Computational studies of the E3 carboxylesterase from *Lucilia cuprina*

Tamara Daniela Meirelles Betancur

A thesis submitted for the degree of Doctor of Philosophy at the Australian National University

Submitted for examination in November 2015

Australian National University

Declaration

The material presented in this thesis is, to the best of my knowledge, my own original work, except where acknowledged in the customary manner. This work has not been previously submitted for a degree at any institution.

Tamara Meirelles

Date:11/11/2015

Acknowledgements

First of all I should thank the two people who gave me the opportunity of an adventure in the land of kangaroos and koalas, my late parents.

I thank my supervisors, Dr Colin Jackson and Prof Michelle Coote, for their efforts in guiding my project and for academic supervision.

Help from the NCI helpdesk is gratefully acknowledged, especially from Rika Kobayashi, Ching Yeh (Leaf) Lin and Vladislav Vassiliev.

Thanks to Paul Carr for his time and patience teaching me how to solve crystal structures. Thanks also to Thomas Huber for computing time granted in his cluster. Also thanks to everyone who proofread chapters of this thesis for their helpful comments.

Finally, I would like to thank all of my ANU friends for being the awesome people that they are; and to say a big *gracias* to Andrew for so many things that would be enough to fill another thesis if I were to write them.

Abstract

The Australian sheep blowfly, *Lucilia cuprina*, has evolved resistance to organophosphate insecticides *via* a single point mutation in a carboxylesterase (E3). In addition to their use as insecticides, organophosphates have been synthesised as chemical warfare agents, posing a threat to humans. Thus, an efficient organophosphate detoxification agent, such as E3, has potential value as a prophylactic that could break down organophosphates before they cause intoxication. The work in this thesis aims at understanding the mechanism of organophosphate catalysis by E3 to reveal how insecticide resistance evolved and to evaluate whether E3 could be a useful prophylactic to prevent organophosphate poisoning.

The mechanism by which the naturally occurring mutant E3-Gly137Asp catalyses the hydrolysis of organophosphates, as predicted through quantum-cluster calculations, is presented here. Whereas the initial phosphorylation of the active site serine (Ser218) occurs in the mutant the same way as in the wild-type enzyme, the results presented here suggest that the enzyme plays two key roles in the second dephosphorylation step. First, the new Asp137 residue in the active site acts as a general base in the initial nucleophilic attack of a water molecule on the phosphorylated serine. Second, the catalytic histidine residue of E3 (His471) acts as a general acid in the dephosphorylation step, donating a proton to the departing phosphodiester. Additionally, the role of the oxyanion in lowering activation energy barriers was identified. The accuracy of computational methods for the prediction of turnover rates of enzymes is assessed. Despite the limitations present in the methods used, practically useful predictions were achieved.

The *in silico* binding of a range of different substrates, both carboxylester and organophosphate, to E3 is analysed in the light of experimental data. Predictions are made about possible natural substrates and the potential uses of E3 to degrade organophosphate chemical warfare agents. This is additionally investigated through the use of X-ray crystallographic data. Molecular dynamics simulations have been used to investigate the effects of substrate presence and mutations on the sampling of different rotamers of amino acids within the active site of E3. The implications of these results on enzyme engineering are discussed. It is suggested that the E3-Gly137Asp mutation might be detrimental to the catalytic activity of E3, owing to its sampling of non-productive conformations. The additional mutations that have accumulated in a laboratory-evolved mutant of E3 (Gly137Asp/Lys306Met/Met308Val/Ser470Gly), which has further enhanced catalytic activity, have apparently reduced the sampling of non-productive states.

Altogether, this work has utilized a range of computational techniques, from ligand docking to molecular dynamics simulations and quantum chemical simulations, to generate several novel insights into the catalytic mechanism of the recently evolved organophosphate-degrading E3-Gly137Asp mutant from *L. cuprina.* As part of a larger research program, many of these hypotheses have been tested, and supported, through enzyme kinetics and protein X-ray crystallography. The new knowledge that has been gained from this work will hopefully aid in the further improvement of this enzyme as an organophosphate detoxification agent and in better understanding how insecticide resistance can evolve.

Table of contents

Chapter 1

Introduction

1.1 The sheep blowfly *Lucilia cuprina*

"Fly strike" is a common name for cutaneous myiases (skin infections). It is a major problem for Australian sheep and is generally caused by the sheep blowfly *Lucilia cuprina*[1]. Illness and death as a result of fly strike have a large impact on the farming industry of Australia and other countries[2,3]. The sheep blowfly *L. cuprina* was introduced into Australia unintentionally in the last century, and as susceptible breeds of sheep were introduced to the country as well, the blowfly found favorable conditions to reproduce[1]. It then spread to New Zealand[4]. The infection occurs as a result of flies laying their eggs (approximately 200-250 in one batch) on the fleece of sheep; when the larvae are born they feed on the skin and flesh of the sheep causing lesions and even death[5]. Usually, rainy weather causes bacterial infections on the skin of some sheep and these generate wounds that attract female blowflies to lay their eggs. Warm weather is also a requirement for a strike to occur because the pupae need a minimum temperature to develop[3]. The most important species in blowfly strikes in Australia is *L. cuprina*[1], but other species can cause fly strike in other locations[1,6,7]. A detailed account of the biochemical and morphological changes in the cycle of *L. cuprina* can be found in the work of Barritt *et al.*[8].

1.2 Organophosphates

Phosphate monoesters and diesters occur in nature, and often have specific biological functions related to their high stability (nucleic acids and ATP are good examples of the widespread use of phosphate esters by living beings)[9]. Besides being stable, phosphate mono- and diesters have negative charge that helps protect the molecules from hydrolysis and confines them to intracellular compartments[10]. Phosphotriesters, on the other hand, are more reactive and do not occur naturally, but can be synthesized and are commonly used as pesticides[9]. Organophosphates are hydrophobic compounds that consist of a phosphoryl or thiophosphoryl group, two ester substituents and a leaving group[11] (see examples in Figure 1.1). Organophosphates make up approximately 38% of the total volume of insecticides used worldwide[12].

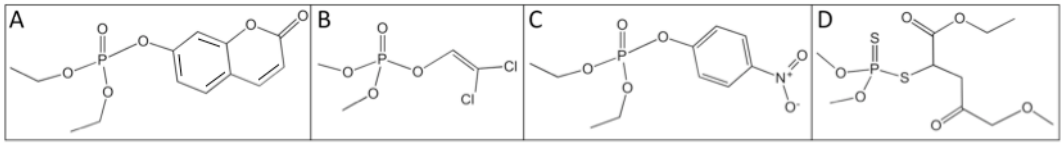


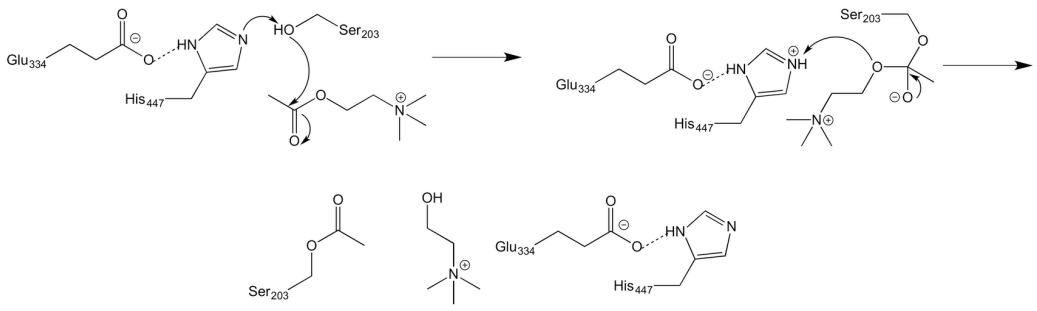
Figure 1.1. *Structures of some commonly used organophosphate pesticides.* **A.** Diethyl coumarinyl phosphate **(**dECP). **B.** Dichlorvos. **C.** Paraoxon. **D.** Malathion.

1.3 Acetylcholinesterase and the action of organophosphates

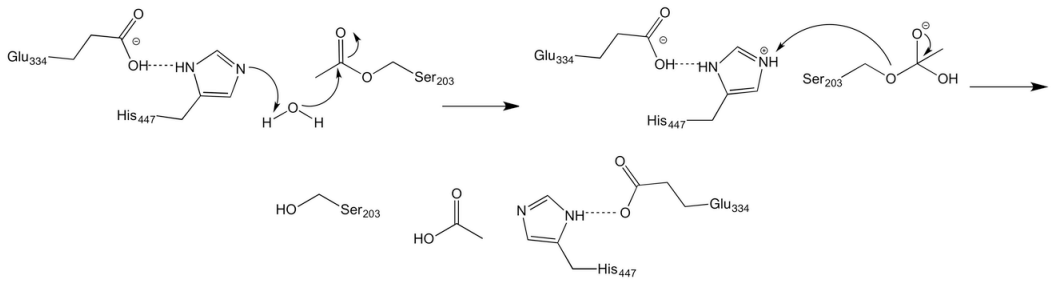
Organophosphates function as insecticides through their inhibition of insect acetylcholinesterase (AChE), but also inhibit human acetylcholinesterase[13]. Because of this, they pose a risk to human health. Indeed, human intoxication (accidental and suicidal), is frequent[14]. Acetylcholine (ACh) is a neurotransmitter present in the neuromuscular synapses. The release of acetylcholine causes muscular contraction. This action is terminated by AChE, which hydrolyses acetylcholine, producing inactive breakdown products. This enzyme has been studied extensively

and has been shown to have other roles besides terminating the nerve signal, both physiological and pathological[15]. Structures of AChE have been solved for several species, such as the fly *Drosophila melanogaster*[16], the eel *Electrophorus electricus*[17], the electric ray *Torpedo californica*[18-21], the house mouse *Mus musculus*[22,23], and human *Homo sapiens*[24-27].

Experimental[28-30] and computational[31,32] studies have shown that the hydrolysis of acetylcholine by AChE is a two-stage mechanism (Schemes 1.1 and 1.2). In the first stage of the reaction, the electrophilic carbonyl carbon of acetylcholine is attacked by the nucleophilic side chain oxygen of Ser203, which loses a proton to His447, which is in turn stabilized through its interaction with Glu334. The Glu334 residue, as was initially observed in the structurally related serine protease subtilisin[33], forms a hydrogen bond with the Nε-H of His447 while the latter takes a proton from Ser203. In the second stage, a water molecule carries out nucleophilic attack on the acylated serine that releases it, aided by general base catalysis performed by His447.



Scheme 1.1. *First stage (acylation) of the enzymatic hydrolysis of acetylcholine by acetylcholinesterase.* The histidine residue of the catalytic triad activates the serine, which attacks the substrate, acetylcholine. A negatively charged glutamate makes a hydrogen bond with the histidine, which bears a positive charge in the intermediate[28-32].



Scheme 1.2. *Second stage (deacylation) of the enzymatic hydrolysis of acetylcholine by acetylcholinesterase.* The histidine residue activates a water molecule that, in turn, attacks the acylated serine. The serine residue is released from the acyl intermediate and regains its proton from the histidine. The glutamate residue makes a hydrogen bond with the charged histidine, as in the first stage[28-32].

Acetylcholinesterase is a monomeric enzyme consisting of over five hundred amino acid residues[34] (see Figure 1.2). Its structure features an α/β hydrolase fold (this feature is described in Ollis *et al.*[35]), which consists of β-sheets connected by α-helices. The active site of this serine hydrolase lies at the bottom of a narrow gorge that leads into the protein, and consists of a catalytic triad that is made of Ser-His-Glu (rather than Ser-His-Asp) and a substrate-binding pocket[34,36]. Its mechanism of action (which will be described in more detail in Chapter 3) involves a nucleophilic attack of the active site serine on the carbonyl carbon of acetylcholine, with elimination of the choline moiety[28,29,31,32]. AChE can undergo a similar reaction with organophosphate compounds, which causes enzyme inhibition. This has been reviewed extensively (see for example reviews by Marrs[37], Pope[38], Costa[39], King *et al.*[40]). The mechanistic aspects of AChE inhibition will also be discussed in more detail in Chapter 3. Essentially inhibition of the enzyme occurs because organophosphates bind it the way the substrate does but they do not undergo the release step. Inhibition of AChE causes accumulation of acetylcholine in the neuromuscular synapses, leading to prolonged muscle contraction and death[41]. AChE also undergoes a slower process, namely an aging reaction, that was first observed decades ago[42] and renders the enzyme irreversibly

inactivated. The mechanism of aging of phosphorylated AChE is still being studied, but it involves the formation of a highly stable bond with the inhibitor. Alternative mechanisms have been proposed that would depend on the stereochemistry of the substrate[43,44], as well as a possible migration mechanism[45].

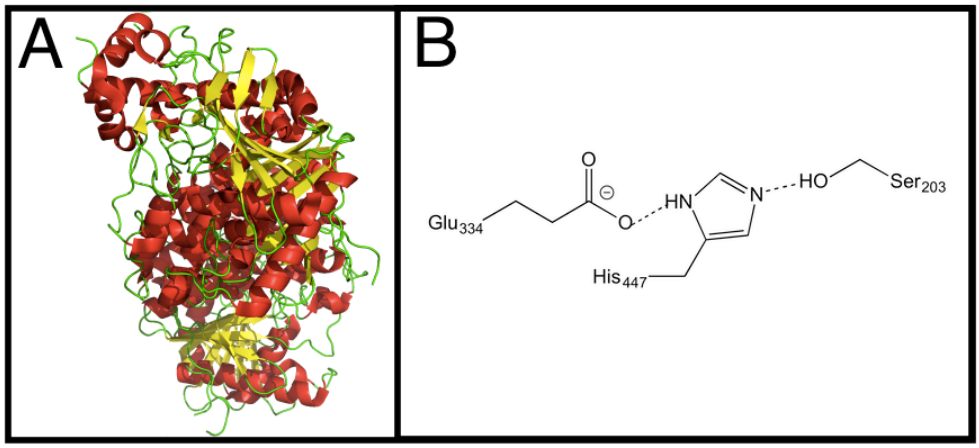


Figure 1.1. *Active site of AChE.* **A.** Human AChE (PDB ID 4MOE). **B.** Active site in detail.

1.4 Health risks associated with organophosphate pesticides

Besides the risks associated with organophosphate insecticides, such as accidental and deliberate acute intoxication, the lack of proper disposal and contamination can have consequences that are hard to measure. Besides acute effects such as nausea, skin problems and pain[46], organophosphate pesticides cause organophosphate-induced delayed polyneuropathy[47] and intermediate syndrome[48] amongst other neuropsychiatric syndromes[49,50]. It has been suggested that organophosphates may also be implied in diseases like Bovine Spongiform Encephalopathy and the Creutzfeldt Jakob disease[51].

Other types of organophosphate compounds have been developed as warfare nerve agents[52]. These include a G-series - tabun (GA), sarin

(GB), soman (GD), cyclosarin (GF) - and a V-series (VE, VG, VM, VR, VX) (see Figure 1.3). These compounds react with AChE very quickly[53] and may undergo an aging reaction, as discussed above, that makes it impossible to reactivate the enzyme. Use of warfare agents in terrorist attacks[54] and the existence of stockpiles – mainly in the U.S. and Russia[55] require the development of prophylactics for intoxication as well as safe means of destruction. Engineered enzymes could provide such means in a clean and efficient way.

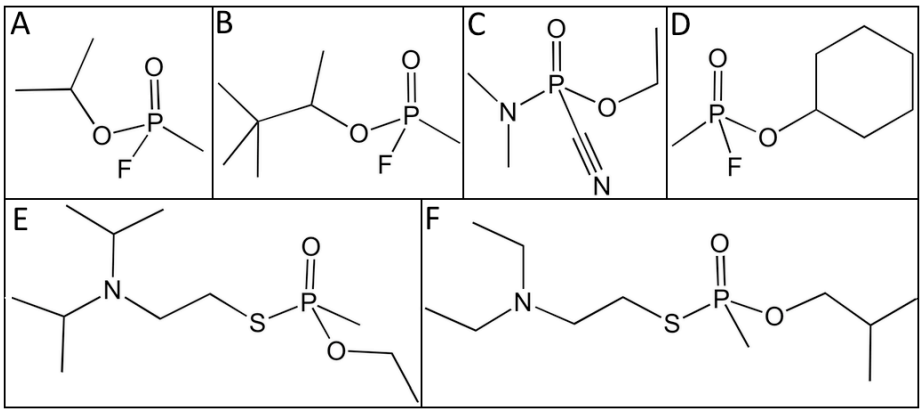


Figure 1.3. Nerve agents. **A.** Sarin. **B.** Soman. **C.** Tabun. **D.** Cyclosarin. **E.** VX. **F.** VR.

1.5 Treatment of organophosphate poisoning

Traditional treatment of organophosphate intoxication consists of prophylactic inhibition of AChE and administration of acetylcholine antagonists[56]. The first is achieved by the use of compounds that are known to be reversible inhibitors of AChE such as the carbamates aminostigmine, pyridostigmine and physostigmine, which carbamylate AChE and temporarily inactivate it. AChE then undergoes spontaneous decarbamylation, which renders the enzyme active again, resulting in an increase in the effective concentration of functional AChE. Other reversible (non-covalent) inhibitors of AChE, like galanthamine and huperzine, which are currently used in the treatment of Alzheimer's disease, have also been shown to protect AChE from irreversible

inactivation by warfare agents[56]. Again, these reversible inhibitors compete with organophosphates for access to the active site, preventing irreversible inhibition by organophosphates, and can later be displaced to restore functionality. Anticholinergics, such as atropine, ameliorate the effects of excess acetylcholine on muscarinic receptors and are used for post-exposure treatment to counteract the accumulation of acetylcholine upon inhibition of AChE[57].

Oximes are another class of therapeutic agents that deserves special attention (see Figure 1.4 for an example of the most commonly studied ones and Van Helden *et al.*[58] for a review). They reactivate organophosphate-inhibited AChE by nucleophilic attack at the P atom of the phosphorylated serine, breaking the phosphor-serine bond that keeps the enzyme inactivated. So far, antidotes tested on animals prevented death but failed to prevent the serious sequelae of organophosphate poisoning[14]. This is, in part, due to the fact that oximes do not cross the blood-brain barrier readily, which limits their ability to reactivate AChE molecules in the brain[57]. Recently, human serum albumin has been tested as an oxime carrier through the blood-brain barrier, with relative success[59]. Another problem presented by oximes is that none of them are broad-spectrum reactivators[60]. Among oximes, HI-6 is effective in reactivating VR-[61] and sarin-inhibited[62] AChE, but is ineffective against soman and tabun[62]. Recently, non-quaternary pyridine aldoximes have been described as a promising new group of AChE reactivators[63].

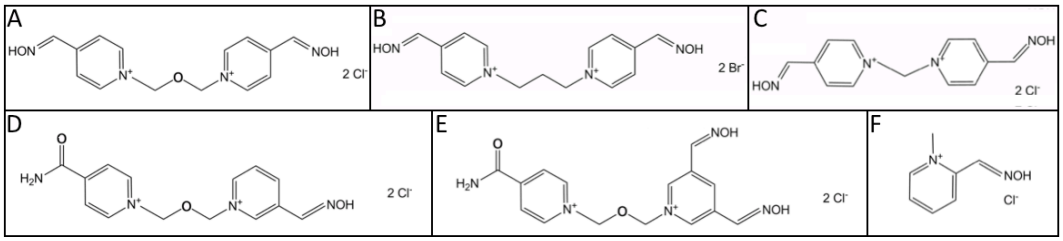


Figure 1.4. *Some well-studied oximes.* **A.** Obidoxime. **B.** Trimedoxime bromide (TMB-4). **C.** Methoxime. **D.** HI-6. **E.** HLo-7. **F.** Pralidoxime (2-PAM). Based on the work of Shrot *et al.*[64].

Other treatment strategies aim to prevent absorption of warfare agents after contact with the skin, such as through the cooling of the affected area[61] (which increases the after-exposure window for decontamination or other treatment), or the Reactive Skin Decontaminant Lotion (the latter is especially effective in preventing absorption of chemical warfare agents if administered on time)[65].

1.6 Protein-based organophosphate scavengers and detoxification enzymes

Although effective at the level of the peripheral nervous system – especially on respiration – treatment with oximes and anti-cholinergics deals with already inhibited AChE molecules rather than preventing their inactivation, and so do not protect the central nervous system from permanent damage[14,66]. New approaches to organophosphate poisoning prophylaxis and treatment are of great interest, and particular attention has been focused on engineering proteins such as antibodies, scavenging enzymes, and enzymes with organophosphate hydrolase activity[56]. Genes for *in situ* production of bioscavengers and cell therapy to regenerate neurons have also been studied[14]. *Toxicology Letters* recently dedicated a special issue (206, 2011) to the treatment of organophosphate intoxications in humans.

Bioremediation is a relatively new trend that offers a safe and efficient way of destroying potentially harmful contaminants[67,68]. Phosphotriesterases have been used for environmental remediation of organophosphate insecticides[69], and different types of phosphotriesterase have been mutated and selected for enhanced hydrolytic activity[70-73]. Recently, bacterial phosphotriesterase (PTE) was computationally redesigned and evolved in the laboratory for chemical warfare agent destruction[74,75]; human butyrylcholinesterase has been shown to act as a scavenger for organophosphates[14]; and serum paraoxonase 1 (PON1) was selectively evolved for the hydrolysis of sarin, cyclosarin and soman[76].

Phosphotriesterases (PTEs) and organophosphate-sequestering proteins are widespread in nature: they have been found in animals, plants, bacteria and archaea. The following sections provide a summary of the main classes that have been reported.

1.6.1 Human and animal phosphotriesterases

Butyrylcholinesterase (BuChE, a serum cholinesterase present in humans) has been shown to protect AChE from organophosphate inactivation by scavenging organophosphate compounds[77]. This sequestration ability has led to studies that show that it can confer soman protection to humans. However, large doses of BuChE would need to be administered, and this would have a high financial cost given that the enzyme is derived from human serum, which is costly to stock[14]. A polyethylene glycol (PEG)-ylated recombinant form of human BuChE was shown to protect mice from VX exposure[78]. Mutants of this enzyme have been designed and tested, of which Gly117His is the most active (although not active enough for protection of humans[14]). There is no agreement on the role that the mutant histidine plays in the hydrolysis reaction, but different hypotheses

have been postulated: that this histidine acts as a hydrogen bond acceptor to activate a water molecule[79], that it causes a structural rearrangement of the oxyanion hole (which in turn facilitates the access of a water molecule to the phosphorus)[80], that the histidine performs nucleophilic attack on the phosphorus[81], that it would promote a proton abstraction on a water molecule by another histidine[82], and that His117 would strain the adduct conformation therefore destabilizing it[83].

Serum paraoxonase (PON) is a soluble enzyme that is associated with high density lipoproteins (HDL) and plays a role in lipid metabolism; its native role is believed to be as a lactonase[84,85]. Three enzymes of this kind are known, PON1, PON2 and PON3[86]. A recently published study[76] on PON1 enzymes with multiple mutations found $k_{cat}/K_M = 10^7$ $M^{-1}s^{-1}$ for cyclosarin and soman, and 10^6 for sarin. These properties are enough, according to the authors, for protection of AChE *in vivo*. This PON1 mutant also has activity against VX. It has S_P stereospecificity, although it would be due to reduced R_P isomer hydrolysis more than to an increase in S_P hydrolysis. Recently, multiple mutants with catalytic enhancements of 2 and 3 orders of magnitude for S_P isomers have been described[76].

Diisopropyl fluorophosphatase (DFPase), present in species like the squid *Loligo vulgaris*, is another enzyme with the ability to destroy organophosphate compounds[87]. The mechanism that has been proposed involves a Ca^{2+} ion that coordinates an aspartate residue in the active site (depicted in Figure 1.5(A)) and which would activate the substrate for water attack[87]. DFPase laboratory-made mutants with more space in the binding pocket for a Walden inversion around the phosphorus atom have shown reversed enantioselectivity. These mutants, such as Glu37Ala(Asp)/Tyr144Ala/Arg146Ala/Thr195Met, have similar catalytic

activity to wild type but hydrolyze the S_P isomers more readily than the R_P ones[88].

E3 is an insect carboxylesterase that can also hydrolyse and detoxify organophosphorus insecticides. It is encoded by the *LcaE7* gene of the blowfly *L. cuprina* and confers resistance to organophosphate pesticides owing to some low-level organophosphate-hydrolase activity[89,90]. Mutations that confer on E3 an enhanced ability to hydrolyze organophosphate compounds have been classified in two types[3]: the *diazinon resistant* flies harbour a Gly137Asp mutation,[89] which is more active against diethyl than dimethyl organophosphates (these flies are not resistant to malathion)[91]. This same mutation confers resistance to the house fly *Musca domestica*[92]. The second type of resistant fly, *malathion resistant*, is more resistant to dimethyl organophosphates and highly resistant to malathion[91]. In this mutant, organophosphate resistance is associated to the Trp251Leu mutation[93]. Similar mutations have been described in other species[94,95], which has been interpreted as an indication of constrained possibilities for organophosphate resistance[96], which would have implications for insecticide design. Double resistance (to malathion and diazinon) has also been described[97]. The rapid evolution of resistance is thought to be due to preexisting mutations in the blowfly population[96].

The active site residues of E3 are depicted in Figure 1.5(B). The mechanism by which E3 hydrolyzes organophosphates and the role of the Gly317Asp mutation have not yet been studied in molecular detail and is the focus of this thesis. This enzyme, and its mutants that are able to hydrolyse organophosphates, are promising candidates for the detoxification task as they are structurally very similar to AChE and have high affinity for organophosphates[98]. The active site of E3 is similar to that of AChE, with a catalytic triad made of Ser-His-Glu (residues number 218, 471 and 351,

respectively), and an oxyanion hole composed of the backbone NH groups of Gly136, Gly137 and Ala219[98]. The naturally occurring E3 mutations Gly137Asp and Trp251Leu enable the enzyme to hydrolyze some common pesticides[89,93]; and laboratory redesign for hydrolysis of pyrethroids has been successfully carried out[90,99]. Other natural and artificial mutations have also been tested that increase hydrolysis of pesticides[90,99-101]. Among these, Phe309Leu and the double mutation Trp251Leu/Phe309Leu increases hydrolysis of malathion and of organophosphates dECP (diethyl coumarinyl phosphate) and dMUP (dimethyl umbelliferyl phosphate) by an order of magnitude[99]. Data of some mutations and their effect on catalysis is shown in Table 1.1.

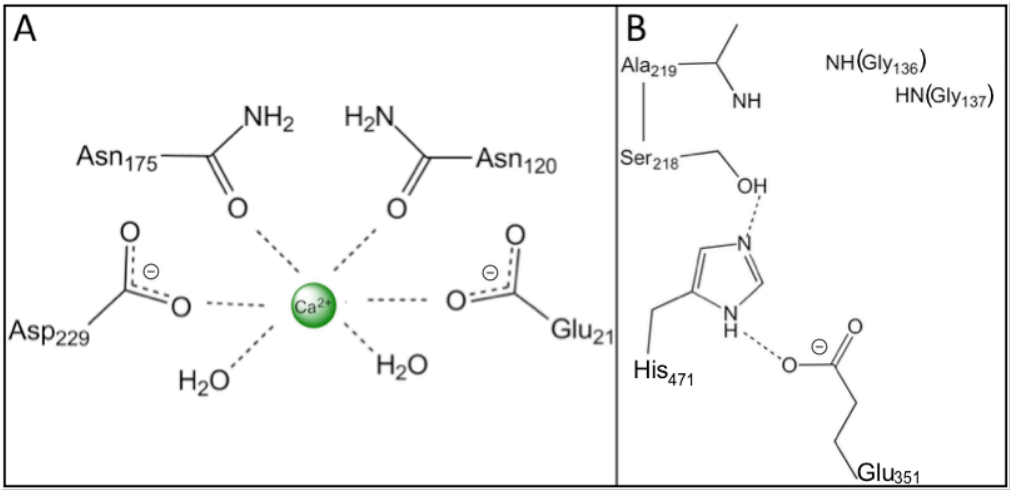


Figure 1.5. *Active site structures of DFPase and E3.* **A.** Active site of DFPase (PDB ID 2GVV). **B.** Active site of *L. cuprina* E3 (PDB ID 4FNG).

	dECP	dMUP	Malathion
Mutant	k_{cat} min^{-1}		
Wild type E3	9.0×10^{-4}	1.8×10^{-3}	54
W251L	9.2×10^{-3}	6.1×10^{-2}	220
W251G	3.0×10^{-3}	4.6×10^{-3}	307
W251A	1.9×10^{-2}	8.1×10^{-2}	68
F309L	4.7×10^{-3}	3.4×10^{-3}	172
W251L/F309L	4.9×10^{-3}	1.6×10^{-2}	147
G137D	5.0×10^{-2}	5.7×10^{-2}	7.8

Table 1.1. *Kinetic data for wild type E3 and mutants with different organophosphate substrates.* Data from Heidari *et al*[99]. dECP = Diethyl coumarinyl phosphate; dMUP = Dimethyl umbelliferyl phosphate.

1.6.2 Bacterial phosphotriesterases

Bacterial PTEs can be classified as organosphorus hydrolase (OPH)[102,103], methyl parathion hydrolase (MPH)[104], PTE-like lactonases (PPL) and organophosphorus acid anhydrase (OPAA)[105]. Bacteria that are able to hydrolyze organophosphate pesticides were first identified decades ago[106] and have lately regained attention for bioremediation of pesticides among other uses[69].

Organophosphorus hydrolases (also usually termed "bacterial phosphotriesterases" or just "PTE") have attracted attention for use in organophosphate detoxification due to their broad substrate specificity and ability to hydrolyse the O-P bond in compounds with an S-substituted leaving group[12,107]. These enzymes are members of the amidohydrolase superfamily and share a characteristic α/β barrel fold, although they have a unique reaction mechanism[12]. The X-ray crystallographic structure of the PTE from the bacteria *Brevundimonas diminuta* shows it is a homodimeric enzyme with a binuclear zinc center in each subunit[108]. This enzyme catalyzes the hydrolysis of organophosphate pesticides like as paraoxon[109], and was also shown to hydrolyze warfare agents such as soman and sarin[110]. Several bacteria that contain *opd* (organophosphorus detoxification) genes have been identified[111]. The natural substrate of these enzymes remains unknown[112]. The active site (depicted in Figure 1.6) features a carboxylated lysine residue (Lys169) and two Zn^{2+} ions. One of these is ligated to Lys169, His201, His230 and a bridging water or hydroxide forming a tetrahedral geometry. The other zinc ion, 3.3 Å apart from the first, is buried among His55, His57, Lys169 and Asp301 and also coordinates the bridging water, having a trigonal bipyramidal geometry[108,113-115]. The two metal ions have also been described coordinating the phosphoryl oxygen or phosphonyl sulphur in the

substrate[116]. The mechanism of action relies on a water molecule, coordinated by the two metal ions, that is activated to attack the phosphorus atom[117]. The PTE-catalyzed reaction has been observed to involve a Walden inversion around the phosphorus center, which is consistent with this proposal[103]. Wild type PTE has a k_{cat} for the hydrolysis of R_P isomers that is 10 times higher than that for the S_P isomers[118]. The same study shows that the steric preferences of PTE are $R_PR_C<R_PS_C<<S_PR_C<S_PS_C$.

Directed evolution experiments have produced mutant OPH enzymes with increased activity towards pesticides (parathion, methyl parathion, paraoxon, coumaphos)[119-123]. In that study, most mutations that increase catalytic activity are away from the active site and are likely to exert their effect through interactions with residues closer to it. Other mutations are close to the leaving group site or to the binding pocket. A triple mutant has been constructed that shows improved efficiency for the hydrolysis of the S_PS_C isomer of a soman analog[73]. This improvement is achieved via residue substitutions at or near the active site that create a new metal binding site. Another mutant which has preference for S_P isomers was identified by Nowlan *et al.*[124]. The authors suggest that the changes in catalysis are due to steric effects originated from the positioning of the substrate leaving group in the enzyme active site. An engineered triple mutant has a k_{cat} for the S_P isomer that is 10 times larger than its rate for the R_P isomer, and about 1.5 times that of the wild type for the R_P isomer[118]. These results are summarized in Table 1.2.

PTE-like lactonases (PPL) have been isolated from thermophilic microorganisms and shown to have sequences similar to that of bacterial PTE and a conserved active site, which is one of the reasons why they were postulated to be the ancestors of actual PTEs[126,127]. They have low

promiscuous phosphoesterase activity (four and five orders of magnitude lower than OPH)[11] and are thought to have a role in quorum sensing[128]. Promiscuity in enzymes is a hot topic and its role in evolution is currently being investigated[129]. Among PLL, SsoPox and SacPox are archaeal enzymes that show exceptional thermostability[128]. Their properties have recently been reviewed by Del Vecchio *et al.*[130].

Organophosphorus acid anhydrase (OPAA**)** is another bacterial enzyme with a bi-metallic center bridged by a OH^- or water molecule, which, although initially thought to be a homodimer[131], is now proposed to be a tetramer[132]. It belongs to the dipeptidase family[12] and its active site (see Figure 1.7) harbours two metal ions that are believed to be Co^{2+} due to the enzyme's activity dependence on cobalt[133]. However, crystal structures have Zn^{2+} instead of Co^{2+}, which is thought to be due to the experimental procedure[131]. Tested with sarin and soman analogs, OPAA shows preference for the R isomers. It displays similar stereoselectivity to OPH and a comparable rate for soman, but is 3 orders of magnitude slower than that of OPH for sarin[134].

Finally, methyl parathion hydrolase is present in various bacteria, and is active against diverse organophosphate compounds. It is a dimer with a binuclear center that belongs to the β-lactamase superfamily and evolved independently from the *opd* gene family[12].

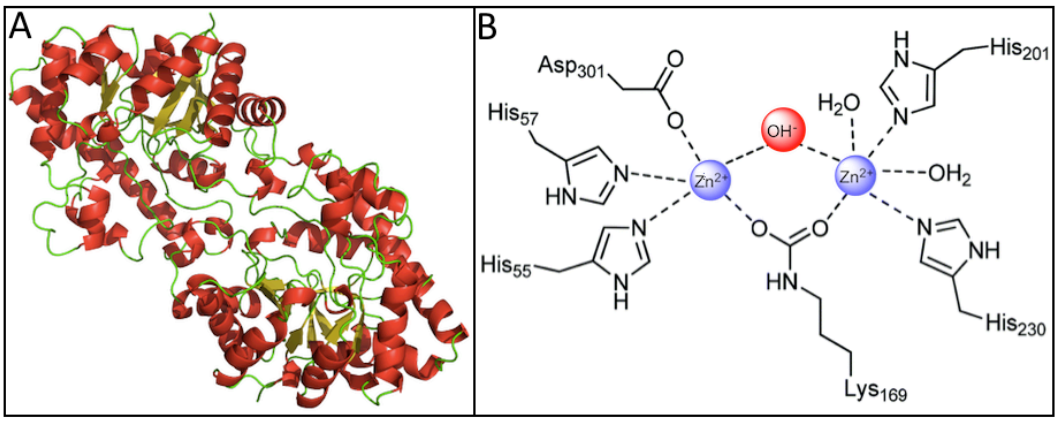


Figure 1.6. *Active sie of PTE.* **A.** Phosphotriesterase from *P. diminuta* (PDB ID: 1DPM). **B.** Active site of the PTE which includes a carbamylated lysine (based on Vanhooke *et al.*[113]).

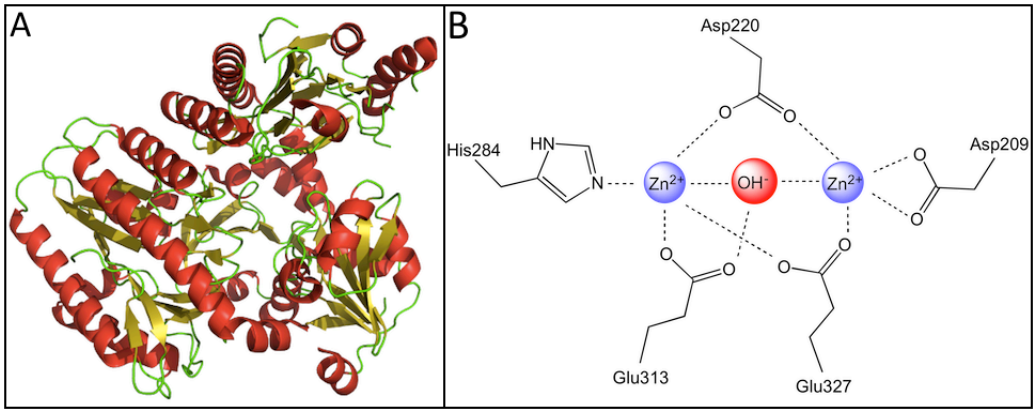


Figure 1.7. *Active site of OPAA.* **A.** OPAA from *P. furiosus* (PDB ID: 1PV9) **B.** Active site of OPAA (based on Maher *et al.*[131]).

	Sarin	Soman	Stereoselectivity
	k_{cat}/K_M ($\times 10^4$) ($M^{-1}s^{-1}$)		
Wild type OPH	9	0.26	R_PR_S,R_PR_C,S_PR_C >> S_PS_C
G60A	12	2.5	R_PS_C,R_PR_C > S_PS_C,S_PR_C
H257Y/L303T	30	8.9	S_PS_C,S_PR_C > R_PS_C,R_PR_C
H254G/H257Y/L303T	1.5	0.22	S_PS_C,S_PR_C >> R_PS_C,R_PR_C

Table 1.2. *Stereospecificity data of OPH.* Values for *B. diminuta.* Data extracted from Tsai *et al.*[135].

1.7 Challenges and aims

Organophosphate insecticide resistance is problematic for farmers, however, it could be turned into useful applications if these enzymes can be engineered to become efficient catalysts for breaking down

organphosphates dangerous to humans (see the United Nations Chemical Weapons Convention at http://www.opcw.org/chemical-weapons-convention/).

Designing a suitable enzyme for destruction of organophosphates is not trivial, as an efficient catalyst must meet a number of requirements. Some are functional, such as a high turnover rate and broad substrate specificity, while others relate to properties that make it manageable like a long shelf life, stability over a range of conditions and low cost of production[107]. Promiscuous enzymes should nevertheless be studied carefully, since they may destroy other molecules giving rise to important side effects[14]. Immune reactions are also a problematic issue if administering non-human enzymes to humans[14]. Studies aiming to determine what makes an enzyme stable and how to engineer stability have been reviewed by Goldman[136]. Applications recently developed in the field of cholinesterases and phosphoesterases include AChE nanosheets for biosensors[137] and bioremediation (for a review see Alcalde *et al.*[138]). Metal complexes have been designed as analogues of phosphoesterases, although their utility still has severe limitations (see review by Gahan *et al.*[139]). Carrier-bound enzymes and genetically modified plants that can break down pesticides are some others of the recent developments in the area[12]. Although significant progress has been made, so far engineered cholinesterases have insufficient efficiency, making it necessary to use large amounts of them[140].

The *L. cuprina* enzyme E3 is a candidate bioscavenger for the prophylactic prevention of organophosphate poisoning and bioremediation applications. This enzyme has not yet been studied in detail with regards to detoxification of organophosphate compounds toxic to humans, and its mechanism has not been fully elucidated. This thesis aims to clarify those points and to provide a rational basis for the computational prediction of

rate constants of enzymatic catalysis. In the following chapters, studies of the mechanistic aspects of E3 hydrolysis of organophosphates and of the effects of mutations on the structure and dynamics of the enzyme will be described.

1.8 References

(1) Tellam, R. L.; Bowles, V. M. *International Journal for Parasitology* **1997**, *27*, 261.

(2) Heath, A. C. G.; Levot, G. W. *New Zealand Veterinary Journal* **2015**, *63*, 199.

(3) Levot, G. W. *International Journal for Parasitology* **1995**, *25*, 1355.

(4) Sandeman, R. M.; Levot, G. W.; Heath, A. C. G.; James, P. J.; Greeff, J. C.; Scott, M. J.; Batterham, P.; Bowles, V. M. *International Journal for Parasitology* **2014**, *44*, 879.

(5) Ashworth, J. R.; Wall, R. *Medical and Veterinary Entomology* **1994**, *8*, 303.

(6) de Carvalho, R. A.; Torres, T. T.; de Azeredo-Espin, A. M. L. *Veterinary Parasitology* **2006**, *140*, 344.

(7) Salisbury, R. H.; Barrowman, P. R. *Journal of the South African Veterinary Association* **1984**, *55*, 147.

(8) Barritt, L. C.; Birt, L. M. *Journal of Insect Physiology* **1971**, *17*, 1169.

(9) Cleland, W. W.; Hengge, A. C. *Chemical Reviews* **2006**, *106*, 3252.

(10) Westheimer, F. H. *Science* **1987**, *235*, 1173.

(11) Mandrich, L.; Merone, L.; Manco, G. *Environmental Technology* **2010**, *31*, 1115.

(12) Singh, B. K. *Nature Reviews Microbiology* **2009**, *7*, 156.

(13) Goel, A.; Aggarwal, P. *National Medical Journal of India* **2007**, *20*, 182.

(14) Masson, P. *Toxicology Letters* **2011**, *206*, 5.

(15) Soreq, H.; Seidman, S. *Nature Reviews Neuroscience* **2001**, *2*, 294.

(16) Harel, M.; Kryger, G.; Rosenberry, T. L.; Mallender, W. D.; Lewis, T.; Fletcher, R. J.; Guss, J. M.; Silman, I.; Sussman, J. L. *Protein Science* **2000**, *9*, 1063.

(17) Bourne, Y.; Grassi, J.; Bougis, P. E.; Marchot, P. *Journal of Biological Chemistry* **1999**, *274*, 30370.

(18) Legler, P. M.; Soojhawon, I.; Millard, C. B. *Acta Crystallographica Section D: Biological Crystallography* **2015**, *71*, 1788.

(19) Bartolucci, C.; Stojan, J.; Yu, Q. S.; Greig, N. H.; Lamba, D. *Biochemical Journal* **2012**, *444*, 269.

(20) Sanson, B.; Colletier, J. P.; Xu, Y.; Lang, P. T.; Jiang, H.; Silman, I.; Sussman, J. L.; Weik, M. *Protein Science* **2011**, *20*, 1114.

(21) Sanson, B.; Nachon, F.; Colletier, J. P.; Froment, M. T.; Toker, L.; Greenblatt, H. M.; Sussman, J. L.; Ashani, Y.; Masson, P.; Silman, I.; Weik, M. *Journal of Medicinal Chemistry* **2009**, *52*, 7593.

(22) Andersson, C. D.; Forsgren, N.; Akfur, C.; Allgardsson, A.; Berg, L.; Engdahl, C.; Qian, W.; Ekström, F.; Linusson, A. *Journal of Medicinal Chemistry* **2013**, *56*, 7615.

(23) Artursson, E.; Andersson, P. O.; Akfur, C.; Linusson, A.; Börjegren, S.; Ekström, F. *Biochemical Pharmacology* **2013**, *85*, 1389.

(24) Kryger, G.; Harel, M.; Giles, K.; Toker, L.; Velan, B.; Lazar, A.; Kronman, C.; Barak, D.; Ariel, N.; Shafferman, A.; Silman, I.; Sussman, J. L. *Acta Crystallographica Section D: Biological Crystallography* **2000**, *56*, 1385.

(25) Carletti, E.; Colletier, J. P.; Dupeux, F.; Trovaslet, M.; Masson, P.; Nachon, F. *Journal of Medicinal Chemistry* **2010**, *53*, 4002.

(26) Nachon, F.; Carletti, E.; Ronco, C.; Trovaslet, M.; Nicolet, Y.; Jean, L.; Renard, P. Y. *Biochemical Journal* **2013**, *453*, 393.

(27) Cheung, J.; Rudolph, M. J.; Burshteyn, F.; Cassidy, M. S.; Gary, E. N.; Love, J.; Franklin, M. C.; Height, J. J. *Journal of Medicinal Chemistry* **2012**, *55*, 10282.

(28) Stein, S. S.; Koshland Jr, D. E. *Archives of Biochemistry and Biophysics* **1953**, *45*, 467.

(29) Krupka, R. M. *Biochemistry* **1967**, *6*, 1183.

(30) Colletier, J. P.; Fournier, D.; Greenblatt, H. M.; Stojan, J.; Sussman, J. L.; Zaccai, G.; Silman, I.; Weik, M. *EMBO Journal* **2006**, *25*, 2746.

(31) Zhang, Y.; Kua, J.; McCammon, J. A. *Journal of the American Chemical Society* **2002**, *124*, 10572.

(32) Vagedes, P.; Rabenstein, B.; Åqvist, J.; Marelius, J.; Knapp, E. W. *Journal of the American Chemical Society* **2000**, *122*, 12254.

(33) Jordan, F.; Polgár, L. *Biochemistry* **1981**, *20*, 6366.

(34) Sussman, J. L.; Harel, M.; Frolow, F.; Oefner, C.; Goldman, A.; Toker, L.; Silman, I. *Science* **1991**, *253*, 872.

(35) Ollis, D. L.; Cheah, E.; Cygler, M.; Dijkstra, B.; Frolow, F.; Franken, S. M.; Harel, M.; Remington, S. J.; Silman, I.; Schrag, J.; Sussman, J. L.; Verschueren, K. H. G.; Goldman, A. *Protein Engineering* **1992**, *5*, 197.

(36) Massoulié, J.; Pezzementi, L.; Bon, S.; Krejci, E.; Vallette, F. M. *Progress in Neurobiology* **1993**, *41*, 31.

(37) Marrs, T. C. *Pharmacology and Therapeutics* **1993**, *58*, 51.

(38) Pope, C. N. *Journal of Toxicology and Environmental Health - Part B: Critical Reviews* **1999**, *2*, 161.

(39) Costa, L. G. *Clinica Chimica Acta* **2006**, *366*, 1.

(40) King, A. M.; Aaron, C. K. *Emergency Medicine Clinics of North America* **2015**, *33*, 133.

(41) Mileson, B. E.; Chambers, J. E.; Chen, W. L.; Dettbarn, W.; Ehrich, M.; Eldefrawi, A. T.; Gaylor, D. W.; Hamernik, K.; Hodgson, E.; Karczmar, A. G.; Padilla, S.; Pope, C. N.; Richardson, R. J.; Saunders, D. R.; Sheets, L. P.; Sultatos, L. G.; Wallace, K. B. *Toxicological Sciences* **1998**, *41*, 8.

(42) Berends, F.; Posthumus, C. H.; Sluỳs, I. v. d.; Deierkauf, F. A. *BBA - Biochimica et Biophysica Acta* **1959**, *34*, 576.

(43) Nachon, F.; Asojo, O. A.; Borgstahl, G. E. O.; Masson, P.; Lockridge, O. *Biochemistry* **2005**, *44*, 1154.

(44) Carletti, E.; Li, H.; Li, B.; Ekström, F.; Nicolet, Y.; Loiodice, M.; Gillon, E.; Froment, M. T.; Lockridge, O.; Schopfer, L. M.; Masson, P.; Nachon, F. *Journal of the American Chemical Society* **2008**, *130*, 16011.

(45) Sirin, G. S.; Zhou, Y.; Lior-Hoffmann, L.; Wang, S.; Zhang, Y. *Journal of Physical Chemistry B* **2012**, *116*, 12199.

(46) McCauley, L. A.; Anger, W. K.; Keifer, M.; Langley, R.; Robson, M. G.; Rohlman, D. *Environmental Health Perspectives* **2006**, *114*, 953.

(47) Lotti, M.; Moretto, A. *Toxicological Reviews* **2005**, *24*, 37.

(48) Jamal, G. A. *Adverse Drug Reactions and Toxicological Reviews* **1997**, *16*, 133.

(49) Chen, Y. *NeuroToxicology* **2012**, *33*, 391.

(50) Jamal, G. A.; Hansen, S.; Julu, P. O. O. *Toxicology* **2002**, *181-182*, 23.

(51) Ragnarsdottir, K. V. *Journal of the Geological Society* **2000**, *157*, 859.

(52) Chauhan, S.; D'Cruz, R.; Faruqi, S.; Singh, K. K.; Varma, S.; Singh, M.; Karthik, V. *Environmental Toxicology and Pharmacology* **2008**, *26*, 113.

(53) Bernhard, S. A.; Orgel, L. E. *Science* **1959**, *130*, 625.

(54) Okumura, T.; Hisaoka, T.; Yamada, A.; Naito, T.; Isonuma, H.; Okumura, S.; Miura, K.; Sakurada, M.; Maekawa, H.; Ishimatsu, S.; Takasu, N.; Suzuki, K. *Toxicology and Applied Pharmacology* **2005**, *207*, S471.

(55) Kim, K.; Tsay, O. G.; Atwood, D. A.; Churchill, D. G. *Chemical Reviews* **2011**, *111*, 5345.

(56) Bajgar, J.; Fusek, J.; Kassa, J.; Kuca, K.; Jun, D. *Current Medicinal Chemistry* **2009**, *16*, 2977.

(57) Tenberken, O.; Worek, F.; Thiermann, H. *Toxicology Letters* **2011**, *206*, 3.

(58) Van Helden, H. P. M.; Busker, R. W.; Melchers, B. P. C.; Bruijnzeel, P. L. B. *Archives of Toxicology* **1996**, *70*, 779.

(59) Dadparvar, M.; Wagner, S.; Wien, S.; Kufleitner, J.; Worek, F.; von Briesen, H.; Kreuter, J. *Toxicology Letters* **2011**, *206*, 60.

(60) Mercey, G.; Verdelet, T.; Renou, J.; Kliachyna, M.; Baati, R.; Nachon, F.; Jean, L.; Renard, P. Y. *Accounts of Chemical Research* **2012**, *45*, 756.

(61) Mikler, J.; Tenn, C.; Worek, F.; Reiter, G.; Thiermann, H.; Garrett, M.; Bohnert, S.; Sawyer, T. W. *Toxicology Letters* **2011**, *206*, 47.

(62) Seeger, T.; Niessen, K. V.; Langer, P.; Gerhardus, J.; Worek, F.; Friess, H.; Bumm, R.; Mihaljevic, A. L.; Thiermann, H. *Toxicology Letters* **2011**, *206*, 72.

(63) Mercey, G.; Verdelet, T.; Saint-André, G.; Gillon, E.; Wagner, A.; Baati, R.; Jean, L.; Nachon, F.; Renard, P. Y. *Chemical Communications* **2011**, *47*, 5295.

(64) Shrot, S.; Markel, G.; Dushnitsky, T.; Krivoy, A. *NeuroToxicology* **2009**, *30*, 167.

(65) Bjarnason, S.; Mikler, J.; Hill, I.; Tenn, C.; Garrett, M.; Caddy, N.; Sawyer, T. W. *Human and Experimental Toxicology* **2008**, *27*, 253.

(66) Lenz, D. E.; Yeung, D.; Smith, J. R.; Sweeney, R. E.; Lumley, L. A.; Cerasoli, D. M. *Toxicology* **2007**, *233*, 31.

(67) Timmis, K. N.; Pieper, D. H. *Trends in Biotechnology* **1999**, *17*, 201.

(68) Helbling, D. E. *Current Opinion in Biotechnology* **2015**, *33*, 142.

(69) Sutherland, T. D.; Horne, I.; Weir, K. M.; Coppin, C. W.; Williams, M. R.; Selleck, M.; Russell, R. J.; Oakeshott, J. G. *Clinical and Experimental Pharmacology and Physiology* **2004**, *31*, 817.

(70) Chen-Goodspeed, M.; Sogorb, M. A.; Wu, F.; Raushel, F. M. *Biochemistry* **2001**, *40*, 1332.

(71) Griffiths, A. D.; Tawfik, D. S. *EMBO Journal* **2003**, *22*, 24.

(72) Hawwa, R.; Larsen, S. D.; Ratia, K.; Mesecar, A. D. *Journal of Molecular Biology* **2009**, *393*, 36.

(73) Hill, C. M.; Li, W. S.; Thoden, J. B.; Holden, H. M.; Raushel, F. M. *Journal of the American Chemical Society* **2003**, *125*, 8990.

(74) Bigley, A. N.; Xu, C.; Henderson, T. J.; Harvey, S. P.; Raushel, F. M. *Journal of the American Chemical Society* **2013**, *135*, 10426.

(75) Khare, S. D.; Kipnis, Y.; Greisen, P. J.; Takeuchi, R.; Ashani, Y.; Goldsmith, M.; Song, Y.; Gallaher, J. L.; Silman, I.; Leader, H.; Sussman, J. L.; Stoddard, B. L.; Tawfik, D. S.; Baker, D. *Nature Chemical Biology* **2012**, *8*, 294.

(76) Goldsmith, M.; Ashani, Y.; Simo, Y.; Ben-David, M.; Leader, H.; Silman, I.; Sussman, J. L.; Tawfik, D. S. *Chemistry and Biology* **2012**, *19*, 456.

(77) Doctor, B. P.; Saxena, A. *Chemico-Biological Interactions* **2005**, *157-158*, 167.

(78) Mumford, H.; Troyer, J. K. *Toxicology Letters* **2011**, *206*, 29.

(79) Schopfer, L. M.; Ticu Boeck, A.; Broomfield, C. A.; Lockridge, O. *J Med Chem Def* **2004**, *2*, 1.

(80) Yao, Y.; Liu, J.; Zhan, C. G. *Biochemistry* **2012**, *51*, 8980.

(81) Albaret, C.; Masson, P.; Broomfield, C. A.; El Kaim, L.; Fortier, P. L. In *Structure and Function of Cholinesterases and Related Proteins*; Doctor, B. P., Ed.; Plenum Press: New York, 1998, p 399.

(82) Millard, C. B.; Lockridge, O.; Broomfield, C. A. *Biochemistry* **1995**, *34*, 15925.

(83) Nachon, F.; Carletti, E.; Wandhammer, M.; Nicolet, Y.; Schopfer, L. M.; Masson, P.; Lockridge, O. *Biochemical Journal* **2011**, *434*, 73.

(84) Valiyaveettil, M.; Alamneh, Y. A.; Doctor, B. P.; Nambiar, M. P. *Toxicology Letters* **2012**, *210*, 87.

(85) Khersonsky, O.; Tawfik, D. S. *Biochemistry* **2005**, *44*, 6371.

(86) Harel, M.; Aharoni, A.; Gaidukov, L.; Brumshtein, B.; Khersonsky, O.; Meged, R.; Dvir, H.; Ravelli, R. B. G.; McCarthy, A.; Toker, L.; Silman, I.; Sussman, J. L.; Tawfik, D. S. *Nature Structural and Molecular Biology* **2004**, *11*, 412.

(87) Blum, M. M.; Löhr, F.; Richardt, A.; Ruterjans, H.; Chen, J. C. H. *Journal of the American Chemical Society* **2006**, *128*, 12750.

(88) Melzer, M.; Chen, J. C. H.; Heidenreich, A.; Gäb, J.; Koller, M.; Kehe, K.; Blum, M. M. *Journal of the American Chemical Society* **2009**, *131*, 17226.

(89) Newcomb, R. D.; Campbell, P. M.; Ollis, D. L.; Cheah, E.; Russell, R. J.; Oakeshott, J. G. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 7464.

(90) Heidari, R.; Devonshire, A. L.; Campbell, B. E.; Dorrian, S. J.; Oakeshott, J. G.; Russell, R. J. *Insect Biochemistry and Molecular Biology* **2005**, *35*, 597.

(91) Bell, J. D.; Busvine, J. R. *Entomologia Experimentalis et Applicata* **1967**, *10*, 263.

(92) Claudianos, C.; Russell, R. J.; Oakeshott, J. G. *Insect Biochemistry and Molecular Biology* **1999**, *29*, 675.

(93) Campbell, P. M.; Newcomb, R. D.; Russell, R. J.; Oakeshott, J. G. *Insect Biochemistry and Molecular Biology* **1998**, *28*, 139.

(94) Guerrero, F. D. *Insect Biochemistry and Molecular Biology* **2000**, *30*, 1107.

(95) Zhu, Y. C.; Dowdy, A. K.; Baker, J. E. *Pesticide Science* **1999**, *55*, 398.

(96) Hartley, C. J.; Newcomb, R. D.; Russell, R. J.; Yong, C. G.; Stevens, J. R.; Yeates, D. K.; La Salle, J.; Oakeshott, J. G. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 8757.

(97) Smyth, K. A.; Boyce, T. M.; Russell, R. J.; Oakeshott, J. G. *Heredity* **2000**, *84*, 63.

(98) Jackson, C. J.; Liu, J. W.; Carr, P. D.; Younus, F.; Coppin, C.; Meirelles, T.; Lethier, M.; Pandey, G.; Ollis, D. L.; Russell, R. J.; Weik, M.; Oakeshott, J. G. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 10177.

(99) Heidari, R.; Devonshire, A. L.; Campbell, B. E.; Bell, K. L.; Dorrian, S. J.; Oakeshott, J. G.; Russell, R. J. *Insect Biochemistry and Molecular Biology* **2004**, *34*, 353.

(100) Devonshire, A. L.; Heidari, R.; Bell, K. L.; Campbell, P. M.; Campbell, B. E.; Odgers, W. A.; Oakeshott, J. G.; Russell, R. J. *Pesticide Biochemistry and Physiology* **2003**, *76*, 1.

(101) Devonshire, A. L.; Heidari, R.; Huang, H. Z.; Hammock, B. D.; Russell, R. J.; Oakeshott, J. G. *Insect Biochemistry and Molecular Biology* **2007**, *37*, 891.

(102) Munnecke, D. M. *Applied and Environmental Microbiology* **1976**, *32*, 7.

(103) Lewis, V. E.; Donarski, W. J.; Wild, J. R.; Raushel, F. M. *Biochemistry* **1988**, *27*, 1591.

(104) Rani, N. L.; Lalithakumari, D. *Canadian Journal of Microbiology* **1994**, *40*, 1000.

(105) Anderson, R. S.; Durst, H. D.; Landis, W. G. *Comparative Biochemistry and Physiology. Part C, Comparative* **1988**, *91*, 575.

(106) Sethunathan, N.; Yoshida, T. *Canadian Journal of Microbiology* **1973**, *19*, 873.

(107) Di Sioudi, B. D.; Miller, C. E.; Lai, K.; Grimsley, J. K.; Wild, J. R. *Chemico-Biological Interactions* **1999**, *119-120*, 211.

(108) Benning, M. M.; Shim, H.; Raushel, F. M.; Holden, H. M. *Biochemistry* **2001**, *40*, 2712.

(109) Dumas, D. P.; Caldwell, S. R.; Wild, J. R.; Raushel, F. M. *Journal of Biological Chemistry* **1989**, *264*, 19659.

(110) Dumas, D. P.; Durst, H. D.; Landis, W. G.; Raushel, F. M.; Wild, J. R. *Archives of Biochemistry and Biophysics* **1990**, *277*, 155.

(111) Harper, L. L.; McDaniel, C. S.; Miller, C. E.; Wild, J. R. *Applied and Environmental Microbiology* **1988**, *54*, 2586.

(112) Raushel, F. M. *Current Opinion in Microbiology* **2002**, *5*, 288.

(113) Vanhooke, J. L.; Benning, M. M.; Raushel, F. M.; Holden, H. M. *Biochemistry* **1996**, *35*, 6020.

(114) Chae, M. Y.; Omburo, G. A.; Lindahl, P. A.; Raushel, F. M. *Journal of the American Chemical Society* **1993**, *115*, 12173.

(115) Omburo, G. A.; Mullins, L. S.; Raushel, F. M. *Biochemistry* **1993**, *32*, 9148.

(116) Hong, S. B.; Raushel, F. M. *Biochemistry* **1996**, *35*, 10904.

(117) Aubert, S. D.; Li, Y.; Raushel, F. M. *Biochemistry* **2004**, *43*, 5707.

(118) Li, W. S.; Lum, K. T.; Chen-Goodspeed, M.; Sogorb, M. A.; Raushel, F. M. *Bioorganic and Medicinal Chemistry* **2001**, *9*, 2083.

(119) Cho, C. M. H.; Mulchandani, A.; Chen, W. *Protein Engineering, Design and Selection* **2006**, *19*, 99.

(120) Cho, C. M. H.; Mulchandani, A.; Chen, W. *Applied and Environmental Microbiology* **2002**, *68*, 2026.

(121) Cho, C. M. H.; Mulchandani, A.; Chen, W. *Applied and Environmental Microbiology* **2004**, *70*, 4681.

(122) Yang, H.; Carr, P. D.; McLoughlin, S. Y.; Liu, J. W.; Horne, I.; Qiu, X.; Jeffries, C. M. J.; Russell, R. J.; Oakeshott, J. G.; Ollis, D. L. *Protein Engineering* **2003**, *16*, 135.

(123) Amitai, G.; Gaidukov, L.; Adani, R.; Yishay, S.; Yacov, G.; Kushnir, M.; Teitlboim, S.; Lindenbaum, M.; Bel, P.; Khersonsky, O.; Tawfik, D. S.; Meshulam, H. *FEBS Journal* **2006**, *273*, 1906.

(124) Nowlan, C.; Li, Y.; Hermann, J. C.; Evans, T.; Carpenter, J.; Ghanem, E.; Shoichet, B. K.; Raushel, F. M. *Journal of the American Chemical Society* **2006**, *128*, 15892.

(125) Tsai, P. C.; Fan, Y.; Kim, J.; Yang, L.; Almo, S. C.; Gao, Y. Q.; Raushel, F. M. *Biochemistry* **2010**, *49*, 7988.

(126) Afriat, L.; Roodveldt, C.; Manco, G.; Tawfik, D. S. *Biochemistry* **2006**, *45*, 13677.

(127) Afriat-Jurnou, L.; Jackson, C. J.; Tawfik, D. S. *Biochemistry* **2012**, *51*, 6047.

(128) Merone, L.; Mandrich, L.; Rossi, M.; Manco, G. *Current Chemical Biology* **2008**, *2*, 237.

(129) Khersonsky, O.; Roodveldt, C.; Tawfik, D. S. *Current Opinion in Chemical Biology* **2006**, *10*, 498.

(130) Del Vecchio, P.; Elias, M.; Merone, L.; Graziano, G.; Dupuy, J.; Mandrich, L.; Carullo, P.; Fournier, B.; Rochu, D.; Rossi, M.; Masson, P.; Chabriere, E.; Manco, G. *Extremophiles* **2009**, *13*, 461.

(131) Maher, M. J.; Ghosh, M.; Grunden, A. M.; Menon, A. L.; Adams, M. W. W.; Freeman, H. C.; Guss, J. M. *Biochemistry* **2004**, *43*, 2771.

(132) Vyas, N. K.; Nickitenko, A.; Rastogi, V. K.; Shah, S. S.; Quiocho, F. A. *Biochemistry* **2010**, *49*, 547.

(133) Ghosh, M.; Grunden, A. M.; Dunn, D. M.; Weiss, R.; Adams, M. W. W. *Journal of Bacteriology* **1998**, *180*, 4781.

(134) Hill, C. M.; Li, W. S.; Cheng, T. C.; DeFrank, J. J.; Raushel, F. M. *Bioorganic Chemistry* **2001**, *29*, 27.

(135) Tsai, P. C.; Bigley, A.; Li, Y.; Ghanem, E.; Cadieux, C. L.; Kasten, S. A.; Reeves, T. E.; Cerasoli, D. M.; Raushel, F. M. *Biochemistry* **2010**, *49*, 7978.

(136) Goldman, A. *Structure* **1995**, *3*, 1277.

(137) Zhang, L.; Long, L.; Zhang, W.; Du, D.; Lin, Y. *Electroanalysis* **2012**, *24*, 1745.

(138) Alcalde, M.; Ferrer, M.; Plou, F. J. *Biocatalysis and Biotransformation* **2007**, *25*, 113.

(139) Gahan, L. R.; Smith, S. J.; Neves, A.; Schenk, G. *European Journal of Inorganic Chemistry* **2009**, 2745.

(140) Trovaslet-Leroy, M.; Musilova, L.; Renault, F.; Brazzolotto, X.; Misik, J.; Novotny, L.; Froment, M. T.; Gillon, E.; Loiodice, M.; Verdier, L.; Masson, P.; Rochu, D.; Jun, D.; Nachon, F. *Toxicology Letters* **2011**, *206*, 14.

Chapter 2

Theory background

2.1 Introduction

Computer simulations complement and provide insight into experimental data, and allow in some cases for quick and inexpensive testing of large numbers of systems[1]. Deciphering reaction mechanisms in atomic detail (a focal point of this thesis) is a great strength of *in silico*[2] experiments. Some structural studies also benefit from or can only be done with the help of computational models[3]. Although it is often difficult to make use of computational methods when there is little or no experimental data available, limited data can easily be expanded and meaning extracted by their use. Because of these reasons, collaborative research that encompasses experimental and computational approaches has become very common in molecular biology and biochemistry[4,5].

How enzymes achieve the enhancement of chemical reaction rates is an active field of research, catalysis being the interplay of many different factors. It has been shown, for example, that the active site of enzymes is electrostatically preorganised to stabilise the transition state, thereby lowering the activation energy barrier[6]. However, regions of the enzyme far from the active site can also have an important influence on catalysis. When the reaction involves tunnelling of a particle (electron, proton or hydrogen atom), for example, collective motions of protein domains can have a gating effect which increases the probability of tunneling[7].

2.2 Approximations to the computational study of biological molecules

Computational methods can be used to obtain structural, dynamical and kinetic information about an enzyme performing its catalytic cycle. Studying enzyme catalysis computationally allows dissection of the chemistry of the reaction, but it also poses special difficulties. As computational power increases, the spectrum of reactions and the size of the systems that can be studied grows, but enzymes and proteins remain too big to be treated entirely with quantum mechanics methods even today. Different approaches have been developed to deal with the size problem.

One such approach is the use of semi-empirical methods, in which calculation is speeded up by replacing some integrals with parameters. These, although fast, are not reliable for the prediction of energies[8]. Efforts to improve semi-empirical methods are still on-going, with recent work from T. Clark *et al.*[9] and K. Merz *et al*[10]. among others to develop better parameters. Another approach to semi-empirical methods is the Empirical Valence Bond (EVB), which uses diabatic states to describe the system and provides better accuracy than the semi-empiricals described above[6,11,12].

Another approach is to study the reaction on a fragment of the active site that includes the amino acids involved in catalysis either *in vacuo* or accounting for the enzyme environment with a low dielectric constant. This method is known as a Quantum Cluster, and it is described in detail in Siegbahn *et al.*[13]. The main advantage of this method is the possibility to investigate the chemistry of the reaction at a high level of theory. The systems used typically include the reacting chemical groups and any part of the surrounding environment that has strong polar effects, and the reaction pathway is mapped in this reduced model. This will reproduce the actual enzymatic reaction as long as the fragment has been chosen

carefully (problems generated by too small a fragment can be seen in Wang *et al.*[14]). Although the whole enzyme environment is not considered, the smaller system size in QC models allows the use of higher levels of theory in these calculations, which is fundamental to obtain accurate energies.

Another popular approach is the ONIOM method[15], which consists of different zones ("layers") within a system and different levels of theory (such as QM, MM, semiempirical) are used on them. QM/MM methods[16,17] could be considered a particular subset of these. They deal with large systems by splitting them in two zones. The smallest of them consists of the reacting molecules/residues and is treated at QM level, while the rest of the enzyme is treated with an MM force field. However, because of the restrictions imposed by limited computational resources, the QM region is frequently described with semiempirical, DFT or HF methods, which are arguably not accurate enough[18]. A good description of this method can be found in a review by Groenhof *et al.*[16] and references therein.

The coupling between the MM and QM parts in these methods is an interesting subject that deserves special attention. In additive schemes, the total energy of the system is the sum of the MM component, the QM component and a QM/MM coupling term. In subtractive schemes, the total energy is calculated as the MM energy of the whole system plus the QM energy of the QM component, minus the MM energy of the QM part of the system. More detail can be obtained in Refs. [16] and [19].

2.3 Quantum mechanics

The aim of the following sections is to summarise the fundamental concepts used in this thesis and to provide a background to the methodology used. A more detailed presentation can be found in

textbooks such as Atkins' Physical Chemistry[20] and in the articles cited in the text.

The cornerstone of quantum chemistry is the Schrödinger equation[21] which states that the energy of a system can be obtained by applying the Hamiltonian operator to the system's wavefunction:

$$H\Psi = E\Psi \tag{2.1}$$

The Hamiltonian is made of operators for the potential and kinetic energies. Solving the Schrödinger equation for a given system allows one to obtain the energy of that system. However, because there is no exact solution for any system bigger than the hydrogen atom, approximations must be made when computing the properties of larger molecules and especially biomolecules. Different sorts of approximations can be made, which give rise to a variety of methods. In the Self-Consistent Field (SCF) Hartree-Fock (HF) method[22,23], the orbitals and energies are obtained by making an initial guess of the orbital wavefunctions and iteratively improving the guess until self-consistency is reached. The HF method was one of the milestones in the history of computational chemistry but is little used these days because it does not include electron correlation, which is extremely important for understanding chemical reactions. Newer methods account for this in different ways. Examples of such are Density Functional Theory (DFT) and Møller-Plesset (MP) methods, which were used in this work and are discussed in the following sections.

The electronic wavefunctions in molecular orbital methods are represented making use of what is called a *basis set*. These are a group of mathematical functions that represent the electronic orbitals. Slater Type Orbitals are approximated by Gaussian basis functions because the molecular integrals are difficult to solve. For that reason, the most

commonly used basis sets today model orbitals using a series of Gaussian functions (primitives), which are simple to use and computationally affordable. These basis sets follow the form 6-31G, in this case a linear combination of 6 primitive Gaussian functions represents the core electrons and two basis functions that represent the valence electrons, one made of three primitive functions and the other one of only one primitive (more primitive functions are used for the core electrons because the cusp of the function is hard to reproduce). This is a split-valence double-zeta basis set (triple-zeta basis sets will have three numbers after the hyphen). Diffuse functions can also be added; they are especially important when the system has a net charge, and are represented by a "+". Polarization functions of the *d* type are normally added to *p* elements, and *p* polarization functions are added to hydrogen atoms when these are transferred in the reaction or form hydrogen bonds. The Dunning (or correlation-consistent, cc) basis sets[24] are designed to account for the correlation energy. They include polarization and can be "augmented" with diffuse functions. For instance, the aug-cc-pvtz basis set indicates an augmented triple zeta basis set. As is obvious, they must be used with a correlation method, and they are usually paired with MP2 (this method will be described next).

2.3.1 Perturbation Theory

In Rayleigh-Schrödinger perturbation theory, the Hamiltonian operator **H** is written in terms of an operator $\mathbf{H}^{(0)}$ for which eigenfunctions can be found, and a perturbing operator **V**:

$$\mathbf{H} = \mathbf{H}^{(0)} + \lambda\mathbf{V} \tag{2.2}$$

where λ is the perturbation parameter that maps $\mathbf{H}^{(0)}$ onto $\mathbf{H}$. The perturbed wavefunction and its eigenvalue are written as a Taylor expansion (the subscript 0 indicates that this is the ground-state eigenfunction of $\mathbf{H}^{(0)}$):

$$\Psi = \Psi_0 + \lambda \frac{\partial \Psi}{\partial \lambda} \Big|_{\lambda=0} + \frac{1}{2!} \lambda^2 \frac{\partial^2 \Psi}{\partial \lambda^2} \Big|_{\lambda=0} + \frac{1}{3!} \lambda^3 \frac{\partial^3 \Psi}{\partial \lambda^3} \Big|_{\lambda=0} + \ldots \quad (2.2a)$$

and the energy $\mathbf{E}$ is written as

$$E = E_0 + \lambda \frac{\partial E}{\partial \lambda} \Big|_{\lambda=0} + \frac{1}{2!} \lambda^2 \frac{\partial^2 E}{\partial \lambda^2} \Big|_{\lambda=0} + \frac{1}{3!} \lambda^3 \frac{\partial^3 E}{\partial \lambda^3} \Big|_{\lambda=0} + \ldots \quad (2.2b)$$

these are usually written as a zeroth-order term and a series of *n*th-order corrections:

$$\Psi = \Psi_0 + \lambda \Psi^{(1)} + \lambda^2 \Psi^{(2)} + \lambda^3 \Psi^{(3)} + \ldots \quad (2.2c)$$

and

$$E = E_0 + \lambda E^{(1)} + \lambda^2 E^{(2)} + \lambda^3 E^{(3)} + \ldots \quad (2.2d)$$

The Møller-Plesset methods are denominated MP*n*, where *n* is the order at which the Taylor expansion of the ground state eigenfunctions and eigenvalues is truncated. MBPT (Many Body Perturbation Theory) is another acronym that refers to these methods. The "perturbation" in this case is the electron-electron repulsion energy, and it is the reason why lower-order truncations will not describe the correlation energy properly. The computational cost of MP2 (Ref. [25]) scales as N^5 (where N is the number of basis functions). MP3 provides little improvement over MP2; MP4 – which accounts for >95% of the correlation energy if an adequate basis set is used – scales as N^7. For this reason, a usual approximation is MP4SDQ, which omits the triple excitations calculation and can be

accurate as long as the molecule is a closed-shell singlet with a large separation between frontier orbitals[26]. A modification known as RI (Resolution of Identity), also gives good results and will be discussed in detail later in this chapter.

2.3.2 Density Functional Theory

DFT was formulated in the 1960s based on the Hohenberg-Kohn theorem[27,28], which states that the energy of a system is a functional of the electron density, and therefore the electron density is the only requirement to calculate the energy of the system. This reduces the dimensionality of the problem (the dimension of the wavefunction is 3N while the density is three-dimensional), but the exact functional is not known, for which reason approximations are derived by assuming that the electronic part of the system can be seen locally as a uniform electron gas, from which all properties can be derived.

