Theoretical Characterization of Pentacovalent Oxyphosphorane Intermediate Structures of the Hydrolysis of RNA Catalyzed by RNase A

eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

Elena Formoso Estensoro

2010

Theoretical Characterization of Pentacovalent Oxyphosphorane Intermediate Structures of the Hydrolysis of RNA Catalyzed by RNase A

Elena Formoso Estensoro

(elena.formoso@ehu.eus)

Doctoral Dissertation

Supervisor: X. Lopez Pestaña

2010

- ISBN: 978-84-9082-454-2

Mila esker guztioi!

Azkenean iritsi da eguna. Bazen garaia, ezta? Hainbeste urte pasa dira tesi honekin hasi nintzenetik, iada ez naizela zehazki gogoratzen. Hala ere, inguruan izan deten jende guztiari eskerrak eman nahi nizkioke. Denak gogoan zaituztet.

Comienzo agradeciendo a mi familia el apoyo recibido durante todos estos años. Ama, ya acabo. No se te ocurra preguntarme, "¿y ahora qué?".

Ezagutzen nauzuenak burugogorra naizela dakizue, eta hori hemen agerian geratu da. Karrerako 3. mailan erabaki nuen Xabikin tesia egin nahi nuela, nahiz eta ez jakin ze demontre suposatzen zuen horrek. Beraz Xabi, mila esker nigan jarritako konfidantzagatik eta azkeneko urteetan erakutsi didazun guztiagatik. Badakit ez dela erraza izan nire tesiko zuzendaria izatea, baina azkenean lortu dugu. Zorionak! Taldearekin jarraituz, Jesus, eskerrik asko taldean sartzeko aukera eman zenidalako, eta behar izan dudanean hor egon zarelako laguntzeko prest. Ezin det ahaztu ere, zure "erruz" orain ardo beltza gustokoa dudala. Aitonak 20 urtetan lortu ez zuena, zuk lortu duzu! Hurrengo aipamen berezia Txemarentzako da (el crack informático), aspaldi ikasi genuen zu gabe mintegia desastre hutsa izango zela. Mila esker zure denbora eta pazientziagatik. Azkenik, talde osoari, nahiz eta batzuk iada alde egin izana, Mario, Joni, Txoni, Eider, Eli, Iñaki, Julen, Elisa, Oier, Jon Mikel, Ivan eta azken inkorporazioak Andreas eta Slawek, eskerrak zueri ere. Zuen ondoan lan egitea plazer bat izan da.

Denek dakigunez, lan honek bidaiatzeko "aukera" eskeintzen digu, lanean aberasgarria izateaz gain, lagunak eta leku berriak ezagutzeko aukera eskeiniz. Beraz, eskerrak nire Minneapoliseko eta Bristoleko egonaldietan inguruan izan ditudan pertsona guztiei. Thanks to all the people I met when I visited Darrin M. York's group in Minneapolis and Adrian Mulholland's group in Bristol for their hospitality and friendship. Hope to see you soon.

Eskerrik asko unibertsitateko lagunei ere. Azkenik, ezin ditut

ahaztu tesia aurretik inguruan nituen pertsonak, nire bizitzan aspalditik daudenak, beti hor daudenak: Ainara, Maider B., Oihana, Martineli, Maider A., eta Bego. Ezta azkeneko urteetan agertu diranak ere: Amaia, Txetxu, Luis, Eli, David, Laura .... ezinezkoa da denen izenak jartzea. Muchas gracias por soportarme y sobre todo por hacerme comprender que cada uno ve la vida desde un punto de vista diferente.
Como se suele decir: Quién bien te quiere, te hará llorar.

Acknowledgments

This research was funded by the Spanish Ministerio de Ciencia e Innovación (grant BES2005-6803).

Technical and human support provided by IZO-SGI SGIker (UPV/EHU, MICINN, GV/EJ, ESF) and the Minnesota Supercomputing Institute for Advanced Computational Research (MSI) are gratefully acknowledged.

Contents

ABBREVIATIONS

3',5'-CpA = cytidilyl-3',5'-adenosine
3'-Cp = cytidine 3'-phosphate
A = Adenosine
ABNR = Adopted-Basis set Newton-Raphson
Ade = Adenine
AM1 = Austen Model 1
AMBER = Assisted Model Building with Energy Refinement
Arg = Arginine
Asn = Asparagine
Asp = Aspartic Acid
C = Cytidine
CASSCF = Complete Active Space Self Consistent Field
CC = Coupled-Cluster
CHARMM = Chemistry at HARvard Macromolecular Mechanics
CI = Configuration Interaction
CMAP = Correction Map
CNDO = Complete Neglect of Differential Overlap
Cyt = Cytosine
C >p = cytidine 2',3'-cyclic phosphate
DFT = Density Functional Theory
DNA = Deoxyribonucleic Acid
ECP = Eosinophil-Cationic Protein
EDN = Eosinophil-Derived Neurotoxin
G = Guanosine
GGA = Generalized Gradient Approximation
GHO = Generalized Hybrid Orbital
Gln = Glutamine
Glu = Glutamic Acid
GROMOS = GROningen MOlecular Simulation
HF = Hartree-Fock
His = Histidine
HSP = Doubly Protonated state of histidine
HSD = His protonated at Nδ_1 position
[Hyd]^{oph} = Hydrolysis' Oxyphosphorane Intermediate
[Hyd]^{TS} = Hydrolysis' Transition State Mimic
INDO = Intemediate Neglect of Differential Overlap
LJ = Lennard-Jones
LSCF = Local Self-Consisten Field
Lys = Lysine
MD = Molecular Dynamics
MM = Molecular Mechanics
MMFF = Merck Molecular Force Field

MM-PBSA = Molecular Mechanics-Poisson Boltzmann Surface Area
MNDO = Modified Neglect of Diatomic Overlap
MP = Møller-Plesset
MRCI = Multi-Referential Configuration Interaction
NDDO = Neglect of Differential Diatomic Overlap
NMR = Nuclear Magnetic Resonance
NOE = Nuclear Overhauser Effects
$[\text{native}]^{\text{Hyd}}$ = Hydrolysis' Transition State oxyphosphorane mimic
$[\text{native}]^{\text{Trans}}$ = Transphosphorylation's Transition State oxyphosphorane mimic
Occ = Occupancy
OPLS = Optimized Potentials for Liquid Simulations
PBC = Periodic Boundary Conditions
PDB = Protein Data Bank
PES = Potential Energy Surface
Phe = Phenilalanine
PME = Particle Mesh Ewald
PMF = Potential of Mean Force
PM3 = Parametric Model 3
QM = Quantum Mechanics
RC = Reaction Coordinate
RDF = Radial Distribution Function
RESD = REStrained Distances
RMS = Root Mean Squared
RMSD = Root Mean Squared Deviation
RNA = Ribonucleic Acid
RNase = Ribonuclease
RNase A = Bovine Pancreatic Ribonuclease A
SBC = Stochastic Boundary Conditions
SD = Steepest Descent
STO = Slater Type Orbital
$[\text{S:O}_{1\text{P}}]^{\text{Hyd}}$ = Monothio substitution of $[\text{native}]^{\text{Hyd}}$ at position O_{1P}
$[\text{S:O}_{2\text{P}}]^{\text{Hyd}}$ = Monothio substitution of $[\text{native}]^{\text{Hyd}}$ at position O_{2P}
$[\text{S:O}_{1\text{P}},\text{O}_{2\text{P}}]^{\text{Hyd}}$ = Dithio substitution of $[\text{native}]^{\text{Hyd}}$ at positions O_{1P} and O_{2P}
$[\text{S:O}_{1\text{P}}]^{\text{Trans}}$ = Monothio substitution of $[\text{native}]^{\text{Trans}}$ at position O_{1P}
$[\text{S:O}_{2\text{P}}]^{\text{Trans}}$ = Monothio substitution of $[\text{native}]^{\text{Trans}}$ at position O_{2P}
$[\text{S:O}_{1\text{P}},\text{O}_{2\text{P}}]^{\text{Trans}}$ = Dithio substitution of $[\text{native}]^{\text{Trans}}$ at positions O_{1P} and O_{2P}
TBP = Trigonal bipyramidal
Thr = Threonine
$[\text{Trans}]^{oph}$ = Transphosphorylation's Oxyphosphorane Intermediate
$[\text{Trans}]^{TS}$ = Transphosphorylation's Transition State Mimic
U = Uridine
UB = Urey-Bradley
WHAM = Weighted Histogram Analysis Method

ZDO = Zero Differential Overlap

Part I

General Introduction
Sarrera Orokorra

Chapter 1

Erribonukleasak

1.1 Sarrera

Gure planetan Darwin-en eboluzio teoria bermatzeko, DNAk eta RNAk hirugarren biomolekula mota batekin lan egiten dute, proteinak. Elkar lan honek, izugarrizko bizitza mota barietatea sortu du, gaur egun planetako ia zirrikitu guztietan aurkitu daitekeena. DNAk informazio biologikoa biltzen duen bitartean, proteinak agerian jartzen du. Informazioa biltegitik agertzeko erara eramaten duen bidea, RNA da:[1]

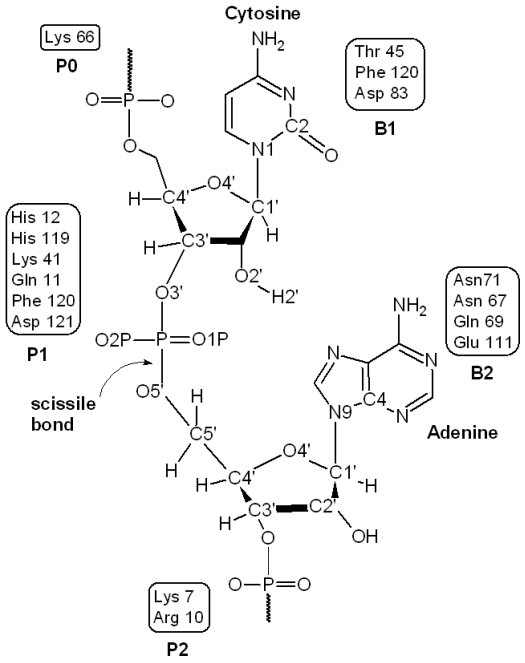
$$\text{DNA} \rightleftharpoons \text{RNA} \rightarrow \text{proteina}$$

Bizitza ezagutzen dugun bezala izateko, funtsezkoa da informazio fluxua RNAren bidez ematea. Bi entzima mota fluxu hori kontrolatzeaz arduratzen dira, RNAren sintesia edo degradazioa katalizatuz. Sintesia RNA polimerasak katalizatzen duen bitartean, degradazioa RNA depolimerasak katalizatzen du. Azken honi, normalki "Erribonukleasa" deitzen zaio (RNasa laburtuta). Erribonukleasa guztiak ondorio zitotoxikoak dituzte. Lehenik erribonukleasak zelula jakin batzuetan adsorbatzen dira, ondoren zelulen zitosolan barneratzen dira, eta azkenik erribonukleasek RNA degradatzen dute. Ondorio bezala, erribonukleasek proteinen sintesia inhibitzen dute zelulen heriotza eraginez.[2] Gainera, erribonukleasak minbizi tratamendu terapeutikotan erabil daitezke. RNasak RNAri sortzen dioten kaltea, DNA alterazio kimioterapia arrunten alternatiba bat izan daiteke.[3,4]

Erribonukleasak exoerribonukleasa eta endoerribonukleasa motetan banatzen dira. Exoerribonukleasek,[5] RNA degradatzen dute nukleotido terminalak RNAren 5' edo 3' bukaeratik eliminatuz. Bestalde, endoerribonukleasek apurketa internoak sortzen dituzte. Azken mota honetako erribonukleasetatik gehien ikertutako familia *A erribonukleasa pankreatikoa* da.[6]

1.2 A Erribonukleasa Pankreatiko Familia

Izugarri dibertsifikatu den entzima superfamilia interesgarri honen ekintzak ornodunen eboluzioarekin oso erlazionatuta daude.[6–11] A RNasa familiako proteinek RNA polimero kateak deskonposatzen dituzte. Famili honetako kide askok, beraien jarduera entzimatikoaz dependatzen duten aktibitate biologiko oso indartsuak dituzte. Adibidez, gizaki angiogenina[12] (RNasa 5), gaixotasun baskularretan eta erreumatoidetan, baita minbizian erlazionatua dago;[13–16] Neurotoxina eosinofil-deribatua (EDN, RNasa 2) eta proteina eosinofil-kationikoa (ECP, RNasa 3) *in vivo* neurotoxikoak dira eta alergia eta sindrome hipereosinofilikoekin erlazionatuta daude;[17,18] Abelgorri RNasa pankreatiko seminalak, jarduera antiespermatogenikoa eta inmunosupresiboa du;[19,20] azkenik, Onconasak aktibitate antitumoral oso indartsuak ditu.[21–23]



Scheme 1.1: A RNasaren gune aktiboaren eskema RNA ligando bati lotuta. RNAren baseekin, B_i, eta fosfato taldeekin, P_i, interakzionatzen duten azpiguneak adierazita daude.

RNAren depolimerazio prozesua gertatzeko, entzima eta substratu polimerikoaren gune aktiboko interakzioaz gain, entzimaren azpigune multzoen loturen bitartez substratuaren kokapen egokia eman behar da (begiratu 1.1 Eskema). Ezarritako nomenklaturarekin bat eginik,[24] gune aktiboa B_i eta P_i azpigunetan banatzen da. Hauek RNAren baseekin eta fosfato taldeekin elkartuko dira hurrenez hurren.[25] B_1 azpiguneak pirimidina baseekiko es-

pezifikotasuna du. Zitidinarekiko duen lehentasun zinetikoa uridinarekiko duena baino 30 aldiz handiagoa da.[26] Espezifikotasun hori Thr45 erresiduoaren eskutik dator abelgorri A erribonukleasa pankreatikoan. Erresiduo hau RNasa guztietan agertzen da.[27] Proteinek, P-$O_{5'}$ (scissile) lotura hausten dute B_1-ari lotutako base pirimidikoen 3' aldean. P_1 azpigunean erreakzioan paper garrantzitsua duten eta RNasa homologo guztietan mantentzen diren bi histidina kataliko eta lisina bakarra daude (His12, His119 eta Lys41 erresiduoak abelgorri A erribonukleasa pankreatikoan).[24,28,29] B_2 azpigunea RNasa guztietan mantentzeaz gain, base guztiak antzematen ditu, adeninarekiko lehentasuna ageriz.[30] Azpigune honek lotura poltsatxo antzeko bat eskeintzen dio askatzen den nukleotidoari. Bestalde, P_2k substratu-entzima konplexuaren sorkuntzan parte hartzeaz gain, mekanismo katalitikoan zeharkako papera du.[31] Badirudi gune honek rol berezia duela A RNasaren aktibitate endonukleolitikoan (kate barneko apurketa).[32–34] Gune honetan, interakzio elektrostatikoaren gabeziak substratuaren haustura exonukleolitikoa izatera bultzatzen du. Beste gune batzuk (adibidez P_0 eta B_3) aldakorragoak dira gisakoen artean.

1.3 Abelgorri A Erribonukleasa Pankreatikoa

Abelgorri A erribonukleasa pankreatikoa (A RNasa) [EC 3.1.27.5], abelgorri pankreak sortzen duen entzima nagusia izateaz gain, gaur egun A erribonukleasa supermafiliaren arketipoa dela onartzen da. 60 urte baino gehiagoz ikertua izan da,[24,28,35] A RNasaren garbiketa emanez, aminoazido sekuentzia osoa argituz, egitura kristalinoa lortuz, proteina plegapen bideak argituz eta mekanismo katalikoan parte hartzen duten elementuen identifikazioa lortuz.[36–42] Entzima honi buruz egin zituzten lanengatik,[43–45] 1972an hiru ikertzailek eta laugarren batek 1984an kimikako Nobel Sariak jaso zituzten (1.1 Taula). Stanford Moore eta William H. Steinek lehengo aldiz proteina baten aminoazido konposizioa identifikatu zuten, A RNasa, eta urte batzuk beranduago honen aminoazido sekuentzia lortu zuten. Berriro ere, lehenengo aldia izan zen entzima baterako eta hirugarren aldia proteinentzat.[44] Saria Christian B. Anfisenekin banatua izan zen, proteinen estruktura tridimentsionala aminoazidoen oinarrian datzala aurkitu baitzuen.[43] 1984ko kimikako Nobel Sarituak, Bruce Merrieldek, entzimaren sintesi kimiko osoa lortu zuen.[45] Beraz, endoerribonukleasa hau katalisi entzimatikoaren oinarri estrukturalak eta dinamikoak maila teorikoan ikertzeko oso aproposa da.

1.3.1 Egitura

A RNasa 13.7kDa masa molekularreko eta 124 aminoazidoz osatutako kate polipeptidiko bakuna da. 20 amino azido naturaletatik, triptofanoa bakarrik du faltan. β-lamina tolestura zentral bati elkartutako helize motxez eta N-terminal helize luze batez osatuta dagoen giltzurrun formako molekula da

Nobel Saritua	Urtea	Nobel Lana
Christian B. Anfinsen (1916-1995)	1972	Studies on the Principles that Govern the Folding of Protein Chains [43]
Stanford Moore (1913-1982)	1972	The Chemical Structures of Pancreatic Ribonuclease and Deoxyribonuclease [44]
William H. Stein (1911-1980)	1972	The Chemical Structures of Pancreatic Ribonuclease and Deoxyribonuclease [44]
Bruce Merrifield (1921-2006)	1984	Solid Phase Synthesis [45]

Table 1.1: A RNasaren gainean lan egindako kimika Nobel Sarituak.

(1.1 Irudia). Estruktura orokorra zortzi zisteina erresiduek sortzen duten lau disulfuro zubien bitartez egonkortua egoteaz gain, loop gehinak disolbatzailearekin interakzionatzen dute.[46] Lotura gunea (gune aktiboa) koba forma ematen dioten lau β-lamina antiparaleloz eratua dago, eta sarrera, N-terminal helizeaz (α1) eta C-terminal laminaz (β6) babestuta dago.[47] Bestalde, A RNasa entzimaren malgutasuna oso garrantzisua izateaz gain, katalisi abiadura mugatzen du: urrats mantsoan, proteinaren aldaketa konformazionala ematen da produktoa askatuz. Entzimak konformazio askea eta lotuaren arteko oreka mantentzen du, nahiz eta bi egoeren artean kobaren estrukturan aldaketa egon. Honek, entzimak substratua ez dagoenean ere, katalitikoki garrantzitsuak diren konformazioak lagintzen dituela adierazten du.[46] Bandaren irekiera eta itxiera abiadura konstantea $\sim$ 1000-3000 s^{-1}koa da. Substratua batzen denean, banda itxi egiten da ingurune hidrofobiko bat sortuz. Ondoren produktuak askatzeko banda ireki egiten da. Batzearen ondorioz dagoen aldaketa konformazional hau, koba tamainaren txikitzearekin laburtu daiteke.[48] Aldaketa honek loop batzuk nahasten ditu (4 loopa eta 1 loopa) hala nola gune aktiboko erresiduoak. 4 loopak (64-71 erresiduoak) purina 5' aldetik lotzera behartzen du, 1loopak (14-24 aminoazidoak) gune aktibotik $>$ 20 Åetara dauden erresiduo malgukor asko dituen bitartean.[49] Azken honek α1-helizea eta α2-helizea elkartzen ditu. Hala eta guztiz ere, A RNasak egonkortasun konformazional oso handia aurkezten du, nahikoa geldoa baita pikosegundo-nanosegundo denbora tartean.[47]

Aminoazido bakoitzaren ezaugarriek, proteinaren konformazioan eta funtzioan paper garrantzitsu bat jokatzen dute. Emaitza kristalografikoetatik entzima honen 6 lotura azpigune identifikatu dira[24,50–54] (begiratu 1.1 Eskema): P_{-1}-en Arg85; P_0-n Lys66; P_1-en Gln11, His12, Lys41, His119, Phe120 eta Asp121; P_2-n Lys7 eta Arg10; B_1-an Thr45, Asp83 eta Phe120; B_2-n Asn67, Gln69, Asn71 eta Glu111. P_1 azkiguneko His12, His119 eta Lys41 erresiduo katalitikoen mutazioak, A RNasaren aktibitate gehiengoaren

Figure 1.1: A RNasaren egitura kristalinoa (PDB ID= 1RPG [50])

galera dakar.[41,55] His12 eta His119 azido eta base moduan jokatzen dutela suposatzen den bitartean, Lys41 erresiduoak trantsizio egoera egonkortzen duela usten da.[24,28,56,57] Hala ere, ikerketa batzuk erresiduo hauei beste eginkizun desberdin batzuk ematen dizkiete.[29,58–61] Gainera, His119 erresiduoaren albo-kateak bi konformazio har ditzake, C_α-C_β loturaren inguruan 142°ko errotazioarekin eta C_β-C_γ loturan 38°ko errotazioarekin erlazionatzen direnak: A, $\chi_{1(N-C_\alpha-C_\beta-C_\gamma)} \sim 160°$, eta B, $\chi_1 \sim -80°$.[62] A konformazioa aktiboa kontsideratzen da eta katalisia ahalegintzen duen bitartean, B egitura konformazio inaktiboa kontsideratzen da.[50]

B_1 azpigunea nukleotido pirimidinikoekin bakarrik elkartzen da, zitidinarekiko lehentasuna izanik uridinaren aurrean. Thr45k espezifikotasun hortan zerikusia du, nukleotido pirimidikoekin hidrogeno loturak sortzeaz gain, adenina nukleobasearekin eragozpen esterikoak baititu.[63] Gainera, oinarrizko egoera ezegonkortzen duenez, katalisian paper oso garrantzitsua du.[26,27] Bestalde, Asp83k ezin du zuzenean lotutako substratuarekin interakzionatu, baina hidrogeno lotura bat sortzen du Thr45 erresiduoaren talde hidroxiloarekin, B_1 azpigunean espezifikotasuna sortzen duen aminoazidoa. B_1-aren azken erresiduoa, Phe120, substratuaren lotura indartzen du bere π elektroi eta hidrofobizitatearen bitartez, baina ez du purina/pirimidina espezifikotasunean parte hartzen.[64,65] Hala eta guztiz ere, His119 erresiduoaren kokapen zorrotzaren arduraduna da.[66] B_2 azpigunea base guztiekin batu daiteke hurrengo abiadura konstante ordena aurkeztuz A > G > C > U.[25,67] Beste alde batetik, P_2 azpiguneko erresiduen mutazioak A

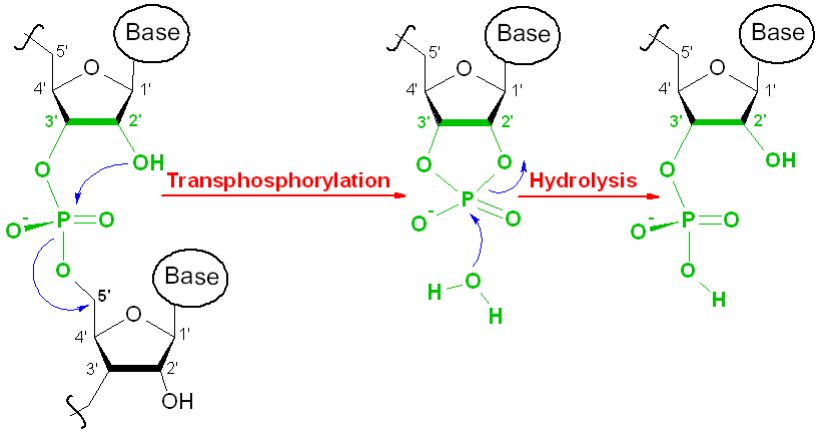
RNasaren endonukleasa aktibitatearen galera eta exonukleasa aktibitatearen indartzea dakar.[32–34] Gainera, P_0 eta P_2-ko aminoazidoek substratuaren lotura zuzenaz gain, katalisian zeharkako ekarpena dute. Hauen interakzio coulombikoak His12 eta His119ren pK_a-k balio egokietan mantentzen dituzte.[68] P_1 azpiguneko Gln11 erresiduoak paper oso garrantzitsua jokatzen du katalisian, hidrogeno lotura bat eratzen baitu bere albo-kateko nitrogenoa eta P_1-eko talde fosforilikoaren artean, substratua modu egokian orientatuz eta honen lotura era ez emankor batean gerta ez dadin eragotziz.[69] Gainera, Gln11 erresiduoak entzima-substratu konplexuaren energi askea handiagotzen du.[26] Bestalde, Asp121 eta His119 erresiduek bikote kataliko bat eratzen dute, His119ren tautomero egokia ipintzeaz gain, A RNasaren egonkortasun konformazionala handitzen duena.[70,71] Gune aktiboa eta loop 4aren arteko lotura nagusia erresiduo honek, Asp121, egiten du.

1.3.2 Erreakzio Mekanismoa

A RNasa endoerribonukleasa, urdaileko mikroorganismoek sortzen duten RNA kantitate handiak digeritzeko beharrezkoa da.[2] Kate bakuneko RNAren 3',5'-fosfodiester lotura apurketa katalizatzen du. Entzima honi buruzko datu biokimiko, fisiko eta kristalografiko aukera zabala dago.[24,35,72] Horietatik, RNAren hidrolisiaren katalisia bi urratsetako mekanismo bat izan daitekeela ondorioztatu da,[73] transfosforilazioa eta hidrolisia, nun bitartekari/produktu fosfodiester zikliko bat eratzen den gero hidrolizatua izateko (begiratu 1.2 Eskema). 3',5'-fosfodiester lotura aktibo gehiago geratzen ez direnean bakarrik hasten da 2',3'-fosfodiester ziklikoaren hidrolisia 3'-fosfato terminal produktua emateko.[74,75] Beraz, A RNasaren eraginkortasun katalitikoa askoz handiagoa da transfosforilazio erreakzioan hidrolisian baino. Substratuaren kate luzera transfosforilazio erreakzio abiaduran kontutan hartu behar den beste faktore bat da. Ikerketa zinetiko batzuk ondorengoa ziurtatu dute:[25]

- Entzimaren eraginkortasun katalitikoa substratu kate luzerarekin handitzen da.
- Gune aktiboko 5' aldean dauden base nitrogenatuek eraginkortasun katalitikoa dute.
- A RNasak substratu polinukleotidoak oligonukleotidoak baino nahiago ditu.

Bestalde, mekanismoaren bi urratsak base orokor batek eta azido orokor batek modu kontzertatu batean katalizatzen dituztela usten da. Bi urratsak fosforoaren gain desplazamendu lerrokatua gertatu ondoren, bipiramide trigonaleko geometria duen trantsio egoeretatik igarotzen direla suposatzen da.[73,76–78] Mekanismo honentzako lau postulatu daude:



Scheme 1.2: RNA hidrolisiaren katalisia bi-urratsetako mekanismo batean, transfosforilazioa eta hidrolisia. Lehenego fosfodiester zikliko bat eratzen da bitartekari/produktu moduan, zeina gero hidrolizatua izaten den.

- Bost kideko eraztun batean posizio apikal eta ekuatorial bat luzatzerakoan, eraztunaren tentsioa minimizatzen da.
- Talde elektronegatiboenek posizio apikalak betetzen dituzte.
- Bitartekariaren pseudoerrotazioa gertatu daiteke.
- Taldeak bitartekariaren posizio apikaletatik sartu eta atera behar dira.

Nahiz eta informazio experimental asko izan, oraindik ez dago mekanismoaren ulermen orokor bat, ezta RNasak sortzen duen katalisi abiadura handitzearen analisi zehatzik. Gaur egun, abelgorri A RNasa pankreatikoak RNA hidrolizatzeko beste mekanismo desberdin batzuk jarraitu ditzakela proposatu da.[29,60,61,73,76,79–81] Gainera, trantsizio egoeren naturari eta protonazio egoerari buruz eztabaida handia dago oraindik.[28,82]

Lehenago aipatu dugun moduan, azido-base katalisi orokor baten bitartez, A RNasak RNA substratua bi urratsetako prozesu baten bitartez apurtzen du.[73] Lehenengo urratsan (transfosforilazioan), RNA katea fosfoester truke batez apurtzen da: ribosaren 2'-hidroxil taldeak fosfoesterraren lotura erasotzen du (P-$O_{2'}$ lotura eratzen da) eta hurrengo nukleotidoaren $O_{5'}$ oxigenoa kanporatua izaten da (P-$O_{5'}$ lotura apurtzen da). Honen ondorioz, 2',3'-fosfato zikliko bat sortzen da. Berriz, bigarren urratsan (hidrolisian), aktibatutako ur batek 2',3'-fosfato ziklikoa erasotzen du, fosfato 3'-terminal bat eskuratuz. Horretarako, P-$O_{2'}$ lotura apurtzen du eta 2'-hidroxil taldea berreskuratzen da (begiratu 1.2 Eskema). Erreakzio hauen talde katalitiko nagusiak His12 eta His119aren albo-kateak eta Lys41aren amino talde kationikoa dira. Kontutan izan behar da ere Gln11aren albo-katea eta Phe120aren kate nagusia tartean sartua daudela, bigarren mailako paper garrantzitsua jokatuz.

Mekanismo Klasikoa

Mekanismo klasikoa[73,76] katalisi azido-basiko kontzertatu bat da. Bertan, histidina katalitikoetako bat (His12) neutroa da eta base orokor bezala jokatzen duen bitartean, bestea (His119) protonatua dago eta azido orokor bezela jokatzen du. Transfosforilazio urratsan (2',3'-fosfato zikliko bat sortzen da, $PO_{2'}$ lotura eratuz eta P-$O_{5'}$ apurtuz), His12ren imidazol albo-kateak base moduan jokatzen du 2'-hidroxil hidrogenoa substratutik abstraituz, eta honen erasoa fosforo atomoan erraztuz. 5' nukleosidoaren desplazamendua lortzeko eraso hau lerrokatua izan behar da. Bitartean, His119ren imidazolium albo-kateak azido bezela jokatzen du $O_{5'}$ oxigenoa protonatuz, eta ondorioz atera behar den taldearen desplazamendua errazten du. Bukatzeko, lortutako bi produktuak, 2',3'-fosfato ziklikoa eta nukleosidoa disolbatzailean askatuak dira.

2',3'-fosfato ziklikoaren hidrolisiaren urrats geldoa beste prozesu batean gertatzen da. Prozesu hau transfosforilazioaren alderantzizkoa dela dirudi (3'-fosfato terminal bat lortzen da, P-$O_{5'}$ lotura eratuz eta P-$O_{2'}$ apurtuz). Oraingoan His119 base moduan jokatzen du nukleofiloaren (ur molekula) protoi bat hartuz eta honen eraso nukleofilikoa fosforoaren gain erraztuz. Bitartean, His12k azido moduan parte hartzen du $O_{2'}$ oxigenoa protonatuz.

Bi erreakzio hauek oxifosforano dianioniko trantsizio egoeratik igarotzen dira. Bertan fosforo atomoa pentakobalentea da. Trantsizio egoera hauek hidrogeno loturen bitartez egonkortzen dituzte Lys41 eta Gln11ren albo-kateek eta Phe120aren kate nagusiak. Ondorioz, erresiduo hauek katalisia areagotzen dute.

Breslow-ren Mekanismoa

Breslow eta bere kolaboratzaileek[79,83] proposatutako mekanismoa, mekanismo klasikoa[73,76] moduan, azido-base katalisi kontzertatu bat da. Baina oraingoan, transfosforilazio urratseko eraso nukleofilikoa bi protoi mugimendurekin kontzertatua dago: lehenengoa substratuaren 2'-hidroxilotik His12 base orokorrari, eta bigarrena His119 azido orokorretik O_{2P} oxigeno fosforilikora, $O_{2'}$ren eraso nukleofilikoa fosforoaren gain errazten duena. Hurrengo urratsan, His119k O_{2P} oxigenoaren hidrogenoa berreskuratzen du eta $O_{5'}$ri ematen dio, askatzen den taldearen mugimendua erraztuz. Eraso nukleofilikoaren aurretik zubirik sortzen ez duen oxigeno fosforiliko bat protonatzen denez, triester moduko mekanismoa deitzen zaio honi ere. Urrats honetan sortzen den trantsizio egoera oxifosforano monoanioniko bat da. Azkenik, bi produktuak, 2',3'-fosfodiester ziklikoa eta nukleosidoa, disolbatzailera askatzen dira.

2',3'-fosfodiester ziklikoaren hidrolisi geldoa beste prozesu batean ematen da eta mekanismo klasikoarekin[73,76] antza handia hartzen du. Urrats honen trantsizio egoera oxifosforano dianioniko bat da. Ez da beharrezkoa

bitartekariaren protonazioa gertatzea transfosforilazioan bezala, hala ere, dianioiak protoi bat eskuratu dezake geroago.

2',3'-fosfodiester ziklikoaren hidrolisi geldoa beste prozesu batean ematen da eta mekanismo klasikoarekin[73,76] antza handia hartzen du. 2',3'-zikloa nahikoa erreaktiboa danez,[84] ez da beharrezkoa honen protonazioa gertatzea eraso nukleofilikoa aurretik, transfosforilazioan bezala. Hala ere, dianioiak protoi bat eskuratu dezake geroago.

Haydock-en Mekanismoa

Haydock eta bere lankideek,[80,81] Breslow-en mekanismoak[79] lehengo urratsarako azaltzen duen mekanismoaren beste bide desberdin bat proposatu zuten. Oraingo honetan, His119k bere protoia eman beharrean, hidrogeno lotura bat sortzen du oxigeno fosforiloarekin. Berriz, His12k 2'-hidroxiloaren hidrogenoa oxigeno fosforilora desplazatzen laguntzeaz gain (konbertsio tautomerikoa), $O_{2'}$-ren erasoa fosforoaren gain ahalegintzen du. Guzti honen ondoren, His119k $O_{5'}$ protonatzen du ateratzen den taldearen mugimendua erraztuz. Azkenik, sortutako bi produktuak disolbatzailera askatuak dira. Oraingoan trantsizio egoera oxifosforano monoanioniko bat da.

2',3'-fosfodiester ziklikoaren hidrolisi geldoa, beste prozesu batean ematen da eta mekanismo klasikoarekin[73] antza handia du. Gln11aren albo-kateak eta Phe120aren kate nagusiak fosforo atomoaren karakter elektrofilikoa handituz, katalisia areagotzen duten bitartean, Lys41k trantsizio egoera egonkortzen du ur molekula baten bitartez oxigeno fosforiliko batekin interakzio ez zuzen bat sortuz.

Wladkowski-ren Mekanismoa

Wladkowski eta bere lankideek[60,85–87] proposatutako mekanismoan Lys41 erresiduoa base moduan jokatzen du transfosforilazio urratsan eta azido moduan hidrolisian. Hauek guztiz protonatutako bipiramide trigonal pentakobalente (TBP) bitartekari oso egonkor bat proposatu zuten, nun His12 eta Lys41 erresiduoak neutroak izan behar diren. His119 erresiduoak azido moduan jokatzen du transfosforilazioan $O_{5'}$ oxigenoa protonatuz eta askatzen den taldearen desplazamendua erraztuz, berriz base moduan jokatzen du hidrolisian talde nukleofilikoaren protoi bat abstraituz eta honen eraso nukleofilikoa fosforo atomoaren gain erraztuz.

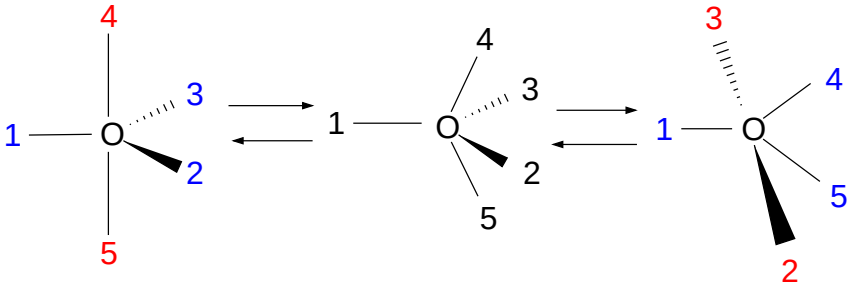
Lopez-en Mekanismoa

Lopez eta bere lankideek[29] hidrolisi urratsaren beste mekanismo bat proposatu zuten. Oraingoan, $O_{2'}$ oxigenoa protonatzen duen azidoa Lys41 da, His12 beharrean mekanismo klasikoak[73,76] adieratzen duen moduan. Wladkowski[87] eta honen mekanismoaren arteko desberdintasuna trantsizio egoeraren protonazioan datza. Wladkowskik bitartekari neutro bat proposatzen

duen bitartean, Lopezek bitartekari dianioniko bat iradokitzen du. His12 eta Gln11ren albo-kateek eta Phe120aren kate nagusiak hidrogeno loturen bitartez, oxigeno ekuatorialen gehiegizko karga negatiboa egonkortzen dute.

1.4 Oxifosforanoak

Oxifosforanoak garrantzi biologiko handiko molekula bipiramide trigonal pentakobalenteak dira (TBP). Hauen ligandoek bi posizio har ditzakete, apikala edo ekuatoriala (begiratu 1.3 Eskema). Oxifosforanoak transesterifikazio eta hidrolisi erreakzioen trantsizio egoerak edo bitartekariak izaten dira. Lehenago aipatu dugun moduan, RNAren transfosforilazioa RNAren ribosa eraztunean aktibatutako 2'-hidroxi taldearen eraso lerrokatutik fosfatoaren gain abiatzen da, TBP trantsizio egoera/bitartekaria sortuz. Ondoren P-$O_{5'}$ loturaren apurketa gertatzen da eta 2',3'-fosfato zizklikoa sortzen da.



Scheme 1.3: Estruktura bipiramide trigonal pentakobalenteen irudikapen eskematikoa. TBP-k bi ligando mota dituzte: ligando ekuatorialak (urdinez) eta ligando axialak (gorriz). Berry pseudoerrotazio mekanismoan, piramide tetragonal estruktura bitartez ligando posizioen aldaketa bat gertatzen da. Bi ligando ekuatorial posizio axialetara igarotzen diren bitartean, bi ligando axialak posizio ekuatorialetara mugitzen dira. Atomo pibotea, 1, posizio ekuatorialean mantentzen da, baina piramide tetragonal trantsizio egoeran posizio apikala hartzen du.

Urte askotan zehar transferentzi fosforiliko erreakziotan bitartekari egonkorrak dauden edo ez eztabaidatu da.[28,82,88] Bitartekariaren erdibizitza denborak, protonazio egoerarekin eta fosforanoaren tamainarekin handitzen dena, funtsezko garrantzia du. Hala ere, egoera experimental desberdinetan fosfodiester eta bitartekariak izan dezaketen protonazio egoerari buruzko eztabaida dago. Fosfodiester zikliko eta ez-ziklikoen hidrolisia gas eta medio homogeneoetan (A RNasaren katalisi erreakzioaren berdintsua) aztertzen duten lan askok, oxifosforano bitartekari egonkor bat proposatzen dute erreakzio-bidean zehar.[29,89–92] Fosforano espezien protonazioak, pH eta fosforanoaren pK_a-ren menpe dagoena, egonkortasunean eta erreakzioetan lortuko den produktuan zerikusi handia du. Baldintza azidoetan oxifosforano bitartekari monoanioniko batek pseudoerrotazioa izan dezake. Ligando posizioen aldaketa piramide tetragonal baten bitartez, bi ligando ekuatorial posizio

axialetara igarotzen diren bitartean, bi ligando axialak posizio ekuatorialetara mugitzen dira. Atomo pibotea posizio ekuatorialean mantentzen da, baina piramide tetragonal trantsizio egoeran posizio apikala hartzen du (begiratu 1.3 Eskema). Berry pseudoerrotazioa dagoen edo ez jakin daiteke 2',5'-lotura duen migraketa produktua identifikatzen bada. Bestalde, baldintza basikoetan oxifosforano dianioniko trantsizio egoera/bitartekaria eratzen da. Honen pseudoerrotazioa ezinezkoa da, elektroi-emaile oxianioi bat posizio apikalean jarri beharko baitzen, espezie oso ezegonkor bat eratuz.[84] Ondorioz, ez da migrazio produkturik aurkitzen oxifosforano dianionikoarentzako.

Datu experimental eta teorikoak transfosforilazioaren bitartekari oxifosforano dianioniko bat zinetikoki trantsizio egoeratik bereiztu ezina izateaz gain, protonazioa edo pseudoerrotazioa emateko bizitza denbora oso motza duela diote gas fasean.[82,89–94] Hala ere, orain dela gutxi emaitza konputazionalak[61] kontutan hartuz oxifosforano dianioniko bitartekari egonkor bat proposatu da A RNasa hidrolisi erreakzioarentzako. Bestalde, bitartekariaren existentziatik baztertuz, lotura endoziklikoa, P-$O_{2'}$, exoziklikoa, P-$O_{5'}$, baino ahulagoa da.[95,96] Beraz, oxifosforanoaren oxigeno exoziklikoaren irteera izan behar du urrats mugatzailea.

Estrukturaren hidrogeno lotura intramolekularrei esker lortzen da oxifosforano neutro/diprotiko eta monoanioniko/monoprotikoen egonkortasuna. Lotura intramolekular hauek P-O lotura axialen elongazioa eta ahulezia sortzen dute, batez ere apurketa endoziklikoarekin erlazionatua dagoenarena, P-$O_{2'}$.

1.5 Tesi honen helburua

Gaur egun ez da lortu oraindik RNAren hidrolisi mekanismoaren erabateko ulermena ezta A RNasak sortzen duen abiadura hobetzearen analisi zehatzik, nahiz eta entzima honen datu ugari ezagutzen diren. Gainera, oxifosforano pentakobalente bitartekari/trantsizio egoeren natura eta protonazioa oraindik ez dago argi. Honako tesi honetan ezagutzen den datu multzoari atal bat gehitzen, eta oraindik argi ez dauden arloei ikuspuntu berri bat ematen saiatzen gara. Tesia lau ataletan banatzen da:

- Part I: Ikertutako sistemaren eta erabilitako metodoen sarrera orokorra.
- Part II: RNAren transfosforilazio eta hidrolisi erreakziotan sortzen diren A RNasa konplexu desberdinen dinamika konformazionala, estrukturen erlaxazioa eta solbatazio desberdinak, dinamika molekular klasikoa erabiliz ikertzen dira. 3. Kapituluak 3',5'-CpA dinukleotidoa eta A RNasaren arteko konplexua aztertzen du erreaktibo, eta transfosforilazio eta hidrolisi trantsizio egoeren mimika modeloak erabiliz.

4. Kapituluan transfosforilazio eta hidrolisien fosforano bitartekariek eta A RNasak sortzen duten konplexuen thio ordezkapenak, dinamika eta solbatazioan duen eragina aztertzen da.

- Part III: QM/MM metodologia erabili da oxifosforano bitartekarien protonazio egoera desberdinen egonkortasun erlatiboa eta A RNasa entzimak katalizatzen duen RNAren hidrolisiaren profil energetiko desberdinak aztertzeko. Hamiltoniano semienpiriko berri bat erabili da QM zatiarentzako, AM1/d-PhoT.[97] Hau transferentzia fosforilikoa duten erreakzioetarako parametrizatua izan da.

- Part IV: Aurreko kapituluen informazio gehigarria.

Chapter 2

Ribonucleases

2.1 Introduction

In modern life on earth, DNA and RNA collaborate with a third encoded class of biomolecules, proteins, to support Darwinian evolution. This collaboration, generated the enormous range of life that today occupies nearly every habitable niche on the planet. DNA stores biological information and proteins manifest it. The conduit between the storage and the manifestation is RNA.[1]

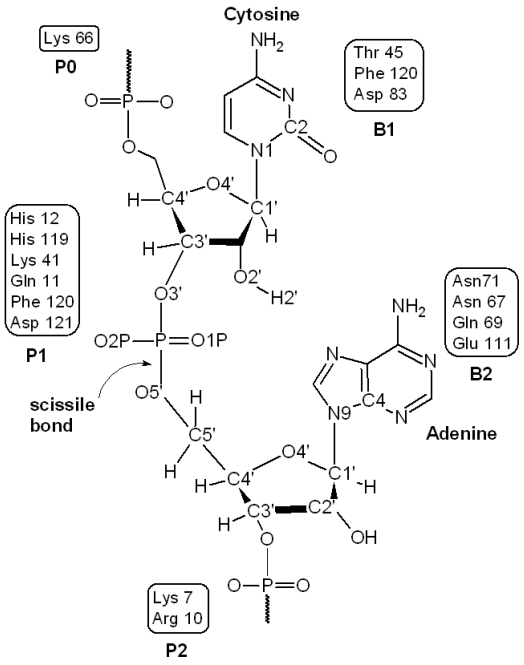
$$\text{DNA} \rightleftharpoons \text{RNA} \rightarrow \text{protein}$$

The flow of information through RNA is essential for known life. Two classes of enzymes control this flow by catalyzing the synthesis or degradation of RNA. The synthesis is catalyzed by RNA polymerases, while RNA degradation is catalyzed by RNA depolymerases, which are most often called *"ribonucleases"* (abbreviated as RNase). All ribonucleases exhibit cytotoxic effects. They first are adsorbed specifically to certain cells, then RNases enter cells' cytosol, and finally degrade the RNA. Consequently, ribonucleases inhibit protein synthesis and cause cell death.[2] In addition, RNases possess therapeutic opportunities for cancer treatment. The damage cause to RNA by RNases could be an important alternative to usual DNA-damaging chemotherapeutics.[3,4]

Ribonucleases can be divided into exoribonucleases and endoribonucleases. The exoribonucleases[5] degrade RNA by removing terminal nucleotides from either 5' end or 3' end of RNA molecule. Whereas, endoribonucleases create internal breaks. One of the most studied families of this last type ribonucleases is the *pancreatic ribonuclease A* family.[6]

2.2 Pancreatic Ribonuclease A family

This intriguing enzyme superfamily actions are closely linked to the evolution of vertebrates and has undergone tremendous diversification.[6–11] RNase A family contains proteins, which decompose the RNA polymer chain. Many members of the superfamily have potent biological activities that are dependent on their enzymatic activity. For example, human angiogenin[12] (RNase 5) is implicated in cancer and in vascular and rheumatoid diseases;[13–16] eosinophil-derived neurotoxin (EDN, RNase 2) and eosinophil-cationic protein (ECP, RNase 3) are neurotoxic *in vivo* and are involved in hypereosinophilic syndromes and allergy;[17,18] bovine pancreatic seminal RNase has antispermatogenic and immunosuppressive activity;[19,20] and Onconase has powerful antitumor activities.[21–23]



Scheme 2.1: Representation of Bovine Pancreatic Ribonuclease A active site with a bound RNA ligand. The subsites interacting with RNA bases, B_i, and phosphate groups, P_i, are indicated.

The process of RNA depolymerization involves the interaction of the enzyme with the polymeric substrate in the active site and its correct alignment on the surface of the enzyme through multiple binding subsites (see Scheme 2.1). According to the established nomenclature,[24] the active site is partitioned into subsites B_i and P_i, which interact, respectively, with the RNA bases and phosphate groups.[25] Subsite B_1 has a strong specificity for

pyrimidine bases, with a 30-fold kinetic preference for cytidine versus uridine.[26] The specificity is conferred by residue Thr45 in bovine pancreatic ribonuclease A, which is strictly conserved in all RNases.[27] The protein cleaves the P-$O_{5'}$ (scissile) bond on the 3' side of pyrimidine bases bound at B_1. Two catalytic histidines and a single lysine in subsite P_1 (residues His12, His119, and Lys41 in bovine pancreatic ribonuclease A) are strictly conserved among RNase homologs and play key roles in the reaction.[24,28,29] Subsite B_2 is highly conserved and recognizes all bases, having a preference for adenine.[30] It provides a binding pocket for the leaving group nucleotide. On the other hand, P_2 subsite contributes to the binding efficiency in the formation of the enzyme-substrate complex, and plays an indirect role in the catalytic mechanism.[31] It seems that this binding site is involved and plays a key role in the *endonucleolytic activity* (cleavage at an internal part) of RNase A.[32–34] The variants that lack electrostatic interaction in this site show a clear shift toward the exonucleolytic cleavage of substrate. Other subsites (e.g., P_0 and B_3) are more variable among homologs.

2.3 Bovine Pancreatic Ribonuclease A

Bovine pancreatic ribonuclease A (RNase A) [EC 3.1.27.5], the predominant form of enzyme produced by the bovine pancreas, is now increasingly recognized as the archetype of ribonuclease A superfamily. It has been studied for over 60 years,[24,28,35] which had lead to the purification of RNase A, the elucidation of the complete amino acid sequence, solution of the crystal structure, clarification of protein folding pathways and identification of the elements underlying the catalytic mechanims.[36–42] In 1972 three researches and a fourth one in 1984 were awarded with the Nobel Prize in chemistry for work on this enzyme[43–45] (see Table 2.1). Stanford Moore and William H. Stein were the first researches on providing for the first time the amino acid composition of a protein, RNase A, and some years later its amino sequence, another first time for an enzyme and a third for a protein.[44] The prize was shared with Christian B. Anfinsen who discovered that the three-dimensional structure of a protein is based on its amino acid sequence.[43] The Nobel laureate of 1984, Bruce Merrifield, achieved the complete chemical synthesis of the enzyme.[45] Therefore, this endoribonuclease is an ideal system for a theoretical study of the structural and dynamic basis of enzyme catalysis.

2.3.1 Structure

RNase A is a single polypeptide of molecular mass 13.7 kDa and a sequence length of 124 amino acids. It contains 19 of the 20 natural amino acids, lacking only tryptophan. The molecule is kidney-shaped with a long N-terminal helix and shorter helices that are packed against a central β-pleated sheet (see Figure 2.1). The overall structure is stabilized by four disulfide bonds

Nobel Laureate	Year	Nobel Lecture
Christian B. Anfinsen (1916-1995)	1972	Studies on the Principles that Govern the Folding of Protein Chains [43]
Stanford Moore (1913-1982)	1972	The Chemical Structures of Pancreatic Ribonuclease and Deoxyribonuclease [44]
William H. Stein (1911-1980)	1972	The Chemical Structures of Pancreatic Ribonuclease and Deoxyribonuclease [44]
Bruce Merrifield (1921-2006)	1984	Solid Phase Synthesis [45]

Table 2.1: Nobel prizes in Chemistry for work on Bovine Pancreatic Ribonuclease A

that involve all eight of its cysteine residues, and large solvent-exposed loops are presented.[46] Besides, the binding pocket is formed by three antiparallel β-sheets, which have groove shape. The entrance to the active site cleft is flanked by the N-terminal helix (α1) and the C-terminal strand (β6).[47] RNase A is an enzyme in which flexibility is important and rate limiting to catalysis: the slow step involves a protein conformational change that gates the product release. This enzyme presents an equilibrium between the free and bound conformation, which brings a change in the groove's structure, indicating that the enzyme samples the catalytically relevant conformations even in the absence of substrate.[46] The rate constant of this opening and closing of the hinge is of $\sim$ 1000–3000 s^{-1}. When the substrate binds, the hinge closes, creating a hydrophobic environment. Later, the hinge opens to release the products. The conformational change can be summarized by the reduction in the size of the active site groove upon binding.[48] This change involves several loop regions (loop 4 and loop 1) as well as residues at active site. Loop 4 (residues 64–71) provides specificity of purine binding at 5' side, and loop 1 (residues 14–24) contains several flexible residues that are $>$ 20 Å from the active site.[49] It connects α1-helix and α2-helix. Nevertheless, RNase A has high conformational stability. It is quite immobile in the picosecond to nanosecond time window.[47]

The different characteristics of each amino acid residue play key roles in the conformation and function of the protein. From the results of crystallographic studies, six substrate binding subsites have been identified for this enzyme[24,50–54] (see Scheme 2.1): Arg85 at P_{-1}; Lys66 at P_0; Gln11, His12, Lys41, His119, Phe120 and Asp121 at P_1; Lys7 and Arg10 at P_2; Thr45, Asp83 and Phe120 at B_1; Asn67, Gln69, Asn71 and Glu111 at B_2. The modification of catalytic residues His12, His119 and Lys41 results in the loss of most of RNase A activity.[41,55] The putative role of His12 and His119 is to act as an acid and a base, while Lys41 acts stabilizing the tran-

Figure 2.1: Crystal structure of RNase A (PDB ID= 1RPG [50])

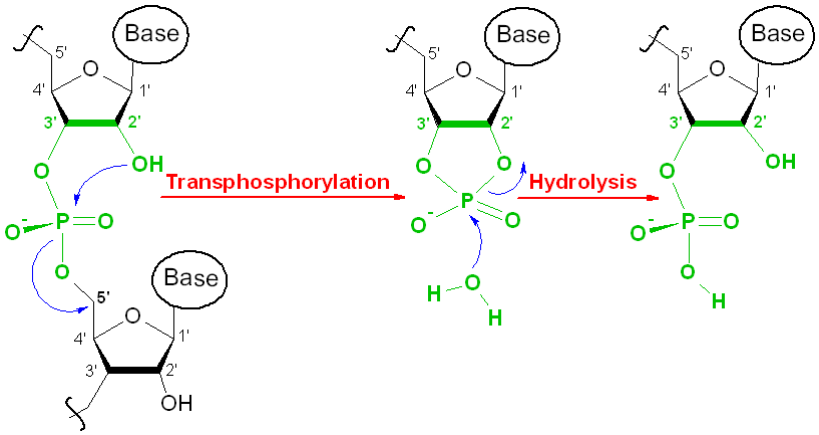
sition state.[24,28,56,57] However, some studies present different roles for those residues.[29,58–61] In addition, the side chain of residue His119 can adopt two conformations denoted as A, $\chi_{1(N-C_\alpha-C_\beta-C_\gamma)} \sim 160°$, and B, $\chi_1 \sim -80°$ [62], which are related by a 142° rotation about $C_\alpha - C_\beta$ bond and 38° rotation about the $C_\beta - C_\gamma$ bond. Conformation A is considered as the active conformation, which promotes catalysis, whereas conformation B is considered as the inactive conformation.[50]

The B_1 subsite binds only to pyrimidine nucleotides with a preference for cytidine versus uridine. Thr45 contributes to this specificity by forming hydrogen bonds with pyrimidine nucleobases and by steric exclusion of adenine nucleobase.[63] Moreover, Thr45 plays an important role in catalysis by destabilizing the ground state.[26,27] Besides, Asp83 cannot interact directly with the bound substrate. It forms a hydrogen bond with the hydroxyl group of Thr45, which gives substrate specificity at B_1 subsite. The other residue of B_1 subsite, Phe120, is likely to enhance substrate binding with the aid of its π electron and hydrophobicity, but it does not contribute to purine/pyrimidine specificity.[64,65] However, it is responsible for the strict positioning of His119.[66] The B_2 subsite bind all bases, and presents a rate constant order of A > G > C > U.[25,67] On the oder hand, the mutation of P_2 subsite residues, results in a loss of endonuclease activity and increase of exonuclease activity of RNase A.[32–34] Besides, amino acid residues of P_2 and P_0 subsites make some indirect contribution to activity, other than direct substrate binding. Their coulombic interactions maintain His12 and His119

at their optimum pK_a values.[68] Residue Gln11 of subsite P_1 plays an important role in catalysis, it forms a hydrogen bond between its side chain nitrogen and the phosphoryl group of P_1 subsite orienting the substrate and preventing it from binding in a nonproductive mode.[69] Moreover, Gln11 increases the free energy of the enzyme–substrate complex.[26] Whereas, Asp121 makes a hydrogen bond interaction with His119, forming a catalytic dyad which positions the proper tautomer of His119 and enhances the conformational stability of RNase A.[70,71] This residue provides the main link between the active site pocket and loop 4.

2.3.2 Reaction Mechanism

RNase A distributive endoribonuclease is needed to digest the large amounts of RNA produced by microorganisms present in forestomachs.[2] It catalyzes the breakdown of 3',5'-phosphodiester linkage of single stranded RNA. A particularly wide range of biochemical, physical, and crystallographic data are available for this enzyme.[24,35,72] These have led to proposals for the catalysis of the hydrolysis of RNA by a two-step mechanism,[73] transphosphorylation and hydrolysis, in which a cyclic phosphodiester intermediate/product is formed and subsequently hydrolysed (see Scheme 2.2). Only when no susceptible 3',5'-phosphodiester bonds are left, the hydrolysis of the 2',3'-cyclic phosphodiester bonds to a 3'-phosphate product takes place.[74,75] Actually, the catalytic efficiency of RNase A for the transphosphorylation reaction is much higher than that for the hydrolysis reaction. The length of the substrate is another factor that affects the rate of transphosphorylation. Some kinetic studies have demostrated that:[25]



Scheme 2.2: Catalysis of the hydrolysis of RNA by a two-step mechanism, transphosphorylation and hydrolysis. A cyclic phosphodiester intermediate/product is formed and subsequently hydrolysed.

1. The catalytic efficiency of the enzyme increases with the length of the

substrate

2. The catalytic efficiency depends on the nitrogenated bases adjacent at the 5'-side of the active site

3. RNase A prefers polynucleotide substrates rather than oligonucleotides

On the other hand, both steps, transphosphorylation and hydrolysis, are thought to involve in-line displacement at the phosphorus, to pass through trigonal bipyramid (TBP hereafter) transition states and to be catalyzed by the concerted action of a general acid and a general base.[73,76–78] There are four postulates for this mechanism:

1. If a five-membered ring is present, ring strain is minimized if it spans one apical and one equatorial position.

2. Apical positions are occupied by the more electronegative groups.

3. Pseudorotation of the intermediate may occur.

4. Groups must enter and leave the intermediate from apical positions.

In spite of the accumulated experimental information, a full understanding of the mechanism and a detailed analysis of the rate enhancement produced by ribonuclease has not been achieved, and nowadays, there are different mechanism proposed[29,60,61,73,76,79–81] for RNA hydrolysis catalyzed by bovine pancreatic RNase A. Moreover, there remains considerable debate concerning the nature of the structures of the transitions states along the reaction path and their protonation states.[28,82]

As mentioned above, RNase A cleaves RNA substrates in a two-step process, by general acid-base catalysis.[73] In the first step (transphosphorylation), the RNA chain is cleaved by a phosphate ester exchange, in which the 2'-hydroxyl group of ribose attacks the phosphate ester linkage ($P–O_{2'}$ bond is formed) and the $O_{5'}$ oxygen of the next nucleotide is ejected ($P–O_{5'}$ bond is cleaved). As a result, a 2',3'-cyclic phosphate is formed. In the second step (hydrolysis), the previously formed $P–O_{2'}$ bond of the 2',3' cyclic phosphate is cleaved by an attacking activated water molecule to generate a 3'-terminal phosphate group and regenerate the 2'-hydroxyl group (see Scheme 2.2). The principal catalytic groups for these reactions are the side chains of His12 and His119 and the cationic amino group of Lys41, whereas in addition, the side chain of Gln11 and the main chain of Phe120 are also implicated as having important secondary roles.

Classical Mechanism

The classical mechanism[73,76] is a concerted acid-base catalysis, where one of the catalytic histidines (His12) is neutral and acts as a general base,

while the other (His119) is protonated and acts as a general acid. In the transphosphorylation step (P–$O_{2'}$ bond is formed and P–$O_{5'}$ bond is cleaved, generating a 2',3'-cyclic phosphate), the imidazole side chain of His12 acts as a base that abstracts the 2'-hydroxyl hydrogen of the substrate, and thereby facilitates its attack on the phosphorus atom. This attack proceeds in line to displace the 5' nucleoside. Meanwhile, the imidazolium side chain of His119 acts as an acid that protonates $O_{5'}$ oxygen, facilitating the displacement of the leaving group. Finally, both products, the 2',3'-cyclic phosphodiester and the nucleoside, are released to solvent.

The slow hydrolysis of the nucleoside 2',3'-cyclic phosphodiester occurs in a separate process that resembles the reverse of transphosphorylation (P–$O_{5'}$ bond is formed and P–$O_{2'}$ bond is cleaved, creating a 3'-terminal phosphate). Now, His119 acts as a base abstracting the proton from the nucleophile, a water molecule, and facilitating the nucleophilic attack on the phosphorus atom. While His12 acts as an acid, protonating $O_{2'}$ oxygen.

Both reactions occur via dianionic oxyphosphorane transition states, where the phosphorus is a pentacovalent atom. The side chain of Lys41 and Gln11, and the main chain of Phe120 enhance catalysis by stabilizing these transition states via hydrogen bonds.

Breslow's Mechanism

As the general accepted mechanism, classical mechanism,[73,76] this one proposed by Breslow and co-workers[79,83] is a concerted acid-base catalysis. Although in this mechanism, the nucleophilic attack in the transphosphorylation step is concerted with two proton transfers: a first one from the 2'-hydroxyl hydrogen of the substrate to the general base His12; and a second one from the general acid His119 to O_{2P} phosphoryl oxygen, facilitating the attack of $O_{2'}$ on the phosphorus atom. In the next step, imidazole side chain of His119 removes the proton from the phosphoryl oxygen O_{2P} and transfers it to $O_{5'}$, facilitating the displacement of the leaving group. This mechanism is also called the triester like mechanism, because a non-bridging oxygen is protonated prior to the nucleophilic attack. The transition state form in this step is a pentacovalent monoanionic phosphorane. Finally, both products, the 2',3'-cyclic phosphodiester and the nucleoside, are released to solvent.

The slow hydrolysis of the nucleoside 2',3'-cyclic phosphodiester occurs in a separate process, which is similar to the classical mechanism.[73,76] It is not necessary the protonation of the 2',3'-cyclic phosphodiester prior to the nucleophilic attack, like in the transphosphorylation step, because it is sufficiently reactive.[84] Nevertheless, the dianion can pick the proton later.

Haydock's Mechanism

Haydock and co-workers proposed an alternative mechanism[80,81] for the first step of Breslow's mechanism.[79] Here, the imidazolium side chain of His119 hydrogen bonds instead of giving up its proton, to a phosphoryl oxygen, while imidazole side chain of residue His12 facilitates the intramolecular transfer of the 2'-hydroxyl hydrogen to a phosphoryl oxygen (tautomeric conversion) and the attack of $O_{2'}$ on the phosphorus atom. This is followed by a proton transfer from imidazolium side chain of His119 to $O_{5'}$ oxygen, which facilitates the displacement of the leaving group. Then, both products, the 2',3'-cyclic phosphodiester and the nucleoside, are released to solvent. The transition state form in this step is a monoanionic pentacovalent phosphorane.

The slow hydrolysis step of the nucleoside 2',3'-cyclic phosphodiester occurs in a separate process similar to the classical mechanism.[79] The side chain of Gln11 and the main chain of Phe120 enhance catalysis by increasing the electrophilicity of the phosphorus atom, while the side chain of Lys41 stabilize the transition state by an indirect interaction with a phosphoryl oxygen via a water molecule.

Wladkowski's Mechanism

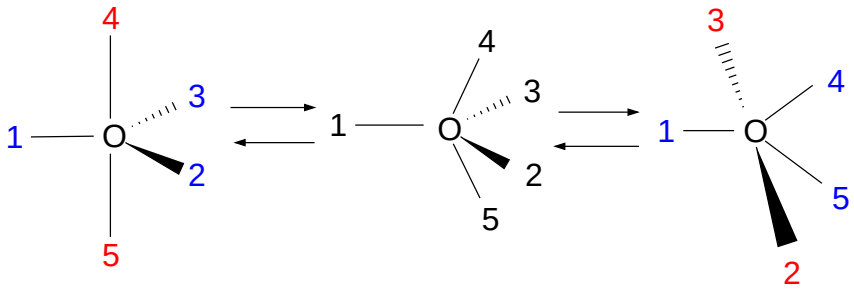
Wladkowski and co-workers[60,85–87] proposed a mechanism where Lys41 acts as the catalytic base in the transphosphorylation step and as the general acid in the hydrolysis step. They suggest a fully protonated and highly stable pentacovalent trigonal bipyramidal (TBP) phosphorane intermediate, where residues His12 and Lys41 are neutral. Residue His119 acts as the general acid in the transphosphorylation step, protonating the $O_{5'}$ axial oxygen and facilitating the displacement of the leaving group, while in the hydrolysis step His119 acts as the general base abstracting the proton from the nucleophile and facilitating the nucleophilic attack on the phosphorus atom.

Lopez's Mechanism

Lopez and co-workers proposed an alternative hydrolysis step,[29] where Lys41 act as the acid protonating $O_{2'}$ oxygen rather than His12 as proposed in the standard mechanism.[73,76] The difference between this mechanism and the one proposed by Wladkowski[87] is in the protonation state of the phosphorane intermediate. Lopez suggests a dianionic intermediate, while Wladkowski suggest a neutral one. Imidazolium side chain of residue His12, side chain of Gln11 and the main chain of Phe120 stabilizes the excess of negative charge of the equatorial oxygens through hydrogen bond interactions.

2.4 Oxyphosphoranes

The oxyphosphoranes are pentavalent trigonal bipyramidal (TBP) molecules of important biological significant, where the ligands occupy two different positions, the apical or the equatorial (see Scheme 2.3). They are the transition states or intermediates in biological transesterification and hydrolysis reactions. As mention above, the dominant reaction path for the transphosphorylation in RNA proceeds via an in-line attack of an activated 2'-hydroxy group of the RNA sugar ring on the reactive phosphate group to produce a TBP phosphorane transition state/intermediate, followed by the cleavage of the P-$O_{5'}$ bond to produce a 2',3'-cyclic phosphate.



Scheme 2.3: Schematic representation of the structure of pentacoordinate trigonal bypiramidal species. The TBP species have two types of ligand: equatorial ligands (in blue) and axial ligands (in red). At Berry pseudorotation mechanism, an exchange of ligand positions via square pyramidal structure occurs. Two equatorial ligands move to axial positions, and the two axial ligands move to equatorial positions. The pivot atom, 1, remains in the equatorial position and it occupies the apical position in the square pyramidal transition state.

The possible presence of stable intermediates in phosphoryl transfer reactions has been actively debated for many years.[28,82,88] The lifetime of the intermediate, which increases with the protonation and the size of the phosphorane, is of paramount importance. In addition, controversy exists concerning the protonation states of the phosphodiesters and intermediate under various experimental conditions. A large number of studies of the hydrolysis of cyclic and non-cyclic phosphodiester in gas and in homogeneous media exist, analogues to RNase A catalysis reaction, where the existence of a stable phosphorane intermediate along the reaction path has been predicted.[29,89–92] The protonation of the phosphorane species, which depends upon the pH and the pK_a of the phosphorane, plays a central role in its stability and the outcome of these reactions. Under acidic conditions, a monoanionic oxyphosphorane intermediate can undergo pseudorotation. An exchange of ligand positions via a square pyramidal structure, where two equatorial ligands move to the axial positions, and the two axial ligands move to the equatorial positions. The pivot atom remains in the equatorial position, and it occupies the apical position in the square pyramidal transi-

tion state (see Scheme 2.3). The Berry pseudorotation can be deduced by identification of the migration product with 2',5'-linkage. By contrast, under basic conditions a dianionic oxyphosphorane transition state/intermediate is formed, which pseudorotation is prohibited because an electron-donating oxyanion must be placed at the apical position that would result in an extremely unstable species.[84] Therefore, no migration product is found for the dianionic oxyphosphorane.

Experimental and theoretical data suggest that a dianionic oxyphosphorane transphosphorylation intermediate is kinetically indistinguishable from a transition state and is too short-live to undergo other processes as protonation or pseudorotation in the gas phase.[82,89–94] However, recently a stable dianionic oxyphosphorane intermediate for the RNase A hydrolysis reaction has been proposed taking into account computational results.[61] Irrespective of whether or not the intermediate exists, the endocyclic bond, P–$O_{2'}$, is intrinsically weaker than the exocyclic bond, P–$O_{5'}$.[95,96] Therefore, the departure of exocyclic oxygen from the oxyphosphorane should be the rate-limiting step.

The stability of the neutral/diprotic and monoanionic/monoprotic oxyphosphorane is due to the intramolecular hydrogen bonding on the structure. This intramolecular hydrogen bonding cause considerable elongation and weakening of the axial P-O bond, in particular the one involved in the endocyclic cleavage, P-$O_{2'}$.

2.5 Scope of this thesis

A full understanding of the reaction mechanism and a detailed analysis of the rate enhancement produced by RNase A has not been achieved for RNA hydrolysis, despite the large amounts of data available for this enzyme. In addition, the protonation states and the nature of the pentacovalent oxyphosphorane transition states/intermediates' structures along the reaction path remain being unclear today. The present thesis is a contribution to the extended range of data and tries to give some clue to the unresolved issues. The thesis work is divided in 4 parts:

- Part I: A General introduction to the system under study and to the applied methodology.
- Part II: Different RNase A complexes' conformational dynamics, structural relaxation, and different solvation that occurs at discreet stages of the transesterification and cleavage reaction of RNA are studied by classical molecular dynamics. Chapter 4 analyzes RNase A complexed with the dinucleotide cytidilyl-3',5'-adenosine (CpA) in the reactant, and transphosphorylation's and hydrolysis' transition state analogs. In Chapter 5, the influence of thio substitutions in the dynamics and

solvation of RNase A complexed with transphosphorylation's and hydrolysis' phosphorane intermediates are studied.

- Part III: QM/MM methodology has been applied to study the relative stability of different protonation states of oxyphosphorane intermediates and the energetic profiles of RNA hydrolysis catalyzed by RNase A enzyme. A new semiempirical Hamiltonian, AM1/d-PhoT ,[97] parametrized for phosphoryl transfer reactions has been used in the QM part.

- Part IV: Supporting Information of previous chapters

Chapter 3

Methods

3.1 Quantum Theory

Computational chemistry and molecular modeling are used to characterize and to predict the behavior of small and macro-molecules by evaluating their energy. The most precise method to calculate the energy of molecular systems is by solving the Schrödinger equation, the basis of the Quantum Theory. This theory, where the energy is quantized, was developed in the twenties of the last century.[98–102] Although at the beginning Quantum Theory was mainly the playground for the physicist, it soon found applications in chemistry, creating the so call Quantum Chemistry. Over the years quantum chemistry has produced important tools for chemists in order to calculate, understand and predict molecular properties.

Quantum Theory is unfortunately limited to systems with a small number of atoms, because this methodology treats the electrons explicity, which is computationally very expensive. For large molecules like proteins, the description of energetics of the system relies on simpler methods, namely molecular mechanics. These methods are based on the principles of the classical instead of quantum mechanics, treating the position of the atoms and not of the electrons. This methodology will be explain in Section 3.3.

In principle, all the amenable information of a system, may be obtained from the arrangement of its nuclei (protons and neutrons) and electrons. The Quantum Theory describes this arrangement with a mathematical function called the *wavefunction*, Ψ. In order to obtain it, the **Schrödinger** equation needs to be solved,

$$\widehat{H}\Psi = E\Psi \tag{3.1}$$

where $\widehat{H}$ denotes the Hamiltonian operator, which, when applied to the wavefunction, Ψ, outputs the energy of the system, E, times the wavefuntion itself. For a system of N_e electrons and N_N nuclei, the Hamiltonian in atomic

units would be written as in Equation 3.2, aside from relativistic effects [1] and spin-orbit coupling [2] .

$$\begin{aligned} \hat{H} &= \hat{T}_e + \hat{T}_N + \hat{V}_{e-e} + \hat{V}_{e-N} + \hat{V}_{N-N} \\ &= -\sum_{i}^{N_e} \frac{\nabla_i^2}{2} - \sum_{A}^{N_N} \frac{1}{2M_A} \nabla_A^2 \\ &\quad - \sum_{A}^{N_N} \sum_{i}^{N_e} \frac{Z_A}{r_{iA}} + \sum_{i>j}^{N_e} \frac{1}{r_{ij}} + \sum_{A>B}^{N_N} \frac{Z_A Z_B}{R_{AB}} \end{aligned} \tag{3.2}$$

where $\hat{T}_e$ and $\hat{T}_N$ represent the kinetic energy of each electron and each nucleus, respectively; $\hat{V}_{e-e}$, $\hat{V}_{e-N}$ and $\hat{V}_{N-N}$ represent the electron-electron, electron-nucleus and nucleus-nucleus interaction terms, respectively; M_A is the mass of nucleus A, Z_A and Z_B are the atomic number of nucleus A and B; and $R_{AB} = |\vec{R}_B - \vec{R}_A|$, $r_{ij} = |\vec{r}_j - \vec{r}_i|$ and $r_{iA} = |\vec{R}_A - \vec{r}_i|$ are the distance between nuclei A and B, electrons i and j, and electron i and nucleus A, respectively.

Equation 3.2 can be simplified supposing that, since nuclei are so heavy with respect to electrons, their movement will be much slower, and hence the nuclear kinetic term, $\hat{T}_N$, can be neglected. As a consequence, the nucleus-nucleus potential energy term, $\hat{V}_{N-N}$, can be regarded as constant, due to the fact that it only depends on internuclear distances, which are fixed in the approximation. This is referred to as the **Born-Oppenheimer approximation**,[103] and results in the approximate Hamiltonian shown in Equation 3.3.

$$\begin{aligned} \hat{H}_e &= \hat{T}_e + \hat{V}_{e-e} + \hat{V}_{e-N} + \hat{V}_{N-N} \\ &= -\sum_{i}^{N_e} \frac{\nabla_i^2}{2} - \sum_{A}^{N_N} \sum_{i}^{N_e} \frac{Z_A}{r_{iA}} + \sum_{i>j}^{N_e} \frac{1}{r_{ij}} + \sum_{A>B}^{N_N} \frac{Z_A Z_B}{R_{AB}} \end{aligned} \tag{3.3}$$

Here $\hat{H}_e$ is the electronic Hamiltonian operator. Therefore, the electronic wavefunction, Ψ_e, depends explicity on the electronic coordinates, and parametrically on the position of the nuclei. It is worth mentioning that for any arrangement of nuclei (that is, each molecular geometry), one obtains

[1] The Relativistic Effects are the change in the inherent properties of the mass, when it moves at a speed approaching the speed of light. For the quantum description of the heaviest atoms in the periodic table, it becomes crucial.

[2] Magnetic interaction between the spin of the electron and the angular momentum of the orbital.

an electronic energy, and also a nucleus-nucleus interaction. This builds up an effective $3N_N$-dimensional energy potential (for N_N nuclei), the topology of which will provide the energies of different conformations, and the paths through which two energetically local minima structures are connected. This effective potential is often called **Potential Energy Surface** (PES hereafter). The PES of a set of atoms is a vital feature that provides the information needed to discern the mechanisms of chemical transformations, and their energetics.

Many mathematical methods have been devised to provide a way of finding the electronic wavefunction for any given atomic arrangement. One of the first developed molecular orbital method was the Hartree-Fock method (HF hereafter),[104,105] where the wavefunction is an antisymmetrized product of one-electron orbitals. The electrons are treated as moving in a mean field due to the nucleus and the remaining electrons. The main drawback of this method is that correlation of electrons with opposite spins is neglected. There are different ways in which this correlation can be taken into account. One of them are the perturbational methods such as n^{th} order Møller-Plesset theory (MPn hereafter).[106] In these methods the electron correlation is treated as a perturbation of the HF problem. In the Configuration-Interaction method (CI hereafter)[107,108] the wavefunction is expressed as a linear combination of configurations to provide a better variational solution to the exact many-electron wavefunction. There are other more sophisticated methods such as Coupled Cluster (CC hereafter)[109–111], Multi-Reference Configuration Interaction (MRCI hereafter) or Complete-Active-Space Self Consistent Field (CASSCF hereafter) methods,[112,113] which are very useful tools to study electronic properties of both ground and excited states. These methods are called *ab initio* methods, because they calculate the wavefunction from first principles, according to the raw Quantum Theory.

A more modern approach, and highly promising, is the Density Functional Theory (DFT hereafter).[114–117] Although not strictly *ab initio*, it is nonetheless very closely related to the HF method (simplest of *ab initio* methods), but modified in such a way that its accuracy is substantially improved.

Other approach is the **Semiempirical** methods, which are much faster but less accurate than any *ab initio* or DFT calculations. They are less rigorous in their treatment of the underlying physics.

3.2 Semiempirical Methods

The Semiempirical methods are a simplification of the HF method, to beef up its performance. Only the valence electrons are explicity considered, the core electrons are accounted for by reducing the nuclear charge or introducing functions to model the combined repulsion due to the nuclei and core

electrons. Furthermore, only a minimum basis set, the minimum number of functions necessary for accommodating the electrons in the neutral atom, is used for the valence electrons. *Slater type orbitals* (STO), exponential functions, are used as basis functions,

$$\chi_{\zeta,n,l,m}(r,\theta,\varphi) = NY_l^m(\theta,\varphi)r^{n-l}e^{-\zeta r} \tag{3.4}$$

where N is a normalization constant, Y_l^m are the spherical harmonic functions, n, m and l are quantum numbers, principal, magnetic and azimuthal respectively, θ is the polar angle and φ is the azimuthal angle, r is the distance of the electron from the nucleus and ζ is a constant related to the effective charge of the nucleus. The exponential dependence ensures a fairly rapid convergence with increasing number of functions.

The semiempirical methods make use of the Zero Differential Overlap (ZDO) approximation, in which integrals containing the product $\chi_\mu^A(1)\chi_\lambda^B(1)$ where $\mu \neq \lambda$ are neglected (Equation 3.6). Therefore, certain two-electron repulsion integrals (see Equation 3.5) are ignored.

$$\langle \mu_A \nu_B | \lambda_C \sigma_D \rangle = \int\int \chi_\mu^A(1)\chi_\lambda^C(1)\frac{1}{r_{12}}\chi_\nu^B(2)\chi_\sigma^D(2)d\tau d\tau \tag{3.5}$$

$$\langle \mu_A \nu_B | \lambda_C \sigma_D \rangle = \delta_{\mu\lambda}\delta_{\nu\sigma}\langle \mu_A \nu_B | \mu_A \nu_B \rangle \tag{3.6}$$

$$\text{where } \delta_{\mu\lambda} = \begin{cases} 1 & \mu = \lambda \\ 0 & \mu \neq \lambda \end{cases}$$

Here μ, ν, λ and σ denote the basis functions, and A, B, C and D are the atoms where the basis functions are centered on. One of the main consequences of this assumption, is that all two-electron integrals involving three- and four-center charge distributions are neglected. The remaining integrals are made into parameters, which are fitted from experimental results, or higher level calculations. This is why these methods are called Semiempirical. Different semiempirical methods exist depending on how many integrals are neglected and how the parametrization is done.

Neglect of Diatomic Differential Overlap Approximation (NDDO)

All the terms arising from the overlap of two atomic orbitals which are on different centers are set to zero. The approximation is defined by the following equations:

Overlap integral and one-electron operator:

$$S_{\mu\nu} = \langle \mu_A | \nu_B \rangle = \delta_{\mu\nu}\delta_{AB} \tag{3.7}$$

$$\hat{h} = -\frac{1}{2}\nabla^2 - \sum_A^{N_N} \frac{Z'_A}{|\vec{R}_A - \vec{r}|} = -\frac{1}{2}\nabla^2 - \sum_A^{N_N} \mathbf{V}_A \tag{3.8}$$

Here Z_a' denotes that the nuclear charge has been reduced by the number of core electrons.

One-electron integrals:

$$\begin{aligned}
\langle\mu_A|\hat{h}|\nu_A\rangle &= \delta_{\mu\nu}\langle\mu_A| - \frac{1}{2}\nabla^2 - \mathbf{V}_A|\mu_A\rangle - \sum_{a\neq A}^{N_N}\langle\mu_A|\mathbf{V}_a|\nu_A\rangle \\
\langle\mu_A|\hat{h}|\nu_B\rangle &= \langle\mu_A| - \frac{1}{2}\nabla^2 - \mathbf{V}_A - \mathbf{V}_B|\nu_B\rangle \qquad (3.9)\\
\langle\mu_A|\mathbf{V}_C|\nu_B\rangle &= 0
\end{aligned}$$

In the case of the two-electron integrals, all the products of basis functions depending on the same electron when located at different atoms are neglected.

$$\langle\mu_A\nu_B|\lambda_C\sigma_D\rangle = \delta_{AC}\delta_{BD}\langle\mu_A\nu_B|\lambda_A\sigma_B\rangle \qquad (3.10)$$

Intermediate Neglect of Differential Overlap Approximation (INDO)

In addition to the NDDO approximations, INDO neglects all two-center two-electron integrals which are not of the Coulomb type. Furthermore, the total energy is independent of a rotation of the coordinate system, some of the integrals are independent of the orbital type. Therefore, one-electron integrals involving two different functions on the same atom and a $\mathbf{V}_a$ operator from another atom disappear.

One-electron integrals:

$$\langle\mu_A|\hat{h}|\nu_A\rangle = -\delta_{\mu\nu}\sum_{a}^{N_N}\langle\mu_A|\mathbf{V}_a|\mu_A\rangle \qquad (3.11)$$

Two-electron integrals:

$$\langle\mu_A\nu_B|\lambda_C\sigma_D\rangle = \delta_{\mu_A\lambda_C}\delta_{\nu_B\sigma_D}\langle\mu_A\nu_B|\mu_A\nu_B\rangle \qquad (3.12)$$

Those two terms ($\langle\mu_A|\mathbf{V}_a|\mu_A\rangle$ and $\langle\mu_A\nu_B|\mu_A\nu_B\rangle$) are independent to the orbital type, s or p. Moreover, the two-electron integrals are reduced to just two parameters.

$$\begin{aligned}
\langle\mu_A\nu_A|\mu_A\nu_A\rangle &= \langle\mu_A\mu_A|\mu_A\mu_A\rangle = \gamma_{AA} \\
\langle\mu_A\nu_B|\mu_A\nu_B\rangle &= \gamma_{AB} \qquad (3.13)
\end{aligned}$$

Complete Neglect of Differential Overlap Approximation (CNDO)

In this approximation only the Coulomb one-center and two-center two-electron integrals remain.

