelty.[9] The use of a polarizable force field for such study, or for any study of AChE has never been conducted before. Those simulations are still ongoing due to the computational cost of polarizable force field simulations for large enzymatic systems. In this section only the setup will be discussed. This part of the thesis project was the result of a collaboration with the team of Jean-Philip Piquemal at LCT lab from UPMC university.

The molecular dynamic simulations have been performed using the high performance version of Tinker, Tinker-HP[10] with the 2009 AMOEBA biopolymer force field (often dubbed

[9] J. Wiesner, Z. Kříž, K. Kuča, D. Jun, J. Koča, *J. Biomol. Struct. Dyn.* **2010**, *28*, 393–403

[10] a) F. Lipparini, L. Lagardère, B. Stamm, E. Cancès, M. Schnieders, P. Ren, Y. Maday, J.-P. Piquemal, *J. Chem. Theory Comput.* **2014**, *10*, 1638–1651. b) L. Lagardère, F. Lipparini, É. Polack, B. Stamm, É. Cancès, M. Schnieders, P. Ren, Y. Maday, J.-P. Piquemal, *J. Chem. Theory Comput.* **2015**, *11*, 2589–2599. c) Q. Wang, J. A. Rackers, C. He, R. Qi, C. Narth, L. Lagardere, N. Gresh, J. W. Ponder, J.-P. Piquemal, P. Ren, *J. Chem. Theory Comput.* **2015**, *11*, 2609–2618.

amoebabio09).[11] The setup was performed by the author with the guidance of both Lea El Khoury and Louis Lagardère from the LCT lab. The setup was based on the previous enzyme setup described in **Chapter 4**. The simulations were performed in periodic boundary conditions which required to put the system in a 100 Å by 100 Å by 100 Å solvent box using Chimera 1.10.1.[1] The Chemshell/CHARMM compatible .pdb file was then manually converted to the Tinker .xyz format.

The most delicate aspect of the setup was the modification of the amoeba force field to add parameters for the covalent VX adduct. The difficulty of this step is that the existing protocol to generate parameters from *ab-initio* calculations describes only the parametrization of substrates and non-covalent adducts.[12] The strategy for the parametrization of the covalent adduct was to start from the regular parameters for serine residues and to replace the parameters for the CH_2-OH side chain of a regular serine with the parameters for the P(O)(OEt)(CH_3)(CH_2) moiety of P(O)(OEt)$_2CH_3$ (see **Figure III-1**). The parameters for P(O)(OEt)$_2CH_3$ would be obtained through the existing protocol.

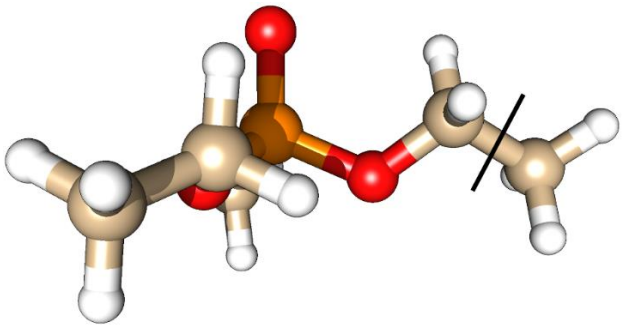


Figure III-1. Parametrized molecule used to create the Ser203-VX adduct.

The protocol for parametrization is described in the supporting information of the 2011 *J. Chem. Theory Comput.* article of P. Ren, C. Wu and J. W. Ponder and briefly summarized here.[12] The first step is the optimization of the substrate using Gaussian09[13] at the MP2 level of theory[14], using the 6-311G(d,p) basis set.[15] A distributed multipole analysis is then performed using the program GDMA.[16] The poledit utility of Tinker is then used to derive monopoles and preliminary multipole moments from the distributed multipole analysis. The potential utility of Tinker is finally used to fit the atomic multipoles to the electrostatic potential of the MP2/6-311G(1d,1p) calculation. The parameters generated using this protocol were then added to the standard AMOEBAbio09 parameters. The parameters generated this way however, once combined with the regular parameters for serine produced a non-neutral residue. To circumvent this problem, a new molecule was parametrized,

[11] Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, P. Ren, *J. Chem. Theory Comput.* **2013**, *9*, 4046–4063.
[12] P. Ren, C. Wu, J. W. Ponder, *J. Chem. Theory Comput.* **2011**, *7*, 3143–3161.
[13] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, et al., *Gaussian 09 Revision A.01*, Gaussian Inc., Wallingford CT, **2009**.
[14] J. A. Pople, J. S. Binkley, R. Seeger, *Int. J. Quantum Chem.* **1976**, *10*, 1–19.
[15] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, *72*, 650–654.
[16] A. J. Stone, *J. Chem. Theory Comput.* **2005**, *1*, 1128–1132.

resembling more closely the VX-serine adduct (see **Figure III-2**). From this molecule $P(O)(OEt)(CH_3)_2$-$CH_2CH(CO\text{-}NH\text{-}CH_3)(NH\text{-}CO\text{-}CH_3)$, the parameters for $P(O)OEtCH_3$ where used to replace the CH_2-OH side chain of a regular serine.

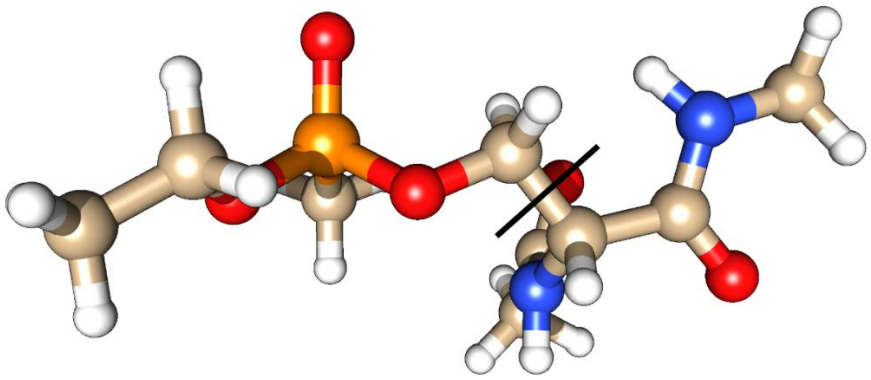


Figure III-2. Parametrized molecule used to create the Ser203-VX adduct.

The parametrization protocol was used again on this molecule to obtain parameters for the $P(O)OEtCH_3$ moiety. This moiety was combined to standard serine parameters, to produce a set for parameters for the Ser203-VX adduct. Finally, the monopoles were adjusted to obtain a neutral residue. The final parameters are presented in **Appendix 3**.

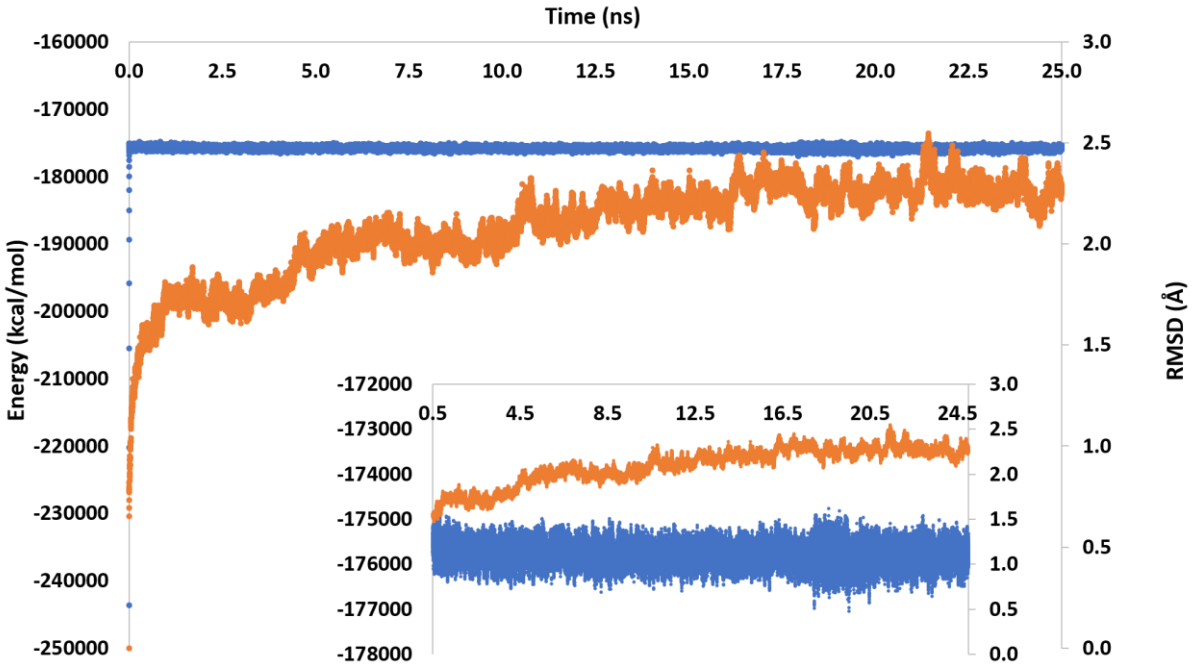


Figure III-3. Energy of the equilibration and following molecular dynamic trajectory in blue ($kcal.mol^{-1}$). RMSD in orange (in Å). The lower segment of the figure is occupied by the same data reframed without the first 500 ps.

Once the VX adduct has been parametrized for the two VX-inhibited AChE systems, the simulations have been performed. The system was first minimized and then heated up to 300 K. This was followed by 25 ns long trajectories for all four systems, VX-inhibited AChE in both protonation states of the side chain of Glu202 and non-inhibited AChE in both protonation states of the side chain of Glu202. The total energy of the VX-inhibited AChE with protonated

Glu202 system as well as the RMSD of non-hydrogen atoms for the same system are presented in **Figure III-3**.

As can be seen in **Figure III-3** the energy of the system is very rapidly equilibrated and remains stable throughout. The RMSD however stabilizes only after 15 ns and remain stable for the last 10 ns of the dynamic. The other trajectories will need to be checked when finished for their stability and probably extended by an additional 15 ns to reach 25 ns of equilibrated molecular dynamic.

The data and the setup used to produce it will be a starting point for the PhD project of Frederic Célerse. This project is to pursue the study of AChE with polarized force fields under the joint supervision of Etienne Derat at IPCM and Jean-Philip Piquemal at LCT lab.

IV. Constant pH molecular dynamic simulations

Constant pH molecular dynamic (CpHMD) simulations have been used primarily to evaluate the protonation state of key residues of AChE at physiological pH of 7.2.[17] The results obtained at different pH values allowed to draw titration curves for the studied residues and to obtain the pK_a. The CpHMD simulations have been performed in Amber16[18], with the Generalized Amber force field (GAFF) FF14SB.[19] The principle for CpHMD simulations as implemented in Amber16 is first described by Mongan, Case and McCammon[20] and modified for explicit solvents by Swails, York and Roitberg[21]. This section of the thesis project is the result of a collaboration with Professor Kamerlin and most of the simulations were performed during a 3 months stay in her research group at Uppsala University.

The parameters for the Ser203-VX adduct were constructed using the tutorials available on the Amber website.[22] The Ser203-VX adduct (see **Figure IV-1**) was first fully optimized with DFT functional M06-2X[23] in the 6-31G(d) basis set[24] in Gaussian09.[13] This calculation is use to get the electrostatic potential according to the RESP procedure.[25] Antechamber was then used to calculate the partial charges and parameters for the Ser203-VX adduct. Those parameters can be found in **Appendix 4**.

[17] R. Vroman, L. J. Klaassen, M. H. C. Howlett, V. Cenedese, J. Klooster, T. Sjoerdsma, M. Kamermans, *PLoS Biol.* **2014**, *12*, e1001864.

[18] D. A. Case, R. M. Betz, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, et al., *Amber 2016*, University of California, San Francisco, **2016**.

[19] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

[20] J. Mongan, D. A. Case, J. A. McCammon, *J. Comput. Chem.* **2004**, *25*, 2038–2048.

[21] J. M. Swails, D. M. York, A. E. Roitberg, *J. Chem. Theory Comput.* **2014**, *10*, 1341–1352.

[22] a) J. Swails, D. Case, T.-S. Lee, "Amber Basic Tutorials - Tutorial B5," can be found under http://ambermd.org/tutorials/basic/tutorial5/, **2015**. b) R. Walker, S. Tang, "Antechamber Tutorial," can be found under http://ambermd.org/tutorials/basic/tutorial4b/, **2015**.

[23] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215–241.

[24] V. A. Rassolov, M. A. Ratner, J. A. Pople, P. C. Redfern, L. A. Curtiss, *J. Comput. Chem.* **2001**, *22*, 976–984.

[25] P. Cieplak, W. D. Cornell, C. Bayly, P. A. Kollman, *J. Comput. Chem.* **1995**, *16*, 1357–1377.

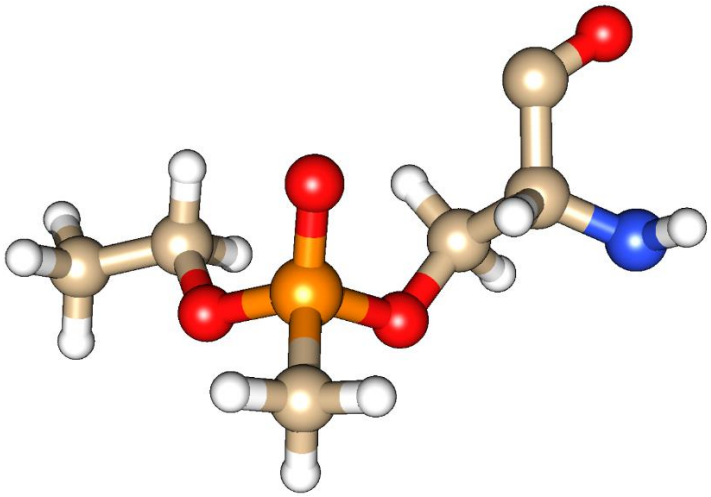


Figure IV-1. Parametrized Ser203-VX adduct

Once the VX-Ser203 adduct has been parametrized, the CpHMD simulations have been prepared as per the Amber CpHMD tutorial available on the Amber website.[26] The 45 residues susceptible to protonation state change (glutamates, aspartates and histidines) closest to the active site were selected. The system was then put in a solvent octahedron, minimized, heated and finally equilibrated for 109 ps at 300 K with a 2 fs time step and a write out every 100 ps under periodic boundary conditions. After the first 34 ns a snapshot was drawn every 5 ns to generate 15 seeds. Each of those 15 seeds was used to start eight 100 ns long constant pH molecular dynamic simulations at pH values 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0 and 10.0 to have simulations at 3 pH units around the physiological pH of 7.2. The range was increased from 4 to 3 in the acids because of the standard pKa of glutamates and aspartates (around 4). In those simulation, a protonation state change was attempted every 100 step which is every 0.2 ps.

Besides the Cartesian coordinates of all atoms and the total energy of the system at every point, CpHMD simulations yield for every residue selected as titratable the fraction of the simulation time this residue was protonated. From this fraction at all 8 pH values over all 15 seeds, a titration curve can be drawn for each of the 45 selected titratable residues using the Hill equation. The calculated pK_a are listed in **Table IV-1**.

Table IV-1. Calculated pK_a values with CpHMD, standard error of the pK_a determination from the titration curves, and proportion protonated at physiological pH 7.2.

Residue	Asp74	Glu81	Glu84	Glu91	Glu94	Asp95	Asp131	Asp134	Glu142
Calculated pK_a	4.41	5.23	5.04	4.73	4.24	3.85	3.05	3.70	3.97
Standard Error	0.03	0.04	0.03	0.01	0.01	0.00	0.03	0.18	0.02

[26] J. Swails, T. D. McGee Jr, "AMBER Advanced Tutorial 18 - Constant pH MD: Introduction," can be found under http://ambermd.org/tutorials/advanced/tutorial18/, **2013**.

Proportion protonated	0.16%	**1.07%**	0.69%	0.34%	0.11%	0.04%	0.01%	0.03%	0.06%

Asp175	**Glu202**	**His212**	**His223**	**Glu243**	**Asp280**	**Asp283**	**His284**	**Glu285**
null	7.59	6.57	5.25	2.14	1.18	3.84	8.08	4.05
null	0.15	0.02	0.06	0.03	0.05	0.00	0.07	0.02
null	**71.00%**	**19.08%**	**1.10%**	0.00%	0.00%	0.04%	**88.40%**	0.07%
His287	**Glu292**	**Asp304**	**Asp306**	**Asp310**	**Glu313**	**Asp320**	**Asp323**	**Asp333**
6.61	5.83	3.01	4.16	4.10	4.38	4.18	3.53	2.49
0.01	0.02	0.01	0.00	0.00	0.03	0.04	0.03	0.24
20.42%	**4.10%**	0.01%	0.09%	0.08%	0.15%	0.10%	0.02%	0.00%
Glu334	**Asp349**	**Glu351**	**His381**	**His393**	**Asp396**	**Asp404**	**His405**	**Glu431**
2.41	3.98	4.74	7.05	8.65	3.61	null	6.04	0.72
0.04	0.00	0.01	0.01	0.03	0.01	null	0.03	0.09
0.00%	0.06%	0.35%	**41.26%**	**96.55%**	0.03%	null	**6.48%**	0.00%
His432	**His447**	**Glu450**	**Glu452**	**Asp460**	**Glu469**	**Asp488**	**Asp491**	**Glu519**
3.73	2.97	1.15	6.15	2.69	6.06	3.00	2.16	3.28
0.09	0.43	0.82	0.02	0.04	0.04	0.00	0.03	0.01
0.03%	0.01%	0.00%	**8.24%**	0.00%	**6.74%**	0.01%	0.00%	0.01%

For two residues, Asp175 and Asp404, the pK_a value could not be calculated. Those two residues remained unprotonated more than 99.99% of the simulation time at all pH values. A majority of glutamates and all aspartate residues are in their expected basic form due to their high acidity. It is the case for Asp74, Glu84, Glu91, Glu94, Asp95, Asp131, Glu142, Asp175, Glu243, Asp280, Asp283, Glu285, Asp304, Asp306, Asp310, Glu313, Asp320, Asp323, Asp333, Glu334, Asp349, Glu351, Asp396, Asp404, Glu431, Asp460, Asp488, Asp491, and Glu519. Some glutamate residues have pK_as indicating a possible ability to be transient proton acceptors at pH 7.2. For Glu81, Glu292, Glu452 and Glu469 their probability to be protonated at pH 7.2 based on their calculated pK_a is superior to 1% and inferior to 10%. All four are solvent accessible and so are available for protonation and subsequent deprotonation by the solvent (see **Figure IV-2**). Finally, based on CpHMD, a single glutamate is likely to remain protonated most of the time, Glu202. With a pK_a of 7.59 ± 0.15 it is likely to be protonated more than 70% of the time.

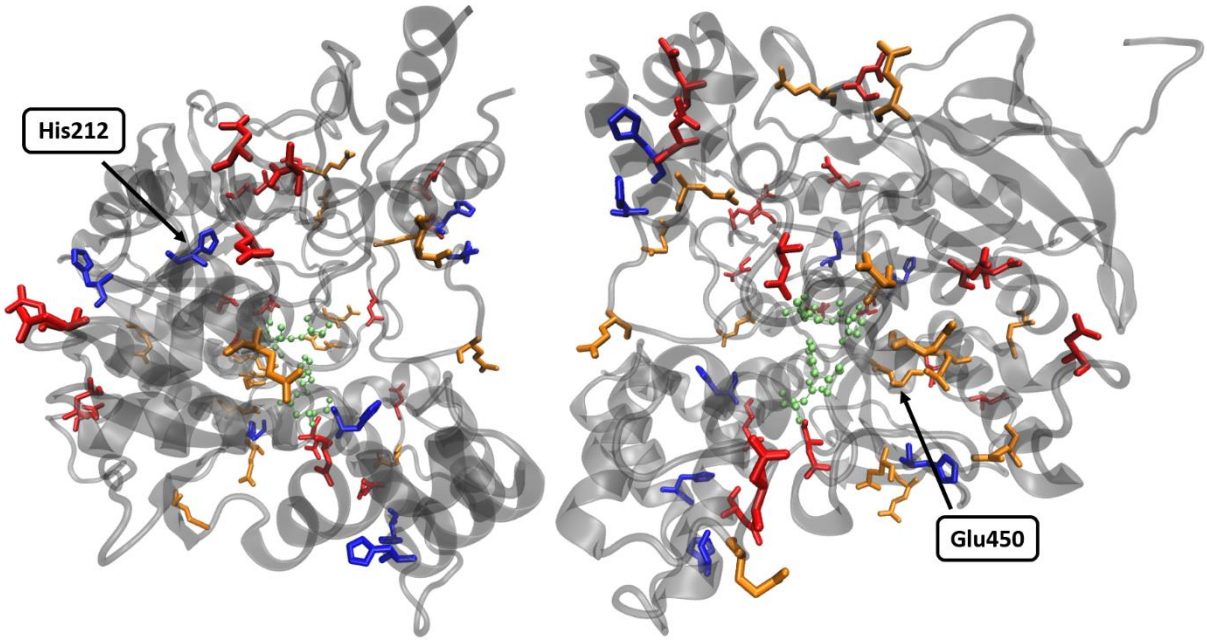


Figure IV-2. Two views of AChE illustrating the position of titratable residues. Aspartates are in red, Glutamates in orange and histidines in blue, all in licorice representation. His447, Glu334, Glu202 and the Ser03-VX adduct are in the centre in lime and in Ball&Stick representation. The two labelled residues are those that are not solvent accessible.

The standard water solution pK_a of histidines being 6.0, they are expected to often be transient proton acceptors. For two histidines, His432 and His447, the CpHMD simulations predict their imidazole side chain remains constantly singly-protonated in their imidazole form. Five histidines have calculated pK_as that enable them to be transient proton acceptors. These histidines are His212, His223, His287, His381, and His405 with respective pK_as of 6.57, 5.25, 6.61, 7.05, and 6.04. They are respectively susceptible to be protonated 19.08, 1.10, 20.42, 41.26, and 6.48 percent of the time. Four out of five of these histidines, His223, His287, His381, and His405, are in contact with bulk solvent and thus in positions where they can be protonated or deprotonated by the solvent. Only His212 appears to be inaccessible to solvent and thus less accessible to a protonation state change. Two histidines, His284, and His393 are likely to be protonated most of the time with pK_as of 8.08 and 8.65 respectively. Their side chain is likely to remain in its imidazolium form during most of the enzyme's activity but to occasionally become unprotonated as both are surface residues and thus in contact with bulk solvent.

Five residues were of special interest. Glu334 and His447 are both residues of the catalytic triad. Their CpHMD obtained titration curves are presented in **Figure IV-3**.

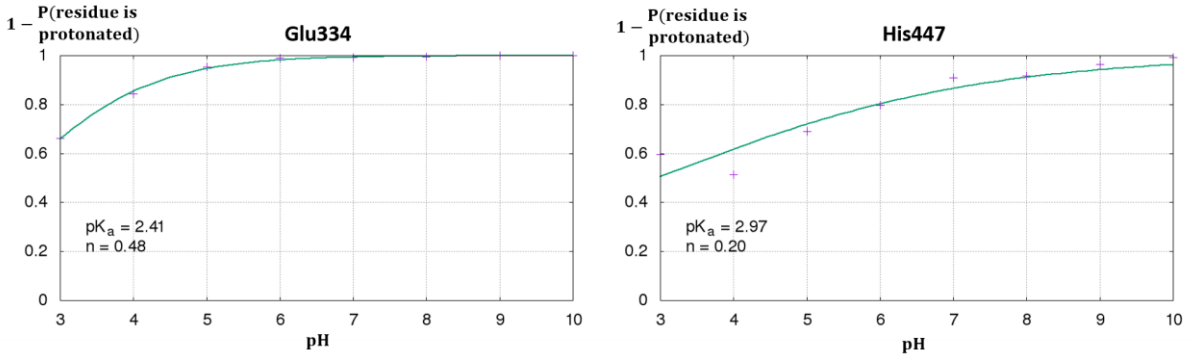


Figure IV-3. Titration curves of Glu334 and His447 obtained from CpHMD calculations. The inverse probability for the residue to be protonated is shown as a function of pH.