This initial approach, called Local Density Approximation[29] (LDA) did not consider different α and β spin densities (open shell systems), and was later replaced by the Local Spin Density Approximation[30] (LSDA). In the Generalised Gradient Approximation (GGA) methods[31] the exchange and correlation energy depends on the electron density and its first derivative. Second order derivatives of the electron density (represented by the Laplacian, $\nabla^2\rho$) have also been implemented in the so-called "meta-GGA methods", which also include the orbital kinetic energy density[32]. Hybrid functionals are built to improve the exchange term. They contain a mix of GGA and LSDA exchange and correlation terms plus exact HF exchange. One example of this is the very popular B3LYP functional[33,34] which combines the Becke88 exchange functional and the Lee-Yang-Parr correlation functional with 20% of exact HF exchange. Although DFT was

initially found to perform well for a variety of circumstances[35], failure has since become evident when calculating energies (see for example Izgorodina *et al.*[36] and references therein) although they remain good methods for geometries. The main reason why DFT is inaccurate for energy calculations is the exchange and correlation (XC) functional, for which no exact form is known and therefore needs to be approximated. There are two main problems of the approximations. The first of these is the self-interaction (SI) error[37], which is a spurious interaction of each electron with itself that causes large errors in the computation of energies. The second problem is the improper description of the tail of the electronic wavefunction[38,39]. The first problem arises from the construction of the Coulombic energy functional and results in unreliable energies[32], the second affects also geometries, and will be discussed next.

Dispersion forces, also known as London forces, are attractive forces that originate when a fluctuation in the electronic density in one molecule creates an instantaneous dipole, which in turn induces a dipole moment in another molecule. Although weak, these interactions play an important role in the structure of bigger molecules and complexes, such as proteins and nucleic acids. They decay as R^{-6} and are therefore long-range forces (Grimme *et al.*[40] reviews dispersion-corrected DFT-D methods and the importance of dispersion in biomolecules). Recent work has pointed out the importance of dispersion energy when analyzing protein-substrate binding[41] and protein folding[42]. There is evidence[43] that the recently developed DFT-D methods provide satisfactory results for the interaction energies of biologically relevant groups.

Pauli's exclusion principle states that two fermions cannot occupy the same space at a time, which implies that the two-particle wavefunction must be antisymmetric to exchange of the particles. This is the origin of

the exchange energy. This is part of the HF method but it needs to be added in the case of DFT. HF, however, includes no electron correlation – which DFT does by construction. Because of the lack of dynamic effects, the electron repulsion is underestimated by HF. These effects give a reduced probability of finding a second electron close to where a first electron is, and a higher chance for this event far from the first electron in HF methods. This is known as the exchange-correlation hole, which is non-local – that is to say, it is not limited to one atom but rather delocalized over the entire system. The exchange and correlation (XC) functional in DFT is local, and so classical DFT methods fail to describe long-range interactions properly. This inability of DFT methods to reproduce dispersion effects was noted at the beginning of this century[39,44]. Since then, new functionals with this added capability have been created[45-48] which show improved performance[49]. In these methods, the total energy $E_{DFT\text{-}D}$ is the DFT (Kohn-Sham, KS) energy plus a correction for the dispersion energy E_{disp}[47].

$$E_{DFT\text{-}D} = E_{DFT(KS)} + E_{disp} \quad (2.3)$$

where E_{disp} is calculated as

$$E_{disp} = -s_6 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{C_6^{ij}}{R_{ij}^6} f_{damp}(R_{ij}) \quad (2.4)$$

The dispersion correction is scaled by the scaling factor s_6, which is empirically determined for each density functional. N is the number of atoms, C_6^{ij} is the dispersion coefficient for a given atom pair (ij) and R_{ij} is the distance between atoms i and j. The analytical form of the dispersion energy was shown by Alonso *et al*[50] and to have asymptotic behaviour,

which will produce unphysical effects at short distances. This was addressed long time ago by Ahlrichs *et al.*[51] who added dispersion corrections to a semi-empirical method, and who proposed the use of a damping function to eliminate the spurious effects of adding dispersion energies at short distances[51]. The damping function f_{damp} is as follows:

$$f_{damp}(R) = \frac{1}{1+e^{-\alpha(\frac{R}{R_0}-1)}} \qquad (2.5)$$

An analytical expression for the dispersion forces has been derived for use with DFT functionals[50] although the empirical correction just discussed is the one of choice for dispersion-corrected DFs so far – likely due to its lower cost as indicated in Chai *et al.*[46].

In this work, the following functionals are used. ωB97XD[46] is a hybrid functional based on the ωB97X long-range corrected functional[52] (that was developed from B97[53] with long-range corrections) with an unscaled dispersion correction, which is equivalent to setting the factor s_6 to 1. This functional is claimed to be free of long-range self-interaction, although there is some degree of self-interaction at short range. It is important to note that the long-range correlation is only based on *empirical* dispersion corrections and therefore the dispersion effects do not enter the KS orbitals. The M06 suite of methods[54] are hybrid meta-GGAs with exchange and correlation functionals based on the M05 suite[55]. All these include a percentage of HF exchange, which is around 25-27% for M05 and M06 and double that for the -2X version of the functional. DFT-D is still a work in progress, with more accurate and widely applicable functionals expected in the future.

2.3.3 Accurate calculation of energies

As the size of the system increases, the most exact methods for energy calculation become increasingly unaffordable. For systems of the size of those used for modelling biomolecules, these types of calculations are currently not feasible and therefore a compromise solution must be found.

2.3.3.1 Gaussian-*n* methods

The Gaussian-*n* series methods (G*n*)[56-59], use B3LYP or MP2 (depending on the series) geometries and B3LYP or HF harmonic frequencies and thermal correction. Single point calculations are performed at the QCISD(T) or CCSD(T) level with the 6-31G(d) basis set, and the large basis set energy is approximated by other methods using large basis sets that contain diffuse and higher polarization functions.

The Gaussian-*n* strategy is summarized below for G1, and further improvements on it are detailed afterwards.

- Geometry optimization at HF/6-31G(d). Harmonic frequencies and Zero Point Energies are obtained from this calculation.
- Further geometry optimization at MP2/6-31G(d).
- The MP2/6-31G(d) geometry obtained in the previous step is used for a series of single point calculations at higher levels of theory, the first of which is MP4(FC)/6-311G(d,p).
- A correction for the addition of diffuse functions (+) is calculated as DE(+) = E(MP4(FC)/6-311G(d,p)) – E(MP4/6-311G(d,p).
- A correction for higher polarization functions, DE(2df) = E(MP4/6-311G(3df,p)) – E(MP4/6-311G(d,p).

- A correction for truncation of the Møller-Plesset expansion at fourth order, DE(QCI) = E(QCISD(T)/6-311G(d,p)) – E(MP4/6-311G(d,p)).
- A higher-level correction (HLC) for incompleteness of the basis set is made, which consists of empirical terms and depends on the basis sets used for the previous calculations.

G2 adds a MP2/6-311+G(3df,2p) single point energy calculation as a correction to G1. G3 uses a MP2 calculation with a basis set specifically designed, named G3large. It also includes a spin-orbit correction. G4 introduces a few novelties: B3LYP/6-31G(2df,p) geometries and ZPE, a HF/limit calculation and a new G3 basis set, G3LargeXP, which has more polarization functions than the previous G3large. G3 energies have a mean average deviation from experiment of about 1 kcal/mol[60]. G4 has additional parameters to the HLC term, which although developed with the motivation of describing radicals and lone pairs better[58], was recently found to introduce errors in the calculation of reaction energies[61].

2.3.3.2 Resolution of the Identity (RI) methods

The computation of four-index two-electron integrals is a demanding step in *ab initio* calculations, as these methods do not include parameters. The RI approximation introduces an identity transformation to write the product of basis functions in an auxiliary basis set[62], replacing the four-index two-electron integrals for three-index one- and two-electron integrals. This approximation allows for computational savings as it scales as N^2 while *ab initio* methods scale as N^4. RI-MP2[63] provides a speed-up of MP2 energy calculations by one order of magnitude by using an approximate resolution of identity, while retaining accuracy[64].

2.4 Solvent models

Different solvation models are available for QM calculations. One of the most popular is the Polarizable Continuum Model (PCM)[65], used in this thesis with the Integral Equation Formalism (IEF)[66]. In continuum solvent models a cavity consisting of overlapping spheres is created for the solute (at a certain energetic cost) and the molecule(s) are placed in it. The interactions between solute and solvent are computed and enter the Hamiltonian in the form of an interaction potential. Analytical gradients are available in PCM for HF and DFT[67].

Another important choice (especially in cluster models) is that of the dielectric constant to apply to the active site of the enzyme. The larger the size of the cluster, the lesser this choice will affect the results[68] (also see a review of this issue in Ref. 13). In this work, the enzyme active site was mimicked by placing the cluster in a solvent with a dielectric constant of 4. This value is standard in quantum cluster calculations of enzymes[13].

2.5 Molecular Mechanics

Molecular Mechanics (MM) methods are computationally fast and cheap because they rely on classical mechanics (Newton's equations of motion) to describe molecular structures and dynamics. This section will describe the basis of the technique and the applications of it used in this thesis. More details on the topics discussed below can be found in textbooks[20] and research articles (see, for example, Adcock *et al.*[69] and references therein).

2.5.1 Force fields

To describe a molecule in MM a set of parameters known as a force field is used. Atoms are assigned a "type" depending on the chemical element and bonds (e.g. an sp3 carbon has a different atom type than an sp2 carbon, atoms in aliphatic chains are different from the same atom bound to an electron-withdrawing group, etc.). Each atom type has a specified mass, charge and van der Waals radius; and the bonds, angles and dihedrals formed by different atom types are given an equilibrium value and a stretch, bend or torsion constant. This has led the model to be described as "balls and springs". All these values come from experiment and/or high-level QM calculations. Different sets of parameters can be obtained from different sources (and for different types of molecules) that will be more or less expensive to acquire.

The relationship between structure and energy of the system is given by the potential energy function *U(R)* (Eq. 2.6) which describes the forces in the system according to the position of the atoms (see Mackerell *et al.*[70] and references therein). In this equation, b is the bond length between two atoms any given time, b_0 is the equilibrium bond length and K_b is a force constant; angles are described in a similar manner (θ is the angle value, θ_0 is the equilibrium value and K_θ is the force constant); dihedrals are represented by a sinusoid function in which the dihedral force constant is K_χ, $n\chi$ is the value of the dihedral and δ the phase angle; improper dihedrals are treated in a similar manner to angles, with φ being the improper angle and φ_0 the equilibrium improper angle, and K_{imp} the improper force constant. The last term of Eq. 2.6 accounts for non-bonded interactions between atoms and it is known as a Lennard-Jones potential[71]. In this term, q_i and q_j are the charges of atoms i and j, respectively, ε_{ij} is the depth of the well of the potential, R is the minimum interaction radius, ε is

the dielectric constant and r_{ij} is the distance between the two interacting atoms.

$$U(R) = \sum_{bond} K_b(b - b_0)^2 + \sum_{angle} K_\theta(\theta - \theta_0)^2 + \sum_{dihedral} K_x(1 + \cos(n\chi - \delta)) + \sum_{improper} K_{imp}(\varphi - \varphi_0)^2 + \sum_{nonbond} \left(\varepsilon_{ij} \left[\left(\frac{R_{ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{ij}}{r_{ij}} \right)^{6} \right] \right) + \frac{q_i q_j}{\varepsilon r_{ij}}$$

(Eq. 2.6)

MM is not suitable for the study of chemical reactions because electrons are not explicitly described (and therefore forming and breaking bonds is not possible for these types of calculations), but due to high parameterisation they can describe protein structures and molecular dynamics very well. In commonly used force fields, amino acids are parameterised in a standard way that can be applied to most proteins without deriving new parameters. A review of the most popular force fields is given in Mackerell *et al.*[70].

2.5.2 Molecular Dynamics Simulations

Molecules are not static. Atoms in them vibrate and rotate with respect to each other depending on the degrees of freedom in a particular molecule. In proteins, whole domains move with respect to each other (such movements are key to the native role of a protein) and amino acid residue side chains also change positions as their environment fluctuates. Molecular dynamics (MD) provides a means to analyse the

conformational energy landscape (the different geometrical configurations and their respective energies) of a protein[72,73]. While MD simulations are widely used to study the structural properties of biological macromolecules, sampling the conformational space is an added difficulty and several techniques have been developed to improve the results. Given the size of proteins, MD simulations with *ab initio* QM methods can only be run for a very short time, for this reason MM force fields are used. Appropriate sampling of the conformational space is an important part of MD simulations and different approaches have been developed (see Liwo *et al*.[74], and references therein). In this thesis, adequate sampling was ensured by running multiple trajectories for each system (unless otherwise specified). This has been shown to provide significantly better sampling than running one long trajectory[75,76].

2.5.3 Docking

Docking techniques are another application of MM methodologies. They allow *in silico* prediction of whether a substrate will bind a certain site in a protein in a productive configuration, for this reason these methods are very popular in drug design. Multiple conformations of the ligand (and, if required, of the protein too) are tested and a scoring function determines which are acceptable bindings and which are not. More detail of the aspects mentioned below can be found in books[77,78]. Good reviews of the techniques in their current state and are given also available[79-84].

There are different methods that can be used to perform a conformational search of the substrate. Systematic methods, as their name implies, work by testing all possible conformations, which results in a problem known as combinatorial explosion (in which the calculation becomes unaffordable because of the large number of possible conformations available).

Stochastic methods use as a starting point a random configuration of the ligand, which is accepted or rejected with a certain probability (popular examples of these are Monte Carlo and genetic algorithms). Molecular Dynamics in which different parts of the system are simulated at different temperatures are also used for conformational searches in docking[79].

The scoring function ranks the docked structures, and it is a key aspect of the method as the accuracy of the results depends on it. It can be based on Molecular Mechanics force fields (described in the previous section), or it can be fitted to reproduce experimental binding energies or geometries. The first approach will reproduce the limitations of the force field chosen, while the other two will depend on the data set used. For these reasons, the scoring functions are the main weakness of docking techniques and where improvement efforts are directed[80].

2.6 Thermal factors

B-factors (or thermal factors) quantify the thermal motion of atoms based on the attenuation of x-ray scattering. They have been used to study protein flexibility[85] and thermostability[86,87], among other properties. B-factors can also be extracted from MD simulations making use of the relationship between these, RMSD (Root Mean Square Deviation) from the initial coordinates and RMSF (Root Mean Square Fluctuations)[88]. The following equations summarize the basis of this relationship.

$$\mathrm{RMSF}^2 = \frac{3B}{8\pi^2} \tag{2.7}$$

$$\langle RMSD^2 \rangle^{\frac{1}{2}} = \sqrt{\frac{2N_S}{N_S - 1}} \langle RMSF^2 \rangle^{\frac{1}{2}} \tag{2.8}$$

(where B is the B-factor and N_s is the number of structures in the ensemble).

For $N_S \gg 1$,

$$\langle RMSD^2 \rangle^{\frac{1}{2}} \approx \sqrt{\frac{2}{N_a} \sum_{k=1}^{N_a} \frac{3B_k}{8\pi^2}} \quad (2.9)$$

(where N_a is the number of atoms in the structure).

2.7 Computational resources

2.7.1 Hardware

This research was undertaken with the assistance of resources provided at the National Computing Infrastructure (NCI) National Facility (NF) systems at the Australian National University through the National Computational Merit Allocation Scheme supported by the Australian Government. Part of the Molecular Dynamics simulations were done on a Quad-Core AMD Opteron(TM) Processor 2356 with 8GB of RAM with time generously provided by Professor Thomas Huber.

2.7.2 Software

QM geometry optimisations and MP2 single point energy calculations were carried out with Gaussian09[89]. Scaling factors for wB97XD and M062X were taken from Alecu *et al.*[90]. RI-MP2 single point energies were calculated with QChem4.1[91,92]. CCSD(T) energy calculations were done with Molpro2012.1[93]. Preparation of input geometries was done using GaussView5[94]. Visualisation of results was done with GaussView or Molden5.0[95]. Docking was done using AutoDock Vina[96]. Visualization of

results was done with Pymol[97]. Molecular dynamics simulations were performed using AMBER12[98]; generation of parameters and trajectory analysis were carried out using the AmberTools included in the package.

2.8 References

(1) Van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; Van Der Vegt, N. F. A.; Yu, H. B. *Angewandte Chemie - International Edition* **2006**, *45*, 4064.

(2) Cheng, G. J.; Zhang, X.; Chung, L. W.; Xu, L.; Wu, Y. D. *Journal of the American Chemical Society* **2015**, *137*, 1706.

(3) Shakhnovich, E. *Chemical Reviews* **2006**, *106*, 1559.

(4) Di Ventura, B.; Lemerle, C.; Michalodimitrakis, K.; Serrano, L. *Nature* **2006**, *443*, 527.

(5) Bugrim, A.; Nikolskaya, T.; Nikolsky, Y. *Drug Discovery Today* **2004**, *9*, 127.

(6) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. *Chemical Reviews* **2006**, *106*, 3210.

(7) Masgrau, L.; Basran, J.; Hothi, P.; Sutcliffe, M. J.; Scrutton, N. S. *Archives of Biochemistry and Biophysics* **2004**, *428*, 41.

(8) Schenker, S.; Schneider, C.; Tsogoeva, S. B.; Clark, T. *Journal of Chemical Theory and Computation* **2011**, *7*, 3586.

(9) Winget, P.; Horn, A. H.; Selçuki, C.; Martin, B.; Clark, T. *Journal of molecular modeling (Online)* **2003**, *9*, 408.

(10) Dixon, S. L.; Merz Jr, K. M. *Journal of Chemical Physics* **1996**, *104*, 6643.

(11) Vardi-Kilshtain, A.; Roca, M.; Warshel, A. *Biotechnology Journal* **2009**, *4*, 495.

(12) Liu, H.; Warshel, A. *Biochemistry* **2007**, *46*, 6011.

(13) Siegbahn, P. E. M.; Himo, F. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 323.

(14) Wang, J.; Gu, J.; Leszczynski, J. *Journal of Physical Chemistry B* **2008**, *112*, 3485.

(15) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *Journal of Molecular Structure: THEOCHEM* **1999**, *461-462*, 1.

(16) Groenhof, G. *Methods in molecular biology (Clifton, N.J.)* **2013**, *924*, 43.

(17) Guallar, V.; Wallrapp, F. H. *Biophysical Chemistry* **2010**, *149*, 1.

(18) Warshel, A. 2003; Vol. 32, p 425.

(19) Warshel, A. *Angewandte Chemie - International Edition* **2014**, *53*, 10020.

(20) Atkins, P. W.; de Paula, J. *Atkins' Physical Chemistry*; Oxford University Press: Oxford ; New York 2014.

(21) Schrödinger, E. *Physical Review* **1926**, *28*, 1049.

(22) Hartree, D. R. *Mathematical Proceedings of the Cambridge Philosophical Society* **1928**, *24*, 89.

(23) Fock, V. *Zeitschrift für Physik* **1930**, *61*, 126.

(24) Dunning Jr, T. H. *The Journal of Chemical Physics* **1989**, *90*, 1007.

(25) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. *Chemical Physics Letters* **1990**, *166*, 275.

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

(27) Hohenberg, P.; Kohn, W. *Physical Review* **1964**, *136*, B864.

(28) Kohn, W.; Sham, L. J. *Physical Review* **1965**, *140*, A1133.

(29) Langreth, D. C.; Mehl, M. J. *Physical Review Letters* **1981**, *47*, 446.

(30) Von Barth, U.; Hedin, L. *Journal of Physics C: Solid State Physics* **1972**, *5*, 1629.

(31) Langreth, D. C.; Mehl, M. J. *Physical Review B* **1983**, *28*, 1809.

(32) Jensen, F. *Introduction to Computational Chemistry*; 2nd ed.; John Wiley & Sons Ltd, 2006.

(33) Becke, A. D. *The Journal of Chemical Physics* **1993**, *98*, 1372.

(34) Lee, C.; Yang, W.; Parr, R. G. *Physical Review B* **1988**, *37*, 785.

(35) Kohn, W.; Becke, A. D.; Parr, R. G. *Journal of Physical Chemistry* **1996**, *100*, 12974.

(36) Izgorodina, E. I.; Brittain, D. R. B.; Hodgson, J. L.; Krenske, E. H.; Lin, C. Y.; Namazian, M.; Coote, M. L. *Journal of Physical Chemistry A* **2007**, *111*, 10754.

(37) Dutoi, A. D.; Head-Gordon, M. *Chemical Physics Letters* **2006**, *422*, 230.

(38) Pérez-Jordá, J. M.; San-Fabián, E.; Pérez-Jiménez, A. J. *Journal of Chemical Physics* **1999**, *110*, 1916.

(39) Zimmerli, U.; Parrinello, M.; Koumoutsakos, P. *Journal of Chemical Physics* **2004**, *120*, 2693.

(40) Grimme, S. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 211.

(41) Antony, J.; Grimme, S.; Liakos, D. G.; Neese, F. *Journal of Physical Chemistry A* **2011**, *115*, 11210.

(42) He, X.; Fusti-Molnar, L.; Cui, G.; Merz Jr, K. M. *Journal of Physical Chemistry B* **2009**, *113*, 5290.

(43) Morgado, C.; Vincent, M. A.; Hillier, I. H.; Shan, X. *Physical Chemistry Chemical Physics* **2007**, *9*, 448.

(44) Vondrášek, J.; Bendová, L.; Klusák, V.; Hobza, P. *Journal of the American Chemical Society* **2005**, *127*, 2615.

(45) Grimme, S. *Journal of Computational Chemistry* **2006**, *27*, 1787.

(46) Chai, J. D.; Head-Gordon, M. *Physical Chemistry Chemical Physics* **2008**, *10*, 6615.

(47) Grimme, S. *Journal of Computational Chemistry* **2004**, *25*, 1463.

(48) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *Journal of Chemical Physics* **2010**, *132*.

(49) Grimme, S.; Hujo, W.; Kirchner, B. *Physical Chemistry Chemical Physics* **2012**, *14*, 4875.

(50) Alonso, J. A.; Mañanes, A. *Theoretical Chemistry Accounts* **2007**, *117*, 467.

(51) Ahlrichs, R.; Penco, R.; Scoles, G. *Chemical Physics* **1977**, *19*, 119.

(52) Chai, J. D.; Head-Gordon, M. *Journal of Chemical Physics* **2008**, *128*.

(53) Becke, A. D. *Journal of Chemical Physics* **1997**, *107*, 8554.

(54) Zhao, Y.; Truhlar, D. G. *Theoretical Chemistry Accounts* **2008**, *120*, 215.

(55) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *Journal of Chemical Physics* **2005**, *123*.

(56) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *The Journal of Chemical Physics* **1991**, *94*, 7221.

(57) Curtiss, L. A.; Raghavach Ari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *Journal of Chemical Physics* **1998**, *109*, 7764.

(58) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *Journal of Chemical Physics* **2007**, *126*.

(59) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. *The Journal of Chemical Physics* **1989**, *90*, 5622.

(60) Curtiss, L. A.; Raghavachari, K. *Theoretical Chemistry Accounts* **2002**, *108*, 61.

(61) Lin, C. Y.; Hodgson, J. L.; Namazian, M.; Coote, M. L. *Journal of Physical Chemistry A* **2009**, *113*, 3690.

(62) Vahtras, O.; Almlöf, J.; Feyereisen, M. W. *Chemical Physics Letters* **1993**, *213*, 514.

(63) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chemical Physics Letters* **1993**, *208*, 359.

(64) Weigend, F.; Häser, M. *Theoretical Chemistry Accounts* **1997**, *97*, 331.

(65) Miertš, S.; Scrocco, E.; Tomasi, J. *Chemical Physics* **1981**, *55*, 117.

(66) Tomasi, J.; Mennucci, B.; Cancès, E. *Journal of Molecular Structure: THEOCHEM* **1999**, *464*, 211.

(67) Cossi, M.; Barone, V.; Mennucci, B.; Tomasi, J. *Chemical Physics Letters* **1998**, *286*, 253.

(68) Liao, R. Z.; Yu, J. G.; Himo, F. *Journal of Chemical Theory and Computation* **2011**, *7*, 1494.

(69) Adcock, S. A.; McCammon, J. A. *Chemical Reviews* **2006**, *106*, 1589.

(70) Mackerell Jr, A. D. *Journal of Computational Chemistry* **2004**, *25*, 1584.

(71) Jones, J. E. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **1924**, *106*, 441.

(72) Karplus, M.; Kuriyan, J. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 6679.

(73) Kamerlin, S. C. L.; Warshel, A. *Proteins: Structure, Function and Bioinformatics* **2010**, *78*, 1339.

(74) Liwo, A.; Czaplewski, C.; Ołdziej, S.; Scheraga, H. A. *Current Opinion in Structural Biology* **2008**, *18*, 134.

(75) Caves, L. S. D.; Evanseck, J. D.; Karplus, M. *Protein Science* **1998**, *7*, 649.

(76) Ramanathan, A.; Agarwal, P. K. *Journal of Physical Chemistry B* **2009**, *113*, 16669.

(77) Webster, D. M. *Protein structure prediction : methods and protocols*; Humana Press: Totowa, N.J., 2000; Vol. 143.

(78) Holtje, H.-D.; Sippl, W.; Rognan, D.; Folkers, G. *Molecular modeling : basic principles and applications*; Wiley-VCH: Weinheim, 2008.

(79) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nature Reviews Drug Discovery* **2004**, *3*, 935.

(80) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *Proteins: Structure, Function and Genetics* **2006**, *65*, 15.

(81) Guedes, I. A.; de Magalhães, C. S.; Dardenne, L. E. *Biophysical Reviews* **2014**, *6*, 75.

(82) Ewing, T. J. A.; Kuntz, I. D. *Journal of Computational Chemistry* **1997**, *18*, 1175.

(83) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins: Structure, Function and Genetics* **2002**, *47*, 409.

(84) Michel, J.; Essex, J. W. *Journal of Computer-Aided Molecular Design* **2010**, *24*, 639.

(85) Karplus, P. A.; Schulz, G. E. *Naturwissenschaften* **1985**, *72*, 212.

(86) Vihinen, M. *Protein Engineering* **1987**, *1*, 477.

(87) Parthasarathy, S.; Murthy, M. R. N. *Protein Engineering* **2000**, *13*, 9.

(88) Kuzmanic, A.; Zagrovic, B. *Biophysical Journal* **2010**, *98*, 861.

(89) Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; Gaussian, I., Ed. Wallingford CT, 2009.

(90) Alecu, I. M.; Zheng, J.; Zhao, Y.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2010**, *6*, 2872.

(91) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; DiStasio Jr, R. A.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Yeh Lin, C.; Van Voorhis, T.; Hung Chien, S.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C. P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Min Rhee, Y.; Ritchie, J.; Rosta, E.; David Sherrill, C.; Simmonett, A. C.; Subotnik, J. E.; Lee Woodcock Iii, H.; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer Iii, H. F.; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Physical Chemistry Chemical Physics* **2006**, *8*, 3172.

(92) Krylov, A. I.; Gill, P. M. W. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2013**, *3*, 317.

(93) Werner, H. J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 242.

(94) Dennington, R.; Keith, T.; Millam, J. 2009.

(95) Schaftenaar, G.; Noordik, J. H. *Journal of Computer-Aided Molecular Design* **2000**, *14*, 123.

(96) O. Trott, A. J. O. *Journal of Computational Chemistry* **2010**, *31*, 455.

(97) Schröedinger, L. 2010.

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

Chapter 3

Quantum cluster models of phosphorylation and dephosphorylation reactions

3.1 Introduction

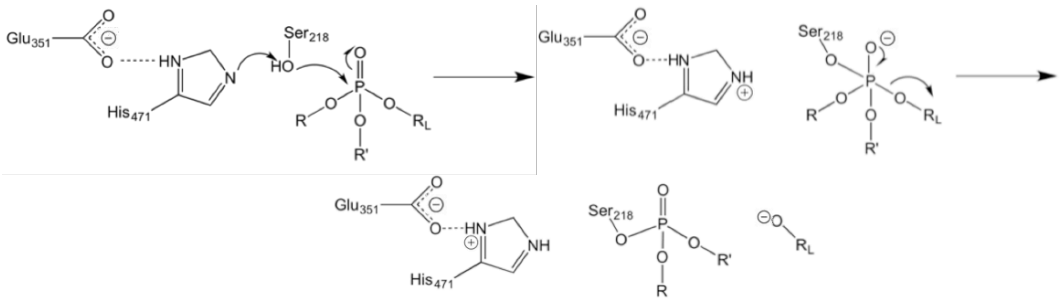
The use of organophosphorus pesticides has led to the development of resistance to these compounds in populations of the blowfly *Lucilia cuprina* (see Levot *et al.*[1] and references therein). The molecular basis of this resistance is a mutation in the gene encoding α-esterase 7, also referred to as E3, that converts an active site glycine residue (Gly137) to an aspartic acid[2]. How the addition of this aspartic acid changes the catalytic mechanism of the enzyme to allow it to catalyse the hydrolysis of organophosphates had not been examined in detail until now. In this chapter, high level quantum chemical calculations, using the quantum cluster (QC) method (for a description and review, see Mulholland et al.[3] and Lovell et al[4]), have been carried out to analyse the role of the aspartate in the catalytic mechanism. The data indicate that the aspartate increases the rate of dephosphorylation by functioning as a general base that activates a water molecule (by proton abstraction) to attack the phosphoserine adduct. This is similar to one of the mechanisms proposed for the dephosphylation of a butyrylcholinesterase mutant, in which a histidine introduced at position 117 acts as a general base[5] (see Chapter 1 for more detail).

Benchmarking studies enable evaluation of a range of potential computational methods for the study of this enzyme. QC calculations

were also performed to analyse the energy barriers to phosphorylation and dephosphorylation of E3 by several substrates, with the aim of evaluating the potential use of E3 to degrade chemical warfare agents. Unexpectedly, the data obtained during this work highlighted a number of serious obstacles to the accurate use of QC methods.

3.2 Summary of the current understanding of the catalytic mechanism of the E3 carboxylesterase

It has been suggested that E3 and acetycholinesterase (AChE) are structurally related carboxylesterases[6]. The mechanism of action of AChE with its physiological substrate, acetylcholine, consists of two steps: acylation and deacylation (described in Chapter 1). The residue responsible for attack of the nucleophilic substrate is the serine of the catalytic triad (Ser218). This catalytic serine is also capable of nucleophilic attack at the electrophilic phosphorus of organophosphate compounds[7,8]. The mechanism of phosphorylation of serine hydrolases such as AChE and E3 by organophosphate substrates has been described previously[9-13] and is presented in Scheme 3.1. The binding of acetylcholine and other molecules to the AChE active site sub-sites is presented in detail in Kua *et al.*[14].



Scheme 3.1 *The two steps of the phosphorylation reaction.* First the side chain oxygen of Ser218 attacks the phosphorus atom of an organophosphate concertedly with a proton transfer to His471. A pentacoordinated intermediate is formed. In the second step the leaving group of the substrate is cleaved and the enzyme is permanently phosphorylated.

In contrast to the initial phosphorylation reaction, which is analogous to carboxylation, dephosphorylation is extremely slow, and can result in an aging reaction (discussed later), leading to irreversible inhibition[15]. Given the physiological importance of AChE to humans and insects[16], this inhibition is the primary basis for the acute toxicity of organophosphates[17]. According to both experimental[18] and computational[9] studies, the (S) isomers of some chiral organophosphates are the most toxic for AChE. The reason why dephosphorylation cannot proceed via the same route as deacylation – that is, with the histidine residue acting as a base to enhance the nucleophilicity of a water molecule – was first discussed decades ago for another serine hydrolase, chymotrypsin, and is due to the steric hindrance that the tetrahedral geometry of the intermediate imposes on the incoming water molecule that is activated by the histidine[19] (also see Jarv *et al*.[20] and references therein). Phosphorylated or phosphonylated AChE can be reactivated by nucleophilic compounds such as oximes if it has not yet undergone secondary irreversible aging reactions[21].

It is important at this stage to point out the difference between covalent inhibition and aging, which was first introduced in Chapter 1. Covalent inhibition refers to the reaction of the enzyme with an inhibitor to form a covalent bond, which prevents the enzyme from catalysing the reaction with its natural substrate. However, slow dephosphorylation can occur, making inhibition at least partially reversible. Aging is a slower process than the initial inhibition, it was first observed decades ago and renders the enzyme irreversibly inactivated. In 1959, Jansz *et al*.[22] observed that DFP-bound Butyrylcholinesterase (BuChE) would age spontaneously by loss of an alkyl group, which was further confirmed by Berends *et al*[23]. While inhibition of AChE does not alter its conformational stability, aging does[24]. It has been observed that the aged enzyme is stabilized by

electrostatic interactions between a phosphonyl oxygen of the inhibitor and the H-Nε of the catalytic triad histidine residue, as well as with the oxyanion hole[25,26]. Early evidence for the participation of the catalytic histidine in the aging process can be found in Keijer *et al*.[27]. A study[28] carried out on three different serine-proteases (subtilisin, trypsin and chymotrypsin) found that the pK_a of the histidine member of the catalytic triad increases from about 7.4 in the inhibited enzyme – comparable to that in the native enzyme – to between 9.7 and 11.4 in the aged enzymes, which points to interactions between the anionic aged residue and the catalytic histidine. Lending support to this idea, the same work determined that aged enzymes, unlike inhibited (but not aged) enzymes, could not be reactivated by oximes. The aging of phosphorylated AChE is still being studied and recently a migration mechanism for the dealkylation was proposed[29]. Computational studies of the dephosphorylation of AChE with water[30] and with the aid of oxymes[31] have recently been carried out, as well as of the hydrolysis of paraoxon by bacterial PTE[32,33]. Studies on engineered butyrylcholinesterase enzymes[34,35] show that the Gly117His mutant can hydrolyse organophosphate compounds and retain its original esterase activity, that is, not becoming irreversibly phosphorylated in the process. The exact role that this histidine plays is still being investigated (see discussion in Chapter 1).

3.3 Computational Methods

Calculations of activation barriers of enzyme-catalysed reactions and of catalytic rates are typically carried out using low levels of theory (such as semi-empirical methods and the very popular but inaccurate DFT functional B3LYP[36]) to compensate for the computational burden of studying such large systems. Here, a different approach was used in

which higher and more accurate levels of theory were selected from benchmarking studies. The resolution of identity approximation (RI) permitted the use of a high-level *ab initio* method (RI-MP2/aug-cc-pvtz//M062X/6-31+G(d,p)) on a relatively large model of the active site (between 70 and 95 atoms depending on the system). For the benchmark, Gibbs free energies were calculated using ZPVE, entropy and thermal corrections calculated at the same level as the corresponding geometry optimisation. Benchmark geometries were optimised in gas phase. All other geometries were optimised in a low dielectric constant (ε=4) to better mimic the active site of the enzyme, using the Polarizable Continuum Model formalism as implemented in Gaussian09[37]. Gibbs free energies in solution are generally obtained by making use of a thermocycle as described by Ho *et al*.[38]. In this work, many calculations had frozen atoms, for which reason it was not possible to obtain vibrational frequencies (and hence vibrational contributions to the enthalpy and entropy) and therefore the energies reported are electronic energies. This issue was found during the calculations, for this reason the benchmark presents Gibbs free energies instead. It was decided that these would be more valuable to the reader as well, as free energies are most commonly the aim of enzyme calculations. The solvent was not included in the calculation of these electronic energies – which were obtained as gas-phase RI-MP2/aug-cc-pvtz single point energies – because continuum solvents are parameterised based on DFT calculations and would therefore make the calculation unbalanced. QChem[39] was used for all RI-MP2 single point energy calculations. CCSD(T) calculations were performed with Molpro[40].

The model systems were built from PDB structures of wild type E3[6] and Gly137Asp E3 [Drs Peter Mabbit & Colin Jackson, unpublished data, Table 3.1] (PDB ID 4FNG and 5C8V, respectively). The substrates were placed

with Autodock Vina[41] on the PDB structure (see Chapter 4) and moved manually to a distance suitable for attack from Ser218 before beginning the transition state search. The active site residues of interest were selected and the rest of the atoms were deleted.

***Lucilia cuprina E3*-G137D (PDB 5C8V)**	
Data Processing	
Space group	C2221
Cell dimensions (Å) a,b,c	49.67, 102.58, 225.48
Resolution range (Å)	44.7- 2.01 (2.06–2.01)[1]
Total number of reflections	146001(9661)
Number of unique reflections	37150 (2490)
Multiplicity	3.9 (3.9)
Completeness (%)	95.4 (97.8)
Mean I/σ(I)	7.7(1.55)
Wilson B factor ($Å^2$)	24.6
[2] $CC_{1/2}$	0.992 (0.581)
R_{merge}	0.132 (0.906)
Refinement	
R_{work}/R_{free}	0.185/0.245 (0.302/0.333)
Total number of atoms	4795
Number of macromolecules	1
No. of waters	208
No of protein residues	566
RMSD for bonds (Å)	0.0149
RMSD for angles (deg)	1.666
Ramachandran favored (%)	96.6
Ramachandran outliers (%)	0
Clashscore	2.09
Average B factor ($Å^2$)	27

Table 3.1. Data processing and refinement data for the X-ray crystal data structure used in this work. [1] Values in parentheses are for the highest resolution shell. [2]Pearson's correlation coefficient calculated from two half-sets of the data[42,43].

Different levels of restraints on these systems were tested. When no restraints were imposed on the system, and when the Cα atoms were frozen to the positions observed in the X-ray crystal structures, the two transition states of the reaction were located and then the corresponding minima were found by following the negative eigenvector using the

Intrinsic Reaction Coordinate (IRC) formalism. For the set of systems in which all atoms (except for the substrate atoms) were subject to constraints, the minima as well as the transition states were obtained by starting from the geometry of the corresponding stationary point of the diethyl 7-hydroxycoumarinyl phosphate (dECP) system, replacing dECP by alternative substrates, and letting the substrate (only) move during the optimisation.

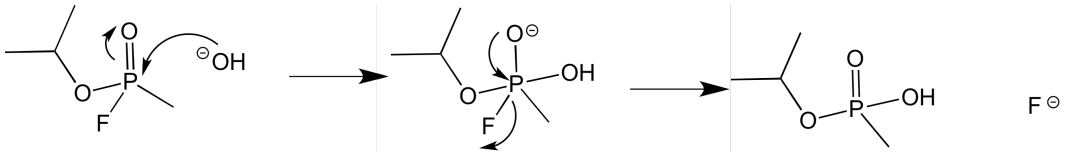
3.4 Benchmarking studies

As detailed in Chapter 2, highly accurate *ab initio* methods are too expensive to be used on entire protein systems of the size considered here, and highly parameterised methods do not always reproduce the energetics of the reaction correctly. Therefore, a benchmarking study was conducted with the purpose of determining the levels of theory that should be used for geometry optimisation and energy calculation in these reaction studies and to obtain the best compromise between accuracy and computational cost. All geometry optimizations and MP2 energy calculations were performed with the Gaussian09[37] package. RI-MP2 energies were calculated using QCHEM[39], and CCSD(T) energies were calculated with MOLPRO[40].

3.4.1 Geometry benchmark

The performance of different methods was assessed using the hydrolysis of isopropyl-methylphosphonofluoridate (sarin) by a hydroxide anion as a model reaction (Scheme 3.2). This compound was selected for being a small organophosphate, which makes calculations less expensive while being chemically similar to the compounds of interest. Geometry optimizations were performed with DFT (using B3LYP) and DFT-D (using

M062X and ωB97XD). The standards for comparison were MP2-optimized geometries. In all cases, the 6-31+G(d) basis set was used, given that the presence of an anion requires the use of diffuse functions, and that the presence of hydrogen bonds (highly polarised hydrogen atoms) makes polarisation functions necessary. After the first order saddle points converged, IRC calculations followed by geometry optimizations led to the minima connecting to the transition states. Single point energy calculations with G3(MP2)-CC were carried out for each of the geometries. G3(MP2)-CC was chosen as the benchmark standard for its high accuracy[44]. As mentioned earlier, Gibbs free energies were calculated using the G3(MP2)-CC electronic energy and the entropy and thermal corrections estimated from the vibrational frequencies calculated with the method used for optimization. The frequency scaling factors used were those published by Alecu *et al.*[45]. M062X performed most similarly to MP2 for geometry optimization followed by B3LYP, wB97XD was the most different from MP2. Table 3.2 and Figure 3.1 summarise the results.



Scheme 3.2. *Model reaction used in benchmark.* The hydrolysis of sarin by a hydroxide anion was the reaction used to compare different methods.

	ΔG1	ΔG2	R - I	I - P
MP2	17.3	19.9	46.8	5.1
M062X	16.5	19.4	44.2	6.5
B3LYP	19.2	23.1	50.3	0.9
ωB97XD	18.7	26.4	50.8	-1.1
[M062X]	[-0.8]	[-0.4]	[-2.6]	[1.4]
[B3LYP]	[1.8]	[3.2]	[3.5]	[-4.2]
[ωB97XD]	[1.3]	[6.5]	[4.0]	[-6.2]

Table 3.2. *G3(MP2)-CC free energy barriers of geometries optimized with different methods (kJ/mol).* ΔG1 and ΔG2 denote the activation barriers; R - I is the result of subtracting the intermediate species's energy from that of the reactant; I - P is the subtraction of the product's energy from that of the

intermediate. In brackets, the differences between the barrier calculated by a particular method and MP2.

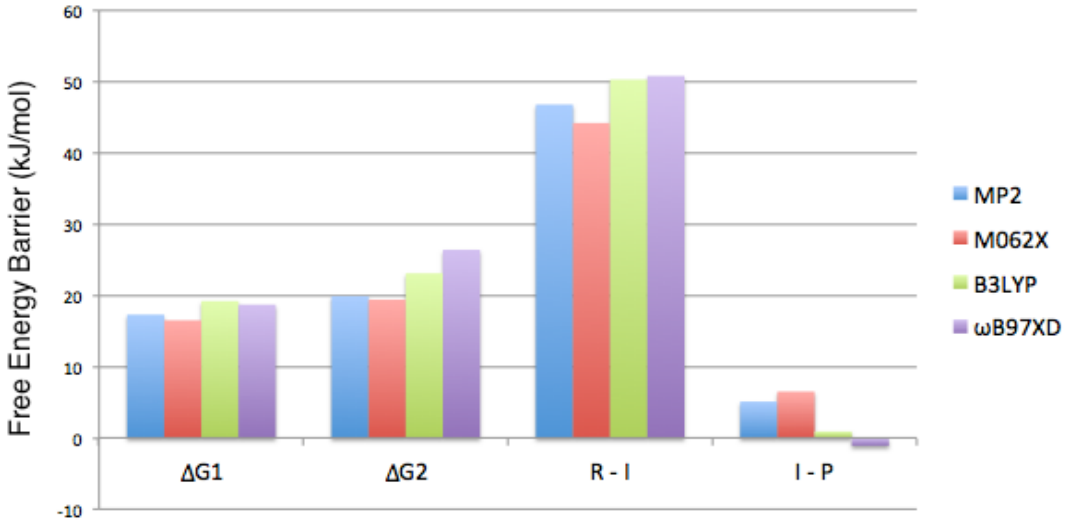


Figure 3.1. *G3(MP2)-CC free energy barriers of systems for which different methods were used for geometry optimisation.* Graphical representation of the data contained in Table 3.2. MP2 is the reference method for geometry optimisation. It is observed that M062X is the most similar to the reference, followed by B3LYP and (last) ωB97XD.

3.4.2 Energy Benchmark

The M062X-optimized geometries of the stationary points in the sarin reaction path were then used for a benchmark of single point energies using different methods and basis sets, to determine the most cost-effective way to obtain reaction energies. The results are collected in Table 3.3 and Figure 3.2. The closest agreement to G3(MP2)-CC was obtained with the MP2 method and aug-cc-pvtz basis set (MP2/aug-cc-pvqz results are within 2kJ/mol of these). The performance of RI-MP2 was virtually identical to that of MP2 in all geometries studied. Therefore RI-MP2/aug-cc-pvtz and RI-MP2/aug-cc-pvqz are the best methods for energies here studied, as compared to G3(MP2)-CC. The DFT methods were all tested using the 6-31+G(d) basis set. Relative to the MP2 benchmark level, wB97XD showed the best performance, followed by B3LYP, M11 and M062X in that order. M062X consistently performed at the bottom of the

group, scoring the largest maximum deviation (24.9 kJ/mol) and largest average deviation (15.6 kJ/mol) among all methods. The deviation from G3(MP2)-CC results for each barrier and method is presented in Table 3.4 and Figure 3.3.

	ΔG1	ΔG2	R - I	I - P
G3(MP2)-CC	16.5	19.4	44.2	6.5
MP2/aug-cc-pvdz	17.6	24.2	51.8	3.0
RI-MP2/aug-cc-pvdz	17.6	24.3	52.0	3.0
MP2/aug-cc-pvtz	19.5	23.4	41.7	6.8
RI-MP2/aug-cc-pvtz	19.4	23.4	41.8	6.8
MP2/aug-cc-pvqz	20.3	23.1	40.1	8.4
RI-MP2/aug-cc-pvqz	20.3	22.8	39.8	8.7
MP2/6-311+G(3df,2p)	8.8	26.4	63.1	6.1
RI-MP2/6-311+G(3df,2p)	8.7	26.5	63.2	6.4
MP2/6-311+G(3df,3pd)	9.7	26.1	60.7	5.0
RI-MP2/6-311+G(3df,3pd)	9.8	26.3	60.8	5.2
ωB97X/6-31+G(d)	17.2	18.2	47.6	15.4
B3LYP/6-31+G(d)	17.4	14.1	37.9	20.6
M11/6-31+G(d)	8.8	14.8	60.4	2.0
M062X/6-31+G(d)	9.7	30.9	69.1	-13.0

Table 3.3. *Free energy barriers calculated by different methods (in kJ/mol).* ΔG1 and ΔG2 refer to the activation barriers; R - I is the result of subtracting the intermediate species energy from that of the reactant; I - P is the subtraction of the product's energy from that of the intermediate. Geometries optimised with M062X/6-31G(d).

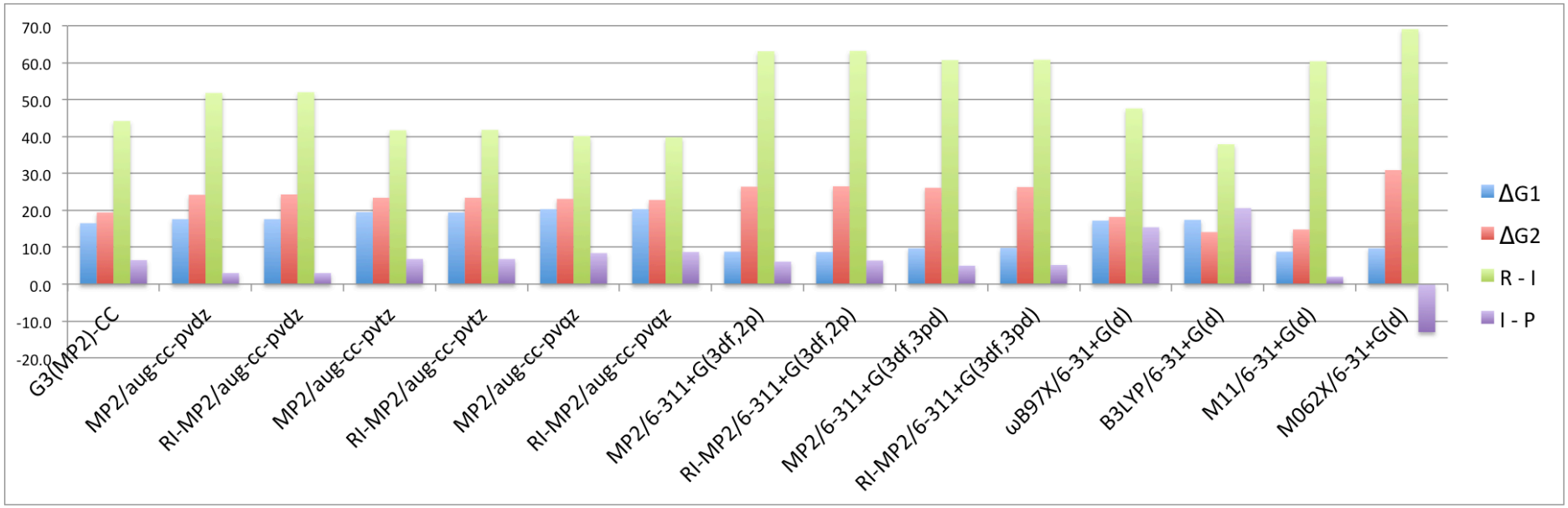


Figure 3.2. *Free energy barriers calculated with different methods.* Graphic representation of the data contained in Table 3.3. ΔG1 and ΔG2 refer to the activation barriers; R - I is the result of subtracting the intermediate species energy from that of the reactant; I - P is the subtraction of the product's energy from that of the intermediate.

	ΔG1	ΔG2	R - I	I - P	Average deviation	Maximum deviation
MP2/aug-cc-pvdz	1.1	4.8	7.6	-3.5	4.3	7.6
RI-MP2/aug-cc-pvdz	1.1	4.8	7.8	-3.5	4.3	7.8
MP2/aug-cc-pvtz	3.0	3.9	-2.5	0.3	2.4	3.9
RI-MP2/aug-cc-pvtz	3.0	3.9	-2.3	0.4	2.4	3.9
MP2/aug-cc-pvqz	3.8	3.7	-4.1	1.9	3.4	4.1
RI-MP2/aug-cc-pvqz	3.9	3.3	-4.4	2.2	3.4	4.4
MP2/6-311+G(3df,2p)	-7.7	7.0	18.9	-0.3	8.5	18.9
RI-MP2/6-311+G(3df,2p)	-7.8	7.0	19.0	-0.1	8.4	19.0
MP2/6-311+G(3df,3pd)	-6.8	6.7	16.5	-1.5	7.9	16.5
RI-MP2/6-311+G(3df,3pd)	-6.7	6.8	16.7	-1.3	7.9	16.7
ωB97X/6-31+G(d)	0.7	-1.2	3.5	8.9	3.6	8.9
B3LYP/6-31+G(d)	0.9	-5.3	-6.3	14.1	6.7	14.1
M11/6-31+G(d)	-7.7	-4.6	16.2	-4.5	8.3	16.2
M062X/6-31+G(d)	-6.8	11.4	24.9	-19.5	15.6	24.9

Table 3.4. *Deviation from G3(MP2)-CC barriers for each method (in kJ/mol).* The average deviation is calculated using absolute values. The maximum deviations are also expressed in absolute values.

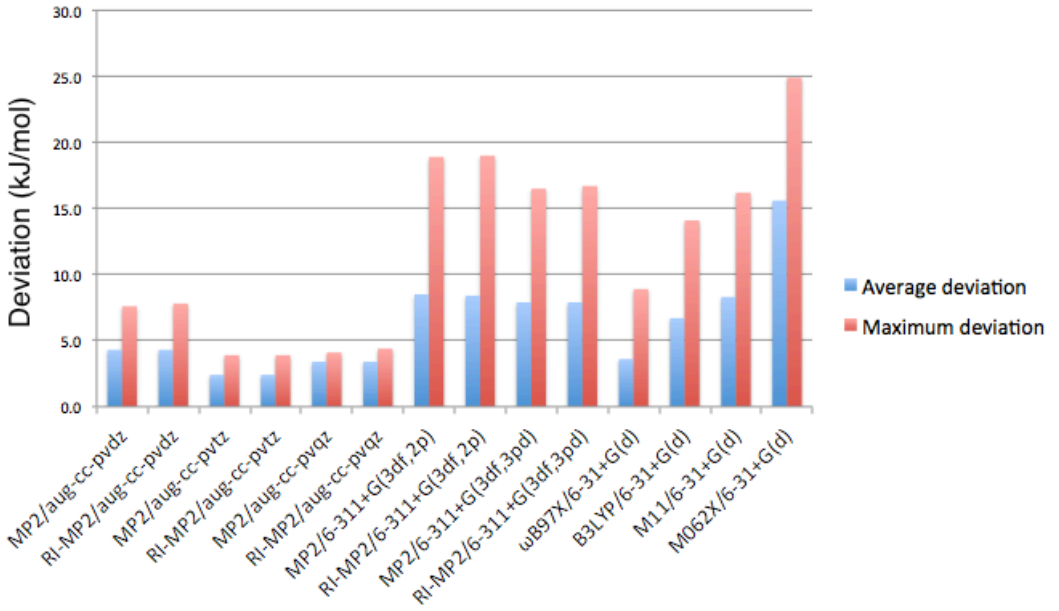


Figure 3.3. *Deviation from G3(MP2)-CC energies of the energies calculated with lower-level methods.* Graphic summary of the information contained in Table 3.4. The average deviation is calculated using absolute values. The maximum deviations are also expressed in absolute values.

Given that the costs of MP2 are high when used with relatively large systems such as biological molecules, the use of DFT geometries is the best option currently available. Careful consideration must always be taken when choosing the functional that will be used. The dispersion corrections included in the newer functionals add an important element that was so far missing in calculations involving biological molecules – namely accounting for the long-range interactions that dominate in molecules such as proteins. It is clear from the data presented here that the calculation of accurate energies requires at the very least MP2 (or RI-MP2) methods with a correlation-consistent basis set. DFT energies – even those calculated with the more modern DFT-D methods – are highly inaccurate[36], however, they are still routinely used in QM/MM studies[46-48]. Unfortunately, the computational cost of the methods selected for the benchmarking studies proved prohibitively expensive for QM/MM simulations when attempted, so a Quantum Cluster (QC) approach was

selected. In fact, many previously published studies using QM/MM on systems similar to E3 have used DFT energies to predict barriers[9,49-52]. These benchmarking studies suggest such methods are so inaccurate that serious doubt must be raised regarding their conclusions. This is discussed in more detail later in this chapter.

3.5 Quantum cluster studies of reaction mechanisms

The mechanism and reaction path for the phosphorylation and dephosphorylation of Ser218 in the E3 active site were analysed. The dephosphorylation of E3 mutant Gly137Asp was initially studied using an active site model that included the side chains of residues Asp137, Ser218 (phosphorylated with dECP, from the structure with PDB ID 4NFM[6]), Glu351 and His471, and a water molecule. In the models used in this section, the residues were represented by their functional groups anchored to a methyl group, which were frozen to the position of the corresponding carbon in the crystallographic structure. Electronic energies were calculated with a single point RI-MP2/aug-cc-pvtz calculation on M062X structures, as determined in the previous section. The phosphorylation reaction was also probed and compared to the existing model[9] with a similar model system that included Ser218, Glu351, His471 and dECP. The choice of which functional groups to include in the QC model and which to leave out, and whether and where to place restraints, can have a big impact on the accuracy of the model. To study the effect of system size and restraints on the calculated energies, a variety of combinations were tested. Very different results were obtained from these combinations, which are described in the corresponding section.

3.5.1 Mechanism of phosphorylation of wild-type E3

The initial phosphorylation of E3 corresponded with what has been previously postulated and presented in Scheme 3.1[9-13]. The system used to probe this reaction is shown in Figure 3.4. To prevent artificial movement of the residues, the carbon atoms where the residues were truncated were frozen to their respective X-ray structure positions while the substrate was allowed to move freely. In the first step, Ser218 attacks dECP and donates a proton to His471 in a concerted manner. In the second step, the leaving group of dECP is cleaved. The energy barriers for these steps were calculated to be ΔE1 = 29 kJ/mol and ΔE2 = 22 kJ/mol. These results are presented in Figure 3.5 and Table 3.5.

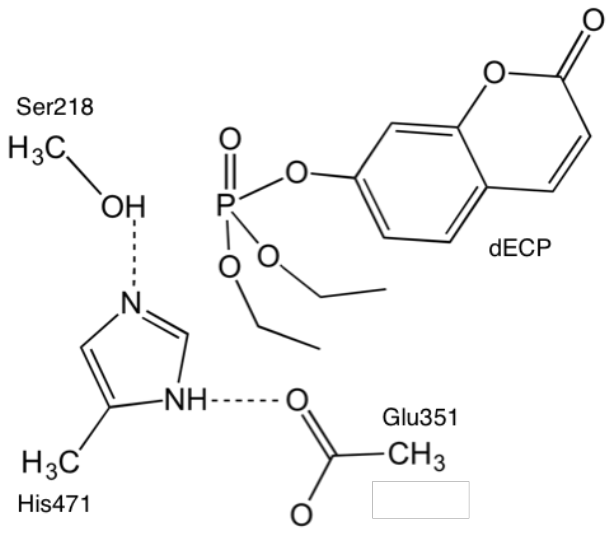


Figure 3.4. *Schematic representation of the system used to study the phosphorylation of E3 with dECP.* The system included the catalytic triad and substrate but omitted the oxyanion hole for simplicity.

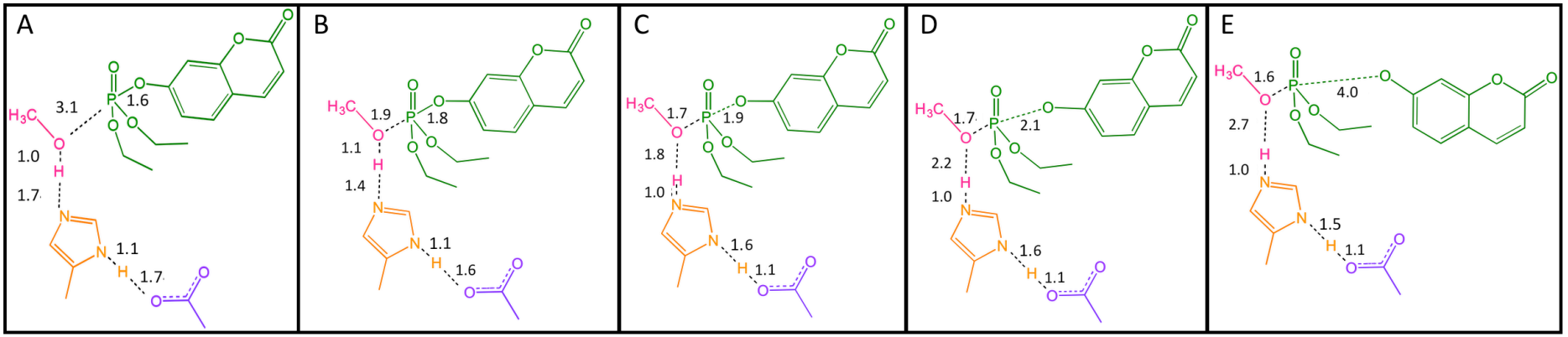


Figure 3.5. *Reaction coordinates of phosphorylation of E3 (dECP).* **A.** Reactant. **B.** TS1. **C.** Intermediate. **D.** TS2. **E.** Product. Green: dECP. Magenta: Ser218. Orange: His471. Purple: Glu351. Distances in Å.

	P - Oγ	Hγ - Oγ	Hγ – Nε	P - O_{dECP}	Hδ - OD	Hδ - Nδ
Reactant	3.1	1.0	1.7	1.6	1.7	1.1
TS1	1.9	1.1	1.4	1.8	1.6	1.1
Intermediate	1.7	1.8	1.0	1.9	1.1	1.6
TS2	1.7	2.2	1.0	2.1	1.1	1.6
Product	1.6	2.7	1.0	4.0	1.1	1.5

Table 3.5. *Principal geometry parameters of the phosphorylation reaction (dECP).* The data show a concerted attack on the phosphorus by Ser218 and proton transfer to His471 (TS1) and breakage of the phosphorus-oxygen (TS2). Distances in Å. O_{dECP} = oxygen of leaving group that participates in the bond that breaks (with P).

3.5.2 Mechanism of dephosphorylation with Asp137 acting as a general base

The mechanism of dephosphorylation in the Gly137Asp mutant E3 of the blowfly *L. cuprina* is discussed here in detail for the first time. As discussed in Section 3.2, phosphorylation inactivates the wild type enzyme because it lacks a general base in the right position to activate a water molecule for attack on the phosphorylated serine. Figure 3.6 shows a schematic representation of the system used for probing this reaction, and Scheme 3.3 shows the hypothesised reaction mechanism. The calculations detailed below followed the same protocol as the phosphorylation calculations in the previous section, with the additional functional group of Asp137 which was truncated to its side chain carbon atom and is also frozen to its X-ray position. These results show that the glycine to aspartate mutation (Gly137Asp) introduces a general base in the active site, in the form of the aspartate side chain, which can adopt the correct position to abstract a proton from a water molecule to yield a nucleophilic hydroxide that can attack the phosphorylated serine. This confers to the mutant E3 the ability to break down organophosphates by making dephosphorylation (and hence reactivation) of the enzyme possible.

In the first step Asp137 activates a water molecule to attack the phosphorus centre of the phospho-serine adduct. The second step consists of the release of Ser218, which becomes protonated again as it abstracts a proton from His471. Table 3.6 and Figure 3.7 record the most relevant geometric parameters of the stationary points in this reaction path. The electronic energy barriers calculated were $\Delta E1$ = 22 kJ/mol and $\Delta E2$ = 60 kJ/mol. This would suggest that the dephosphorylation reaction happens

in the mutant enzyme at a rate much lower than the phosphorylation reaction (as a reminder, the predicted barriers of phosphorylation obtained with an equivalent system were ΔE1 = 29 kJ/mol and ΔE2 = 22 kJ/mol). While this thesis was being written, an article was published[53] that indicates that the much higher second barrier may be an artefact of the calculation. If the barriers obtained are correct, the rate-limiting step in this reaction is the second step: the release of the serine, which takes a proton from His471 (this may need to be revised in a future project in the light of the new information available).

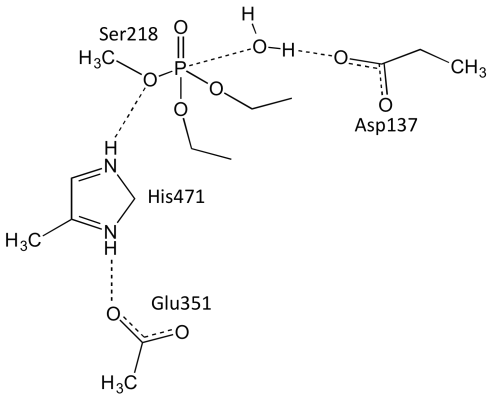
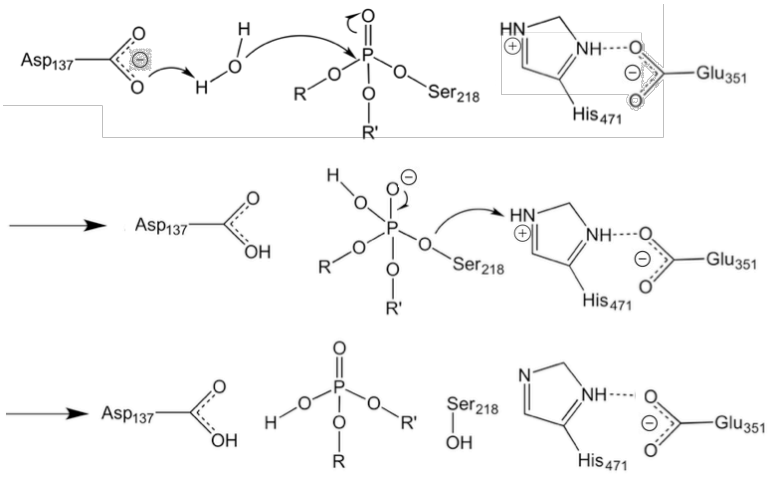


Figure 3.6. *Schematic representation of the system used to study the dephosphorylation of E3 (dECP).* The system included the catalytic triad and mutant Asp137 but omitted the oxyanion hole for simplicity.



Scheme 3.3. *Potential mechanism of dephosphorylation of the E3 Gly137Asp mutant.* Asp137 abstracts a proton from a water molecule, making the oxygen atom more nucleophilic to attack the phosphorus atom. A pentacoordinate intermediate is formed which breaks in a subsequent step when Ser218 is released and reprotonated by His471.

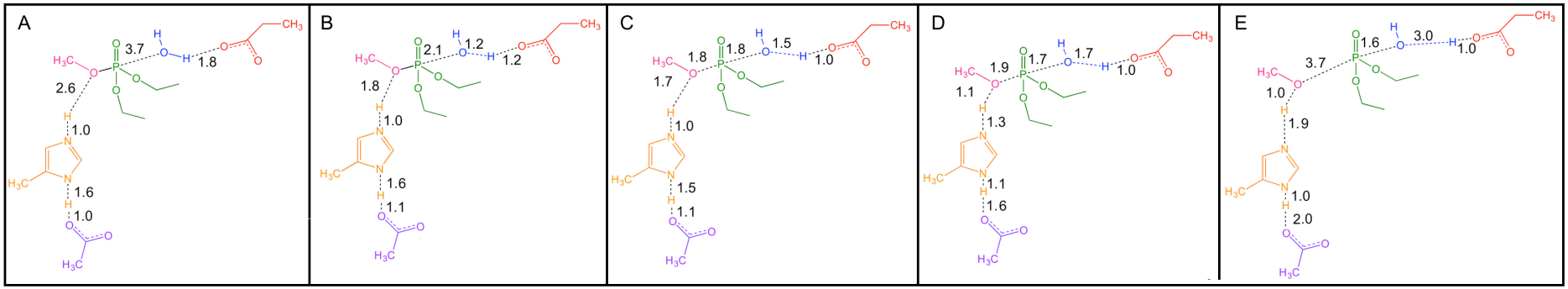


Figure 3.7. *Reaction coordinates of dephosphorylation of E3 (dECP).* **A.** Reactant. **B.** TS1. **C.** Intermediate. **D.** TS2. **E.** Product. Red: Asp137. Blue: water molecule. Green: dECP. Magenta: Ser218. Orange: His471. Purple: Glu351. Distances in Å.

	O_{wat} - P	P - Oγ	H_{wat} - OD	H_{wat} - O_{wat}	Hγ - Oγ	Hγ - Nε	Hδ - OD	Hδ - Nδ
Reactant	3.7	1.6	1.8	1.0	2.6	1.0	1.0	1.6
TS1	2.1	1.7	1.2	1.1	1.8	1.0	1.1	1.6
Intermediate	1.8	1.8	1.0	1.5	1.7	1.0	1.1	1.5
TS2	1.7	1.9	1.0	1.7	1.1	1.3	1.6	1.1
Product	1.6	3.7	1.0	3.0	1.0	1.9	2.0	1.0

Table 3.6. *Principal geometry parameters of the dephosphorylation reaction (dECP).* The data show a concerted attack of the water molecule on the phosphorus and proton transfer to Asp137 (TS1) and phosphorus-oxygen bond break with concerted transfer of a proton from His471 to Ser218 (TS2). Distances in Å.

This result provided a hypothesis that could then be tested experimentally. Specifically, (i) if the second dephosphorylation step was slower than phosphorylation, we would expect to be able to observe a 'burst' phase in the enzyme kinetics and (ii) if the dephosphorylation rate was increased by the Gly137Asp mutation, we would expect the slower, steady state, rate to be increased in the mutant. Data provided by other members of the Jackson laboratory (Mr G. Correy and Dr P. Mabbit, personal communication) confirms this (Table 3.7 and Figure 3.8), showing that the burst (phosphorylation) rate (k_3 in Scheme 3.4) is indeed approximately 1600-fold faster than the steady state (k_{cat}) rate of dephosphorylation. It is worth noting that the calculations presented here and in the previous section would give a 272,000-fold difference between phosphorylation and dephosphorylation (for phosphorylation k = 5.12E+07 s^{-1}; for dephosphorylation 1.89E+02 s^{-1}). These rates were calculated with the equation presented below (Eq. 1.1) and using 298 K as the reference temperature.

$$\mathbf{E} + \mathbf{S} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathbf{ES} \overset{k_3}{\longrightarrow} \mathbf{EP} \overset{k_5}{\longrightarrow} \mathbf{E} + \mathbf{P}$$

Scheme 3.4. *Kinetic equation of E3.* The burst phosphorylation rate is indicated by k_3.

E3 variant	k_3 (s^{-1})	K_M (μM)	k_{cat} (s^{-1}) E-4	k_{cat}/K_M ($M^{-1}s^{-1}$)
Wild type	1.8±0.2	< 1.5	1.0 ± 0.1	>70
Gly137Asp	1.3±0.1	7±2	11.0 ± 1	160±50

Table 3.7. *Kinetic parameters for the hydrolysis of diethylumbelliferyl phosphate at pH 7.5 by E3-WT and E3- Gly137Asp.* Values are the mean ± standard error for 8 replicates. Data provided by G. Correy and P. Mabbit.

$$k = \frac{K_b T}{h}\, e^{-\Delta E/RT} \qquad \text{(Eq. 1.1)}$$

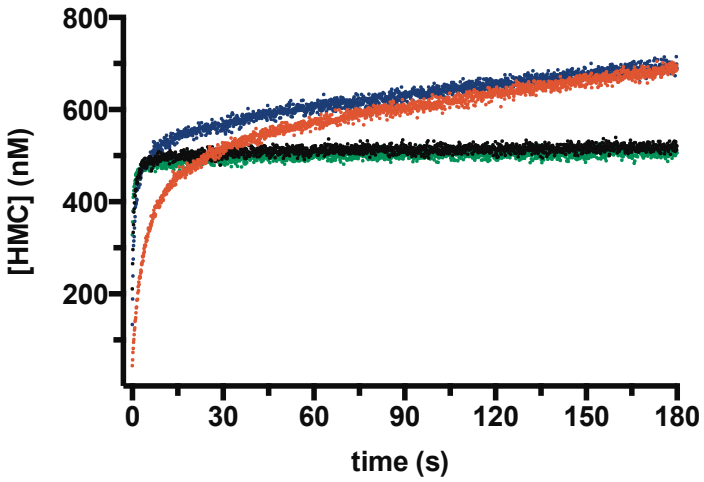


Figure 3.8. *The rapid phosphorylation and slow dephosphorylation of an organophosphate compound by E3 can be monitored by stopped-flow fluorescence.* Representative progress curves of diethylumbelliferyl phosphate hydrolysis by *E3*- WT (green 3 μM substrate, black 24 μM substrate) and *E3*-Gly137Asp (red 3 μM substrate, blue 24 μM substrate) at pH 7.5. Points indicate the concentration of the product 7-hydroxy-4-methyl coumarin (HMC). Data provided by G. Correy and P. Mabbit. The leaving group is 4-methyl umbelliferone.

3.5.3 The importance of the protonation of His471

In the system described in the previous section, His471 is protonated at both nitrogens. In the first step of the reaction Asp137 activates a water molecule to attack on the phosphorus centre of the phospho-serine adduct. In the second step, Ser218 is released as it becomes protonated again by abstracting a proton from His471(Nε) (while Nδ forms a hydrogen bond with Glu351). The electronic energy barriers calculated were ΔE1 = 22 kJ/mol and ΔE2 = 60 kJ/mol, where re-protonation and release of Ser218 is the slowest, rate-limiting, step.

The barrier of dephosphorylation in a system in which the histidine residue is only protonated in Nδ (the group that acts as a hydrogen bond donor to Glu351) was also calculated. In this system the serine residue leaves the phosphate group without abstracting a proton as this happens.

The electronic energy barrier for this step (serine leaving the phosphorus group) was 146 kJ/mol, which is substantially higher than the 60 kJ/mol calculated for the system in which this histidine is doubly protonated. Again, this provided a hypothesis that could be tested experimentally. As shown in Figure 3.9, a strong pH dependence on the reaction was observed: as the pH increases from 5.8 to 8.5 (the stability limits of the protein within which accurate measurements could be made), the catalytic efficiency of the enzyme declines rapidly. This is consistent with a shift in the protonation state of His471 from primarily doubly-protonated in the phosphorylated intermediate, to primarily singly-protonated, i.e. as the second proton is lost through the increase in pH, the energy barrier increases and the rate is reduced. Interestingly, the same trend is not observed in the wild-type protein, suggesting that whatever mechanism is used for dephosphorylation, protonation of the histidine is not required in the rate-limiting step.

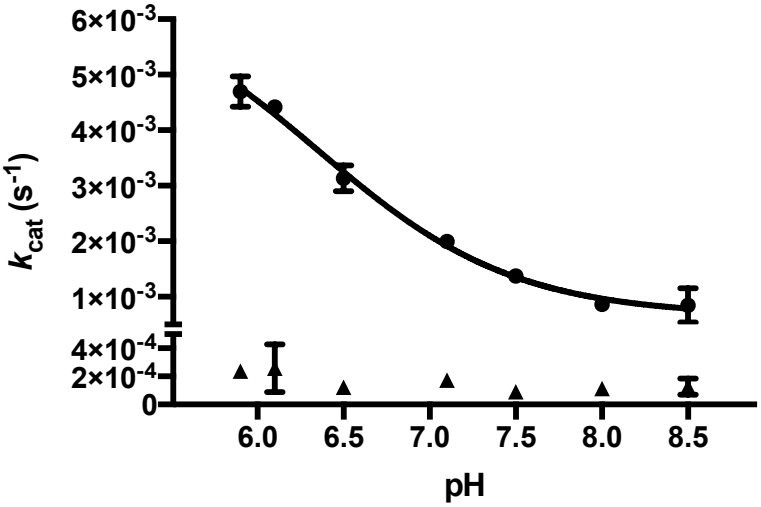


Figure 3.9. *The pH-activity profiles of wild-type E3 and the Gly137Asp mutant for the steady-state dephosphorylation reaction.* Diethylumbelliferyl phosphate hydrolase activity of WT (▲) and Gly137Asp (●) *E3* as a function of pH. Values of k_{cat} are the mean ± *standard deviation for four replicates. The apparent* pK_A *of the* diethylumbelliferyl phosphate hydrolase activity of *E3*-Gly137Asp *was 6.4 ± 0.4* (mean ± *standard deviation).* Data from Ref. 54.