Modified methods used molecular experimental data such as geometries, heats of formation, dipole moments and ionization potentials for fitting the parameters. In the next subsection we are going to explain a bit the Modified NDDO Models, due to the fact that we have employed one of them in our calculations.

3.2.1 Modified NDDO Models

MNDO, AM1 and PM3 are derived from the same approximation, NDDO, and differ only in the way the core-core repulsion is treated, and how the parameters are assigned. Those methods only consider s- and p-functions, which are taken as Slater Type Orbitals (STO) with corresponding exponents, ζ_s and ζ_p (Equation 3.4). The parametrization is done in terms of atomic variables, only the nature of a single atom is taking into account.

The *one-center one-electron integrals* have a value corresponding to the energy of a single electron experiencing the full nuclear charge (U_s or U_p) plus terms from the potential due to all the other nuclei in the system (first Equation in 3.9).

$$\begin{aligned} h_{\mu\nu} &= \langle \mu_A | \hat{h} | \nu_A \rangle = \delta_{\mu\nu} U_\mu - \sum_{a \neq A}^{N_N} Z_a' \langle \mu_A s_a | \nu_A s_a \rangle \\ U_\mu &= \langle \mu_A | -\frac{1}{2}\nabla^2 - \mathbf{V}_A | \mu_A \rangle \end{aligned} \tag{3.14}$$

Here μ_A and ν_A are s- or p-symmetry orbitals centered on atom A, while s_a is a s-type function centered on an atom different to A.

The *two-center one-electron integrals* given by the second Equation in 3.9 are written as a product of the corresponding overlap integral times the average of the two atomic "resonance" parameters, β.

$$\begin{aligned} \langle \mu_A | \hat{h} | \nu_B \rangle &= \langle \mu_A | -\frac{1}{2}\nabla^2 - \mathbf{V}_A - \mathbf{V}_B | \nu_B \rangle \\ &= S_{\mu\nu} \frac{1}{2}(\beta_\mu + \beta_\nu) \\ S_{\mu\nu} &= \langle \mu_A | \nu_B \rangle \end{aligned} \tag{3.15}$$

There are only five types of *one-center two-electron integrals* surviving the NDDO approximation (Equation 3.10)

$$\begin{aligned}
\langle ss|ss\rangle &= G_{ss} \\
\langle sp|sp\rangle &= G_{sp} \\
\langle ss|pp\rangle &= H_{sp} \\
\langle pp|pp\rangle &= G_{pp} \\
\langle pp'|pp'\rangle &= G_{p2}
\end{aligned}$$

The G-type parameters are Coulomb terms, while H parameter is an exchange integral. The G_{p2} integral involves two different types of p-functions (i.e. p_x, p_y or p_z).

On the other hand, there are a total of 22 different *two-center two-electron integrals* arising from a sp-basis, which are modelled as interactions between multipoles (Table 3.1). They represent the energy of an electron density distribution, electron 1, arising from the product of the first two atomic orbitals interacting with the electron density distribution of electron 2, which arises from the product of the second two atomic orbitals (Table 3.2).

Table 3.1: Different types of electron density distribution

Atomic Orbitals	**Multipole distribution**	**# charges**	**Charge**
$\langle ss\|$	Monopole	1	-1 centered on the nucleus
$\langle sp\|$	Dipole	2	+1/2 located at position (x,y,z) -1/2 located at position (-x,-y,-z)
$\langle pp\|$	Monopole + Linear Quadrupole	3	+1/2 located at the nucleus -1/4 at positions (x,y,z) and at (-x,-y,-z)
$\langle pp'\|$	Square Quadrupole	4	+1/4, -1/4, +1/4 and -1/4 forming a square centered on the nucleus

Each of the MNDO, AM1 and PM3 methods involves at least 12 parameters per atom: orbital exponents, $\zeta_{s/p}$; one-electron terms, $U_{s/p}$ and $\beta_{s/p}$; two-electron terms, $G_{ss}, G_{sp}, G_{pp}, G_{p2}, H_{sp}$; parameters used in the core-core repulsion, α; and for the AM1 and PM3 methods a, b and c constants.

Modified Neglect of Diatomic Overlap (MNDO)

The core-core repulsion has the next form in the MNDO model,[118,119]

$$\begin{aligned}
E_N^{MNDO}(A,B) &= Z_A' Z_B' \langle s_A s_B | s_A s_B \rangle \\
&\times (1 + e^{-\alpha_A R_{AB}} + e^{-\alpha_B R_{AB}})
\end{aligned} \tag{3.16}$$

1	$\langle ss\|ss\rangle$	12	$\langle sp_\sigma\|p_\pi p_\pi\rangle$
2	$\langle ss\|p_\pi p_\pi\rangle$	13	$\langle sp_\sigma\|p_\sigma p_\sigma\rangle$
3	$\langle ss\|p_\sigma p_\sigma\rangle$	14	$\langle ss\|sp_\sigma\rangle$
4	$\langle p_\pi p_\pi\|ss\rangle$	15	$\langle p_\pi p_\pi\|sp_\sigma\rangle$
5	$\langle p_\sigma p_\sigma\|ss\rangle$	16	$\langle p_\sigma p_\sigma\|sp_\sigma\rangle$
6	$\langle p_\pi p_\pi\|p_\pi p_\pi\rangle$	17	$\langle sp_\pi\|sp_\pi\rangle$
7	$\langle p_\pi p_\pi\|p_\pi^{'} p_\pi^{'}\rangle$	18	$\langle sp_\sigma\|sp_\sigma\rangle$
8	$\langle p_\pi p_\pi\|p_\sigma p_\sigma\rangle$	19	$\langle sp_\pi\|p_\pi p_\sigma\rangle$
9	$\langle p_\sigma p_\sigma\|p_\pi p_\pi\rangle$	20	$\langle p_\pi p_\sigma\|sp_\pi\rangle$
10	$\langle p_\sigma p_\sigma\|p_\sigma p_\sigma\rangle$	21	$\langle p_\pi p_\sigma\|p_\pi p_\sigma\rangle$
11	$\langle sp_\sigma\|ss\rangle$	22	$\langle p_\pi p_\pi^{'}\|p_\pi p_\pi^{'}\rangle$

Table 3.2: The 22 different two-electron two-center integrals arising from a sp-basis

where the α exponents are taken as fitting parameters. They account for decreased screening of nucleus by the electrons at small interatomic distances. $\langle s_A s_B|s_A s_B\rangle$ is a Coulomb repulsion integral between a s-symmetry orbital centered in A and a s-symmetry orbital centered on B. For O–H and N–H bonds, a modified form of the screening term is used,

$$E_N(A,H) = Z_A^{'} Z_H^{'} \langle s_A s_H|s_A s_H\rangle \times\Big(1+\frac{e^{-\alpha_A R_{AH}}}{R_{AH}}+e^{-\alpha_H R_{AH}}\Big) \tag{3.17}$$

In certain distances, such as noncovalent intermolecular interactions (i.e. hydrogen bonds), the MNDO model is problematic.

Austin Model 1 (AM1)

The AM1 model[120] includes a set of Gaussian core-core terms that alleviate the excessive repulsion just outside the bonding distances, which was observed by MNDO model. The core-core repulsion has the following form,

$$E_N(A,B) = E_N^{MNDO}(A,B) + \frac{Z_A^{'} Z_B^{'}}{R_{AB}} \times\left(\sum_k a_{kA} e^{-b_{kA}(R_{AB}-c_{kA})^2} + \sum_k a_{kB} e^{-b_{kB}(R_{AB}-c_{kB})^2}\right) \tag{3.18}$$

where k is between 2 and 4 depending on the atom (different number of Gaussians is used for each atom), and a_k, b_k and c_k are constants fitted to

molecular data. The Gaussian core-core terms are empirical adjustments to the potential, devoid of rigorous physical meaning. However, a much better description of hydrogen bonds and proton-transfer reactions are done with them.

Parametric Method Number 3 (PM3)

In the PM3 method[121,122] the AM1 expression for the core-core repulsion is kept, except that only 2 Gaussian are assigned to each atom.

AM1 and PM3 models offer a significant improvement for hydrogen bonding relative to MNDO but have the problem that they over-stabilize hypervalent structures because of the artificially attractive core-core interactions.

In addition, with only s- and p-functions included, the MNDO/AM1/PM3 methods are unable to treat a large part of the periodic table. Furthermore, it is known that d-orbitals significantly improve the results for compounds involving second row elements, especially hypervalent species as a pentavalent phosphorane. The main problem is the increase of two-electron integrals: one-center two-electron integrals raises from 5 in a sp-basis to 17 in a spd-basis; while two-center two-electron integrals raises from 22 to 491 when d-functions are included.

Thiel and Voityuk[123,124] constructed a new model called MNDO/d, which includes d-orbitals in the MNDO model. Nevertheless, it retains the same problem as MNDO for modeling hydrogen bonds and proton-transfer reactions. Some years later, Lopez and York[125] developed a new Hamiltonian, AM1/d, that departs from d-orbitals formalism of MNDO/d and introduces the Gaussian core-core terms as in AM1 and PM3, however, the over-stabilization of the hypervalent structures continue. This Hamiltonian was reparametrizated for phosphoryl transfer reactions for H, O and P atoms some years later by Nam et al., **AM1/d-PhoT**.[97] They kept the core-core interactions for hydrogen bonding but turn those interactions off for phosphorus bonding where the d-orbitals allow proper hybridization and accurate representation of hypervalent species. A scale factor was introduced into the Gaussian core-core terms in the AM1/d-PhoT model,

$$E_N(A,B) = E_N^{MNDO}(A,B) + \frac{Z'_A Z'_B}{R_{AB}} G^A_{scale} G^B_{scale} \times \left(\sum_k a_{kA} e^{-b_{kA}(R_{AB}-c_{kA})^2} + \sum_k a_{kB} e^{-b_{kB}(R_{AB}-c_{kB})^2} \right) \tag{3.19}$$

where G^A_{scale} and G^B_{scale} are scaling parameters for atoms A and B, which vary from zero to one (values of 0 recover the conventional MNDO core-core model, whereas values of 1 recover AM1 core-core model). The scaling

factor provide flexibility to attenuate or even shut off Gaussian core-core interactions between certain atoms and offers a simple mechanism for interconverting between AM1-like model and MNDO-like models.

3.3 Molecular Mechanics

As mentioned before, for large biological molecules like enzymes, the description of energetics of the system relies on methods simpler than *ab initio*, DFT or semiempirical methods, namely molecular mechanics (MM hereafter). In these methods, the electronic structure is not explicitly calculated, but it is implicitly built up in the model by a set of fixed parameters (atomic charges, van der Waals parameters, force constants...), carefully obtained to reproduce experimental or *ab initio* data. Molecular mechanical methods provide an analytical form for the potential energy of a given system as a function of the atomic coordinates, which can be efficiently computed and, therefore, opens the possibility to treat much bigger systems than those attainable by the *ab initio* or DFT methods.

3.3.1 Force Fields

The specific form of the analytical function and the corresponding set of parameters used in Molecular Mechanics are referred to as the force field. A common pictorial representation of a force field is one in which atoms are represented by spheres and the covalent bonds by springs, ball and spring model. The value of the energy is calculated as a sum of several terms: *i)* internal or bonded terms (E_{bond}), and *ii)* a sum of external or non-bonded terms ($E_{non-bond}$). The former describes the covalent bonds, the angular deformations between covalently bound atoms and torsional energies in a molecule. The latter accounts for the interactions between non-covalently bonded atoms or atoms separated by three or more covalent bonds. In the molecular mechanics program used in this work, the CHARMM biomolecular program[126,127] (Chemistry at HARvard Molecular Mechanics), the potential energy function ($U(\vec{r}^N)$) is a function of the positions ($\vec{r}$) of the N atoms of the system, and can be written as in Equation 3.20:

$$
\begin{aligned}
U(\vec{r}^N) &= E_{bond} + E_{non-bonded} \\
U(\vec{r}^N) &= \sum_{\text{bonds}} K_b \left(b - b_0\right)^2 \\
&\quad + \sum_{\text{angles}} K_\theta \left(\theta - \theta_0\right)^2 + \sum_{\text{Urey-Bradley}} K_{UB} \left(S - S_0\right)^2 \\
&\quad + \sum_{\text{dihedrals}} K_\varphi \left(1 + cos(n\varphi - \delta)\right) + \sum_{\text{impropers}} K_\omega \left(\omega - \omega_0\right)^2
\end{aligned}
$$

$$+ \sum_{\text{residues}} U_{CMAP}(\varphi, \psi) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(\underbrace{4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right]}_{\text{Lennard-Jones Potential}} + \underbrace{\frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}}_{\text{Coulomb Potential}} \right) \tag{3.20}$$

Among the eight terms showed in Equation 3.20, the first six correspond to the so-called *"bonded, internal"* terms, whereas the last two terms correspond to the so-called *"non-bonded"* terms. All of them are explained below.

Bonded terms

The first term accounts for the change in the potential energy when the distance between covalently bonded atom pairs is changed. Since it is modelled through a harmonic potential, if the distance deviates b from the equilibrium bond length b_0, a quadratic penalty is introduced. The magnitude of this penalty is controlled through the K_b force constant, specific for each bond type, and determines its flexibility or rigidity.

The angle term, describes the changes in energy upon variations in the angle θ formed by three atoms, $A - B - C$, where A and C are covalently bonded to B. Since this term is also modelled by a harmonic potential, it is required the definition of an equilibrium angle θ_0, and K_θ is the force constant associated to a specific angular term.

The third term, or Urey-Bradley (UB hereafter) term, is also associated with the angles. It is used to optimize the fit to vibrational spectra and out of plane motions. UB term takes into account the harmonic stretching of the $A \cdots C$ distance S, with a K_{UB} force constant and S_0 equilibrium distance between atoms A and C. The UB term is optional, it is only used in special cases.

The fourth term describes the energy required for a torsion around a bond. This potential is periodic, it has a sinusoidal expression, K_φ being the barrier controlling the amplitude of the torsional motion, n the periodicity of the dihedral angle and δ the phase shift that determines the values of the angles at which the energy vanishes.

Improper dihedrals constitute the fifth term. They are used for the description of out of plane motions. ω is the improper dihedral angle, ω_0 the equilibrium angle and K_ω the force constant assigned to a particular improper dihedral angle.

The sixth term, is a numerical correction to the dihedral angle of the main chain, called CMAP.[128] It corrects certain systematic errors in the description of the protein backbone.

Non-bonded terms

The energy term representing the contribution of non-bonded interactions in the CHARMM potential function has two components, the Coulombic interactions term (electrostatic interactions) between the point charges (q_i and q_j), and the Lennard-Jones 12-6 (LJ hereafter) term, which is used for the treatment of the core-core repulsion and the attractive van der Waals dispersion interaction. Although some other potential functions also may be used in CHARMM.

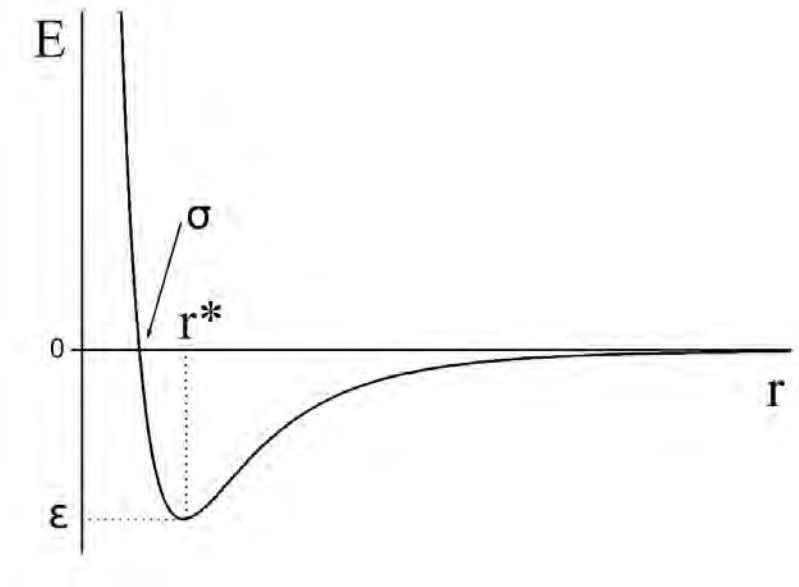


Figure 3.1: Lennard-Jones 12-6 potential. $r^\star$ is the optimal separation of atoms, ϵ is the value of the energy at the minimum (the well depth), and σ is the separation for which the energy is zero (the collision parameter).

The van der Waals interactions are one of the most important forces for the stability of the biological macromolecules, and CHARMM computes it according to the seventh term in Equation 3.20. The van der Waals interaction is most often modeled using the Lennard-Jones 12-6 potential which expresses the interaction energy using the atom-type dependent constants, the collision parameter σ (the separation for which the energy is zero) and the well depth ϵ[129] (see Figure 3.1). This interaction between two atoms arises from a balance between repulsive and attractive forces. The repulsive force dominates at short distances (nucleus-nucleus repulsion, electron-electron Pauli repulsion, etc.), while the attractive forces are long-range. The attractive forces can be considered as arising from fluctuations in the charge distribution in the electron clouds. The fluctuation in the electron distribution on one atom or molecule gives rise to an instantaneous dipole which, in turn, induces a dipole in a second atom or molecule giving rise to

an attractive interaction. Each of these two effects is equal to zero at infinite atomic separation r_{ij} and become significant as the distance decreases. The attractive interaction is longer range than the repulsion but as the distance becomes shorter, the repulsive interaction becomes dominant. This gives rise to a minimum in the energy, the well depth. The value of the energy at the minimum (ϵ), and the optimal separation of atoms ($r^\star$), which is roughly equal to the sum of van der Waals radii of the atoms, depend on the atom chemical type.

The electrostatic interactions between a pair of atoms is represented by the Coulomb potential, as shown in the eighth term of Equation 3.20; $4\pi\epsilon_0$ is the term for the effective dielectric moment of the medium, r_{ij} is the distance between two atoms having charges q_i and q_j and ϵ_0 is the dielectric constant *in vacuo*.[130]

3.3.2 Treatment of the Non-bonded Interactions

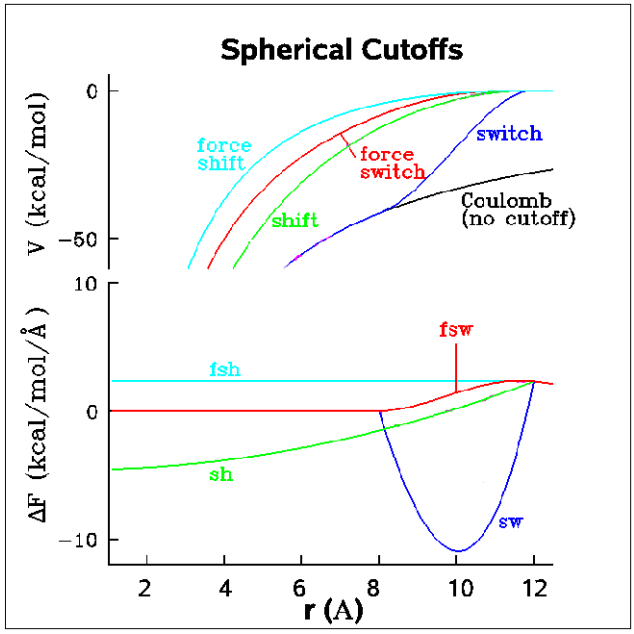
The most time consuming part of an energy evaluation is the calculation of the non-bonded terms in the potential energy function. In principle, the non-bonded energy terms between every pair of atoms should be evaluated; in this case, the number of computation steps increases as the square of the number of atoms for a pairwise model (N^2). To speed up the computation, the interactions between two atoms separated by a distance greater than a predefined distance, the cutoff distance r_{off}, are ignored, and are therefore equal to cero. Several different ways to terminate the interaction have been developed over the years, depending on whether the distance is calculated between the interacting atoms (atom-based) or between two groups of atoms (group-based). Furthermore, the interaction energy or force can be truncated abruptly at the cutoff distance, *truncation method*,[131] or some kind of smoothing scheme can be applied, *spherical truncation methods*.[132] The *truncation method* leads to large fluctuations in the energy and is therefore not often used. However, *spherical truncation methods* are a better option, which include energy shifting and switching as well as force shifting and switching approaches.[131,132] The Shift function modifies the entire potential surface such that at the cutoff distance, r_{off}, the interaction potential is zero (Figure 3.2), while the Switch function tapers over a predefined range of distances.

There are several methods to go beyond the cutoff scheme in order to handle all long-range electrostatic interactions. These methods include the Ewald summation method,[133–138] the Extended Electrostatics approach[139] or Fast Multipole Methods.[140,141]

Ewald summation

Ewald[133] developed a method which transforms the conditionally and slowly convergent electrostatic energy term lattice sum over all pair interactions

Figure 3.2: The effects of the spherical cutoff functions on the electrostatic potential.



and over all lattice vectors if Periodic Boundary Conditions (PBC hereafter) are imposed, Equation 3.21, into the summation of a smoothly varying long-range term and a short-range term, which are more complicated but absolutely and rapidly convergent sums, plus a "self-energy" and a "dipole" term, Equation 3.22 (see Section 3.3.6 for PBC information). In the Ewald method, each charge is considered to be surronded by a neutralising charge Gaussian distribution of equal magnitude but opposite sign.

$$U(\vec{r}^N) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\vec{m}}{}' \frac{q_i q_j}{4\pi\epsilon_0 |r_{ij}+\vec{m}|} \tag{3.21}$$

$$U(\vec{r}^N) = \underbrace{\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\pi V}\frac{q_i q_j}{4\pi\epsilon_0}\sum_{\vec{k}\neq 0}\frac{e^{(-\pi^2 k^2/\kappa^2)}}{k^2}e^{(2\pi \mathbf{I}\vec{k}r_{ij})}}_{\text{Reciprocal space}} + \underbrace{\sum_{i=1}^{N}\sum_{j=i+1}^{N}\sum_{\vec{m}=0}^{\infty}\frac{q_i q_j}{4\pi\epsilon_0}\frac{\mathrm{erfc}(\kappa|r_{ij}+\vec{m}|)}{|r_{ij}+\vec{m}|}}_{\text{Direct space}} - \frac{\kappa}{\sqrt{\pi}}\sum_{i=1}^{N}\frac{q_i^2}{4\pi\epsilon_0} + \frac{2\pi}{3V}\left|\sum_{i=1}^{N}\frac{q_i}{4\pi\epsilon_0}r_i\right|^2 \tag{3.22}$$

where $\vec{m}$ is the lattice vector (real space) of the periodic array of unit cells,

the prime on the sum of Equation 3.21 indicates that $j \neq i$ when $\vec{m} = 0$, erfc is the complementary error function $\left(\text{erfc}(x) = 2/\sqrt{\pi}\int_x^\infty e^{-t^2}dt\right)$, $\vec{k}$ is the reciprocal space lattice vector, V is the volume of the unit cell, I is the imaginary unit, and κ is a constant (determines the width of the Gaussian distribution) which adjusts the relative rates of convergence of the real and reciprocal space sums. The former converges more rapidly for large κ, whereas the latter converges more rapidly for small κ. It is usually chosen so as to optimize the balance of accuracy and efficiency of the calculations.

The first term of Equation 3.22 is a reciprocal space sum over all pairwise interactions in the infinitive lattice. The second term is a direct sum over all short-range pairs. It is a sum of the interactions between the point-charges and the neutralising distributions. The third term of the equation, the self-energy term, cancels the charge distribution for each Gaussian in the unit cell interacting with itself. Finally, the fourth term, the dipole term, must be included if the surrounding medium is vacuum.

Variants on the Ewald method like the employment of pairwise cutoff lists for direct sum, charges on grids, and fast Fourier transforms, greatly enhance the computational performance. Part of the work present in this thesis has been carried out applying Particle Mesh Ewald method[135–138] (PME hereafter). This method introduced a fast $N \cdot log(N)$ algorithm for computing the Ewald sums.

3.3.3 Limitations of the Empirical Potential Energy Function

In molecular mechanics the electronic structure is not explicitly taken into account, which leads to certain limitations. For instance, since covalent bonds are represented as harmonic springs, the creation or breaking of covalent bonds is not possible. Another limitation is due to the fixed set of atom types employed when determining the parameters for the force field. Atom types are used to define an atom in a particular bonding situation, for example an aliphatic carbon atom in a sp3 bonding situation has different properties than a carbon atom found in the Histidine (His hereafter) ring. Instead of presenting each atom in the molecule as an unique one described by unique set of parameters, there is a certain amount of grouping in order to minimize the number of atom types. This can lead to type-specific errors. The properties of certain atoms like aliphatic carbon or hydrogen atoms, are less sensitive to their surroundings and a single set of parameters may work quite well, while other atoms like oxygen, nitrogen and sulfur are much more influenced by their environment. These atoms require more types and parameters to account for the different bonding environments.

An approximation introduced to decrease the computational demand is the pairwise additive approximation, i.e., interaction energy between one atom and the rest of the system is calculated as a sum of pairwise (one

atom to one atom) interactions, or as if the pair of atoms are invisible to other atoms in the system (see Equation 3.20). The simultaneous interaction between three or more atoms is not calculated in most contemporary force fields, so polarization effects are not explicitly included in the force field. This can lead to subtle differences between calculated and experimental results, for example, in the calculation of experimentally observable pK_a shifts of ionizable amino acid residue side chains induced by the electrostatic field of the whole protein. The development of polarizable force fields is currently an active area of research.[142]

Another important point to take into consideration is that the energy based methods do not explicitly include entropic effects. Thus, a minimum value of energy does not necessarily correspond to the equilibrium, or the most probable state. Because of the fact that experiments are generally carried out under isothermal-isobaric conditions (constant number of particles, constant pressure and constant temperature, NPT) the equilibrium state corresponds to the minimum of Gibbs Free Energy, G. While just an energy calculation ignores entropic effects, these are included in a molecular dynamics simulations[129] (see Section 3.3.5).

Despite their limitations, current force fields are calibrated towards experimental results and quantum mechanical calculations of small model compounds. Force fields differ by the definition of their mathematical potential energy function and the values used for their associated parameters, such as: b_0, K_b, θ_0, K_θ, K_φ, n, δ, ω_0 and K_ω (see Equation 3.20). Among the most commonly used potential energy functions are CHARMM (Chemistry at HARvard Macromolecular Mechanics),[143–146] AMBER (Assisted Model Building with Energy Refinement),[147–150] GROMOS (GROningen MOlecular Simulation),[151–153] OPLS (Optimized Potentials for Liquid Simulations)[154,155] and MMFF (Merck Molecular Force Fiel)[156] force fields. The ability of these force fields to reproduce physical properties measurable by experiment has been critically tested. The properties include structural data obtained from X-ray crystallography and NMR, dynamic data obtained from spectroscopy, inelastic neutron scattering and thermodynamic data. Force fields are also developed and adapted to different types of systems; there are for example all atom force fields for proteins,[144] nucleic acids,[143] organic molecules[157] and lipids[144] developed for the CHARMM biomolecular program,[126,127] and similarly for the other force fields. The continuing development of force fields remains an intense area of research with implications for both fundamental research as well as for applied research in the pharmaceutical industry.

3.3.4 Energy Minimization

The potential energy plotted as a function of the atomic coordinates yields the potential energy surface (PES). For a particular molecule, the energy

minima on this surface correspond to stable conformations. Knowledge of the potential energy surface combined with determining the relative populations in the various energy minima provides an understanding of the energetics and eventually the thermodynamics of the molecular system. Sampling the PES can be done by several different techniques.

For a given molecule one would like to find low energy conformations. For most molecules, except for very simple ones, the PES is a complicated multi-dimensional function that depends on the coordinates of the N atoms of the system and to find the global potential energy minimum would require a search of the entire potential energy surface. Another fundamental point on the PES is the lowest-energy *saddle point* connecting two local minima. The saddle point is the highest point on a lowest total energy pathway connecting these minima. It has the property that is a maximum with respect to one degree of freedom and is a minimum with respect to all other degrees of freedom. The significance of the saddle point is that it is an approximate dynamic bottleneck, a point of no return, for a transformation from the vicinity of one local minimum or valleys to another. The probability, rate, of reaction is very strongly influenced by the height of this saddle point. Both the minima and the saddle point are stationary points on the PES.

All energy minimization methods are used to find local minima, thus the structures found by this approach are always close to the initial structure.

There are several different algorithms for minimizing the energy of the system.[158–160] They all involve calculating the first derivative and possibly the second derivative of the potential energy (Equation 3.20), and using this information to guide the coordinates to a lower energy conformation.

One of the simplest minimization algorithms is the *steepest descent, SD,* method,[158,160] a first derivative method. The coordinates are adjusted at each step of this iterative method in the negative direction of the gradient $\left(-\frac{dE}{dx}\right)$ where E is the energy and x refers to the coordinates. The step size is an adjustable parameter which determines how far the coordinates are shifted at each step. If the energy decreases, the step size is increased by 20 % to accelerate the convergence. If the energy rises, a minimum has been neglected and the step size thus halved. Although the steepest descent method does not converge readily for complicated functions, it rapidly improves very poor conformations.[126,127]

Another first derivative method is the *conjugate gradient* technique which has better convergence characteristics[159] than SD. The method is iterative and remembers the previous minimization steps as well as the current gradient to determine the next step. The conjugate gradient technique requires less energy and gradient evaluations to achieve the same reduction in energy as the steepest descent method.

A second derivative method is the *Adopted-Basis set Newton-Raphson, ABNR,* technique[161] which provides better convergence to the local minimum

when there are no large gradient components.[126,127] This routine performs energy minimization using a Newton-Raphson algorithm applied to a subspace of the coordinate vector spanned by the displacement coordinates of the latest positions. The second derivative matrix is constructed numerically from the change in the gradient vectors, and is inverted by an eigenvector analysis allowing the routine to recognize and avoid saddle points in the energy surface. At each step the residual gradient vector is calculated and used to add a steepest descent step onto the Newton-Raphson step, incorporating new direction into the basis set. In most circumstances, this method is the best.

Typically, initial minimizations are done using the first derivative methods, usually with SD, and then are followed by second derivative methods.

3.3.5 Molecular Dynamics Simulations

One of the principal tools in the theoretical study of biological molecules is the method of molecular dynamics simulations (MD simulations hereafter). This computational method calculates the time dependent behavior of a molecular system. Like Monte Carlo methods,[162–165] MD provides a means to sample the potential energy surface of a molecule but in addition, this computational method calculates the time dependent behavior of the system. Classical MD simulations are now routinely used to investigate the structure, dynamics and thermodynamics of biological molecules and their complexes.[166] They are also used in the optimization of structures from X-ray crystallography and NMR experiments.

Molecular dynamics simulations are based on Newton's second law or equation of motion,

$$\vec{F} = m\vec{a} \tag{3.23}$$

where $\vec{F}$ is the force exerted on the particle, m is its mass and $\vec{a}$ is its acceleration. From a knowledge of the force on each atom, it is possible to determine the acceleration of each atom in the system. Integration of the equations of motion then yields a trajectory that describes the positions, velocities and accelerations of the particles as they vary with time. From this trajectory, the average values of properties can be determined. The method is deterministic; once the positions and velocities of each atom are known, the state of the system can be predicted at any time in the future or the past.

More precisely, the trajectory is obtained by solving the differential equations embodied in Newton's second law

$$\frac{F_{x_i}}{m_i} = \frac{d^2x_i}{dt^2} \tag{3.24}$$

This equation describes the motion of a particle of mass m_i along one coordinate (x_i) with F_{x_i} being the force on the particle in that direction. The

force can also be expressed as the gradient of the potential energy (U), as shown in Equation 3.25.

$$\vec{F_i} = -\vec{\nabla}_i U \tag{3.25}$$

Combining Equations 3.24 and 3.25, Equation 3.26 is obtained, where U is the potential energy of the system. Newton's equation of motion can then relate the derivative of the potential energy to the changes in position as a function of time.

$$\frac{-dU}{dx_i} = m_i \frac{d^2 x_i}{dt^2} \tag{3.26}$$

The initial distribution of velocities is usually determined from a random distribution with the magnitudes conforming to the required temperature and corrected so there is no overall momentum (P) as expressed in Equation 3.27.

$$P = \sum_{i=1}^{N} m_i v_i = 0 \tag{3.27}$$

Usually the velocities v_i are chosen randomly from a Maxwell-Boltzmann or Gaussian distribution at a given temperature, which gives the probability (p_{ix}) that an atom i has a velocity v_x in the x direction at a temperature T (Equation 3.28). Here k_B is the Boltzman constant.

$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_B T}\right)^{1/2} e^{\left(-\frac{1}{2}\frac{m_i v_{ix}^2}{k_B T}\right)} \tag{3.28}$$

The temperature can then be calculated if we consider that at thermal equilibrium, the average kinetic energy per degree of freedom can be expressed as in Equation 3.29, where v_α is the α component of the velocity of a given particle. We can use this relation to define an instantaneous temperature at time t, $T(t)$, as shown in Equation 3.30.

$$\left\langle \frac{1}{2} m v_\alpha^2 \right\rangle = \frac{1}{2} k_B T \tag{3.29}$$

$$T(t) = \sum_{i=1}^{N} \frac{m_i v_i^2(t)}{k_B N_f} \tag{3.30}$$

where N_f is the number of degrees of freedom, which is $3N - 3$ for a system of N particles with fixed total momentum. The relative fluctuations in the temperature will be of order $1/\sqrt{N_f}$. As N_f is typically on the order of $10^2 - 10^3$, the statistical fluctuations in the temperature are on the order of 5-10%.

Integration Algorithms

The potential energy is a function of the atomic positions (3N) of all the atoms in the system (N). Due to the complicated nature of this function, there is no analytical solution to the equations of motion, they must be solved numerically. Numerous numerical algorithms have been developed for integrating the equations of motion. The algorithm must conserve energy and momentum, should be computationally efficient and permit long integration time steps.[129,167,168] All integration algorithms assume the positions, velocities and accelerations can be approximated by a Taylor series expansion (Equations 3.31, 3.32 and 3.33):

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \frac{1}{3!}b(t)\delta t^3 + ... \quad (3.31)$$

$$v(t+\delta t) = v(t) + a(t)\delta t + \frac{1}{2}b(t)\delta t^2 + ... \quad (3.32)$$

$$a(t+\delta t) = a(t) + b(t)\delta t + ... \quad (3.33)$$

where r is the position, v is the velocity (the first derivative of r with respect to time), a is the acceleration (second derivative of r with respect to time), $b(t)$ is the third derivative and δt the time step in the Molecular Dynamics simulation. The algorithm used in the present work was mostly the *Verlet Leap-Frog* algorithm.

Verlet Algorithm: The *Verlet* algorithm[169,170] is derived as shown in Equations 3.34 and 3.35. When summing these two equations one gets Equation 3.36

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \frac{1}{3!}b(t)\delta t^3 + \mathcal{O}(\delta t^4) \quad (3.34)$$

$$r(t-\delta t) = r(t) - v(t)\delta t + \frac{1}{2}a(t)\delta t^2 - \frac{1}{3!}b(t)\delta t^3 + \mathcal{O}(\delta t^4) \quad (3.35)$$

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + a(t)\delta t^2 + \mathcal{O}(\delta t^4)$$

$$\approx 2r(t) - r(t-\delta t) + \frac{f(t)}{m}\delta t^2 + \mathcal{O}(\delta t^4) \quad (3.36)$$

The *Verlet* algorithm uses positions and accelerations at time t and the positions from time $t-\delta t$ to derive new positions at time $t+\delta t$, with an error that is of order δt^4. The Verlet algorithm uses no explicit velocities. The advantages of the Verlet algorithm are its straightforwardness and the modest store requirements. However, only moderate precision is obtained.

Leap-Frog Algorithm: In this algorithm,[171] the velocities are first calculated at time $t+1/2\delta t$ (Equation 3.37). They are then used to calculate

the positions r at time $t+\delta t$ (Equation 3.38). In this way, the velocities leap over the positions, then the positions leap over the velocities. The advantage of this algorithm is that the velocities are explicitly calculated, however, the disadvantage is that they are not calculated at the same time as the positions. The velocities at time t can be approximated by the relationship shown in Equation 3.39,

$$v(t+\frac{1}{2}\delta t) = v(t-\frac{1}{2}\delta t)+a(t)\delta t \tag{3.37}$$

$$r(t+\delta t) = r(t)+v(t+\frac{1}{2}\delta t)\delta t \tag{3.38}$$

$$v(t) = \frac{1}{2}[v(t-\frac{1}{2}\delta t)+v(t+\frac{1}{2})] \tag{3.39}$$

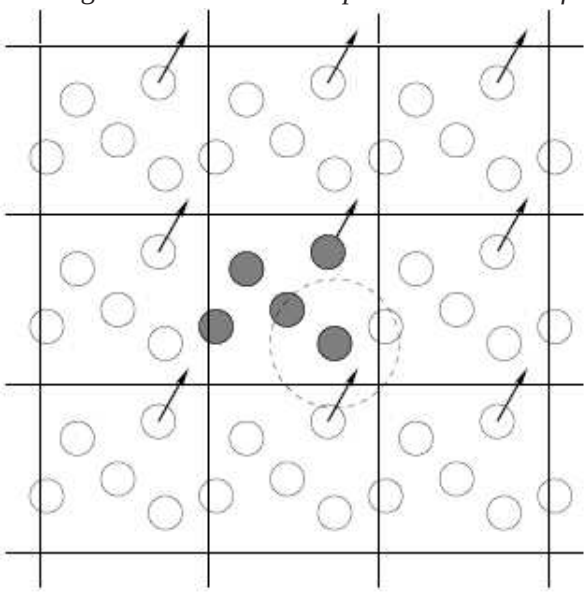
3.3.6 Boundary Conditions

A major obstacle for a simulation is the large fraction of molecules that lie on the surface of any small sample. For example, in the case of 1000 molecules arranged in a 10x10x10 $Å^3$ cube, 488 molecules are located at the cube faces,[167,168] and these will have a large effect on the measured properties. Molecules on the surface will experience completely different forces than those molecules inside the bulk.

Periodic Boundary Conditions

One way to overcome the problems linked to the finite arrangement in molecular dynamics simulations is the use of periodic boundary conditions. This method allows for the use of a relatively small number of particles in a simulation in such a way that the particles experience forces as though they were in a bulk solution. The central box is surrounded with replicas of itself. See, for example, the two dimensional box in Figure 3.3. The coordinates of the image particles, those found in the surrounding box are related to those in the primary box by simple translations. The box is replicated throughout space infinitely. Forces on primary particles are calculated from particles within the same box as well as in the neighboring image box. The cutoff is chosen such that a particle in the primary box is not affected by its image in the surrounding boxes. As a molecule moves in the original box during a simulation, the periodic image in each of the neighboring boxes moves in exactly the same way. If a molecule leaves the central box, one of its images will enter through the opposite face. There are no walls at the boundary of the central box and no surface molecules. The particle density in the central box is conserved. It is not necessary to store the coordinates of all image boxes, only the the coordinates in the central box are stored.[167] The simplest box is the cubic box, although different shapes exist, i.e. rhombic

Figure 3.3: Representation of a two-dimensional periodic system for a two dimensional box with only the first neighbors of the central box shown. As a particle moves out of the simulation box, an imagine particle moves in to replace it. When calculating particle interactions within the cutoff range, both real and image neighbors are included. Image from the book *Computer Simulation of Liquids*.[167]



dodecahedron and the truncated octahedron. Both of them are more spherical than the cube, which makes the distance between periodic images larger than in the cube.

Limitations of this method are mainly patent for very small system size ($\sim$ 100 atoms) and for properties with important long-range contributions. PBCs however, have little effect on the equilibrium thermodynamic properties and structures.[167]

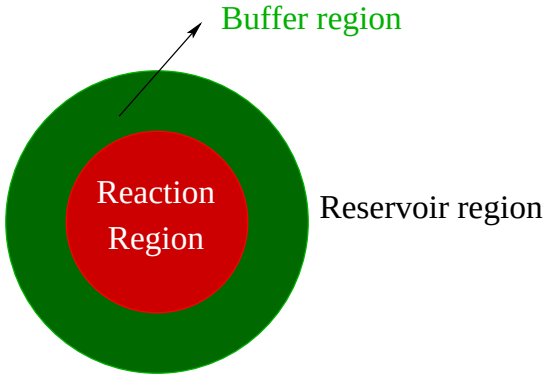
In recent years, a number of models have been introduced which permit the inclusion of long-range electrostatic interactions in molecular dynamics simulation. For simulations of proteins and enzymes in a crystalline state, the Ewald summation[133] is considered to be the correct treatment for long range electrostatic interactions[167] (see Section 3.3.2). Variations of the Ewald method for periodic systems include the particle-mesh Ewald method.[135–138] To treat non-periodic systems, such as an enzyme in solution, methods based on multipole expansions have been developed. Many of these methods partition the electrostatic interaction into a long-range component and a short-range component. The short-range component is treated in the usual pairwise fashion while a multipole approximation is introduced to approximate the long-range electrostatic interaction.[127,139–141] Although these methods require more computer time than if one simply neglects the long-range part, they are significantly faster than if one does the N^2 summation

of all the interactions, and the results can be significantly improved.

Solvent Boundary Potentials

There exist numerous cases where one may not wish to use periodic boundary conditions. In some cases, the use of PBC requires the use of a prohibitively large number of water molecules. The simplest way to incorporate water molecules in the simulation is to surround the protein or just a part of the protein with a sphere of water. Boundary potentials have been developed which restrain the water molecules to a sphere while maintaining a strong semblance to bulk water. Structural and thermodynamics properties when calculated under these conditions indicate that the water still behaves as bulk water. This usually involves much fewer water molecules than in a PBC simulation and is often sufficient. Other times, one may be interested in studying the motion of one part of the protein as moving in a "solvent bath" provided by the remainder of the protein. One way to deal with these issues is, of course, to explicitly include all the additional components in the simulation. This solution is computationally very expensive, because much work is done on parts of the system that are of no direct interest to the study. Another way is to reduce the magnitude of the problem by eliminating the explicit solvent degrees of freedom from the calculation and representing them by simplified models. An alternative approach is to move away from Newtonian molecular dynamics toward stochastic dynamics.

Figure 3.4: Partitioning of the system in a SBC method. The reaction region is treated by Newtonian dynamics, while the buffer region is treated by Langevin dynamics.



One of the solvent boundary potentials method is the **Stochastic Boundary Conditions** (SBC hereafter)[172–175] method. The system is devided into three regions, a *boundary* or *reservoir* region and a *reaction zone*. The reaction zone is partitioned into a *reaction region* and a *buffer region*, Figure 3.4. In the reaction region, which contains the site of major dynamic interest, the particles are treated explicitly by molecular dynamics, Newton's 2nd

law, Equation 3.23. The surrounding buffer region molecules represent a simplified heat bath and are treated explicity by stochastic dynamics based on the Langevin equation,[173, 176]

$$m_i \frac{d^2 r_i}{dt^2} = -\vec{\nabla}_i U(r) - m_i \beta_i v_i(t) + f_i(t) \tag{3.40}$$

where the first term of the right-hand side is the same as that used in Newtonian dynamics, Equations 3.23, 3.25 and 3.26. The other two terms are unique to stochastic dynamics and represent solvent effects. The second term is a dissipative drag force reflecting friction due to the solvent. The friction coefficient, β, is related to the diffusion constant, D, via the Einstein relationship.[129] The third term, f_i is a random force that represent stochastic collisions between solvent molecules and the solute. The stochastic force introduces energy into the system, while the friction force removes it. The Langevin random force is obtained from a random Gaussian distribution of zero mean and a variance related to the friction coefficient,

$$\langle f_i(t) \rangle = 0 \tag{3.41}$$

$$\langle f_i(t) f_i(0) \rangle = 6 m_i k_B T \delta(t) = 2 D_i \delta(t) \tag{3.42}$$

Finally, the boundary or reservoir region provides a static force field that helps to insure that the correct structural and dynamic properties will be maintained within the reaction zone. Due to this force field, each atom in the buffer region in addition to the stochastic forces is subject to an average boundary force. An additional F_i^{mean} term is added to Equation 3.40.

3.4 Quantum Mechanical/Molecular Mechanical Methods

Because the QM treatment of an entire biological macromolecule requires very large amounts of computer time, and because the MM methods can not study bond cleavage/formation processes, hybrid QM/MM potentials are commonly used to study chemical and biological processes involving bond cleavage and formation, such as enzymatic reactions.[177–179] In this approach, a small region of the system (the QM region, the active-site), which electronic structural changes are of interest, is treated quantum mechanically and the remainder of the system (the MM region, protein and solvent which is not involved in the reaction and is believed to change little) is represented by a classical MM force field (see Section 3.3). In this approach, electrostatic effects as well as steric contributions from the environment are incorporated directly into the electronic structure calculations of the reactive

region, affecting its charge polarization and chemical reactivity.[180] Owing to the strong QM-MM interactions, the total energy of the system cannot be writting as the sum of the energies of the subsystems. Coupling terms have to be considered, and special precautions need to be taken at the boundary between the subsytems, especially if it cuts through covalent bonds. The effective Hamiltonian would be written as

$$\hat{H}_{eff} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM-MM} + \hat{H}_{Boundary} + \hat{H}_{Restrains} \tag{3.43}$$

where $\hat{H}_{QM}$ is the Hamiltonian of the QM region. Any *ab initio*, DFT or semiempirical QM method can be used to generate the appropiate electronic description (see Sections 3.1 and 3.2). The $\hat{H}_{MM}$ details the classical interactions in the MM region, and can be generated using any standard molecular simulation program (see Equation 3.20). The third term, $\hat{H}_{QM-MM}$ describes the interaction between the QM region atoms and the MM region particles (see Equation 3.44 in atomic units). This Hamiltonian includes van der Waals and electrostatic interactions between both regions. Van der Waals contributions are treated classically, while electrostatic interactions are described by treating the MM atoms as point charges in an electrostatic field which interacts with the QM atoms.

$$\hat{H}_{QM-MM} = -\sum_{i,M} \frac{q_M}{r_{iM}} + \sum_{\alpha,M} \frac{Z_\alpha q_M}{R_{\alpha M}} + \sum_{\alpha,M} \left(\frac{A_{\alpha M}}{R_{\alpha M}^{12}} - \frac{B_{\alpha M}}{R_{\alpha M}^{6}} \right) \tag{3.44}$$

Here i and α represents the QM electrons and nuclei, respectively, and M corresponds to the MM atoms. The first term, the electrostatic interaction between a MM atom of partial charge q_M and electron i, is included in the SCF calculation, whereas the other two terms do not depend on the electronic coordinates. If semiempirical methods are used, the first term is reduced to the interaction between the valence electrons of QM atoms and the partial charge of MM atoms, while the second term, the Coulombic interaction, can be expressed as Equation 3.45 if AM1/d-PhoT semiempirical potential is being used,

$$\begin{aligned} E_{QM-MM}^{\text{core}} &= Z_Q' q_M \langle s_Q s_M | s_Q s_M \rangle \times \left(1 + e^{-\alpha_Q R_{QM}} + e^{-\alpha_M R_{QM}} \right) \\ &+ \frac{Z_Q' q_M}{R_{QM}} G_{scale}^{Q} G_{scale}^{M} \\ &\times \left(\sum_k a_{kQ} e^{-b_{kQ}(R_{QM} - c_{kQ})^2} + \sum_k a_{kM} e^{-b_{kM}(R_{QM} - c_{kM})^2} \right) \end{aligned} \tag{3.45}$$

where Z_Q' is the core charge of the quantum atom Q, q_M is the point charge of atom M in MM region, and s_Q and s_M are s orbitals associated with the nuclei an MM atom, respectively (see Equations 3.16 and 3.19 for futher explanation).

3.4.1 Treatment of Boundary atoms

The separation into quantum region and classical regions is not straighforward in enzyme reactions,[179,181] because the QM region is frequently bonded through covalent bonds to the MM region, and is not clear how to define uniquely the boundary conditions for the electronic structure calculations of the QM region, nor how to incorporate the electrostatic and van der Waals effects of the classical region into the QM energy expression. A number of different approaches have been proposed to deal with this situation. We will explain only two of them, the simplest approach and the approach we have used in the work done in this thesis project.

Link atom Method

The link atom method[182–184] is the simplest approach proposed to circumvent the problem. In this approach a dummy atom called a link atom, usually hydrogen, is introduced into the quantum system along the covalent bond between the QM and MM regions, to saturate the valency of the QM fragment. It is usually positioned at 1 Å along the original bond but it is not constrained during the simulations. For each link atom (hydrogen) one nucleus and one electron are added to the QM system. Since the link atoms are not part of the actual chemical system under study, they introduce additional degrees of freedom into the system. In addition, partial charges on the MM atoms that are closest to the link atoms must be removed to avoid convergence difficulties.

Generalized Hybrid Orbital Method

This method developed by Gao and co-workers[185] is based on local self-consisten field[186–189] (LSCF) algorithm developed by Rivail and co-workers.[186] The General Hybrid Orbital method[185,190–194] (GHO hereafter) makes use of hybrid atomic orbitals as basis function on the boundary atoms of the MM fragment. The hybrid orbitals are devided into three auxiliary and one active orbitals (see Figure 3.5). The active orbital is included in the SCF calculations, while the remainder auxiliary orbitals are kept frozen and are not allow to mix with the other orbitals. Consequently, the chemical bond connecting the QM and MM regions is explicity treated without introducing spurious "link atoms" and additional degrees of freedom. Furthermore, in contrast to the LSCF approach, the GHO method does not need to be reparametrized every time a new system is studied. The point charge of the MM atom is distributed equally over the three frozen auxiliary hybrid orbitals. They thus provide a type of pseudopotential that mimic the electronic character at the link.

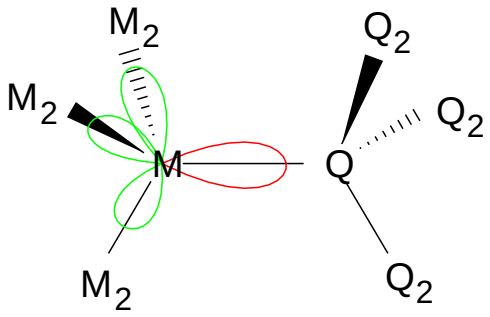


Figure 3.5: Partition between the QM and MM region in the GHO method. The boundary atom of the MM fragment, M, is defined by four hybrid atomic orbitals, where three of them are called auxiliary orbitals (in green) and are pointing towards MM atoms, M_2. The last hybrid orbital, the active orbital (in red), is pointing towards the QM region Q atom.

3.5 Free Energy Calculations

The free energy difference between states of a system indicates the relative stability. The state corresponding to the global minimum in free energy is the most stable, and a system will be in this state with the highest probability unless it becomes trapped in a local minimum due to high energy barriers. At a given temperature, the free energy is a function of all the positions and momenta in the system. Calculation of the free energy as a function of all these variables is practically very difficult. To simplify things, everything can be integrated out except one or a few internal coordinates, reaction coordinates. Integrating everything except the reaction coordinates gives a free energy profile called the *potential of mean force*[195] (PMF hereafter).

The PMF difference between two values of a reaction coordinate is related to the relative probability of observing those values of the reaction coordinate.[196] This requires sampling of all values of the reaction coordinate between those two values. MD simulations can be extended to facilitate the calculation of free energies. They provide a direct link between the microscopic structure and fluctuations of a system and the free energy.

Free energy changes associated with conformational changes are one of the most important applications of free energy calculations. To obtain reliable free energy estimates from simulations, sufficient conformational sampling must be achieved, not only of the starting and final states but also of many intermediate states. The system must be driven along the reaction coordinate with an appropiate set of constraints or restraints. Usually, this represents a considerable challenge. In addition, many techniques to enhance sampling have been proposed: thermodynamic integration using constrained MD,[197,198] umbrella sampling,[162,199] Jarzynski's equality,[200,201] steered MD[202,203] and adaptive force biasing.[204]

Umbrella Sampling

The Helmholtz free energy (free energy of canonical ensemble NVT) along the reaction coordinate ξ is a configuration integral over all other degrees of freedom and takes the form

$$\begin{aligned} A(\xi) &= -k_B T \; ln \left[\frac{\int \delta[\xi - \xi(\mathbf{r})] \; e^{[-\beta H(\mathbf{r})]} \; d\mathbf{r}}{\int e^{[-\beta H(\mathbf{r})]} \; d\mathbf{r}} \right] \\ &= -k_B T \; ln \, P(\xi) \end{aligned} \tag{3.46}$$

where δ is the Dirac delta function, which is 1 when $\xi(\mathbf{r}) = \xi$ and is 0 otherwise, $P(\xi)$ is the reaction coordinate probability density, H is the total energy of the system, $\mathbf{r}$ represents all coordinates for the system and $\beta = 1/k_B T$. When the comparing conformations are separated by energy barriers greater than $k_B T \sim 1$ Kcal/mol, barrier crossing in a simulation will be rare and $P(\xi)$ statistically unreliable. To bias the sampling toward a region of interest that would not otherwise be significantly populated, a restraining potential, *umbrella potential* $U_{bias}(\xi)$, is added to the potential energy of the system (Equation 3.47), which can be described as in Equation 3.48 if a harmonic form is assumed

$$V' = V + U_{bias} \tag{3.47}$$

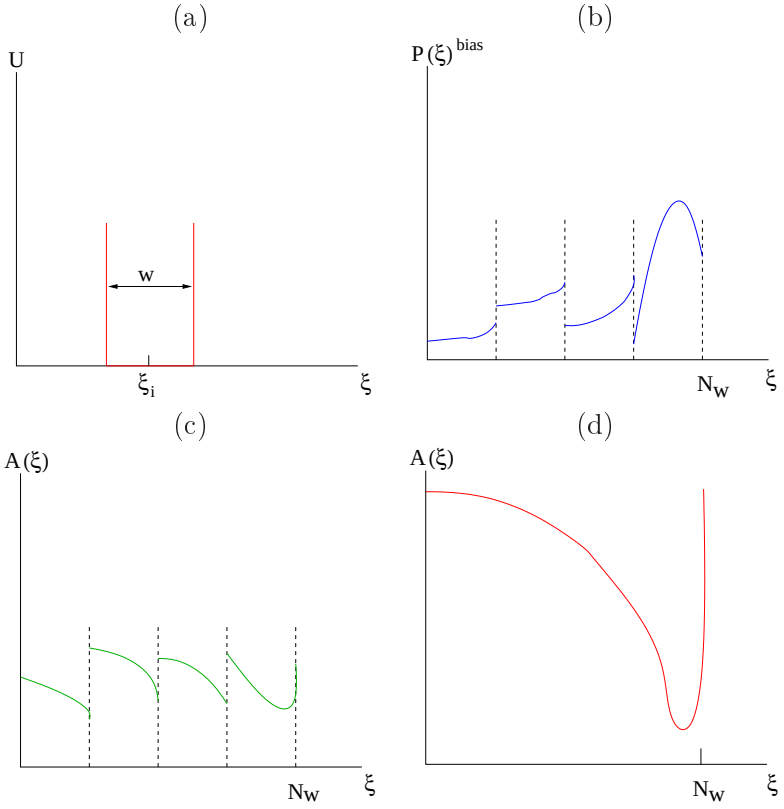
$$U(\xi)^{bias} = \frac{1}{2} K \, (\xi - \xi_i)^2 \tag{3.48}$$

where K is a force constant and ξ_i is a target value of ξ. The biasing potential serves to confine the variations of the coordinate ξ within a small interval around ξ_i, helping to achieve a more efficient configurational sampling in this region. By making the biasing potential sufficiently steep, large K, the energy of V' surface far from ξ_i will become so high in energy that only the region near ξ_i will be sampled at ambient temperatures.

The *Umbrella Sampling* technique[162, 199] choses a set of umbrella or window potentials over the whole range of interest ξ, and performed a simulation with each of them, biasing the configurational sampling around a different region of ξ. The analysis of a histogram is done within each window, to determine the probability of ξ_i in the range $\xi_i \pm w/2$ (see Figure 3.6). After the full range of the reaction coordinate is studied in that way, the results of various windows are unbiased and recombined together to obtain the final estimate free energy along the reaction path, the PMF.

Taking into account Equation 3.46 the biased probability density obtain from the ith biased ensemble is

Figure 3.6: Picture (a) shows one of the umbrella potentials, one of the windows. (b) and (c) shows a set of probabilities and Helmholtz free energies, respectively, that would be acquired from each window. Finally, (d) depicted the continuos curve that would be obtained by connecting the $A'(q)$ from each window assuming that umbrellas have points in common with their neighbors.



$$P(\xi)_i^{bias} = \frac{\int \delta\left[\xi - \xi(\mathbf{r})\right]\ e^{\{-\beta[H(\mathbf{r})+U_i(\xi(\mathbf{r}))]\}}\ d\mathbf{r}}{\int e^{\{-\beta[H(\mathbf{r})+U_i(\xi(\mathbf{r}))]\}}\ d\mathbf{r}} \tag{3.49}$$

$$= e^{-\beta U_i(\xi(\mathbf{r}))}\ P(\xi)_i^{unbias}\ \left\langle e^{-\beta U_i(\xi(\mathbf{r}))}\right\rangle^{-1}$$

where $< \cdots >$ indicates the average over the system ensemble. Therefore, the unbias probability can be expressed as

$$P(\xi)_i^{unbias} = e^{\beta U_i(\xi(\mathbf{r}))}\ P(\xi)_i^{bias}\ \left\langle e^{-\beta U_i(\xi(\mathbf{r}))}\right\rangle \tag{3.50}$$

based on Equation 3.46 the PMF for the ith window is

$$A(\xi)_i = -k_B T \; ln \, P(\xi)_i^{bias} - U(\xi)_i^{bias} + F_i \tag{3.51}$$

where F_i is the free energy constant which represents the free energy associated with the introduction of the window potential and is defined as

$$e^{-\beta F_i} = \left\langle e^{-\beta U_i(\xi(\mathbf{r}))} \right\rangle \tag{3.52}$$

The unknown free energy constants F_i are obtained by adjusting the various $A(\xi)_i$ of adjacent windows in the region in which they overlap until they match,[162,199] the matching can be done manually or automatically. Once F_i values are obtained, the PMF for the whole range of interest is generated by connecting the various $A(\xi)_i$ together. The process of matching the adjacent windows is not easy, is somewhat arbitrary, so the uncertainty involved in the process results in a global error that grows with the number of windows. The *weighted histogram analysis method*[205–208] (WHAM hereafter) is an approach designed to minimized propagation error by making optimal use of the information from multiple simulations. The central idea, the maximum overlap method, consists in constructing an optimal estimate of the $P(\xi)^{unbias}$ as a weighted sum over the data extracted from all the simulations and determining the functional form of the weight factors that minimizes the statistical error.

$$P(\xi)^{unbias} = \sum_{i=1}^{N_w} P(\xi)_i^{unbias} \times \left[\frac{n_i \; e^{-\beta[U_i(\xi)-F_i]}}{\sum_{j=1}^{N_w} n_j \; e^{-\beta U_j(\xi)-F_j]}} \right] \tag{3.53}$$

where n_i is the number of data points used to construct the biased distribution function and N_w is the number of windows used to cover the whole range of the reaction coordinate. Based on Equations 3.50 and 3.52, Equation 3.53 can be re-written as

$$\begin{aligned} P(\xi) &= \sum_{i=1}^{N_w} \frac{n_i \; P(\xi)_i^{bias} \; e^{-\beta F_i} \; e^{-\beta[U_i(\xi)-F_i]}}{e^{-\beta U_i(\xi)} \; \sum_{j=1}^{N_w} n_j \; e^{-\beta[U_j(\xi)-F_j]}} \\ &= \sum_{i=1}^{N_w} \frac{n_i \; P(\xi)_i^{bias}}{\sum_{j=1}^{N_w} n_j \; e^{-\beta[U_j(\xi)-F_j]}} \end{aligned} \tag{3.54}$$

Part II

Characterization of Oxyphosphorane Intermediates by Classical Molecular Dynamic Simulations

Introduction

As has been mention in the General Introduction (see Section 2.3) *Bovine pancreatic ribonuclease A* is an ideal system for a theoretical study of the structural and dynamic basis of enzyme catalysis. This endoribonuclease catalyzes the breakdown of 3',5'-phosphodiester linkage of single stranded RNA. The wide range of biochemical, physical, and crystallographic data available for this enzyme[24,28,35,57,72] have led to proposals for the catalysis of the hydrolysis of RNA by a two-step mechanism,[76] in which a cyclic phosphate intermediate is formed and subsequently hydrolyzed. Both steps are thought to involve in-line displacement at the phosphorus and to be catalyzed by the concerted action of a general acid and a general base. In spite of the accumulated experimental information, a detailed understanding of the origin of the rate enhancement produced by RNase A has not been achieved, and there remains some debate regarding several aspects of the chemical mechanism.[29,60,61,73,79,80,87,209]

Theoretical calculations are potentially powerful tools that, together with experiment, can provide deep insight into the details of catalytic mechanisms. In order to provide insight into mechanism, it is necessary to characterize and understand the changes in enzyme conformation and contacts that occur along the reaction coordinate, particularly at the transition state. It is difficult to obtain direct structural information about a transition state from experiment. In the case of phosphoryl transfer reactions, for example, vanadate transition state mimics have been commonly used and characterized with X-ray crystallography.[210–213] Borah *et al.*[212] studied the structure of RNase A with an uridine vanadate (U>v) transition state analog by ^{51}V and proton NMR spectroscopy in solution, and by joint neutron and X-ray diffraction (2.0 Å resolution) in a crystalline environment. Some years later Ladner *et al.*[213] studied the same complex at 1.3 Å resolution. The overall enzyme structure and the relative position of the key active-site residues, His12, His119 an Lys41, are similar, while the V-O bond distances and angles present significant differences (see Tables A.1 and A.2). However, the results obtained solely based on analyses of a complex with a pentavalent organovanadate have to be taken with some skepticism, as vanadate transition state mimics might be only marginally relevant to the true oxyphosphorane transition state.[214] Very recently, X-ray crystallography was used to study the

influence of four naturally occurring 5-pyrophosphate linked substituents on the binding of adenylic RNase A inhibitors, the analysis of which enabled kinetic data on the binding to be rationalized.[215] In solution, Loria *et al.* have performed extensive NMR spin-relaxation experiments on the μs-ms dynamics of RNase A,[47] studies of the apo and substrate-mimicked states,[48] and characterization of the dynamics at different stages along the reaction coordinate[216] including the transition state.[49] These and other experimental studies provide a wealth of information about RNase A that are linked, albeit sometimes indirectly, to its structure and binding at different stages of the reaction. Nonetheless, to date a detailed study of the structure and dynamics of RNase A at various critical points along the catalytic chemical pathway in solution has not been reported.

On the other hand, a powerful experimental technique used to probe the mechanistic details is the analysis of "thio effects",[217] i.e., the change in the reaction rate that occurs when oxygen atoms are substituted by sulfur at selected positions. Thio-substitutions may simultaneously affect several structural parameters, such as chain folding, hydrogen bonding, metal-ion binding, solvation, and van der Waals interactions, and, hence, interpretation of the observed change in the reaction rate is not always unambiguous. Theoretical studies can provide valuable aid in the interpretation of the thio effect experiments. In addition, electronic and solvation effects of the thio-substitution are opposite. While the soft sulfur atoms stabilize anionic phosphorane-like transition states by more facile delocalization of the negative charge, this rate-accelerating effect is largely compensated by less efficient solvation of the thiophosphorane intermediates compared to their oxygen counterparts. Therefor, the cleavage of phosphodiester via a dianionic transition state usually exhibits a rather small thio effect. Bearing the importance of solvation in mind, stabilization of a phosphorane intermediate or leaving group by intramolecular hydrogen bonding could well be reflected in the magnitude of the thio effect.

Chapter 4

Bovine Pancreatic Ribonuclease A - CpA and Transition State-like Complexes

Abstract

The mechanisms of enzymes are intimately connected with their overall structure and dynamics in solution. Experimentally, it is considerably challenging to provide detailed atomic level information about the conformational events that occur at different stages along the chemical reaction path. Here, theoretical tools may offer new potential insights that complement those obtained from experiments that may not yield an unambiguous mechanistic interpretation. In this study, we apply molecular dynamics simulations of bovine pancreatic ribonuclease A, an archetype ribonuclease, in order to study the conformational dynamics, structural relaxation, and differential solvation that occurs at discreet stages of the transesterification and cleavage reaction. Simulations were performed with explicit solvation with rigorous electrostatics, and utilize recently developed molecular mechanical force field parameters for transphosphorylation and hydrolysis transition state analogs. Herein, we present results for the enzyme complexed with the dinucleotide substrate cytidilyl-3',5'-adenosine (CpA) in the reactant, and transphosphorylation and hydrolysis transition states. A detailed analysis of active site structures and hydrogen bond patterns are presented and compared. The integrity of the overall backbone structure is preserved in the simulations, and support a mechanism whereby His12 stabilizes accumulating negative charge at the transition states through hydrogen bond donation to the non-bridge oxygens. Lys41 is shown to be highly versatile along the reaction coordinate, and can aid in the stabilization of the dianionic transition state, while being

poised to act as a general acid catalyst in the hydrolysis step.

4.1 Introduction

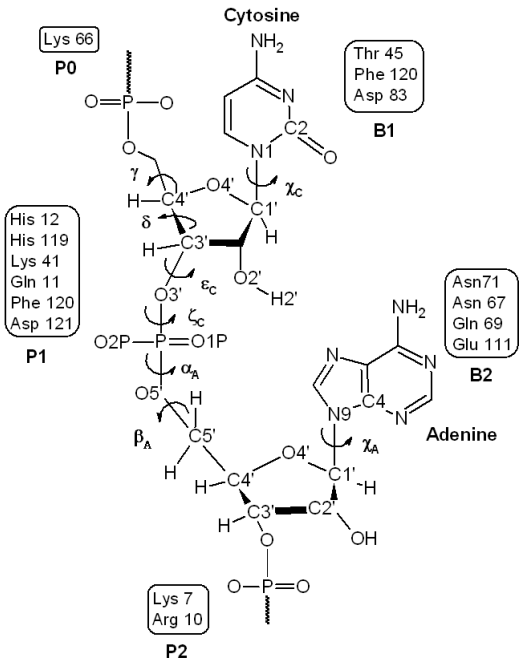
In this chapter we investigate the structure and dynamics of RNase A - 3',5'-CpA in solution with MD simulations of the reactant state, and transphosphorylation and hydrolysis transition state mimics (Scheme 4.1). Transition state mimics were modeled by the CHARMM force field parameters for reactive phosphoryl transfer intermediates.[218] Simulations of RNase A have been previously reported by other authors,[80,219–224] however, the present work is, to our knowledge, the first time that simulations of oxyphosphorane transition states for RNase A have been studied. The simulations provide insight into the flexibility of the bound ligands and the interaction of the surrounding active site residues to stabilize the trigonal bipyramidal geometry of the transition state/intermediates of the transphosphorylation and hydrolysis reactions.

The chapter is organized as follows: The Computational Details section describes the force field models and simulation set up. The Results section presents simulation results for the reactant model, transphosphorylation transition state mimic model ($[\text{Trans}]^{TS}$) and hydrolysis transition state mimic model ($[\text{Hyd}]^{TS}$). The Discussion section places the simulation results into a broader biological context, and makes detailed comparison with experiments. The Conclusion section summarizes the main points of the chapter and outlines future research directions.

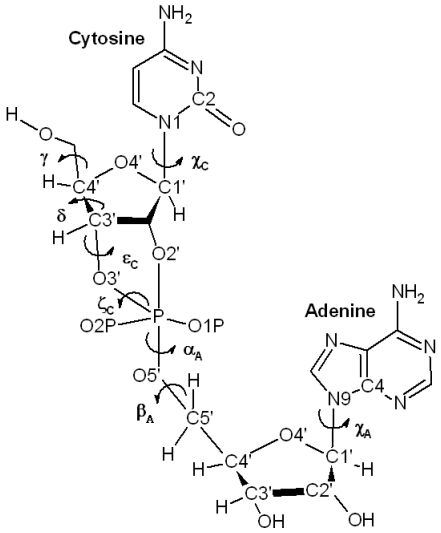
4.2 Computational details

The starting structure for all-atom molecular dynamic simulations of RNase A complexed with cytidilyl-3',5'-adenosine (CpA) in solution was taken from Protein Data Bank with accesion number 1RPG,[50] where the structure was determined at a resolution of 1.4 Å. The CpA structure was built from the crystal results for the inhibitor complex, d(CpA), by introducing the 2' oxygen into the deoxy sugar. The transphosphorylation and hydrolysis transition state mimics were built from the previous complex using a "patch" to introduce a dianionic pentavalent cyclic oxyphosphorane transition state mimic with 2' (nucleophile) and 5' (leaving group) positions oriented axially.[218] The position of the hydrogen atoms were determined using HBUILD facility in the program CHARMM,[126,127] Ver. c32a1. Atomic charges, van der Waals, and force-field parameters corresponded to the all-atom CHARMM27 nucleic acid force field[144–146] with extension to transition state mimics[218] and TIP3P water model.[225] In the reactant state, His12 and His119 are neutral and protonated, respectively, in states appropriate for their proposed roles as general base and acid catalysts for the transphosphorylation reactions. In

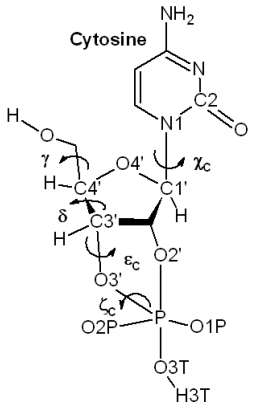
Scheme 4.1: Schematic representation of the RNase A active site with a bound RNA ligand (CpA). The subsites interacting with RNA bases (B_i) and phosphate groups (P_i) are indicated for the reactant state (left). The transphosphorylation ([Trans]^{TS}) and hydrolysis ([Hyd]^{TS}) transition state mimics are shown in the middle and right, respectively. The atomic names and torsional angles discussed in the text are indicated. [Trans]^{TS} and [Hyd]^{TS} are renamed as [native]$^{\text{Trans}}$ and [native]$^{\text{Hyd}}$ in Section 5.



Reactant



[Trans]^{TS}



[Hyd]^{TS}

the transition state mimics, both of these residues are protonated since the general base has already accepted a proton from the nucleophile, and the general acid is loaded to donate a proton to the leaving group. All other residues were assumed to be in their standard ionization states in solution at neutral pH.

The aqueous solvation environment was created by retaining the 164 crystallographic water positions, and further solvating the system with bulk water molecules and ions. Additional water molecules were included by overlaying a pre-equilibrated rhombic dodecahedral cell of 8,346 TIP3P water molecules, centering it at the center of mass of the complex using statistics methodology, and deleting water molecules with oxygen atom lying in a radius of 2.8 Å from any non-hydrogen atom of the protein, enzyme or crystallographic waters. The ion atmosphere consisted of Na^+ and Cl^- ions that were added at random positions to neutralize the net +6 charge of the solute and reach the physiologic extracellular concentration of 0.15 M. The reactant simulation corresponded to the entire protein (1,859 atoms), the ligand (63 atoms), 7,548 water molecules, 6 Cl^- ions to neutralize the net positive charge of the solute and 42 additional counter and coions (21 Cl^{-1} and 21 Na^{+1} ions). The transition state mimic simulations were prepared in an analogous fashion.

Periodic boundary conditions (see Section 3.3.6) were used along with the isothermal-isobaric ensemble (NPT)[196] at 1 atm and 298 K via an extended system pressure algorithm[226] and Nosé-Hoover thermostat.[227,228] Electrostatic interactions were treated rigorously using the smooth particle mesh Ewald method (PME)[135,137,138] with a κ value of 0.35 $Å^{-1}$ and direct-space cut-off of 12 Å, 72 FFT grid points for each of the lattice directions, and 4th-order B-spline interpolation for spreading the atomic charges to the FFT grid (see Section 3.3.2). Nonbonded interactions were performed using a cutoff of 12 Å with shifted van der Waals potential to smooth the Lennard-Jones term. Nonbond pair lists were maintained out to 13.0 Å and updated heuristically. Newton's equations of motion were integrated numerically using the leapfrog Verlet algorithm within a timestep of 1 fs,[167] while atomic coordinates were saved for analyses every 0.5 ps (see Section 3.3.5). Covalent bond lengths involving hydrogen were constrained using the SHAKE algorithm.[229]

Water molecules were initially relaxed for 5000 steps of steepest descents[158,160] energy minimization keeping all solute atom and ion positions restrained to their initial coordinates using a harmonic potential constant of 10 $kcal{\cdot}mol^{-1}{\cdot}Å^{-2}$. The restraints on the ions were then released, and the solvent (water and ions) were relaxed with 5000 steps of conjugate gradient minimization,[159] keeping the solute positions restrained. From this starting point, restrained molecular dynamics was performed starting at 0 K and heating to 298 K over 20 ps, and carried out to 220 ps, during the course of which the harmonic restraints on the solute atoms were slowly released. The production phase molecular dynamics then began, and covered 6.0 ns, the

last 5 ns of which is used for analysis. All the simulations were performed with the CHARMM biomolecular program[126,127], Ver. c32a1.

4.3 Results

This section presents molecular dynamics simulation results for RNase A in three different model states along the catalytic reaction: the reactant state complex (3',5'-CpA), the transphosphorylation transition state mimic ([Trans]^{TS}) and hydrolysis transition state mimic ([Hyd]^{TS}) models (Scheme 4.1). Representative structures from the simulation are shown in Figure 4.1. Results are compared with available crystallographic data in Tables A.1 and A.2. The results are divided into different subsections for each model structure.

First, the root-mean-square deviation (RMSD) and the ligand conformation are shown. The glycosyl dihedral angles $\chi_{C(O_{4'}-C_{1'}-N_1-C_2)}$ and $\chi_{A(O_{4'}-C_1-N_9-C_4)}$ describe, respectively, the orientation of the cytidine and adenine rings with respect to the sugar,[230,231] whereas the pseudorotation phase angle P[232] describes the sugars ring conformation, and the backbone torsion angles ($\alpha, \beta, \gamma, \delta, \epsilon$ and ζ) characterize the conformation of the sugar-phosphate backbone (Scheme 4.1). Then, we discuss the geometry of the active-site, focusing attention on the conformation of key residues and their interaction with the substrate and solvent molecules. The side chain of residue His119 can adopt two conformations denoted as A ($\chi_{1(N-C_\alpha-C_\beta-C_\gamma)} \sim 160°$) and B ($\chi_1 \sim -80°$),[62] which are related by a 142° rotation about $C_\alpha - C_\beta$ bond and 38° rotation about the $C_\beta - C_\gamma$ bond. Conformation A is the "active" conformation that promotes catalysis, whereas conformation B is an "inactive" conformation.[50] In the crystallographic structure of the complex, conformation A is observed.[50] The criteria for the existence of hydrogen bonds of the form $D - H \cdots A$, where D and A are the hydrogen bond donor and acceptor, respectively, are 1) a maximum $H \cdots A$ distance R of 2.4 Å, 2) a minimum $\widehat{DHA}$ angle of 120° and 3) a hydrogen bond lifetime $\tau \geq 5.0$ ps. Finally radial distribution functions (RDFs) of water molecules around phosphoryl and oxyphosphorane oxygens are calculated. The RDFs are denoted $g_{XY}(r)$ where the subscript XY refers to the distribution of Y atoms around X atoms. Analysis for all trajectories was over the last 5 ns of production simulation, following pre-equilibration and 1 ns of unrestrained equilibration dynamics.

4.3.1 Reactant model

In this subsection, results are presented for the simulations of the RNase A - 3',5'-CpA complex in the reactant state (Scheme 4.1 and Figure 4.1). The RMSD with respect to the initial conformation as a function of the simulation

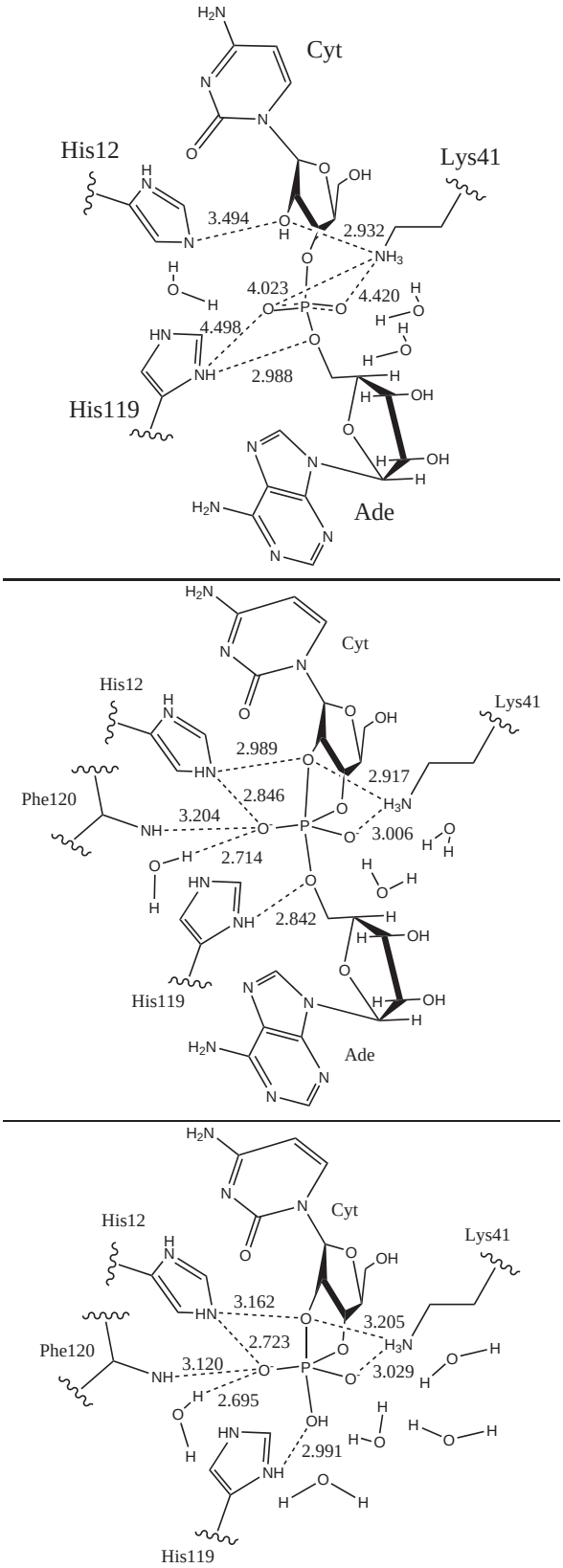


Figure 4.1: Structure of the Reactant model at top, $[\text{Trans}]^{TS}$ in the middle and $[\text{Hyd}]^{TS}$ at the bottom. Some MD average distances between heavy atoms are depicted.

time is shown in Appendix Figure A.1. The RMSD reaches equilibrium after 1 ns and fluctuates stably with an average value of 1.46 ± 0.11 Å.

Conformational dynamics

The time evolution of key dihedral angles that define the conformation of the substrate and the His119 residue are shown in Supporting Information (Figure A.2) and the average values are listed in Table A.3. The glycosyl dihedral angles, χ_C and χ_A, are in the anti conformation in the crystallographic structure.[50] The fluctuations of these dihedral angles are very small, signifying that the χ_C and χ_A anti conformations are stable throughout the simulation. The average values are -161.0° and -65.7° respectively. However, the conformation of the ribose shows some differences with respect to the crystallographic structure. In particular, the cytidine ribose essentially stays at North $C_{3'}$-endo throughout the simulation in contrast with the South $C_{2'}$-endo conformation encountered at the crystallographic structure.[50] On the other hand, the adenosine ribose is observed in the simulations to sample both the South $C_{2'}$-endo conformation as in the crystal structure, and the North $C_{3'}$-endo pucker. The average values are 23.2° and 140.8° respectively (Figure A.2 and Table A.3). The tendency of the simulation to occupy a $C_{3'}$-endo pucker is due to the fact that the sugar rings are the native ribose sugars of RNA, whereas in the crystal structure, the substrate is d(CpA), having the ribose replaced with deoxyribose in order to render the complex inactive (recall, it is the 2'OH group that acts as nucleophile in the transphosphorylation). Whereas classical B-form DNA tends to have $C_{2'}$-endo sugar pucker, RNA typically has a $C_{3'}$-endo sugar pucker similar to an A-form helix that reduces the steric repulsion between the 2' and 3' oxygens.[233] In addition, the dinucleotide phosphodiester backbone torsional angle undergoes a transition from gauche$^+$ (γ_C), trans (γ_A) in the crystallographic structure to the more stable gauche$^+$, gauche$^+$ in the simulation, which was also observed by Brünger *et al.*[219] Consequently, these differences in pucker between the simulation and crystal structure have relevance for the native reactive substrate. Finally, it is noteworthy that the side chain of His119 maintains its active A conformation of the crystallographic structure throughout the simulation, with a mean χ_1 value of 166.9° and χ_2 of -104.2° (see Figure A.2 in Supporting Information).

Hydrogen bond interactions

The average distances between atoms participating in protein-substrate hydrogen bonds, with the hydrogen-bond occupancies, average time and number of events are listed in Table 4.1. Time evolution of selected distances between protein and ligand atoms is plotted in Figure A.3. In the crystal structure of RNase A - 3',5'-d(CpA) complex[50] the ligand phosphate

group interacts with His12, His119, Gln11, Phe120 and a water molecule (Table A.1). In our reactant model simulation, this group makes a strong hydrogen bond with the imidazolium side chain of His119, and weaker hydrogen bond interactions with Gln11, Lys41, His12, Phe120 and solvent water (Table 4.1).