As can be seen in **Table IV-1**, from their CpHMD calculated pK_a value, Glu334 and His447 are very unlikely to be protonated at physiological pH of 7.2. For Glu334, the inverse probability of having the residue protonated during the simulation as a function of pH (**Figure I-1**) is in agreement with the prediction based on pK_a obtained through the Hill equation. It shows that the probability for Glu334 to be protonated is close to zero at pH superior to 6.0. For His447, the titration curve shows a non-zero probability of being protonated at physiological pH. The Hill coefficient of 0.20 for the titration curve of His447 indicates a high level of cooperativity for His447 and a probable link between the protonation state of His447 and nearby residues. According to CpHMD simulations at pH values 7 and 8, His447 has a 9% probability of being protonated. This data is consistent with the experimental observation that Glu334 remains unprotonated and forms a strong hydrogen bond with His447.[27] This data also concurs with the role of His447 as a possible transient proton acceptor in the regular catalytic activity of AChE, the inhibition of AChE by nerve agents, and potentially the reactivation as was show in **Chapter 4**.[28]

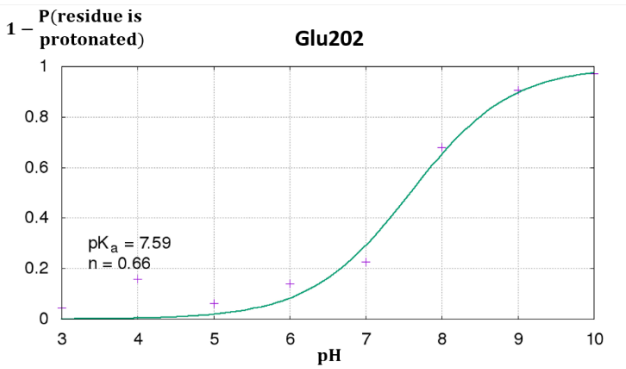


Figure IV-4. Titration curves of Glu202 obtained from CpHMD calculations. The inverse probability for the residue to be protonated is shown as a function of pH.

[27] a) H. Tsukada, D. M. Blow, *J. Mol. Biol.* **1985**, *184*, 703–711. b) M. A. Massiah, C. Viragh, P. M. Reddy, I. M. Kovach, J. Johnson, T. L. Rosenberry, A. S. Mildvan, *Biochemistry* **2001**, *40*, 5682–5690. c) H. Tsukada, D. M. Blow, *J. Mol. Biol.* **1985**, *184*, 703–711.

[28] a) Y. Zhou, S. Wang, Y. Zhang, *J. Phys. Chem. B* **2010**, *114*, 8817–8825. b) G. S. Sirin, Y. Zhang, *J. Phys. Chem. A* **2014**, *118*, 9132–9139. c) A. da Silva Gonçalves, T. C. C. França, M. S. Caetano, T. C. Ramalho, *J. Biomol. Struct. Dyn.* **2014**, *32*, 301–307.

The plot of the probability for Glu202 to be unprotonated as a function of pH is presented in **Figure IV-4**. The simulations at pH 7 indicate a protonated fraction of 77% and the simulations at pH 8 a protonated fraction of 32%. This seems to be in agreement with the pK_a value calculated from the fitting to the Hill equation. There is however a divergence from the fitted titration curve at low pH where Glu202 retains some ability to be unprotonated. The divergence of the Hill coefficient n from 1 reduces the steepness of the curve and conveys a possible influence of the protonation state of nearby residues (His447 or Glu450 for instance) on the protonation state of Glu202. This data is coherent with the literature as well as the work presented in this manuscript in **Chapter 4**.[29] Two other residues of AChE were of special interest, Glu450 and Glu452. Glu450 is a buried residue close to Glu202 and Glu452 is a residue in contact with bulk solvent in the vicinity of Glu450 (see **Figure IV-5**).

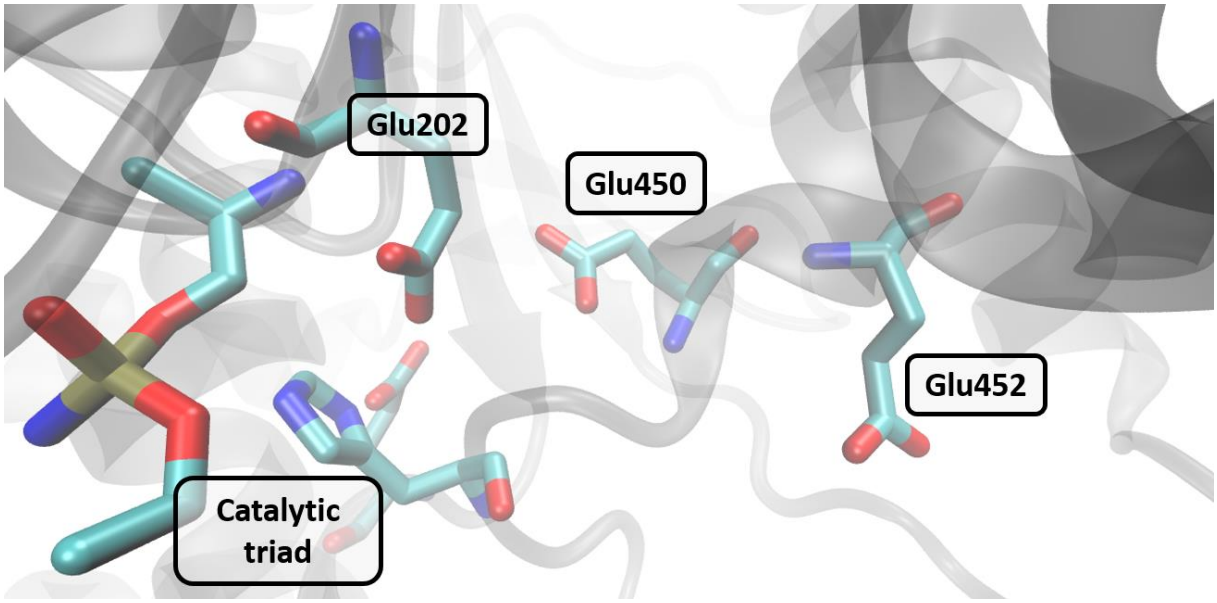


Figure IV-5. Position of Glu450 and Glu452 relative to the catalytic triad and Glu202

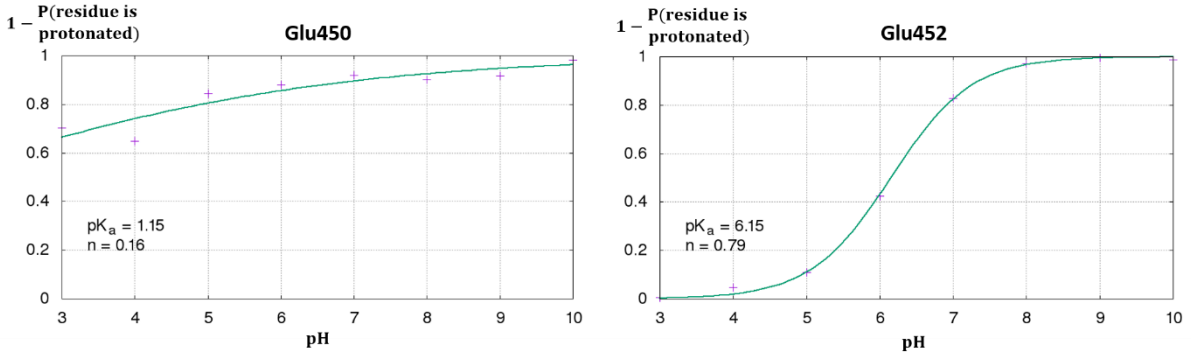


Figure IV-6. Titration curves of Glu450 and Glu452 obtained from CpHMD calculations. The inverse probability for the residue to be protonated is shown as a function of pH.

[29] a) P. Vagedes, B. Rabenstein, J. Åqvist, J. Marelius, E.-W. Knapp, *J. Am. Chem. Soc.* **2000**, *122*, 12254–12262. b) S. V. Lushchekina, I. A. Kaliman, B. L. Grigorenko, A. V. Nemukhin, S. D. Varfolomeev, *Russ. Chem. Bull.* **2011**, *60*, 2196–2204.

The titration curves of Glu450 and Glu452 obtained from fitting of the CpHMD data to the Hill equation are presented in **Figure IV-6**. The fitted curve for Glu450 is very flat and shows that the probability for this residue to be protonated is almost stable around 10 % for pH between 6 and 9. The CpHMD data contradicts the protonation probability based on the pK_a due to the low Hill coefficient (see subsection **II.4** of **Chapter 2** where the Hill coefficient is introduced) of 0.16 indicating that the protonatability of Glu450 is cooperative and not strictly set by its pK_a. For Glu452, the CpHMD data indicates a probability of the side chain to be protonated in the range of 10% which is coherent with the probability that could be calculated from the pK_a obtained through the fitting to the Hill equation. It confirms that Glu452 can be transiently protonated at physiological pH which is a possibility as Glu452 is in contact with bulk solvent. The Hill coefficient indicates much less cooperativity for the protonation state of Glu452 compared to Glu450.

CpHMD is only as good as the force field used for the simulations and in the case of protonation states, the polarizability a force field like AMOEBA could bring is sorely lacking. Yet these results confirm the expected trends for Glu334, Glu202 and His447 and seem to indicate two residues, Glu450 and Glu452 with a respective secondary and tertiary connection to the triad are capable of being protonated. It indicates a potential avenue of proton diffusion in and out of the active site whose evaluation is presented in the next two sections.

V. A novel mechanism for proton diffusion in and out of the active site by QM/MM

The hypothesis that was formulated based on the study of the environment of Glu202 and Glu450 is that a proton could be transferred through several steps from Glu452 to Glu202 which is the door to the active site. The proposed mechanism for this transfer of protons is that first a proton is transferred from Glu452 to the nitrogen of the amide between Tyr449 and Glu450 (see **Figure V-1**). The hydrogen already bonded to the nitrogen is consecutively or concurrently transferred to Glu450. This proton would be then accessible to be transferred to Glu202. The possibility of a proton transfer between Glu450 and Glu202 is not a shocking proposition as both residues are commonly either in direct hydrogen bond or indirectly hydrogen bonded through a water molecule. The transfer of a proton through an amide bond has also been discussed in the literature but is not considered a standard feature of enzymes and has never been discussed for AChE.[30] The current conscensus for the transfer of a proton through amide bonds occurs via O-protonation. In AChE, however, for the amide bond between Tyr449 and Glu450, the oxygen is not available to transfer a proton to Glu450. We will now demonstrate that the proton transfer through the amide occurs via N-protonation.

[30] a) C. L. Perrin, *Accounts of Chemical Research* **1989**, *22*, 268–275. b) K. Shimokata, Y. Katayama, H. Murayama, M. Suematsu, T. Tsukihara, K. Muramoto, H. Aoyama, S. Yoshikawa, H. Shimada, *Proc. Nat. Ac. Sc. USA* **2007**, *104*, 4200–4205. c) K. Kamiya, M. Boero, M. Tateno, K. Shiraishi, A. Oshiyama, *J. Am. Chem. Soc.* **2007**, *129*, 9663–9673.

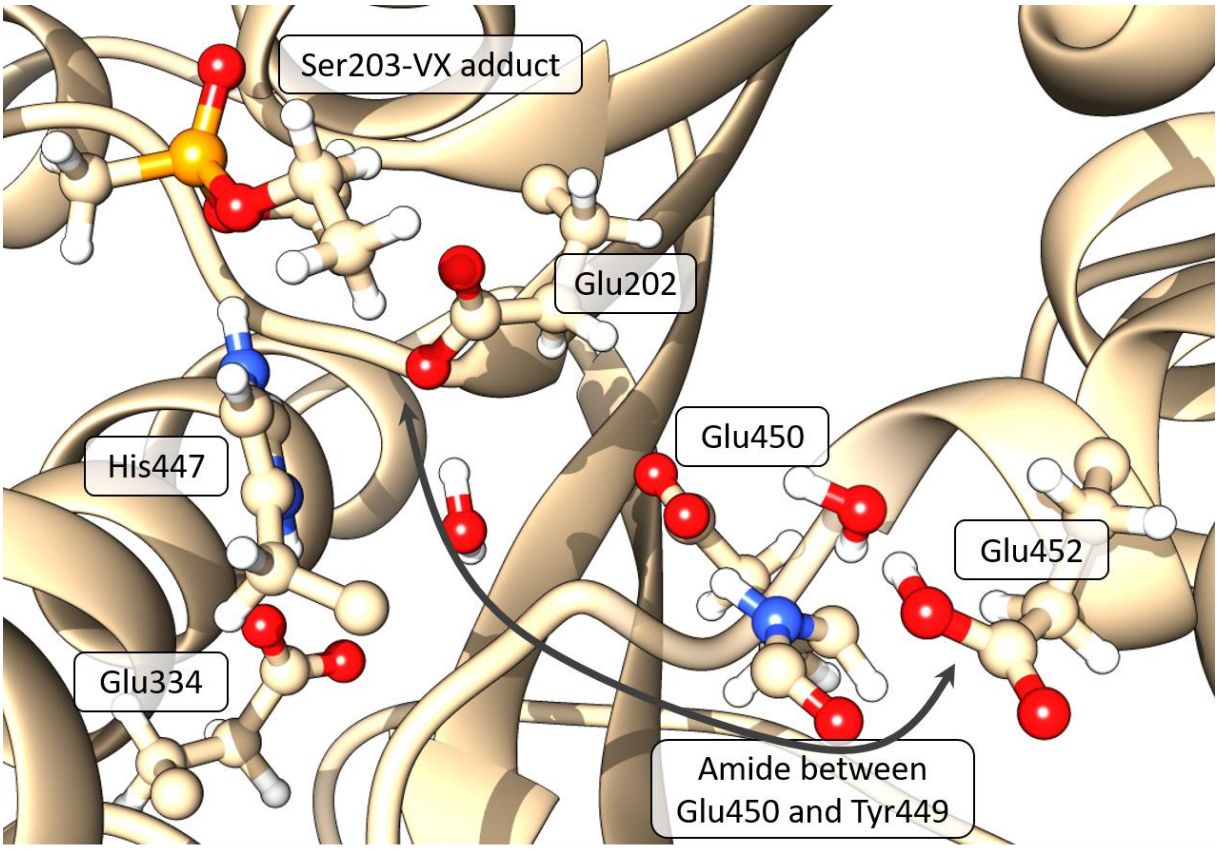


Figure V-1. Putative pathway for a proton entrance or exit from the active site of AChE, circumventing the channel of AChE.

This putative pathway was studied in VX-inhibited and non-inhibited AChE. The setup for VX-inhibited AChE is the one described in section **II**, page **227** of this chapter. For VX free AChE, the setup from section **II** was used as a starting point. A 16 Ångström thick shell was added around the system to the existing 8 Ångström thick shell to bring the water molecule shell around the enzyme to 24 Ångström thick using Chimera 1.10.1.[1] The system was then equilibrated for 10 ns, with the outer 16 Å thick water molecules shell kept frozen, using NAMD 2.9.[2] A snapshot was extracted from this dynamic to perform QM/MM on. Finally, the outer 16 Å of water molecules used to keep the molecular dynamics in a bubble are removed. The QM and MM regions used in these simulations are presented in **Figure V-1bis** and **Figure V-1ter** respectively. The QM region is composed of the catalytic triad with the VX adduct, the side chains of Glu20, Glu450 and Glu452, the amide bonds between Gly448 and Tyr449 and between Tyr449 and Glu450, Tyr428, Ser229 and three water molecules involved in hydrogen bonds with either Glu02, Glu450 or Glu452. The MM region includes all residues and water molecules in the vicinity of the QM region.

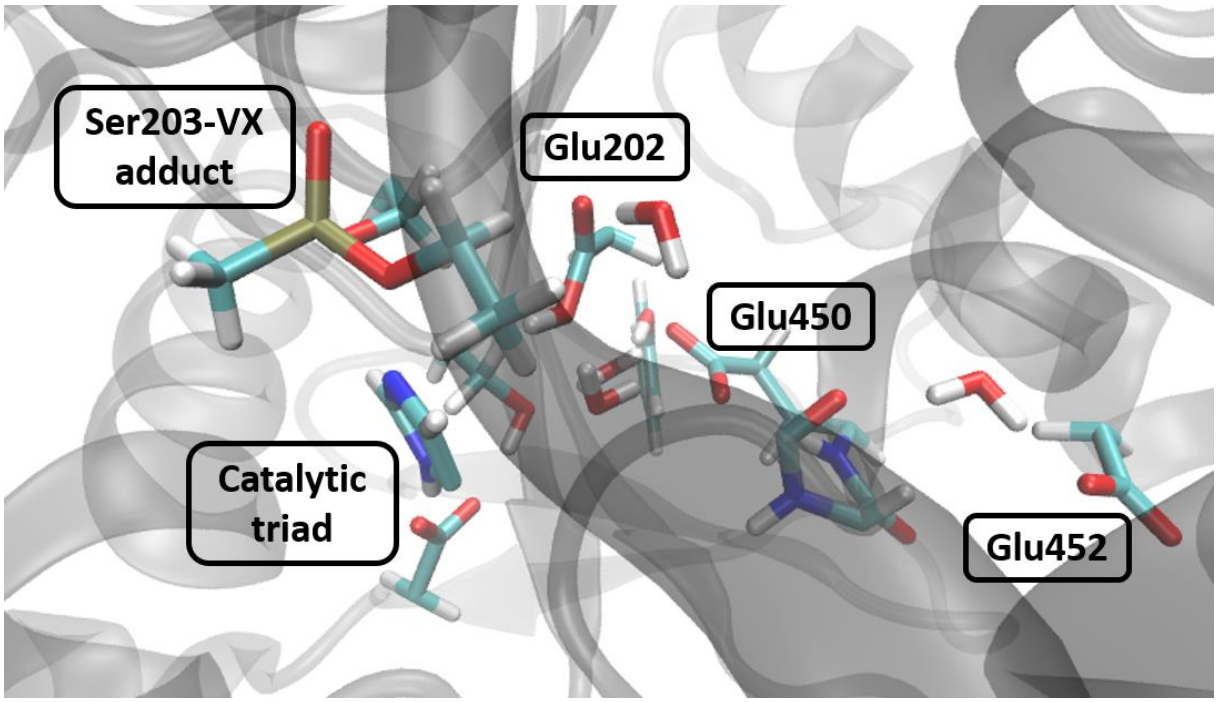


Figure V-1bis. QM region for the QM/MM simulations of the proton relay mechanism. Residues included in the QM region are in licorice representation.

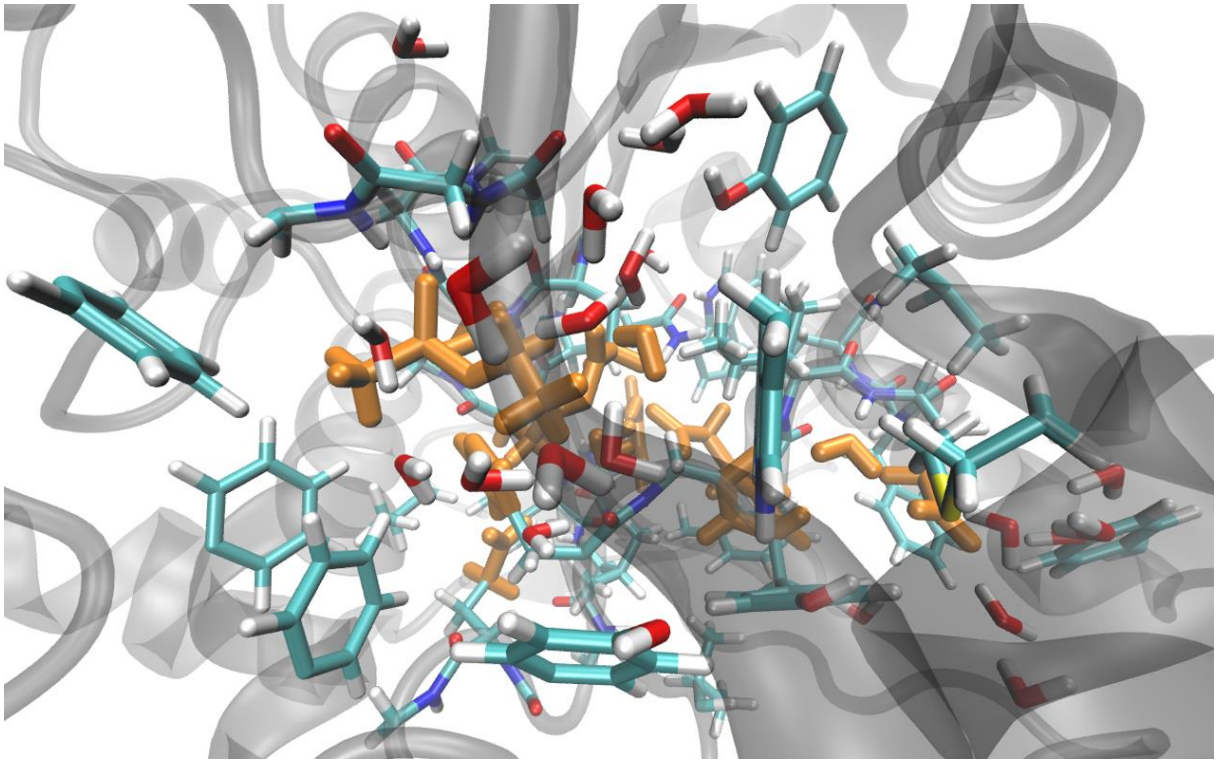


Figure V-1ter. MM region for the QM/MM simulations of the proton relay mechanism. Residues included in the MM region are in licorice representation, residues of the QM region are in orange.

All the transition states were located though mono-dimensional scans with a single distance under constraint, between the proton and the receiving atom. This reaction distance was made to vary in 0.05 Å increments and when the area of the TS was located, the precise TS was obtained by a finer scan with 0.01 Å steps.

V.1. VX-inhibited AChE with bis-protonated His447

The simulations of the proton transfer from Glu452 to Glu202 in VX-inhibited AChE have been performed with a bis protonated His447 with the side chain in imidazolium form. The energy profile of this process is presented in **Figure V-2** and the corresponding structures are presented in **Figure V-3**.