3.6 The impact of restraints on quantum cluster calculations

As it was mentioned previously, the choice of system size, and whether and where to place restraints on the system's geometry, could drastically affect the results of a QC calculation. In order to determine how such changes would affect the results obtained for the system at hand, different options were tested – namely, freezing the first carbon after the functional group of a residue (named Restraint 1 or R1, as was done for the systems described in the previous section), and restraining all but the substrate atoms to the geometry already established for the reaction of phosphorylation with dECP (R2). A free optimisation of all atoms – that is, no restraints at all – was also attempted but the option was discarded because the amino acid side chains moved to positions that would be unrealistic in the enzyme (i.e. residue backbone moving too far apart). System size will be analysed in the next section of this chapter. Another possibility (not analysed in this work) in the search for minimum energy paths would be to use snapshots from Molecular Dynamics simulations as initial structures for these calculations.

3.6.1 Phosphorylation system

Table 3.8 shows the electronic energy barriers (RI-MP2/aug-cc-pvtz) for the phosphorylation reaction with the different substrates that were tested. From that table, it is clear that reactions with different substrates show very different ΔE1 energy barriers, and that a change in the restraints used can alter the results (or the ability to obtain them) dramatically. The reason for the first observation is that side chain groups can adopt slightly different positions in the presence of different substrates, which causes large differences in the calculated energies. This is not necessarily a problem, as such effects will also be present in the

enzyme (and is likely the reason, or part of it, why different substrates are metabolised at different rates), but whether a cluster of this size reproduces these effects correctly needs to be assessed (this is described in Section 3.7). The second observation occurs because, while lack of restraints may lead to artificial movements, an excess of restraints prevents the system from reaching a real minimum or a first-order saddle point. Indeed, several of these calculations did not converge after several attempts and hundreds of hours of computer wall time used.

	R1		R2	
	ΔE1	**ΔE2**	**ΔE1**	**ΔE2**
dECP	29	22	n/a	n/a
Diazinon	17	n/d	n/d	n/d
Dichlorvos	44	63[41]	31	n/d
Parathion	24	15	n/d	n/d

Table 3.8. *Phosphorylation reactions with different substrates and restraints.* Electronic energy (RI-MP2/aug-cc-pvtz) barriers of the four reactions tested. Restraints were R1 (freezing the first carbon after the functional group of a residue) or R2 (restraining all but the substrate to the geometry already established for the reaction that has dECP as a substrate). The number in brackets was obtained by IRC from a modified geometry of TS1 instead of from TS2 (detailed below). R2 was not applicable (n/a) to the dECP system. For several systems transition states were not determined (n/d) as transition states were not located after several attempts.

It was observed that the use of restraints, especially freezing all atoms except for the substrate, causes difficulty in locating transition states. For this reason Table 3.8 (and also Table 3.9 in the dephosphorylation section) contains several blank spaces that correspond to reactions for which it was not possible to locate the transition state after many attempts. As detailed below, at least part of the catalytic triad (the atoms that take part of the reaction) needs to be allowed to move during optimisation, which brings us back to the issues faced with R1.

An unexpected problem faced with R1 restraints (see Figures 3.10–3.14) was that IRC calculations from TS1 and TS2 led in some cases to different intermediate geometries and therefore different energies, i.e. multiple reaction coordinates with slightly different energy barriers were possible. This is especially true in the dephosphorylation reaction (described in the next section). The complexity and relative flatness of the PES of a system held together by multiple hydrogen bonds is the reason why small changes – such as different substrates – cause the system to fall into different energy wells in these optimisations. Examination of the converged structures of the phosphorylation system that had dichlorvos as a substrate (Figure 3.10), showed that in TS2 the residues His471 and Glu351 move closer to each other in order to facilitate the transfer of a proton between the two. This contact persists as a hydrogen bond in the intermediate. However, this contact is non-existent in the X-ray structure of the free enzyme and is not necessary in TS1, so it does not form during the TS1 search and is therefore also absent in the intermediate found following the IRC from this TS. Hence, two different intermediate geometries with different energies are obtained from IRC of TS1 and TS2. Changing the position of these residues in the initial structure used for TS1 search to match the position in TS2 results in a geometry of the intermediate (obtained by IRC) that resembles that of the intermediate obtained by IRC from TS2 more closely than that obtained when the residues in TS1 were not moved from the X-ray position (see Figure 3.10). This intermediate, however, is still more than 20 kJ/mol lower in energy than the intermediate obtained by IRC from TS2, as is evidenced by the electronic energy barriers: 63 kJ/mol from TS2 and 41 kJ/mol from TS1 (the number in brackets in Table 3.8). Reoptimisation of the reactant from this modified TS1 did not cause significant changes to the energy barrier (the difference was 2 kJ/mol).

Overlaying the different phosphorylation TS1 structures (Figure 3.12) shows that the geometries of all systems studied present differences with respect to dECP, although these are small in some cases. Only the geometry of parathion differs significantly from dECP, while diazinon and dichlorvos are a very close match. Examination of the reactant structures shows that the reactants differ more from each other than the transition states. Dichlorvos matches the reactant geometry of dECP better than parathion and diazinon. These differences in geometry result in large differences in the energy barriers seen in Table 3.8.

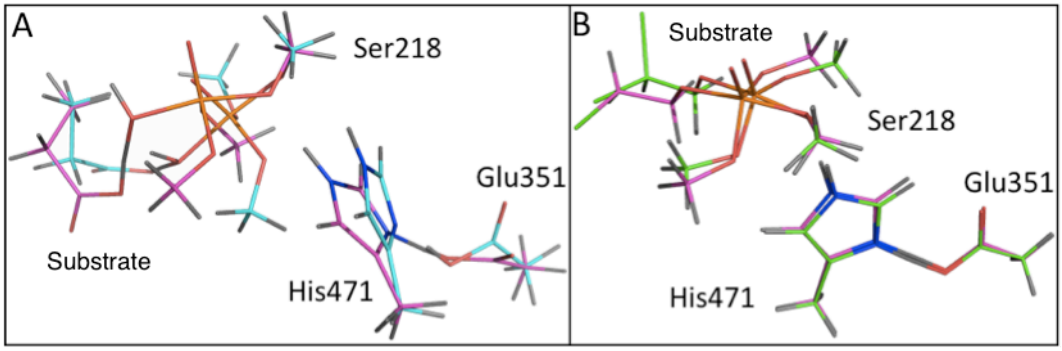


Figure 3.10. *Various geometries of intermediate obtained in the phosphorylation reaction of dichlorvos.* **A.** Intermediate geometry optimised after IRC from TS2 (magenta) and intermediate geometry optimised after IRC from the TS1 geometry initially found (blue). **B.** Intermediate geometry optimised after IRC from TS2 (pink) and intermediate optimised from the TS2 obtained with the modified geometry in which Glu351 and His471 are at a closer distance (green).

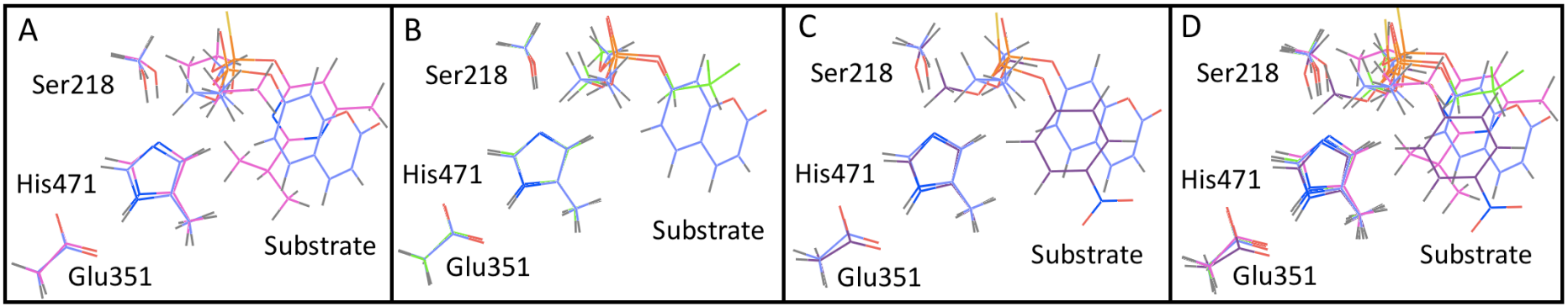


Figure 3.11. *Phosphorylation reaction (R1 systems), reactant geometries.* **A.** dECP (blue) and diazinon (magenta). **B.** dECP (blue) and dichlorvos (green) **C.** dECP (blue) and parathion (purple). **D.** All the previous systems, overlapped. Where multiple geometries were obtained (as described above), the geometries presented here and in all other figures are, unless otherwise specified, those that correspond to the lowest energy barriers.

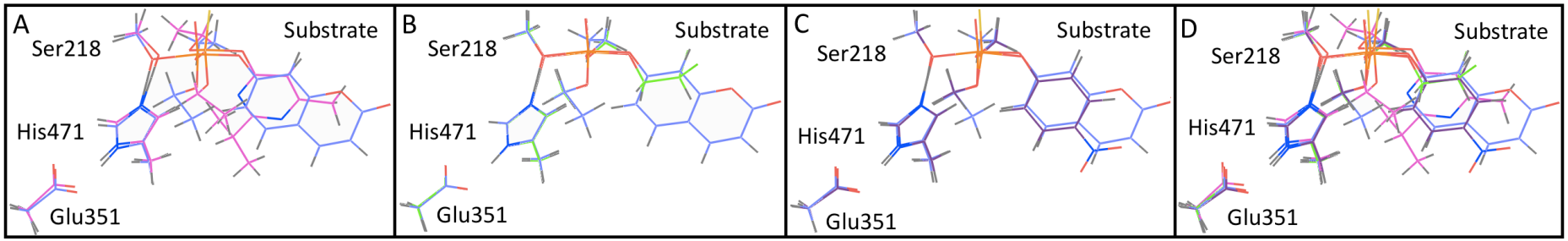


Figure 3.12. *Phosphorylation reaction (R1 systems), TS1 geometries.* **A.** dECP (blue) and diazinon (magenta). **B.** dECP (blue) and dichlorvos (green) **C.** dECP (blue) and parathion (purple). **D.** Overlap of all systems.

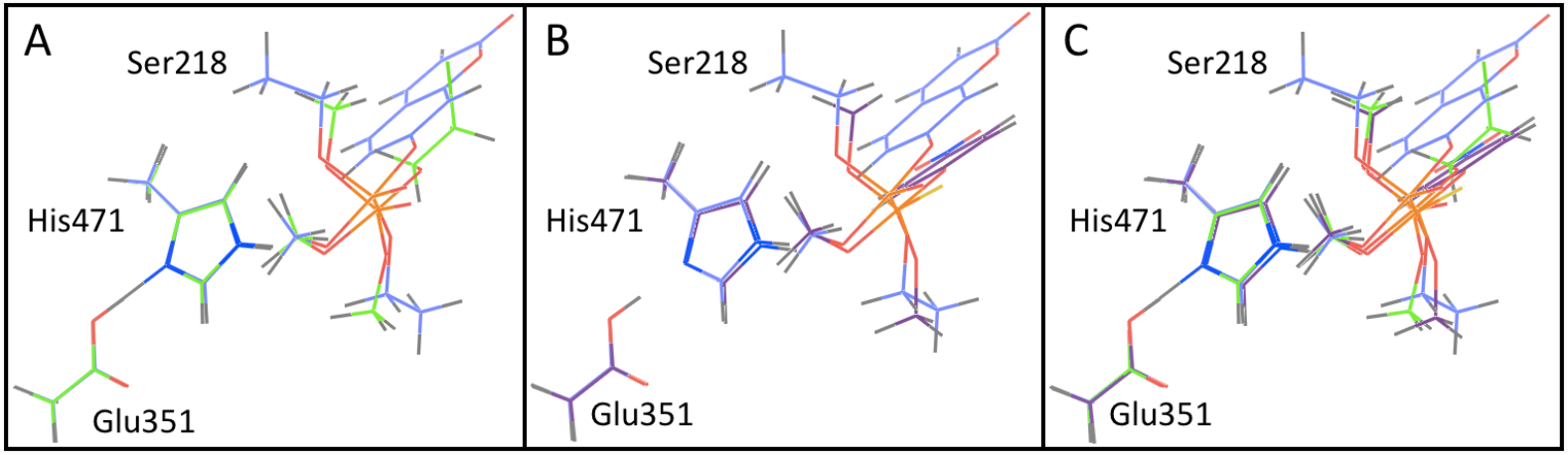


Figure 3.13. *Phosphorylation reaction, intermediate geometries.* **A.** dECP (blue) and dichlorvos (green). **B.** dECP (blue) and parathion (purple). **C.** Overlap of all systems. TS2 and intermediate were not located for diazinon.

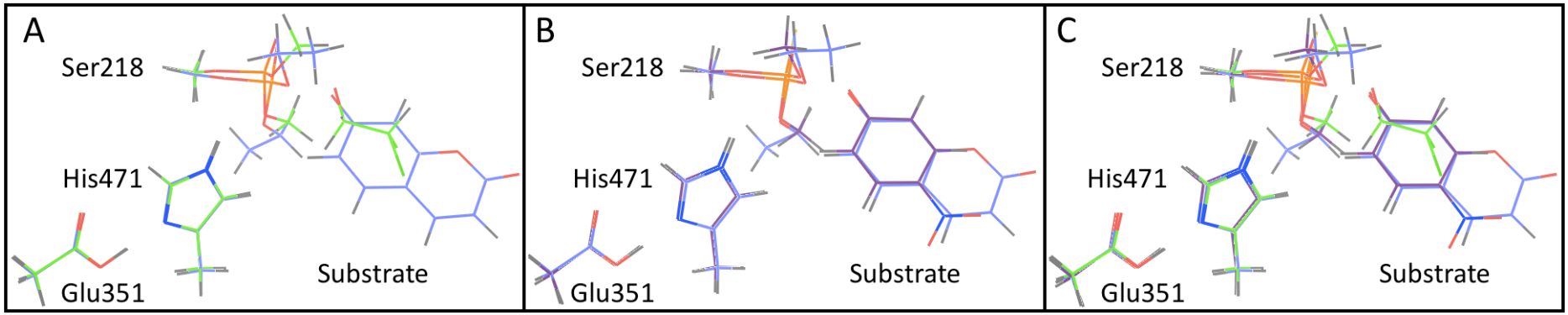


Figure 3.14. *Phosphorylation reaction (R1 systems), TS2 geometries.* **A.** dECP (blue) and dichlorvos (green). **B.** dECP (blue) and parathion (purple). **C.** Overlap of all systems. TS2 and intermediate were not located for diazinon.

3.6.2 Dephosphorylation system

Parathion and dichlorvos have dimethyl side chains, therefore the phosphoadduct has dimethyl side chains, whereas dECP and diazinon have diethyl side chains, and therefore the phosphoadduct has diethyl side chains. For this reason, the dephosphorylation study would be made of two systems instead of four. In order to have more data points for comparison, two new dephosphorylation systems, the E3-VX(R) and E3-VX(S) adducts, were introduced. For these, VX was docked to E3 (see Chapter 4) and manually moved to a distance suitable for attack.

During examination of the converged structures of the dephosphorylation system for the diethyl and dimethyl substrates it was noticed that in TS2 the His471 residue moves closer to the Ser218, in order to transfer a proton to the serine oxygen atom. This contact still exists, as a hydrogen bond, in the intermediate. There is no such hydrogen bond, however, in the X-ray structure of the free enzyme. It is also not necessary in TS1, and it is therefore also absent from the intermediate as determined by following the IRC from TS1. Hence, two different intermediate geometries with different energies are obtained from IRC of TS1 and TS2.

In the case of the diethyl substrate, the barriers initially obtained were 52 kJ/mol for the first step of the reaction and 60 kJ/mol for the second. It is worth noting that, because of the occurrence of different intermediate geometries as described in the previous paragraph, calculating ΔE2 using the intermediate obtained from TS1 resulted in an erroneous negative barrier, which means this intermediate geometry has higher energy than the transition state. Because of this error, the position of His471 in the initial structure used for TS1 search was changed to match its position in TS2 and a new TS1 geometry (labelled $TS1_2$) was obtained. IRC from $TS1_1$

leads to a geometry of the intermediate that is virtually identical to that obtained by IRC from TS2 and which therefore has its same energy. Furthermore, reoptimisation of the reactant was attempted by IRC from $TS1_1$ and the geometry obtained substantially decreased the calculated electronic energy barrier, from 52 kJ/mol (this number is shown in brackets in Table 3.9) to 22 kJ/mol (see Figures 3.15 and 3.16). In this case, the optimisations based on modified geometries allowed for a whole reaction path to be constructed. This highlights the need for special care in quantum cluster calculations, in which there exists a complex PES with multiple possible pathways.

For the dimethyl substrate, the first set of barriers obtained were ΔE1 = 26 kJ/mol and ΔE2 = 14 kJ/mol. The intermediates obtained by IRC from TS1 and TS2 were different, and calculating ΔE2 using the intermediate obtained from TS1 yielded and erroneous value of -41 kJ/mol. As discussed for the diethyl adduct before, this error occurs because the two halves of the reaction path do not match. Therefore, an optimisation of TS1 was carried out with a modified geometry in which His471 was moved to its position in TS2. The subsequent IRC calculation and optimisation of the intermediate, however, led to a higher energy barrier for the second step (33 kJ/mol, shown in brackets in table 3.9) than that obtained when the intermediate is optimised from TS2 (14 kJ/mol). The barrier for the first step of the reaction also increased, from 26 kJ/mol to 44 kJ/mol when this new TS1 geometry was used to obtain a new reactant geometry. Therefore, in the case of the dimethyl adduct, the geometries used to try to improve the path led to worse results instead. The time available for this project did not make it possible to attempt more path searches for this system, and a continuous path is yet to be found. Yet another problem faced with the use of restraints was that in one case, namely the dephosphorylation of the E3-

VX(R) adduct, they led to artificial 'negative' barriers, which means that the geometry of the enzyme needs to relax in order to produce realistic energies for TS and minima. For a comparison of the geometries produced by different substrates, see Figures 3.17 – 3.20.

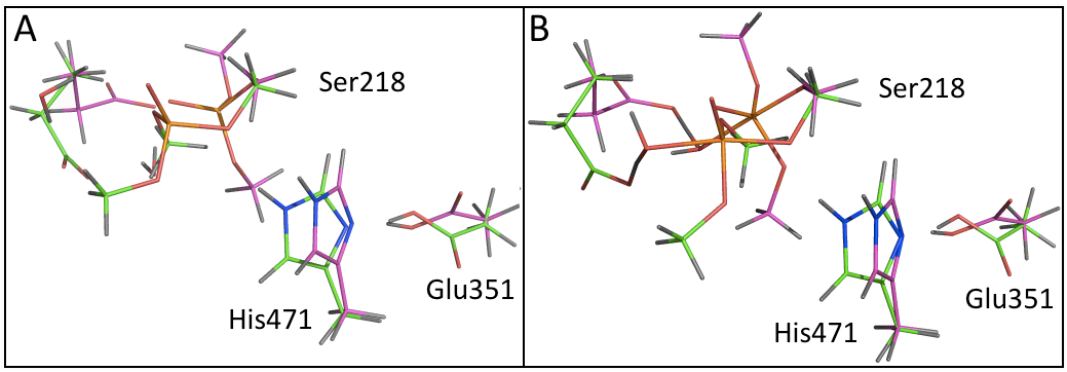


Figure 3.15. *Effect of the position of His471 on the geometry of TS1 and reactant on the dephosphorylation geometries of dimethyl adducts.* **A.** Reactant obtained via IRC from the TS1 geometry initially optimised (magenta) and the reactant obtained from the modified TS1 geometry (green). **B.** Geometry of TS1 optimised initially (magenta) and TS1 reoptimised with a modified geometry (see text for details) (green).

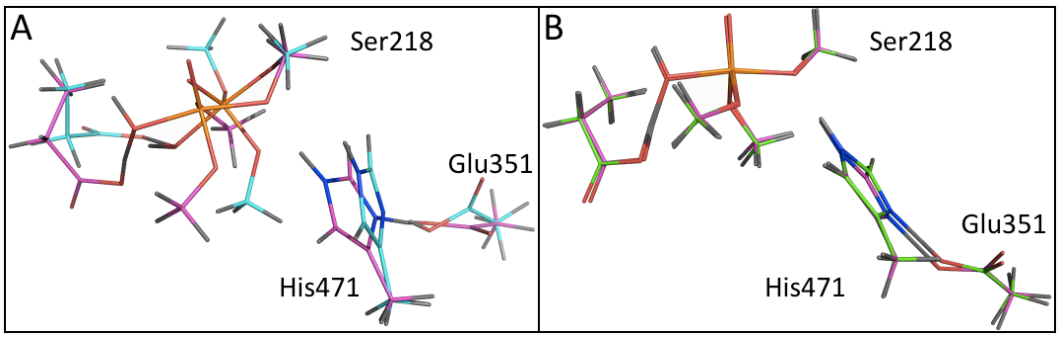


Figure 3.16. *Effect of the position of His471 on the geometry the intermediate geometry on the dephosphorylation geometries of dimethyl adducts.* **A.** Geometry of the intermediate obtained *via* IRC from TS2 (magenta) and geometry obtained from TS1 initially optimised (blue). **B.** Geometry of the intermediate obtained *via* IRC from TS2 (magenta) and geometry obtained from TS1 reoptimised with a modified geometry (see text for details) (green).

	R1		R2	
	ΔE1	**ΔE2**	**ΔE1**	**ΔE2**
Diethyl adduct	22[52]	60	n/a	n/a
Dimethyl adduct	26[44]	14[33]	7	57
VX(R) adduct	n/d	-27	-10	n/d
VX(S) adduct	24	n/d	n/d	24

Table 3.9. *Dephosphorylation reactions with different substrates and restraints.* Electronic energy (RI-MP2/aug-cc-pvtz) barriers for the four dephosphorylation reactions studied. Restraints were R1 (freezing the first carbon after the functional group of a residue) or R2 (restraining all but the substrate to the geometry already established for the reaction that has dECP as a substrate). R2 was not applicable (n/a) to the dECP system. For several systems transition states were not determined (n/d) as they were not located after several attempts. All values are in kJ/mol. Numbers in brackets indicate a less-favourable energy barrier obtained with other geometries (as described in the text).

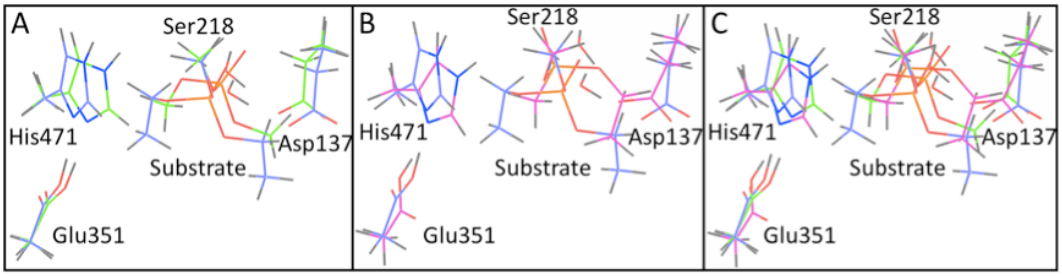


Figure 3.17. *Dephosphorylation reaction, reactant geometries.* **A.** E3-diethyl adduct (blue) and E3-dimethyl adduct (green). **B.** E3-diethyl (blue) and E3-VX(S) adduct (purple). **C.** Overlap of all the systems (dimethyl, diethyl and E3-VX(S)). The geometries used in this and all other captions, unless otherwise specified, are those that lead to a lower energy barrier, as these would be the preferred conformations.

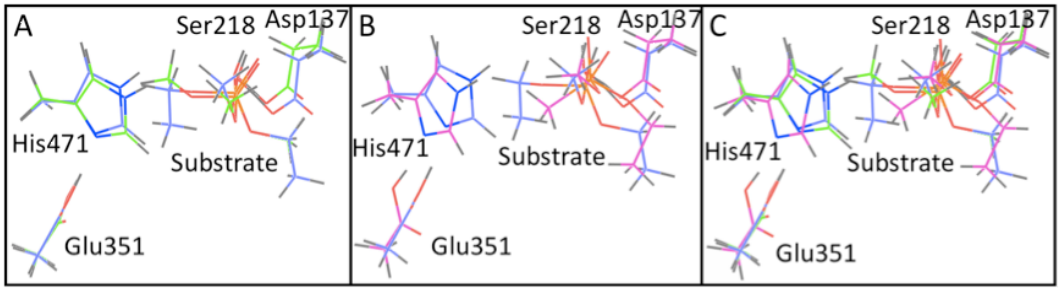


Figure 3.18. *Dephosphorylation reaction, TS1 geometries.* **A.** Diethyl (blue) and dimethyl (green). **B.** Diethyl (blue) and VX(S) (purple). **C.** Overlap of all the systems.

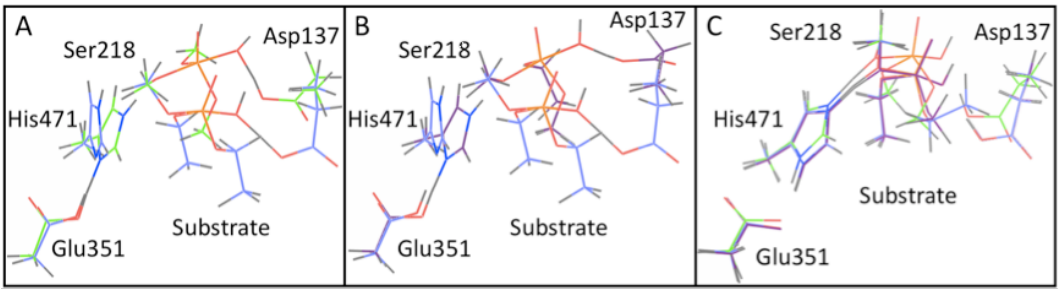


Figure 3.19. *Dephosphorylation reaction, intermediate geometries.* **A.** E3-diethyl adduct (blue) and E3-dimethyl adduct (green). **B.** E3-diethyl (blue) and E3-VX(R) adduct (purple). **C.** Overlap of these intermediate structures.

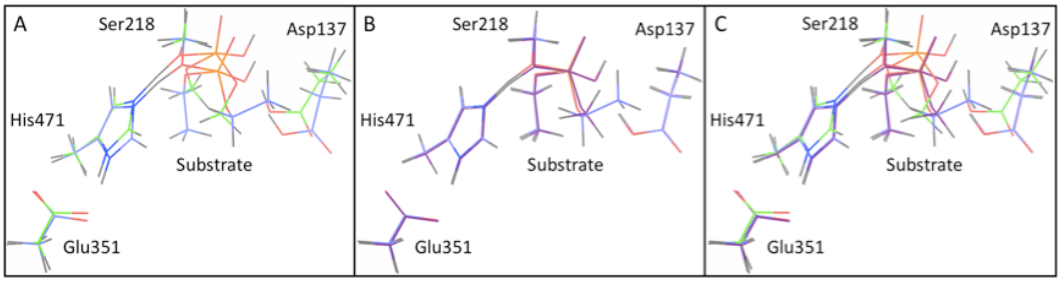


Figure 3.20. *Dephosphorylation reaction, TS2 geometries.* **A.** E3-diethyl adduct (blue) and E3-dimethyl adduct (green). **B.** E3-diethyl (blue) and E3-VX(R) adduct (purple). **C.** Overlap these TS2 geometries.

3.7 Analysis of the size of the system

In addition to the small systems described above, which consist of the side chains of the catalytic triad anchored to a methyl group plus the substrate; a larger model was built that includes the catalytic triad groups up to the alpha carbon besides the oxyanion hole (Figure 3.21). For the dephosphorylation reaction, the residue Asp137 was added as well (it is only CH_3COO^- in the small model but includes the side chain up to Cα in the large model). Only the Cα methyl group was kept frozen in these calculations, so they are effectively of the same type as R1.

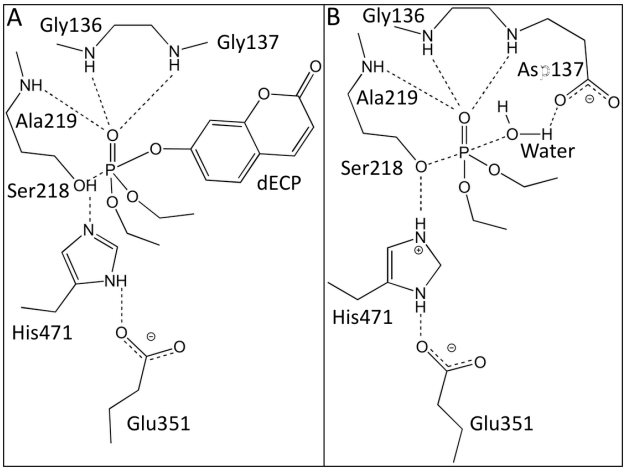


Figure 3.21. *Large system models.* **A.** Phosphorylation reaction. Active site (catalytic triad and oxyanion hole) with dECP as substrate. E3 wild type. **B.** Dephosphorylation reaction. Active site (catalytic triad, oxyanion hole and mutant aspartate), enzyme phosphorylated with dECP.

3.7.1 Phosphorylation reaction (dECP)

The main geometry parameters of the large phosphorylation system are summarised in Table 3.10. Transition states that led to the correct minima after IRC and optimisation were obtained. However, and similarly to what was observed for the small-sized systems, the phosphorylation reaction produced two different conformers of the intermediate when following the IRC from the two transition states – that is to say, the geometry of the intermediate obtained by IRC from TS1 was different from that of the intermediate obtained by IRC from TS2. It was noted that the main difference in the geometries of TS1 and TS2 was the position of His471. An attempt to re-optimise TS2 adjusting the position of His471 to resemble that observed in TS1 (that is, making a hydrogen bond with the serine residue) was made, which yielded a TS2 and intermediate both with the same geometry and energy as the ones initially found. The calculated energy barriers of the reaction are ΔE1 = 21 kJ/mol and ΔE2 = 7 kJ/mol (using the intermediate found following the IRC from TS2). These differ significantly from the value obtained previously with the small system, in which ΔE1 = 29 kJ/mol and ΔE2 = 22 kJ/mol (see summary in Table 3.11

and Figure 3.22). It is clear that the presence of the oxyanion hole reduces the energy barriers of the reaction considerably, especially that of the second step. The low value of ΔE2 = 7 kJ/mol might indicate a concerted process instead of a two-step reaction, but more work is needed to clarify this aspect.

	Reactant	TS1	Intermediate	TS2	Product
P - Oγ	3.1	1.9	1.7	1.7	1.6
Hγ - Oγ	1.0	1.1	1.8	2.2	2.7
Hγ – Nε	1.7	1.4	1.0	1.0	1.0
P - O_{dECP}	1.6	1.8	1.9	2.1	4.0
Hδ - OD	1.7	1.6	1.1	1.1	1.1
Hδ - Nδ	1.1	1.1	1.6	1.6	1.5
O_P - H_{219}	2.6	2.3	2.5	2.4	2.8
O_P - H_{136}	3.3	3.3	3.4	3.4	3.2
O_P - H_{137}	2.0	2.2	2.3	2.4	2.3

Table 3.10. *Principal geometry parameters of the phosphorylation reaction with dECP (large system).* O_{dECP} = oxygen of leaving group that participates in the bond that breaks (with P). O_P – H_{219}: distance between the phosphoryl oxygen of dECP and the backbone hydrogen of residue 219. The same applies for the other members of the oxyanion hole. Distances in Å.

Reading Table 3.10 it is clear that the relevant distances involved in the reaction coordinate are exactly the same for the large and small systems, despite the different number of groups included. Figure 3.23 shows that the angles that determine the relative orientation of the chemical groups, however, do change. The energies calculated indicate that these different conformations have very different energies and give different results for the calculated barriers.

	Large system	**Small system**
ΔE1	21	29
ΔE2(2)	7	22
ΔE2(1)	-6	-

Table 3.11. *Phosphorylation, large system.* Electronic energy barriers (kJ/mol) calculated with different conformers of the intermediate. (1) indicates that the intermediate was obtained by IRC from TS1 and then optimisation. (2) indicates that the intermediate was located by the same procedure but starting from TS2 rather than TS1. The corresponding values for the small system are included for the purpose of comparison.

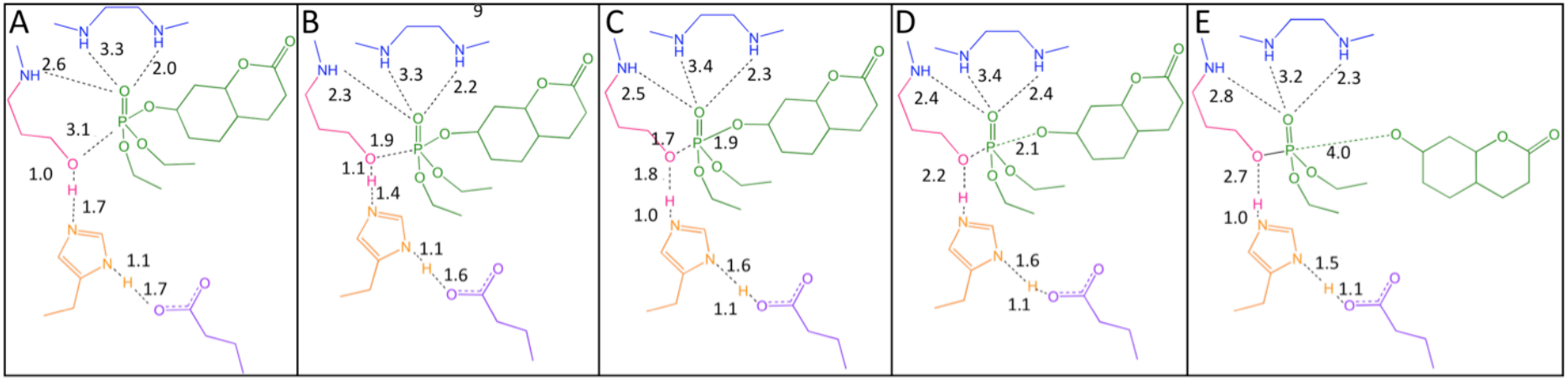


Figure 3.22. *Geometry coordinates of the phosphorylation of the large system.* **A.** Reactant. **B.** TS1. **C.** Intermediate. **D.** TS2. **E.** Product.

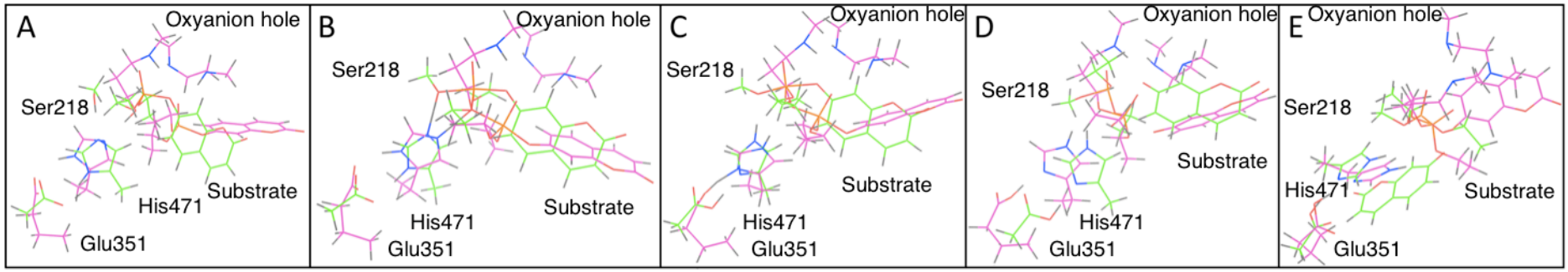


Figure 3.23. *Comparison of structures of the phosphorylation reaction for different size of systems.* **A.** Reactant. **B.** TS1. **C.** Intermediate. **D.** TS2. **E.** Product. Green = small system, magenta = large system.

3.7.2 Dephosphorylation reaction

In the dephosphorylation reaction (of which the geometry parameters are presented in Table 3.12 and Figure 3.25) with the diethyl-substituted phosphor-adduct (from reaction with dECP), the landscape was more complex. Three different reactant conformers, three TS1 conformers and four intermediate conformers were identified – and in all probability many more exist. The multiple conformers (see Figure 3.24) were found during an attempt to make the TS1 geometry more similar to that of TS2 with the aim of obtaining a single intermediate that linked TS1 and TS2 into a single reaction path. The TS1 initially optimised was labelled $TS1_1$; a modification of its geometry to place His471 in a similar position to TS2 was labelled $TS1_2$, and a subsequent TS search in which the active site structure obtained for TS2 was used – modifying the position of the reacting atoms and then optimising it with only Cα frozen – yielded $TS1_3$. Each intermediate and reactant geometry were located from an IRC search from the respective TS1 geometry. As was expected, $Intermediate(1)_1$ – the number between parenthesis indicates whether it was optimised from TS1 or TS2 – was the most different in terms of geometry from Intermediate(2), and $Intermediate(1)_3$ was the most similar, since the active site conformation of TS2 was used to identify $TS1_3$. The RMSD between conformers ranges between 0.4 and 1.0 Å (Table 3.13). However, the intermediate conformers more similar to Intermediate(2) resulted in higher energy barriers than this conformer, rather than more similar ones (see Table 3.14). Clearly, obtaining a complete reaction path is more difficult than simply adjusting the position of one residue. The modified TS1 structures and their corresponding reactant geometries did not lead to lower energy barriers for the first step, which suggests that what really happens is a rearrangement of the intermediate, from the structure

obtained from TS1, into the structure that precedes TS2. Unfortunately, time constraints did not make it possible to search for a transition between these two intermediates as part of this project. The larger system, however, showed that the overall barrier for dephosphorylation was overestimated by the small system, as is discussed next (also see Figure 3.26 for a comparison of geometries).

	Reactant	TS1	Intermediate	TS2	Product
P - O_{wat}	3.8	2.0	1.8	1.7	1.6
P - Oγ	1.6	1.7	1.7	2.0	3.8
H_{wat} - OD	1.8	1.2	1.0	1.0	1.0
H_{wat} - O_{wat}	1.0	1.2	1.5	1.6	2.6
Hγ - Oγ	3.9	3.7	2.1	1.3	1.0
Hγ – Nε	1.0	1.0	1.0	1.3	2.6
Hδ - OD	1.0	1.1	1.1	1.5	2.3
Hδ - Nδ	1.6	1.6	1.6	1.1	1.1
O_P - H_{219}	3.3	3.5	4.4	4.7	4.7
O_P - H_{136}	4.0	3.5	5.4	5.8	5.9
O_P - H_{137}	2.6	2.7	4.2	4.6	5.3

Table 3.12. *Principal geometry parameters of the dephosphorylation reaction (dECP).* The data show a concerted attack of the water molecule on the phosphorus and proton transfer to Asp137 (TS1) and phosphorus-oxygen bond break with concerted transfer of a proton from His471 to Ser218 (TS2). The geometries in this reaction path correspond to the species that gave the lowest energy barriers. Distances in Å.

Conformer of intermediate	RMSD (all atoms)
$Intermediate(1)_1$ - $Intermediate(2)_1$	1.0
$Intermediate(1)_2$ - $Intermediate(2)_1$	0.7
$Intermediate(1)_3$ - $Intermediate(2)_1$	0.5
$Intermediate(1)_1$ - $Intermediate(1)_2$	1.0
$Intermediate(1)_1$ - $Intermediate(1)_3$	1.0
$Intermediate(1)_2$ - $Intermediate(1)_3$	0.4

Table 3.13. RMSD of intermediate conformers (dephosphorylation, large system). The all-atoms RMSD between different conformers of the intermediate shows significant variations ranging between 0.4 and 1.0 Å. The subscripts have the same meaning as detailed above.

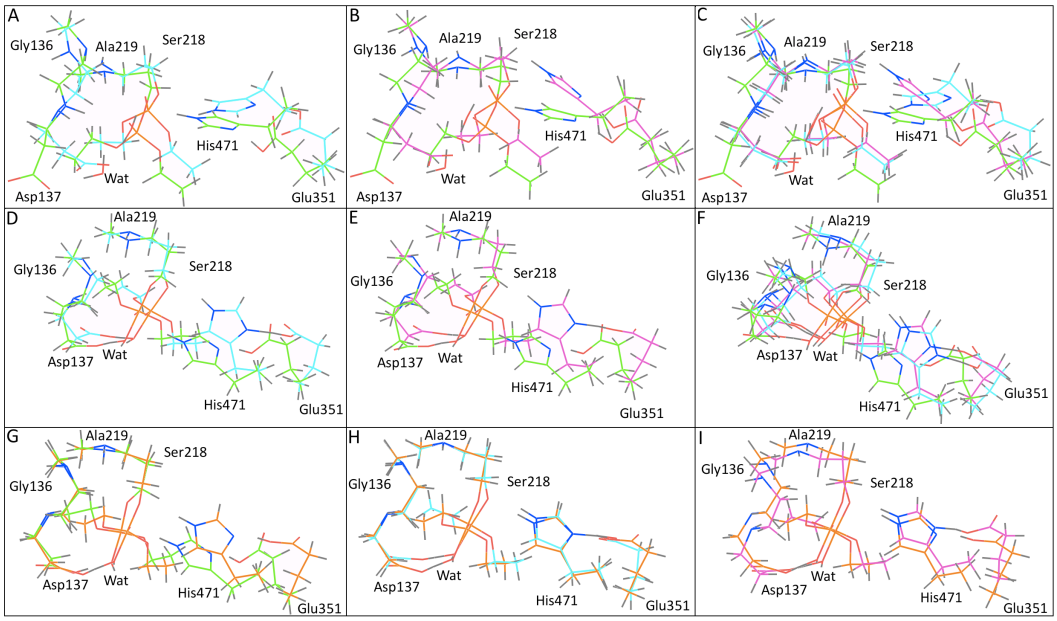


Figure 3.24. *Multiple structures of stationary points of the dephosphorylation reaction.* **A.** Reactant structures R_1 (green) and R_2 (blue). **B.** Reactant structures R_1 (green) and R_2 (magenta). **C.** Superimposition of R_1, R_2 and R_3. **D.** TS1 structures $TS1_1$ (green) and $TS1_2$ (blue). **E.** TS1 structures $TS1_1$ (green) and $TS1_3$ (magenta). **F.** Superimposition of $TS1_1$, $TS1_2$ and $TS1_3$. **G.** Intermediate structures I(2) (orange) and $I(1)_1$ (green). **H.** Intermediate structures I(2) (orange) and $I(1)_2$ (blue). **I.** Intermediate structures I(2) (orange) and $I(1)_3$ (magenta).

Species considered	Barrier	Large system	Small system
$TS1_1$ – $Reactant_1$	ΔE1	36	22
$TS1_2$ – $Reactant_2$	ΔE1	70	-
$TS1_3$ – $Reactant_3$	ΔE1	63	-
TS2 - $Intermediate(1)_1$	ΔE2	37	-
TS2 - Intermediate(2)	ΔE2	23	60
TS2 - $Intermediate(1)_2$	ΔE2	42	-
TS2 - $Intermediate(1)_3$	ΔE2	46	-

Table 3.14. *Energy barriers (dephosphorylation reaction of large system) calculated using different geometries.* The subscripts indicate the following: $_1$Geometries optimised using the X-ray structure as a starting point. $_2$Geometries obtained after placing His471 in the same position as in TS2. $_3$Using the optimised geometry of TS2, the position of the reacting atoms was modified and after a TS search a new TS1 conformer was obtained. (1) and (2) have the same meaning as in Table 3.13 (above). Here too the energy barriers calculated for the small system are included for comparison.

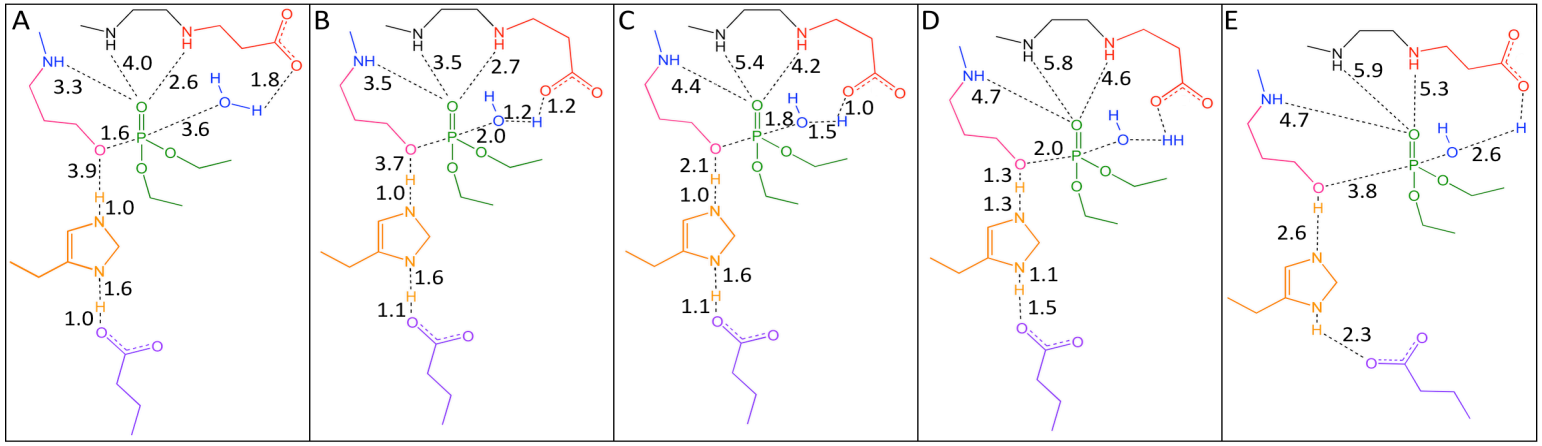


Figure 3.25. *Geometry coordinates of the dephosphorylation of the large system.* **A.** Reactant. **B.** TS1. **C.** Intermediate. **D.** TS2. **E.** Product. The geometries presented here, and in all other figures (unless otherwise specified) correspond to the conformers that gave the lowest energy barriers. Distances in Å.

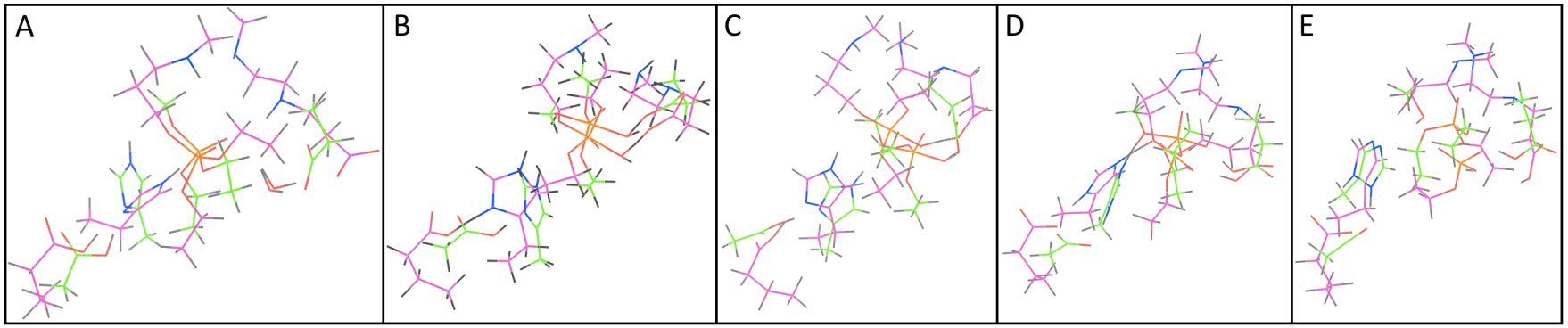


Figure 3.26. *Comparison of structures of the dephosphorylation reaction for large and small systems.* **A.** Reactant. **B.** TS1. **C.** Intermediate. **D.** TS2. **E.** Product. Green = small system, magenta = large system.

The energy barriers calculated between the various reactants, transition states and intermediates depend on the conformer chosen and range between 36 and 70 kJ/mol (ΔE1) and between 23 and 46 kJ/mol (ΔE2) (see Table 3.14 for energy comparison and Table 3.12 for geometry). The lowest barriers calculated by this method are 36 kJ/mol for the first step and 23 kJ/mol for the second.

Evidence of the importance of the oxyanion hole in decreasing the energy barrier of these reactions comes, rather unexpectedly, from the multiple geometries of TS1 and reactant discussed earlier. A big difference between the conformers is the distance between the phosphoryl oxygen and the oxyanion hole (Table 3.15). The lowest ΔE1 energy barrier is that provided by $TS1_1$ and the corresponding reactant, R_1. The second lowest barrier is that of $TS1_3$ with R_3, and the highest barrier is that calculated with the conformers $TS1_2$ and R_2 (Table 3.16). As is can be seen in Table 3.15, the phosphoryl oxygen is in much closer contact with the oxyanion hole in $TS1_1$ than in $TS1_2$ and $TS1_3$. This would stabilise the transition state and explain the lower energy barrier. Furthermore, although $TS1_2$ and $TS1_3$ have virtually the same parameters for these distances, in R_2 the phosphoryl oxygen is much closer to one of the three members of the oxyanion hole than in R_3 (while the other two are at the same distance). If the reactant is stabilised by the oxyanion hole, the energy barrier for the pair $TS1_2 – R_2$ would be higher than that of the pair $TS1_3 – R_3$ (although not as large a difference as with respect to $TS1_1$ since in this case only one distance is affected). This is indeed what is observed in the energy barriers calculated. Therefore, although a complete reaction path on a single coordinate could not be obtained, an attempt to solve the problems caused by multiple possible pathways led to evidence that the oxyanion hole plays a role in stabilising the system.

$TS1_1$	$TS1_2$	$TS1_3$		$Reactant_1$	$Reactant_2$	$Reactant_3$
2.0	2.0	2.0	**P - O_{wat}**	3.6	3.8	3.8
1.7	1.7	1.7	**P - Oγ**	1.6	1.6	1.6
1.2	1.2	1.2	**H_{wat} - OD**	1.8	1.8	1.8
1.2	1.2	1.2	**H_{wat} - O_{wat}**	1.0	1.0	1.0
3.7	2.3	1.9	**Hγ - Oγ**	3.9	2.5	2.9
1.0	1.0	1.0	**Hγ – Nε**	1.0	1.0	1.0
1.1	1.1	1.1	**Hδ - OD**	1.0	1.1	1.0
1.6	1.5	1.5	**Hδ - Nδ**	1.6	1.5	1.6
3.5	4.5	4.5	**O_P - H_{219}**	3.3	4.4	4.4
3.5	5.4	5.4	**O_P - H_{136}**	4.0	5.3	5.3
2.7	4.3	4.4	**O_P - H_{137}**	2.6	2.6	4.3

Table 3.15. *Relevant geometry (distance) parameters of the different conformers of TS1 and reactant.* Distance between different atoms in the multiple conformers described in the text. Distances in Å.

Conformers	ΔE1 (kJ/mol)
$TS1_1$ - R_1	36
$TS1_2$ - R_2	70
$TS1_3$ -R_3	54

Table 3.16. *Energy barriers obtained with the multiple conformers.* The subscript indicates the respective conformer (as detailed in text).

The geometries of the small and large system resemble each other more closely in the dephosphorylation reaction than they did in phosphorylation. Dephosphorylation was found to have very similar distance coordinate parameters in the small and large systems (see Tables 3.9 and 3.16). A few exceptions are the distances between Hγ and Oγ in TS1 and intermediate, those between Hγ and Nε in TS1 and product, and those between H_{wat} and O_{wat} in the product. The barriers calculated for each system, however, are highly different, which translates to very different calculated reaction rates (summarised in Table 3.17). The small systems led to a difference in predicted reaction rates of 272,000-fold between phosphorylation and dephosphorylation (with phosphorylation being the fastest), while the large system predicts that the difference in rate between phosphorylation and dephosphorylation is a factor of 426.

This is in reasonable agreement with the experimental result (ca. 1600 fold) if it is considered that entropic differences are neglected in the quantum cluster calculations. These results are encouraging, indicating that the larger system is the appropriate size to reproduce experimental results.

	Small system	Large system
Phosphorylation	5.12E+07	1.29E+09
Dephosphorylation	1.89E+02	3.04E+06

Table 3.17. *Enzyme reaction rates obtained with different size of systems.* Rates in units of s^{-1}.

Attempts to obtain a lower energy path by modifying the structures were unsuccessful, however, this does not mean that alternative paths do not exist. Many pathways are indeed possible, but structural constraints imposed by the enzyme can limit pathway choice. One possibility for future work is to explore different conformations (e.g. using umbrella sampling methods) and the barriers they are associated to, with the aim of finding a conformation, if possible, that makes the reaction faster and finding out whether the enzyme can be forced into such conformation by engineering.

3.8 Conclusions

This chapter presents an analysis of the molecular basis of pesticide resistance in the blowfly *L. cuprina*. It was established that the Gly137Asp mutation provides an active site base that activates a water molecule for attack on the phosphorylated serine. Alternative possibilities, such as strain, appear to be ruled out by the similarity between the observed rates and the calculated rates (experimentally, phosphorylation is ~1600-fold faster than dephosphorylation, while the calculated difference is 426-fold). The position of the aspartate residue is suitable for the formation of the

trigonal bipyramidal intermediate that forms during the general-base catalyse concerted nucleophilic substitution reaction. It has also been established that the presence of the oxyanion hole and its position with respect to the substrate have a strong impact on the energy barriers calculated, as it stabilises these structures.

Several problems were found with the application of QC methods, which will require further work to optimize so that this method can become widely applicable. The decision of how many (and which) atoms to freeze has a large impact on the calculation. Trying to realistically reproduce the flexibility of the enzyme in a model of reduced size can be challenging, and the choices made have profound effects on the barriers. This work has looked at a few of the countless possible scenarios. While some options are clearly unrealistic – namely using no restraints – it is not always trivial to decide what is right and what is not. A future project could look into how to give the atom coordinates some flexibility without allowing them to move to unrealistic positions. Frozen atoms come with problems such as multiple conformers being found, more than one negative frequency and difficulty locating transition states. Such problems can in principle be avoided by modelling the whole enzyme making use of QM/MM methods. However, with the currently available techniques and typically available computational resources, the QM region must be fairly small and the QM methods must be low-cost ones – reporting DFT energies is standard practice in the field. The benchmarking studies presented in this chapter have shown that these low-cost methods are inaccurate and fail to reproduce the energies predicted by higher-level methods. DFT and DFT-D methods consistently show significant deviations of at least 10 kJ/mol and up to 25 kJ/mol with respect to the energies predicted by G3-MP2-CC. Chemical accuracy requires a deviation of no more than 5 kJ/mol (one

order of magnitude in reaction rates). It is noteworthy that MP2 with double and triple zeta Dunning basis sets reproduces G3-MP2-CC energies to an accuracy of about 4 kJ/mol, and that, even though MP2 is too expensive for the systems at hand, RI-MP2 is just as accurate but so cheap it can – and should – be used routinely in QC systems of the size of the ones presented in this work. It is noteworthy that the proper solvation of negatively charged hydroxyl groups can be a challenge and affect the energies obtained[55]. In this work, however, a very low dielectric was used, and therefore such effects would not be significant.

Calculating Gibbs free energies of reaction profiles inside an enzyme is fraught with problems. Clearly use of ideal gas partition functions is likely inaccurate and in any case difficult to apply for species that, due to geometry constraints, have additional imaginary frequencies. Thus in the present work only electronic energies were used. As a result, the absolute reaction barriers are unlikely to quantitatively meaningful but, given the likely similarities in entropic contributions, the relative barriers were expected to be useful, as was proven by comparison with experiment.

The need to either compromise on the size (and therefore accuracy) of the system in order to use high-level theoretical procedures that yield accurate energies, or use low-level theory in order to allow for a better chemical model, still remains. It is not clear that a suitable compromise is possible for the system at hand, and more work is needed to develop cost-effective theory – for which the RI methods are a promising candidate. Given the aforementioned problems, the best approach with the present resources seems to be to use accurate methods on the largest possible model systems – as has been done in this work. The large phosphorylation model showed a behaviour that was similar to that of the smaller model, which is encouraging, but basing the calculations on certain backbone conformers

could lead to the wrong stationary points. If enzyme fragment models are going to be used, at least some experimental data should ideally be available of a highly similar system to be sure that this is not going to be a problem.

The methods currently available to model enzyme catalysis suffer severe problems – i.e. fragment models do not necessarily reproduce the conformation of the enzyme correctly, and systems that include the whole enzyme are too large to obtain accurate energies. Future work should aim to develop reliable, accurate and less expensive methods that allow biochemistry research to benefit more from *in silico* studies.

3.9 References

(1) Levot, G. W. *International Journal for Parasitology* **1995**, *25*, 1355.

(2) Newcomb, R. D.; Campbell, P. M.; Ollis, D. L.; Cheah, E.; Russell, R. J.; Oakeshott, J. G. *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 7464.

(3) Mulholland, A. J.; Grant, G. H.; Richards, W. G. *Protein Engineering* **1993**, *6*, 133.

(4) Lovell, T.; Himo, F.; Han, W. G.; Noodleman, L. *Coordination Chemistry Reviews* **2003**, *238-239*, 211.

(5) Nachon, F.; Carletti, E.; Wandhammer, M.; Nicolet, Y.; Schopfer, L. M.; Masson, P.; Lockridge, O. *Biochemical Journal* **2011**, *434*, 73.

(6) Jackson, C. J.; Liu, J. W.; Carr, P. D.; Younus, F.; Coppin, C.; Meirelles, T.; Lethier, M.; Pandey, G.; Ollis, D. L.; Russell, R. J.; Weik, M.; Oakeshott, J. G. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 10177.

(7) Main, A. R.; Iverson, F. *Biochemical Journal* **1966**, *100*, 525.

(8) Chiu, Y. C.; Dauterman, W. C. *Biochemical Pharmacology* **1969**, *18*, 1665.

(9) Kwasnieski, O.; Verdier, L.; Malacria, M.; Derat, E. *Journal of Physical Chemistry B* **2009**, *113*, 10001.

(10) Lombardo, D. *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular* **1982**, *700*, 67.

(11) Bernhard, S. A.; Orgel, L. E. *Science* **1959**, *130*, 625.

(12) Järv, J. *Bioorganic Chemistry* **1984**, *12*, 259.

(13) Sultatos, L. G. In *Toxicology of Organophosphate & Carbamate Compounds*; Gupta, R. C., Ed.; Academic Press: Burlington, 2006, p 209.

(14) Kua, J.; Zhang, Y.; Eslami, A. C.; Butler, J. R.; McCammon, J. A. *Protein Science* **2003**, *12*, 2675.

(15) Soreq, H.; Seidman, S. *Nature Reviews Neuroscience* **2001**, *2*, 294.

(16) Rosenberry, T. L. *Advances in enzymology and related areas of molecular biology* **1975**, *43*, 103.

(17) Karami-Mohajeri, S.; Nikfar, S.; Abdollahi, M. *Human and Experimental Toxicology* **2014**, *33*, 92.

(18) Lenz, D. E.; Yeung, D.; Smith, J. R.; Sweeney, R. E.; Lumley, L. A.; Cerasoli, D. M. *Toxicology* **2007**, *233*, 31.

(19) Steitz, T. A.; Hendekson, R.; Blow, D. M. *Journal of Molecular Biology* **1969**, *46*, 337.

(20) Jarv, J. *Bioorganic Chemistry* **1984**, *12*, 259.

(21) Masson, P. *Toxicology Letters* **2011**, *206*, 5.

(22) Jansz, H. S.; Brons, D.; Warringa, M. G. P. J. *BBA - Biochimica et Biophysica Acta* **1959**, *34*, 573.

(23) Berends, F.; Posthumus, C. H.; Sluỳs, I. v. d.; Deierkauf, F. A. *BBA - Biochimica et Biophysica Acta* **1959**, *34*, 576.

(24) Masson, P.; Goasdoue, J. L. *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology* **1986**, *869*, 304.

(25) Millard, C. B.; Kryger, G.; Ordentlich, A.; Greenblatt, H. M.; Harel, M.; Raves, M. L.; Segall, Y.; Barak, D.; Shafferman, A.; Silman, I.; Sussman, J. L. *Biochemistry* **1999**, *38*, 7032.

(26) Masson, P.; Cléry, C.; Guerra, P.; Redslob, A.; Albaret, C.; Fortier, P. L. *Biochemical Journal* **1999**, *343*, 361.

(27) Keijer, J. H.; Wolring, G. Z.; De Jong, L. P. A. *Biochimica et Biophysica Acta* **1974**, *334*, 146.

(28) Adebodun, F.; Jordan, F. *Journal of Cellular Biochemistry* **1989**, *40*, 249.

(29) Sirin, G. S.; Zhou, Y.; Lior-Hoffmann, L.; Wang, S.; Zhang, Y. *Journal of Physical Chemistry B* **2012**, *116*, 12199.

(30) Li, Y.; Sun, X.; Du, L.; Zhang, Q.; Wang, W. *Computational and Theoretical Chemistry* **2012**, *980*, 108.

(31) Li, Y.; Du, L.; Hu, Y.; Sun, X.; Hu, J. *Canadian Journal of Chemistry* **2012**, *90*, 376.

(32) Wong, K. Y.; Gao, J. *Biochemistry* **2007**, *46*, 13352.

(33) Zhang, X.; Wu, R.; Song, L.; Lin, M.; Cao, Z.; Wu, W.; Mo, Y. *Journal of Computational Chemistry* **2009**, *30*, 2388.

(34) Millard, C. B.; Lockridge, O.; Broomfield, C. A. *Biochemistry* **1995**, *34*, 15925.

(35) Amitay, M.; Shurki, A. *Proteins: Structure, Function and Bioformatics* **2009**, *77*, 370.

(36) Izgorodina, E. I.; Brittain, D. R. B.; Hodgson, J. L.; Krenske, E. H.; Lin, C. Y.; Namazian, M.; Coote, M. L. *Journal of Physical Chemistry A* **2007**, *111*, 10754.

(37) Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; Gaussian, I., Ed. Wallingford CT, 2009.

(38) Ho, J.; Coote, M. L. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 649.

(39) Krylov, A. I.; Gill, P. M. W. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2013**, *3*, 317.

(40) Werner, H. J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 242.

(41) O. Trott, A. J. O. *Journal of Computational Chemistry* **2010**, *31*, 455.

(42) Diederichs, K.; Karplus, P. A. *Acta Crystallographica Section D: Biological Crystallography* **2013**, *69*, 1215.

(43) Karplus, P. A.; Diederichs, K. *Science* **2012**, *336*, 1030.

(44) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 810.

(45) Alecu, I. M.; Zheng, J.; Zhao, Y.; Truhlar, D. G. *Journal of Chemical Theory and Computation* **2010**, *6*, 2872.

(46) Piazzetta, P.; Marino, T.; Russo, N. *Physical Chemistry Chemical Physics* **2015**, *17*, 14843.

(47) Ribeiro, A. J. M.; Yang, L.; Ramos, M. J.; Fernandes, P. A.; Liang, Z. X.; Hirao, H. *ACS Catalysis* **2015**, *5*, 3740.

(48) Blomberg, L. M.; Blomberg, M. R. A.; Siegbahn, P. E. M.; Van der Donk, W. A.; Tsai, A. L. *Journal of Physical Chemistry B* **2003**, *107*, 3297.

(49) Beck, J. M.; Hadad, C. M. *Chemico-Biological Interactions* **2010**, *187*, 220.

(50) Nemukhin, A. V.; Lushchekina, S. V.; Bochenkova, A. V.; Golubeva, A. A.; Varfolomeev, S. D. *Journal of Molecular Modeling* **2008**, *14*, 409.

(51) Li, J.; Du, L.; Wang, L. *Journal of Physical Chemistry B* **2010**, *114*, 15261.

(52) Zhou, Y.; Wang, S.; Zhang, Y. *Journal of Physical Chemistry B* **2010**, *114*, 8817.

(53) Duarte, F.; A˚qvist, J.; Williams, N. H.; Kamerlin, S. C. L. *Journal of the American Chemical Society* **2015**, *137*, 1081.

(54) Mabbitt, P. D.; Correy, G. J.; Meirelles, T.; Fraser, N. J.; Coote, M. L.; Jackson, C. J. *Biochemistry* **2016**, *55*, 1408.

(55) Duarte, F.; Geng, T.; Marloie, G.; Al Hussain, A. O.; Williams, N. H.; Kamerlin, S. C. L. *Journal of Organic Chemistry* **2014**, *79*, 2816.

Chapter 4

Investigating the substrate specificity of E3

4.1 Introduction

In the present chapter, the binding of potential substrates to the E3 active site cavity is analysed, with the aim of helping to elucidate the physiological function of E3, and as a first step in determining the substrate range for potential organophosphates substrates. Some of these structures were subsequently used in molecular dynamics simulations to study how mutations alter the conformational landscape of the enzyme and to identify motions that are involved in substrate binding (Chapter 5).

The physiological function of E3 is unknown, but recent results have shown that it plays a role in lipid metabolism as well as resistance to insecticides[1,2], which suggests that the potential substrate range of E3 could be narrowed down to cholesterol/lipid substrates and xenobiotics. Another point of interest is the activity of E3 against organophosphate compounds[3], for which there are no bound structures available (although Dr. Colin Jackson has structures of E3 phosphorylated with dECP and VX). Such structures would be of interest to analyse how substrates bind E3 and possibly identify (or rule out) new substrates. This would help to prevent resistance to pesticides by identifying susceptible compounds early on, and it could also shed light on whether E3 could be used to detoxify organophosphates dangerous to humans. Among these, the nerve agent VX is of special interest given its high toxicity to humans and the

existence of stockpiles that need to be destroyed (see the United Nations Chemical Weapons Convention).

Ligand docking methodologies are particularly useful because they allow the estimation of the structure of enzyme-substrate Michaelis complexes, which are intrinsically short-lived and difficult to capture experimentally. These theoretical complexes allow analysis of the various molecular interactions between enzyme and substrate and prediction of whether potential substrates will bind in catalytically productive orientations[4-6]. - This chapter analyses how Fatty Acid Methyl Esters (FAMEs) as well as organophosphates bind to E3, providing structures of E3-substrate complexes for other calculations.

4.2 Methods

4.2.1 Docking

The structure of E3 from *L. cuprina* was obtained from the Protein Data Bank (PDB ID 4FNG). Rigid docking was performed using Autodock Vina[7]. The input files for docking were creating with AutoDock Tools[8]. The size of the search space was 36 x 52 x 48 Å (xyz respectively) in each case. Docking grids were centred on x = 19.321, y = -9.76, z = -25.543, corresponding to the active site of E3. The substrates tested were the Fatty Acid Methyl Esters methyl hexanoate, methyl octanoate, methyl decanoate, methyl laurate and methyl myristate; and the organophosphates dECP, diazinon, sarin and VX. Analysis of the docked structures was performed using PyMOL[9], and the most suitable ones (substrate phosphorus centre closer to serine 218 and phosphoryl oxygen oriented towards the oxyanion hole) are presented here.

4.2.2 Molecular Dynamics simulations

The coordinates of E3 with VX(S) in the active site obtained by the docking procedures described above were used for a Molecular Dynamics (MD) simulation to test the structural stability of the binding pose. Simulations were performed with the AMBER12 software package[10], using the force field ff12SB[11] for the protein and GAFF[12] parameters for the substrate, VX(S). The protonation state of ionisable residues was decided based on the pKa of the residues (calculated with PROPKA[13-16]) and on visual inspection of local hydrogen bond networks. Coordinate and topology files were generated with the module *tleap* of AMBER12. The system was solvated in explicit water of the TIP3P model[17] and placed in a truncated octahedron periodic box with its walls no less than 10 Å away from the solute, these procedures were also carried out with *tleap*. The system was then energy minimised, first with constraints on the backbone based on the x-ray structure and using a harmonic potential of 500.0 kcal/molÅ^2 to remove bad contacts, and then without restraints. The system was subsequently heated carefully in steps of 50 K for 10ps each (to a total of 50ps) with harmonic restraints of 10.0 kcal/molÅ^2 in the NVT ensemble. The Andersen temperature coupling scheme, as implemented in AMBER12, was used. The system was equilibrated and the simulation was run at a temperature of 300 K in the NPT ensemble with isotropic position scaling, pressure relaxation time of 1.0 ps and compressibility of 44.6E-6 bar^{-1}. One trajectory of 15 ns length was produced, of which the first 6 ns were discarded as equilibration. Analysis of the data produced was carried out with the module *ptraj* of AMBER12. This simulation followed the same protocol as those describe in Chapter 5, where more details of it are given.

4.2.3 X-ray crystallography

Crystals were grown by Dr Colin Jackson at 290 K using the hanging-drop vapor diffusion method, with reservoir solutions of 100 mM sodium-acetate (pH 4.6) and 20% PEG 2K MME. Crystals were soaked in 100 mM sodium-acetate (pH 4.6) and 20% PEG 2K MME with 5 mM racemic VX by Dr Florian Nachon. Diffraction data were collected at beamline ID 14-4 of the European Synchrotron Radiation Facility with wavelength of 0.8726 Å. Diffraction data were indexed, integrated, and scaled using the XDS package[18]. The resolution limits of the data were assessed on the basis of the significance of the $CC_{1/2}$ at the P=0.001 level[19,20] Phases were obtained by molecular replacement with the apo-structure of *Lc* α E7-4[21] (PDB ID 4FNG) using the PhaserMR program implemented in the CCP4 software package[22]. Refinement was carried out using the CCP4 software[23] and the model was built using *Coot*[24] and Phenix[25].

4.3 Investigating the potential substrate range of E3 through *in silico* docking experiments

The potential substrate range of E3 is of great interest for two reasons. First, its physiological role in the fly is currently unknown, despite its high expression in many tissues throughout all stages of the lifecycle of the fly[26]. Testing natural substrates, such as fatty acid esters, is difficult, and often requires laborious experimental approaches, such as the use of gas chromatography with standards, which takes several months to accurately develop and calibrate and cannot assay more than ~10 samples per day. Thus, it is not possible to experimentally test all possible physiological substrates; molecular docking can narrow down the range of substrates significantly. Secondly, E3 has significant potential as a biotherapeutic to treat individuals that have been poisoned by

organophosphates. For this reason, it is essential to gain an idea of which organophosphates are possible substrates. In the case of chemical warfare agents, it is not a simple matter to test these compounds, with only two-three laboratories in the world having access. Thus, preliminary screening through molecular docking experiments is an essential first step. Parts of the following work has been published in "Structure and function of an insect α-carboxylesterase (αEsterase7) associated with insecticide resistance" (Jackson *et al.*) in the Proceedings of the National Academy of Sciences, U. S. A.

4.3.1 Docking FAMEs to E3 (wild type)

Fatty acid methyl esters (FAMEs) are some of the most plausible substrates for E3, given their co-location in fat bodies and the fact that they are carboxylesters, which is known to be the broad substrate class for E3[2,27,28]. In this section, the results of docking different FAMES to E3 (wild type) to test the possibility of a productive binding to the active site are presented. A series of potential fatty acid methyl ester substrates were docked into the active site of E3 using Autodock Vina. Each docking run produced several poses: only the lowest energy (i.e. most stable) poses, which were assessed by visual inspection, are discussed here.

Examination of the enzyme:substrate complexes shows that all substrates are coordinated in a way that positions them correctly for nucleophilic attack by Ser218 of the catalytic triad of E3. Five FAMEs of different length were investigated: methyl hexanoate (6C), methyl octanoate (8C), methyl decanoate (10C), methyl laurate (12C) and methyl myristate (14C). In each case, the carbonyl carbon atom of the FAME is between 3.4 and 3.6 Å from the nucleophilic Oγ of Ser218, a distance that indicates the substrate is well positioned for attack (for geometry parameters see Table 4.3). FAMEs with

a long chain are positioned with it folded rather than extended along the active site gorge, most likely to shield it from polar solvent.

Methyl hexanoate is positioned within the active site with its short (6C) chain along the active site cleft. Its carbonyl carbon is positioned 3.5 Å away from Ser218-Oγ, which is an appropriate distance for approach and attack by Ser218. It is also well positioned with respect to the oxyanion hole, having its carbonyl oxygen pointing towards it at a distance of 4.5 Å from the backbone nitrogen of Ala219, 3.0 Å from N(G136) and 3.2 Å from N(G137) (Figure 4.1).

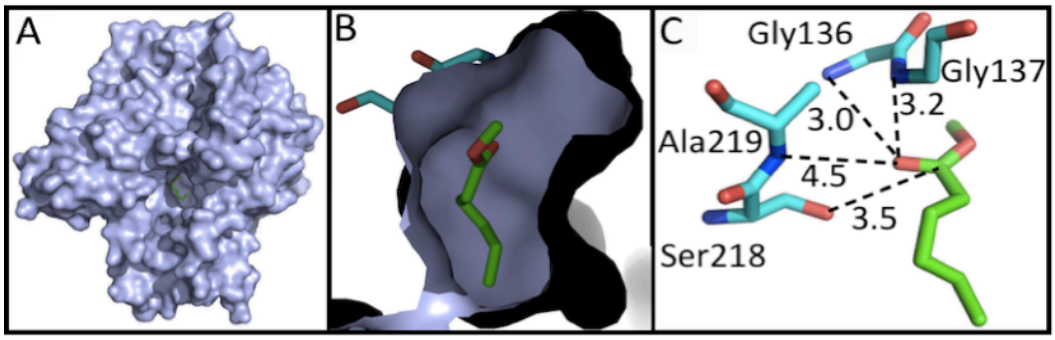


Figure 4.1. *Methyl hexanoate docked in the active site of E3(wt).* **A.** View of the active site gorge of E3, with methyl hexanoate lying in it. **B.** Close-up of the gorge, which shows the chain of methyl hexanoate extended along it. In the background Ser218 is shown in cyan. **C.** Detail of the active site with the substrate placed in it. Distances in Å. Hydrogen atoms have been omitted for clarity.

