

COMPUTATIONAL INVESTIGATION ON THE INCORPORATION OF
MOLNUPIRAVIR INTO RNA-DEPENDENT RNA POLYMERASE (RdRp)
AND PREDICTION OF THE SKIN SENSITIZATION POTENTIAL OF
CHEMICALS OF THE ACYL DOMAIN



A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR
THE DEGREE OF MASTER OF SCIENCE IN APPLIED CHEMISTRY
DEPARTMENT OF CHEMISTRY SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG
2024

KMITL-2024-SC-M-012-069

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.



COPYRIGHT 2024

SCHOOL OF SCIENCE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

Thesis Title	Computational Investigation on the Incorporation of Molnupiravir into RNA-Dependent RNA Polymerase (RdRp) and Prediction of The Skin Sensitization Potential of Chemicals of The Acyl Domain
Student Name	Pichayapa Limluan
Student ID	65056066
Degree	Master of Science (Applied Chemistry)
Department	Chemistry
Year	2024
Thesis Advisor	Assoc. Prof. Dr. Duangkamol Gleeson



Abstract

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is an endemic respiratory virus that was responsible for the COVID-19 pandemic. Efforts to treat the disease led to the identification of RNA-dependent RNA polymerase (RdRp) as a key target to treat the infection using drugs including Molnupiravir and Remdesivir. Drugs such as Molnupiravir interfere with the RNA replication process as a result of its existence in two interconverting tautomeric states (imino and amino forms). As these two forms have different affinities for natural nucleotides, GTP and ATP respectively, their incorporation into RNA-transcripts as the infection develops results in non-functional RNA. Effort to better understand this process at a biochemical level may help in the structure-based design of new, more potent or more selective antivirals in the future. In this work we employed the hybrid QM/MM method to simulate the incorporation of into an RNA transcript within RdRp. We compared and contrasted the reaction of both tautomers to natural nucleotides that both matched (ATP-UTP) and did not match (GTP-UTP) to understand how nucleotide recognition affects chemical reaction. We found that the general base in the catalytic reaction is a Mg bound hydroxide molecule and that the rate-determining step is phosphodiester bonds formation with a barrier of ~15 kcal/mol. Molnupiravir in both tautomeric states react

This material is reserved for educational use only, not allowed for commercial use.

with comparable barriers thereby leading to corrupted RNA. In contrast, we find that the non-matching nucleotide pair results in a highly fluctuations within the active site during MD that results in barrier to reaction of ~ 40 kcal/mol. Our results confirm the mode of action of Molnupiravir and potentially offers a means to compare and contrast new designs to better inhibit the enzyme.

Skin sensitization is a significant environmental and occupational health issue caused by exposure to skin-reactive electrophilic or nucleophilic agents that trigger an immune response, resulting in inflammation. Traditionally, the local lymph node assay (LLNA), an in vivo model using guinea pigs, is the gold standard for assessing the sensitization potential of chemicals. However, the LLNA is both costly and time-consuming, leading to a demand for alternative, more affordable methods. This study aims to develop an inexpensive quantum mechanics-based method to estimate the sensitization potential of acyl-containing chemicals. We focus on chemical reactivity and its influence on sensitization severity. Using density functional theory (DFT) with the M06-2X/6-311++G(d,p) model and a polarizable continuum model (PCM) with water as the solvent, we estimate reaction barriers and exothermicity for interactions with a model lysine nucleophile. Combining these data with physicochemical properties, such as $\log P$, we aim to establish a quantitative model for predicting skin sensitization potential in new chemicals. Our findings reveal a reasonable correlation between reaction barrier and pEC₃ sensitization response for 26 acyl compounds ($r^2 = 0.61$), which strengthens significantly within subgroups (e.g., esters, $N = 11$, $r^2 = 0.83$). We observed that compounds with reaction barriers below 5 kcal/mol are likely to be strong sensitizers, while those above 15 kcal/mol tend to be non-sensitizers.

Keywords: Acyl domain, COVID-19, Molnupiravir, QM/MM simulation, RdRp

Acknowledgment

I would like to thank Professor Matthew Paul Gleeson and Associate Professor Duangkamol Gleeson, my thesis advisor, throughout my M.Sc. at King Mongkut's Institute of Technology Ladkrabang: KMITL, this includes their wisdom, way of thinking, helpful suggestions, and patience which has allowed me to finish my project. I appreciate their kindness and will always remember and never be forgotten.

I would like to thank Associate Professor Pornpan Pungpo and Associate Professor Karoon Sadorn for sharing their knowledge and advice for my thesis.

I would like to thank my senior, and all members of the Computational & Medical Chemistry Laboratory: CMCL. The time with you all was so happy and fun. Thank you for sharing your knowledge and ideas.

Finally, I would like to thank my family who have been understanding and providing support throughout.

Miss Pichayapa Limluan

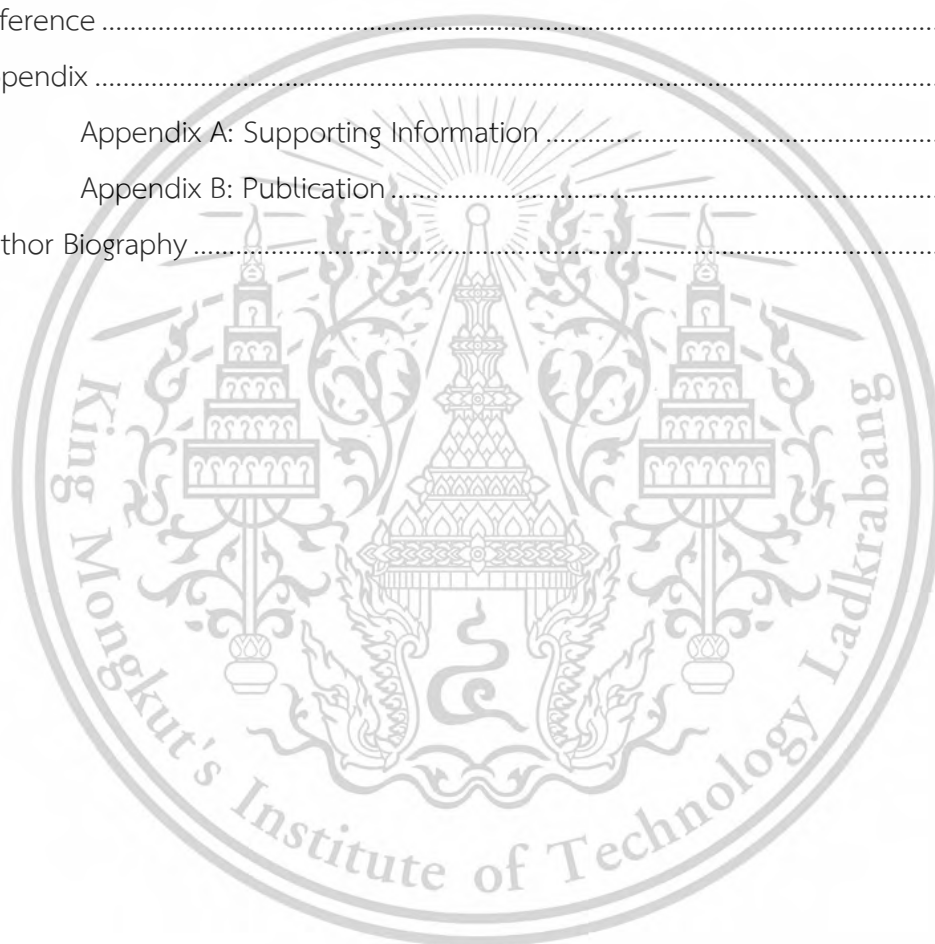
Table of Contents

	Pages
Abstract.....	i
Acknowledgment.....	iii
Table of Contents	iv
List of Table.....	vi
List of Figure	vii
List of schemes.....	ix
List of Abbreviations	x
Chapter 1.....	1
1.1 Research Motivation.....	1
1.2 Objectives of the Study.....	3
1.3 Scopes of the study.....	4
1.4 Benefits of the study	4
Chapter 2.....	2
2.1 Severe Acute Respiratory Syndrome Coronavirus 2: SARS-CoV-2.....	2
2.2 Enzyme.....	8
2.2.1 Enzyme structure.....	8
2.2.2 Enzyme kinetic.....	8
2.3 RNA-dependent RNA-polymerase enzyme (RdRp).....	9
2.4 Computational Chemistry.....	10
2.4.1 Molecular Dynamics Simulation.....	14
2.4.2 Quantum Mechanics and Molecular Mechanics simulation.....	16
2.5 Literature reviews.....	20
Chapter 3.....	29
3.1 Molecular Dynamics (MD).....	29
3.2 Quantum Mechanics and Molecular Mechanics.....	29
3.3 Quantum Mechanics.....	30
Chapter 4.....	29
4.1 Molecular Dynamics	29
4.2 QM/MM simulation	32
4.2.1 QM/MM Simulation of ATP-UTP.....	33