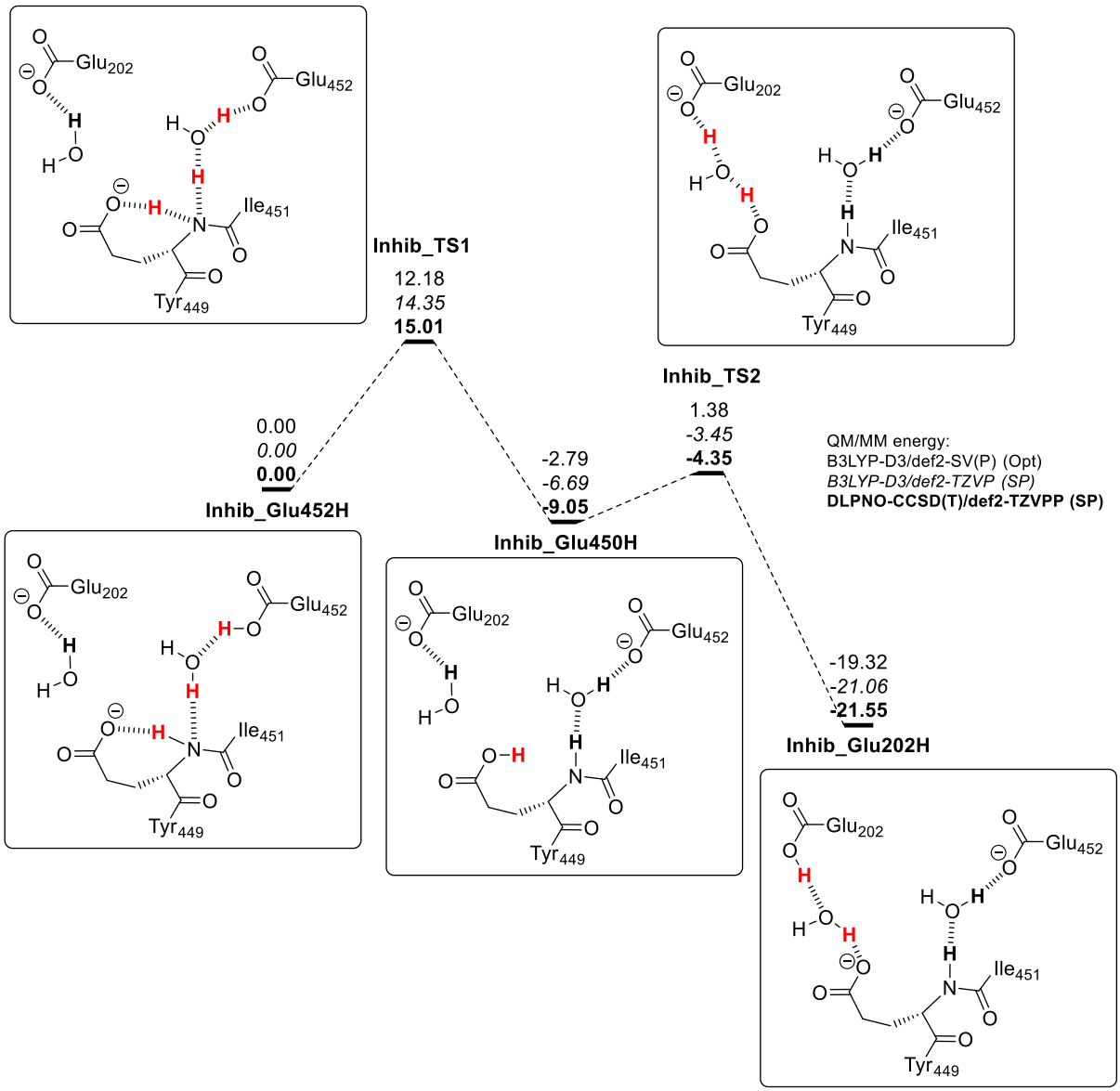


Figure V-2. Energy profile of the proton transfer from Glu452 to Glu202 in VX-inhibited AChE. Energies in kcal.mol^{-1}.

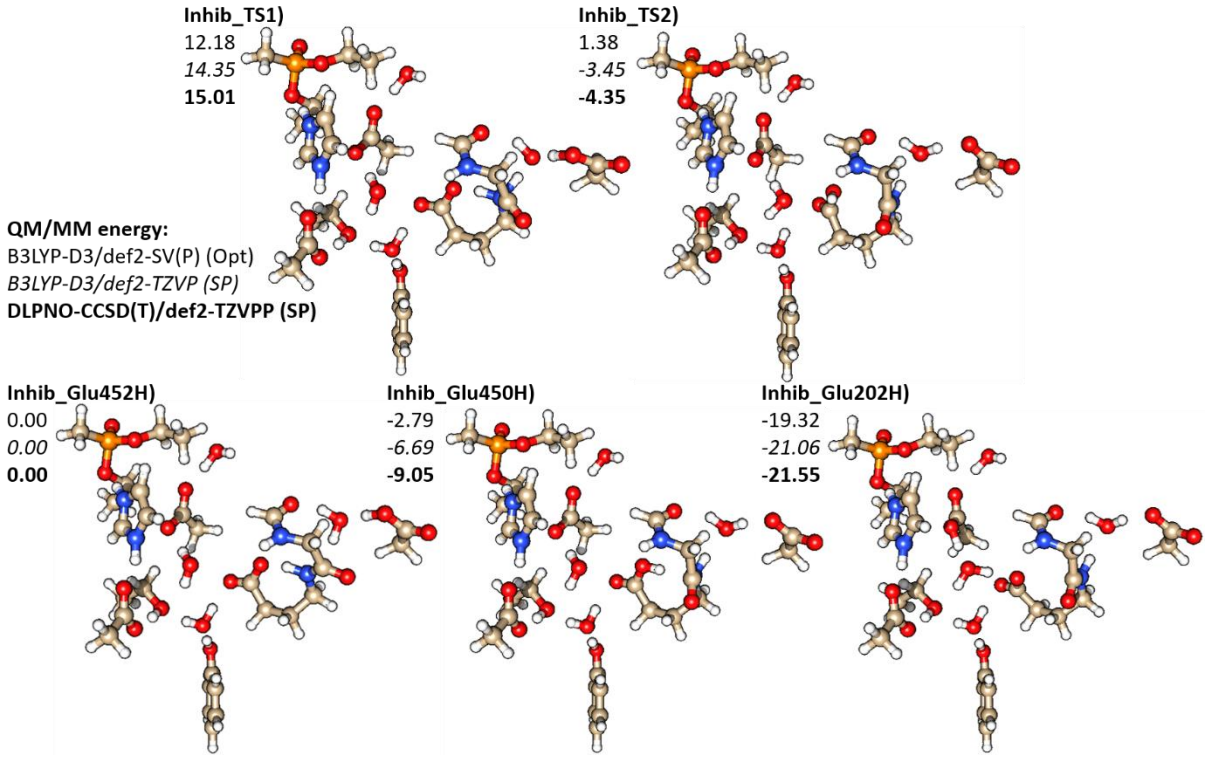


Figure V-3. Optimized structures of (Inhib_Glu452H) the reactant, (Inhib_TS1) the first transition state, (Inhib_Glu450H) the stable intermediate, (Inhib_TS2) the second transition state, and (Inhib_Glu202H) the product for the transfer of a proton from Glu452 to Glu202 in VX-inhibited His447H$^+$ AChE.

The proton transfer from Glu452 to Glu202 appears to be a two-step process. The first step involves the transfer of three protons from Glu452 to a nearby water molecules, from the same water molecule to the nitrogen of the amide between Tyr449 and Glu450, and from this amide to Glu450. This process was found to have and energy barrier of 12.18 kcal.mol^{-1} at the B3LYP-D3/def2-SV(P) level that was used for geometry optimization. This step produces an energy difference compared to the proton on Glu452 of -2.79 kcal.mol^{-1}. Single points using the def2-TZVP basis set give and energy barrier of 14.35 kcal.mol^{-1} and an energy difference of -6.69 kcal.mol^{-1}. Those energies are confirmed by the DLPNO-CCSD(T)/def-TZVPP single points which give similar energies with a barrier of -15.01 kcal.mol^{-1} and an energy difference -9.05 kcal.mol^{-1}. The geometry of the transition state of this step is presented in **Figure V-4**. The peptide bond between Tyr449 and Glu450 is a tetracoordinated and tetrahedral amidium in the transition state, having received a proton from a nearby water molecule and not yet ceded a proton to Glu450. The bond between the proton the nitrogen received from the water molecule is 1.27 Å long. The bond between the nitrogen and the proton to be ceded to Glu450 is 1.11 Å long. This step in concurrent to a rotation of the amide bond. In **Inhib_Glu452H** the oxygen of the amide is pointed towards Glu452, but in **Inhib_Glu450H**, after the proton transfer through the amide, and more visibly in **Inhib_Glu202H**, it is oriented towards Glu450.

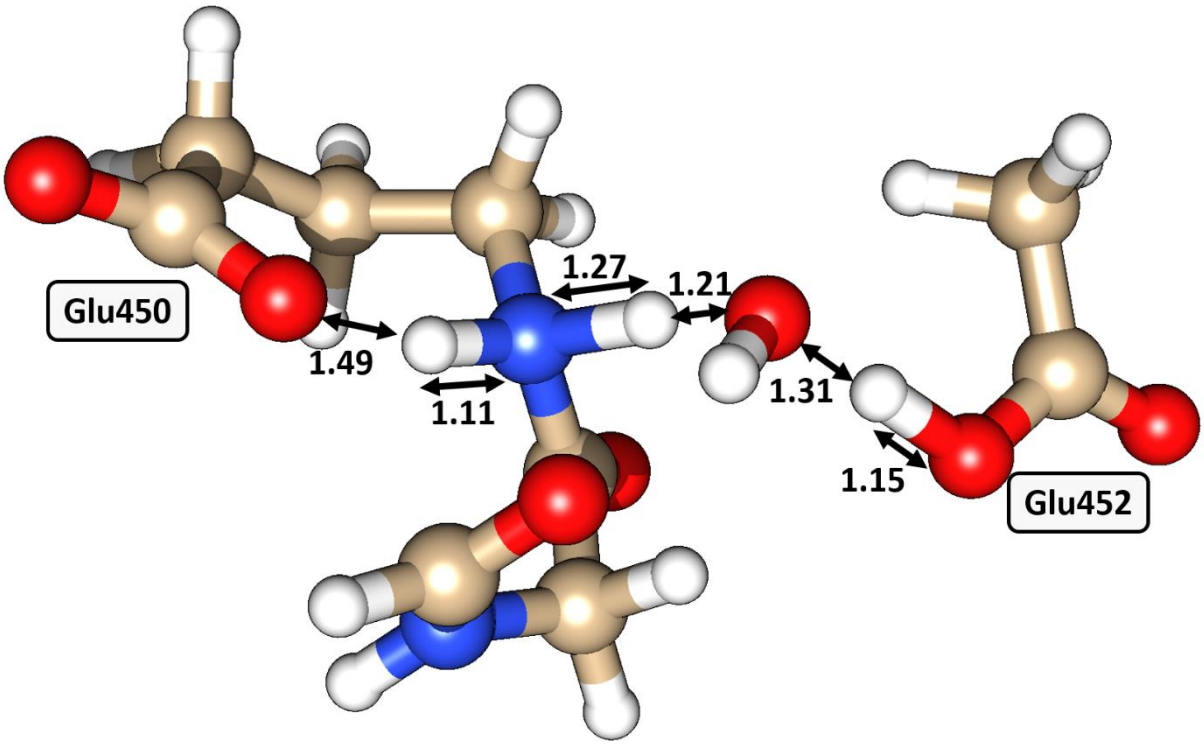


Figure V-4. Structure of the transition state for the proton transfer step through an amide bond. Case of Vx-inhibited AChE with a bis-protonated His447.

The second step of the proton transfer between Glu452 and Glu202 involves the transfer of a proton between Glu450 and a nearby water molecule which can then in turn transfer a proton to Glu202. Before this double proton transfer however, the proton on Glu450 must change position to be accessible to the water molecule (see **Inhib_Glu450H** in **Figure V-3**). Attempts to optimize a structure with the proton of Glu450 oriented towards Glu202 or rotated in any way failed. In those cases, the reaction spontaneously occurs with both proton transfers. The rotation of the proton to be accessible to the water molecule is the energetically costly process of this step. The rest is very energetically favoured and occurs spontaneously. This is illustrated by the low energy barrier and high exothermicity of this step. In B3LYP-D3/def2-SV(P), this step has an energy barrier of 4.17 kcal.mol^{-1} with an energy difference of -16.53 kcal.mol^{-1} (with **Inhib_Glu450H**) as the zero. The def2-TZVP single points give and energy barrier of 3.24 kcal.mol^{-1} and an exothermicity of -14.37 kcal.mol^{-1}. The DLPNO-CCSD(T)/def2-TZVPP single points finally, give similar energies with an energy barrier of 4.70 kcal.mol^{-1} and an exothermicity of -12.50 kcal.mol^{-1}. The overall energy difference between **Inhib_Glu452H** and **Inhib_Glu202H** is -19.32 kcal.mol^{-1} at the B3LYP-D3/def2-SV(P) level. With the def2-TZVP basis set the energy difference is -21.06 kcal.mol^{-1}. DLPNO-CCSD(T)/def2-TZVPP calculation give something similar with an overall energy difference of -21.55 kcal.mol^{-1}. The second set of simulations focused on a non-inhibited AChE to gage the role this proton relay mechanism could have in the regular catalytic activity of the enzyme.

V.2. Non-inhibited AChE with mono-protonated His447

The simulations of the proton relay mechanism in non-inhibited AChE were performed with His447^{0}. The imidazolium side chain of His447 carries a single proton on protonation site δ oriented to form a hydrogen bond with Glu334. The protonation site ε of His447 is free of any proton and in a hydrogen bond with Ser203 able to receive its proton. The results of these simulations are presented in **Figure V-5** with the optimized structures in **Figure V-7**.

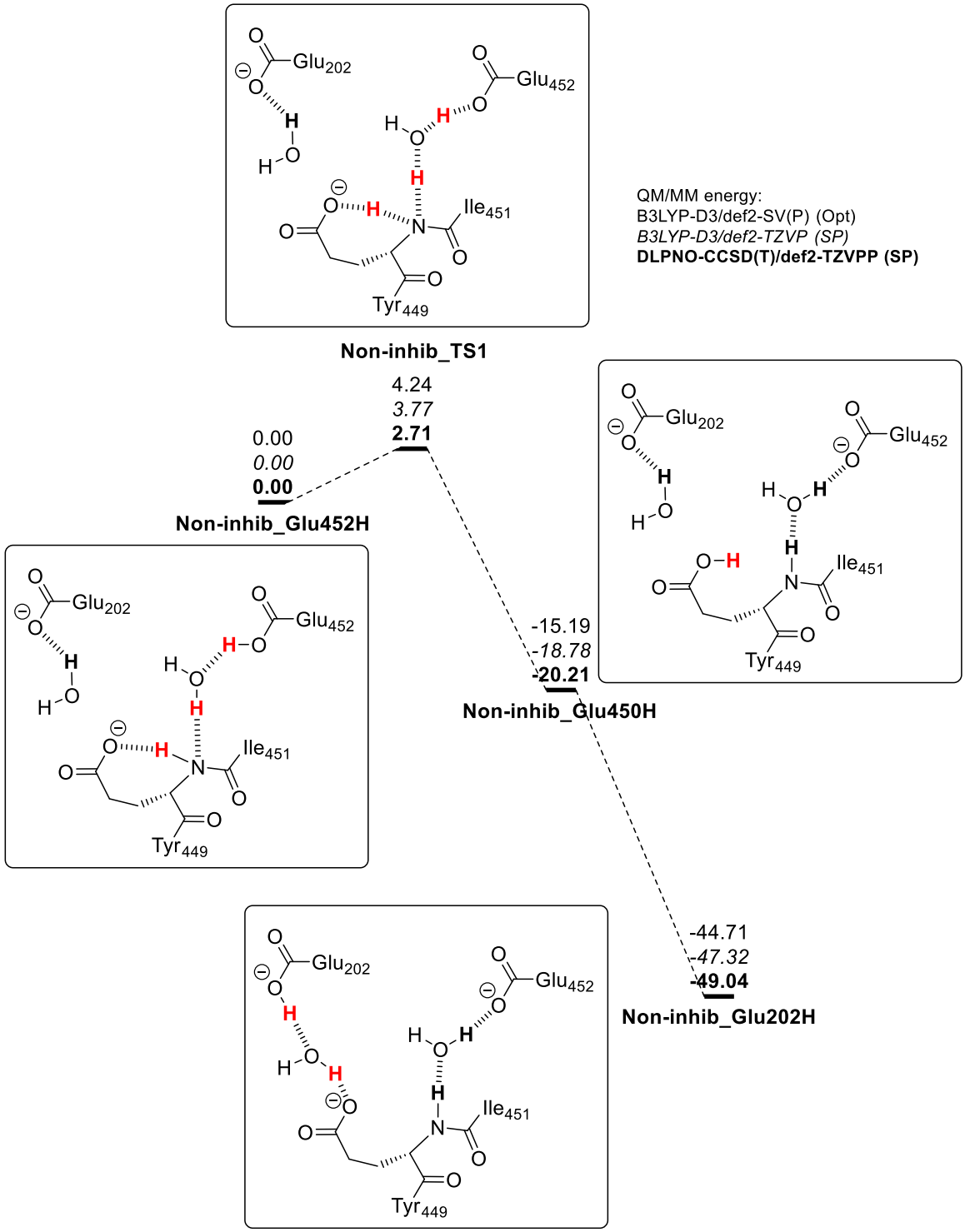


Figure V-5. Energy profile of the proton transfer from Glu452 to Glu202 in VX-inhibited AChE. Energies in kcal.mol^{-1}.

In non-inhibited AChE, the proton transfer from Glu452 to Glu202 is a two-step process. The first step, the proton transfer from Glu452 to Glu450 involves three proton transfers. The proton of Glu452 is transferred to a nearby water molecule which transfers another proton to the nitrogen of the amide between Tyr449 and Glu450. The proton already involved in the amide is transferred to Glu450. This step has an energy barrier of 4.24 kcal.mol^{-1} at the B3LYP-D3/def2-SV(P) level with an energy difference of -15.19 kcal.mol^{-1}. The def2-TZVP single points give similar results with an energy barrier of 3.77 kcal.mol^{-1} and an exothermicity of -18.78 kcal.mol^{-1}. Finally, at the DLPNO-CCSD(T) level, the energy barrier is 2.71 kcal.mol^{-1} and the exothermicity at -20.21 kcal.mol^{-1}. The geometry of the transition state is presented in **Figure V-6**.

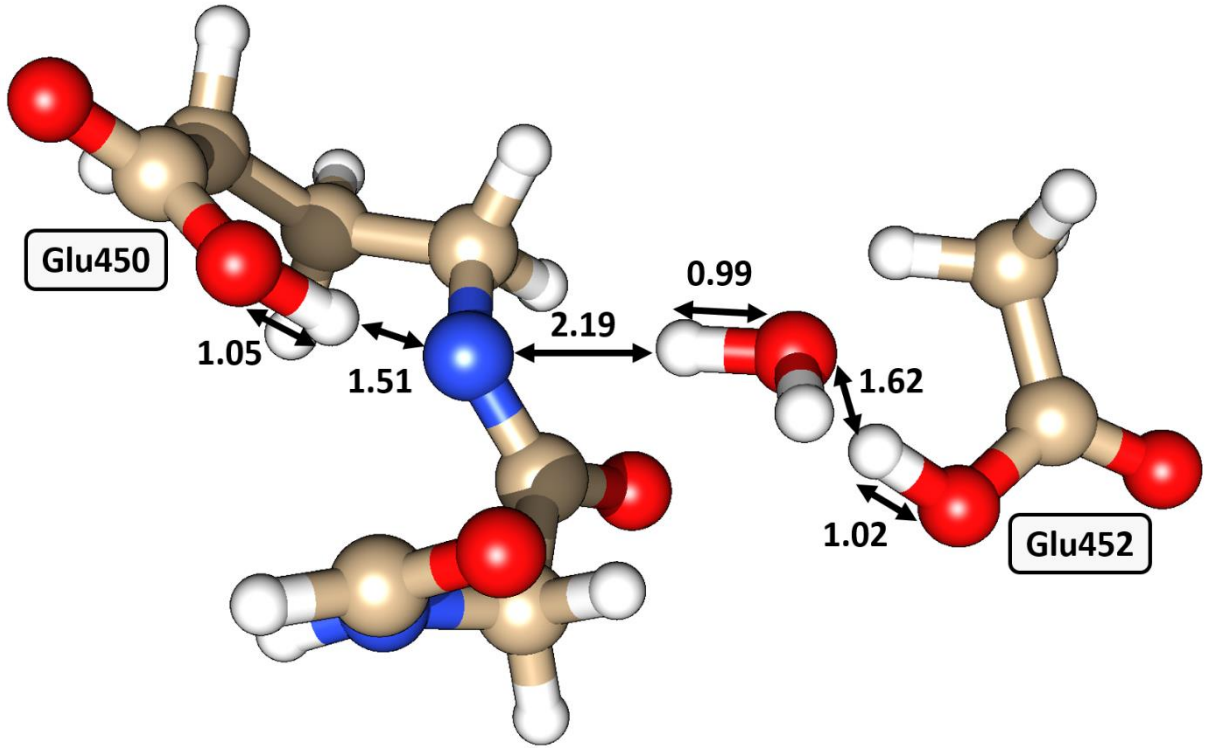


Figure V-6. Structure of the transition state for the proton transfer step through an amide bond. Case of non-inhibited His447^{0} AChE.

This transition state is completely different to the one obtained with VX-inhibited AChE. The peptide bond between Tyr449 and Glu450 is an amidate, having already transferred its proton to Glu450 and not yet received the proton of the water molecule near Glu452. The distance between the nitrogen and the proton it transferred to Glu450 is 1.51 Å long. The distance between the nitrogen and the proton of the water molecule is 2.19 Å long. Once again, this step produces a rotation of the peptide bond with the oxygen pointing towards Glu452 in **Non-inhib_Glu452H** and then rotate toward Glu450 is **Non-inhib_Glu450H**.

The second step of the proton transfer from Glu452 to Glu202 is the transfer of a proton from Glu450 to Glu202. It involves two simultaneous proton transfers, from Glu450 to a nearby water molecule which transfer one of its proton to Glu202. The transition state could not be located for this step as no energy barrier was observed in any scan. This step is highly

exothermic with an energy difference at the B3LYP-D3/def2-SV(P) level of -29.52 kcal.mol^{-1}. The B3LYP-D3/def2-TZVP and DLPNO-CCSD(T)/def2-TZVPP single points confirm this exothermicity with respective energy differences of -28.54 and -28.83 kcal.mol^{-1}. The overall process is found to be very exothermic with the energy difference between **Non-inhib_Glu452H** and **Non-inhib_Glu202H** being of -44.71, -47.32, and -49.04 kcal.mol^{-1} at the B3LYP-D3/def2-SV(P) level, the B3LYP-D3/def2-TZVP level, and the DLPNO-CCSD(T)/def2-TZVPP level respectively.

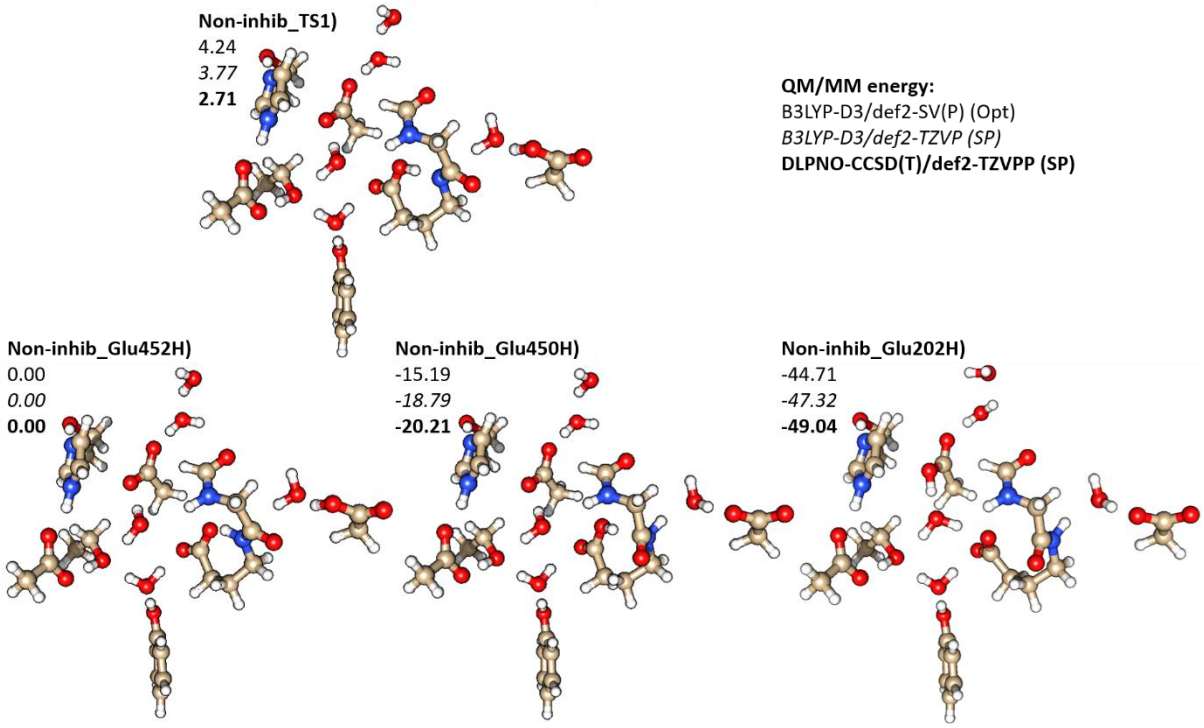


Figure V-7. Optimized structures of (Non-inhib_Glu452H) the reactant, (Non-inhib_TS1) the first transition state, (Non-inhib_Glu450H) the stable intermediate, and (Non-inhib_Glu202H) the product for the transfer of a proton from Glu452 to Glu202 in non-inhibited His447 AChE.