Methyl octanoate (Figure 4.2) was docked into the active site in a position very similar to that of methyl hexanoate. The carbonyl carbon is 3.4 Å away from the hydroxyl oxygen of Ser218, while the carbonyl oxygen is 4.8 Å away from N(Ala219), 4.3 Å from N(Gly136) and 3.2 Å from N(Gly137).

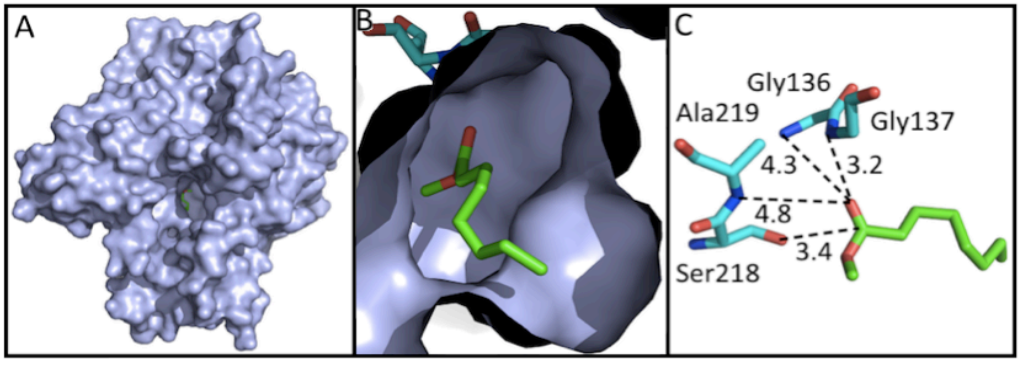


Figure 4.2. *Methyl octanoate docked in the active site of E3(wt).* **A.** View of the active site gorge of E3, with methyl octanoate lying in it. **B.** Close-up of the gorge, which shows the chain of methyl octanoate extended along it. In the background Ser218 is shown in cyan. **C.** Detail of the active site with methyl octanoate in it. Distances in Å. Hydrogen atoms have been omitted for clarity.

Methyl decanoate was also positioned within the active site in a good orientation for attack, with its carbonyl carbon 3.5 Å away from Oγ (see Figure 4.3). The distances between the carbonyl oxygen of the substrate and the oxyanion hole are as follows: 4.5 Å from N(Ala219), 3.0 Å from N(Gly136) and 3.2 Å from N(Gly137). Its chain lies in a folded conformation rather than extended along the cleft.

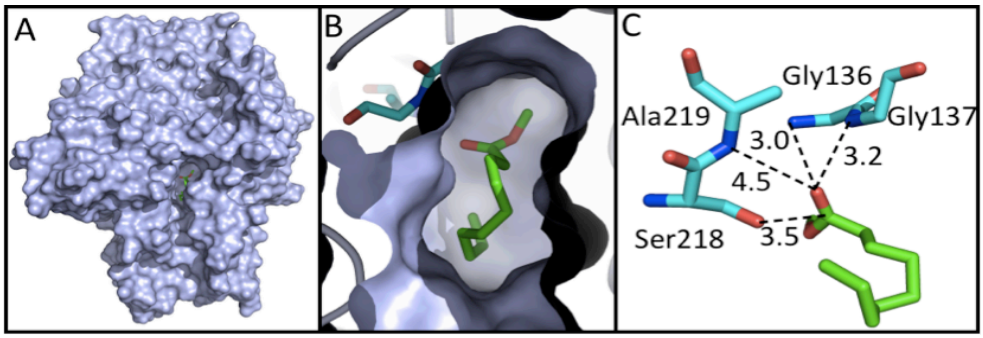


Figure 4.3. *Methyl decanoate docked in the active site of E3(wt).* **A.** Image of E3 that shows the active site cleft and methyl decanoate in it. **B.** Close-up of the cleft and substrate. **C.** Image of the active site with methyl decanoate docked. Distances in Å. All hydrogen atoms are omitted for clarity. In the background Ser218 is shown in cyan.

The long chain (12C) of methyl laurate is also folded like in the case of methyl decanoate, rather than extended along the cleft. The position of methyl laurate in the active site is similar to that of the previous substrates, in which the carbonyl oxygen is tilted towards the members of the oxyanion hole. The distances are as follows: 4.5 Å to N(Ala219), 3.1 Å to N(Gly136), and 3.2 Å to N(Gly137). The carbonyl carbon is 3.5 Å away from Oγ. The data are presented in Figure 4.4.

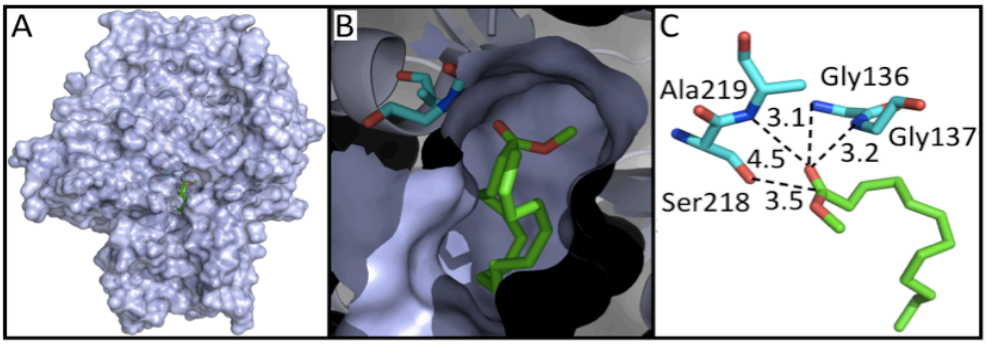


Figure 4.4. *Methyl laurate docked in the active site of E3(wt).* **A.** Image of the enzyme where the gorge and substrate are visible. **B.** Closer image of methyl laurate positioned in the active site and gorge (Ser218 shown in magenta). **C.** Active site and substrate in detail. Distances in Å. Hydrogens are omitted for clarity. In the background Ser218 is shown in cyan.

Methyl myristate also has a long chain (14C) that lies in a folded conformation in the active site cleft (Figure 4.5). Its carbonyl group lies 3.7 Å away from Oγ (a bit further than the other FAMEs that were tested). The carbonyl oxygen is 4.5 Å away from N(A219), 4.1 Å from N(Gly136) and 3.1 Å from N(Gly137). Overall, the position of methyl myristate is good for attack but less optimal than that of the other FAMEs discussed here.

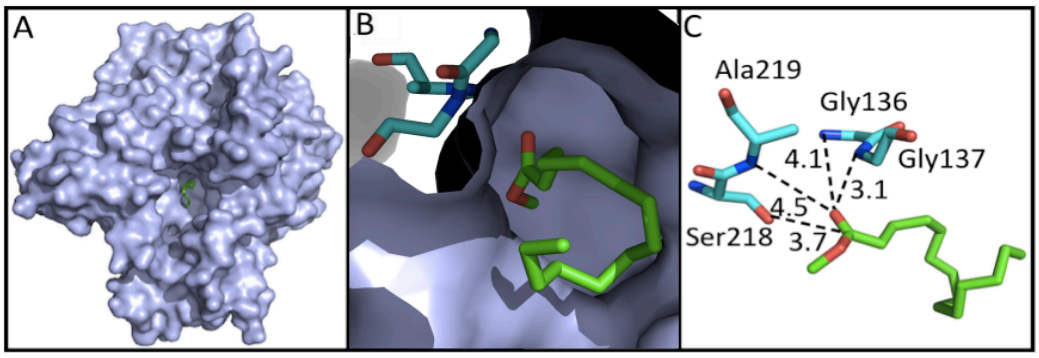


Figure 4.5. *Methyl myristate docked in the active site of E3(wt).* **A.** E3 active site cleft and substrate. **B.** Close-up of methyl myristate positioned in the active site and gorge. **C.** Geometry details of the active site. Distances in Å. All hydrogens are omitted for clarity.

All the FAMEs studied were docked in the active site of wild-type E3 in the correct position for attack. The position of the carboxyl centre with respect to the catalytic triad and oxyanion hole was very similar for all substrates, only methyl myristate was placed a bit further from these centres than the other substrates but still in a good position. It is clear that

all the FAMEs tested fit correctly in the active site cavity. Experimental data taken by Mr Faisal Younis at the CSIRO shows that catalytic rates are very different for different FAMEs (see Table 4.2). Although substrate binding parameters, such as distances and angles, are not enough to predict reaction kinetics, docking techniques have been useful in showing that all these compounds can bind in the right conformation to undergo attack, and to analyse how FAMEs are positioned with respect to the active site groups. This information will be used in the next section to compare to how non-physiological substrates bind E3. The structures obtained by docking organophosphates to E3 were used as a starting point for MD simulations (presented later in this chapter). Unfortunately, time constraints did not allow for simulations of E3 with FAMEs bound to be performed, and so the comparison of natural and unnatural substrates will be part of another project.

Substrate	k_{cat}/K_M (10^6 M^{-1} s^{-1})
Methyl hexanoate	0.28 ± 0.02
Methyl octanoate	0.83 ± 0.07
Methyl decanoate	1.38 ± 0.02
Methyl laurate	0.2 ± 0.03
Methyl myristate	0.061 ± 0.001
Diethyl 4-methylumbelliferyl phosphate	0.05 ± 0.005

Table 4.2. *Kinetic parameters of substrate hydrolysis by E3.* Data from Ref. 29.

	C_{sub}-Oγ	O_{sub}-N_{Ala219}	O_{sub}-N_{Gly136}	O_{sub}-N_{Gly137}
Methyl hexanoate	3.5	4.5	3.0	3.2
Methyl octanoate	3.4	4.8	4.3	3.2
Methyl decanoate	3.5	4.5	3.0	3.2
Methyl laurate	3.5	4.5	3.1	3.2
Methyl myristate	3.7	4.5	4.1	3.1

Table 4.3. *Key parameters of the E3-FAME complexes.* Distances presented are those between the reacting oxygen (Oγ) of the serine and the carbonyl carbon of the substrate (C_{sub}), and those between the carbonyl oxygen of the substrate (O_{sub}) and the backbone nitrogen of given oxyanion hole residues. Distances in Å.

4.3.2 Docking organophosphates to E3 (wild type)

The binding of organophosphate pesticides to E3 is of interest as a first step in analysing whether the blowfly can destroy other, hitherto unstudied, compounds – in particular, warfare agents. Diazinon, dECP, VX, and both isomers of sarin were docked into the active site cavity of wild type E3. Furthermore, chemical warfare agents may only be tested by a limited number of military research institutes, meaning that establishing the potential for chemical warfare organophosphates is essential before proceeding with experimental studies. The docking results indicate that dECP, diazinon, sarin (both isomers) and VX(S) fit in the active and are oriented correctly for catalysis to occur. The isomer VX(R) could not be placed in the active site cavity in the correct orientation. The distance Oγ - $P_{organophosphate}$ for these compounds is in general longer than those observed for FAMEs Oγ - P_{FAME} (between 3.3 and 4.2 Å). This is probably because the leaving group of organophosphates is more bulky than those of shorter chain FAMEs, which suggests the active site cavity may be optimised to accommodate the latter.

The docked structure of dECP in the active site of E3 (Figure 4.6) had its phosphorus atom 3.5 Å away from Oγ(Ser218). The phosphoryl oxygen of dECP was placed 5.1 Å away from the backbone nitrogen of Ala219, 5.2 Å from N(Gly136) and 3.6 Å from N(Gly137). The phosphorus atom was at a distance of 3.5 Å from Oγ(Ser218). The position for attack by the nucleophilic serine was good, but the phosphoryl oxygen is too far from the oxyanion hole for hydrogen bonds to form with two of the three residues in the hole. This is unlike the situation observed when docking FAMEs, for which these distances were significantly shorter.

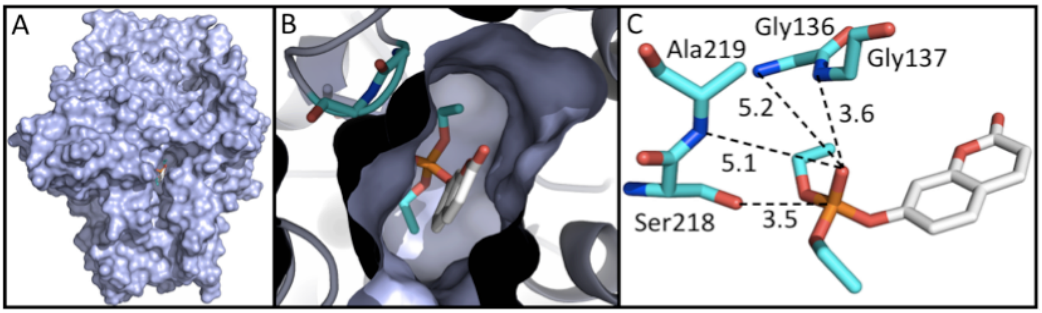


Figure 4.6. *dECP docked in the active site of E3(wt).* **A.** E3 active site cleft and substrate. **B.** Close-up of dECP as docked in the active site gorge. **C.** Geometry parameters of the active site. In the background Ser218 is shown in cyan. Distances in Å.

The other pesticide tested, diazinon, was placed in the active site cavity in a position similar to that of dECP with respect to Ser218, at a distance of 3.6 Å from Oγ. Distances from the phosphoryl sulphur to the oxyanion holes are as follows: 5.4 Å to N(Ala219), 6.2 Å to N(Gly136), and 4.4 Å to N(Gly137) (see Figure 4.7). These distances are longer than those observed for FAMEs and also longer than those of dECP. This is possibly due to the lower electronegativity of sulphur compared to oxygen.

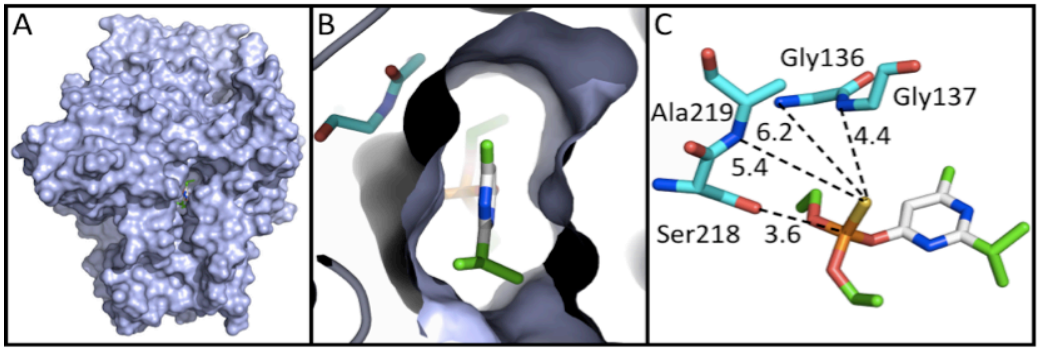


Figure 4.7. *Diazinon docked in the active site of E3(wt).* **A.** Image of the enzyme and active site gorge with diazinon docked. **B.** Close-up of diazinon in the gorge that leads to the active site. **C.** Image detailing the geometry of the docked complex. In the background Ser218 is shown in cyan. Distances in Å.

After the pesticides, this work focused on organophosphates poisonous to humans such as sarin and VX. Each has two isomers, all of which were tested. Sarin(S) (Figure 4.8) bound in a good position for attack by Ser218, although its phosphorus is slightly distant (4.2 Å) from Oγ. This may

reflect a sub-optimal binding, yet it may be easily solved by the enzyme's motions accommodating the substrate in the cavity. The phosphoryl oxygen of the substrate is 6.2 Å away from N(Ala219), 4.8 Å from N(Gly136) and 3.1 Å from N(Gly137). Sarin(R) (Figure 4.9) was docked with its phosphorus atom closer to Oγ (3.7 Å, compared to 4.2 Å for sarin(S)), but it was similarly far from the oxyanion hole. The phosphoryl oxygen was 5.8 Å away from N(Ala219), 4.8 Å from N(Gly136) and 4.2 Å from N(Gly137).

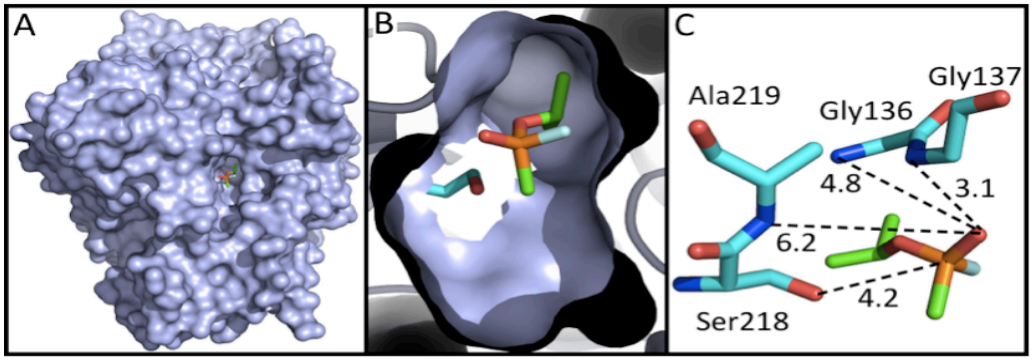


Figure 4.8. *Sarin(S) docked in the active site of E3(wt).* **A.** E3 active site gorge with sarin(S) docked. **B.** Close-up of sarin(S) placed in the gorge (Ser218 is shown in the background). **C.** Geometry of the active site. In the background Ser218 is shown in cyan. Distances in Å.

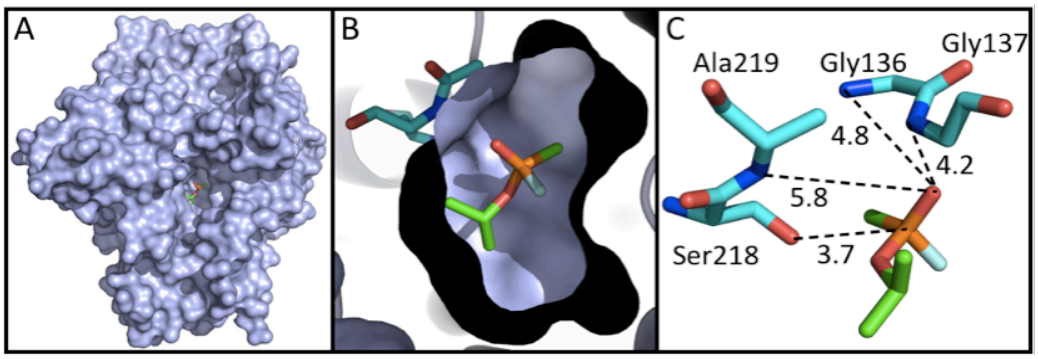


Figure 4.9. *Sarin(R) docked in the active site of E3(wt).* **A.** E3 active site gorge with Sarin(R) docked. **B.** Close-up image of Sarin(R) placed in the cleft that leads to the active site. **C.** Geometry details of the active site. In the background Ser218 is shown in cyan. Distances in Å.

The docked structure of VX(S) had the phosphorus atom placed 3.3 Å away from Oγ, while the phosphoryl oxygen was 4.2 Å away from N(Ala219), 3.6 Å from N(Gly136) and 2.8 Å from N(Gly137) (Figure 4.10).

However, VX(R) (the isomer less toxic to acetylcholinesterase) could not be placed in the active site cavity in a position and orientation suitable for attack by Ser218. This raises the possibility that E3 may have a similar stereoselectivity to human AChE.

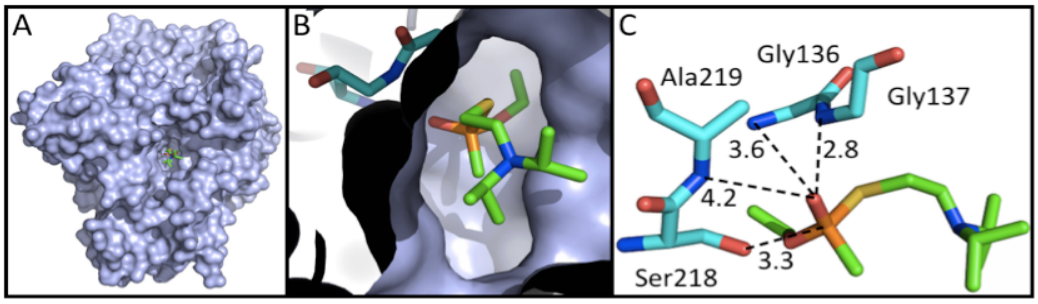


Figure 4.10. *VX(S) docked in the active site of E3(wt).* **A.** Image of the enzyme and active site gorge with VX(S) docked. **B.** Close-up of VX(S) placed in the gorge. **C.** Geometry detail of the active site and docked substrate. In the background Ser218 is shown in cyan. Distances in Å.

The organophosphate compounds here analysed bind wild-type E3 in the a productive orientation for reaction, but they do not bind the oxyanion hole well and the phosphorus is at a longer distance from Oγ than any of FAMEs studied. This is not surprising, as organophosphates are not the natural substrate of E3. It was surprising, however, to find that VX(S) docked into a position with reference parameters (distances) very similar to those observed for FAMEs. This is the most toxic isomer of VX for humans and animals, as it is a potent inhibitor of acetylcholinesterase[30]. This raises the question of whether wild-type E3 can metabolise this compound. Table 4.4 presents a summary of the distances observed for each compound.

	P_{sub}-Oγ	O/S_{sub}-N_{Ala219}	O/S_{sub}-N_{Gly136}	O/S_{sub}-N_{Gly137}
dECP	3.5	5.1	5.2	3.6
Diazinon	3.6	5.4	6.2	4.4
Sarin(S)	4.2	6.2	4.8	3.1
Sarin(R)	3.7	5.8	4.8	4.2
VX(S)	3.3	4.2	3.6	2.8

Table 4.4. *Key parameters of E3-organophosphate complexes.* Distances presented are those between the reacting oxygen (Oγ) of the serine and the

reacting phosphorus atom of the substrate (P_{sub}), and those between the phosphoryl oxygen (or sulphur, in the case of diazinon) of the substrate (O_{sub} or S_{sub}) and the backbone nitrogen of given oxyanion hole residues. Distances in Å.

4.4 Defining the substrate binding pocket of E3

From these analysis and those described in Jackson *et al.*[29], it can be observed that the active site of E3 contains two substrate binding pockets (Figures 4.11 and 4.12): a large one made of residues Trp251, Met308, Phe309, Phe355 and Phe421; and a small pocket that consists residues Phe354, Tyr457, Met460 and Thr472[29].

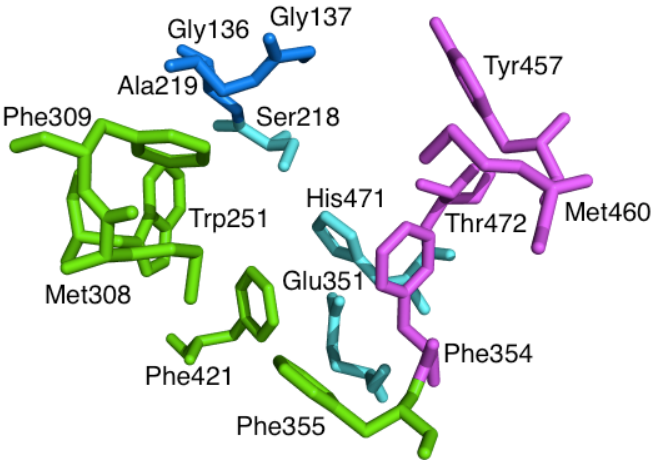


Figure 4.11. *E3 active site and substrate binding pockets.* The large pocket (left) is depicted in green, the small pocket (right) is shown in magenta. Dark blue corresponds to the oxyanion hole, and light blue to the catalytic triad.

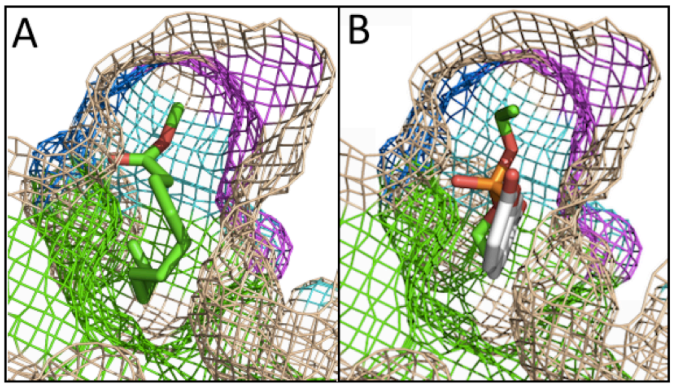


Figure 4.12. *Binding pockets and active site with substrates docked.* **A.** Methyl decanoate. **B.** dECP. The enzyme (represented as mesh) has the large binding pocket coloured in green, the small binding pocket in magenta, the catalytic triad in light blue, and the oxyanion hole in dark blue.

4.5 Molecular Dynamics Simulations of VX(S) bound to E3

There are some limitations to the use of docking algorithms to study substrate binding to E3. The changes in residue conformation necessary to adjust to the presence of the substrate cannot be observed in a rigid docking approach. Furthermore, the persistence over time of the changes introduced by the substrate cannot be quantified, and multiple conformations cannot be assessed. Because of the difficulty in experimentally analysing VX(S) binding to E3, it was sought to better establish whether the substrate binding mode that was observed in the docking pose was stable for a sufficient period of time (15 ns) for initiation of nucleophilic attack by making use of MD simulations. The results show a stable binding, with the phosphorus atom of VX(S) at a distance of 3.9 Å (average) from the side chain oxygen atom of Ser218, with a minimum distance of 3.1 Å and a maximum distance of 5.5 Å during the simulation (see Figure 4.13).

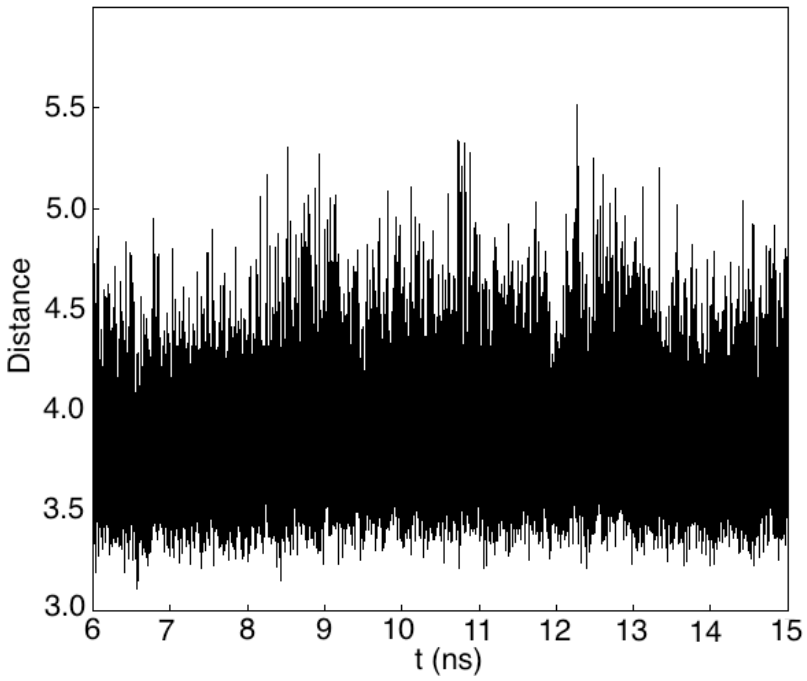


Figure 4.13. *Position of VX(S) along the simulation.* Distance (in Å) between the phosphorus atom of VX(S) and the hydroxyl oxygen of Ser218.

4.6 X-ray crystallographic studies of VX(S) binding by E3

The docking results demonstrated that the toxic isomer of VX, VX(S), bound at the active site of E3 in a productive orientation, while VX(R) did not. Molecular dynamics simulations then established that the binding mode of VX(S) was stable for several ns. Together these computational/theoretical results led to a hypothesis that E3 might be able to bind to and hydrolyse, and therefore detoxify, this dangerous compound. This was tested experimentally by Dr Colin Jackson in collaboration with Dr Florian Nachon from the Centre de Recherche du Service de Santé des Armées, Grenoble, France. Crystals of E3 grown by Dr Jackson were soaked by Dr Nachon in a racemic mixture of VX for ten minutes. X-ray diffraction data were then collected at the European Synchrotron Radiation Facility (Grenoble, France) by Dr Nachon at beamline 14-4. Subsequent data reduction, refinement and analysis of this crystal structure form part of this thesis and are described in the methods section. Data collection and refinement statistics are provided in Table 4.5. These crystals diffracted to 1.75 Å resolution, which is sufficient to observe any modification of the catalytic serine residue due to interaction with VX.

As shown in Figure 4.14, omit difference density at the catalytic serine of E3 is consistent with phosphorylation. Moreover, the adduct clearly has a short sidechain in the x axis and a longer sidechain in the y axis, which is consistent with phosphorylation of the catalytic serine by VX(S), rather than by VX(R). Effectively, refinement of the structure with the VX(S) adduct results in a correct fit of the electronic density around the serine (see Figure 4.15). Thus, this experimental analysis has verified the data from the substrate docking and molecular dynamics simulations, and indicates that E3 may be a potentially useful biotherapeutic in the treatment and detoxification of VX.

Lucilia cuprina αE7-VX	
Data Processing	
Space group	C2221
Cell dimensions (Å) a,b,c	50.58, 102.64, 226.22
Resolution range (Å)	35-1.75 (1.78-1.75)[1]
Total number of reflections	870397 (43888)
Number of unique reflections	59578 (3212)
Multiplicity	14.6 (13.7)
Completeness (%)	99.6 (99.0)
Mean I/σ(I)	15.1 (2.0)
Wilson B factor (Å^2)	24.6
[2] $CC_{1/2}$	0.999 (0.584)
R_{merge}	0.186 (1.704)
Refinement	
R_{work}/R_{free}	0.175/0.208 (0.275/0.291)
Total number of atoms	4813
Number of macromolecules	1
RMSD for bonds (Å)	0.022
RMSD for angles (deg)	2.115
Ramachandran favored (%)	96.6
Ramachandran outliers (%)	0

Table 4.5. *Data processing and refinement data for the X-ray crystal data structure described in this work.* [1]Values in parentheses are for the highest resolution shell. [2]Pearson's correlation coefficient calculated from two half-sets of the data[31,32].

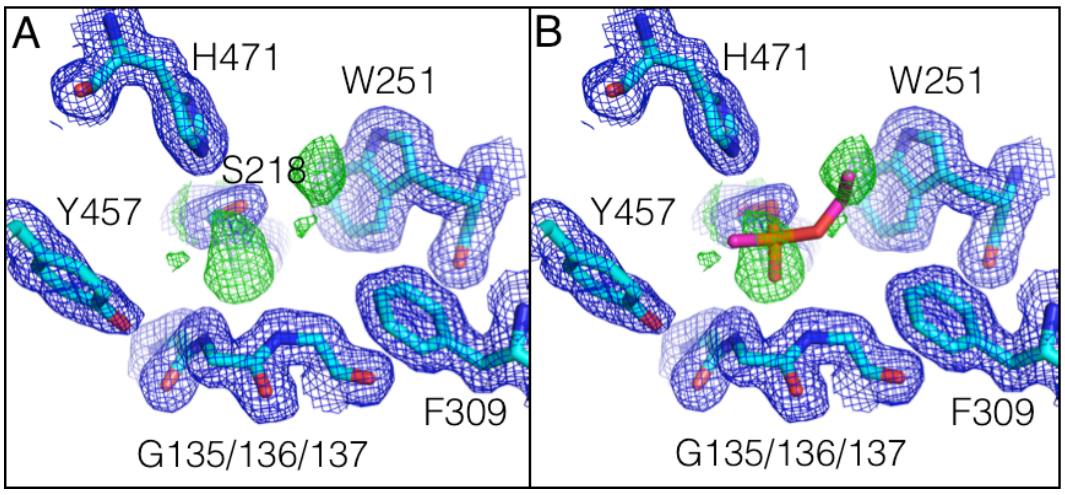


Figure 4.14. *Fitting VX into the active site of the E3 X-ray structure.* **A.** Positive difference density (in green). **B.** VX(S) fits into the green density.

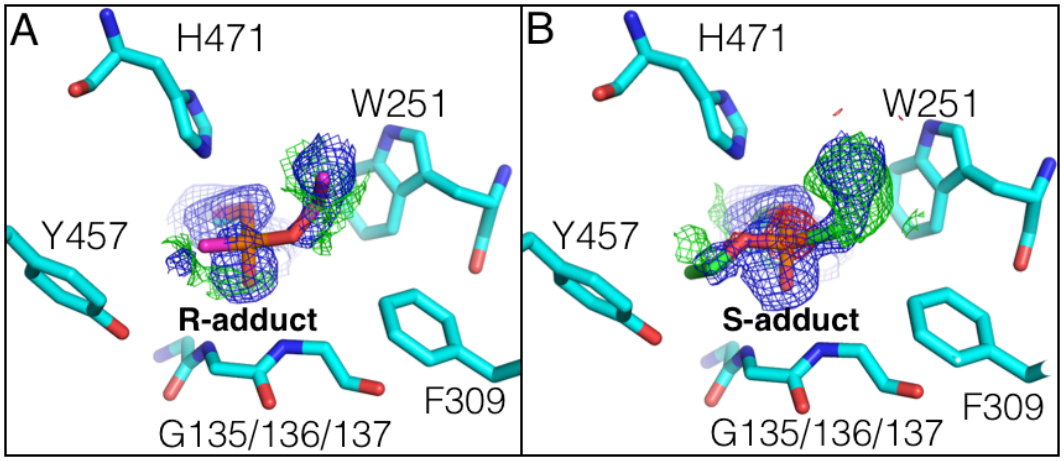


Figure 4.15. *Refinement of the structure with the different isomers of VX.* **A.** Serine-adduct with the VX(R) isomer. **B.** Serine-adduct with the VX(S) isomer.

4.7 Conclusions

Docking techniques have been highly valuable in analysing the way different substrates bind E3, which cannot be done with high throughput experimentally given the difficulty in assaying some substrates. This data lent support to the idea that mid-length chain FAMEs are the physiological substrate of E3. Subsequent experimental work verified this; therefore, a plausible native substrate of E3 has been established for the first time. These techniques also provided structures for simulation and data that predicted the binding of non-physiological substrates. The studies here reported raised the hypothesis that E3 could hydrolyse VX(S) and therefore be useful as a detoxification agent. MD simulations performed with the docked structure of VX(S) confirmed that this binding is stable over a 15 ns period of time. The X-ray structure of E3 phosphorylated with VX(S) presented here then shows that E3 is indeed preferentially phosphorylated by the VX(S) isomer.

4.8 References

(1) Parker, A. G.; Campbell, P. M.; Spackman, M. E.; Russell, R. J.; Oakeshott, J. G. *Pesticide Biochemistry and Physiology* **1996**, *55*, 85.

(2) Birner-Gruenberger, R.; Bickmeyer, I.; Lange, J.; Hehlert, P.; Hermetter, A.; Kollroser, M.; Rechberger, G. N.; Kühnlein, R. P. *Insect Biochemistry and Molecular Biology* **2012**, *42*, 220.

(3) Levot, G. W. *International Journal for Parasitology* **1995**, *25*, 1355.

(4) Wester, M. R.; Johnson, E. F.; Marques-Soares, C.; Dansette, P. M.; Mansuy, D.; Stout, C. D. *Biochemistry* **2003**, *42*, 6370.

(5) Chen, H.; Zhang, Y.; Li, L.; Han, J. G. *Journal of Physical Chemistry B* **2012**, *116*, 10219.

(6) Szklarz, G. D.; Graham, S. E.; Paulsen, M. D. In *Vitamins and Hormones* 2000; Vol. 58, p 53.

(7) O. Trott, A. J. O. *Journal of Computational Chemistry* **2010**, *31*, 455.

(8) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. *Journal of computational chemistry* **2009**, *30*, 2785.

(9) Schröedinger, L. 2010.

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

(11) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. *Journal of Chemical Theory and Computation* **2015**, *11*, 3696.

(12) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *Journal of Computational Chemistry* **2004**, *25*, 1157.

(13) Olsson, M. H. M.; SØndergaard, C. R.; Rostkowski, M.; Jensen, J. H. *Journal of Chemical Theory and Computation* **2011**, *7*, 525.

(14) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. *Journal of Chemical Theory and Computation* **2011**, *7*, 2284.

(15) Bas, D. C.; Rogers, D. M.; Jensen, J. H. *Proteins: Structure, Function and Genetics* **2008**, *73*, 765.

(16) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins: Structure, Function and Genetics* **2005**, *61*, 704.

(17) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926.

(18) Kabsch, W. *Acta Crystallographica Section D: Biological Crystallography* **2010**, *66*, 125.

(19) Diederichs, K.; Karplus, P. A. *Acta Crystallographica Section D: Biological Crystallography* **2013**, *69*, 1215.

(20) Karplus, P. A.; Diederichs, K. *Science (New York, N.Y.)* **2012**, *336*, 1030.

(21) Kaithamanakallam, R. P.; Karunakaran, R.; Srikumar, P. S. *International Journal of Pharmacy and Pharmaceutical Sciences* **2014**, *6*, 662.

(22) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. *Acta Crystallographica Section D: Biological Crystallography* **2011**, *67*, 235.

(23) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. *Acta Crystallographica Section D: Biological Crystallography* **2011**, *67*, 235.

(24) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallographica Section D: Biological Crystallography* **2010**, *66*, 486.

(25) Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Crystallographica Section D: Biological Crystallography* **2010**, *66*, 213.

(26) Parker, A. G.; Russell, R. J.; Delves, A. C.; Oakeshott, J. G. *Pesticide Biochemistry and Physiology* **1991**, *41*, 305.

(27) Spackman, M. E.; Oakeshott, J. G.; Smyth, K. A.; Medveczky, K. M.; Russell, R. J. *Biochemical Genetics* **1994**, *32*, 39.

(28) Gilbert, L. I.; Chino, H. *Journal of Lipid Research* **1974**, *15*, 439.

(29) Jackson, C. J.; Liu, J. W.; Carr, P. D.; Younus, F.; Coppin, C.; Meirelles, T.; Lethier, M.; Pandey, G.; Ollis, D. L.; Russell, R. J.; Weik, M.; Oakeshott, J. G. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 10177.

(30) Benschop, H. P.; De Jong, L. P. A. *Accounts of Chemical Research* **1988**, *21*, 368.

(31) Diederichs, K.; Karplus, P. A. *Acta Crystallographica Section D: Biological Crystallography* **2013**, *69*, 1215.

(32) Karplus, P. A.; Diederichs, K. *Science* **2012**, *336*, 1030.

Chapter 5

Molecular dynamics studies of wild-type and variant *Lucilia cuprina* E3

5.1 Introduction

Computational methods based on molecular mechanics, i.e. not accounting for quantum effects, are valuable tools for studying proteins because the number of atoms in biomolecules makes them impossible to study using quantum chemistry for the whole molecule[1,2]. The ligand docking techniques used in Chapter 4 permit analysis of the binding of possible substrates and inhibitors at the active site of an enzyme[3], while molecular dynamics (MD) simulations can provide insight into the motions that affect substrate binding and catalysis[3].

It is well known that the molecular motions of enzymes are essential to their function (see for example Singh *et al.*[4] and references therein). In this work, MD simulations were used to study the conformational landscape of wild-type E3, the naturally occurring E3-Generation 1 (E3-G1; Gly137Asp) mutant, and a lab-evolved E3-Generation 4 (E3-G4; Gly137Asp, Lys306Met, Met308Val and Ser470Gly) mutant (see Figure 5.1). One benefit of MD simulations is the ability to extract large amounts of data about the position of different parts of the protein over time, and it has been used extensively in the analysis of enzyme function[5]. In the present work, attention was paid to the influence of substrates, phosphorylation and mutations on the conformational sampling of key amino acid side-chains within the active site. E3-G4 is a laboratory-evolved variant that was generated by Dr C. J. Jackson and Dr Jian-Wei

Liu at the CSIRO (Black Mountain, Canberra) to have higher activity with the fluorescent organophosphate diethyl umbelliferyl-phosphate (DEUP). The steady state, and pre-steady-state, kinetic parameters of E3, E3-G1 and E3-G4 are shown in Table 5.1. These measurements were performed by Mr Galen Correy within the Jackson laboratory as part of this project, and show how G1 can metabolise dECP an order of magnitude faster than the wild type, while G4 provides an improvement of almost double that rate.

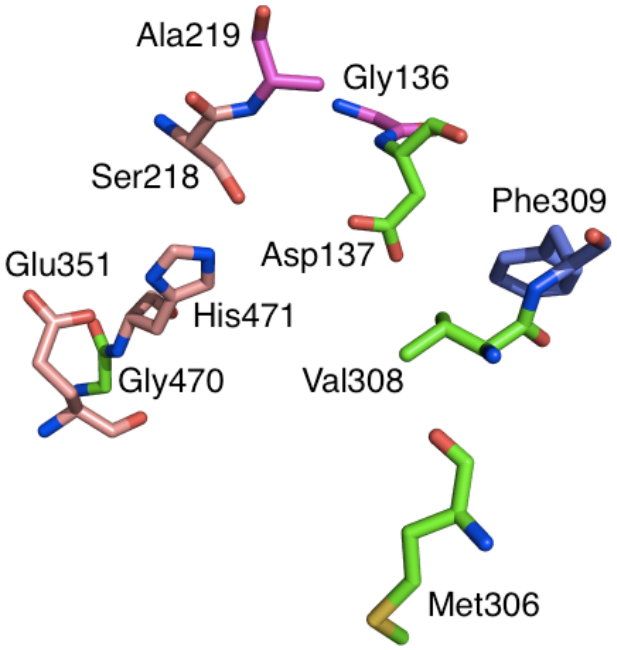


Figure 5.1. *Active site of E3-G4.* The mutated residues are highlighted in green, the catalytic triad in orange and the oxyanion hole in magenta.

Variant	k_3 (s^{-1})	k_{cat} $\times 10^{-4}$ (s^{-1})	Km (μM)	k_{cat}/K_M ($M^{-1}.s^{-1}$)
E3	1.8 ± 0.2	0.7 ± 0.1	< 1.5	>50
E3-G1 (G137D)	1.3 ± 0.1	11±1	7±2	160±50
E3-G4 (G137D, K306M, M308V, S470G)	0.45 ± 0.03	56±1	20±3	280±40

Table 5.1. Kinetic parameters for the hydrolysis of diethylumbelliferyl phosphate at pH 7.5 by *Lcα*E7 variants. (mean ± SEM n=4).

Previous works have shown that molecular dynamics simulations can, in combination with analysis of experimentally determined atomic

displacement parameters (B-factors) from crystal structures, provide substantial insight into the structural and conformational changes that occurs in highly mobile regions of proteins, such as active site loops[6]. Normal Mode Analysis (NMA) is commonly performed on MD trajectories to analyse large-scale motions of a protein based on the information the simulation provides of the movement of all atoms and residues. This has been used to describe changes caused by mutations[7,8], as well as analysing the evolution of enzyme function[9], identification of gating (opening/closing) motions[10], and to describe slow-timescale (ms) motions and conformational changes that cannot be easily observed through short-timescale MD simulations[11]. Molecular dynamics simulations are used in the present chapter to study the conformational space and motions of amino acid side chains of E3 and how mutations and substrates affect these.

5.2 Methods

The initial coordinates for the MD simulations were taken from the X-ray structures with PDB codes 4FNG (free wild type E3) and 4FNM (phosphorylated wild type E3) and the X-ray structure of the naturally occurring Gly137Asp mutant (PDB ID 5C8V). No crystal structure of E3-G4 was available. For this structure, three additional point mutations (Lys306Met, Met308Val and Ser470Gly) were made by Dr Colin Jackson using default protocols, beginning with the E3-G1 structure, using the Swiss-Model server[12] (www.swissmodel.expasy.org). The initial structure of wild-type E3 with dECP bound was that obtained by docking procedures as described in Chapter 4. Protonation states of ionisable residues were selected according to pKa values determined by PROPKA[13-16] and by visual inspection of hydrogen bond networks. The AMBER12

software package[17] and its tools were used to prepare, run and analyse all simulations. The module *tleap* of AMBER12 was used to prepare the coordinate and topology files of each system, consisting of the protein solvated in a truncated octahedron box of TIP3P[18] water in which the walls of the box were no closer than 10 Å from the solute. The ff12SB[19] force field was used for the protein and GAFF[20] for the substrate. Where there was a non-standard residue – the phosphorylated serine of 4FNM – its parameters were derived to be consistent with ff12SB using the module *antechamber* of AMBER and the *resp* fitting procedure for charges. The solvated system was energy minimised with restraints on the backbone to reduce bad contacts due to the position of hydrogen atoms and water molecules. The system was then carefully heated in steps of 50 K temperature increases for 10ps each (to a total of 50ps) with a harmonic restraining potential of 10.0 kcal/molÅ^2 using SHAKE[21] (with the x-ray or docked structure as reference) in the NVT ensemble. Heating was carried out to a set of final temperatures of 296 K, 298 K, 300 K, 302 K and 304 K. Each of these was used as the starting point of one of a set of five trajectories in which the SHAKE algorithm as implemented on AMBER12 was used to constrain hydrogens, with a tolerance of 0.00001 Å. The starting structure of each trajectory was then equilibrated in the NVT ensemble at 300 K and the production runs were run in the NPT ensemble (with isotropic position scaling) at 300 K, using the *sander* module, for a total of 15 ns each, of which the first 5 ns were discarded as equilibration time. A Langevin[22] thermostat as implemented in AMBER was used with a time constant of 1.0 ps and compressibility of 44.6×10^{-6} bar^{-1}. The AMBER module *ptraj* was used to extract simulation data for analysis. In the analysis of hydrogen bonds, the cutoff distance for a bond to exist was 3.5 Å between donor and acceptor, and the cutoff for the angle between acceptor, hydrogen and donor was 135°. Analysis was done with the *tleap*

module of amber and visualisation was done using PyMOL[23] and VMD[24]. Normal Mode Analysis was performed with the NMWizard tool of VMD.

5.3 Investigating the conformational landscape of amino acid side chains in E3 and its variants

The naturally occurring mutant of E3-Gly137Asp (G1) confers to *L. cuprina* resistance to organophosphorus pesticides. The mutant E3-G4, evolved in the laboratory, catalyses the same reaction as G1, but faster (Table 5.1). In this section, the conformational landscape of the residues that together comprise the active site is analysed in order to draw conclusions about how these mutants work and what sets them apart. The dihedral angles χ_1 and χ_2 can describe the position of the side chain of amino acid residues. All angles reported are in degrees unless otherwise specified. This provides a useful means to reduce the three dimensional conformation of a side chain to numerical values that can easily be analysed or illustrated. χ_1 is defined as the dihedral N-Cα-Cβ-Cγ for most amino acids except a few. Of the residues studied here, it is worth mentioning that serine lacks a Cγ and therefore χ_1 becomes N-Cα-Cβ-Oγ, and so does threonine, in which χ_1 is defined as N-Cα-Cβ-Oγ1. Methionine has a sulphur atom instead of Cγ, so χ_1 is N-Cα-Cβ-Sδ. χ_2 changes according to the residue, but it can be defined as 'one atom down the chain' with respect to χ_1 (see Figure 5.2).

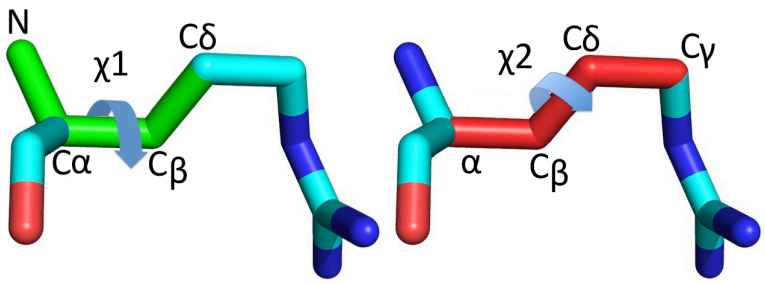


Figure 5.2. *χ_1 and χ_2 exemplified for an arginine residue.* To the left (in green) the atoms that define the torsion χ_1. To the right (in red) the atoms that define χ_2.

For the rest of this chapter, G0 will refer to the system of wild type E3 with nothing bound (PDB code 4FNG), G0+dECP is a system of wild type E3 with dECP bound in a Michaelis complex, G0pho is a system comprising E3 phosphorylated by dECP (PDB code 4FNM), G1 is the naturally occurring E3-Gly137Asp mutant, and G4 is a laboratory-made E3 with the mutations Gly137Asp/Lys306Met/Met308Val/Ser470Gly.

5.3.1 The conformational stability of the catalytic triad across G0 and variants

In G0, the side chains of the catalytic triad residues adopt very specific conformations, which are stable throughout the course of the simulation. Likewise, when compared with the simulations of the other systems (G0-dECP, G0pho, G1 and G4) it is clear that the catalytic triad is not affected by any of these mutations, nor by the presence of substrate or intermediate. Glu351, which coordinates to His471 in the catalytic triad, has only one side chain conformation in all systems, where $\chi_1 \approx -50$ and $\chi_2 \approx -65$ (see Figure 5.3). This is most likely due to the hydrogen bond it forms with His471-Nδ, which exists in over 80% of the simulation time (see Table 5.2 in the next section) and would stabilize the conformation of this residue. The side chain of His471 also has a well-defined conformation ($\chi_1 \approx 10$; $\chi_2 \approx 60$) with almost no variability throughout the systems studied (Figure 5.4). Ser218, the nucleophilic group that attacks the substrate, also has a very stable conformation with little change. As the side chain of a serine is short, only the dihedral $\chi 1$ exists in this residue. Indeed, Ser218 spends all of the simulation time around the value of 180 degrees. This is true for all systems except G1, where two short-lived jumps to the region between -50 and -100 are observed (Figure 5.5).

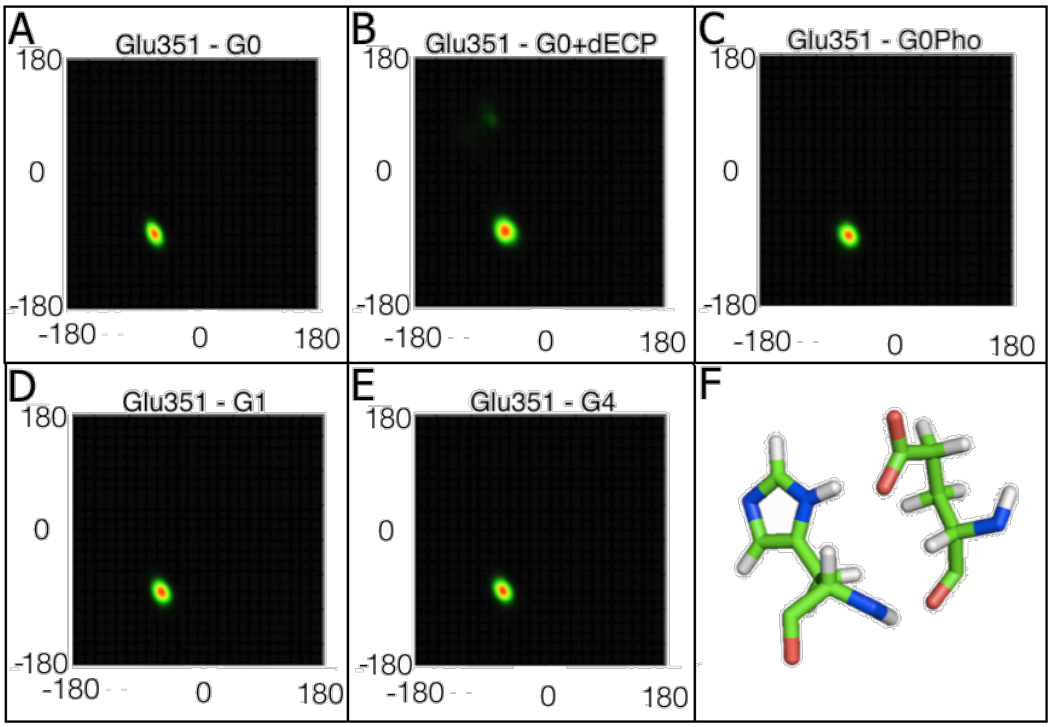


Figure 5.3. *χ_1/χ_2 plot of Glu351.* *x* axis = χ_1, *y* axis = χ_2. **A.** G0. **B.** G0+dECP. **C.** G0pho. **D.** G1. **E.** G4. **F.** A snapshot of Glu351 and His471 during the simulation, showing the close contact between the histidine hydrogen and a glutamate oxygen.

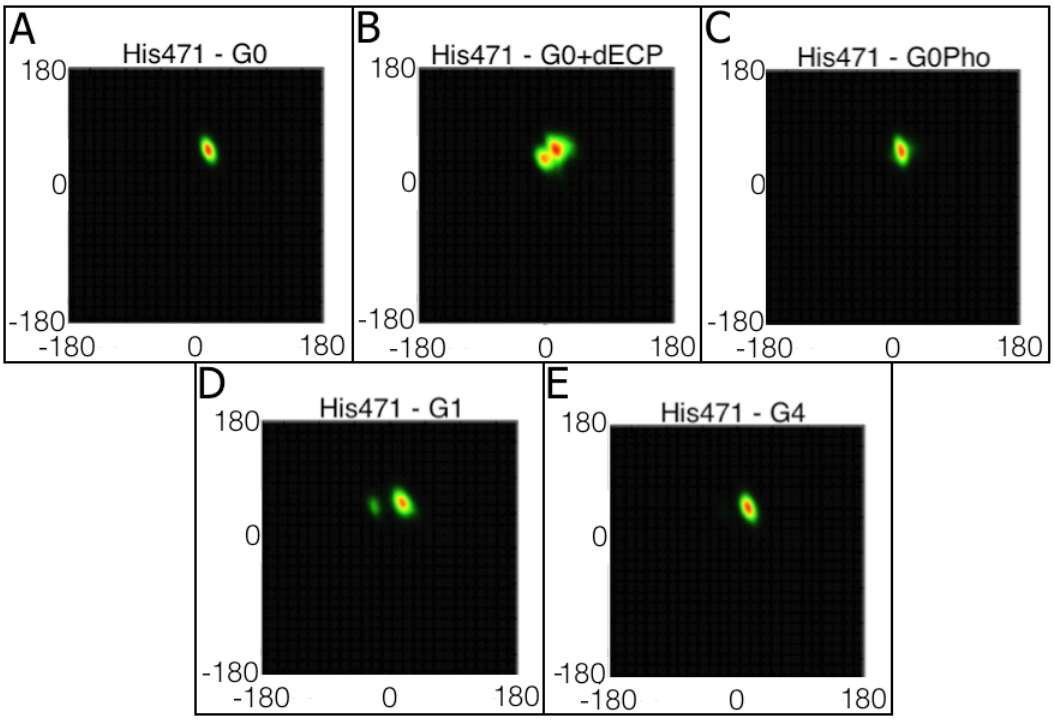


Figure 5.4. *χ_1/χ_2 plot of His471. x* axis = χ1, *y* axis = χ2. **A.** G0. **B.** G0+dECP. **C.** G0pho. **D.** G1. **E.** G4.

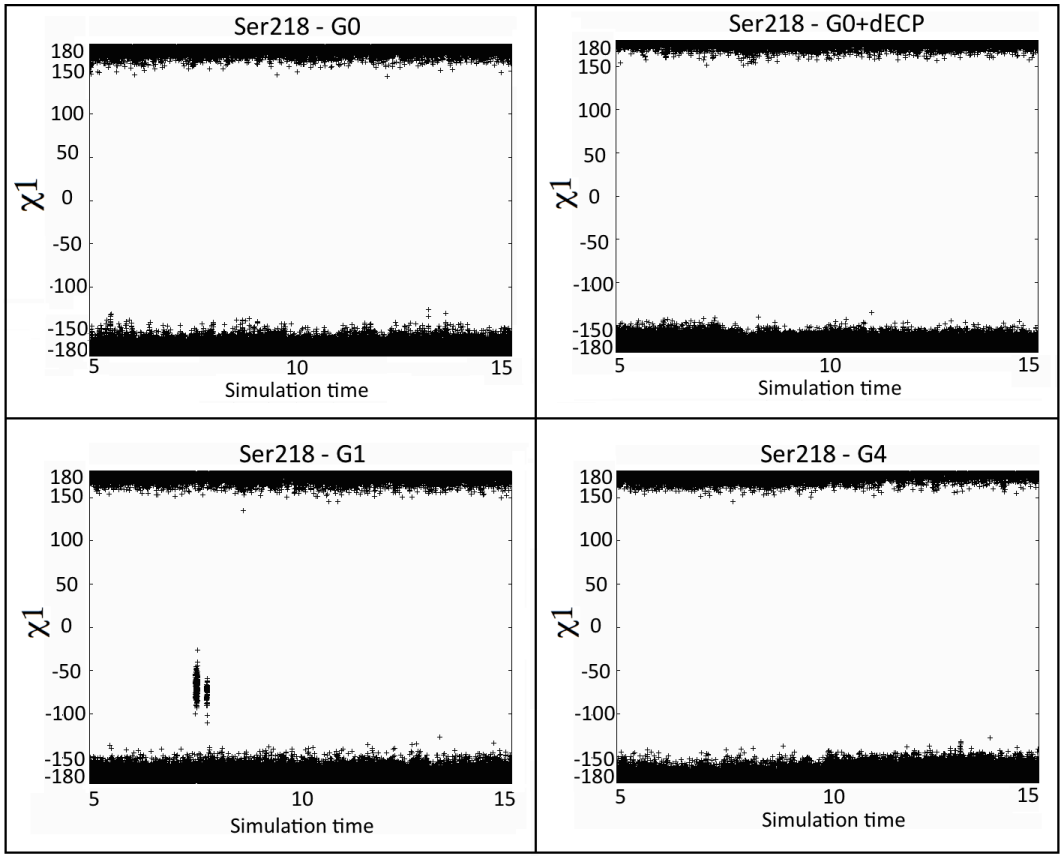


Figure 5.5. *Dihedral χ_1 of Ser 218.* Angles in degrees. Simulation time in ns. These plots show that the catalytic serine residue is highly organised as part of the catalytic triad in both wild-type and variant E3.

5.3.2 Analysis of hydrogen bond network of the catalytic triad

Inspection of the hydrogen bond network around the catalytic triad during the MD simulations shows that stable contacts persist during substrate binding and phosphorylation (see Figure 5.6 for a definition). Other contacts are observed that are disrupted by the presence of the substrate (See Table 5.2). The contact between the hydroxyl hydrogen (Hγ) of Ser218 and Nδ of His471 exists over a relatively low proportion (about 20%) of the simulation time in all free enzymes (G0 and mutants G1 and G4). As expected, substrate binding increases the persistence of this contact – to 65% of the simulation time. There is no hydrogen bond between Oγ and Hγ (the latter is the proton donated to His471 by Ser218 when phosphorylation occurs) in the phosphorylated enzyme. Glu351

makes multiple hydrogen bonds where its two oxygen atoms (OD) act as acceptors. It forms a contact with His471(Hε) that exists throughout the majority of the simulation time (~80% of the total) in all systems, regardless of the presence of the substrate or mutations. Glu351(OD) also accepts a hydrogen bond from Ser244(Oγ). This contact is present virtually all of the simulation time in all systems except for G0+dECP. This is because a new contact, namely Ser244(Hγ) – Glu474(OD), is formed in the presence of dECP which displaces the former. The carboxyl oxygen atoms of Glu351 also make a persistent hydrogen bond with Asn347(Hδ), which exists during more than 90% of the simulation time in all systems. Another persistent hydrogen bond made by Glu351(OD) is that with Thr348(Hγ), which is not affected by mutations, phosphorylation or substrate binding. Most of the active site contact network is unaltered by mutations, while binding dECP stabilizes a contact between Ser218 and His471 (which will act as a general base and abstract a proton from Ser218).

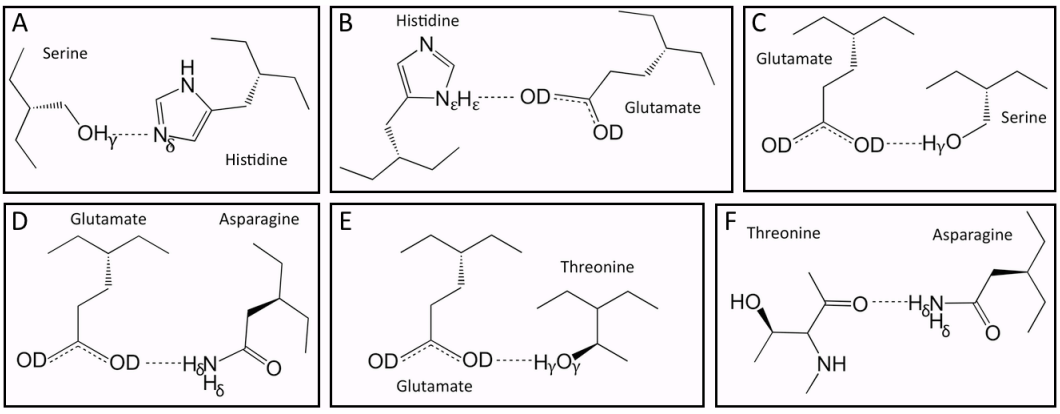


Figure 5.6. *Nomenclature of atoms in amino acid residues.* Atom nomenclature in this thesis follows that used in AMBER[17].

	G0	G0+dECP	G0pho	G1	G4
Ser218(Oγ) - His471(Hδ)	16	65	0	24	22
Glu351(OD) - His471(Hε)	83	83	80	76	90
Glu351(OD) - Ser244(Hγ)	96	40	97	97	98
Glu474(OD) - Ser244(Hγ)	0	58	0	0	0
Glu351(OD) - Asn347(Hδ)	94	91	93	95	95
Glu351(OD) - Thr348(Hγ)	86	90	88	87	89

Table 5.2. *Hydrogen bonding at or near the active site.* Percentage of the simulation time that a hydrogen bond exists between donor-acceptor pairs. For glutamate residues, this is the sum of the two OD atoms.

5.3.3 Analysis of the conformational variability of Asp137

The crystal structure of the Gly137Asp mutant of E3 was solved by Dr Peter Mabbit from the Jackson Laboratory (Table 3.1 in Chapter 3). Unexpectedly, the conformation of Asp137 was such that substrate binding and phosphorylation would not be possible without significant rearrangement of this residue. This is shown in Figure 5.7.

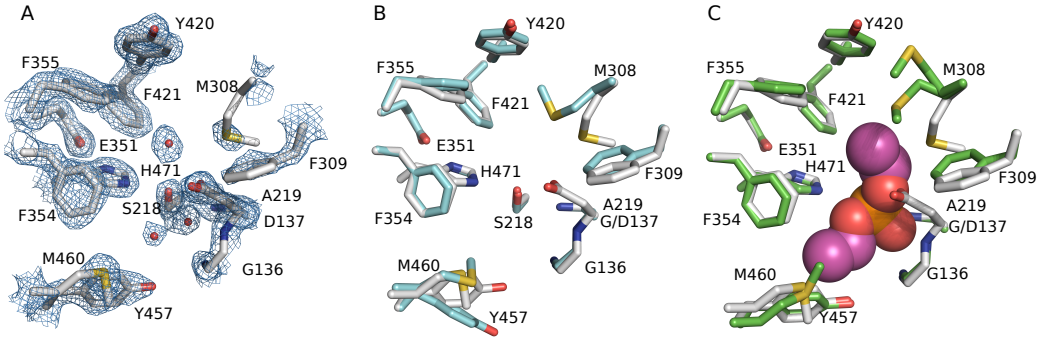


Figure 5.7. *The Gly137Asp mutation in the active site of E3 (from Mr Galen Correy).* **A:** Close up of the *Lc α* E7-Gly137Asp active site with catalytic triad (Ser218, Glu351 His471), oxyanion hole (Gly136, Asp137, Ala219), and binding pocket residues (Met308, Phe309, Phe354, Phe355, Tyr420, Phe421, Met460, Tyr457) shown as grey sticks. The 2|Fo|-|Fc| map centred on these residues is shown as blue lines, contoured at 1.5 σ. **B:** Comparison of the binding pocket of G0 (blue sticks) and G1 (white sticks). **C**: Comparison of the binding pocket of G0Pho (green sticks) and G1 (grey sticks).

In the E3-Gly137Asp crystal structure, the carboxylate group of Asp137 is rotated such that it actually blocks the substrate binding site for OPs (Figure 5.8). The MD simulations performed revealed that in G1 and G4,

the side chain of Asp137 most often samples this crystallographic conformation, at $\chi 1$ = -67; $\chi 2$ = -20, although its carboxyl headgroup rotates frequently, sampling $\chi 2$ values from approximately 20 to –80 (Figure 5.8). To allow ideal binding of OPs and to minimise a steric clash, this carboxylate group should rotate to $\chi 2$ = -100. Interestingly, in the enhanced variant E3-G4, we observe significantly greater sampling of $\chi 2$ values closer to –100 (especially from –75 to –90 and –105 to –120).

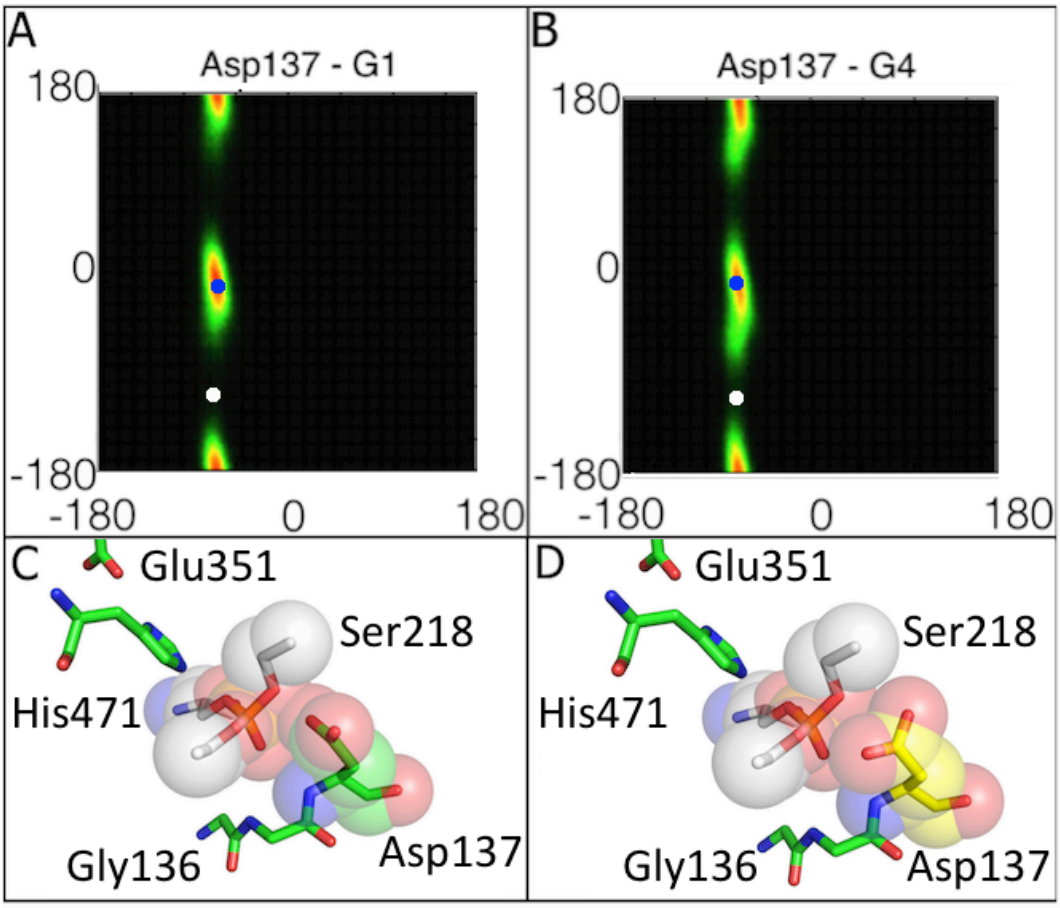


Figure 5.8. *χ_1/χ_2 plot of mutant Asp137.* **A.** G1. **B.** G4. The blue point in A and B represents the (closed) conformation of this residue in the crystal structure of G1, the white point is the open conformation observed in the simulations. **C.** Closed conformation of Asp137, shown with phosphorylated Ser218. **D.** Open conformation of Asp137. Residues and substrate shown in sticks, volume represented in spheres.

These results suggest that although Asp137 clearly enhances activity by introducing a new general base into the active site (which was discussed in detail in Chapter 3), the conformational sampling of active site residues

is far from optimal, as they frequently adopt conformations not suitable for substrate binding. Indeed, this is consistent with the observation that the K_M of G1 is at least 4.5-fold higher than G0, which is consistent with reduced substrate binding. As shown in Table 5.3, the G4 mutant displays a significantly faster rate of dephosphorylation than G0 (56×10^{-4} s^{-1} *vs.* 11×10^{-4} s^{-1}, yet the mutations in this variant are remote from the active site. Molecular dynamics simulations on G4 were used to investigate whether the increase in rate for this variant could be related to improved conformational sampling of Asp137. As shown in Figure 5.8, although the sampling of this residue in G4 is similar to that observed in G1, a change is observed in its conformational landscape by which sampling of a more open is significantly increased. Thus, it seems that some neighbouring residues, potentially Phe309 as shown in Figure 5.1, could be constraining Asp137 in a sub-optimal conformation and this constraint could be partially relieved by the additional mutations in E3-G4 (Lys306Met, Met308Val and Ser470Gly)

5.3.4 Conformational diversity within the substrate binding pockets of E3

Based on the substrate docking results presented in Chapter 4, for this analysis three regions of the substrate binding site were defined (see Figure 5.9): the oxyanion hole (Gly136, Gly/Asp137 and Ala219), the small binding pocket (Phe354, Tyr457, Met460 and Thr472) and the large binding pocket (Trp251, Met308, Phe309, Phe355 and Phe421). All of the residues of the binding pockets that have side chains long enough to define $\chi 1$ and $\chi 2$ were analysed.

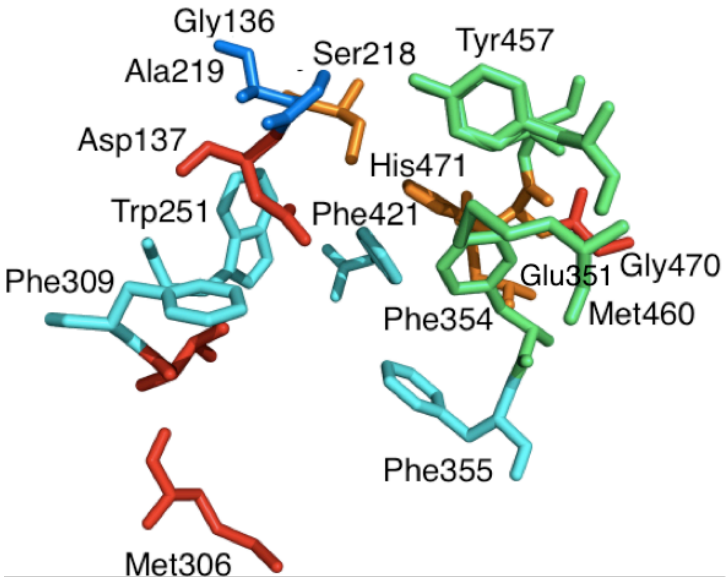


Figure 5.9. *Mutations in the active site of E3-G4.* The active site cavity can be observed, with the small binding pocket (green), the large binding pocket (light blue), the catalytic triad (orange) and the oxyanion hole residues (dark blue). The mutated residues Asp137, Met306, Val308 and Gly470 are depicted in red.

5.3.4.1 Conformational diversity within the small binding pocket

The small binding pocket residues undergo changes in terms of the conformational space that is sampled and substrate binding and/or mutations dictate these changes. The side chain of has only one conformation in G0, centred on χ_1 = 170; χ_2 = 80. This does not change significantly upon binding of dECP, although a new conformation, which is centred on χ_1 = -70; χ_2 = 80, is infrequently sampled. This new conformation is sampled more often in the two mutants, G1 and G4, although it is still less preferred than χ_1 = -160; χ_2 = 80. Phosphorylation does not alter the conformation of this residue. See Figure 5.10. These results suggest that the mutations (Gly137Asp, Lys306Met, Met308Val and Ser470Gly) provide additional freedom/flexibility for Phe354 to move.

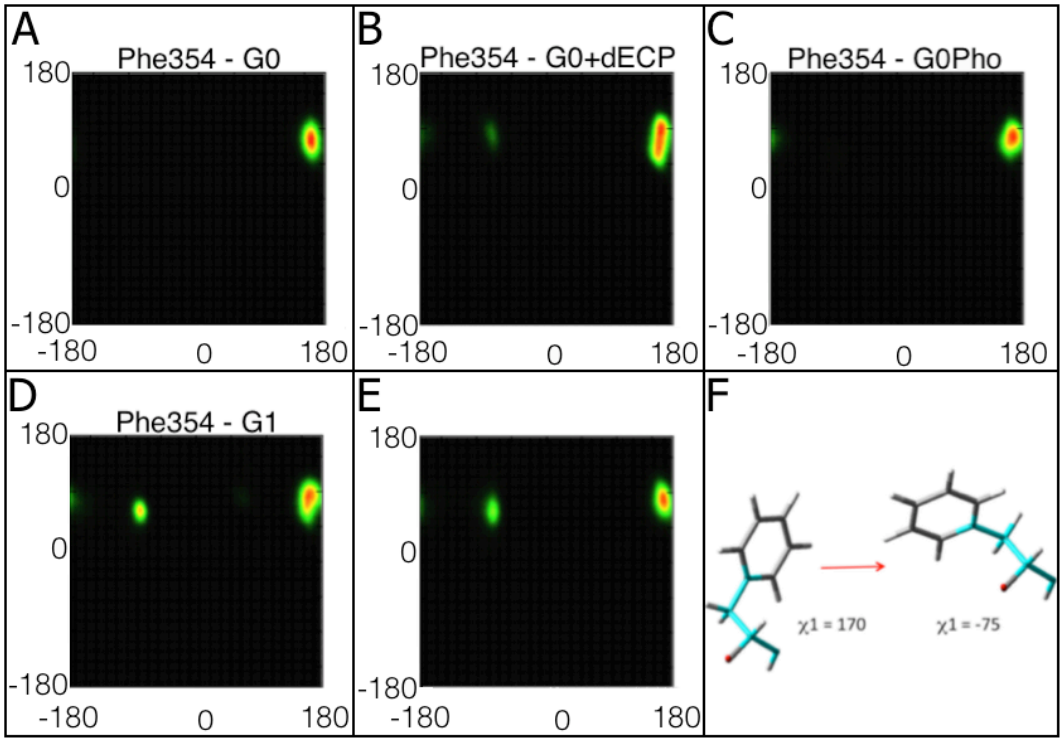


Figure 5.10. *χ1/χ2 plot of Phe354.* The side chain of this residue has one main conformation, a second much less favoured one appears when dECP binds and in the two mutants. **A.** G0. **B.** G0+dECP. **C.** G0pho. **D.** G1. **E.** G4. **F.** A snapshot that shows the affected dihedral angle.