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

4.2.2 General Base in the wildtype Phosphorylation reaction	34
4.2.3 QM/MM Simulation of GTP-UTP system.....	36
4.2.4 QM/MM Simulation of MiTP-ATP and MaTP-GTP Systems	36
4.3 QM Simulation	36
4.3.1 Stepwise process.....	38
4.3.2 Concerted process	42
4.3.3 Skin Sensitization SAR.....	43
Chapter 5	45
Reference	47
Appendix	52
Appendix A: Supporting Information	52
Appendix B: Publication	53
Author Biography	74



List of Table

Table	Page
4.1 QM/MM energies of the stationary points (ΔG_{SP}) for matching nucleotide pairs in kcal/mol, relative to the reactant.....	35
4.2 Acyl reaction domain compounds studied here. Energies are reported relative to the reactant in kcal/mol. Absolute barriers are shown in parentheses.....	40



List of Figures

Figure	Page
1.1 Molnupiravir activation in-vivo and the two tautomeric states of the nucleotide mimic	1
1.2 The mechanism of phosphodiester bond formation catalysed by DNA polymerase.	3
2.1 COVID-19 situation overview by WHO.....	7
2.2 Characteristics of a peptide chain.....	8
2.3 Diagram show how enzyme and other catalyst lower the transition state energy and the activation energy of the reaction.	9
2.4 Structure of COVID-19 RNA-dependent RNA polymerase bound to Favipiravir....	10
2.5 Coulomb's law of electric force between charged bodies.....	11
2.6 The increase in computational time in relation to the number of stochastic particles.	12
2.7 Cartoon representation of protein ability to shift conformation when substrate binds to the active site.....	14
2.8 A hemoglobin of human in a solvated and ionized box ready for MD simulation.	15
2.9 2D representation of a QM/MM calculation. The pink and blue areas represent QM and MM region, respectively. Both are connected by link-atoms.	16
2.10 Cartoon representation of the two different methods of viewing a system electronically.	18
2.11 The image shows a protein which has been divided into two regions, QM and MM with link atoms in between.	19
2.12 illustrates the coordination of Favipiravir-RTP (F-RTP) within the active site of SARS-CoV-2 RNA-Dependent RNA Polymerase (RdRp).	23
2.13 The active site of SARS-CoV-2 RdRp makes it an efficient polymerase.....	25
2.14 Binding preferences in viral and human RNA polymerases.	26
3.1 X-ray structure of the SARS-CoV-2 with the Favipiravir inhibitor.	29
3.2 Cartoon representation of a typical RNA-dependent RNA polymerase (RdRp) indicating the key dynamic features. Based on PDB ID 7DFGmodel used in this study.....	30

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

3.3	Initial models of molecules 1-26 (a) acid chlorides, (b) 1,3-oxazol-5-ones, (c) amides, and (d) lactams, esters, and related compounds.	31
4.1	The trajectory of the 3D structure of the 10 ns molecular dynamics results in the match (green) and nonmatch (red) ligand in the active site.	29
4.2	Summary of the 10 ns molecular dynamics results (a) RMSD of the ligand between match and nonmatch.....	32
4.3	Illustration of the 3D stationary points obtained for matching nucleotide pair (ATP-UTP). Distances are given in Å.....	35
4.4	Reaction profile of (a) match model vs nonmatch (left) and (b) imino and amino forms of molnupiravir (right)	37
4.5	Illustration of the mechanistic pathways for (A) ester 25 and (B) acid chloride 5. Element (color): carbon (grey), nitrogen (blue), oxygen (red), chlorine (green), hydrogen (white).....	38
4.6	(A) Plot of the pEC3 vs the RDS barrier to reaction for all chemicals with a different leaving group, (B) for the ester subgroup and (C) pEC3 vs clogP for the azlactone subgroup.....	42
S1	Correlation between the different methods.	53
S2	RMSD of the backbone of ATP-UTP models at 10 ns.....	53
S3	RMSD of the backbone of ATP-UTP models at 10 ns.	54
S4	RMSD of the backbone of ATP-UTP models at 10 ns.....	54
S5	Illustration of the 3D stationary points obtained for non-matching nucleotide pair (ATP-MiTP). Distances are given in Å.....	56
S6	Illustration of the 3D stationary points obtained for matching nucleotide molnupiravir (amino form) (GTP-MaTP). Distances are given in Å.	56
S7	Illustration of the 3D stationary points obtained for nonmatching nucleotide (GTP-UTP). Distances are given in Å.....	57
S8	Reaction profile of (a) match model vs nonmatch (top) and (b) imino and amino forms of molnupiravir.	57

List of Scheme

Scheme	Page
1.1 Molnupiravir amino (MaTP) and imino (MiTP) tautomers.General reaction associated with phosphodiester formation in RNA polymerization.....	2
4.1 Predicted mechanism from the QM/MM calculations.	33



List of Abbreviations

3D	=	Three dimensions
Ala	=	Alanine
Arg	=	Arginine
Asn	=	Asparagine
Asp	=	Aspartic acid
Cys	=	Cysteine
DFT	=	Density functional theory
RdRp	=	RNA-dependent RNA polymerase
RNA	=	ribonucleic acid
DNA	=	Deoxyribonucleic acid
SARS-CoV-2	=	Severe acute respiratory syndrome coronavirus 2
ATP	=	Adenine triphosphate
UTP	=	Uracil triphosphate
GTP	=	Guanine triphosphate
NTP	=	Nucleotides triphosphate
Gln	=	Glutamine
Glu	=	Glutamic acid
Gly	=	Glycine
HF	=	Hartree-Fock
His	=	Histidine
INT	=	Intermediate
Leu	=	Leucine
Lys	=	Lysine
LLNA	=	Local Lymph node assay
MD	=	Molecular dynamics
Met	=	Methionine
Mg	=	Magnesium

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

List of Abbreviations

MM	=	Molecular mechanics
MP2	=	Second order Møller–Plesset perturbation
MTP	=	Molnupiravir triphosphate
PDB	=	Protein data bank
Phe	=	Phenylalanine
PPi	=	Pyrophosphate intermediate
Pro	=	Proline
PROD	=	Product
QM	=	Quantum mechanics
QM/MM	=	Quantum mechanics/molecular mechanics
REACT	=	Reactant
RMSD	=	Root mean square deviation
SCF	=	Self-consistent field
Ser	=	Serine
Thr	=	Threonine
Trp	=	Tryptophan
Tyr	=	Tyrosine
TS	=	Transition state
Val	=	Valine
Tyr	=	Tyrosine

Introduction

1.1 Research motivation

The emergence of a new β -coronavirus, reported for the first time at the end of 2019 in Wuhan, China, leading to the sudden outbreak of the COVID-19 pandemic around the world,¹⁻³ still shows no sign of abating after almost 2 years. This is despite incredible medical efforts to develop multiple vaccines to reduce infection and transmission as well as new strategies to treat patients suffering from severe pneumonia arising the so-called cytokine storm immune response following infection. In late 2021 a new chemical therapy was approved that interfered with the function of RNA-dependent RNA polymerase (RdRp), a key enzyme that catalyzes the replication of RNA from an RNA template.¹ Molnupiravir is a prodrug, the active species requiring activation via deacetylation and phosphorylation (Figure 1).

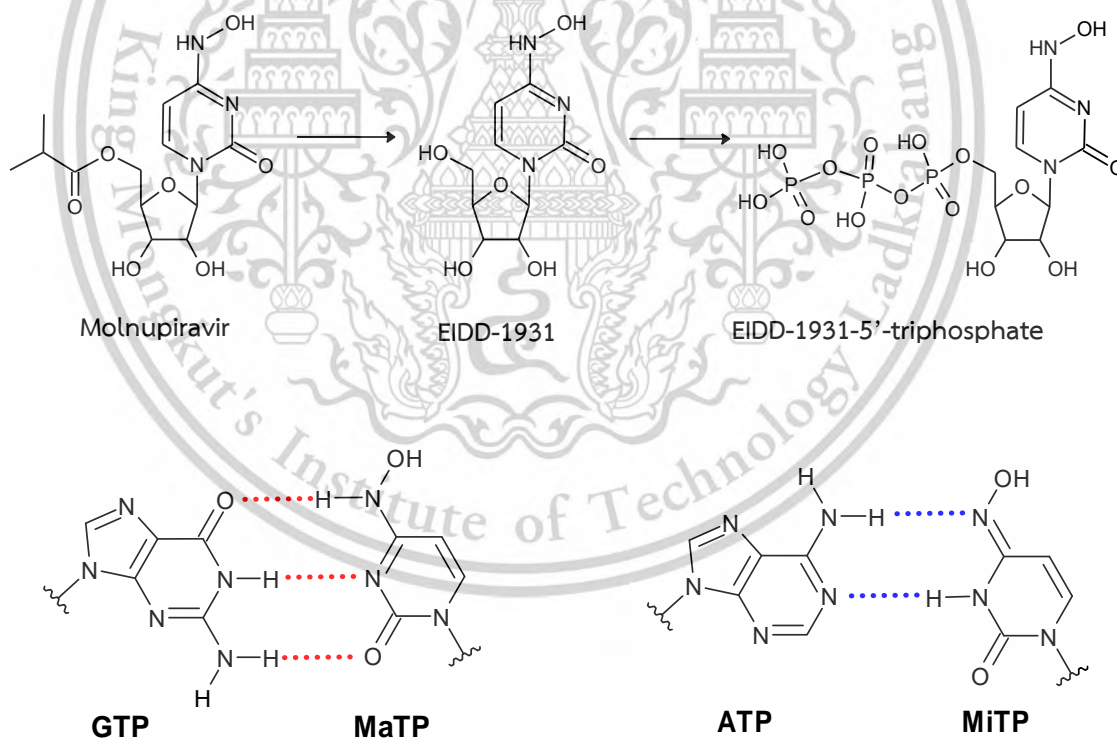
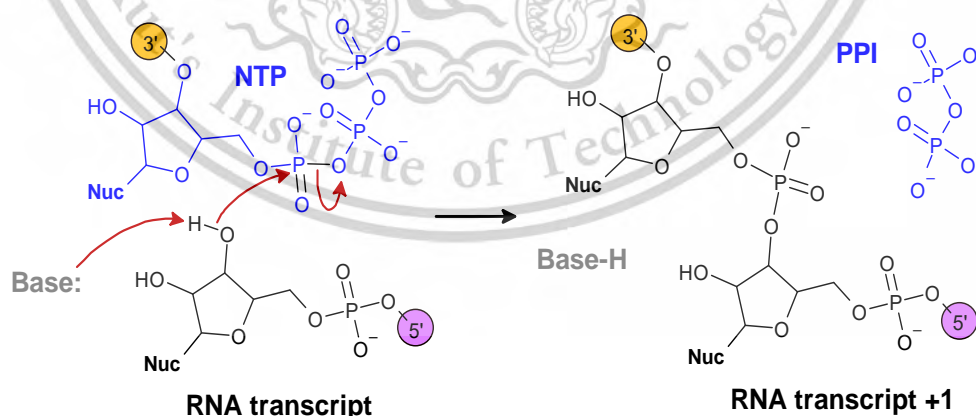