This process involves multiple fast proton transfer steps and while QM/MM provides a good description of both the reaction with QM treatment and the environment, the static nature of the QM/MM methodology used leaves out not only the entropic contributions to the process, but also all dynamical aspect. For these reasons, it was chosen to study this reaction with a different method.

VI. EVB simulations on proton diffusion in and out of the active site

To study dynamical aspects of the proton relay mechanism uncovered with QM/MM, it was decided to use EVB simulations. The simulations were run using the software Q.[31] They were performed during a 3 month stay in the group of Professor Lynn Kamerlin at Uppsala University. The simulations have been performed under her guidance and the setup prepared by the author with help from Paul Bauer. Finally, the setup and the preparation of calculations

[31] J. Marelius, K. Kolmodin, I. Feierberg, J. Åqvist, *J. Mol. Graphics Modell.* **1998**, *16*, 213–225.

was greatly simplified and accelerated by a set of python scripts developed by Miha Purg called qtools.

The EVB simulations were performed using the Q implementation of the OPLSAA force field.[32] Parameters are already available for standard residues but the parametrization of the Ser203-VX adduct was necessary. To parametrize this modified residue a truncated model of the VX adduct (see **Figure VI-1**) was optimized with DFT functional M06-2X[23] in the 6-31G* basis set[24] in Gaussian09.[13] Schrödinger ffld_server was used to produce parameters in OPLSAA format for this molecule.[33] One of the script in qtools was used to convert the parameters to the Q format. Finally, those parameters were mixed with the classical parameters for serine by replacing the methyl at the end of an ethoxy substituent of VX with the backbone of serine. The final parameters are presented in **Appendix 5**.

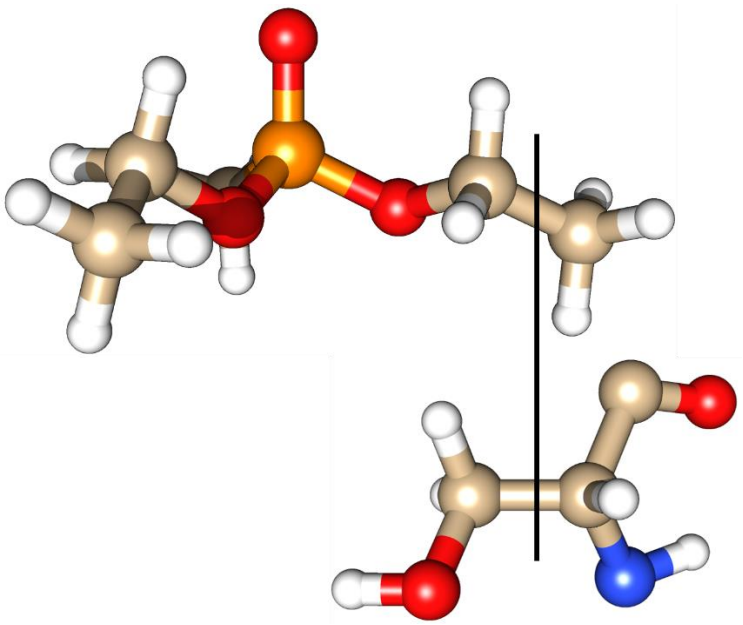


Figure VI-1. Parametrized truncated model of the VX adduct lined with the serine residue the adduct will be added to.

Now that the parameters are complete, the system was prepared for EVB simulations. The starting point was the setup described in section **II** of this chapter. This setup, fully solvated and protonated is completely stripped of non-crystalline water molecules and hydrogens. Q is used, following the topology dictated by the OPLSAA force field, to re-add hydrogens and to add a water bubble with a radius of 26 Å around the residues involved in the proton relay mechanism. The preparation of the setup is the step where variations in the environment of the reaction can be introduced. In the case at hands, the reaction was modelled in VX-inhibited His447H$^+$ AChE, in VX-inhibited His447^0 AChE, and in non-inhibited His447^0 AChE.

[32] a) W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236. b) G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, W. L. Jorgensen, *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

[33] Schrödinger, *MacroModel Release 2013-3*, Schrödinger LLC: New York, **2013**.

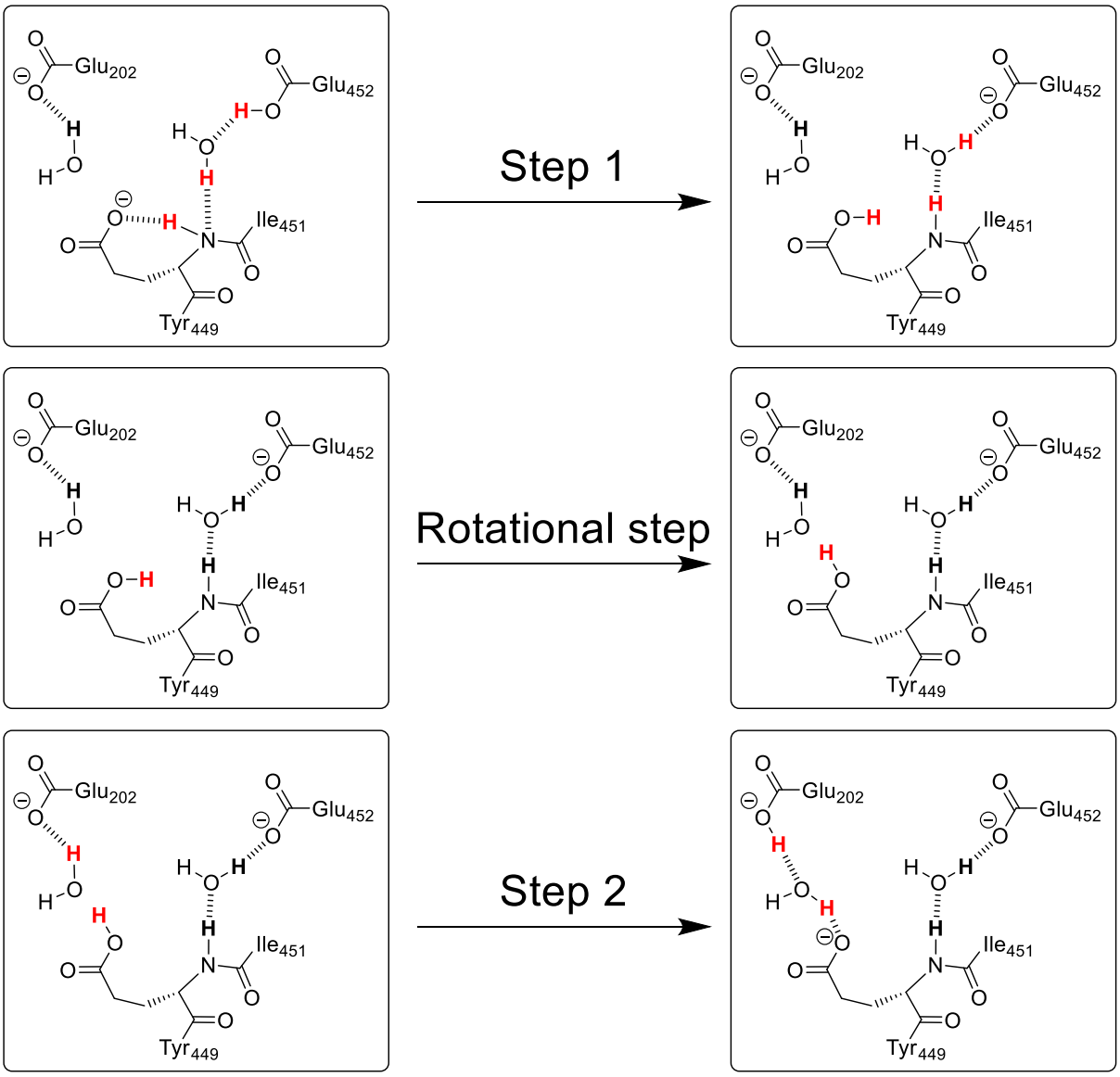


Figure VI-2. Steps of the transfer of a proton from Glu452 to Glu202.

To study the reaction with EVB it was broken down in three steps, the proton transfer from Glu452 to Glu450, the rotation of the proton around Glu450 and the proton transfer from Glu450 to Glu202 (see **Figure VI-2**). The first step, step 1, involves three concurrent proton transfers. The proton on Glu452 is transferred to a water molecule. A second proton on that same water molecule is transferred to the nitrogen of the amide between Tyr449 and Glu450. A third proton is transferred from the same amide to Glu450. The second step, step 2, involves the simultaneous proton transfer to a water molecule from Glu450 and from that same water molecule to Glu202. In-between these two step, there is an extra rotational step which involves the rotation of a proton from the *cis* position relative to the side chain of Glu450 to the *trans* position.

Those steps are parametrized into the EVB simulation by establishing a list of all changes in the topology between the reactant state and the product state. Even for simple steps many changes will have to be made in the topology because when a bond in broken, not only will this bond have to be removed from the topology, but also will all the angles, dihedral

angles, and impropers going through this bond. The same is true for the formation of a new bond.

Once all the setup has been prepared, the adduct has been parametrized, and the steps to be modelled have been defined, the EVB simulations can be performed. The system is minimized, slowly heated and then equilibrated for 10ns. Out of this equilibration dynamic, 3 snapshots were extracted at 8 ns, 9 ns and 10 ns. From each of those three snapshots five dynamics where started with different starting velocities for a total of 15 sets of EVB simulations. For every one of the 15 sets of EVB stimulations, the simulations where initiated with a 0.2 ns equilibration dynamic.

The core of the EVB method is the molecular dynamic simulations of the system at various states of parameter mixing between the parameters of the product and the parameters of the reactant according to a mapping parameter λ. The system was made to vary from the reactant to the product in 50 steps. At every λ state, the contribution of parameters based on the topology of the reactant was reduced by 2% and the contribution of parameters based on the topology of the product was increased by 2%. The λ state of the reactant is thus λ = 1.00, with 100% contribution of parameters based on the topology of the reactant an 0% contribution of parameters based on the topology of the product. The following λ state is λ = 0.98 with 98% contribution of parameters based on the topology of the reactant an 2% contribution of parameters based on the topology of the product. At every λ state a molecular dynamic trajectory was computed for 0.1 ns.

In addition to the simulations in enzyme, the reaction steps were modelled using EVB with truncated models in water bubbles. The truncated models used for every reaction are presented in **Figure VI-3**. For step 1, the truncated model includes butanoic acid, a water molecule, and an amino acid model with a full glutamate, with full peptide bonds at both ends connected to a methyl group. For step 2, the truncated model includes a butanoic acid, a butanoate, and a water molecule in-between. For the rotational step only a butanoic acid is included.

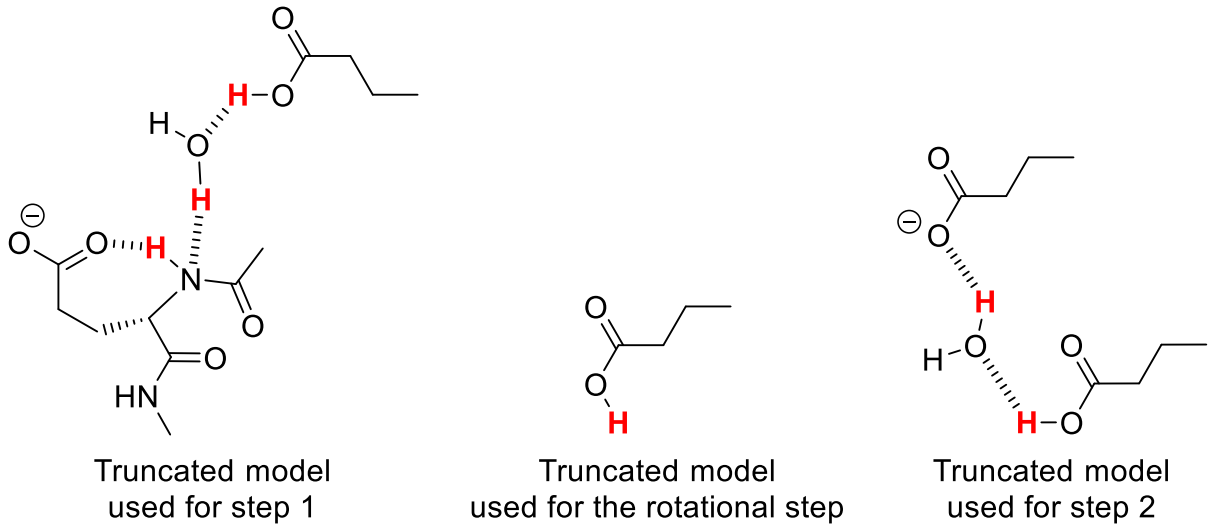


Figure VI-3. Truncated models used for all three steps.

The EVB simulation of those truncated models follow the same protocol as describe for the reaction steps in enzyme. The selected truncated models are stripped of hydrogens which are then re-added using Q. They are solvated in water bubble with a radius of 26 Å, minimized, heated, and equilibrated. Finally, molecular dynamic trajectories of 0.02 ns are performed for each λ state for 5 replicates coming a single snapshot for all three reaction steps. Like the in-enzyme simulations, the λ state is made to vary in 0.2 increments over 50 steps.

Q maps free energy profiles for a reaction step based on the energy write-outs from the trajectory as a function of the reaction coordinate $X = (\varepsilon_{product} - \varepsilon_{reactant})$. Two numerical parameters are necessary for this mapping, the H_{ij} coupling constant, and the gas phase shift α_{gas}^{i}. Those two numerical parameters are necessary to obtain the energetics of a reaction step but they can also be obtained by fitting the energy of a reaction step to a known reference energy. This is the purpose of the reference truncated models, with a reference energy for those models, they are used to obtain H_{ij} coupling constant, and the gas phase shift α_{gas}^{i} constant. Those are then transferable to the reaction in enzyme and can be used to calculate the energetics of the reaction steps in enzyme.

The choice of a proper reference is thus a key point of the EVB method and great care should be taken to ensure quality of the reference energies. A protocol to obtain QM/MM reference energies was designed. The aim was to extract snapshots from the equilibration dynamic of the model systems in a water bubble and to evaluate each reaction step in the truncated model in QM/MM. 25 snapshots were originally intended for all three reaction steps, step 1, step 2 and the rotational step. Step 2, the proton transfer from Glu450 to Glu202 through a water molecule was the first to be initiated. The QM/MM methodology was simple, the truncated model as well as the 10 nearest water molecules would compose the QM region. The truncated model plus the 125 nearest water molecules composed the optimized MM region. The first results highlighted the excessive rigidity of the water bubble around the system favouring the reactant by 2.18 kcal.mol^{-1} when the reactant and product are identical and thus the energy difference for the reaction was expected to be close to 0. An additional 25 snapshots were performed in reverse, going from the product to the reactant. On average those snapshots favoured the product by 4.34 kcal.mol^{-1}. It appears that the rigidity of the optimized water network around the system systematically favours the state initial state. For step 2, after 50 snapshots, 25 forward, 25 in reverse, the $\Delta G^{\ddagger}$ was 8.40 kcal.mol^{-1} and ΔG_0 was -1.29 kcal.mol^{-1}. Step 1 proved to be even more challenging because three proton transfers are involved in this step. An initial assessment over a single snapshot by scanning along three reaction coordinates, one for every proton transfer show that the proton transfer Glu452 to a water molecule and the transfer of a proton of this water molecule to the amide to be concurrent, followed by the spontaneous proton transfer from the amide to Glu450 (see **Figure VI-4**).

This initial assessment opened the possibility to search for the transition state by scanning along only two of the three reaction coordinates, dividing the computational cost at least 10-fold. When work was started on the snapshots however, the variations in the water

network around the system induced changes in the reaction. It appears now that scanning along all three reaction coordinates for every snapshot will most likely be mandatory to study the reaction. The human time to prepare and analyse those calculations, as well as the computational time required to run them compelled an alternate approach.

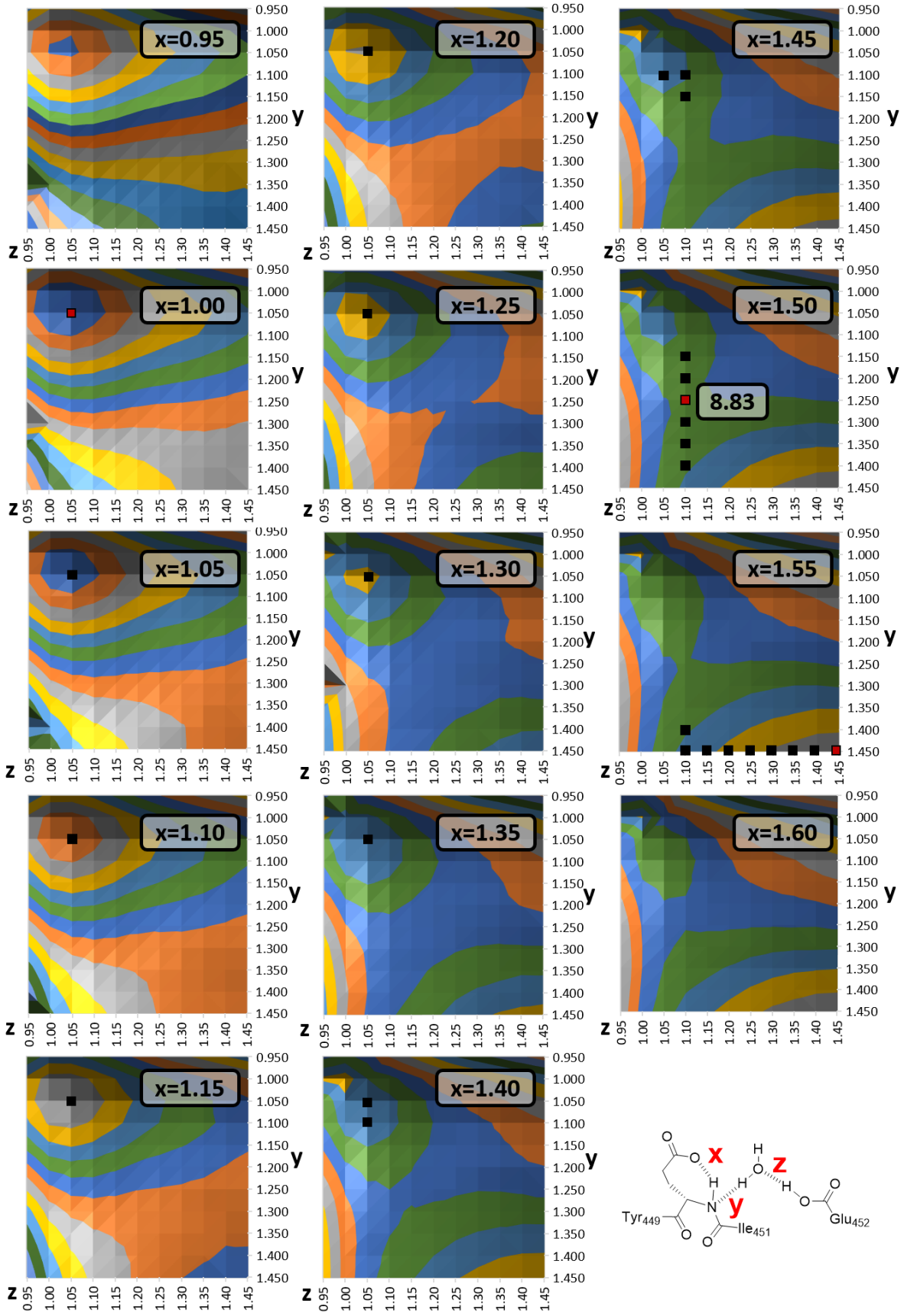


Figure VI-4. Results of a three-dimensional scans. Each surface corresponds to a set x value, the y coordinate is the ordinate axis, the z coordinate is the abscise axis. The black marks one the potential energy surfaces materialize the minimum energy path across all three coordinates. The red marks indicate the reactant, product and transition state position. Every colour band corresponds to a 2 kcal.mol^{-1} range.

Experimentally obtained reference energies were sought in the literature for steps 1 and 2. Step 2 is a proton transfer between a butanoic acid and a butanoate through a water molecule and thus is equivalent to the exchange of a proton between butanoic acid a water molecule from the solvent. In the literature this proton exchange is known to have a rate constant of 9.5 x 10^7 s^{-1}.[34] Using the Eyring equation (equations **VI-I** and **VI-II**), the free energy barrier $\Delta G^{\ddagger}$ for this process could be calculated and is 6.6 kcal.mol^{-1}. The ΔG_0 of the reaction is 0.0 kcal.mol^{-1} as the reactant and product are strictly equivalent.

$$k = \frac{k_b T}{h} e^{-\frac{\Delta G^{\ddagger}}{RT}} \qquad \textbf{VI-I}$$

$$\Delta G^{\ddagger} = ln\left(\frac{k_b T}{h}\right) RT - ln(k)\, RT \qquad \textbf{VI-II}$$

The reference free energy barrier for step 1 was assumed to be tied to the activation energy of proton exchange between an amide and water. A reference rate constant was found for this process.[35] In the article the rate constants for proton exchange on the amide for three systems equivalent to an amide bond between two amino acids is presented. The average of those three rate constants is 2.5 x 10^9 s^{-1} which corresponds to a free energy barrier $\Delta G^{\ddagger}$ of 4.6 kcal.mol^{-1}. As step 1 involves both the exchange of a proton on an amide and the exchange of a proton on a carboxylic acid, the highest energy barrier was used for the process and thus, the reference $\Delta G^{\ddagger}$ for step 1 was chosen to be 6.6 kcal.mol^{-1} with a ΔG_0 of 0.0 kcal.mol^{-1} as the reactant and product states are equivalent.

No reference energy was sought for the rotational step as this step was not well modelled in EVB. The parameter mixing approach of EVB prove ineffective to sample the rotation of the proton and only the reactant and product states of the rotation were sufficiently sampled with the λ states in-between severely lacking. Many attempts were made to achieve proper sampling but proved fruitless. As a result, it was decided to take another approach to the study of this step and due to a lack of time this part of the project was handed over to Paul Bauer at Kamerlin lab.

In the next sub-sections the results obtained for different setups of AChE are presented. The first two are VX-inhibited AChE with either His447H$^+$ or His447^0. The third and last one is non-inhibited AChE.

VI.1. VX-inhibited AChE with a bis-protonated His447

The results obtained from EVB simulations of the proton relay mechanism in VX-inhibited His447H$^+$ AChE are presented in **Figure VI-5**. Those results are average energies from the 15 replicates performed for steps 1 and 2.

[34] a) J. E. Crooks, *Comprehensive Chemical Kinetics* **1977**, *8*, 197–250. b) Z. Luz, S. Meiboom, *J. Am. Chem. Soc.* **1963**, *85*, 3923–3925.

[35] M. Sheinblatt, *Journal of the American Chemical Society* **1970**, *92*, 2505–2509.

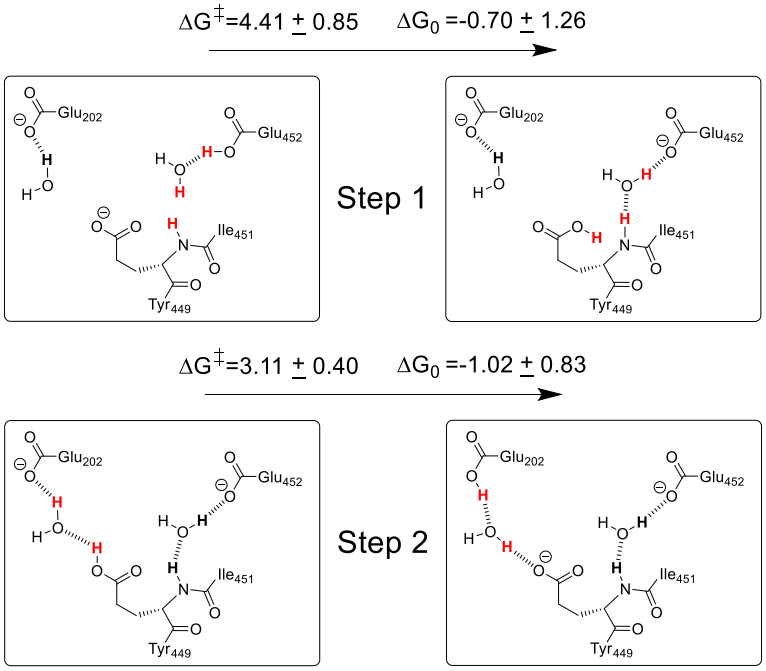


Figure VI-5. Energetic of the proton transfer steps modelled in EVB for VX-inhibited His447H^{+}.