The side chain of Tyr457 adopts one unique conformation with centre on χ_1 = -80; χ_2 = -80 in G0. The tyrosine ring rotates (180°) around a symmetric plane to the position χ_2 = 100, however this rotated conformation is only present a small fraction of the time (Figure 5.11). Interestingly, the centre point between these two configurations is sampled, which is consistent with slight rotation of the tyrosine side chain. This is particularly noticeable when dECP binds, as compared with G0, especially in the regions χ_1 = 60-90 and χ_2 = -100-120. In other words, phosphorylation appears to shift the equilibrium to favour a slightly rotated conformation of this residue.

This analysis generated a hypothesis that was tested experimentally by Mr Galen Correy from the Jackson Laboratory. Namely, that organophosphate binding by E3 involves some 'induced-fit' in which a pre-existing conformation of Tyr457 is stabilized by the substrate. Based

on this analysis, the previously solved[25] structure of apo-G0 was investigated in more detail. This revealed that Tyr457 does indeed exist in two main orientations, as predicted by the MD (see Figure 5.12). These two orientations effectively 'open' and 'close' the active site of E3. Such movements would be consistent with a model in which the apo-E3 is predominantly 'open', maximizing substrate binding, but once substrate binds, Tyr457 rotates to stabilize it and maximises the rate of phosphorylation. To test this, Mr Galen Correy made a Tyr457Ala mutant of E3-G0. As shown in Table 5.3, this variant has significantly lower affinity for organophosphates (as shown by the increased dissociation constant), as well as a significant reduction in the rate of phosphorylation. Thus, the conformational dynamics of E3, within the binding site, are surprisingly sophisticated and contribute to its high affinity for organophosphates.

Interestingly, in the G1 and G4 variants, the landscape is changed slightly, with the more closed form being more commonly sampled, as in the phosphorylated enzyme. It is difficult to predict (or test) exactly what the effects of these changes in the conformational landscape of this residue might involve.

	K_d (nM)	k_2 (s^{-1})
WT	25 ± 3	0.43 ± 0.06
Y457A	120 ± 20	0.26 ± 0.02

Table 5.3. *Kinetic constants for the dissociation constant and rate of phosphorylation of E3 and E3-Tyr457Ala by the organophosphate diethylumbelliferyl phosphate.* K_d (dissociation constant) and k_2 (phosphorylation rate constant) were determined by a double-reciprocal method. Standard errors are given (N=6). (Data provided by Mr Galen Correy).

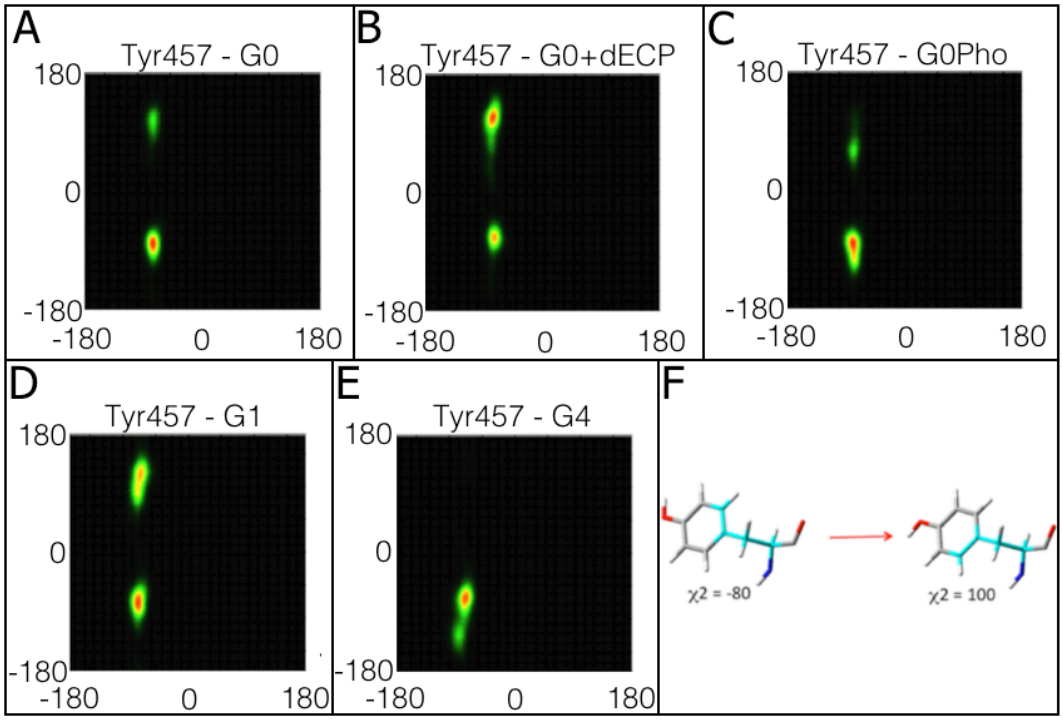


Figure 5.11. *χ_1/χ_2 plot of Tyr457.* This residue has only one side chain conformation, with the tyrosine ring rotating around its axis. **A.** G0. **B.** G0+dECP. **C.** G0pho. **D.** G1. **E.** G4. **F.** Depiction of the rotating angle (cyan).

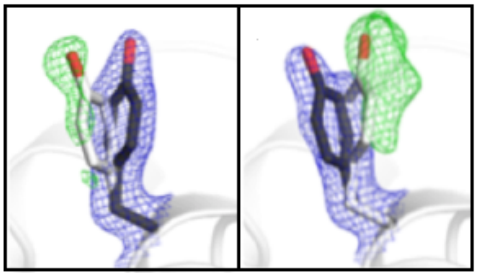


Figure 5.12. *Conformational heterogeneity in the apo-enzyme active site (from Mr Galen Correy)* . The two Tyr457 conformations as predicted by qFit. The modelled conformations (black sticks) are shown with the 2mFo-dFc density in mesh (blue) representation (contoured at 1 σ). Alternative conformations are depicted in white, and their mFo-DFc difference density is shown as mesh (green) (contoured at 3 σ).

Methionine residues are highly mobile because of their long, linear side chain. Met460 takes mainly one configuration space in the dynamics of the free wild type enzyme (G0), with centre on χ_1 = -70 and χ_2 = 180. Another orientation, present but much less preferred, is that centred at χ_1 = 180; χ_2 = 180. With dECP bound, at χ_1 = 180; χ_2 = 180 becomes the most

sampled space of the Met460 side chain, together with that centred on χ_1 = 180; χ_2 = -90. In G0pho, χ_1 = 180; χ_2 = 180 is the preferred configuration. In G1, the most sampled conformations for this side chain are those of G0 together with those of G0-dECP (which also gives the hint that this residue is fairly mobile). A new configuration with centre on χ_1 = 180; χ_2 = -90 appears in G0+dECP and the G1 mutant, although it is not visited frequently. In G4 Met460 is mostly in the configuration with χ_1 = 180; χ_2 = 180, a strong difference with respect to G0. There is only one more conformation sampled by this residue (albeit with lower probability), χ_1 = 65; χ_2 = 175. See Figure 5.13. Thus, in the case of Met460, there is a significant change in the average conformation both as a result of substrate binding and due to the mutations that increased activity. Specifically, the conformation at at χ_1 = 180; χ_2 = 180, which dominates when the substrate/intermediate is present, is enriched at the expense of the original conformer at χ_1 = -70 and χ_2 = 180.

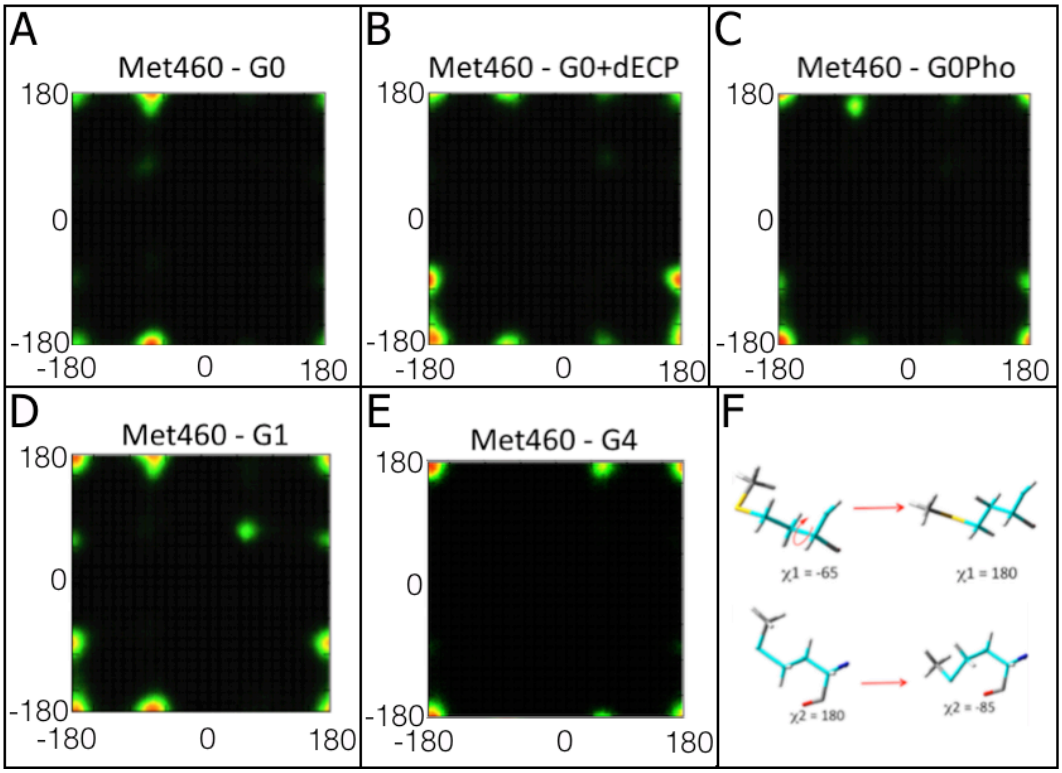


Figure 5.13 *χ_1/χ_2 plot of Met460.* **A.** G0. **B.** G0+dECP. **C.** G0pho. **D.** G1. **E.** G4. Panel F shows the two main conformations of this residue.

5.3.4.2 Conformational diversity within the large substrate binding pocket

Large pocket residue Trp251 is also very stable with one single conformation (χ_1 = 50; χ_2 = -95) in all the systems studied (Figure 5.14). Met308 (Figure 5.15) is preferentially in the regions centred on χ_1 = 180; χ_2 = 180 and χ_1 = -70; χ_2 = 180 in G0. There are some scarcely populated areas with centre on χ_1 = -70; χ_2 = -60 and χ_1 = -170; χ_2 = 70. The configuration of Met308 is slightly modified by the binding of dECP to E3. The two most favoured regions remain the same as in G0, but that around χ_1 = -70; χ_2 = -60 becomes more relevant. Interestingly, this is region is favoured even more in G1, although the most important conformation is still χ_1 = -70; χ_2 = 180 as in G0. There only change in the phosphorylated system is that the less favoured regions of G0 are sampled even less frequently in this system, which indicates more rigidity of this residue in G0pho. As a reminder, Met308 is a valine in the G4 mutant.

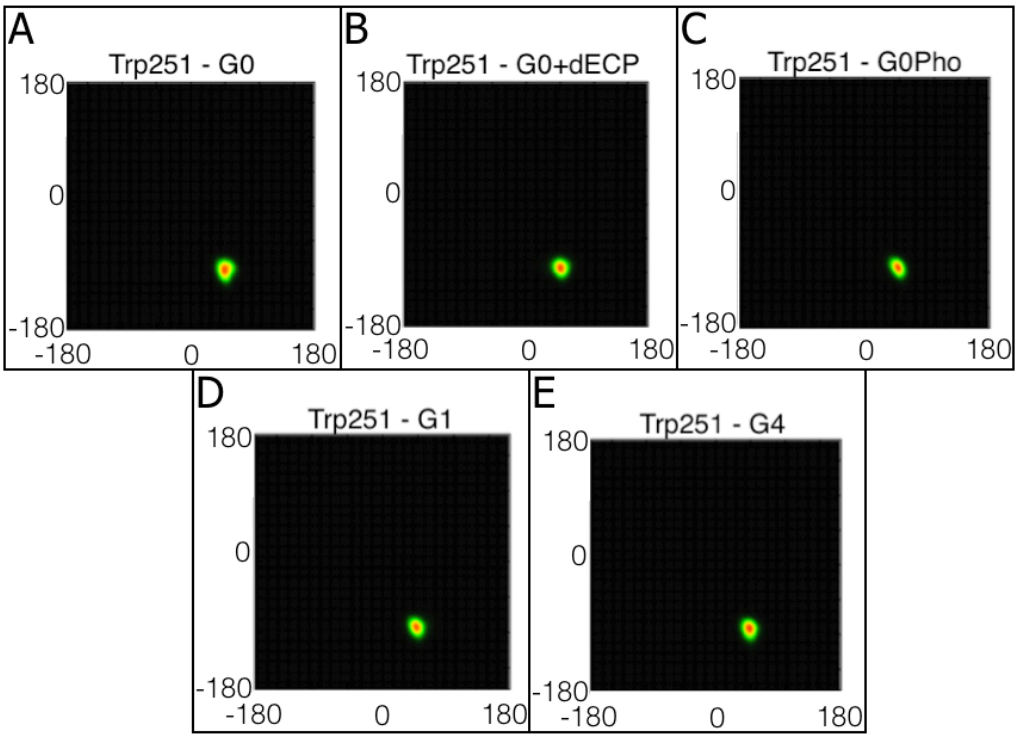


Figure 5.14. *χ_1/χ_2 plot of Trp251.* This residue has only one conformation, stable throughout the systems tested.

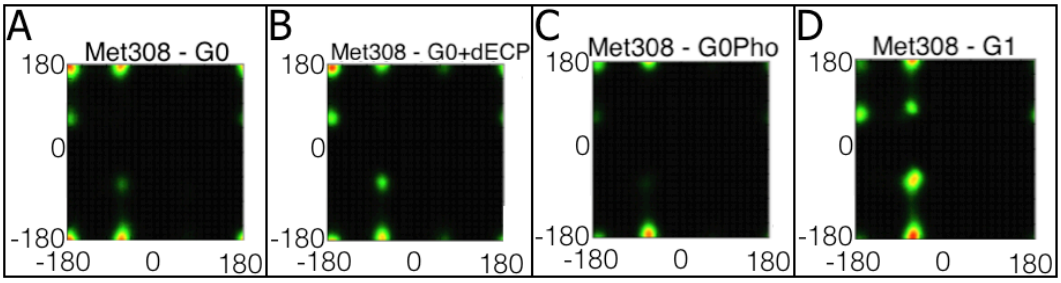


Figure. *χ_1/χ_2 plot plot of Met308.* The preferred configuration changes when dECP binds. A new conformation appears in the mutant G1. In G4 this residue is a valine instead.

Phe309 (see Figure 5.16) has one preferred conformation in G0, that with centre on $\chi_1 = 40$; $\chi_2 = -100$, and another (less favoured) one with centre on $\chi_1 = -50$; $\chi_2 = -70$. This latter conformation is the only one present when dECP is bound. It is also the preferred conformation in the G0pho, although the position $\chi_1 \approx 40$; $\chi_2 \approx -100$ is also present. In G1, Phe309 behaves the same way as it does in the phosphorylated enzyme. In G4, the side chain of Phe309 adopts the same conformation as it does in the dECP-bound enzyme, which is very similar to that seen in G0pho and G1, namely $\chi_1 \approx -50$; $\chi_2 \approx -70$.

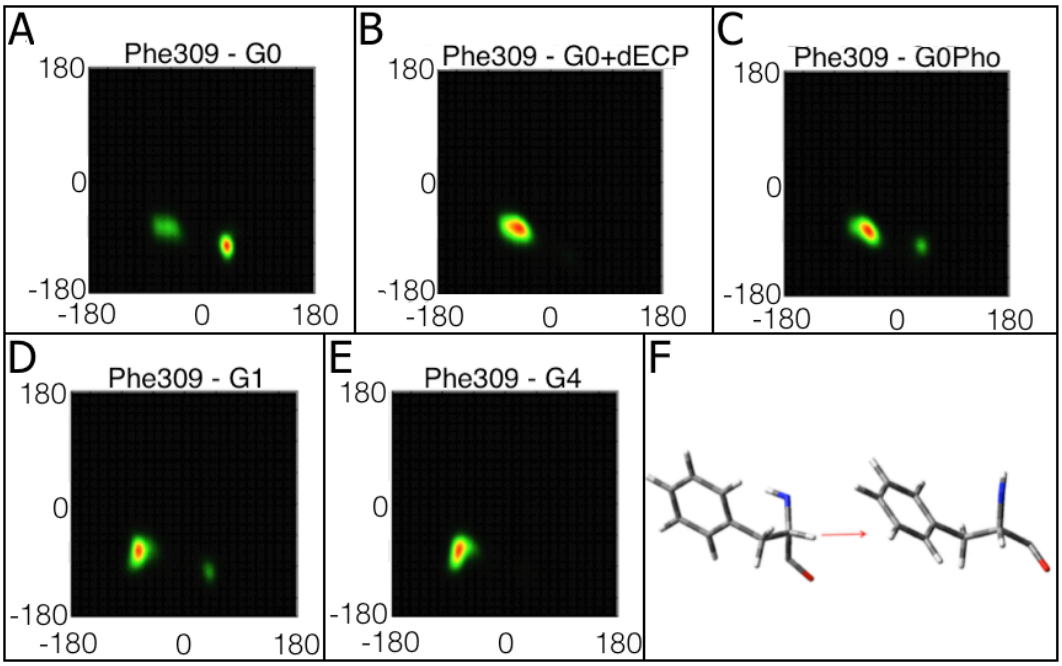


Figure 5.16. *χ_1/χ_2 plot of Phe309.* The sampled conformations of the ring of Phe309 are the same in all systems, but the preferred one changes when it is phosphorylated or binds a substrate.

Phe355 (Figure 5.17) samples one unique conformation in G0, with centre on χ_1 = -70; χ_2 = -70. This is not significantly affected by the presence of dECP or the reaction intermediate. The conformation of Phe355 in G1 is slightly different from that seen in G0, it is centred on χ_1 = -70; χ_2 = -90. The phenyl ring rotates more in this system, to the equivalent region $\chi_1 \approx$ -70; $\chi_2 \approx$ 90. In the mutant G4, this residue adopts the same configuration as in G0, G0+dECP and G0pho.

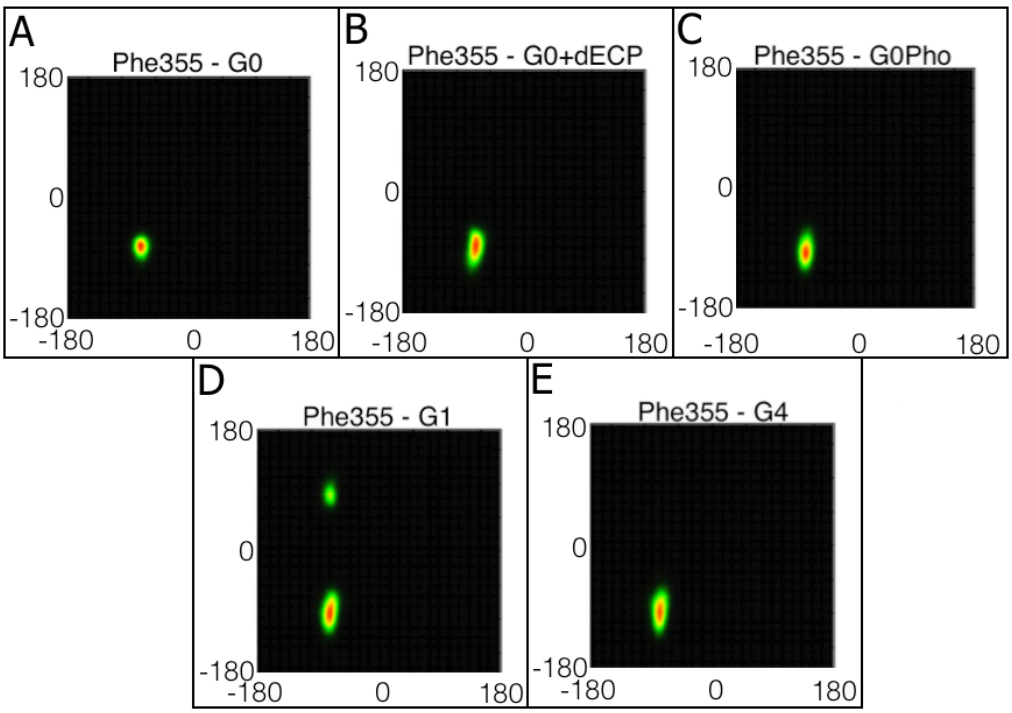


Figure 5.17. *χ_1/χ_2 plot of Phe355.* Another stable residue with only one conformation. The ring appears to have more freedom to flip in the G1 mutant than in the other systems.

Phe421 (Figure 5.18) has only one conformation in G0, G1 and G4, which is centred on χ_1 = 165; χ_2 = 75. With dECP bound, this residue has the same preferred conformation as in G0, but a second configuration, with centre on χ_1 = -60; χ_2 = -70, is sampled. This latter configuration is the only one adopted by this residue in the phosphorylated enzyme (G0pho).

From all this, it is clear the active site of G1 is shows more mobility than that of G0. Moreover, this single point mutation causes several residues in the binding pockets to behave the way they do when the substrate dECP is

bound. This suggests that, besides introducing a base in the active site, the Gly137Asp mutation could also work by making the active site adopt the conformation of the bound enzyme.

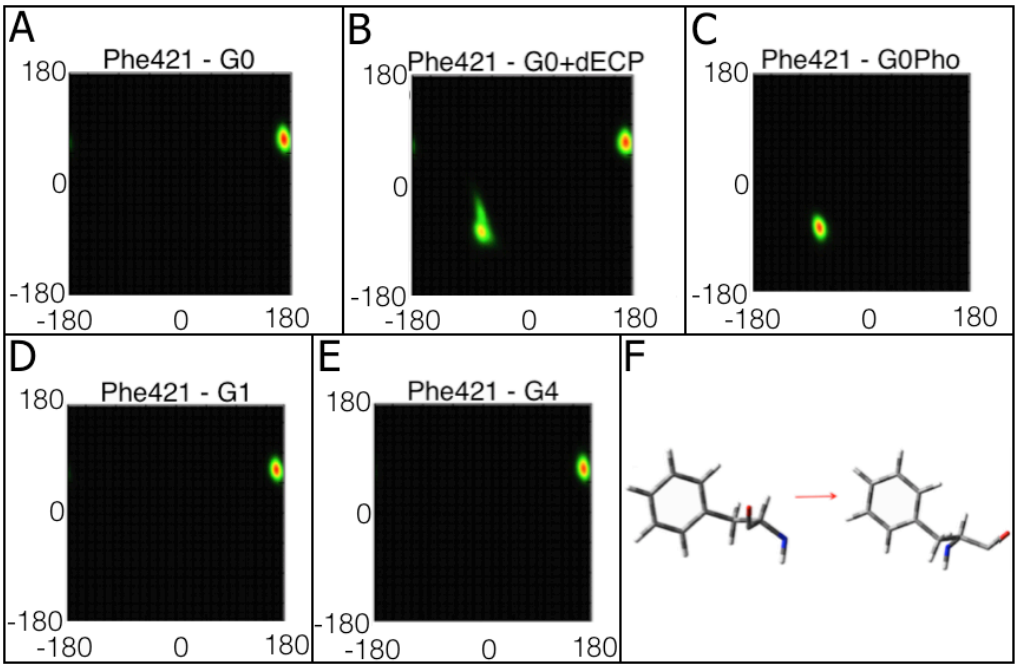


Figure 5.18. *χ_1/χ_2 plot of Phe421.* Mutations do not alter the conformational space of this residue, but it is when substrates bind or it becomes phosphorylated. The caption in panel G shows the difference between these two conformations.

Phosphorylation has mixed effects on the active site of E3. While most residues, including Glu351 and His471, do not change with respect to the free enzyme, some go through the same changes observed in dECP-bound E3 and in the G1 mutant. The mutations present in G4 have a similar (and at times the same) effect on the enzyme as the only mutation in G1, which prearranges the active site to the bound conformation. The mutations in G4 revert the changes present in two residues of G1 to the G0 configuration, namely Phe354 and Tyr457. The latter adopts the same orientation in G1 as it does in G0+dECP, but in G4 it reverts to the orientation observed in G0. More work is needed to determine whether new mutations that make this residue take the same orientation as in G1 would improve catalysis in G4. Mutations have altered the conformational

landscape of E3 (for a summary, see Figure 5.19), which is likely to be a major reason for its enhanced catalytic specificity towards organophosphates. The extent of conformational change shown in this figure could also be one of the reasons why the quantum cluster calculations described in Chapter 3 were difficult to converge when using restraints.

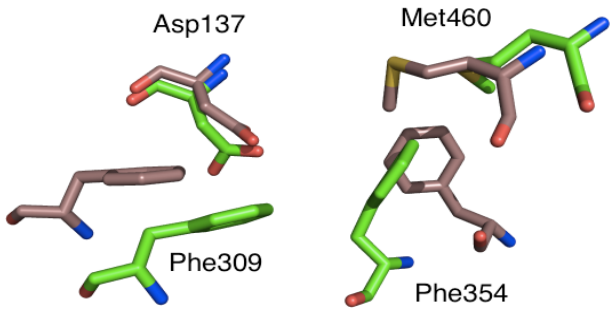


Figure 5.19. *Conformations of the E3-G4 active site predicted by MD simulations.* The residues that have the most different conformations across systems are shown. In brown, the predicted structure of E3-G4; in green a different conformation of these residues predicted by MD and not found in the X-ray structure.

5.4 Analysis of B-factors

B-factors are a measure of thermal fluctuations and provide information about the flexibility of a protein[26] (for more detail see Chapter 2). In this work, B-factors calculated from the atomic fluctuations obtained from the MD simulations were used to compare dynamical information of different systems and probe whether mutations, phosphorylation and substrate binding alter the thermal stability of E3. The results of the analysis are summarised in Figure 5.20 and Table 5.4. It is noted that the simulated B-factors follow a similar trend to those observed in the x-ray experiments.

The mobility of the active site is not significantly affected by mutations or the presence of substrate or intermediate. The B-factor of Ser218 increases slightly upon binding dECP. The mobility of this residue in the two mutants is similar to that observed in G0. Residues His471 and Glu351 do

not differ significantly between systems either, except for a slight increase in the B-factor of Glu351 when dECP binds. The B-factor of mutant residue Asp137 is similar in the two mutants, G1 and G4. The effects of mutations and substrate presence on the mobility of small binding pocket residues is mixed (see Table 5.4). The B-factor of Phe354 increases significantly in the mutant G4 (although not in G1), it also increases, although less, when dECP or the intermediate are bound. The B-factor of Tyr457 does not experience significant changes after phosphorylation, binding dECP or mutations. Met460 shows the largest change (increase) in the mutant G1. With respect to G0, the B-factor of this residue is unaltered in G4, it increases slightly when dECP is bound and decreases slightly when E3 is phosphorylated. The B-factor of Thr472 increases in the dECP-bound system, G1 and G4, but remains unchanged with respect to G0 in the phosphorylated system. Overall, mutations seem to mimic the changes introduced by the substrate in the small pocket.

	G0	G0+dECP	G0pho	G1	G4
Ser218	4 [17]	11	-	6 [10]	6 [16]
His471	6 [22]	7	6 [14]	7 [11]	8 [21]
Glu351	6 [19]	10	7 [12]	6 [7]	6 [15]
Asp137	-	-	-	17 [13]	18 [21]
Phe354	18 [28]	31	27 [19]	20 [20]	35 [24]
Tyr457	17 [30]	17	16 [16]	16 [19]	18 [22]
Met460	37 [39]	42	32 [24]	48 [31]	38 [39]
Thr472	7 [20]	12	8 [16]	13 [10]	13 [19]
Trp251	9 [25]	10	9 [13]	14 [14]	9 [20]
Met308	31 [25]	32	27 [16]	64 [25]	-
Phe309	40 [24]	23	29 [12]	39 [20]	32 [30]
Phe355	14 [35]	23	24 [19]	19 [25]	23 [31]
Phe421	7 [19]	8	6 [11]	8 [9]	7 [16]

Table 5.4. *B-factors of selected active site residues.* B-factors (Å^2) from the simulations. In brackets, B-factors of the Cα from the x-ray structures.

In the large binding pocket, changes in B-factors caused by mutations resemble very closely the changes in this parameter introduced by substrate binding. Trp251 is fairly stable to changes in the enzyme. Its B-factor only affected significantly in the mutant G1, in which it has a small

increase with respect to its value in G0. The mobility of Met308 is not affected by substrate binding, it decreases slightly in the phosphorylated system. However, the B-factor of this residue is greatly increased in the mutant G1 (reminder: this residue is a valine in the mutant G4). Phe309 is less mobile when the substrate or intermediate are bound. Its B-factor does not change significantly in G1 but it decreases in G4, accompanying to some extent what happens in the substrate or intermediate-bound systems. The B-factor of Phe355 increases significantly both when dECP is bound and when the enzyme is phosphorylated. The same extent of change is observed in G4, while the increase is less but still present in G1. Finally, residue Phe421 does not experience significant changes in its B-factor in any of the systems.

In the light of these results, the simulations carried out here indicate that binding of the substrate dECP makes the enzyme overall more flexible than the free form, while phosphorylation makes it less flexible (see Figure 5.20), which is consistent with experimental findings that show that human butyrylcholinesterase is stabilised by phosphorylation[27]. Mutations do not alter the B-factor of the catalytic triad residues, Ser218, His471 and Glu351. This indicates that mutations do not affect the catalytic machinery of E3 directly (except, of course, for the introduction of a base in the active site) and is most likely due to the hydrogen bond network that positions the catalytic triad and restricts its mobility. Thr472 and Phe421 have low mobility and do not change significantly across systems. These residues lie in the inner parts of the pockets and their rigidity is probably conserved so they can selectively bind the substrate. The changes in B-factor observed in G4 resemble those in the dECP-bound enzyme more closely than those observed in G1 (which does, however, follow the same trend). This is likely part of the reason why G4 is a better catalyst than G1.

increase with respect to its value in G0. The mobility of Met308 is not affected by substrate binding and decreases slightly in the phosphorylated system. However, the B-factor of this residue is greatly increased in the mutant G1 (as a reminder, this residue is a valine in the mutant G4). Phe309 is less mobile when the substrate or the intermediate are bound. Its B-factor does not change significantly in G1 but it decreases in G4, accompanying to some extent what happens in the substrate or intermediate-bound systems. The B-factor of Phe355 increases significantly both when dECP is bound and when the enzyme is phosphorylated. The same extent of change is observed in G4, while the increase is less but still present in G1. Finally, residue Phe421 does not experience significant changes in its B-factor in any of the systems.

In the light of these results, the simulations carried out here indicate that binding of the substrate dECP makes the enzyme overall more flexible than the free form, while phosphorylation makes it less flexible (the latter is consistent with experimental findings that show that human butyrylcholinesterase is stabilised by phosphorylation[22]). Mutations do not alter the B-factor of the catalytic triad residues, Ser218, His471 and Glu351. This indicates that mutations do not affect the catalytic machinery of E3 directly (except, of course, for the introduction of a base in the active site) and is most likely due to the hydrogen bond network that positions the catalytic triad and restricts its mobility. Thr472 and Phe421 have low mobility and do not change significantly across systems. These residues lie in the inner parts of the pockets and their rigidity is probably conserved so they can selectively bind the substrate. The changes in B-factor observed in G4 resemble those in the dECP-bound enzyme more closely than those observed in G1 (which does, however, follow the same trend). This is likely part of the reason why G4 is a better catalyst than G1.

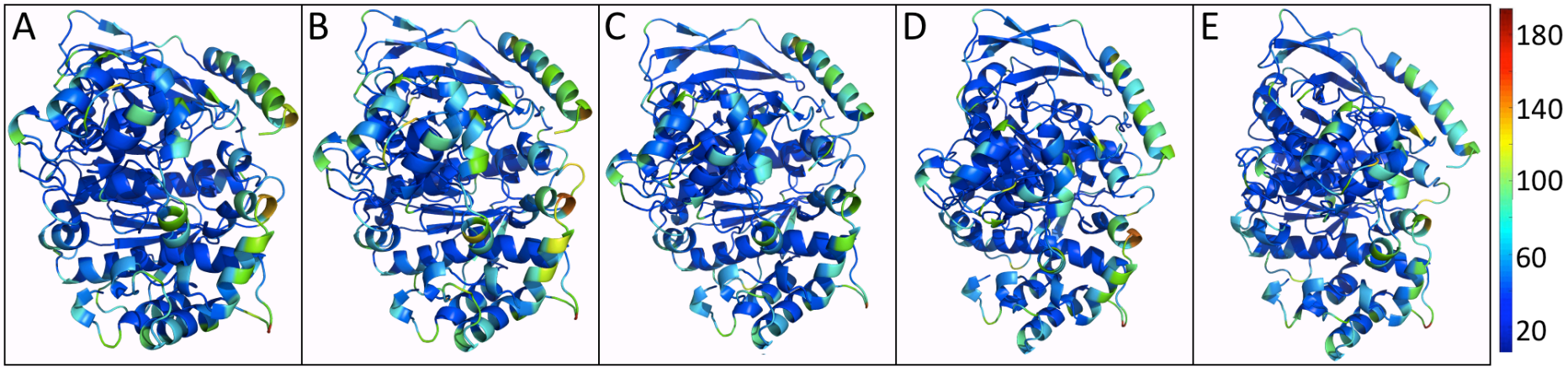


Figure 5.20. *B-factors plotted on the enzyme structure.* The colour bar to the right indicates the scale used in these captions. The minimum was 3 Å^2 and the maximum 194 Å^2. **A.** G0. **B.** G0+dECP. **C.** G0pho. **D.** G1. **E.** G4. Based on the initial structures for simulation.

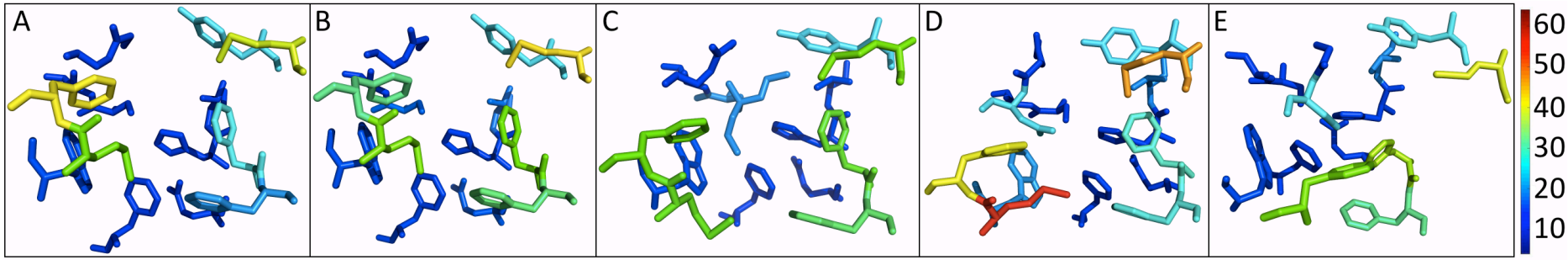


Figure 5.21. *B-factors of the active site residues.* The colour bar to the right indicates the scale used for colouring residues in this figure. The minimum value was 4 Å^2 and the maximum was 64 Å^2. **A.** G0. **B.** G0+dECP. **C.** G0pho. **D.** G1. **E.** G4. Based on the initial structures for simulation.

Some loops of the enzyme also exhibit different mobility in the different systems (Figures 5.22 – 5.25). The loop that comprises residues 402 – 408 is highly mobile in all systems. Other mobile regions include the loop that spans residues 464 – 467 (which is close in sequence to the catalytic residue His471), the loop and helix of residues 358 – 363 (which is less mobile in the phosphorylated system, while it has the highest mobility in the dECP-bound E3 and in G1), and residues 270 and 272 (which are part of a loop that spans residues 268 – 274). The dynamics of the different domains of E3 will be analysed in more detail in the next section.

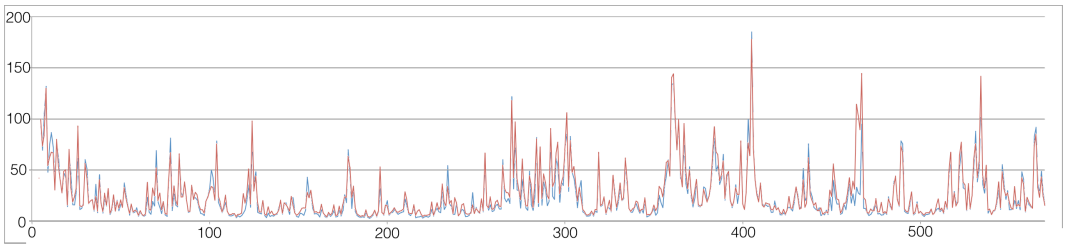


Figure 5.22. *B-factor of backbone atoms for G0 and G0+dECP. x* axis: residue number. *y* axis: B-factor (Å²). Color code: G0 = blue; G0+dECP = red.

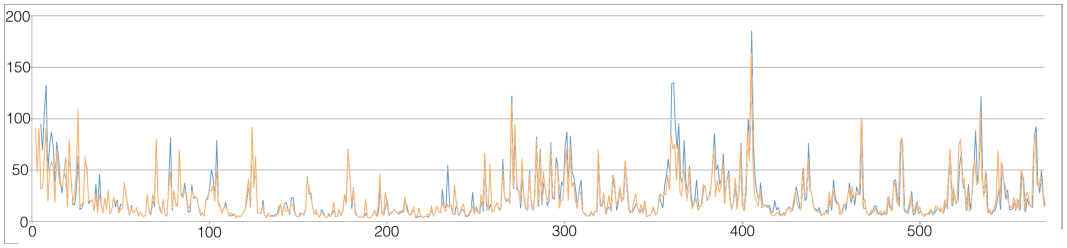


Figure 5.23. *B-factor of backbone atoms for G0 and G0pho. x* axis: residue number. *y* axis: B-factor (Å²). Color code: G0 = blue; G0Pho = orange.

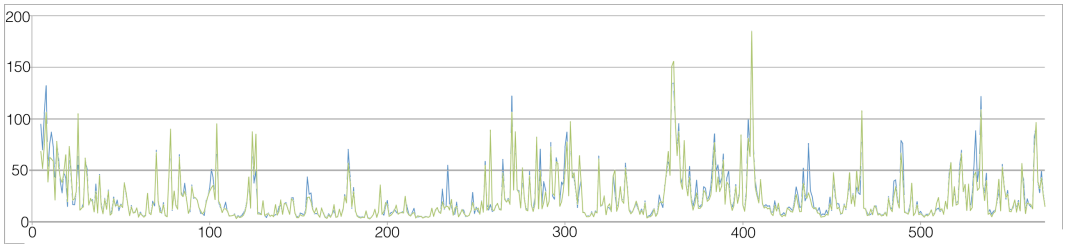


Figure 5.24. *B-factor of backbone atoms for G0 and G1. x* axis: residue number. *y* axis: B-factor (Å²). Color code: G0 = blue; G1 = green.

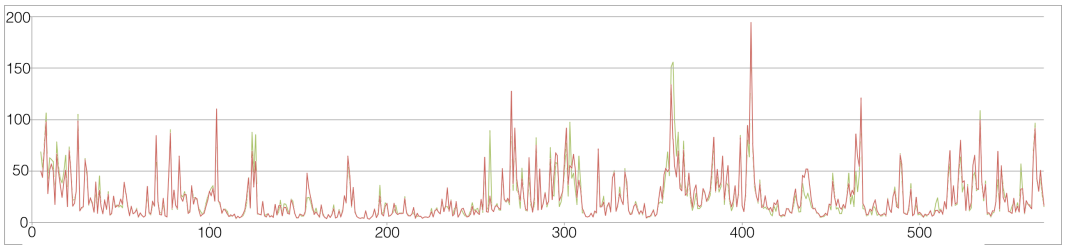


Figure 5.25. *B-factor of backbone atoms for G0 and G4. x* axis: residue number. *y* axis: B-factor (Å^2). Color code: G1 = green; G4 = red.

5.5 Analysis of correlated and anti-correlated motions

5.5.1 Correlated motions

Analysis of the correlation of motions between pairs of residues shows that most correlated and anti-correlated motions (where segments of the enzyme move in consonance) are conserved across systems (see Figure 5.26), with some exceptions. Highly correlated motions present in all systems will be described first. The movement of residues 475-480 correlates with the movements of residues 140-150. These are two adjacent regions in the tertiary structure that are also close in sequence to part of the catalytic triad (see Figure 5.27A). Evidence was found that distant regions of the enzyme could affect substrate binding. One such example is the correlation between the movements of helix 199-221 and sheet 117-137 (see Figure 5.27B), which is of great significance because the first of these regions includes active site residues Ser218 and Ala219 (the latter one is part of the oxyanion hole). The more external part of it interacts with the sheet and loop that comprise residues 117-137, some of which are solvent exposed, and which includes the remaining part of the oxyanion hole. This mode shows correlated motions of the catalytic triad and oxyanion hole domains. Regions 250-261 and 168-176 also show correlated motions. The former contains large binding pocket residue Trp251, (see Figure 5.27C) some of the other residues are quite far from the active site. The segment 306-323, which includes large binding pocket residues Met308 and Phe309,

correlates to region 166-185 (see Figure 5.27D) which is remote from the active site. Part of this region makes up the surface of the enzyme, again suggesting that the surface can affect the active site. The residues 307-310 also correlate with residues 248-255, which shows how the large substrate binding pocket moves as a whole despite being made of distant parts of the residue sequence (see Figure 5.27E). It can be observed in Figure 5.26 that there is much greater low-level correlation in G4 than in G1, which suggests that the G4 mutations increase the cooperative motions.

5.5.2 Anti-correlated motions and active site gate

Anti-correlated motions are of special importance because they may indicate that a gate mechanism is operating. Anti-correlation is observed between regions 255-305 and 350-390, which form part of the cleft that leads the substrate to the active site. This motion is present in all systems. Normal mode analysis and visualisation of the first low frequency normal mode shows an opening/closing mechanism operating within this region. These regions approach and move away, opening and closing access to the active site (see Figure 5.28). The region 350-390 also shows movements that anti-correlate with those of the area 165-185. The latter correlates with the movements of the region 250-261 (as discussed above), therefore the movement – which is observed in all systems – is probably part of the gate described before.

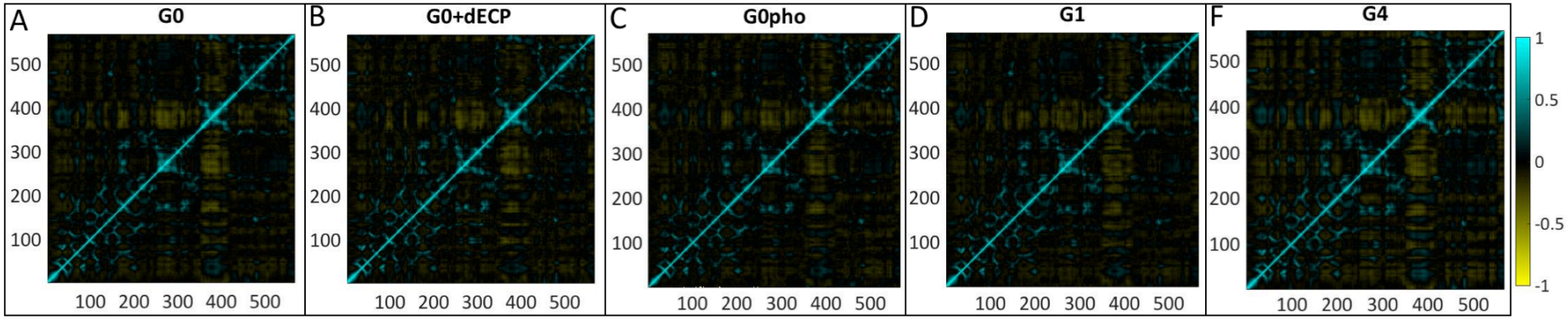


Figure 5.26. *Correlated motions of residues.* A totally correlated motion (value on the scale of 1) the area is coloured blue (as happens in the diagonal line that correlates a residue to itself). On the other hand, intense yellow is anticorrelated motion (value of -1). **A.** G0. **B.** G0+dECP. **C.** G0pho. **D.** G1. **E.** G4. All plots have the same colour bar as **A**.

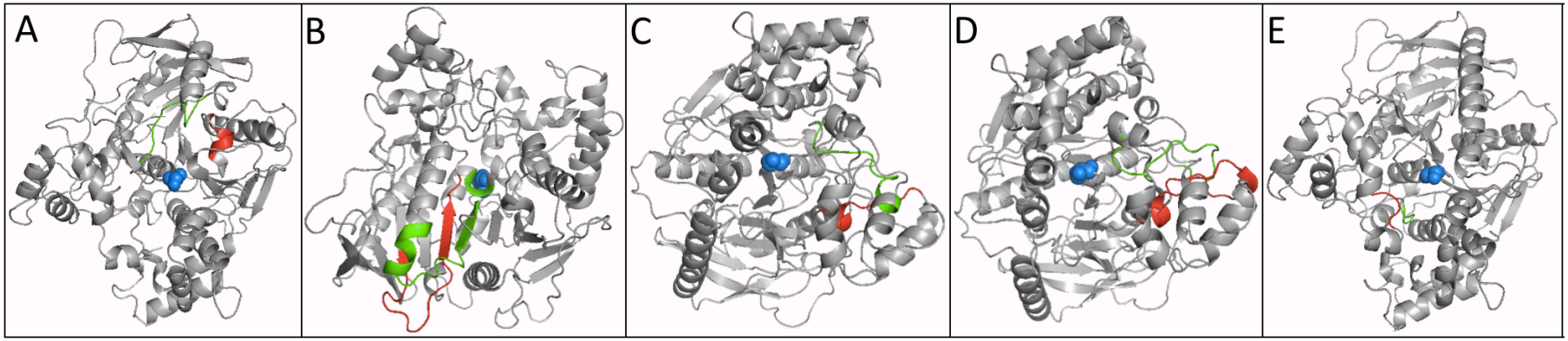


Figure 5.27. *Correlated motions.* The most relevant correlated motions are presented in this figure. **A.** residues 475-480 (red) and 140-150 (green). **B.** 117-137 (red) and 199-221 (green). **C.** 168-176 (green) and 250-261 (red). **D.** 306-323 (green) and 166-185 (red). **E.** 248-255 (green) and 307-310 (red). The catalytic triad is shown in blue spheres, for reference.

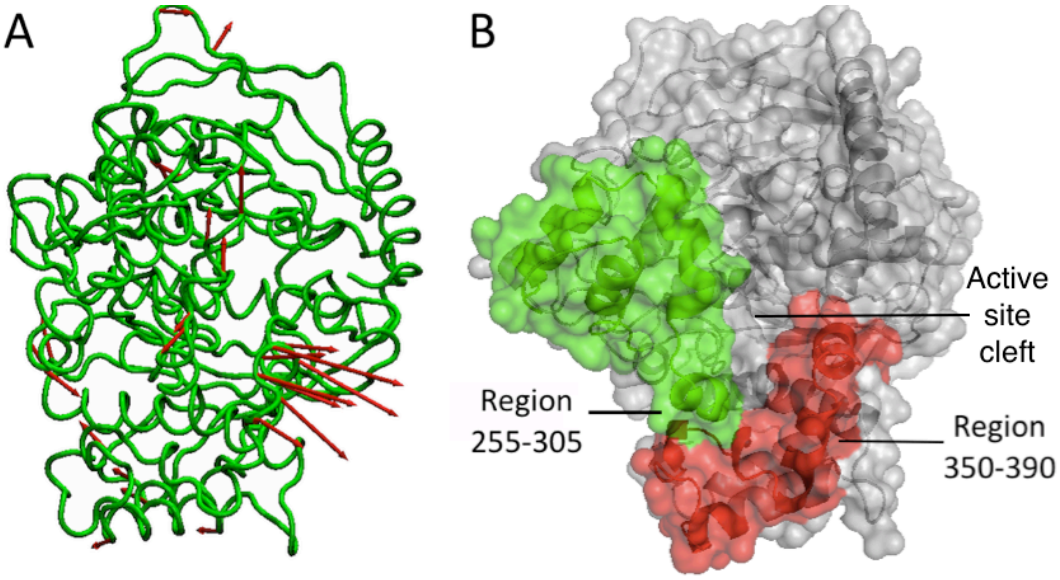


Figure 5.28. *First normal mode (anti-correlated).* **A.** Representation of E3 showing the motions of the first normal mode (in red arrows). **B.** Surface representation depicting the areas concerned in the motion, which surround the active site cleft.

5.5.3 The effect of mutations, substrate binding and phosphorylation on the active site gating motion

Inspection of the plots presented in Figure 5.29 reveals that the mutants have a very similar anti-correlation map to G0. The most significant difference is observed in G1. In this system the residues that make the active site gating motion are less strongly anti-correlated than in G0. This suggests that the Gly137Asp mutation may have advantageously introduced a base in the active site of E3 but, at the same time, caused negative effects on the dynamics of the enzyme. The fact that this anti-correlation is not affected in G4 (as compared to G0) seems to indicate that mutations to positions other than 137 can be useful to restore functionality lost previously. More significant differences are observed after the substrate binds and, especially, after phosphorylation. The gating mode (regions 350-390 and 255-305) is less strong in simulations of G0pho, which suggests that phosphorylation may induce a change of motions that

makes the active site less accessible. dECP binding has a similar but less marked effect than phosphorylation on the gating motion.

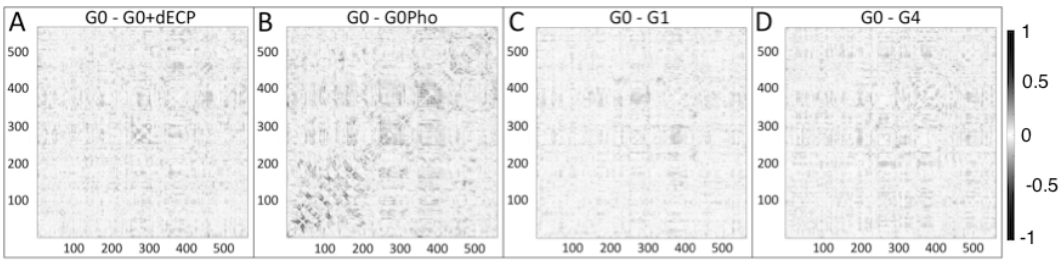


Figure 5.29. *Differences in cross-correlated motions between systems.* The maps were obtained by subtracting the cross-correlation values of one system from the other. **A.** Effects of binding dECP on wild-type E3. **B.** Effects of phosphorylation. **C.** Effect of the G1 mutation on the cross-correlations map. **D.** Effect of the G4 mutations. Residue numbers in the *x* and *y* axes.

5.6 Conclusion

The results presented in this chapter contain important evidence that the mutation that introduced a base in the active site and gave E3 the ability to break down organophosphate pesticides, namely Gly137Asp, may have been detrimental to the dynamics of the enzyme in several ways, and that the mutations present in the lab-evolved G4 may be reversing these problems. This residue is in a 'closed' conformation in the X-ray structure of G1, but the MD simulations reveal that it can adopt other, more favourable conformations, not present in the crystal structure.

NMA studies show that there is a gating motion present in the gorge that leads to the active site of E3. This motion is to some extent reduced in the G1 mutant, although it is restored in G4. Phosphorylation reduces this movement significantly. An extensive network of hydrogen bonds was found around the active site, which is highly stable. As expected, substrate binding causes the Ser218-His471 hydrogen bond to become more persistent during the simulation, as Ser218 is now ready to attack the substrate and donate a proton to His471. Analysis of B-factors reveals a

flexible enzyme that becomes less so upon phosphorylation. This supports previously existing experimental results[27]. Mutations are also detrimental to flexibility, although much less so than phosphorylation. G4 seems to partially restore the flexibility lost in G1. This raises the question of whether catalysis could be improved further by developing a more flexible mutant.

The analysis of $\chi1/\chi2$ dihedrals shows that the mutations to E3 studied here could mimic active site changes introduced by substrate binding, which could at least in part explain the enhanced catalysis observed in the mutants. Although some residues are highly stable to the changes introduced (substrates and mutations), other residues were found to suffer conformational changes in either or both mutants that resemble their conformation when the substrate dECP is bound. Phe354 only has an altered configuration in G1, which suggests that the G4 mutant may be a way to compensate for this change.

Future work should aim at determining which of the changes observed in the mutants are responsible for the increased catalytical rate of G4 and which are detrimental to enzyme activity and need further modification *via* new mutations.

4.8 References

(1) Kang, J.; Hagiwara, Y.; Tateno, M. *Journal of Biomedicine and Biotechnology* **2012**, *2012*.

(2) Harris, S. A.; Kendon, V. M. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2010**, *368*, 3581.

(3) Ferreira, L. G.; Dos Santos, R. N.; Oliva, G.; Andricopulo, A. D. *Molecules* **2015**, *20*, 13384.

(4) Singh, P.; Abeysinghe, T.; Kohen, A. *Molecules* **2015**, *20*, 1192.

(5) Karplus, M.; Kuriyan, J. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 6679.

(6) Bruice, T. C. *Chemical Reviews* **2006**, *106*, 3119.

(7) Lee, H. S.; Son, Y. J.; Chong, S. H.; Bae, J. Y.; Leem, C. H.; Jang, Y. J.; Choe, H. *Archives of Biochemistry and Biophysics* **2009**, *486*, 35.

(8) Laine, E.; de Beauchêne, I. C.; Perahia, D.; Auclair, C.; Tchertanov, L. *PLoS Computational Biology* **2011**, *7*.

(9) Armenta-Medina, D.; Pérez-Rueda, E.; Segovia, L. *Proteins: Structure, Function and Bioinformatics* **2011**, *79*, 1662.

(10) Lu, S.; Jiang, Y.; Lv, J.; Zou, J.; Wu, T. *Biopolymers* **2011**, *95*, 669.

(11) Ramanathan, A.; Agarwal, P. K. *Journal of Physical Chemistry B* **2009**, *113*, 16669.

(12) Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Cassarino, T. G.; Bertoni, M.; Bordoli, L.; Schwede, T. *Nucleic Acids Research* **2014**, *42*, W252.

(13) Olsson, M. H. M.; SØndergaard, C. R.; Rostkowski, M.; Jensen, J. H. *Journal of Chemical Theory and Computation* **2011**, *7*, 525.

(14) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. *Journal of Chemical Theory and Computation* **2011**, *7*, 2284.

(15) Bas, D. C.; Rogers, D. M.; Jensen, J. H. *Proteins: Structure, Function and Genetics* **2008**, *73*, 765.

(16) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins: Structure, Function and Genetics* **2005**, *61*, 704.

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

(18) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926.

(19) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. *Journal of Chemical Theory and Computation* **2015**, *11*, 3696.

(20) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *Journal of Computational Chemistry* **2004**, *25*, 1157.

(21) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *Journal of Computational Physics* **1977**, *23*, 327.

(22) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. *Biopolymers* **1992**, *32*, 523.

(23) Schröedinger, L. 2010.

(24) Humphrey, W.; Dalke, A.; Schulten, K. *J. Molec. Graphics* **1996**, *14*, 33.

(25) Jackson, C. J.; Liu, J. W.; Carr, P. D.; Younus, F.; Coppin, C.; Meirelles, T.; Lethier, M.; Pandey, G.; Ollis, D. L.; Russell, R. J.; Weik, M.; Oakeshott, J. G. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 10177.

(26) Kuzmanic, A.; Zagrovic, B. *Biophysical Journal* **2010**, *98*, 861.

(27) Gabel, F.; Weik, M.; Masson, P.; Renault, F.; Fournier, D.; Brochier, L.; Doctor, B. P.; Saxena, A.; Silman, I.; Zaccai, G. *Biophysical Journal* **2005**, *89*, 3303.

Chapter 6

Conclusions

6.1 Summary of the results presented in this thesis

The possibility of using the resistance of the sheep blowfly *L. cuprina* to organophosphate pesticides, which is detrimental to farming activities, in a beneficial way to detoxify organophosphates that are poisonous to humans, has been investigated. The results presented in this thesis shed light on the mechanistic, structural and dynamical aspects of *L. cuprina* carboxylesterase E3. Computational studies have provided a relatively inexpensive tool to generate *in silico* predictions and hypotheses that could then be confirmed experimentally.

The use of quantum cluster methods has permitted us to clarify the role of the naturally occurring Gly137Asp E3 mutation, which provides an active site general base in the right position to activate a water molecule for nucleophilic attack on the phospho-serine adduct. This is important because the catalytic histidine residue, which could also play the role of a base, is not in the right position to do so[1]. Other hypotheses previously advanced about the mechanism of dephosphorylation in the related enzyme human butyrylcholinesterase, such as strain[2], are also not strongly supported by the results presented in this thesis.

The activation energy barrier for this reaction was calculated to be approximately an order of magnitude higher if His471 is singly protonated (which would not allow for Ser218 to abstract a proton from it when it is released). Therefore, it is established that His471 is most likely

doubly protonated, bearing a single positive charge, during the dephosphorylation step of this reaction. The phosphorylation reaction mechanism found in this work coincided with that described by earlier works on related enzymes[3-7]. It was also observed that the hydrogen bonds that the substrate makes with the oxyanion hole residues serve to lower the activation energy barriers of the reaction.

Although the quantum cluster methods used in this thesis were invaluable in generating testable hypotheses relating to the catalytic mechanism, it was noted that the technique does suffer from problems due to the need to impose artificial restraints on the system to prevent the atoms drifting from the biologically possible conformations. This makes accurate prediction of reaction rates troublesome. Both the restraints applied, and the size of the system chosen, affect not only the calculation of energies, but also the ability to locate stationary points on the reaction path. Convergence of the calculations was extremely difficult when atoms corresponding to everything other than the substrate were frozen. Restraints applied on the Cα atoms of a larger model system were assessed to be the best option. This is not free of problems either, however, as in some cases multiple reaction pathways are possible and care must be taken to ensure a single low-energy reaction pathway is found.

Despite the many problems encountered, the barriers obtained in this work matched experimental results reasonably well. The quantum cluster calculations predict that the difference in rate between phosphorylation and dephosphorylation is a factor of 426. This is in very reasonable agreement with the experimental result (ca. 1600 fold), particularly considering that entropic differences are neglected in the quantum cluster calculations. This highlights the importance of the choice of level of theory to be used. The results presented in Chapter 3 indicate that RI-MP2 with a

Dunning basis set (aug-cc-pvtz) provides a good compromise between cost and accuracy for energy calculations of large systems such as the ones studied here. It is clear that methods frequently used to obtain energy barriers in QM/MM calculations, such as DFT (and in particular the functional B3LYP), are not accurate. This is, to the best of the author's knowledge, one of only few studies available of a system of this size at a high level of theory (see for example Navrátil *et al.*[8] and Bauzá *et al.*[9] for other examples).

Structural studies show that fatty acid methyl esters (FAMEs) fit extremely well in the active site cavity of E3, in close contact with the catalytic serine and within hydrogen bond distance from the oxyanion hole, which makes them likely natural substrates of E3. Experimental data obtained by the Jackson laboratory lends support to this idea, as FAMEs are metabolised by E3 at rates typical of a natural substrate[10].

Some organophosphates also fit in the E3 active site cavity, of which the toxic isomer of VX - VX(S) - was docked in a productive conformation (within attack distance from the catalytic serine, and within hydrogen bond distance of the oxyanion hole). Other organophosphate compounds were placed at a distance from Ser218 suitable for attack but not in contact with the oxyanion hole. Attention was focused on VX(S), and a molecular dynamics simulation showed that the binding was stable over a 15 ns trajectory. Resolution of crystallographic data of E3 phosphorylated with racemic VX confirms preferential phosphorylation of E3 by the S isomer rather than to the R isomer. Quantum cluster calculations suggest that the E3-Gly137Asp adduct with VX(S) could dephosphorylate at a rate similar to that of methyl and dimethyl adducts, which will rely upon future experimental studies to can confirm or disprove this hypothesis. The

potential effectiveness of E3 at hydrolysing and detoxifying this highly toxic molecule is exciting.

Molecular dynamics simulations indicate that substrate binding and phosphorylation change the conformational landscape of the active site of E3. It was noted that Tyr457 plays a special role, helping the induced fit of organophosphate substrates by moving from an open to a closed conformation when dECP binds. The catalytic triad (Glu351-His471-Ser218) itself is highly stable, held together by a network of hydrogen bonds, and its motions are not affected by the presence of substrate nor by mutations. The mutated residue Gly137Asp, which endows E3 with the organophosphate hydrolase activity, behaves similarly in G1 and G4. However, in the latter mutant this residue shows an increased sampling of a more open conformation, leading to the hypothesis that it is constrained to suboptimal positions in G1, possibly by Phe309. Simulations reveal that mutations mimic some of the conformational changes introduced by substrate binding. However, the presence of the mutant aspartate may be detrimental to the dynamics of residues near the active site. These studies have shown that the mutations in G4 are likely to provide a form of reversing these detrimental effects to the enzyme.

Normal mode analysis performed on the simulation data revealed a meaningful gating motion in the active site gorge of E3, which is significantly reduced when the enzyme is phosphorylated. This is consistent with analysis of the flexibility of E3 in different states, which showed that the enzyme becomes more rigid after phosphorylation. This had been previously observed in experimental studies[11]. Finally, although the G1 mutant loses some flexibility, the mutations in G4 restore it at least partially. Molecular mechanics simulations have been extremely valuable

to provide insight into E3 and its function, especially in the light of the issues faced with quantum cluster calculations.

6.2 Future directions

The results summarised above answer many questions asked at the beginning of this project, but also pose new ones. Time constraints did not allow for exploration of a number of side-paths to this project that would have been of interest. The results of the *in silico* and X-ray crystallography studies presented indicate that E3 and its mutants are promising candidates for the detoxification of organophosphate compounds, in particular of VX(S). Experimental kinetic confirmation, however, remains to be obtained.

Molecular dynamics studies of E3 with different organophosphates bound (and also phosphorylated with them) would make for interesting comparison of their effects on the conformational landscape of the enzyme. In particular, it would be of interest to compare the effects of the presence of organophosphate pesticides to that of organophosphate nerve agents.

Although light was shed on the effect that mutations have on E3, questions remain about the specific role and importance of active site (and nearby) residues in isolation. *In silico* mutation of residues, one or more at a time, followed by MD simulation runs would provide important information about these points. QM/MM calculations and/or EVB simulations would also prove highly useful to tackle the problems faced in this work and to better analyse this enzyme. The ultimate goal of this work was to understand the mechanism of E3, and the effects of mutations, sufficiently for the prediction of beneficial mutations to increase organophosphate hydrolase activity of E3 towards chemical

warfare agents. Although this ambitious goal remains out of reach, we have made significant progress towards it, establishing a foundation of detailed understanding of the process of substrate binding, the catalytic mechanism and the effects of mutations on the conformational landscape of the protein.

6.3 References

(1) Steitz, T. A.; Hendekson, R.; Blow, D. M. *Journal of Molecular Biology* **1969**, *46*, 337.

(2) Nachon, F.; Carletti, E.; Wandhammer, M.; Nicolet, Y.; Schopfer, L. M.; Masson, P.; Lockridge, O. *Biochemical Journal* **2011**, *434*, 73.

(3) Kwasnieski, O.; Verdier, L.; Malacria, M.; Derat, E. *Journal of Physical Chemistry B* **2009**, *113*, 10001.

(4) Lombardo, D. *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular* **1982**, *700*, 67.

(5) Bernhard, S. A.; Orgel, L. E. *Science* **1959**, *130*, 625.

(6) Jarv, J. *Bioorganic Chemistry* **1984**, *12*, 259.

(7) Sultatos, L. G. In *Toxicology of Organophosphate & Carbamate Compounds*; Gupta, R. C., Ed.; Academic Press: Burlington, 2006, p 209.

(8) Navrátil, V.; Klusák, V.; Rulíšek, L. *Chemistry - A European Journal* **2013**, *19*, 16634.

(9) Bauzá, A.; Quiñonero, D.; Deyà, P. M.; Frontera, A. *Chemistry - A European Journal* **2014**, *20*, 6985.

(10) Jackson, C. J.; Liu, J. W.; Carr, P. D.; Younus, F.; Coppin, C.; Meirelles, T.; Lethier, M.; Pandey, G.; Ollis, D. L.; Russell, R. J.; Weik, M.; Oakeshott, J. G. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 10177.

(11) Gabel, F.; Weik, M.; Masson, P.; Renault, F.; Fournier, D.; Brochier, L.; Doctor, B. P.; Saxena, A.; Silman, I.; Zaccai, G. *Biophysical Journal* **2005**, *89*, 3303.