Table 4.1: Statistics of protein-ligand and water-ligand hydrogen-bonds observed in the simulation of the reactant model. The criteria for the existence of hydrogen bonds are 1) a maximum H $\cdots$ A (acceptor) distance of 2.4 Å 2) a minimum $\widehat{DHA}$ of 120° and 3) a hydrogen bond lifetime $\tau \geq 5.0$ ps. The occupancy, *Occ*, is defined as the total lifetime of each hydrogen bond by the production time. The events are the number of times each hydrogen bond is formed. All distances are in Å. HSP stands for a doubly protonated His, while HSD stands for a His protonated at Nδ position.

	Distance			Hydrogen bond		
Atom pair	MD[†]	MD[‡]	X-Ray[§]	*Occ*	$\langle\tau\rangle$(ps)	# Events
LYS 41 Hζ_1 - ADE O$_{1P}$	4.294(0.937)	4.420(0.693)	4.722	0.03	33.02	5
LYS 41 Hζ_2 - ADE O$_{1P}$	4.260(0.877)			0.01	25.00	2
LYS 41 Hζ_3 - ADE O$_{1P}$	4.311(0.827)			0.01	15.00	4
ADE 2 H3T - ADE O$_{1P}$	5.400(0.616)	5.198(0.522)	5.062	0.01	30.00	1
WATER - ADE O$_{1P}$			3.148	1.72	25.70	335
GLN 11 Hϵ_{21} - ADE O$_{2P}$	2.419(0.571)	3.212(0.455)	4.181	0.57	37.24	76
LYS 41 Hζ_1 - ADE O$_{2P}$	4.254(0.863)	4.023(0.736)	4.938	0.02	22.50	4
LYS 41 Hζ_2 - ADE O$_{2P}$	4.138(0.855)			0.01	15.00	4
LYS 41 Hζ_3 - ADE O$_{2P}$	4.313(0.789)			0.00	20.00	1
HSP 119 Hδ_1 - ADE O$_{2P}$	3.696(0.312)	4.498(0.318)	3.803	0.00	10.00	1
CYT 1 H$_{2'}$ - ADE O$_{2P}$	1.916(0.394)	2.741(0.236)		0.88	79.56	55
WATER - ADE O$_{2P}$			3.062	1.19	43.46	137
LYS 41 Hζ_1 - CYT O$_{2'}$	2.900(0.754)	2.932(0.253)		0.32	56.96	28
LYS 41 Hζ_2 - CYT O$_{2'}$	2.867(0.819)			0.36	152.06	12
LYS 41 Hζ_3 - CYT O$_{2'}$	3.093(0.725)			0.20	100.51	10
WATER 11 - CYT O$_{2'}$		3.836(0.344)		0.00	10.00	2
HSP 119 Hδ_1 - CYT O$_{3'}$	2.901(0.405)	3.318(0.241)	4.429	0.03	12.47	12
WATER - CYT O$_{3'}$				0.01	11.67	3
HSP 119 Hδ_1 - ADE O$_{5'}$	2.141(0.394)	2.988(0.242)	2.656	0.81	59.87	68
WATER 5892 - ADE O$_{5'}$				0.00	10.00	1
HSD 12 Nϵ_2 - CYT H$_{2'}$	3.250(0.202)	3.494(0.279)		0.03	14.98	9
WATER 11 - CYT H$_{2'}$		3.836(0.344)		0.24	12.82	92

[†] Arithmetic average hydrogen bond distance over the last 5 ns of the production simulation, with standard deviation in parentheses. [‡] Arithmetic average distance between the corresponding heavy atoms of the last 5 ns of the production trajectory, with standard deviation in parentheses.
[§] X-Ray data from Zegers *et al.*[50]

The O_{2P} phosphate oxyanion (pro-S oxygen) makes a strong hydrogen bond interaction with cytosine $H_{2'}$ (the nucleophile) and side chain of residue Gln11 with average distances $< R >$ of 1.916 Å and 2.419 Å, and hydrogen bond occupancies *Occ* of 0.88 and 0.57, respectively (Table 4.1). For the last 2 ns of the production dynamics, the interaction between O_{2P} and Gln11 becomes slightly weaker (Figure A.3). The O_{1P} oxyanion (pro-R) does not interact directly with any residue of the enzyme. The O_{2P} and O_{1P} oxygens interact with water molecules throughout the simulation. The imidazole side chain of His12 forms a water bridge with O_{2P} for most of the simulation. This same water molecule forms a hydrogen bond with the 2'OH of cytosine. Together, these interactions form a hydrogen bond network that stabilizes the orientation of the nucleophile. The $O_{3'}$ phosphoryl oxygen is hydrogen bonded to the imidazolium side chain of protonated His119 in the last 1.5 ns, while the interaction between $O_{5'}$ phosphoryl oxygen and His119 remains strong throughout the simulation. The average distances and hydrogen bond occupancies are 2.901 Å and 0.03 for $O_{3'}$ and 2.141 Å and 0.81 for $O_{5'}$ (Table 4.1 and Supporting Information Figure A.3).

Residue Lys41 is positioned at a distance of 3.7 Å from $O_{3'}$ oxygen in the crystal structure,[50] whereas in the reactant state simulation, the average distance is 5.00 Å (Table A.1). The side chain of Lys41 forms weak hydrogen bond interactions with O_{1P} and O_{2P} phosphoryl oxygens, and a strong hydrogen bond interaction with the 2'-hydroxyl oxygen, $O_{2'}$ (Table 4.1 and Figure A.3). The average heavy-atom distances and hydrogen bond group occupancies involving Lys41 are 4.420 Å and 0.05, respectively, for O_{1P}, 4.023 Å and 0.03, respectively, for O_{2P}, and 2.932 Å and 0.88, respectively, for $O_{2'}$.

Solvation of the substrate

Determination of ligand-solvent interactions in the active site is of paramount importance to reveal the role of specific water molecules and bulk solvent effects in the enzymatic mechanism. We analyzed two aspects of ligand-solvent interactions: specific hydrogen bond interactions with active site water molecules, and radial distribution functions of water solvent around key atoms of the substrate.

The CpA ligand makes a variety of hydrogen bonds with various water molecules in the active site (Table 4.1). Each atom of the ligand capable of hydrogen bonding was observed to form hydrogen bonds with several different water molecules in the course of the production simulation, with the exception of the 2'-hydroxyl group and $O_{5'}$ oxygen that only interact transiently with a single water molecule in the course of the simulation. The longest average hydrogen bond lifetime τ is 43.46 ps and corresponds to the O_{2P} phosphoryl oxygen, whereas the largest "occupancy", 1.72, corresponds to water interactions with O_{1P} phosphoryl oxygen. Note that in this case

the hydrogen bond occupancy is > 1 because it accounts for the possibility of making simultaneous hydrogen bonds with multiple water molecules, and in this sense is more similar to an average coordination number.

Radial distribution functions are shown in Figure 4.2. Broadly speaking, the RDF's describe three kind of situations:

- *i)* a very sharp peak at the first solvation layer followed by a deep minimum with a very low probability. This is indicative of tightly bound and highly ordered water molecules interacting directly with the atom, with slow or negligible water exchange. We will refer to this regime as a *highly ordered water interaction*. This is the case for O_{1P} and O_{2P}. The first one shows the largest peak among the RDF's for the reactant complex simulation, 3.4 at 2.6 Å and falls to 0.3 at 3.2 Å while O_{2P} RDF is 3.1 at 2.7 Å and falls to 0.4 at 3.1 Å.

- *ii)* a pronounced peak at the first solvation layer followed only by a modest decrease to a local minimum that still has significant probability. This is indicative of a reasonably ordered first solvation layer that is able to readily exchange with bulk solvent. We will denote this regime as a *fully-exposed solvent interaction*. This is the case for g_{P-O_W} in the reactant simulation.

- *iii)* a weak first peak followed by a fairly disordered fluctuations in the RDF that gradually relax to bulk values. This behavior is indicative of poorly solvated areas, or limited solvent accessibility. We will classify these atoms as *poorly solvated atoms*. This is the regimen we found for $g_{O_{2'}-O_W}$, $g_{O_{5'}-O_W}$ and $g_{O_{3'}-O_W}$. Those RDF start to increase for distances > 3.0 Å (Figure 4.2).

4.3.2 Transphosphorylation transition state mimic model, [Trans]^{TS}

In this subsection we show the results obtained for the RNase A - 3',5'-CpA transphosphorylation transition state mimic model (Scheme 4.1 and Figure 4.1). The RMSD with respect to the initial conformation as a function of the simulation time is shown in Figure A.1. The RMSD reaches equilibrium after 1 ns and fluctuate stably with an average value of 1.24 ± 0.10 Å.

Conformational dynamics

The time evolution of χ_C and χ_A glycosyl dihedral angles are shown in Figure A.2 and the average values are listed in Table A.4. The fluctuations of both dihedral angles are small, signifying that the χ_C and χ_A anti conformations remain stable throughout the simulation as in the reactant model. The average values are -158.9° and -63.1°, respectively. The cytidine ribose stays

in the North $C_{3'}$-endo throughout the simulation, while the adenosine ribose occupies mainly a South $C_{2'}$-endo conformation with infrequent and short-lived transitions to a North $C_{3'}$-endo pucker. The average values are 10.1° and 149.4° respectively. From $\sim$ 5300 ps to $\sim$ 5700 ps the α and γ_A phosphodiester dihedral angles undergo a conformational transition from trans to gauche$^+$ and from gauche$^+$ to trans, respectively (Figure A.4) known as a crankshaft motion, and is often observed in $B_I \leftrightarrow B_{II}$ transitions .[234] Finally, the imidazolium side chain of residue His119 retains an active A-type conformation as in the reactant model, with a mean χ_1 value of 165.9° and χ_2 value of -108.6° (Figure A.2.)

Hydrogen bond interactions

Hydrogen bond formation between the ligand and the protein residues in the active site is described in Table 4.2, in which the average distances between atoms participating in protein-ligand hydrogen bonds, with the hydrogen-bond occupancies, average time and number of events are listed. The time evolution of selected distances between protein and ligand atoms is plotted in Supporting Information Figure A.4. In the $[\text{Trans}]^{TS}$ simulation, the phosphodiester group makes strong hydrogen bond interactions with the two catalytic histidines, His12 and His119, Gln11, Phe120 and a water molecule, and somewhat weaker interactions with Lys41 and some water molecules.

The O_{1P} equatorial non-bridge oxygen makes a strong hydrogen bond interaction with the side chain of Gln11, which was not seen in the reactant model, in addition to hydrogen bonding with water molecules. Also observed at this position was a weaker hydrogen bond interaction with the side chain of Lys41 (Table 4.2). The average O_{1P} hydrogen bond distance and occupancy with Gln11 are 1.984 Å and 0.94, respectively, while the average distance between O_{1P} and Nζ of Lys41 is 3.006 Å with a hydrogen bond group occupancy of 0.30. Equatorial oxygen O_{2P} makes strong hydrogen bond interactions with imidazolium side chain of His12 and a water molecule, and somewhat weaker interactions with the main chain of Phe120 (Table 4.2 and Figure A.4). The average O_{2P} hydrogen bond distances and occupancies are 1.877 Å and 0.99 with His12, 2.246 Å and 0.73 with Phe120, and 2.714 Å and 1.00 with the water molecule. As in the reactant model, O_{1P} and O_{2P} interact with water molecules throughout the simulation. In the $[\text{Trans}]^{TS}$ simulation, His12 interacts directly with the O_{2P} equatorial oxygen, instead of via a water bridge as in the reactant state simulation. The side chain of His12 also hydrogen bonds with the $O_{2'}$ axial oxygen with an average distance of 2.374 Å and hydrogen bond occupancy of 0.36.

The imidazolium side chain of His119 forms a strong hydrogen bond interaction throughout the simulation with the $O_{5'}$ axial oxygen and a weak interaction with $O_{3'}$ equatorial oxygen, similar to the reactant model. However, the His119 residue, which is protonated in all of the simulations, is

closer to the dianionic phosphorane in the [Trans]^{TS} simulation than to the monoanionic phosphate in the reactant state simulation. The average hydrogen bond distances are 1.851 Å and 2.602 Å respectively. The hydrogen bond between His119 and $O_{5'}$ is observed to exchange only eight times during the [Trans]^{TS} simulation, and has an average lifetime of 638.7 ps. The cationic Lys41 side chain forms strong interactions with the $O_{2'}$ axial oxygen ($< R >$=2.917 Å and Occ=0.80) and weaker interactions with O_{1P} equatorial oxygen ($< R >$=3.006 Å and Occ=0.30).

Table 4.2: Statistics of protein-ligand and water-ligand hydrogen-bonds observed in the simulation of the [Trans]^{TS} mimic model. The criteria for the existence of hydrogen bonds are 1) a maximum H $\cdots$ A (acceptor) distance of 2.4 Å 2) a minimum $\widehat{DHA}$ of 120° and 3) a hydrogen bond lifetime $\tau \geq$ 5.0 ps. The occupancy, Occ, is defined as the total lifetime of each hydrogen bond by the production time. The events are the number of times each hydrogen bond is formed. All distances are in Å. HSP stands for a doubly protonated His.

	Distance		Hydrogen bond		
Atom pair	MD†	MD‡	Occ	$\langle\tau\rangle$(ps)	# Events
GLN 11 Hϵ_{21} - CYT O_{1P}	1.984(0.246)	2.917(0.183)	0.94	85.81	56
LYS 41 Hζ_1 - CYT O_{1P}	2.956(0.640)	3.006(0.235)	0.09	14.49	30
LYS 41 Hζ_2 - CYT O_{1P}	2.938(0.579)		0.10	13.73	39
LYS 41 Hζ_3 - CYT O_{1P}	2.811(0.571)		0.11	14.87	37
WATER - CYT O_{1P}			1.09	48.64	114
HSP 12 Hϵ_2 - CYT O_{2P}	1.877(0.192)	2.846(0.157)	0.99	218.90	23
PHE 120 HN - CYT O_{2P}	2.246(0.246)	3.204(0.224)	0.73	28.35	131
WATER 11 - CYT O_{2P}		2.714(0.117)	1.00	391.92	13
HSP 12 Hϵ_2 - CYT $O_{2'}$	2.374(0.205)	2.989(0.189)	0.36	14.44	127
LYS 41 Hζ_1 - CYT $O_{2'}$	3.040(0.717)	2.917(0.183)	0.23	23.21	50
LYS 41 Hζ_2 - CYT $O_{2'}$	3.071(0.721)		0.25	24.07	53
LYS 41 Hζ_3 - CYT $O_{2'}$	2.908(0.778)		0.32	28.01	58
HSP 119 Hδ_1 - CYT $O_{3'}$	2.602(0.230)	3.155(0.166)	0.05	11.44	24
WATER - CYT $O_{3'}$			0.52	18.46	143
HSP 119 Hδ_1 - ADE $O_{5'}$	1.851(0.117)	2.842(0.104)	1.00	638.75	8
WATER - ADE $O_{5'}$			0.14	15.12	46

† Arithmetic average hydrogen bond distance for 5 ns of the production simulation, with standard deviation in parentheses. ‡ Arithmetic average distance between the corresponding heavy atoms in 5 ns of the production trajectory, with standard deviation in parentheses.

Solvation of the substrate

The [Trans]^{TS} simulation exhibits a variety of interactions between the ligand substrate and solvent (Table 4.2). Each hydrogen-bonding atom of the CpA ligand interacts with several different water molecules in the course of the [Trans]^{TS} simulation. However, it is noteworthy that the O_{2P} equatorial oxygen only interacts with a single water molecule. The largest hydrogen bond group occupancy (1.09) corresponds to hydrogen bond interactions with O_{1P} equatorial oxygen. On the other hand, $O_{2'}$ axial oxygen does not form any hydrogen bond with water molecules, instead interacting almost exclusively with the side chain of Lys41.

Inspection of the RDF's (Figure 4.2) indicate the O_{2P} and O_{1P} present signatures of *highly ordered water interactions*. The former exhibits the largest peak among the RDF's of the [Trans]^{TS} complex, 4.3 at 2.7 Å that drops to 0.02 at 3.2 Å while O_{1P}'s RDF is 3.2 at 2.7 Å and falls to 0.2 at 3.1 Å. The observed peak value for O_{2P} is significantly larger than the peak value in the reactant model (3.1), and suggest an important solvent stabilization of the negative charge accumulated at the non-bridge oxygens in the dianionic transition state. On the other hand, the fact that the peak falls to a near 0 value in the case of O_{2P} indicates that exchange is very slow. With respect to the rest of the phosphorane oxygen atoms (axial $O_{5'}$ and $O_{2'}$ oxygens and equatorial $O_{3'}$), the RDFs present signatures of a *poorly solvated* situation that indicate less ordered solvent and more rapid exchange relative to the RDFs for the equatorial O_{2P} and O_{1P} atoms. In the [Trans]^{TS} simulation, the $O_{3'}$ and $O_{5'}$ RDFs begin to increase earlier than the corresponding reactant state simulation RDFs, while the opposite occurs for the $O_{2'}$ RDFs. The latter begins to increase for distances > 4.0 Å and presents a maximum value of 1.6 at 4.7 Å which falls to 0.6 at 5.8 Å.

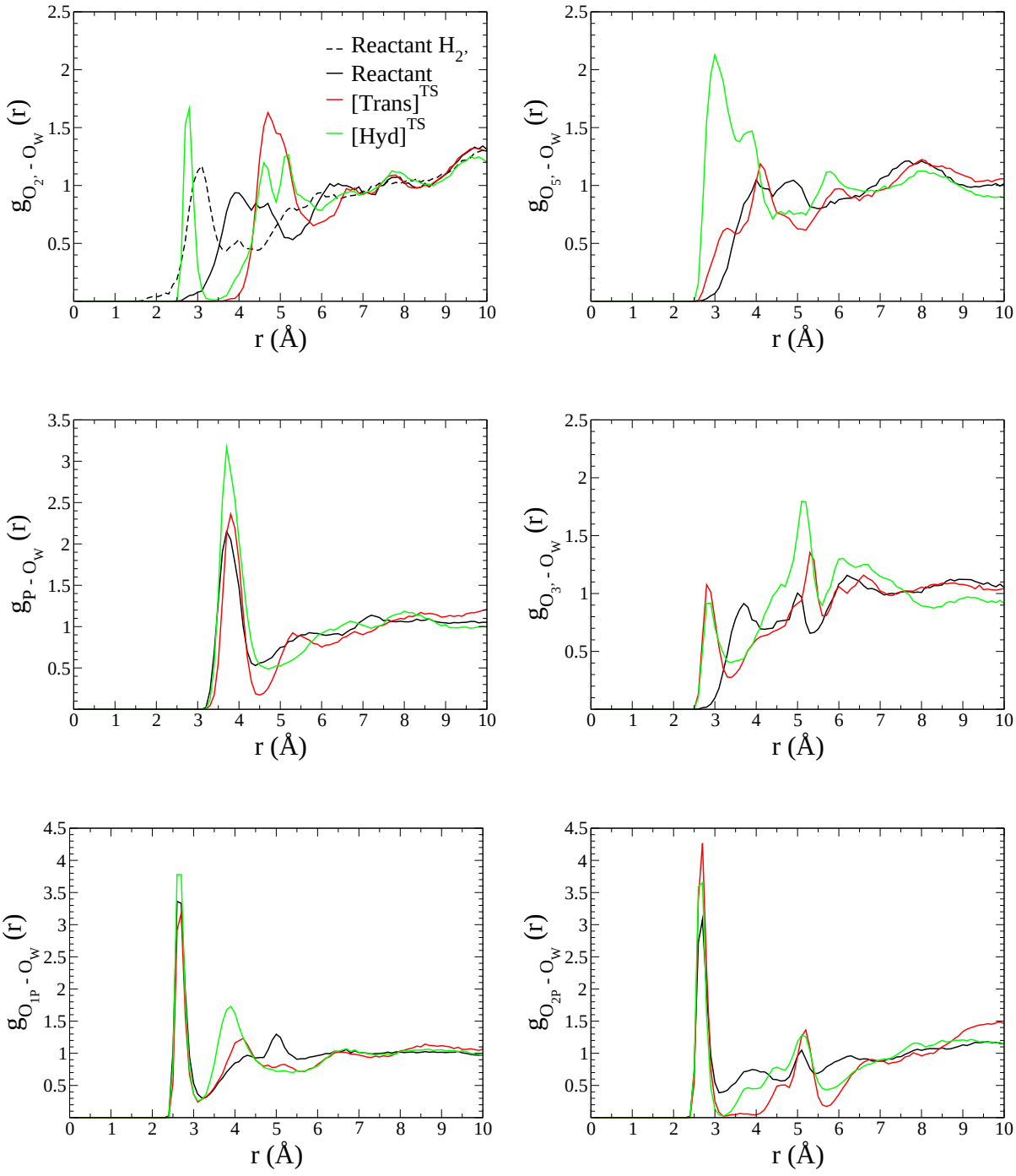
4.3.3 Hydrolysis transition state mimic, [Hyd]^{TS}

In this section we present the results for the RNase A - 3',5'-CpA hydrolysis transition state mimic model (Scheme 4.1 and Figure 4.1). The RMSD with respect to the initial conformation as a function of the simulation time is shown in Figure A.1. The RMSD reaches equilibrium after 1 ns and fluctuate stably with an average value of 1.27 ± 0.11 Å, very similar to the [Trans]^{TS} simulation.

Conformational dynamics

The time evolution of this dihedral angle is shown in Supporting Information Figure A.2 and the average value is listed in Table A.4. The fluctuation of the dihedral angle is very small signifying that the χ_C anti conformation is very stable throughout the simulation, like in the reactant model and [Trans]^{TS} model simulations. The average value is -161.6°. The cytidine ribose stays

Figure 4.2: Distribution of water around phosphorus atom, phosphate and phosphorane oxygens's atoms is shown for reactant model in black, [Trans]^{TS} model in red and [Hyd]^{TS} model in green. $O_{2'}$ is at top left corner, at top right corner $O_{5'}$/O_{3T} is depicted, phosphorus atom is in the middle left and on the right side is $O_{3'}$ oxygen, finally O_{1P} is at the botton left corner and at the right corner O_{2P} are shown.



in the North $C_{3'}$-endo pucker throughout the simulation with an average value of 15.9°. Finally, the imidazolium side chain of His119 is retained in the active A conformation like in the reactant and $[Trans]^{TS}$ models, with a mean χ_1 value of 154.5° and χ_2 value of -105.8°. The torsional dihedral angles do not undergo any significant conformational transitions (Figure A.5).

Hydrogen bond interactions

Hydrogen bond formation between the ligand and the protein residues in the active site is described in Table 4.3, in which the average distances between atoms participating in protein-ligand hydrogen bonds, with the hydrogen-bond occupancies, average time and number of events are listed. The time evolution of selected distances is plotted in Supporting Information Figure A.5 . In the $[Hyd]^{TS}$ simulation, the phosphodiester group makes strong hydrogen bond interactions with both catalytic histidines, His12 and His119, the backbone of Phe120, the side chains of Gln11 and Lys41, and solvent molecules.

The equatorial non-bridge O_{1P} atom forms a strong hydrogen bond interaction with the side chain of residue Gln11. After 2 ns, the interaction weakens and is partially replaced with an interaction with the side chain of Lys41 that gets stronger (see Table 4.3 and Figure A.5). The hydrogen bond occupancies are 0.86 and 0.57 respectively. The reactant state, $[Trans]^{TS}$ and $[Hyd]^{TS}$ simulations all show strong interactions between O_{1P} and the side chain of Gln11, but only the $[Trans]^{TS}$ and $[Hyd]^{TS}$ simulations predict a hydrogen bond interaction between O_{1P} and Lys41, this interaction being most prominent in the hydrolysis transition state mimic. The $[Trans]^{TS}$ and $[Hyd]^{TS}$ simulations also both show strong interactions between O_{1P} and side chain of Gln11, being stronger in the $[Trans]^{TS}$ model. The O_{2P} equatorial oxygen makes 3 strong interactions with residues His12, Phe120 and a water molecule. One of them, His12 $H\epsilon_2$-O_{2P}, is essentially locked for over the last 5 ns of simulation, with an average hydrogen bond distance of 1.729 Å and occupancy of 1.00 with only 9 observed exchange events and an average lifetime of 573.33 ps. On the other hand, the hydrogen bond between O_{2P} and the backbone of Phe120 is weaker, with a hydrogen bond occupancy of 0.85 and an average distance average of 2.157 Å. The O_{1P} and O_{2P} equatorial oxygens interact with water molecules throughout the simulation as in the reactant and $[Trans]^{TS}$ models, but quite remarkably, the O_{2P} interacts exclusively with a single water molecule as in the $[Trans]^{TS}$ simulation. On the other hand, the axial $O_{2'}$ oxygen makes weak interactions with the side chain of His12, Lys41 and a water molecule. The average $O_{2'}$-His12 distance is 2.488 Å hydrogen bond occupancy is 0.13, and average lifetime is 11.67 ps. The interaction between $O_{2'}$ and Lys 41 is stronger, with hydrogen bond group occupancy of 0.46. Analysis of the interactions of Lys41 suggest this residue interacts with the $O_{2'}$ axial oxygen over the first

Table 4.3: Statistics of protein-ligand and water-ligand hydrogen-bonds observed in the simulation of the $[Hyd]^{TS}$ model. The criteria for the existence of hydrogen bonds are 1) a maximum H $\cdots$ A (acceptor) distance of 2.4 Å 2) a minimum $\widehat{DHA}$ of 120° and 3) a hydrogen bond lifetime $\tau \geq 5.0$ ps. The occupancy, Occ, is defined as the total lifetime of each hydrogen bond by the production time. The events are the number of times each hydrogen bond is formed. All distances are in Å. HSP stands for a doubly protonated His.

	Distance		Hydrogen Bond		
Atom pair	MD†	MD‡	Occ	$\langle\tau\rangle$(ps)	# Events
GLN 11 Hϵ_{21} - CYT O_{1P}	2.053(0.388)	2.963(0.284)	0.86	62.54	71
LYS 41 Hζ_1 - CYT O_{1P}	2.957(0.776)	3.029(0.385)	0.21	62.65	17
LYS 41 Hζ_2 - CYT O_{1P}	2.988(0.627)		0.14	46.56	16
LYS 41 Hζ_3 - CYT O_{1P}	2.814(0.668)		0.22	59.48	19
WATER - CYT O_{1P}			1.47	39.38	193
HSP 12 Hϵ_2 - CYT O_{2P}	1.729(0.119)	2.723(0.099)	1.00	573.33	9
HSP 119 Hδ_1 - CYT O_{2P}	3.102(0.308)	3.863(0.288)	0.00	10.00	1
PHE 120 HN - CYT O_{2P}	2.157(0.259)	3.120(0.231)	0.85	50.17	88
WATER 397 - CYT O_{2P}		2.695(0.131)	0.99	341.33	15
HSP 12 Hϵ_2 - CYT $O_{2'}$	2.488(0.197)	3.162(0.171)	0.13	11.67	57
LYS 41 Hζ_1 - CYT $O_{2'}$	3.173(0.683)	3.205(0.383)	0.12	55.47	11
LYS 41 Hζ_2 - CYT $O_{2'}$	3.274(0.821)		0.15	83.88	9
LYS 41 Hζ_3 - CYT $O_{2'}$	3.135(0.839)		0.19	64.65	15
WATER 6130 - CYT $O_{2'}$		13.333(13.670)	0.55	188.34	15
HSP 119 Hδ_1 - CYT $O_{3'}$	2.165(0.230)	2.961(0.149)	0.79	31.77	128
WATER - CYT $O_{3'}$			0.40	14.48	144
HSP 119 Hδ_1 - CYT O_{3T}	2.060(0.296)	2.991(0.252)	0.88	76.77	59
WATER - CYT O_{3T}			0.19	12.47	79

† Arithmetic average hydrogen bond distance for 5 ns of the production simulation, with standard deviation in parentheses. ‡ Arithmetic average distance between the corresponding heavy atoms in the production trajectory, with standard deviation in parentheses.

2110 ps, while in last 2890 ps it interacts with the O_{1P} equatorial oxygen. In turn, during the last 2890 ps of the simulation, $O_{2'}$ hydrogen bonds with a water molecule (average heavy atom distance of 2.859 Å), while there was no interaction between $O_{2'}$ and water molecules during the first 2110 ps. The entry of this water molecule into the active site is thus connected with the movement of Lys41 toward the O_{1P} equatorial oxygen. Finally, the side chain of residue His119 makes 2 strong interactions with the $O_{3'}$ equatorial oxygen and the terminal O_{3T} axial oxygen, with hydrogen bond occupancies of 0.79 and 0.88, respectively, and average distances of 2.165 Å and 2.060 Å respectively.

Solvation of the substrate

Each of the phosphoryl oxygens of the CpA ligand form hydrogen bonds with residues of the enzyme during the course of the simulation (Table 4.3), with the exception of the O_{2P} equatorial oxygen and $O_{2'}$ axial oxygen. The overall hydrogen bond pattern is similar to that of the $[Trans]^{TS}$ simulation, with a few notable exceptions. The hydrogen bonding interactions between O_{1P} and Lys41 is strengthened in the $[Hyd]^{TS}$ simulation (group *Occ* 0.57) relative to the $[Trans]^{TS}$ simulation (group *Occ* 0.30), whereas the interaction of O_{1P} with Gln11 is reduced. The O_{2P} was observed to make transient hydrogen bonding interaction with the protonated His119 residue not present in the $[Trans]^{TS}$ simulation, whereas the interaction of $O_{2'}$ with the protonated His12 and Lys41 residues are less pronounced. Moreover, in the $[Hyd]^{TS}$ simulation, the $O_{3'}$ atom of cytidine acquires a more significant hydrogen bond with His119 with *Occ* 0.79, whereas this interaction is greatly reduced in the $[Trans]^{TS}$ simulation (*Occ* 0.05). In the $[Hyd]^{TS}$ simulation, the proposed general acid His119 is observed to make a strong hydrogen bond with the activated water nucleophile/terminal hydroxyl group of cytidine (O3T in Table 4.3). The O_{1P}, O_{2P}, $O_{2'}$ and $O_{3'}$ atoms maintain additional strong hydrogen bond interactions with water molecules in the $[Hyd]^{TS}$ simulation. Unlike the $[Trans]^{TS}$ simulation, the $[Hyd]^{TS}$ simulation exhibits a significant solvent interaction with the $O_{2'}$ axial oxygen. The solvent interaction with the largest lifetime average (341.33 ps) corresponds to a hydrogen bond between a water molecule and the O_{2P} equatorial oxygen, while the largest hydrogen bond group occupancy (1.47) corresponds to the O_{1P} equatorial oxygen.

Radial Distribution Function analysis reveal a *highly ordered water interaction* for the O_{1P}, O_{2P} and $O_{2'}$ oxygens. The non-bridging equatorial oxygens O_{1P} and O_{2P} have the highest probability in their RDF with water oxygens, with maximum value of 3.8 and 3.7, respectively, both located at 2.7 Å. The minima for O_{2P} is considerably lower than that for O_{1P}, consistent with the larger number of solvent exchange events for the latter. The $g_{O_{2'}-O_W}$ RDF for the $[Hyd]^{TS}$ model has an ordered first solvation peak (1.7 at 2.8 Å), unlike that of the reactant and $[Trans]^{TS}$ models. Additionally, the axial oxygen in the $[Hyd]^{TS}$ model (shown in Figure 4.2 as $O_{5'}$) is solvent exposed, and exhibits a broad solvation peak with maximum of 2.1 at 3.0 Å which falls to 0.7 at 4.4 Å. Comparison of these peak values with the RDF's of the reactant and $[Trans]^{TS}$ simulations indicates the preferential solvent stabilization at this position in the hydrolysis transition state mimic. Together with the g_{P-O_W} RDF, the hydrolysis transition state is more highly solvated than the reactant state and $[Trans]^{TS}$ model.

4.4 Discussion

The present work presents a series of molecular dynamics simulations of RNase A - 3',5'-CpA complex and its transphosphorylation and hydrolysis transition state mimic models. These simulations employ a set of recently developed force field parameters for oxyphosphorane transition states or intermediates which are consistent with the CHARMM27 all-atom empirical force field for nucleic acids.[218] The parameters were developed based on density-functional calculations, with partial atomic charges derived from electrostatic potential fitting and Lennard-Jones parameters reproducing interaction energies with water molecules. Bonded energy terms (bond, angle and torsion parameters) were also derived from the density-functional calculations and renormalized to maintain compatibility with the existing CHARMM27 parameters for standard residues. Thus, the parameters used in this work reproduce the main geometrical and electronic characteristics of important oxyphosphorane intermediates based on density-functional theory, but within a molecular mechanical approximation. This has allowed us to provide key structural information on how these reactive intermediates are structurally recognized in the active site of RNase A, using classical MD simulations with reasonable time scales. One has to consider that this type of data can be difficult to be accessed from experiments, due to the difficulty of trapping reactive intermediates, and to the inherent limitations of using vanadate species as oxyphosphorane analogs.[214] The main advantage of the use of a classical force field is the computational efficiency that allows for sufficiently long time scales as to get statistically relevant results. However, there are also important limitations of the present approach that should be remarked upon. For instance, a non-polarizable force field is used in a system significantly charged, which results in a lack of significant dispersion effects in the present model. In addition, due to the use of a classical force field, there is no possibility of providing with a reaction profile as a function of covalent bond formation and cleavage, and one is limited to the analysis of structural information. Still important insights on the reaction mechanism can be inferred from the structural data presented here. In the present section, we make a discussion on the relevance of our structural findings for the classical catalytic reaction mechanism. In this so-called classical mechanism, RNase A cleaves RNA substrates in a two-step process, by general acid-base catalysis. In the first step (transphosphorylation, $[\text{Trans}]^{TS}$ structure), the RNA chain is cleaved by a phosphate ester exchange, in which the 2'-hydroxyl group of ribose attacks the phosphate ester linkage (P–$O_{2'}$ bond is formed) and the $O_{5'}$ oxygen of the next nucleotide is ejected (P–$O_{5'}$ bond is cleaved). In the second step (hydrolysis $[\text{Hyd}]^{TS}$ structure), the P–$O_{2'}$ bond of the 2',3' cyclic phosphate is cleaved by an attacking activated water molecule to generate a 3'-terminal phosphate group and regenerate the 2'-hydroxyl group. The principal catalytic groups for these reactions are the side chains of His12

and His119 and the cationic amino group of Lys41. In the so-called classical mechanism, the most accepted one among experimentalists, His12 acts as a general base, while His119 acts as an acid in the transphosphorylation step, and in the hydrolysis step His12 acts as an acid, while His119 acts as a base. However, the exact role of each of these residues is still a matter of discussion, and different mechanisms have been proposed.[29,60,61,73,79,80,87,209] We analyze now the structural information relevant from our dynamics with respect to these three important residues and the role of solvent molecules in the stabilization of oxygen oxyphosphoranes.

Regarding Lys41, our MD simulations demonstrate that this residue is more flexible in the reactant model than in the transition state mimic models. The cationic side chain of Lys41 hydrogen bonds with O_{1P} and O_{2P} phosphate oxygens as well as cytosine $O_{2'}$ oxygen in the reactant model, while in the $[Trans]^{TS}$ and $[Hyd]^{TS}$ models Lys41 only forms significant hydrogen bonds to the $O_{2'}$ axial oxygen and O_{1P} equatorial oxygen (Tables 4.1, 4.2 and 4.3). The interaction with the equatorial oxygen gets stronger proceeding from the reactant model to the transphosphorylation and hydrolysis transition state models (hydrogen bond group occupancy of 0.03-0.05, 0.30 and 0.57 respectively). This is consistent with the early supposition that Lys41 provides preferential electrostatic stabilization of the reactive intermediate/transition state .[214,235] Whereas the interaction of Lys41 with the non-bridge equatorial oxygen becomes stronger in proceeding from the reactant to the transition states, the interaction with the $O_{2'}$ oxygen becomes slightly weaker. The interaction between Lys41 and the $O_{2'}$ position is of interest, as Lys41 is known to have a reduced pK_a (8.6-9.1) ,[236] has been observed crystallographically to undergo structural change in the pH range 8.0-8.8,[237] and has even been suggested to act as general base/acid catalyst instead of His12 .[60,61,87] It is possible that Lys41 might have a depressed pK_a due to interaction with nearby hydrophobic groups, as has been observed for a lysine residue in enolase .[238] However, to reconcile this question would require the calculation of the pK_a value of Lys41 (and possibly other residues) using a combined quantum mechanical/molecular mechanical (QM/MM) simulation and free energy perturbation approach. This is beyond the scope of the current work that focuses on the structure and dynamics at different stages along the reaction coordinate. It is, nonetheless, of importance to note that the simulations predict a stable hydrogen bond interaction with the $O_{2'}$ nucleophile in the reactant state that is preserved in proceeding to the transition state. In the ionized form, Lys41 could assist in catalysis by increasing the acidity of the nucleophile and facilitating proton transfer to His12. Alternately, the interaction of Lys41 with the nucleophile might implicate a more active role as a general base catalyst, although this would require further simulation and QM/MM calculations to fully test.

On the other hand, His12 has a very weak interaction with the CpA reactant model. Instead, in both transition state mimics, His12, in protonated

form, makes strong hydrogen bond interactions with the O_{2P} equatorial oxygen. The hydrogen bond occupancy is 0.99 for the $[Trans]^{TS}$ mimic and 1.0 for $[Hyd]^{TS}$ mimic (Tables 4.2 and 4.3). Thus, the simulation results suggest that interaction with His12 may be important to stabilize the excess of negative charge at O_{2P} as the reaction proceeds. In addition, the O_{2P} equatorial oxygen makes two strong hydrogen bond interactions in the transition state mimic models that provide stabilization of its negative charge: one hydrogen bond interaction with the main chain of residue Phe120, and the other with a water molecule. On the other hand, the interaction between His12 and the $O_{2'}$ axial oxygen is not strong in either of the transition state mimic simulations. The hydrogen bond occupancies are 0.36 for the $[Trans]^{TS}$ and 0.13 for the $[Hyd]^{TS}$. This last occupancy is not what one would expect if His12 were to play the role of a general acid catalyst as in the classical mechanism.[73] Nonetheless, this small occupancy might be partially offset by an indirect interaction via a bridging water molecule between His12 and the $O_{2'}$ axial oxygen that becomes evident in the last 3 ns of the simulation.

The results for Lys41 and His12 suggest that Lys41 may play a more versatile, and possibly more active role in RNase A catalysis, which could lead to its use as a base catalyst in the transphosphorylation step and an acid catalyst in the hydrolysis step[29,60,87] instead of His12. A recent QM/MM study on the hydrolysis step of the reaction by Elsässer *et al.*[61] lends further support to the interpretation of our results, pointing to Lys41 as the acid catalyst of the reaction and His12 as a hydrogen-bond stabilizer of the excess of negative charge at the equatorial phosphoryl oxygen. Our results with simulations on the nanoscale support these conclusions and predict a similar behavior for the transphosphorylation step of the reaction.

The interaction between imidazolium side chain of His119 and the $O_{3'}$ equatorial oxygen increases as the reaction proceeds in the forward direction, from a hydrogen bond occupancy of 0.03 in the reactant model to 0.79 for the $[Hyd]^{TS}$ mimic. In the case of the axial $O_{5'}$ oxygen, the interaction increases from a hydrogen bond occupancy of 0.81 for the reactant model to 1.0 for the $[Trans]^{TS}$ mimic, and it decreases slightly to 0.88 for the $[Hyd]^{TS}$ model. Finally, it is remarkable that the side chain of residue His119 remains in its active A conformation that promotes catalysis[50,62] in each simulation. In this sense, His119 and the adenine rings form continuous $\pi - \pi$ stacking interactions that contribute to the stabilization of the His119 A orientation and the adenine ring anti orientation, and these interaction therefore seems to be fundamental for catalysis.

On the other hand, there are important differences in the solvation of the two transition state mimics. The g_{P-O_W} RDF (Fig. 4.2) shows a higher degree of solvation (as reflected by the peak height of the RDF and first coordination number) for the $[Hyd]^{TS}$ mimic simulation than for the other two simulations. The O_{3T} axial oxygen is clearly more accessible to solvent in the $[Hyd]^{TS}$ simulation. Its RDF maximum (2.1) is at 3.0 Å while $[Trans]^{TS}$

model's maximum (1.2) is at 4.1 Å. This is also seen if we analyze the running coordination number up to the first minimum in the oxygen-RDF:

$$n(R) = 4\pi\rho_W \int_0^R g(r) r^2 dr \,;\, \rho_W \equiv \text{Water average density} \tag{4.1}$$

In the case of $g_{O_{5'}/O_{3T}-O_W}$ RDF, n(R) is 0.75 for $[\text{Trans}]^{TS}$ and 3.19 for $[\text{Hyd}]^{TS}$. Except for O_{2P} (see below) the rest of the atoms follow the same tendency with running coordination numbers up to the first shell (around 3.1-3.6 Å) larger for $[\text{Hyd}]^{TS}$ than for $[\text{Trans}]^{TS}$, namely, 1.59/1.20 (O_{1P}), 0.55/0.00 ($O_{2'}$) and 0.84/0.79 ($O_{3'}$) for $[\text{Hyd}]^{TS}$ /$[\text{Trans}]^{TS}$ respectively. The only exception is found for O_{2P} for which the same first-shell running coordination number is found at both transition state mimics, namely, 1.00.

Finally, we would like to highlight the behavior of the RDF's of O_{2P}. In both transition state mimic models, its RDF is 0 after the large peak between 2.5 and 3.0 Å. Also remarkable, is the zero probability to find a water molecule between 3 and 4 Å in the $[\text{Trans}]^{TS}$ model, indicative of a rigid structural solvation shell in the vicinity of this atom. Thus, based on these simulations it seems that this water molecule could play a structural role in the stabilization of oxyphosphorane species. In this sense, Pelmenschikov *et al.*[239] has demonstrated by DFT calculations that for the stabilization of the salt bridge between deprotonated DNA phosphate groups and protonated basic groups, such as histidines, requires the direct interactions of these groups with water molecules. In particular, the specific interaction of a water molecule with the phosphoryl oxygen is needed to stabilize the negative charge at this oxygen and avoid proton transfer from the positively charged histidine. In oxyphosphoranes this tendency is probably enhanced due to the built up of negative charge at the phosphoryl oxygens, and our results points to this interpretation of an enhanced necessity of stabilization of the salt-bridge between the histidine and the oxyphosphorane by a direct interaction of O_{2P} with a water molecule.

4.5 Conclusions

In the present paper, we report the results of MD simulations of RNase A - 3',5'-CpA complex and its transphosphorylation and hydrolysis transition state mimic models using recently developed force field parameters for phosphoryl transfer intermediates.[218] The aim of these studies is to determine structural relaxation, and differential solvation that occurs at discreet stages of the transesterification and cleavage reaction, in particular, at the CpA phosphate reactant, and at the two pentacovalent oxyphosphorane structures that represent the transition states or high-energy intermediates in the transphosphorylation and hydrolysis steps of the reaction. Simulations were performed with explicit solvation, with rigorous electrostatics, and for

several nanoseconds, which allow to make a detailed analysis of the main changes in hydrogen bond patterns and interaction with the solvent among the three structures.

Based on our simulations we can conclude that there are no major differences between the backbone structure of the studied three structures. Besides, the relative orientation of the bases, denoted by their glycosyl dihedral angle conformations, is very stable and it is not affected by the nature of the substrate. In addition, it is quite remarkable that the active A conformation of His119 is maintained for the three structures studied and along the whole production run.

Analysis of the hydrogen bond occupancies between the substrates and RNase A shows that the residues directly interacting with the substrates are His12, His119, Lys41, Gln11, Phe120 and a variety of water molecules. However, there are important changes in the hydrogen bond pattern between the reactant and transition state analogs that shed light on the reaction mechanism and role of specific residues in transition state stabilization. The formation of pentacovalent phosphorane anchors His12 to interact with the equatorial O_{2P} oxygen, especially in the case of the $[Hyd]^{TS}$ simulation, and causes Lys41 to loosen the interaction with the equatorial O_{2P} oxygen and strengthen the interaction with the equatorial O_{1P} oxygen while maintaining significant hydrogen bond occupancy with axial $O_{2'}$ atom. This result suggests that the role of His12 and Lys41 in the transphosphorylation and hydrolysis reactions could be interchanged to the ones proposed in the classical mechanism, and is in qualitative agreement with conclusions of other theoretical calculations[29,60,61,87] in which Lys41 has been claimed to possibly act as the acid catalyst in the hydrolysis reaction.

Analysis of substrate interaction with water molecules has been accomplished by the calculation of hydrogen bond occupancies between the substrate oxygen atoms and specific water molecules and by the analysis of the radial distribution functions around key atoms of the phosphate and phosphorane moieties. Simulations results point to a higher solvation of the $[Hyd]^{TS}$, compared to $[Trans]^{TS}$ or the reactant model. In all models, the equatorial O_{1P} and O_{2P} atoms are the most solvated phosphate/phosphorane atoms. It is remarkable that the formation of phosphorane structures leads to the encapsulation of a water molecule hydrogen bonded directly to O_{2P} that shows no exchange with other water molecules in the time scale of the simulations, and suggests an important role of this water molecule in the stabilization of the negative charge accumulated at this equatorial phosphorane oxygen in both $[Trans]^{TS}$ and $[Hyd]^{TS}$ simulations.[239]

The work presented in this paper could also be taken as the base for the future analysis of MD simulations of this type of oxyphosphorane transition state analogues bound to RNase A mutants that are known to affect the enzyme turnover rate, namely, H119N, H119A, H119D, H12A, H12E, H12D, K41R or K41A[58,240–242] which decrease k_{cat} between two and four orders

of magnitude. In addition, the present data could also be used as reference structures to analyze how chemical changes at the oxyphosphorane itself (like sulfur substitution, i.e., thio effects) affects their recognition by RNase A. Work along these lines is under progress.

Chapter 5

Thio effects on Bovine Pancreatic Ribonuclease A - CpA Transition State-like Complexes

Abstract

The mechanisms of enzymes are intimately connected with their overall structure and dynamics in solution. Experimentally, it is considerably challenging to provide detailed atomic level information about the conformational events that occur at different stages along the chemical reaction path. Here, theoretical tools may offer new potential insights that complement those obtained from experiments that may not yield an unambiguous mechanistic interpretation. In this study, we apply thio-effects to molecular dynamics simulations of bovine pancreatic ribonuclease A, an archetype ribonuclease, in order to study the conformational dynamics, structural relaxation, and differential solvation that occurs at discreet stages of the transesterification and cleavage reaction. Simulations were performed with explicit solvation with rigorous electrostatics, and utilize recently developed molecular mechanical force field parameters for transphosphorylation and hydrolysis transition state analogs. Herein, we present results for the enzyme complexed with the dinucleotide substrate cytidilyl-3',5'-adenosine (CpA) in transphosphorylation and hydrolysis transition thio-substituted states. A detailed analysis of active site structures and hydrogen bond patterns are presented and compared. The integrity of the overall backbone structure is preserved in the simulations, and support a mechanism whereby His12 stabilizes accumulating negative charge at the transition states through hydrogen bond donation to the non-bridge oxygens. Lys41 is shown to be highly versatile along the reaction coordinate, and can aid in the stabilization of the

dianionic transition state, while being poised to act as a general acid catalyst in the hydrolysis step.

5.1 Introduction

In this work we investigate for the first time the structural and dynamical behavior of RNase A - CpA transition state-like complexes' thio substitutions in solution with molecular dynamics simulations: 1) two monothio substitutions of non-bridging phosphorane oxygens of transphosphorylation transition state/intermediate mimic model, namely, $[S{:}O_{1P}]^{Trans}$ and $[S{:}O_{2P}]^{Trans}$ (see Scheme 5.1); 2) a dithio substitution of non-bridging phosphorane oxygens of transphosphorylation transition state/intermediate mimic model, $[S{:}O_{1P},O_{2P}]^{Trans}$; 3) two monothio substitutions of non-bridging phosphorane oxygens of hydrolysis transition state/intermediate mimic model, namely, $[S{:}O_{1P}]^{Hyd}$ and $[S{:}O_{2P}]^{Hyd}$; and 4) a dithio substitution of non-bridging phosphorane oxygens of hydrolysis transition state/intermediate mimic model, $[S{:}O_{1P},O_{2P}]^{Hyd}$. Transition state mimics were modeled by the CHARMM force field parameters for reactive phosphoryl transfer intermediates.[218] The simulations provide insights on the flexibility of the bound ligands and the surrounding active site residues in stabilizing the trigonal bipyramidal geometry of the key transition state/intermediates of the reaction. Moreover, the recent characterization of a stable dianionic phosphorane intermediate in RNase A[61] remarks the importance of studying pentavalent oxyphosphorane structures and its derivatives within RNase A, with methods that allows for sufficient long-time dynamics to evaluate potential changes in the structure of key active site residues upon phosphorane formation, and the corresponding thio substitutions.

The chapter is organized as follows: the Computational Details section describes the force field models and simulation set up. The Results section shows the simulation results obtained for the transphosphorylation's and hydrolysis' transition state mimic models. Three types of thio substitutions have been considered for each case: *i)* thio substitution at the O_{1P} position, denoted as monothio $S{:}O_{1P}$ substitution, *ii)* thio substitution at the O_{2P} position, referred as monothio $S{:}O_{2P}$ substitution, and *iii)* double substitution at the O_{1P} and O_{2P} positions, denoted as the dithio $S{:}O_{1P},O_{2P}$ substitution. The Discussion section places the simulation results into a broader biological context, and compares them. Finally, the Conclusion section summarizes the main points of the paper and outlines future research directions.

5.2 Computational details

The starting structure for all-atom periodic boundary conditions Molecular Dynamic simulations of RNase A complexed with cytidilyl-3',5'-adenosine

Scheme 5.1: Schematic representation of the studied models. On the top row, transition state mimic thio substituted models are represented ($[S:O_{1P}]^{Trans}$, $[S:O_{2P}]^{Trans}$ and $[S:O_{1P},O_{2P}]^{Trans}$), while on the second row the hydrolysis transition state mimic thio substituted complexes are shown ($[S:O_{1P}]^{Hyd}$, $[S:O_{2P}]^{Hyd}$ and $[S:O_{1P},O_{2P}]^{Hyd}$).

(CpA) was taken from Protein Data Bank with accession number 1RPG,[50] where the structure was determined at a resolution of 1.4 Å. The CpA structure was built from the crystal results for the inhibitor complex, d(CpA), by introducing the 2' oxygen into the deoxy sugar. The transphosphorylation and hydrolysis transition state mimics were built from the previous complex using a "patch" to introduce a dianionic pentavalent cyclic phosphorane transition state mimic with 2' (nucleophile) and 5' (leaving group) positions oriented axially.[218] As well as, thio substitutions. The position of the hydrogen atoms were determined using HBUILD facility in the program CHARMM,[126,127] Ver. c32a1. Atomic charges, van der Waals, and force-field parameters corresponded to the all-atom CHARMM27 nucleic acid force field[144–146] with extension to transition state mimics[218] and TIP3P water model.[225] Residues His12 and His119 are protonated, in the states appropriate for their proposed roles for the transphosphorylation reactions. The general base has already accepted a proton from the nucleophile, and the general acid is ready to donate a proton to the leaving group. All other residues were assumed to be in their standard ionization sates in solution at neutral pH.

The aqueous solvation environment was created by retaining the 164 crystallographic water positions, and further solvating the system with bulk water molecules and ions. Additional water molecules were included by overlaying a pre-equilibrated rhombic dodecahedral cell of 8,346 TIP3P water molecules, centering it at the center of mass of the complex using statistics methodology, and deleting water molecules with oxygen atom lying in a radius of 2.8 Å from any non-hydrogen atom of the protein, enzyme or crystallographic waters. The ion atmosphere consisted of Na^+ and Cl^- ions that were added at random positions to neutralize the net +6 charge of the solute and reach the physiologic extracellular concentration of 0.15 M. The transphosphorylation transition state mimic models simulation corresponded to the entire protein (1,860 atoms), the ligand (62 atoms), 7,384 water molecules, 6 Cl^- ions to neutralize the net positive charge of the solute and 42 additional counter and co-ions (21 Cl^{-1} and 21 Na^{+1} ions). A total of 24,206 atoms formed the systems under study. The hydrolysis transition state mimic models simulations were prepared in an analogous fashion.

Periodic boundary conditions were used along with the isothermal-isobaric ensemble (NPT) at 1 atm and 298 K via an extended system pressure algorithm[226] and Nosé-Hoover thermostat.[227,228] Electrostatic interactions were treated rigorously using the smooth particle mesh Ewald method (PME)[135,137,138] with a κ value of 0.35 $Å^{-1}$ and direct-space cut-off of 12 Å, 72 FFT grid points for each of the lattice directions, and 4th-order B-spline interpolation for spreading the atomic charges to the FFT grid. Nonbonded interactions were performed using a cutoff of 12 Å with shifted van der Waals potential to smooth the Lennard-Jones term. Nonbond pair lists were maintained out to 13.0 Å and updated heuristically. Newton's equations of motion were in-

tegrated numerically using the leapfrog Verlet algorithm within a timestep of 1 fs,[167] while atomic coordinates were saved for analysis every 0.5 ps. Covalent bond lengths involving hydrogen were constrained using the SHAKE algorithm.[229]

Water molecules were initially relaxed for 5000 steps of steepest descents energy minimization keeping all solute atom and ion positions restrained to their initial coordinates using a harmonic potential constant of 10 kcal·mol^{-1}·Å^{-2}. The restraints on the ions were then released, and the solvent (water and ions) were relaxed with 5000 steps of conjugate gradient minimization, keeping the solute positions restrained. From this starting point, restrained molecular dynamics was performed starting at 0 K and heating to 298 K over 20 ps, and carried out to 220 ps, during the course of which the harmonic restraints on the solute atoms were slowly released. The production phase molecular dynamics then began, and covered 6.0 ns. All the simulations were performed with the CHARMM biomolecular program,[126,127] Ver. c32a1.

5.3 Results

This section presents molecular dynamics simulation results for thio-substituted oxyphosphorane compounds with RNase A, taking into account thio substitutions of the non-bridging equatorial phosphorane oxygens for the transphosphorylation and hydrolysis phosphorane structures, that can be taken as a prototype of the transition state structure or high-energy phosphorane reaction intermediates that RNase A has to stabilize in order to catalyze this reaction. The combination of these possibilities leads to a total of six different RNase A – thio-oxyphosphorane complexes treated in this work, namely: *i)* the transphosphorylation's transition state mimic monothio substituted ([S:O$_{1P}$]^{Trans} and [S:O$_{2P}$]^{Trans}) complexes, and the dithio substituted one ([S:O$_{1P}$,O$_{2P}$]^{Trans}), and *ii)* the hydrolysis' transition state mimic monothio substituted [S:O$_{1P}$]^{Hyd} and [S:O$_{2P}$]^{Hyd} complexes, and the dithio substituted [S:O$_{1P}$,O$_{2P}$]^{Hyd} (see Scheme 5.1). First, we analyze the results for the transphosphorylation's transition state/intermediate mimic complexes, and then, the hydrolysis' transition state mimic models. The results are compared with those for the non-substituted phosphorane-like transition state mimics characterized previously[243] (see Table 5.1 and Section 4), which are referred as [native]^{Trans} and [native]^{Hyd} structures.

First, the root-mean-square deviation (RMSD) and the ligand conformation are shown. Special attention is focused in describing the ligand conformation, using the glycosyl dihedral angles $\chi_{C(O_{4'}-C_{1'}-N_1-C_2)}$ and $\chi_{A(O_{4'}-C_1-N_9-C_4)}$, which describe respectively, the orientation of the cytosine and adenine rings with respect to the sugar,[230,231] whereas, the pseudorotation phase angle P[232] describes the sugars' ring conformation, and

Table 5.1: Phosphorane structure. All distances are in Å and angles in degrees. Standard deviation is in parentheses

Atom pair	[native]$^{\text{Trans}}$ [243]	[S:O_{1P}]$^{\text{Trans}}$	[S:O_{2P}]$^{\text{Trans}}$	[S:O_{1P},O_{2P}]$^{\text{Trans}}$	[native]$^{\text{Hyd}}$ [243]	[S:O_{1P}]$^{\text{Hyd}}$	[S:O_{2P}]$^{\text{Hyd}}$	[S:O_{1P},O_{2P}]$^{\text{Hyd}}$
P - $O_{2'}$	1.97(0.04)	1.979(0.041)	2.008(0.043)	2.014(0.042)	1.96(0.04)	1.992(0.042)	2.001(0.061)	2.017(0.042)
P - $O_{5'}$	1.76(0.05)	1.767(0.045)	1.770(0.045)	1.787(0.045)	1.72(0.05)	1.770(0.046)	1.766(0.061)	1.779(0.046)
P - $O_{3'}$	1.74(0.04)	1.746(0.038)	1.748(0.039)	1.753(0.038)	1.73(0.04)	1.751(0.038)	1.747(0.055)	1.760(0.036)
P - X_{2P}	1.53(0.02)	1.525(0.022)	2.064(0.037)	2.074(0.037)	1.52(0.02)	1.529(0.022)	2.071(0.059)	2.075(0.036)
P - X_{1P}	1.52(0.02)	2.059(0.037)	1.525(0.022)	2.072(0.037)	1.52(0.02)	2.076(0.037)	1.527(0.041)	2.084(0.037)
$O_{5'}$ - P - $O_{2'}$	164.2(2.4)	164.1(2.5)	164.7(2.5)	163.9(2.4)	164.2(2.6)	164.9(2.5)	163.7(2.4)	164.9(2.6)
$O_{5'}$ - P - $O_{3'}$	81.8(2.3)	82.2(2.3)	81.9(2.4)	82.3(2.3)	81.7(2.6)	83.2(2.5)	81.9(2.3)	84.2(2.5)
$O_{5'}$ - P - X_{2P}	95.8(3.4)	95.4(3.4)	95.5(3.5)	95.2(3.3)	94.3(4.1)	97.1(3.6)	96.0(3.3)	94.8(3.4)
$O_{5'}$ - P - X_{1P}	98.7(3.4)	97.5(3.3)	96.6(3.5)	98.0(3.2)	96.4(3.7)	98.4(3.5)	100.3(3.1)	101.0(3.0)
$O_{2'}$ - P - $O_{3'}$	83.2(1.5)	82.6(1.6)	83.6(1.6)	82.4(1.5)	83.5(1.6)	82.6(1.5)	82.9(1.5)	82.1(1.5)
$O_{2'}$ - P - X_{2P}	89.1(2.8)	90.2(2.8)	88.6(2.8)	89.4(2.7)	89.7(2.8)	88.9(2.9)	86.6(2.8)	87.4(2.8)
$O_{2'}$ - P - X_{1P}	89.5(2.8)	90.9(2.7)	92.6(2.8)	91.6(2.7)	92.5(2.8)	89.4(3.1)	91.1(2.8)	89.8(2.6)
X_{2P} - P - X_{1P}	127.4(1.9)	124.6(4.7)	123.5(4.4)	123.9(4.4)	127.7(1.9)	121.5(5.3)	122.5(4.3)	121.1(4.9)
X_{2P} - P - $O_{3'}$	116.7(3.1)	122.3(3.6)	113.5(4.1)	119.2(4.1)	117.2(3.2)	122.9(3.9)	115.1(3.9)	118.0(3.8)
X_{1P} - P - $O_{3'}$	115.2(3.2)	112.6(4.3)	122.6(3.9)	116.4(4.0)	114.7(3.2)	114.7(4.9)	121.5(3.9)	119.7(4.5)
$O_{3'}$ - X_{2P} - X_{1P} - P	-6.0	-4.2	-3.3	-4.1	-2.9	-6.1	-6.3	-6.7

the backbone torsion angles ($\alpha, \beta, \gamma, \delta, \epsilon$ and ζ) characterize the conformation of the sugar-phosphate backbone (see Scheme 5.1). Then, we discuss the geometry of the active site, focusing attention on the conformation of key residues and their interaction with the substrate and solvent molecules. The side chain of residue His119 can adopt two conformations denoted as A ($\chi_{1(N-C_\alpha-C_\beta-C_\gamma)} \sim 160°$) and B ($\chi_1 \sim -80°$),[62] which are related by a 142° rotation about $C_\alpha - C_\beta$ bond and 38° rotation about the $C_\beta - C_\gamma$ bond. Conformation A is the "active" conformation that promotes catalysis, whereas conformation B is an "inactive" conformation.[50] On the other hand, the criteria for the existence of hydrogen bonds of the form $D - H \cdots A$, where D and A are the hydrogen bond donor and acceptor, respectively, are 1) a maximum $H \cdots O$ (oxygen) distance R of 2.4 Å or $H \cdots S$ (sulfur) distance R of 3.0 Å, 2) a minimum $\widehat{DHA}$ angle of 120° and 3) a hydrogen bond lifetime $\tau \geq 5.0$ ps. Finally, Radial Distribution Functions (RDFs) of water molecules around phosphorane oxygens and sulfurs are also calculated, due the importance of the role of specific water molecules and bulk solvent effects in stabilizing phosphorane structures .[244] The RDF's are denoted as g_{XY}(r) where the subscript XY refers to the distribution of Y atoms around X atoms.

5.3.1 Transphosphorylation's transition state mimic [S:O_{1P}]^{Trans} model

In this subsection, results are presented for the simulations of RNase A - 3',5'-CpA transphosphorylation phosphorane transition state/intermediate mimic [S:O_{1P}]^{Trans} complex (see Scheme 5.1), in which the non-bridging phosphorane oxygen at position O_{1P} is substituted by a sulfur atom. We show the results obtained for the first 3807 ps of the production run, since for the last 2193 ps there is a drastic conformational change that we consider unrealistic. The RMSD with respect to the initial conformation as a function of the simulation time is shown in Figure B.1. The RMSD reveals that for the first 3807 ps, the complex is stable with an average RMSD of 0.86 ± 0.11 Å. Then, the structure is not equilibrated for the last part of the dynamic, evolving to a conformation that implies several structural changes of both the phosphorane structure and of the protein, with the partial disappearance of one of the β-sheets. This structural change seems too drastic to be consider realistic, and moreover, it is only observed for this MD simulation, and not for the rest of the five MD simulations with the other thio-substituted cases. In our opinion, there is not an obvious reason for this to be the case, and therefore we omit from the present analysis this last 2193 ps of the production run, although a complete analysis of these structural changes are provided in the supplementary material (see Section B).

Conformational dynamics

The time evolution of key dihedral angles that define the conformation of the substrate and the His119 residue are shown in Supporting Information (Figure B.2) and the average values are list in Table B.1. The glycosyl dihedral angles, χ_C and χ_A, are in the anti conformation in the native transition state mimic model simulation[243]([native]$^{\text{Trans}}$ structure, hereafter). The fluctuations of these dihedral angles are very small, signifying that both glycosyl dihedral angles anti and high-anti conformation are very stable throughout the simulation. The average values are -159.58° and -66.83° respectively. The cytidine ribose stays in the North $C_{3'}$-endo conformation throughout the simulation, while the adenosine ribose occupies mainly a South $C_{2'}$-endo conformation with infrequent transitions to the North $C_{3'}$-endo pucker. The average values are 12.50° and 155.85° respectively (see Figure B.3 and Table B.1). Besides, α and γ_A phosphodiester dihedral angles undergo a conformational transition at $\sim$ 1400 ps, from gauche$^+$ to trans and from trans to gauche$^+$, respectively, known as crankshaft motion. Whereas, γ_C phosphodiester dihedral angle has some infrequent and short-lived conformational transitions from gauche$^+$ to trans and gauche$^-$ (see Figure B.5). Finally, it is noteworthy that the side chain of residue His119 maintains its active A conformation of the transphosphorylation transition state mimic model of the native reaction[243] ([native]$^{\text{Trans}}$), with a mean χ_1 value of 165.83° and χ_2 value of -107.75° (see Figure B.2).

Hydrogen bond interactions

The hydrogen bond formation between the ligand and the protein residues in the active site is described in Table 5.2, in which the average distances between atoms participating in protein-ligand hydrogen bonds, with the hydrogen-bond occupancies, average time and number of events are listed. The time evolution of selected distances between protein and ligand atoms is plotted in Supporting Information Figure B.6. In the simulation of the [native]$^{\text{Trans}}$ model,[243] the oxyphosphorane group makes strong hydrogen bond interactions with the two catalytic histidines, His12 and His119, Gln11, Phe120 and a water molecule, and somewhat weaker interactions with Lys41 and some water molecules. In [S:O_{1P}]$^{\text{Trans}}$ model, the interaction with Gln11 and Phe120 are much weaker and some new weak interactions appear (see Table 5.2).

The equatorial O_{2P} non-bridge oxygen makes a strong hydrogen bond interaction with imidazolium side chain of residue His12, which average distance is $< R >$ 1.807 Å and the hydrogen bond occupancy Occ is 0.99, in addition to hydrogen bonding with a water molecule (see Figure B.6). Surprisingly, the interaction between residue Phe120 and O_{2P} equatorial oxygen is not that strong in comparison with [native]$^{\text{Trans}}$ model[243] (Occ 0.42 vS

Table 5.2: Statistics of protein-ligand and water-ligand hydrogen-bonds observed in the simulation of $[S{:}O_{1P}]^{Trans}$ model for the first 3807 ps of the production time. The criteria for the existence of hydrogen bonds are 1) a maximum H $\cdots$ A(acceptor) distance of 2.4 Å for oxygens and 3.0 Å for sulfurs, 2) a minimum $\widehat{DHA}$ of 120° and 3) a hydrogen bond lifetime $\tau \geq 5.0$ ps. The occupancy, Occ, is defined as the total lifetime of each hydrogen bond by the production time. The events are the number of times each hydrogen bond is formed. All distances are in Å.

	Distance		Hydrogen Bond		
Atom pair	MD†	MD‡	Occ	$\langle\tau\rangle$(ps)	# Events
LYS 7 Hζ_2 - CYT S_{1P}	5.245(1.255)	5.097(0.770)	0.00	10.00	1
LYS 7 Hζ_3 - CYT S_{1P}	5.263(1.122)		0.00	12.50	2
GLN 11 Hϵ_{21} - CYT S_{1P}	3.587(1.655)	4.077(0.934)	0.47	47.34	38
LYS 41 Hζ_1 - CYT S_{1P}	3.484(0.349)	3.885(0.671)	0.08	15.01	19
LYS 41 Hζ_2 - CYT S_{1P}	3.375(0.475)		0.13	15.94	32
LYS 41 Hζ_3 - CYT S_{1P}	3.561(0.440)		0.13	20.00	24
WATER - CYT S_{1P}			1.44	16.81	325
HSP 12 Hϵ_2 - CYT O_{2P}	1.807(0.028)	2.788(0.019)	0.99	289.62	13
HSP 119 Hδ_1 - CYT O_{2P}	3.098(0.048)	3.789(0.046)	0.00	10.00	1
PHE 120 HN - CYT O_{2P}	2.408(0.163)	3.365(0.145)	0.42	61.12	26
WATER 2082 - CYT O_{2P}		2.820(0.115)	0.91	60.52	57
HSP 12 Hϵ_2 - CYT $O_{2'}$	2.477(0.050)	3.151(0.036)	0.18	12.05	56
LYS 41 Hζ_1 - CYT $O_{2'}$	3.684(1.420)	3.816(0.739)	0.03	23.02	5
LYS 41 Hζ_2 - CYT $O_{2'}$	3.696(1.199)		0.23	145.00	6
LYS 41 Hζ_3 - CYT $O_{2'}$	3.558(0.097)		0.15	50.45	11
WATER 2455 - CYT $O_{2'}$		9.853(143.629)	0.56	75.71	28
HSP 119 Hδ_1 - CYT $O_{3'}$	2.623(0.010)	3.178(0.031)	0.02	11.28	8
WATER - CYT $O_{3'}$			0.27	15.77	65
HSP 119 Hδ_1 - ADE $O_{5'}$	1.843(0.013)	2.836(0.010)	1.00	473.75	8
WATER - ADE $O_{5'}$			0.33	26.26	47

† Arithmetic average hydrogen bond distance in the production simulation, 3807 ps, with square standard deviation in parentheses.

‡ Arithmetic average distance between the corresponding heavy atoms in the production trajectory, with square standard deviation in parentheses.

0.73), with an hydrogen bond average distance of 2.408 Å. Residue Phe120 is slightly displaced from the vicinity of O_{2P} equatorial oxygen at $\sim$ 1400 ps, which coincided with the crankshaft motion of α and γ_A phosphodiester dihedral angles. Also observed at this position was a really weak hydrogen bond interaction with imidazolium side chain of residue His119 with only 1 short event and an average distance of 3.098 Å (see Figure B.6 and Table 5.2). This last hydrogen bond interaction was not observed in $[native]^{Trans}$ model.

On the other hand, the equatorial S_{1P} atom makes strong hydrogen bond interaction throughout the simulation time with different water molecules. However, the hydrogen bond interactions between this atom and the enzyme residues are quite weak (see Table 5.2). The stronger ones correspond to the interaction with residue Gln11 and the amino group of residue Lys41. It only has an occupancy of 0.47 with a distance average of 3.587 Å for residue Gln11 and a group occupancy of 0.34, with an average distance between S_{1P} and Lys41's Nζ of 3.885 Å for residue Lys41 (see Table 5.2 and Figure B.6). For the first $\sim$ 1612 ps, residue Lys41 is hydrogen bonded to S_{1P}, but after, Gln11 gets closer S_{1P} and Lys41 is displaced. Besides, in $[S{:}O_{1P}]^{Trans}$ model, the interactions between imidazolium side chain of residue His12 and the amino group of residue Lys41 with $O_{2'}$ axial oxygen are weaker than in $[native]^{Trans}$ model[243] (*Occ* 0.18 vS 0.36 and *Occ* 0.41 vS 0.80 respectively, see Table 5.2 and Figure B.6). The cationic residue Lys41 is more mobile ($< R > = 3.816$ Å and $Occ = 0.41$) than residue His12 ($< R > = 2.477$ Å and $Occ = 0.18$). We can conclude that this residue Lys41 is hydrogen bonding to $O_{2'}$ for the first $\sim$ 1580 ps during all the time, while in the rest of the simulation it makes sporadic hydrogen bonds. Finally, in contrast to the $[native]^{Trans}$ model, $O_{2'}$ axial oxygen interacts with a water molecule (see below).

The imidazolium side chain of residue His119 forms a strong hydrogen bond interaction throughout the simulation with $O_{5'}$ axial oxygen and a weak interaction with $O_{3'}$ equatorial oxygen, similar to the $[native]^{Trans}$ model[243] (see Table 5.2). The average distances and hydrogen bond occupancies are 1.843 Å, 1.00 and 2.623 Å, 0.02 respectively.

Solvation of the substrate

Determination of ligand-solvent interactions in the active site is of paramount importance to reveal the role of specific water molecules and bulk solvent effects in enzymatic mechanism. We analyze two aspects of ligand-solvent interactions: specific interactions with water molecule through the formation of hydrogen bonds, and statistical distribution of water molecules around key atoms of the substrate through the analysis of the corresponding radial distribution functions.

The $[S{:}O_{1P}]^{Trans}$ simulation exhibits a variety of interactions between the ligand substrate and solvent (see Table 5.2). Each hydrogen-bonding atom of the ligand interacts with several different water molecules in the course of the $[S{:}O_{1P}]^{Trans}$ simulation. However, it is remarkable that O_{2P} equatorial oxygen and $O_{2'}$ axial oxygen only interact with one water molecule, in contrast to the $[native]^{Trans}$ model,[243] in which $O_{2'}$ did not show any specific hydrogen bond interaction with water molecules. The largest hydrogen bond group occupancy (1.44) corresponds to hydrogen bond interactions with S_{1P} equatorial sulfur, while the largest lifetime-average of solvent hydrogen bonds

(75.71 ps) corresponds to hydrogen bond interactions with $O_{2'}$ axial oxygen.

Radial distribution functions (RDF) can be found in Figure 5.1. Broadly speaking, the RDFs describe three kind of situations:

- *i)* a very sharp peak at the first solvation layer followed by a minimum with a negligible or very small value of RDF (i.e. $g_{O_{2P}-O_W}$ and $g_{O_{2'}-O_W}$). This is an indicative of tightly bound and highly ordered water molecules interacting directly with the atom with slow or negligible water exchange. We will refer to this regime as a *Highly Ordered Water Interaction*. This is specially the case for O_{2P}, which shows the largest peak among the RDFs of $[S{:}O_{1P}]^{Trans}$ complex, 4.11 at 2.7 Å to fall to 0.12 at 3.2 Å. To a lesser extent, $O_{2'}$ follows a similar pattern with a peak of 1.36 at 2.8 Å, which falls to 0.07 at 4.0 Å. Comparing these RDFs with the ones found for the native structure (dotted lines in Figure 5.1), we can say that whereas $g_{O_{2P}-O_W}$ in $[S{:}O_{1P}]^{Trans}$ is very similar to the one in $[native]^{Trans}$, $g_{O_{2'}-O_W}$ is qualitatively different, in the sense that the presence of the first peak is lacking in $[native]^{Trans}$.

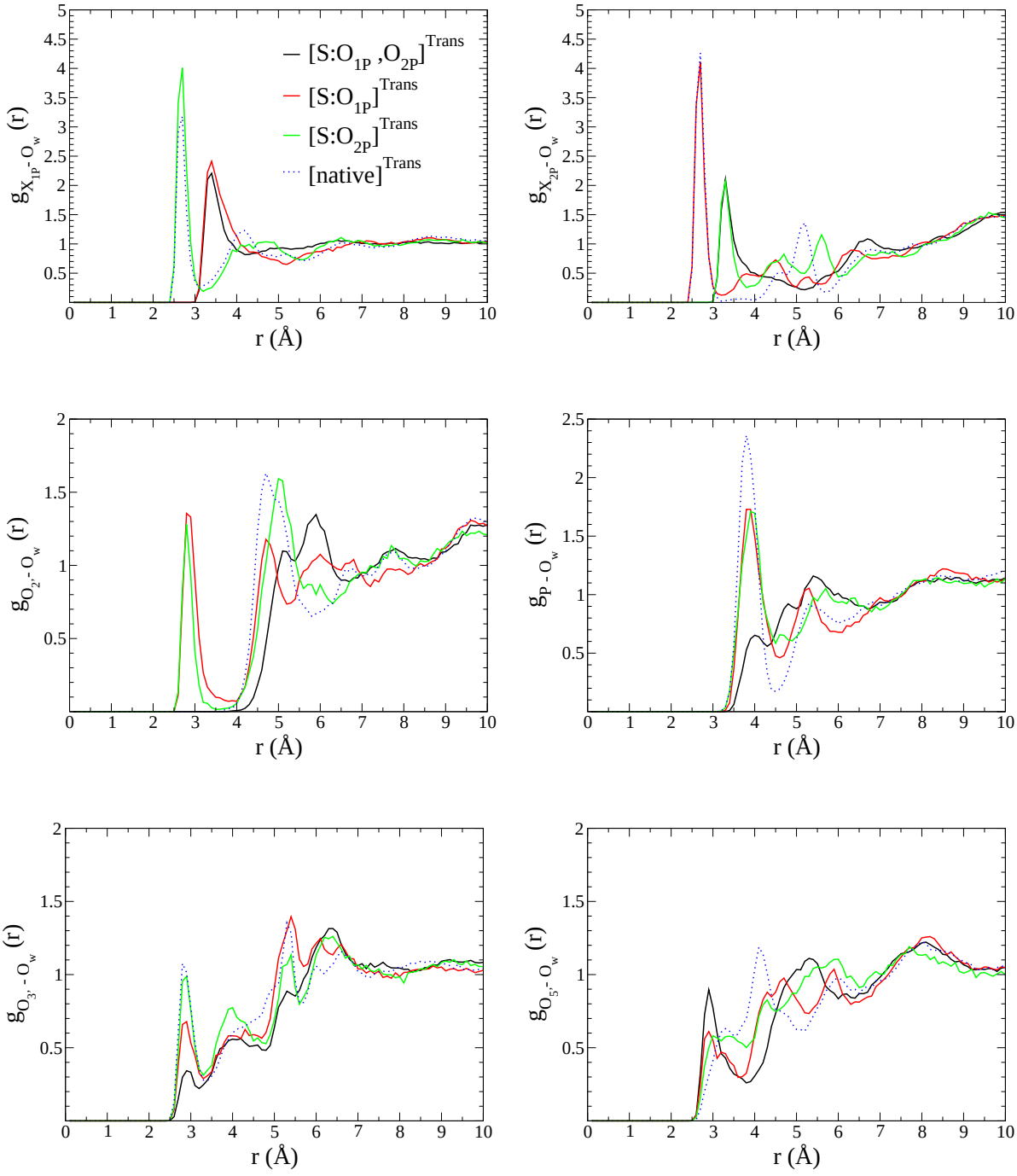
- *ii)* a sharp peak followed by a smooth decrease towards bulk values (i.e. $g_{S_{1P}-O_W}$), which is indicative of a formed first solvation coordination shell that rapidly exchanges waters with bulk solvent. We will denote this regime as a *Fully-exposed Solvent Interaction*. Notice that this thio substitution implies a significant reduction and shift to longer distances of the first peak of $g_{S_{1P}-O_W}$ (2.4 at 3.4 Å) with respect to $[native]^{Trans}$ (3.2 at 2.7 Å), with a reduction in the order of the first solvation shell at this position.

- *iii)* small peaks and fluctuations of RDFs that increases up to bulk values, which are indicative of poorly solvated areas, or small solvent accessibility. We will classified this as *Poorly Solvated Atoms*. In the case of $[S{:}O_{1P}]^{Trans}$ complex, this is the regime we find for $g_{O_{3'}-O_W}$ and $g_{O_{5'}-O_W}$. The RDF for $O_{3'}$ has its first peak at 2.9 Å with a small 0.68 value, and $g_{O_{5'}-O_W}$ starts to increase for distances > 2.6 Å. It is also significant the reduction in the values of these RDFs with respect to the ones found for $[native]^{Trans}$, indicating the lower solvation of the active site upon sulfur substitution. This is also noticeable in the behavior of g_{P-O_W}.

5.3.2 Transphosphorylation transition state mimic $[S{:}O_{2P}]^{Trans}$

In this subsection we show and discuss the results obtained for RNase A - 3',5'-CpA transphosphorylation transition state/intermediate model with thio substitution at 2P position, i.e., the O_{2P} non-bridging equatorial oxygen of the phosphorane is replaced by a sulfur atom in this model, $[S{:}O_{2P}]^{Trans}$ (see Figure 5.1). The RMSD with respect to the initial conformation as a

Figure 5.1: Radial distribution functions of water oxygens around phosphoryl atoms, for the transphosphorylation transition state mimic thio substituted models. Equatorial phosphoryl atoms are at the top, on the left corner 1P and on the right 2P. At the center $O_{2'}$ axial oxygen is on the left and P atom on the right, while at the bottom $O_{3'}$ equatorial oxygen is on the left and $O_{5'}$ axial oxygen on the right.



function of the simulation time is shown in Figure B.1. The RMSD fluctuates stably with an average value of 0.97 ± 0.13 Å.

Conformation dynamics

The time evolution of χ_C and χ_A glycosyl dihedral angles are shown in Figure B.2 and the average values are listed in Table B.2. The fluctuations of both dihedral angles are small, signifying that the χ_C and χ_A anti conformations remain stable throughout the simulation as in the [native]$^{\text{Trans}}$ model. The average values are -155.59° and -67.25°, respectively. Very similar to the ones obtained in the [native]$^{\text{Trans}}$ simulation[243] (χ_C = -158.85° and χ_A = -63.14°). The cytidine ribose stays at North $C_{3'}$-endo conformation throughout the simulation, however, adenosine ribose is mainly at South $C_{2'}$-endo conformation, although it sometimes shows transitions to North $C_{3'}$-endo pucker. The average values are 11.40° and 130.03° respectively (see Figure B.3 and Table B.2). Besides, after the first ∼ 454 ps a crankshaft motion is observed for α and γ_A phosphodiester dihedral angles, they undergo a conformational transition from gauche$^+$ to trans and from trans to gauche$^+$ respectively (see Figure B.5). Whereas, γ_C is mainly at gauche$^+$ conformation with some infrequent and short-lived transitions to trans conformation. The rest of torsional dihedral angles stay stable at their conformations. Finally, imidazolium side chain of residue His119 retains an active A-type conformation as in the [native]$^{\text{Trans}}$ model, with a mean χ_1 value of 164.9° and χ_2 value of -101.65° (see Figure B.2).

Hydrogen bond interactions

Hydrogen bond formation between the ligand and the protein residues in the active site is described in Table 5.3, in which the average distances between atoms participating in protein-ligand hydrogen bonds, with the hydrogen-bond occupancies, average time and number of events are listed. The time evolution of selected distances between protein and ligand atoms is plotted in Supporting Information Figure B.7. As commented before, in [native]$^{\text{Trans}}$ the oxyphosphorane group makes strong hydrogen bond interactions with the two catalytic histidines, His12 and His119, Gln11, Phe120 and a water molecule, and somewhat weaker interactions with Lys41 and some water molecules.[243] However, in [S:O_{2P}]$^{\text{Trans}}$, the interaction with residue Gln11 is much weaker, *Occ* 0.19 vS 0.94 (see Table 5.3).