The first step of the proton relay mechanism in this setup has a free energy barrier of 4.41 ± 0.85 kcal.mol^{-1}. This step is exergonic with a free energy difference of -0.70 ± 1.26 kcal.mol^{-1}. The second step has a free energy barrier of 3.11 ± 0.40 kcal.mol^{-1} and a free energy difference of -1.02 ± 0.83 kcal.mol^{-1}.

VI.2. VX-inhibited AChE with a mono-protonated His447

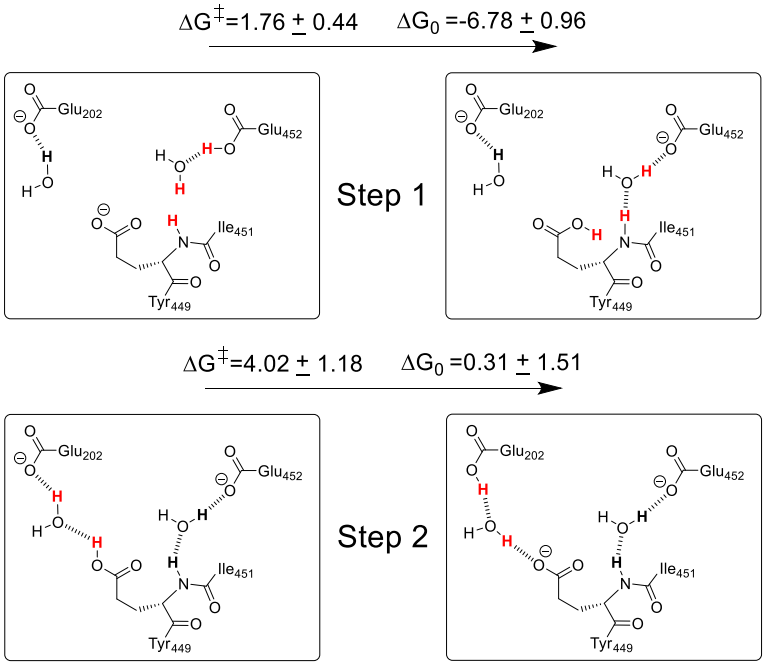


Figure VI-6. Energetic of the proton transfer steps modelled in EVB for VX-inhibited His447^{0} AChE.

Similar simulations were performed with a mono-protonated His447, with an imidazole side chain protonated a the δ site. The results for this setup are presented in **Figure**

VI-6. Those results are average energies from the 15 replicates performed for steps 1 and 2. The first step of the proton relay mechanism has a free energy barrier of 1.76 ± 0.44 kcal.mol^{-1}. This step is exergonic with a free energy difference of -6.78 ± 0.96 kcal.mol^{-1}. The second step has a free energy barrier of 4.02 ± 1.18 kcal.mol^{-1} and a free energy difference of 0.31 ± 1.51 kcal.mol^{-1} indicative of a slight endergonicity for this step.

VI.3. Non-inhibited AChE

The same reaction was modelled in non-inhibited AChE with two different setups, identical but differing by their starting point. Setup **A** has the same starting structure as all the other setups from which the Ser20-VX adduct is replaced by a regular serine. It originates from the setup presented in section **II** of this chapter. Setup **B** is generated using the protocol described at the beginning of this section with the setup for non-inhibited AChE presented in section **V** of this chapter.

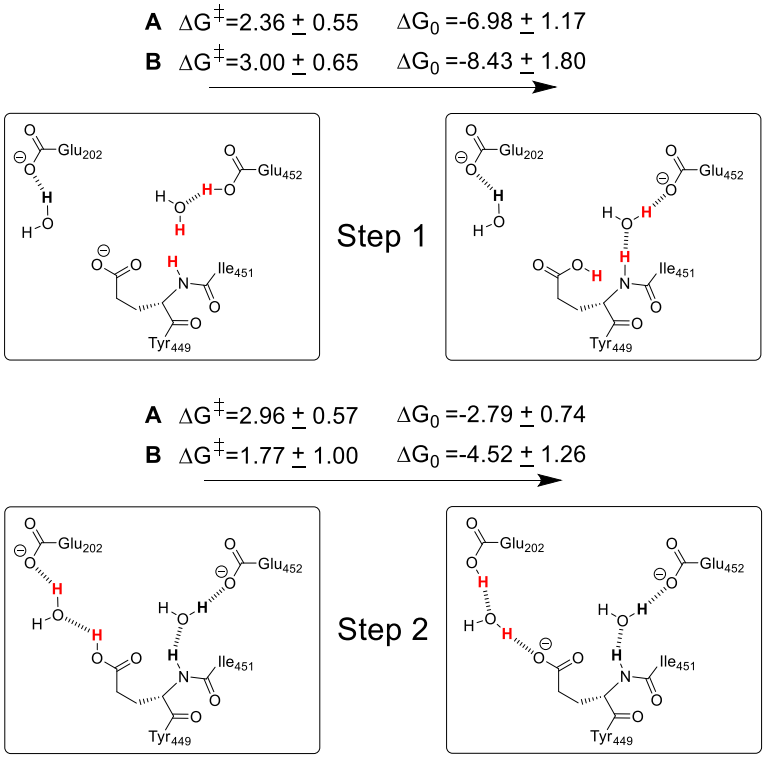


Figure VI-7. Energetic of the proton transfer steps modelled in EVB for non-inhibited AChE.

The first step of the proton relay has a free energy barrier of 2.36 ± 0.55 kcal.mol^{-1} for setup **A**, and a free energy barrier of 3.00 ± 0.65 kcal.mol^{-1} for setup **B**. The difference between the two energies is 0.74 kcal.mol^{-1}. The free energy difference for this step is -6.98 ± 1.17 kcal.mol^{-1} for setup **A** and -8.43 ± 1.80 kcal.mol^{-1} for setup **B**. The difference between the two setups is 1.45 kcal.mol^{-1}.

The second step of the proton relay has a free energy barrier of 2.96 ± 0.57 kcal.mol^{-1} for setup **A**, and a free energy barrier of 1.77 ± 1.00 kcal.mol^{-1} for setup **B**. The difference between the two energies is 1.29 kcal.mol^{-1}. The free energy difference for this step is -2.79 ±

0.74 kcal.mol^{-1} for setup **A** and -4.52 ± 1.26 kcal.mol^{-1} for setup **B**. The difference between the two setups is 1.73 kcal.mol^{-1}.

VII. Discussion

First and foremost, the energies discussed in this section for QM/MM simulations are B3LYP-D3/def2-TZVP single point energies. Those energies have systematically been confirmed by DLPNO-CCSD(T)/def2-TZVPP single points which give energies in line with the B3LYP-D3/def2-TZVP single points in all the cases presented in this manuscript.

The direct exchange of protons between the water solvent and residues of the active site His447 and Glu202 has been modelled and is summarized in **Figure VII-1**. From this figure, it clearly appears the proton transfer reaction between Glu202 of His447 in either state of protonation with water is very endothermic. The exchange of a proton, from or towards the solvent is only possible if a hydronium or a hydroxide ion is involved.

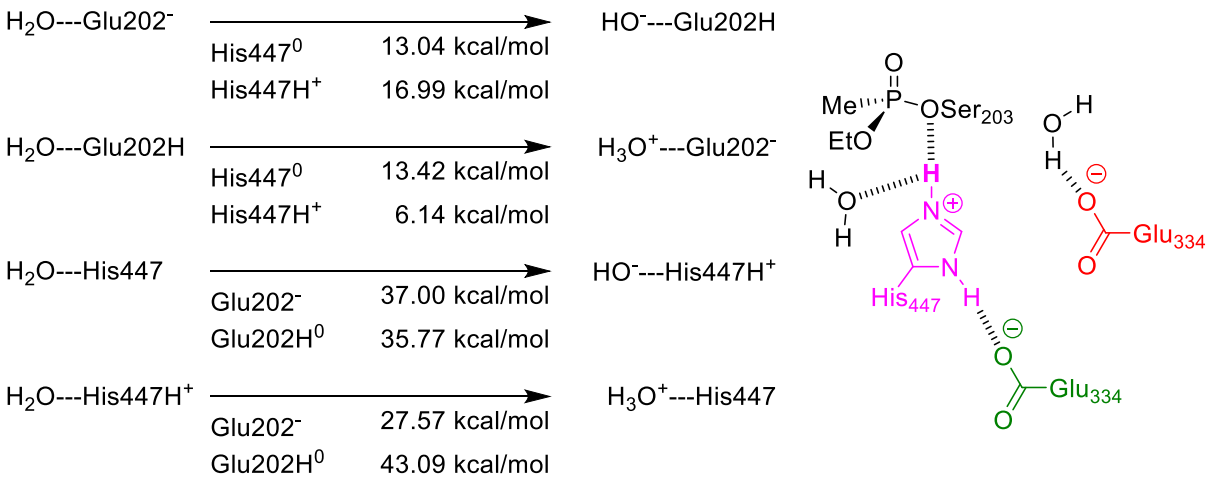


Figure VII-1. Energy difference of proton exchange reactions of His447 and Glu202 with water.

In an open active site, these residues would be in contact with bulk solvent and thus would be in the protonation state dictated by their pK_a. In AChE however, those residues are in a buried active site accessible to the solvent, but not necessarily accessible to a hydronium or a hydroxide. In those conditions, His447 and Glu202 are unlikely to be able to freely exchange protons with the solvent.

There is however, a way for those two residues of the active site to exchange protons with the solvent, through a proton relay mechanism uncovered by our QM/MM calculations. This proton relay mechanism involves two glutamates, Glu452 and Glu452. QM/MM showed that if Glu452 is protonated and Glu202 is not protonated, it is possible that the proton on Glu452 could be transferred to Glu202. This proton transfer involves the residue Glu450 as a relay and goes through an amide bond, requiring the exchange of a proton on the nitrogen moiety of the amide. The entrance of a proton in the active site through this proton relay mechanism in VX inhibited AChE has been shown to be exothermic with QM/MM and EVB simulations (see **Figure VII-2**). There however strong divergence between the QM/MM and EVB energetics for the reaction. For step 1, while both methods show that a proton transfer

through the amide is possible, in the same His447H$^+$ case as QM/MM was performed in, the free energy barrier is almost 10 kcal.mol^{-1} lower in EVB compared to the energy barrier from QM/MM. The free energy difference from EVB is also much lower than the energy difference obtained with QM/MM, reduced from -6.69 kcal.mol^{-1} to -0.70 kcal.mol^{-1} by almost 6 kcal.mol^{-1}. This divergence might be due to the lack of entropic contribution and the absence of dynamical effects from the static QM/MM simulations. EVB only accounting for electronic effects through its parametrization on an experimental value could also underestimate the stabilization of the product or the instability of the transition state.

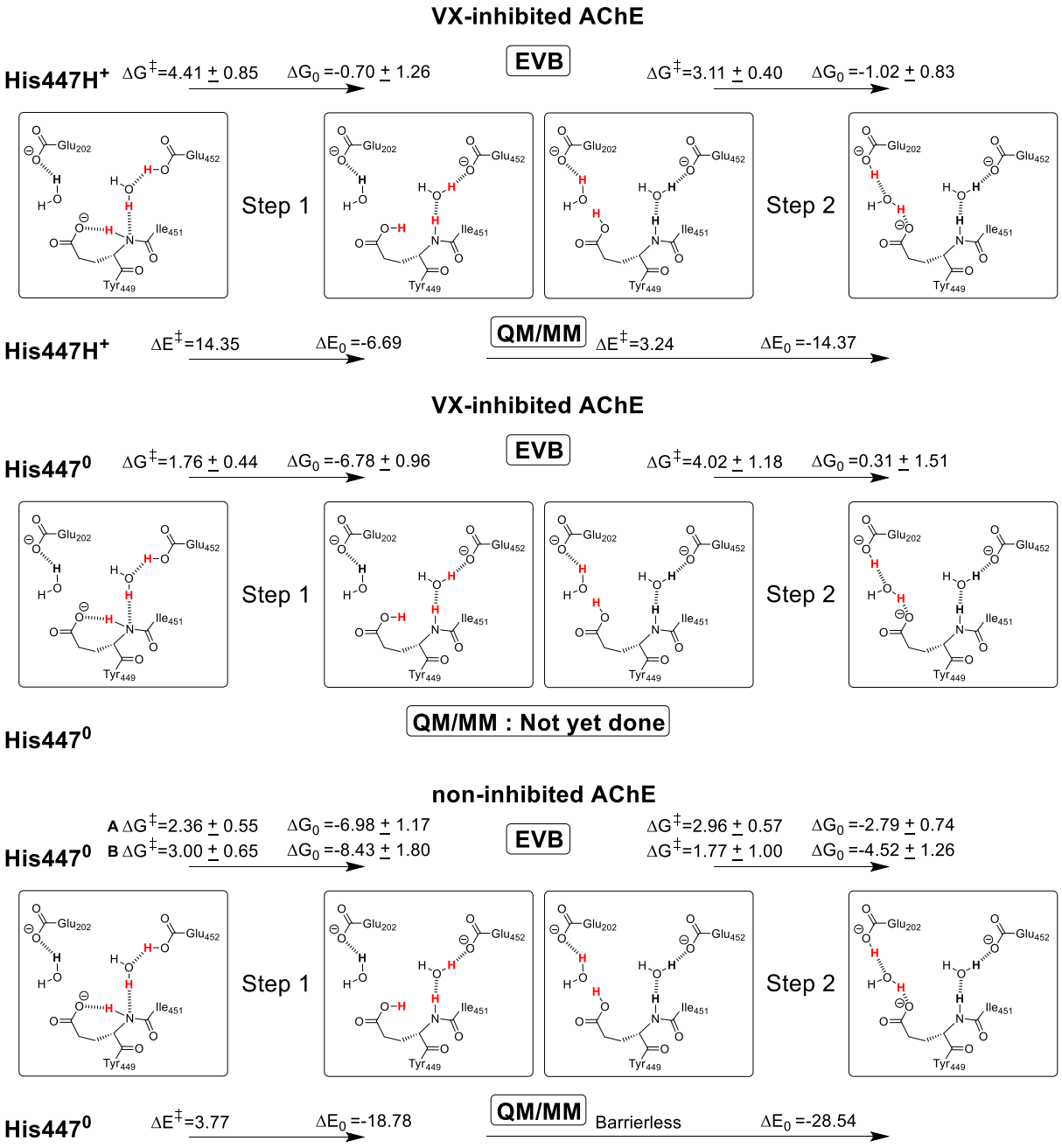


Figure VII-2. Summary of the energetics of the proton relay mechanism for the transfer of a proton from Glu452 to Glu202 obtained with EVB and QM/MM with VX-inhibited and non-inhibited AChE. Energies are in kcal.mol^{-1}.

For step 2 in VX-inhibited AChE, the QM/MM and EVB simulations indicate similar free energy, and energy barriers although in the QM/MM simulations, the rotational step is included in step 2 which is not the case with EVB. The exothermicity of this step however is 14 times greater in QM/MM compared to the EVB, although once again the exothermicity of the rotational step is included. This divergence might again be explained by the specificity of each method, the absence of dynamical effects and entropic contribution in QM/MM, and the classical nature of EVB in addition to the fact that both steps are not identical in both methods.

When the protonation state of the imidazole side chain of His447 is changed from an imidazolium to an imidazole protonated at the δ protonation site in VX-inhibited AChE, EVB simulations show a facilitated proton entrance through the proton relay mechanism (see **Figure VII-2**). The free energy barrier of the first step is reduced from 4.41 kcal.mol^{-1} to 1.76 kcal.mol^{-1}. The exergonicity of this step also increases from -0.70 to -6.78 kcal.mol^{-1}. This facilitated proton entrance might be due to the electrostatic attraction of the charge of Glu202, unshielded by the positive charge of His447 now absent. The second step shows a slightly different trend. When His447 is mono-protonated the free energy barrier slightly increases to 4.02 kcal.mol^{-1} and the step is no longer exergonic but endergonic, by 0.31 kcal.mol^{-1}, although in this step the standard error is much greater than for His447H^{+}. The standard error goes from ± 0.40 for the free energy barrier for His447^{0} to ± 1.18 for His447H^{+}. Similarly, the standard error goes from ± 0.83 for the free energy difference for His447^{0} to ± 1.51 for His447H^{+}. A greater conformational flexibility induced by this protonation state might explain both the increased standard deviation and the observed trend.

When AChE is not inhibited by VX, in the standard mono-protonated at the δ site protonation state for His447, both the EVB and the QM/MM simulations indicate and increased exergonicity and exothermicity respectively for the entrance of a proton through the proton relay mechanism (see **Figure VII-2**). In this case, contrary to the case of VX-inhibited His447^{0} AChE, the energetics of both steps 1 and 2 show a more favoured proton entrance. The divergence between QM/MM and EVB remains however present even though their evolution compared to VX-inhibited His447H^{+} AChE QM/MM and EVB values follow a similar trend. In the discussion of this step, no difference is made between the two setups as they produce the same trends when compared to the EVB simulations of VX-inhibited AChE and besides different starting points there is no clear point of divergence between the two.

For step 1, the exothermicity in QM/MM is much greater that the free energy difference obtained in EVB while the energy barrier obtained through QM/MM is quite close to the free energy barrier obtained from EVB. In step 2, the QM/MM indicates a barrierless step while the EVB free energy barriers are quite small but present. The energy difference however in massively different between the QM/MM exothermicity of -28.54 kcal.mol^{-1} and the free energy differences obtained in EVB, closer to 0 by more than 24 kcal.mol^{-1}.

The comparison of both QM/MM simulations in VX-inhibited and non-inhibited AChE show that the proton entrance is favoured in non-inhibited AChE. The energy, or free energy difference of both step is increased by more than 12 kcal.mol^{-1}. The energy barriers are also reduced. For step 1 it is reduced by 10.58 kcal.mol^{-1} while the energy barrier of step 2

disappears entirely in non-inhibited AChE compared to VX-inhibited AChE. This effect might be attributed to the absence of the covalent inhibitor on Ser203, but also to the protonation state of His447.

The comparison of EVB results can be used to address this question. For step 1, the free energy difference of non-inhibited His447^0 AChE is very close to the free energy difference obtained with VX-inhibited His447^0 AChE and much more exergonic that what was obtained with VX-inhibited His447H$^+$ AChE. The free energy barriers, although higher than for VX-inhibited His447^0 AChE by 0.92 kcal.mol^{-1} on average, are lower than for VX-inhibited His447H$^+$ AChE by 1.73 kcal.mol^{-1} on average. For step 2, non-inhibited His447^0 AChE shows significative divergence from both VX-inhibited His447^0 and His447H$^+$ AChE. The exergonicity is significantly increased and the energy barrier are lower than either VX-inhibited His447^0 AChE and VX-inhibited His447H$^+$ AChE. There are several possible explanations for this behaviour. The similarities evidenced for step 1 between VX-inhibited His447^0 AChE and non-inhibited His447^0 AChE could mean that the protonation state of His447 has the main influence on the favourability of proton entrance in both cases. Step 2 for non-inhibited His447^0 AChE is however significantly different than for VX-inhibited His447^0 AChE. His447 could still explain these differences as in non-inhibited His447^0 the hydrogen bond between His447 and Ser203 maintains the position of His447 which can then assist Glu202 in attracting a proton to the active site. In VX-inhibited His447^0 AChE however, this hydrogen bond is absent and the side chain of His447 is likely to have greater mobility as evidenced by X-ray structural analysis of inhibited AChE.[36] This mobility could limit the role of His447 in assisting Glu202 to attract a proton to the active site.

The question of proton diffusion in and out of the active site does not have to be considered from the perspective of a proton entering the active site. Thus, the results of the same simulations can be rearranged, as per **Figure VII-3** to show the energetics of a proton leaving the active site of AChE. What this perspective shows is that according to QM/MM, it is impossible for a proton in the active site of either VX-inhibited His447H$^+$ AChE or non-inhibited His447^0 AChE to leave the active site of AChE through the proton relay mechanism. The EVB simulations however, paint a different picture. For VX-inhibited or non-inhibited His447^0 AChE the proton exchange between Glu450 and Glu452 has both a high energy barrier and an endergonicity around 7 kcal.mol^{-1}. For VX-inhibited His447H$^+$ AChE however, the energy difference of both step 1 and 2 remain close to 0 and the energy barriers are both low at 4.13 kcal.mol^{-1} for step 2 and 5.11 kcal.mol^{-1} for step 1. EVB indicates that the exit of a proton through the proton relay mechanism is possible in a specific protonation state for the active site.

[36] A. Hörnberg, E. Artursson, R. Wärme, Y.-P. Pang, F. Ekström, *Biochem. Pharmacol.* **2010**, *79*, 507–515.

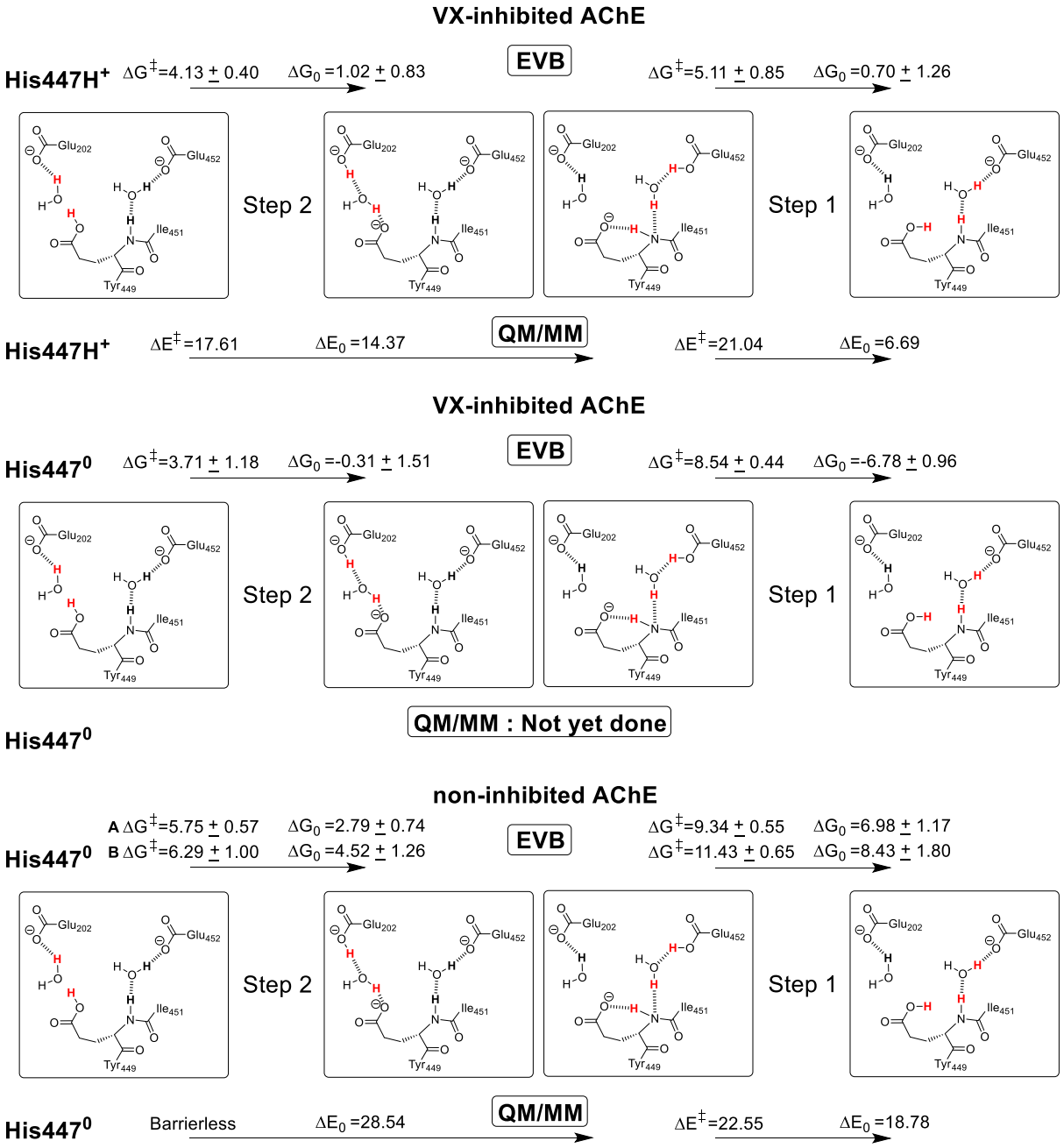


Figure VII-3. Summary of the energetics of the proton relay mechanism for the transfer of a proton from Glu202 to Glu452 obtained with EVB and QM/MM with VX-inhibited and non-inhibited AChE. Energies are in kcal.mol^{-1}

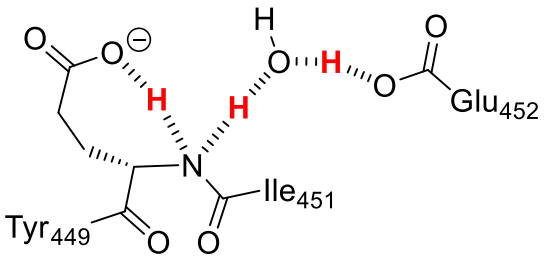
This proton relay mechanism is part of a broader picture here the entering proton, or a proton already in the active site could be transferred from Glu202 to His447, although not favourably as QM/MM simulations have shown. The transfer of a proton from Glu202 to His447 has an energy barrier of 6.43 kcal.mol^{-1} and is slightly endothermic by 3.17 kcal.mol^{-1}. This result would tend to indicate that if a single proton is in the active site it would most likely tend to be transferred to and remain on Glu202. It will be a crucial point to address this point by EVB calculations as it is the last step of the proton relay mechanism. In this context, the question of the entrance of a proton through the proton relay mechanism when Glu202 is protonated and His447 is not remains. Would an extra proton be pumped in through the proton relay while the proton on Glu202 would be transferred to His447, would the entering

proton remain on Glu450 ? The matter of the contribution of this proton relay mechanism to other processes of AChE is also left open. There is however some experimental indication that it may be the case as it was shown that mutations of Glu202 or Glu450 decrease significantly the rate of O-dealkylation.[37]

Constant pH molecular dynamics cannot answer this question but they did confirm the possibility for Glu452 and Glu450 to be transiently protonated. The reliability of CpHMD, despite the use of a classical non-polarizable force field, is confirmed by the other computational results described in this thesis. The CpHMD simulations are also confirmed by experimental data. In non-inhibited AChE, CpHMD indicates that at physiological pH His447 is in its mono-protonated state and Glu334 is unprotonated.[27],[28] CpHMD also gives a similar trend as experimental data for the protonation state of Glu202.[29] Thus, the possibility of the proton relay mechanism, is confirmed by three different simulation methods.