Appendix

A.1 Computational data of Quantum Cluster calculations

A.1.1 Phosphorylation reaction – small systems

dECP – Reactant

1\1\GINC-R2937\FOpt\RM062X\6-31+G(d,p)\C20H28N2O9P1(1-)\ROOT\16-Aug-2014\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight) IOP(2/17=4) Freq=noraman scrf=(pcm,read)\\opt fwd\\-1,1\C,8.302786904,-2.897633137,0.1631980027\C,2.0765879491,-3.0441850564,-0.4959635953\C,1.8058721469,2.9804041934,2.2976475925\H,2.424249439,-3.3948065247,-1.4742466489\H,2.4873662083,-3.7185601713,0.26361152\ H,0.9855151882,-3.1175566527,-0.4683869281\O,2.1062891298,3.1424353907,0.9289765017\H,2.3090553928,2.2377320809,0.5667703352\C,2.4928385956,-1.6317909273,-0.245762045\C,1.7697074646,-0.4813014839,-0.0363194087\N,2.6262768734,0.5829818645,0.1513785032\C,3.8448589176,0.0742058484,0.0600050826\N,3.8105261612,-1.252912596,-0.180314391\H,4.6484254564,-1.8806058981,-0.2811495636\H,4.7767354239,0.6181301396,0.1659645032 \H,0.6939317436,-0.354225501,-0.001298247\C,6.9670200767,-2.1787778504, 0.0007886953\O,5.9817809802,-2.8982532966,-0.360625025\O,6.905681911,-0.9479521314,0.2310659877\P,-0.7709526246,2.8209654147,-0.2500173825\O,-0.9291438643,4.2113244893,0.2176856748\O,-0.4617800531,1.6828912189,0.810662745\C,-1.3174493792,0.0025972441,2.2718425409\H,-1.9575607806,-0.1950321062,3.1376723765\H,-1.6761085836,-0.5945565247,1.4251027888\H,-0.2938311187,-0.3044378258,2.5105136258\O,-2.1880800294,2.3217956447,-0.9005581888\C,-2.5035313331,0.9921886465,-1.077815361\C,-1.6073977331,0.0920928941,-1.6731169909\C,-1.9872028693,-1.2313864723,-1.8155771032\C,-3.2499693304,-1.6740617664,-1.3886498135\C,-4.1208102029,-0.7381962863,-0.8159268423\C,-3.764179174,0.5959613424,-0.651284807\O,-5.3629328538,-1.1049458843,-0.3827136469\C,-5.8286087535,-2.3914846268,-0.4703111986\O,-6.9440638166,-2.6088665513,-0.0465638354\C,-4.9401185262,-3.3824874372,-1.0665260963\C,-3.711817606,-3.0359295625,-1.5031272796\H,-5.3311856293,-4.3955481998,-1.1297839929\H,-3.0484650786,-3.7776540009,-1.9477320875\H,-1.3010236273,-1.946835416,-2.2664726599\H,-0.6276891712,0.4318644629,-2.0038529631\H,-4.4510608876,1.3069651683,-0.1926540298\O,0.2879989879,2.5304592291,-1.3802000279\C,1.2146741781,3.5450812206,-1.8568535144\H,1.2829351944,4.3389420762,-1.1087328588\H,2.1807393807,3.0346750528,
-1.919555186\C,0.7408540821,4.0562752915,-3.2002779488\H,-0.2314273446,
4.5510891167,-3.1021259648\H,1.4620880582,4.7833251654,-3.5885008599\H,0.6508368292,3.2366262561,-3.9199170397\C,-1.356033157,1.4759026931,1.930641457\H,-2.3707616746,1.7984664815, 1.6593441238\H,-0.9983190265,2.1003456182,2.7575154209\H,0.9489740921,
3.6159753662,2.5562959202\H,1.5431778672,1.9377287511,2.5253393097\H,2.6565357203,3.2688929537,2.9310092528\H,8.2099111838,-3.650814034,0.9527933017\H,8.5491523258,-3.4241933754,-0.7641834848\H,9.1001992964,-2.1968574042,0.4171218012\\Version=ES64L-G09RevD.01\State=1-A\HF=-1906.55335\RMSD=4.040e-09\RMSF=6.414e-08\Dipole=-5.9173371,-0.7389039,-

1.4197667\Quadrupole=-87.1309184,26.0592755,61.0716429,17.531858,-2.9068253,-0.3539509\PG=C01 [X(C20H28N2O9P1)]\\@

dECP – TS1

1\1\GINC-R2396\FTS\RM062X\6-31+G(d,p)\C20H28N2O9P1(1-)\ROOT\08-Aug-2014\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultrafine) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(pcm,read)\\title\\-1,1\C,8.97677111,-2.35370772,-1.45066956\C,4.59223,-2.789983,3.000379\C,1.800106,2.724484, 0.556912\H,4.6998018974,-3.7834950903,2.5517064693\H,5.5598823262,-2.5120893929, 3.4320565184\H,3.8566241603,-2.8502263935,3.8067230072\O,1.4215783698, 1.4561676733,0.0375562956\H,2.1814181003,0.6790936102,0.3703605524\C,4.1495226742,-1.7879033967,1.986480788\C,3.035192153,-0.990062429,1.9226138186\N,3.0931134213,-0.2325265821,0.7730468232\ C,4.2206705154,-0.5677501754,0.1580642461\N,4.8823024407,-1.500461511, 0.8589793441\H,5.8162388127,-1.9363344281,0.5741446606\H,4.5864390304,-0.1631392887,-0.7793863738\H,2.204825004,-0.9076553719,2.6111148499\ C,7.5972506168,-1.9788775092,-0.9213608546\O,7.2273044694,-2.5624022328, 0.1500139327\O,6.9142731945,-1.1310309731,-1.5401531787\P,-0.3720943825, 0.8688438262,0.461954371\O,-0.9680118105, 2.1626719399,0.0114757675\O,0.2583003854,0.6737108849,1.9509769694\C,0.4428884646,1.2916470795,4.2440203942\H,-0.1137982975,1.5549565549,5.1493723634\H,1.0521457588, 0.4073398046,4.4571291102\H,1.1102949707,2.1207837767,3.9891081763\O,-1.8850765694, 0.1161720222,0.9072184586\C,-2.0809216937,-1.1332523989, 1.353086644\C,-1.0474432695,-2.0630376105,1.6217493337\C,-1.3645912467,-3.3178070961,2.1069124953\C,-2.6960811048,-3.7018972439,2.3414126495\C,-3.7002572781,-2.7622607881,2.0658953446\C,-3.4151288443,-1.4957046177, 1.5812845179\O,-5.0190985384,-3.0694436704,2.2726330402\C,-5.4328902054,-4.2843421116,2.750096584\O,-6.6294520664,-4.4444658097,2.8971551264\C,-4.4056020781,-5.2729783151,3.039178041\C,-3.0989834708,4.986872947, 2.8408520459\H,-4.755816999,-6.2312644425,3.4149411944\H,-2.3303780979,-5.7289465077,3.0573813894\H,-0.5695220994,-4.0336595494,2.3110373024\H,-0.010611774,-1.7977097172,1.4388546019\H,-4.2153179815,-0.7857158567, 1.3770957489\O,-0.0309621753,-0.3767216921,-0.5004632662\C,0.5841068786,-0.252249208,-1.800071722\H,0.3671288751,0.7402884562,-2.2107175852 \H,1.6682897282,-0.356492287,-1.6826800188\C,0.0147681204,-1.3541909882,-2.669348045\H,-1.0670583027,-1.2353339884,-2.7824809354\H,0.4804838642,-1.3180948428,-3.6597900709\H,0.216187391,-2.3342159877,-2.2254793008\C,-0.5251441451,1.0133443908,3.1132588656\H,-1.1874820972,0.1752508878, 3.3613255639\H,-1.141882308,1.8931343307,2.8877727902 \H,1.2530593029,3.5028416036,0.0215846496\H,1.5734403169,2.7864486791,1.6285015543\H,2.8773647912,2.8502996952,0.4057331283\H,9.7332689223,-2.0485810077,-0.7198180077\H,9.0454580336,-3.440486257,-1.5592403688\H,9.1756411867,-1.8683079667,-2.4076925399\\Version=ES64L-G09RevD.01\HF=-1906.5404959 \RMSD=4.360e-09\RMSF=3.445e-07\Dipole=-1.1498025,-0.8969757,2.7270591 \Quadrupole=-98.8654883,42.116679,56.7488093,2.8201014,29.9809129, 3.4092147\PG=C01 [X(C20H28N2O9P1)]\\@

dECP – Intermediate

1\1\GINC-R2525\FOpt\RM062X\6-31+G(d,p)\C20H28N2O9P1(1-)\ROOT\04-Sep-20 14\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight) IOP

(2/17=4) Freq=noraman geom=check Guess=Read scrf=(pcm,read)\\title\\-1
,1\C,8.5434361092,-2.2175311905,-1.0178104204\C,2.9675782337,-3.068946
1571,1.7027727175\C,1.8219216316,3.440872666,1.0106915512\H,3.23490133
75,-3.908661275,1.0516808272\H,3.7020816636,-3.0294946857,2.514838141\
H,1.9834391695,-3.2669658146,2.136419144\O,1.1910190714,2.9970535645,-
0.1737969908\H,1.9140290165,0.9699273628,-0.3141286438\C,2.9495730025,
-1.784930558,0.9386435419\C,1.9225104068,-0.893370064,0.7621752334\N,2
.4313523201,0.1269200325,-0.0067990162\C,3.7171852188,-0.1610409739,-0
.277607534\N,4.0653917751,-1.3100439968,0.2802656435\H,5.4682829191,-1
.9301411152,0.1201158226\H,4.371057722,0.469096506,-0.8711293685\H,0.8
959966979,-0.8919228055,1.1088095957\C,7.172491404,-1.6322764482,-0.78
69903012\O,6.4160211246,-2.3668302657,0.0025133442\O,6.7994229507,-0.5
784320115,-1.287062605\P,-0.4827970034,2.6335092403,-0.0944498453\O,-1
.0502947028,3.9205736888,0.4343365248\O,-0.114214043,1.4334182123,0.96
92654708\C,-0.1491485313,1.1176004814,3.3422458351\H,-0.7687039728,0.8
085558191,4.1913154656\H,0.7000713363,0.430222839,3.2607404502\H,0.236
3017329,2.1234012989,3.5394132547\O,-2.2062211773,1.8806505203,-0.2189
807269\C,-2.5169999875,0.648323522,-0.5608533329\C,-1.571997728,-0.346
9486959,-0.9517902809\C,-1.9896021916,-1.6230897514,-1.2651720196\C,-3
.3491799209,-1.9881685258,-1.2202361769\C,-4.2686174298,-0.9960690564,
-0.8391739343\C,-3.8815090719,0.2912872207,-0.5137182207\O,-5.60887870
62,-1.2838320457,-0.7729403426\C,-6.1217283673,-2.5189964263,-1.066327
8926\O,-7.3289692451,-2.6549910623,-0.9691034513\C,-5.1849181041,-3.55
32755936,-1.4642685663\C,-3.8574709985,-3.2880284267,-1.5346237193\H,-
5.6118038821,-4.5261166849,-1.6950763667\H,-3.1547936541,-4.0663189065
,-1.8340130033\H,-1.2569776202,-2.3741975637,-1.5597139892\H,-0.517951
2852,-0.0958637158,-1.0046882859\H,-4.6221093476,1.0336641015,-0.21892
85496\O,-0.441392459,2.2559994702,-1.6848647604\C,0.4282528396,2.88474
97193,-2.6385067232\H,0.4761981463,3.9615285003,-2.4377713487\H,1.4395
046525,2.4736841697,-2.5422403443\C,-0.1489420811,2.6048017057,-4.0120
817721\H,-1.1535424334,3.0296785884,-4.1004650653\H,0.4915006432,3.048
8230132,-4.781837934\H,-0.2117869029,1.5260565471,-4.1894764997\C,-0.9
773438136,1.1129626957,2.0709381239\H,-1.4185701905,0.1236237806,1.882
6399435\H,-1.7898870673,1.8448807727,2.1328406309\H,1.2353918176,4.231
295855,1.4946819138\H,1.95683469,2.6102712291,1.7169264366\H,2.8035816
089,3.8389978976,0.7339890219\H,8.4415851187,-3.2054590868,-1.47843079
92\H,9.1271932943,-1.5606695217,-1.6633115761\H,9.049063086,-2.3501400
501,-0.0560676189\\Version=ES64L-G09RevD.01\State=1-A\HF=-1906.5477067
\RMSD=1.167e-09\RMSF=3.652e-09\Dipole=3.7083158,-2.9864257,0.0871296\Q
uadrupole=-61.9978833,15.8505392,46.1473441,-1.9120566,-1.3882476,-3.6
207555\PG=C01 [X(C20H28N2O9P1)]\\@

dECP – TS2

1\1\GINC-R2535\FTS\RM062X\6-31+G(d,p)\C20H28N2O9P1(1-)\ROOT\28-Aug-201
4\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultrafine
) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(pcm
,read)\\title\\-1,1\C,8.4981425898,-2.3356400622,-0.9816320838\C,3.055
8982082,-2.8802372332,2.0683374775\C,1.851175202,3.5151632954,0.720485
6063\H,3.3165040958,-3.8012012444,1.5350419218\H,3.821039909,-2.709986
5933,2.8338768525\H,2.0946262781,-3.0300019196,2.5677587222\O,1.058563
7738,2.8959947851,-0.280546008\H,1.8142781587,0.8002384097,-0.46760400

83\C,2.976668429,-1.7216100342,1.1278729285\C,1.9183166724,-0.89071986
01,0.8651508166\N,2.3683028544,0.02042302,-0.0616216205\C,3.6526460289
,-0.2720939517,-0.337401494\N,4.0542160973,-1.3194446712,0.3655039366\
H,5.4702974034,-1.9274397743,0.2404960136\H,4.2655660979,0.2795149641,
-1.0422091568\H,0.9074810845,-0.8640016171,1.2528763175\C,7.1218844455
,-1.7500733217,-0.7860540177\O,6.4234827089,-2.3566625006,0.1520679977
\O,6.6955220583,-0.8041627373,-1.4366096024\P,-0.5383168186,2.54192287
19,0.1038228183\O,-1.1095946242,3.8458051565,0.5546974797\O,-0.1032402
481,1.4245557593,1.1853873917\C,0.2082237,1.2136086218,3.5495160593\H,
-0.296793408,1.0080645147,4.4995980145\H,0.9670261119,0.4414905016,3.3
832249607\H,0.7104522705,2.1836585648,3.6262033265\O,-2.5037674771,1.7
525075579,0.3817910605\C,-2.9336563022,0.5631781045,0.1129419834\C,-2.
0754371838,-0.5331241074,-0.2480965031\C,-2.5886837899,-1.7842086365,-
0.5017176146\C,-3.973517471,-2.04763346,-0.4235414883\C,-4.8102576971,
-0.9707612209,-0.0699541319\C,-4.3282003877,0.2953276189,0.1935390162\
O,-6.1680822813,-1.1597947137,0.0243994068\C,-6.7765038747,-2.36256368
17,-0.2192771642\O,-7.9920546332,-2.3966217748,-0.1085150608\C,-5.9274
256097,-3.4767988711,-0.5824747704\C,-4.5813990111,-3.311202856,-0.677
0679021\H,-6.4259507637,-4.4241009257,-0.770121093\H,-3.9437714633,-4.
1522155154,-0.9522893935\H,-1.9199443401,-2.6007116196,-0.7745444571\H
,-1.0065664533,-0.3570917414,-0.3273130372\H,-5.0115861337,1.099454252
1,0.4629114151\O,-0.7825139422,1.9001361625,-1.3572808196\C,-0.4910734
264,2.6854467665,-2.5311360217\H,-0.7137716023,3.7409289863,-2.3264355
367\H,0.5765611736,2.5928188289,-2.7618110753\C,-1.3593420275,2.150428
5656,-3.6509669963\H,-2.4177687939,2.2550932895,-3.3937499052\H,-1.160
569176,2.7054717646,-4.5739243725\H,-1.1469623449,1.0908979142,-3.8270
883764\C,-0.8101894722,1.2323795264,2.4267949866\H,-1.3556775679,0.282
5532336,2.3554578773\H,-1.5372946055,2.0389352245,2.5619388786\H,1.362
4046924,4.4204542589,1.0985831767\H,2.03018151,2.8210610496,1.55252065
36\H,2.8076991706,3.7805561522,0.2619706295\H,8.4059993775,-3.38974908
45,-1.2625236866\H,9.0323482087,-1.7856559466,-1.7569114642\H,9.049931
3011,-2.2906414604,-0.037328468\\Version=ES64L-G09RevD.01\HF=-1906.546
1705\RMSD=7.428e-09\RMSF=7.578e-07\Dipole=4.7353566,-2.0020109,0.03261
38\Quadrupole=-77.5034574,26.167583,51.3358744,-1.8917645,7.772898,-3.
4404697\PG=C01 [X(C20H28N2O9P1)]\\@

dECP – Product

1\1\GINC-R3527\FOpt\RM062X\6-31+G(d,p)\C20H28N2O9P1(1-)\ROOT\09-Sep-
2014\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight) IOP(2/17=4)
Freq=noraman geom=check scrf=(pcm,read)\\opt fwd\\-
1,1\C,6.3849639098,2.8128254292,-1.5110234993\C,3.074900163,-
0.435327013,2.6995934721\C,-2.1946127086,3.3340585485,1.218404471\H,3.9643082138,-
0.82553
76599,2.1923701378\H,3.4057229633,0.2721015319,3.4684354878\H,2.569498
8391,-1.2731726905,3.1887009748\O,-2.6334334135,2.588692454,0.06892368
35\H,-0.458030059,0.9355291267,0.0283064443\C,2.1608467235,0.224255426
3,1.716202378\C,0.8141850791,0.0497079234,1.5170755845\N,0.4651263335,
0.8975619305,0.4925134285\C,1.5778416895,1.5330740987,0.0851040311\N,2
.6178992598,1.1625878621,0.8140534424\H,4.0465110809,1.5300374293,0.42
09214314\H,1.6099434247,2.2412023588,-0.735299934\H,0.0984843564,-0.61
37575923,1.9842462924\C,5.0104858572,2.4775243631,-0.9888232267\O,5.02

86710386,1.7541735665,0.1101330575\O,3.9779924667,2.8434426351,-1.5385
769523\P,-3.5800513955,1.332421578,0.3309191514\O,-4.7759564052,1.5513
436734,1.1813160639\O,-2.5504048276,0.2709543607,0.9319348351\C,-2.840
46796,-0.3409363809,3.2608954559\H,-3.1067445522,-1.1640818873,3.93251
55185\H,-1.8081653506,-0.041268952,3.4716660415\H,-3.5042267985,0.5057
453301,3.4636845113\O,-2.8054639613,-2.3028486302,-1.088552183\C,-1.55
29601506,-2.0792640869,-1.1523108176\C,-0.9980635323,-1.1465459776,-2.
1142327486\C,0.352807218,-0.9256204531,-2.2096713917\C,1.2740199809,-1
.5832236272,-1.3568569975\C,0.7372485253,-2.4777397618,-0.4063436298\C
,-0.6140351697,-2.7325397114,-0.2905419223\O,1.5766218594,-3.133764331
6,0.4613965112\C,2.9384278674,-2.9583965961,0.4688019476\O,3.565149100
6,-3.6043187322,1.2990254652\C,3.4944449178,-2.0387198848,-0.490936748
1\C,2.6781410557,-1.3786414555,-1.3619506456\H,4.5708111753,-1.8928906
251,-0.459990948\H,3.1006927081,-0.6674489682,-2.0737153202\H,0.745380
6773,-0.2200142797,-2.9429681523\H,-1.6987316488,-0.6298786483,-2.7694
797923\H,-0.9727762649,-3.4470266475,0.4494633278\O,-3.8683622065,0.93
26418082,-1.1752309979\C,-4.9718951552,0.0199164762,-1.4332938504\H,-4
.8164582559,-0.8897537138,-0.8408024369\H,-5.8985854316,0.5190438111,-
1.1273997731\C,-4.9578116172,-0.3045265698,-2.9078493891\H,-4.03760510
39,-0.8371438144,-3.163300185\H,-5.808718328,-0.9520881938,-3.14309425
66\H,-5.0344433706,0.6086601724,-3.5068388817\C,-2.9811364479,-0.78960
03013,1.821038059\H,-2.3323917815,-1.6368328221,1.5799926717\H,-4.0140
666486,-1.067558038,1.5816001904\H,-3.0572994679,3.7458913018,1.750256
2988\H,-1.6089474433,2.6892628071,1.8841681933\H,-1.5658708594,4.13916
25981,0.8369880847\H,6.9231518789,1.8858350328,-1.7335075774\H,6.30565
34505,3.4257152688,-2.4092460188\H,6.9481823041,3.3450340047,-0.738023
3117\\Version=ES64L-G09RevD.01\State=1-A\HF=-1906.5712165\RMSD=3.778e-
09\RMSF=6.063e-08\Dipole=-0.3010365,3.3493345,0.4653221\Quadrupole=-16.8125553,-
8.660281,25.4728363,10.2932146,-9.6176933,2.9994049\PG=C01 [
X(C20H28N2O9P1)]\\@

Diazinon – Reactant

1\1\GINC-R3587\FOpt\RM062X\6-31+G(d,p)\C19H34N4O6P1S1(1-)\ROOT\06-Nov-
2014\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight) I
OP(2/17=4) Freq=noraman scrf=(pcm,read)\\opt fwd\\-1,1\C,8.720369,-0.8
81739,-0.199861\C,2.742497,-2.743864,-0.336681\C,0.96546,3.381487,1.53
3517\H,3.4347679949,-3.1046287338,-1.1054977464\H,3.078000486,-3.14503
89068,0.6269005271\H,1.7497914391,-3.1478942569,-0.5504157281\O,0.8397
66232,3.2520521713,0.1354938658\H,1.3147705707,2.4152218093,-0.1211573
416\C,2.6984934778,-1.2517500159,-0.3182926463\C,1.6614147261,-0.34867
35989,-0.3557950819\N,2.1633312343,0.9354984232,-0.3143098905\C,3.4778
159606,0.8041639463,-0.2457340293\N,3.8428369556,-0.4946831156,-0.2473
891904\H,4.8234713608,-0.8587446888,-0.2057870097\H,4.2011525317,1.610
8662997,-0.1924145236\H,0.5906576317,-0.5331186397,-0.4025873344\C,7.2
323715964,-0.5585742993,-0.0688764632\O,6.4286348179,-1.5308948774,-0.
2198572502\O,6.8901328626,0.6252146612,0.167017722\H,8.9779687647,-1.7
031285441,0.4759717871\H,8.9219137544,-1.2219781907,-1.2212347738\H,9.
3355759665,-0.0078264797,0.0221649877\P,-1.9830705213,1.9062250605,-0.
052270986\O,-1.2275636057,1.3395269646,1.2275906364\C,-0.9401320068,0.
9527828383,3.5697778414\H,-1.4636929742,0.7143093172,4.501241696\H,-0.
269736616,0.124485486,3.3193664381\H,-0.3415120033,1.854115679,3.72415

36367\O,-3.3323888413,0.9541198182,-0.1845393069\C,-3.3660758745,-0.40
54015014,-0.202232825\C,-4.6254185863,-1.0024921726,-0.3155696227\C,-4
.6386501357,-2.3874481649,-0.3065675589\C,-2.3587507709,-2.4295504351,
-0.0946291024\H,-5.5278425309,-0.4014704169,-0.40321499\O,-1.054465617
,1.4114454221,-1.2287234297\C,-0.5698007777,2.2690515657,-2.2944540541
\H,-0.4739902765,3.2886194433,-1.9116425128\H,0.4292155845,1.880258677
2,-2.5173745217\C,-1.5027816902,2.1796598407,-3.4832508986\H,-2.495536
6922,2.5631167095,-3.2227357375\H,-1.1045333627,2.7825050527,-4.306391
9487\H,-1.599445472,1.1438976855,-3.8234767756\C,-1.9520226814,1.16147
31748,2.4649433092\H,-2.6029255814,0.2834178367,2.3522423535\H,-2.5721
888756,2.0492176016,2.6505861029\H,0.0047271911,3.7069310703,1.9546599
939\H,1.2339984388,2.4225438111,1.9959414888\H,1.7315485405,4.12453171
94,1.7984975019\S,-2.6069217631,3.735420698,-0.0152801078\N,-2.2468343
155,-1.0876897541,-0.0993290734\N,-3.4988172895,-3.103512926,-0.191075
0945\C,-1.0536798839,-3.1736009,0.0680522407\H,-0.3726837462,-2.735383
4682,-0.6809261514\C,-0.4809255194,-2.8723630851,1.4616460043\H,0.4914
92348,-3.3610284096,1.5852127004\H,-0.3468032017,-1.794988745,1.604655
1378\H,-1.1565822959,-3.2527722444,2.238261817\C,-1.1710491466,-4.6733
090523,-0.1759615652\H,-1.5916437749,-4.8887962059,-1.1629259517\H,-0.
1786769066,-5.1326917036,-0.1111052861\H,-1.8177987531,-5.1411968371,0
.5740269419\C,-5.9132430221,-3.1707170902,-0.4203110274\H,-6.0284047,-
3.8163940902,0.4568186362\H,-6.7813847352,-2.5123787188,-0.5005931608\
H,-5.8687110828,-3.8186901915,-1.3022888374\\Version=ES64L-G09RevD.01\
State=1-A\HF=-2154.0826459\RMSD=7.564e-09\RMSF=2.287e-08\Dipole=-8.500
1559,-2.5968388,0.3744928\Quadrupole=-58.1386535,10.32314,47.8155134,1
0.5507499,-2.3283585,-0.8265479\PG=C01 [X(C19H34N4O6P1S1)]\\@

Diazinon – TS1

1\1\GINC-R2621\FTS\RM062X\6-31+G(d,p)\C19H34N4O6P1S1(1-)\ROOT\24-Oct-2
014\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultrafi
ne) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(p
cm,read) geom=check guess=read\\title\\-1,1\C,8.976608639,-2.353682110
9,-1.4505434878\C,4.5923313574,-2.7899544299,3.0002592616\C,1.80013000
36,2.7244755408,0.5568202262\H,4.7809635201,-3.7551931566,2.5170391573
\H,5.5193884546,-2.4764704397,3.4930028817\H,3.8255756743,-2.930795412
9,3.7658436591\O,1.2224741797,1.4563477316,0.2773493331\H,1.9869984751
,0.7268152496,0.4584015115\C,4.1371345726,-1.7705906513,2.0098369762\C
,2.9792980923,-1.0390797737,1.9157351283\N,3.0482143492,-0.2234981136,
0.8064136389\C,4.225446857,-0.4601791614,0.2448065639\N,4.9101017422,-
1.385489475,0.9397537533\H,5.8698554483,-1.7538700682,0.6900889407\H,4
.6172136641,0.0056403799,-0.652663222\H,2.1040674737,-1.0445350667,2.5
555838791\C,7.6726164667,-1.8579695789,-0.8312017426\O,7.3510325784,-2
.3704863058,0.2883589539\O,6.9955898646,-0.9977200237,-1.4419461673\H,
9.7939653109,-2.2216085121,-0.7347323527\H,8.8898329695,-3.42600787,-1
.6550148303\H,9.2012479146,-1.8205121005,-2.37605686\P,-0.4401924663,0
.8693216854,1.3649519856\O,0.5927600867,1.1385007627,2.5735394278\C,1.
3145136474,2.2184950784,4.5595945727\H,1.0386407348,2.4982681251,5.581
6184736\H,2.171863764,1.5382339923,4.6007620995\H,1.6096786003,3.12137
1428,4.0161385841\O,-1.6816693491,0.1644619907,2.3526162717\C,-1.62069
03847,-1.0246413116,2.9669283906\C,-2.8407113478,-1.6351659771,3.30181
65117\C,-2.7675985551,-2.8427324923,3.9699455649\C,-0.4831348449,-2.74

53524709,3.9275116841\H,-3.7846270127,-1.1657506001,3.0337985553\O,-0.
1450111062,-0.5263462578,0.6312535325\C,-0.02314751,-0.6735353111,-0.8
012590813\H,-0.3474020484,0.2500872392,-1.2930435386\H,1.0372920143,-0
.8413867024,-1.0209075738\C,-0.8787343063,-1.8538983418,-1.2115161451\
H,-1.9316461287,-1.6593042837,-0.9833417354\H,-0.7769069854,-2.0249389
984,-2.2884552062\H,-0.5641748227,-2.7590458988,-0.6827479778\C,0.1412
980411,1.5441762089,3.8806975936\H,-0.179289265,0.6531017957,4.4319811
394\H,-0.7099100875,2.229388444,3.7744712288\H,1.0156720044,3.48232239
92,0.491460893\H,2.2429718091,2.7325676252,1.5593920948\H,2.5727982103
,2.934369712,-0.1906393269\S,-1.5409302217,2.3048139458,0.5756150916\N
,-0.4527874778,-1.5712304254,3.2743671674\N,-1.579420727,-3.4041054869
,4.2878056237\C,0.8724279282,-3.3053417472,4.3052707018\H,1.4912739935
,-3.2376842883,3.3943574725\C,1.4967136351,-2.4022560897,5.3809374549\
H,2.4836779627,-2.7792977242,5.6711568449\H,1.6098004679,-1.3749653684
,5.0169283132\H,0.8623914901,-2.3881412521,6.2762518885\C,0.8228160838
,-4.7572530724,4.7684485746\H,0.3526868625,-5.4009754439,4.0188167614\
H,1.8410278738,-5.1186650759,4.94953494\H,0.2505702352,-4.8516841644,5
.6974112821\C,-3.9963210526,-3.6018709121,4.3797946763\H,-3.9960726482
,-3.7476822867,5.4653628851\H,-4.9075269197,-3.074299479,4.0878072808\
H,-3.9895792231,-4.5935784836,3.9147682335\\Version=ES64L-G09RevD.01\H
F=-2154.0725686\RMSD=5.449e-09\RMSF=4.191e-07\Dipole=-4.4931565,-0.898
2731,3.7647253\Quadrupole=-66.9819202,26.3720933,40.6098269,31.1481163
,10.8268777,5.3080331\PG=C01 [X(C19H34N4O6P1S1)]\\@

Dichlorvos – Reactant – R1

1\1\GINC-R3389\FOpt\RM062X\6-31+G(d,p)\C11H22Cl2N2O7P1(1-)\ROOT\18-Oct
-2014\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight)
IOP(2/17=4) Freq=noraman geom=check Guess=Read scrf=(pcm,read)\\title\
\-1,1\C,-8.5646983244,-0.8851690265,-0.2266836245\C,-2.6812274644,-2.9
827949176,0.2231375187\C,-0.5643942056,3.089420768,-1.4564997637\H,-3.
2731542477,-3.367577692,1.0610346411\H,-3.1376950469,-3.3506031631,-0.
7025081597\H,-1.6703354273,-3.3929624547,0.2968362091\O,-0.7176220736,
3.03474361,-0.0553678145\H,-1.2129065722,2.1970575775,0.1507086149\C,-
2.6252244245,-1.4900248317,0.2412648158\C,-1.5759651397,-0.6057202517,
0.3262993735\N,-2.0494430929,0.6885733384,0.2979827637\C,-3.3646482296
,0.5806440315,0.1948353537\N,-3.7525888124,-0.710888479,0.1586416731\H
,-4.7419444564,-1.0557752462,0.0808115179\H,-4.0740410362,1.3992172501
,0.1397703324\H,-0.5156564472,-0.8174984867,0.4012963431\C,-7.06392553
64,-0.6279652519,-0.1408584891\O,-6.3227446226,-1.6547731431,-0.022852
3878\O,-6.6472255904,0.5536489597,-0.1893413053\P,1.9805827949,1.57634
49924,0.6937341886\O,2.4797918865,2.9406729119,0.4391960542\O,1.243245
5944,0.8239522731,-0.5058099996\O,3.2110053011,0.5925357911,1.04445733
82\O,0.9845675285,1.3319942072,1.8958276624\H,0.4255000712,3.500580761
,-1.6908372047\H,-0.6382253305,2.0886031745,-1.9036579424\H,-1.3264945
026,3.7320361193,-1.9198618778\H,-8.892511466,-1.4143665092,0.67404170
24\H,-9.1205894926,0.04852796,-0.3296147891\H,-8.771728548,-1.53645841
51,-1.0820792477\C,0.30014565,2.3961397173,2.5912980884\H,0.5936466523
,2.3314951452,3.6409598768\H,0.569805766,3.36390501,2.1658346497\H,-0.
7717017828,2.2257751315,2.4752782207\C,1.9251635323,0.7529248822,-1.76
94836342\H,2.8394422289,0.1563440262,-1.6732141248\H,1.2318577214,0.27
13126564,-2.4606481259\H,2.1703266701,1.7591209278,-2.123059838\C,2.94

65531421,-0.7599586547,1.3896090913\H,2.2973632874,-1.2366511366,0.635
4121586\H,2.463613482,-0.8128014967,2.3786827124\C,4.2490988763,-1.526
45268,1.4462182449\H,4.0557357597,-2.577982221,1.720570568\Cl,5.057559
6547,-1.5426705452,-0.1505534366\Cl,5.3451729564,-0.8535984955,2.68706
72134\\Version=ES64L-G09RevD.01\State=1-A\HF=-2330.006379\RMSD=1.652e-
09\RMSF=1.170e-07\Dipole=6.3292884,-2.6079665,0.681621\Quadrupole=-74.
1293722,26.0544669,48.0749054,-16.8224451,-11.3488157,-0.6871054\PG=C0
1 [X(C11H22Cl2N2O7P1)]\\@

Dichlorvos – TS1 – R1

1\1\GINC-R3111\FTS\RM062X\6-31+G(d,p)\C11H22Cl2N2O7P1(1-)\ROOT\10-Oct-
2014\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultraf
ine) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(
pcm,read) geom=check guess=read\\title\\-1,1\C,8.9764442476,-2.3536611
493,-1.4504106073\C,4.5924513326,-2.7899462714,3.0001409106\C,1.800211
4199,2.7244004207,0.5568906967\H,4.7508483793,-3.7601675291,2.51700871
11\H,5.5277329266,-2.5057830349,3.4944877062\H,3.816952273,-2.90303691
17,3.762011912\O,1.4237948307,1.4439918791,0.0848291151\H,2.279366451,
0.6249595249,0.4404284418\C,4.1753518183,-1.7580603482,2.0065531906\C,
3.041802384,-0.9925088297,1.9258403309\N,3.1364381366,-0.1891099427,0.
8116509949\C,4.3008011977,-0.4624616691,0.2330590614\N,4.9527827533,-1
.4028091762,0.9273822669\H,5.9186934518,-1.8013536059,0.6629278957\H,4
.6915421861,-0.0081701229,-0.6710173329\H,2.1748015529,-0.9581362814,2
.5711504679\C,7.6637217143,-1.855525256,-0.8563185956\O,7.3133445362,-
2.3801503857,0.253000976\O,7.0038645775,-0.985577358,-1.4677086623\P,-
0.3359355224,0.8537971007,0.5873824449\O,-0.9650023455,2.1280259789,0.
1114623375\O,0.3619094635,0.7282616377,2.0713293989\O,-1.7742130646,0.
0872844591,1.1124424159\O,0.0208260178,-0.4312257854,-0.3419208157\H,1
.225511627,3.4921505673,0.0323939085\H,1.6205310203,2.8145684034,1.636
7905003\H,2.8696705946,2.8670798471,0.3613601817\H,8.8839327063,-3.421
3918794,-1.6754025879\H,9.2272723213,-1.8068121114,-2.3609052399\H,9.7
772287593,-2.2419828773,-0.7128426574\C,0.6193008348,-0.3294827237,-1.
6376700699\H,0.2077944977,-1.1552658146,-2.2231976363\H,0.3688459543,0
.623537143,-2.110567073\H,1.7052134013,-0.4288319619,-1.5649323731\C,-
0.3790642628,1.1917611222,3.2039168615\H,-1.039230001,0.4106795232,3.5
904726817\H,0.3613415147,1.4639693154,3.9606901678\H,-0.9813248529,2.0
691397914,2.9443225101\C,-1.7444955371,-1.2402348949,1.547569666\H,-1.
0377990802,-1.3894433723,2.386096984\H,-1.4597406178,-1.9277880314,0.7
340077878\C,-3.1138625767,-1.6756336449,2.0287836995\H,-3.0856222008,-
2.7206398262,2.3790307597\Cl,-3.6854709125,-0.6893168013,3.4147029045\
Cl,-4.3182141042,-1.6096608175,0.7051748533\\Version=ES64L-G09RevD.01\
HF=-2329.987966\RMSD=5.898e-09\RMSF=2.510e-07\Dipole=-1.970376,-1.5893
182,2.7198502\Quadrupole=-74.9333413,32.3949315,42.5384098,25.3990071,
14.9690666,-2.4616495\PG=C01 [X(C11H22Cl2N2O7P1)]\\@

Dichlorvos – Intermediate – R1

1\1\GINC-R2546\FOpt\RM062X\6-31+G(d,p)\C11H22Cl2N2O7P1(1-)\ROOT\30-Oct
-2014\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight)
IOP(2/17=4) Freq=noraman geom=check Guess=Read scrf=(pcm,read)\\opt fw
d\\-1,1\C,7.752456989,-2.7673061801,-0.770205398\C,2.3102356245,-3.311
8344724,2.2797297587\C,1.1055212121,3.0834346925,0.9319130393\H,2.4603

315105,-4.1759171357,1.622686301\H,3.1136339927,-3.3165367694,3.024994
2206\H,1.3563924574,-3.4363004519,2.7999783581\O,0.6019562883,2.276572
391,-0.1121890626\H,1.3212446943,0.7482156644,0.2568073973\C,2.3117724
42,-2.0386550821,1.4984753265\C,1.3245029236,-1.094215168,1.379118067\
N,1.819354394,-0.118142762,0.5485966482\C,3.0632214239,-0.4812496886,0
.1912217702\N,3.3969722128,-1.6380897935,0.7465641949\H,4.7387108089,-
2.3211716338,0.5023993503\H,3.7016630706,0.10080248,-0.465299943\H,0.3
303430195,-1.0293510109,1.797897784\C,6.4193697897,-2.1169629087,-0.48
82847661\O,5.6620131588,-2.8087167445,0.3363529166\O,6.0786714434,-1.0
483510305,-0.9815202088\P,-1.1499382737,2.0327125219,-0.1341829295\O,-
1.5459210469,3.4798838407,0.0313741428\O,-0.8971356227,1.0768610184,1.
2029619319\O,-0.9347336649,1.2582886271,-1.5672758444\H,0.7531985904,4
.1152212297,0.8299924768\H,0.7980669212,2.6995096766,1.9155146279\H,2.
2003985553,3.0636837097,0.8738433923\H,8.3332077721,-2.1490719583,-1.4
554587899\H,8.2962382044,-2.9021510498,0.1702954098\H,7.5873784108,-3.
7589181804,-1.2036909964\C,0.1565811775,1.4298907688,-2.4658032331\H,1
.0484742144,0.9049415685,-2.1123973552\H,-0.1732579488,0.9957290442,-3
.414002036\H,0.3933346088,2.4878175801,-2.6048706151\C,-1.8318393872,1
.1153469469,2.278107542\H,-2.5736624336,0.3164320866,2.1855278109\H,-1
.253498348,0.9756670806,3.1975194922\H,-2.3534158537,2.077615018,2.314
2516698\O,-2.8149110201,1.4721569081,-0.226113007\C,-3.1111133101,0.13
70172842,-0.4934828592\H,-2.8146380669,-0.1511717626,-1.5162070504\H,-
2.6299312971,-0.5655075726,0.2146312072\C,-4.6216353336,-0.0248414597,
-0.4238969719\H,-5.1379596462,0.6937402443,-1.0844482565\Cl,-5.2625580
801,0.2517838665,1.2280408009\Cl,-5.0722837083,-1.6764820899,-0.970113
3382\\Version=ES64L-G09RevD.01\State=1-A\HF=-2329.9941199\RMSD=9.430e-
09\RMSF=6.047e-08\Dipole=0.9407887,-2.7688734,0.5464369\Quadrupole=-24
.4752827,-3.6908137,28.1660965,13.838792,2.8503861,0.8362227\PG=C01 [X
(C11H22Cl2N2O7P1)]\\@

Dichlorvos – TS2 – R1

1\1\GINC-R3170\FTS\RM062X\6-31+G(d,p)\C11H22Cl2N2O7P1(1-)\ROOT\23-Oct-
2014\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultraf
ine) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(
pcm,read)\\title\\-1,1\C,8.497898,-2.33561,-0.981512\C,3.056096,-2.880
267,2.068242\C,1.851219,3.515171,0.720459\H,3.2998419663,-3.8249378694
,1.5699979775\H,3.8403357841,-2.6787978389,2.8061431122\H,2.106867308,
-3.0036144267,2.5967375076\O,1.3297332004,3.0570310574,-0.5227488469\H
,1.8050203868,0.6396847875,-0.6850599935\C,2.9609262934,-1.7669467131,
1.0755758908\C,1.9037283695,-0.9461398005,0.7727243235\N,2.3516931564,
-0.1013629527,-0.2150610866\C,3.6306727855,-0.4160396328,-0.484682366\
N,4.0323016843,-1.4183583443,0.2796964138\H,5.4415109019,-2.0068969686
,0.1924421909\H,4.238620377,0.0879522902,-1.2275702387\H,0.8956782452,
-0.8899948551,1.1595717239\C,7.0969048928,-1.8031491829,-0.8217059818\
O,6.411839261,-2.4091601736,0.1257709402\O,6.642085567,-0.8963921714,-
1.5079260664\P,-0.2739086507,2.6635201377,-0.5548272367\O,-1.067106964
5,3.8807374923,-0.2451705572\O,-0.1450984023,1.4908269472,0.5379731046
\O,-0.1653652395,2.0705210198,-2.039336482\H,1.2506708564,4.3448562173
,1.1121267845\H,1.8722949972,2.6982812942,1.4521775399\H,2.8699540556,
3.8613331645,0.5321903604\H,9.0204103493,-1.7869340075,-1.7655518769\H
,9.0339723487,-2.2408017945,-0.0319743785\H,8.453727587,-3.4002566423,

-1.2324828756\C,-1.0186178085,2.5801341354,-3.0762078726\H,-0.76219538
96,3.6237376026,-3.2846804511\H,-0.817225038,1.9670858825,-3.956493609
7\H,-2.0613256531,2.4960371675,-2.7626264676\C,-1.0992063682,1.3410129
673,1.5995074193\H,-1.7187231017,0.4642241186,1.406896486\H,-0.5149692
203,1.214474717,2.5165981082\H,-1.7417920565,2.2207299298,1.6672594946
\O,-2.4251841918,1.5856384524,-0.7587699908\C,-2.3199803314,0.31086534
85,-1.2035343605\H,-1.9882446702,0.216500545,-2.294857622\H,-1.5947184
243,-0.346308468,-0.6069269698\C,-3.6769989935,-0.4006034113,-1.183673
8744\H,-4.4463956765,0.1613636179,-1.7369571469\Cl,-4.3121249151,-0.58
95576826,0.4923703301\Cl,-3.5616144501,-2.0333952048,-1.9513634384\\Ve
rsion=ES64L-G09RevD.01\HF=-2329.9815473\RMSD=7.687e-09\RMSF=4.186e-07\
Dipole=2.6466531,-1.6917821,0.7280957\Quadrupole=-25.6786893,-2.796526
3,28.4752156,9.1753052,5.7957105,-0.2498272\PG=C01 [X(C11H22Cl2N2O7P1)
]\\@

Dichlorvos – Reactant – R2

1\1\GINC-R3569\FOpt\RM062X\6-31+G(d,p)\C11H22Cl2N2O7P1(1-)\ROOT\19-Nov
-2014\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT IOP(2/17=4) Fr
eq=noraman scrf=(pcm,read)\\opt fwd\\-1,1\C,8.278832,-2.816232,0.01279
6\C,2.029763,-3.154469,-0.224402\C,1.7669454024,2.9637230762,2.3590169
165\H,2.322644,-3.532974,-1.2102\H,2.508987,-3.78795,0.530264\H,0.9454
57,-3.255885,-0.121551\O,1.97177519,3.0806527587,0.9682818591\H,2.1769
348147,2.1686927306,0.6267296696\C,2.420061,-1.722881,-0.054268\C,1.67
8900314,-0.5849525524,0.1606925694\N,2.5144578636,0.5082852264,0.25189
10837\C,3.7387768827,0.0291527888,0.0982054882\N,3.727567,-1.30666,-0.
090623\H,4.575041,-1.915126,-0.224035\H,4.659299,0.601473,0.121917\H,0
.60448,-0.485444,0.262615\C,6.914853,-2.140129,-0.086447\O,5.929491,-2
.899047,-0.354898\O,6.832795,-0.90345,0.102267\P,-0.9657894503,2.63751
43971,-0.0039013647\O,-1.1334716757,4.0400989152,0.4222276464\O,-0.554
5642195,1.549706174,1.074751615\O,0.0249966939,2.3323757133,-1.1906381
435\H,0.9104482176,3.5853896518,2.6507176788\H,1.5502050759,1.92373402
25,2.6416220635\H,2.648344318,3.2988304542,2.9233097557\H,8.259972,-3.
540751,0.833822\H,8.479157,-3.371238,-0.90907\H,9.070351,-2.085145,0.1
87261\O,-2.4072674273,2.0760366435,-0.5398888375\C,-2.4704854391,0.757
6665998,-1.0627641352\H,-2.058147761,0.0287964392,-0.3442274214\H,-1.9
125792243,0.6919783391,-2.0113622492\C,-3.9137972518,0.3911530923,-1.3
27049678\H,-3.9760623979,-0.6339664893,-1.7310649132\Cl,-4.8711062987,
0.4164384684,0.1853074288\Cl,-4.6483215004,1.4785673676,-2.540872668\C
,-1.3303843346,1.4645627299,2.2828330148\H,-0.8480009162,0.703168325,2
.8975532136\H,-1.3238154209,2.4275390861,2.8025088973\H,-2.3601260257,
1.1704085938,2.0504684219\C,0.9395861733,3.3119934597,-1.7253386657\H,
0.9167986017,4.2218710637,-1.1242535065\H,1.9365954604,2.8696140286,-1
.6917425522\H,0.6361194817,3.511795005,-2.7551663668\\Version=ES64L-G0
9RevD.01\State=1-A\HF=-2330.0052125\RMSD=6.479e-09\RMSF=4.158e-06\Dipo
le=-6.5248525,-0.5820178,-0.4157963\Quadrupole=-54.9238902,8.0662433,4
6.8576469,41.996131,-3.7373881,-3.8830418\PG=C01 [X(C11H22Cl2N2O7P1)]\
\@

Dichlorvos – TS1 – R2

1\1\GINC-R3588\FTS\RM062X\6-31+G(d,p)\C11H22Cl2N2O7P1(1-)\ROOT\20-Nov-
2014\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultraf

ine) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(
pcm,read)\\title\\-1,1\C,8.976771,-2.353708,-1.45067\C,4.59223,-2.7899
83,3.000379\C,1.800106,2.724484,0.556912\H,4.699802,-3.783495,2.551706
\H,5.559882,-2.512089,3.432057\H,3.856624,-2.850226,3.806723\O,1.42881
54977,1.4526569749,0.045732699\H,2.2472710875,0.6205527193,0.402132084
4\C,4.149523,-1.787903,1.986481\C,3.035192,-0.990062,1.922614\N,3.0931
13,-0.232527,0.773047\C,4.220671,-0.56775,0.158064\N,4.882302,-1.50046
2,0.858979\H,5.816239,-1.936334,0.574145\H,4.586439,-0.163139,-0.77938
6\H,2.204825,-0.907655,2.611115\C,7.597251,-1.978878,-0.921361\O,7.227
304,-2.562402,0.150014\O,6.914273,-1.131031,-1.540153\P,-0.3438587699,
0.8674664427,0.4737985516\O,-0.9519084139,2.1629809967,0.0283443023\O,
0.3228615189,0.6764978531,1.9643709214\O,0.0148080784,-0.3852107647,-0
.5043984135\H,1.253059,3.502842,0.021585\H,1.57344,2.786449,1.628502\H
,2.877365,2.8503,0.405733\H,9.733269,-2.048581,-0.719818\H,9.045458,-3
.440486,-1.55924\H,9.175641,-1.868308,-2.407693\O,-1.800903223,0.09420
30016,0.934756401\C,-1.8045800681,-1.2559785512,1.3104497301\H,-2.4614
66901,-1.3875842743,2.1866315339\H,-0.7948028025,-1.626502127,1.560483
4923\C,-2.3352432581,-2.1259973932,0.1773123953\H,-1.7548604314,-2.009
6901957,-0.749610081\Cl,-4.0326871996,-1.7091814052,-0.2109873123\Cl,-
2.2304822128,-3.8501806909,0.6676915712\C,-0.4686689892,0.9126598645,3
.1307213272\H,-1.0460680641,0.0246936572,3.4046025272\H,0.2379149621,1
.1511108486,3.9299126498\H,-1.1506687388,1.7552650346,2.97781353\C,0.6
702726097,-0.2493192016,-1.7687347242\H,1.7499292674,-0.3747186571,-1.
6559397062\H,0.2700414613,-1.0460481048,-2.401490269\H,0.45766397,0.72
43565958,-2.217437069\\Version=ES64L-G09RevD.01\HF=-2329.9909861\RMSD=
9.486e-09\RMSF=1.325e-06\Dipole=-1.756344,-0.7080237,2.4032618\Quadrup
ole=-68.5267651,20.4300468,48.0967183,22.5645996,12.0508359,-0.169208\
PG=C01 [X(C11H22Cl2N2O7P1)]\\@

Parathion – Reactant – R1

1\1\GINC-R3464\FOpt\RM062X\6-31+G(d,p)\C15H23N3O8P1S1(1-)\ROOT\18-Oct-
2014\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight) I
OP(2/17=4) Freq=noraman geom=check Guess=Read scrf=(pcm,read)\\opt fwd
\\-1,1\C,-7.9226506152,0.9122948864,-0.2310175923\C,-2.0864747517,2.65
00288348,1.2304176962\C,-0.5057413492,-3.7758867978,1.8481690444\H,-2.
1827896492,3.3069106655,0.3584377119\H,-2.9074626788,2.8856627093,1.91
66068486\H,-1.1403640568,2.8784100228,1.7296812704\O,-0.8298765832,-3.
5742929647,0.4897195176\C,-2.1126555368,1.2102564272,0.8301915624\C,-1
.1891882393,0.2002324268,0.9647822633\N,-1.6816280708,-0.9670940677,0.
4202125056\C,-2.8886362462,-0.6616633327,-0.0279858403\N,-3.1861750072
,0.6359508814,0.1945609188\H,-4.0953226252,1.1004228806,-0.038391295\H
,-3.5847940019,-1.3384537551,-0.5115689916\H,-0.202259927,0.2487409548
,1.4115098644\C,-6.4230729648,0.780652623,-0.4816626305\O,-5.696229241
2,1.7591830921,-0.1231893922\O,-5.9925186862,-0.2824499057,-0.99078051
36\H,0.5462054292,-4.0793981318,1.9475760523\H,-0.6548791534,-2.856913
3216,2.4320012004\H,-1.135882503,-4.5642684081,2.27965818\H,-8.4847585
1,0.1939935453,-0.8312419013\H,-8.1134285627,0.7108802781,0.8294971531
\H,-8.2568626271,1.9316879719,-0.4415113908\H,-1.100706703,-2.62326909
61,0.3949114153\P,1.9976193386,-2.2590270178,-0.495751794\O,1.37280644
92,-1.6701252471,0.8498219891\O,3.1003652544,-1.1159920152,-0.90035760
99\C,2.9135458773,0.2405425104,-0.7725857946\O,0.8504886619,-1.9902020

545,-1.5470770888\S,2.8370845176,-3.9915175646,-0.4271750657\C,0.07457
86778,-2.9999509696,-2.2257819871\H,0.3559716096,-2.9705058149,-3.2802
434057\H,0.2594991968,-3.9831666924,-1.7919889329\H,-0.9740372617,-2.7
2858413,-2.0920744721\C,2.1852517164,-1.5796703906,2.0299066192\H,3.00
55960652,-0.8694627918,1.8719855946\H,1.5246551408,-1.2189700234,2.820
1638324\H,2.5825079172,-2.5656665734,2.2915765733\N,2.6344827972,4.414
7726408,-0.3445351087\O,3.6707942351,5.0549039639,-0.2429374164\O,1.52
26528397,4.9182312995,-0.3188007222\C,3.9935832769,2.3708124378,-0.474
4670586\H,4.8833280782,2.9793003456,-0.3427950872\C,1.6558811396,0.846
4244456,-0.8182317699\H,0.7510815055,0.2600585406,-0.9604530827\C,4.07
99112016,0.9934512086,-0.6068437417\H,5.0413072437,0.4817822579,-0.582
150181\C,1.5671772984,2.2256547677,-0.6718630347\H,0.5993021604,2.7197
517571,-0.6960485758\C,2.7332924405,2.9641645014,-0.5020206493\\Versio
n=ES64L-G09RevD.01\State=1-A\HF=-2090.6330549\RMSD=2.605e-09\RMSF=4.42
7e-08\Dipole=6.5488251,-0.3181561,1.1528097\Quadrupole=-38.1059334,-9.
8460106,47.951944,5.9332652,-2.5942298,-0.0471987\PG=C01 [X(C15H23N3O8
P1S1)]\\@

Parathion – TS1 – R1

1\1\GINC-R3409\FTS\RM062X\6-31+G(d,p)\C15H23N3O8P1S1(1-)\ROOT\14-Oct-2
014\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultrafi
ne) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(p
cm,read) geom=check guess=read\\title\\-1,1\C,8.9763293669,-2.35364028
12,-1.450323415\C,4.5925150141,-2.7899167399,3.0000551278\C,1.80026261
91,2.724350021,0.5568892872\H,4.6512704938,-3.8027631014,2.5867756428\
H,5.5856887761,-2.5265526822,3.3799890871\H,3.89097591,-2.7981511428,3
.8385130928\O,1.4050438162,1.5040287039,-0.0518889106\C,4.1358067016,-
1.8131168621,1.9669935417\C,3.0369435648,-0.9913422363,1.9039384448\N,
3.0653924561,-0.277965123,0.7247420125\C,4.1654463585,-0.6631015346,0.
0930514847\N,4.8353214537,-1.5847037306,0.8069843035\H,5.7499628127,-2
.0391992622,0.5181724617\H,4.513716052,-0.30308232,-0.8688294391\H,2.2
347175192,-0.8692498804,2.6208922744\C,7.5602212966,-2.0681149301,-0.9
642458354\O,7.1977425468,-2.6677516218,0.0989826423\O,6.8467257293,-1.
2634941181,-1.607809067\H,1.3342258861,3.5522563323,0.0190737778\H,1.4
908964896,2.753323622,1.6098462709\H,2.8909299732,2.8127883798,0.49858
09274\H,9.145875553,-1.927363696,-2.4407742551\H,9.6864407843,-1.91471
74347,-0.7407323078\H,9.1535469524,-3.4330008045,-1.4673846513\H,2.083
021169,0.7570667523,0.2647617799\P,-0.5243084236,0.7399722592,0.428556
3504\O,0.2086210892,0.6298409505,1.8709930448\O,-1.9291923745,-0.11069
76502,0.9138386122\C,-1.9952984696,-1.3606875156,1.4220885719\O,-0.019
7100372,-0.4559358505,-0.5111779676\S,-1.379936521,2.3803597389,-0.225
4902531\C,0.5485427361,-0.2820969707,-1.8195042898\H,0.1475413766,-1.0
955534818,-2.4269900857\H,0.2612037415,0.6847825998,-2.2403223305\H,1.
6357595893,-0.3551873719,-1.7525919059\C,-0.5169118486,0.9643809081,3.
0634318192\H,-1.149228259,0.1284859265,3.3780068886\H,0.2411250472,1.1
653706569,3.8230020238\H,-1.1294052526,1.858485989,2.903816109\N,-2.57
51318666,-5.1881938929,3.1063420096\O,-3.7171852091,-5.5408366593,3.37
08126567\O,-1.5926503671,-5.8878825168,3.3104798329\C,-3.4883168412,-3
.0568481583,2.2953496423\H,-4.4845412196,-3.4075137682,2.547971743\C,-
0.8899770235,-2.1966471744,1.6543665176\H,0.1159767054,-1.87705074,1.3
970576672\C,-3.2904354689,-1.8030241028,1.7457132441\H,-4.1284697705,-

1.1347061052,1.5527157642\C,-1.0874184227,-3.4544678233,2.208116181\H,
-0.2428989872,-4.1126613303,2.3920617793\C,-2.3779564029,-3.8701690542
,2.5230998577\\Version=ES64L-G09RevD.01\HF=-2090.6224823\RMSD=4.343e-0
9\RMSF=4.387e-07\Dipole=-3.5570681,1.2880723,2.2555497\Quadrupole=-53.
5997207,4.0365279,49.5631928,17.7070271,19.2406293,21.0536325\PG=C01 [
X(C15H23N3O8P1S1)]\\@

A.1.2 Dephosphorylation reactions – small systems

dECP – Reactant

1\1\GINC-R2526\FOpt\RM062X\6-31+G(d,p)\C14H30N2O9P1(1-)\ROOT\17-Oct-20
14\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight) IOP
(2/17=4) Freq=noraman geom=check Guess=Read scrf=(pcm,read)\\opt fwd\\
-1,1\C,7.5850454665,2.9237773236,-0.5780997053\C,4.1286791882,-0.94539
26184,-1.9093549759\C,-5.0998002355,3.1133839618,1.5463189438\C,-1.443
3225479,0.5383958838,2.6986710519\C,6.5825714478,1.8302670549,-0.30625
56123\O,6.8957865278,0.6714370778,-0.0739775539\O,5.3323772846,2.25400
17057,-0.3435141368\H,4.663619279,1.4903991028,-0.133781919\N,3.604080
7755,0.3529271173,0.1747708741\C,3.4619443516,-0.792073456,-0.58188450
53\C,2.8844803829,0.1748861337,1.2692607039\C,2.6395839393,-1.66271084
92,0.0856775759\H,2.769885662,0.8890992447,2.07734577\N,2.2879046975,-
1.032809097,1.2554834615\H,2.2716587671,-2.64791777,-0.1728994065\H,1.
6136961933,-1.3866077579,1.9562908091\O,-0.8434697566,-0.5518279996,1.
9630260042\P,-1.4684877275,-0.9019188096,0.5360018712\O,-1.2388289921,
0.3624960874,-0.3888628059\O,-0.4227503508,-2.0299170189,0.1224984072\
O,-2.8964687021,-1.3027018101,0.5017595819\C,0.0687908674,0.985884401,
-0.4824751742\C,-0.6546302263,-2.809351056,-1.0792220622\H,0.830903228
1,0.1981236536,-0.5162672116\C,0.0925422913,1.8338550413,-1.7315441344
\H,0.2094921178,1.5990664258,0.41752161\C,-0.3103275486,-2.0215923444,
-2.3293813773\H,-1.6997793846,-3.1436713193,-1.084282318\H,-0.00608303
02,-3.6828873187,-0.963158908\H,-0.6996925892,2.5868660548,-1.68683937
93\H,-0.0585418783,1.2111384305,-2.6197077051\H,1.0663555721,2.3297707
522,-1.8075798509\H,-0.3771775534,-2.6823070323,-3.2002802019\H,0.7106
884567,-1.6275728765,-2.264873671\H,-1.0052370479,-1.1880675611,-2.478
2167615\H,-3.6866008382,-0.4072795862,-1.1786547594\O,-3.9127122906,0.
0083304672,-2.0253263884\H,-3.6527273415,0.9452673041,-1.9145139768\O,
-3.2488017727,2.7325033337,-1.383280931\C,-3.2171820567,2.7549371276,-
0.1190788647\C,-4.3938872251,2.1121806614,0.6279852301\O,-2.3013623211
,3.2673305766,0.5792101991\H,-4.0005760333,1.2740432186,1.2216038596\H
,-5.0976974372,1.6970738356,-0.1012749402\H,7.3812553957,3.3691047692,
-1.5569707645\H,7.4733602952,3.7116384457,0.1736520706\H,8.5966867622,
2.518300277,-0.5508003168\H,3.7884960675,-0.1690920314,-2.6042815844\H
,5.2160033168,-0.8558288269,-1.8079201179\H,3.8971018654,-1.9224175747
,-2.3426067814\H,-5.9282556593,2.6393267916,2.0832449961\H,-4.39707567
29,3.5205977658,2.2800605805\H,-5.5055262691,3.9501454679,0.9666632071
\H,-0.778706998,0.7369905489,3.5401185371\H,-1.5317034333,1.4299403485
,2.0658021635\H,-2.4300197703,0.2347929173,3.0611942633\\Version=ES64L
-G09RevD.01\State=1-A\HF=-1679.2231653\RMSD=2.362e-09\RMSF=1.493e-07\D
ipole=5.1416678,-4.0084401,2.1350502\Quadrupole=-50.77739,16.3933927,3
4.3839974,19.9180734,-17.6585551,6.2886421\PG=C01 [X(C14H30N2O9P1)]\\@

dECP – TS1

1\1\GINC-R3263\FTS\RM062X\6-31+G(d,p)\C14H30N2O9P1(1-)\ROOT\11-Oct-201
4\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultrafine
) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(pcm
,read)\\title\\-1,1\C,7.9086739711,2.2958475755,-0.1300679914\C,4.7256
276275,-1.8810782226,-1.1755180302\C,-4.8591851078,2.0251721745,1.3953
336034\C,-1.1524974908,-0.2017545274,2.9977534183\C,6.7883278105,1.305
1136883,0.0616466089\O,6.9410511769,0.214006012,0.5934959015\O,5.63189
47419,1.7368683573,-0.4015563633\H,4.8629043268,1.0421763547,-0.245456
0751\N,3.6543228453,0.0932572539,-0.0601667787\C,3.6009141951,-1.22961
73301,-0.4440001126\C,2.5179033786,0.3484094791,0.5695370632\C,2.40349
40917,-1.7585937812,-0.0331005713\H,2.2340217548,1.3016240649,1.003123
6317\N,1.7357328354,-0.7437851502,0.6068441478\H,1.9792282459,-2.75140
8171,-0.1314917407\H,0.803376068,-0.8110318569,1.0388252242\O,-0.80833
92981,-0.7796747784,1.7474914354\P,-2.0677752063,-1.1986081379,0.71068
27529\O,-2.3632823,0.2943211365,0.1698828255\O,-0.9915580104,-1.981639
5046,-0.1916781919\O,-3.0270112359,-1.8670398709,1.6495062409\C,-1.306
2130977,1.2499219152,-0.0312766715\C,-1.0343431619,-2.3761550972,-1.57
51814454\H,-0.4140464654,0.7349500371,-0.4104060945\C,-1.8015319793,2.
2811625522,-1.0238091305\H,-1.0541506311,1.7092585504,0.9346672867\C,-
0.9879887921,-1.1806907792,-2.5056908479\H,-1.9183132402,-2.9901841964
,-1.7525070255\H,-0.135490765,-2.9914446691,-1.6931913758\H,-2.6932008
103,2.7919339339,-0.6471465413\H,-2.0660715916,1.8047963632,-1.9733160
258\H,-1.0169260164,3.0247190757,-1.2015995563\H,-0.9528011888,-1.5296
852622,-3.5436033278\H,-0.0913781313,-0.5799081113,-2.3138910094\H,-1.
8782041243,-0.5545285839,-2.383726629\H,-4.1379393921,-2.1746824223,-0
.2846766798\O,-3.485300979,-1.6442126434,-0.7766909806\H,-4.0035848967
,-0.7813162965,-1.1724482259\O,-4.6398327811,0.2786469849,-1.762192725
4\C,-5.0766238832,1.2300591164,-1.0205465375\C,-5.3440441292,0.9222569
889,0.456233349\O,-5.3409080724,2.3702183248,-1.4443124244\H,-4.894365
1137,-0.03695277,0.7306622881\H,-6.4352239219,0.8140639232,0.548910458
4\H,7.6660153257,3.2226797766,0.3998963955\H,8.8440474535,1.8818982766
,0.2477398524\H,8.0036012675,2.5376409416,-1.1934295874\H,4.9356081896
,-1.3568282676,-2.1147977555\H,5.6399069711,-1.8628963403,-0.572657728
\H,4.4777602762,-2.9204192412,-1.4084950817\H,-5.1515275581,1.81053344
92,2.4288226961\H,-3.7669479946,2.1002187075,1.3618881271\H,-5.2802917
961,2.9922776091,1.1049950731\H,-0.2251919167,0.157699654,3.453375648\
H,-1.8427490749,0.6431548874,2.8657076021\H,-1.6248904174,-0.945778621
6,3.6460793669\\Version=ES64L-G09RevD.01\HF=-1679.2028226\RMSD=5.517e-
09\RMSF=1.025e-06\Dipole=4.3140892,-1.0179467,1.0586823\Quadrupole=-62
.1880597,31.5352106,30.6528491,21.0447779,-24.6712186,18.4098657\PG=C0
1 [X(C14H30N2O9P1)]\\@

dECP – Intermediate

1\1\GINC-R3592\FOpt\RM062X\6-31+G(d,p)\C14H30N2O9P1(1-)\ROOT\13-Sep-20
14\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight) IOP
(2/17=4) Freq=noraman scrf=(pcm,read)\\opt fwd\\-1,1\C,7.6584226546,1.
6273188129,2.0636777003\C,4.7288098805,-0.2531087748,-2.0046585725\C,-
5.0883455744,-0.0841344258,2.1606451298\C,-1.278421359,-2.6237418702,1
.6157163678\C,6.6153136374,0.8004751389,1.3536476697\O,6.8615047394,-0

.2721043298,0.817957226\O,5.4188564131,1.3511783932,1.370200001\H,4.69
8654633,0.7509347614,0.8873650803\N,3.5747995885,-0.0337164577,0.21358
73698\C,3.5775354794,-0.4991765661,-1.0839824298\C,2.426801646,-0.4227
832464,0.750446363\C,2.4019334115,-1.172506,-1.30735186\H,2.1028662147
,-0.218111558,1.7655542298\N,1.6902814218,-1.1139550288,-0.1345389729\
H,2.0185015117,-1.6822708119,-2.1810089835\H,0.7412873182,-1.503514416
9,0.0310346273\O,-0.8591836711,-1.9940807151,0.4244382553\P,-2.1093956
433,-1.420057175,-0.6638836043\O,-2.3678577077,-0.0903582133,0.2723136
891\O,-0.854086517,-1.2318875052,-1.7012912781\O,-2.9725215523,-2.6665
216418,-0.5759819818\C,-1.3500239748,0.5605912755,1.0342116269\C,-0.79
58252799,-0.3569413683,-2.832047946\H,-0.4115397502,0.571531142,0.4652
327104\C,-1.8254397042,1.9740856172,1.311114176\H,-1.1744716823,0.0126
699869,1.9700766445\C,-0.687244898,1.0959424125,-2.4051155054\H,-1.660
2381254,-0.5168329025,-3.4805934021\H,0.1104184128,-0.66573801,-3.3667
00796\H,-2.7680801501,1.9673337574,1.8688594973\H,-1.9924009933,2.5140
164756,0.3730685502\H,-1.0748186874,2.5115910804,1.9009399187\H,-0.575
1516405,1.7382624973,-3.285780603\H,0.1879393587,1.2362110177,-1.75934
16221\H,-1.5865910277,1.4055959171,-1.8614667643\H,-3.8304765944,-1.32
70376251,-2.1320062101\O,-3.2356125335,-0.621541512,-1.8011342103\H,-3
.9790316452,0.5925846736,-1.3102692232\O,-4.4303109898,1.4755283646,-1
.0380709215\C,-5.070860491,1.4039506257,0.1132226622\C,-5.3952816946,0
.0276825173,0.6683112395\O,-5.4443033937,2.4230125507,0.678481235\H,-4
.870446655,-0.7511153213,0.1089757697\H,-6.4741798324,-0.106858486,0.4
991722219\H,7.3622552943,1.7640454846,3.1087399764\H,8.629218303,1.133
5929918,2.0100375876\H,7.7125120904,2.618673827,1.602488139\H,4.914378
8104,0.8211693199,-2.1169241491\H,5.6431602945,-0.7099305119,-1.611035
4544\H,4.5223070905,-0.6733243585,-2.993007888\H,-5.4351267173,-1.0477
811262,2.545894455\H,-4.0085642664,-0.0191639441,2.3296701785\H,-5.580
4890713,0.7164348734,2.7208905995\H,-0.411504159,-2.6877327534,2.28433
10537\H,-2.0707721644,-2.0489952069,2.119143482\H,-1.6632880579,-3.629
0376516,1.4136749452\\Version=ES64L-G09RevD.01\State=1-A\HF=-1679.2101
047\RMSD=3.420e-09\RMSF=5.815e-06\Dipole=2.1713201,0.0616921,-0.017876
9\Quadrupole=-47.8699944,5.5619931,42.3080013,20.9263299,-2.4856923,-2
.4039338\PG=C01 [X(C14H30N2O9P1)]\\@

dECP – TS2

1\1\GINC-R2519\FTS\RM062X\6-31+G(d,p)\C14H30N2O9P1(1-)\ROOT\03-Sep-201
4\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultrafine
) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(pcm
,read) geom=check\\title\\-1,1\C,6.8137095695,3.0140470528,0.405824201
\C,4.6577426598,-1.2222744213,-2.5986771585\C,-5.8511978577,-0.7182027
502,-0.3776411928\C,-0.2123993716,-2.2782908812,2.5557991503\C,6.03412
79586,1.9273868077,-0.3208679651\O,6.490362114,0.7637344897,-0.3424014
369\O,4.9268903891,2.2854974092,-0.8462179975\H,4.0269664225,0.9902583
323,-0.9635596881\N,3.4396900279,0.1027368456,-0.8505656234\C,3.642472
6819,-1.0901879297,-1.5027399698\C,2.4792556064,-0.0774008516,0.066853
2783\C,2.7561527643,-1.9759429021,-0.9353652447\H,2.1230679634,0.70651
68561,0.7259929867\N,2.0379143235,-1.3284362267,0.0493118133\H,2.58175
80774,-3.0180199098,-1.1707781312\H,0.9467463322,-1.7209263502,0.70712
13758\O,-0.0855432537,-1.9493883798,1.1795424585\P,-1.5793579859,-1.87
01113995,-0.0430105347\O,-1.3994284334,-0.2584586051,-0.0688980132\O,-

0.4828837816,-2.5709929693,-1.0118910993\O,-2.4310282251,-2.6400388608
,0.9310951869\C,-1.2256810458,0.6008419267,1.0616460562\C,-0.414318511
7,-2.4096737152,-2.439004157\H,-0.9765829007,1.5742972525,0.6254647756
\C,-2.4982626853,0.6915698407,1.8861216149\H,-0.376553205,0.260129526,
1.6624726602\C,-0.1218475248,-0.979679362,-2.8564425136\H,-1.339355807
6,-2.7743665077,-2.8951504352\H,0.410293194,-3.0687962227,-2.732260658
9\H,-2.7691817199,-0.2871818503,2.2966913716\H,-3.3245349642,1.0492029
537,1.2623453963\H,-2.3558537362,1.3936225651,2.7149620984\H,0.0588706
704,-0.9509635884,-3.9368208182\H,0.7709894354,-0.6060710696,-2.343462
77\H,-0.9614915012,-0.318215253,-2.625594925\H,-3.4308718451,-2.387880
2684,-1.1475271146\O,-2.750963459,-1.696745655,-1.2990756087\H,-3.2576
491857,-0.1366998719,-1.5410712892\O,-3.5531861304,0.7904246061,-1.770
6773087\C,-4.8501363642,0.8702766184,-2.0417152834\C,-5.6626956317,-0.
3948845032,-1.8667799589\O,-5.34328171,1.9330689679,-2.3754143774\H,-5
.1522626341,-1.2221197804,-2.3768964979\H,-6.6298366274,-0.2315654414,
-2.35106225\H,6.2932705316,3.2500217626,1.3409734761\H,7.8282115358,2.
6828445611,0.6353785309\H,6.8383628281,3.9267799858,-0.1962630432\H,4.
4087236389,-0.5641212217,-3.4380699107\H,5.6504621334,-0.9443480806,-2
.2297983763\H,4.684064994,-2.253726477,-2.9596587823\H,-6.4756620691,-
1.6089931971,-0.2623897783\H,-4.8927833165,-0.9103639955,0.117894657\H
,-6.3444543504,0.1137388825,0.135742368\H,0.7876284321,-2.2491596534,3
.0030838246\H,-0.8594261356,-1.5558050867,3.0670135577\H,-0.6434744056
,-3.2763626157,2.6661333034\\Version=ES64L-G09RevD.01\HF=-1679.1934725
\RMSD=3.583e-09\RMSF=6.054e-08\Dipole=-3.1372129,-2.7497155,0.600447\Q
uadrupole=-57.4823726,18.8519395,38.6304331,-0.7604674,-2.6786104,21.4
588611\PG=C01 [X(C14H30N2O9P1)]\\@

dECP – Product

1\1\GINC-R2576\FOpt\RM062X\6-31+G(d,p)\C14H30N2O9P1(1-)\ROOT\01-Nov-20
14\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT IOP(2/17=4) Freq=
noraman scrf=(pcm,read)\\opt fwd\\-1,1\C,7.6946190318,-1.7012329325,-1
.8299974694\C,4.5409273932,0.0336292149,2.4904143574\C,-5.4178330871,0
.0122356426,-1.5649126895\C,0.1840006621,3.3476230749,-2.1393291985\C,
6.2429227365,-1.6697017734,-1.3543050715\O,6.0386222951,-1.8078700959,
-0.1129150723\O,5.3400978176,-1.4777103705,-2.2117803609\H,4.201779959
1,-0.9244349593,-0.1355163489\N,3.6026484339,-0.1249473736,0.163041125
7\C,3.6391031909,0.488355453,1.3901230612\C,2.6992662452,0.5254672956,
-0.6027949184\C,2.7222680225,1.5105383015,1.304336531\H,2.4961490633,0
.2394983064,-1.6296504479\N,2.1360930659,1.5241168006,0.0566408945\H,2
.4547049361,2.2349894195,2.0668173996\H,0.6124974962,2.5173379116,-0.4
242508873\O,-0.1778168232,2.9315176707,-0.8390006033\P,-2.2690179096,0
.4673211928,0.9386039483\O,-1.8847871011,-0.5868227022,-0.2035172959\O
,-0.9112806781,0.8953058819,1.6381921144\O,-3.0926761804,1.6221530201,
0.5282745403\C,-0.8996066605,-0.2548018015,-1.22807466\C,-0.1656831781
,0.0135659872,2.5196585298\H,-0.4172754826,-1.2093830655,-1.4632708119
\C,-1.5796760645,0.3577669357,-2.4315982742\H,-0.1616304719,0.42667803
13,-0.7917015798\C,0.4725860238,-1.149784317,1.7877820708\H,-0.8421549
935,-0.3285684096,3.3114732306\H,0.5965338535,0.664572823,2.9544150226
\H,-2.053338865,1.3065041574,-2.1570369426\H,-2.3289560259,-0.32617793
21,-2.8448736111\H,-0.8298314547,0.5560823131,-3.2051929831\H,1.121525
5084,-1.6908270759,2.4853280185\H,1.0892782746,-0.7888816187,0.9597729

226\H,-0.2777983714,-1.8469258783,1.4017022272\H,-3.6959067344,-0.1281
716606,2.4572747019\O,-2.9947373972,-0.5634729575,1.9316199634\H,-3.22
68129005,-1.8240848202,-0.7789966881\O,-3.6484646051,-2.54335534,-1.29
34031761\C,-4.9844417636,-2.4454118735,-1.3257236715\C,-5.597499212,-1
.2266858092,-0.6752590666\O,-5.6284925641,-3.2987157082,-1.899268145\H
,-5.1241503134,-1.059564018,0.3013164978\H,-6.6585706332,-1.4417552256
,-0.5188572514\H,8.133280569,-0.7108016523,-1.6610896109\H,8.271706516
6,-2.4262316394,-1.2503262756\H,7.7536391337,-1.9300641014,-2.89626183
98\H,4.3051960959,-0.992079554,2.7960267966\H,5.587996666,0.0490429544
,2.1697013599\H,4.4259052263,0.6907019979,3.3566922278\H,-5.9213559443
,0.870349514,-1.1117175133\H,-4.3618116819,0.2753444677,-1.672658181\H
,-5.8442892867,-0.1602491332,-2.5582546661\H,0.8487068694,4.2228234032
,-2.121419196\H,0.6875777819,2.5453762235,-2.6990074843\H,-0.731636079
,3.6173614264,-2.6747820181\\Version=ES64L-G09RevD.01\State=1-A\HF=-
1679.2103745\RMSD=4.930e-09\RMSF=2.554e-06\Dipole=-
4.3839646,1.8532653,3.3276467\Quadrupole=-50.596169,4.5668361,46.0293329,6.1294922,-
0.8217288,-14.5680788\PG=C01 [X(C14H30N2O9P1)]\\@

Dichlorvos – Reactant – R1

1\1\GINC-R3298\FOpt\RM062X\6-31+G(d,p)\C12H26N2O9P1(1-)\ROOT\20-Oct-20
14\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight) IOP
(2/17=4) Freq=noraman geom=check Guess=Read scrf=(pcm,read)\\opt fwd\\
-1,1\C,6.8153901501,4.4620625847,0.870534414\C,4.6595145309,0.22583479
09,-2.133978052\C,-5.8494432704,0.7298715029,0.0870702322\C,-0.2106475
887,-0.8301861888,3.0205323731\C,6.1855206661,3.1418256081,0.492923174
3\O,6.8283181558,2.1627745272,0.1476188045\O,4.8648976781,3.1546636191
,0.5745594191\H,4.4439086564,2.2514609236,0.3224876946\N,3.5209072596,
0.9224130368,-0.0022442303\C,3.6248757069,0.0603658239,-1.0748821171\C
,2.5061913284,0.4877944074,0.7264105243\C,2.6530878296,-0.8998401587,-
0.9721231839\H,2.1470647488,0.9417636095,1.6435401267\N,1.954867457,-0
.6128690771,0.175499627\H,2.399750459,-1.7412318319,-1.606468397\H,1.1
625940189,-1.1487652227,0.5332301742\O,-0.2662630507,-1.6266435637,1.8
254721337\P,-1.6812663741,-2.2180528147,1.3491980957\O,-2.5686404358,-
0.9654878178,0.9323750571\O,-1.1544955387,-2.9161351791,0.0275715891\O
,-2.4018403842,-3.0511841884,2.3344306292\H,-4.6497052971,-2.514357166
8,1.2066418351\O,-5.1916354443,-2.8020981966,0.4550726086\H,-4.8827477
544,-2.2747634559,-0.3177028079\O,-4.2918148577,-1.4002677595,-1.73404
82472\C,-4.749055284,-0.2348855336,-1.9490293545\C,-6.0895981293,0.114
4594328,-1.2966410821\O,-4.1615207055,0.6566041708,-2.60946931\H,-6.69
14318242,-0.796141706,-1.1968615729\H,-6.622705368,0.8278823349,-1.935
0301463\H,6.4273888123,5.2497051226,0.2170069257\H,6.5375756652,4.7157
881383,1.8984345258\H,7.8995354678,4.3950600997,0.7787071919\H,4.56767
69138,1.2039499217,-2.6192737838\H,5.6673180047,0.1604033013,-1.709936
287\H,4.5489201712,-0.5493794779,-2.8970282226\H,-6.7960182239,0.96629
09135,0.5838787499\H,-5.2964390492,0.0325989182,0.7272511087\H,-5.2687
266606,1.6553953131,0.0004594217\H,0.8477594073,-0.6964190947,3.250169
5312\H,-0.6881879854,0.1401302652,2.8461956588\H,-0.7056299728,-1.3526
627836,3.8447782904\C,-2.0824163432,-0.035761842,-0.0605266978\H,-2.82
52230835,0.7600431601,-0.1217410193\H,-1.1140274972,0.371010114,0.2525
245561\H,-2.013087963,-0.5304618253,-1.0325646643\C,-2.1027598746,-3.6
505664203,-0.7903098257\H,-1.4998054764,-4.2067231168,-1.5080937567\H,

-2.678712121,-4.3368925103,-0.1630213608\H,-2.7645362111,-2.9469491987
,-1.3005944678\\Version=ES64L-G09RevD.01\State=1-A\HF=-
1600.6151721\RMSD=8.216e-09\RMSF=2.256e-
07\Dipole=5.1935929,1.1970903,2.8391697\Quadrupole=-
58.8965181,29.1211289,29.7753892,-23.9297452,-16.8552813,16.6542516\PG=C01
[X(C12H26N2O9P1)]\\@

Dichlorvos – TS1 – R1

1\1\GINC-R3371\FTS\RM062X\6-31+G(d,p)\C12H26N2O9P1(1-)\ROOT\17-Oct-201
4\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultrafine
) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(pcm
,read)\\title\\-1,1\C,6.8136429106,3.0139780321,0.4058189779\C,4.65778
70984,-1.2222345935,-2.5986981373\C,-5.8511825236,-0.7182045281,-0.377
6429719\C,-0.2123914854,-2.2782599105,2.5558261314\C,6.1894040599,1.69
31352201,0.0211023814\O,6.836307465,0.733325505,-0.3727933184\O,4.8761
069536,1.6836052347,0.151906878\H,4.4495872391,0.7755055124,-0.1209459
099\N,3.5590812056,-0.5046716751,-0.4546354643\C,3.6223595791,-1.36069
90797,-1.5341104417\C,2.5347640373,-0.9104108661,0.2812523786\C,2.6145
656731,-2.2851958608,-1.4251436557\H,2.2010649236,-0.4512555219,1.2060
492496\N,1.9386942513,-1.9831473392,-0.2693345104\H,2.3241231635,-3.11
21422261,-2.0630993645\H,1.1162246072,-2.4755510329,0.1124051146\O,-0.
2781033967,-2.7899275007,1.2375829857\P,-1.8196202724,-3.0746243705,0.
5750158511\O,-2.2097360759,-1.4951264691,0.4874629354\O,-1.0261050636,
-3.6157871895,-0.7263613954\O,-2.4631567385,-3.8795915119,1.6677145084
\H,-4.1785837561,-3.3399522791,0.1613902875\O,-3.4614814929,-3.3309570
823,-0.498367316\H,-3.626785452,-2.5356824801,-1.3105516936\O,-3.65787
31934,-1.7266490431,-2.2786056713\C,-4.7909292389,-1.199202803,-2.6030
923794\C,-5.9718521219,-1.5080964275,-1.687689196\O,-4.9391193122,-0.4
469488429,-3.5746893502\H,-5.9889082221,-2.5860935681,-1.4766826759\H,
-6.8960076118,-1.2432815352,-2.2110716864\H,6.3984559504,3.807458441,-
0.2238385332\H,6.5597975297,3.2470312081,1.4448288311\H,7.8960293975,2
.9640883227,0.2841375451\H,4.6057666334,-0.2331410106,-3.0676769221\H,
5.6645548324,-1.3376778454,-2.1826575877\H,4.5096088647,-1.9796553288,
-3.3734900771\H,-6.6865804956,-0.9450519293,0.2922850017\H,-4.91581322
15,-0.9574157946,0.1416716053\H,-5.8580100278,0.3585543298,-0.57936364
69\H,0.8460505993,-2.1393642016,2.7989977264\H,-0.7320240346,-1.312861
6232,2.6348755635\H,-0.6608772789,-2.9813462351,3.2658416561\C,-1.3205
544013,-0.5726500017,-0.1455682302\H,-1.9019474676,0.3327476387,-0.331
552934\H,-0.4693744115,-0.347999178,0.506474362\H,-0.9588438032,-0.970
2515072,-1.1003558988\C,-1.5296375061,-4.3982797994,-1.812839759\H,-0.
6419759202,-4.7762429485,-2.3254871101\H,-2.1339999404,-5.2300856204,-
1.4458536122\H,-2.1197756847,-3.7808132821,-2.4934933856\\Version=ES64
L-G09RevD.01\HF=-1600.5966542\RMSD=4.959e-09\RMSF=3.341e-
06\Dipole=2.1316032,0.1023863,1.5750214\Quadrupole=-
35.9828741,23.1542381,12.828636,-9.7376655,-21.4551704,13.3964338\PG=C01
[X(C12H26N2O9P1)]\\@

Dichlorvos – Intermediate – R1

1\1\GINC-R2447\FOpt\RM062X\6-31+G(d,p)\C12H26N2O9P1(1-)\ROOT\14-Oct-20
14\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight) IOP
(2/17=4) Freq=noraman scrf=(pcm,read)\\opt fwd\\-1,1\C,7.3655294828,4.