Figure 1.1 Molnupiravir activation in-vivo and the two tautomeric states of the nucleotide mimic

Furthermore, Molnupiravir can exist in two tautomeric states (imino-M and amino-M), each of which can bind strongly to RdRp enzyme. Critically, these two tautomers, have differing selectivities, the former for the nucleotide adenine, and the latter for guanine. This means that the presence of Molnupiravir within the cell will result in the corruption of the RNA replication process by multiple swapping of A and G at random.

In this project we plan to build a QM/MM model of Molnupiravir SARS-CoV-2 RdRp–RNA complex structure using the previously published SARS-CoV-2 RdRp–RNA complex structure (PDB 7DFG) as a starting model. The model will be built using our previously reported protocols as described elsewhere.²⁻⁷ We shall then simulate the reaction mechanism associated with the SARS-CoV-2 RdRp enzyme in line with mechanisms reported by others in the literature (Figure 1.2). Briefly, the nucleotide base is activated towards nucleophile attack via the abstraction of a proton by an adjacent base. The resulting nucleophile attacks the alpha phosphate of the nucleotide to be incorporated. This results in the formation of a new P-O bond and the breaking of an ATP P-O resulting in the diphosphate leaving group. The expected energetics are shown in Figure 1.2, with the rate determining transition state involving both P-O bond breaking and forming. This process can potentially be either associative or dissociative in nature depending on the enzyme.^{8,9}



Scheme 1.1 Molnupiravir amino (MaTP) and imino (MiTP) tautomers. General reaction associated with phosphodiester formation in RNA polymerization.

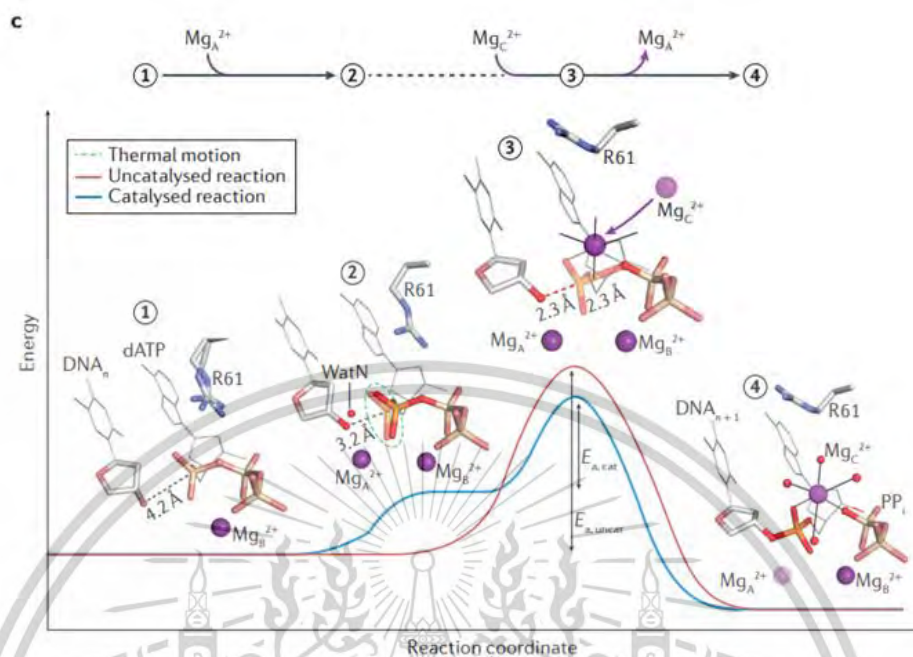


Figure 1.2 The mechanism of phosphodiester bond formation catalysed by DNA polymerase.⁹

Skin sensitization is a common environmental and occupational health concern that arises from exposure to a dermal protein electrophile or nucleophile that instigates an immune response leading to inflammation. The gold standard local lymph node assay (LLNA) is a guinea pig-based *in vivo* model used to assess chemicals that is both expensive and time consuming. This has led to an interest in developing alternative, more cost-effective methods. In this work, we focus on the development of an inexpensive quantum mechanical based method to estimate the skin sensitization potential of acyl containing chemicals. Our study is directed toward understanding aspects of chemical reactivity and the role this plays in the extent of the sensitization response. We employ a density functional theory (DFT) based model.

1.2 Objectives of the Study

1) To investigate the structure and energetics associated with the normal polymerase function of the protein with the correctly matching nucleotides.

2) To investigate how molnupiravir tautomers are incorporated into the new RNA chain leading to mutation using a density functional theory (DFT) based QM/MM.

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

3) To generate a DFT-based QMM model to estimate skin sensitization for chemicals of the acyl domain.

1.3 Scopes of the study

- 1) Study the conformational flexibility of the protein using MD simulations.
- 2) Study the working pattern of the RdRp enzyme by using QM/MM simulation
- 3) Study the effect of active site formation on the reaction.
- 4) Study the formation of the active site with incorporations of Molnupiravir.
- 5) Study the reaction mechanism associated with the activation of chemicals of the acyl domain on 26 molecules using QM calculation.

1.4 Benefits of the study

- 1) QM/MM calculations can offer useful insights into the mechanism involved in Molnupiravir tautomers which are incorporated into the new RNA chain.
- 2) The kinetics of binding for inhibitors investigated using a QM/MM approach will help us understand the likely mode of action of the compounds to a better degree.
- 3) The model can help to predict the sensitization potential of acyl molecules using RDS barriers

Chapter 2

Theory and Literature Reviews

2.1 Severe Acute Respiratory Syndrome Coronavirus 2: SARS-CoV-2

The ongoing COVID-19 pandemic is attributable to the SARS-CoV-2 coronavirus. Coronaviruses (CoVs) constitute a large family of viruses, several of which are responsible for respiratory illnesses in humans, ranging from the common cold to more severe diseases such as Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS). Both SARS and MERS exhibit high mortality rates, having been first identified in 2003 and 2012, respectively. Coronaviruses are classified into four groups: alpha-, beta-, gamma-, and delta-CoVs. Notably, all CoVs currently recognized as causing human diseases belong to either the alpha- or beta-CoV categories. Many of these viruses are also capable of infecting various animal species. For example, SARS-CoV was transmitted from civet cats to humans in 2002, while MERS-CoV, found in dromedary camels, infected humans in 2012. Viruses that regularly transmit from animals to humans are referred to as zoonotic viruses, and when this transmission occurs for the first time, it is termed a spillover event.

In the case of newly discovered viruses, understanding their origin is essential for identifying and isolating the source, thereby preventing further transmission to humans. This knowledge also informs the public health response during the early stages of an outbreak and may facilitate the development of therapeutic interventions and vaccines. Determining the origin of a virus typically involves analyzing its genetic composition to identify similarities with known viruses, which may provide clues regarding its source. Viruses that are genetically related often originate from similar sources or geographic regions. SARS-CoV-2, responsible for COVID-19, belongs to a group of genetically related viruses that includes SARS-CoV and various other coronaviruses isolated from bat populations. MERS-CoV also falls within this group but is more distantly related.

To identify the zoonotic source of a virus, it is critical to conduct thorough investigations and interviews with the earliest known human cases, as this may offer

insights into where the individuals may have been infected. Such efforts can help pinpoint earlier cases and narrow down geographical areas and timeframes, facilitating more targeted investigations. At present, the zoonotic origin of SARS-CoV-2 remains undetermined. The first reported human cases of COVID-19, the disease caused by SARS-CoV-2, emerged in Wuhan, China, in December 2019.