VIII. Conclusion

There is a mechanism that allows AChE to have regulated protonation states for active site residues, despite having a buried active site. Through a proton relay mechanism, protons can enter and maybe even leave the active site, circumventing the main channel of AChE. This mechanism involves two glutamates which have been demonstrated as able to transiently receive a proton and end on Glu202, which is favourably protonated at physiological pH. This mechanism can have a role in the regulation of the protonation state of active site residues, to maintain them in protonation states that allow for the enzymatic catalysis to take place. It could also be involved in dynamical protonation state changes allowing to transiently protonate or deprotonate Glu202 and His447 to allow them, for instance, to deprotonate a reactivator entering the active site in oxime form, or to deprotonate a water molecule during the O-dealkylation of AChE-nerve agent adducts (the aging process). Finally, this research shows the transfer of a proton between two residues through N-protonation of an amide bond (see **Scheme VIII-1**).



Scheme VIII-1. Transition state of the proton transfer step through N protonation of an amide bond.

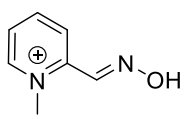
[37] A. Ordentlich, C. Kronman, D. Barak, D. Stein, N. Ariel, D. Marcus, B. Velan, A. Shafferman, *FEBS Lett.* **1993**, *334*, 215–220.

General Conclusion and Perspectives

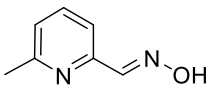
General Conclusion and Perspectives

The aim of this PhD research project was to investigate the reactivation process and the active site of nerve agent inhibited AChE by computational methodologies to gain insight towards the rational design of new reactivators.

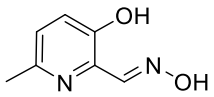
An initial study using a truncated QM model showed that Glu334 could block the reactivation if not properly compensated by surrounding residues. It also confirmed the role of the oxyanionic hole in the stabilization of the transition state of the reactivation. With this finding in mind, QM/MM simulations were prepared. First the simulations focused on the interactions of classical reactivator **2-PAM** with the active site of AChE. It was shown that Glu202, a residue near the catalytic triad, needs to be protonated for the reactivation to occur. Those simulations also showed that the reactivator can be deprotonated in the active site of AChE by His447. More generally they served to highlight the role of protonation states and proton transfers in the reactivation of AChE.



2-PAM
2-pyridiniumaldoxime methyl iodide



6M-2PA
6-methyl-2-pyridinealdoxime



6M-3H-2PA
6-methyl-3-hydroxy-2-pyridinealdoxime

Scheme 1. Classical reactivator and non-pyridinium reactivator prototypes.

After classical reactivator **2-PAM** was studied, new reactivators, characterized by a non-pyridinium core were studied under the same lens. It was showed that non-pyridinium reactivator **6M-2PA** is an excellent reactivator but its deprotonation in the active site of AChE has an extremely high energetic cost. Non-pyridinium **6M-3H-2PA** is a more potent reactivator than **2-PAM** albeit not as potent as **6M-2PA.** It is, however, of the three reactivators the more easily deprotonated in the active site of AChE. These simulations rationalized the potential of 3-hydroxy-2-pyridinealdoxime as a reactivator prototype. Since these studies demonstrate that the deprotonation of the reactivator is more important than the nucleophilicity, the last part of my PhD project was dedicated to understanding proton flow in and out of the active site of AChE. The results show that for active site residues to directly exchange protons with the solvent, hydronium or hydroxy ions need to access the active site through the main channel of AChE. More interestingly, a proton relay mechanism was identified through QM/MM simulations, involving two glutamate residues, Glu450 and Glu452 positioned behind the active site. The potential for these two residues to be transiently protonated and thus involved in a proton relay was confirmed by CpHMD simulations. The proton relay mechanism allows protons to enter and maybe exit the active site of AChE. It involves the transfer of a proton between Glu450, in contact with the solvent, and Glu202, through multiple proton transfers, including the N-protonation of an amide which is a novel mechanism. To ascertain this proton transfer, free energy perturbation simulations based on the EVB approach were performed. They confirm the potentiality of this proton relay mechanism

This work participated in expanding the frame of key AChE residues and structures. The role of Glu202 in the reactivation is now clearer and two previously unstudied residues have been identified as having a role in maintaining the protonation state of active site residues (see **Figure 1**). This mechanism might not be specific to AChE. It may be a common way for other enzymes, especially other serine proteases, to maintain their active site protonation state, especially when buried and not easily accessible to solvent. It could be interesting to dedicate some effort into the search for similar mechanisms or residue arrangements in other enzymes.

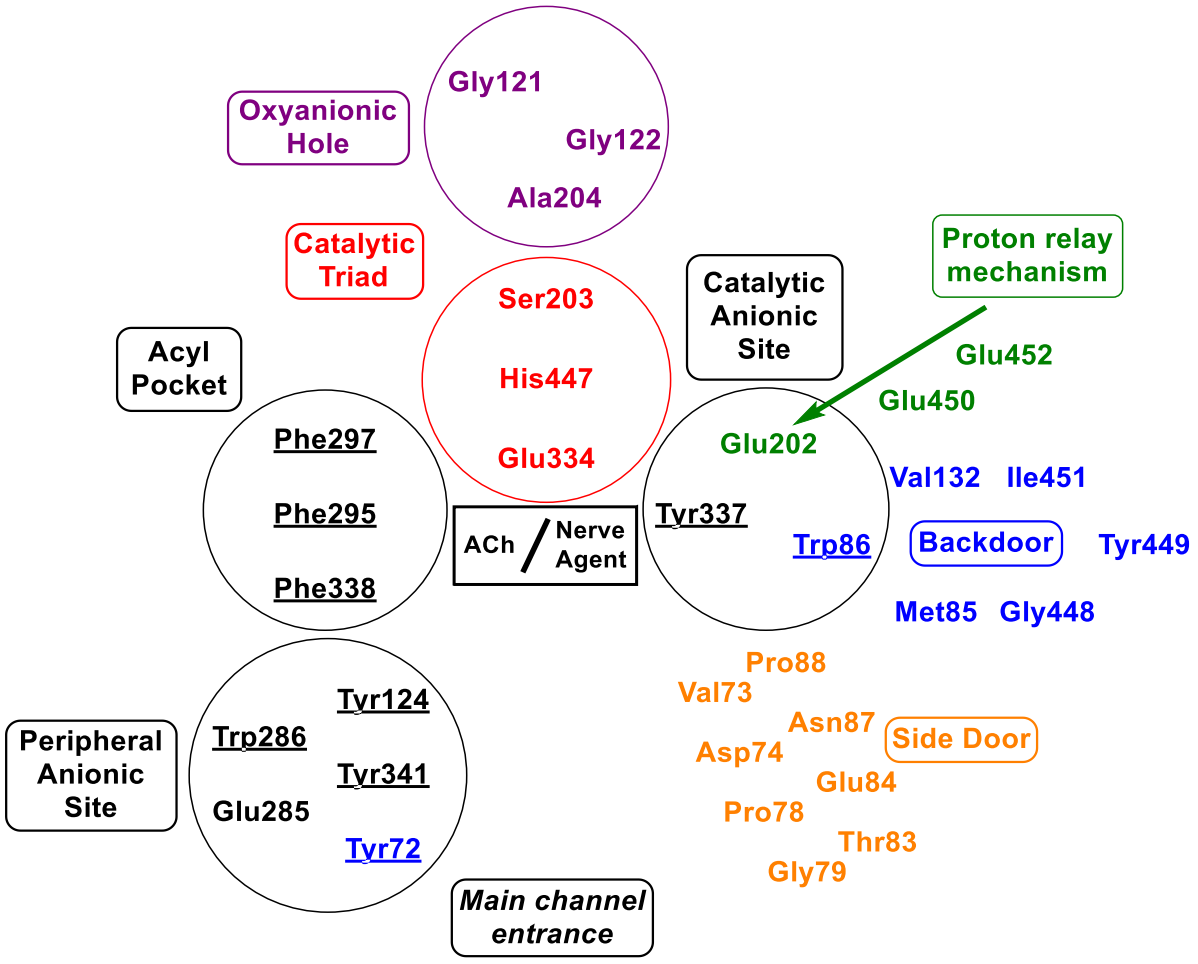


Figure 1. AChE residues grouped in their respecting subunits in a similar presentation as in the end of section **I** of **Chapter 1**, on page **37**. Underlined residues are aromatic residues of the channel. Residues in blue are involved in the backdoor, residues in orange are involved in the side door. In red and purple are the residues of the catalytic triad and the oxyanionic hole respectively. Residues surrounded by a circle belong to a stable and recognised substructure. The proton relay mechanism and the residues involved in it are added to the original scheme in green.

This work is a progress towards the investigation of AChE reactivation to guide reactivator design. It has been demonstrated that the nucleophilic power of the reactivator and its affinity for AChE are not the only factors in its ability to restore AChE functions. The ability for the reactivator to be deprotonated inside the active site of the enzyme is a key factor in its efficiency. The importance of the deprotonability of the reactivator in the active

site of the enzyme is further confirmed by the low barrier to deprotonate **6M-3H-2PA** which is known to be a potent reactivator.

Both the role of protonation states in the reactivation process and the newly discovered proton relay mechanism suggest that dynamical proton transfers might play a role in the reactivation of the enzyme:

- before, during, and after the entrance of an oxime reactivator in the active site of AChE;
- before, during, and after the deprotonation of the oxime to an oximate;
- before, during, and after the reactivation;

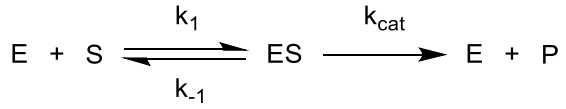
All these possibilities offered by such dynamical proton transfers between active site residues open new pathways for future studies, and thus better understanding of the reactivation process. Other AChE processes, such as inhibition, aging, or even the standard catalytic activity, could be studied from this point of view.

Methodological challenges are raised by this concomitance of multiple proton transfers shaping the active site and the reactivity of an enzyme, transiently generating the conditions for enzymatic reactivity. The theoretical methods currently used to study enzymes and enzymatic reactivity are either temporally static or chemically static. Even dynamical reactive methods, like QM/MM molecular dynamics, EVB, or free energy perturbation QM/MM are restrained, either by the computational cost or by the limitations of the methodologies. They cannot simulate dynamic proton transfers in the whole enzyme, and on the time scale of enzymatic reactivity. Hopefully, my PhD. work will highlight the necessity to accept the dynamical proton exchanges as a non-negligible factor in future simulations of AChE and other enzymes.

Appendices

Appendix 1: Enzyme Kinetics

Enzymatic catalysis can be represented by **Scheme 1**. It describes an enzyme E, binding to a substrate S to form the enzyme substrate complex ES. The reaction catalysed by enzyme E then takes place concluding with the release of product P.



Scheme 1. Equation for the enzymatic catalysis of the reaction turning substrate S into product P

The rates constants for the binding process are k_1 for the forward rate and k_{-1} for the reverse rate, the binding being a reversible process. Forward, reverse and catalytic constants are defined by the reaction rate equations for the formation (r_f) of the enzyme substrate complex, its dissociation (r_d) and product formation (r_r).

$$r_f = k_1[E][S] \quad \textbf{I}$$

$$r_d = k_{-1}[ES] \quad \textbf{II}$$

$$r_r = k_{cat}[ES] \quad \textbf{III}$$

As such k_1 is expressed in $M^{-1}.s^{-1}$ and both k_{-1} and k_{cat} in s^{-1}. This allows the calculation of the change in concentration of all chemical species of the system over time.

$$\frac{d[E]}{dt} = -k_1[E][S] + k_{-1}[ES] + k_{cat}[ES] \quad \textbf{IV}$$

$$\frac{d[S]}{dt} = -k_1[E][S] + k_{-1}[ES] \quad \textbf{V}$$

$$v = \frac{d[P]}{dt} = k_{cat}[ES] \quad \textbf{VI}$$

$$\frac{d[ES]}{dt} = k_1[E][S] - k_{-1}[ES] - k_{cat}[ES] \quad \textbf{VII}$$

Out of those four equations two have a particular meaning. Equation **VI**, the rate of product formation, can be thought of as the reaction velocity. The assumption can also be made that after an initial increase the concentration of the enzyme substrate complex will have no significant change and that equation **VII** equals zero. This is steady-state

approximation, brought forth by Briggs and Haldane in a 1925 article.[1] This approximation is only valid if the substrate concentration is large over the enzyme concentration pushing the equilibrium of the complexation state towards the products. Knowing that the total amount of enzyme is the sum of free enzyme E and complexed enzyme ES allows us to obtain equation **XII** from equation **VII**.

$$[E]_0 = [ES] + [E] \quad \textbf{VIII}$$

$$k_1[E][S] = (k_{-1} + k_{cat})[ES] \quad \textbf{IX}$$

$$(k_{-1} + k_{cat})[ES] + k_1[ES][S] = k_1[E]_0[S] \quad \textbf{X}$$

$$[ES] = \frac{k_1[E]_0[S]}{(k_{-1} + k_{cat}) + k_1[S]} = \frac{[E]_0[S]}{\left(\frac{k_{-1} + k_{cat}}{k_1}\right) + [S]} \quad \textbf{XI}$$

$$v = \frac{d[P]}{dt} = k_{cat}[ES] = \frac{k_{cat}[E]_0[S]}{\left(\frac{k_{-1} + k_{cat}}{k_1}\right) + [S]} \quad \textbf{XII}$$

This formulation allows to define V_{max} and K_M in the original Michaelis-Menten equation for **Scheme 1**, equation **XIII**, as presented in their seminal 1913 article.[2]

$$v = \frac{d[P]}{dt} = k_{cat}[ES] = \frac{V_{max}[S]}{K_M + [S]} \quad \textbf{XIII}$$

$$V_{max} = k_{cat}[E] \quad \textbf{XIV}$$

$$K_M = \frac{k_{-1} + k_{cat}}{k_1} \quad \textbf{XV}$$

Thus, V_{max} is the maximum velocity, the velocity achieved if the totality of the enzyme is involved in an enzyme substrate complex, completely saturated in substrate. The Michaelis constant K_M is the substrate concentration needed to reach half velocity. As such it has the same dimensions as [S], M. It is also an inverse measure of the affinity of the substrate for the enzyme. The lower it is, the lower the concentration in substrates needs to be for the catalysis to reach maximum velocity (see **Figure 1**).

[1] G. E. Briggs, J. B. S. Haldane, *Biochem J* **1925**, *19*, 338–339.

[2] a) L. Michaelis, M. L. Menten, *Biochem. Z.* **1913**, *49*, 333–369.b) K. A. Johnson, R. S. Goody, *Biochemistry* **2011**, *50*, 8264–8269.

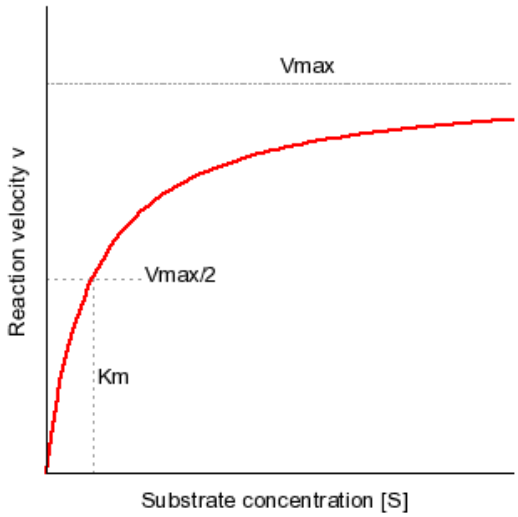


Figure 1. Reaction velocity as a function of substrate concentration

A different approximation can be made than the steady state approximation if the complexation states reaches an equilibrium much faster than the rate of product formation, formally if both k_{cat} over k_1 and k_{cat} over k_{-1} are very inferior to one. This is usually associated to a low substrate concentration relative to the enzyme concentration thus breaking the assumption made for the steady state approximation. The constant of this chemical equilibrium is called the dissociation constant and defined by equation **XVI**.

$$K_d = \frac{[E][S]}{[ES]} \quad \textbf{XVI}$$

$$[E]_0 = \frac{K_d[ES]}{[S]} + [E] \quad \textbf{XVII}$$

$$[ES] = \frac{[E]_0[S]}{K_d + [S]} \quad \textbf{XVIII}$$

$$v = \frac{d[P]}{dt} = k_{cat}[ES] = \frac{k_{cat}[E]_0[S]}{K_d + [S]} = \frac{V_{max}[S]}{K_d + [S]} \quad \textbf{XIX}$$

$$K_D = \frac{k_{-1} + k_{cat}}{k_1} \approx \frac{k_{-1}}{k_1} \quad \textbf{XX}$$

As both approximations rely on different assumptions both could be used. In practice however, the K_M is most commonly used in the case of enzymatic catalysis when K_D is used in the case of non-reactive or slow-reactive ligands and inhibitors. An enzymatic reaction is commonly characterized in the literature by its affinity (the K_M) and the catalytic rate constant,

(the k_{cat}). Often, a catalytic efficiency constant, also called specificity constant is defined by k_{cat} over K_M (or K_D), expressed in $M^{-1}s^{-1}$ (equation **XXI**), usually called k_s or k_e.

$$k_e = \frac{k_{cat}}{K_M} \qquad \textbf{XXI}$$

Appendix 2: CHARMM Parameters for Ser203-VX adduct and 2-PAM

CHARMM topology for Ser203-VX adduct

```
RESI SRT          0.00
GROUP
ATOM N    NH1    -0.47  !     |
ATOM HN   H       0.31  !  HN-N
ATOM CA   CT1     0.07  !     |   HB1
ATOM HA   HB      0.09  !     |   |
GROUP                   !  HA-CA--CB--OG
ATOM CB   CT2    -0.22  !     |   |     \
ATOM HB1  HA      0.19  !     |   HB2    PL
ATOM HB2  HA      0.19  !   O=C
ATOM OG   OH1    -0.57  !     |
ATOM P    PL      1.17 !
ATOM C4   CT3    -0.68
ATOM H41  HA      0.19
ATOM H42  HA      0.19
ATOM H43  HA      0.19
ATOM O4   O2L    -0.56 !
ATOM O1   OSL    -0.54 !
ATOM C3   CTL2   -0.04 !
ATOM H3A  HAL2    0.19 !
ATOM H3B  HAL2    0.19 !
ATOM C1   CTL3   -0.46 !
ATOM H11  HAL3    0.19 !
ATOM H12  HAL3    0.19 !
ATOM H13  HAL3    0.19 !
GROUP
ATOM C    C       0.51
ATOM O    O      -0.51
BOND CB CA   OG CB  N HN  N  CA
BOND C  CA  C +N  CA HA  CB HB1
BOND CB HB2  OG P
BOND P    O1   P C4    P    O4    O1 C3 C3 C1
BOND C1   H11    C1   H12   C1   H13   C3 H3A C3 H3B
BOND C4 H41 C4 H42 C4 H43
DOUBLE   O  C
IMPR N -C CA HN  C CA +N O
DONOR HN N
ACCEPTOR O C
IC -C   CA   *N   HN    1.3474 124.3700  180.0000 114.1800  0.9999
IC -C   N    CA   C     1.3474 124.3700  180.0000 105.8100  1.5166
IC N    CA   C    +N    1.4579 105.8100  180.0000 117.7200  1.3448
IC +N   CA   *C   O     1.3448 117.7200  180.0000 120.2500  1.2290
IC CA   C    +N   +CA   1.5166 117.7200  180.0000 124.6300  1.4529
IC N    C    *CA  CB    1.4579 105.8100  124.7500 111.4000  1.5585
IC N    C    *CA  HA    1.4579 105.8100 -115.5600 107.3000  1.0821
IC N    CA   CB   OG    1.4579 114.2800  180.0000 112.4500  1.4341
IC OG   CA   *CB  HB1   1.4341 112.4500  119.3200 108.1000  1.1140
IC OG   CA   *CB  HB2   1.4341 112.4500 -123.8600 110.3800  1.1136
```

CHARMM topology for 2-PAM

```
RESI PAM 0.00
GROUP
ATOM C1 CA -0.20191
ATOM H1 HP 0.21170
ATOM C5 CA -0.09002
ATOM H5 HP 0.15523
ATOM C4 CA -0.17256
ATOM H4 HP 0.15125
ATOM C3 CA 0.05738
```

```
ATOM H3 HP 0.16786
ATOM N1 NPH -0.42134
ATOM C7 CT3 -0.33720
ATOM H7 HA 0.17312
ATOM H2 HA 0.19598
ATOM H8 HA 0.19587
ATOM C2 CA 0.36790
ATOM C6 CA -0.15712
ATOM H6 HP 0.11930
ATOM N2 NR3 -0.01537
ATOM O1 OC -0.40006
BOND C1 C5 C4 C3 N1 C2 N1 C3 N1 C7 C2 C6
BOND C1 H1 C5 H5 C4 H4 C3 H3 C6 H6 C7 H7 C7 H2 C7 H8
BOND N2 O1 C1 C2 C4 C5 C6 N2
ACCEPTOR O1 N2
PATCHING FIRS NONE LAST NONE
```