The equatorial non-bridge O_{1P} oxygen makes strong hydrogen bond interactions with some water molecules, and somewhat weaker interactions with the amino group of Lys41 ($< R >$ = 2.880 Å Occ = 0.64) and residue Gln11 ($< R >$ = 2.863 Å and Occ = 0.19), see Table 5.3 and Figure B.7). The other non-bridging equatorial atom, S_{2P} sulfur, makes strong hydrogen bond interactions with the main chain of residue Phe120, imidazolium side

chain of residue His12 and some water molecules, and weak interactions with the side chain of residue Gln11 and imidazolium side chain of residue His119 (see Table 5.3 and Figure B.7). The average distances and hydrogen bond occupancies are 2.529 Å and 0.99 for Phe120, 2.430 Å and 0.97 for His12, 3.044 Å and 0.24 for His119 and 3.986 Å and 0.01 for Gln11. This last interaction is not seen in [native]$^{\text{Trans}}$ or [S:O$_{1P}$]$^{\text{Trans}}$. In this model, [S:O$_{2P}$]$^{\text{Trans}}$, His12 interacts nearly throughout the simulation directly with S$_{2P}$, the atom at 2P position, as in [native]$^{\text{Trans}}$ and [S:O$_{1P}$]$^{\text{Trans}}$. Imidazolium side chain of residue His12 also interacts with O$_{2'}$ axial oxygen ($< R >$ = 2.667 Å Occ = 0.20). In contrast to [native]$^{\text{Trans}}$ model,[243] but similar to [S:O$_{1P}$]$^{\text{Trans}}$ thio substitution (see above), O$_{2'}$ has a hydrogen bond interaction with only one water molecule from $\sim$ 2857 ps to $\sim$ 5356 ps (see Figure B.7).

The imidazolium side chain of residue His119 forms a strong hydrogen bond interaction throughout the simulation with O$_{5'}$ axial oxygen and a very weak interaction with O$_{3'}$ equatorial oxygen, similar to [native]$^{\text{Trans}}$ and [S:O$_{1P}$]$^{\text{Trans}}$ complexes (see Tables 5.2, 5.3 and Figures B.6, B.7). The hydrogen bond occupancies and average distances are 0.97 and 1.979 Å for O$_{5'}$ axial oxygen, and 0.01 and 2.818 Å for O$_{3'}$ equatorial oxygen. On the other hand, the amino group of the cationic Lys41 forms quite strong hydrogen bond interactions with O$_{1P}$ equatorial oxygen ($< R >$ = 2.880 Å and Occ = 0.64) and weaker interactions with O$_{2'}$ axial oxygen ($< R >$ = 3.251 Å and Occ = 0.36).

Solvation of the substrate

Next, we describe in detail ligand solvent interactions through the analysis of specific hydrogen bond interactions (see Table 5.3), and the analysis of RDFs of solvent molecules around selected ligand atoms (see Figure 5.1). Each hydrogen-bonding atom of the ligand form hydrogen bonds with several water molecules throughout the production simulation. Nevertheless, it is remarkable that O$_{2'}$ axial oxygen only interacts with one water molecule. Furthermore, that interaction only happens between $\sim$ 2587 ps and $\sim$ 5356 ps. The largest hydrogen bond group occupancy (1.35) corresponds to the interactions with O$_{1P}$ equatorial oxygen, similar to [native]$^{\text{Trans}}$ transition state mimic model,[243] while the largest hydrogen bond lifetime-average (77.93 ps) corresponds to O$_{2'}$ axial oxygen, as in [S:O$_{1P}$]$^{\text{Trans}}$. S$_{2P}$ equatorial atom makes hydrogen bond interactions with the solvent nearly during all the production simulation, while O$_{3'}$ equatorial oxygen and O$_{5'}$ axial oxygen have hydrogen bond interactions with group occupancies of 0.44 and 0.24 respectively.

Comparison of the RDFs for [S:O$_{2P}$]$^{\text{Trans}}$ and [native]$^{\text{Trans}}$ reveal important changes upon thio substitution (Figure 5.1). There is a very significant reduction of the first peak of $g_{S_{2P}-O_W}$ in [S:O$_{2P}$]$^{\text{Trans}}$ with respect to the $g_{O_{2P}-O_W}$ in [native]$^{\text{Trans}}$, a value of 2.07 at 3.3 Å vS a peak of 4.31 at 2.7

Table 5.3: Statistics of protein-ligand and water-ligand hydrogen-bonds observed in the simulation of the $[S{:}O_{2P}]^{Trans}$ model. The criteria for the existence of hydrogen bonds are 1) a maximum H $\cdots$ A(acceptor) distance of 2.4 Å for oxygens and 3.0 Å for sulfur, 2) a minimum $\widehat{DHA}$ of 120° and 3) a hydrogen bond lifetime $\tau \geq 5.0$ ps. The occupancy, *Occ*, is defined as the total lifetime of each hydrogen bond by the production time. The events are the number of times each hydrogen bond is formed. All distances are in Å.

	Distance		Hydrogen Bond		
Atom pair	MD†	MD‡	*Occ*	$\langle\tau\rangle$(ps)	# Events
GLN 11 Hϵ_{21} - CYT O_{1P}	2.863(0.624)	3.690(0.448)	0.19	19.92	56
LYS 41 Hζ_1 - CYT O_{1P}	2.591(0.529)	2.880(0.064)	0.36	47.02	45
LYS 41 Hζ_2 - CYT O_{1P}	2.848(0.380)		0.15	34.22	26
LYS 41 Hζ_3 - CYT O_{1P}	2.938(0.344)		0.13	23.75	32
WATER - CYT O_{1P}			1.35	39.09	204
GLN 11 Hϵ_{21} - CYT S_{2P}	3.986(0.645)	4.280(0.484)	0.01	10.00	3
HSP 12 Hϵ_2 - CYT S_{2P}	2.430(0.029)	3.318(0.012)	0.97	144.12	40
HSP 119 Hδ_1 - CYT S_{2P}	3.044(0.071)	3.710(0.056)	0.24	13.02	109
PHE 120 HN - CYT S_{2P}	2.529(0.025)	3.482(0.018)	0.99	189.03	31
WATER - CYT S_{2P}			0.85	49.50	101
HSP 12 Hϵ_2 - CYT $O_{2'}$	2.667(0.212)	3.477(0.160)	0.20	17.41	68
LYS 41 Hζ_1 - CYT $O_{2'}$	3.153(0.571)	3.251(0.113)	0.13	17.15	46
LYS 41 Hζ_2 - CYT $O_{2'}$	3.226(0.480)		0.11	23.53	27
LYS 41 Hζ_3 - CYT $O_{2'}$	3.523(0.640)		0.12	20.71	35
WATER 2143 - CYT $O_{2'}$		16.371(230.032)	0.38	77.93	29
HSP 119 Hδ_1 - CYT $O_{3'}$	2.818(0.070)	3.275(0.042)	0.01	10.00	4
WATER - CYT $O_{3'}$			0.44	17.16	150
HSP 119 Hδ_1 - ADE $O_{5'}$	1.979(0.030)	2.917(0.019)	0.97	109.52	52
WATER - ADE $O_{5'}$			0.24	15.43	92

† Arithmetic average hydrogen bond distance in the production simulation, 5985 ps, with square standard deviation in parentheses.
‡ Arithmetic average distance between the corresponding heavy atoms in the production trajectory, with square standard deviation in parentheses.

Å in $[native]^{Trans}$. In the case of O_{1P}, there is no big changes in $g_{O_{1P}-O_W}$ with respect to the native oxyphosphorane, although there is a slight increase in the height of the first peak, 4.01 at 2.7 Å which falls to 0.19 at 3.2 Å, indicative of a quite ordered water interaction with O_{1P}. On the other hand, $O_{2'}$ axial oxygen shows a similar pattern as in $[S{:}O_{1P}]^{Trans}$, with a peak of 1.28 at 2.8 Å that drops to a value of 0.02 at 3.5 Å, suggesting an interaction with a highly ordered water molecule, an interaction not present in $[native]^{Trans}$. Finally, $O_{3'}$ and $O_{5'}$ resembles a *poorly-solvated* situation as in the case of $[S{:}O_{1P}]^{Trans}$, although there is a substantial increase in

the height of the first peak of the $g_{O_{3'}-O_w}$ when passing from [S:O_{1P}]^{Trans} to [S:O_{2P}]^{Trans} , reaching to a level similar to the one found in [native]^{Trans} . Inspection of the g_{P-O_W}'s suggests a lower solvation of the active site upon monothio substitution.

5.3.3 Transphosphorylation transition state mimic model with dithio substitution [S:O_{1P},O_{2P}]^{Trans}

In this subsection, we present the results for RNase A - 3',5'-CpA transphosphorylation transition state/intermediate mimic model with dithio substitutions, [S:O_{1P},O_{2P}]^{Trans} model, in which both non-bridging oxyphosphorane equatorial oxygens, O_{1P} and O_{2P}, are substituted by sulfur atoms, S_{1P} and S_{2P} (see Figure 5.1). The RMSD with respect to the initial conformation as a function of the simulation time is shown in Figure B.1. The RMSD fluctuates stably with an average value of 0.97 ± 0.14 Å.

Conformational dynamics

The time evolution of key dihedral angles that define the conformation of the substrate and residue His119 are shown in Figure B.2 and the average values are list in Table B.3. The average values of glycosyl dihedral angles χ_C and χ_A are -152.76° and -69.56°, respectively, with small fluctuations, denoting that anti and high-anti conformations are very stable, and in concordance with the values obtained in the simulations for the [native]^{Trans} model,[243] and [S:O_{1P}]^{Trans} and [S:O_{2P}]^{Trans} thio-substituted complexes. The cytidine ribose stays at North $C_{3'}$-endo conformation throughout the simulation, however, adenosine ribose is occupies mainly a South $C_{2'}$-endo conformation with infrequent transitions to a North $C_{3'}$-endo pucker. The average values are 9.40° and 152.77° respectively (see Figure B.3 and Table B.3). Besides, after the first $\sim$ 4142 ps a crankshaft motion is observed for α and γ_A phosphodiester dihedral angles, they undergo a conformational transition from gauche$^+$ to trans and from trans to gauche$^+$ respectively (see Figure B.5). Whereas, γ_C occupies mainly a gauche$^+$ conformation with infrequent and short-lived transitions to a trans conformation. The rest of torsional dihedral angles do not undergo any conformational change along the production run. Finally, the imidazolium side chain of residue His119 retains an active A-type conformation, with a mean χ_1 value of 166.19° and χ_2 value of -105.13° (see Figure B.2), similar values to the ones obtained in the previous simulations of mono-thio substituted models and in [native]^{Trans} .

Hydrogen bond interactions

Hydrogen bond formation between the ligand and the protein residues in the active site is described in Table 5.4, in which the average distances between atoms participating in the protein-ligand hydrogen bonds, with the

hydrogen-bond occupancies, average time and number of events are listed. The time evolution of selected distances between protein and ligand atoms is plotted in Supporting Information Figure B.8 . Similar to the simulation of the [native]$^{\text{Trans}}$ model[243] in [S:O_{1P},O_{2P}]$^{\text{Trans}}$, the phosphorane group makes strong hydrogen bond interactions with the two catalytic histidines, His12 and His119, Gln11 and Phe120, but somewhat weaker interactions with Lys41 and water molecules (see Table 5.4).

Table 5.4: Statistics of protein-ligand and water-ligand hydrogen-bonds observed in the simulation of the [S:O_{1P},O_{2P}]$^{\text{Trans}}$ model. The criteria for the existence of hydrogen bonds are 1) a maximum H $\cdots$ A(acceptor) distance of 2.4 Å for oxygens and 3.0 Å for sulfur, 2) a minimum $\widehat{DHA}$ of 120° and 3) a hydrogen bond lifetime $\tau \geq 5.0$ ps. The occupancy, *Occ*, is defined as the total lifetime of each hydrogen bond by the production time. The events are the number of times each hydrogen bond is formed. All distances are in Å.

	Distance		Hydrogen Bond		
Atom pair	MD†	MD‡	*Occ*	$\langle\tau\rangle$(ps)	# Events
GLN 11 Hϵ_{21} - CYT S_{1P}	2.708(0.175)	3.602(0.085)	0.87	61.20	86
LYS 41 Hζ_1 - CYT S_{1P}	3.386(0.398)	3.484(0.032)	0.10	13.16	46
LYS 41 Hζ_2 - CYT S_{1P}	3.484(0.469)		0.12	14.42	52
LYS 41 Hζ_3 - CYT S_{1P}	3.439(0.367)		0.06	11.94	31
WATER - CYT S_{1P}			1.30	14.17	557
HSP 12 Hϵ_2 - CYT S_{2P}	2.422(0.024)	3.326(0.011)	0.99	238.78	25
HSP 119 Hδ_1 - CYT S_{2P}	3.134(0.075)	3.805(0.058)	0.17	13.40	75
PHE 120 HN - CYT S_{2P}	2.553(0.031)	3.499(0.020)	0.98	197.99	30
WATER - CYT S_{2P}			0.88	47.67	112
HSP 12 Hϵ_2 - CYT $O_{2'}$	2.772(0.156)	3.565(0.117)	0.11	12.11	57
LYS 41 Hζ_1 - CYT $O_{2'}$	3.077(0.639)	3.110(0.064)	0.28	30.10	56
LYS 41 Hζ_2 - CYT $O_{2'}$	3.027(0.585)		0.25	24.02	63
LYS 41 Hζ_3 - CYT $O_{2'}$	3.232(0.460)		0.14	17.97	47
HSP 119 Hδ_1 - CYT $O_{3'}$	2.822(0.065)	3.294(0.043)	0.00	10.00	1
WATER - CYT $O_{3'}$			0.14	12.84	65
HSP 119 Hδ_1 - ADE $O_{5'}$	1.982(0.025)	2.935(0.018)	0.99	299.24	20
WATER - ADE $O_{5'}$			0.42	16.73	153

† Arithmetic average hydrogen bond distance in the production simulation, 6000 ps, with square standard deviation in parentheses.
‡ Arithmetic average distance between the corresponding heavy atoms in the production trajectory, with square standard deviation in parentheses.

The S_{1P} equatorial sulfur atom makes strong hydrogen bond interactions with some water molecules and residue Gln11 ($< R > = 2.708$ Å and $Occ = 0.87$), and somewhat weaker hydrogen bond interaction with the amino group

of Lys41 ($< R >$ = 3.484 Å and Occ = 0.28), see Table 5.4 and Figure B.8. The other sulfur atom, S_{2P}, has a strong hydrogen bond interaction with imidazolium side chain of residue His12 ($< R >$ = 2.422 Å and Occ = 0.99), with the main chain of Phe120 ($< R >$ = 2.553 Å and Occ = 0.98) and with some water molecules, while the interaction with imidazolium side chain of residue His119 is much weaker ($< R >$ = 3.134 Å and Occ = 0.17), like in $[S{:}O_{2P}]^{Trans}$ model (see Table 5.4 and Figure B.8). In comparison to $[S{:}O_{1P}]^{Trans}$ model, $[S{:}O_{1P},O_{2P}]^{Trans}$ model's hydrogen bond interaction with His119 is stronger, while $[native]^{Trans}$ lacks this interaction, in addition, some water molecules interact with $O_{2'}$ instead of one. Also observed at this position was a really weak hydrogen bond interaction with imidazolium side chain of residue His12, hydrogen bond occupancy of 0.11 and an average distance of 2.772 Å.

The imidazolium side chain of residue His119 makes a strong hydrogen bond interaction with $O_{5'}$ axial oxygen throughout the simulation. The average distance is 1.982 Å with a hydrogen bond occupancy of 0.99 and only 20 events (see Table 5.4 and Figure B.8). On the other hand, the amino group of Lys41 forms quite strong hydrogen bond interaction with $O_{2'}$ axial oxygen ($< R >$ = 3.110 Å and Occ = 0.67), while the interaction with S_{1P} equatorial sulfur is less strong ($< R >$ = 3.484 Å and Occ = 0.28), similar to $[native]^{Trans}$ model.

Solvation of the substrate

As in the previous subsections, we describe in detail ligand solvent interactions through the analysis of specific hydrogen bond interactions (see Table 5.4), and the use of RDFs of solvent molecules around selected ligand atoms. Each hydrogen-bonding atom of the ligand interacts with several different water molecules in the course of $[S{:}O_{1P},O_{2P}]^{Trans}$ simulation. However, it is noteworthy that the $O_{2'}$ axial oxygen does not interact with any water molecule. Quite interestingly, this situation for $O_{2'}$ is similar to the one found in the $[native]^{Trans}$ simulation,[243] and it differs to the previously described one for monothio substitutions. The largest hydrogen bond group occupancy corresponds to the interactions with S_{1P} (1.30), while the largest hydrogen bond lifetime-average corresponds to S_{2P} (47.67 ps). $O_{3'}$ equatorial and $O_{5'}$ axial oxygens have much weaker hydrogen bond interactions, with group occupancies of 0.14 and 0.42 respectively. Quite interestingly, the occupancy for $O_{3'}$ is lower than for $O_{5'}$, contrary to the behavior of $[native]^{Trans}$ and $[S{:}O_{2P}]^{Trans}$.

Inspection of RDFs (Figure 5.1) reveals several interesting aspects on the solvation of the dithio phosphorane complex. Non-bridging equatorial atoms S_{1P} and S_{2P}, have the largest probability to find water molecules in their surroundings. S_{1P} has a maximum value of 2.21 at 3.4 Å while S_{2P}'s maximum is 2.10 at 3.3 Å. Nevertheless, these values are the lowest among

the whole transphosphorylation transition states investigated, and indicative of the profound effect that thio substitution has on the interaction with solvent molecules. In addition, notice that these RDFs are significantly less structured than the ones for $[\text{native}]^{\text{Trans}}$, and suggest a loss in the specificity of the interaction with water molecules, or in other words, these water molecules are less ordered. There is also a significant effect on $g_{O_{2'}-O_W}$ upon double thio substitution. Notice that the first peak that appears in $[\text{S:O}_{1\text{P}}]^{\text{Trans}}$ and $[\text{S:O}_{2\text{P}}]^{\text{Trans}}$ now disappears, and we find a situation similar to the one found in $[\text{native}]^{\text{Trans}}$. This indicates the lack of a direct solvent interaction with $O_{2'}$ in both native oxyphosphorane and double thio substituted phosphorane, the direct interaction is only present for monosubstituted transphosphorylation structures. It is also quite remarkable that for the less solvated $O_{3'}$ equatorial atoms and $O_{5'}$ axial atoms, double thio substitution has a contrary effect. Whereas for $O_{3'}$ double thio substitution almost leads to the disappearance of the first peak, for $O_{5'}$ there is an increase of this first peak.

5.3.4 Hydrolysis transition state mimic model $[\text{S:O}_{1\text{P}}]^{\text{Hyd}}$

The next three subsections are dedicated to discuss the results for RNase A - 3',5'-CpA hydrolysis transition state/intermediate mimic thio-substituted models. First, we show the results for the $[\text{S:O}_{1\text{P}}]^{\text{Hyd}}$ complex, in which the oxygen at O1P position is substituted by a sulfur atom (see Figure 5.1). The RMSD with respect to the initial conformation as a function of time is shown in Figure B.1. The RMSD fluctuates stably with an average value of 0.89 ± 0.16 Å, very similar to the $[\text{S:O}_{1\text{P}}]^{\text{Trans}}$ simulation.

Conformational dynamics

As in previous subsections, we first focus on the values of key dihedral angles that determine the conformation of the substrate and residue His119 (Figure B.2 and Table B.4). The fluctuation of χ_C angle is very small, signifying that its anti conformation remains stable throughout the simulation as in the hydrolysis transition state mimic model of the native reaction[243] ($[\text{native}]^{\text{Hyd}}$ complex hereafter). The average value of χ_C is -159.71°. Besides, the cytidine ribose occupies a North $C_{3'}$-endo pucker throughout the simulation (see Figure B.3). Torsional dihedral angles do not undergo any conformational transition (see Figure B.9), δ_C and ϵ stays at gauche$^+$ conformation, ζ is at trans conformation, while γ_C has some infrequent short-lived transitions from gauche$^+$ to trans, known as crankshaft motion. Finally, the imidazolium side chain of residue His119 retains an active A-type conformation as in the $[\text{native}]^{\text{Hyd}}$ model, with a mean χ_1 value of 159.03 and χ_2 value of -107.44 (see Figure B.2).

Hydrogen bond interactions

The average distances between atoms participating in protein-ligand hydrogen bonds, with the hydrogen bond occupancies, average time and number of items are listed in Table 5.5, while, the time evolution of selected distances between protein and ligand atoms is plotted in Figure B.10. The $[\text{native}]^{\text{Hyd}}$ complex makes strong hydrogen bond interactions with both catalytic histidines, His12 and His119, residues Gln11, Lys41, Phe120 and a water molecule, and weaker interactions with some other water molecules.[243] However, in $[\text{S:O}_{1\text{P}}]^{\text{Hyd}}$ model the interaction with Gln11 is much weaker, 0.39 vS 0.86 (see Table 5.5).

The equatorial non-bridging atom at 1P (S_{1P}) makes hydrogen bond interactions with 3 different residues, the amino group of Lys7 and Lys41 and with Gln11. In comparison to $[\text{native}]^{\text{Hyd}}$, in $[\text{S:O}_{1\text{P}}]^{\text{Hyd}}$, there is a new interaction with residue Lys7 (Occ = 0.06), however, the other hydrogen bond interactions, with Lys41 (Occ = 0.36) and Gln11 (Occ = 0.39), decrease. Equatorial O_{2P} non-bridging oxygen makes strong hydrogen bond interactions with imidazolium side chain of residue His12 throughout the simulation and with the main chain of residue Phe120, and somewhat weaker interactions with some water molecules (see Table 5.5 and Figure B.10). The hydrogen bond occupancies and the average distances are 1.00, 1.750 Å and 0.90, 2.086 Å respectively. The imidazolium side chain of residue His12 forms a very weak hydrogen bond interaction with $O_{2'}$ axial oxygen ($< R >$ = 2.560 Å and Occ = 0.06). Nevertheless, also observed in this position was a strong hydrogen bond interaction with ammonium ion of residue Lys41 ($< R >$ = 3.066 Å and Occ = 0.87), similar to the $[\text{native}]^{\text{Hyd}}$ simulation.

The imidazolium side chain of residue His119 forms a strong hydrogen bond with O_{3T} axial oxygen ($< R >$ = 1.922 Å and Occ = 0.96), in addition to hydrogen bonding with $O_{3'}$ equatorial oxygen ($< R >$ = 2.318 Å and Occ = 0.51), see Table 5.5 and Figure B.10.

Solvation of the substrate

The $[\text{S:O}_{1\text{P}}]^{\text{Hyd}}$ simulation exhibits a variety of interactions between the ligand substrate and solvent (Table 5.5 and Figure 5.2). Each hydrogen-bonding atom of the ligand interacts with several different water molecules throughout the $[\text{S:O}_{1\text{P}}]^{\text{Hyd}}$ simulation, except $O_{2'}$ axial oxygen which only interacts with two different water molecules. The largest hydrogen bond group occupancy (1.17) corresponds to S_{1P} equatorial atom, while the largest hydrogen bond lifetime-average (42.70 ps) corresponds to the other non-bridging equatorial atom, O_{2P}, similar to $[\text{native}]^{\text{Hyd}}$ model. Axial oxygen $O_{2'}$ has only a group hydrogen bond occupancy of 0.08, while $O_{3'}$ occupancy is 0.37 and O_{3T} is 0.29.

The analysis of RDFs reveal interesting effects on solvation upon thio sub-

Table 5.5: Statistics of protein-ligand and water-ligand hydrogen-bonds observed in the simulation of the [S:O_{1P}]^{Hyd} model. The criteria for the existence of hydrogen bonds are 1) a maximum H $\cdots$ A(acceptor) distance of 2.4 Å for oxygens and 3.0 Å for sulfur, 2) a minimum $\widehat{DHA}$ of 120° and 3) a hydrogen bond lifetime $\tau \geq 5.0$ ps. The occupancy, *Occ*, is defined as the total lifetime of each hydrogen bond by the production time. The events are the number of times each hydrogen bond is formed. All distances are in Å.

	Distance		Hydrogen Bond		
Atom pair	MD†	MD‡	*Occ*	$\langle\tau\rangle$(ps)	# Events
LYS 7 Hζ_1 - CYT S_{1P}	5.237(2.001)	5.102(1.561)	0.01	12.50	6
LYS 7 Hζ_2 - CYT S_{1P}	5.266(2.064)		0.02	14.26	7
LYS 7 Hζ_3 - CYT S_{1P}	5.251(2.114)		0.03	13.46	13
GLN 11 Hϵ_{21} - CYT S_{1P}	3.479(1.003)	3.962(0.459)	0.39	30.00	78
LYS 41 Hζ_1 - CYT S_{1P}	3.464(0.558)	3.609(0.150)	0.16	15.22	63
LYS 41 Hζ_2 - CYT S_{1P}	3.644(0.577)		0.10	16.42	35
LYS 41 Hζ_3 - CYT S_{1P}	3.542(0.490)		0.10	13.98	44
WATER - CYT S_{1P}			1.17	14.62	481
HSP 12 Hϵ_2 - CYT O_{2P}	1.750(0.018)	2.739(0.012)	1.00	665.56	9
PHE 120 HN - CYT O_{2P}	2.086(0.051)	3.052(0.041)	0.90	65.61	82
WATER - CYT O_{2P}			0.73	42.70	102
HSP 12 Hϵ_2 - CYT $O_{2'}$	2.560(0.042)	3.211(0.031)	0.06	11.24	32
LYS 41 Hζ_1 - CYT $O_{2'}$	2.873(0.720)	3.066(0.146)	0.40	88.69	27
LYS 41 Hζ_2 - CYT $O_{2'}$	3.062(0.540)		0.24	67.63	21
LYS 41 Hζ_3 - CYT $O_{2'}$	3.111(0.518)		0.23	86.22	16
WATER - CYT $O_{2'}$			0.08	35.40	13
HSP 119 Hδ_1 - CYT $O_{3'}$	2.318(0.061)	3.058(0.031)	0.51	20.50	150
WATER - CYT $O_{3'}$			0.37	13.87	160
HSP 119 Hδ_1 - CYT O3T	1.922(0.049)	2.877(0.032)	0.96	125.21	46
LYS 7 Hζ_3 - CYT O3T	5.822(2.617)	5.748(2.062)	0.00	10.00	1
WATER - CYT O3T			0.29	12.43	140

† Arithmetic average hydrogen bond distance in the production simulation, 6000 ps, with square standard deviation in parentheses.
‡ Arithmetic average distance between the corresponding heavy atoms in the production trajectory, with square standard deviation in parentheses.

stitution at 1P. Comparing the RDFs with those for the [native]^{Hyd} suggest a general decrease on solvation upon sulfur substitution, specially evident from the analysis of g_{P-O_W}. The first peak of $g_{S_{1P}-O_W}$ is reduced in height to 2.26 and shifted to longer distances 3.4 Å. Moreover, the shape of this RDFs also suggest a loss of order in first solvation shell around this atom with respect to the native structure. S_{1P} thio substitution also reduces the first peak of

$g_{O_{2P}-O_W}$ to 2.67 at 2.7 Å, but the overall shape of the RDFs is maintained and it corresponds to a highly-ordered water interaction as in [native]$^{\text{Hyd}}$. $O_{2'}$ axial oxygen has the lowest probability to find water molecules around it between all the phosphorane oxygen atoms, at distances < 3 Å. Here sulfur substitution at O_{1P} has a dramatic effect on the disappearance of the first peak.

Finally RDFs of $O_{3'}$ and O_{3T} atoms are qualitatively similar to those found for the native structure, although specially for O_{3T} again there is a significant reduction of the first peak. These atoms as in transphosphorylation structures has a poorer solvation than the non-bridging equatorial atoms, but this solvation is higher than that found for the analogous transphosphorylation structures.

5.3.5 Hydrolysis transition state mimic model [S:O_{2P}]$^{\text{Hyd}}$

The [S:O_{2P}]$^{\text{Hyd}}$ complex is formed by sulfur substitution of the O_{2P} oxygen in the hydrolysis-type oxyphosphorane (see Figure 5.1). The RMSD with respect to the initial conformation as a function of the simulation time is shown in Figure B.1. The RMSD fluctuates stably with an average value of 1.03 ± 0.21 Å.

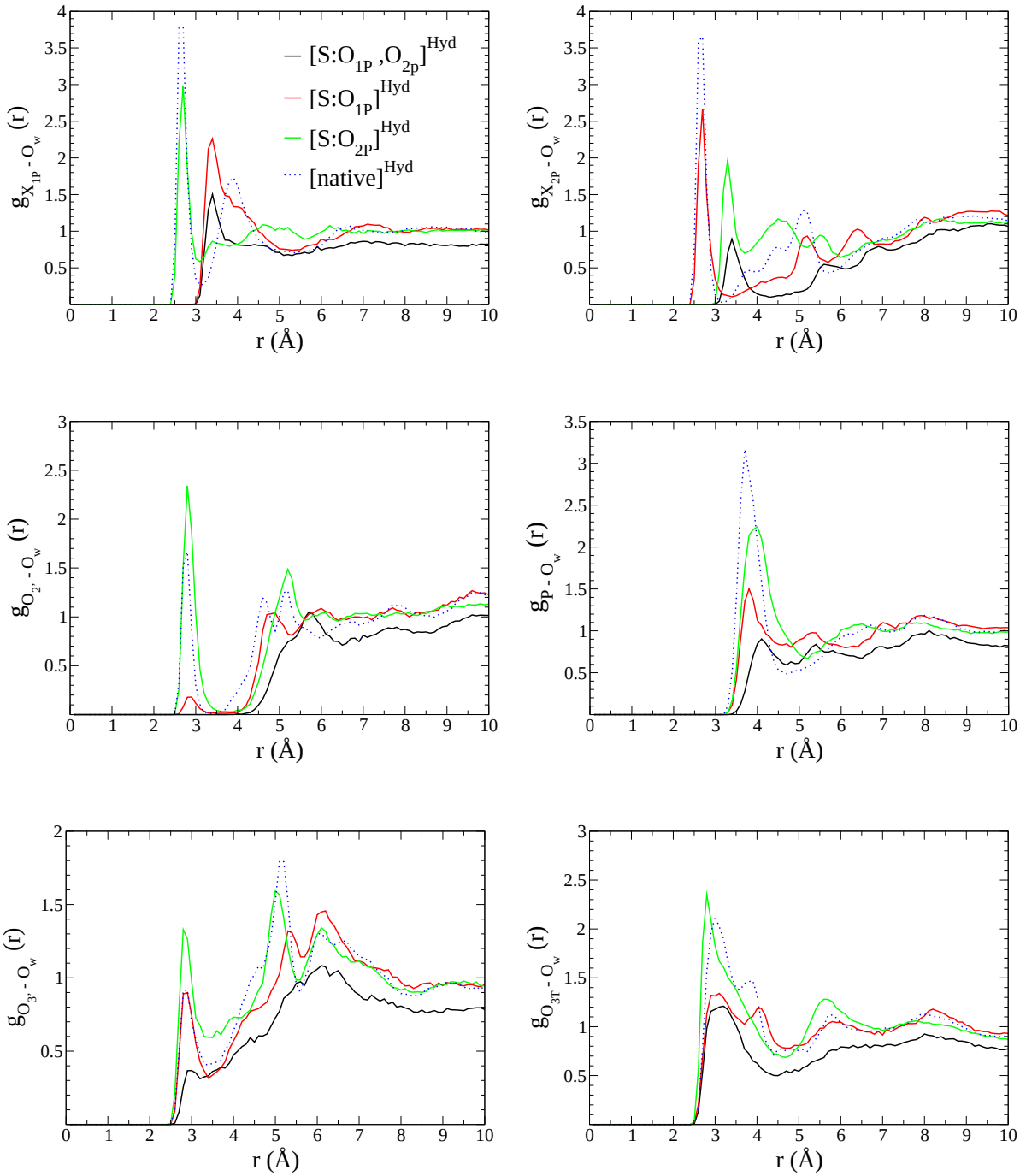
Conformational dynamics

The time evolution of χ_C glycosyl dihedral angle is shown in Supporting Information Figure B.2 and the average value is listed in Table B.4. The fluctuation of χ_C dihedral angle is small, similar to the rest of the simulations, i.e., anti-conformation is again very stable and is not affected by this thio substitution. The average value is -153.44°. Besides, the cytidine ribose essentially occupies a North $C_{3'}$-endo pucker throughout the simulation, although it sometimes suffer some infrequent short-lived transitions to East $O_{4'}$-endo conformation (see Figure B.3). γ_C presents some crankshaft motion from gauche$^+$ to trans, while the rest of torsional dihedral angles do not undergo any conformational transition (see Figure B.9). Finally, imidazolium side chain of residue His119 has a transition from its active A-type conformation to the inactive B-type conformation after the first $\sim$ 2154 ps. The mean values of χ_1 and χ_2 in the A conformation are 170.70° and -102.33° respectively, and -64.77° and -64.84° in B conformation (see Figure B.2). This is the first time we see this kind of transition in our simulations.

Hydrogen bond interactions

Table 5.6 summarizes the hydrogen bond formation between the ligand and the protein residues in the active site in the [S:O_{2P}]$^{\text{Hyd}}$ complex. Time evolution of selected distances between protein and ligand atoms is plotted in Figure B.11. The oxyphosphorane in [native]$^{\text{Hyd}}$ makes strong hydrogen

Figure 5.2: Radial distribution functions of water oxygens around phosphoryl atoms, for the hydrolysis transition state mimic thio substituted models. Equatorial phosphoryl atoms are at the top, on the left corner 1P and on the right 2P. At the center $O_{2'}$ axial oxygen is on the left and P atom on the right, while at the bottom $O_{3'}$ equatorial oxygen is on the left and O_{3T} axial oxygen on the right.



bond interactions with both catalytic histidines, His12 and His119, residues Gln11, Lys41, Phe120 and a water molecule, and weaker interactions with some other water molecules.[243] However, the $[S{:}O_{2P}]^{Hyd}$ complex maintains strong hydrogen bond interactions with residues His12, Lys41, Phe120 and a water molecule, but there is a weaker interaction with Gln11 and His119 (see Table 5.6).

The equatorial non-bridging O_{1P} atom makes a strong hydrogen bond interaction with the amino group of Lys41, group hydrogen bond occupancy of 0.89, and a weaker interaction with residue Gln11, $Occ = 0.46$. Quite interestingly, the reverse trend was observed for $[native]^{Hyd}$, with occupancies of 0.57 and 0.86, respectively.[243] The average distance is 2.524 Å for Gln11 and 2.825 Å between O_{1P} and Lys41's Nζ (see Table 5.6 and Figure B.11). The other equatorial non-bridging atom, S_{2P}, makes strong interactions with imidazolium side chain of residue His12 ($< R > = 2.698$ Å and $Occ = 0.88$), the main chain of residue Phe120 ($< R > = 2.745$ Å and $Occ = 0.87$) and some water molecules (see Table 5.6 and Figure B.11). Also observed in this position there is a weak hydrogen bond interaction with imidazolium side chain of residue His119, $Occ = 0.14$ and $< R > = 5.198$ Å, like in $[native]^{Hyd}$. A remarkable point, is that the amino group of Lys41 does not make any important hydrogen bond interaction with $O^{2'}$ axial oxygen (see Table 5.6 and Figure B.11). The group hydrogen bond occupancy is only 0.01 and the average distance between the heavy atoms is 4.071 Å. However, the hydrogen bond interaction between $O_{2'}$ atom and imidazolium side chain of residue His12 is stronger than in the $[native]^{Hyd}$ model simulation[243] (Occ 0.28 vS 0.13).

Surprisingly, imidazolium side chain of residue His119 does not make any strong hydrogen bond interaction with the ligand atoms. We can say that this residue is very mobile in this model, because it makes weak hydrogen bond interactions with S_{2P} and $O_{3'}$ equatorial atoms, in addition to hydrogen bonding with $O_{2'}$ and O_{3T} axial atoms, in contrast to $[native]^{Hyd}$ and $[S{:}O_{1P}]^{Hyd}$, where it only interacts with $O_{3'}$ and O_{3T} strongly. The reason of the weak interactions of $[S{:}O_{2P}]^{Hyd}$ is due to the conformational change that takes place after the first $\sim$ 2154 ps, from active to inactive conformation (see Figure B.2 and B.11 and Table 5.6). Furthermore, because His119 does not stabilize so much the phosphodiester group, Lys41 takes a more central conformation, making a strong hydrogen bond interaction with O_{1P} equatorial oxygen and a null interaction with $O_{2'}$ axial oxygen.

Solvation of the substrate

The $[S{:}O_{2P}]^{Hyd}$ simulation exhibits a variety of interactions between the ligand substrate and solvent (Table 5.6 and Figure 5.2). Each hydrogen-bonding atom of the ligand, interacts with several different water molecules throughout the $[S{:}O_{2P}]^{Hyd}$ simulation. However, it is noteworthy that the $O_{2'}$

Table 5.6: Statistics of protein-ligand and water-ligand hydrogen-bonds observed in the simulation of the $[S{:}O_{2P}]^{Hyd}$ model. The criteria for the existence of hydrogen bonds are 1) a maximum H $\cdots$ A(acceptor) distance of 2.4 Å for oxygens and 3.0 Å for sulfur, 2) a minimum $\widehat{DHA}$ of 120° and 3) a hydrogen bond lifetime $\tau \geq 5.0$ ps. The occupancy, *Occ*, is defined as the total lifetime of each hydrogen bond by the production time. The events are the number of times each hydrogen bond is formed. All distances are in Å.

	Distance		Hydrogen Bond		
Atom pair	MD†	MD‡	*Occ*	$\langle\tau\rangle$(ps)	# Events
GLN 11 Hϵ_{21} - CYT O_{1P}	2.524(0.384)	3.354(0.270)	0.46	26.32	106
HSP 12 Hϵ_2 - CYT O_{1P}	3.935(0.395)	4.829(0.365)	0.04	75.00	3
LYS 41 Hζ_1 - CYT O_{1P}	2.755(0.591)	2.825(0.131)	0.32	60.00	32
LYS 41 Hζ_2- CYT O_{1P}	2.676(0.532)		0.33	73.33	27
LYS 41 Hζ_3 - CYT O_{1P}	2.790(0.501)		0.24	66.35	22
WATER - CYT O_{1P}			1.21	28.40	259
HSP 12 Hϵ_2 - CYT S_{2P}	2.698(0.428)	3.602(0.430)	0.88	243.63	22
HSP 119 Hδ_1 - CYT S_{2P}	5.198(2.419)	4.952(0.783)	0.14	32.12	26
PHE 120 HN - CYT S_{2P}	2.745(0.262)	3.679(0.209)	0.87	105.08	50
WATER - CYT S_{2P}			0.89	43.55	124
HSP 12 Hϵ_2 - CYT $O_{2'}$	2.621(0.340)	3.395(0.318)	0.28	16.61	104
LYS 41 Hζ_1 - CYT $O_{2'}$	4.166(0.827)	4.071(0.337)	0.00	10.00	2
LYS 41 Hζ_2 - CYT $O_{2'}$	3.972(0.685)		0.00	10.00	2
LYS 41 Hζ_3 - CYT $O_{2'}$	4.089(0.768)		0.01	13.33	3
HSP 119 Hδ_1 - CYT $O_{2'}$	7.263(4.920)	6.986(2.780)	0.02	13.90	9
WATER (4491) - CYT $O_{2'}$		2.974(0.180)	0.93	82.57	68
HSP 119 Hδ_1 - CYT $O_{3'}$	6.052(4.968)	5.793(3.175)	0.01	10.00	4
HSP 119 Hϵ_2 - CYT $O_{3'}$	6.232(1.627)	5.865(1.627)	0.00	10.00	1
WATER - CYT $O_{3'}$			0.65	16.38	242
LYS 7 Hζ_3 - CYT O3T	5.381(2.099)	5.386(1.441)	0.00	20.00	1
GLN 11 Hϵ_{21} - CYT O3T	4.111(0.382)	4.930(0.333)	0.02	16.25	8
HSP 119 Hδ_1 - CYT O3T	5.244(3.461)	5.020(1.492)	0.13	24.84	31
WATER - CYT O3T			1.00	14.88	407

† Arithmetic average hydrogen bond distance in the production simulation, 6000 ps, with square standard deviation in parentheses.
‡ Arithmetic average distance between the corresponding heavy atoms in the production trajectory, with square standard deviation in parentheses.

axial oxygen only interacts with a single water molecule , as in $[native]^{Hyd}$, during all the simulation, which has a hydrogen bond occupancy of 0.93, and an average distance of 2.974 Å. The largest hydrogen bond group occupancy corresponds to hydrogen bond interaction with O_{1P} equatorial oxygen (1.21), although O_{3T} occupancy is very high too (1.00). The largest hydrogen bond

lifetime-average (82.57 ps) corresponds to $O_{2'}$ axial oxygen. S_{2P} equatorial atom has a group hydrogen bond occupancy of 0.89.

Regarding RDF analysis (Figure 5.2), O_{1P} equatorial oxygen has the largest probability to find water molecules around it (with a peak of 2.98), but there is a significant decrease of the height of the first peak with respect to the one found in $[native]^{Hyd}$. Notice as well that the $g_{O_{1P}-O_W}$ is less structured than in $[native]^{Hyd}$. As expected, there is a very high decrease in solvation at the 2P position with a peak of 1.97 (almost half value than in $[native]^{Hyd}$) and a shift to longer distances. On the other hand, the RDFs for $O_{2'}$, $O_{3'}$ and O_{3T} are qualitatively similar to the ones found in $[native]^{Hyd}$, with an enhanced solvation at short distances for $O_{2'}$ and $O_{3'}$, peaks of 2.34 and 1.33, respectively. Notice that, as inspection of g_{P-O_W} reveals, thio substitution at 2P position leads to an overall lower solvation of the phosphorane moiety (lower peaks in the RDFs) with respect to $[native]^{Hyd}$, but the effect is significantly less pronounced than in $[S{:}O_{1P}]^{Hyd}$.

5.3.6 Hydrolysis transition state mimic model with dithio substitution $[S{:}O_{1P},O_{2P}]^{Hyd}$

In this section, we present the results for RNase A - 3',5'-CpA hydrolysis transition state/intermediate mimic model with both non-bridging equatorial oxygens substituted by sulfur, denoted as the $[S{:}O_{1P},O_{2P}]^{Hyd}$ complex (see Figure 5.1). The RMSD with respect to the initial conformation as a function of the simulation time is shown in Figure B.1. The RMSD fluctuates stably with an average value of 0.98 ± 0.14 Å, very similar to $[S{:}O_{1P},O_{2P}]^{Trans}$ simulation.

Conformational dynamics

The time evolution of key dihedral angles that define the conformation of the substrate and His119 residue are shown in Figure B.2 and average values are list in Table B.4. Again, it is found that the fluctuation of glycosyl dihedral angle χ_C is very small, being the average value -156.43° and denoying a very stable χ_C anti conformation. These values are similar to the ones obtained in the monothio substituted simulations (see above). Cytidine ribose occupies a North $C_{3'}$-endo conformation throughout the simulation, with an average value of 8.61° (see Figure B.3 and Table B.4). Besides, rotational dihedral angles do not undergo any conformational change. γ_C has infrequent and short-lived transitions from gauche$^+$ to trans conformation, ϵ and δ_C stays at gauche$^+$ during all the simulation, and ζ stays at trans conformation (see Figure B.9). Finally, imidazolium side chain of residue His119 retains an A-type conformation throughout the simulation, with a mean χ_1 value of 170.04° and χ_2 value of -105.35° (see Figure B.2).

Hydrogen bond interactions

Hydrogen bond formation between the ligand and the protein residues in the active site is described in Table 5.7, in which the average distances between atoms participating in the protein-ligand hydrogen bonds, with the hydrogen-bond occupancies, lifetime-average and number of events are list in. The time evolution of selected distances is plotted in Supporting Information Figure B.12. As in $[\text{native}]^{\text{Hyd}}$, dithio substituted $[\text{S:O}_{1\text{P}},\text{O}_{2\text{P}}]^{\text{Hyd}}$ complex makes strong hydrogen bond interactions with both catalytic histidines, His12 and His119, Gln11, and Phe120, but the interaction with Lys41 and some water molecules are weakened (see Table 5.7).

Similar to $[\text{S:O}_{1\text{P}}]^{\text{Hyd}}$ simulations, the equatorial non-bridging atom at 1P position makes hydrogen bond interactions with 3 different residues, the amino group of Lys7 and Lys41, and with Gln11. In this model, $[\text{S:O}_{1\text{P}},\text{O}_{2\text{P}}]^{\text{Hyd}}$, S_{1P}'s group hydrogen bond occupancies are 0.06 for the amino group of Lys7, as in $[\text{S:O}_{1\text{P}}]^{\text{Hyd}}$, and 0.29 for Lys41, while for Gln11 the occupancy is 0.82. Thus, S_{1P} makes a strong hydrogen bond interaction with residue Gln11, similar to $[\text{native}]^{\text{Hyd}}$, with an average distance of 2.730 Å, while the average distance with Lys41 and Lys7 is much higher, 3.544 Å, and 5.479 Å, respectively (see Table 5.7 and Figure B.12). The other sulfur atom, S_{2P}, makes hydrogen bond interactions with 3 different residues as the atom at 2P position in $[\text{native}]^{\text{Hyd}}$ and $[\text{S:O}_{2\text{P}}]^{\text{Hyd}}$ models. S_{2P} makes strong hydrogen bond interactions with imidazolium side chain of residue His12, *Occ* of 1.00 and $< R >$ of 2.381 Å, and the main chain of residue Phe120, *Occ* of 0.86 and $< R >$ of 2.704 Å. Also observed in this position there was a really weak interaction with imidazolium side chain of His119, *Occ* is only 0.03 and $< R > = 3.547$ Å (see Figure B.12). In addition, His12 makes a weak hydrogen bond interaction with $O_{2'}$ axial oxygen ($Occ = 0.10$ and $< R > = 2.758$ Å), whereas $O_{2'}$ has a quite strong hydrogen bond interaction with the amino group of Lys41 ($Occ = 0.58$ and $< R > = 3.213$ Å).

The imidazolium side chain of residue His119 makes a strong hydrogen bond interaction with O_{3T} axial oxygen and weak hydrogen bond interactions with $O_{3'}$ and S_{2P} equatorial atoms. The hydrogen bond occupancies are 0.74, 0.13 and 0.03 respectively (see Table 5.7 and Figure B.12).

Solvation of the substrate

A detail description of the ligand solvent interactions through the analysis of specific hydrogen bond interactions, Table 5.7, and the use of RDFs of solvent molecules around selected ligand atoms, Figure 5.2 is presented. Each hydrogen-bonding atom of the ligand, interacts with several different water molecules throughout the production simulation, except for the $O_{2'}$ axial oxygen, which does not interact with any water molecule. The largest hydrogen bond group occupancy corresponds to the interactions with S_{1P}

Table 5.7: Statistics of protein-ligand and water-ligand hydrogen-bonds observed in the simulation of the [S:O_{1P},O_{2P}]^{Hyd} model. The criteria for the existence of hydrogen bonds are 1) a maximum H $\cdots$ A(acceptor) distance of 2.4 Å for oxygens and 3.0 Å for sulfur, 2) a minimum $\widehat{DHA}$ of 120° and 3) a hydrogen bond lifetime $\tau \geq 5.0$ ps. The occupancy, Occ, is defined as the total lifetime of each hydrogen bond by the production time. The events are the number of times each hydrogen bond is formed. All distances are in Å.

	Distance		Hydrogen Bond		
Atom pair	MD†	MD‡	Occ	$\langle\tau\rangle$(ps)	# Events
LYS 7 Hζ_1 - CYT S_{1P}	5.668(2.170)	5.479(1.803)	0.01	15.00	3
LYS 7 Hζ_2 - CYT S_{1P}	5.683(2.385)		0.03	21.25	8
LYS 7 Hζ_3 - CYT S_{1P}	5.757(2.515)		0.02	16.11	9
GLN 11 Hϵ_{21} - CYT S_{1P}	2.730(0.266)	3.585(0.095)	0.82	96.85	51
LYS 41 Hζ_1 - CYT S_{1P}	3.491(0.376)	3.544(0.043)	0.10	15.25	38
LYS 41 Hζ_2 - CYT S_{1P}	3.451(0.379)		0.10	13.92	42
LYS 41 Hζ_3 - CYT S_{1P}	3.436(0.377)		0.09	13.03	43
WATER - CYT S_{1P}			1.04	13.36	465
HSP 12 Hϵ_2 - CYT S_{2P}	2.381(0.022)	3.305(0.010)	1.00	497.92	12
HSP 119 Hδ_1 - CYT S_{2P}	3.547(0.188)	4.182(0.123)	0.03	11.06	14
PHE 120 HN - CYT S_{2P}	2.704(0.065)	3.630(0.043)	0.86	59.94	86
WATER - CYT S_{2P}			0.49	17.32	170
HSP 12 Hϵ_2 - CYT $O_{2'}$	2.758(0.153)	3.528(0.116)	0.10	13.71	43
LYS 41 Hζ_1 - CYT $O_{2'}$	3.273(0.574)	3.213(0.082)	0.16	20.00	48
LYS 41 Hζ_2 - CYT $O_{2'}$	3.190(0.594)		0.20	19.68	61
LYS 41 Hζ_3 - CYT $O_{2'}$	3.182(0.604)		0.22	21.02	63
HSP 119 Hδ_1 - CYT $O_{3'}$	2.694(0.309)	3.151(0.073)	0.13	15.99	50
WATER - CYT $O_{3'}$			0.10	12.24	49
LYS 7 Hζ_1 - CYT O3T	5.799(3.938)	5.775(3.684)	0.01	20.00	3
LYS 7 Hζ_2 - CYT O3T	5.936(4.701)		0.02	14.38	8
LYS 7 Hζ_3 - CYT O3T	6.017(4.772)		0.00	12.50	2
HSP 119 Hδ_1 - CYT O3T	2.159(0.159)	3.039(0.080)	0.74	47.17	94
WATER - CYT O3T			0.29	11.99	143

† Arithmetic average hydrogen bond distance in the production simulation, 6000 ps, with square standard deviation in parentheses.
‡ Arithmetic average distance between the corresponding heavy atoms in the production trajectory, with square standard deviation in parentheses.

(1.04), while the largest hydrogen bond lifetime-average corresponds to S_{2P} (17.32 ps), but in general this model's active site is the less solvated one compared to the monothio substituted and native structures. This is specially evident from the analysis of the corresponding RDFs.

The RDFs show important differences with respect to the ones found in [native]^{Hyd} . The decrease of the heights of the peaks for S_{1P} and S_{2P} is quite dramatic compared to the peaks found for O_{1P} and O_{2P} in [native]^{Hyd} , and indicates a reduced solvation of the active site. Moreover, the rest of oxygen atoms also show an important decrease in their solvation. Notice the disappearance of the first peak in $g_{O_{2'}-O_W}$ and $g_{O_{3'}-O_W}$, indicating the loss of specific water interactions with these atoms. Notice that $g_{O_{2'}-O_W}$ has its first maximum, 1.05, at 5.7 Å. Comparing the various RDFs in [S:O_{1P},O_{2P}]^{Hyd} , S_{1P} atom shows the largest probability to find water molecules around it. Its maximum value is 1.50 at 3.4 Å. The next bigger probability corresponds to O_{3T} axial oxygen, with a maximum of 1.21 at 3.2 Å. S_{2P} equatorial atom has a maximum of 0.89 at 3.4 Å. Finally, analysis of g_{P-O_W} evidences the loss of solvent interactions in the active site as compared with [native]^{Hyd} and the monothio [S:O_{1P}]^{Hyd} and [S:O_{2P}]^{Hyd} complexes.

5.4 Discussion

The present work presents a series of molecular dynamics simulations of RNase A - 3',5'-CpA transphosphorylation's and hydrolysis' transition state/intermediate mimic thio substituted models (see Figure 5.1). In each case, transphosphorylation and hydrolysis, three different thio substituted models have been studied, two mono-thio substitutions at the oxyphosphorane's non-bridging oxygens positions (O_{1P} and O_{2P}), and a dithio substitution at both positions. These simulations employ a set of recently developed force field parameters for oxyphosphorane transition states or intermediates which are consistent with the CHARMM27 all-atom empirical force field for nucleic acids .[218] The parameters were developed based on density-functional calculations, with partial atomic charges derived from electrostatic potential fitting and Lennard-Jones parameters reproducing interaction energies with water molecules. Bonded energy terms (bond, angle and torsion parameters) were also derived from the density-functional calculations and renormalized to maintain compatibility with the existing CHARMM27 parameters for standard residues. Thus, the parameters used in this work reproduce the main geometrical and electronic characteristics of important phosphorane intermediates based on density-functional theory, but within a molecular mechanical approximation. This has allowed us to provide key structural information on how these reactive intermediates are structurally recognized in the active site of RNase A, using classical MD simulations with reasonable time scales. Due to the use of a classical force field, there is no possibility of providing with a reaction profile as a function of covalent bond formation and cleavage, and one is limited to the analysis of structural information. Still important insights on the reaction mechanism can be inferred from the structural data presented here. In the present section, we make a discussion

on the relevance of our structural findings for the classical catalytic reaction mechanism (see Section 2.3.2).

This section is divided into two parts, 1) discussion of the results obtained for the different studied thio models of the transphosphorylation's transition state mimic; and 2) discussion of thio substitutions on hydrolysis' transition state mimic model.

5.4.1 Thio substitutions at transphosphorylation's transition state mimic models

In this subsection we analyze and compare structural information relevant from our transphosphorylation's transition state mimic thio substituted models' dynamics (see Figure 5.1), with respect to the three most important residues of the classical reaction mechanism[73,76] (His12, His119, Lys41) and the role of solvent molecules in the stabilization of phosphoranes.

Similar to the results obtained for the native oxyphosphorane's model,[243] $[\text{native}]^{\text{Trans}}$, cytidine ribose stays at North $C_{3'}$-endo pucker, which is consistent with a canonical A-form RNA pucker, while the adenosine ribose sugar pucker shows more variability, and visits both the South $C_{2'}$-endo pucker (essentially) and North $C_{3'}$-endo pucker. Moreover, the orientation of the bases about the glycosidic bonds of the substrates fluctuate stably around an anti conformation (Tables B.1, B.2 and B.3). Often observed α and γ_A crankshaft motion in $B_I \leftrightarrow B_{II}$ transitions[234] is also presented in these simulations. Finally, it is remarkable that the side chain of residue His119 remains in its active A conformation that promotes catalysis[50,62] in each simulation. In this sense, His119 and adenine rings form continuous $\pi - \pi$ stacking interactions, that contribute to the stabilization of His119 A orientation and the adenine ring anti orientation. Therefore, overall we can say that thio substitution in non-bridging equatorial positions at the transphosphorylation's transition state, has a small effect on the conformation of the ligand.

Regarding Lys41, its side chain hydrogen bonds with $O_{2'}$ axial oxygen and the atom at 1P position, as in $[\text{native}]^{\text{Trans}}$ simulation[243] (Table 4.2). However, we observe that the strength of these interactions is very sensible to thio substitutions (see Tables 5.2, 5.3 and 5.4). The interaction between Lys41 and the $O_{2'}$ atom is of interest, as Lys41 is known to have a reduced pK_a (8.6-9.1),[236] and has been observed crystallographically to undergo structural change in the pH range 8.0-8.8,[237] and has even been suggested to act as general base/acid catalyst instead of His12.[60,61,87] It is possible that Lys41 might have a depressed pK_a due to interaction with nearby hydrophobic groups, as has been observed for a lysine residue in enolase.[238] However, to reconcile this question would require the calculation of the pK_a value of Lys41 (and possibly other residues) using a combined quantum mechanical/molecular mechanical (QM/MM) simulation and free

energy perturbation approach. This is beyond the scope of the current work that focuses on the structure and dynamics of thio substitutions at different stages along the reaction coordinate. It is of importance to note that thio substitutions weaken the interaction with $O_{2'}$. However, there is an interesting rescue effect of this interaction upon dithio substitution (Occ = 0.8 in $[native]^{Trans}$, Occ = 0.41 $[S{:}O_{1P}]^{Trans}$, Occ = 0.36 $[S{:}O_{2P}]^{Trans}$, Occ = 0.61 $[S{:}O_{1P},O_{2P}]^{Trans}$). This lower interaction is balanced in $[S{:}O_{2P}]^{Trans}$ with a larger interaction of Lys41 with the equatorial O_{1P}, 0.64 occupancy versus 0.3 in $[native]^{Trans}$ structure. The end result is that while for the native structure there is a preference for the interaction of Lys41 with $O_{2'}$, thio substitutions can affect these interactions and even reverse the trend in $[S{:}O_{2P}]^{Trans}$ (Tables 4.2, 5.2, 5.3 and 5.4).

An interesting rescue effect is also observed for Gln11 upon dithio substitution. Monothio compounds leads to a weakening of this interaction, while double sulfur substitution shows a much higher occupancy, more similar to the strong hydrogen bond interaction found for the $[native]^{Trans}$ model (Occ = 0.94 in $[native]^{Trans}$, Occ = 0.47 $[S{:}O_{1P}]^{Trans}$, Occ = 0.19 $[S{:}O_{2P}]^{Trans}$, Occ = 0.87 $[S{:}O_{1P},O_{2P}]^{Trans}$).

On the other hand, His12 forms strong hydrogen bond interactions with the phosphorane's equatorial non-bridging atom at 2P position (Occ values between 0.97 and 0.99 for all cases). Thus, the simulations results suggest that interaction with His12 is key to stabilize the excess of negative charge at the 2P non-bridging position. In addition, the atom at 2P equatorial position makes another two strong hydrogen bond interactions that provide stabilization of its negative charge: one hydrogen bond interaction with the main chain of residue Phe120, and the other one with a water molecule. Finally, and quite interestingly, a weak hydrogen bond interaction is observed for each simulation, between the atom at 2P equatorial position and residue His119, which was not present in $[native]^{Trans}$ simulation. On the other hand, the interaction between His12 and $O_{2'}$ axial oxygen is not strong in either of the models. These last occupancies are not what one would expect if His12 were to play the role of general base catalyst as in the classical mechanism.[73] Nonetheless, it is interesting to remark that in those models where $O_{2'}$ shows weak interaction with both Lys41 and His12 ($[S{:}O_{1P}]^{Trans}$ and $[S{:}O_{2P}]^{Trans}$), there are interactions of $O_{2'}$ with the solvent, that were not present in $[native]^{Trans}$ model.

The interaction between imidazolium side chain of His119 and axial oxygen $O_{5'}$ presents strong hydrogen bond interactions for each simulation (Occ = 1.00 in $[native]^{Trans}$, Occ = 1.00 $[S{:}O_{1P}]^{Trans}$, Occ = 0.97 $[S{:}O_{2P}]^{Trans}$, Occ = 0.99 $[S{:}O_{1P},O_{2P}]^{Trans}$). Therefore, it seems that thio effects has a low influence in the interaction with the leaving group of the transphosphorylation reaction.

On the other hand, the solvation of the ligand is very sensible to thio substitution and leads to a very meaningful reduction of the solvation of the

active site. This is very apparent in the shape of g_{P-O_W}, Figure 5.1, with a very critical reduction of its first peak with the number of sulfur atoms contained in the phosphorane. As expected, thio substitution at a given position leads to lower and less structured peaks at the thio-substituted positions, revealing a loss of solvent interactions with the thio-substituted atom and structure of the first solvation layer. However, thio substitutions can also affect the solvation of the atoms of the phosphorane that are not directly substituted, and in ways that are not easily predicted. For instance, in the case of $O_{2'}$ specific interactions with solvent appear in mono-thio substituted compounds that disappear again upon double thio substitution (see the first peak of the corresponding $g_{O_{2'}-O_W}$ Figure 5.1 distribution functions). Finally and quite surprisingly, an increase in the solvent interactions with $O_{5'}$ is revealed in the dithio model.

5.4.2 Thio substitutions at hydrolysis' transition state mimic models

In this subsection we discuss the structural information relevant from our hydrolysis' transition state mimic thio substituted models' dynamics (see Figure 5.1).

Similar to the results obtained for the native oxyphosphorane's model,[243] $[\text{native}]^{\text{Hyd}}$, cytidine ribose stays at North $C_{3'}$-endo pucker, which is consistent with a canonical A-form RNA pucker. Moreover, the orientation of the base about the glycosidic bond of the substrates fluctuates stably around an anti conformation (Tables B.4). Finally, the conformation of His119 seems to be a bit more sensible to the specific thio substitutions in this models than in the transphosphorylation's thio substituted models. In both $[\text{S:O}_{1\text{P}}]^{\text{Hyd}}$ and $[\text{S:O}_{1\text{P}},\text{O}_{2\text{P}}]^{\text{Hyd}}$ models residue His119 remains in its active A conformation that promotes catalysis,[50,62] similar to the $[\text{native}]^{\text{Hyd}}$ model, whereas in $[\text{S:O}_{2\text{P}}]^{\text{Hyd}}$ model, there is a conformational change to the inactive B-type conformation, after the first 2154 ps. When His119 is in the inactive conformation, it does not interact with the active site, therefore, this change could be significant in the catalytic stabilization of the corresponding oxyphosphoranes by RNase A. Nevertheless, overall we can say that thio substitution in non-bridging equatorial positions at the hydrolysis transition states, has a small effect on the conformation of the substrate.

Regarding Lys41, its side chain hydrogen bonds with $O_{2'}$ axial oxygen and the atom at 1P position, as in $[\text{native}]^{\text{Hyd}}$ and all the transphosphorylation's transition state mimic simulations.[243] However, we observe that the strength of these interactions is very sensible to thio substitutions (Tables 5.5, 5.6, 5.7 4.3 5.2, 5.3 5.4 and 4.2). As has been mention above, the interaction between Lys41 and the $O_{2'}$ atom is of interest. Lys41 is known to have a reduced pK_a (8.6-9.1),[236] and has been observed crystallographically to undergo structural change in the pH range 8.0-8.8,[237] and has even been

suggested to act as general base/acid catalyst instead of His12.[60,61,87] It is possible that Lys41 might have a depressed pK_a due to interaction with nearby hydrophobic groups, as has been observed for a lysine residue in enolase.[238] However, to reconcile this question would require the calculation of the pK_a value of Lys41 (and possibly other residues) using a combined QM/MM simulation and free energy perturbation approach. This is beyond the scope of the current work that focuses on the structure and dynamics of thio substitutions at different stages along the reaction coordinate. It is of importance to note that thio substitutions at 1P position, $[S{:}O_{1P}]^{Hyd}$ and $[S{:}O_{1P},O_{2P}]^{Hyd}$ complexes, strengthens the interaction of Lys41 with $O_{2'}$ axial oxygen (Occ = 0.46 $[native]^{Hyd}$, Occ = 0.87 $[S{:}O_{1P}]^{Hyd}$, Occ = 0.58 $[S{:}O_{1P},O_{2P}]^{Hyd}$) and concomitantly, weakens the interaction with S_{1P} (Occ = 0.57 $[native]^{Hyd}$ with O_{1P}, Occ = 0.36 $[S{:}O_{1P}]^{Hyd}$, Occ = 0.29 $[S{:}O_{1P},O_{2P}]^{Hyd}$). Nevertheless, thio substitution at 2P position has the reverse effect. It weakens extremely the hydrogen bond interaction with $O_{2'}$ in $[S{:}O_{2P}]^{Hyd}$, and strengthens the interaction with the equatorial O_{1P}, Occ = 0.01 and Occ = 0.89, respectively. The end result is that while for the native structure there is a preference for the interaction of Lys41 with O_{1P}, thio substitutions can affect these interactions and even reverse the trend in $[S{:}O_{1P}]^{Hyd}$ and $[S{:}O_{2P}]^{Hyd}$ (Tables 4.3, 5.5, 5.6 and 5.7). Besides, an interesting rescue effect in this position is observed for Gln11 upon dithio substitution. Monothio compounds leads to a weakening of this hydrogen bond interaction (Occ = 0.39 $[S{:}O_{1P}]^{Hyd}$, Occ = 0.46 $[S{:}O_{2P}]^{Hyd}$), while double sulfur substitution shows a much higher occupancy, 0.82, more similar to the strong hydrogen bond interaction found for the $[native]^{Hyd}$, Occ = 0.86.

On the other hand, the interaction of residue His12 with the oxyphosphorane is less sensible to thio effects. His12 forms strong hydrogen bond interactions with the phosphorane's equatorial non-bridging atom at 2P position (Occ = 1.00 for $[S{:}O_{1P}]^{Hyd}$, $[S{:}O_{1P},O_{2P}]^{Hyd}$ and $[native]^{Hyd}$, Occ = 0.88 $[S{:}O_{2P}]^{Hyd}$). Thus, the simulations results suggest that interaction with His12 is key to stabilize the excess of negative charge at the 2P non-bridging position, similar to the transphosphorylation's transition state models. In addition, atom at 2P equatorial position makes another two strong hydrogen bond interactions that provide stabilization of its negative charge: one hydrogen bond interaction with the main chain of residue Phe120, and a second one with solvent. Also observed in this position, there is a very weak hydrogen bond interaction with the imidazolium side chain of residue His119, which disappears when the sulfur atom is only at 1P position, $[S{:}O_{1P}]^{Hyd}$. On the other hand, the lower occupancy found for His12 and 2P position interaction in $[S{:}O_{2P}]^{Hyd}$ is balanced with a higher hydrogen bond interaction with $O_{2'}$, Occ = 0.28, the highest among the hydrolysis-type models (Tables 5.5, 5.6 5.7 and 4.3). These last occupancies are not what one would expect if His12 were to play the role of general acid catalyst as in the classical mechanism.[73]

Finally, it should be also highlighted that when His119 is in its active A-type conformation, it makes strong hydrogen bond interactions with O_{3T}, irrespective of the thio substitution. However, when His119 passes to the inactive B-type conformation in $[S{:}O_{2P}]^{Hyd}$, these hydrogen bond interactions are lost (see Figure B.11).

On the other hand, there are important differences in the solvation of the active site upon thio substitution (see Figure 5.2). In general there is a reduction in the solvation of the active site as sulfur atoms are introduced. This is very apparent in the shape of g_{P-O_W} with a critical reduction of its first peak upon thio substitution. Quite interestingly, among monothio complexes $[S{:}O_{2P}]^{Hyd}$ shows a larger solvation than $[S{:}O_{1P}]^{Hyd}$, partially motivated by the transition of His119 from A to B conformation, which allows the entrance of more solvent molecules in the active site. For instance, both $g_{O_{3T}-W}$ and $g_{O_{3'}-W}$ denote a higher solvation of these atoms in $[S{:}O_{2P}]^{Hyd}$ than even in $[native]^{Hyd}$. As expected, thio substitution at a given position leads to lower and less structured peaks at the thio-substituted positions, revealing a loss of solvent interactions with the thio-substituted atom and structure of the first solvation layer. However, thio substitutions can also affect the solvation of the atoms of the phosphorane that are not directly substituted, and in ways that are not easily predicted. For instance, in the case of $O_{2'}$ specific interactions with solvent are enhanced for $[S{:}O_{2P}]^{Hyd}$, highly reduced for $[S{:}O_{1P}]^{Hyd}$ and vanished for $[S{:}O_{1P},O_{2P}]^{Hyd}$, Figure 5.2.

5.5 Conclusions

In the present paper we report results of MD simulations of RNase A - 3',5'-CpA transphosphorylation's and hydrolysis' transition state or high-energy intermediate mimic thio substituted models, using recently developed force field parameters for phosphoryl transfer intermediates.[218] The aim of these studies is to determine how the structural relaxation, and differential solvation that occurs at discreet stages of the transesterification and cleavage reaction, are affected by a perturbation of the system in the form of thio substitution. Simulations were performed with explicit solvation, with rigorous electrostatics, and for several nanoseconds, which allow to make a detailed analysis of the main changes in hydrogen bond patterns and interaction with the solvent among the three pentacovalent phosphorane structures that represent the transition states or high-energy intermediates in the transphosphorylation and hydrolysis steps of the reaction.

Based on our simulations we can conclude that there are no major differences between the backbone structure of the studied six structures. Besides, the relative orientation of the bases, denoted by their glycosyl dihedral angle conformations, is very stable and it is not affected by the thio substitutions at non-bridging equatorial positions. In addition, it is quite remarkable that

the active A conformation of His119 is maintained for the five of the six structures studied and along the whole production run. $[S{:}O_{2P}]^{Hyd}$ model presents a transition from the active A conformation to the inactive B-type conformation after the first 2154 ps.

Analysis of the hydrogen bond occupancies between the substrates and RNase A shows that the residues directly interacting with the substrates are His12, His119, Lys41, Gln11, Phe120 and a variety of water molecules. However, the hydrogen bond pattern is sometimes highly affected by the thio substitutions, which shed light on the reaction mechanism and role of specific residues in transition state stabilization.

Our previous simulations of $[native]^{Trans}$ and $[native]^{Hyd}$ models[243] showed that the formation of pentacovalent phosphorane anchors His12 to interact with the equatorial O_{2P} oxygen. Thio effects conserve the strength of this interaction in both transphosphorylation and hydrolysis steps, underlining the robustness of this type of stabilization in RNase A. On the other hand, the interactions of residues Lys41 and His12 with $O_{2'}$ axial oxygen is highly influenced by the type of thio effect included, and therefore, seems more labile. This could have mechanistic consequences. In the transphosphorylation transition state, all the thio compounds show reduced interactions of Lys41 and His12 with $O_{2'}$. However, the hydrolysis case shows a more complicate picture, in some cases, as in $[S{:}O_{1P}]^{Hyd}$, the interaction of $O_{2'}$ with Lys41 is strengthened and the interaction with His12 is weakened. However, thio substitution at 2P position, $[S{:}O_{2P}]^{Hyd}$, leads to the reverse effect and induces the disappearance of the interaction with Lys41. These differences suggest that thio at these positions could alter the ability of Lys41/His12 to stabilize this position or even interchanged their roles to the ones proposed in the classical mechanism. This is in qualitative agreement with conclusions of other theoretical calculations[29,60,61,87] in which Lys41 has been claimed to possibly act as the acid catalyst in the hydrolysis reaction.

On the other hand, thio effects could be relevant in the propensity of His119 to be positioned as a good base catalyst in the hydrolysis reaction. The interaction of His119 with the attacking O_{3T} nucleophile presents a high sensibility to the thio substitution in the hydrolysis-type transition states. whereas this interaction is quite well conserved in $[S{:}O_{1P}]^{Hyd}$, $Occ = 0.96$, in $[S{:}O_{1P},O_{2P}]^{Trans}$ is reduced to $Occ = 0.74$, and in $[S{:}O_{2P}]^{Trans}$ is further reduced to $Occ = 0.13$ due to a conformational change of His119 from the active to the inactive conformation.

Analysis of substrate interaction with water molecules has been accomplished by the calculation of hydrogen bond occupancies between the substrate oxygen and sulfur atoms and specific water molecules and by the analysis of the radial distribution functions around key atoms of the phosphorane moieties. Simulations results point to a poorer solvation and to a less structured first solvation layer of the thio substituted models. This could lead to a lower solvent stabilization of these structures, that balance

the higher stabilization of oxyphosphoranes upon sulfur substitution. The changes in the shape of g_{P-O_W} are a clear indicative of this behavior. It is remarkable that the encapsulated water molecule hydrogen bonded directly to O_{2P} equatorial oxygen shown in [native]$^{\text{Trans}}$ and [native]$^{\text{Hyd}}$ simulations,[243] is maintained upon thio substitution at 1P position, confirming the important role of this water molecule in the stabilization of the negative charge accumulated at this equatorial phosphorane oxygen in both transphosphorylation and hydrolysis transitions states.[239] It is also interesting to remark that thio effects can also affect qualitatively the solvation of atoms that are not directly substituted, leading in some cases to a reduced solvation, as for the $O_{3'}$ atom in the transphosphorylation structures, and $O_{2'}$ in [S:O_{1P}]$^{\text{Hyd}}$ and [S:O_{1P},O_{2P}]$^{\text{Hyd}}$, or to an enhanced solvation in other cases, $O_{2'}$ in [S:O_{1P}]$^{\text{Trans}}$ and [S:O_{2P}]$^{\text{Trans}}$, and $O_{2'}$, $O_{3'}$ and O_{3T} for the [S:O_{2P}]$^{\text{Hyd}}$ complex. All these effects will certainly stabilize/destabilize these phosphorane structures. Whether all these effects can balance each other and lead to similar overall energetic barriers is an interesting question that can only be elucidated by QM/MM type of calculations. These will be the subject of further work in this area.

Further Work

In this part of the Thesis, we have characterized by Classical Molecular Dynamic Simulations, the structure of the reactant and intermediates of the hydrolysis reaction of RNase A - 3',5'-CpA complex, and how chemical changes at the oxyphosphorane itself affects their recognition by RNase A. The data collected here, could be taken as the base for future analysis of MD simulations of this type of oxyphosphorane intermediates. An extension of the present work of considerable interest could therefore be the analysis of oxyphosphorane intermediate analogues bound to RNase A mutants, that are known to affect the enzyme turnover rate, namely, H119N, H119A, H119D, H12A, H12E, H12D, K41R or K41A [58,240–242] which decrease k_{cat} between two and four orders of magnitude.

On the other hand, an analysis based on Molecular Mechanics-Poisson Boltzmann Surface Area, MM-PBSA[245] could be of great interest to analyze the binding free energy of the protein and the substrate. A free energy decomposition could be done to analyze the interaction of each residue with the substrate along the reaction mechanism, to estimate the magnitude and nature of the interaction at each step.

Part III

Reaction Mechanism Analysis by QM/MM methods

Introduction

As has been mention previously (see Section 2.3) *Bovine pancreatic ribonuclease A* is an ideal system for a theoretical study of the structural and dynamic basis of enzyme catalysis. This endoribonuclease catalyzes the breakdown of 3',5'-phosphodiester linkage of single stranded RNA. The wide range of biochemical, physical, and crystallographic data available for this enzyme[24,28,35,57,72] have led to proposals for the catalysis of the hydrolysis of RNA by a two-step mechanism,[76] in which a cyclic phosphate intermediate is formed and subsequently hydrolyzed (see Scheme 2.2). Both steps are thought to involve in-line displacement at the phosphorus, to pass through trigonal bipyramid (TBP) transition states and to be catalyzed by the concerted action of a general acid and a general base. In spite of the accumulated experimental information, a full understanding of the mechanism and a detailed analysis of the rate enhancement produced by ribonuclease has not been achieved, and nowadays, there are different mechanisms proposed[29,60,61,73,79,80,87,209] for RNA hydrolysis catalyzed by bovine pancreatic RNase A. Moreover, there remains considerable debate concerning the nature of the structures of the transitions states along the reaction path and their protonation states.[28,82]

It is difficult to obtain direct structural information about a transition state or intermediate from experiment. In the case of phosphoryl transfer reactions, for example, vanadate transition state mimics have been commonly used and characterized with X-ray crystallography.[211–213] However, the results obtain solely based on analysis of a complex with a pentavalent organo-vanadate have to be taken with some skepticism, because vanadate transition state mimics might be only marginally relevant,[214] and the relative stability of their different protonation states could be different to the ones of an oxyphosphorane. Theoretical calculations are potentially powerful tools that, together with experiment, can provide deep insight into the details of catalytic mechanisms. In order to reliably model the bond formation and cleavage that occurs in chemical reactions, a quantum electronic structure method is required. Among the most attractive methods for systems of intermediate size are modern density-functional methods.[246] For very large systems, however, these methods can be considerably computationally intensive, especially when combined with a high degree of conformational

sampling that is required to determine converged reaction free energy profiles. A strategy in such calculations is to define a subset of atoms (the active site, e.g. substrates and co-factor) that can be treated quantum mechanically (QM) while the remainder of the system (e.g. protein and solvent) is treated with molecular mechanics (MM) models, and in this way develop a combined QM/MM[177–179] potential that is appropriate for the reaction of interest (see Section 3.4). In the field of biological phosphoryl transfer reactions, recent advances have been obtained from the accurate parameterization of a semiempirical hamiltonian (AM1/d-PhoT)[97] which combined the possibility of efficient QM/MM calculations with the required degree of chemical accuracy to simulate reaction barriers and relative energies between isomers. In order to obtain a meaningful free energy profile that provides insight into mechanism, it is necessary to characterize the changes in enzyme conformation and contacts that occur along the reaction coordinate (variations in specific conformational variables), and especially at the transition state.

Separation into quantum and classical regions

Previously characterized transphosphorylation's oxyphosphorane model, $[\text{Trans}]^{oph}$, and hydrolysis' oxyphosphorane model, $[\text{Hyd}]^{oph}$, are partitioned into an inner region that is treated quantum-mechanically by a new semiempirical Hamiltonian AM1/d-PhoT[97] (QM region), and an outer region, which is described by the all-atom CHARMM27 nucleic acid force field[144–146] with extension to oxyphosphorane models[218] and TIP3P water model[225] (MM region).

The separation into quantum and classical regions is not straightforward in enzyme reactions, because the QM region is frequently bonded to MM region, as in this case, and is not clear how to define uniquely the boundary conditions for the electronic structure calculations of the QM region, nor how to incorporate the electrostatic and van der Waals effects of the classical region into the quantum region energy expression. We have chosen the Generalized Hybrid Orbitals (GHO)[185,190–194] method to connect the two regions (see Section 3.4.1).

The QM region includes 10 atoms from each of the imidazolium rings of residues His12 and His119, where the boundary atom is placed at Cβ atom; 8 atoms from the side chain of residue Lys41, where the GHO atom is at Cδ atom; and 16 atoms from the transphosphorylation's oxyphosphorane model of CpA, $[\text{Trans}]^{oph}$, where the boundary atoms are situated at cytidine ribose ring $C_{4'}$ and $C_{1'}$ and adenosine ribose ring $C_{4'}$, or 13 atoms from the hydrolysis' oxyphosphorane model of CpA, $[\text{Hyd}]^{oph}$, where the GHO atoms are placed at cytidine ribose ring $C_{4'}$ and $C_{1'}$ (see Figure 6.1). That makes up a total of 44 QM atoms for $[\text{Trans}]^{oph}$ and 41 QM atoms for $[\text{Hyd}]^{oph}$.

Chapter 6

Dianionic versus Monoanionic Oxyphosphorane Intermediates in the Hydrolysis of RNA Catalyzed by RNase A

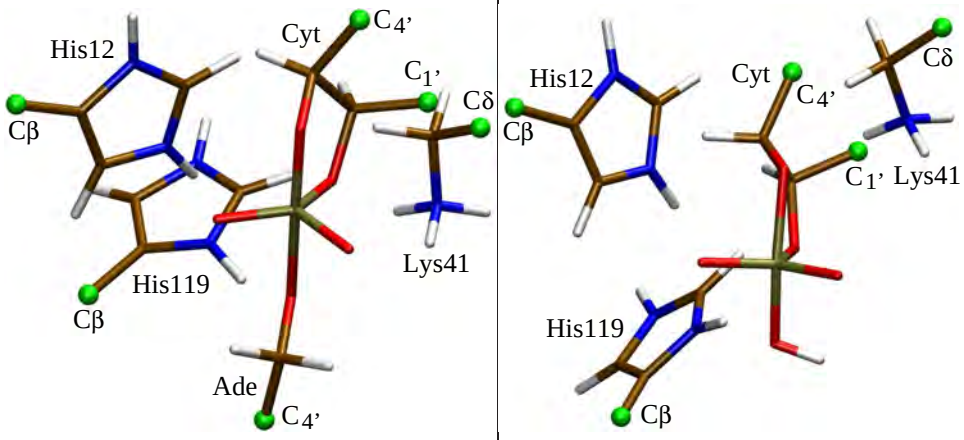
6.1 Introduction

Bovine pancreatic ribonuclease A has been studied for over 60 years. It is an ideal system for a theoretical study of the structural and dynamic basis of enzyme catalysis. This endoribonuclease catalyzes the breakdown of 3',5'-phosphodiester linkage of single stranded RNA. A particularly wide range of biochemical, physical, and crystallographic data are available for this enzyme.[24,35,72] These have led to proposals for the catalysis of the hydrolysis of RNA by a two-step mechanism,[76] in which a cyclic phosphate intermediate is formed and subsequently hydrolyzed. Both steps are thought to involve in-line displacement at the phosphorus, to pass through trigonal bipyramid (TBP) transition states and to be catalyzed by the concerted action of a general acid and a general base. In spite of the accumulated experimental information, a full understanding of the mechanism and a detailed analysis of the rate enhancement produced by ribonuclease has not been achieved, and nowadays, there are different mechanisms proposed[29,60,73,79,80] for RNA hydrolysis catalyzed by bovine pancreatic RNase A. Moreover, there remains considerable debate concerning the nature of the structures of the transitions states along the reaction path and their protonation states.[28,82] One aspect of considerable importance is the possible presence of stable intermediates in phosphoryl transfer reactions. This topic has been debated for many years[28,82,88,247] and it is intricate with the protonation state of the phosphorane intermediate. It was proposed that without proton transfer, no stable intermediate could be formed.[28,80,82,85,88–90,93] However, recently a

stable unprotonated (dianionic) intermediate has been found.[61] Estimation of the pK_a of model phoshoranes in solution[248] lead to values of pK_a^1 and pK_a^2 of 7.94 and 14.27. That means that in order to be able to have a dianionic oxyphosphorane intermediate at physiological pH, there should be a significant stabilization of dianionic structures in the active site of RNase A. The nature of the protonation state of the oxyphosphorane intermediates and how this protonation state affects the internal structure of the oxyphosphorane and the interaction with the active site of the protein is thus a fundamental problem and its understanding is highly relevant to analyze the various alternative mechanisms proposed.