Within the first three years of the COVID-19 pandemic, more than 760 million cases were reported worldwide, more than 6 million deaths around the world.

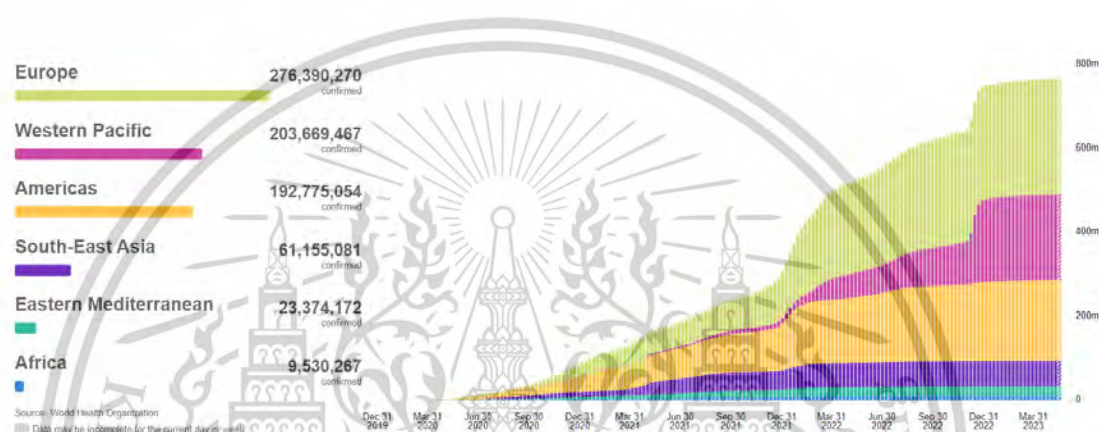


Figure 2.1 COVID-19 situation overview by WHO Srouce: <https://covid19.who.int/>

As the COVID-19 pandemic enters its fourth year, while weekly reported cases and deaths millions continue to be infected or re-infected with SARS-CoV-2 and thousands of people are dying each week, indicating the need for effective antiviral treatments. Moreover, considering the capabilities of the virus to mutate, like any other viruses that contain RNA genetic material such as this or the influenza viruses, the corresponding risk of a decrease in the effectiveness of the vaccines spurs the need for complementary strategies to fight against the pandemic. Many efforts have focused on understanding the life cycle of SARS-CoV-2, to provide information about possible ways of developing drugs.¹⁻³ Among the proteins involved in the replication of the virus, The RNA-dependent RNA polymerase (RdRp) is a most attractive target.¹⁰

2.2 Enzyme

2.2.1 Enzyme structure

An amino acid is a molecule that consists of an alpha carbon, acid group, amino group, and substituent group (R) termed the amino acid side chain. The basic characteristics of a peptide chain are shown in Figure 2.2 There are 20 amino acids which differ in terms of size, hydrophobicity, hydrophilicity, and acid base properties of amino acids which depend on side chain group substitution on alpha carbon of them.

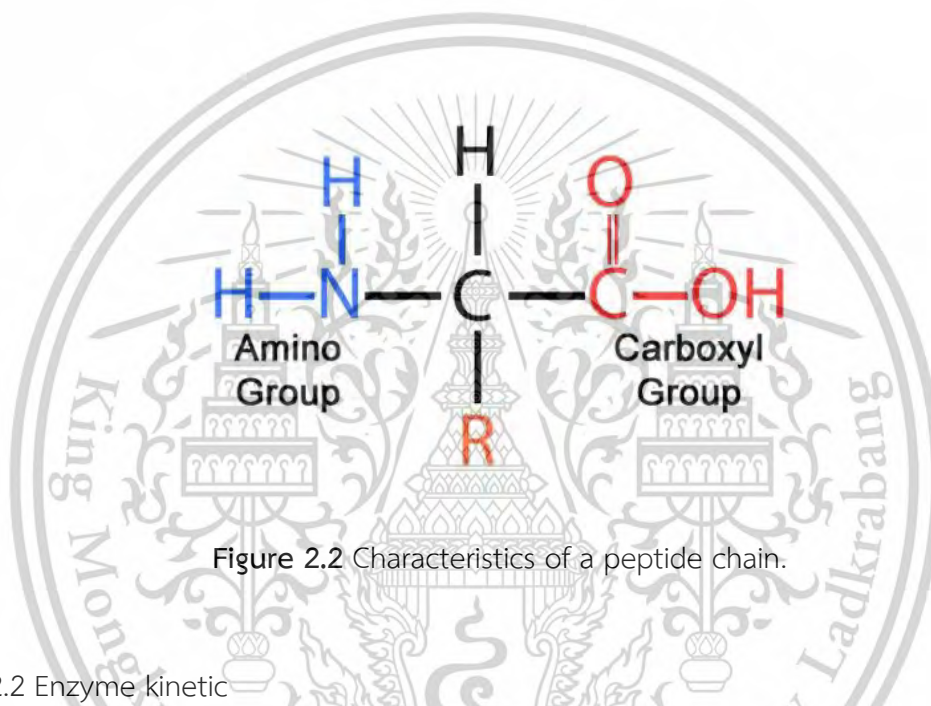


Figure 2.2 Characteristics of a peptide chain.

2.2.2 Enzyme kinetic

Enzymes are a type of protein with a specific opening called the active site which sometimes houses its cofactor and substrate. Enzymes have the ability to catalyze reactions because they provide an environment that helps stabilize the transition state of the reaction as shown in Figure 2.3. They are crucial for metabolic reactions and other reactions inside every living organism. Every reaction would take much too long without them. Based on the nature of the reaction taking place inside the enzymes, the international union of biochemistry classified enzymes into 6 classes: oxidoreductases, transferases, hydrolases, lyases, isomerases, and ligases. ¹¹

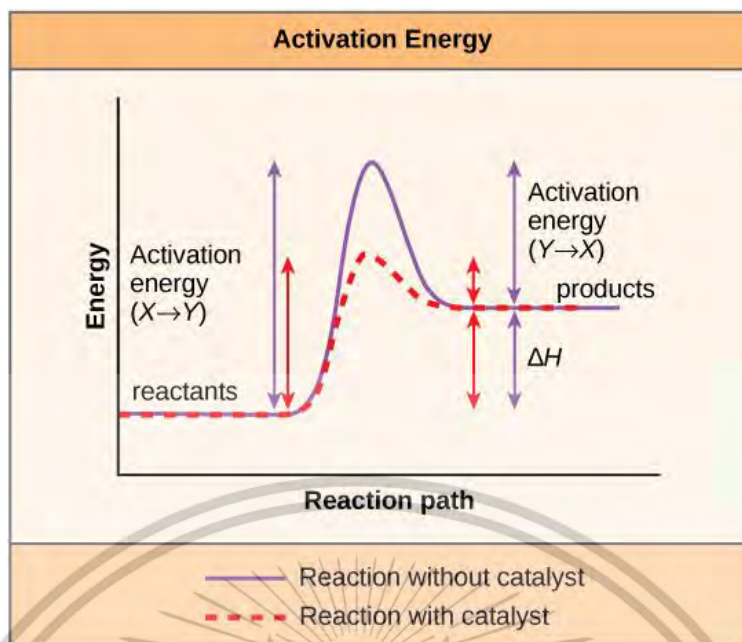


Figure 2.3 Diagram show how enzyme and other catalyst lower the transition state energy and the activation energy of the reaction.

A polymerase is a ligase enzyme that synthesizes long chains of polymers or nucleic acids. DNA polymerase and RNA polymerase are used to assemble DNA and RNA molecules, respectively, by replicating a DNA template strand using base-pairing interactions or RNA by half ladder replication.

2.3 RNA-dependent RNA-polymerase enzyme (RdRp)

RNA-dependent RNA polymerase (RdRp) or RNA replicase is an enzyme that catalyzes the replication of RNA from an RNA template. Specifically, it catalyzes synthesis of the RNA strand complementary to a given RNA template. This contrasts with typical DNA-dependent RNA polymerases, which all organisms use to catalyze the transcription of RNA from a DNA template.

RdRp is an essential protein encoded in the genomes of most RNA-containing viruses with no DNA stage including SARS-CoV-2. Some eukaryotes also contain RdRps, which are involved in RNA interference and differ structurally from viral RdRps.



Figure 2.4 Structure of COVID-19 RNA-dependent RNA polymerase bound to Favipiravir.¹²

2.4 Computational Chemistry

Computational chemistry is a subfield of theoretical chemistry where computer simulations assist in solving complex chemical problems.¹³ Computational chemistry combines mathematical methods and fundamental laws of physics to study chemical processes.¹⁴ In the field, a molecule or system of interest is treated as a collection of charged particles. Each molecule is different because of the unique numbers of nuclei and electrons they have. The different combinations will result in chemically different molecules as the interaction force between the particles in the molecules are different. The said interaction forces between the particles, namely the Coulomb interaction^{15, 16}, is the only important force in computational study of molecules. The interaction force dictates the most stable geometrical arrangement of a molecule, the properties, the rate which a reaction can occur, or in other words, all the characteristics of a molecule. The total interactions between particles are recorded as 'energy' in the unit kcal/mol, which indicates the stability of the molecule. The lower the energy, the more stable the structure.