Appendix 3: AMOEBA Parameters for Ser203-VX adduct

```
      ##############################
      ##                          ##
      ##  Atom Type Definitions   ##
      ##                          ##
      ##############################

atom        414   85    P     "VX_2"                       15    30.974    4
atom        415   86    O     "VX_2"                        8    15.999    1
atom        417   88    O     "VX_2"                        8    15.999    2
atom        418   89    C     "VX_2"                        6    12.011    4
atom        419   90    H     "VX_2"                        1     1.008    1
atom        420   91    H     "VX_2"                        1     1.008    1
atom        421   92    H     "VX_2"                        1     1.008    1
atom        422    8    C     "VX_2"                        6    12.011    4
atom        423    9    H     "VX_2"                        1     1.008    1
atom        424    9    H     "VX_2"                        1     1.008    1
atom        425   96    C     "VX_2"                        6    12.011    4
atom        426   97    H     "VX_2"                        1     1.008    1
atom        427   98    H     "VX_2"                        1     1.008    1
atom        428   99    C     "VX_2"                        6    12.011    4
atom        429  100    H     "VX_2"                        1     1.008    1
atom        430  101    H     "VX_2"                        1     1.008    1
atom        431  102    H     "VX_2"                        1     1.008    1
atom        432  103    C     "VX_2"                        6    12.011    4
atom        433  104    H     "VX_2"                        1     1.008    1
atom        434  105    H     "VX_2"                        1     1.008    1
atom        435  106    H     "VX_2"                        1     1.008    1
atom         33    7    CA    "PhosphrylatedSerine CA"      6    12.000    4
atom         34    8    C     "PhosphrylatedSerine CB"      6    12.000    4
atom         35    9    H     "PhosphrylatedSerine HB"      1     1.008    1
atom        416   87    O     "PhosphrylatedSerine OG"      8    15.999    2

      ################################
      ##                            ##
      ##  Van der Waals Parameters  ##
      ##                            ##
      ################################

vdw          85              4.450      0.3900
vdw          86              3.360      0.1120
vdw          87              3.405      0.1120
vdw          88              3.405      0.1120
vdw          89              3.820      0.1010
vdw          90              2.780      0.0260     0.91
vdw          91              2.780      0.0260     0.91
vdw          92              2.780      0.0260     0.91
vdw          93              3.820      0.1010
vdw          94              2.780      0.0260     0.91
vdw          95              2.780      0.0260     0.91
vdw          96              3.820      0.1010
vdw          97              2.780      0.0260     0.91
vdw          98              2.780      0.0260     0.91
vdw          99              3.820      0.1010
vdw         100              2.780      0.0260     0.91
vdw         101              2.780      0.0260     0.91
vdw         102              2.780      0.0260     0.91
vdw         103              3.820      0.1010
vdw         104              2.780      0.0260     0.91
vdw         105              2.780      0.0260     0.91
vdw         106              2.780      0.0260     0.91

      ##################################
      ##                              ##
```

```
      ##  Bond Stretching Parameters  ##
      ##                              ##
      ##################################

bond         85   86          775.0      1.4798
bond         85   87          450.0      1.6267
bond         85   88          450.0      1.6032
bond         85   89          350.0      1.7910
bond         87   93          465.0      1.4311
bond         87    8          465.0      1.4311
bond         88   96          465.0      1.4444
bond         89   90          400.0      1.0912
bond         89   91          400.0      1.0924
bond         89   92          400.0      1.0902
bond         93   94          400.0      1.0952
bond         93   95          400.0      1.0924
bond         93   99          385.0      1.5245
bond         96   97          400.0      1.0944
bond         96   98          400.0      1.0939
bond         96  103          385.0      1.5130
bond         99  100          400.0      1.0960
bond         99  101          400.0      1.5443
bond         99  102          400.0      1.4522
bond        103  104          400.0      1.0922
bond        103  105          400.0      1.0939
bond        103  106          400.0      1.0920

      ##################################
      ##                              ##
      ##  Angle Bending Parameters    ##
      ##                              ##
      ##################################

angle        86   85   87      75.86     113.74
angle        86   85   88      75.86     117.56
angle        86   85   89      80.00     116.39
angle        87   85   88      65.58     100.07
angle        87   85   89      80.00     105.40
angle        88   85   89      80.00     101.57
angle        85   87   93      80.30     117.04
angle        85   87    8      80.30     117.04
angle        85   88   96      80.30     118.30
angle        85   89   90      35.00     109.30
angle        85   89   91      35.00     108.71
angle        85   89   92      35.00     109.73
angle        90   89   91      34.50     109.10
angle        90   89   92      34.50     110.26
angle        91   89   92      34.50     109.71
angle        87   93   94      51.50     108.07
angle        87   93   95      51.50     111.51
angle        87   93   99      88.00     107.61
angle        94   93   95      34.50     109.31
angle        94   93   99      38.00     110.29
angle        95   93   99      38.00     110.01
angle        87    8    9      51.50     108.07
angle        87    8    9      51.50     111.51
angle        87    8    7      88.00     107.61
angle         9    8    9      34.50     109.31
angle         9    8    7      38.00     110.29
angle         9    8    7      38.00     110.01
angle        88   96   97      51.50     108.54
angle        88   96   98      51.50     109.09
angle        88   96  103      88.00     107.43
angle        97   96   98      34.50     108.94
angle        97   96  103      38.00     111.37
angle        98   96  103      38.00     111.39
angle        93   99  100      38.00     107.20
angle        93   99  101      38.00     116.12
```

```
angle        93   99  102      38.00     111.26
angle       100   99  101      34.50     105.97
angle       100   99  102      34.50     106.63
angle       101   99  102      34.50     109.09
angle        96  103  104      38.00     110.33
angle        96  103  105      38.00     109.84
angle        96  103  106      38.00     110.26
angle       104  103  105      34.50     108.78
angle       104  103  106      34.50     108.82
angle       105  103  106      34.50     108.77

      ################################
      ##                            ##
      ##  Stretch-Bend Parameters   ##
      ##                            ##
      ################################

strbnd       86   85   87      14.40      14.40
strbnd       86   85   88      14.40      14.40
strbnd       86   85   89      14.40      14.40
strbnd       87   85   88      14.40      14.40
strbnd       87   85   89      14.40      14.40
strbnd       88   85   89      14.40      14.40
strbnd       85   87   93      38.00      38.00
strbnd       85   87    8      38.00      38.00
strbnd       85   88   96      38.00      38.00
strbnd       85   89   90      18.70      11.50
strbnd       85   89   91      18.70      11.50
strbnd       85   89   92      18.70      11.50
strbnd       87   93   94      18.70      11.50
strbnd       87   93   95      18.70      11.50
strbnd       87   93   99      18.70      18.70
strbnd       94   93   99      11.50      18.70
strbnd       95   93   99      11.50      18.70
strbnd       87    8    9      18.70      11.50
strbnd       87    8    9      18.70      11.50
strbnd       87    8    7      18.70      18.70
strbnd        9    8    7      11.50      18.70
strbnd        9    8    7      11.50      18.70
strbnd       88   96   97      18.70      11.50
strbnd       88   96   98      18.70      11.50
strbnd       88   96  103      18.70      18.70
strbnd       97   96  103      11.50      18.70
strbnd       99   96  103      11.50      18.70
strbnd       93   99  100      18.70      11.50
strbnd       93   99  101      18.70      11.50
strbnd       93   99  102      18.70      11.50
strbnd       96  103  104      18.70      11.50
strbnd       96  103  105      18.70      11.50
strbnd       96  103  106      18.70      11.50

      ############################
      ##                        ##
      ##  Torsional Parameters  ##
      ##                        ##
      ############################

torsion       1    7    8   87      0.192 0.0 1  -0.115 180.0 2  -0.309 0.0 3
torsion       3    7    8   87     -1.664 0.0 1  -0.531 180.0 2   0.000 0.0 3
torsion       6    7    8   87      0.000 0.0 1   0.000 180.0 2   0.341 0.0 3
torsion      86   85   87   93     -2.000 0.0 1  -1.680 180.0 2  -0.800 0.0 3
torsion      88   85   87   93     -2.000 0.0 1  -1.680 180.0 2  -0.800 0.0 3
torsion      89   85   87   93     -2.000 0.0 1  -1.680 180.0 2  -0.800 0.0 3
torsion      86   85   87    8     -2.000 0.0 1  -1.680 180.0 2  -0.800 0.0 3
torsion      88   85   87    8     -2.000 0.0 1  -1.680 180.0 2  -0.800 0.0 3
torsion      89   85   87    8     -2.000 0.0 1  -1.680 180.0 2  -0.800 0.0 3
torsion      86   85   88   96     -2.000 0.0 1  -1.680 180.0 2  -0.800 0.0 3
torsion      87   85   88   96     -2.000 0.0 1  -1.680 180.0 2  -0.800 0.0 3
```

```
torsion      89   85   88   96     -2.000 0.0 1  -1.680 180.0 2  -0.800 0.0 3
torsion      86   85   89   90      0.000 0.0 1   2.500 180.0 2   0.500 0.0 3
torsion      86   85   89   91      0.000 0.0 1   2.500 180.0 2   0.500 0.0 3
torsion      86   85   89   92      0.000 0.0 1   2.500 180.0 2   0.500 0.0 3
torsion      87   85   89   90      0.000 0.0 1   2.500 180.0 2   0.500 0.0 3
torsion      87   85   89   91      0.000 0.0 1   2.500 180.0 2   0.500 0.0 3
torsion      87   85   89   92      0.000 0.0 1   2.500 180.0 2   0.500 0.0 3
torsion      88   85   89   90      0.000 0.0 1   2.500 180.0 2   0.500 0.0 3
torsion      88   85   89   91      0.000 0.0 1   2.500 180.0 2   0.500 0.0 3
torsion      88   85   89   92      0.000 0.0 1   2.500 180.0 2   0.500 0.0 3
torsion      85   87   93   94      0.000 0.0 1   0.000 180.0 2   0.750 0.0 3
torsion      85   87   93   95      0.000 0.0 1   0.000 180.0 2   0.750 0.0 3
torsion      85   87   93   99      2.000 0.0 1  -1.500 180.0 2   0.890 0.0 3
torsion      85   87    8    9      0.000 0.0 1   0.000 180.0 2   0.750 0.0 3
torsion      85   87    8    9      0.000 0.0 1   0.000 180.0 2   0.750 0.0 3
torsion      85   87    8    7      2.000 0.0 1  -1.500 180.0 2   0.890 0.0 3
torsion      85   88   96   97      0.000 0.0 1   0.000 180.0 2   0.750 0.0 3
torsion      85   88   96   98      0.000 0.0 1   0.000 180.0 2   0.750 0.0 3
torsion      85   88   96  103      2.000 0.0 1  -1.500 180.0 2   0.890 0.0 3
torsion      87   93   99  100      0.000 0.0 1   0.000 180.0 2   0.300 0.0 3
torsion      87   93   99  101      0.000 0.0 1   0.000 180.0 2   0.300 0.0 3
torsion      87   93   99  102      0.000 0.0 1   0.000 180.0 2   0.300 0.0 3
torsion      94   93   99  100      0.000 0.0 1   0.000 180.0 2   0.300 0.0 3
torsion      94   93   99  101      0.000 0.0 1   0.000 180.0 2   0.300 0.0 3
torsion      94   93   99  102      0.000 0.0 1   0.000 180.0 2   0.300 0.0 3
torsion      95   93   99  100      0.000 0.0 1   0.000 180.0 2   0.300 0.0 3
torsion      95   93   99  101      0.000 0.0 1   0.000 180.0 2   0.300 0.0 3
torsion      95   93   99  102      0.000 0.0 1   0.000 180.0 2   0.300 0.0 3
torsion      88   96  103  104      0.000 0.0 1   0.000 180.0 2   0.300 0.0 3
torsion      88   96  103  105      0.000 0.0 1   0.000 180.0 2   0.300 0.0 3
torsion      88   96  103  106      0.000 0.0 1   0.000 180.0 2   0.300 0.0 3
torsion      97   96  103  104      0.000 0.0 1   0.000 180.0 2   0.300 0.0 3
torsion      97   96  103  105      0.000 0.0 1   0.000 180.0 2   0.300 0.0 3
torsion      97   96  103  106      0.000 0.0 1   0.000 180.0 2   0.300 0.0 3
torsion      98   96  103  104      0.000 0.0 1   0.000 180.0 2   0.300 0.0 3
torsion      98   96  103  105      0.000 0.0 1   0.000 180.0 2   0.300 0.0 3
torsion      98   96  103  106      0.000 0.0 1   0.000 180.0 2   0.300 0.0 3

      ####################################
      ##                                ##
      ##  Atomic Multipole Parameters   ##
      ##                                ##
      ####################################

multipole    12   33   422              0.09030
                                        0.00560    0.00000    0.05742
                                       -0.00170
                                        0.00000    0.00174
                                        0.01050    0.00000   -0.00004
multipole   414  415  416               1.69045
                                        0.24267   -0.45214    0.29883
                                        0.11758
                                        0.32225   -0.74558
                                       -0.31590    0.82818    0.62800
multipole   415  414  416              -1.02875
                                       -0.01494    0.01429   -0.39608
                                       -0.03170
                                       -0.01391   -0.04015
                                       -0.00248    0.06130    0.07185
multipole   416  414  422              -0.52259
                                        0.39351    0.13488   -0.13860
                                        0.30309
                                        0.07187   -0.51356
                                       -0.36841    0.19479    0.21047
multipole   417  414  425              -0.54893
                                        0.25928   -0.03213   -0.19711
                                        0.40317
                                        0.01086   -0.63018
```

```
                                    -0.67023   -0.07801    0.22701
multipole   418  414  419           -0.32762
                                    -0.00957   -0.10117   -0.12667
                                     0.18434
                                     0.05021    0.25543
                                     0.00411   -0.14122   -0.43977
multipole   419  418  414            0.09691
                                     0.00397   -0.00238   -0.11980
                                     0.09296
                                     0.00386    0.02741
                                    -0.02781    0.01110   -0.12037
multipole   420  418  414            0.08168
                                    -0.02036   -0.01537   -0.13351
                                    -0.02445
                                     0.02042    0.06811
                                    -0.00551    0.01003   -0.04366
multipole   421  418  414            0.09379
                                     0.01422    0.00277   -0.11660
                                     0.07425
                                    -0.00898    0.00102
                                    -0.02497   -0.00608   -0.07527
multipole   422  416  428            0.11651
                                     0.26067    0.06540    0.25276
                                     0.06007
                                    -0.00427   -0.78194
                                    -0.55390    0.10667    0.72187
multipole   422  416  33             0.11651
                                     0.26067    0.06540    0.25276
                                     0.06007
                                    -0.00427   -0.78194
                                    -0.55390    0.10667    0.72187
multipole   423  422  416            0.05046
                                    -0.00001    0.00337   -0.10981
                                     0.03289
                                     0.01632    0.01812
                                    -0.01314   -0.02439   -0.05101
multipole   424  422  416            0.07208
                                    -0.01441    0.01687   -0.14681
                                     0.03649
                                    -0.01396    0.01191
                                    -0.04783    0.03341   -0.04840
multipole   425  417  432            0.10122
                                     0.12662    0.00763    0.34207
                                    -0.27176
                                     0.00154   -0.50481
                                    -0.22569    0.02187    0.77657
multipole   426  425  417            0.06040
                                     0.01258   -0.01273   -0.12481
                                     0.06715
                                    -0.01949    0.06381
                                    -0.03715    0.00388   -0.13096
multipole   427  425  417            0.06813
                                     0.01105    0.00543   -0.13082
                                     0.06946
                                     0.00720    0.06074
                                    -0.02681   -0.00048   -0.13020
multipole   428  422  429           -0.16466
                                    -0.02016    0.00271    0.21650
                                    -0.20003
                                    -0.00010   -0.21820
                                    -0.01903    0.01731    0.41823
multipole   429  428  422            0.06251
                                     0.00346    0.00033   -0.10650
                                     0.04572
                                    -0.00787    0.04055
                                     0.00031   -0.00082   -0.08627
multipole   430  428  422            0.06631
                                     0.01153   -0.00246   -0.10819
```

```
                                     0.04894
                                     0.00285    0.04716
                                    -0.01999   -0.00534   -0.09610
multipole   431  428  422             0.06709
                                     0.01031    0.00080   -0.10849
                                     0.05960
                                    -0.00327    0.03473
                                    -0.01622   -0.00197   -0.09433
multipole   432  425  433            -0.16464
                                     0.00874    0.01354    0.21515
                                    -0.20881
                                     0.01034   -0.20927
                                     0.00475    0.01571    0.41808
multipole   433  432  425             0.06764
                                     0.01148    0.00291   -0.10918
                                     0.06487
                                    -0.00268    0.03285
                                    -0.01733   -0.00021   -0.09772
multipole   434  432  425             0.06296
                                     0.00267   -0.00054   -0.10602
                                     0.04882
                                    -0.00025    0.03486
                                     0.00864    0.00124   -0.08368
multipole   435  432  425             0.07117
                                     0.01127   -0.00160   -0.11053
                                     0.05916
                                     0.00375    0.03602
                                    -0.01443   -0.00052   -0.09518

      ########################################
      ##                                    ##
      ##  Dipole Polarizability Parameters  ##
      ##                                    ##
      ########################################

polarize     34          1.3340     0.3900     35   36  416  423  424  428
polarize    414          1.8280     0.3900    415  416  417  418
polarize    415          0.8370     0.3900    414
polarize    416          0.8370     0.3900    414  422
polarize    417          0.8370     0.3900    414  425
polarize    418          1.3340     0.3900    414  419  420  421
polarize    419          0.4960     0.3900    418
polarize    420          0.4960     0.3900    418
polarize    421          0.4960     0.3900    418
polarize    422          1.3340     0.3900    416  423  424  428
polarize    423          0.4960     0.3900    422
polarize    424          0.4960     0.3900    422
polarize    425          1.3340     0.3900    417  426  427  432
polarize    426          0.4960     0.3900    425
polarize    427          0.4960     0.3900    425
polarize    428          1.3340     0.3900    422  429  430  431
polarize    429          0.4960     0.3900    428
polarize    430          0.4960     0.3900    428
polarize    431          0.4960     0.3900    428
polarize    432          1.3340     0.3900    425  433  434  435
polarize    433          0.4960     0.3900    432
polarize    434          0.4960     0.3900    432
polarize    435          0.4960     0.3900    432
```

Appendix 4: GAFF Parameters for Ser203-VX adduct

```
MASS
ct 12.010         0.878
os 16.000         0.465
p  30.970         1.538
o2 16.000         0.434
h1 1.008          0.135
hc 1.008          0.135

BOND
CT-ct  300.90   1.538       same as c3-c3, penalty score=  0.0
ct-ct  300.90   1.538       same as c3-c3, penalty score=  0.0
ct-h1  330.60   1.097       same as c3-h1, penalty score=  0.0
ct-os  308.60   1.432       same as c3-os, penalty score=  0.0
os-p   330.60   1.615       same as os-p5, penalty score=  0.0
o2-p   479.50   1.487       same as  o-p5, penalty score=  0.0
ct-p   243.30   1.839       same as c3-p5, penalty score=  0.0
ct-hc  330.60   1.097       same as c3-hc, penalty score=  0.0

ANGLE
ct-CT-N    80.000     109.700
C -CT-ct   63.000     111.100
ct-ct-os   68.000     107.970   same as c3-c3-os, penalty score=  0.0
CT-ct-os   68.000     107.970   same as c3-c3-os, penalty score=  0.0
ct-ct-h1   46.390     109.560   same as c3-c3-h1, penalty score=  0.0
CT-ct-h1   46.390     109.560   same as c3-c3-h1, penalty score=  0.0
ct-CT-H1   46.390     109.560   same as c3-c3-h1, penalty score=  0.0
ct-os-p    77.720     119.540   same as c3-os-p5, penalty score=  0.0
h1-ct-os   50.800     109.780   same as h1-c3-os, penalty score=  0.0
o2-p -os   43.850     115.460   same as o -p5-os, penalty score=  0.0
os-p -os   45.000     101.840   same as os-p5-os, penalty score=  0.0
ct-p -os   40.700     100.770   same as c3-p5-os, penalty score=  0.0
hc-ct-p    53.310     108.430   same as hc-c3-p5, penalty score=  0.0
ct-p -o2   39.440     112.500   same as c3-p5-o , penalty score=  0.0
ct-ct-hc   46.340     109.800   same as c3-c3-hc, penalty score=  0.0
h1-ct-h1   39.240     108.460   same as h1-c3-h1, penalty score=  0.0
hc-ct-hc   39.400     107.580   same as hc-c3-hc, penalty score=  0.0

DIHE
N -CT-ct-os   9    1.400         0.000          3.000
h1-ct-CT-N    9    1.400         0.000          3.000
CT-ct-os-p    3    1.150         0.000         -3.000      same as c3-c3-os-p5
CT-ct-os-p    1    3.950       180.000          1.000      same as c3-c3-os-p5,
penalty score=  0.0
ct-ct-os-p    3    1.150         0.000         -3.000      same as c3-c3-os-p5
ct-ct-os-p    1    3.950       180.000          1.000      same as c3-c3-os-p5,
penalty score=  0.0
ct-os-p -o2   1    0.800         0.000         -2.000      same as c3-os-p5-o
ct-os-p -o2   1    0.550         0.000          3.000      same as c3-os-p5-o ,
penalty score=  0.0
ct-os-p -os   1    0.250         0.000         -3.000      same as os-p5-os-c3
ct-os-p -os   1    1.200         0.000          2.000      same as os-p5-os-c3,
penalty score=  0.0
ct-os-p -ct   3    2.400         0.000          2.000      same as X -os-p5-X ,
penalty score=  0.0
H1-CT-ct-os   1    0.000         0.000         -3.000      same as h1-c3-c3-os
H1-CT-ct-os   1    0.250         0.000          1.000      same as h1-c3-c3-os,
penalty score=  0.0
h1-ct-ct-os   1    0.000         0.000         -3.000      same as h1-c3-c3-os
h1-ct-ct-os   1    0.250         0.000          1.000      same as h1-c3-c3-os,
penalty score=  0.0
C -CT-ct-os   9    1.400         0.000          3.000      same as X -c3-c3-X ,
penalty score=  0.0
```

```
hc-ct-p -os    9     0.200          0.000            3.000       same as X -c3-p5-X ,
penalty score=  0.0
h1-ct-os-p     3     1.150          0.000            3.000       same as X -c3-os-X ,
penalty score=  0.0
hc-ct-p -o2    9     0.200          0.000            3.000       same as X -c3-p5-X ,
penalty score=  0.0
hc-ct-ct-os    1    0.000         0.000           -3.000      same as hc-c3-c3-os
hc-ct-ct-os    1     0.250          0.000            1.000       same as hc-c3-c3-os,
penalty score=  0.0
h1-ct-ct-hc    9     1.400          0.000            3.000       same as X -c3-c3-X ,
penalty score=  0.0
h1-ct-ct-h1    9     1.400          0.000            3.000       same as X -c3-c3-X ,
penalty score=  0.0
H1-CT-ct-h1    9     1.400          0.000            3.000       same as X -c3-c3-X ,
penalty score=  0.0
C -CT-ct-h1    9     1.400          0.000            3.000       same as X -c3-c3-X ,
penalty score=  0.0

IMPROPER

NONBON
  ct          1.9080  0.1094
  os          1.6837  0.1700
  p           2.1000  0.2000
  o2          1.6612  0.2100
  h1          1.3870  0.0157
  hc          1.4870  0.0157
```