As has been mentioned previously, it is difficult to obtain direct structural information about a transition state or intermediate from experiment. However, theoretical calculations are potentially powerful tools that, together with experiment, can provide deep insight into the details of catalytic mechanisms.

Figure 6.1: a) Representation of the 44 atoms which make up the [Trans]^{oph} model QM region. b) Representation of the 41 atoms which make up the [Hyd]^{oph} model QM region. The GHO boundary atoms are depicted in green.



(a) [Trans]^{oph} (b) [Hyd]^{oph}

In this work we investigate the structure and energetic profile of the dianionic vs monoanionic transphosphorylation's and hydrolysis' high-energy phosphate reaction intermediates that RNase A has to stabilize in order to catalyze this reaction, by QM/MM approach. Oxyphosphorane structures have been used to model these intermediates. We denote them as [Hyd]^{oph} , the oxyphosphorane model which mimics the hydrolysis' intermediate, and [Trans]^{oph} , the oxyphosphorane model which resembles the transphosphorylation's intermediate. The most plausible path towards the formation of a monoanionic oxyphosphorane is through a proton transfer between the positively charged His12 and one of the non-bridging phosphoryl oxygens,

O_{2P}, of the pentacovalent oxyphosphorane intermediate.[60] In addition, we discuss the effect that the protonation state of the oxyphosphorane has on important structural parameters of the oxyphosphorane itself, and how the protonation state affects the interactions of the intermediate with the active site residues and solvent molecules.

The chapter is organized as follows. The Computational Details section describes the QM/MM approach and the simulation set up. The Results section presents umbrella sampling results for the hydrolysis' oxyphosphorane model, which resembles the hydrolysis' intermediate (denoted as $[Hyd]^{oph}$) and transphosphorylation's oxyphosphorane model, which resembles the transphosphorylation's intermediate (denoted as $[Trans]^{oph}$). The Discussion section compares the obtain results. Finally, the Conclusion section summarizes the main points of the paper and outlines future research directions.

6.2 Computational details

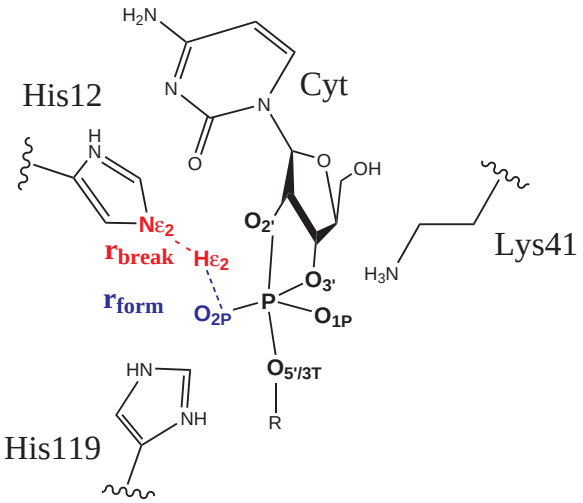
The starting structures for QM/MM calculations are taken from our previous work on *"Molecular Dynamics Simulation of RNase A - CpA transition state-like complexes"* [243] (see Section 4). The reaction pathway is studied using QM/MM molecular dynamics. The entire system is partitioned as described in the introduction of Part III. The reactive region of $[Trans]^{oph}$ has 44 atoms and the MM region consists on 6958 atoms, while the $[Hyd]^{oph}$ model has 41 atoms in the QM region and 6995 in the MM region.

The QM/MM simulations are carried out using Stochastic Boundary Conditions[172–175] (see Section 3.3.6). For an accurate description of the electrostatic forces on the QM subsystem arising from the environment, no electronic cutoff have been used. The protocol includes an initial equilibration of the configuration produced by MD simulations,[243] followed by a short run where only the MM part is free to move, while the QM part is kept frozen. Water molecules are initially relaxed for 300 steps of SD energy minimization[158,160] keeping all solute atoms and ion positions restrained to their initial coordinates. The restraints over the ions are then released, and the solvent (water and ions) is relaxed with 300 steps of SD energy minimization. Then the restraints over the solute except for the QM region are released and another relaxation of 300 steps of SD energy minimization is done. Once the MM region is minimized, another set of minimizations is done with a RESD[183] restrained on the combination of the axial distances of the oxyphosphoranes to maintain their structure. First a 500 steps of SD minimization is done, followed by two more minimizations of 500 steps using the adopted-basis set Newton-Raphson (ABNR) method.[126,127,161] From this point, restrained MD with the QM part fix, is performed starting at 0 K and heating to 298 K over 18 ps, and carried out to 133 ps. Finally, the

whole system is allowed to move for 40 ps.

From this starting structure, the *potential of mean force* (PMF)[195] is determined using the *weight histogram analysis method* (WHAM)[205,207,208] to combine the data obtain in the simulations executed applying umbrella sampling technique[162,199,249] along the reaction coordinate, RC, which is defined geometrically as the difference between the breaking bond, $r_{break} = r(N\epsilon_2 - H\epsilon_2)^{His12}$, and the forming bond, $r_{form} = r(H\epsilon_2^{His12} - O_{2P})$ (see Scheme 6.1 and Section 3.5).

$$RC = \ r_{break} - r_{form} \ = r(N\epsilon_2 - H\epsilon_2)^{His12} - \ r(H\epsilon_2^{His12} - O_{2P})$$



Scheme 6.1: Schematic representation of the reaction coordinate, $RC = r(N\epsilon_2 - H\epsilon_2)^{His12} - r(H\epsilon_2^{His12} - O_{2P})$. R= H in [Hyd]^{oph} and R= Ade in [Trans]^{oph} .

To span the entire range of the reaction coordinate from the dianionic oxyphosphorane to the monoanionic oxyphosphorane a total of 61 separate simulations (windows) are executed in the [Trans]^{oph} model and 56 in the [Hyd]^{oph} model. Each simulation is performed with a harmonic restraining potential centered at the location of that particular window (RC_i^0) and a force constant of 150 kcal·mol·Å^{-2}. For each window a set of three minimizations are done with Nuclear Overhauser effects, NOE, distance constraint method on the P–$O_{5'/3T}$ and P–$O_{2'}$ axial bonds and RESD[183] restrained on the combination of those axial distances. First, a 500 steps SD minimization is done, which is followed by two other 500 steps ABNR minimizations. After the minimizations, each window is heated from 110 K to 298 K over 25 ps, and carried out to 40 ps. Finally, the probability density of configurations along RC is collected for an additional 15 ps and is sorted into bins of 0.2 Å width.

Newton's equation is integrated using leapfrog Langevin[167] algorithm within a time step of 0.5 fs (Section 3.3.5).

6.3 Results

This section presents results for the studied two different oxyphosphorane structure simulations (see Scheme 4.1). One of them, $[\text{Trans}]^{oph}$, resembles the structure of transphosphorylation's intermediate state of RNase A - CpA complex, and the second one, $[\text{Hyd}]^{oph}$, resembles the structure of the hydrolysis' intermediate. The results are presented into different subsections for each structure. First, the PMF[195] of the reaction coordinate, $RC = r(N\epsilon_2 - H\epsilon_2)^{His12} - r(H\epsilon_2^{His12} - O_{2P})$, from a dianionic oxyphosphorane to a monoanionic oxyphosphorane is studied, see Scheme 6.1. Secondly, the ligand conformation is shown. The glycosyl dihedral angles $\chi_{C(O_{4'}-C_{1'}-N_1-C_2)}$ and $\chi_{A(O_{4'}-C_1-N_9-C_4)}$ describe, respectively, the orientation of the cytidine and adenine rings with respect to the sugar,[230,231] whereas, the pseudorotation phase angle[232] describes the sugars ring conformation, and the backbone torsion angles ($\alpha, \beta, \gamma, \delta, \epsilon$ and ζ) characterize the conformation of the phosphate group (see Scheme 4.1). Finally, we discuss the geometry of the active-site, focusing our attention on the conformation of residue His119, the structure of the oxyphosphorane, and the interaction between the ligand and the most important residues of the active site (His12, His119, and Lys41). The side chain of residue His119 can adopt two conformations denoted as A ($\chi_{1(N-C_\alpha-C_\beta-C_\gamma)} \sim 160°$) and B ($\chi_1 \sim -80°$),[62] which are related by a 142° rotation about $C_\alpha - C_\beta$ bond and 38° rotation about the $C_\beta - C_\gamma$ bond. Conformation A is considered as the active conformation, which promotes catalysis, whereas conformation B is considered as the inactive conformation.[50] In the crystallographic structure of the complex[50] and in previous MD simulations,[219,222,224,243] conformation A is observed.

6.3.1 Hydrolysis' Oxyphosphorane model, $[\text{Hyd}]^{oph}$

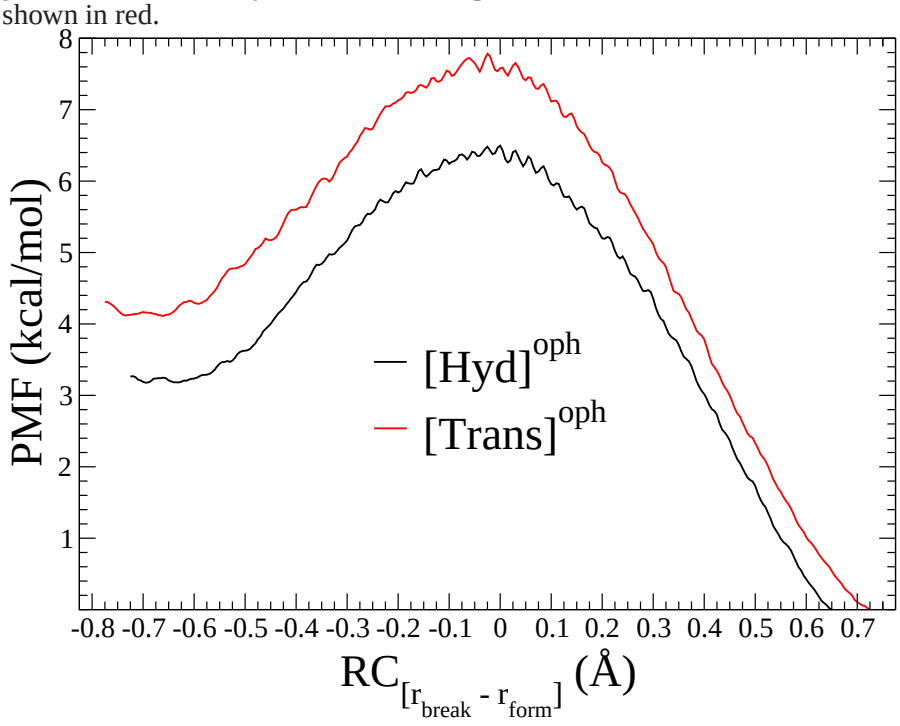
In this subsection we show the results obtain for the $[\text{Hyd}]^{oph}$ model, the structure of RNase A - 3',5'-CpA hydrolysis' oxyphosphorane intermediate (see Scheme 4.1). The QM region consists on 41 atoms. 5 of them are GHO[185,190–194] boundary atoms situated at C_β of residues His12 and His119, C_δ of residue Lys41 and cytidine ribose ring $C_{4'}$ and $C_{1'}$ (Figure 6.1). A total of 56 separate simulations (windows) were executed to span the entire range of the reaction coordinate, $RC = r(N\epsilon_2 - H\epsilon_2)^{His12} - r(H\epsilon_2^{His12} - O_{2P})$ from the dianionic oxyphosphorane to the monoanionic oxyphosphorane, [-0.725,0.650] Å (see Scheme 6.1).

Potential of Mean Force (PMF)

The PMF[195] shows two minima, dianionic [Hyd]^{oph} and monoanionic [Hyd]^{oph} , separated by a transition state maximum (see Figure 6.2). The first minimum has a RC value of -0.675 Å and a PMF value of 3.2 kcal/mol, the breaking bond is r_{break} = 1.065(0.023) Å while the forming bond is $rform$ = 1.734(0.032) Å. The transition state is reach at RC = 0.0 Å, where the length of both r_{break} and r_{form} is 1.29(0.03) Å, while the PMF value is 6.5 kcal/mol. After this point, the system evolves downhill towards the monoanionic [Hyd]^{oph} . At RC = 0.650 Å, we find the second minimum of the PMF, which is taken as the reference, r_{break} = 1.750(0.037) Å and r_{form} = 1.059(0.027) Å.

In addition, one-step mechanism is observed to evolve from a dianionic oxyphosphorane to a monoanionic one, with a ΔPMF = 3.3 kcal/mol (see Figure 6.2). The monoanionic [Hyd]^{oph} is energetically 3.2 kcal/mol more stable than the dianionic one.

Figure 6.2: PMF of [Hyd]^{oph} model is depicted in black while [Trans]^{oph} model is shown in red.



Ligand conformation and geometry of the active-site

The time evolution at each simulation window of key dihedral angles that define the conformation of the substrate and the His119 residue are shown

in Appendix Figure C.1 and the average values are listed in Table C.1. The fluctuations of χ_C dihedral angle and phosphodiester torsional angles are very small (see Figures C.1 and C.2), signifying that the anti conformation of χ_C is very stable. Besides, the cytidine ribose stays essentially at North $C_{3'}$-endo conformation throughout the simulation (see Figure C.1), which is the RNA's typical pucker conformation.[233] Finally, imidazolium side chain of residue His119 is retained in A conformation in all the windows. The mean χ_1 and χ_2 values are 179.5° and -113.15° (see Figure C.1).

On the other hand, the fluctuation of the oxyphosphoranes' bonds, P–O_X, are small. The oxyphosphorane structure is formed in all the studied simulations except in the window centered at -0.05 Å. In that window, P-$O_{3'}$ bond is broken, $O_{3'}$ phosphoryl oxygen is protonated by H_{3T} and the distance between the phosphorus and the oxygen atom is > 3.0 Å. Therefore, we have a 2'-terminal phosphate. The average and standard deviation values of the oxyphosphorane P-O_X distances without taking into account the previous cited window are 1.915(0.050) Å for P-$O_{2'}$, 1.719(0.039) Å for P-O_{3T}, 1.568(0.036) Å for P-O_{2P}, 1.515(0.022) Å for P-O_{1P} and 1.708(0.044) Å for P-$O_{3'}$. The equatorial bonds and the P-O_{3T} bond are shorter than the compare 2 uridine-vanadate crystallographic structures[212,213] (see Table 6.1). Besides, the fluctuation of dihedral angle $O_{3'}$-O_{2P}-O_{1P}-P is very small when the phosphorane structure is formed, its mean value is -3.72°.

Now we are going to focus our attention in the two windows where PMF has minimum values. First, the window centered at -0.675 Å, which corresponds to a dianionic $[Hyd]^{oph}$, and secondly, the window centered at 0.650 Å, which is a monoanionic-type $[Hyd]^{oph}$.

Dianionic Structure: window centered at RC = -0.675 Å

We start analyzing the window centered at RC = -0.675 Å. The oxyphosphorane structure average values are listed in Table 6.1. The reaction coordinate mean values are 1.065(0.023) Å for r_{break}, and 1.734(0.032) Å for r_{form}. Therefore, in this simulation we are studying a dianionic oxyphosphorane, that has a pentacovalent phosphorus atom with a distorted trigonal-bipyramid geometry.

The average and standard deviation values of the oxyphosphorane P-O_X distances are 1.962(0.045) Å for P-$O_{2'}$, 1.757(0.038) Å for P-O_{3T}, 1.536(0.021) Å for P-O_{2P}, 1.512(0.020) Å for P-O_{1P} and 1.703(0.039) Å for P-$O_{3'}$. The equatorial bonds and the P-O_{3T} axial bond are shorter than the distances in the two uridine-vanadate crystallographic structures[212,213] (see Table 6.1). The longest P-O distance corresponds to P-$O_{2'}$, and there is only small differences in the P-O bond length between non-bridging phosphoryl oxygens. Various angles characterizing the trigonal-bipyramid structure, can also be found in Table 6.1. In particular, the angle O_{3T}-P-$O_{2'}$ is often taken to characterize the distortion of the phosphorane with respect to the ideal linearity

Table 6.1: Comparison of two uridine-vanadate crystallographic structures (mimic of the hydrolysis transition state) and our simulations' oxyphosphorane structures, [Trans]^{TS} and [Hyd]^{TS} . All distances are in Å and angles in degrees. Standard deviation is in parentheses.

Atom pair	6RSA[212]	1RUV[213]	[Hyd]^{TS}		[Trans]^{TS}	
			z=-0.675	z=0.65	z=-0.725	z=0.725
P/V - $O_{3T/O5'}$	1.91	1.91	1.76(0.04)	1.69(0.03)	1.94(0.04)	1.82(0.04)
P/V - $O_{2'}$	1.88	2.00	1.96(0.04)	1.88(0.03)	1.83(0.04)	1.73(0.03)
P/V - $O_{3'}$	1.91	2.00	1.70(0.04)	1.70(0.03)	1.71(0.04)	1.72(0.04)
P/V - O_{2P}	1.75	1.78	1.54(0.02)	1.61(0.02)	1.53(0.03)	1.62(0.02)
P/V - O_{1P}	1.75	1.34	1.51(0.02)	1.51(0.02)	1.52(0.02)	1.51(0.02)
$O_{3T/O5'}$ - P/V - $O_{2'}$	165.6	149.5	166.7(4.0)	168.3(4.1)	165.6(3.3)	162.9(4.1)
$O_{3T/O5'}$ - P/V - $O_{3'}$	77.8	73.1	85.1(3.0)	85.6(2.6)	77.8(2.6)	78.9(2.5)
$O_{3T/O5'}$ - P/V - O_{2P}	91.7	96.0	95.1(3.2)	91.1(3.3)	92.1(3.1)	87.3(3.4)
$O_{3T/O5'}$ - P/V - O_{1P}	83.7	100.6	94.8(3.4)	98.5(3.6)	91.6(2.9)	97.0(3.4)
$O_{2'}$ - P/V - $O_{3'}$	91.0	76.9	83.2(2.6)	85.7(2.3)	88.4(2.1)	88.0(2.2)
$O_{2'}$ - P/V - O_{2P}	101.5	104.9	93.5(3.4)	87.6(2.6)	94.2(3.1)	90.9(3.2)
$O_{2'}$ - P/V - O_{1P}	97.5	96.0	87.8(2.9)	91.1(3.2)	94.7(2.8)	98.0(3.5)
O_{2P} - P/V - O_{1P}	107.5	103.1	124.3(4.4)	125.2(5.3)	125.6(4.1)	121.3(4.3)
O_{2P} - P/V - $O_{3'}$	117.9	123.2	114.7(4.0)	115.9(5.3)	119.7(4.8)	119.6(5.5)
O_{1P} - P/V - $O_{3'}$	131.3	133.6	120.6(5.2)	118.3(5.3)	113.9(4.3)	118.4(5.0)
$O_{3'}$ - O_{2P} - O_{1P} - P/V	10.9	2.2	-4.0	-4.3	5.8	4.8

between axial P-O bonds. An average value of 166.7° is observed from our simulations, similar to the one observed in 6RSA[212] and significantly more linear than the value of 149.5 of the 1RUV vanadate structure.[213]

Hydrogen bond formation between the ligand and the protein residues in the active site for this window is described in Table 6.2, in which the average distances between atoms participating in protein-ligand hydrogen bonds, with the hydrogen bond occupancies, average time and number of events are listed. The time evolution of selected distances between the protein and ligand atoms is plotted in Appendix Figure C.3.

O_{1P} equatorial oxygen is hydrogen bonded by the amino side chain of residue Lys41, with a heavy-atom average distance of 2.761 Å and a group occupancy of 1.0 (see Table 6.2 and Figure C.3). Besides, residue Gln11 is very close to O_{1P} and in some other dianionic windows is hydrogen bonding O_{1P}, which is also hydrogen bonded at least by one water molecule in some windows, but not in this one. On the other hand, equatorial O_{2P} oxygen makes one strong hydrogen bond interaction with imidazolium side chain of His12, and two quite strong hydrogen bond interactions with the main chain of Phe120 and a water molecule. The average distances are $< R > = 1.734$

Å, $< R >$ = 2.308 Å and $< R >$ = 2.246 Å, respectively, with a hydrogen bond occupancy of Occ = 1.00 for His12 and Occ = 0.6 for Phe120 and the water molecule. It worth mentioning, that in other dianionic windows residue His119 makes a fourth hydrogen bond interaction with this equatorial oxygen. Interestingly, His12 interacts with $O_{2'}$ axial oxygen with $H\epsilon_1$ instead of with $H\epsilon_2$ as was seen in previous MD simulations.[80,219–221,243] The average distance is $< R >$ = 1.922 Å with a hydrogen bond occupancy of Occ = 1.0, while residue Lys41 makes a not strong hydrogen bond interaction with it, $< R >$ = 2.278 Å and Occ = 0.4 (Figure C.3).

Table 6.2: Statistics of protein-ligand and water-ligand hydrogen-bonds observed in window centered at -0.675 Å of [Hyd]^{oph} model. The criteria for the existence of hydrogen bonds are 1) a maximum H $\cdots$ A(acceptor) distance of 2.4 Å, 2) a minimum $\widehat{DHA}$ of 120° and 3) a hydrogen bond lifetime $\tau \geq$ 5.0 ps. The occupancy, Occ, is defined as the total lifetime of each hydrogen bond by the production time. The events are the number of times each hydrogen bond is formed. All distances are in Å. HSP stands for a doubly protonated His.

	Distance		Hydrogen Bond		
Atom pair	MD†	MD‡	Occ	$\langle\tau\rangle$(ps)	# Events
LYS 41 $H\zeta_2$ - CYT O_{1P}	2.045(0.258)	2.761(0.102)	0.60	9.00	1
LYS 41 $H\zeta_3$ - CYT O_{1P}	2.739(0.394)	2.761(0.102)	0.40	6.00	1
HSP 12 $H\epsilon_2$ - CYT O_{2P}	1.734(0.032)	2.756(0.049)	1.00	15.00	1
PHE 120 HN - CYT O_{2P}	2.308(0.219)	3.220(0.177)	0.60	9.00	1
WATER 397 - CYT O_{2P}	2.246(0.330)	3.095(0.234)	0.60	9.00	1
HSP 12 $H\epsilon_1$ - CYT $O_{2'}$	1.922(0.143)	2.778(0.099)	1.00	15.00	1
LYS 41 $H\zeta_2$ - CYT $O_{2'}$	2.278(0.395)	3.209(0.300)	0.40	6.00	1
HSP 119 $H\epsilon_1$ - CYT $O_{3'}$	2.377(0.252)	2.794(0.175)	0.80	12.00	1
HSP 119 $H\delta_1$ - CYT O_{3T}	2.031(0.216)	2.828(0.134)	1.00	15.00	1
WATER 1432 - CYT O_{3T}	1.612(0.135)	2.521(0.120)	1.00	15.00	1
WATER 3991 - CYT O_{3T}	2.592(0.531)	3.167(0.362)	0.40	6.00	1
WATER 4218 - CYT H_{3T}	2.161(0.454)	2.925(0.315)	0.40	6.00	1

† Arithmetic average hydrogen bond distance for 15.0 ps, with standard deviation in parentheses.

‡ Arithmetic average distance between the corresponding heavy atoms in 15.0 ps, with standard deviation in parentheses.

Equatorial oxygen $O_{3'}$ makes a strong hydrogen bond interaction with the imidazolium side chain of His119, average distance $< R >$ = 2.377 Å and Occ = 0.80. The imidazolium side chain of His119 is also making a strong hydrogen bond interaction with O_{3T} axial oxygen, $< R >$ = 2.031 Å and Occ = 1.0, which is hydrogen bonded as well by two water molecules.

Finally, H_{3T} makes a not strong hydrogen bond interaction with a water molecule ($< R > = 2.161$ Å, $Occ = 0.40$). More water molecules are at a hydrogen bond distance from this atom, however, the interactions they make are < 5.0 ps (see Figure C.3).

In Figure 6.3, the RDFs between solvent's oxygen atoms and selected atoms of the oxyphosphorane species can be found. The analysis of RDFs can give interesting insight into the solvation characteristics of the oxyphosphorane moiety. Briefly, O_{3T} axial oxygen and O_{1P} equatorial oxygen present a fully-exposed solvent interaction, high peak followed by a decrease towards bulk values. The $g_{O_{3T}-O_W}$ has a maximum value of 2.9 at 2.5 Å and $g_{O_{1P}-O_W}$ is 2.1 at 2.9 Å. On the other hand, O_{2P} equatorial oxygen presents a well defined first peak and decreases almost to 0.0 at a distance of 4.3 Å. With respect to the rest of the oxyphosphorane's oxygen atoms (equatorial $O_{3'}$ and axial $O_{2'}$), the RDFs present signatures of a poorly solvated situation that indicate less ordered solvent and more rapid exchange relative to the RDFs for the axial O_{3T}.

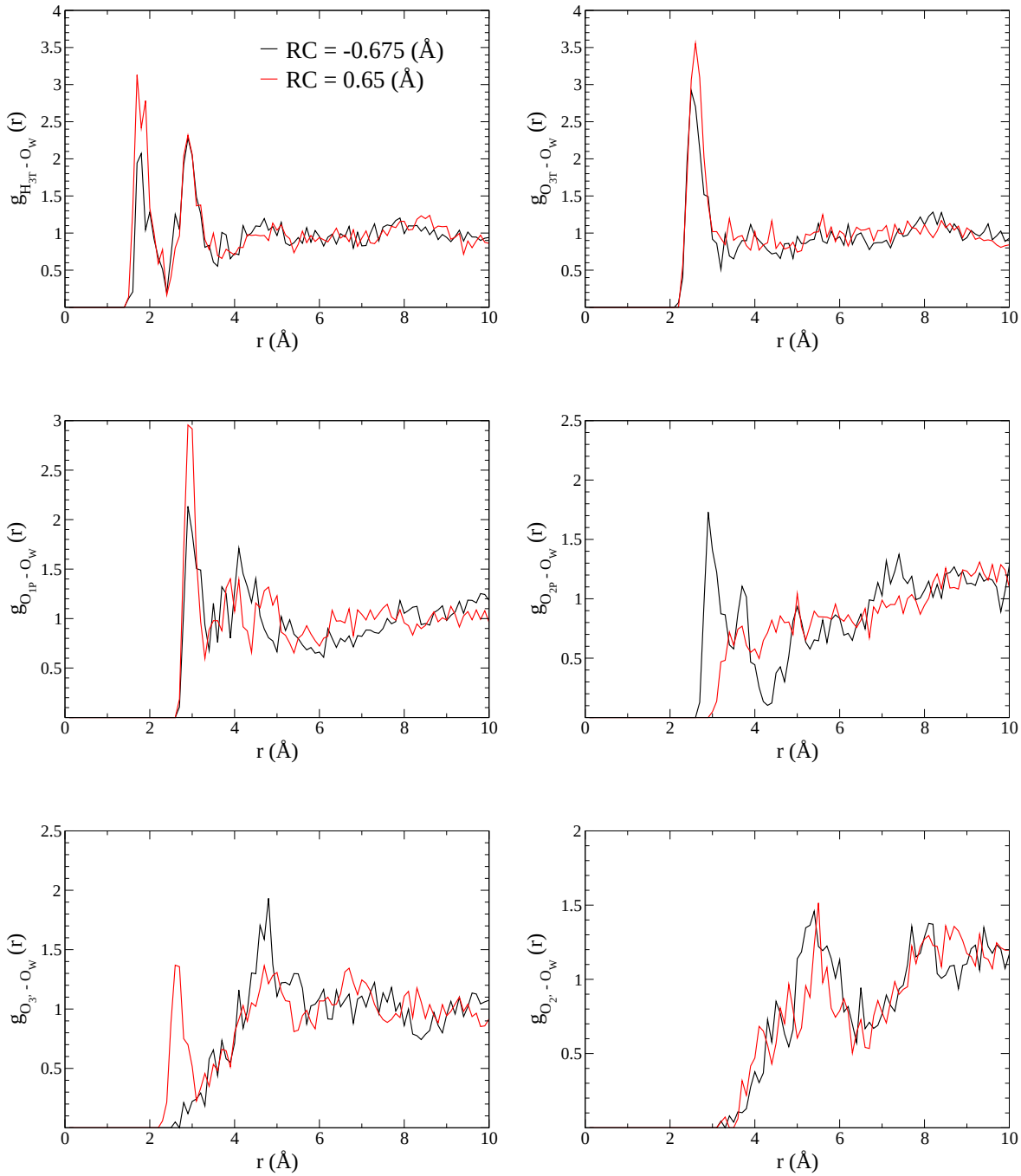
Monoanionic Structure: window centered at RC = 0.650 Å

Next we analyze the window centered at the reaction coordinate $RC = 0.650$ Å. The oxyphosphorane structure average values are list in Table 6.1. The reaction coordinate mean values are 1.723(0.037) Å for r_{break}, and 1.059(0.027) Å for r_{form}. Therefore, in this simulation we are studying a monoanionic oxyphosphorane, that has a pentacovalent phosphorus atom with a distorted trigonal-bipyramid geometry.

The average and standard deviation values of the oxyphosphorane's P-O_X distances for this window are 1.876(0.034) Å for P-$O_{2'}$, 1.685(0.028) Å for P-O_{3T}, 1.608(0.023) Å for P-O_{2P}, 1.510(0.017) Å for P-O_{1P} and 1.701(0.035) Å for P-$O_{3'}$. Compared to the two uridine-vanadate crystallographic structures,[212,213] all bonds are shorter (see Table 6.1). With respect to the previous dianionic structure, there is a significant shortening in the two axial P-O bond lengths, 0.07 Å, for P-O_{3T} and 0.08 Å, for P-$O_{2'}$, and a lengthening of the P-O_{2P} bond by 0.06 Å. Therefore, now the two non-bridging phosphoryl oxygens show a difference in bond length of 0.1 Å. Notice that this anisotropy in phosphoryl oxygen bond lengths have been used to identify a monoanionic uridine vanadate in the past.[60] The longest P-O distance still corresponds to P-$O_{2'}$ in the monoanionic phosphorane. Quite interestingly, the angles defining the trigonal-bipyramid structure of the oxyphosphorane show little change from the dianionic to the monoanionic structure. This is specially interesting for the O_{3T}-P-$O_{2'}$ angle, which goes from 166.7° in the dianionic structure to 168.3° in the monoanionic oxyphosphorane.

In Table 6.3, we can find selected data for the hydrogen bond formation between the oxyphosphorane and the protein residues, namely, average distances between atoms participating in protein-ligand hydrogen bonds, with

Figure 6.3: Distribution of water around oxyphosphorane's oxygens is shown for two windows of [Hyd]^{oph} model, RC = -0.675 Å in black, and RC = 0.65 Å in red. H_{3T} is at top left corner, at top right corner O_{3T} is depicted, O_{1P} and O_{2P} non-bridging phosphoryl atoms are in the middle left and right, respectively, and finally $O_{3'}$ and $O_{2'}$ are at the bottom left and right corner.



the hydrogen bond occupancies, average time and number of events.

Table 6.3: Statistics of protein-ligand and water-ligand hydrogen-bonds observed in window centered at 0.65 Å of [Hyd]^{oph} model. The criteria for the existence of hydrogen bonds are 1) a maximum H $\cdots$ A(acceptor) distance of 2.4 Å, 2) a minimum $\widehat{DHA}$ of 120° and 3) a hydrogen bond lifetime $\tau \geq 5.0$ ps. The occupancy, Occ, is defined as the total lifetime of each hydrogen bond by the production time. The events are the number of times each hydrogen bond is formed. All distances are in Å. HSP stands for a doubly protonated His.

	Distance		Hydrogen Bond		
Atom pair	MD†	MD‡	Occ	$\langle\tau\rangle$(ps)	# Events
LYS 41 Hζ_3 - CYT O_{1P}	2.307(0.363)	2.929(0.193)	0.40	6.00	1
LYS 41 Hϵ_2 - CYT O_{1P}	2.448(0.269)	3.053(0.226)	0.60	9.00	1
WATER 3229 - CYT O_{1P}	3.191(1.339)	3.612(1.052)	0.60	9.00	1
WATER 5311 - CYT O_{1P}	2.268(0.365)	3.086(0.238)	1.00	15.00	1
LYS 41 Hζ_3 - CYT $O_{2'}$	2.003(0.224)	2.952(0.116)	1.00	15.00	1
HSP 12 H$\epsilon_2^{(a)}$ - CYT $O_{2'}$	2.403(0.158)	3.416(0.173)	0.40	6.00	1
WATER 1344 - CYT $O_{3'}$	2.086(0.546)	2.787(0.264)	1.00	15.00	1
WATER 5573 - CYT O_{3T}	2.395(0.780)	2.626(0.194)	0.40	6.00	1
WATER 1432 - CYT O_{3T}	3.690(1.388)	4.626(1.067)	0.40	6.00	1
WATER 5573 - CYT H_{3T}	2.511(0.478)	2.626(0.194)	0.40	6.00	1
WATER 4970 - CYT H_{3T}	2.498(0.917)	3.206(0.674)	0.40	6.00	1
HSP 12 H$\epsilon_2^{(a)}$ - HSP 12 Nϵ_2	1.750(0.037)	2.750(0.056)	1.00	15.00	1

† Arithmetic average hydrogen bond distance for 15.0 ps, with standard deviation in parentheses.

‡ Arithmetic average distance between the corresponding heavy atoms in 15.0 ps, with standard deviation in parentheses.

Lys41 makes two quite strong hydrogen bond interactions with O_{1P} equatorial oxygen, with Hϵ_2 and with Hζ_3 ($< R > = 2.448$ Å, $Occ = 0.60$ for Hϵ_2 and $< R > = 2.307$ Å, $Occ = 0.40$ for Hζ_3). Two water molecules make also a hydrogen bond interaction with this equatorial oxygen. Besides, residue Gln11 is very close to O_{1P} equatorial oxygen and in some other windows is making a hydrogen bond interaction (see Figure C.4). In addition, Lys41 is making a strong hydrogen bond interaction with $O_{2'}$ axial oxygen ($< R > =$ 2.003 Å, $Occ = 1.00$), which is also hydrogen bonded by His12's Hϵ_2, $< R >$ = 2.403 Å and $Occ = 0.4$ (remember that Hϵ_2 is bonded to equatorial O_{2P} in this window). This Lys41-$O_{2'}$ interaction is stronger than in the dianionic structure. On the other hand, the hydrogen bond between His12 and O_{2P} equatorial oxygen is maintained in the monoanionic simulation, although now His12 is neutral and O_{2P} is the donor atom instead of the acceptor. Hϵ_2

is hydrogen bonding His12 during all the simulation, with an occupancy of 1.0 and average distance of 1.750 Å. Furthermore, there is a significant loss of interactions with the protein for for $O_{3'}$ and O_{3T}, and now these atoms only form hydrogen bond interactions with water molecules (see Table 6.3). In addition, O_{3T} presents a water bridge with imidazolium side chain of His119. Although in some other windows, His119 interacts directly with O_{3T}.

The analysis of the RDFs of Figure 6.3 can shed light on the changes in solvation structure of the oxyphosphorane upon protonation. The RDFs show similar characteristics as the ones commented above for the dianionic case, except for O_{2P} and $O_{3'}$. In the case of O_{2P} equatorial non-bridging oxygen, there is a significant reduction of the first peak in the RDFs. This suggests a loss of direct O_{2P}-solvent interactions upon O_{2P} protonation. The reverse is true for $O_{3'}$ equatorial oxygen, and now there is an increase in the solvation of this atom upon formation of the monoanionic phosphorane.

6.3.2 Transphosphorylation's Oxyphosphorane model, [Trans]^{oph}

In this subsection we show the results obtain for the [Trans]^{oph} model, which resembles the structure of RNase A - 3',5'-CpA transphosphorylation's oxyphosphorane intermediate (see Scheme 4.1). The QM region consists on 44 atoms. Six of them GHO[185,190–194] boundary atoms situated at C_β of residues His12 and His119, C_δ of residue Lys41, cytidine ribose ring $C_{4'}$ and $C_{1'}$, and adenosine ribose ring $C_{4'}$ (see Figure 6.1). A total of 61 separate simulations (windows) were executed to span the entire range of the reaction coordinate, $RC = r(N\epsilon_2 - H\epsilon_2)^{His12} - r(H\epsilon_2^{His12} - O_{2P})$, from the dianionic oxyphosphorane to the monoanionic oxyphosphorane, [-0.775,0.725] Å (see Scheme 6.1).

Potential of Mean Force (PMF)

The potential of mean force (PMF) shows two minima, dianionic [Trans]^{oph} and monoanionic [Trans]^{oph} , separated by a transition state maximum (see Figure 6.2). The first minimum has a RC value of -0.725 Å and a PMF value of 4.1 kcal/mol, the breaking bond is r_{break} = 1.064(0.036) Å while the forming bond is r_{form} = 1.787(0.036) Å. The transition state is reach at RC = -0.025 Å, where the length of r_{break} is 1.271(0.031) Å and r_{form} is 1.305(0.030) Å, while the PMF value is 7.8 kcal/mol. After this point, the system evolves downhill towards the monoanionic [Trans]^{oph} . At RC = 0.725 Å, we find the second minimum of the PMF, which is taken as the reference, r_{break} = 1.787(0.033) Å and r_{form} = 1.046(0.025) Å.

In addition, one-step mechanism is observed to evolve from a dianionic oxyphosphorane to a monoanionic one, with a ΔPMF = 3.7 kcal/mol (see Figure 6.2). The monoanionic [Hyd]^{oph} is energetically 4.1 kcal/mol more stable than the dianionic one.

Ligand conformation and geometry of the active-site

The time evolution at each simulation window of key dihedral angles that define the conformation of the substrate and the His119 residue are shown in Supporting Information Figure C.1 and the average values are listed in Table C.1. The fluctuations of χ_C and χ_A dihedral angles are very small, signifying that their anti conformations are very stable. The adenine ribose stays essentially at South $C_{2'}$-endo pucker, while cytidine ribose stays essentially at North $C_{3'}$-endo conformation throughout the simulation (see Figure C.1). γ_C torsional angle presents some conformational transitions from gauche$^+$ to trans (Figure C.2) known as crankshaft motion, which is often observed in $B_I \leftrightarrow B_{II}$ transitions.[234] Finally, imidazolium side chain of residue His119 is retained in A conformation in all the windows. The mean χ_1 and χ_2 values are 174.1° and -114.5° (see Figure C.1).

On the other hand, the fluctuation of the oxyphosphoranes' bonds, P–O_X, are small. The oxyphosphorane structure is formed in all the studied simulations. The average and standard deviation values of the oxyphosphorane P-O_X distances are 1.782(0.048) Å for P-$O_{2'}$, 1.883(0.056) Å for P-$O_{5'}$, 1.571(0.040) Å for P-O_{2P}, 1.513(0.022) Å for P-O_{1P} and 1.709(0.038) Å for P-$O_{3'}$. Compared to the two uridine-vanadate crystallographic structures,[212,213] all bonds are shorter (see Table A.2). Besides, the fluctuation of dihedral angle $O_{3'}$-O_{2P}-O_{1P}-P is very small, its mean value is 5.3°.

Now we are going to focus our attention in the two windows where PMF has minimum values. First, the window centered at -0.725 Å, which corresponds to a dianionic [Trans]^{oph} , and secondly, the window centered at 0.725 Å, which is a monoanionic-type [Trans]^{oph} .

Dianionic Structure: window centered at RC = -0.725 Å

We start analyzing the window centered at RC = -0.725 Å. The oxyphosphorane structure average values are listed in Table 6.1. The reaction coordinate mean values are 1.064(0.025) Å for r_{break}, and 1.787(0.036) Å for r_{form}. Therefore, in this simulation we are studying a dianionic oxyphosphorane, that has a pentacovalent phosphorus atom with a distorted trigonal-bipyramid geometry.

The average and standard deviation values of the oxyphosphorane P-O_X distances are 1.827(0.040) Å for P-$O_{2'}$, 1.938(0.045) Å for P-$O_{5'}$, 1.526(0.026) Å for P-O_{2P}, 1.519(0.025) Å for P-O_{1P} and 1.706(0.038) Å for P-$O_{3'}$. Compared to the two uridine-vanadate crystallographic structures,[212,213] all bonds are shorter (see Table 6.1). The longest P-O distance corresponds to P-$O_{5'}$, and there is only small differences in the P-O bond length between non-bridging phosphoryl oxygens. Various angles characterizing the trigonal-bipyramid structure, can also be found in Table 6.1. In particular, the angle $O_{5'}$-P-$O_{2'}$ is often taken to characterize the distortion of the phosphorane

with respect to the ideal linearity between axial P-O bonds. An average value of 165.6° is observed from our simulations, identical to the one observed in 6RSA[212] and significantly more linear than the value of 149.5 of the 1RUV vanadate structure.[213]

Hydrogen bond formation between the ligand and the protein residues in the active site for this window is described in Table 6.4, in which the average distances between atoms participating in protein-ligand hydrogen bonds, with the hydrogen bond occupancies, average time and number of events are listed. The time evolution of selected distances between the protein and ligand atoms is plotted in Supporting Information Figure C.5.

Table 6.4: Statistics of protein-ligand and water-ligand hydrogen-bonds observed in window centered at -0.725 Å of [Trans]^{oph} model. The criteria for the existence of hydrogen bonds are 1) a maximum H $\cdots$ A(acceptor) distance of 2.4 Å, 2) a minimum $\widehat{DHA}$ of 120° and 3) a hydrogen bond lifetime $\tau \geq 5.0$ ps. The occupancy, Occ, is defined as the total lifetime of each hydrogen bond by the production time. The events are the number of times each hydrogen bond is formed. All distances are in Å. HSP stands for a doubly protonated His.

Atom pair	Distance		Hydrogen Bond		
	MD†	MD‡	Occ	$\langle\tau\rangle$(ps)	# Events
GLN 11 Hϵ_{21} - CYT O$_{1P}$	2.105(0.148)	3.035(0.119)	0.80	12.00	1
LYS 41 Hζ_1 - CYT O$_{1P}$	2.030(0.413)	2.839(0.109)	0.80	12.00	1
WATER 179 - CYT O$_{1P}$	2.157(0.383)	3.042(0.279)	0.80	6.00	2
HSP 12 Hϵ_2 - CYT O$_{2P}$	1.787(0.036)	2.821(0.057)	1.00	15.00	1
PHE 120 HN - CYT O$_{2P}$	2.324(0.157)	3.287(0.153)	0.40	6.00	1
WATER 11 - CYT O$_{2P}$	2.053(0.153)	2.984(0.138)	1.00	15.00	1
WATER 628 - CYT O$_{3'}$		2.555(0.110)	1.40	10.50	2
HSP 119 Hδ_1 - CYT O$_{5'}$	1.837(0.115)	2.805(0.092)	1.00	15.00	1
WATER 628 - CYT O$_{5'}$	1.648(0.141)	2.522(0.107)	1.00	15.00	1

† Arithmetic average hydrogen bond distance for 15.0 ps, with standard deviation in parentheses.

‡ Arithmetic average distance between the corresponding heavy atoms in 15.0 ps, with standard deviation in parentheses.

O$_{1P}$ equatorial oxygen makes 3 strong hydrogen bond interactions with the amino side chain of residue Lys41, Gln11 and a water molecule. The occupancy of each hydrogen bond is 0.8 while the average distances are $< R > = 2.030$ Å, $< R > = 2.105$ Å and $< R > = 2.157$ Å, respectively (see Table 6.4 and Figure C.5). On the other hand, equatorial O$_{2P}$ oxygen makes two strong hydrogen bond interactions with imidazolium side chain of His12 and a water molecule, and one weaker hydrogen bond interaction with the main chain of Phe120. The average distances are $< R > = 1.787$ Å, $< R > = 2.053$ Å and $< R > = 2.324$ Å, respectively, with a hydrogen

bond occupancy of Occ = 1.00 for His12 and the water molecule and Occ = 0.4 for Phe120. It worth mentioning, that in some dianionic windows residue His119 makes a fourth hydrogen bond interaction with this equatorial oxygen. However, in this window the average distance between heavy atoms is 3.422 Å. Interestingly, $O_{2'}$ does not make any hydrogen bond interaction in this window. Although, His12's $H\epsilon_2$ is at a hydrogen bond distance from this atom, however, the interactions are < 5.0 ps.

Furthermore, equatorial oxygen $O_{3'}$ makes a strong hydrogen bond interaction with a water molecule (Occ = 1.00), and somewhat a weaker interaction with the other hydrogen of the same water molecule (Occ = 0.40). The heavy atoms average distance is 2.555 Å.

Finally, $O_{5'}$ makes two strong hydrogen bond interactions, with the imidazolium side chain of residue His119 and with a water molecule ($< R >$ = 1.837 Å and $< R >$ = 1.648 Å, respectively). The occupancy is 1.0 for both of them.

In Figure 6.4, the RDFs between solvent's oxygen atoms and selected atoms of the oxyphosphorane species can be found. The analysis of RDFs can give interesting insight into the solvation characteristics of the oxyphosphorane moiety. Briefly, $O_{5'}$ axial oxygen, O_{2P} and $O_{3'}$ equatorial oxygens present a highly ordered water interaction, high peak followed by a decrease towards a minimum with negligible or very small value of RDF, indicating the presence of a strong interaction of the oxygen atoms with a very structurally ordered water molecule. The $g_{O_{5'}-O_W}$ has a maximum value of 3.2 at 2.5 Å which drops till 0.1 at 2.9 Å, while $g_{O_{3'}-O_W}$ maximum is 2.7 at 2.6 Å and falls till 0.0 at 2.9 Å. $g_{O_{2P}-O_W}$ maximum is 3.1 at 2.9 Å and decreases to 0.0 at 3.5 Å. With respect to axial $O_{2'}$, the RDF presents signatures of a poorly solvated situation that indicate less ordered solvent and more rapid exchange relative to the RDFs for the axial $O_{5'}$.

Monoanionic Structure: window centered at RC = 0.725 Å

Next we analyze the window centered at the reaction coordinate RC = 0.725 Å. The oxyphosphorane structure average values are list in Table 6.1. The reaction coordinate mean values are 1.787(0.033) Å for r_{break}, and 1.046(0.025) Å for r_{form}. Therefore, in this simulation we are studying a monoanionic oxyphosphorane, that has a pentacovalent phosphorus atom with a distorted trigonal-bipyramid geometry.

The average and standard deviation values of the oxyphosphorane's P-O_X distances for this window are 1.733(0.032) Å for P-$O_{2'}$, 1.822(0.036) Å for P-$O_{5'}$, 1.620(0.024) Å for P-O_{2P}, 1.511(0.020) Å for P-O_{1P} and 1.733(0.032) Å for P-$O_{3'}$. Compared to the two uridine-vanadate crystallographic structures,[212,213] all bonds are shorter (see Table 6.1). With respect to the previous dianionic structure, there is a significant shortening in the two axial P-O bond lengths, 0.08 Å, for P-$O_{5'}$ and 0.10 Å, for P-$O_{2'}$, and a lengthening

of the P-O_{2P} bond by 0.09 Å. Therefore, now the two non-bridging phosphoryl oxygens show a difference in bond length of 0.11 Å. Notice that this anisotropy in phosphoryl oxygen bond lengths have been used to identify a monoanionic uridine vanadate in the past.[60] The longest P-O distance still corresponds to P-$O_{5'}$ in the monoanionic phosphorane. Quite interestingly, the angles defining the trigonal-bipyramid structure of the oxyphosphorane show little change from the dianionic to the monoanionic structure. This is specially interesting for the $O_{5'}$-P-$O_{2'}$ angle, which goes from 166.5° in the dianionic structure to 162.9° in the monoanionic oxyphosphorane.

In Table 6.5, we can find selected data for the hydrogen bond formation between the oxyphosphorane and the protein residues, namely, average distances between atoms participating in protein-ligand hydrogen bonds, with the hydrogen bond occupancies, average time and number of events.

O_{1P} equatorial oxygen makes 2 strong hydrogen bond interactions with the amino side chain of residue Lys41, $< R > = 2.908$ Å, and with Gln11, $< R > = 3.128$ Å, which occupancies are 1.0, and somewhat a not that strong hydrogen bond interaction with a water molecule, $Occ = 0.6$, (the same water molecule which was making a hydrogen bond interaction with O_{1P} in the dianionic structure) and a weaker interaction with other water molecule. These water molecules do not interact simultaneously with O_{1P}, first one of them makes the hydrogen bond interaction and after the other one makes it (see Table 6.5 and Figure C.6). On the other hand, the hydrogen bond between His12 and O_{2P} equatorial oxygen is maintained in the monoanionic simulation, although now His12 is neutral and O_{2P} is the donor atom instead of the acceptor. Hϵ_2 is hydrogen bonding His12 during all the simulation, with an occupancy of 1.0 and average distance of 1.787 Å. Interestingly, there is no hydrogen bond interactions between this equatorial atom and the solvent in the monoanionic structure. $O_{2'}$ axial oxygen is also hydrogen bonded by His12's Hϵ_2, $< R > = 2.401$ Å and $Occ = 0.4$. Furthermore, equatorial oxygen $O_{3'}$ makes a weak hydrogen bond interaction with imidazolium side chain of residue His119 ($Occ = 0.4$, $< R > = 2.401$ Å) and a strong interaction with a water molecule ($Occ = 1.0$, $< R > = 1.883$ Å). This water molecule is also hydrogen bonding $O_{5'}$ axial oxygen ($Occ = 0.40$). Finally, $O_{5'}$ makes a strong hydrogen bond interaction with imidazolium side chain of His119 ($Occ = 1.0$, $< R > = 1.892$ Å).

The analysis of the RDFs of Figure 6.4 can shed light on the changes in solvation structure of the oxyphosphorane upon protonation. The RDFs show a less solvated oxyphosphorane compared to the dianionic structure. It presents similar characteristics as the ones commented above for the dianionic case, except for O_{2P} and $O_{5'}$. O_{2P} equatorial oxygen and $O_{5'}$ axial oxygen show a loss of direct solvent interactions upon O_{2P} protonation.

Figure 6.4: Distribution of water around oxyphosphorane's oxygens is shown for two windows of [Trans]^{oph} model, RC = -0.725 Å in black, and RC = 0.725 Å in red. $O_{2'}$ is at top left corner, at top right corner $O_{5'}$ is depicted, O_{1P} and O_{2P} non-bridging phosphoryl atoms are in the middle left and right, respectively, and finally $O_{3'}$ atom is at the bottom left corner.

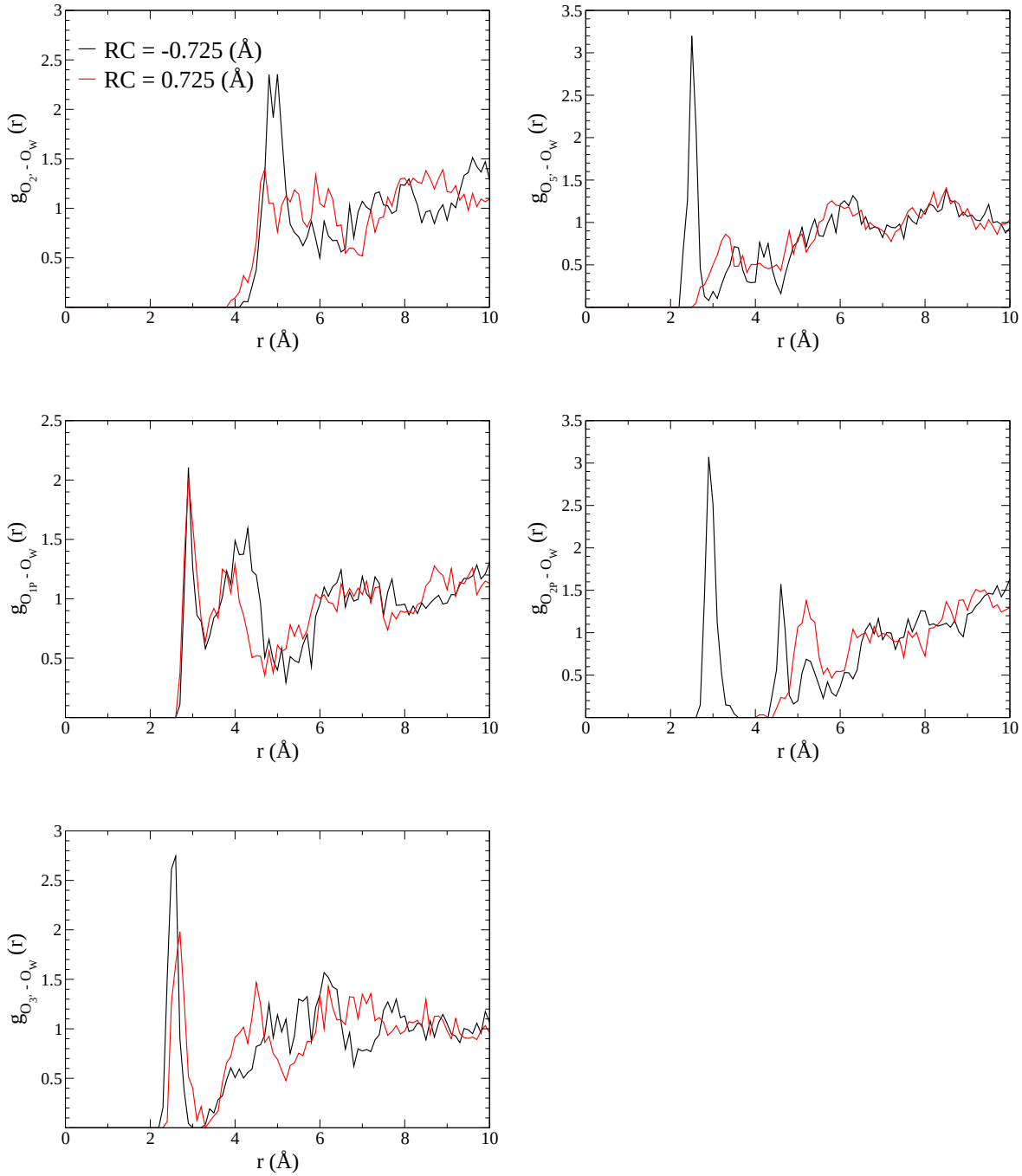


Table 6.5: Statistics of protein-ligand and water-ligand hydrogen-bonds observed in window centered at 0.725 Å of $[\text{Trans}]^{oph}$ model. The criteria for the existence of hydrogen bonds are 1) a maximum H $\cdots$ A(acceptor) distance of 2.4 Å, 2) a minimum $\widehat{DHA}$ of 120° and 3) a hydrogen bond lifetime $\tau \geq 5.0$ ps. The occupancy, Occ, is defined as the total lifetime of each hydrogen bond by the production time. The events are the number of times each hydrogen bond is formed. All distances are in Å. HSP stands for a doubly protonated His.

	Distance		Hydrogen Bond		
Atom pair	MD†	MD‡	Occ	$\langle\tau\rangle$(ps)	# Events
GLN 11 Hϵ_{21} - CYT O_{1P}	2.227(0.245)	3.128(0.181)	1.00	15.00	1.00
LYS 41 Hζ_3 - CYT O_{1P}	2.191(0.434)	2.908(0.128)	1.00	15.00	1.00
WATER 179 - CYT O_{1P}	2.592(0.930)	3.318(0.610)	0.60	9.00	1.00
WATER 4851 - CYT O_{1P}	3.910(1.273)	4.468(1.107)	0.40	6.00	1.00
HSP 12 $H\epsilon_2^{(a)}$ - CYT $O_{2'}$	2.401(0.139)	3.378(0.158)	0.40	6.00	1.00
HSP 119 Hϵ_1 - CYT $O_{3'}$	2.401(0.268)	3.091(0.199)	0.40	6.00	1.00
WATER 5618 - CYT $O_{3'}$	1.883(0.227)	2.721(0.160)	1.00	15.00	1.00
HSP 119 Hδ_1 - CYT $O_{5'}$	1.892(0.155)	2.789(0.095)	1.00	15.00	1.00
WATER 5618 - CYT $O_{5'}$	2.460(0.300)	3.250(0.269)	0.40	6.00	1.00
HSP 12 $H\epsilon_2^{(a)}$ - HSP 12 Nϵ_2	1.787(0.033)	2.807(0.055)	1.00	15.00	1.00

† Arithmetic average hydrogen bond distance for 15.0 ps, with standard deviation in parentheses.

‡ Arithmetic average distance between the corresponding heavy atoms in 15.0 ps, with standard deviation in parentheses.

6.4 Discussion

The present work presents a series of QM/MM molecular dynamic simulations of RNase A - 3',5'-CpA transphosphorylation's and hydrolysis' phosphorane intermediate models, with the aim of comparing the relative energy of a dianionic and a monoanionic oxyphosphorane within the structure of RNase A. These oxyphosphoranes are important intermediates in the hydrolysis reaction of phosphodiester bonds by RNase A. Two models have been taken into account, one $[\text{Hyd}]^{oph}$, related to the intermediate of the hydrolysis step of the reaction, and $[\text{Trans}]^{oph}$ corresponding to the transphosphorylation's intermediate. The monoanionic oxyphosphoranes are formed by a proton transfer from His12, which is hydrogen bonded to O_{2P} with high stability.

For both oxyphosphoranes, the monoanionic structure is lower in energy than the dianionic intermediate by -3.2 kcal/mol and -4.1 kcal/mol for $[\text{Hyd}]^{oph}$ and $[\text{Trans}]^{oph}$, respectively. We will analyze these numbers taking into account the value of the pK_a^2 for oxyphosphoranes and the value of the pK_a for a histidine. The experimental pK_a of a histidine is 6.04.[250]

There are difficulties to estimate experimental pK_as of oxyphosphoranes due to their inherent instability. However, theoretical estimations[248] points to values of 7.9 and 14.3 for pK_a^1 and pK_a^2, respectively. This implies that in solution the reaction energy for proton transfer between an oxyphosphorane and a histidine can be estimated as -11.4 kcal/mol.

$$\begin{aligned} His^+ + [Phos]^{-2} &\rightarrow His + [Phos]^{-1} \\ \Delta G &= \Big(pK_a(His) - pK_a^2(Phos)\Big) \times 2.303RT \\ &= -11.4\ kcal/mol \end{aligned}$$

This reaction energy for proton transfer is significantly lower than the values obtained from our PMF in RNase A, namely, -3.2 kcal/mol ([Hyd]^{oph}) and -4.1 kcal/mol ([Trans]^{oph}). This indicates that compared to solution values, RNase A is introducing a significant stabilization of the dianionic [Hyd]^{oph} and [Trans]^{oph} with respect to the monoanionic structures. The stabilization is higher for [Hyd]^{oph} than for [Trans]^{oph} , by around 1 kcal/mol. The stabilization of dianionic oxyphosphorane structures has also been pointed out by Elsässer *et al.*[61] in the context of the hydrolysis reaction in RNase A. However, we find than an extra stabilization of these intermediates can be attained by protonation of equatorial O_{2P} oxygen, and this protonation may be involved in the corresponding reaction mechanism. However, protonation of O_{2P} could also lead to higher barriers in the breaking of the oxyphosphoranes, and therefore, our present calculations can not assure whether the reaction goes through a monoanionic type mechanism. Glennon *et al.*[251] found in the transphosphorylation step an almost equal probability for both protonation states with a slightly stronger support for the mechanism with a dianionic structure.

Our QM/MM simulations can also give relevant results with respect to the structural changes with respect to the protonation state of the oxyphosphoranes. This type of data have been considered in the past[60] as guidance to infer the protonation state for a given X-Ray structure. Our results, albeit corresponding to short simulations, can shed light on specific structural parameters affected by the protonation of the oxyphosphorane. We consider three types of structural changes: i) the ones caused in the internal coordinates of the oxyphosphorane itself, ii) the ones caused in the interaction between the oxyphosphoranes and the protein and iii) changes in the solvation shell around the oxyphosphoranes.

The largest changes in P-O bond distances upon protonation corresponds to the equatorial P-O_{2P} and the two axial P-$O_{3T/5'}$ and P-$O_{2'}$ distances. As expected, the protonation of O_{2P} oxygen leads to a lengthening of the P-O bond by almost 0.1 Å. This induces an asymmetry of 0.1 Å ([Hyd]^{oph}) and 0.11 Å ([Trans]^{oph}) in the P-O bonds with non-bridging phosphoryl oxygens. This asymmetry has been taken as signature in the past for monoanionic vanadates.[60] The axial P-O bond distances are also highly affected by the

protonation state of the oxyphosphorane. In both oxyphosphorane structures, there is a significant shortening of axial P-O bond distances in the monoanionic phosphorane, being the tendency more enhanced in $[Trans]^{oph}$ than in $[Hyd]^{oph}$. Irrespective of the protonation state, the largest P-O distance corresponds to P-$O_{2'}$ in $[Hyd]^{oph}$, but to P-$O_{5'}$ in $[Trans]^{oph}$. Thus, based on these structural data the P-$O_{2'}$ bond is more activated in the $[Hyd]^{oph}$ model than in $[Trans]^{oph}$, which shows a higher activation for P-$O_{5'}$. Other geometrical parameters show only small changes upon protonation of the oxyphosphorane. One should emphasize the small change in the departure of the axial $O_{3T/5'}$-P-$O_{2'}$ angle from linearity, which can be taken as a representative of the degree of distortion of the oxyphosphorane structure from an idealized trigonal bipyramid structure. The changes in this angle upon protonation is only of 2-3°, which fall within the fluctuation value of this angle in each dynamics. It should be emphasized that neutral oxyphosphorane structures show axial O-P-O angles close to 180°, indicating the close relation between distortion of the trigonal bipyramid and accumulation of negative charge in the oxyphosphorane. Range *et al.*[252] found a neutral oxyphosphoranes axial angle range between 175.2° and 176.6°, and a range between 160.4° and 168.4° for the monoanionic oxyphosphoranes, similar to the one found by Lopez *et al.*,[253] 160° and 162°. These last co-workers also report an angle $\sim$ 163° for dianionic oxyphosphoranes. Our results are in good agreement with the cited ones.

The hydrogen bond pattern between the oxyphosphorane substrate and residues of the active site is sensible to the protonation state of the phosphoranes, specially in the case of $[Hyd]^{oph}$. We have found that the position of Lys41 is sensible to the protonation state of the oxyphosphorane. In the case of $[Hyd]^{oph}$, protonation at O_{2P} leads to a stronger interaction between the amino group of Lys41 and the $O_{2'}$ leaving group, and a weaker interaction between this amino group and O_{1P} equatorial oxygen. However in $[Trans]^{oph}$ the position of Lys41 shows little changes upon formation of the monoanionic phosphorane. In addition, imidazolium side chain of His119 is directly hydrogen bonding O_{3T} axial oxygen in the dianionic $[Hyd]^{oph}$, but it makes a water bridge interaction in the monoanionic one. Again little changes are observed in the case of $[Trans]^{oph}$, the interaction between His119 and $O_{5'}$ is conserved in both dianionic and monoanionic oxyphosphoranes. Besides, the imidazolium side chain of residue His119 is observed in its active, A, conformation[62] for all oxyphosphoranes in any protonation state. However, the position of His12 in $[Hyd]^{oph}$ and $[Trans]^{oph}$ is almost unaffected by the protonation of O_{2P} equatorial oxygen, the only change is that Nδ_1 and O_{2P} will switch the proton donor/acceptor character upon protonation. Other residues, like Gln11 in $[Hyd]^{oph}$ and in $[Trans]^{oph}$ are less affected by the protonation state of the oxyphosphoranes, whereas Phe120 in $[Hyd]^{oph}$ and in $[Trans]^{oph}$ shows sensible loss of interactions with the substrate when going from the dianionic to monoanionic oxyphosphoranes. The end result is

that upon formation of a monoanionic phosphorane, there are on one hand, less direct interactions between the substrate and the residues of the active site (O_{2P} and Phe120, $O_{3'}$ and His119, and O_{3T} and His119 in $[Hyd]^{oph}$, and O_{2P} and Phe120 in $[Trans]^{oph}$), and others show a change in the hydrogen bonding pattern, like the strengthening of Lys41 and $O_{2'}$ interaction in $[Hyd]^{oph}$.

The pass from dianionic to monoanionic oxyphosphorane has a sizable effect on its solvation. However its effect can vary in $[Hyd]^{oph}$ and $[Trans]^{oph}$ structures. In $[Trans]^{oph}$, the formation of a monoanionic phosphorane leads to loss of waters at the first shell of oxyphosphorane oxygens. This is specially dramatic for O_{2P}, for which the first peak is completely lost when passing from dianionic to monoanionic structure. In $[Hyd]^{oph}$, however, the situation is more mixed, with some oxygens as O_{2P} losing again its direct interaction with a water molecule upon protonation, but with other oxygens, such as equatorial $O_{3'}$ and O_{1P}, enhancing the interaction with solvent. In our opinion, the main result from all the RDF's analysis is the loss of direct water interaction with O_{2P} upon its protonation. Our recent MM dynamics[243] showed that dianionic structures show a very highly ordered water molecule hydrogen bonded to O_{2P}, which has also been detected in X-Ray structures.[50,60,213] In a recent publication,[243] we assumed that the role of this water molecule was to stabilize the salt bridge with the positively charged His12, as observed to be the case of phosphates interactions with imidazole.[239] The present calculations confirm this hypothesis and nicely shown how upon protonation at O_{2P} and disappearance of the salt bridge, the strength of the interaction with this water molecule is lost. In other words, the presence of a highly order water molecule at the O_{2P} position seems to be a clear signature for a dianionic oxyphosphorane structure in RNase A.

6.5 Conclusions

In the present paper, we present QM/MM umbrella sampling simulations with the aim to study the relative energy of dianionic and monoanionic oxyphosphoranes that are intermediates in the transphosphorylation and hydrolysis reactions of RNA by RNase A. The formation of the monoanionic oxyphosphorane is done from a proton transfer from His12 to O_{2P}.

For both $[Hyd]^{oph}$ and $[Trans]^{oph}$ models we have found that the formation of a monoanionic phosphorane from a proton transfer from His12 leads to a lower energy structure than the formation of a salt bridge between a dianionic phosphorane and a positively charged histidine. The relative free energies are -4.1 kcal/mol and -3.2 kcal/mol for $[Trans]^{oph}$ and $[Hyd]^{oph}$ models. These values are lower in absolute value than the estimated values in solution based on pK_a's of histidines and oxyphosphoranes, which indicates

a significant stabilization of RNase A of the dianionic oxyphosphorane structure. Our calculations demonstrate that further stabilization of the structure can be done by a proton transfer from His12.

Structural data of the oxyphosphorane and the hydrogen bond pattern of the substrate and residues and water molecules of the active site are also sensible to the protonation state of the oxyphosphorane. In this sense, we have found that the protonation of O_{2P} leads to an asymmetry in the P-O bonds with non-bridging phosphoryl oxygens and to a shrinking of P-O axial distances. Other internal parameters, like the axial O-P-O angle, are less affected.

We have also found that interaction of $[Hyd]^{oph}$ with the residues of the active site is specially sensible to the protonation state of the oxyphosphorane. For instance, the interaction of Lys41 with $O_{2'}$ is significantly reinforced upon protonation at O_{2P}. Finally, our calculations confirm previous theoretical findings of the role of a specific water molecule and O_{2P} interaction in the stabilization of a salt bridge between dianionic phosphoranes and positively charged histidines, finding a dramatic loss of this interaction upon disappearance of the salt bridge due to monoanionic oxyphosphorane formation. In this sense, the presence of a highly ordered water molecule at the O_{2P} position can be taken as a signature for the presence of a dianionic oxyphosphorane intermediate.

Chapter 7

Potential Energy Surfaces of RNase A - CpA Complex

7.1 Introduction

The *potential energy surfaces* (see Section 3.3.4) of polyatomic molecules are very hard to visualize, since they involve a large number of dimensions. Usually, slices through PES's that involve only a single or two coordinates are taken, which show the relevant reduced-dimensionality energy curves or surfaces.

Different PES's of RNase A - 3',5'-CpA hydrolysis reaction have been studied analyzing different reaction coordinates. Several adiabatic maps have been generated with the aim to have a better understanding of the steps of the reaction. An adiabatic map is a simple way to examine molecular motion by characteristic low-energy paths along a prescribed reaction coordinate. For each combination of these conformational coordinates, the entire potential energy of the system is minimized to approximate behavior for the motion under study, the reaction coordinate is fixed while all other degrees of freedom are fully minimized.

Here only some of the studied PES are going to be exposed.

7.2 Computational Details

Stochastic Boundary Conditions without electrostatic cutoff are applied in these QM/MM calculations (see Section 3.3.6). The system is partitioned into three zones. The reaction zone from 0 Å to 21 Å was surrounded by a buffer region of 4 Å, which used Langevin dynamics (β=62.0 ps^{-1} for water and β=250.0 ps^{-1} for the enzyme) to serve as a constant temperature bath. Besides, a deformable boundary potential was used to reproduce the containment effect of bulk solvent outside the buffer zone.[174]

The starting structure is taken from our previous calculations on MD simulations of RNase A - CpA transition-state like complexes[243] (see Chapter 4 and page 126). The protocol includes an initial equilibration of the configuration produced by MD simulations, followed by a short run where only the MM part is free to move, while the QM part is kept frozen. Water molecules were initially relaxed for 300 steps of steepest descents[158,160] energy minimization keeping all solute atoms and ion positions restrained to their initial coordinates (see Section 3.3.4). The restraints on the ions were then released, and the solvent (water and ions) were relaxed with 300 steps SD energy minimization. Then the restraints on the solute except for the QM region were released and another relaxation of 300 steps of SD energy minimization was done. Once the MM region is minimized another set of minimizations were done with a RESD[183] restrained on the combination of the axial distances of the oxyphosphorane, to maintain its structure. First a 500 step SD minimization was done, followed by two more minimizations of 500 steps using the adopted-basis set Newton-Raphson (ABNR) method.[126,127,161] From this starting point, restrained MD with the QM part fix, was performed starting at 0 K and heating to 298 K over 18 ps, and carried out to 125 ps. Then, the whole system is allowed to move for 40 ps. Finally, a set of SD and ABNR minimizations were performed until the tolerance average gradient was less than 0.0001 kcal/mol·Å^{-1}.

Once we got the initial structure, we performed different RESD minimizations to get the adiabatic maps. Each of the structures is minimized first by 100 steps SD minimization and a tolerance average gradient of 0.1 kcal/mol·Å^{-1}, continued by two ABNR minimizations of 10000 steps each and tolerance average gradient of 0.01 kcal/mol·Å^{-1} and 0.001 kcal/mol·Å^{-1}, respectively.

On the other hand, after the PES are constructed and the lower-energy path selected, some DFT single point energy calculations were performed on those structures.

DFT Single Point Calculations

DFT methodology[114–117] with two different hybrid functionals, the gradient corrected hybrid B3LYP [254–256] and the hybrid meta GGA MPWB1K,[257] with Pople's 6-311++G(2df,2p) triple-zeta basis set are applied to calculate single point energies of the previously defined low-energy path structures. All the energy calculations were carried out with the GAUSSIAN03 package.[258] As mention in the General Introduction (see Section 3.1), it is impossible to make a DFT calculation of such big system ($>$ 7000 atoms), therefore two different options have been used:

QM(1)^{DFT}: The previously mentioned QM part (44 atoms for [Trans]^{oph} and 41 atoms for [Hyd]^{oph} systems, see Figure 6.1) is taken into ac-

count for the DFT calculations. Using the link atom method (see Section 3.4.1) to saturate the valency of the QM part, our new systems have a total of 62 atoms for the transphosphorylation step and 56 for the hydrolysis step. Once the energy of those systems is calculated, the environment effect is added. Where the environment effect is defined as the difference between the total energy of the whole system (> 7000 atoms) and the energy of the QM part at AM1/d-PhoT//MM level, with MM charges set to zero.

QM(2)^{DFT}: We consider that the residues in a radius ≤ 4.5 Å from the phosphorus atom are the most influenced atoms in the reaction. Hence, only the previously mentioned QM part (44 atoms for [Trans]^{oph} and 41 atoms for [Hyd]^{oph} systems) and the residues at a distance ≤ 4.5 Å from the phosphorus atom are taken into account in these calculations, which makes up a new system of 154 QM atoms for the transphosphorylation step and 131 QM atoms for the hydrolysis step. The atoms outside this sphere of 4.5 Å are not taken into account.

7.3 Transphosphorylation step

7.3.1 Classical Mechanism

The classical mechanism[73,76] suggests a concerted acid-base catalysis, where His12 acts as a base and His119 acts as an acid in the transphosphorylation step (see Section 2.3.2), therefore, we studied the reaction PES involving those proton transfers as reaction coordinates (see Figure 7.1).

$$\begin{aligned} R_a &= (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'}) \\ R_b &= (N\delta_1^{His119} - H\delta_1^{His119}) - (H\delta_1^{His119} - O_{5'}) \end{aligned}$$

where the reaction coordinates range between

$$\begin{aligned} -1.1 \text{ Å} \leq \quad R_a \quad \leq 2.0 \text{ Å} \\ -1.8 \text{ Å} \leq \quad R_b \quad \leq 2.0 \text{ Å} \end{aligned}$$

Imidazole side chain of His12 abstracts the 2'-hydroxyl proton and His119 protonates one of the oxyphosphorane's axial oxygens, $O_{5'}$.

The reactant (3',5'-CpA, neutral His12 and cationic His119) and the product (C >p and cationic His12 and His119) states have the minimum energy values in this PES, see Figure 7.1 and Table 7.1. The two states are separated by an energy barrier $\sim$ 38.0 kcal/mol. A similar value is obtained at QM(1)^{DFT} level of theory, 37.7 - 40.1 kcal/mol, however, it decreases till 16.1 - 17.8 kcal/mol when QM(2)^{DFT} is applied. Although the

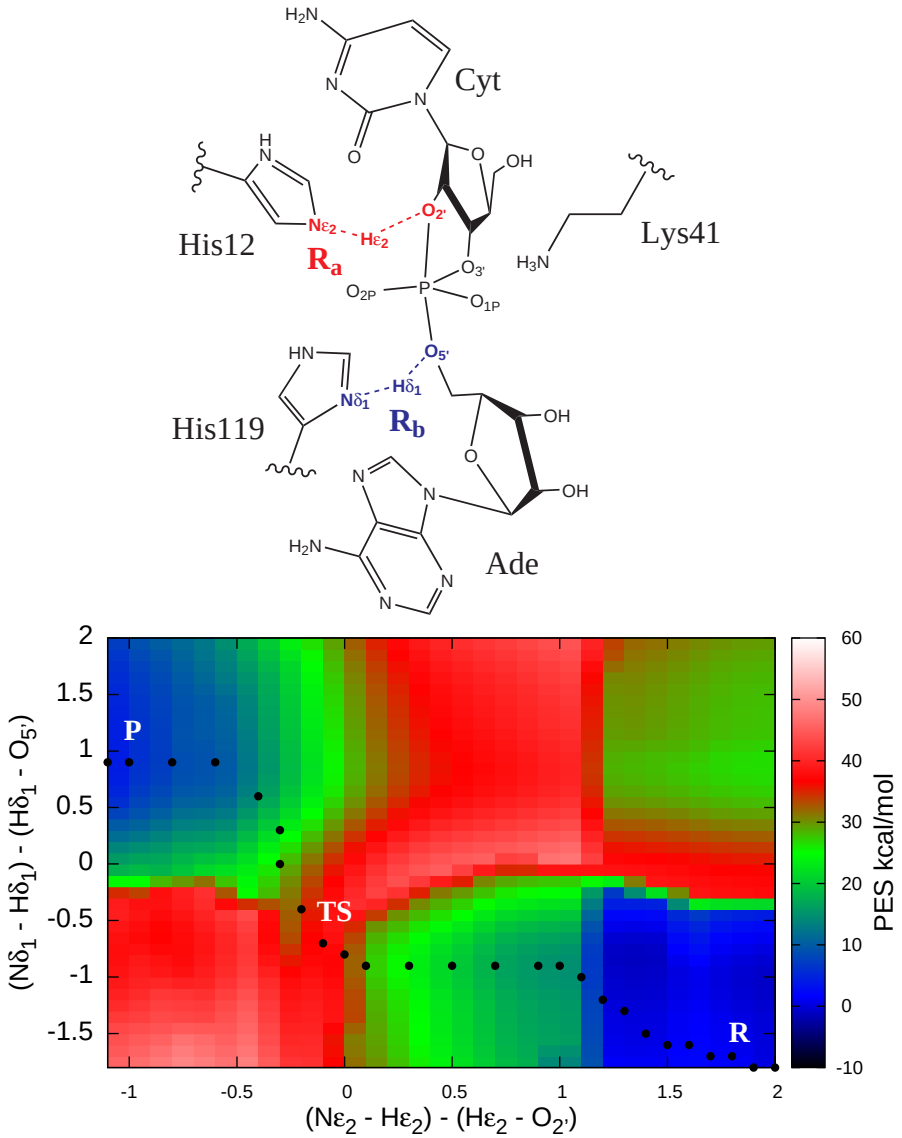


Figure 7.1: The top Figure is a scheme of the reaction coordinates. $R_a = (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'})$ reaction coordinate represents a proton transfer from His12 to $O_{2'}$ oxyphosphorane's axial oxygen, and $R_b = (N\delta_1^{His119} - H\delta_1^{His119}) - (H\delta_1^{His119} - O_{5'})$ represents a proton transfer from His119 to $O_{5'}$ oxyphosphorane's axial oxygen. The Figure at the bottom is the Potential Energy Surface of the cited reaction coordinates. The black circles represent the proposed reaction path, for which single point energy calculations have been done at DFT level of theory. R, P and TS denote the reactant, product and transphosphorylation transition state, respectively. The distances are in Å and the energies in kcal/mol.

reaction path suggests an one step mechanism, it is a reflection of the chosen coordinates (we can not control the axial bond distances with them). It is known that it is a two step mechanism, since there is a stable oxyphosphorane

intermediate in the reaction[28,61,82,88,253] (see Part II). The first step is a concerted nucleophilic attack (formation of P-$O_{2'}$ bond) and proton transfer from $O_{2'}$ to His12, and the second step is the cleavage and protonation of the leaving group (see Figure D.1).

The energy barrier corresponds to a dianionic oxyphosphorane structure (both axial bonds are formed, P-$O_{2'}$=1.86 Å and P-$O_{5'}$=1.70 Å), where O_{1P} equatorial oxygen is hydrogen bonded by the side chain of Lys41, and both axial oxygens, $O_{2'}$ and $O_{5'}$, are hydrogen bonded by His12 and His119, respectively (see Figure D.1). It is important to remark, that the hydrogen bond formed in the reactant state between His119 and O_{2P} equatorial oxygen, disappears as the reaction proceeds. The energies obtained for the product structure are also sensible to the type of QM(#)^{DFT} used, both functionals give similar results.

Table 7.1: Single point energies in kcal/mol for the TS and product states of the reaction coordinates $R_a = (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'})$ and $R_b = (N\delta_1^{His119} - H\delta_1^{His119}) - (H\delta_1^{His119} - O_{5'})$. The 6-311++G** basis set is applied for DFT calculations.

		QM(1)^{DFT}		QM(2)^{DFT}	
	AM1/d-PhoT//MM	B3LYP	MPWB1K	B3LYP	MPWB1K
TS	38.0	40.1	37.7	16.1	17.8
Product	2.6	10.9	11.1	-2.1	-2.9

7.3.2 Alternative mechanism: Lys41 as base catalyst

On the other hand, Wladkowski and co-workers[60,85–87] (see Section 2.3.2) proposed a mechanism where Lys41 acts as the catalytic base in the transphosphorylation step of the reaction, instead of His12. Therefore, we decided to analyze the energy difference between the 2'-hydroxyl hydrogen abstraction by His12 or by Lys41 in the reactant structure (3',5'-CpA), see Figure 7.2. The reaction coordinates are

$$\begin{aligned} R_1 &= (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'}) \\ R_2 &= (N\zeta^{Lys41} - H\zeta^{Lys41}) - (H\zeta^{Lys41} - O_{2'}) \end{aligned}$$

where

$$\begin{aligned} 1.4 \text{ Å} &\geq R_1 \geq -1.2 \text{ Å} \\ -1.9 \text{ Å} &\leq R_2 \leq 1.2 \text{ Å} \end{aligned}$$

First a set of RESD minimizations over R_1 reaction coordinate is carried out. It is important to remark that as the 2'-hydroxyl proton is transferred to

His12, the oxyphosphorane gets formed, P-$O_{2'}$ axial distance decreases until the bond is formed (see Figure D.2). Then, once the 2'-hydroxyl oxygen is deprotonated by His12 a new set of RESD minimizations over R_2 is carried out to transfer $H\zeta^{Lys41}$ to $O_{2'}$ oxygen. Now, as the proton is being transferred from Lys41 to $O_{2'}$ oxygen, the oxyphosphorane gets broken by the cleavage of the P-$O_{2'}$ bond.

The contour of these RESD minimizations is shown in Figure 7.2. The reactant dinucleotide (monoanionic) is energetically more stable when His12 is neutral and Lys41 is cationic (left hand side), rather than when Lys41 is neutral and His12 is cationic (right hand side). The energy difference between both structures is 15.7 kcal/mol, which decreases tilt $\sim$ 10 kcal/mol when the energy is recalculated applying QM(1)^{DFT} level of theory, and until $\sim$ 9 kcal/mol when QM(2)^{DFT} is used (see Table 7.2). This is in good agreement with the fact that at neutral pH, Lys41 is more likely to be protonated than unprotonated. If we take the literature standard values of histidine (pK_a 6.04) and lysine (pK_a 10.53),[234] the energetic cost of the deprotonation of Lys41 by His12 is $\sim$ 6.24 kcal/mol.