3429217507,0.6491605191\C,4.9761672513,0.1761212463,-1.4001131455\C,-5
.7361809194,0.5403462295,0.6020397663\C,-0.5237944947,-1.4008736158,2.
9331022372\C,6.4348647806,3.1338210553,0.7157191114\O,6.9232220531,2.0
029099952,0.9423358096\O,5.1988316917,3.3719152299,0.5276323143\H,4.24
13280922,2.0049628888,0.5865536962\N,3.6378215802,1.1459392203,0.48846
28896\C,3.8493508269,0.1409297021,-0.4209262782\C,2.5590575446,0.81079
06102,1.2269949626\C,2.8564856015,-0.780786481,-0.1841384959\H,2.17825
23874,1.4349888143,2.0278440524\N,2.0527609015,-0.3525835094,0.8511001
045\H,2.6809791508,-1.7257768868,-0.6883560332\H,0.6706221736,-1.33120
08935,1.3835639795\O,-0.122441379,-1.8570171558,1.6593522643\P,-2.6084
907817,-2.2096462554,0.1334961119\O,-2.4341425908,-0.636202636,0.18752
84575\O,-1.484828408,-2.7820219565,-0.8334856922\O,-2.8033498086,-2.97
00430923,1.3806081505\H,-4.5371378283,-2.9842469777,-0.5777008048\O,-3
.9635814718,-2.2168132838,-0.775480431\H,-4.2929858178,-0.6448010176,-
1.6555283348\O,-4.3855439913,0.1757435939,-2.1795714472\C,-5.517916615
,0.8263371614,-1.8788794143\C,-6.3575868546,0.2352983348,-0.7694673764
\O,-5.7949112954,1.8489370804,-2.4688418008\H,-6.4387139561,-0.8498885
883,-0.9152409537\H,-7.3567225988,0.6738076389,-0.8481044356\H,7.28688
68138,4.8033002004,-0.3412634498\H,7.043948196,5.0889885535,1.38287480
95\H,8.4010055161,4.0558700197,0.8412043346\H,4.8909045124,1.040015985
2,-2.069003643\H,5.9380229529,0.2511270286,-0.8810733094\H,4.969023314
1,-0.7340546817,-2.0060123203\H,-6.3668129039,0.1291931748,1.395159531
2\H,-4.7382520684,0.1030875044,0.6962010246\H,-5.6546267084,1.62148740
83,0.7529959682\H,0.2986010573,-1.4495343626,3.6604895081\H,-0.8931374
719,-0.3628935215,2.8978400125\H,-1.3410778878,-2.0414159901,3.2736441
74\C,-1.2448572941,0.1174539633,-0.1155076565\H,-1.5406601531,0.886785
9987,-0.8331633484\H,-0.8813349435,0.5690656026,0.8086519285\H,-0.4632
417915,-0.5235034079,-0.5267584071\C,-1.4786856948,-2.4189898325,-2.22
68891138\H,-0.4527663046,-2.5525722536,-2.5728442303\H,-2.1563875633,-
3.0778240711,-2.7756724885\H,-1.7831752842,-1.3759605208,-2.3705761076
\\Version=ES64L-G09RevD.01\State=1-A\HF=-1600.600826\RMSD=5.048e-
09\RMSF=4.519e-06\Dipole=-6.9290526,-3.0442499,-1.7041145\Quadrupole=-
55.4905713,18.1502389,37.3403324,-15.6422398,-26.0362981,18.9446431\PG=C01
[X(C12H26N2O9P1)]\\@

Dichlorvos – TS2 – R1

1\1\GINC-R3122\FTS\RM062X\6-31+G(d,p)\C12H26N2O9P1(1-)\ROOT\17-Sep-201
4\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultrafine
) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(pcm
,read)\\title\\-1,1\C,6.8137050334,3.0140367389,0.4058158587\C,4.65764
39972,-1.2222816743,-2.5985715011\C,-5.8511774736,-0.7182048906,-0.377
6417415\C,-0.212315557,-2.278271174,2.5557023839\C,6.0415619639,1.8697
407638,-0.2311558501\O,6.5948001671,0.7592839661,-0.3777949196\O,4.832
9514679,2.1279402437,-0.5547717139\H,4.0814505996,0.7911107464,-0.6863
82528\N,3.5043699438,-0.1184515967,-0.6592657114\C,3.631772569,-1.2000
503115,-1.4978582539\C,2.5307895066,-0.3804526409,0.2221757713\C,2.684
4825415,-2.1057700392,-1.0731357185\H,2.2239355558,0.3127233964,0.9981
120482\N,2.0032682094,-1.5796024142,0.0080274644\H,2.4490186042,-3.087
5091704,-1.4640346543\H,0.9029412323,-2.0482650763,0.641266732\O,-0.07
33928842,-2.4017537912,1.1430465561\P,-1.6689480331,-2.6399962637,0.08
91733926\O,-1.9939261279,-1.0573883632,0.2785997395\O,-0.5354822571,-2

.8665328986,-1.0487338148\O,-2.1674831712,-3.6628136385,1.0715477791\H
,-3.6188879777,-3.4088614959,-0.6667746205\O,-2.9734484979,-2.76541335
82,-1.0295838228\H,-3.4158296184,-1.4408828715,-1.8339198987\O,-3.5497
648982,-0.6470876837,-2.4341624447\C,-4.8289980111,-0.3507130306,-2.63
17805755\C,-5.8382704723,-1.1614112541,-1.8477837122\O,-5.1411433208,0
.5578675623,-3.380173359\H,-5.5804314896,-2.2260988056,-1.9205701688\H
,-6.8176879166,-1.0059445835,-2.3095261175\H,6.6688459889,3.9312985089
,-0.172074255\H,6.411263772,3.1901559915,1.4096719769\H,7.8768260602,2
.7781090263,0.4818766397\H,4.4902905287,-0.3970286401,-3.2987135219\H,
5.6636519045,-1.1127745461,-2.1812020361\H,4.59522325,-2.1660005957,-3
.1465735532\H,-6.5971094276,-1.2898521488,0.1818099999\H,-4.8732747931
,-0.8786125525,0.0902542195\H,-6.1022102822,0.344479137,-0.2993479274\
H,0.7519917042,-1.9610150005,2.9683482784\H,-0.9777824556,-1.529942193
,2.7976914288\H,-0.5049665298,-3.2414015343,2.9792218024\C,-1.00642509
75,-0.0310938101,0.1983060541\H,-1.5454496978,0.8868960768,-0.04596153
89\H,-0.4859269478,0.0843883362,1.1543656052\H,-0.2760332895,-0.247851
0415,-0.5882712644\C,-0.750261909,-2.6017633237,-2.4383723705\H,0.2399
855533,-2.6409553801,-2.8975096533\H,-1.4072145406,-3.3528642483,-2.88
12290605\H,-1.1792595387,-1.6055536948,-2.5879646246\\Version=ES64L-G0
9RevD.01\HF=-1600.5938312\RMSD=6.267e-09\RMSF=4.857e-07\Dipole=-3.2114818,-
1.8025695,0.6917381\Quadrupole=-49.2746728,19.6160714,29.6586014,
-8.1863607,-9.5960773,18.8031336\PG=C01 [X(C12H26N2O9P1)]\\@

Dichlorvos – Product – R1

1\1\GINC-R3011\FOpt\RM062X\6-31+G(d,p)\C12H26N2O9P1(1-)\ROOT\17-Oct-20
14\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight) IOP
(2/17=4) Freq=noraman scrf=(pcm,read)\\opt fwd\\-1,1\C,6.5688975595,4.
7580350251,1.0514429257\C,4.7883475912,0.5489113374,-1.6513934157\C,-5
.1044776045,1.3076889521,0.3347756035\C,-0.1309014188,-1.6175727365,2.
8496714937\C,5.9135298472,3.3878999221,0.8846539538\O,6.6454274978,2.3
813173936,0.7400139084\O,4.6416135387,3.3674439064,0.9070989117\H,4.02
1719899,1.7960636838,0.7412292934\N,3.5694477207,0.8724512668,0.518505
6314\C,3.8083977484,0.1265662892,-0.6072723257\C,2.6491680079,0.217222
1341,1.2574797596\C,2.9951983879,-0.9770111439,-0.492538315\H,2.281843
9344,0.594114818,2.2059645332\N,2.2734201412,-0.9128383734,0.680313993
2\H,2.8968317813,-1.8108732496,-1.1798950111\H,0.8473078067,-1.8994661
719,1.2001052214\O,0.0261201135,-2.27015451,1.6064134899\P,-2.71030894
04,-2.2142311841,0.4309136309\O,-2.215504995,-0.7383788293,0.137558271
2\O,-1.8857874456,-3.1859595149,-0.5194130221\O,-2.8559544196,-2.68630
07211,1.8189017199\H,-4.8442557903,-2.6287973362,0.1179556518\O,-4.160
1913282,-2.0673577151,-0.2991510216\H,-4.2665749276,-0.7291358615,-1.5
425110654\O,-4.2518850658,-0.0940550501,-2.2860463791\C,-5.1768182884,
0.8644747922,-2.1342958452\C,-5.9665540235,0.8347061944,-0.8455270204\
O,-5.3129466168,1.7052996419,-2.9972456648\H,-6.32347023,-0.1885534358
,-0.6667658089\H,-6.8338321319,1.4876489167,-0.9798197739\H,6.19672755
61,5.4356580313,0.2763411728\H,6.2802781702,5.1793280782,2.0200754802\
H,7.6562813232,4.6854281167,0.990581374\H,4.5146838854,1.5204179731,-2
.0782448599\H,5.7927310916,0.6467280813,-1.2255461472\H,4.8131994641,-
0.190241319,-2.4567587363\H,-5.6996782785,1.3078144314,1.2521260475\H,
-4.2401939891,0.6552536232,0.4878136685\H,-4.7415628252,2.3256569081,0
.1605779458\H,0.7822792086,-1.682697998,3.4579412699\H,-0.3860678722,-

0.5528584645,2.7195741997\H,-0.9535797438,-2.1004449956,3.3810477469\C
,-0.9171797795,-0.3380199269,-0.3424417173\H,-1.0973525082,0.376310695
4,-1.1492765172\H,-0.3726939093,0.1319574475,0.4773180574\H,-0.3485342
607,-1.198009647,-0.7011457374\C,-2.003286956,-3.0604477459,-1.9489647
989\H,-1.1372510033,-3.5717562182,-2.3708075172\H,-2.9284335574,-3.537
0273741,-2.2834836202\H,-1.9978533649,-2.0087131372,-2.2578936348\\Ver
sion=ES64L-G09RevD.01\State=1-A\HF=-1600.6013266\RMSD=5.498e-09\RMSF=3.412e-
06\Dipole=-6.9805107,-2.9869529,-1.2726962\Quadrupole=-
44.3490954,15.3171722,29.0319232,-21.0751051,-26.0156996,17.1210463\PG=C01
[X(C12H26N2O9P1)]\\@

Dichlorvos – Reactant – R2

1\1\GINC-R3579\FOpt\RM062X\6-31+G(d,p)\C12H26N2O9P1(1-)\ROOT\19-Nov-20
14\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT IOP(2/17=4) Freq=
noraman scrf=(pcm,read)\\opt fwd\\-1,1\C,7.17425,2.603116,-0.44224\C,3
.88208,-1.517713,-1.374749\C,-5.621194,2.754621,0.864696\C,-1.99788081
69,0.4859522044,2.6119940973\C,6.177098,1.535311,-0.067943\O,6.496482,
0.433886,0.356088\O,4.92411,1.913392,-0.244012\H,4.2577371835,1.171610
8742,0.0388701316\N,3.2021131412,0.0642947508,0.4526280877\C,3.1299097
587,-1.1851798639,-0.1282018166\C,2.4180082338,0.0329033446,1.51642742
52\C,2.2824060406,-1.9679213933,0.6134447216\H,2.2389976011,0.85711863
03,2.1981949724\N,1.8459876262,-1.1792379587,1.6504511181\H,1.95128354
77,-2.9908965137,0.4853834717\H,1.1354665186,-1.442421932,2.3553201203
\O,-1.3323420503,-0.6853145525,2.0878560307\P,-1.8584901543,-1.2632906
265,0.6955703541\O,-1.5946546502,-0.1469044109,-0.3951548595\O,-0.7686
767655,-2.4115578562,0.5221825506\O,-3.2740692142,-1.7048646965,0.6373
915948\H,-3.972491,-1.095067,-1.203965\O,-4.152108,-0.818135,-2.115973
\H,-3.917191,0.131723,-2.133452\O,-3.581185,1.989171,-1.85675\C,-3.630
302,2.202511,-0.611044\C,-4.839871,1.646789,0.153149\O,-2.770366,2.839
643,0.054878\H,-4.469482,0.918909,0.889732\H,-5.48811,1.107078,-0.5454
08\H,7.024743,2.890143,-1.487849\H,7.000404,3.491708,0.172993\H,8.1895
,2.234891,-0.293066\H,3.572286,-0.8647,-2.198698\H,4.958914,-1.383454,
-1.22313\H,3.696724,-2.554923,-1.667583\H,-6.473081,2.343744,1.416806\
H,-4.974176,3.287176,1.568809\H,-6.004832,3.482895,0.141354\H,-1.39183
6,0.82749,3.451785\H,-2.062512,1.26925,1.846845\H,-2.999769,0.212832,2
.956658\C,-0.81558663,-3.175973243,-0.7004499902\H,-0.1292527902,-4.01
22762408,-0.5646706446\H,-1.8299258409,-3.5474789513,-0.8764298677\H,-
0.4906205445,-2.5524582712,-1.5400332783\C,-0.331423986,0.5387529701,-
0.4895003606\H,-0.4285302149,1.2132758958,-1.3406207408\H,-0.155611995
8,1.1205128298,0.4204003078\H,0.4796044591,-0.1821227294,-0.6467499064
\\Version=ES64L-G09RevD.01\State=1-A\HF=-1600.6166809\RMSD=4.161e-
09\RMSF=1.857e-06\Dipole=5.2318039,-3.3230767,3.4368647\Quadrupole=-
48.6523164,20.330147,28.3221695,13.6895857,-28.0137523,9.680734\PG=C01
[X(C12H26N2O9P1)]\\@

Dichlorvos – TS1 – R2

1\1\GINC-R2817\FTS\RM062X\6-31+G(d,p)\C12H26N2O9P1(1-)\ROOT\23-Dec-201
4\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultrafine
) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(pcm
,read)\\title\\-1,1\C,7.908674,2.295848,-0.130068\C,4.725628,-1.881078
,-1.175518\C,-4.859185,2.025172,1.395334\C,-1.152497,-0.201753,2.99775

2\C,6.788328,1.305114,0.061647\O,6.941051,0.214006,0.593496\O,5.631895
,1.736868,-0.401556\H,4.862904,1.042176,-0.245456\N,3.654323,0.093257,
-0.060167\C,3.600914,-1.229617,-0.444\C,2.517903,0.348409,0.569537\C,2
.403494,-1.758594,-0.033101\H,2.234022,1.301624,1.003124\N,1.735733,-0
.743785,0.606844\H,1.979228,-2.751408,-0.131492\H,0.803376,-0.811031,1
.038825\O,-0.808347,-0.779691,1.747507\P,-2.0821267175,-1.1709431864,0
.6948632201\O,-2.3606440881,0.3515985411,0.2105734367\O,-0.9935381549,
-1.8823796517,-0.2647657815\O,-2.9954619027,-1.886047858,1.6494789118\
C,-1.001524013,-1.7892127087,-1.6936250384\H,-1.5646258386,-2.62083814
76,-2.1204917268\H,-1.4475972503,-0.8469139353,-2.0269476085\H,0.04753
41848,-1.833928406,-1.999137092\C,-1.3014864209,1.2987929327,0.0608182
834\H,-1.7185710042,2.1157072222,-0.5335113287\H,-0.9694059407,1.67421
59626,1.0341462129\H,-0.4488339999,0.856243499,-0.4677498153\O,-3.4399
518315,-1.5683813459,-0.7326023254\H,-4.0778571845,-2.1511161121,-0.28
09474212\H,-4.0015734324,-0.6990132284,-1.1734729185\O,-4.6171158107,0
.2829027737,-1.7623821099\C,-5.076653,1.230043,-1.020544\C,-5.344044,0
.922257,0.456233\O,-5.340908,2.370218,-1.444312\H,-4.894365,-0.036953,
0.730662\H,-6.435224,0.814064,0.54891\H,7.666015,3.22268,0.399896\H,8.
844047,1.881898,0.24774\H,8.003601,2.537641,-1.19343\H,4.935608,-1.356
828,-2.114798\H,5.639907,-1.862896,-0.572658\H,4.47776,-2.920419,-1.40
8495\H,-5.151528,1.810533,2.428823\H,-3.766948,2.100219,1.361888\H,-5.
280292,2.992278,1.104995\H,-0.225192,0.1577,3.453376\H,-1.842749,0.643
155,2.865708\H,-1.62489,-0.945779,3.646079\\Version=ES64L-G09RevD.01\H
F=-1600.5987673\RMSD=7.648e-09\RMSF=4.524e-06\Dipole=4.220512,-
0.7254112,1.2610408\Quadrupole=-59.0500833, 29.2251379,29.8249454,20.1239151,-
23.8465943,17.9214396\PG=C01 [X(C12H26N2O9P1)]\\@

Dichlorvos – Intermediate – R2

1\1\GINC-R3567\FOpt\RM062X\6-31+G(d,p)\C12H26N2O9P1(1-)\ROOT\19-Nov-20
14\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT IOP(2/17=4) Freq=
noraman scrf=(pcm,read)\\opt fwd\\-1,1\C,7.90866,2.295857,-0.13005\C,4
.726321,-1.881307,-1.176006\C,-4.859207,2.025169,1.395363\C,-1.153156,
-0.201532,2.998194\C,6.808584,1.285526,0.081973\O,6.997714,0.184393,0.
581711\O,5.630127,1.710228,-0.325365\H,4.870158,1.00008,-0.151439\N,3.
692467,0.048916,0.051134\C,3.616743,-1.252621,-0.396764\C,2.562806,0.2
89149,0.702014\C,2.413239,-1.780503,0.001012\H,2.296391,1.226006,1.180
125\N,1.763701,-0.790095,0.695677\H,1.972208,-2.757639,-0.143399\H,0.8
13921,-0.850642,1.112826\O,-0.777515,-0.759885,1.757826\P,-2.064877,-1
.178138,0.642147\O,-2.238915,0.396235,0.192229\O,-0.858936,-1.939806,-
0.165933\O,-2.956474,-1.839274,1.678651\H,-3.854792,-2.165948,-0.33078
3\O,-3.224738,-1.501314,-0.680581\H,-3.908637,-0.318492,-1.313777\O,-4
.321068,0.470593,-1.828151\C,-4.904296,1.35893,-1.046098\C,-5.238031,0
.950059,0.378307\O,-5.220446,2.456881,-1.484296\H,-4.76382,-0.001794,0
.631158\H,-6.327401,0.79546,0.388503\H,7.669609,3.212032,0.419534\H,8.
861524,1.890652,0.211921\H,7.966653,2.552777,-1.192611\H,4.935509,-1.3
08277,-2.086495\H,5.646473,-1.911895,-0.58267\H,4.458592,-2.901874,-1.
463795\H,-5.212484,1.743004,2.39179\H,-3.770879,2.136176,1.439384\H,-5
.300135,2.989476,1.125759\H,-0.255625,0.230118,3.457325\H,-1.90289,0.5
94493,2.872356\H,-1.575307,-0.964983,3.660382\C,-1.1579266698,1.317955
8453,0.1313329686\H,-1.5209264806,2.1656167676,-0.4572923739\H,-0.8628
93768,1.6600616247,1.1288341789\H,-0.288517472,0.874314355,-0.36803710

17\C,-0.8519706884,-2.1327330512,-1.5791120623\H,0.1832244597,-2.36883
41925,-1.8408313471\H,-1.5109538214,-2.955190911,-1.8668803518\H,-1.16
44673247,-1.2251193367,-2.1051092287\\Version=ES64L-G09RevD.01\State=1
-A\HF=-1600.6050179\RMSD=6.678e-09\RMSF=8.868e-07\Dipole=2.3905023,-
0.1895167,0.2574859\Quadrupole=-47.267174,24.2534842,23.0136898,15.8334737,-
17.1487422,19.0110577\PG=C01 [X(C12H26N2O9P1)]\\@

Dichlorvos – TS2 – R2

1\1\GINC-R2532\FTS\RM062X\6-31+G(d,p)\C12H26N2O9P1(1-)\ROOT\20-Nov-201
4\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultrafine
) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(pcm
,read)\\title\\-1,1\C,6.81371,3.014047,0.405824\C,4.657743,-1.222274,-
2.598677\C,-5.851198,-0.718203,-0.377641\C,-0.212399,-2.278291,2.55579
9\C,6.034128,1.927387,-0.320868\O,6.490362,0.763734,-0.342401\O,4.9268
9,2.285497,-0.846218\H,4.026966,0.990258,-0.96356\N,3.43969,0.102737,-
0.850566\C,3.642473,-1.090188,-1.50274\C,2.479256,-0.077401,0.066853\C
,2.756153,-1.975943,-0.935365\H,2.123068,0.706517,0.725993\N,2.037914,
-1.328436,0.049312\H,2.581758,-3.01802,-1.170778\H,0.9551632487,-1.719
6243697,0.7003098377\O,-0.0868885625,-1.9519604337,1.1783008917\P,-1.5
855123583,-1.8875925534,-0.038864795\O,-1.4120404791,-0.2744719719,-0.
063629827\O,-0.4959049752,-2.5780927999,-1.0257375176\O,-2.4307435199,
-2.6507058256,0.9430014188\H,-3.4293553386,-2.4160396227,-1.156372185\
O,-2.7585532388,-1.7128514206,-1.2910905132\H,-3.2595854608,-0.1358454
565,-1.5434022488\O,-3.553186,0.790425,-1.770677\C,-4.850136,0.870277,
-2.041715\C,-5.662696,-0.394885,-1.86678\O,-5.343282,1.933069,-2.37541
4\H,-5.152263,-1.22212,-2.376896\H,-6.629837,-0.231565,-2.351062\H,6.2
93271,3.250022,1.340973\H,7.828212,2.682845,0.635379\H,6.838363,3.9267
8,-0.196263\H,4.408724,-0.564121,-3.43807\H,5.650462,-0.944348,-2.2297
98\H,4.684065,-2.253726,-2.959659\H,-6.475662,-1.608993,-0.26239\H,-4.
892783,-0.910364,0.117895\H,-6.344454,0.113739,0.135742\H,0.787628,-2.
24916,3.003084\H,-0.859426,-1.555805,3.067014\H,-0.643474,-3.276363,2.
666133\C,-0.3796013499,-2.1801816088,-2.396005969\H,0.5824945625,-2.56
80751922,-2.7360983865\H,-1.1931600333,-2.597474275,-2.9923262453\H,-0
.3810724658,-1.0882185762,-2.4832109827\C,-1.2423048433,0.5087299255,1
.1143940977\H,-1.4799988292,1.5352516685,0.8275457053\H,-1.9327447355,
0.1799893058,1.8992963022\H,-0.2144699263,0.4465628672,1.4805961581\\V
ersion=ES64L-G09RevD.01\HF=-1600.5893803\RMSD=2.065e-09\RMSF=4.593e-07
\Dipole=-2.8423916,-2.7994681,0.5982635\Quadrupole=-57.1493848,18.5155
356,38.6338492,-0.6624569,-2.9049811,21.0931907\PG=C01 [X(C12H26N2O9P1
)]\\@

Dichlorvos – Product – R2

1\1\GINC-R3070\FOpt\RM062X\6-31+G(d,p)\C12H26N2O9P1(1-)\ROOT\20-Nov-20
14\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT IOP(2/17=4) Freq=
noraman scrf=(pcm,read)\\opt fwd\\-1,1\C,7.914458,0.59842,1.342228\C,3
.595985,4.352325,1.838353\C,-4.663115,-0.90393,-1.761482\C,0.811983371
6,0.6946152669,-3.3968150267\C,6.407649,0.734012,1.562375\O,6.009896,1
.711918,2.259833\O,5.649161,-0.113458,1.016783\H,4.109819,1.630781,1.3
78215\N,3.38997,2.085488,0.781482\C,3.01864,3.40477,0.840647\C,2.70207
6,1.50035,-0.224442\C,2.094899,3.559663,-0.166994\H,2.836446,0.457195,
-0.489432\N,1.902038781,2.3637538384,-0.8242879757\H,1.559611,4.459539

,-0.451452\H,0.4785149568,1.9911249252,-2.0054730255\O,-0.0522070749,1
.5928499102,-2.7322522837\P,-2.2846915602,1.2894650708,0.3943843079\O,
-1.9726459424,-0.2894650189,0.3153407817\O,-0.9011731582,2.0300721436,
0.6284021424\O,-3.0262347719,1.8554526642,-0.7471889497\H,-3.712199,1.
98353,1.871968\O,-3.0521176409,1.2641831816,1.7972973459\H,-3.23409831
42,-1.581730622,0.7673915387\O,-3.7655998693,-2.3995903652,0.844209745
9\C,-4.9932747786,-2.2558570934,0.3246314179\C,-5.291658,-0.942468,-0.
36106\O,-5.781198,-3.175318,0.397942\H,-4.905145,-0.112331,0.244595\H,
-6.37954,-0.846398,-0.425592\H,8.200288,1.256758,0.514064\H,8.463838,0
.919642,2.230276\H,8.181041,-0.426699,1.077523\H,3.355536,4.046739,2.8
62934\H,4.687354,4.390146,1.753776\H,3.193345,5.354857,1.671186\H,-4.8
87677,0.051436,-2.241637\H,-3.574553,-0.998453,-1.712774\H,-5.056554,-
1.715051,-2.381973\H,1.7196264932,1.1944098135,-3.7604207592\H,1.12062
00162,-0.137561642,-2.7447817378\H,0.2738554918,0.2786989271,-4.253861
522\C,-0.0622191299,1.6370998005,1.7360852813\H,0.8136533104,2.2836382
172,1.692098031\H,-0.6001769109,1.7810605872,2.6774324235\H,0.23917206
89,0.5899946655,1.6304290535\C,-1.0511764812,-0.7240309281,-0.71855950
81\H,-0.0344049108,-0.4185968393,-0.4487927186\H,-1.128413114,-1.81204
66456,-0.7505340463\H,-1.3241293716,-0.2744004753,-1.6777418885\\Versi
on=ES64L-G09RevD.01\State=1-A\HF=-1600.6079495\RMSD=3.423e-09\RMSF=5.410e-
06\Dipole=-4.3512822,1.7076818,-0.1500109\Quadrupole=-
57.0399956,21.4157896,35.624206,-25.8832281,-24.1461463,14.6901357\PG=C01
[X(C12H26N2O9P1)]\\@

VX(R) – Intermediate – R1

1\1\GINC-R2415\FOpt\RM062X\6-31+G(d,p)\C13H28N2O9P1(1-)\ROOT\19-Sep-20
14\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight) IOP
(2/17=4) Freq=noraman scrf=(pcm,read) geom=check guess=read\\opt fwd\\
-1,1\C,-7.1784529577,1.61098625,-1.9521807837\C,-4.2599747194,0.070232
6065,2.6008338743\C,5.9342043846,0.3067901379,-0.8118739979\C,0.480661
618,-3.2107879544,-1.6608156384\C,-6.1929440391,0.9251915531,-1.038759
7335\O,-6.4422523447,-0.1162241932,-0.4455424905\O,-5.0386389555,1.553
6797686,-0.9544361741\H,-4.3402543028,1.002435213,-0.38355533\N,-3.247
9775932,0.211288252,0.3097074604\C,-3.2076809689,-0.2736084686,1.59873
59357\C,-2.1967877772,-0.2996236169,-0.3162557318\C,-2.1023767795,-1.0
786926734,1.7271511351\H,-1.934667426,-0.1252109701,-1.3548429811\N,-1
.4790735105,-1.0813689757,0.5043852874\H,-1.7189213103,-1.6448922332,2
.5657920092\H,-0.6260988469,-1.6143062097,0.2276031749\O,0.6358431454,
-2.364695355,-0.5447429881\P,2.2069857773,-2.1677424957,0.206458862\O,
2.5873053793,-0.9222127396,-0.8028202403\O,1.3088137611,-1.6883615819,
1.4949499958\O,2.6654070096,-3.5965853906,-0.0162667186\C,1.6759355472
,-0.7521261804,2.5117601398\C,1.6638690904,0.6706612537,1.9810861658\H
,2.6456534578,-1.0134874707,2.9437498293\H,0.9073656466,-0.875404315,3
.2842827196\H,1.9064726396,1.3730493717,2.7858973748\H,0.6712232884,0.
9148958345,1.5843000036\H,2.4064778612,0.79647476,1.1860470389\H,4.163
9758629,-2.5744992134,1.2414192232\O,3.7256722862,-1.7303125042,1.0346
50034\H,4.4017533415,-0.5275624532,1.5923402751\O,4.842814142,0.259552
0616,2.0902155241\C,6.0628313009,0.5525486756,1.6751619663\C,6.5860959
569,-0.2126165661,0.4776861804\O,6.7081387768,1.4355854371,2.220586710
6\H,6.3641647582,-1.2803840519,0.5970794661\H,7.6713404857,-0.07549697
83,0.4488183159\H,-6.7834962446,1.6014867813,-2.9736233348\H,-8.139967

922,1.0974261407,-1.92092239\H,-7.2947932121,2.656685122,-1.6513678283
\H,-4.3036788056,1.1521882469,2.7689221056\H,-5.2455782059,-0.25487081
43,2.2492412331\H,-4.051112958,-0.4191052663,3.5562829875\H,6.35047224
32,-0.2146717421,-1.6788460675\H,4.8543911008,0.1274421892,-0.79534178
95\H,6.1184209004,1.3796323638,-0.9327503981\H,-0.5331767025,-3.059320
1122,-2.0508172207\H,1.2045278746,-2.9725272721,-2.4540932731\H,0.6159
05733,-4.2608058237,-1.3814769121\C,1.6242213669,-0.1364889184,-1.4939
421133\H,2.1718617453,0.7257046113,-1.8860668217\H,1.168223227,-0.6879
926719,-2.3213659989\H,0.8370173703,0.2125648793,-0.8168677208\\Versio
n=ES64L-G09RevD.01\State=1-A\HF=-1639.9007413\RMSD=3.033e-09\RMSF=1.459e-
07\Dipole=-1.9132383,0.7120194,-1.3023089\Quadrupole=-
49.3920967,11.0005471,38.3915496,-11.7958813,-18.2003997,-11.050348\PG=C01
[X(C13H28N2O9P1)]\\@

VX(R) – TS2 – R1

1\1\GINC-R3330\FTS\RM062X\6-31+G(d,p)\C13H28N2O8P1(1-)\ROOT\18-Nov-201
4\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultrafine
) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(pcm
,read)\\title\\-1,1\C,6.8136780641,3.0140163858,0.4058000525\C,4.65753
99123,-1.2222900196,-2.5984343097\C,-5.8511746308,-0.7181934403,-0.377
6468951\C,-0.2121873457,-2.2782539258,2.5555861523\C,6.0382808359,1.88
92327308,-0.2636031109\O,6.5661270427,0.7624838417,-0.3731588619\O,4.8
55589158,2.1793508821,-0.6515399258\H,4.0481878013,0.8764471473,-0.791
0154093\N,3.4569053067,-0.0256941393,-0.7444852879\C,3.625912538,-1.15
71721187,-1.5082576119\C,2.4794890445,-0.2559434016,0.1404731958\C,2.7
015826279,-2.0618560812,-1.0382723726\H,2.1419228375,0.4708106029,0.87
15825553\N,1.9939113617,-1.4834019483,-0.0037498357\H,2.494016422,-3.0
714887988,-1.3686496047\H,0.965759188,-1.9290049609,0.6305935329\O,-0.
0574961134,-2.2906039643,1.1458980686\P,-1.6223484011,-2.3097051702,-0
.0020674661\O,-0.4999639085,-2.8470279316,-1.0590413984\O,-2.334300986
1,-3.2649174428,0.9271366428\C,-0.527529188,-2.6517134538,-2.479718499
3\C,-0.3444647736,-1.1950160727,-2.8672979887\H,-1.4535144806,-3.05865
96188,-2.896576825\H,0.3149482229,-3.2499473146,-2.8443979025\H,-0.231
0893421,-1.1165714567,-3.9540611813\H,0.5545709824,-0.7846363602,-2.39
32414797\H,-1.2083725735,-0.5890815103,-2.5773348211\H,-3.4423617065,-
3.0316788149,-1.1165419476\O,-2.8936953978,-2.2223399687,-1.2081218151
\H,-3.4843291572,-1.0841445047,-2.0814899695\O,-3.7961816385,-0.321043
2626,-2.6693128703\C,-5.1165173271,-0.1955526327,-2.7211304616\C,-5.92
74298596,-1.1354119955,-1.8532870194\O,-5.6257025983,0.6664447615,-3.4
15514975\H,-5.5518913994,-2.1593128791,-1.9759651496\H,-6.9606142014,-
1.0989869608,-2.2110189387\H,7.8682608522,2.753706191,0.5138149473\H,6
.7066570736,3.9388226091,-0.1681236452\H,6.3839414909,3.1899598978,1.3
984312928\H,4.4586152404,-0.4646633652,-3.3638369113\H,5.654878291,-1.
035922188,-2.1872917091\H,4.6398132491,-2.2088125251,-3.0687177683\H,-
6.4889709179,-1.3677034866,0.2286056863\H,-4.8282304512,-0.8008367983,
0.0035710317\H,-6.1923090762,0.314413871,-0.2512837735\H,0.7808521198,
-2.2041762591,3.0137429265\H,-0.8173745754,-1.4223032805,2.8842070384\
H,-0.7017807422,-3.2006435552,2.8780960742\C,-1.6112890065,-0.49570462
08,0.2031969556\H,-1.7442653606,-0.2720244934,1.2649949788\H,-0.637162
9865,-0.1114631678,-0.1106719601\H,-2.4043505852,-0.015825034,-0.37187

78357\\Version=ES64L-G09RevD.01\HF=-1564.6835927\RMSD=9.353e-09\RMSF=9.131e-07\Dipole=-2.6354172,-1.98401,0.4102894\Quadrupole=-54.0667279,22.8804235,31.1863044,-7.7395071,-13.0034241,20.7705606\PG=C01 [X(C13H28N2O8P1)]\\@

VX(R) – Reactant – R2

1\1\GINC-R3080\FOpt\RM062X\6-31+G(d,p)\C13H28N2O8P1(1-)\ROOT\20-Nov-2014\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT IOP(2/17=4) Freq=noraman scrf=(pcm,read)\\opt fwd\\-1,1\C,7.17425,2.603116,-0.44224\C,3.8821073727,-1.5177241111,-1.3747295029\C,-5.621194,2.754621,0.864696\C,-1.9978725313,0.4859596692,2.6119876129\C,6.177098,1.535311,-0.067943\O,6.496482,0.433886,0.356088\O,4.92411,1.913392,-0.244012\H,4.257737,1.171611,0.03887\N,3.2021141537,0.0642961767,0.4526288889\C,3.1296198915,-1.1850220852,-0.128366409\C,2.4180334417,0.0328754998,1.5163895339\C,2.2830901379,-1.968133016,0.6132441467\H,2.2389926292,0.8571242716,2.1981868433\N,1.8460187915,-1.1792669113,1.6506025071\H,1.9506211015,-2.9909042549,0.4857342525\H,1.1354403632,-1.4423540147,2.3553440075\O,-1.3323185206,-0.6853303426,2.0878536953\P,-1.8589234165,-1.2635947814,0.6958397084\O,-1.594566559,-0.1469084936,-0.3951248562\O,-3.2740721534,-1.704837406,0.6372563536\C,-0.2954925891,0.4917554816,-0.5037995573\H,0.481784,-0.270451,-0.371768\C,-0.208304,1.142375,-1.863543\H,-0.223653,1.237104,0.299344\H,-1.015677,1.871026,-1.981345\H,-0.291071,0.389081,-2.654197\H,0.758959,1.648301,-1.955565\H,-3.9724901131,-1.095069682,-1.2039662246\O,-4.152108,-0.818135,-2.115973\H,-3.917191,0.131723,-2.133452\O,-3.581185,1.989171,-1.85675\C,-3.630302,2.202511,-0.611044\C,-4.839871,1.646789,0.153149\O,-2.770366,2.839643,0.054878\H,-4.469482,0.918909,0.889732\H,-5.48811,1.107078,-0.545408\H,7.024743,2.890143,-1.487849\H,7.000404,3.491708,0.172993\H,8.1895,2.234891,-0.293066\H,3.572286,-0.8647,-2.198698\H,4.958914,-1.383454,-1.22313\H,3.696724,-2.554923,-1.667583\H,-6.473081,2.343744,1.416806\H,-4.974176,3.287176,1.568809\H,-6.004832,3.482895,0.141354\H,-1.391836,0.82749,3.451785\H,-2.062512,1.26925,1.846845\H,-2.999769,0.212832,2.956658\C,-0.7740565443,-2.6643165648,0.3680630921\H,-0.3428412662,-3.0621390462,1.2893994559\H,-1.4111537325,-3.4275860632,-0.083510059\H,0.0153680678,-2.4014685468,-0.3344478759\\Version=ES64L-G09RevD.01\State=1-A\HF=-1564.698949\RMSD=5.153e-09\RMSF=1.590e-06\Dipole=5.803568,-2.9017215,3.8481621\Quadrupole=-50.3163052,21.3418873,28.9744179,10.5151487,-27.9318171,3.8257182\PG=C01 [X(C13H28N2O8P1)]\\@

VX(R) – TS1 – R2

1\1\GINC-R2410\FTS\RM062X\6-31+G(d,p)\C13H28N2O8P1(1-)\ROOT\04-Dec-2014\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultrafine) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman\\title\\-1,1\C,7.94622,2.15138,-0.669333\C,4.645228,-2.064632,-0.681485\C,-4.828365,2.619934,0.748248\C,-1.193046,0.756931,2.888109\C,6.796807,1.270263,-0.249851\O,6.916005,0.340863,0.536731\O,5.654832,1.60582,-0.81695\H,4.86539,0.992893,-0.50165\N,3.629022,0.153126,-0.099613\C,3.537995,-1.221585,-0.145007\C,2.498775,0.587477,0.436074\C,2.324438,-1.599122,0.371919\H,2.241693,1.625807,0.61769\N,1.684859,-0.439463,0.734412\H,1.871567,-2.573285,0.517788\H,0.749708,-0.371913,1.16026\O,-0.862408,-0.121669,1.82306\P,-2.130598,-0.748997,0.909548\O,-1.1986126589,-2.0332763634,0

.5558709821\O,-3.232175153,-0.8138216591,1.9247342206\C,-1.3300868213,
-2.9691777462,-0.5146070678\C,-1.0233734478,-2.3472779503,-1.864653662
\H,-2.3285120597,-3.4066404407,-0.5085666089\H,-0.5946715181,-3.745251
5792,-0.2752205188\H,-0.9787317812,-3.1297429547,-2.6293710327\H,-0.05
60648119,-1.8340177078,-1.8378703348\H,-1.8056680138,-1.6391112471,-2.
1463547534\H,-4.2373267417,-1.8030561178,0.2250909024\O,-3.6087282437,
-1.5218100467,-0.4517397787\H,-4.1666577194,-0.7258384468,-1.110517976
3\O,-4.6526641939,0.1245957103,-1.8729554048\C,-5.062153,1.258076,-1.3
98446\C,-5.342649,1.332759,0.105785\O,-5.291715,2.26482,-2.09355\H,-4.
922026,0.459176,0.613361\H,-6.436785,1.281064,0.211469\H,7.729296,3.18
6861,-0.387402\H,8.867995,1.818055,-0.191585\H,8.051208,2.119854,-1.75
846\H,4.873136,-1.795078,-1.719066\H,5.55793,-1.923167,-0.092724\H,4.3
67699,-3.122034,-0.652752\H,-5.12981,2.675663,1.799425\H,-3.734322,2.6
5397,0.707494\H,-5.220158,3.496322,0.223307\H,-0.256914,1.192135,3.249
974\H,-1.857887,1.561666,2.544395\H,-1.688862,0.209796,3.695494\C,-1.8
982186615,0.5459674127,-0.3600261573\H,-0.8564915396,0.4673464291,-0.6
856304836\H,-2.5640051813,0.490577875,-1.2158111345\H,-2.0153182941,1.
5142009282,0.135077942\\Version=ES64L-G09RevD.01\State=1-A\HF=-
1564.6031517\RMSD=3.188e-09\RMSF=8.457e-06\Dipole=3.3618287,-
1.1122588,1.3444114\Quadrupole=-54.5946771,34.3642525,20.2304246, 13.7559211,-
23.8370442,13.5738711\PG=C01[X(C13H28N2O8P1)]\\@

VX(S) – Reactant – R1

1\1\GINC-R2612\FOpt\RM062X\6-31+G(d,p)\C13H28N2O8P1(1-)\ROOT\29-Oct-
2014\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight) IOP(2/17=4)
Freq=noraman geom=check Guess=Read scrf=(pcm,read)\\opt fwd\\
-1,1\C,-1.3409547824,5.9927434185,4.5354294713\C,-0.9779256241,0.45011
65208,5.4205429671\C,-2.5909200519,-1.8589350226,-1.5353114786\C,3.093
5994347,-3.2578304536,-3.8162155919\C,-1.7166920236,4.7520531234,3.766
9243269\O,-2.3598823743,4.7666025205,2.7272452677\O,-1.2768888654,3.64
53832723,4.3348425276\H,-1.5390642501,2.8035405073,3.7918803895\N,-1.8
019876017,1.3411379906,3.2295657908\C,-1.4889324585,0.2605618871,4.028
0660891\C,-2.2364503149,0.8559230642,2.0807945636\C,-1.7425561726,-0.8
924774931,3.3299160775\H,-2.5585599284,1.4387817743,1.2243910125\N,-2.
2189057124,-0.4922134451,2.1019776541\H,-1.6154565651,-1.9377882809,3.
5866877855\H,-2.5006826792,-1.1148889583,1.3287391523\H,-3.4579405751,
-1.2877798925,-1.8703786355\H,-2.1377241616,-2.379550429,-2.3860816637
\H,-2.8873510472,-2.5941430774,-0.7803638582\O,-1.6699132632,-0.908767
7448,-0.9680647623\P,-0.1920741995,-1.4728927229,-0.5991687602\O,0.522
7198793,-1.6902881146,-2.0140739273\O,-0.2310357882,-2.7573738854,0.15
55756512\C,0.5825926636,-0.0998314916,0.2550164262\H,0.3799719897,-0.1
975111803,1.3247046799\H,1.6600772139,-0.1681489903,0.0772682678\H,0.2
070958764,0.8562062123,-0.1126398697\C,0.3265372184,-0.8510783719,-3.1
799549542\C,0.5174095001,0.6305889587,-2.9267536274\H,-0.6746532659,-1
.0620931082,-3.5782323356\H,1.07596045,-1.207342901,-3.8918390161\H,0.
5104171899,1.1445091488,-3.89380755\H,-0.2877236687,1.0473500828,-2.31
31750239\H,1.4910874033,0.8106198508,-2.4588443228\O,3.0445211474,-2.5
07765284,0.0019789055\H,2.2600981539,-2.998177821,-0.2911422826\H,3.24
0325051,-1.8462569938,-0.7047741056\O,3.5630105215,-0.590132099,-1.852
4892405\C,3.7617183839,-0.9464490579,-3.0545576795\C,4.2408018948,-2.3
836273711,-3.2962495964\O,3.5713635615,-0.2222454606,-4.0638506227\H,4
.6368240817,-2.7982840138,-2.3627272471\H,5.0474933832,-2.356329474,-4

.038924225\H,-1.7262452346,5.9191265053,5.5571063326\H,-0.2500916571,6
.0620945391,4.5957672368\H,-1.745580042,6.8766573856,4.0417739575\H,-0
.0491723651,1.0308051001,5.4176363158\H,-1.7090060423,0.9918446658,6.0
304586686\H,-0.7804241774,-0.5172658621,5.8900136657\H,3.4296333736,-4
.2855963431,-3.9890552051\H,2.2700314967,-3.2854892146,-3.0929350782\H
,2.707383313,-2.8601066635,-4.7611906454\\Version=ES64L-G09RevD.01\Sta
te=1-A\HF=-1564.7053937\RMSD=5.042e-09\RMSF=8.430e-08\Dipole=-6.2623563,-
0.2512795,2.8575524\Quadrupole=-10.0546615,6.7630627,
3.2915988,22.7264214,36.8565841,-1.7557983\PG=C01[X(C13H28N2O8P1)]\\@

VX(S) – TS1 – R1

4\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultrafine
) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(pcm
,read)\\title\\-1,1\C,44.1759564368,53.3520979524,52.2232564272\C,44.5
402698882,47.8105324964,53.1098899586\C,42.9266142181,45.5007352653,46
.1506464524\C,48.6099199269,44.102127446,43.8726111419\C,43.5927242454
,52.1174371078,51.5882012146\O,42.877751815,52.1367837543,50.595931887
7\O,43.940773145,51.0087897611,52.2113743405\H,43.5816887383,50.167403
9891,51.7199767042\N,43.2418259712,48.7303383226,51.1822792156\C,43.74
05959614,47.6383004678,51.8594985732\C,42.5918606484,48.2669527099,50.
1312509727\C,43.3735315055,46.4991032194,51.1907906488\H,42.0837349205
,48.8682066398,49.3845281083\N,42.6446773194,46.9200310574,50.10082778
31\H,43.5703463269,45.4511020172,51.3859123823\H,42.2124422228,46.3077
663557,49.382397588\H,41.9992738336,45.840796975,45.677561375\H,42.984
8005254,44.407705318,46.1097298376\H,42.9045680668,45.7995159229,47.21
20130389\O,44.0028152157,46.0982441605,45.4696649746\P,45.5680163007,4
5.8625878091,46.1323004713\O,46.276069345,46.4043347581,44.7675803861\
O,45.4728824703,44.3889147844,46.4883689507\C,45.3895035808,47.1638027
154,47.4046564454\H,45.6799076692,46.7636272602,48.3786465828\H,46.080
0054892,47.9735251054,47.1531821605\H,44.3689845271,47.5508311951,47.4
379015594\C,45.6482958541,47.0551345544,43.6574576422\C,45.3142268487,
48.5001495445,43.9832954239\H,44.7611093702,46.499477268,43.3425823865
\H,46.4028214839,47.0074521008,42.8629969722\H,44.9308350713,49.007189
2264,43.0906447994\H,44.5509870896,48.5504131871,44.7661438863\H,46.21
22385394,49.0261538824,44.3248112017\O,47.4470001925,45.8361968713,46.
7275142196\H,47.592648392,44.9117645111,46.9930213981\H,48.3495107322,
46.2456101824,46.0199725605\O,49.2459558808,46.7353689668,45.410203931
5\C,50.0589050249,45.9499412924,44.7781630722\C,49.7619047484,44.45610
37714,44.8231988559\O,51.025587111,46.3705114418,44.1323918618\H,49.49
96943118,44.1677360347,45.8493485589\H,50.6742517386,43.9230517244,44.
5370931015\H,43.9192557893,53.3754287904,53.2867137187\H,45.2675076125
,53.3102764409,52.1451572989\H,43.801657578,54.2455555267,51.722470958
5\H,45.4399029353,48.4044017539,52.9146761901\H,43.9555867255,48.33195
98373,53.8757010139\H,44.8458889933,46.8386880496,53.507017736\H,48.42
22968671,43.0237788949,43.8844092857\H,47.6909733242,44.6156023693,44.
1668558688\H,48.8591290177,44.3951852841,42.8464871029\\Version=ES64L-
G09RevD.01\HF=-1564.6774238\RMSD=4.914e-09\RMSF=2.269e-06\Dipole=-
5.2874769,0.0149461,4.0834406\Quadrupole=-426.5060871,-96.4810277,522.9871147,-
8151.3106024,-7835.8220136,-7718.4757908\PG=C01 [X(C13H28N2O8P1)]\\@

VX(S) – Intermediate – R2

1\1\GINC-R2464\FOpt\RM062X\6-31+G(d,p)\C13H28N2O8P1(1-)\ROOT\20-Nov-20
14\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT IOP(2/17=4) Freq=
noraman scrf=(pcm,read)\\opt fwd\\-1,1\C,7.90866,2.295857,-0.13005\C,4
.726321,-1.881307,-1.176006\C,-4.8592075288,2.024734592,1.3965078853\C
,-1.1531609455,-0.2016274278,2.9982354552\C,6.808584,1.285526,0.081973
\O,6.997714,0.184393,0.581711\O,5.630127,1.710228,-0.325365\H,4.870158
,1.00008,-0.151439\N,3.692467,0.048916,0.051134\C,3.616743,-1.252621,-
0.396764\C,2.562806,0.289149,0.702014\C,2.413239,-1.780503,0.001012\H,
2.296391,1.226006,1.180125\N,1.763701,-0.790095,0.695677\H,1.972208,-2
.757639,-0.143399\H,0.8139156726,-0.850559531,1.1128012544\O,-0.777519
4746,-0.7599697775,1.7578629221\P,-2.065534731,-1.1779697638,0.6423904
625\O,-0.8588047876,-1.9397260753,-0.1658123517\O,-2.9563974464,-1.839
4177842,1.6786250617\C,-0.8338533149,-2.2773518805,-1.5562458557\C,-0.
663819,-1.043502,-2.424376\H,-1.734037,-2.833344,-1.828127\H,0.035337,
-2.938039,-1.658645\H,-0.578752,-1.334602,-3.477342\H,0.24644,-0.50253
2,-2.139023\H,-1.525642,-0.37566,-2.317093\H,-3.8547614884,-2.16604578
13,-0.3309138764\O,-3.2246911682,-1.5012474502,-0.6805812077\H,-3.9086
608016,-0.318460043,-1.3133981592\O,-4.3208238854,0.4707793908,-1.8280
606049\C,-4.9043923277,1.3588741739,-1.0461064867\C,-5.2380444802,0.95
02784952,0.3780801639\O,-5.220446,2.456881,-1.484296\H,-4.76382,-0.001
794,0.631158\H,-6.327401,0.79546,0.388503\H,7.669609,3.212032,0.419534
\H,8.861524,1.890652,0.211921\H,7.966653,2.552777,-1.192611\H,4.935509
,-1.308277,-2.086495\H,5.646473,-1.911895,-0.58267\H,4.458592,-2.90187
4,-1.463795\H,-5.2124489115,1.7426888953,2.3917131184\H,-3.7706245037,
2.1365346598,1.4376673396\H,-5.3001448771,2.9895143077,1.1259118792\H,
-0.255625,0.230118,3.457325\H,-1.90289,0.594493,2.872356\H,-1.575307,-
0.964983,3.660382\C,-1.9396555612,0.6039007797,0.1406772759\H,-2.07535
60887,1.2471501218,1.0091201641\H,-0.8977931875,0.7148944748,-0.170771
4962\H,-2.5708869698,0.9482684824,-0.6678639395\\Version=ES64L-
G09RevD.01\State=1-A\HF=-1564.6787598\RMSD=2.695e-09\RMSF=4.718e-
06\Dipole=1.6591478,-0.2831938,-0.0881616\Quadrupole=-
42.5824715,21.4588636,21.1236079,14.2241526,-15.9564971,19.4271078\PG=C01
[X(C13H28N2O8P1)]\\@

VX(S) – TS2 – R2

1\1\GINC-R2687\FTS\RM062X\6-31+G(d,p)\C13H28N2O8P1(1-)\ROOT\20-Nov-201
4\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultrafine
) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(pcm
,read)\\title\\-1,1\C,7.155304,2.067851,1.506884\C,4.391116,-0.456263,
-2.689513\C,-5.906761,0.381135,0.289988\C,-0.46134,-2.800477,2.039171\
C,6.222731,1.429902,0.487086\O,6.520085,0.315053,0.006051\O,5.161444,2
.083294,0.209669\H,4.095387,1.038915,-0.313851\N,3.398266,0.247445,-0.
495374\C,3.4267,-0.628924,-1.554001\C,2.442866,-0.151859,0.355688\C,2.
443488,-1.554727,-1.292807\H,2.208349,0.363216,1.280654\N,1.839289,-1.
245594,-0.091099\H,2.12734,-2.405209,-1.883137\H,0.8156376794,-1.69739
40989,0.4215197773\O,-0.3141494161,-2.014217578,0.8680494354\P,-1.8511
265611,-1.3241100676,-0.1219772706\O,-1.2942656684,0.1617788434,0.2639
410701\O,-2.8115030035,-2.1448385029,0.7072102097\C,-1.0042366191,0.60
56222969,1.5926022501\H,-0.5275836624,1.5826388314,1.4556662698\C,-2.2
650219696,0.737131223,2.4322273769\H,-0.2834710829,-0.0738573209,2.060
4611748\H,-2.7454690774,-0.237609438,2.56725351\H,-2.9730819568,1.4143

554272,1.9442098143\H,-2.011382708,1.144202185,3.4173974474\H,-3.82035
04038,-1.0826262723,-1.1876519953\O,-3.0338189855,-0.5001763793,-1.127
950143\H,-3.2860410792,1.0469859841,-0.7674306106\O,-3.460269,2.027389
,-0.605496\C,-4.740836,2.351253,-0.736555\C,-5.709898,1.214993,-0.9841
89\O,-5.095961,3.510664,-0.61939\H,-5.324997,0.58652,-1.797793\H,-6.65
7279,1.657711,-1.305181\H,6.691463,1.994236,2.497074\H,8.121751,1.5607
02,1.526547\H,7.287478,3.130362,1.28373\H,4.213124,0.492368,-3.207371\
H,5.419798,-0.452723,-2.314538\H,4.272508,-1.272947,-3.406349\H,-6.641
116,-0.409922,0.111952\H,-4.971603,-0.091333,0.611383\H,-6.273945,1.01
0743,1.107154\H,0.543327,-3.056074,2.394532\H,-0.995459,-2.253135,2.82
4697\H,-1.018392,-3.710607,1.803468\C,-1.0115626318,-1.9800485388,-1.6
017924915\H,-0.3637280795,-2.8204551786,-1.3583519937\H,-1.7731440118,
-2.2768150585,-2.3248266115\H,-0.4205014007,-1.1679873879,-2.038287747
1\\Version=ES64L-G09RevD.01\HF=-1564.6782643\RMSD=5.471e-09\RMSF=2.883e-
07\Dipole=-2.4546028,-3.0190966,-0.1410217\Quadrupole=-
56.5469091,8.3017515,48.2451576,8.06022,2.6720135,9.6370928\PG=C01
[X(C13H28N2O8P1)]\\@

A.1.3 Phosphorylation reaction, large system

dECP – Reactant

1\1\GINC-R3113\FOpt\RM062X\6-31+G(d,p)\C30H53N5O9P1(1-)\ROOT\18-Sep-20
14\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight) IOP
(2/17=4) Freq=noraman geom=check Guess=Read scrf=(pcm,read)\\title\\-1
,1\C,8.1654691539,-3.6346445191,-0.5851170542\C,5.9043676573,-1.696177
0884,2.5185214317\C,-2.3958751495,5.4017395026,-1.2038303314\C,-3.6252
665292,4.438292696,3.4618294322\C,-5.5937282123,1.6151081436,-0.506073
7167\C,1.0538922563,4.1957174033,-0.0812251169\C,9.5021752205,-2.95593
35807,-0.2833232729\C,9.3771773822,-1.7923288008,0.6966629776\C,8.6545
974762,-0.5369866827,0.1779101795\O,8.3061982614,0.298191165,1.0674826
897\O,8.4719560542,-0.3991644308,-1.0561511465\H,7.4939483503,-2.94754
1222,-1.1078315718\H,8.3043620761,-4.5159710567,-1.2202749998\H,7.6738
459964,-3.9620732233,0.3391436692\H,10.1974045181,-3.6960243373,0.1363
378036\H,9.9444897802,-2.5911611014,-1.218749633\H,10.3741543486,-1.45
30047242,1.0142814246\H,8.8619526562,-2.1113936701,1.6147750483\H,6.44
7061025,0.84864201,0.5148237173\N,5.4415828715,0.6342810493,0.33600152
49\C,4.8480230098,-0.5963376364,0.4861325312\C,4.527560667,1.476057807
9,-0.1929584453\C,5.5685369558,-1.788278464,1.0260867878\C,3.561908979
3,-0.4295543733,0.0313529073\N,3.3713185841,0.8687553317,-0.3908886261
\H,4.7468603891,2.5126761271,-0.4226060336\H,6.4919568209,-1.917254908
,0.4556970563\H,4.9481646649,-2.6721363202,0.8299092927\H,2.7722416215
,-1.1723564475,-0.0346035952\H,6.4201147011,-2.6047679617,2.8484311747
\H,6.5644082251,-0.8436034728,2.7096212256\H,4.994124294,-1.5788896844
,3.1170625944\H,2.0271426019,4.2446301872,-0.6060728245\H,1.1109508589
,4.8512765764,0.809037189\C,0.8119485986,2.7657365961,0.3847248198\C,-
0.0424389505,4.7296039153,-1.0186837953\H,-0.1680934821,2.6864243504,0
.8773964346\H,1.5767912758,2.4877146163,1.1282848939\H,0.1865749296,5.
7766472135,-1.2711382783\H,-0.0216295821,4.1593038571,-1.9695981059\N,
-1.377540479,4.7068200094,-0.4057349233\H,1.7501765716,1.4982525955,-0
.763451571\H,-1.6742036033,3.7211095354,-0.3104312626\O,0.8246073211,1
.8433722492,-0.6899472654\H,-3.3778277855,5.2490205575,-0.7437586778\H

,-2.178307014,6.4768021248,-1.2161450673\H,-2.4413691392,5.0501379902,
-2.2502394957\H,0.7317687238,0.2614219171,-4.5223474549\C,0.5500434903
,1.1847372031,-3.9637921299\C,-0.7062873891,1.0725064666,-3.1323064534
\H,0.4417859477,2.0091357734,-4.6764016373\H,1.4084583916,1.3868165041
,-3.3164259787\O,-0.5304182931,-0.0495997189,-2.2249467624\H,-0.870249
2705,1.9749001976,-2.5354367568\H,-1.5899381594,0.8623827432,-3.747437
7714\P,-1.317618718,-0.0951571062,-0.8630126486\O,-1.9467771581,1.1388
799522,-0.3414768773\O,-2.4423654515,-1.2139383434,-1.2665755846\O,-0.
4059983538,-0.9162111457,0.1371626512\C,-3.5278267298,-1.5015774822,-0
.4811295672\C,0.1419675999,-0.3298667875,1.3494491699\C,-3.5568382891,
-1.2540305496,0.9013641914\C,-4.6087394929,-2.082586102,-1.1339321082\
C,0.3570693339,-1.4539302123,2.3376290063\H,-0.5619450929,0.423659418,
1.7256896509\H,1.0785927874,0.1677454991,1.0875455488\C,-4.6928978737,
-1.5823434261,1.6200818442\H,-2.7030519582,-0.806483729,1.4046479081\C
,-5.7327662273,-2.4076059917,-0.3848153285\H,-4.5771463667,-2.27234224
85,-2.2061494224\H,-0.5876867948,-1.950793629,2.5811863807\H,0.7896115
226,-1.0469236458,3.2574030462\H,1.0491316259,-2.1970525526,1.92897733
08\C,-5.8053251524,-2.1653257614,0.9932775412\H,-4.7313618829,-1.38790
69984,2.6909606353\O,-6.7808370207,-2.9713762121,-1.0548544919\C,-7.02
40360515,-2.5232541187,1.6744617087\C,-7.9535802948,-3.3223361062,-0.4
371739269\C,-8.0522615632,-3.0740543344,0.9963989331\H,-7.0985598211,-
2.3389865651,2.7460307545\O,-8.8204624225,-3.8137964412,-1.1287488524\
H,-8.9911810548,-3.3592764019,1.4655809026\H,-5.6177735934,1.135691413
5,2.0186655133\C,-5.0541522051,2.0807438388,1.8268439177\H,-5.74746272
59,2.9166128427,2.0463233629\N,-4.6197939589,2.1622068188,0.4343989258
\C,-3.8534062405,2.0928099588,2.7809911819\H,-3.7505135701,1.638820873
4,0.3291410002\H,-4.2155600975,1.9133836816,3.8116593942\H,-3.20693222
63,1.233395068,2.5104683366\N,-3.0107468081,3.2855387993,2.8089648636\
H,-5.1623333939,1.6033921936,-1.5134273288\H,-5.9051884373,0.585439990
6,-0.2460842341\H,-6.493263838,2.2440078269,-0.5253917841\H,-2.7309799
895,3.5408256464,1.8465944653\H,-3.8083810262,4.1987721888,4.517172235
4\H,-2.9324189185,5.285242225,3.4202219386\H,-4.5816778606,4.756345907
,3.0130831457\\Version=ES64L-G09RevD.01\State=1-A\HF=-
2466.7011944\RMSD=8.510e-10\RMSF=1.543e-07\Dipole=-
9.214633,0.0368755,2.277422\Quadrupole=-147.1675428,62.9267758,84.2407671,-
16.1123281,-17.2318535,-21.6664551\PG=C01 [X(C30H53N5O9P1)]\\@

dECP – TS1

1\1\GINC-R3536\FTS\RM062X\6-31+G(d,p)\C30H53N5O9P1(1-)\ROOT\11-Sep-201
4\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultrafine
) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(pcm
,read) geom=check guess=read\\Title Card Required\\-1,1\C,-10.44108,-1
6.977172,-23.351514\C,-12.020406,-13.931675,-25.946385\C,-21.834019,-9
.055902,-22.333816\C,-20.032956,-5.867653,-25.620105\C,-19.016824,-5.3
51127,-20.492651\C,-19.480952,-11.510068,-24.082023\C,-10.9110760509,-
18.2131527795,-24.1204424122\C,-11.853835005,-17.8769613162,-25.272324
355\C,-13.2273792187,-17.3282312852,-24.8684122525\O,-13.9042859992,-1
6.7953705082,-25.807889991\O,-13.6089155328,-17.427495556,-23.68115574
27\H,-11.2786370339,-16.5113653153,-22.8253914801\H,-9.680495891,-17.2
378478557,-22.6082021746\H,-10.0040058486,-16.2354410564,-24.031729004
7\H,-10.0383128888,-18.743350638,-24.5253416287\H,-11.4154160789,-18.9

009630938,-23.4303578586\H,-12.0493016778,-18.7695281044,-25.8845275269\H,-11.3935790105,-17.1418301445,-25.9490779452\H,-14.7289134883,-15.4597642875,-25.0861656797\N,-14.8525869471,-14.5322120538,-24.6018737005\C,-13.814326821,-13.7312575081,-24.1815738592\C,-16.0052134842,-13.9727805082,-24.2161054422\C,-12.3834144676,-14.0523954528,-24.4619957765\C,-14.3986954083,-12.6741936431,-23.5348505658\N,-15.7651319692,-12.8415368884,-23.5670660324\H,-16.9844867373,-14.3968916034,-24.4073702065\H,-12.1822695791,-15.0691448078,-24.1123106727\H,-11.7624920435,-13.3778031829,-23.8597299558\H,-13.9436425607,-11.8269998017,-23.0370189293\H,-10.9640075092,-14.1768902539,-26.1000070949\H,-12.6209923509,-14.6245633162,-26.5447754382\H,-12.1940555566,-12.9122376496,-26.3076729203\H,-19.3117572281,-12.589642575,-23.9063570337\H,-19.9313524711,-11.3936500007,-25.0865258777\C,-18.1440015607,-10.7846516709,-24.0793928012\C,-20.4726665438,-10.9741456036,-23.0335202671\H,-18.2885239646,-9.6986012749,-24.1341408926\H,-17.5576478273,-11.0891819585,-24.9617496769\H,-21.4265326879,-11.5135451034,-23.1431819879\H,-20.0927750523,-11.1973056611,-22.0175162729\N,-20.7474499966,-9.542027075,-23.1967002454\H,-16.5974571375,-11.9985421705,-23.1768289335\H,-19.8931540228,-9.0298620279,-22.9294642939\O,-17.3875682573,-11.1068379498,-22.9157753895\H,-21.907503606,-7.9675177294,-22.4297446176\H,-22.7843298889,-9.5007657432,-22.6539267022\H,-21.6795521539,-9.3001852136,-21.267925168\H,-16.8996652545,-13.7887619542,-20.3521734038\C,-17.8293270807,-13.2841131116,-20.6346552454\C,-17.8016136935,-11.8354870273,-20.1931686375\H,-18.6675005168,-13.7996652632,-20.153148711\H,-17.9481687819,-13.3524974673,-21.7192510435\O,-16.6155163268,-11.1876014206,-20.6937641752\H,-18.6900200026,-11.2966391501,-20.5417164647\H,-17.753879957,-11.7552179153,-19.1006989174\P,-16.7419269929,-9.865818297,-21.6063054046\O,-17.9413955674,-8.9823738694,-21.8051652215\O,-16.0157145331,-9.0011906261,-20.256126338\O,-15.3695985769,-9.6977138648,-22.4518490865\C,-15.7391588164,-7.6951805766,-20.2359157013\C,-15.3659584165,-9.4219452147,-23.8630175466\C,-15.6669882322,-6.8750385178,-21.3875867279\C,-15.4802810085,-7.1324767098,-18.9779190029\C,-14.0153213775,-8.8275602622,-24.2038915868\H,-16.1819603538,-8.7244032681,-24.1031431883\H,-15.5336019283,-10.3503693685,-24.4186898105\C,-15.3235003625,-5.5425685338,-21.2707867084\H,-15.8837972327,-7.2902412949,-22.3643568862\C,-15.1447289544,-5.7909543465,-18.8897795978\H,-15.5379003997,-7.7491154801,-18.0822787584\H,-13.8435553509,-7.9028208462,-23.6437705123\H,-13.9697948979,-8.6080430113,-25.2754734225\H,-13.2153828197,-9.5338939217,-23.9586341106\C,-15.0504574308,-4.9638333479,-20.0193010905\H,-15.261144571,-4.9196340457,-22.162173975\O,-14.90136999,-5.2897587105,-17.6383085607\C,-14.698088495,-3.5872411647,-19.8159554913\C,-14.5655059105,-3.9818104835,-17.4129366829\C,-14.4653281033,-3.109017405,-18.5723623804\H,-14.6229329013,-2.932084185,-20.6840055168\O,-14.3752693675,-3.6487719362,-16.2586436467\H,-14.1973926188,-2.0746985051,-18.3713934004\H,-17.6875878276,-4.5053401257,-22.5359225356\C,-18.5563091698,-5.1218372814,-22.8683414786\H,-19.3679428222,-4.4134002732,-23.1280214575\N,-19.0193608402,-6.0048745056,-21.8006729056\C,-18.1095320435,-5.9182603199,-24.0996624271\H,-18.389371301,-6.8165762232,-21.7466124038\H,-17.6342594237,-5.2255123919,-24.8205997121\H,-17.3265031516,-6.6290449763,-23.7699685616\N,-19.1217744096,-6.6843786528,-24.8229630867\H,-19.2923732173,-6.0773442862,-19.7205374287\H,-18.0314684684,-4.9205969336,-

20.234263342\H,-19.7552941835,-4.5389425059,-20.4811960918\H,-19.66570
86323,-7.2559722916,-24.1562302801\H,-19.4633311014,-5.3562771478,-26.
4066061005\H,-20.769389051,-6.5188587787,-26.1027063207\H,-20.57857745
07,-5.1025223492,-25.0422549158\\Version=ES64L-G09RevD.01\HF=-
2466.6866275\RMSD=5.205e-09\RMSF=5.124e-07\Dipole=-1.6138661,4.9805744,-
1.9953353\Quadrupole=225.3381919,392.5097719,-617.8479638,-659.3545179,-1162.773436,-
1046.0963168\PG=C01 [X(C30H53N5O9P1)]\\@

dECP – Intermediate

1\1\GINC-R3416\FOpt\RM062X\6-31+G(d,p)\C30H53N5O9P1(1-)\ROOT\27-Nov-20
14\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight) IOP
(2/17=4) Freq=noraman geom=check Guess=Read scrf=(pcm,read)\\opt fwd\\
-1,1\C,5.5376010524,-6.8887080118,-0.909425039\C,3.9582030252,-3.84321
43207,-3.5042984728\C,-5.8554581088,1.0325364423,0.1082426994\C,-4.054
2272635,4.2203859173,-3.1776643609\C,-3.0381827081,4.7372154238,1.9491
475562\C,-3.5023199163,-1.4210018451,-1.6400000361\C,5.2025217935,-8.3
775740251,-0.7958553223\C,4.1127581393,-8.8210377915,-1.7910777698\C,2
.9246228251,-7.8984771072,-1.6487784908\O,2.8527685696,-6.9753118247,-
2.5857129552\O,2.1492979796,-7.9628198712,-0.7020533479\H,4.6851351489
,-6.2747366004,-0.599011826\H,6.3855310725,-6.6281079673,-0.2692878152
\H,5.7862531048,-6.6142750795,-1.9410849483\H,6.0995474919,-8.98337440
47,-0.974817677\H,4.8601917907,-8.5978152822,0.223546014\H,3.79812636,
-9.8473498703,-1.5714631231\H,4.4920566175,-8.7668300756,-2.8174860818
\H,2.2673144283,-6.157191235,-2.2661486946\N,1.6991274739,-4.858126333
7,-1.7239074\C,2.4340360753,-3.710914233,-1.4977505444\C,0.5206020562,
-4.6732068322,-1.1519977262\C,3.8344412645,-3.5496505725,-2.0058438822
\C,1.6635795016,-2.832816409,-0.7781434179\N,0.4598796685,-3.461690485
2,-0.5734085949\H,-0.2919035403,-5.391032339,-1.1343734132\H,4.5072948
517,-4.2138769616,-1.4484487963\H,4.160363991,-2.5241842199,-1.7906121
177\H,1.855669966,-1.8367019506,-0.4005798969\H,4.9912681362,-3.705657
09,-3.8403228533\H,3.6628291435,-4.8730289962,-3.7250151328\H,3.314680
602,-3.1710113184,-4.0832151807\H,-3.2999969795,-2.4886651527,-1.43050
91643\H,-4.0589940557,-1.3594418234,-2.5954246406\C,-2.1739121978,-0.6
975850612,-1.8157469627\C,-4.3931259026,-0.8439379541,-0.5246722631\H,
-2.3294965639,0.3896655246,-1.8657840478\H,-1.7159776747,-1.013491991,
-2.7655542884\H,-5.3138533637,-1.4458231809,-0.4647041106\H,-3.8746416
058,-0.9406232589,0.4510946171\N,-4.788268784,0.5496326049,-0.78155577
19\H,-0.3631056319,-3.0338796624,-0.1055660392\H,-3.9519290174,1.13598
77681,-0.5972155753\O,-1.2861598415,-1.0376017091,-0.7619323348\H,-6.0
126348489,2.1017795945,-0.0694702481\H,-6.7884327917,0.5009487559,-0.1
173029634\H,-5.6251158867,0.8896072863,1.1785578276\H,-0.27988464,-3.1
313072749,2.9898455227\C,-0.9860292024,-2.3311267759,3.2372138433\C,-0
.4821409176,-0.9986975765,2.7168261105\H,-1.0934987981,-2.2897290013,4
.3264661045\H,-1.9593123444,-2.5743359726,2.7986317184\O,-0.3211953932
,-1.129123347,1.2976464995\H,-1.1951939902,-0.1929049675,2.931916624\H
,0.4818059516,-0.7316587627,3.1617452192\P,-0.5315181203,0.1094975384,
0.2535554875\O,-1.6256288343,1.1470886885,0.3408049425\O,0.555409543,1
.0656503635,1.4960701216\O,0.780580174,0.2941536255,-0.7163884795\C,0.
5543803147,2.3849573053,1.5283178687\C,0.7469733392,0.0734498424,-2.13
15770924\C,0.9089267935,3.1747806224,0.3984363449\C,0.2129908257,3.034
6437611,2.7279718354\C,2.1379733846,0.3591175673,-2.6616523499\H,0.007