Appendix 5: Q-OPLSAA Parameters for Ser203-VX adduct

Q-OPLSAA topology for Ser203-VX adduct

```
{SRT}
    [atoms]
        1  N      N             -0.500000 #  -0.500000 (dq=+0.000000)
        2  H      H              0.300000 #   0.300000 (dq=+0.000000)
        3  CA     CT             0.140000 #   0.140000 (dq=+0.000000)
        4  HA     HC             0.060000 #   0.060000 (dq=+0.000000)
        5  P      P_P1_450       1.197076 #   1.245930 (dq=-0.048854)
        6  O1     O2Z_O8_451    -0.739797 #  -0.711884 (dq=-0.027913)
        7  OG     OS_O1_180     -0.518279 #  -0.498724 (dq=-0.019555)
        8  O4     OS_O1_180     -0.518279 #  -0.498724 (dq=-0.019555)
        9  C1     CT_C1_135     -0.678675 #  -0.741540 (dq=+0.062865)
       10  H11    HC_H1_140      0.226225 #   0.208545 (dq=+0.017680)
       11  H12    HC_H1_140      0.226225 #   0.208545 (dq=+0.017680)
       12  H13    HC_H1_140      0.226225 #   0.208545 (dq=+0.017680)
       13  C3     CT_C1_135      0.269405 #   0.280400 (dq=-0.010995)
       14  H31    HC_H1_140      0.011528 #   0.011998 (dq=-0.000470)
       15  H32    HC_H1_140      0.011528 #   0.011998 (dq=-0.000470)
       16  C4     CT_C1_135     -0.208704 #  -0.221176 (dq=+0.012472)
       17  H41    HC_H1_140      0.069568 #   0.065855 (dq=+0.003713)
       18  H42    HC_H1_140      0.069568 #   0.065855 (dq=+0.003713)
       19  H43    HC_H1_140      0.069568 #   0.065855 (dq=+0.003713)
       20  CB     CT_C1_135      0.227488 #   0.236772 (dq=-0.009284)
       21  HB1    HC_H1_140      0.029665 #   0.030876 (dq=-0.001211)
       22  HB2    HC_H1_140      0.029665 #   0.030876 (dq=-0.001211)
       23  C      C              0.500000 #   0.500000 (dq=+0.000000)
       24  O      O             -0.500000 #  -0.500000 (dq=+0.000000)
    [bonds]
        N     H
        N     CA
        CA    HA
        P     OG
        P     O1
        P     O4
        P     C1
        OG    CB
        O4    C3
        C1    H11
        C1    H12
        C1    H13
        C3    H31
        C3    H32
        C3    C4
        C4    H41
        C4    H42
        C4    H43
        CB    HB1
        CB    HB2
        CA    CB
        CA    C
        C     O
    [impropers]
        H     N     -C    CA
        O     C     CA    +N
    [charge_groups]
        N H CA HA
        P OG O1 O4 C3 H31 H32 CB HB1 HB2
        C4 H41 H42 H43
        C1 H11 H12 H13
        C O
    [connections]
        head N
```

```
        tail C
```

Q-OPLSAA parameters for Ser203-VX adduct

```
[atom_types]
CT_C1_135        944.518     944.518     22.0296    667.8751     15.5773      12.011  #
FFLD: C: alkanes
HC_H1_140        84.5728     84.5728      5.4127      59.802      3.8274      1.0079  #
FFLD: H: alkanes
O2Z_O8_451      873.7925    873.7925     27.9561    617.8646     19.7679      15.999  #
FFLD: O= in RPO3-- WD default
OS_O1_180        445.125     445.125     18.2511    314.7509     12.9055      15.999  #
FFLD: O: OX2 default
P_P1_450       2447.7925   2447.7925     46.7907   1730.8507      33.086      30.974  #
FFLD: P in RPO3-- WD default

[bonds]
CT_C1_135    CT_C1_135          536.0       1.529  # FFLD: high 139 0 CT -CT ==> CT -
CT
CT_C1_135    HC_H1_140          680.0        1.09  # FFLD: high 140 0 CT -HC ==> CT -
HC
CT_C1_135    OS_O1_180          640.0        1.41  # FFLD: high 152 0 OS -CT ==> CT -
OS
CT_C1_135    P_P1_450           424.0       1.843  # FFLD: high 192 0 P -CT ==> CT -P
O2Z_O8_451   P_P1_450          1050.0        1.48  # FFLD: high 184 0 P -O2Z ==> O2 -P
OS_O1_180    P_P1_450           460.0        1.61  # FFLD: high 189 0 P -OS ==> OS -P

[angles]
CT_C1_135    CT_C1_135    HC_H1_140           75.0       110.7  # FFLD: high 376 0 CT
-CT -HC ==> CT -CT -HC
CT_C1_135    CT_C1_135    OS_O1_180          100.0       109.5  # FFLD: high 395 0 OS
-CT -CT ==> CT -CT -OS
CT_C1_135    OS_O1_180    P_P1_450           200.0       120.5  # FFLD: high 609 0 P -
OS -CT ==> CT -OS -P
CT_C1_135    P_P1_450     O2Z_O8_451          90.0       109.5  # FFLD: high 631 0 O2Z
-P -CT ==> CT -P -O2
CT_C1_135    P_P1_450     OS_O1_180           90.0       109.5  # FFLD: high 628 0 OS
-P -CT ==> CT -P -OS
HC_H1_140    CT_C1_135    HC_H1_140           66.0       107.8  # FFLD: high 410 0 HC
-CT -HC ==> HC -CT -HC
HC_H1_140    CT_C1_135    OS_O1_180           70.0       109.5  # FFLD: high 430 0 OS
-CT -HC ==> HC -CT -OS
HC_H1_140    CT_C1_135    P_P1_450            82.0       109.5  # FFLD: high 437 0 P -
CT -HC ==> HC -CT -P
O2Z_O8_451   P_P1_450     OS_O1_180          200.0      108.23  # FFLD: high 619 0 O2Z
-P -OS ==> O2 -P -OS
OS_O1_180    P_P1_450     OS_O1_180           90.0       102.6  # FFLD: high 625 0 OS
-P -OS ==> OS -P -OS

[torsions]
CT_C1_135    CT_C1_135    OS_O1_180    P_P1_450          1.495     2       180.0   1.0
# FFLD: high 619 0 P -OS -CT -CT ==> CA -CA -OS -P
CT_C1_135    OS_O1_180    P_P1_450     CT_C1_135          0.75    -3         0.0   1.0
# FFLD: high 608 0 CT -P -OS -CT ==> CT -P3 -OS -CT
CT_C1_135    OS_O1_180    P_P1_450     CT_C1_135         -1.65    -2       180.0   1.0
# FFLD: high 608 0 CT -P -OS -CT ==> CT -P3 -OS -CT
CT_C1_135    OS_O1_180    P_P1_450     CT_C1_135          1.75     1         0.0   1.0
# FFLD: high 608 0 CT -P -OS -CT ==> CT -P3 -OS -CT
CT_C1_135    OS_O1_180    P_P1_450     O2Z_O8_451        0.281     3         0.0   1.0
# FFLD: high 617 0 O2Z -P -OS -CT ==> O2 -P -OS -CT
CT_C1_135    OS_O1_180    P_P1_450     OS_O1_180         0.281     3         0.0   1.0
# FFLD: high 617 0 OS -P -OS -CT ==> O2 -P -OS -CT
HC_H1_140    CT_C1_135    CT_C1_135    HC_H1_140          0.15     3         0.0   1.0
# FFLD: high 5 0 HC -CT -CT -HC ==> HC -CT -CT -HC
HC_H1_140    CT_C1_135    CT_C1_135    OS_O1_180         0.234     3         0.0   1.0
# FFLD: high 42 0 OS -CT -CT -HC ==> HC -CT -CT -OS
HC_H1_140    CT_C1_135    OS_O1_180    P_P1_450         0.1785     3         0.0   1.0
# FFLD: high 618 0 P -OS -CT -HC ==> P -OS -CT -HC
HC_H1_140    CT_C1_135    P_P1_450     O2Z_O8_451        0.125     3         0.0   1.0
# FFLD: high 615 0 O2Z -P -CT -HC ==> HC -CT -P3 -O2
```

```
HC_H1_140    CT_C1_135    P_P1_450     OS_O1_180         0.125     3         0.0   1.0
# FFLD: high 616 0 OS -P -CT -HC ==> HC -CT -P3 -OS
```

Appendix 6: Python scripts

I. get_atomindex.py

This script is used to generate a list of atom indexes. Those indexes are necessary to define the QM or MM region in Chemshell. The script needs two input PDBs and will output a third PDB file. The first PDB input must contain the atoms selected for the desired QM or MM region and only those atoms. This PDB must be inputted as an argument when launching the script. The script will then ask for a reference PDB, the second input PDB. This reference PDB is the PDB for the full system used as input for Chemshell should be used. The script will then ask how to name the file containing the indexes and a verification PDB. The verification PDB is a debugging tool and part of the output of the script. It uses the selected atom indexes to regenerate a PDB of the desired QM or MM region. It should include all the atoms and only the atoms present in the PDB used as argument.

```
#!/usr/bin/python

import sys
import subprocess

def main():
    x = len(sys.argv[1:])
    if x == 0:
        input("Press any key to continue")
    else:
        arg = sys.argv[1]
        list_coordinates = make_list(arg)
        ref = raw_input ("to what pdb file do you want to compare
" + arg + " to ? ")
        dict_coordinates = make_dict(make_list(ref), ref)
        index = compare_dict(list_coordinates, dict_coordinates)
        print_list(index)
        test_script(index, ref)
    raw_input("Press any key to continue")

def make_list(arg):
    b = open(arg, "r")
    lines = b.readlines()
    true_lines = []
    for i in lines:
        if i[0:4] == "ATOM":
            true_lines.append(i)
    li_coordinates = []
    for i in true_lines :
        shortlines = i[31:54]
        li_shortlines = shortlines.split(' ')
        coordinates = []
        for i in li_shortlines:
            if i != '':
                coordinates.append(i)
        li_coordinates.append(coordinates)
    return li_coordinates

def make_dict(tot_coordinates, arg):
```

```
        b = open(arg, "r")
        lines = b.readlines()
        true_lines = []
        for i in lines:
            if i[0:4] == "ATOM":
                true_lines.append(i)
        li_index = []
        for i in true_lines :
            shortlines = i[5:11]
            li_shortlines = shortlines.split(' ')
            index = []
            for i in li_shortlines:
                if i != '':
                    index.append(i)
            li_index.append(index[0])
        dict_output = {}
        for i in range(0, (len(li_index) - 1)):
            dict_output[li_index[i]] = tot_coordinates[i]
        return dict_output

    def compare_dict(list_coordinates, dict_coordinates):
        index = ""
        for i in dict_coordinates:
            for j in list_coordinates:
                if j == dict_coordinates[i]:
                    index += str(i) + " "
        return index

    def print_list(index):
        name = raw_input("what should we name the file ? ")
        li_index = index.split(" ")
        b = open(name + ".txt", "w+")
        count2 = 0
        for i in li_index:
            b.write(str(i) + " ")
            count2 += 1
            if count2 == 15:
               b.write("\\\n")
               count2 = 0
        b.close()

    def test_script(index, arg):
        name2 = raw_input("How to name the verification/preview pdb
? ")
        c = open(name2 + ".pdb", "w+")
        li_index = index.split(" ")
        for i in li_index:
            if len(i) == 1:
               j = "M      " + i + " "
            elif len(i) == 2:
               j = "M     " + i + " "
            elif len(i) == 3:
               j = "M    " + i + " "
            elif len(i) == 4:
               j = "M   " + i + " "
            elif len(i) == 5:
```

```
                j = "M  " + i + " "
            else:
                j = "pass"
            print(i)
            print(len(i))
            print(j)
            k = "/" + j + "/ {print $0}"
            line    =    subprocess.Popen(['awk',    k,    arg],
stdout=subprocess.PIPE).communicate()[0]
            c.write(line)
        c.close()

    if __name__ == "__main__":
        main()
```

II. turbomole_scan.py

This script has been used to scan along two reaction coordinates with Turbomole. The input for the scan must be manually edited in the top rows of the script. The script scan be used to scan along one or two reactive distance. It also tolerates extra non-scanned constraints if they are defined after the reactive constraints in the “define” programme included in Turbomole, and if the *contraintes_additionelles* variable is properly set.

```
#!/usr/bin/python

import pexpect
import os
import time
import subprocess

def main():

#
# Parameters for the scan
# for convenience, the starting and end distances are in Ax10^2
# Dimensions indicates whether the scan is mono- or bi dimensional
# mono-dimensional, dimensions = "1"
# bi-dimensional, dimensions = "2"
# contraintes_additionelles is mandatory if non-reactive distances are
also une contraints
#

    dimensions = 2
    start_line = 150
    end_line = 250
    start_column = 150
    end_column = 250
    step = 10
    home = str(os.getcwd() + "/")
    print(home)
    contraintes_additionelles = 0

#
# Prepare and repare list of values
```

```
#

    lines_list = make_list(start_line, end_line, step)
    columns_list = make_list(start_column, end_column, step)
    decimals = len(str(start_line)) - 1
    corrected_lines_list = transform_list(lines_list, decimals)
    corrected_columns_list = transform_list(columns_list, decimals)
    corrected_start_column   =   str(start_column)[0]   +   "."   +
str(start_column)[-decimals:]
    corrected_columns_list.insert(0, corrected_start_column)
    corrected_start_line    =    str(start_line)[0]    +    "."    +
str(start_line)[-decimals:]
    print("corrected_lines_list :", corrected_lines_list)
    print("corrected_columns_list :", corrected_columns_list)
    print("corrected_start_column :", corrected_start_column)
    print("corrected_start_line :", corrected_start_line)

#
# Run first job
#

    print("now to run the first job\n")
    run_job(home)
    status = check_run_status()
    print("status after first step", status)
    if status == "Converged":

#
# If first job worked the script will launch the remaining jobs on the
same calculation node
#

        if dimensions == 2:
            make_2D_columns(corrected_columns_list,
corrected_lines_list,          home,          corrected_start_line,
corrected_start_column, dimensions, contraintes_additionelles)
        if dimensions == 1:
            current_column = "empy"
            run_columns(corrected_lines_list,  home,  current_column,
dimensions, contraintes_additionelles)
        else:
            print("Wrong dimensions parameter, try again")


#
# If the first job fails the script exits
#

    elif status == "Failed":
        print("Scan failed at first step for some reason")
    else:
        print("something strange happened for some reason, check it
!")
```

```
def make_2D_columns(corrected_columns_list, corrected_lines_list,
home, corrected_start_line, corrected_start_column, dimensions,
contraintes_additionelles):
    for i in range(len(corrected_columns_list)):
        print("we are working on column ", i)
        print("we are working on column ", corrected_columns_list[i])
        current_column = corrected_columns_list[i]
        print("Now we run the first column with the following values")
        print("corrected_lines_list :", corrected_lines_list)
        print("home :", home)
        print("current_column :" + current_column)
        run_columns(corrected_lines_list, home, current_column,
dimensions, contraintes_additionelles)
        print("now the column is done")
        try:
            return_directory = "../" + corrected_columns_list[i] + "-
" + corrected_start_line
            print("\nreturn_directory :", return_directory)
            print("go to return directory\n")
            os.chdir(return_directory)
            dir_path = "../../" + corrected_columns_list[i + 1]
            print("\ndir_path :", dir_path)
            try:
                print("make the directory\n")
                os.mkdir(dir_path)
            except OSError:
                pass
            new_column_directory = dir_path + "/" +
corrected_columns_list[i + 1] + "-" + corrected_start_line
            column_cpc = "cpc " + new_column_directory
            print("\nnew_column_directory :", new_column_directory)
            print("cpc to this directory\n")
            print(os.getcwd())
            attempt = subprocess.Popen(column_cpc.split(),
stdout=subprocess.PIPE).communicate()[0]
            os.chdir(new_column_directory)
            print("go to it")
            print(os.getcwd())
            print("\nrun the define with the following parameter :",
corrected_start_line, corrected_columns_list[i + 1])
            print("\n")
            make_define(corrected_start_line,
corrected_columns_list[i + 1], dimensions, contraintes_additionelles)
            run_job(home)
            status = check_run_status()
            print("status after first step", status)
        except IndexError:
            print("Well, this is probably done, needs checking")
            break

def read_machines(path):
    a = open(path, "r")
    lines = a.readlines()
    node = (str(lines[0][4] + lines[0][5]), str(lines[4][4] +
lines[4][5]), str(lines[8][4] + lines[8][5]), str(lines[12][4] +
lines[12][5]))
```

```
    return node

def  run_columns(corrected_lines_list,  home,  column,  dimensions,
contraintes_additionelles):
    for i in corrected_lines_list:
        if dimensions == 2:
            directory = "../" + column + "-" + i
        if dimensions == 1:
            directory = "../" + i
        print(directory)
        cpc = "cpc " + directory
        print(cpc)
        attempt            =            subprocess.Popen(cpc.split(),
stdout=subprocess.PIPE).communicate()[0]
        os.chdir(directory)
        print("the define will be performed with :", i, column)
        make_define(i, column, dimensions, contraintes_additionelles)
        run_job(home)
        status = check_run_status()
        if status == "Converged":
            print("step ", i, " done")
        elif status == "Failed":
            print("Scan failed at step " + i + " for some reason")
            break
        else:
            print("something strange happenned at step " + i + " for
some reason, check it !")
            break
    print("column ended")

#
#make_define interacts with the define tool of Turbomole to change the
constrained coordinates
#

def      make_define(i,      corrected_start_column,      dimensions,
contraintes_additionelles):
    dist_1 = corrected_start_column + "A"
    dist_2 = i + "A"
    print(dist_1)
    print(dist_2)
    child3 = pexpect.spawn('define')
    child3.expect("INPUT TITLE")
    child3.sendline('')
    child3.expect("DO YOU WANT TO CHANGE THE GEOMETRY DATA ?")
    child3.sendline('y')
    child3.expect("OF THAT COMMAND MAY BE GIVEN")
    child3.sendline('i')
    child3.expect("ENTER COMMAND OR HIT")
    child3.sendline('imanat')
    child3.expect("TO GET THIS MENU")
    if dimensions == 2:
        child3.sendline(dist_1)
        child3.sendline(dist_2)
    if dimensions == 1:
        child3.sendline(dist_2)
```

```
    while contraintes_additionelles != 0:
        child3.sendline('')
        contraintes_additionelles -= 1
    child3.expect("TO CONTINUE, ENTER")
    child3.sendline('')
    child3.expect("ENTER COMMAND OR HIT")
    child3.sendline('')
    child3.expect("OF THAT COMMAND MAY BE GIVEN")
    child3.sendline('ired')
    child3.expect("OF THAT COMMAND MAY BE GIVEN")
    child3.sendline('*')
    child3.expect("DO YOU WANT TO CHANGE THESE DATA ?")
    child3.sendline('')
    child3.expect("DO YOU WANT TO CHANGE THESE DATA ?")
    child3.sendline('')
    child3.expect("LEFT OVER FROM PREVIOUS CALCULATIONS ?")
    child3.sendline('')
    child3.sendline('')
    child3.sendline('')
    child3.sendline('')
    child3.sendline('')
    child3.expect("END OF DEFINE SESSION")
    child3.sendline('')
    child3.expect("END OF DEFINE SESSION")
    child3.sendline('')
    child3.sendline('*')
    child3.sendline('')
    child3.close()

def make_list(start, end, step):
    list = []
    current = start
    if end - start > 0:
        while current < end:
            current += step
            list.append(current)
        return list
    elif end - start < 0:
        while current > end:
            current -= step
            list.append(current)
        return list
    else:
        return list

def transform_list(list, decimals):
    corrected_list =[]
    for i in list:
        corrected_list.append(str(i)[0] + "." + str(i)[-decimals:])
    return corrected_list

def run_job(home):
    destination = home + "machines"
    nodes = read_machines(destination)
    child = pexpect.spawn('parwrite')
    child.expect("Nombre de noeuds a utiliser:")
```

```
    child.sendline('4')
    child.expect("Premier noeud:")
    child.sendline(str(nodes[0]))
    child.expect("Deuxieme noeud:")
    child.sendline(str(nodes[1]))
    child.expect("Troisieme noeud:")
    child.sendline(str(nodes[2]))
    child.expect("Quatrieme noeud:")
    child.sendline(str(nodes[3]))
    child.sendline()
    child.close()
    child2 = pexpect.spawn('gt')
    child2.expect("4: escf")
    child2.sendline('2')
    child2.expect("numero du noeud a utiliser")
    child2.sendline(str(nodes[0]))
    child2.sendline()
    child2.close()

def check_run_status():
    c = 0
    status = ""
    while c != 4:
        time.sleep(30)
        b = os.listdir("./")
        print("\nCheck_Geo_Opt\n")
        y = 0b1
        for i in b:
            if "GEO_OPT_CONVERGED" in i:
                y += 1
            if "GEO_OPT_FAILED" in i:
                y += 3
            if "GEO_OPT_RUNNING" in i:
                y += 7
        print("\nresult of Check\n")
        mask1 = 0b1110
        mask2 = 0b1101
        mask3 = 0b1011
        mask4 = 0b0111
        if y & mask2 == 0:
            print("Converged")
            c = 4
            status = "Converged"
        if y & mask3 == 0:
            print("Failed")
            c = 4
            status = "Failed"
        if y & mask4 == 0:
            print("Running")
            c = 0
            status = "Running"
        if y & mask1 ==0:
            print("GEO_OPT_* file is missing, call Dale Cooper")
            c += 1
            status = "Not There"
    return status
```

```
if __name__ == "__main__":
    main()
```

Thomas Driant, Méthodes et Applications en Chimie Organique, IPCM, UMR 8232, UPMC

On the Role of Protons in the Reactivation of Acetylcholinesterase: Quantum and Molecular Mechanics Studies

Abstract: The project of this PhD was to investigate the reactivation process and the active site of nerve agent inhibited AChE by computational methodologies to gain insight about the rational design of new reactivators. An initial truncated QM model study provided some insight in the necessary compensation of Glu334 by the enzyme. It also confirmed the role of the oxyanionic hole in the stabilization of the transition state of the reactivation. QM/MM simulations of the reactivation with classical reactivator 2-PAM, as well as two non-pyridinium reactivators, were performed. It was shown that Glu202, a residue near the catalytic triad of AChE, needs to be protonated for the reactivation to occur. Those simulations also showed that the reactivator can be deprotonated in the active site of AChE by His447. Non-pyridinium reactivator were found to have a greater nucleophilicity than 2-PAM and, for one of them, to be easily deprotonated in the active site. Our results indicate that the capacity of a reactivator to be deprotonated in the active site of the enzyme is more important than its nucleophilicity. Finally, a proton relay mechanism was identified through QM/MM and EVB simulations. It involves two glutamate residues, Glu450 and Glu452, positioned behind the active site. The potential for these two residues to be transiently protonated and thus involved in a proton relay was confirmed by CpHMD simulations. This proton relay mechanism relies on the N-protonation of an amide which is a novel mechanism.

Keywords: Acetylcholinesterase, Proton transfer, Reactivation, QM/MM, EVB

Du rôle des protons dans la réactivation de l'acétylcholinestérase : Etudes en mécanique quantique et mécanique moléculaire

Résumé : Le projet de cette thèse était l'évaluation du processus de réactivation et l'étude du site actif de l'AChE inhibée par un agent neurotoxique par des méthode computationnelles. L'objectif était de guider le design rationnel de nouveau réactivateurs. Une étude initiale avec un modèle QM tronqué a indiqué la nécessité de modéliser l'environnement enzymatique pour compenser la charge du Glu334. Elle a aussi confirmé le rôle du trou oxyanionique dans la stabilisation des états de transition de la réactivation. Des simulations QM/MM de la réactivation par le réactivateur classique 2-PAM, ainsi que par deux réactivateurs au cœur aromatique non chargé ont été effectuées. Il a été démontré que le Glu202, un résidu à proximité de la triade catalytique de l'AChE, doit être protoné pour que la réactivation ait lieu. Ces simulations ont aussi montré que le réactivateur peut être déprotoné dans le site actif de l'AChE par His447. Les réactivateurs au cœur aromatique non chargé sont plus nucléophiles que la 2-PAM et l'un d'entre eux est plus aisément déprotoné dans le site actif. Nos résultats indiquent que la capacité d'un réactivateur à être facilement déprotoné est plus importante que sa nucléophilie. Enfin, un mécanisme de migration de protons a été identifié par des calculs QM/MM et EVB. Il implique deux glutamates derrière le site actif, Glu450 et Glu452. La possibilité que ces deux protons soient temporairement protonés et donc impliqués dans une migration de protons a été confirmé par des calculs CpHMD. La migration de proton passe par la N-protonation d'une liaison amide, ce qui constitue un nouveau mécanisme.

Mots-clés : Acétylcholinestérase, Transfert de protons, Réactivation, QM/MM, EVB