Nevertheless, the energy barrier of the deprotonation of 2'-hydroxyl oxygen is smaller in R_2 (Lys41 acts as the base) than in R_1 (His12 acts as the base). The energy difference in R_2 ranges from $\sim$ 7 kcal/mol with MPWB1K functional to 11.4 kcal/mol when the semiempirical methodology is applied, see Table 7.2. Taking into account the energy penalty to pay for deprotonation of Lys41 ($\sim$ 16 kcal/mol) and the energy barrier we get for 2'-hydroxyl proton abstraction by Lys41 ($\sim$ 11 kcal/mol), the overall process is competitive with His12 acting as the base catalyst (see Table 7.2), i.e., the overall energy barriers are similar in both reaction coordinates, R_1 and R_2. Therefore, we can not rule out that Lys41 acts as the base catalyst.

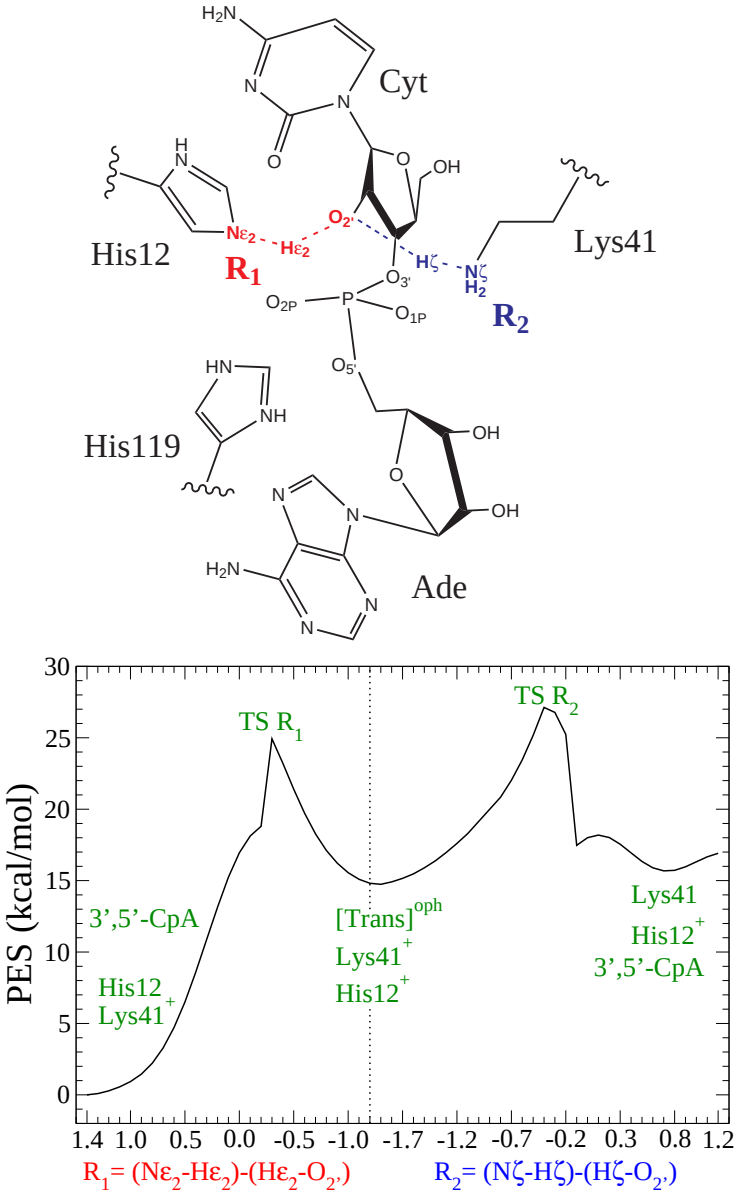
Table 7.2: Single point energies in kcal/mol for RNase A - 3',5'-CpA complex, neutral Lys41 and cationic His12 structure; the [Trans]^{oph} , cationic Lys41 and cationic His12 structure; and for the transition states, in the reaction coordinates $R_1 = (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'})$ and $R_2 = (N\zeta^{Lys41} - H\zeta^{Lys41}) - (H\zeta^{Lys41} - O_{2'})$. The 6-311++G** basis set is applied for DFT calculations.

		QM(1)^{DFT}		QM(2)^{DFT}	
	AM1/d-PhoT//MM	B3LYP	MPWB1K	B3LYP	MPWB1K
TS R_1	24.9	11.0	12.1	16.9	13.7
[Trans]^{oph}	14.8	5.1	2.7	9.9	6.4
TS R_2	27.1	20.7	17.7	18.8	15.4
§	15.7	10.7	10.2	8.7	8.6

§ RNase A - 3',5'-CpA complex, neutral Lys41 and cationic His12

TS R_1 corresponds to a dianionic oxyphosphorane structure with three

Figure 7.2: The top figure is a scheme of the reaction coordinates $R_1 = (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'})$ which represents 2'-hydroxyl hydrogen abstraction by His12, and $R_2 = (N\zeta^{Lys41} - H\zeta^{Lys41}) - (H\zeta^{Lys41} - O_{2'})$ that represents a proton transfer from Lys41 to $O_{2'}$ oxygen. The figure at the bottom is the Potential Energy Surface of the cited reaction coordinates. Distances are in Å and energy in kcal/mol.



hydrogen bond interactions: one between $H\epsilon_2^{His12}$ and $O_{2'}$ axial oxygen, a second one between O_{1P} equatorial oxygen and residue Lys41, and a last one between His119 and $O_{5'}$ axial oxygen. The energy increment found

at that point corresponds to the decrease of P-$O_{2'}$ distance from 2.5 Å to 1.8 Å (see Figures 7.2 and D.2). Similarly, TS R_2 point corresponds to a dianionic oxyphosphorane structure with three hydrogen bond interactions: one between Lys41 and $O_{2'}$ axial oxygen, a second one between $H\epsilon_2^{His12}$ and O_{2P} equatorial oxygen, and the last one between $H\delta_1^{His119}$ and $O_{5'}$ axial oxygen. The energy drop found in this point is related to the P-$O_{2'}$ bond cleavage. It is interesting to remark that when R_1 is analyzed, $H\zeta^{Lys41}$ is at a hydrogen bond distance from O_{1P} non-bridging equatorial phosphoryl oxygen, while when R_2 is analyzed $H\epsilon_2^{His12}$ is at a hydrogen bond distance from the other non-bridging equatorial phosphoryl oxygen, O_{2P}. This suggests, that those two residues can stabilize the excess of negative charge at the oxyphosphorane non-bridging positions. In addition, $H\delta_1^{His119}$ is at a hydrogen bond distance from $O_{5'}$ oxygen in both reaction coordinates, in a correct position to act as the acid catalyst. Nevertheless, it is not at a hydrogen bond distance from O_{2P} oxygen, as proposed by Haydock[80,81] and Breslow[79] (see Section 2.3.2).

7.3.3 Other reaction mechanisms

Haydock and Breslow have proposed (see Section 2.3.2) that there is an activation of the phosphate previous to the nucleophilic attack by either protonating one of the equatorial phosphoryl oxygens, or by specific hydrogen bonding from His119 to one of the phosphoryl oxygens. In order to analyze these possibilities, we investigated another two PES's, corresponding to, *i)* a proton transfer from the cationic His12 to O_{2P} equatorial oxygen as the nucleophilic attack of the unprotonated $O_{2'}$ axial oxygen occurs (see Figure 7.3), and *ii)* the 2'-hydroxyl proton abstraction by His12 as His119 donates $H\delta_1$ proton to O_{2P} equatorial oxygen (see Figure 7.4).

In case *i)* we do not see the proton transfer of $H\epsilon_2^{His12}$ to O_{2P} oxygen previous to the oxyphosphorane formation proposed by Haydock (Figure D.3). That mechanism is energetically more unfavorable, energy barrier $\sim$ 26 kcal/mol, than the concerted mechanism, energy barrier $\sim$ 6 kcal/mol (see Figure 7.3). Moreover, during the nucleophilic attack, both studied PES's do not show the hydrogen bond interaction between $H\delta_1^{His119}$ and O_{2P} oxygen proposed in the mechanisms of Haydock[80,81] and Breslow[79] (Figures D.3 and D.4). Case *ii)* presents an energy barrier $\sim$ 25 kcal/mol for the nucleophilic attack without the hydrogen bond interaction of $H\delta_1^{His119}$ and O_{2P}, and a barrier of $\sim$ 36 kcal/mol when that hydrogen bond exists, see Figure 7.4.

On the other hand, it is important to remark, that Lys41 is at a hydrogen bond distance from O_{1P} equatorial oxygen when the oxyphosphorane is formed in both PES's, *i)* $R_b = (P - O_{2'}) - (P - O_{5'}) \leq 0.3$ Å or *ii)* $R_a = (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'}) \leq 0.1$ Å (Figures D.3 and D.4). Besides, in the case *ii)* $H\delta_1^{His119}$ is at hydrogen bond distance

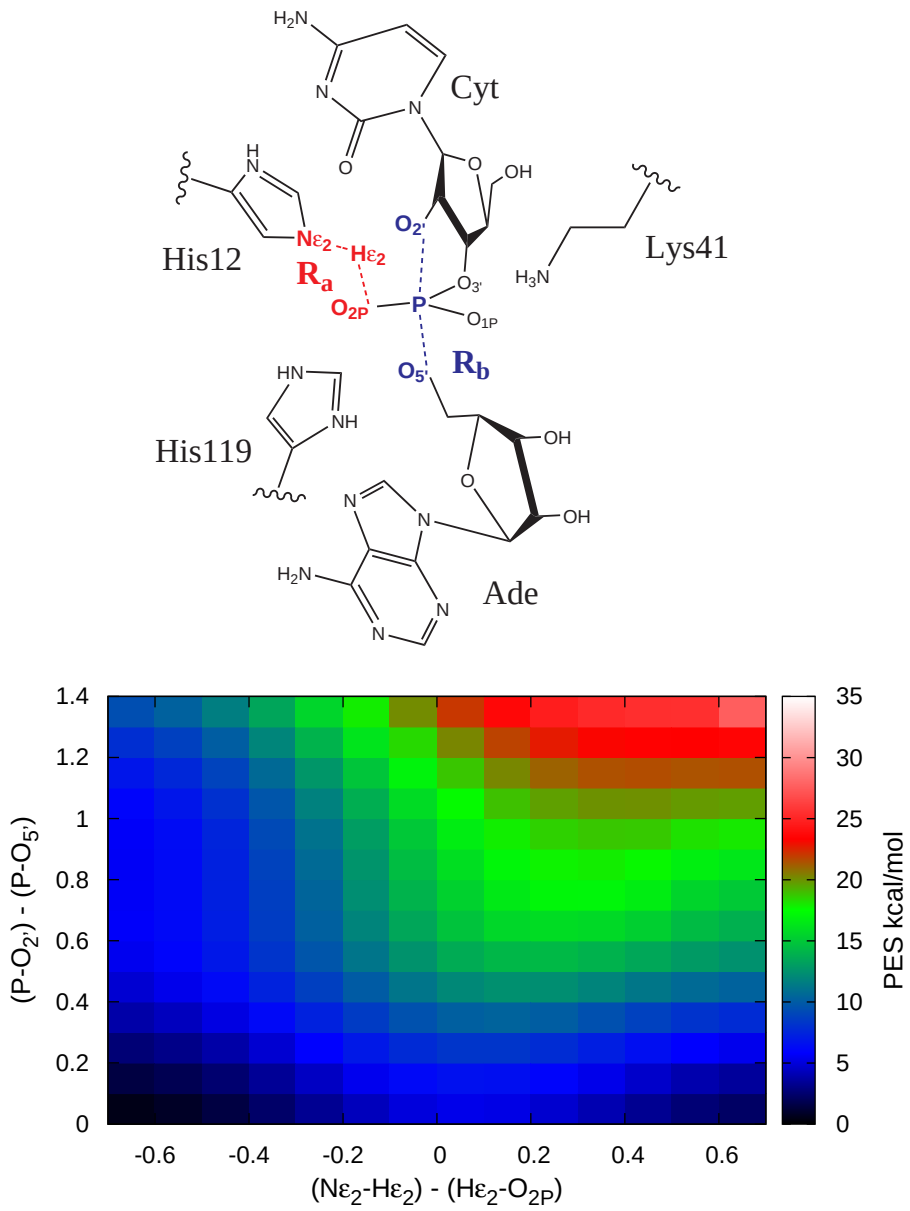


Figure 7.3: The top figure is a scheme of the applied reaction coordinates. $R_a = (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2P})$ reaction coordinate represents a proton transfer from His12 to O_{2P} phosphoryl oxygen, while $R_b = (P - O_{2'}) - (P - O_{5'})$ represents the formation/cleavage of the oxyphosphorane. The figure at the bottom is the Potential Energy Surface of the combination of cited reaction coordinates. The distances are in Å and the energies in kcal/mol.

from $O_{5'}$ axial oxygen, when the oxyphosphorane is formed and His119 is cationic, $R_a = (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'}) \leq 0.1$ Å and $R_b = (N\delta_1^{His119} - H\delta_1^{His119}) - (H\delta_1^{His119} - O_{2P}) \leq$ -1.0 Å.

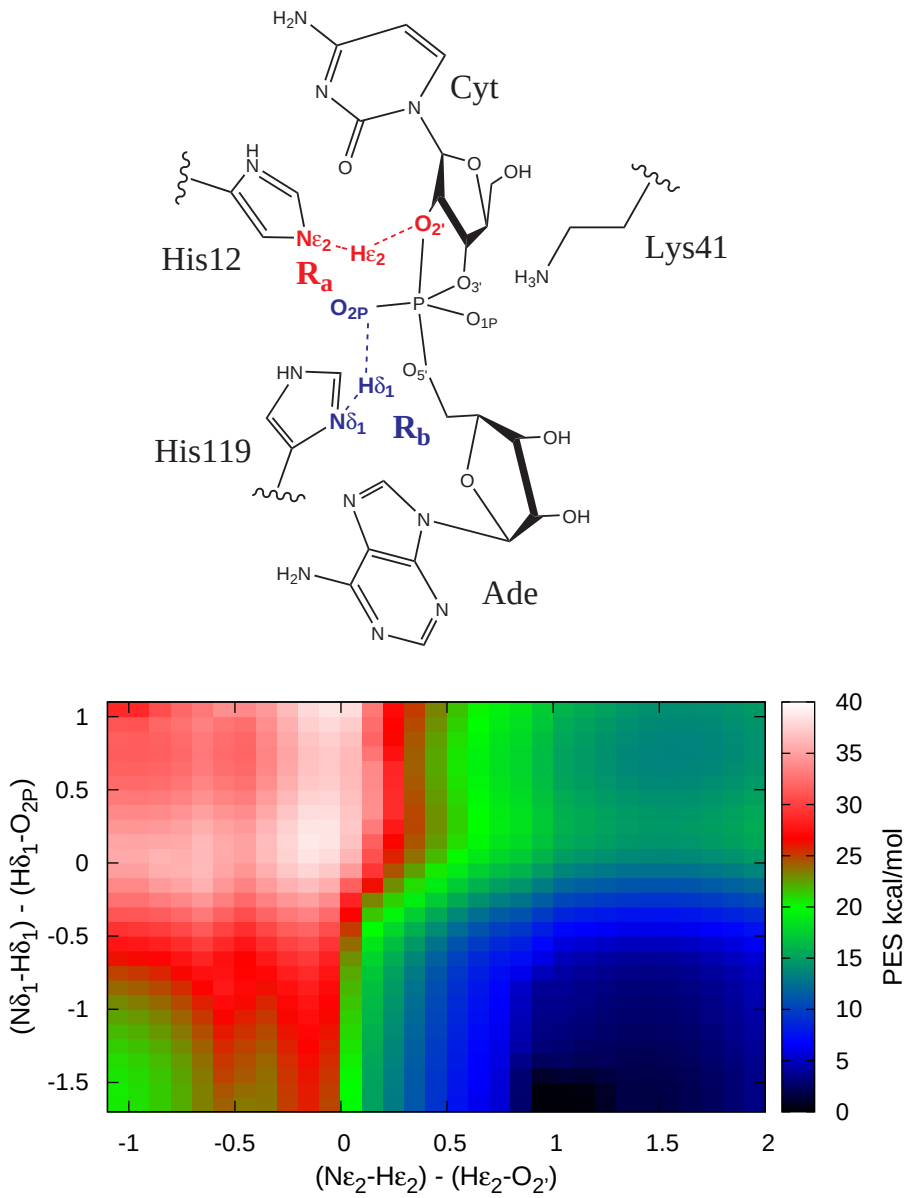


Figure 7.4: The top figure is a scheme of the applied reaction coordinates. $R_a = (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'})$ reaction coordinate represents a proton transfer from His12 to $O_{2'}$ oxyphosphorane's axial oxygen, while $R_b = (N\delta_1^{His119} - H\delta_1^{His119}) - (H\delta_1^{His119} - O_{2P})$ represents the proton transfer from His119 to O_{2P} phosphoryl oxygen. The bottom figure is the Potential Energy Surface of the cited reaction coordinate. The distances are in Å and the energies in kcal/mol.

7.4 Hydrolysis step

7.4.1 Classical Mechanism

The classical mechanism[73,76] suggests a concerted acid-base catalysis, where His12 acts as an acid and His119 acts as a base in the hydrolysis step (see

Section 2.3.2), therefore, we studied the reaction PES involving those proton transfers as reaction coordinates (see Figure 7.5).

$$\begin{aligned} R_a &= (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'}) \\ R_b &= (N\delta_1^{His119} - H\delta_1^{His119}) - (H\delta_1^{His119} - O_{3T}) \end{aligned}$$

where the reaction coordinates range between

$$\begin{aligned} -1.0 \text{ Å} \leq \quad R_a \quad &\leq 2.0 \text{ Å} \\ -0.8 \text{ Å} \leq \quad R_b \quad &\leq 1.5 \text{ Å} \end{aligned}$$

Imidazole side chain of His119 abstracts a proton from the nucleophile's oxygen (O_{3T}) and imidazolium side chain of His12 protonates $O_{2'}$ axial oxygen.

The reactant (C >p, cationic His12 and neutral His119) and the product (3'-Cp, neutral His12 and cationic His119) states have the minimum energy values in this PES, see Figure 7.5 and Table 7.3. The two states are separated by an energy barrier $\sim$ 23 kcal/mol. A similar value is obtained at QM(1)^{DFT} level of theory, however, it decreases till 15.4 kcal/mol when QM(2)^{DFT} with MPWB1K functional is applied. Those values are above and below the observable value of 19 kcal/mol.[259] The energies obtained for the product structure are also sensible to the QM(#)^{DFT} and functional used, but all are in good agreement with observations, -4 kcal/mol.[259] Although the reaction path suggests an one step mechanism, it is a reflection of the chosen coordinates (we can not control the axial bond distances with them). It is known that it is a two step mechanism, since there is a stable oxyphosphorane intermediate in the reaction[28,61,82,88,253] (see Part II). The first step is a concerted nucleophilic attack (formation of P-O_{3T} bond) and proton transfer from the nucleophile's oxygen O_{3T} to His119, and the second step is the cleavage of P-$O_{2'}$ bond and protonation of $O_{2'}$ by His12.

The energy barrier corresponds to a dianionic oxyphosphorane structure (P-$O_{2'}$=1.73 Å and P-O_{3T}=1.89 Å), where the side chain of Lys41 is hydrogen bonding O_{1P} equatorial oxygen, 1.8 Å (see Figure D.5). Besides, imidazolium side chain of His12 is hydrogen bonding $O_{2'}$ axial oxygen (1.4 Å) and His119 is abstracting the proton from O_{3T} axial oxygen (1.5 Å). Furthermore, we do not see any proton transfer or hydrogen bond between His119 and O_{2P} equatorial oxygen and the cleavage of P-$O_{2'}$ bond occurs before a total proton transfer from His12 to $O_{2'}$ oxygen (P-$O_{2'}$=2.5 Å and $O_{2'}$-Hϵ_2=1.4 Å).

7.4.2 Alternative mechanism: Lys41 as acid catalyst

On the other hand, Wladkowski and co-workers[60,85–87] and Lopez *et al.*[29] (see Section 2.3.2) proposed a mechanism where Lys41 acts as the acid in the

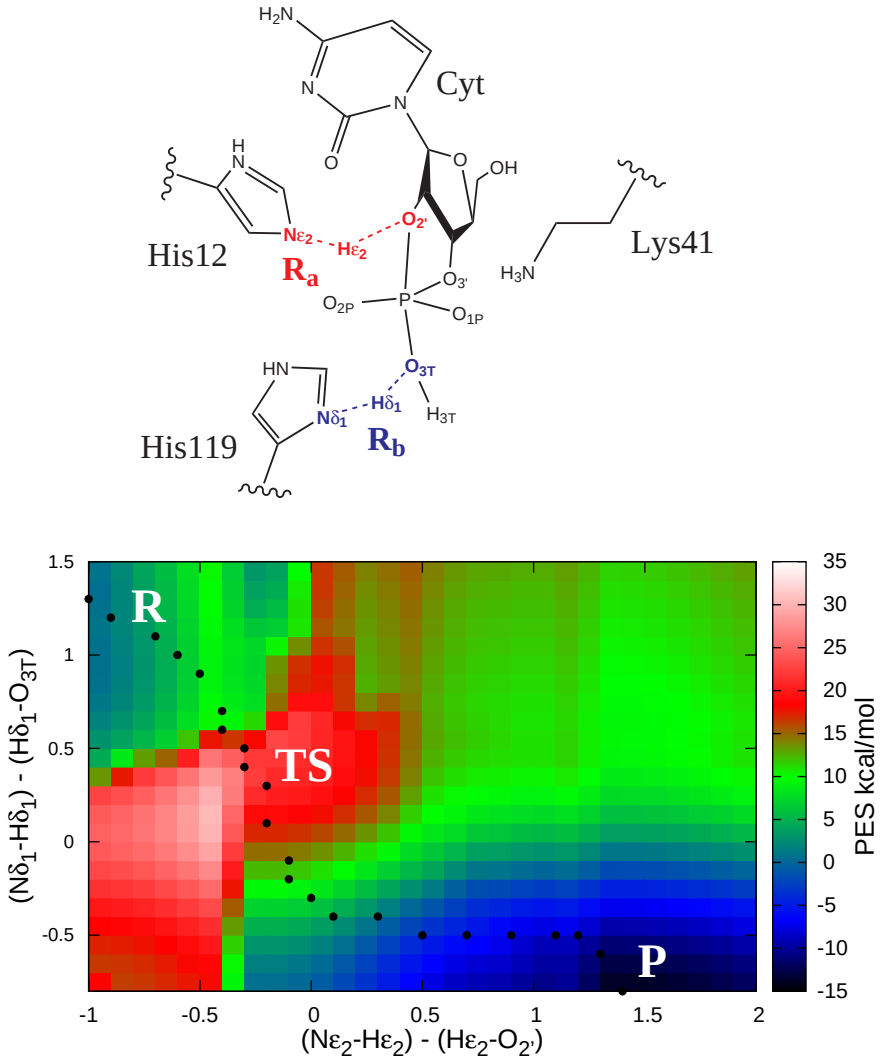


Figure 7.5: The top figure is a scheme of the reaction coordinates used to analyze the classical mechanism in the hydrolysis step. $R_a = (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'})$ reaction coordinate represents a proton transfer from His12 to $O_{2'}$ oxyphosphorane's axial oxygen, while $R_b = (N\delta_1^{His119} - H\delta_1^{His119}) - (H\delta_1^{His119} - O_{3T})$ represents a proton transfer from O_{3T} oxyphosphorane's axial oxygen to His119. The bottom figure is the Potential Energy Surface of the cited reaction coordinates. The black circles represent the proposed reaction path, for which single point energy calculations have been done at DFT level of theory. R, P and TS denote the reactant, product and hydrolysis transition state, respectively. The distances are in Å and the energies in kcal/mol.

hydrolysis step of the reaction, instead of His12. Therefore, we decided to analyze the energy difference between the protonation of 2'-hydroxyl oxygen by His12 or Lys41 in the product structure (3'-Cp), see Figure 7.6. The

Table 7.3: Single point energies in kcal/mol for different reaction coordinates. RC_1 corresponds to the combination of the reaction coordinates $R_a = (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'})$ and $R_b = (N\delta_1^{His119} - H\delta_1^{His119}) - (H\delta_1^{His119} - O_{3T})$. RC_2 corresponds to the combination of the reaction coordinates $R_1 = (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'})$ and $R_2 = (N\zeta^{Lys41} - H\zeta^{Lys41}) - (H\zeta^{Lys41} - O_{2'})$.

			QM(1)^{DFT}		QM(2)^{DFT}	
		AM1/d-PhoT//MM	B3LYP	MPWB1K	B3LYP	MPWB1K
RC_1	TS	22.665	22.563	22.627	20.929	15.454
	Product	-14.084	-4.313	-7.167	-1.504	-6.363
RC_2	¶	11.151	15.160	15.415	8.295	8.479
	§	9.617	6.459	6.179	-0.557	1.734

¶ RNase A - 3'-Cp^{-2} complex, cationic His12 and Lys41.
§ RNase A - 3'-Cp^{-1} complex, neutral Lys41 and cationic His12.

reaction coordinates are

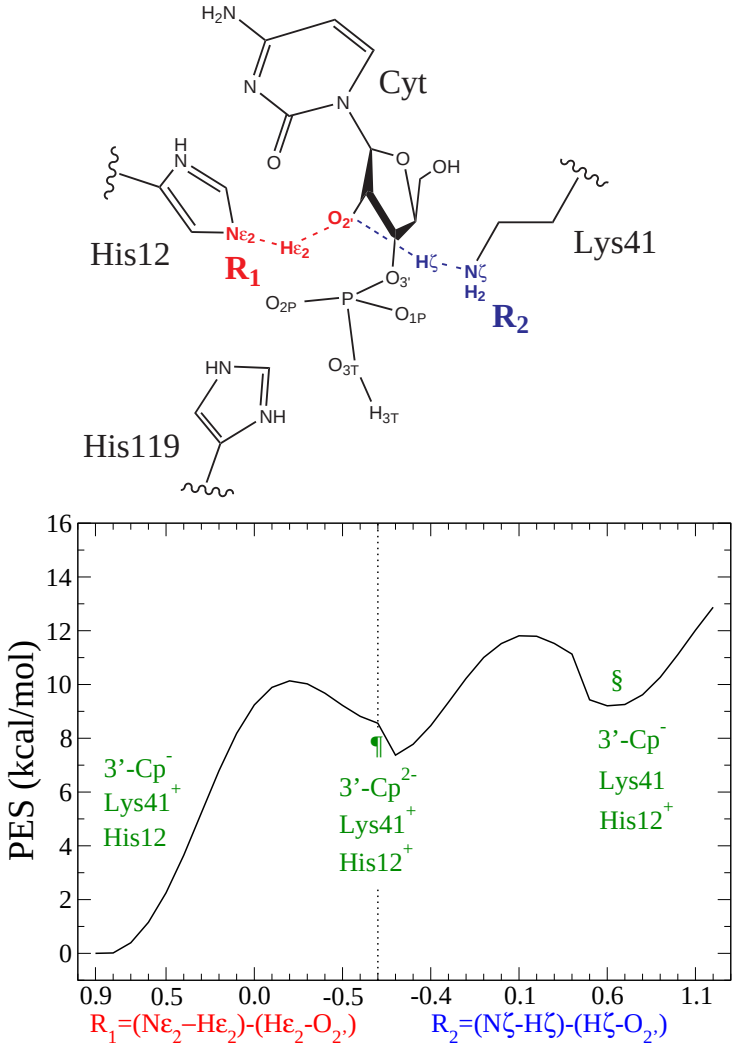
$$\begin{aligned} R_1 &= (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'}) \\ R_2 &= (N\zeta^{Lys41} - H\zeta^{Lys41}) - (H\zeta^{Lys41} - O_{2'}) \end{aligned}$$

where

$$\begin{aligned} 0.9 \text{ Å} &\geq R_1 \geq -0.7 \text{ Å} \\ -0.7 \text{ Å} &\leq R_2 \leq 1.2 \text{ Å} \end{aligned}$$

First a set of RESD minimizations over R_1 reaction coordinate is carried out. As the proton is being transfer to His12, the distances between P and $O_{2'}$ and between $H\zeta^{Lys41}$ and $O_{2'}$ decrease a little bit, from 2.8 Å to 2.6 Å , and from 2.0 Å to 1.8 Å , respectively (see Figure D.6). Then, once His12 has being protonated a new set of RESD minimizations over R_2 is carried out to transfer $H\zeta^{Lys41}$ to the unprotonated 2'-hydroxyl oxygen. Now, as the proton is transferred from Lys41, the distance between P and 2'-hydroxyl oxygen increases 0.1 Å , and the distance between $H\epsilon_2^{His12}$ and $O_{2'}$ increases from 1.9 Å to 2.2 Å. On the other hand, taking into account both sets of minimizations, P-O_{3T} axial bond and the distance between $H\delta_1^{His119}$ and O_{3T} remain constant. The distance between $H\epsilon_2^{His12}$ and O_{2P} phosphoryl oxygen decreases from 2.4 Å to 1.8 Å , while $H\zeta^{Lys41}$ O_{1P} and $H\delta_1^{His119}$ O_{2P} phosphoryl oxygens distances increase from 2.0 Å to 2.3 Å and from 3.3 Å to 3.5 Å, respectively.

Figure 7.6: The top figure is a scheme of the reaction coordinates $R_1 = (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'})$ which represents 2'-hydroxyl proton abstraction by His12, while $R_2 = (N\zeta^{Lys41} - H\zeta^{Lys41}) - (H\zeta^{Lys41} - O_{2'})$ represents a proton transfer from Lys41 to $O_{2'}$ oxygen. The bottom figure is the Potential Energy Surface of the cited reaction coordinates. Distances are in Å and energy in kcal/mol. ¶ represents RNase A - 3'-Cp^{-2} complex, with cationic His12 and Lys41, and § represents RNase A - 3'-Cp^{-} complex, with neutral Lys41 and cationic His12.



The contour of these RESD minimizations is shown in Figure 7.6. The product (3'-Cp, monoanionic) is more stable when His12 is neutral and Lys41 is cationic (left hand side), rather than when Lys41 is neutral and His12 is cationic (right hand side). The energy difference between both states is 9.6 kcal/mol if we take as reference $R_1 = 0.9$ Å and as the last point $R_2 = 0.8$ Å.

However, the energy contour of the second set of minimizations, the one corresponding to R_2 reaction coordinate, is not very reliable. The DFT energy recalculations show that the energy should decrease as Lys41 protonates $O_{2'}$ oxygen. QM(1)^{DFT} decreases the energy till $\sim$ 6 kcal/mol, and QM(2)^{DFT} till $\sim$ -0.6 kcal/mol if B3LYP functional is applied (see Table 7.3). These last results are in concordance with the fact that Lys41 may act as the acid in the hydrolysis step, as proposed by Wladkowski,[60,85–87] Lopez[29] and Elsässer.[61]

7.5 Associative versus dissociative mechanism

The hydrolysis of phosphodiesters is generally thought to follow an associative mechanism.[260–262] Early experimental studies remarked that dissociative mechanism is highly unlikely for the hydrolysis of phosphodiesters.[260] More recent ^{18}O kinetic isotope studies[90] and determination of β_{lg} Brønsted-type slopes for the cleavage of hydroxide-ion-catalyzed RNA phosphodiester bonds[263–266] also support an associative mechanism. The results obtained here from the analysis of different reaction coordinates, clearly agree with that view. We can concluded that the reaction mechanism can be classify as associative. In the transphosphorylation step, before the cleavage of P-$O_{5'}$ bond, P-$O_{2'}$ must be formed, and in the hydrolysis step before the cleavage of P-$O_{2'}$ bond, P-O_{3T} must be formed.

The transphosphorylation classical mechanism subsection (Section 7.3.1) shows that there is a first concerted mechanism step between the nucleophilic attack of $O_{2'}$ oxygen and the 2'-hydroxyl proton abstraction by His12. This step is followed by the cleavage and protonation of P-$O_{5'}$ bond by Hδ_1^{His119} (see Figure D.1). The P-O axial distances of the TS are around 1.8 Å for P-$O_{2'}$ and 1.7 Å for P-$O_{5'}$. Lys41 as base catalyst subsection (Section 7.3.2) shows similar results: 2'-hydroxyl proton abstraction by His12 or Lys41 occurs simultaneously to P-$O_{2'}$ bond cleavage. The P-O axial distances of the [Trans]^{oph} are around 1.7 Å for P-$O_{2'}$ and 1.8 Å for P-$O_{5'}$. Finally, the hydrolysis classical mechanism subsection (Section 7.4.1) shows a first concerted step between the nucleophilic attack and the nucleophile's proton abstraction by His119, which is followed by the cleavage and protonation of P-$O_{2'}$ bond by Hϵ_2^{His12} (see Figure D.5). The P-O axial distances of the TS are around 1.7 Å for P-$O_{2'}$ and 1.9 Å for P-$O_{5'}$.

Further Work

QM/MM methodology has been applied in this part of the Thesis to analyze the hydrolysis reaction mechanism of RNase A. The first chapter of this part, Chapter 6, presents the structures and an energetic profile of the dianionic versus monoanionic transphosphorylation's and hydrolysis' high-energy phosphate intermediates that RNase A has to stabilize in order to catalyze the reaction. The MD simulations presented for each window are quite short, 15ps. We recognize that longer dynamics should be need to give more representative/realistic pattern of hydrogen bond interactions and solvation of the complex. Work along the elongation of these simulations is under progress.

On the other hand, in Chapter 7 we show PES of some studied reaction coordinates for the transphosphorylation and hydrolysis step of RNase A -3',5'-CpA complex hydrolysis reaction. This chapter has been one of the most complicated ones, due to the difficulty on choosing a good reaction coordinate. One that allows us to follow all the steps of the reaction mechanism. Although, plenty of reaction coordinates have been studied, some of them without any successful results, there are more reaction coordinates one could try. Besides, it would be desirable to run MD simulations in order to give results with a properly equilibrated environment. We have already started work along this line, but there are several difficulties as the proper reaction coordinate to choose. Due to the time limitation of this thesis project, work on this will be continued in the future.

Part IV

Appendix

Chapter A

Supporting Information of Chapter 4

Time evolution of ligand dihedral angles, phase angles, torsional dihedral angles and selected distances are shown in the Supporting Information.

Table A.1: Contact distances between heavy-atoms of the active-site residues and the ligand. All distances are in Å. In the simulations, only the last 5000 ps are taken into account. Standard deviation is in parentheses

	Crystal Structures					Molecular Dynamics		
Atom pair	1RPG[50]	1RCA[267]	5RSA[211]	6RSA[212]	1RUV[213]	3',5'-CpA	$[Trans]^{TS}$	$[Hyd]^{TS}$
GLN 11 O - ASN 44 $N\delta_2$	2.8	3.0	2.8	2.8	2.8	2.86(0.14)	2.84(0.13)	2.88(0.15)
GLN 11 $O\epsilon_1$ - O_{1P}	4.5	3.1	3.5	2.5	3.8	5.84(0.86)	3.72(0.29)	3.94(0.40)
GLN 11 $O\epsilon_1$ - LYS 41 $N\zeta$		3.9	4.2	3.6	2.9	4.32(0.66)	3.27(0.52)	3.25(0.77)
GLN 11 $N\epsilon_2$ - LYS 41 $N\zeta$	4.4	3.0	4.1	3.9	3.9	4.33(0.43)	4.09(0.32)	3.88(0.41)
GLN 11 $N\epsilon_2$ - O_{1P}	2.9	4.0	5.0	2.5	2.9	5.21(0.63)	2.92(0.18)	2.96(0.41)
GLN 11 $N\epsilon_2$ - O_{2P}	4.2	5.8	6.0	4.1	3.9	3.21(0.45)	3.79(0.27)	3.91(0.33)
HIS 12 $N\delta_1$ - ASN 44 $O\delta_1$	3.4	3.6	3.4	3.6	3.5	3.78(0.45)	3.38(0.35)	3.49(0.44)
HIS 12 $N\delta_1$ - THR 45 O	2.7	2.7	2.7	2.5	2.7	3.01(0.19)	2.87(0.15)	2.83(0.13)
HIS 12 $N\delta_1$ - GUA O_6		3.5						
HIS 12 $N\delta_1$ - URI O_2				3.6	3.8			
HIS 12 $N\delta_1$ - CYT O_2	3.8					3.80(0.28)	3.47(0.17)	3.53(0.18)
HIS 12 $N\epsilon_2$ - O_{2P}	2.8	3.0	2.6	2.7	2.7	4.23(0.41)	2.85(0.16)	2.72(0.10)
HIS 12 $N\epsilon_2$ - O_{1P}	4.3	4.5	4.4	4.0	3.9	6.43(0.49)	4.39(0.16)	4.43(0.17)
HIS 12 $N\epsilon_2$ - $O_{2'}$		3.2	3.1	3.0	3.4	3.49(0.28)	3.07(0.15)	3.16(0.17)
LYS 41 $N\zeta$ - $O_{2'}$		2.8	3.0	2.8	2.7	2.93(0.25)	2.99(0.19)	3.20(0.38)
LYS 41 $N\zeta$ - O_{1P}	4.7	3.7	5.2	3.5	3.3	4.42(0.69)	3.01(0.23)	3.03(0.38)
LYS 41 $N\zeta$ - O_{2P}	4.9	5.0	5.2	5.0	4.8	4.02(0.74)	3.80(0.97)	4.97(0.24)
LYS 41 $N\zeta$ - $O_{3'}$	3.7			5.0	4.9	5.00(0.28)	4.42(0.17)	4.59(0.21)

Continued on next page

Table A.1: Contact distances between heavy-atoms of the active-site residues and the ligand. All distances are in Å. In the simulations, only the last 5000 ps are taken into account. Standard deviation is in parentheses

	Crystal Structures					Molecular Dynamics		
Atom pair	1RPG[50]	1RCA[267]	5RSA[211]	6RSA[212]	1RUV[213]	3',5'-CpA	$[Trans]^{TS}$	$[Hyd]^{TS}$
LYS 41 Nζ - ASN 44 Oδ_1	2.6	2.9	2.9	2.8	2.8	4.12(0.60)	4.27(0.64)	4.48(1.11)
VAL 43 O - Ribose1 $O_{4'}$	3.8	6.5		3.5	3.3	3.90(0.34)	3.65(0.25)	3.71(0.25)
THR 45 N - GUA O_6		3.0						
THR 45 N - CYT O_2	3.0					3.02(0.16)	2.92(0.13)	2.93(0.14)
THR 45 N - URI O_2				2.6	2.9			
THR 45 Oγ_1 - URI N_3				2.7	2.8			
THR 45 Oγ_1 - CYT N_3	2.8					2.90(0.12)	2.89(0.12)	2.90(0.12)
THR 45 Oγ_1 - GUA N_7		2.8						
ASN 71 Nδ_2 - ADE N_1	2.9					3.16(0.49)	3.05(0.17)	
ASN 71 Oδ_1 - ADE N_6	2.8					3.18(0.71)	3.12(0.42)	
ARG 85 NH1 - Ribose $O_{5'}$	10.3	3.1		9.0	8.0	8.84(0.73)	9.15(0.82)	9.40(0.85)
GLU 111 Oϵ_2 - ADE N_1	3.2					6.49(1.27)	5.48(1.22)	
GLU 111 Oϵ_2 - ADE C_2	2.4					5.94(1.24)	4.79(1.15)	
GLU 111 Oϵ_2 - ADE N_3	3.1					6.62(1.26)	5.34(1.09)	
HIS 119 Nϵ_2 - ASP 121 Oδ_1	2.6	2.6	2.7	2.9	2.7	3.37(1.02)	3.01(0.69)	2.73(0.19)
HIS 119 Nδ_1 - $O_{5'}$	2.7	2.6	2.4	3.0	2.6	2.99(0.24)	2.84(0.10)	2.99(0.25)
HIS 119 Nδ_1 - O_{2P}	3.8	3.4	3.9	3.9	3.7	4.49(0.32)	3.66(0.20)	3.86(0.29)
HIS 119 Nδ_1 - $O_{3'}$	4.4			2.6	3.2	3.32(0.24)	3.15(0.17)	2.96(0.15)
HIS 119 Nδ_1 - Ribose2 $O_{4'}$	3.3					3.40(0.34)	3.48(0.29)	

Continued on next page

Table A.1: Contact distances between heavy-atoms of the active-site residues and the ligand. All distances are in Å. In the simulations, only the last 5000 ps are taken into account. Standard deviation is in parentheses

	Crystal Structures					Molecular Dynamics		
Atom pair	1RPG[50]	1RCA[267]	5RSA[211]	6RSA[212]	1RUV[213]	3',5'-CpA	$[\text{Trans}]^{TS}$	$[\text{Hyd}]^{TS}$
PHE 120 N - O_{2P}	3.07	2.8	2.8	2.9	2.9	5.35(0.50)	3.20(0.22)	3.12(0.23)
PHE 120 O - $O_{3'}$	4.7			3.5	3.6	5.92(0.99)	3.68(0.24)	3.80(0.24)
GUA N_1 - $O_{2'}$		2.7						
GUA N_2 - $O_{5'}$		3.3						

$[\text{Trans}]^{TS}$ is the transphosphorylation transition state mimic model
$[\text{Hyd}]^{TS}$ is the hydrolysis transition state mimic model
1RPG Deoxycytidylyl-3',5'-deoxyadenosine (d(CpA))
1RCA Deoxycytidylyl-3',5'-guanosine (3',5'-dCpdG = d(CpG))
5RSA Phosphate ion
6RSA Uridine-2',3'-vanadate (U>v)
1RUV Uridine-2',3'-vanadate (U>v)

Table A.2: Phosphorane or vanadate structure. All distances are in Å and angles in degrees. Only the last 5000 ps of the simulations are taken into account. Standard deviation is in parentheses.

	Crystal Structures					Molecular Dynamics		
Atom pair	1RPG[50]	1RCA[267]	5RSA[211]	6RSA[212]	1RUV[213]	3',5'-CpA	$[\text{Trans}]^{TS}$	$[\text{Hyd}]^{TS}$
P/V - $O_{2'}$		1.6	1.5	1.9	2.0	3.24(0.09)	1.97(0.04)	1.96(0.04)
P/V - $O_{5'}$	1.6	1.5	1.5	1.9	1.9	1.57(0.03)	1.76(0.05)	1.72(0.05)
P/V - $O_{3'}$	1.6			1.9	2.0	1.56(0.03)	1.74(0.04)	1.73(0.04)
P/V - O_{2P}	1.5	1.5	1.5	1.7	1.8	1.48(0.02)	1.53(0.02)	1.52(0.02)
P/V - O_{1P}	1.5	1.5	1.5	1.7	1.3	1.48(0.02)	1.52(0.02)	1.52(0.02)
$O_{5'}$ - P/V - $O_{2'}$		100.0	116.8	165.6	149.5		164.2(2.4)	164.2(2.6)
$O_{5'}$ - P/V - $O_{3'}$	102.5			77.8	73.1	100.8(3.3)	81.8(2.3)	81.7(2.6)
$O_{5'}$ - P/V - O_{2P}	106.5	63.5	111.3	91.7	96.0	108.7(2.9)	95.8(3.4)	94.3(4.1)
$O_{5'}$ - P/V - O_{1P}	104.2	118.6	102.3	83.7	100.6	108.9(2.9)	98.7(3.4)	96.4(3.7)
$O_{2'}$ - P/V - $O_{3'}$				91.0	76.9		83.2(1.5)	83.5(1.6)
$O_{2'}$ - P/V - O_{2P}		101.6	108.9	101.5	104.9		89.1(2.8)	89.7(2.8)
$O_{2'}$ - P/V - O_{1P}		112.9	108.0	97.5	96.0		89.5(2.8)	92.5(2.8)
O_{2P} - P/V - O_{1P}	121.8	143.6	109.1	107.5	103.1	117.4(2.7)	127.4(1.9)	127.7(1.9)
O_{2P} - P/V - $O_{3'}$	110.3			117.9	123.2	108.4(2.9)	116.7(3.1)	117.2(3.2)
O_{1P} - P/V - $O_{3'}$	109.6			131.3	133.6	111.0(2.9)	115.2(3.2)	114.7(3.2)
$O_{3'}$ - O_{2P} - O_{1P} - P/V	-31.5			10.9	2.2	-33.6(2.8)	-6.0	-2.9

$[\text{Trans}]^{TS}$ is the transphosphorylation transition state mimic model
$[\text{Hyd}]^{TS}$ is the hydrolysis transition state mimic model
1RPG Deoxycytidylyl-3',5'-deoxyadenosine (d(CpA))
1RCA Deoxycytidylyl-3',5'-guanosine (3',5'-dCpdG = d(CpG))
5RSA Phosphate ion
6RSA Uridine-2',3'-vanadate (U>v)
1RUV Uridine-2',3'-vanadate (U>v)

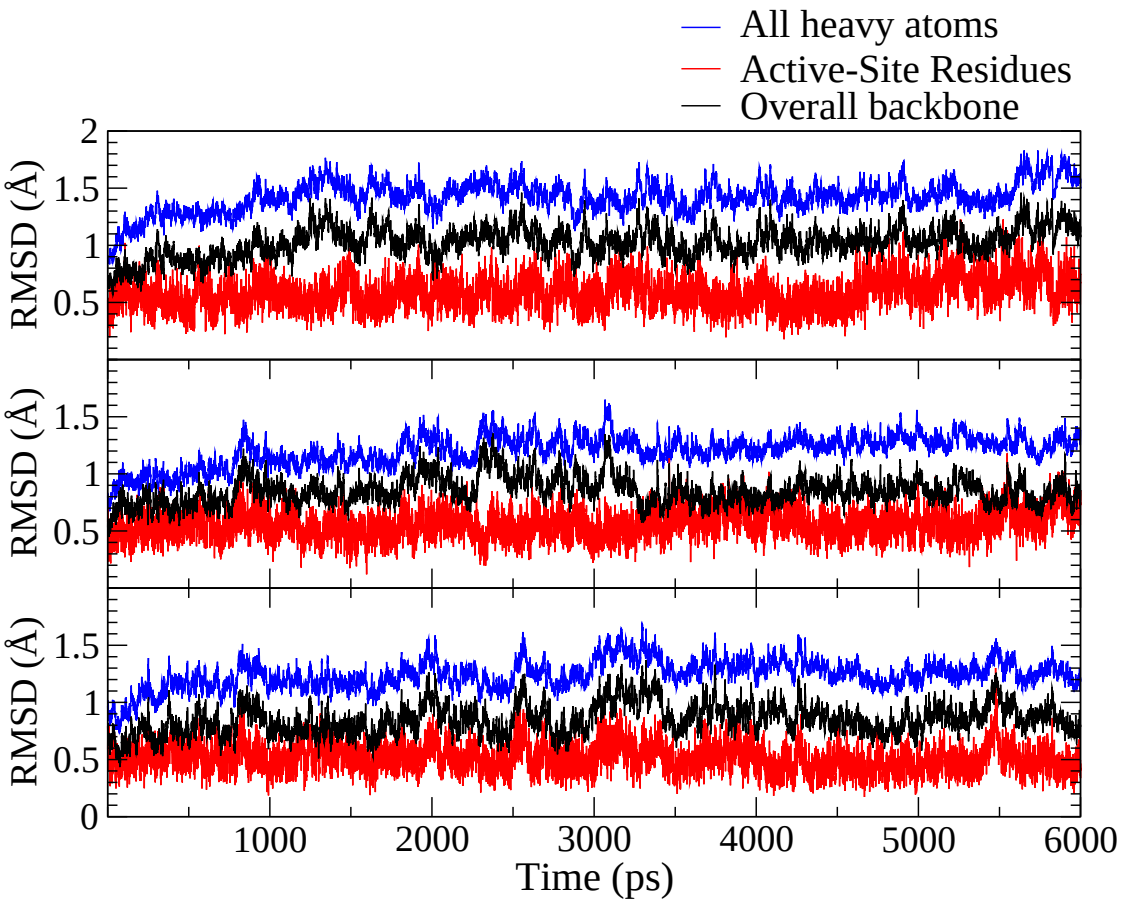


Figure A.1: Root-mean-square deviation of the reactant model at top, the [Trans]^{TS} mimic model in the middle and the [Hyd]^{TS} mimic model at the bottom, are represented. On each plot the all-heavy-atom RMSD is shown in blue, the overall backbone in black and the active site residues backbone (His12, His119, Lys41 and the ligand) in red. The RMSD is from the starting conformation overlapping backbone of residues 3 to 122 , plotted as a function of the simulation time.

Table A.3: Averaged structural parameters of the dinucleotides in the RNase A - 3',5'-CpA complex for the last 5 ns of the production time. All angles are in degrees.

	3',5'-CpA				3',5'-d(CpA) X-Ray[†]	
Angles	Cyt	Range	Ade	Range	Cyt	Ade
Glycosyl torsion						
$O_{4'}$-$C_{1'}$-N_1-C_2 (χ_C)	-160.97(7.02) (anti)	169.87 $\cdots$ -124.30			-155 (anti)	
$O_{4'}$-$C_{1'}$-N_9-C_4 (χ_A)			-65.71(13.53) (anti)	-148.36 $\cdots$ -9.63		-103 (anti)
Sugar torsion						
$C_{4'}$-$O_{4'}$-$C_{1'}$-$C_{2'}$ ($\tau 0$)	1.97	-49.80 $\cdots$ 27.68	-33.35	-60.31 $\cdots$ 13.75	-18	-24
$O_{4'}$-$C_{1'}$-$C_{2'}$-$C_{3'}$ ($\tau 1$)	-25.87	-46.91 $\cdots$ 30.76	39.14	-35.85 $\cdots$ 59.88	25	37
$C_{1'}$-$C_{2'}$-$C_{3'}$-$C_{4'}$ ($\tau 2$)	38.41	-14.67 $\cdots$ 53.65	-29.90	-51.46 $\cdots$ 49.33	-22	-35
$C_{2'}$-$C_{3'}$-$C_{4'}$-$O_{4'}$ ($\tau 3$)	-38.21	-55.26 $\cdots$ -3.09	11.17	-53.47 $\cdots$ 41.16	11	22
$C_{3'}$-$C_{4'}$-$O_{4'}$-$C_{1'}$ ($\tau 4$)	40.31	-2.13 $\cdots$ 75.47	13.66	-25.55 $\cdots$ 53.44	5	1
Phase	23.20 N $C_{3'}$-endo		140.83 S $C_{2'}$-endo		151.4 S $C_{2'}$-endo	159.6 S $C_{2'}$-endo
Backbone torsion						
$O_{5'}$-$C_{5'}$-$C_{4'}$-$C_{3'}$ (γ)	60.55 (g^+)	-56.59 $\cdots$ -179.70	56.33 (g^+)	3.75 $\cdots$ -165.51	31 (g^+)	172 (t)
$C_{5'}$-$C_{4'}$-$C_{3'}$-$O_{3'}$ (δ)	71.96 (g^+)	49.74 $\cdots$ 106.64	125.18 (g^+)	56.35 $\cdots$ 164.09	122 (g^+)	139 (g^+)
Phosphodiester torsion						
$O_{3'}$-P-$O_{5'}$-$C_{5'}$ (α)			-155.72 (t)	-21.47 $\cdots$ -88.37		77 (g^+)
P-$O_{5'}$-$C_{5'}$-$C_{4'}$ (β)			-149.42 (t)	50.50 $\cdots$ -93.08		110 (g^+)
$C_{4'}$-$C_{3'}$-$O_{3'}$-P (ϵ)	50.03 (g^+)	10.03 $\cdots$ -108.48			-160 (t)	
$C_{3'}$-$O_{3'}$-P-$O_{5'}$ (ζ)	-173.81 (t)	14.30 $\cdots$ -72.32			55 (g^+)	

[†] X-Ray data from Zegers *et al.*[50]
The Phase is defined as $\tan(P) = \frac{(\tau_4+\tau_1)-(\tau_3+\tau_0)}{(2*\tau_2*(\sin(36)+\sin(72))}$ If τ_2 has a negative value, then 180° is added to the value of P[230,232]

Figure A.2: Time evolution of selected ligand dihedral angles and phase angle for the simulations corresponding to each complex. On the left column the glycosyl dihedral angles of cytosine (χ_C in black) and adenine (χ_A in red) with His119 χ_1 in green and χ_2 in blue are depicted, while on the right column the phase angle of cytosine in black and adenine in red are shown. The first row correspond to the reactand model, the middle one to the [Trans]^{TS} model and the last one to the [Hyd]^{TS} model.

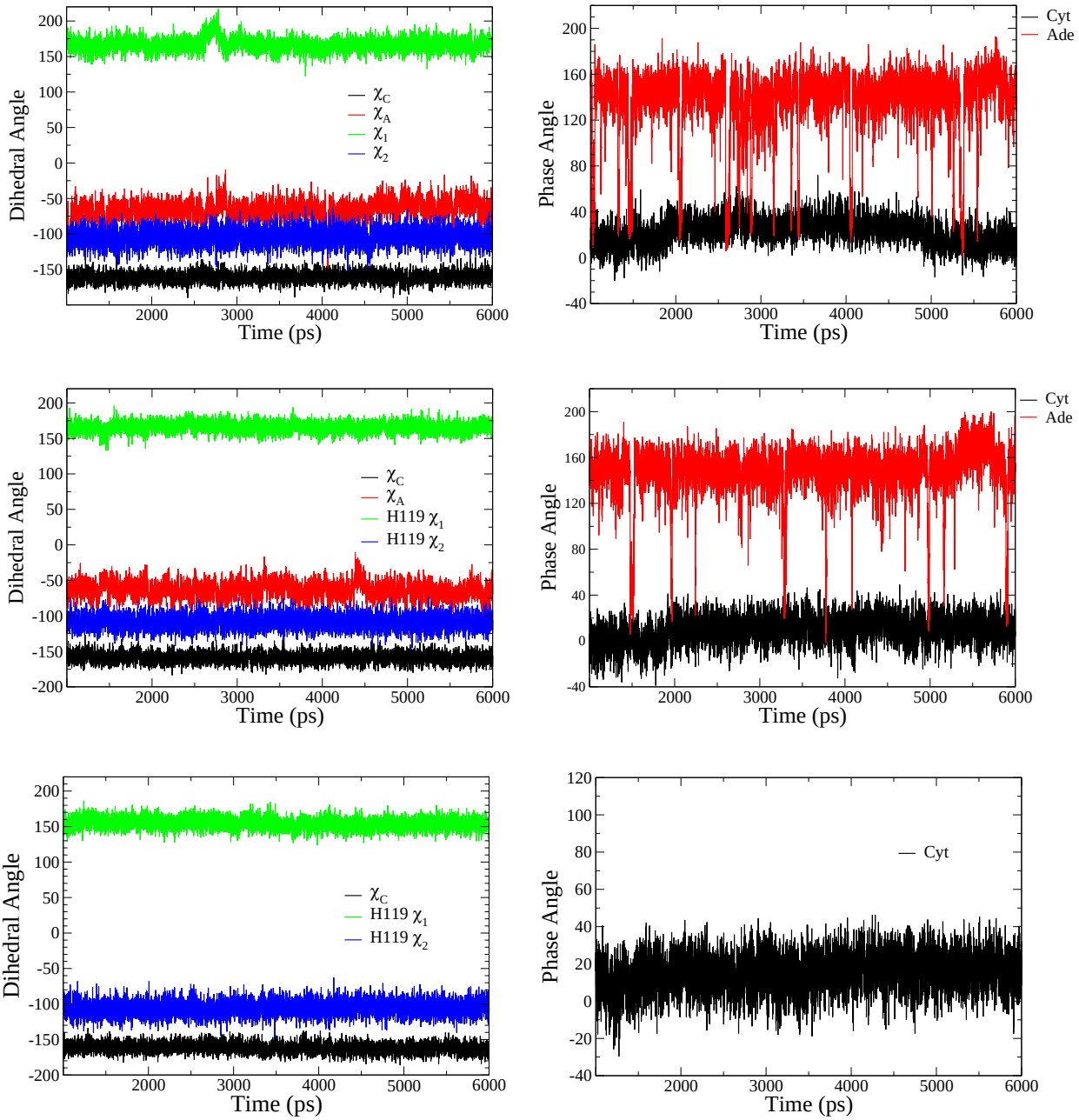


Figure A.3: Time evolution of reactant model's phosphate torsional dihedral angles are shown in the first row. Distances between O_{1P} phosphoryl oxygen and imidazolium side chain of residue His119 (in red), main chain of residue Phe120 (in green), side chain of Gln11 (in blue) and Cyt $H_{2'}$ atom (in black) are depicted in the middle left. In the middle right, the distances between ammonium ion of residue Lys41 and Cyt 2'-hydroxyl oxygen ($O_{2'}$) are shown. Finally, at lower left corner, ditances between side chain of residue His119 and $O_{3'}$ (in red) and $O_{5'}$ (in black) phosphate oxygens are depicted.

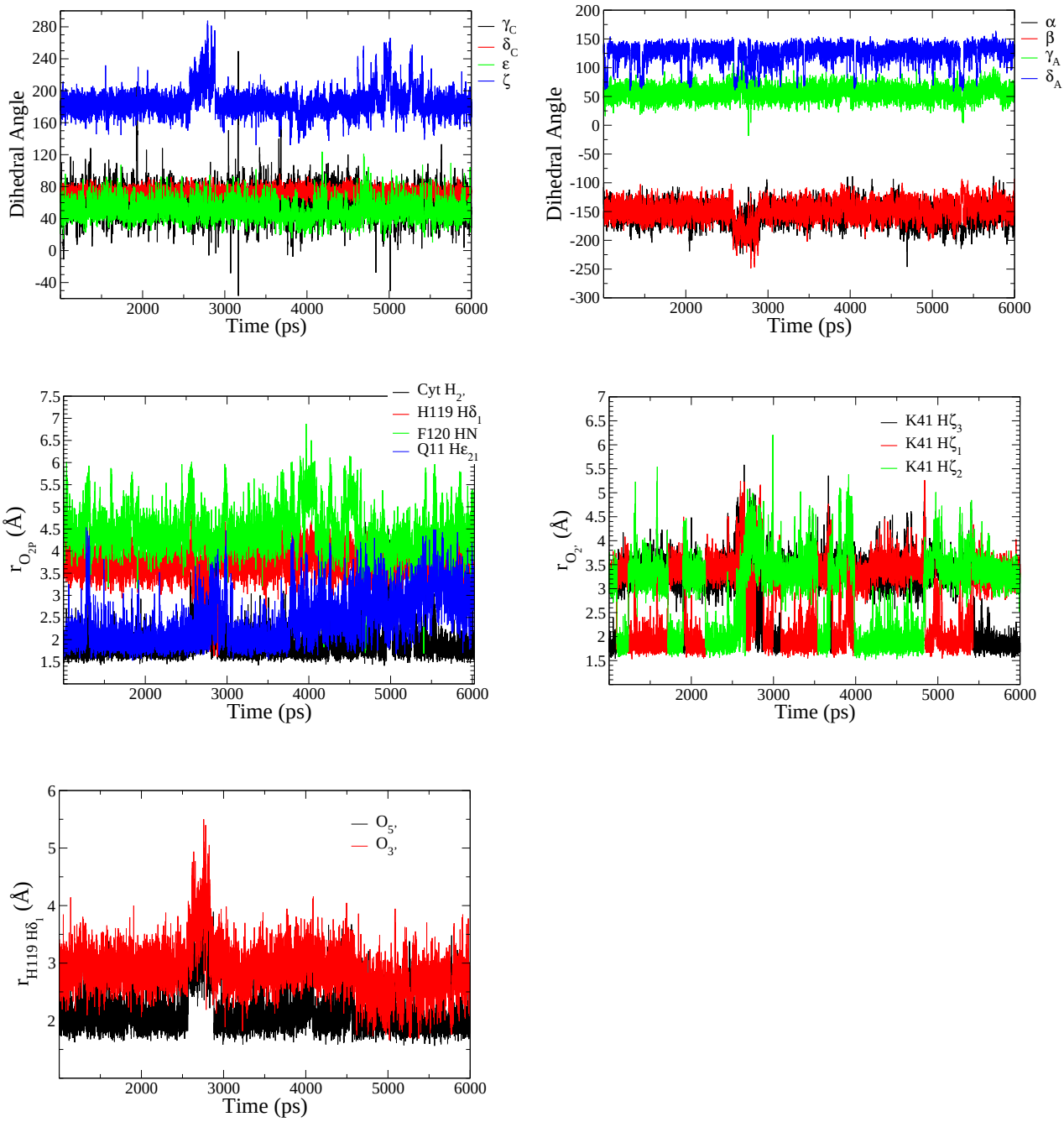


Table A.4: Averaged structural parameters of the dinucleotides in the RNase A - [Trans]^{TS} and [Hyd]^{TS} mimic complexes for the last 5 ns of the production time. All angles are in degrees.

	Transphosphorylation TS				Hydrolysis TS	
Angles	Cyt	Range	Ade	Range	Cyt	Range
Glycosyl torsion						
$O_{4'}$-$C_{1'}$-N_1-C_2 (χ_C)	-158.85 (anti)	176.23 $\cdots$ -126.93			-161.59 (anti)	173.63 $\cdots$ -133.89
$O_{4'}$-$C_{1'}$-N_9-C_4 (χ_A)			-63.14 (high anti)	-126.77 $\cdots$ -10.76		
Sugar torsion						
$C_{4'}$-$O_{4'}$-$C_{1'}$-$C_{2'}$ ($\tau 0$)	3.58	-20.38 $\cdots$ 27.03	-30.19	-54.09 $\cdots$ 15.25	1.65	-22.83 $\cdots$ 26.49
$O_{4'}$-$C_{1'}$-$C_{2'}$-$C_{3'}$ ($\tau 1$)	-28.94	-72.64 $\cdots$ -4.75	39.50	-35.87 $\cdots$ 59.18	-23.49	-39.97 $\cdots$ -6.10
$C_{1'}$-$C_{2'}$-$C_{3'}$-$C_{4'}$ ($\tau 2$)	33.86	18.84 $\cdots$ 48.01	-33.51	-51.90 $\cdots$ 45.04	35.02	19.59 $\cdots$ 47.08
$C_{2'}$-$C_{3'}$-$C_{4'}$-$O_{4'}$ ($\tau 3$)	-32.96	-50.26 $\cdots$ -10.32	16.79	-50.61 $\cdots$ 44.72	-35.22	-51.16 $\cdots$ -13.71
$C_{3'}$-$C_{4'}$-$O_{4'}$-$C_{1'}$ ($\tau 4$)	18.09	-6.89 $\cdots$ 41.54	8.07	-27.03 $\cdots$ 53.03	20.66	-5.25 $\cdots$ 44.39
Phase	10.08 N $C_{3'}$-endo		149.41 S $C_{2'}$-endo		15.92 N $C_{3'}$-endo	
Backbone torsion						
$O_{5'}$-$C_{5'}$-$C_{4'}$-$C_{3'}$ (γ)	60.38 (g^+)	-179.67 $\cdots$ 178.42	57.02 (g^+)	15.30 $\cdots$ -160.98	65.59 (g^+)	0.00 $\cdots$ -0.16
$C_{5'}$-$C_{4'}$-$C_{3'}$-$O_{3'}$ (δ)	85.75 (g^+)	62.20 $\cdots$ 111.80	130.87 (g^+)	59.79 $\cdots$ 164.03	83.92 (g^+)	58.97 $\cdots$ 110.29
Phosphodiester torsion						
$O_{3'}$-P-$O_{5'}$-$C_{5'}$ (α)			176.51 (t)	4.84 $\cdots$ -115.00		
P-$O_{5'}$-$C_{5'}$-$C_{4'}$ (β)			-164.90 (t)	124.33 $\cdots$ -106.28		
$C_{4'}$-$C_{3'}$-$O_{3'}$-P (ϵ)	99.94 (g^+)	72.94 $\cdots$ 126.07			100.82 (g^+)	71.45 $\cdots$ 123.50
$C_{3'}$-$O_{3'}$-P-$O_{5'}$ (ζ)	172.93 (t)	140.89 $\cdots$ -152.83			169.63 (t)	143.19 $\cdots$ -149.74

The Phase is defined as $\tan(P) = \frac{(\tau_4+\tau_1)-(\tau_3+\tau_0)}{(2*\tau_2*(\sin(36)+\sin(72))}$ If τ_2 has a negative value, then 180° is added to the value of $P^{230,232}$

Figure A.4: [Trans]^{TS} model's time evolution of some phosphodiester torsional dihedral angles are shown at upper left corner. Distances between O_{2P} equatorial oxygen and imidazolium side chain of residue His12 (in black), main chain of Phe120 (in green) and a water molecule (in red) are depicted at upper right corner. At lower left corner, the distances between imidazolium side chain of residue His119 and $O_{5'}$ axial oxygen, $O_{3'}$ and O_{2P} equatorial oxygens are shown. Finally, at lower right corner distances between O_{1P} equatorial oxygen and Gln11 (in black) and between $O_{2'}$ axial oxygen and His12 (in red) are plotted.

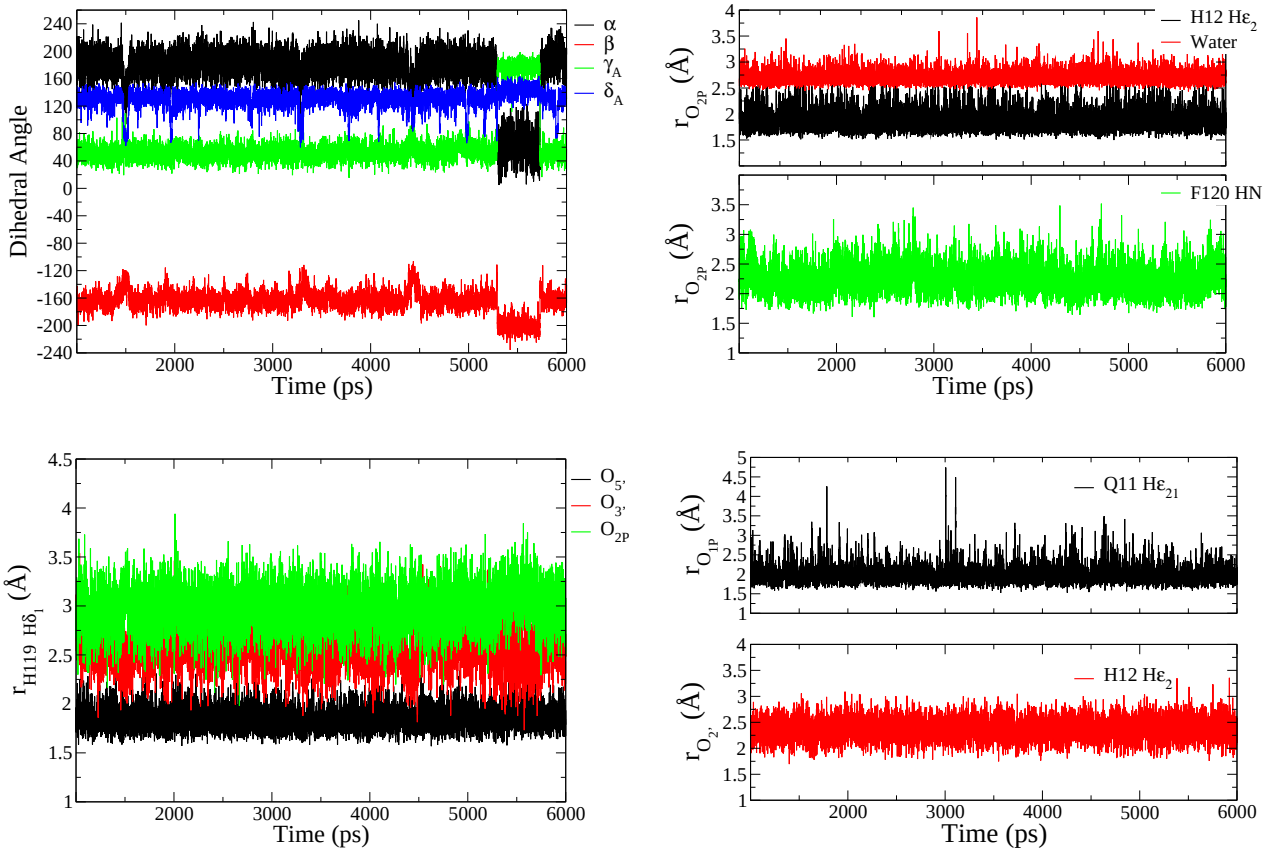
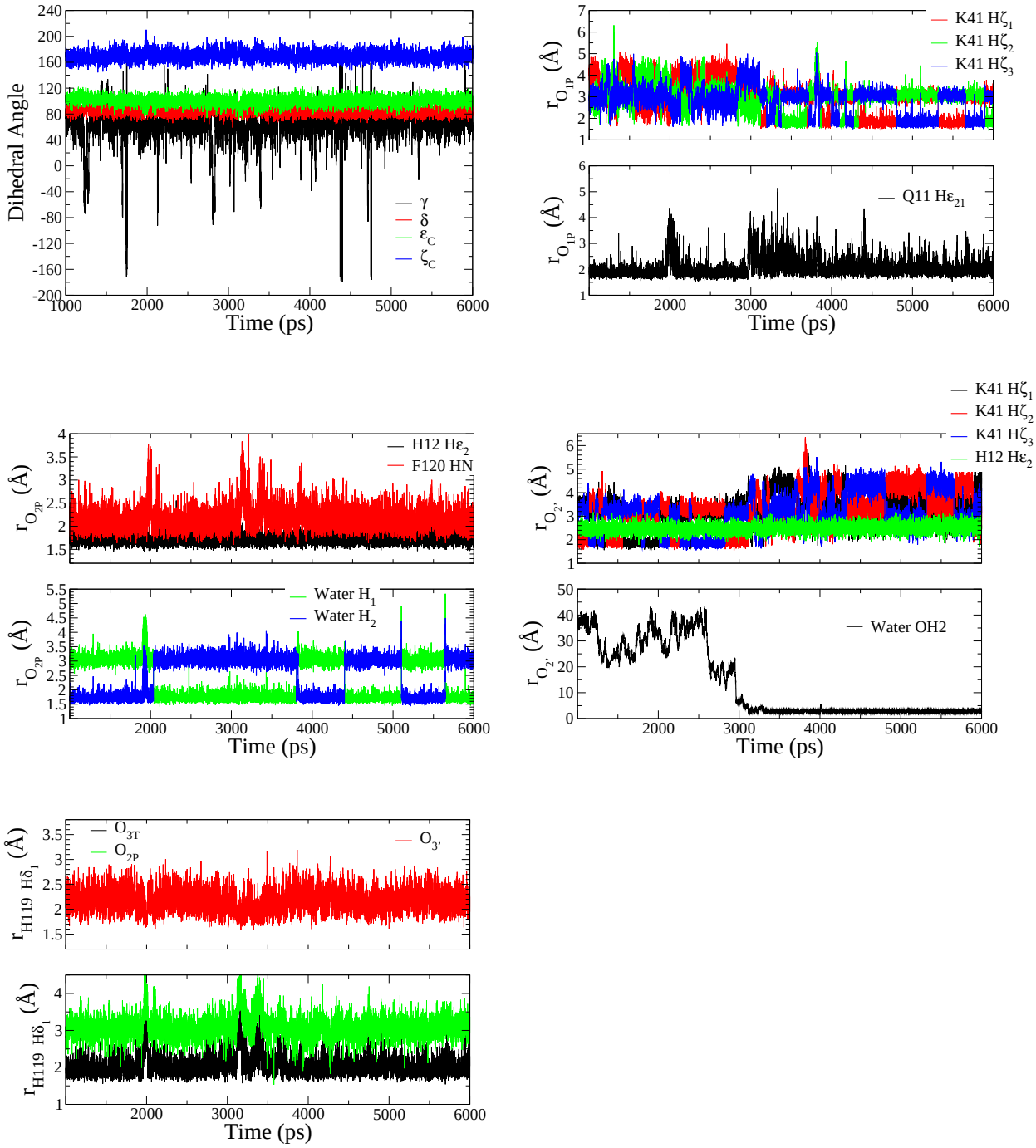


Figure A.5: [Hyd]^{TS} model's time evolution of some phosphodiester torsional dihedral angles are shown at upper left corner. Distances between O_{2P} equatorial oxygen and residues His12 (in blue), Phe120 (in orange) and a water molecule are depicted in the middle left. At upper right corner, the distances between O_{1P} equatorial oxygen and residues Lys41 and Gln11 (in black) are shown. Finally, at lower right corner distances between $O_{2'}$ axial oxygen and His12 (in green), Lys41 and a water molecule (in orange) are plotted.



Chapter B

Supporting Information of Chapter 5

A complete analysis of transphosphorylation transition state mimic $[S{:}O_{1P}]^{Trans}$ model last 2193 ps is done in this supporting information. Time evolution of ligand dihedral angles, phase angles, torsional dihedral angels and selected distances are shown.

Conformational dynamics

The time evolution of key dihedral angles that define the conformation of the substrate and residue His119 are shown in Figure B.2 and average values are list in Table B.1. The fluctuation of glycosyl dihedral angles χ_C and χ_A is small, signifying that both glycosyl dihedral angles syn and high-anti conformation are stable throughout the simulation. The average values are 40.03° and -64.59° respectively. Nevertheless, we have to take into account that χ_C suffers a transition from its usual anti conformation to syn, between the first part of the simulation and this last part. The cytidine ribose essentialy stays in North $C_{3'}$-endo conformation throughout the simulation, while the adenosine ribose occupies mainly a South $C_{2'}$-endo conformation with transitions to other puckers. The average values are 28.73° and 163.98° respectively (see Figure B.3 and Table B.5). Besides, α phosphodiester dihedral angle undergoes a conformational transition at $\sim$ 4240 ps, from trans to gauche$^-$, while from $\sim$ 3907 ps to 4776 ps, γ_C phosphodiester dihedral angle undergoes a conformational transition from gauche$^+$ to gauche$^-$ (see Figure B.5). It is the first time we see transitions to gauche$^-$ conformation. Finally, imidazolium side chain of residue His119 retains an active A-type conformation like in the $[native]^{Trans}$ model,[243] with a mean χ_1 value of -157.43° and χ_2 value of -106.52° (see Figure B.2).

Figure B.1: Root-mean-square deviation of $[S{:}O_{1P}]^{Trans}$ and $[S{:}O_{2P}]^{Trans}$ on the first row are depicted. In the middle, $[S{:}O_{1P},O_{2P}]^{Trans}$ and $[S{:}O_{1P},O_{2P}]^{Hyd}$ dithio models are shown. Finally, in the last row $[S{:}O_{1P}]^{Hyd}$ and $[S{:}O_{2P}]^{Hyd}$ models are depicted. On each plot the all-heavy-atom RMSD is shown in blue, the overall backbone in black and the active site residues backbone (His12, His119, Lys41 and the ligand) in red. The RMSD is from the starting conformation overlapping backbone of residues 3 to 122 , plotted as a function of the simulation time.

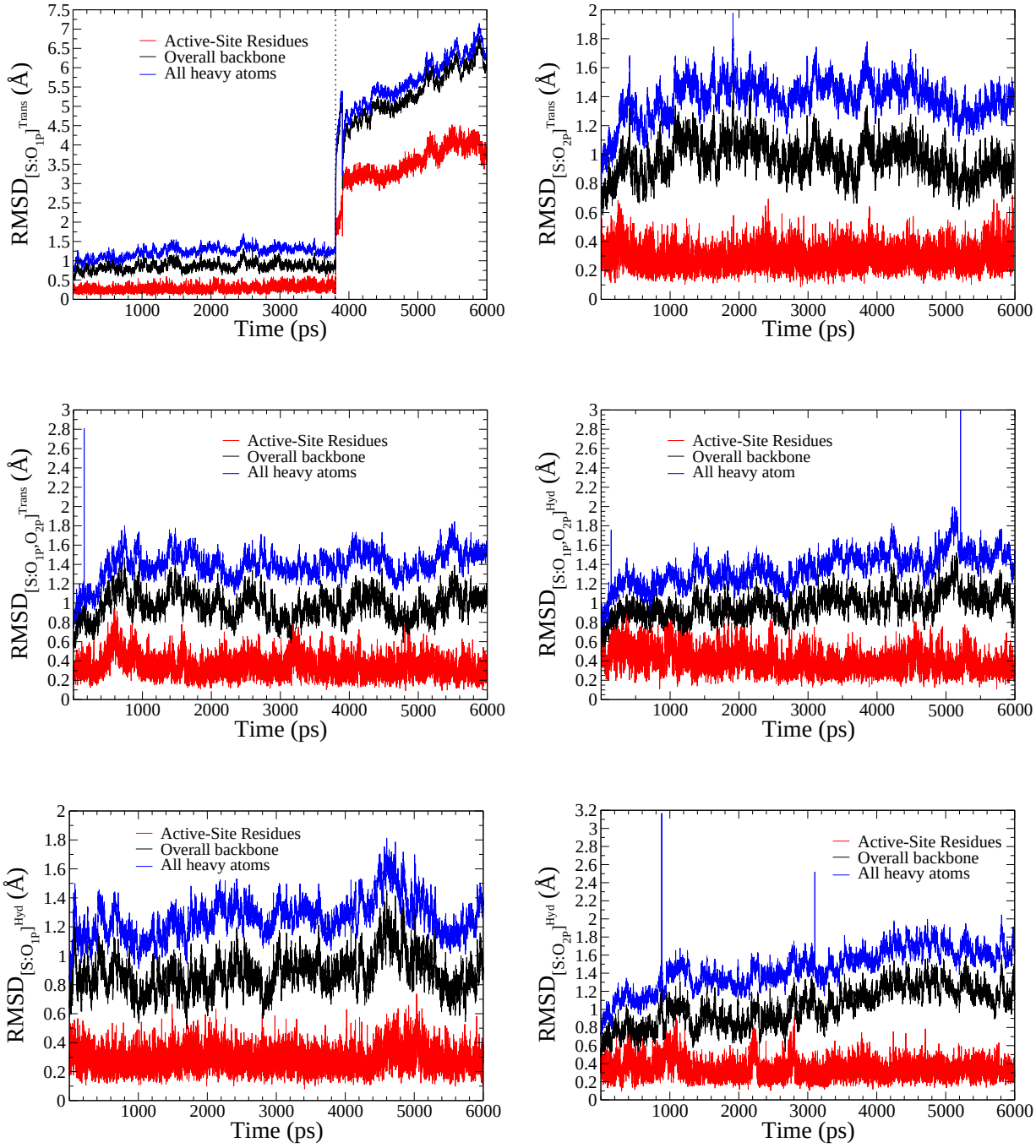


Table B.1: Averaged structural parameters of the dinucleotides in the RNase A - 3',5'-CpA transphosphorylation transition state mimic $[S{:}O_{1P}]^{Trans}$ complex for the first 3807 ps of the production time (6000ps). All angles are in degrees.

	$[S{:}O_{1P}]^{Trans}$ first 3807 ps			
Angles	Cyt	Range	Ade	Range
Glycosyl torsion				
$O_{4'}$-$C_{1'}$-N_1-C_2 (χ_C)	-159.58(anti)	128.70 $\cdots$ -127.89		
$O_{4'}$-$C_{1'}$-N_9-C_4 (χ_A)			-66.83(high anti)	-129.81 $\cdots$ -13.88
Sugar torsion				
$C_{4'}$-$O_{4'}$-$C_{1'}$-$C_{2'}$ ($\tau 0$)	3.80	-17.95 $\cdots$ 30.57	-24.91	-53.98 $\cdots$ 20.73
$O_{4'}$-$C_{1'}$-$C_{2'}$-$C_{3'}$ ($\tau 1$)	-24.49	-40.85 $\cdots$ -8.03	35.92	-39.49 $\cdots$ 55.86
$C_{1'}$-$C_{2'}$-$C_{3'}$-$C_{4'}$ ($\tau 2$)	34.53	19.45 $\cdots$ 48.16	-32.88	-56.61 $\cdots$ 48.88
$C_{2'}$-$C_{3'}$-$C_{4'}$-$O_{4'}$ ($\tau 3$)	-33.53	-50.00 $\cdots$ -6.30	19.02	-51.66 $\cdots$ 47.05
$C_{3'}$-$C_{4'}$-$O_{4'}$-$C_{1'}$ ($\tau 4$)	18.32	-12.06 $\cdots$ 38.88	3.56	-37.89 $\cdots$ 49.05
Phase	12.50 N $C_{3'}$-endo		155.85 S $C_{2'}$-endo	
Backbone torsion				
$O_{5'}$-$C_{5'}$-$C_{4'}$-$C_{3'}$ (γ)	61.55(g^+)	-179.53 $\cdots$ 179.21	79.32(g^+)	17.47 $\cdots$ -156.21
$C_{5'}$-$C_{4'}$-$C_{3'}$-$O_{3'}$ (δ)	85.26(g^+)	62.11 $\cdots$ 119.98	133.89(g^+)	52.98 $\cdots$ 168.62
Phosphodiester torsion				
$O_{3'}$-P-$O_{5'}$-$C_{5'}$ (α)			167.49(t)	-179.97 $\cdots$ 179.97
P-$O_{5'}$-$C_{5'}$-$C_{4'}$ (β)			-170.16(t)	-30.94 $\cdots$ -104.00
$C_{4'}$-$C_{3'}$-$O_{3'}$-P (ϵ)	101.10(g^+)	75.20 $\cdots$ 128.30		
$C_{3'}$-$O_{3'}$-P-$O_{5'}$ (ζ)	170.00(t)	144.20 $\cdots$ -163.51		

The Phase is defined as $\tan(P) = \frac{(\tau_4+\tau_1)-(\tau_3+\tau_0)}{(2*\tau_2*(\sin(36)+\sin(72))}$ If τ_2 has a negative value, then 180° is added to the value of P[230,232]

Table B.2: Averaged structural parameters of the dinucleotides in the RNase A - 3',5'-CpA transphosphorylation transition state mimic $[S{:}O_{2P}]^{Trans}$ complex for the production time (6000ps). All angles are in degrees.