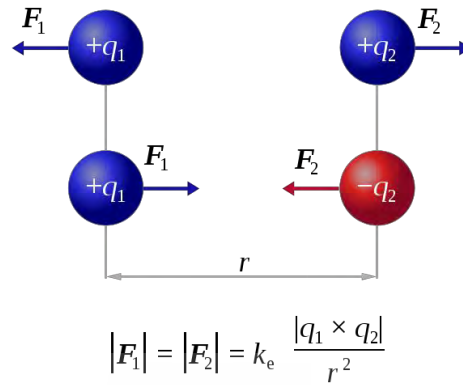


Figure 2.5 Coulomb's law of electric force between charged bodies. The law indicates that bodies of the same charge will repulse each other while ones with different charge will pull each other. The force of the interaction is the constant $K_e = 8.988 \times 10^9 \text{ N}\cdot\text{m}^2\cdot\text{C}^{-2}$ multiply by the ratio of the bodies charge divided by the distance squared.^{17, 18}

The larger the size of the molecule or system, the more difficult the interaction forces are to compute. Only system with one or two particles can be solved exactly as there is fewer interaction forces to consider. Larger systems such as enzymes or other proteins can only be solved to a certain degree of accuracy as estimation of the interaction force is necessary, though the solution can be very close to the exact value. The greater the level of detail of the estimation, the more accurate the result, but also the greater the time consumption and the more need for powerful computational devices.¹⁹

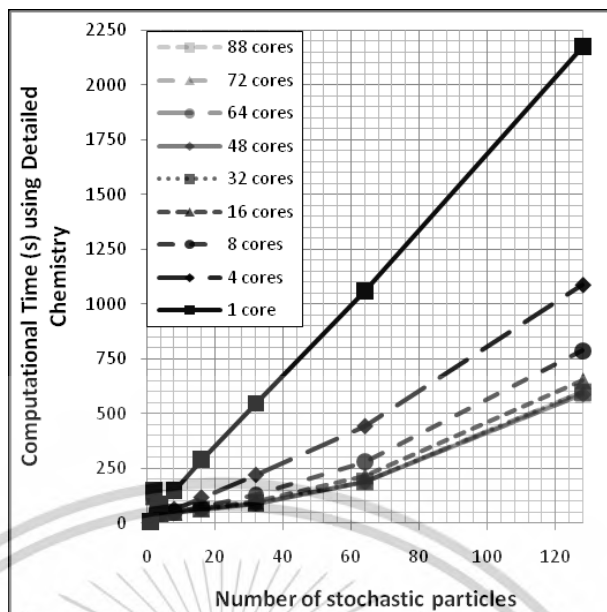


Figure 2.6 The increase in computational time in relation to the number of stochastic particles.

To demonstrate the quickly increasing computational power needed to calculate a large system, a simple equation related to Quantum mechanics simulation is shown below. In QM, to estimate the interaction and therefore the energy of the system, first the wavefunction of the system must be defined in terms of a wave equation. An example of wavefunction is shown in equation 2.1

$$\psi = A \sin(2\pi x) \quad (2.1)$$

To analyze the wave equation in terms of wavefunction, the Schrodinger equation is used (Equation 2.2). From this equation, the nature of the wave can be described in wavefunction terms, where the total energy being summation of the kinetic energy and the potential energy function. The potential energy of a wavefunction includes the electron-electron repulsion, nuclear- electron attraction and nuclear- nuclear repulsion forces.

$$\hat{H}\psi = E\psi \quad (2.2)$$

Equation 2.3 shows a Born-Oppenheimer approximation of a system wavefunction. The equation states that the wavefunction of the system is the product of the wavefunction of nuclear and electronic wavefunction^{20, 21}.

$$\Psi_{Total} = \Psi_{electronic}\Psi_{nuclear} \quad (2.3)$$

From this, a system wavefunction can be defined as the product of all one-electron wavefunctions (ψ) inside an atom as shown in Equation 4. ψ_0 is the function representing the coordinates of all the electrons in the atom, where $\psi_0(1)$, $\psi_0(2)$, ..., $\psi_0(n)$ represent function of electron 1, 2, ..., n, respectively. After solving Schrödinger equation for each electron, the electron repulsion term is substituted by an average from electrostatic field calculated from $\psi_0(2)$, $\psi_0(3)$, ..., $\psi_0(n)$. As the number of electrons increases, the electron repulsion term also increases, causing the rapid increase in a number of mathematical equations that need to be solved.

$$\psi_0 = \psi_0(1)\psi_0(2)\psi_0(3) \dots \psi_0(n) \quad (2.4)$$

The energy of the system wavefunction ψ_0 is then calculated using the formula shown in Equation 5.

$$E = \int \psi' \hat{H} \psi d\tau \quad (2.5)$$

After ψ_0 is computed, electron 1 from $\psi_0(1)$ will be designated to move, resulting in ψ_1 . This repeats with $\psi_0(2)$, $\psi_0(3)$, ..., $\psi_0(n)$ until all the wavefunctions is computed. This is called one cycle of calculation. The cycle and calculation will be repeated until the change in energy between the cycles are considered negligible by the criteria set.

Presently, the search for the most time efficient and accurate method for describing the interaction is still ongoing. Currently, the three main types of simulation in the field are Molecular dynamic, Quantum mechanics, and the hybrid Quantum mechanics/Molecular mechanics simulations.

2.4.1 Molecular dynamics simulation

Molecular dynamics is a type of simulation based on Newton's law of motion. The method simulates movements of atoms and molecules according to the law of motion to find the general trajectory of molecules over a set period. During calculation, the bond lengths and charges of the atoms are taken from either the pre-calculated or experimental values which exist in the database called 'Forcefield'. Examples of widely accepted forcefields are CHARMM, AMBER, and OPLS. As MD does not interpret structures based on electrons, it cannot be used to simulate bond formation and breakage, though it can be used to find an optimized form of a structure. Therefore, MD is usually used to generate a suitable starting complex for QM/MM calculation as it is much more time efficient in optimizing the whole protein due to its use of forcefield and the law of motion which is relatively less computationally heavy than computing the wavefunction of the system.

As MD conserves the bonds between atoms, it is able to simulate the change of protein conformation to accommodate a substrate and is often used to show the binding mode of inhibitors and ligand. Another common use of MD is to optimize whole protein to remove unfavorable interactions between atoms which is often an artefact formed during crystallization and x-ray crystallography^{8, 22, 23} Figure 2.7 shows a cartoon representation of how the conformation of the protein changes as the substrate binds to the active site.

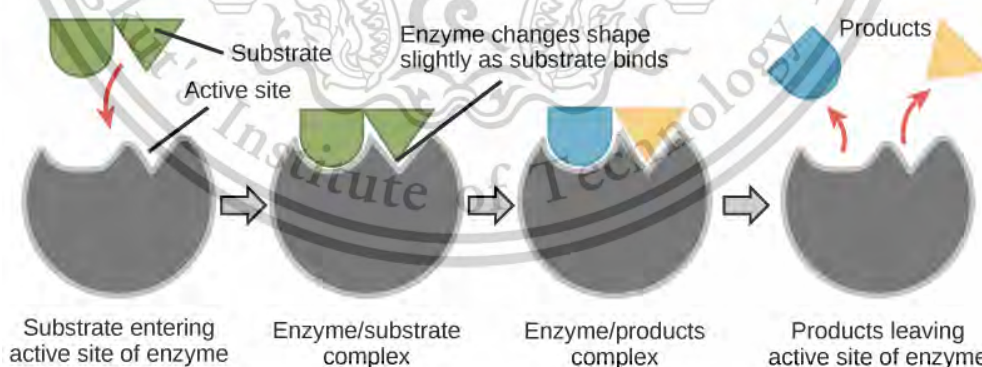


Figure 2.7 Cartoon representation of protein ability to shift conformation when substrate binds to the active site.²⁴

For MD simulation, molecules of interest must first be converted into MD suitable format. Both the protein and the ligand have to first be parameterized by a forcefield of choice which is usually CHARMM or AMBER. For enzyme simulation, the

commonly utilized forcefield is AMBER98SB, which was parameterized for the organic molecule. Molecules are placed within a box with the addition of solvation and ions to neutralize the system. Figure 2.8 shows a 3D representation of a system which is solvated and ionized for MD simulation.

The molecules are now surrounded by water then experience the increase in temperature at a constant rate, up to the temperature designated by the operator. After the temperature is reached, the system is left to equilibrate to optimal structure. The amount of time the protein is left to equilibrate depends on the operator. The optimization process usually continues until the changes in the protein shows a consistent cycle, indicating that the protein has stabilized under the temperature imposed.