0375508,0.7466404974,-2.5883670353\H,0.4573703442,-0.9617196647,-2.341
6885634\C,0.9079422527,4.5502861363,0.477966108\H,1.1695794105,2.66563
43929,-0.5256435277\C,0.2166762899,4.419497564,2.7787588681\H,-0.07707
02977,2.4505878879,3.6004521181\H,2.4368554774,1.3900324303,-2.4467262
493\H,2.1615386364,0.2046733119,-3.7456460135\H,2.8661379672,-0.316000
0653,-2.1988768902\C,0.5598684237,5.2130208126,1.6725266364\H,1.169875
6524,5.1507434866,-0.3931155108\O,-0.1524271282,5.0037976211,3.9633818
713\C,0.5030340805,6.6352004355,1.826245323\C,-0.2330463573,6.35921944
09,4.1334837193\C,0.1203072344,7.1947832497,2.9996994394\H,0.765592655
5,7.2644203025,0.9750963241\O,-0.5910031142,6.760619906,5.226982113\H,
0.0565468561,8.2688583401,3.1556165577\H,-1.7396315785,5.6406205246,-0
.1653632135\C,-2.5524202719,4.9204394118,-0.4295199619\H,-3.454671406,
5.5271685144,-0.6492210862\N,-2.8496083181,4.0259129987,0.6860431206\C
,-2.1115079321,4.1362096488,-1.6727139922\H,-2.0783374603,3.3424811297
,0.7875275077\H,-1.6367805142,4.8355423505,-2.3884192041\H,-1.33472967
01,3.4111496203,-1.353635129\N,-3.1379337573,3.3911649916,-2.400831220
2\H,-3.0602766598,4.0184391237,2.7757459691\H,-2.2431438865,5.47797484
53,2.1530454942\H,-3.9951941479,5.2760927798,1.9346638834\H,-3.6713739
149,2.8049252294,-1.73722902\H,-3.4920599621,4.7344195971,-3.968249531
4\H,-4.8021160393,3.5784598432,-3.6557643321\H,-4.5875104955,4.9866928
487,-2.5889756638\\Version=ES64L-G09RevD.01\State=1-A\HF=-
2466.6966128\RMSD=4.022e-09\RMSF=6.479e-08\Dipole=1.3108453,-0.3101247,-
2.7883842\Quadrupole=32.6189221,-34.3899926,1.7710704, 23.2313271,-1.8871408,-
50.3653837\PG=C01 [X(C30H53N5O9P1)]\\@

dECP – TS2

1\1\GINC-R3303\FTS\RM062X\6-31+G(d,p)\C30H53N5O9P1(1-)\ROOT\19-Nov-201
4\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultrafine
) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(pcm
,read) geom=check guess=read\\Title Card Required\\-1,1\C,-10.44104554
12,-16.977177927,-23.3515116049\C,-12.020461884,-13.9316778521,-25.946
3454124\C,-21.8341375161,-9.0559906254,-22.3338471642\C,-20.0328879818
,-5.8680968909,-25.6196658996\C,-19.0167726865,-5.3511762237,-20.49296
59947\C,-19.4809313903,-11.5094774809,-24.0821579243\C,-10.7392748588,
-18.4759388349,-23.2681007876\C,-11.8216535911,-18.9254671875,-24.2696
188976\C,-13.0220389372,-18.023563713,-24.1082249186\O,-13.1031432229,
-17.0789200098,-25.0239643721\O,-13.795257421,-18.1165816125,-23.16261
13859\H,-11.3077899595,-16.3894880342,-23.0293516358\H,-9.5992870204,-
16.7093028859,-22.7062150604\H,-10.1991666815,-16.6766029479,-24.37750
75891\H,-9.82825905,-19.055748333,-23.4615638711\H,-11.0735382431,-18.
7253385088,-22.2527884252\H,-12.1173356511,-19.9603654499,-24.06501795
81\H,-11.4451428127,-18.8479587538,-25.2955114301\H,-13.6695799208,-16
.2645232729,-24.6709397475\N,-14.1963282736,-14.9655578445,-24.0673810
428\C,-13.4529692496,-13.8187356994,-23.8667021017\C,-15.3420851356,-1
4.792248534,-23.4299867436\C,-12.079108297,-13.6475015875,-24.44162650
75\C,-14.1856418064,-12.9520760779,-23.0947234191\N,-15.3731345592,-13
.5903153777,-22.8284691234\H,-16.151025143,-15.5126055828,-23.37923256
93\H,-11.377523303,-14.3111001446,-23.9203425177\H,-11.7481081276,-12.
6219047876,-24.235151101\H,-13.9757662262,-11.9586060366,-22.718975228
6\H,-11.0042835391,-13.7840599505,-26.3263968509\H,-12.3176025445,-14.
9624107416,-26.1603542208\H,-12.6938421751,-13.2609631772,-26.49211091

18\H,-19.2889550514,-12.5791422769,-23.8727357981\H,-20.0700558428,-11
.4461987815,-25.017893013\C,-18.1485186792,-10.8150376087,-24.31523161
54\C,-20.3283008208,-10.9132056874,-22.9425297591\H,-18.2776104429,-9.
724540201,-24.3558857413\H,-17.721954738,-11.1439344686,-25.2738632844
\H,-21.2443180873,-11.5169452898,-22.8441308779\H,-19.77759878,-10.999
9356344,-21.9834484053\N,-20.7377582245,-9.5231352777,-23.1992067925\H
,-16.1785270139,-13.1687705521,-22.3266038424\H,-19.9130683096,-8.9246
244199,-23.0008847663\O,-17.2268587887,-11.1795456625,-23.2895986601\H
,-21.9984830817,-7.9881257066,-22.5128190929\H,-22.7539320214,-9.59771
68387,-22.5866564036\H,-21.6306625514,-9.2000742378,-21.2584209279\H,-
16.6092486884,-13.1680856733,-19.4457480095\C,-17.3093216528,-12.33111
16405,-19.3478659707\C,-16.6624878681,-11.0382395393,-19.8049221405\H,
-17.601030687,-12.2481326251,-18.2954049132\H,-18.2034693997,-12.54801
2427,-19.9417583976\O,-16.2576190041,-11.2120643143,-21.1764176479\H,-
17.3589104741,-10.1938323916,-19.7355506064\H,-15.771047593,-10.797014
4281,-19.218780377\P,-16.4923658446,-10.061944712,-22.2874306675\O,-17
.5169088457,-8.9739063043,-22.1773902122\O,-15.3115436524,-8.987898430
1,-20.8836541384\O,-15.1355039877,-9.87289358,-23.156208438\C,-15.3375
798352,-7.6900341035,-20.877441892\C,-15.1291730484,-10.06781988,-24.5
81914889\C,-15.0169260311,-6.9162499234,-22.0408054909\C,-15.683868123
3,-6.9893473628,-19.6958362503\C,-13.726537438,-9.7641179827,-25.06640
88029\H,-15.8622493823,-9.3903085669,-25.0434717898\H,-15.407639353,-1
1.102465064,-24.8125147226\C,-15.0491321521,-5.5418657007,-22.01063733
11\H,-14.7542010885,-7.4510695444,-22.9504222729\C,-15.7124669631,-5.6
064768052,-19.6967402114\H,-15.9464232473,-7.5444693033,-18.7961225572
\H,-13.4426639323,-8.7345199856,-24.8273991134\H,-13.6720974919,-9.902
5458748,-26.151326476\H,-13.0078153362,-10.4413761711,-24.5926644942\C
,-15.3991386249,-4.8409264798,-20.835051882\H,-14.8111453555,-4.966409
8234,-22.905629349\O,-16.0794167295,-4.9840103625,-18.528763283\C,-15.
4806418398,-3.4208823537,-20.7282132951\C,-16.1769537232,-3.6239519661
,-18.4032778708\C,-15.8557463132,-2.823642391,-19.5670159206\H,-15.242
0579058,-2.8162266251,-21.6041286525\O,-16.5252172647,-3.1933263469,-1
7.3155138664\H,-15.9345152904,-1.7460908842,-19.4490715209\H,-17.73818
08288,-4.4130589867,-22.612102731\C,-18.5378732734,-5.1492062502,-22.8
716116563\H,-19.4554160065,-4.5626582202,-23.0820224868\N,-18.80669066
6,-6.0532724611,-21.7556497271\C,-18.0910385661,-5.9248503342,-24.1180
38367\H,-18.0044889363,-6.7026778415,-21.6537943293\H,-17.6326931433,-
5.2178772162,-24.836811333\H,-17.2985714121,-6.6345600023,-23.80260611
59\N,-19.1098755864,-6.6867054764,-24.8396198546\H,-19.0273459932,-6.0
721518969,-19.6681514588\H,-18.2375835918,-4.594991582,-20.2831947868\
H,-19.9848157395,-4.8326894032,-20.5111008579\H,-19.6396396676,-7.2687
919447,-24.1699642564\H,-19.4764526496,-5.3590702035,-26.4174294623\H,
-20.7810550167,-6.5165110045,-26.0885489462\H,-20.5656908534,-5.098406
3783,-25.035072842\\Version=ES64L-G09RevD.01\HF=-2466.6961412\RMSD=5.270e-
09\RMSF=1.131e-06\Dipole=0.7273153,-1.4278436,-
3.4003155\Quadrupole=9.7310733,575.1821694,-584.9132427,-527.1981536,-1216.5893974,-
743.2174484\PG=C01 [X(C30H53N5O9P1)]\\@

dECP – Product

1\1\GINC-R2925\FOpt\RM062X\6-31+G(d,p)\C30H53N5O9P1(1-)\ROOT\04-Dec-20
14\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight) IOP

(2/17=4) Freq=noraman geom=check Guess=Read scrf=(pcm,read)\\opt fwd\\
-1,1\C,8.9339221331,-2.5768962321,-0.1603067032\C,6.1987118655,-0.9614
516652,2.740054285\C,-2.6631380327,4.9579958448,-1.6820856982\C,-4.111
1445709,3.835556514,2.8832173246\C,-5.3685145457,0.7754268101,-1.19529
37546\C,0.8182033856,4.2256738553,-0.2846499891\C,9.8245485096,-1.8801
182492,-1.1913519943\C,9.7419569405,-0.3428985552,-1.1145649691\C,8.29
07877986,0.0669946388,-1.199387017\O,7.749644942,0.3466010905,-0.02929
12243\O,7.6622369175,0.0624027551,-2.2504310791\H,7.8743571929,-2.3887
784494,-0.3664719852\H,9.0897037913,-3.659495322,-0.1782072354\H,9.141
9963028,-2.2150078563,0.8530563025\H,10.8712309407,-2.1751894428,-1.04
72407343\H,9.534341234,-2.1982202972,-2.2009678077\H,10.289566804,0.10
45311154,-1.9511772954\H,10.1647279016,0.0124381627,-0.1683933961\H,6.
7073514718,0.2517342352,-0.0664437431\N,5.2294902071,-0.1962506082,-0.
030385891\C,4.7111583048,-1.2455935688,0.7041086835\C,4.3032223981,0.1
476960774,-0.9081837285\C,5.4912674761,-1.9312046195,1.7872629732\C,3.
4462427187,-1.5160293216,0.2459334115\N,3.2100393161,-0.6243921135,-0.
7721987609\H,4.3965552104,0.9318110031,-1.6512372664\H,6.2371331731,-2
.5974349805,1.3315897785\H,4.8011526657,-2.5765621081,2.3460793166\H,2
.7055208149,-2.2464328324,0.5514150754\H,6.7285001197,-1.5141544674,3.
5224771399\H,6.9279345039,-0.3483514772,2.2035268418\H,5.4757811176,-0
.2918532625,3.2193091774\H,1.7851686832,4.1394355459,-0.8139415459\H,0
.8725142203,5.102335225,0.388447809\C,0.6092967578,3.003452366,0.58449
57983\C,-0.2991842259,4.4447049455,-1.3196976926\H,-0.3260642058,3.076
9630078,1.1570035699\H,1.442634272,2.8612851917,1.2835605146\H,-0.1152
988192,5.4060749909,-1.8234208356\H,-0.2359568404,3.6595371289,-2.0989
795691\N,-1.6394520219,4.5075903258,-0.7255051568\H,2.3462655359,-0.55
67002794,-1.3370360824\H,-1.9067844046,3.5590249831,-0.4044434301\O,0.
5462125378,1.8463540005,-0.2808451759\H,-3.6496422431,4.8840575015,-1.
2133461448\H,-2.4778397747,6.006884278,-1.9430583523\H,-2.6732512601,4
.3666084543,-2.6146542367\H,-0.438743387,1.0035877534,-3.2686313323\C,
-1.3605072481,0.7282982424,-2.7455833914\C,-1.168955188,-0.5366245049,
-1.9311926813\H,-2.1501304658,0.5633170529,-3.4865695584\H,-1.66802685
38,1.563791467,-2.1044982397\O,-0.035669766,-0.420620301,-1.0291273704
\H,-2.0565230058,-0.7700513671,-1.3367961631\H,-0.9302425001,-1.400730
5232,-2.5570632519\P,-0.1726204581,0.5215436736,0.2451515333\O,-1.5459
741287,0.7089259414,0.7685702096\O,-1.5406250227,-2.8030360292,0.22515
6037\O,0.8882809103,-0.0336720315,1.287990158\C,-2.7937292978,-2.64660
06223,0.3794076321\C,0.5213046173,-1.116258916,2.1813646231\C,-3.37212
30044,-2.3642453579,1.6777761695\C,-3.7135576653,-2.7358442683,-0.7177
442645\C,1.5135759299,-1.1149124492,3.3231230831\H,0.5374440153,-2.050
4962747,1.6083364191\H,-0.5041237936,-0.9463297683,2.5317709095\C,-4.7
237591565,-2.2122379861,1.8488936353\H,-2.6866723869,-2.2765851171,2.5
201977389\C,-5.067053405,-2.5742172776,-0.511148869\H,-3.3361639544,-2
.9498847491,-1.7169174402\H,2.5351511459,-1.2292544619,2.9460711383\H,
1.2924286098,-1.9456627118,4.0011597866\H,1.4515095853,-0.1777301342,3
.8849886667\C,-5.6257386275,-2.3155618112,0.75987717\H,-5.1320310155,-
1.9997812663,2.8376467815\O,-5.8898914301,-2.6780703779,-1.6077235211\
C,-7.0347959559,-2.1835787384,0.8531573813\C,-7.2533229429,-2.53811648
44,-1.5439094927\C,-7.8343688721,-2.2887386278,-0.2473092102\H,-7.4768
387797,-1.9888240183,1.8313353144\O,-7.8623895313,-2.6395205703,-2.600
1271374\H,-8.915237465,-2.188388903,-0.2093773227\H,-6.0206105597,0.59

22426918,1.334922231\C,-5.2581429825,1.3880092362,1.1554035105\H,-5.81
30403311,2.33986068,1.0331533849\N,-4.5040154666,1.1271651735,-0.07013
85386\C,-4.3416015068,1.4447644432,2.3839432109\H,-3.8438598769,0.3507
796176,0.1107608758\H,-4.9698481298,1.4135068143,3.2948990802\H,-3.718
6182618,0.5293748398,2.3822158766\N,-3.4456141743,2.593412058,2.506630
8207\H,-4.7571739683,0.4190157026,-2.0326482446\H,-6.1061392586,-0.008
0579521,-0.9465082275\H,-5.9234564463,1.6627216529,-1.5280468257\H,-2.
928942191,2.708009207,1.6217399133\H,-4.5340326454,3.7254065323,3.8901
488879\H,-3.3755601716,4.6468072623,2.9112311247\H,-4.9273156307,4.141
2684553,2.2056875589\\Version=ES64L-G09RevD.01\State=1-A\HF=-
2466.7144401\RMSD=1.775e-09\RMSF=1.196e-07\
Dipole=4.0296618,2.9029243,2.2121522\Quadrupole=-60.6428012,23.4348964, 37.2079048,-
36.6352876,-18.9511655,-18.7664055\PG=C01 [X(C30H53N5O9P1)]\\@

A.1.4 Dephosphorylation – large system

dECP – Reactant

1\1\GINC-R2947\FOpt\RM062X\6-31+G(d,p)\C23H51N5O9P1(1-)\ROOT\27-Nov-20
14\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight) IOP
(2/17=4) Freq=noraman scrf=(pcm,read)\\title\\-1,1\C,-8.5396895113,1.9
076553029,-0.323851214\C,-5.598691672,0.7764691592,2.7102878544\C,4.89
62726237,-2.8515003411,-1.9595391633\C,2.4420245986,-3.2659305806,-1.9
536087313\C,5.8071502379,-2.1935230903,2.8618556762\C,6.1074597231,2.3
406315499,0.169782578\H,-7.9126547274,1.5962232562,0.5202401313\H,-8.7
706075107,2.9688512597,-0.1938951606\H,-9.4773510799,1.3443708132,-0.2
669401632\C,-7.8299583757,1.6540390471,-1.6539226767\C,-7.5098678243,0
.1715422013,-1.880868901\H,-8.4647960976,1.9965336569,-2.4807782165\H,
-6.9002171708,2.2340193294,-1.7017913882\C,-6.4972158183,-0.3963670774
,-0.9085352766\H,-8.4157213837,-0.4406380824,-1.8074966169\H,-7.086859
833,0.0356018671,-2.886544702\O,-6.6832480493,-1.4229169522,-0.2704265
774\O,-5.3822319281,0.309561588,-0.8412691119\H,-4.6820413567,-0.10994
16504,-0.2088167877\N,-3.4615253982,-0.4137878473,0.8095640199\C,-3.28
20395075,0.4533213719,1.8692371755\C,-2.4339965524,-1.2426822043,0.824
2875662\C,-4.3240892257,1.4678928878,2.2135517774\C,-2.1225480139,0.12
59685633,2.5222165784\H,-2.2420689233,-2.0333103529,0.1079954342\N,-1.
6086129334,-0.9587120165,1.8512728137\H,-3.9279744859,2.1439443927,2.9
810059036\H,-4.5538348138,2.0762197743,1.3272695871\H,-1.6272080795,0.
5615518158,3.3826912202\H,-0.7125471457,-1.4219211684,2.0717867698\H,-
6.3806060805,1.5082281236,2.9405597998\H,-5.9847352587,0.0882840506,1.
9498006668\H,-5.3922105514,0.195162265,3.6154236667\H,0.2880441142,-3.
0880444815,0.7708259958\C,1.0335408276,-2.7027471551,0.062951419\H,1.9
466291987,-2.4568170926,0.6220212241\O,0.4834593473,-1.4979163297,-0.5
253497045\C,1.2983123586,-3.7009502668,-1.0478888842\P,0.937609275,-0.
0587272339,0.0019250343\H,0.3704717847,-3.817874714,-1.6375458961\H,1.
5363246571,-4.6822539938,-0.5931487711\O,2.1268188316,0.4303762871,-0.
9339325802\O,-0.3526677823,0.7644604042,-0.4075128099\O,1.3332995768,-
0.0394273116,1.4352170935\H,2.3148354421,-2.1913666655,-2.1956130334\H
,2.4011760517,-3.8156298809,-2.9060619108\N,3.7531836049,-3.5341835686
,-1.3324716253\C,1.8499442048,0.9152398098,-2.2804226938\C,-0.46957775
17,2.1516814874,0.0243360485\H,3.7353245254,-3.2597738356,-0.336313648
4\H,1.7049199023,1.9997414504,-2.1954175694\C,3.0308970535,0.593947955

5,-3.1674960906\H,0.9311831939,0.4424127123,-2.6497946861\C,-1.7153622
089,2.7092267439,-0.6238286701\H,0.4346998037,2.6969573299,-0.27754204
9\H,-0.5424904788,2.1566616135,1.1184197635\H,4.7976601208,-1.75157383
95,-1.9639733277\H,5.8078160911,-3.1159239971,-1.4150237761\H,4.997721
6542,-3.1901335346,-2.9985090833\H,3.1739941965,-0.4860639195,-3.26346
20805\H,3.9390572678,1.0542475783,-2.767143456\H,2.8513423792,1.015590
6805,-4.1622813667\H,-1.8684461934,3.7402033515,-0.2891074086\H,-2.590
0800875,2.1113272786,-0.3462851904\H,-1.6165318231,2.7066913516,-1.713
9635292\H,1.7538036928,1.6822795291,2.2727512228\O,1.9352908626,2.6090
279011,2.4971613503\H,2.241432959,2.9955079603,1.647052766\O,2.7071852
87,3.4692647261,-0.0305718715\C,3.8845948921,3.4178781813,-0.496637879
4\C,5.0484135338,3.3996291105,0.5034913182\O,4.1738964688,3.4006113158
,-1.7206380587\H,4.6550946783,3.2686717734,1.5203055529\H,5.5259054905
,4.3922216096,0.4593591505\H,6.5451756651,2.5645862496,-0.8175874212\H
,6.9184458551,2.3954321238,0.9203486601\N,5.6219841139,0.9498568486,0.
1296239185\H,4.9737447681,0.8717443987,-0.6757543019\C,4.8791877803,0.
5373380937,1.327197292\H,5.525279535,0.7047292183,2.2126958119\H,3.954
8747299,1.136968345,1.4736599902\C,4.4682641453,-0.9354071877,1.241388
5625\H,4.0175986944,-1.09081479,0.2376291054\H,3.6672988116,-1.1254740
815,1.9792564282\N,5.5168938439,-1.9383641896,1.4485188017\H,6.3736130
761,-1.6247068223,0.962824516\H,6.6454344085,-2.8939271675,2.940407596
3\H,6.0579928307,-1.2873548989,3.4393170895\H,4.9305976999,-2.65829612
74,3.3294913192\\Version=ES64L-G09RevD.01\State=1-A\HF=-
2198.8948889\RMSD=5.878e-09\RMSF=1.529e-07\Dipole=-3.0878842,-
3.6282344,0.2637264\Quadrupole=-34.0248034,0.4848426,33.5399609,-
35.5879344,13.1934709,0.8329525\PG=C01 [X(C23H51N5O9P1)]\\@

dECP – TS1

1\1\GINC-R3206\FTS\RM062X\6-31+G(d,p)\C23H51N5O9P1(1-)\ROOT\15-Nov-201
4\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultrafine
) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(pcm
,read)\\title\\-1,1\C,-9.06205,0.880357,0.161422\C,-6.135372,-0.882881
,2.892586\C,4.598568,-2.796968,-2.230887\C,2.164716,-3.290222,-2.40540
5\C,5.384218,-3.3674,2.623781\C,5.539694,1.714303,1.192401\H,-8.046139
4976,0.6170428646,0.4738013776\H,-9.1078558376,1.9688204969,0.06053190
88\H,-9.7479260335,0.5868388478,0.9645115028\C,-9.4275259203,0.1855326
17,-1.1497823054\C,-9.4219410629,-1.3406260239,-1.0115750702\H,-10.425
158105,0.5039445585,-1.4780728483\H,-8.7182781907,0.4798685118,-1.9321
924297\C,-8.1078000414,-1.8576200416,-0.4651421489\H,-10.2148601853,-1
.6840131368,-0.3372245563\H,-9.5840426209,-1.8143865146,-1.9905837001\
O,-8.0367457438,-2.7184885785,0.4025376931\O,-7.0570009354,-1.28105639
22,-1.0100628873\H,-6.1609075827,-1.4754264131,-0.4949187142\N,-4.7640
972605,-1.501483061,0.1160209237\C,-4.2437600332,-0.8288366822,1.20400
59565\C,-3.7377519526,-2.0379924793,-0.5240735278\C,-5.1075293717,-0.0
346973742,2.1355992968\C,-2.8802586163,-0.9902978727,1.2028835315\H,-3
.7974157965,-2.6200826738,-1.437202237\N,-2.5842937869,-1.760220624,0.
1046373968\H,-4.4550469604,0.4918757298,2.8435259662\H,-5.6308041827,0
.7367697944,1.5512047844\H,-2.1055454642,-0.6175008354,1.8616965716\H,
-1.6374364222,-1.944204634,-0.2587345198\H,-6.7666250121,-0.250944901,
3.5261347938\H,-6.7813585135,-1.4261876569,2.1949468932\H,-5.634085195
4,-1.6204054081,3.5287026173\H,0.1809207321,-3.4396363522,0.4431976055

\C,0.8270631408,-2.9347630814,-0.2909828558\H,1.7841813357,-2.73983163
17,0.2056301076\O,0.2074879855,-1.7075482803,-0.675869895\C,1.02981062
4,-3.8033079828,-1.5257232278\P,0.6769562764,-0.2612296856,0.109494940
3\H,0.084869436,-3.8256706491,-2.1004785754\H,1.254296857,-4.842487478
2,-1.2120008533\O,1.9891934049,-0.0430092694,-0.8406512861\O,-0.647068
2177,0.3885805883,-0.5458559728\O,0.8347489759,-0.7713990723,1.5196582
698\H,2.0677644823,-2.1915210852,-2.4993176545\H,2.0905452115,-3.71070
8276,-3.4201962096\N,3.4793299814,-3.6663886436,-1.8451984707\C,1.7296
958114,0.537964468,-2.1376348834\C,-1.2686696197,1.6575398201,-0.30552
60653\H,3.4290859349,-3.6718273907,-0.8114146097\H,1.233512625,1.50664
19014,-1.9984387217\C,3.040139775,0.7296777376,-2.8664680295\H,1.05575
57033,-0.1319150047,-2.6899555609\C,-2.5395254117,1.678875443,-1.13220
11704\H,-0.5880036003,2.4635508431,-0.5985504497\H,-1.4831086308,1.759
2825727,0.7647642226\H,4.479496901,-1.7567289412,-1.8834745465\H,5.525
6311817,-3.1987707254,-1.808841065\H,4.6920310144,-2.7832597815,-3.324
7755145\H,3.5308145781,-0.2287779348,-3.0558613234\H,3.7091531492,1.37
70128195,-2.2913893495\H,2.8464572512,1.2149314661,-3.8293043381\H,-3.
0590004733,2.6315462046,-0.9831673586\H,-3.2103341203,0.864750272,-0.8
374049072\H,-2.3050984302,1.5728075973,-2.196470622\H,1.3501872357,1.3
665216675,1.7279310257\O,1.1015585003,1.5423719969,0.8045227053\H,1.91
6788888,2.2336393613,0.2434043725\O,2.6095228602,2.9303145862,-0.43657
07303\C,3.8658434453,3.0714913199,-0.1568416745\C,4.3491873374,2.68617
1828,1.2381525606\O,4.6769547632,3.5311581074,-0.9684019694\H,3.525750
1486,2.2796365963,1.8386428773\H,4.6759200809,3.62493507,1.7143773787\
H,6.2853235463,2.1089206218,0.4788636301\H,6.0164941721,1.6805549613,2
.1897523896\N,5.214061568,0.3352138092,0.8132321187\H,4.7162994695,0.3
44199084,-0.099896471\C,4.3713563736,-0.3699624671,1.7837914893\H,4.91
13365811,-0.3975916783,2.7519458268\H,3.3968718382,0.1369277357,1.9566
256363\C,4.0579804435,-1.787052998,1.3020389915\H,3.6718243752,-1.6956
830456,0.2655354711\H,3.2269000451,-2.195805209,1.9095754256\N,5.15300
49942,-2.7594646289,1.3133356309\H,6.0148200018,-2.291978778,0.9810501
253\H,6.2527289131,-4.032620757,2.568591179\H,5.5596174368,-2.63846965
15,3.4340753085\H,4.5104824614,-3.9711895752,2.8980755066\\Version=ES6
4L-G09RevD.01\HF=-2198.8653355\RMSD=4.721e-09\RMSF=2.123e-06\Dipole=-
1.9494455,-0.5684973,0.6754163\Quadrupole=-45.4132262,2.0626277,43.3505985,-
37.6888311,24.3205852,12.6315323\PG=C01 [X(C23H51N5O9P1)]\\@

dECP – Intermediate

1\1\GINC-R3163\FOpt\RM062X\6-31+G(d,p)\C23H51N5O9P1(1-)\ROOT\14-Sep-20
14\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight) IOP
(2/17=4) Freq=noraman geom=check Guess=Read scrf=(pcm,read)\\title\\-1
,1\C,-8.7308364356,1.256351575,-0.0128152822\C,-5.8041478824,-0.506887
018,2.7183412307\C,4.9298072307,-2.4210007674,-2.4051477032\C,2.495914
642,-2.9141844379,-2.579603213\C,5.7154352335,-2.9913965228,2.44953993
86\C,5.8708797739,2.0902777867,1.0181612214\H,-7.7191894247,0.97729231
59,0.3000673205\H,-8.7488358432,2.341243517,-0.1526333485\H,-9.4158778
337,1.0091360923,0.8063126171\C,-9.1250304442,0.5239527685,-1.29513630
93\C,-9.1446073377,-0.9972401044,-1.1057098688\H,-10.1208106063,0.8486
89953,-1.6227891696\H,-8.420186016,0.7783399344,-2.0953583996\C,-7.832
6438789,-1.5169340149,-0.5572465706\H,-9.9344244477,-1.3028711163,-0.4
100496157\H,-9.3268986184,-1.4994300127,-2.0666969585\O,-7.7623239892,

-2.3211693143,0.3622520445\O,-6.7801142882,-1.0060092482,-1.1636513126
\H,-5.8834173479,-1.1763269022,-0.6544107368\N,-4.4475819138,-1.065115
7944,-0.0681982122\C,-3.9995400573,-0.2341500155,0.9407894787\C,-3.376
7072603,-1.4433906547,-0.7481796958\C,-4.9368653654,0.4505314623,1.890
9241355\C,-2.6329179163,-0.1341589984,0.8497320434\H,-3.3773284778,-2.
0896188562,-1.6190946687\N,-2.2616395288,-0.9063232274,-0.224342832\H,
-4.3404947159,1.0894156384,2.5545127056\H,-5.589619038,1.1208164684,1.
3122590764\H,-1.9031987445,0.4133204702,1.4361414493\H,-1.3016011185,-
0.9650881539,-0.6204097076\H,-6.4884595875,0.055446317,3.3622144374\H,
-6.3987287185,-1.1552343401,2.0673504882\H,-5.178432553,-1.1453457685,
3.3514384835\H,0.3654183196,-2.5694076296,0.1837318654\C,1.2034866563,
-2.2487834186,-0.4550045333\H,2.1028496397,-2.3181105807,0.1666925535\
O,0.9943940151,-0.9166966141,-0.8973779635\C,1.3066307462,-3.184339357
7,-1.6584001094\P,1.2009179722,0.4371881505,0.1786478231\H,0.377437950
9,-3.1051785091,-2.2547755679\H,1.3672501047,-4.2245218857,-1.27890916
28\O,2.5205910681,0.9773066659,-0.6446580826\O,-0.1433051048,1.0412796
963,-0.5301510605\O,1.2127036928,-0.3693789961,1.4673064666\H,2.442470
635,-1.86628542,-2.9225814664\H,2.4111599045,-3.5456820769,-3.47865465
04\N,3.7854094282,-3.2032027879,-1.9266178782\C,2.6174824887,0.8564523
804,-2.0693605056\C,-0.5583053459,2.410033585,-0.5382902897\H,3.710098
1001,-3.0754460315,-0.9057266941\H,1.6184417113,0.9475457236,-2.515438
9304\C,3.541505876,1.9441163214,-2.5802082527\H,3.0042824925,-0.135439
0061,-2.3165598984\C,-1.6949767472,2.5180321901,-1.5372841732\H,0.2848
054441,3.0513041653,-0.8267482783\H,-0.8818318944,2.7050118425,0.46613
29137\H,4.7828177259,-1.3285334478,-2.3266907822\H,5.815196323,-2.6939
106994,-1.8217136922\H,5.1239134913,-2.657882157,-3.4594554726\H,4.534
7358395,1.8655566512,-2.1217836049\H,3.145255392,2.9379300315,-2.35335
15792\H,3.6588807907,1.8435000535,-3.6649550063\H,-2.0593054425,3.5500
726071,-1.5778060453\H,-2.5250771054,1.8645477647,-1.2462067465\H,-1.3
568756089,2.227052266,-2.5373422543\H,1.4246682814,1.7231496872,2.0625
503886\O,1.415840181,1.9809250439,1.1263445126\H,2.2902614621,3.037970
2819,0.6064748836\O,2.7864785762,3.7999416187,0.0998756213\C,4.0983271
9,3.7746528403,0.2097061606\C,4.7023084766,3.0285182667,1.3862403252\O
,4.8027308314,4.4068115875,-0.5675183295\H,3.9227875792,2.4965773123,1
.9438551147\H,5.0851891888,3.8273740069,2.0433089672\H,6.4485928844,2.
5477805691,0.1936426912\H,6.5455125133,2.0191918371,1.891500827\N,5.50
78601557,0.725360117,0.6390869483\H,4.9067106963,0.7616654756,-0.20142
32658\C,4.7680280715,-0.0036482212,1.6724844413\H,5.4052906298,-0.0692
839453,2.5776943797\H,3.823781639,0.5051083247,1.9621096789\C,4.394372
2021,-1.3949661645,1.1684847293\H,3.9904344079,-1.263732651,0.14124533
16\H,3.5637964396,-1.7949440864,1.7812494417\N,5.4603816966,-2.3963592
421,1.1383707463\H,6.3233460319,-1.953007046,0.7771648209\H,6.56505263
94,-3.6793903518,2.3798922077\H,5.9322507686,-2.2557789656,3.243394623
2\H,4.835109398,-3.5686238455,2.7576277428\\Version=ES64L-G09RevD.01\State=1-
A\HF=-2198.8629109\RMSD=5.778e-09\RMSF=7.600e-08\Dipole=-1.1348554,-
0.0368793,0.0134856\Quadrupole=-37.9066536,-5.6025592,43.5092128,-
36.0485416,20.7743718,10.4317035\PG=C01 [X(C23H51N5O9P1)]\\@

dECP – TS2

1\1\GINC-R3581\FTS\RM062X\6-31+G(d,p)\C23H51N5O9P1(1-)\ROOT\10-Sep-201
4\0\\#m062X/6-31+G(d,p) 6D maxdisk=15GB IOP(2/17=4) INT(grid=ultrafine

) OPT=(TS,calcfc,noeigentest,maxcyc=200) freq=noraman nosymm scrf=(pcm
,read) geom=check guess=read\\title\\-1,1\C,-9.0620495614,0.8803565516
,0.1614215374\C,-6.1353715852,-0.8828812557,2.8925863634\C,4.598567866
3,-2.7969682328,-2.2308869472\C,2.1647158212,-3.2902216932,-2.40540497
91\C,5.3842179196,-3.367400384,2.6237810756\C,5.5396935395,1.714303014
1,1.1924009499\H,-8.0819690746,0.6473523524,0.5901549057\H,-9.16014858
12,1.9696393971,0.1136763171\H,-9.8299061538,0.5026738109,0.8467677351
\C,-9.2049329214,0.2418819739,-1.219782659\C,-9.0935677287,-1.28010170
06,-1.1622798823\H,-10.1762231089,0.5143367881,-1.6543202906\H,-8.4278
404703,0.6330932825,-1.8869249464\C,-7.734484628,-1.7920059115,-0.6763
513252\H,-9.8668774162,-1.7073779165,-0.5112476891\H,-9.2473634585,-1.
712159634,-2.1623130899\O,-7.6851324106,-2.905010627,-0.1086556747\O,-
6.7255809259,-1.0488106551,-0.9122421298\H,-5.3947801788,-1.3149753768
,-0.192434396\N,-4.3974525924,-1.3143134253,0.2264591083\C,-4.00620882
2,-0.788164071,1.4383715754\C,-3.3047125266,-1.7449839792,-0.415731510
3\C,-4.9468641473,-0.0662276752,2.3607523862\C,-2.6423100366,-0.945678
8906,1.4897041062\H,-3.3197395478,-2.1850871503,-1.406442424\N,-2.2174
837274,-1.5421725566,0.32073565\H,-4.3502031806,0.3163444482,3.1983266
949\H,-5.3329939402,0.8119936419,1.8213513639\H,-1.9360370858,-0.66188
20354,2.2586302511\H,-1.0414658594,-1.5155225794,-0.1055182392\H,-6.78
16104251,-0.2488152067,3.5071874723\H,-6.7325361829,-1.2878078836,2.06
69000725\H,-5.7896293767,-1.7249494203,3.5010140942\H,0.2962643407,-3.
1329873365,0.5583941634\C,0.837996434,-2.5902321335,-0.2309934804\H,1.
8223606064,-2.3508929289,0.1815535208\O,0.1434875031,-1.3653056617,-0.
4797320943\C,0.9660146443,-3.5039250838,-1.4631705487\P,0.6468076898,0
.1625913526,0.6610059993\H,0.0297069177,-3.4395552355,-2.0490297746\H,
1.0332888145,-4.5432506614,-1.0851325239\O,2.0099319864,0.3233873573,-
0.1928482001\O,-0.5855912877,0.8164700709,-0.1572909874\O,0.524449928,
-0.7083931881,1.8878175668\H,2.1210649006,-2.284468134,-2.8579703629\H
,2.0729267745,-4.0026361494,-3.2410643051\N,3.441099106,-3.5285607875,
-1.7199795017\C,2.1289892644,0.0998046631,-1.6058070954\C,-1.363003222
5,1.928998964,0.3159348608\H,3.3571299429,-3.3629039454,-0.7088462719\
H,1.1370487423,-0.0040836326,-2.0529136871\C,2.8997669456,1.2379877049
,-2.2409642196\H,2.6614783526,-0.8461107876,-1.7250131209\C,-2.6117614
912,1.9917572354,-0.5391664591\H,-0.7670363328,2.844312819,0.231322236
\H,-1.6185170633,1.7747891154,1.3719485428\H,4.4499467914,-1.702651188
7,-2.2869795761\H,5.4569855783,-2.9957675207,-1.5809941482\H,4.8431246
079,-3.1479534907,-3.2419668022\H,3.8949572598,1.3495932272,-1.7963258
804\H,2.3691292153,2.1866869174,-2.1246331886\H,3.0273618119,1.0277476
195,-3.3089602508\H,-3.2346902384,2.8347440888,-0.221450917\H,-3.19538
64256,1.0708927265,-0.4412597827\H,-2.3488737441,2.1305360593,-1.59258
25533\H,0.9821068001,1.4595876153,2.4607084755\O,1.0347082392,1.622651
089,1.495798574\H,1.9835537751,2.6964163469,0.7868987937\O,2.447022147
5,3.3401656554,0.171321217\C,3.7688450105,3.3078315692,0.2638549231\C,
4.380237873,2.6803372674,1.5018358074\O,4.4541397816,3.8343058054,-0.5
967679309\H,3.6075464496,2.1959842041,2.111776841\H,4.7711556636,3.533
2547932,2.0814545182\H,6.1535123807,2.1476619699,0.3821265813\H,6.1811
016481,1.643935078,2.0896125478\N,5.1579394592,0.356452397,0.819786723
4\H,4.5649466658,0.3910616644,-0.0250004377\C,4.4161998905,-0.36776519
33,1.8545241211\H,5.0487794604,-0.424559405,2.762538369\H,3.465782799,
0.1348323056,2.1343426502\C,4.0735133226,-1.7639616436,1.3418864552\H,

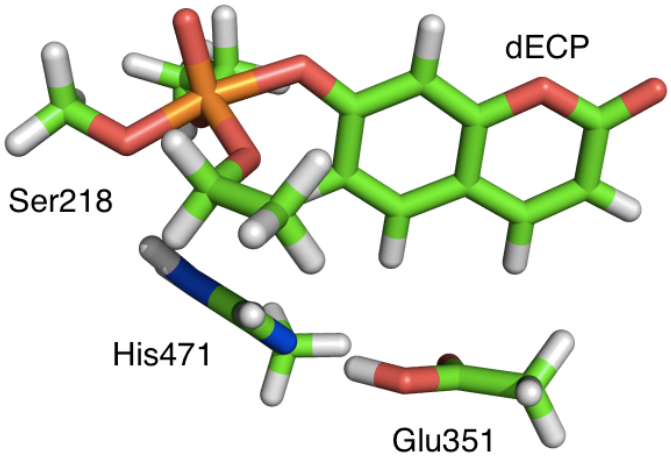
3.7018469545,-1.6294033973,0.3079554089\H,3.2336439408,-2.1832938552,1
.9295324578\N,5.1556592256,-2.7454223616,1.3230613598\H,6.018107464,-2
.2817855655,0.9931984103\H,6.2480927087,-4.0378659323,2.5622647531\H,5
.5646376609,-2.6480295738,3.4415449134\H,4.5072301075,-3.9684894576,2.
8928513709\\Version=ES64L-G09RevD.01\HF=-2198.8538041\RMSD=5.987e-
09\RMSF=1.559e-06\Dipole=4.1561994,1.5410512,1.6341621\Quadrupole=-
72.3455407,10.5418556,61.8036851,-48.0574575,8.221541,11.4018787\PG=C01
[X(C23H51N5O9P1)]\\@

dECP – Product

1\1\GINC-R2785\FOpt\RM062X\6-31+G(d,p)\C23H51N5O9P1(1-)\ROOT\22-Sep-20
14\0\\# m062X/6-31+G(d,p) 6D INT(grid=ultrafine) OPT(ReadFC,Tight) IOP
(2/17=4) Freq=noraman scrf=(pcm,read) geom=check\\title\\-1,1\C,8.6382
600142,-1.7894788536,-0.1781702288\C,5.7645380822,-0.3866241254,2.8063
64308\C,-4.7073790968,3.1620927213,-1.9766230655\C,-2.2440906791,3.518
749105,-2.0273991738\C,-5.5684471236,2.8541467081,2.8889936896\C,-6.00
7079034,-1.8436814669,0.5151885942\H,7.8018249142,-1.604627533,0.50489
4458\H,8.7958504777,-2.8713948372,-0.2346688552\H,9.536332167,-1.33857
30838,0.259582757\C,8.3411814158,-1.1889714517,-1.5513316504\C,8.12466
15759,0.3209976893,-1.4812870637\H,9.1726607488,-1.404060662,-2.236338
8414\H,7.442983624,-1.6577059,-1.9711343183\C,6.87939464,0.7419563802,
-0.6858077717\H,8.9985122158,0.8205106345,-1.0442859494\H,7.9948102138
,0.730845583,-2.4939879127\O,6.9433272131,1.832769696,-0.0495821276\O,
5.8621074432,-0.000285059,-0.7428267104\H,4.7298853279,1.6362940165,0.
6774848948\N,3.8559785944,1.5746963033,1.2480343123\C,3.5141000139,0.5
320128179,2.0732612842\C,2.8487402535,2.4789027083,1.2809374389\C,4.36
814059,-0.6852272656,2.2442768362\C,2.2843026967,0.8787061214,2.584462
3882\H,2.8741065038,3.3980369756,0.706077068\N,1.8716576999,2.09515464
99,2.0831591102\H,3.831952112,-1.3744110224,2.9095936354\H,4.476747274
3,-1.1848975248,1.2695876616\H,1.6627786184,0.3053227439,3.2644317263\
H,-0.3899325656,1.7585510289,0.8746861931\H,6.339843311,-1.3118811768,
2.9152046751\H,6.3154478911,0.2782759821,2.130773328\H,5.6944939761,0.
0985251959,3.7860096323\H,-0.0092300396,3.9740330131,0.6217261528\C,-0
.8937800522,3.5267989866,0.1445675575\H,-1.764006336,3.7560902682,0.78
30096141\O,-0.7334635869,2.1201768271,0.0280350968\C,-1.0760729525,4.1
172128548,-1.2483527817\P,-1.0890180107,-1.5544528767,0.7429288569\H,-
0.1467196777,3.9429469689,-1.8223176052\H,-1.2187455413,5.2102743165,-
1.1540157021\O,-2.2508286061,-1.3342880258,-0.3119804914\O,0.119970712
4,-2.0925301295,-0.1368914894\O,-0.7414662702,-0.4504118032,1.66912035
63\H,-2.0976860011,2.4234961303,-2.1175943556\H,-2.2534369879,3.929034
8033,-3.0489158719\N,-3.5396333626,3.8428765451,-1.4004246714\C,-2.012
8257472,-0.5087046395,-1.4885534602\C,1.4986866453,-1.6873267384,0.093
6713114\H,-3.5063837576,3.6022777159,-0.3961145322\H,-1.2384510375,0.2
326397093,-1.2580283318\C,-1.6633974964,-1.3678180974,-2.6850312328\H,
-2.9553255282,0.0280707917,-1.6443387274\C,1.8136150508,-0.4160804092,
-0.6625053526\H,2.0924089931,-2.5329199419,-0.2658558765\H,1.661031522
7,-1.5606456611,1.1700274927\H,-4.6128016836,2.0601810813,-1.979905680
5\H,-5.5972190048,3.4327824389,-1.4000742298\H,-4.8500141262,3.4896940
799,-3.0142017994\H,-2.4567088771,-2.0942337893,-2.8782162569\H,-0.721
2964707,-1.8970042892,-2.5206179232\H,-1.5533752023,-0.7213241836,-3.5
630101179\H,2.8891078838,-0.210123672,-0.6036929601\H,1.2761176269,0.4

366936649,-0.2323004896\H,1.5366016978,-0.5245006978,-1.7170363164\H,-1.4541082271,-2.847187807,2.4386155305\O,-1.6986917176,-2.8404876199,1.4917273595\H,-3.0528286449,-4.3228662726,-0.1464521351\O,-3.3438881754,-3.9969291603,-1.0302950613\C,-4.340178466,-3.1056051937,-0.9173169075\C,-4.9533092015,-2.9513677377,0.4552484975\O,-4.7297796231,-2.5269258739,-1.9100323656\H,-4.1460402801,-2.78453886,1.1818845085\H,-5.4110202776,-3.9249593955,0.7045559079\H,-6.7118954691,-1.9775569099,-0.3238389486\H,-6.5812398519,-1.9559051067,1.4530175086\N,-5.4960680971,-0.4702824047,0.4688957454\H,-4.9233254339,-0.3540431227,-0.3849470365\C,-4.7101591232,-0.0608807861,1.6360337071\H,-5.3628719766,-0.1300539284,2.5288384735\H,-3.8294587448,-0.71698645,1.8159781402\C,-4.1901214644,1.3731204286,1.4712297536\H,-3.6575113984,1.4299498665,0.49467054\H,-3.4225203571,1.570390851,2.2441223072\N,-5.1871736361,2.446494294,1.5328540979\H,-6.0245219004,2.1532732758,1.0042914692\H,-6.3505133353,3.6180162082,2.828045175\H,-5.9395728589,2.0286654528,3.5207055435\H,-4.6976338336,3.2967795914,3.3875013728\\Version=ES64L-G09RevD.01\State=1-A\HF=-2198.8830998\RMSD=2.600e-09\RMSF=1.876e-07\Dipole=-7.0310361,-3.8683418,1.914187\Quadrupole=-72.6281019,33.9352528,38.6928491,-8.7525014,-16.4778968,-19.0088472\PG=C01 [X(C23H51N5O9P1)]\\@

A.1.5 Unrealistic geometry observed in a QC calculation (TS2 optimisation for phosphorylation).



A.2 Parameters for Molecular Dynamics simulations

Where parameters had to be created for substrates or non-standard residues, they are presented here.

A.2.1 Parameters of phosphorylated serine residue SDX

!!index array str
"SDX"
!entry.SDX.unit.atoms table str name str type int typex int resx int flags int seq int elmnt dbl chg
"C5" "CT" 0 1 131072 1 6 -0.032818

"H5" "HC" 0 1 131072 2 1 0.023795
"H6" "HC" 0 1 131072 3 1 0.023795
"H7" "HC" 0 1 131072 4 1 0.023795
"C4" "CT" 0 1 131072 5 6 0.135916
"H3" "H1" 0 1 131072 6 1 0.069518
"H4" "H1" 0 1 131072 7 1 0.069518
"O3" "OS" 0 1 131072 8 8 -0.28803
"P1" "P" 0 1 131072 9 15 0.642566
"O4" "OS" 0 1 131072 10 8 -0.28803
"C6" "CT" 0 1 131072 11 6 0.135916
"C7" "CT" 0 1 131072 12 6 -0.032818
"H10" "HC" 0 1 131072 13 1 0.023795
"H11" "HC" 0 1 131072 14 1 0.023795
"H12" "HC" 0 1 131072 15 1 0.023795
"H8" "H1" 0 1 131072 16 1 0.069518
"H9" "H1" 0 1 131072 17 1 0.069518
"O5" "O2" 0 1 131072 18 8 -0.50353
"O1" "OS" 0 1 131072 19 8 -0.27620
"C2" "CT" 0 1 131072 20 6 -0.110738
"H1" "H1" 0 1 131072 21 1 0.191036
"H2" "H1" 0 1 131072 22 1 0.191036
"C1" "CT" 0 1 131072 23 6 0.036593
"C3" "C" 0 1 131072 24 6 0.516907
"O2" "O" 0 1 131072 25 8 -0.535811
"H14" "H1" 0 1 131072 32 1 0.136398
"N1" "N" 0 1 131072 33 7 -0.701044
"H13" "H" 0 1 131072 34 1 0.361809
!entry.SDX.unit.atomspertinfo table str pname str ptype int ptypex int pelmnt dbl pchg
"C5" "CT" 0 -1 0.0
"H5" "HC" 0 -1 0.0
"H6" "HC" 0 -1 0.0
"H7" "HC" 0 -1 0.0
"C4" "CT" 0 -1 0.0
"H3" "H1" 0 -1 0.0
"H4" "H1" 0 -1 0.0
"O3" "OS" 0 -1 0.0
"P1" "P" 0 -1 0.0
"O4" "OS" 0 -1 0.0
"C6" "CT" 0 -1 0.0
"C7" "CT" 0 -1 0.0
"H10" "HC" 0 -1 0.0
"H11" "HC" 0 -1 0.0
"H12" "HC" 0 -1 0.0
"H8" "H1" 0 -1 0.0
"H9" "H1" 0 -1 0.0
"O5" "O2" 0 -1 0.0
"O1" "OS" 0 -1 0.0
"C2" "CT" 0 -1 0.0
"H1" "H1" 0 -1 0.0
"H2" "H1" 0 -1 0.0

```
 "C1" "CT" 0 -1 0.0
 "C3" "C" 0 -1 0.0
 "O2" "O" 0 -1 0.0
 "N2" "N" 0 -1 0.0
 "H18" "H" 0 -1 0.0
 "H14" "H1" 0 -1 0.0
 "N1" "N" 0 -1 0.0
 "H13" "H" 0 -1 0.0
!entry.SDX.unit.boundbox array dbl
 -1.000000
 0.0
 0.0
 0.0
 0.0
!entry.SDX.unit.childsequence single int
 2
!entry.SDX.unit.connect array int
 27
 24
!entry.SDX.unit.connectivity table  int atom1x  int atom2x  int flags
 1 2 1
 1 3 1
 1 4 1
 1 5 1
 5 6 1
 5 7 1
 5 8 1
 8 9 1
 9 10 1
 9 18 1
 9 19 1
 10 11 1
 11 16 1
 11 17 1
 11 12 1
 12 13 1
 12 14 1
 12 15 1
 19 20 1
 20 21 1
 20 22 1
 20 23 1
 23 24 1
 23 26 1
 23 27 1
 24 25 1
 27 28 1
!entry.SDX.unit.hierarchy table  str abovetype  int abovex  str belowtype  int belowx
 "U" 0 "R" 1
 "R" 1 "A" 1
 "R" 1 "A" 2
```

"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11
"R" 1 "A" 12
"R" 1 "A" 13
"R" 1 "A" 14
"R" 1 "A" 15
"R" 1 "A" 16
"R" 1 "A" 17
"R" 1 "A" 18
"R" 1 "A" 19
"R" 1 "A" 20
"R" 1 "A" 21
"R" 1 "A" 22
"R" 1 "A" 23
"R" 1 "A" 24
"R" 1 "A" 25
"R" 1 "A" 26
"R" 1 "A" 27
"R" 1 "A" 28
!entry.SDX.unit.name single str
"SDX"
!entry.SDX.unit.positions table dbl x dbl y dbl z
3.536914 1.422858 1.933042E-06
2.539775 1.523975 0.412975
3.880707 0.411825 0.191948
3.484625 1.572795 -1.072304
4.498243 2.412850 0.630122
4.520411 2.316935 1.705626
5.498088 2.278774 0.247273
4.157656 3.763369 0.293094
3.077791 4.599563 1.061312
3.472058 6.089583 0.784272
4.451635 6.785346 1.565189
3.776180 7.635878 2.619965
3.079590 8.328858 2.159785
4.521889 8.209132 3.162821
3.237558 7.015086 3.323689
5.002421 7.389304 0.859307
5.133817 6.075791 2.012091
2.917445 4.241180 2.468569
1.774007 4.407101 0.206241
0.545662 5.067855 0.539297
0.602859 5.508063 1.520559
0.395239 5.849288 -0.186419

-0.583572 4.041607 0.427289
-0.854950 3.143883 1.652563
-1.942561 2.636045 1.758521
-0.293244 3.366953 -0.369961
-1.822428 4.661129 -0.001581
-2.443522 4.028697 -0.455769
!entry.SDX.unit.residueconnect table int c1x int c2x int c3x int c4x int c5x int c6x
27 24 0 0 0 0
!entry.SDX.unit.residues table str name int seq int childseq int startatomx str restype int imagingx
"SDX" 1 29 1 "p" 0
!entry.SDX.unit.residuesPdbSequenceNumber array int
0
!entry.SDX.unit.solventcap array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.SDX.unit.velocities table dbl x dbl y dbl z
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0

0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0

A.2.2 Parameters of dECP

!!index array str
"GRP"
!entry.GRP.unit.atoms table str name str type int typex int resx int flags int seq int elmnt dbl chg
"C5" "ca" 0 1 131072 1 6 -0.099798
"H3" "ha" 0 1 131072 2 1 0.159149
"C2" "ca" 0 1 131072 3 6 -0.314998
"H1" "ha" 0 1 131072 4 1 0.201366
"C1" "ca" 0 1 131072 5 6 0.030430
"C3" "cc" 0 1 131072 6 6 -0.071871
"H2" "ha" 0 1 131072 7 1 0.164068
"C6" "cd" 0 1 131072 8 6 -0.396178
"H4" "ha" 0 1 131072 9 1 0.177932
"C9" "c" 0 1 131072 10 6 0.856509
"O3" "o" 0 1 131072 11 8 -0.587763
"O2" "os" 0 1 131072 12 8 -0.403380
"C4" "ca" 0 1 131072 13 6 0.343268
"C8" "ca" 0 1 131072 14 6 -0.320371
"H5" "ha" 0 1 131072 15 1 0.198170
"C7" "ca" 0 1 131072 16 6 0.237415
"O1" "os" 0 1 131072 17 8 -0.346220
"P1" "p5" 0 1 131072 18 15 0.989433
"O4" "o" 0 1 131072 19 8 -0.629339
"O6" "os" 0 1 131072 20 8 -0.398947
"C11" "c3" 0 1 131072 21 6 0.201237
"C13" "c3" 0 1 131072 22 6 -0.067546
"H13" "hc" 0 1 131072 23 1 0.031033
"H14" "hc" 0 1 131072 24 1 0.031033
"H15" "hc" 0 1 131072 25 1 0.031033
"H8" "h1" 0 1 131072 26 1 0.039123
"H9" "h1" 0 1 131072 27 1 0.039123
"O5" "os" 0 1 131072 28 8 -0.398947
"C10" "c3" 0 1 131072 29 6 0.201237
"H6" "h1" 0 1 131072 30 1 0.039123
"H7" "h1" 0 1 131072 31 1 0.039123
"C12" "c3" 0 1 131072 32 6 -0.067546
"H10" "hc" 0 1 131072 33 1 0.031033
"H11" "hc" 0 1 131072 34 1 0.031033
"H12" "hc" 0 1 131072 35 1 0.031033

```
!entry.GRP.unit.atomspertinfo table  str pname  str ptype  int ptypex  int pelmnt  dbl
pchg
 "C5" "ca" 0 -1 0.0
 "H3" "ha" 0 -1 0.0
 "C2" "ca" 0 -1 0.0
 "H1" "ha" 0 -1 0.0
 "C1" "ca" 0 -1 0.0
 "C3" "cc" 0 -1 0.0
 "H2" "ha" 0 -1 0.0
 "C6" "cd" 0 -1 0.0
 "H4" "ha" 0 -1 0.0
 "C9" "c" 0 -1 0.0
 "O3" "o" 0 -1 0.0
 "O2" "os" 0 -1 0.0
 "C4" "ca" 0 -1 0.0
 "C8" "ca" 0 -1 0.0
 "H5" "ha" 0 -1 0.0
 "C7" "ca" 0 -1 0.0
 "O1" "os" 0 -1 0.0
 "P1" "p5" 0 -1 0.0
 "O4" "o" 0 -1 0.0
 "O6" "os" 0 -1 0.0
 "C11" "c3" 0 -1 0.0
 "C13" "c3" 0 -1 0.0
 "H13" "hc" 0 -1 0.0
 "H14" "hc" 0 -1 0.0
 "H15" "hc" 0 -1 0.0
 "H8" "h1" 0 -1 0.0
 "H9" "h1" 0 -1 0.0
 "O5" "os" 0 -1 0.0
 "C10" "c3" 0 -1 0.0
 "H6" "h1" 0 -1 0.0
 "H7" "h1" 0 -1 0.0
 "C12" "c3" 0 -1 0.0
 "H10" "hc" 0 -1 0.0
 "H11" "hc" 0 -1 0.0
 "H12" "hc" 0 -1 0.0
!entry.GRP.unit.boundbox array dbl
 -1.000000
 0.0
 0.0
 0.0
 0.0
!entry.GRP.unit.childsequence single int
 2
!entry.GRP.unit.connect array int
 1
 32
!entry.GRP.unit.connectivity table  int atom1x  int atom2x  int flags
 1 2 1
 1 3 1
```

1 16 1
3 4 1
3 5 1
5 6 1
5 13 1
6 7 1
6 8 1
8 9 1
8 10 1
10 11 1
10 12 1
12 13 1
13 14 1
14 15 1
14 16 1
16 17 1
17 18 1
18 19 1
18 20 1
18 28 1
20 21 1
21 22 1
21 26 1
21 27 1
22 23 1
22 24 1
22 25 1
28 29 1
29 30 1
29 31 1
29 32 1
32 33 1
32 34 1
32 35 1
!entry.GRP.unit.hierarchy table str abovetype int abovex str belowtype int belowx
"U" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11
"R" 1 "A" 12
"R" 1 "A" 13
"R" 1 "A" 14
"R" 1 "A" 15

"R" 1 "A" 16
"R" 1 "A" 17
"R" 1 "A" 18
"R" 1 "A" 19
"R" 1 "A" 20
"R" 1 "A" 21
"R" 1 "A" 22
"R" 1 "A" 23
"R" 1 "A" 24
"R" 1 "A" 25
"R" 1 "A" 26
"R" 1 "A" 27
"R" 1 "A" 28
"R" 1 "A" 29
"R" 1 "A" 30
"R" 1 "A" 31
"R" 1 "A" 32
"R" 1 "A" 33
"R" 1 "A" 34
"R" 1 "A" 35
!entry.GRP.unit.name single str
"GRP"
!entry.GRP.unit.positions table dbl x dbl y dbl z
3.536914 1.422858 1.933042E-06
2.733899 0.967421 0.544913
4.833271 1.389294 0.463101
5.053218 0.887704 1.388118
5.870981 1.990153 -0.247787
7.260273 1.992772 0.174307
7.522083 1.501468 1.093926
8.190056 2.591442 -0.564610
9.226603 2.620380 -0.296701
7.839786 3.263807 -1.821693
8.607703 3.817193 -2.529642
6.532156 3.234356 -2.175619
5.570187 2.632512 -1.440566
4.275891 2.687090 -1.927674
4.058060 3.187494 -2.850349
3.276698 2.081050 -1.197460
2.014232 2.101683 -1.731685
0.721353 2.680186 -1.004869
0.427480 2.096822 0.290787
-0.374316 2.473092 -2.102979
-1.128078 1.257617 -2.219970
-2.479723 1.413146 -1.555844
-3.067791 0.513413 -1.703767
-2.364051 1.578407 -0.491796
-3.019426 2.250351 -1.985997
-0.570481 0.439561 -1.785883
-1.222691 1.083460 -3.281629
0.996474 4.220656 -1.026901

```
 0.209971 5.141940 -0.263177
 0.175579 4.799950 0.761724
 -0.794622 5.147056 -0.665035
 0.854337 6.504918 -0.378184
 0.892770 6.824015 -1.412393
 1.864341 6.482801 0.014828
 0.279067 7.231733 0.187678
!entry.GRP.unit.residueconnect table  int c1x  int c2x  int c3x  int c4x  int c5x  int c6x
 1 32 0 0 0 0
!entry.GRP.unit.residues table  str name  int seq  int childseq  int startatomx  str restype
int imagingx
 "GRP" 1 36 1 "?" 0
!entry.GRP.unit.residuesPdbSequenceNumber array int
 0
!entry.GRP.unit.solventcap array dbl
 -1.000000
 0.0
 0.0
 0.0
 0.0
!entry.GRP.unit.velocities table  dbl x  dbl y  dbl z
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0
```

0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0
0.0 0.0 0.0

A.2.3 Parameters of VX(S)

!!index array str
"VXX"
!entry.VXX.unit.atoms table str name str type int typex int resx int flags int seq int elmnt dbl chg
"C1" "c3" 0 1 131072 1 6 -0.022495
"H1" "hc" 0 1 131072 2 1 0.018995
"H2" "hc" 0 1 131072 3 1 0.018995
"H3" "hc" 0 1 131072 4 1 0.018995
"C8" "c3" 0 1 131072 5 6 0.149586
"H21" "h1" 0 1 131072 6 1 0.032623
"H22" "h1" 0 1 131072 7 1 0.032623
"O2" "os" 0 1 131072 8 8 -0.413668
"P1" "p5" 0 1 131072 9 15 1.023648
"C7" "c3" 0 1 131072 10 6 -0.447512
"H18" "hc" 0 1 131072 11 1 0.125590
"H19" "hc" 0 1 131072 12 1 0.125590
"H20" "hc" 0 1 131072 13 1 0.125590
"O1" "o" 0 1 131072 14 8 -0.641996
"S1" "ss" 0 1 131072 15 16 -0.403908
"C6" "c3" 0 1 131072 16 6 0.243571
"H16" "h1" 0 1 131072 17 1 0.003874
"H17" "h1" 0 1 131072 18 1 0.003874
"C11" "c3" 0 1 131072 19 6 -0.264202
"H25" "h1" 0 1 131072 20 1 0.113373
"H26" "h1" 0 1 131072 21 1 0.113373
"N1" "n3" 0 1 131072 22 7 -0.022763
"C10" "c3" 0 1 131072 23 6 0.048323
"C4" "c3" 0 1 131072 24 6 -0.180104
"H10" "hc" 0 1 131072 25 1 0.044455
"H11" "hc" 0 1 131072 26 1 0.044455
"H12" "hc" 0 1 131072 27 1 0.044455
"C5" "c3" 0 1 131072 28 6 -0.180104
"H13" "hc" 0 1 131072 29 1 0.044455
"H14" "hc" 0 1 131072 30 1 0.044455
"H15" "hc" 0 1 131072 31 1 0.044455
"H24" "h1" 0 1 131072 32 1 0.078277
"C9" "c3" 0 1 131072 33 6 0.048323
"C3" "c3" 0 1 131072 34 6 -0.180104
"H7" "hc" 0 1 131072 35 1 0.044455
"H8" "hc" 0 1 131072 36 1 0.044455
"H9" "hc" 0 1 131072 37 1 0.044455
"H23" "h1" 0 1 131072 38 1 0.078277
"C2" "c3" 0 1 131072 39 6 -0.180104

"H4" "hc" 0 1 131072 40 1 0.044455
"H5" "hc" 0 1 131072 41 1 0.044455
"H6" "hc" 0 1 131072 42 1 0.044455
!entry.VXX.unit.atomspertinfo table str pname str ptype int ptypex int pelmnt dbl pchg
"C1" "c3" 0 -1 0.0
"H1" "hc" 0 -1 0.0
"H2" "hc" 0 -1 0.0
"H3" "hc" 0 -1 0.0
"C8" "c3" 0 -1 0.0
"H21" "h1" 0 -1 0.0
"H22" "h1" 0 -1 0.0
"O2" "os" 0 -1 0.0
"P1" "p5" 0 -1 0.0
"C7" "c3" 0 -1 0.0
"H18" "hc" 0 -1 0.0
"H19" "hc" 0 -1 0.0
"H20" "hc" 0 -1 0.0
"O1" "o" 0 -1 0.0
"S1" "ss" 0 -1 0.0
"C6" "c3" 0 -1 0.0
"H16" "h1" 0 -1 0.0
"H17" "h1" 0 -1 0.0
"C11" "c3" 0 -1 0.0
"H25" "h1" 0 -1 0.0
"H26" "h1" 0 -1 0.0
"N1" "n3" 0 -1 0.0
"C10" "c3" 0 -1 0.0
"C4" "c3" 0 -1 0.0
"H10" "hc" 0 -1 0.0
"H11" "hc" 0 -1 0.0
"H12" "hc" 0 -1 0.0
"C5" "c3" 0 -1 0.0
"H13" "hc" 0 -1 0.0
"H14" "hc" 0 -1 0.0
"H15" "hc" 0 -1 0.0
"H24" "h1" 0 -1 0.0
"C9" "c3" 0 -1 0.0
"C3" "c3" 0 -1 0.0
"H7" "hc" 0 -1 0.0
"H8" "hc" 0 -1 0.0
"H9" "hc" 0 -1 0.0
"H23" "h1" 0 -1 0.0
"C2" "c3" 0 -1 0.0
"H4" "hc" 0 -1 0.0
"H5" "hc" 0 -1 0.0
"H6" "hc" 0 -1 0.0
!entry.VXX.unit.boundbox array dbl
-1.000000
0.0
0.0

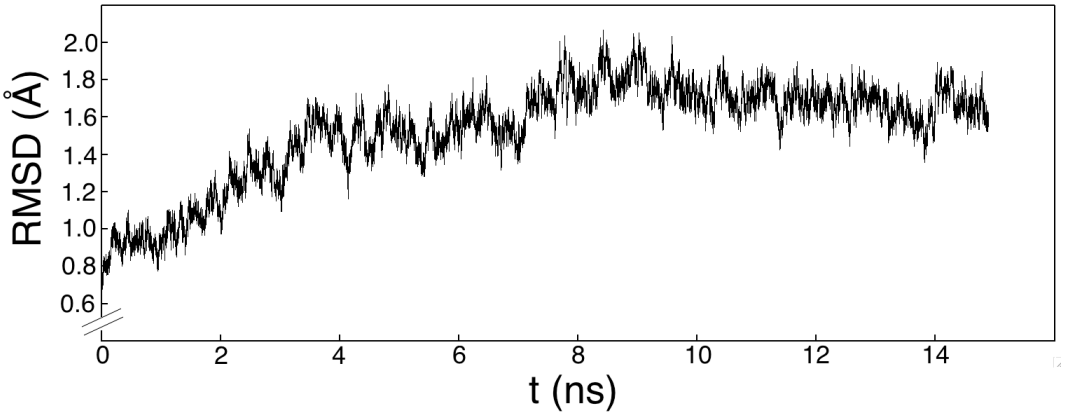
```
 0.0
 0.0
!entry.VXX.unit.childsequence single int
 2
!entry.VXX.unit.connect array int
 1
 39
!entry.VXX.unit.connectivity table  int atom1x  int atom2x  int flags
 1 2 1
 1 3 1
 1 4 1
 1 5 1
 5 6 1
 5 7 1
 5 8 1
 8 9 1
 9 10 1
 9 14 1
 9 15 1
 10 11 1
 10 12 1
 10 13 1
 15 16 1
 16 17 1
 16 18 1
 16 19 1
 19 20 1
 19 21 1
 19 22 1
 22 23 1
 22 33 1
 23 24 1
 23 28 1
 23 32 1
 24 25 1
 24 26 1
 24 27 1
 28 29 1
 28 30 1
 28 31 1
 33 34 1
 33 38 1
 33 39 1
 34 35 1
 34 36 1
 34 37 1
 39 40 1
 39 41 1
 39 42 1
!entry.VXX.unit.hierarchy table  str abovetype  int abovex  str belowtype  int belowx
 "U" 0 "R" 1
```

"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11
"R" 1 "A" 12
"R" 1 "A" 13
"R" 1 "A" 14
"R" 1 "A" 15
"R" 1 "A" 16
"R" 1 "A" 17
"R" 1 "A" 18
"R" 1 "A" 19
"R" 1 "A" 20
"R" 1 "A" 21
"R" 1 "A" 22
"R" 1 "A" 23
"R" 1 "A" 24
"R" 1 "A" 25
"R" 1 "A" 26
"R" 1 "A" 27
"R" 1 "A" 28
"R" 1 "A" 29
"R" 1 "A" 30
"R" 1 "A" 31
"R" 1 "A" 32
"R" 1 "A" 33
"R" 1 "A" 34
"R" 1 "A" 35
"R" 1 "A" 36
"R" 1 "A" 37
"R" 1 "A" 38
"R" 1 "A" 39
"R" 1 "A" 40
"R" 1 "A" 41
"R" 1 "A" 42
!entry.VXX.unit.name single str
"VXX"
!entry.VXX.unit.positions table dbl x dbl y dbl z
3.540003 1.419779 1.933042E-06
3.011051 1.293803 0.938918
4.180285 2.290093 0.074034
2.809035 1.589793 -0.784967
4.356890 0.183878 -0.316970
4.898892 0.301866 -1.245967

```
 3.731972 -0.692948 -0.400953
 5.271137 -0.096224 0.746325
 6.790556 0.370863 0.710330
 7.258459 0.041897 2.417037
 7.023462 -0.979988 2.687992
 8.319407 0.210861 2.537040
 6.709528 0.717841 3.060980
 7.026627 1.734001 0.228302
 7.702675 -1.067355 -0.502868
 8.870927 -0.073187 -1.508042
 8.521986 0.946726 -1.498036
 8.757960 -0.451043 -2.516681
 10.339881 -0.205223 -1.101053
 10.929657 0.161614 -1.943703
 10.560797 -1.259991 -1.016079
 10.695791 0.451541 0.140555
 11.760486 -0.211311 0.907336
 11.187458 -1.352308 1.750345
 10.431488 -0.981294 2.431335
 11.970196 -1.826185 2.337142
 10.729606 -2.116997 1.131598
 12.967257 -0.714241 0.093479
 13.373180 0.050576 -0.555371
 13.756404 -1.025330 0.771615
 12.716222 -1.574259 -0.518547
 12.140590 0.529847 1.599503
 10.770823 1.915282 0.041592
 10.453853 2.588274 1.378570
 11.191161 2.361150 2.142892
 10.437890 3.666164 1.255594
 9.479228 2.281426 1.731431
 9.971868 2.204308 -0.628379
 12.073994 2.476308 -0.549478
 12.919067 2.329268 0.116575
 11.972041 3.545161 -0.712438
 12.315930 2.023404 -1.505286
!entry.VXX.unit.residueconnect table  int c1x  int c2x  int c3x  int c4x  int c5x  int c6x
 1 39 0 0 0 0
!entry.VXX.unit.residues table  str name  int seq  int childseq  int startatomx  str restype
int imagingx
 "VXX" 1 43 1 "?" 0
!entry.VXX.unit.residuesPdbSequenceNumber array int
 0
!entry.VXX.unit.solventcap array dbl
 -1.000000
 0.0
 0.0
 0.0
 0.0
!entry.VXX.unit.velocities table  dbl x  dbl y  dbl z
 0.0 0.0 0.0
```

0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0
0.0	0.0	0.0

A.2.4. RMSD of E3 in complex with VX during MD simulation (Chapter 4).


RMSD (Å)
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0
2
4
6
8
10
12
14
t (ns)