	$[S{:}O_{2P}]^{Trans}$			
Angles	Cyt	Range	Ade	Range
Glycosyl torsion				
$O_{4'}$-$C_{1'}$-N_1-C_2 (χ_C)	-155.59(anti)	177.55 $\cdots$ -120.92		
$O_{4'}$-$C_{1'}$-N_9-C_4 (χ_A)			-67.25(high anti)	-123.16 $\cdots$ -13.14
Sugar torsion				
$C_{4'}$-$O_{4'}$-$C_{1'}$-$C_{2'}$ ($\tau 0$)	4.64	-17.26 $\cdots$ 26.58	-29.70	-54.60 $\cdots$ 20.23
$O_{4'}$-$C_{1'}$-$C_{2'}$-$C_{3'}$ ($\tau 1$)	-25.66	-41.84 $\cdots$ -6.12	34.18	-36.76 $\cdots$ 60.81
$C_{1'}$-$C_{2'}$-$C_{3'}$-$C_{4'}$ ($\tau 2$)	35.56	19.49 $\cdots$ 48.06	-25.56	-52.27 $\cdots$ 49.97
$C_{2'}$-$C_{3'}$-$C_{4'}$-$O_{4'}$ ($\tau 3$)	-34.19	-50.46 $\cdots$ -13.04	8.45	-53.45 $\cdots$ 43.29
$C_{3'}$-$C_{4'}$-$O_{4'}$-$C_{1'}$ ($\tau 4$)	18.18	-7.78 $\cdots$ 41.17	13.32	-22.93 $\cdots$ 49.33
Phase	11.40 N $C_{3'}$-endo		130.03 S $C_{2'}$-endo	
Backbone torsion				
$O_{5'}$-$C_{5'}$-$C_{4'}$-$C_{3'}$ (γ)	62.31($^+$)	-127.26 $\cdots$ -174.76	65.52($^+$)	13.42 $\cdots$ -158.59
$C_{5'}$-$C_{4'}$-$C_{3'}$-$O_{3'}$ (δ)	84.15($^+$)	62.20 $\cdots$ 106.54	121.45($^+$)	57.47 $\cdots$ 162.74
Phosphodiester torsion				
$O_{3'}$-P-$O_{5'}$-$C_{5'}$ (α)			165.61(t)	29.55 $\cdots$ -152.11
P-$O_{5'}$-$C_{5'}$-$C_{4'}$ (β)			-172.25(t)	126.71 $\cdots$ -111.15
$C_{4'}$-$C_{3'}$-$O_{3'}$-P (ϵ)	92.51($^+$)	63.25 $\cdots$ 125.45		
$C_{3'}$-$O_{3'}$-P-$O_{5'}$ (ζ)	178.16(t)	147.33 $\cdots$ -148.49		

The Phase is defined as $\tan(P) = \frac{(\tau_4+\tau_1)-(\tau_3+\tau_0)}{(2*\tau_2*(\sin(36)+\sin(72))}$ If τ_2 has a negative value, then 180° is added to the value of P[230,232]

Table B.3: Averaged structural parameters of the dinucleotides in the RNase A - 3',5'-CpA trans-phosphorylation transition state mimic [S:O_{1P},O_{2P}]^{Trans} complex for the production time (6000ps). All angles are in degrees.

	[S:O_{1P},O_{2P}]^{Trans}			
Angles	Cyt	Range	Ade	Range
Glycosyl torsion				
$O_{4'}$-$C_{1'}$-N_1-C_2 (χ_C)	-152.76(anti)	175.87 $\cdots$ -115.80		
$O_{4'}$-$C_{1'}$-N_9-C_4 (χ_A)			-69.56(high anti)	-140.43 $\cdots$ -14.10
Sugar torsion				
$C_{4'}$-$O_{4'}$-$C_{1'}$-$C_{2'}$ ($\tau 0$)	5.92	-16.42 $\cdots$ 24.24	-24.23	-54.25 $\cdots$ 23.68
$O_{4'}$-$C_{1'}$-$C_{2'}$-$C_{3'}$ ($\tau 1$)	-26.65	-42-22 $\cdots$ -9.72	33.59	-38.14 $\cdots$ 59.81
$C_{1'}$-$C_{2'}$-$C_{3'}$-$C_{4'}$ ($\tau 2$)	35.92	21.34 $\cdots$ 47.90	-29.78	-54.18 $\cdots$ 48.78
$C_{2'}$-$C_{3'}$-$C_{4'}$-$O_{4'}$ ($\tau 3$)	-33.75	-54.95 $\cdots$ -15.49	15.95	-55.09 $\cdots$ 45.72
$C_{3'}$-$C_{4'}$-$O_{4'}$-$C_{1'}$ ($\tau 4$)	17.13	-3.61 $\cdots$ 39.43	5.29	-29.14 $\cdots$ 48.59
Phase	9.40 N $C_{3'}$-endo		152.77 S $C_{2'}$-endo	
Backbone torsion				
$O_{5'}$-$C_{5'}$-$C_{4'}$-$C_{3'}$ (γ)	61.08(g^+)	-179.97 $\cdots$ 179.98	150.68(t)	9.52 $\cdots$ -147.17
$C_{5'}$-$C_{4'}$-$C_{3'}$-$O_{3'}$ (δ)	84.3(g^+)	61.71 $\cdots$ 107.68	130.79(g^+)	51.74 $\cdots$ 165.99
Phosphodiester torsion				
$O_{3'}$-P-$O_{5'}$-$C_{5'}$ (α)			82.15(g^+)	24.87 $\cdots$ -153.07
P-$O_{5'}$-$C_{5'}$-$C_{4'}$ (β)			175.21(t)	123.99 $\cdots$ -120.52
$C_{4'}$-$C_{3'}$-$O_{3'}$-P (ϵ)	96.64(g^+)	74.55 $\cdots$ 120.08		
$C_{3'}$-$O_{3'}$-P-$O_{5'}$ (ζ)	172.49(t)	144.30 $\cdots$ -159.62		

The Phase is defined as $\tan(P) = \frac{(\tau_4+\tau_1)-(\tau_3+\tau_0)}{(2*\tau_2*(\sin(36)+\sin(72))}$ If τ_2 has a negative value, then 180° is added to the value of P[230,232]

Table B.4: Averaged structural parameters of the dinucleotides in the RNase A - 3',5'-CpA hydrolysis transition state mimic $[S{:}O_{1P}]^{Hyd}$, $[S{:}O_{2P}]^{Hyd}$ and $[S{:}O_{1P},O_{2P}]^{Hyd}$ complexes for the production time (6000ps). All angles are in degrees.

	$[S{:}O_{1P}]^{Hyd}$		$[S{:}O_{2P}]^{Hyd}$		$[S{:}O_{1P},O_{2P}]^{Hyd}$	
Angles	Cyt	Range	Cyt	Range	Cyt	Range
Glycosyl torsion						
$O_{4'}$-$C_{1'}$-N_1-C_2 (χ_C)	-159.71(anti)	169.02 $\cdots$ -130.67	-153.44(anti)	162.78 $\cdots$ -86.16	-156.43(anti)	177.58 $\cdots$ -123.73
Sugar torsion						
$C_{4'}$-$O_{4'}$-$C_{1'}$-$C_{2'}$ ($\tau 0$)	2.62	-23.50 $\cdots$ -29.79	-8.01	-56.71 $\cdots$ 27.59	6.54	-21.73 $\cdots$ 28.06
$O_{4'}$-$C_{1'}$-$C_{2'}$-$C_{3'}$ ($\tau 1$)	-24.35	-40.52 $\cdots$ -3.69	-14.33	-37.20 $\cdots$ 43.72	-27.44	-43-13 $\cdots$ -1.52
$C_{1'}$-$C_{2'}$-$C_{3'}$-$C_{4'}$ ($\tau 2$)	35.42	18.25 $\cdots$ 47.51	29.66	-37.57 $\cdots$ 45.85	36.62	16.50 $\cdots$ 48.49
$C_{2'}$-$C_{3'}$-$C_{4'}$-$O_{4'}$ ($\tau 3$)	-34.99	-52.47 $\cdots$ -12.76	-34.87	-55.50 $\cdots$ 32.06	-34.09	-50.02 $\cdots$ -15.05
$C_{3'}$-$C_{4'}$-$O_{4'}$-$C_{1'}$ ($\tau 4$)	20.03	-9.95 $\cdots$ 46.05	26.99	-22.26 $\cdots$ 57.41	16.95	-5.56 $\cdots$ 40.76
Phase	14.43 N $C_{3'}$-endo		31.32 N $C_{3'}$-endo		8.61 N $C_{3'}$-endo	
Backbone torsion						
$O_{5'}$-$C_{5'}$-$C_{4'}$-$C_{3'}$ (γ)	64.34(g^+)	-179.64 $\cdots$ 179.59	61.16(g^+)	-179.84 $\cdots$ 179.53	61.26(g^+)	-179.62 $\cdots$ 179.90
$C_{5'}$-$C_{4'}$-$C_{3'}$-$O_{3'}$ (δ)	84.60(g^+)	59.44 $\cdots$ 106.30	85.77(g^+)	61.68 $\cdots$ 156.36	84.04(g^+)	62.64 $\cdots$ 110.54
Phosphodiester torsion						
$C_{4'}$-$C_{3'}$-$O_{3'}$-P (ϵ)	104.75(g^+)	81.28 $\cdots$ 130.55	85.29(g^+)	61.88 $\cdots$ 165.50	96.63(g^+)	70.04 $\cdots$ 122.24
$C_{3'}$-$O_{3'}$-P-$O_{5'}$ (ζ)	164.77(t)	138.80 $\cdots$ -166.38	-168.33(t)	130.99 $\cdots$ -130.11	173.69(t)	148.13 $\cdots$ -154.31

The Phase is defined as $\tan(P) = \frac{(\tau_4+\tau_1)-(\tau_3+\tau_0)}{(2*\tau_2*(\sin(36)+\sin(72))}$ If τ_2 has a negative value, then 180° is added to the value of P[230, 232]

Table B.5: Averaged structural parameters of the dinucleotides in the RNase A - 3',5'-CpA transphosphorylation transition state mimic $[S{:}O_{1P}]^{Trans}$ complex for the last 2193 ps of the production time (6000ps). All angles are in degrees.

	$[S{:}O_{1P}]^{Trans}$ last 2193 ps			
Angles	Cyt	Range	Ade	Range
Glycosyl torsion				
$O_{4'}$-$C_{1'}$-N_1-C_2 (χ_C)	40.03(syn)	7.34 $\cdots$ 126.40		
$O_{4'}$-$C_{1'}$-N_9-C_4 (χ_A)			-64.59(high anti)	-120.43 $\cdots$ -3.74
Sugar torsion				
$C_{4'}$-$O_{4'}$-$C_{1'}$-$C_{2'}$ ($\tau 0$)	-6.25	-50.02 $\cdots$ 18.88	-19.18	-55.69 $\cdots$ 16.23
$O_{4'}$-$C_{1'}$-$C_{2'}$-$C_{3'}$ ($\tau 1$)	-15.53	-38.90 $\cdots$ 39.43	31.53	-32.51 $\cdots$ 58.05
$C_{1'}$-$C_{2'}$-$C_{3'}$-$C_{4'}$ ($\tau 2$)	29.86	-30.22 $\cdots$ 46.88	-31.57	-55.82 $\cdots$ 46.37
$C_{2'}$-$C_{3'}$-$C_{4'}$-$O_{4'}$ ($\tau 3$)	-34.50	-50.73 $\cdots$ 23.80	21.38	-51.53 $\cdots$ 48.27
$C_{3'}$-$C_{4'}$-$O_{4'}$-$C_{1'}$ ($\tau 4$)	25.16	-11.03 $\cdots$ 50.06	-1.43	-33.49 $\cdots$ 53.95
Phase	28.73 N $C_{3'}$-endo		163.98 S $C_{2'}$-endoo	
Backbone torsion				
$O_{5'}$-$C_{5'}$-$C_{4'}$-$C_{3'}$ (γ)	15.76(g^+)	-94.05 $\cdots$ 97.50	60.53(g^+)	1.36 $\cdots$ 104.07
$C_{5'}$-$C_{4'}$-$C_{3'}$-$O_{3'}$ (δ)	85.33(g^+)	62.88 $\cdots$ 146.67	135.19(g^+)	57.24 $\cdots$ 166.57
Phosphodiester torsion				
$O_{3'}$-P-$O_{5'}$-$C_{5'}$ (α)			-86.93(g^-)	149.15 $\cdots$ -2.16
P-$O_{5'}$-$C_{5'}$-$C_{4'}$ (β)			-156.90(t)	143.34 $\cdots$ -107.42
$C_{4'}$-$C_{3'}$-$O_{3'}$-P (ϵ)	102.56(g^+)	76.69 $\cdots$ 165.17		
$C_{3'}$-$O_{3'}$-P-$O_{5'}$ (ζ)	174.02(t)	125.82 $\cdots$ -155.24		

The Phase is defined as $\tan(P) = \frac{(\tau_4+\tau_1)-(\tau_3+\tau_0)}{(2*\tau_2*(\sin(36)+\sin(72))}$ If τ_2 has a negative value, then 180° is added to the value of P[230]

Figure B.2: χ dihedral angles of all the studied models, plotted as a function of the simulation time. On the first row $[S{:}O_{1P}]^{Trans}$ and $[S{:}O_{2P}]^{Trans}$ model's χ dihedral angles are depicted. In the middle, $[S{:}O_{1P},O_{2P}]^{Trans}$ and $[S{:}O_{1P},O_{2P}]^{Hyd}$ dithio models are shown. Finally, in the last row $[S{:}O_{1P}]^{Hyd}$ and $[S{:}O_{2P}]^{Hyd}$ models are depicted.

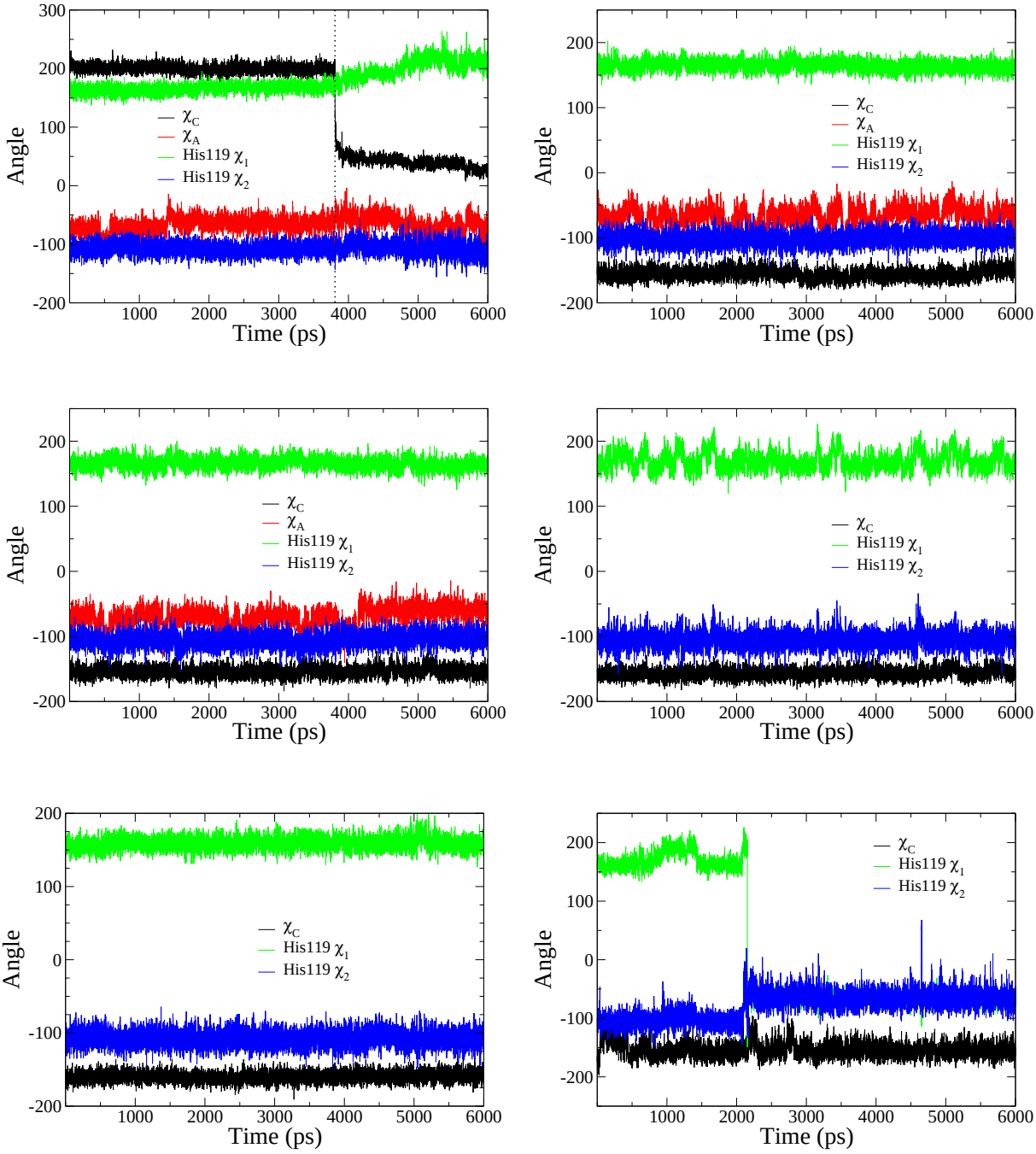
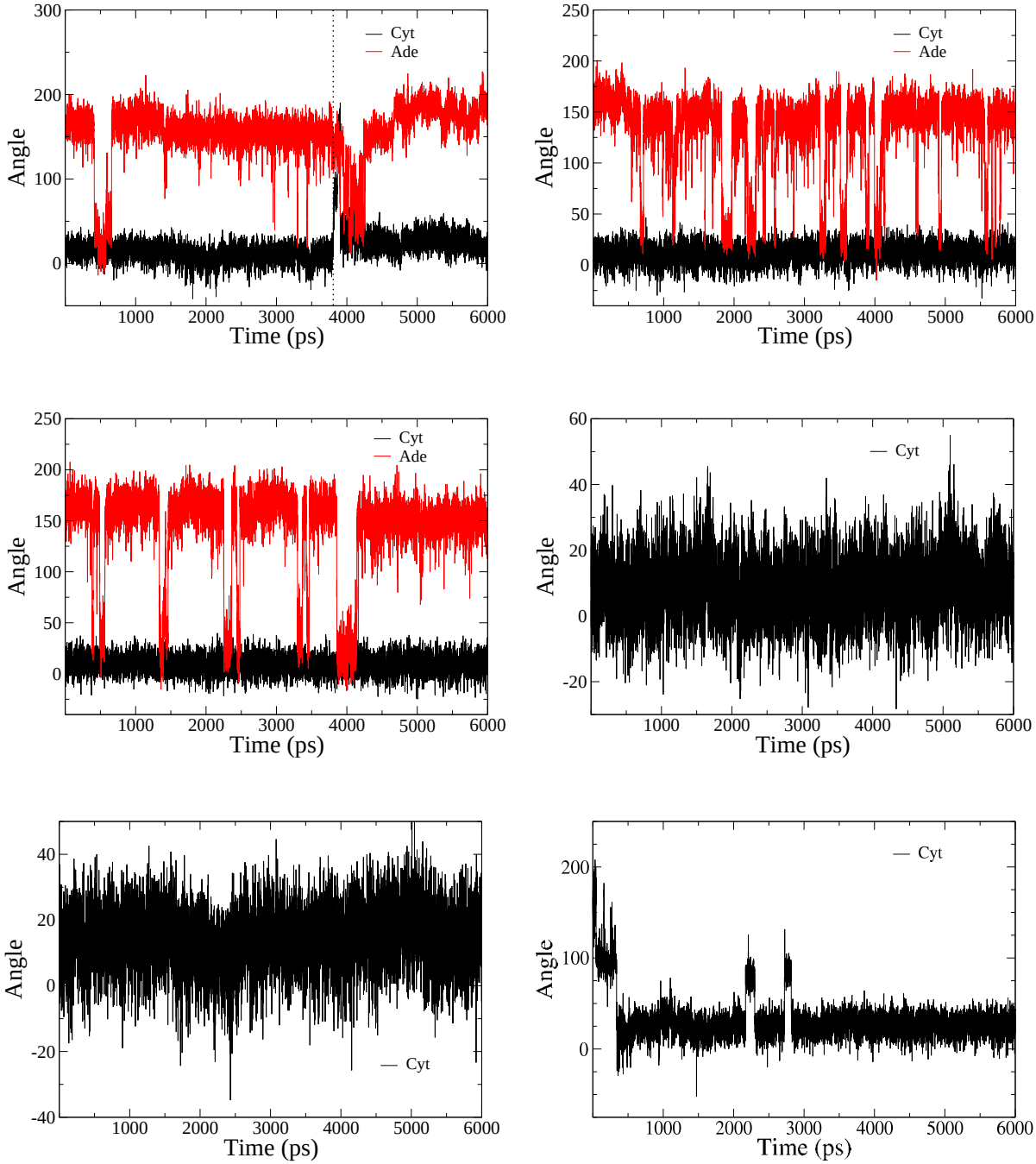


Figure B.3: Phase of all the studied models, plotted as a function of the simulation time. On the first row $[S{:}O_{1P}]^{Trans}$ and $[S{:}O_{2P}]^{Trans}$ model's phase are depicted. In the middle, $[S{:}O_{1P},O_{2P}]^{Trans}$ and $[S{:}O_{1P},O_{2P}]^{Hyd}$ dithio models are shown. Finally, in the last row $[S{:}O_{1P}]^{Hyd}$ and $[S{:}O_{2P}]^{Hyd}$ models are depicted.



Hydrogen bond interactions

The hydrogen bond formation between the ligand and the protein residues in the active site is described in Table B.6, in which the average distances between atoms participating in protein-ligand hydrogen bonds, with the hydrogen-bond occupancies, average time and number of events are listed. The time evolution of selected distances between protein and ligand atoms is plot in Figure B.6. In the simulation of the [native]$^{\text{Trans}}$ model,[243] the phosphodiester group makes strong hydrogen bond interactions with the two catalytic histidines, His12 and His119, Gln11, Phe120 and a water molecule, and somewhat weaker interactions with Lys41 and some water molecules. While, in this part of this thio substitution most of those interactions dissapear and some new one interactions appear (see Table B.6).

The equatorial O_{2P} non-bridge oxygen forms strong hydrogen bond interaction with some water molecules, in addition to hydrogen bonding with imidazolium side chain of residue His12 between $\sim$ 3907 ps and 5107 ps, $< R > = 3.733$ Å and $Occ = 0.50$ (see Figure B.6). In contrast to all studied transphosphorylation-type transition state mimic models ([native]$^{\text{Trans}}$, first 3807ps of [S:O_{1P}]$^{\text{Trans}}$ and [S:O_{2P}]$^{\text{Trans}}$), there is not any hydrogen bond interaction between residue Phe120 and O_{2P} equatorial oxygen. However, there is a weak hydrogen bond interaction between residue His119 and this equatorial oxygen ($Occ = 0.04$ and $< R > = 4.160$ Å), as in the other thio substituted transphosphorylation models, which only occurs between 3807 ps and 3906 ps. The conformational change of cytidine glycosyl dihedral angle from anti to syn conformation and of γ_C from gauche$^+$ to gauche$^-$, induce a structural change in the protein. Residue Phe120 is displaced from the vecinity of O_{2P} equatorial oxygen, while some other residues get closer to the active site, i.e. Ser59, Thr45 (see Figure B.6 and Table B.6). On the other hand, equatorial S_{1P} atom makes strong hydrogen bond interaction with some water molecules. Surprisingly, there is only one residue which forms hydrogen bond interactions with this atom, residue Thr45 with an average distance of 3.896 Å and an occupancy of 0.21 (see Table B.6). Neither Lys41 or Gln11 are near S_{1P} equatorial atom in this part of the simulation. $O_{2'}$ axial oxygen also presents a new weak hydrogen bond interaction with residue Ser59, while noone of the residues present in the other transphosphorylation models appear. This new interaction only occurs between $\sim$ 3807 ps and 3900 ps, and has an average distance of 4.925 Å with an occupancy of 0.04 (see Table 5.2 and Figure B.6). In addition, O2' makes strong hydrogen bond interactions with some water molecules, instead of with only one water molecule as in the first part of [S:O_{1P}]$^{\text{Trans}}$ and [S:O_{2P}]$^{\text{Trans}}$ simulations. This maybe due to the fact that χ_C suffers a conformational change at the begining of this last part of the simulation.

Similar to [native]$^{\text{Trans}}$, [S:O_{1P}]$^{\text{Trans}}$, [S:O_{2P}]$^{\text{Trans}}$ and [S:O_{1P},O_{2P}]$^{\text{Trans}}$ models, imidazolium side chain of residue His119 forms a strong hydrogen

Table B.6: Statistics of protein-ligand and water-ligand hydrogen-bonds observed in the simulation of the $[S{:}O_{1P}]^{Trans}$ mimic model for the last 2193 ps of the production time. The criteria for the existence of hydrogen bonds are 1) a maximum H $\cdots$ A(acceptor) distance of 2.4 Å for oxygens and 3.0 Å for sulfurs, 2) a minimum $\widehat{DHA}$ of 120° and 3) a hydrogen bond lifetime $\tau \geq 5.0$ ps. The occupancy, *Occ*, is defined as the total lifetime of each hydrogen bond by the production time. The events are the number of times each hydrogen bond is formed. All distances are in Å.

	Distance		Hydrogen Bond		
Atom pair	MD†	MD‡	*Occ*	$\langle\tau\rangle$(ps)	# Events
THR 45 Hγ_1 - CYT S_{1P}	3.896(0.876)	4.734(0.384)	0.21	45.00	11.00
WATER - CYT S_{1P}			2.16	17.85	288.00
HSP 12 Hϵ_2 - CYT O_{2P}	3.733(4.547)	4.176(2.470)	0.50	149.36	8.00
HSP 119 Hδ_1 - CYT O_{2P}	4.160(0.288)	5.099(0.277)	0.04	95.00	1.00
WATER - CYT O_{2P}			0.91	60.52	57.00
SER 59 Hγ_1 - CYT $O_{2'}$	4.925(0.940)	5.578(0.997)	0.04	95.00	1.00
WATER - CYT $O_{2'}$			0.99	21.08	112.00 0
HSP 119 Hδ_1 - CYT $O_{3'}$	2.503(0.169)	3.208(0.138)	0.32	21.14	36.00
WATER - CYT $O_{3'}$			0.03	17.50	4.00
HSP 119 Hδ_1 - ADE $O_{5'}$	2.097(0.089)	3.028(0.051)	0.83	50.35	39.00
WATER - ADE $O_{5'}$			0.01	11.67	3.00

† Arithmetic average hydrogen bond distance in the production simulation, 2193 ps, with square standard deviation in parentheses.
‡ Arithmetic average distance between the corresponding heavy atoms in the production trajectory, with square standard deviation in parentheses.

bond interaction with $O_{5'}$ axial oxygen ($< R >$=2.097 Å and $Occ = 0.83$) and a quite weak interaction with $O_{3'}$ equatorial oxygen ($< R >$=2.503 Å and $Occ = 0.32$), stronger than in the other simulations (see Table B.6).

Solvation of the substrate

Determination of ligand-solvent interactions in the active site is of paramount importance to reveal the role of specific water molecules and bulk solvent effects in enzymatic mechanism. We analize two aspects of ligand-solvent interactions: specific interactions with water molecule through the formation of hydrogen bonds, and statistical distribution of water molecules around key atoms of the substrate through the analysis of the corresponding radial distribution functions.

The $[S{:}O_{1P}]^{Trans}$ complex exhibits a variety of interactions between the ligand substrate and solvent (see Table B.6). Each hydrogen-bonding atom of the ligand interacts with several different water molecules in the course

of this last part of the production simulation. It is remarkable that $O_{2'}$ axial oxygen interacts with more than one water molecule, in contrast to the [native]$^{\text{Trans}}$ and [S:O_{1P},O_{2P}]$^{\text{Trans}}$ models, where $O_{2'}$ axial oxygen does not make any hydrogen bond interaction with the solvent, and in contrast to the first 3807 ps of [S:O_{1P}]$^{\text{Trans}}$ and [S:O_{2P}]$^{\text{Trans}}$ models where it interacts with only one water molecule (see above). The largest hydrogen bond group occupancy (2.16) corresponds to hydrogen bond interactions with S_{1P}, while the largest lifetime-average of solvent hydrogen bonds (60.52 ps) corresponds to hydrogen bond interactions with O_{2P} equatorial oxygen. It is remarkable the weak interaction between water molecules and $O_{3'}$ equatorial oxygen and $O_{5'}$ axial oxygen. The hydrogen bond group occupancy are 0.03 and 0.01 with 4 and 3 events respectively.

Regarding radial distribution functions (RDF), this last part of the simulation of the [S:O_{1P}]$^{\text{Trans}}$ model, reveals that non-bridging equatorial atoms and $O_{2'}$ axial oxygen's have the largest probability to find water molecules in their surrondings, between all the studied transphosphorylation-type complexes, while $O_{3'}$ and $O_{5'}$ probabilities decrease (see Figure B.4). S_{1P} equatorial and $O_{2'}$ axial atoms have a *fully-exposed solvent* interaction, while $O_{3'}$ and $O_{5'}$ resembles a *poorly-solvated* situation. Besides, $O_{2'}$ axial oxygen has the greates peak 6.12 at 2.7 Å.

Figure B.4: Radial distribution functions of water oxygens around phosphoryl atoms, for the transphosphorylation transition state mimic thio substituted models. Equatorial phosphoryl atoms are at the top, on the left corner 1P and on the right 2P. At the center $O_{2'}$ axial oxygen is on the left and P atom on the right, while at the bottom $O_{3'}$ equatorial oxygen is on the left and $O_{5'}$ axial oxygen on the right.

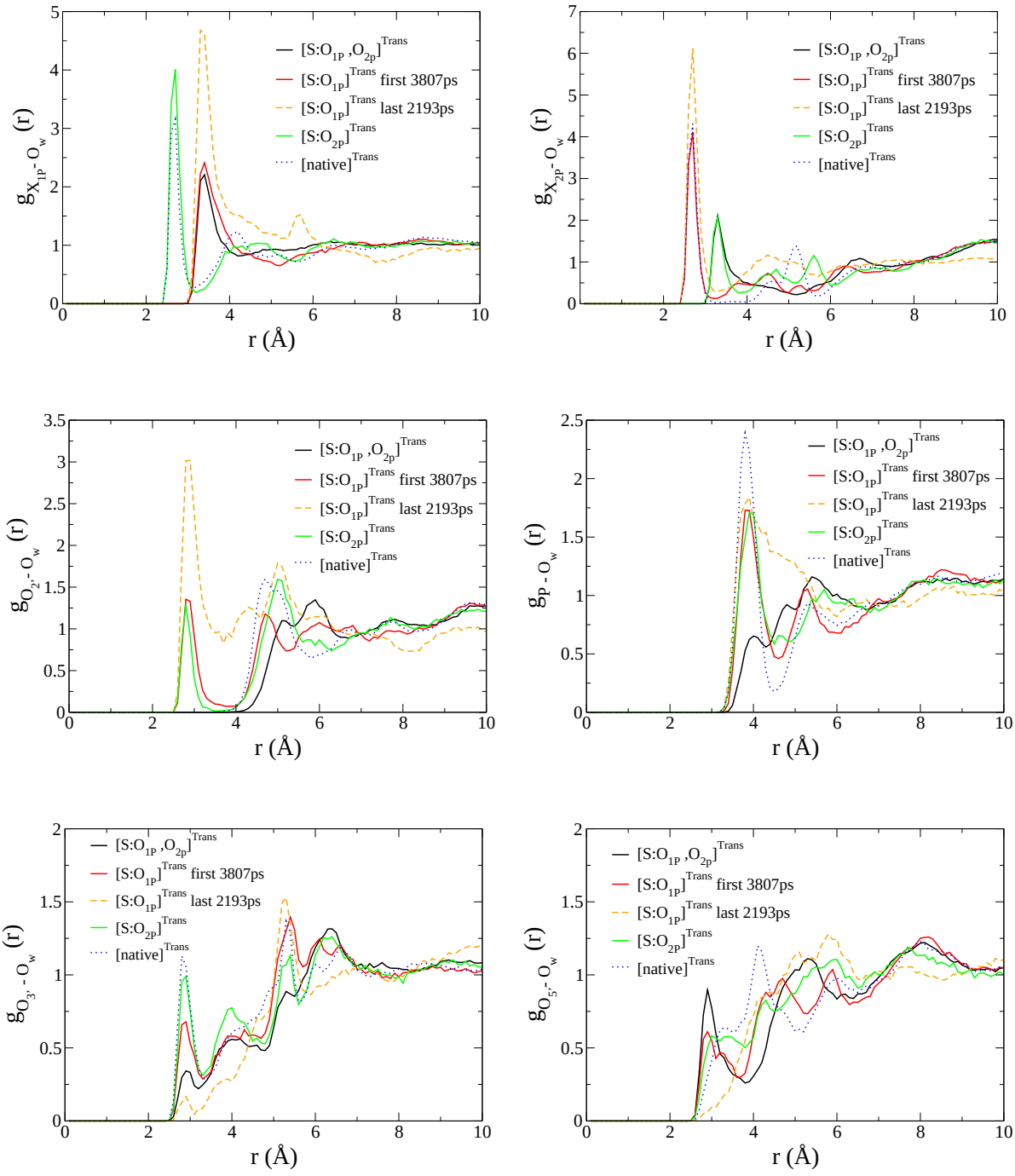


Figure B.5: Backbone torsional dihedral angles of all the transphosphorylation-type studied models, plotted as a function of the simulation time. On the first row $[S{:}O_{1P}]^{Trans}$ model's torsional angles are depicted. In the middle, $[S{:}O_{2P}]^{Trans}$ model is shown. Finally, in the last row $[S{:}O_{1P},O_{2P}]^{Trans}$ model is depicted.

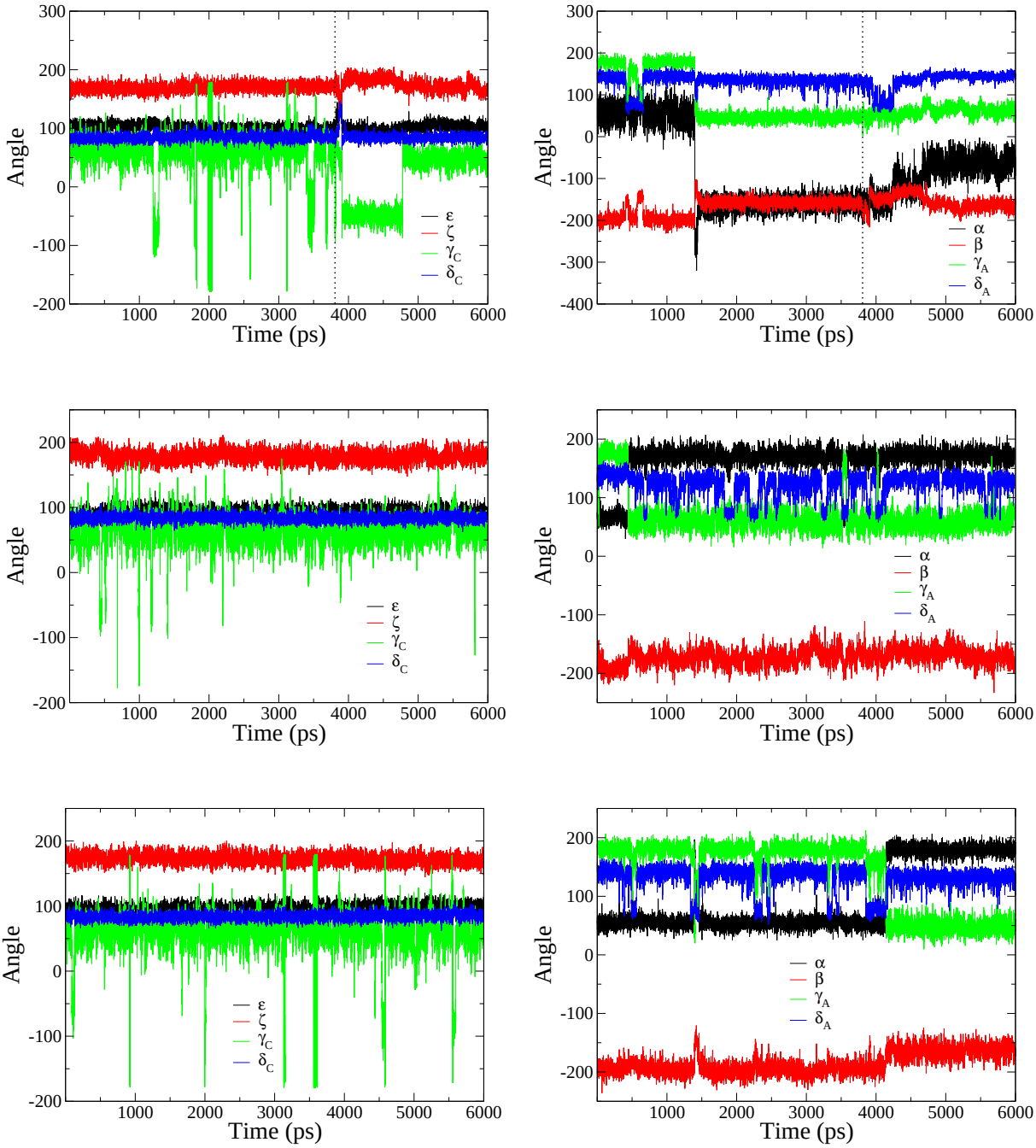


Figure B.6: Time evolution of $[S{:}O_{1P}]^{Trans}$ transphosphorylation transition state model's important distances between the ligand and the enzyme residues.

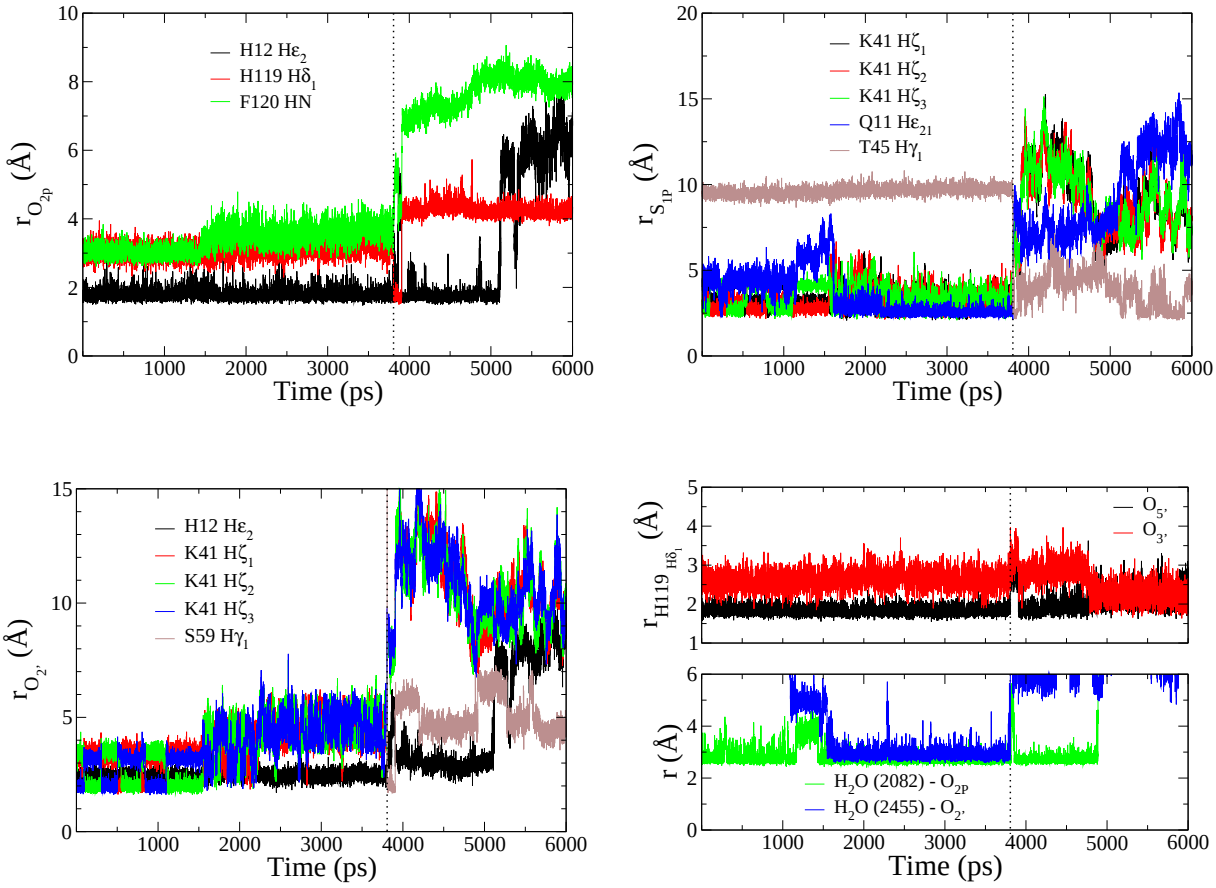


Figure B.7: Transphosphorylation transition state mimic [S:O_{2P}]^{Trans} 's important distances between the ligand and the enzyme residues, plotted as a function of simulation time.

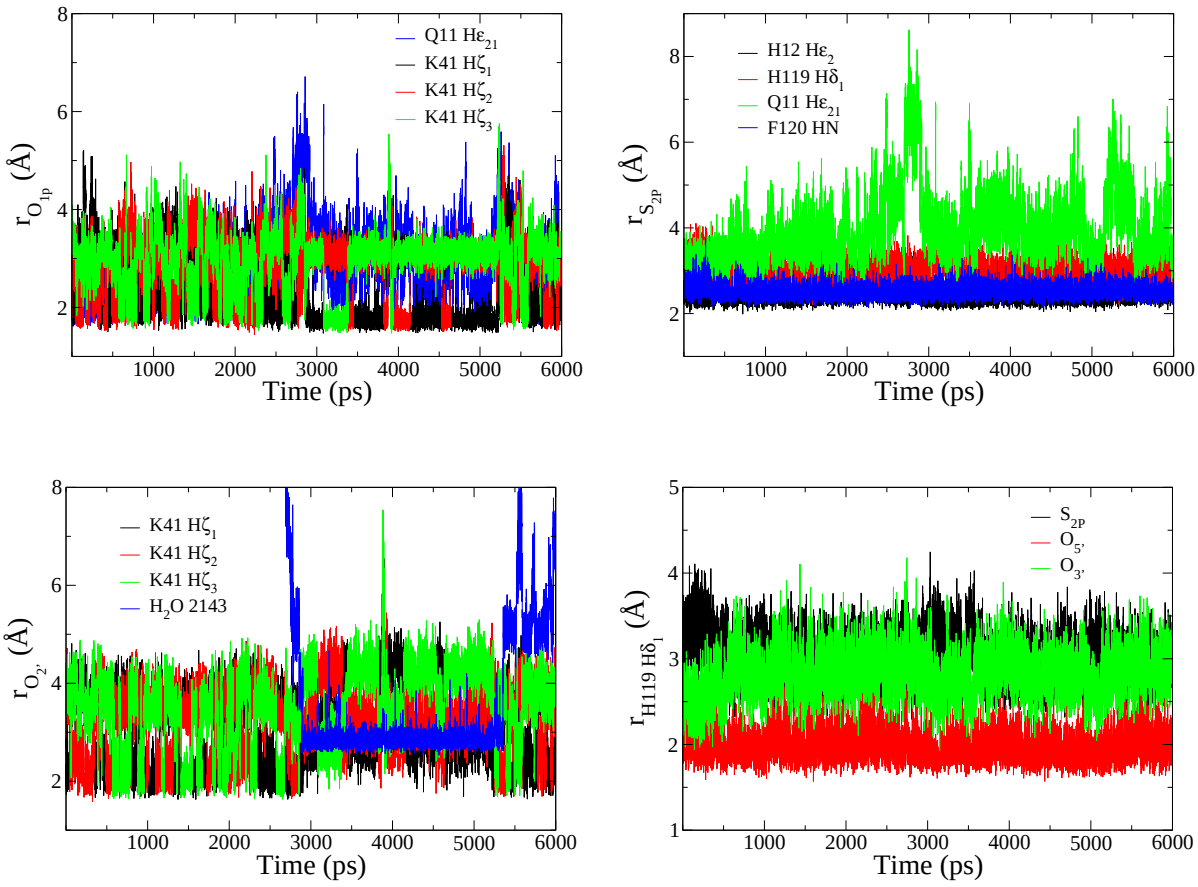


Figure B.8: $[S{:}O_{1P},O_{2P}]^{Trans}$ transphosphorylation transition state mimic's important distances between the ligand and the enzyme residues, plotted as a function of the simulation time.

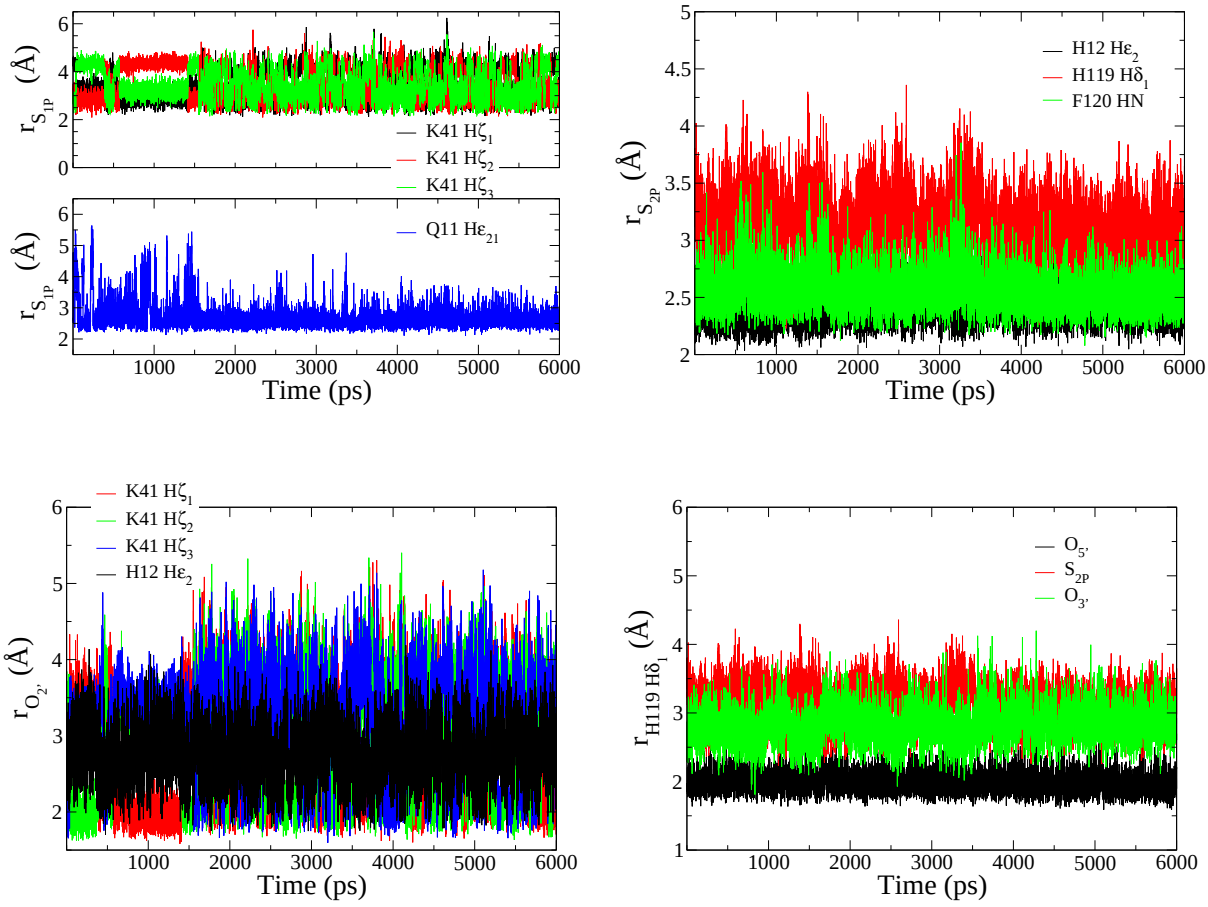


Figure B.9: Backbone torsional dihedral angles of all the hydrolysis-type studied models, plotted as a function of the simulation time. On the first row $[S{:}O_{1P}]^{Hyd}$ model's torsional angles are depicted. In the middle, $[S{:}O_{2P}]^{Hyd}$ model is shown. Finally, in the last row $[S{:}O_{1P},O_{2P}]^{Hyd}$ model is depicted.

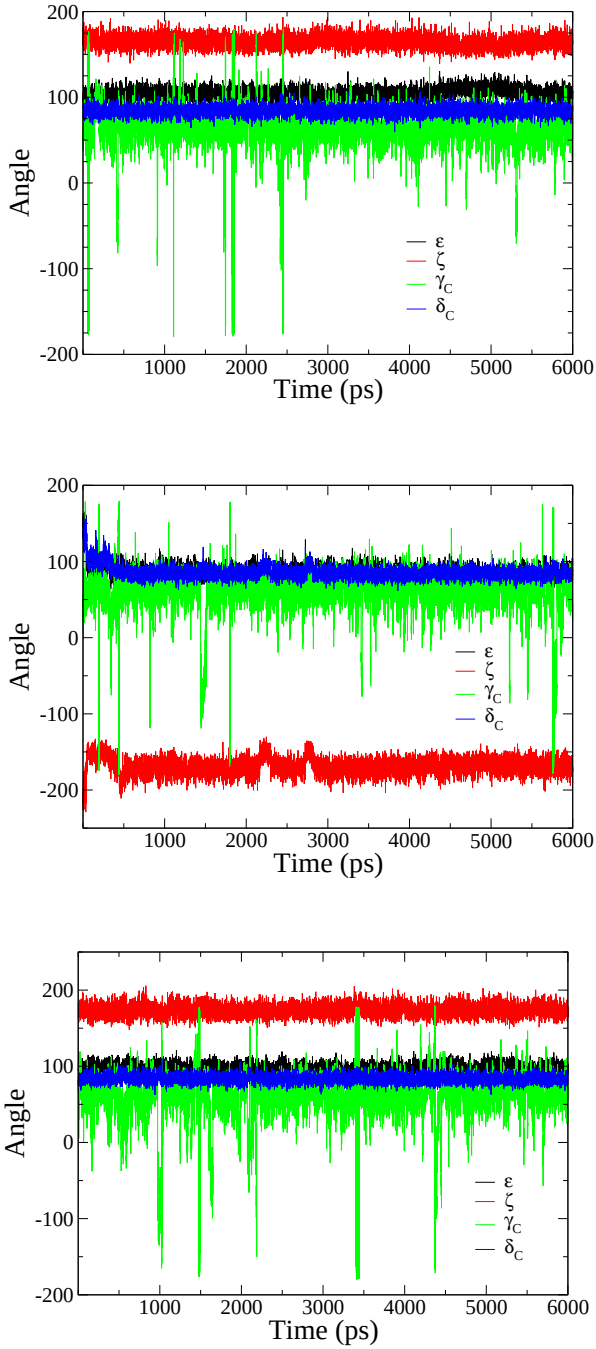


Figure B.10: Time evolution of $[S{:}O_{1P}]^{Hyd}$ hydrolysis transition state mimic important distances between the ligand and the enzyme residues.

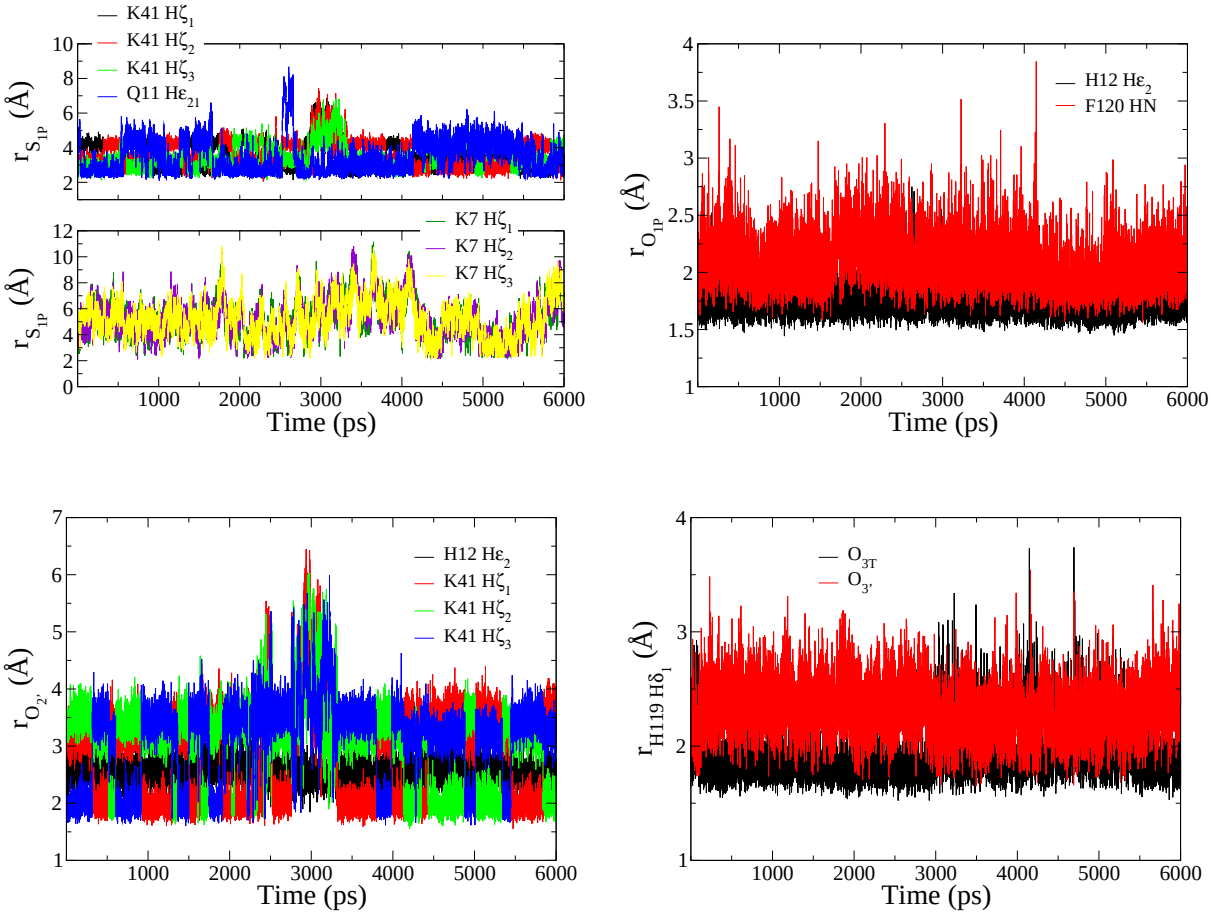


Figure B.11: Time evolution of [S:O_{2P}]^{Hyd} hydrolysis transition state mimic important distances between the ligand and the enzyme residues.

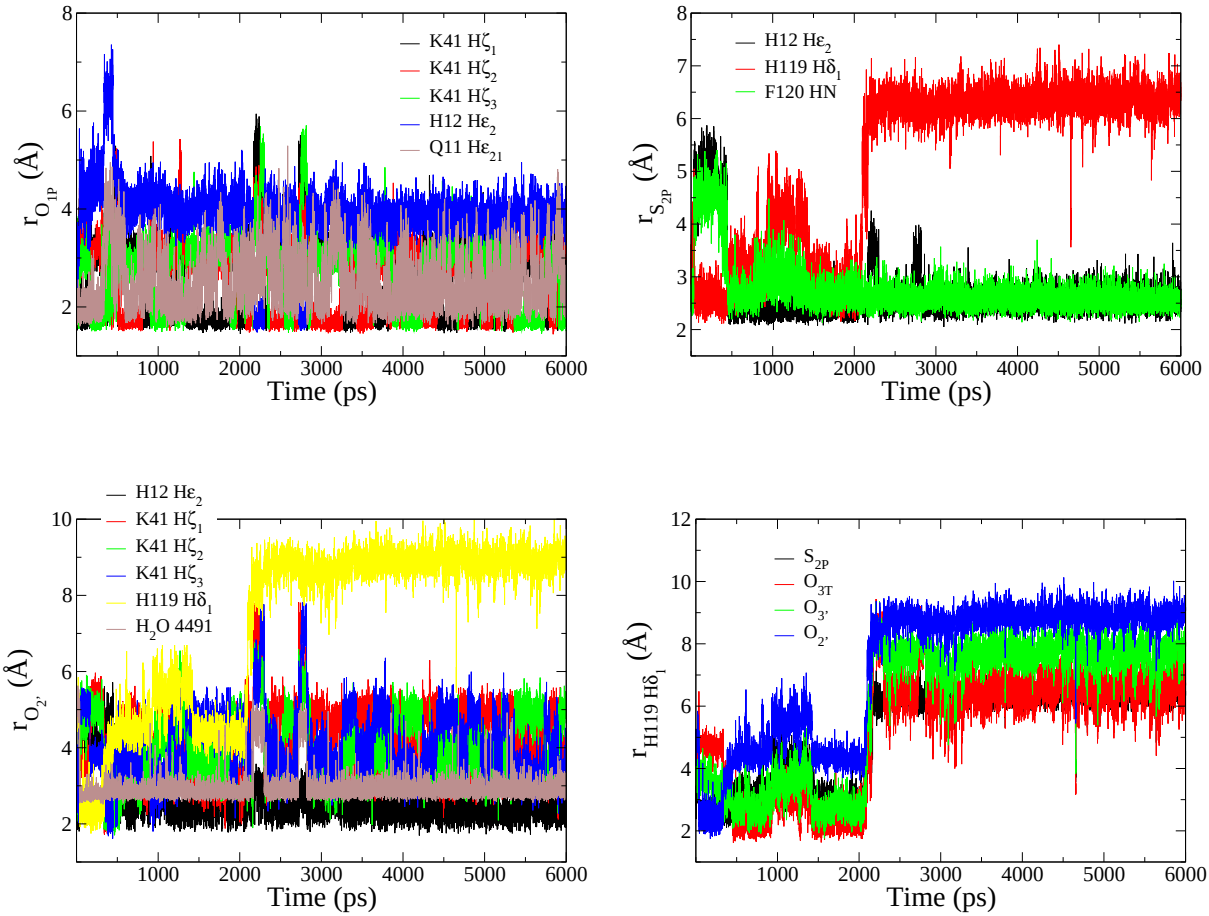
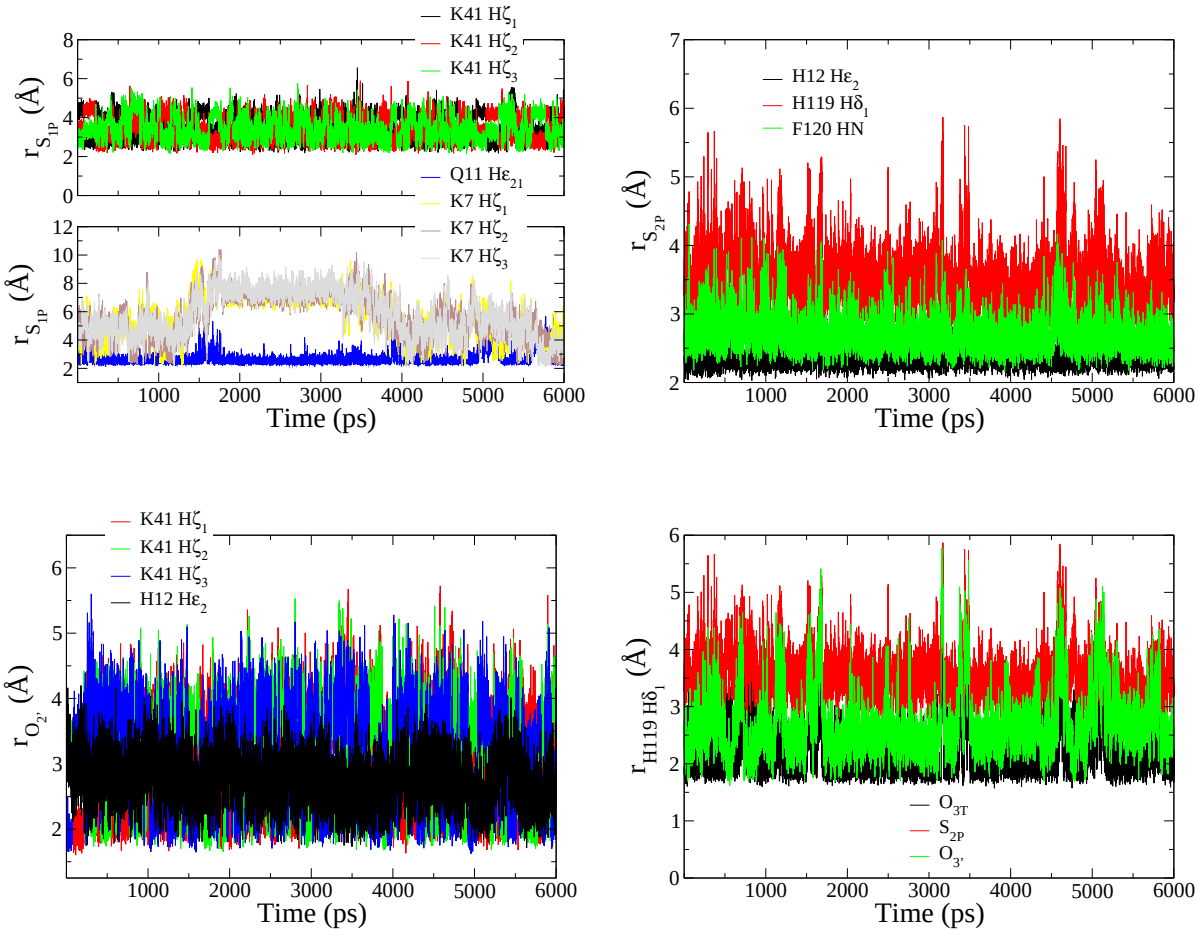


Figure B.12: Time evolution of [S:O_{1P},O_{2P}]^{Hyd} hydrolysis transition state mimic important distances between the ligand and the enzyme residues.



Chapter C

Supporting Information of Chapter 6

Averaged structural parameters of the dinucleotides, time evolution of ligand dihedral angles, phase angles, torsional dihedral angles and selected distances, and the radial distribution of waters around phosphorus atom are depicted in the Supporting Information.

Table C.1: Averaged structural parameters of the dinucleotides in the RNase A - [Trans]^{oph} and [Hyd]^{oph} complexes. All angles are in degrees.

	[Trans]^{TS}				[Hyd]^{TS}	
Angles	Cyt	Range	Ade	Range	Cyt	Range
Glycosyl torsion						
$O_{4'}$-$C_{1'}$-N_1-C_2 (χ_C)	-156.12 (anti)	174.14 $\cdots$ -124.97			154.60 (anti)	-178.40 $\cdots$ -122.66
$O_{4'}$-$C_{1'}$-N_9-C_4 (χ_A)			-72.68 (high anti)	-123.78 $\cdots$ -27.14		
Sugar torsion						
$C_{4'}$-$O_{4'}$-$C_{1'}$-$C_{2'}$ ($\tau 0$)	0.46	-26.34 $\cdots$ 21.18	-37.43	-56.11 $\cdots$ 4.88	1.75	-29.13 $\cdots$ 26.67
$O_{4'}$-$C_{1'}$-$C_{2'}$-$C_{3'}$ ($\tau 1$)	-14.76	-32.83 $\cdots$ 9.44	38.70	-28.48 $\cdots$ 58.16	-15.64	-36.32 $\cdots$ 22.36
$C_{1'}$-$C_{2'}$-$C_{3'}$-$C_{4'}$ ($\tau 2$)	23.53	-0.96 $\cdots$ 40.04	-25.99	-53.03 $\cdots$ 40.63	23.71	-15.38 $\cdots$ 41.58
$C_{2'}$-$C_{3'}$-$C_{4'}$-$O_{4'}$ ($\tau 3$)	-22.73	-41.50 $\cdots$ 2.74	5.05	-46.17 $\cdots$ 36.33	-22.22	-39.86 $\cdots$ 5.42
$C_{3'}$-$C_{4'}$-$O_{4'}$-$C_{1'}$ ($\tau 4$)	14.38	-13.56 $\cdots$ 37.29	19.80	-23.08 $\cdots$ 49.17	13.19	-10.73 $\cdots$ 35.91
Phase	16.82 North $C_{3'}$-endo		131.35 South $C_{2'}$-endo		13.87 North $C_{3'}$-endo	
Backbone torsion						
$O_{5'}$-$C_{5'}$-$C_{4'}$-$C_{3'}$ (γ)	74.14 (g^+)	-149.43 $\cdots$ 15.76	64.86 (g^+)	11.41 $\cdots$ 111.32	67.46 (g^+)	-6.00 $\cdots$ 124.14
$C_{5'}$-$C_{4'}$-$C_{3'}$-$O_{3'}$ (δ)	97.91 (g^+)	69.69 $\cdots$ 125.97	115.70 (g^+)	66.04 $\cdots$ 157.53	97.38 (g^+)	73.43 $\cdots$ 123.53
Phosphodiester torsion						
$O_{3'}$-P-$O_{5'}$-$C_{5'}$ (α)			171.28 (t)	123.32 $\cdots$ -137.39		
P-$O_{5'}$-$C_{5'}$-$C_{4'}$ (β)			-179.12 (t)	137.54 $\cdots$ -129.94		
$C_{4'}$-$C_{3'}$-$O_{3'}$-P (ϵ)	102.24 (g^+)	74.42 $\cdots$ 140.68			109.29 (g^+)	73.45 $\cdots$ 153.72
$C_{3'}$-$O_{3'}$-P-$O_{5'}$ (ζ)	172.01 (t)	134.30 $\cdots$ -149.65			169.73 (t)	128.89 $\cdots$ -143.12

The Phase is defined as $\tan(P) = \frac{(\tau_4+\tau_1)-(\tau_3+\tau_0)}{(2*\tau_2*(\sin(36)+\sin(72))}$ If τ_2 has a negative value, then 180° is added to the value of P[230,232]

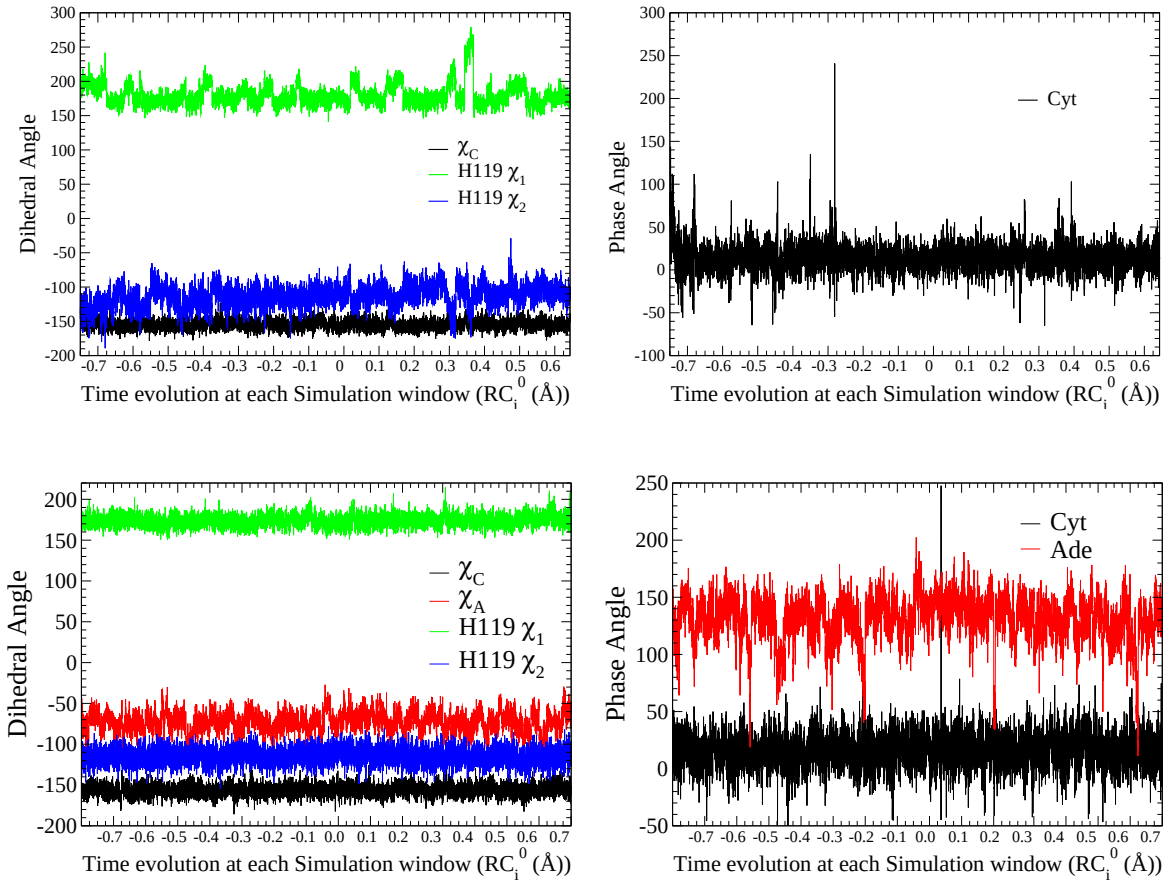
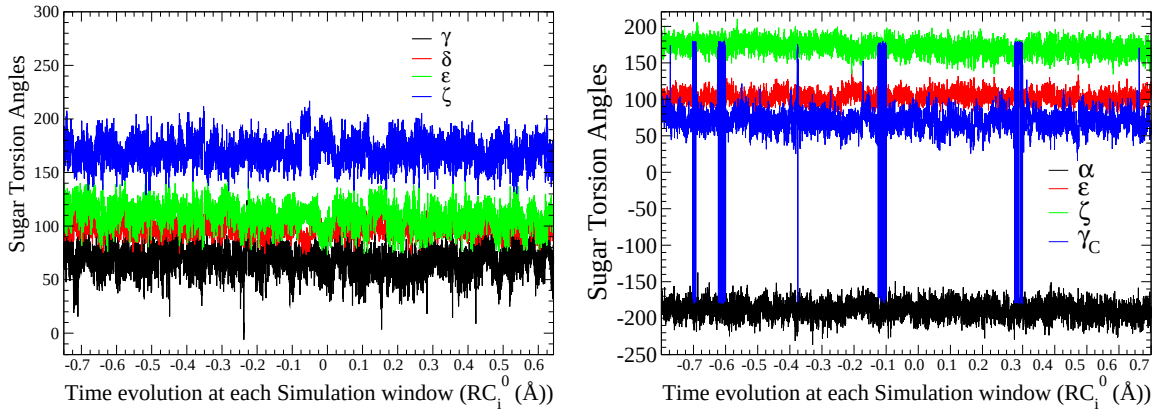


Figure C.1: Time evolution at each window of selected ligand dihedral and phase angles for the two complexes' simulations. On the left column the glycosyl dihedral angles of cytidine (χ_C in black) and adenosine (χ_A in red) with His119 χ_1 (in green) and χ_2 (in blue) are depicted, while on the right column the phase angle of cytidine (in black) and adenosine (in red) are shown. The first row correspond to the [Hyd]^{TS} model and the second one to the [Trans]^{TS} model.

Figure C.2: Phosphodiester torsional dihedral angles' time evolution at each window are shown on the left for [Hyd]^{TS} model and on the right for [Trans]^{TS} .



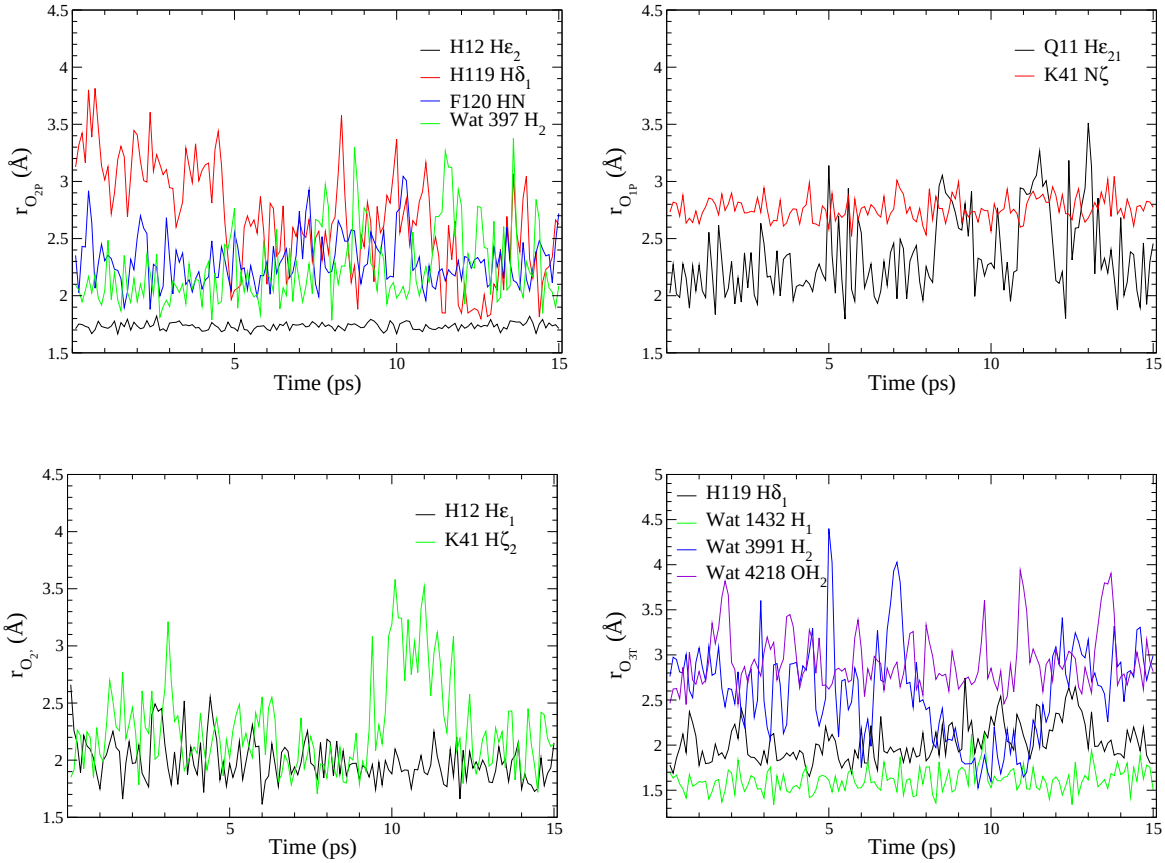


Figure C.3: Dianionic [Hyd]^{oph} model's time evolution of some distances in window centered at reaction coordinate -0.675 Å.

Figure C.4: Dianionic [Hyd]^{oph} model's time evolution of some distances in window centered at reaction coordinate 0.65 Å.

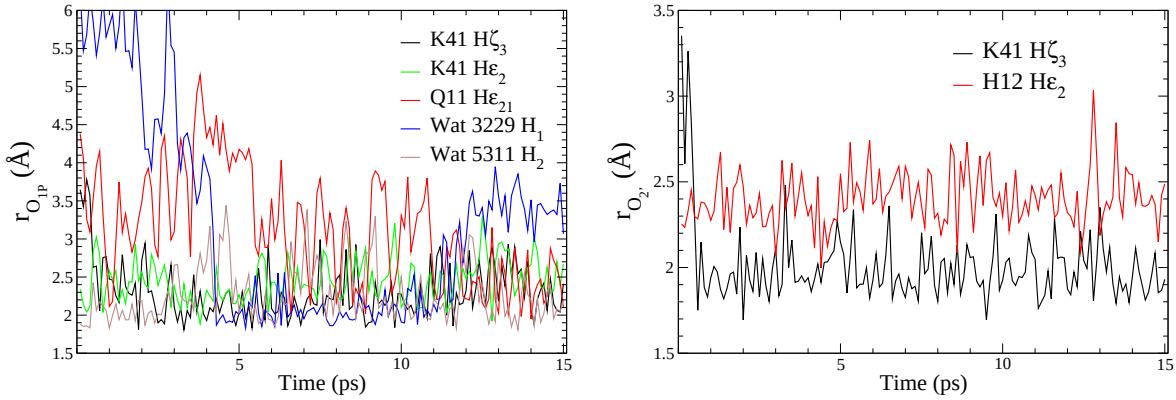


Figure C.5: Dianionic [Trans]^{oph} model's time evolution of some distances in window centered at reaction coordinate -0.725 Å.

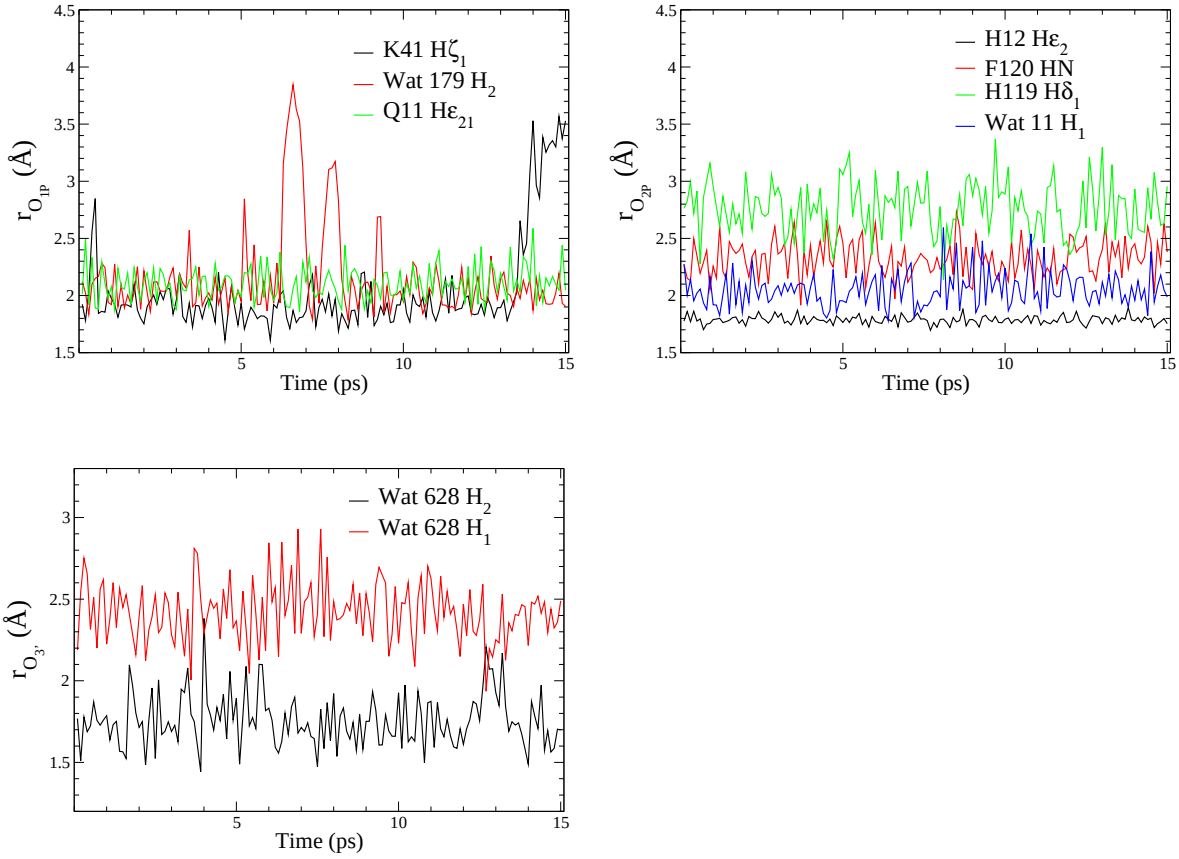


Figure C.6: Monoanionic [Trans]^{oph} model's time evolution of some distances in window centered at reaction coordinate 0.725 Å.