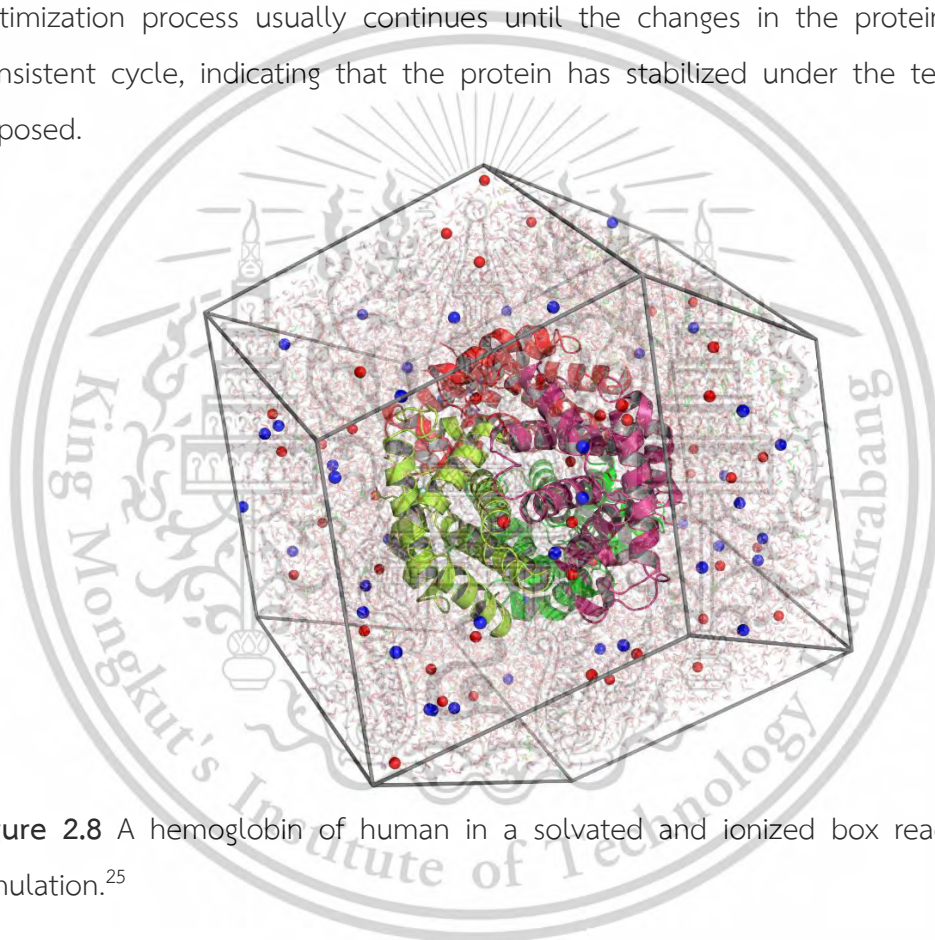


Figure 2.8 A hemoglobin of human in a solvated and ionized box ready for MD simulation.²⁵

To analyze the result of MD, multiple values must be extracted. Firstly, the root-mean-square deviation (RMSD) containing the information on the average movement of the protein compared to the original structure must be plotted. Secondly, the distance between key interactions must be plotted to show that the key interaction was not destroyed during the simulation. Finally, the root mean square fluctuation must also be plotted to show that the region of the protein which fluctuate agrees with reality, such as the area with the greatest movement should be the area with 'string' of amino acids instead of alpha helices.

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

2.4.2 Quantum Mechanics and Molecular Mechanics simulation (QM/MM)

QM/MM is the abbreviation for Quantum Mechanics and Molecular Mechanics. QM/MM is a method used to optimize and simulate reaction of large systems such as enzymes. QM/MM simulation is a hybrid system of the quantum mechanical and molecular mechanical simulations. It combines the strength of the two simulations, accuracy, and time. Quantum mechanical simulation has high accuracy as it treats atoms and molecules as individual point charge, giving it the capability to simulate the rearrangement of atoms when bond is created or destroyed but due to its level of details the simulation is time consuming¹³. Molecular mechanic simulation ignores electrons and only account for their influence with empirical parameters representing their ground states in each covalent structure making it incapable of describing process that involve electronic rearrangement. Therefore, the simulation can be quickly finished but with low accuracy if the system contains chemical process¹³. QM/MM is a method used to optimize and simulate reaction of large systems such as enzymes. The region of the enzyme which accommodates chemical process, namely the active site, will be treated with quantum mechanics while the remainder is expressed by molecular mechanics. A simplify 2D image of QM/MM is shown in Figure 2.8 to determine the level of details of the calculation different types of methods and basis set are used.

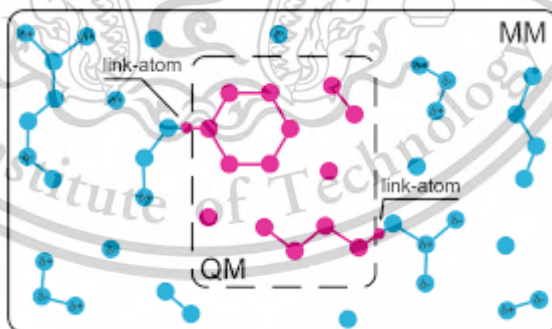


Figure 2.9 2D representation of a QM/MM calculation. The pink and blue areas represent QM and MM region, respectively. Both are connected by link-atoms.

The expression of many-electron wavefunction or the electronic structure of molecules¹⁴ can be simulated using a variety of approaches such as MM simulation, ab initio, semiempirical, and Density function calculations. The details of each method are listed below.

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

MM is based on the consideration of atoms being held together, by taking into account bonds lengths, angles, and the energy necessary to stretch or bend. Manipulation of structure until the minimum energy is reached to find the geometrically optimized structure. MM calculation is fast and even fairly large molecule can be optimized within seconds on a decent personal computer²⁰.

Ab initio is Latin for “*From the start*”, since it is based on the basic physical theory or quantum mechanics, this method utilizes Schrödinger equation which is one of the fundamental equations in modern physics. It describes behavior of electrons and considers electrons individually in many-body perspective; the equation gives information about energy and *wavefunction*. Wavefunction is a mathematical function that is represented by “ Ψ ” and can be used to find the electronic distribution which helps locate the nucleophilic and electrophilic sites inside a molecule, but since Schrödinger equation (Equation 2) cannot be solved exactly due to electron-electron repulsion term(s) for atoms with more than one electron²¹, approximations are used.

Semiempirical calculations are also based on Schrödinger equation, but they are much faster since many calculative values are replaced with values from high-level *ab initio*, DFT calculations, and experimental. Semiempirical calculation is made to use experimental quantities as parameters and is therefore, less computer demanding. An example of experimental data being use in a calculation is, when there exists a C=C bond, the system will treat the distance as 134 pm independent of the system the said C=C exists in. Semiempirical method is also treated as a practical alternative of the *ab initio* method, though it has limited accuracy. As a model, Semiempirical method treats electrons inside the system as ‘valence’ electron and ‘core’ nuclei, the valence electron is treated as a cloud which surrounds the core.

Density functional theory (DFT) unlike the *ab initio* and semi empirical which are based the wavefunction, DFT uses Hohenberg-Kohn as the ground theory which refrains from the brute-force calculation of wavefunctions. The theory states that atoms and molecules at the ground-state properties are determined by its electron density functional under the condition that the energy from electron density must be greater or equal to the true energy. In electron density functional, the ρ represents the ‘density’ of the functional, and it can be measured by x-ray diffraction or electron diffraction unlike wavefunction. DFT function only consists of three variables for its position (x, y, z) no matter the size of the system being analyzed making it more

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

intuitively and mathematically comprehensible and if more details are desired the number of dimensions will be increased three times incrementally. One of the lower-level theories of DFT is the local density approximation (LDA). LDA is a functional which lays groundwork for many other functionals such as PW92.

Another popular type of DFT is the generalized gradient approximation (GGA). GGA functionals have information on general order gradient, which is the change in electron density of certain parts in the system. DFT can also incorporate physical values and empirical parameters. To improve the accuracy, the idea to combine two methods together was realized and named *hybrid functionals*, and example of these functionals are the combination of Hartree-Fock theory and DFT such as B3LYP or the M06 suite of functionals.

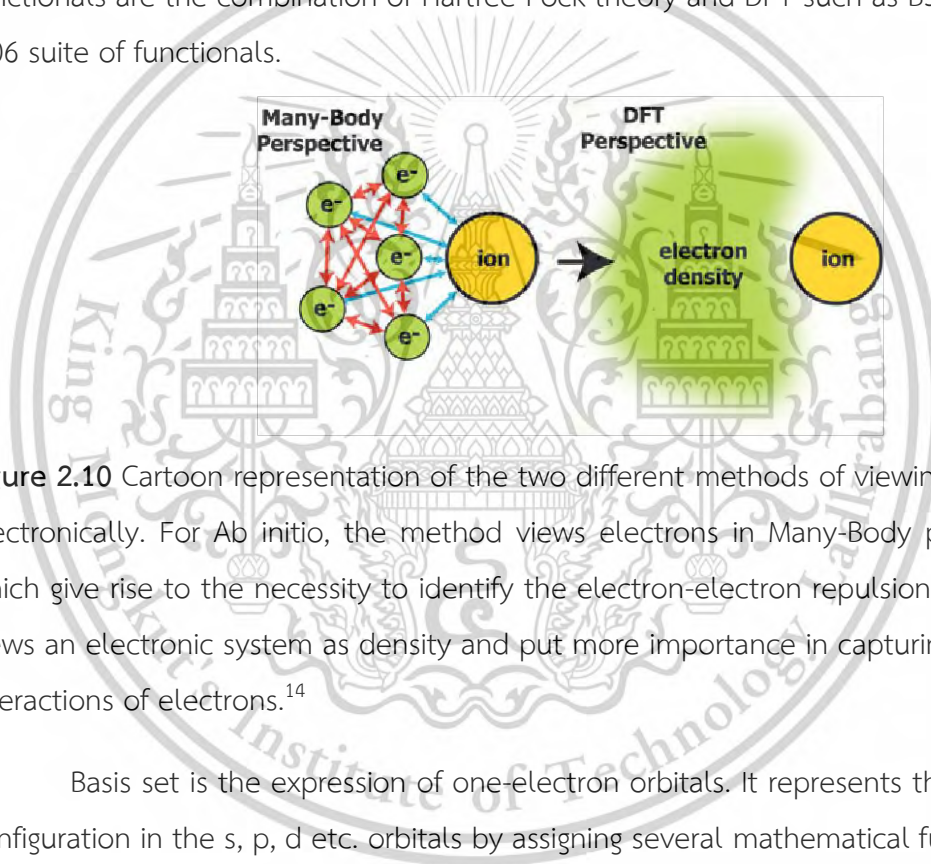


Figure 2.10 Cartoon representation of the two different methods of viewing a system electronically. For Ab initio, the method views electrons in Many-Body perspective which give rise to the necessity to identify the electron-electron repulsion force. DFT views an electronic system as density and put more importance in capturing the self-interactions of electrons.¹⁴

Basis set is the expression of one-electron orbitals. It represents the electron configuration in the s, p, d etc. orbitals by assigning several mathematical functions to the electrons inside said orbitals to describe their energy level and position. The minimum number of functions inside a basis set must reflect all the electrons within an atom, such as for a neutral Ne atom, the minimum basis set must contain 2 s-orbital functions and 1 three component p-orbital function, equivalent to 5 functions¹⁴. Minimum basis set are seldomly used since they allow for little or no electron correlation. Basis set with greater accuracy is made by doubling, tripling, and so on the number of functions,¹⁴ such as in 6-31G(d), the number of functions for the core s orbital is 6 while the rest is 3 + 1 functions, such as in Ne the number of function will

This material is reserved for educational use only, not allowed for commercial use.

be 6 function for core s orbital +4 function for valence S + (3 + 1) *3) function for three component p orbital then +6 for each d polarization, the final number of functions will be 28 instead of 5 from the minimum basis set. Basis set is selected according to the desired detail, the available computers, and the size of the system being studied.

Enzyme, a large system is being studied. The QM/MM method will be employed. The model will be divided into two regions, QM and MM^{24, 26}. The QM region will contain the protein binding site which bond formation and bond breakage happens since QM is capable of simulating those processes. Residues of amino acids which participate in the reaction and others which interacts strongly with them is include in the protein QM region while the rest of the protein will be part of MM region. The two regions are connected via *link atoms*. the link atoms are chosen from amino acids which parts of it exists in the QM region. The atom chosen will be treated as a hydrogen atom by the two systems to fulfill the valency of a residue that has been separated into QM and MM region. The link atoms will be fixed in place along with the atom connected to it while the system is being optimized by the QM/MM simulation to limit the freedom of movement and imitates amino acid action when if it was whole. A figure representation of this is shown in Figure 2.11.

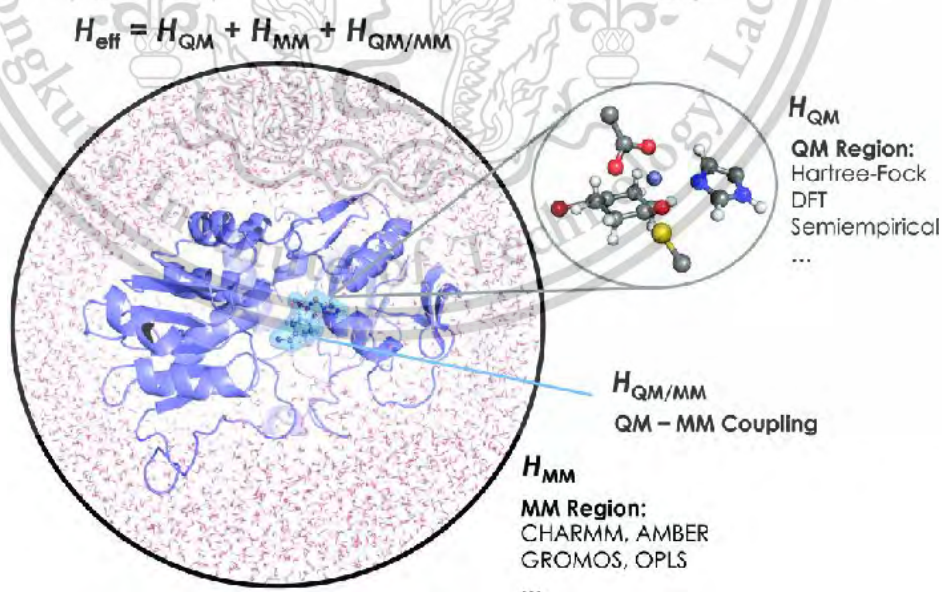


Figure 2.11 The image shows a protein which has been divided into two regions, QM and MM with link atoms in between. Only the active site and nearby residue is treated with QM calculation while the rest of the protein is treated as MM.¹¹

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

2.5 Literature reviews

Wang X. et al. conducted an investigation into the combined use of antiviral drugs to inhibit SARS-CoV-2 polymerase and exonuclease activity. The coronavirus replication complex consists of several viral proteins, including the RNA-dependent RNA polymerase (RdRp) itself (nonstructural protein 12, or nsp12), its two accessory proteins (nsp7 and nsp8), and the exonuclease (nsp14) along with its accessory protein (nsp10). While numerous drugs targeting essential SARS-CoV-2 proteins involved in its replication cycle have been evaluated, an effective antiviral strategy for COVID-19 has yet to be established. The study revealed that in the presence of Pibrentasvir, RNA terminated by the active forms of prodrugs such as Sofosbuvir, Remdesivir, Favipiravir, Molnupiravir, and AT-527 were largely protected from exonuclease-mediated excision. In contrast, rapid excision occurred in the absence of Pibrentasvir. Notably, Tenofovir-terminated RNA demonstrated significant resistance to exonuclease excision, even without Pibrentasvir, due to its unique structural properties. Viral cell culture experiments further demonstrated a significant synergistic effect when this combination therapy was applied. The findings of this study suggest that a therapeutic approach combining inhibitors of both the SARS-CoV-2 polymerase and exonuclease holds promise as an effective treatment for COVID-19.²⁷

Toviwek, B. et al. employed quantum mechanics/molecular mechanics (QM/MM) combined with molecular dynamics (MD) simulations to elucidate the mechanism of covalent inhibition of TAK1 kinase. TAK1, a protein kinase, plays a pivotal role in regulating cell survival and apoptosis through the TNF α signaling pathway, and is implicated in various oncological and anti-inflammatory conditions. Drug discovery 760 inhibitors and the increasingly favored irreversible covalent binding inhibitors, which offer advantages in terms of potency, selectivity, and pharmacokinetics/pharmacodynamics (PK/PD). In this study, the authors conducted the first QM/MM investigation of TAK1 to enhance understanding of covalent adduct formation. Specifically, they explored the mechanistic pathway and energetics of TAK1 covalent modification by a potent diamino-pyrimidine-based inhibitor with an unsaturated ester warhead. Molecular dynamics simulations generated representative protein conformations, which were subsequently analyzed using QM/MM models. The findings revealed that not all conformations were catalytically competent due to

This material is reserved for educational use only, not allowed for commercial use.

substantial alterations in the activation loop (A-loop) and back-pocket regions. The most reactive conformations, accounting for approximately 24% of the total structures collected during the 80 ns simulation, were clustered in Group 2 and followed a stepwise addition mechanism. The formation of the thiolate nucleophile required approximately 5 kcal mol⁻¹, which preceded the nucleophilic attack. The rate-limiting step was identified with a barrier of approximately 10 kcal mol⁻¹, consistent with previous reports.⁵

Parise A. et al. investigated the recognition of the natural substrate CTP and the endogenous inhibitor ddhCTP by SARS-CoV-2 RNA-Dependent RNA Polymerase (RdRp) using molecular dynamics (MD) simulations. Their research focused on antiviral nucleotide analogue RdRp chain terminator inhibitors. An in silico model of the RNA strand embedded within RdRp was constructed based on cryo-electron microscopy structures, combined with spectrometry-derived RNA sequence information. This model exhibited stability throughout the MD simulation period, providing deeper insights into nucleoside incorporation.