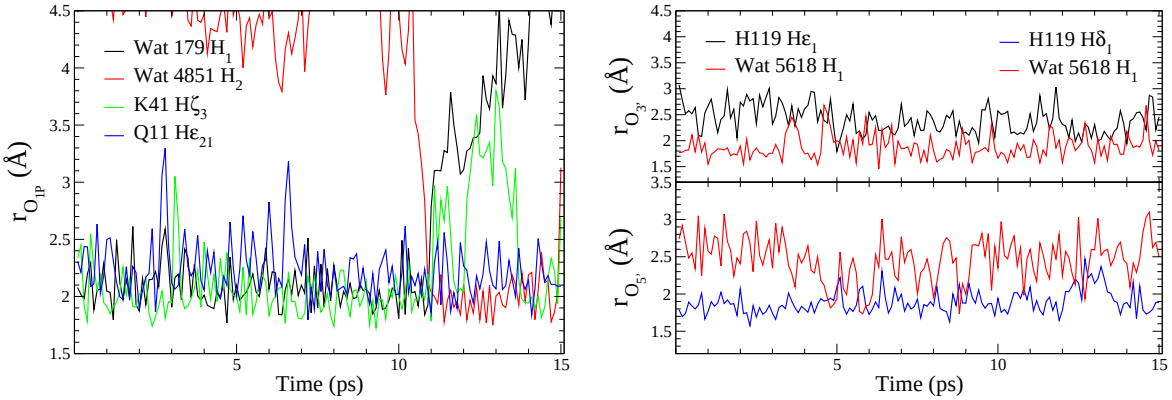
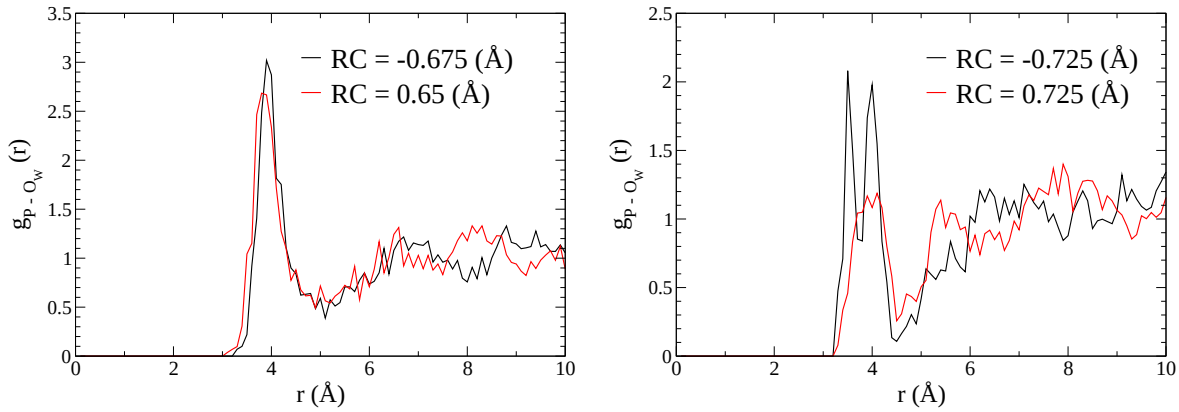


Figure C.7: Distribution of water around phosphorus atom is shown for two windows of [Hyd]^{oph} model in the left hand-side (RC = -0.675 Å in black, RC = 0.65 Å in red) and for other two windows of [Trans]^{oph} in the right hand-side (RC = -0.725 Å in black, RC = 0.725 Å in red).

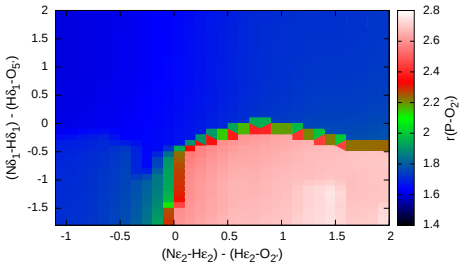


Chapter D

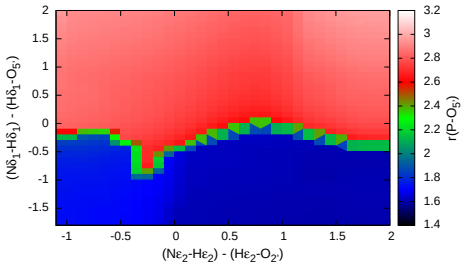
Supporting Information of Chapter 7

Representation of important distances of the studied different reaction coordinates.

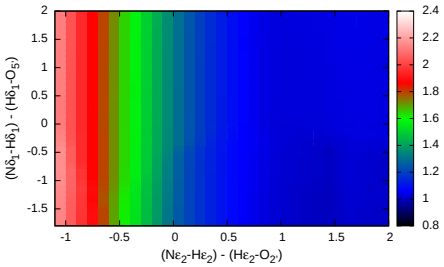
Figure D.1: Representation of some important distances of the QM part for $R_a = (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'})$ and $R_b = (N\delta_1^{His119} - H\delta_1^{His119}) - (H\delta_1^{His119} - O_{5'})$ reaction coordinates in the transphosphorylation step. All the distances are in Å.



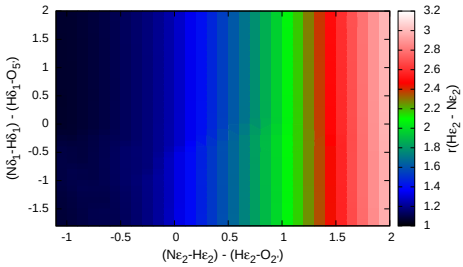
(a) P-$O_{2'}$



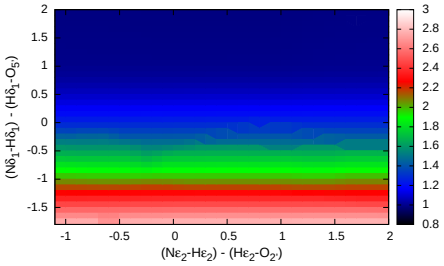
(b) P-$O_{5'}$



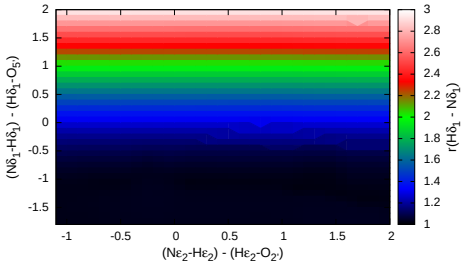
(c) Hϵ_2-$O_{2'}$



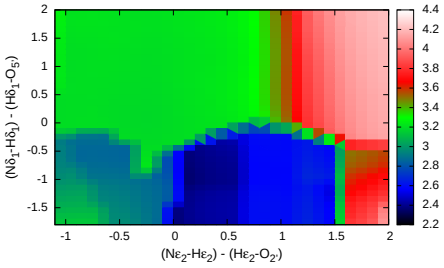
(d) Hϵ_2-Nϵ_2



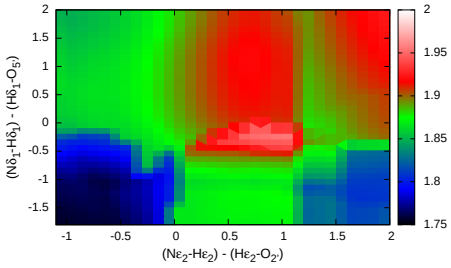
(e) Hδ_1-$O_{5'}$



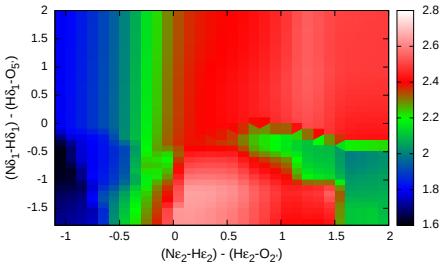
(f) Hδ_1-Nδ_1



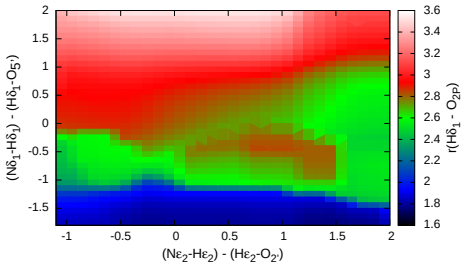
(g) Hζ-$O_{2'}$



(h) Hζ-O_{1P}

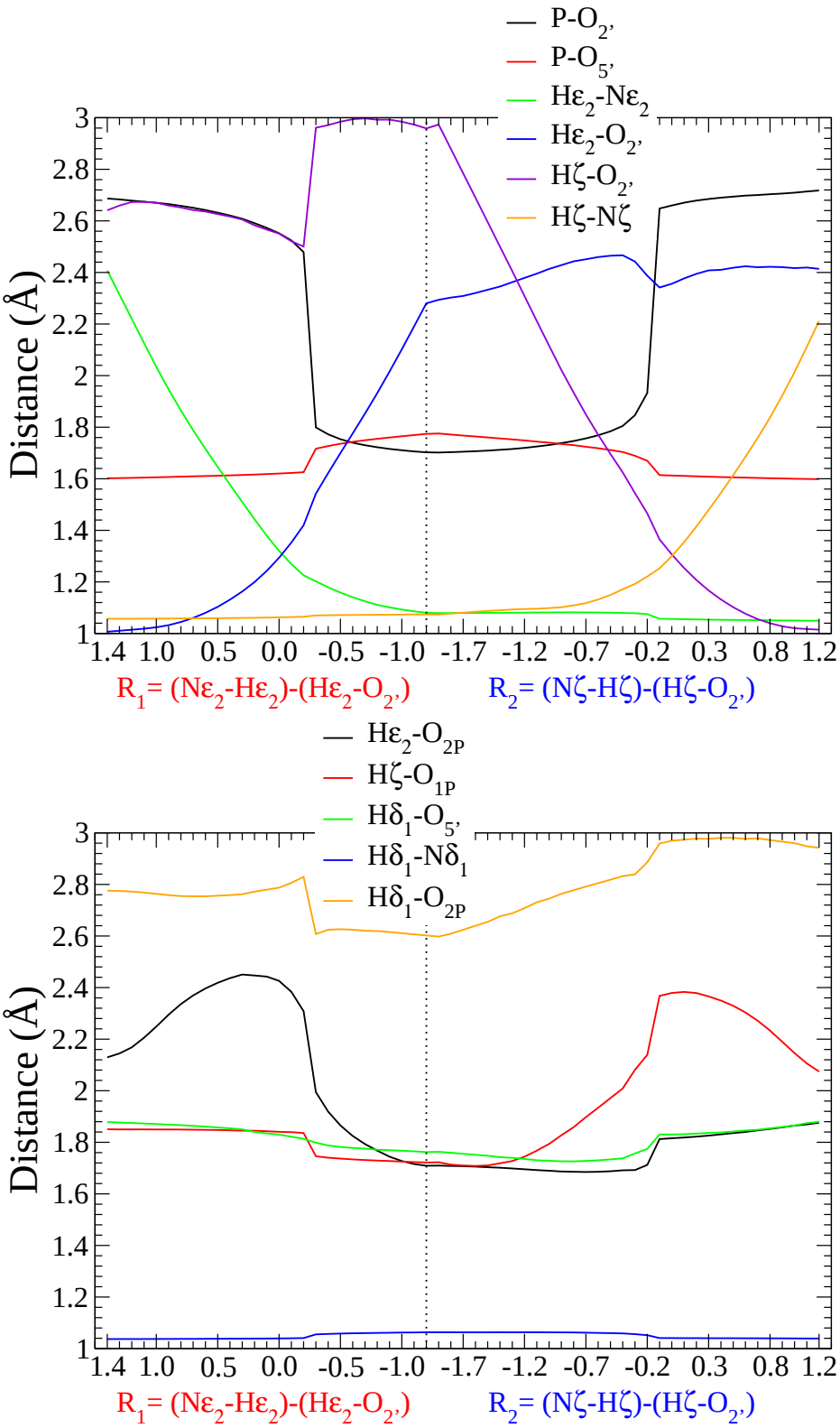


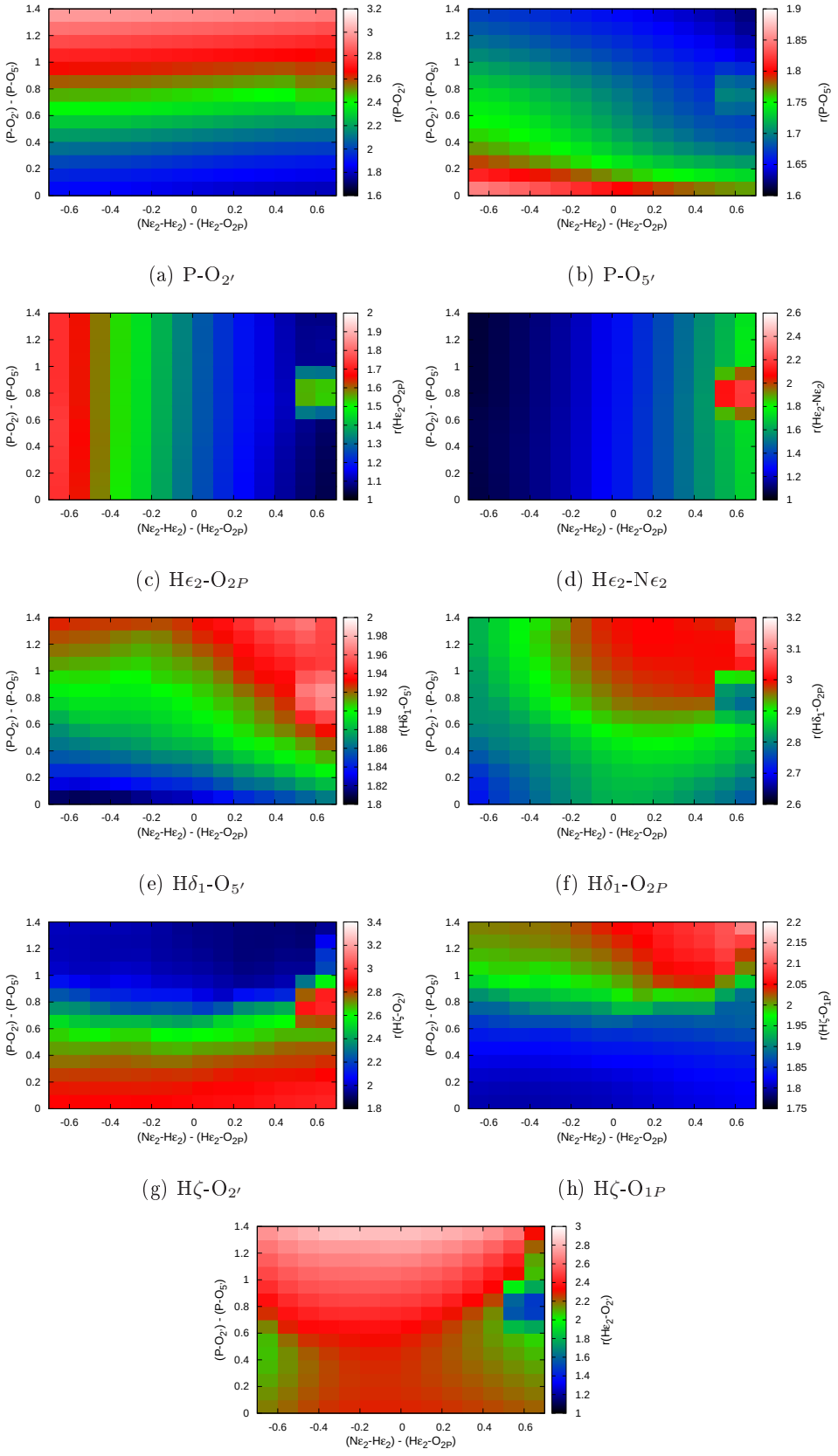
(i) Hϵ_2-O_{2P}



(j) Hδ_1-O_{2P}

Figure D.2: Some important distances of the QM part for the combination of $R_1 = (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'})$ and $R_2 = (N\zeta^{Lys41} - H\zeta^{Lys41}) - (H\zeta^{Lys41} - O_{2'})$ reaction coordinates in the transphosphorylation step. All distances are in Å.





(a) P-$O_{2'}$ (b) P-$O_{5'}$

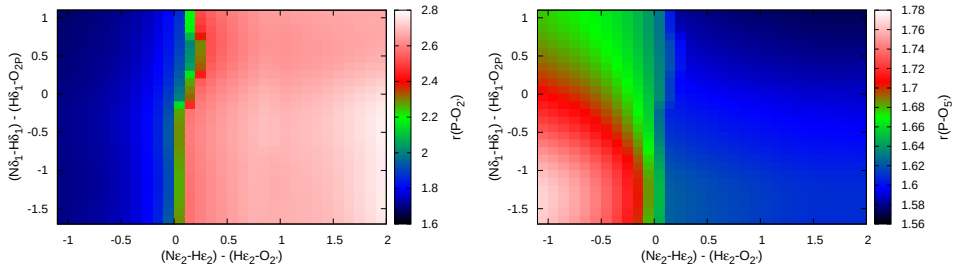
(c) Hϵ_2-O_{2P} (d) Hϵ_2-Nϵ_2

(e) Hδ_1-$O_{5'}$ (f) Hδ_1-O_{2P}

(g) Hζ-$O_{2'}$ (h) Hζ-O_{1P}

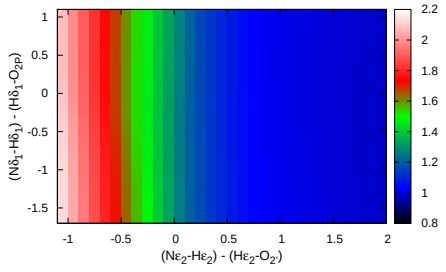
(i) Hϵ_2-$O_{2'}$

Figure D.3: Some important distances of the QM part of $R_a = (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2P})$ and $R_b = (P - O_{2'}) - (P - O_{5'})$ reaction coordinates in the transphosphorylation step. All distances are in Å.

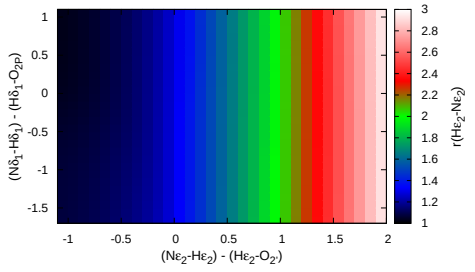


(a) P-$O_{2'}$

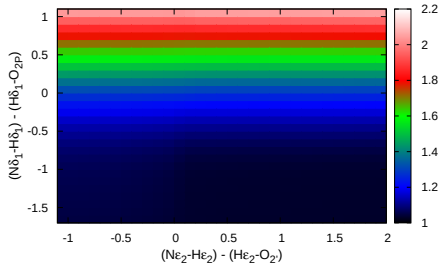
(b) P-$O_{5'}$



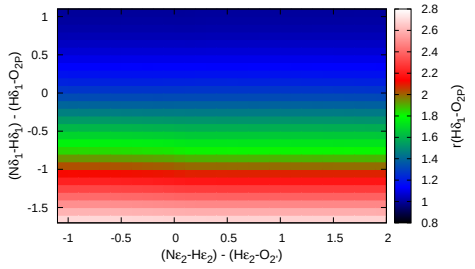
(c) Hϵ_2-$O_{2'}$



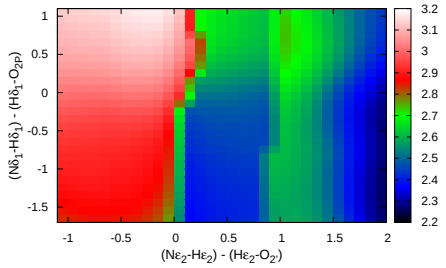
(d) Hϵ_2-Nϵ_2



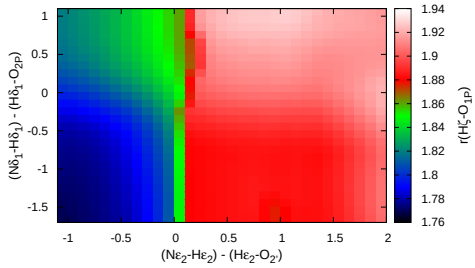
(e) Hδ_1-Nδ_1



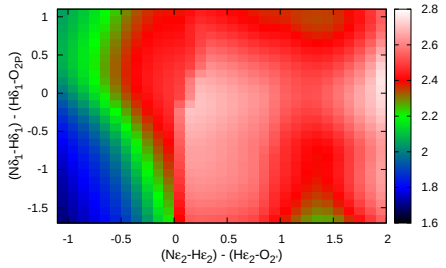
(f) Hδ_1-O_{2P}



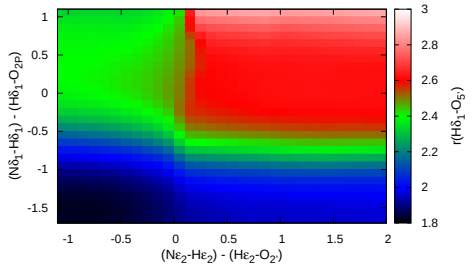
(g) Hζ-$O_{2'}$



(h) Hζ-O_{1P}



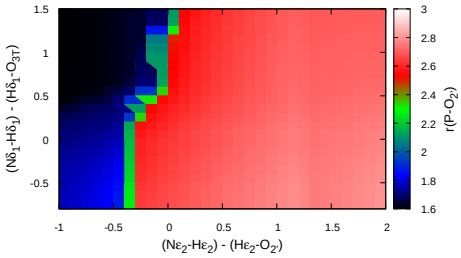
(i) Hϵ_2-O_{2P}



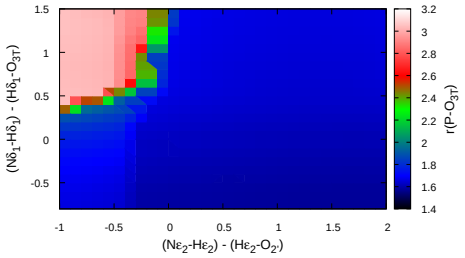
(j) Hδ_1-$O_{5'}$

Figure D.4: Some important distances of the QM part for $R_a = (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'})$ and $R_b = (N\delta_1^{His119} - H\delta_1^{His119}) - (H\delta_1^{His119} - O_{2P})$ reaction coordinates in the transphosphorylation step. All distances are in Å.

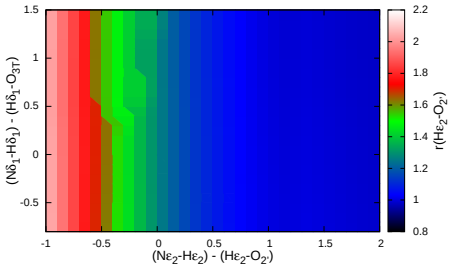
Figure D.5: Representation of some important distances of the QM part for $R_a = (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'})$ and $R_b = (N\delta_1^{His119} - H\delta_1^{His119}) - (H\delta_1^{His119} - O_{3T})$ reaction coordinates in the hydrolysis step. All the distances are in Å.



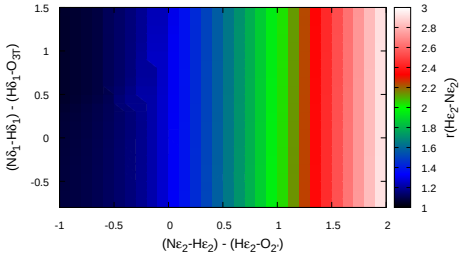
(a) P-$O_{2'}$



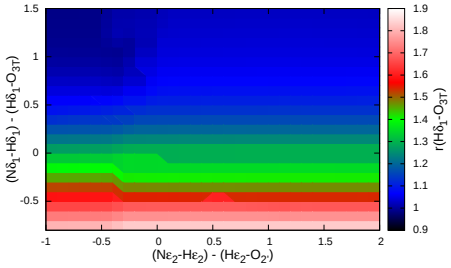
(b) P-O_{3T}



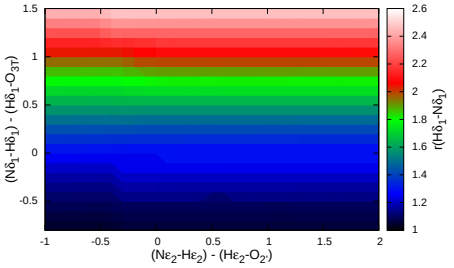
(c) Hϵ_2-$O_{2'}$



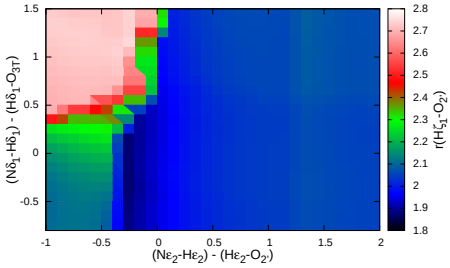
(d) Hϵ_2-Nϵ_2



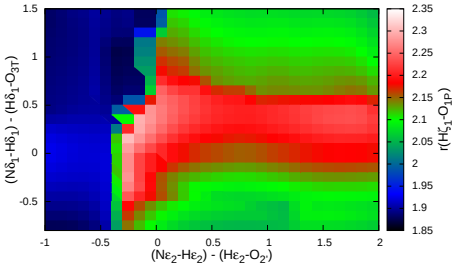
(e) Hδ_1-O_{3T}



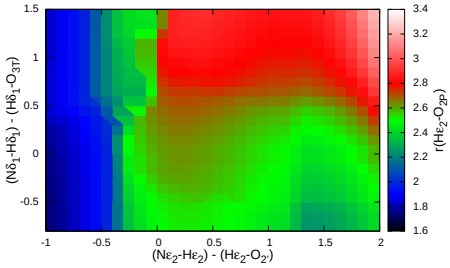
(f) Hδ_1-Nδ_1



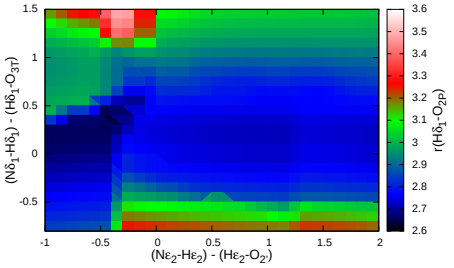
(g) Hζ-$O_{2'}$



(h) Hζ-O_{1P}

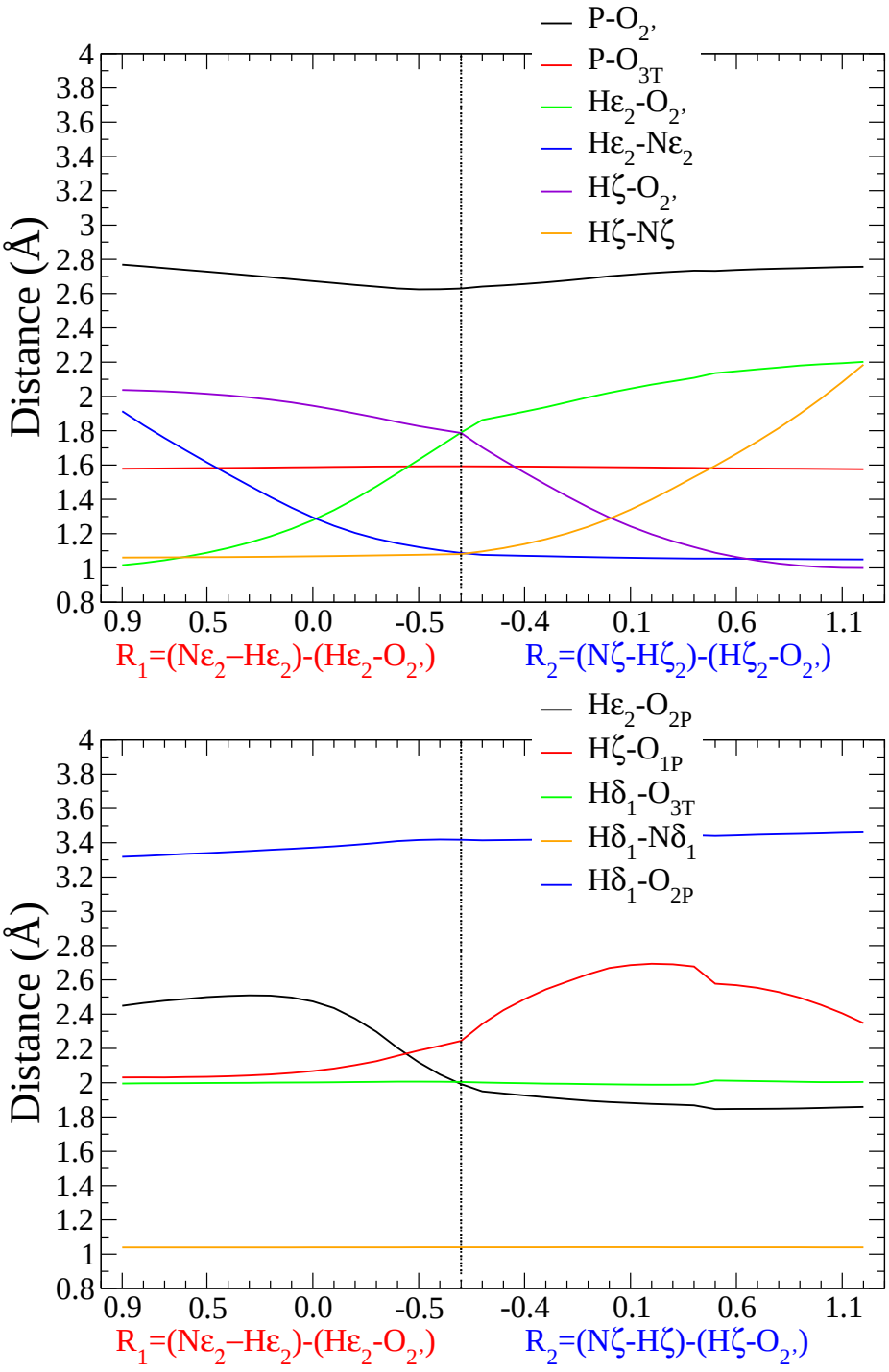


(i) Hϵ_2-O_{2P}



(j) Hδ_1-O_{2P}

Figure D.6: Some important distances of the QM part for the combination of $R_1 = (N\epsilon_2^{His12} - H\epsilon_2^{His12}) - (H\epsilon_2^{His12} - O_{2'})$ and $R_2 = (N\zeta^{Lys41} - H\zeta^{Lys41}) - (H\zeta^{Lys41} - O_{2'})$ reaction coordinates in the hydrolysis step. All distances are in Å.



References

Bibliography

(1) Gesteland, R. F., Cech, T. R., and Atkins, J. F. *The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA*. Cold Spring Harbor Laboratory Press, New York, 1999, 2 edn.

(2) Matousek, J. Ribonucleases and their antitumor activity. *Comp. Biochem. Phys. C* **129** (2001), 175–191.

(3) Leland, P. A. and Raines, R. T. Cancer chemotherapy – ribonucleases to the rescue. *Chem. Biol.* **8** (2001), 405–413.

(4) Makarov, A. and Ilinskaya, O. N. Cytotoxic ribonucleases: molecular weapons and their targets. *FEBS Lett.* **540** (2003), 15–20.

(5) Zuo, Y. and Deutscher, M. P. Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucl. Acids Res.* **29** (2001), 1017–1026.

(6) Beintema, J. J., Breukelman, H. J., Carsana, A., and Furia, A. *Ribonucleases: Structures and Functions*, Academic Press, 1997, chap. 8- Evolution of vertebrate ribonucleases: Ribonuclease A superfamily, pp. 245–269.

(7) Beintema, J. and Kleineidam, R. The Ribonuclease A superfamily: general discussion. *Cell Mol. Life Sci.* **54** (1998), 825–832.

(8) Cho, S., Beintema, J. J., and Zhang, J. The ribonuclease A superfamily of mammals and birds: identifying new members and tracing evolutionary histories. *Genomics* **85** (2005), 208–220.

(9) Pizzo, E. and D'Alessio, G. The success of the RNase scaffold in the advance of biosciences and in evolution. *Gene* **406** (2007), 8–12.

(10) Kazakou, K., Holloway, D. E., Prior, S. H., Subramanian, V., and Acharya, K. R. Ribonuclease A Homologues of the Zebrafish: Polymorphism, Crystal Structures of Two Representatives and their Evolutionary Implications. *J. Mol. Biol.* **380** (2008), 206–222.

(11) Dyer, K. D. and Rosenberg, H. F. The RNase A superfamily: Generation of diversity and innate host defense. *Mol. Diversity* **10** (2006), 585–597.

(12) Riordan, J. F. Angiogenin. *Method. Enzymol.* **341** (2001), 263–273.

(13) Fett, J. W., Strydom, D. J., Lobb, R. R., Alderman, E. M., Bethune, J. L., Riordan, J. F., and Vallee, B. L. Isolation and characterization of angiogenin, an angiogenic protein from human carcinoma cells. *Biochemistry* **24** (1985), 5480–5486.

(14) Shapiro, R. and Vallee, B. L. Site-directed mutagenesis of histidine-13 and histidine-114 of human angiogenin. Alanine derivatives inhibit angiogenin-induced angiogenesis. *Biochemistry* **28** (1989), 7401–7408.

(15) Curran, T. P., Shapiro, R., and Riordan, J. F. Alteration of the enzymic specificity of human angiogenin by site-directed mutagenesis. *Biochemistry* **32** (1993), 2307–2313.

(16) Kao, R. Y. T., Jenkins, J. L., Olson, K. A., Key, M. E., Fett, J. W., and Shapiro, R. A small-molecule inhibitor of the ribonucleolytic activity of human angiogenin that possesses antitumor activity. *Proc. Natl. Acad. Sci. U. S. A.* **99** (2002), 10066–10071.

(17) Gleich, G. J. and Adolphson, C. R. The eosinophilic leukocyte: structure and function. *Adv. Inmunol.* **39** (1986), 177–253.

(18) Rosenberg, H. F. RNase A ribonucleases and host defense: an evolving story. *J. Leokocyte Biol.* **83** (2008), 1079–1087.

(19) Vescia, S., Tramontano, D., Augusti-Tocco, G., and D'Alessio, G. In Vitro Studies on Selective Inhibition of Tumor Cell Growth by Seminal Ribonuclease. *Cancer Res.* **40** (1980), 3740–3744.

(20) Kim, J., Soucek, J., Matousek, J., and Raines, R. T. Mechanism of Ribonuclease Cytotoxicity. *J. Biol. Chem.* **270** (1995), 31097–31102.

(21) Lee, J. E. and Raines, R. T. Contribution of Active-Site Residues to the Function of Onconase, a Ribonuclease with Antitumoral Activity . *Biochemistry* **42** (2003), 11443–11450.

(22) Leland, P. A., Schultz, L. W., Kim, B., and Raines, R. T. Ribonuclease A variants with potent cytotoxic activity. *Proc. Natl. Acad. Sci.* **95** (1998), 10407–10412.

(23) Notomista, E., Catanzano, F., Graziano, G., Piaz, F. D., Barone, G., D'Alessio, G., and Donato, A. D. Onconase: An Unusually Stable Protein. *Biochemistry* **39** (2000), 8711–8718.

(24) Richards, F. M. and Wyckoff, H. W. *The Enzymes*, New York Academic, 1971, vol. IV - Hydrolytic Cleavage. Part A0, chap. Bovine Pancreatic ribonuclease, pp. 647–907. 3rd edn.

(25) Nogués, M. V., Vilanova, M., and Cuchillo, C. M. Bovine pancreatic ribonuclease A as a model of an enzyme with multiple substrate binding sites. *Biochim. Biophys. Acta* **1253** (1995), 16–24.

(26) delCardayré, S. B., Ribó, M., Yokel, E. M., Quirk, D. J., Rutter, W. J., and Raines, R. T. Engineering ribonuclease A: production, purification and characterization of wild-type enzyme and mutants at Gln11. *Protein Eng.* **8** (1995), 261–273.

(27) Kelemen, B. R., Schultz, L. W., Sweeney, R. Y., and Raines, R. T. Excavating an Active Site: The Nucleobase Specificity of Ribonuclease A. *Biochemistry* **39** (2000), 14487–14494.

(28) Raines, R. Ribonuclease A. *Chem. Rev.* **98** (1998), 1045–1065.

(29) Lopez, X., York, D., and Dejaegere, A. Theoretical Studies on the hydrolysis of phosphate diesters in the Gas Phase, Solution, and RNase A. *Int. J. Quant. Chem.* **86** (2002), 10–26.

(30) Witzel, H. and Barnard, E. A. Mechanism and Binding sites in the Ribonuclease Reaction II. Kinetic studies on the first step of the Reaction. *Biochem. Biophys. Res. Commun.* **7** (1962), 295–299.

(31) Boix, E., Nogues, M. V., Schein, C. H., Benner, S. A., and Cuchillo, C. M. Reverse Transphosphorylation by Ribonuclease A Needs an Intact P_2-binding Site. Point mutations at Lys-7 and Arg-10 alter the catalytic properties of the enzyme. *J Biol. Chem.* **269** (1994), 2529–2534.

(32) Moussaoui, M., Guasch, A., Boix, E., Cuchillo, C. M., and Nogués, M. V. The Role of Non-catalytic Binding Subsites in the Endonuclease Activity of Bovine Pancreatic Ribonuclease A. *J. Biol. Chem.* **271** (1996), 4687–4692.

(33) Cuchillo, C. M., Moussaoui, M., Barman, T., Travers, F., and Nogués, M. V. The exo- or endonucleolytic preference of bovine pancreatic ribonuclease A depends on its subsites structure and on the substrate size. *Protein Sci.* **11** (2002), 117–128.

(34) Rodríguez, M., Moussaoui, M., Benito, A., Cuchillo, C., Nogués, M., and Vilanova, M. Human pancreatic ribonuclease presents higher endonucleolytic activity than ribonuclease A. *Arch. Biochem. Biophys.* **471** (2008), 191–197.

(35) Blackburn, P. and Moore, S. *The Enzymes*, New York Academic, 1982, vol. XV- Nucleic Acids. Part B, chap. 12-Pancreatic Ribonuclease, pp. 317–433. 3rd edn.

(36) Hirs, C. H., Moore, S., and Stein, W. A Chromatographic investigation of Pancreatic Ribonuclease. *J. Biol. Chem.* **200** (1953), 493–506.

(37) Crestfield, A. M., Stein, W. H., and Moore, S. On the Preparation of Bovine Pancreatic Ribonuclease A. *J. Biol. Chem.* **238** (1963), 618–621.

(38) Smyth, D. G., Stein, W. H., and Moore, S. The Sequence of Amino Acid Residues in Bovine Pancreatic Ribonuclease: Revisions and Confirmations. *J. Biol. Chem.* **238** (1963), 227–234.

(39) Beers, J., Roland F. Hydrolysis of Polyadenylic Acid by Pancreatic Ribonuclease. *J. Biol. Chem.* **235** (1960), 2393–2398.

(40) Hantgan, R. R. and Scheraga, H. A. Pathways of folding of reduced bovine pancreatic ribonuclease. *Biochemistry* **13** (1974), 3421–3431.

(41) Crestfield, A. M., Stein, W. H., and Moore, S. Alkylation and Identification of the Histidine Residues at the Active Site of Ribonuclease. *J. Biol. Chem.* **238** (1963), 2413–2420.

(42) Walter, B. and Wold, F. The role of lysine in the action of bovine pancreatic ribonuclease A. *Biochemistry* **15** (1976), 304–310.

(43) Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181** (1973), 223–230.

(44) Moore, S. and Stein, W. H. Chemical Structures of Pancreatic Ribonuclease and Deoxyribonuclease. *Science* **180** (1973), 458–464.

(45) Merrifield, R. B. Solid phase synthesis. *Science* **232** (1986), 341–347.

(46) Hammes, G. G. Multiple conformational changes in enzyme catalysis. *Biochemistry* **41** (2002), 8221–8228.

(47) Cole, R. and Loria, J. P. Evidence for Flexibility in the Function of Ribonuclease A. *Biochemistry* **41** (2002), 6072–6081.

(48) Beach, H., Cole, R., Gill, M. L., and Loria, J. P. Conservation of μs-ms Enzyme Motions in the Apo- and Substrate-Mimicked State. *J. Am. Chem. Soc.* **127** (2005), 9167–9176.

(49) Kovrigin, E. L. and Loria, J. P. Characterization of the Transition State of Functional Enzyme Dynamics. *J. Am. Chem. Soc.* **128** (2006), 7724–7725.

(50) Zegers, I., Maes, D., Dao-Thi, M., Poortmans, F., Palmer, R., and Wyns, L. The structures of RNase A complexed with 3'-CMP and d(CpA): active site conformation and conserved water molecules. *Protein Sci.* **3** (1994), 2322–2339.

(51) Fisher, B. M., Grilley, J. E., and Raines, R. T. A New Remote Subsite in Ribonuclease A. *J. Biol. Chem.* **273** (1998), 34134–34138.

(52) Fontecilla-Camps, J., de Llorens, R., le Du, M., and Cuchillo, C. Crystal structure of ribonuclease A·d(ApTpApApG) complex. Direct evidence for extended substrate recognition. *J. Biol. Chem.* **269** (1994), 21526–21531.

(53) McPherson, A., Brayer, G., and Morrison, R. Structure of the Crystalline Complex Between Ribonuclease A and $d(pA)_4$. *Biophys. J.* **49** (1986), 209 – 219.

(54) Aguilar, C. F., Thomas, P. J., Mills, A., Moss, D. S., and Palmer, R. Newly observed binding mode in pancreatic ribonuclease. *J. Mol. Biol.* **224** (1992), 265–267.

(55) Heinrikson, R. L. On the Alkylation of Amino Acid Residues at the Active Site of Ribonuclease. *J. Biol. Chem.* **241** (1966), 1393–1405.

(56) Thompson, J. E. and Raines, R. T. Value of general acid-base catalysis to ribonuclease A. *J. Am. Chem. Soc.* **116** (1994), 5467–5468.

(57) Cuchillo, C. M., Vilanova, M., and Nogués, M. V. *Ribonucleases: Structures and Functions*, Academic Press, 1997, chap. 9-Pancreatic Ribonucleases, pp. 271–300.

(58) Panov, K., Kolbanovskaya, E., Okorokov, A., Panova, T., Terwisscha van Scheltinga, A., Karpeisky, M., and Beintema, J. Ribonuclease A mutant His119Asn: the role of histidine in catalysis. *FEBS Letters* **398** (1996), 57–60.

(59) Breslow, R. and Chapman, J. W. H. On the mechanism of action of ribonuclease A: Relevance of enzymatic studies with a p-nitrophenylphosphate ester and a thiophosphate ester. *Proc. Natl. Acad. Sci. USA* **93** (1996), 10018–10021.

(60) Wladkowski, B. D., Svensson, L. A., Sjolin, L., Ladner, J. E., and Gilliland, G. L. Structure (1.3 Å) and Charge States of Ribonuclease A-Uridine Vanadate Complex: Implications for the Phosphate Ester Hydrolysis Mechanism. *J. Am. Chem. Soc.* **120** (1998), 5488–5498.

(61) Elsässer, B., Valiev, M., and Weare, J. A Dianionic Phosphorane Intermediate and Transition States in an Associative A_N+D_N Mechanism for the Ribonuclease A Hydrolysis Reaction. *J. Am. Chem. Soc.* **131** (2009), 3869–3871.

(62) Borkakoti, N., Moss, D. S., and Palmer, R. A. Ribonuelease-A: Least-Squares Refinement of the Structure at 1.45 Å Resolution. *Acta Crystallogr B* **38** (1982), 2210–2217.

(63) delCardayré, S. B. Structural Determinants of Enzymic Processivity. *Biochemistry* **33** (1994), 6031–6037.

(64) Tanimizu, N., Ueno, H., and Hayashi, R. Role of Phe120 in the Activity and Structure of Bovine Pancreatic Ribonuclease A . *J. Biochem.* **124** (1998), 410–416.

(65) Chatani, E. and Hayashi, R. Functional and structural roles of constituent amino acid residues of bovine pancreatic ribonuclease A. *J. Biosci. Bioeng.* **92** (2001), 98–107.

(66) Chatani, E., Hayashi, R., Moriyama, H., and Ueki, T. Conformational strictness required for maximum activity and stability of bovine pancreatic ribonuclease A as revealed by crystallographic study of three Phe120 mutants at 1.4 Å resolution. *Protein Sci.* **11** (2002), 72–81.

(67) Tarragona-Fiol, A., Eggelte, H., Harbron, S., Sanchez, E., Taylorson, C., Ward, J., and Rabin, B. Identification by site-directed mutagenesis of amino acids in the subsite of bovine pancreatic ribonuclease A. *Protein Eng.* **6** (1993), 901–906.

(68) Fisher, B. M., Schultz, L. W., and Raines, R. T. Coulombic Effects of Remote Subsites on the Active Site of Ribonuclease A. *Biochemistry* **37** (1998), 17386–17401.

(69) Gilliland, G. L. *Ribonucleases: Structures and Functions*, Academic Press, 1997, chap. 10- Crystallographic Studies of Ribonuclease Complexes, pp. 305–341.

(70) Stern, M. and Doscher, M. Aspartic acid-121 functions at the active site of bovine pancreatic ribonuclease. *FEBS Letters* **171** (1984), 253–256.

(71) Quirk, D. J. and Raines, R. T. His . . . Asp Catalytic Dyad of Ribonuclease A: Histidine pK_a Values in the Wild-Type, D121N, and D121A Enzymes. *Biophys. J.* **76** (1999), 1571–1579.

(72) Eftink, M. R. and Biltonen, R. L. *Hydrolytic Enzymes*, Elsevier, 1987, vol. 16 of *New comprehensive biochemistry*, pp. 333–376.

(73) Findlay, D., Herries, D., Mathias, A., Rabin, B., and Ross, C. The active site and mechanism of action of Bovine Pancreatic Ribonuclease. *Nature* **190** (1961), 781–784.

(74) Cuchillo, C., Parés, X., Guasch, A., Barman, T., Travers, F., and Nogués, M. The role of 2',3'-cyclic phosphodiesters in the bovine pancreatic ribonuclease A catalysed cleavage of RNA: intermediates or products? *FEBS Letters* **333** (1993), 207–210.

(75) Thompson, J. E., Venegas, F. D., and Raines, R. T. Energetics of Catalysis by Ribonucleases: Fate of the 2',3'-Cyclic Phosphodiester Intermediate. *Biochemistry* **33** (1994), 7408–7414.

(76) Roberts, G., Dennis, E., Meadows, D., Cohen, J., and Jardetzky, O. The mechanism of Action of Ribonuclease. *Proc. Natl. Acad. Sci. USA* **62** (1969), 1151–1158.

(77) Raines, R. T. *Artificial Nucleases*, Springer-Verlag, 2004, vol. 13 of *Nucleic Acids and Molecular Biology*, chap. Active Site of Ribonuclease A, pp. 19–32.

(78) Holmes, R. R. A model for calculating conformational energies in pentacoordinate phosphorus compounds. *J. Am. Chem. Soc.* **100** (1978), 433–446.

(79) Anslyn, E. and Breslow, R. On the mechanism of catalysis by ribonuclease: cleavage and isomerization of the dinucleotide UpU catalyzed by imidazole buffers. *J. Am. Chem. Soc.* **111** (1989), 4473–4482.

(80) Haydock, K., Lim, C., Brünger, A., and Karplus, M. Simulation Analysis of Structures on the Reaction Pathway of RNase A. *J. Am. Chem. Soc.* **112** (1990), 3826–3831.

(81) Lim, C. and Tole, P. Endocyclic and exocyclic cleavage of phosphorane monoanion: a detailed mechanism of the RNase A transphosphorylation step. *J. Am. Chem. Soc.* **114** (1992), 7245–7252.

(82) Perreault, D. M. and Anslyn, E. V. Unifying the Current Data on the Mechanism of Cleavage-Transesterification of RNA. *Angew. Chem. Int. Ed.* **36** (1997), 432–450.

(83) Breslow, R. and Xu, R. Recognition and catalysis in nucleic acid chemistry. *Proc. Natl. Acad. Sci. USA* **90** (1993), 1201–1207.

(84) Westheimer, F. H. Pseudo-rotation of the hydrolysis of phosphate esters. *Acc. Chem. Res.* **1** (1968), 70–78.

(85) Wladkowski, B. D., Krauss, M., and Stevens, W. Transphosphorylation catalyzed by ribonuclease A: computational study using ab initio Effective Fragment Potentials. *J. Am. Chem. Soc* **117** (1995), 10537–10545.

(86) Wladkowski, B. D., Krauss, M., and Stevens, W. Ribonuclease A catalyzed transphosphorylation: An ab initio theoretical study. *J. Phys. Chem* **99** (1995), 6273–6276.

(87) Wladkowski, B., Ostazeski, P., Chenoweth, S., Broadwater, S., and Krauss, M. Hydrolysis of cyclic phosphates by ribonuclease A: A computational study using a simplified ab initio quantum model. *J. Comput. Chem.* **24** (2003), 1803–1811.

(88) Hengge, A. C. and Onyido, I. Physical organic perspectives on phospho group transfer from phosphates and phosphinates. *Curr. Org. Chem.* **9** (2005), 61–74.

(89) Breslow, R., Dong, S. D., Webb, Y., and Xu, R. Further Studies on the buffer-catalyzed cleavage and isomerization of Uridyluridine. Medium and ionic strength effects on catalysis by morpholine, imidazole, and acetate buffers help clarify the mechanisms involved and their relationship to the mechanism used by the enzyme Ribonuclease and by a Ribonuclease mimic. *J. Am. Chem. Soc.* **118** (1996), 6588–6600.

(90) Gerratana, B., Sowa, G. A., and Cleland, W. W. Characterization of the transition-state structures and mechanisms for the isomerization and cleavage reactions of uridine 3′-*m*-nitrobenzyl phosphate. *J. Am. Chem. Soc.* **122** (2000), 12615–12621.

(91) Sowa, G., Hengge, A., and Cleland, W. ^{18}O Isotope Effects Support a Concerted Mechanism for Ribonuclease A. *J. Am. Chem. Soc.* **119** (1997), 2319–2320.

(92) Kluger, R., Covitz, F., Dennis, E., Williams, L. D., and Westheimer, F. H. pH-Product and pH-Rate Profiles for the Hydrolysis of Methyl Ethylene Phosphate. Rate-Limiting Pseudorotation. *J. Am. Chem. Soc.* **91** (1969), 6066–6072.

(93) Dejaegere, A., Lim, C., and Karplus, M. Dianionic Pentacoordinate Species in the Base-Catalyzed Hydrolysis of Ethylene and Dimethyl Phosphate. *J. Am. Chem. Soc.* **113** (1991), 4353–4355.

(94) Dejaegere, A., Liang, X. L., and Karplus, M. Phosphate ester hydrolysis calculation of gas-phase reaction paths and solvation effects. *J. Chem. Soc. Faraday Trans.* **90** (1994), 1763–1770.

(95) Uchimaru, T., Storer, J., Tanabe, K., Uebayasi, M., Nishikawa, S., and Taira, K. RNA hydrolysis via an oxyphosphorane intermediate. *Biochem. Biophys. Res. Commun.* **187** (1992), 1523–1528.

(96) Taira, K., Uchimaru, T., Storer, J. W., Yliniemela, A., Uebayasi, M., and Tanabe, K. Properties of Dianionic Oxyphosphorane Intermediates: Implication to the Reaction Profile for Base-Catalyzed RNA Hydrolysis. *J. Org. Chem.* **58** (1993), 3009–3017.

(97) Nam, K., Cui, Q., Gao, J., and York, D. M. Specific reaction parametrization of the AM1/d Hamiltonian for phosphoryl transfer reactions: H, O, and P atoms. *J. Chem. Theory Comput.* **3** (2007), 486–504.

(98) Schrödinger, E. Quantisierung als Eigenwertproblem / Quantization as a Problem of Proper Values. Part I. *Ann. Physik* **384** (1926), 361–376.

(99) Schrödinger, E. Über das Verhältnis der Heisenberg-Born-Jordanschen Quantenmechanik zu der meinem. *Ann. Physik* **384** (1926), 734–736.

(100) Schrödinger, E. An Undulatory Theory of the Mechanics of Atoms and Molecules. *Phys. Rev.* **28** (1926), 1049–1070.

(101) Dirac, P. On the Theory of Quantum Mechanics. *R. Soc. Lond. A* **112** (1926), 661–677.

(102) Heisenberg, W. Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen / Quantum-Theoretical Re-Interpretation of Kinematic and Mechanical Relations. *Z. Physik* **33** (1925), 879.

(103) Born, M. and Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann. Physik* **389** (1927), 457–484.

(104) Hartree, D. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods. *Math. Proc. Cambridge Philos. Soc.* **24** (1928), 89–110.

(105) Fock, V. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Z. Physik* **61** (1930), 126–148.

(106) Möller, C. and Plesset, M. S. Note on an Approximation Treatment for Many-electron systems. *Phys. Rev.* **46** (1934), 618–622.

(107) Shavitt, I. The history and evolution of configuration interaction. *Mol. Phys.* **94** (1998), 3–17.

(108) Sherrill, C. D. and Schaefer III, H. F. The Configuration Interaction Method: Advances in Highly Correlated Approaches. *Adv. Quantum Chem.* **34** (1999), 143–269.

(109) Pople, J. A., Krishnan, R., Schlegel, H. B., and Binkley, J. S. Electron correlation theories and their application to the study of simple reaction potential surfaces. *Int. J. Quant. Chem.* **14** (1978), 545–560.

(110) Bartlett, R. J. and Purvis, G. D. Many-body perturbation theory, coupled-pair many-electron theory, and the importance of quadruple excitations for the correlation problem. *Int. J. Quant. Chem.* **14** (1978), 561–581.

(111) Bartlett, R. J. *Modern Electronic Structure Theory. Part II*, 1995, vol. 2 of *Advanced Series in Physical Chemistry*, chap. Coupled-Cluster Theory: An Overview of Recent Developments.

(112) Andersson, K., Malmqvist, P., Roos, B. O., Sadlej, A. J., and Wolinski, K. Second order perturbation theory with a CASSCF reference function. *J. Phys. Chem.* **94** (1990), 5483–5488.

(113) Andersson, K., Malmqvist, P., and Roos, B. O. Second order perturbation theory with a complete active space self-consistent field reference function. *J. Chem. Phys.* **96** (1992), 1218–1226.

(114) Hohenberg, P. and Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **136** (1964), B864.

(115) Kohn, W. and Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140** (1965), A1133.

(116) Parr, R. G. and Yang, W. *Density-Functional Theory of Atoms and Molecules*. International Series of Monographs on Chemistry 16, Oxford University Press, New York, 1989.

(117) Kryachko, E. S. and Ludeña, E. V. *Energy Density Functional Theory of Many-Electron Systems*. Understanding Chemical Reactivity, Kluwer Academic Publishers, London, 1990.

(118) Dewar, M. and Thiel, W. Ground States of Molecules. 38. The MNDO Method. Approximations and Parameters. *J. Am. Chem. Soc.* **99** (1977), 4899–4907.

(119) Dewar, M. and Thiel, W. Ground States of Molecules. 39. MNDO Results for Molecules Containing Hydrogen, Carbon, Nitrogen, and Oxygen. *J. Am. Chem. Soc.* **99** (1977), 4907–4917.

(120) Dewar, M. J. S., Zoebisch, E., Healy, E. F., and Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **107** (1985), 3902–3909.

(121) Stewart, J. J. P. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.* **10** (1989), 209–220.

(122) Stewart, J. J. P. Optimization of parameters for semiempirical methods II. applications. *J. Comput. Chem.* **10** (1989), 221–264.

(123) Thiel, W. and Voityuk, A. A. Extension of MNDO to d Orbitals: Parameters and Results for the Second-Row Elements and for the Zinc Group. *J. Phys. Chem.* **100** (1996), 616–626.

(124) Thiel, W. *Advances in Chemical Physics*, John Wiley and Sons, New York, 1996, vol. 93, chap. Perspectives on semiempirical molecular orbital theory, pp. 703–757.

(125) Lopez, X. and York, D. M. Parameterization of semiempirical methods to treat nucleophilic attacks to biological phosphates: AM1/d parameters for phosphorus. *Theor. Chem. Acc.* **109** (2003), 149–159.

(126) Brooks, B. R., Brooks III, C. L., MacKerell, Jr., A. D., Nilsson, L., J., P. R., Roux, B., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaris, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, W., Yang, W., York, D. M., and Karplus, M. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **30** (2009), 1545–1614.

(127) Brooks, B., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4** (1983), 187–217.

(128) Mackerell, Jr., A. D., Feig, M., and Brooks, III, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25** (2004), 1400–1415.

(129) Leach, A. R. *Molecular Modelling: Principles and Applications.* Prentince-Hall Inc., 2001, 2 edn.

(130) Frenkel, D. and Smit, B. *Understanding Molecular Simulation. From Algorithms to Applications.* Academic Press, 2002, 2nd edn.

(131) Brooks III, C. L., Pettitt, B. M., and Karplus, M. Structural and energetic effects of truncating long ranged interactions in ionic and polar fluids. *J. Chem. Phys.* **83** (1985), 5897–5908.

(132) Steinbach, P. J. and Brooks, B. R. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.* **15** (1994), 667–683.

(133) Ewald, P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* **369** (1921), 253–268.

(134) Kittel, C. *Introduction to Solid State Physics.* John Wiley & Sons, Inc., New York, 2004, 8th edn.

(135) Darden, T., York, D. M., and Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98** (1993), 10089–10092.

(136) York, D. M., Darden, T. A., and Pedersen, L. G. The effect of long-range electrostatic interactions in simulations of macromolecular crystals: A comparison of the Ewald and truncated list methods. *J. Chem. Phys.* **99** (1993), 8345–8348.

(137) Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **103** (1995), 8577–8593.

(138) Sagui, C. and Darden, T. A. Molecular Dynamics Simulations of Biomolecules: Long-Range Electrostatic Effects. *Annu. Rev. Biophys. Biomol. Struct.* **28** (1999), 155–179.

(139) Stote, R., States, D., and Karplus, M. On the treatment of electrostatic interactions in biomolecular simulation. *J. Chem. Phys.* **88** (1991), 2419–2433.

(140) Greengard, L. and Rokhlin, V. A fast algorithm for particle simulations. *J. Comp. Phys.* **73** (1987), 325–348.

(141) Shimada, J., Kaneko, H., and Takada, T. Performance of fast multipole methods for calculating electrostatic interactions in biomacromolecular simulations. *J. Comp. Chem.* **15** (1994), 28–43.

(142) Kaminski, G. A., Stern, H. A., Berne, B. J., Friesner, R. A., Cao, Y. X., Murphy, R. B., Zhou, R., and Halgren, T. A. Development of a polarizable force field for proteins via ab initio quantum chemistry: First generation model and gas phase tests. *J. Comp. Chem.* **23** (2002), 1515–1531.

(143) MacKerell, A. D., Wiórkiewicz-Kuczera, J., and Karplus, M. An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* **117** (1995), 11946–11975.

(144) MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., W. E. Reiher, B. R., III, Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **102** (1998), 3586–3616.

(145) MacKerell, A. D. and Banavali, N. K. All-Atom Empirical Force Field for Nucleic Acids: II. Application to Molecular Dynamics Simulations of DNA and RNA in Solution. *J. Comput. Chem.* **21** (2000), 105–120.

(146) Foloppe, N. and MacKerell, Jr., A. D. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **21** (2000), 86–104.

(147) Cheatham, III, T. E. and A.Young, M. Molecular dynamics simulation of nucleic acids: successes, limitations, and promise. *Biopolymers* **56** (2001), 232–256.

(148) Case, D. A., Cheatham III, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **26** (2005), 1668–1688.

(149) Ponder, J. W. and Case, D. A. Force Fields for Protein Simulations. *Adv. Prot. Chem.* **66** (2003), 27–85.

(150) Case, D., Darden, T., Cheatham, T., III, Simmerling, C., Wang, J., Duke, R., Luo, R., Crowley, M., Walker, R., Zhang, W., Merz, K., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K., Paesani, F., Vanicek, J., Wu, X., Brozell, S., Steinbrecher, T., Gohlke, H., Yang,

L., Tan, C., Mongan, J., Hornak, V., Cui, G., Mathews, D., Seetin, M., Sagui, C., Babin, V., , and Kollman, P. *AMBER 10*. University of California, San Francisco, San Francisco, 2002.

(151) van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hunenberger, P. H., Kruger, P., Mark, A. E., Scott, W. R. P., and Tironi, I. G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*. VDf: Hochschulverlag AG an der ETH Zurich, Gronigen, 1996.

(152) Hermans, J., Berendsen, H., Van Gunsteren, W., and Postma, J. A Consistent empirical potential for water-protein interactions. *Biopolymers* **23** (1984), 1513–1518.

(153) Schuler, L. D., Daura, X., and van Gunsteren, W. F. An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comp. Chem.* **22** (2001), 1205–1218.

(154) Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118** (1996), 11225–11236.

(155) Damm, W., Frontera, A., Tirado-Rives, J., and Jorgensen, W. L. OPLS all-atom force field for carbohydrates. *J. Comput. Chem.* **18** (1997), 1955–1970.

(156) Halgren, T. A. Representation of van der Waals (vdW) interactions in molecular mechanics force fields: Potential form, combination rules, and vdW parameters. *J. Am. Chem. Soc.* **114** (1992), 7827–7834.

(157) Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* **117** (1995), 5179–5197.

(158) Cauchy, A. Méthode générale pour la résolution des systèmes d'équations simultanées. *Compte-Rendu de l'Académie des Sciences, Paris* **27** (1847), 536–538.

(159) Fletcher, R. and Reeves, C. M. Function minimization by conjugate gradients. *The Computer Journal* **7** (1964), 149–154.

(160) Luenberger, D. G. *Linear and Nonlinear programming*. Kluwer Academic, 2003.

(161) Ypma, T. J. Historical development of the Newton-Raphson method. *SIAM Review* **37** (1995), 531–551.

(162) Torrie, G. M. and Valleau, J. P. Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.* **28** (1974), 578–581.

(163) Ferrenberg, A. M. and Swendsen, R. H. Optimized monte carlo data analysis. *Phys. Rev. Lett.* **63** (1989), 1195–1198.

(164) Valleau, J. P. Thermodynamic-scaling methods in Monte Carlo and their application to phase equilibria. *Adv. Chem. Phys.* **105** (1999), 369–404.

(165) Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., and Teller, A. H. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** (1953), 1087–1092.

(166) McCammon, J. A., Gelin, B. R., and Karplus, M. Dynamics of folded proteins. *Nature* **267** (1977), 585–590.

(167) Allen, M. P. and Tildesley, D. J. *Computer Simulation of Liquids.* Oxford University Press, 1987.

(168) Allen, M. P. Introduction to Molecular Dynamics Simulation. *NIC Series* **23**.

(169) Verlet, L. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.* **159** (1967), 98–103.

(170) Verlet, L. Computer "Experiments" on Classical Fluids. II. Equilibrium Correlation Functions. *Phys. Rev.* **165** (1968), 201–214.

(171) Hockney, R. W., Goel, S. P., and Eastwood, J. W. A 10000 particle molecular dynamics model with long range forces. *Chem. Phys. Lett.* **21** (1973), 589–591.

(172) Berkowitz, M. and McCammon, J. A. Molecular Dynamics with Stochastic Boundary Conditions. *Chem. Phys. Letters* **90** (1982), 215–217.

(173) Brünger, A., Brooks III, C. L., and Karplus, M. Stochastic Boundary Conditions force Molecular Dynamics Simulations of ST2 Water. *Chem. Phys. Lett.* **105** (1984), 495–500.

(174) Brooks III, C. L., Brünger, A., and Karplus, M. Active Site Dynamics in Protein Molecules: A Stochastic Boundary Molecular Dynamics Approach. *Biopolymers* **24** (1985), 843–865.

(175) Nadler, W., Brünger, A., Schulten, K., and Karplus, M. Molecular and stochastic dynamics of proteins. *Proc. Natl. Acad. Sci. USA* **84** (1987), 7933–7937.

(176) Adelman, S. A. and Brooks III, C. L. Generalized Langevin models and condensed-phase chemical reaction dynamics. *J. Phys. Chem.* **86** (1982), 1511–1524.

(177) Senn, H. M. and Thiel, W. QM/MM Methods for Biological Systems. *Top. Curr. Chem.* **268** (2007), 173–290.

(178) Senn, H. M. and Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed.* **48** (2009), 1198–1229.

(179) Field, M. J., Bash, P. A., and Karplus, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* **11** (1990), 700–733.

(180) Gao, J. Methods and applications of combined quantum mechanical and molecular mechanical potentials. *Rev. Comput. Chem.* **7** (1995), 119–185.

(181) Reuter, N., Dejaegere, A., Maigret, B., and Karplus, M. Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches. *J. Phys. Chem. A* **104** (2000), 1720–1735.

(182) Warshel, A. and Levitt, M. Theoretic studies of enzymatic reactions:dielectric electrostatic and steric stabilization in the reaction of lysozyme. *J. Mol. Biol.* **103** (1976), 227–249.

(183) Eurenius, K. P., Chatfield, D. C., Brooks, B. R., and Hodoscek, M. Enzyme Mechanisms with Hybrid Quantum and Molecular Mechanical Potentials. I. Theoretical Considerations. *Int. J. Quan. Chem.* **60** (1996), 1189–1200.

(184) Singh, U. C. and Kollman, P. A. A combined *ab initio* quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: applications to the CH_3Cl+Cl^- exchange reaction and gas phase protonation of polyethers. *J. Comput. Chem.* **7** (1986), 718–730.

(185) Gao, J., Amara, P., Alhambra, C., and Field, M. J. A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A* **102** (1998), 4714–4721.

(186) Théry, V., Rinaldi, D., Rivail, J.-L., Maigret, B., and Ferenczy, G. G. Quantum mechanical computations on very large molecular systems: the local self-consistent field method. *J. Comput. Chem.* **15** (1994), 269–282.

(187) Monard, G., Loos, M., Théry, V., Baka, K., and Rivail, J.-L. Hybrid classical quantum force field for modeling very large molecules. *Int. J. Quantum Chem.* **58** (1996), 153–159.

(188) Assfeld, X. and Rivail, J.-L. Quantum chemical computations on parts of large molecules: The ab initio local self-consistent field approach. *Chem. Phys. Lett.* **263** (1996), 100–106.

(189) Ferré, N., Assfeld, X., and Rivail, J.-L. Specific force field parameters determination for the hybrid ab initio QM/MM LSCF method. *J. Comput. Chem.* **23** (2002), 610–624.

(190) Amara, P., Field, M. J., Alhambra, C., and Gao, J. The generalized hybrid orbital method for combined quantum mechanical/molecular mechanical calculations: formulation and tests of the analytical derivatives. *Theor. Chem. Acc.* **104** (2000), 336–343.

(191) Garcia-Viloca, M. and Gao, J. Generalized hybrid orbital for the treatment of boundary atoms in combined quantum mechanical and molecular mechanical calculations using the semiempirical parameterized model 3 method. *Theor. Chem. Acc.* **111** (2004), 280–286.

(192) Pu, J., Gao, J., and Truhlar, D. G. Combining self-consistent-charge density-functional tight-binding (SCC-DFTB) with molecular mechanics by the generalized hybrid orbital (GHO) method. *J. Phys. Chem. A* **108** (2004), 5454–5463.

(193) Pu, J., Gao, J., and Truhlar, D. G. Generalized hybrid orbital (GHO) method for combining ab initio Hartree-Fock wave functions with molecular mechanics. *J. Phys. Chem. A* **108** (2004), 632–650.

(194) Pu, J., Jiali Gao, P., and Donald G. Truhlar, P. Generalized hybrid-orbital method for combining density functional theory with molecular mechanicals. *Chem. Phys. Chem.* **6** (2005), 1853–1865.

(195) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **3** (1935), 300–313.

(196) Chandler, D. *Introduction to Modern Statistical Mechanics.* Oxford University Press, 1987, 1st edn.

(197) den Otter, W. K. and Briels, W. J. The calculation of free-energy differences by constrained molecular-dynamics simulations. *J. Chem. Phys.* **109** (1998), 4139–4146.

(198) Sprik, M. and Ciccotti, G. Free energy from constrained molecular dynamics. *J. Chem. Phys.* **109** (1998), 7737–7744.

(199) Torrie, G. M. and Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **23** (1977), 187–199.

(200) Jarzynski, C. Equilibrium free-energy differences from nonequilibrium measurements: a master-equation approach. *Phys. Rev. E* **56** (1997), 5018–5035.

(201) Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **78** (1997), 2690–2693.

(202) Isralewitz, B., Baudry, J., Gullingsrud, J., Kosztin, D., and Schulten, K. Steered molecular dynamics investigations of protein function. *J. Mol. Graph. Model.* **19** (2001), 13–25.

(203) Isralewitz, B., Gao, M., and Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* **11** (2001), 224–230.

(204) Darve, E. and Pohorille, A. Calculating free energies using average force. *J. Chem. Phys.* **115** (2001), 9169–9183.

(205) Kumar, S., Bouzida, D., Swendsen, R., Kollman, P., and Rosenberg, J. The weighted histogram analysis method for free-energy calculations on biomolecules 1: The method. *J. Comput. Chem.* **13** (1992), 1011–1021.

(206) Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., and Kollman, P. A. Multidimensional Free - Energy Calculations Using the Weighted Histogram Analysis Method. *J. Comp. Chem.* **16** (1995), 1339–1350.

(207) Roux, B. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Commun.* **91** (1995), 275–282.

(208) Souaille, M. and Roux, B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.* **135** (2001), 40–57.

(209) Breslow, R. How do imidazole groups catalyze the cleavage of RNA in enzyme models and in enzymes? Evidence from "Negative Catalysis". *Acc. Chem. Res.* **24** (1991), 317–324.

(210) Wlodawer, A., Miller, M., and Sjölin, L. Active site of RNase: Neutron diffraction study of a complex with uridine vanadate, a transition-state analog. *Proc. Natl. Acad. Sci. USA* **80** (1983), 3628–3631.

(211) Wlodawer, A., Borkakoti, N., Moss, D. S., and Howlin, B. Comparison of Two Independently Refined Models of Ribonuclease-A. *Acta Crystallogr. B* **42** (1986), 379–387.

(212) Borah, B., wan Chen, C., Egan, W., Miller, M., Wlodawer, A., and Cohen, J. S. Nuclear Magnetic Resonance and Neutron Diffraction Studies of the Complex of Ribonuclease A with Uridine Vanadate, a Transition-State Analogue. *Biochemistry* **24** (1985), 2058–2067.

(213) Ladner, J. E., Wladkowski, B. D., Svensson, L. A., Sjölin, L., and Gilliland, G. L. X-ray Structure of a Ribonuclease A–Uridine Vanadate Complex at 1.3 Å Resolution. *Acta Crystallographica Section D* **53** (1997), 290–301.

(214) Messmore, J. and Raines, R. Pentavalent Organo-Vanadates as Transition State Analogues for Phosphoryl Transfer Reactions. *J. Am. Chem. Soc.* **122** (2000), 9911–9916.

(215) Holloway, D. E., Chavali, G. B., and Leonidas, D. D. Influence of Naturally-occurring 5'-Pyrophosphate-linked Substituents on the Binding of Adenylic Inhibitors to Ribonuclease A: an X-Ray Crystallographic Study. *Biopolymers* .

(216) Kovrigin, E. L. and Loria, J. P. Enzyme Dynamics along the Reaction Coordinate: Critical Role of a Conserved Residue. *Biochemistry* **45** (2006), 2636–2647.

(217) Herschlag, D., Piccirilli, J. A., and Cech, T. R. Ribozyme-Catalyzed and Nonenzymatic Reactions of Phosphate Effects upon Substitution of Sulfur for a Nonbridging Phosphoryl Diesters: Rate Oxygen Atom. *Biochemistry* **30** (1991), 4844–4854.

(218) Mayaan, E., Moser, A., Jr., A. D. M., and York, D. M. CHARMM Force Field Parameters for Simulation of Reactive Intermediates in Native and Thio-Substituted Ribozymes. *J. Comput. Chem.* **28** (2007), 495–507.

(219) Brünger, A., Brooks, C., and Karplus, M. Active site dynamics of ribonuclease. *Proc. Natl. Acad. Sci. USA* **82** (1985), 8458–8462.

(220) Seshadri, K., Rao, V. S. R., and Vishveshwara, S. Interaction of Substrate Uridyl 3',5'-Adenosine with Ribonuclease A: A Molecular Dynamics Study. *Biophys. J.* **69** (1995), 2185–2194.

(221) Toiron, C., Gonzalez, C., Bruix, M., and Rico, M. Three-dimensional structure of the complexes of ribonuclease A with 2',5'-CpA and 3',5'-d(CpA) in aqueous solution, as obtained by NMR and restrained molecular dynamics. *Protein Science* **5** (1996), 1633–1647.

(222) Sanjeev, B. and Vishveshwara, S. Protein-water interactions in ribonuclease A and angiogenin: A molecular dynamics study. *Proteins: Structure, Function, and Bioinformatics* **55** (2004), 915–923.

(223) Cotesta, S., Tavernelli, I., and Di lorio, E. E. Dynamics of RNase-A and S-Protein: A Molecular Dynamics Simulation of the Transition Toward a Folding Intermediate. *Biophys. J.* **85** (2003), 2633–2640.

(224) Esposito, L. and Daggett, V. Insight into Ribonuclease A Domain Swapping by Molecular Dynamics Unfolding Simulations. *Biochemistry* **44** (2005), 3358–3368.

(225) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79** (1983), 926–935.

(226) Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **72** (1980), 2384–2393.

(227) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81** (1984), 511–519.

(228) Hoover, W. G. Canonical dynamics: Equilibration phase-space distributions. *Phys. Rev. A* **31** (1985), 1695–1697.

(229) Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **23** (1977), 327–341.

(230) Neidle, S., Schneider, B., and Berman, H. M. *Structural Bioinformatics*, Wiley-Liss, Inc., San Diego Supercomputer Center, Department of Pharmacology, University of California San Diego, La Jolla, CA, USA; Structural Bioinformatics, San Diego SuperComputer Center, University of California San Diego, La Jolla, CA, USA, 2003, vol. 44 of *Methods of Biochemical Analysis*, chap. 3- Fundamentals of DNA and RNA structure, pp. 41–73.

(231) Richardson, J. S., Schneider, B., Murray, L. W., Kapral, G. J., Immormino, R. M., Headd, J. J., Richardson, D. C., Ham, D., Hershkovits, E., Williams, L. D., Keating, K. S., Pyle, A. M., Micallef, D., Westbrook, J., and Berman, H. M. RNA backbone: Consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA Journal* **14** (2008), 465–481.

(232) Altona, C. and Sundaralingam, M. Conformational analysis of the sugar ring in nucleosides and nucleotides. New description using the concept of pseudorotation. *J. Am. Chem. Soc.* **94** (1972), 8205–8212.

(233) Saenger, W. *Principles of nucleic acid structure*, Springer-Verlag, New York, 1984, chap. 17, pp. 368–384.

(234) Bloomfield, V. A., Crothers, D. M., and Tinoco, Jr., I. *Nucleic Acids: Structures, Properties, and Functions*. University Science Books, Sausalito, CA, 2000.

(235) Messmore, J. M., Fuchs, D. N., and Raines, R. Ribonuclease A: Revealing Structure-Function Relationships with Semisynthesis. *J. Am. Chem. Soc* **117** (1995), 8057–8060.

(236) Jentoft, J., Gerken, T., Jentoft, N., and Dearborn, D. [^{13}C]Methylated Ribonuclease A, ^{13}C NMR studies of the interaction of lysine 41 with active site ligands. *J. Biol. Chem.* **256** (1981), 231–236.

(237) Berisio, R., Sica, F., Lamzin, V. S., Wilson, K. S., Zagari, A., and Mazzarella, L. Atomic resolution structures of ribonuclease A at six pH values. *Acta Crystallogr. D.* **58** (2002), 441–450.

(238) Nelson, D. L. and Cox, M. M. *Lehninger Principles of Biochemistry*. W. H. Freeman; 4th Edition, New York, NY, 2005.

(239) Pelmenschikov, A., Yin, X., and Leszczynski, J. Revealing the Role of Water in the Acid-Base Interaction between the Phosphate Groups of DNA and the Amino Acid Side Chains of Proteins: A Density Functional Theory Study of Molecular Models. *J. Phys. Chem. B* **104** (2000), 2148–2153.

(240) Park, C., Schultz, L. W., and Raines, R. T. Contribution of the Active Site Histidine Residues of Ribonuclease A to Nucleic Acid Binding. *Biochemistry* **40** (2001), 4949–4956.

(241) Tanimizu, N., Ueno, H., and Hayashi, R. Replacement of His12 or His119 of Bovine Pancreatic Ribonuclease A with Acidic Amino Acid Residues for the Modification of Activity and Stability. *J. Biosci. Bioeng.* **94** (2002), 39–44.

(242) Trautwein, K., Holliger, P., Stackhouse, J., and Benner, S. A. Site-directed mutagenesis of bovine pancreatic ribonuclease: Lysine-41 and aspartate-121. *FEBS Letters* **281** (1991), 275–277.

(243) Formoso, E., Matxain, J. M., York, D. M., and Lopez, X. Molecular Dynamics Simulation of Bovine Pancreatic Ribonuclease A - CpA and Transition State-like Complexes. *J. Phys. Chem. B* Submited.

(244) Gregersen, B. A., Lopez, X., and York, D. M. Hybrid QM/MM Study of Thio Effects in Transphosphorylation Reactions: The Role of Solvation. *J. Am. Chem. Soc.* **126** (2004), 7504–7513.

(245) Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., and Cheatham III, T. E. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **33** (2000), 889–897.

(246) Cohen, A. J., Mori-Sánchez, P., and Yang, W. Insights into Current Limitations of Density Functional Theory. *Science* **321** (2008), 792–794.

(247) Lahiri, S., Zhang, G., Dunaway-Mariano, D., and Allen, K. The Pentacovalent Phosphorus Intermediate of a Phosphoryl Transfer Reaction. *Science* **299** (2003), 2067–2071.

(248) Lopez, X., Schaefer, M., Dejaegere, A., and Karplus, M. Theoretical Evaluation of pKa in Phosphoranes: Implications for Phosphate Ester Hydrolysis. *J. Am. Chem. Soc* **124** (2002), 5010–5018.

(249) Rajamani, R., Naidoo, K. J., and Gao, J. Implementation of an adaptive umbrella sampling method for the calculation of multidimensional potential of mean force of chemical reactions in solution. *J. Comput. Chem.* **24** (2003), 1775–1781.

(250) Lide, D. R., ed. *CRC handbook of chemistry and physics.* CRC Press LLC, Boca Raton,FL, 2003, 83 edn.

(251) Glennon, T. M. and Warshel, A. Energetics of the Catalytic Reaction of Ribonuclease A: A Computational Study of Alternative Mechanisms. *J. Am. Chem. Soc.* **120** (1998), 10234–10247.

(252) Range, K., McGrath, M. J., Lopez, X., and York, D. M. The Structure and Stability of Biological Metaphosphate, Phosphate, and Phosphorane Compounds in the Gas Phase and in Solution. *J. Am. Chem. Soc.* **126** (2004), 1654–1665.

(253) Lopez, X., Dejaegere, A., Leclerc, F., York, D. M., and Karplus, M. Nucleophilic Attack on Phosphate Diesters: A density Functional Study of In-Line Reactivity in Dianionic, Monoanionic and Neutral Systems. *J. Phys. Chem. B* **110** (2006), 11525–11539.

(254) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A.* **38** (1988), 3098–3100.

(255) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98** (1993), 5648–5652.

(256) Lee, C., Yang, W., and Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional ofthe electron density. *Phys. Rev. B.* **37** (1988), 785–789.

(257) Zhao, Y. and Truhlar, D. G. Hybrid Meta Density Functional Theory Methods for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions: The MPW1B95 and MPWB1K Models and comparative Assessments for Hydrogen Bonding and van der Waals Interactions. *J. Phys. Chem. A* **108** (2004), 6908.

(258) Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, J. A., Jr., Vreven, T., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., 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., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford,

S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C., and Pople, J. A. Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.

(259) Eftink, M. R. and Biltonen, R. L. Energetics of Ribonuclease A Catalysis. 1. pH, Ionic strength, and solvent isotope dependece of the Hydrolysis of cytidine Cyclic 2',3'-Phosphate. *Biochemistry* **22** (1983), 5123–5134.

(260) Kirby, A. J. and Younas, M. The reactivity of phosphate esters. diester hydrolysis. *J. Chem. Soc. B* pp. 510–513.

(261) Åqvist, J., Kolmodin, K., Florian, J., and Warshel, A. Mechanistic alternatives in phosphate monoester hydrolysis: what conclusions can be drawn from avaliable experimental data? *Chem. and Bio.* **3** (1999), 71–80.

(262) Knowles, J. Seeing is Believing. *Science* **299** (2003), 2002–2003.

(263) Kosonen, M., Seppänen, R., Wichmann, O., and Lönnberg, H. Hydrolysis and intramolecular transesterification of ribonucleoside 3'-phosphotriesters: comparison of structural effects in the reactions of asymmetric and symmetric dialkyl esters of 5'-O-piavaloyl-3'-uridylic acid. *J. Chem. Soc., Perkin Trans. 2* pp. 2433–2439.

(264) (the late), S. M., Kosonen, M., and Lönnberg, H. Cleavage and Isomerization of RNA Phosphodiester Bonds: Nucleoside Phosphotriesters and Chimeric Ribo/2Â´-O-Methylribo Oligonucleotides as Tools for Mechanistic Studies. *Current Organic Chemistry* **6** (2002), 523–538.

(265) Lönnberg, H., Strömberg, R., and Williams, A. Compelling evidence for a stepwise mechanism of the alkaline cyclisation of uridine 3'-phosphate esters. *Org. Biomol. Chem.* **2** (2004), 2165–2167.

(266) Lönnberg, T. and Lönnberg, H. Chemical models for ribozyme action. *Current Opinion in Chemical Biology* **9** (2005), 665–673.

(267) Lisgarten, J., Maes, D., and Wyns, L. Structure of the Crystalline Complex of Deoxycytidylyl-3',5'-guanosine (3',5'-dCpdG) Cocrystallized with Ribonuclease at 1.9 A Resolution. *Acta Crystallographica Section D* **51** (1995), 767–771.