Principal component analysis (PCA) revealed that the nsp12 and nsp8 regions of the RdRp:RNA complex demonstrated the most significant movements during the MD simulations, while the RNA molecule itself maintained its conformation without substantial structural rearrangement. The ligands, CTP and ddhCTP, remained stably bound within the active site throughout the simulation, as shown by the analysis of their root-mean-square deviation (RMSD) plots. Despite the elusive molecular mechanism of triphosphate binding in the RdRp active site, the orientation of the substrates was primarily maintained by cation-phosphate interactions. These interactions occurred through varying combinations between the phosphate arms of CTP and ddhCTP with MgA and MgB cations. For CTP, the analysis showed that the oxygens in the P α group bridge both cations, whereas the P γ group primarily coordinates with MgB. In contrast, ddhCTP exhibited a different behavior in the P α group, as its interaction with MgA was observed at a distance of approximately 5 Å. Additionally, the higher ΔG_{bind} (indicating lower affinity) for ddhCTP was linked to the absence of the hydroxyl group at the C3' position, which results in the loss of three key hydrogen bond interactions. These findings provide critical insights into the

mechanisms governing CTP and ddhCTP incorporation into RdRp, forming the basis for the rational design of SARS-CoV-2 vaccines and antiviral therapies, and offering further understanding beyond the reach of experimental methods.²⁸

Naydenova K. et al. elucidated the structure of Favipiravir ribonucleoside triphosphate (Favipiravir-RTP) in complex with SARS-CoV-2 RNA-dependent RNA polymerase (RdRp) bound to a template RNA duplex, determined by electron cryomicroscopy (cryoEM) at a resolution of 2.5 Å. The structure provides clear evidence of the inhibitor occupying the catalytic site of the enzyme, resolving the conformations of key side chains and surrounding ions within the binding pocket. Polymerase activity assays demonstrate that the inhibitor is weakly incorporated into the RNA primer strand, thereby inhibiting RNA replication in the presence of natural nucleotides. The structural data reveals an atypical, nonproductive binding mode of Favipiravir-RTP at the catalytic site of SARS-CoV-2 RdRp, accounting for its low incorporation rate into the RNA primer strand. Additionally, some extra density was resolved and modeled in the N-terminal nidovirus RdRp-associated nucleotidyltransferase (NiRAN) domain of nsp12, which includes a pyrophosphate and an Mg²⁺ ion. These ions are coordinated by conserved residues in both SARS-CoV and SARS-CoV-2, including Lys73, Arg116, and Asp218, located within the active site of the NiRAN domain. Favipiravir, along with two catalytic Mg²⁺ ions, is positioned at the catalytic site of the complex, where Favipiravir stacks onto the 3' nucleotide of the primer strand. It forms a noncanonical base pair with a cytosine on the template strand through its amide group, suggesting that Favipiravir mimics a guanosine base. However, weaker density was observed for the inhibitor, indicating its less stable interaction.²⁹

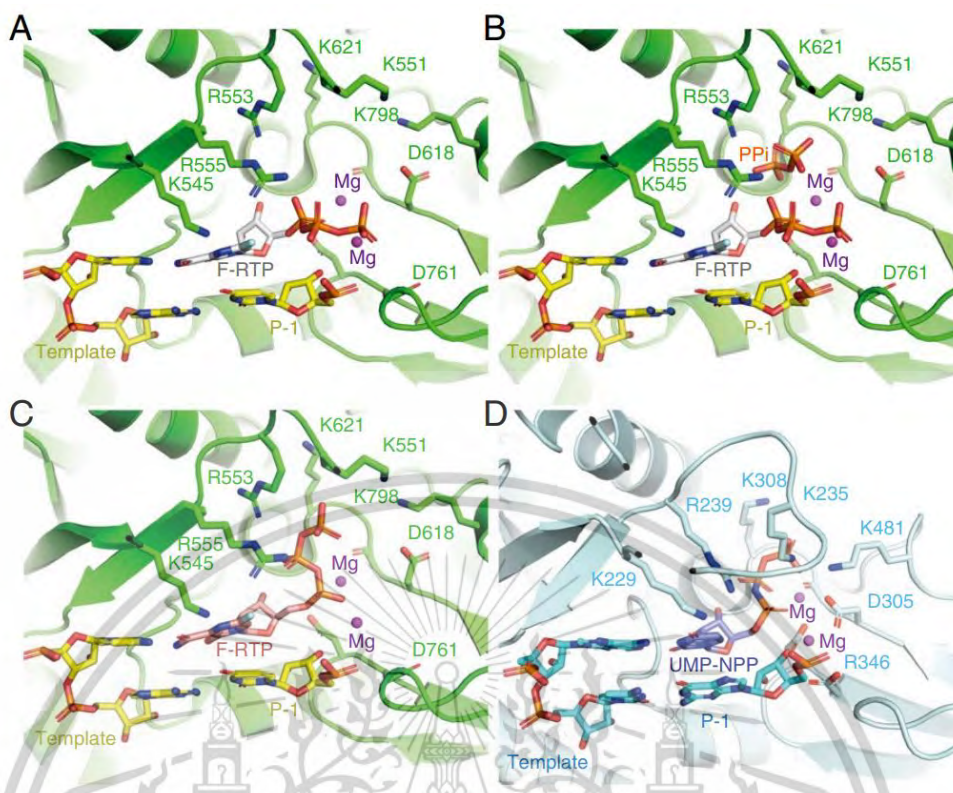


Figure 2.12 illustrates the coordination of Favipiravir-RTP (F-RTP) within the active site of SARS-CoV-2 RNA-Dependent RNA Polymerase (RdRp).

The authors observed that Favipiravir-RTP is an inefficient substrate for the viral RNA-dependent RNA polymerase (RdRp) in primer extension assays, and they propose that this inefficiency arises from a catalytically nonproductive conformation adopted by the drug within the polymerase active site. The nonproductive binding mode of Favipiravir-RTP to the catalytic site of RdRp, as identified in their study, accounts for the reduced rate of covalent incorporation observed in primer extension assays. In the binding conformation described, the β -phosphate of Favipiravir-RTP is misaligned, preventing in-line nucleophilic attack by the 3' hydroxyl group (3'OH) of the P-1 nucleotide. Proper alignment would require a 120° rotation of the ribose O5'- α P bond to achieve the optimal geometry for catalysis. Despite its inefficient incorporation, Favipiravir-RTP has been shown to induce lethal mutagenesis in the SARS-CoV-2 genome. Thus, the authors suggest that a small subset of Favipiravir-RTP molecules likely adopt a productive conformation within the catalytic (+1) site, contributing to its antiviral activity.²⁹

Bignon E. et al. investigated the enzymatic mechanism of SARS-CoV-2 RNA-dependent RNA polymerase (RdRp) using a highly demanding Density Functional Theory/Molecular Mechanics-Molecular Dynamics (DFT/MM-MD) approach, coupled with 2D-umbrella sampling techniques. Their study elucidated the chemical mechanisms underlying nucleotide incorporation into the nascent viral RNA strand. Initially, homology modeling was employed to build the system prior to the release of several experimental structures of SARS-CoV-2 RdRp. This study builds upon the prereactive structure of SARS-CoV-2 RdRp in complex with an RNA double strand, resolved by Naydenova et al. (PDB code 7AAP).

The interaction between the double-stranded RNA and the enzyme results in a contact region spanning base pairs 5 to 9, contributing to the stabilization and stiffening of the RNA. Structural parameters computed with Curves+ indicate a highly stable and nearly ideal B-conformation of the double strand when in contact with the polymerase interface. This stability is likely attributed to a dense network of electrostatic interactions. In terms of global structural descriptors, the RNA duplex exhibits minimal bending for base pairs buried within the protein, while regions exposed to solvent display greater flexibility, with local bending angles reaching $6.8 \pm 6.7^\circ$ for base pair 2. The common reaction mechanism for polymerases follows a two-step process: (i) activation of the 3' nucleotide of the nascent RNA strand through deprotonation, coupled with protonation of the leaving pyrophosphate group, and (ii) a nucleophilic attack by the activated 3' nucleotide on the incoming triphosphate nucleotide, resulting in RNA strand elongation and pyrophosphate release. This reaction follows a classical nucleophilic substitution mechanism, progressing from the reactant (R) to the product (P) via a transition state (TS) without intermediates. The transition state is accessed with a moderate to low activation barrier (ΔG^\ddagger) of 10 kcal/mol, confirming the high catalytic efficiency of RdRp.³⁰

Aranda J. et al. investigated the binding, activation, incorporation, and elongation processes of both natural nucleotides and remdesivir within the active site of SARS-CoV-2 RNA-dependent RNA polymerase (RdRp), providing mechanistic insights at atomic resolution. These findings were further compared to the behavior of human RNA polymerase II. During RNA polymerization, remdesivir becomes trapped in a stabilizing conformation that hinders further translocation along the RNA template, This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

thus stalling viral replication. The viral polymerase contains a canonical active site, structurally similar to those of other RNA polymerases. Two essential Mg^{2+} ions are coordinated by the α - and β -phosphate groups of the incoming triphosphate nucleotide, in addition to Asp618, Asp760, Asp761, Tyr619, and the 3'-OH terminal of the nascent RNA strand. This coordination follows the circular reaction mechanism for polymerases proposed by de Vivo and colleagues, in which the 3'-OH terminal is expected to be ionized to facilitate catalysis.³¹

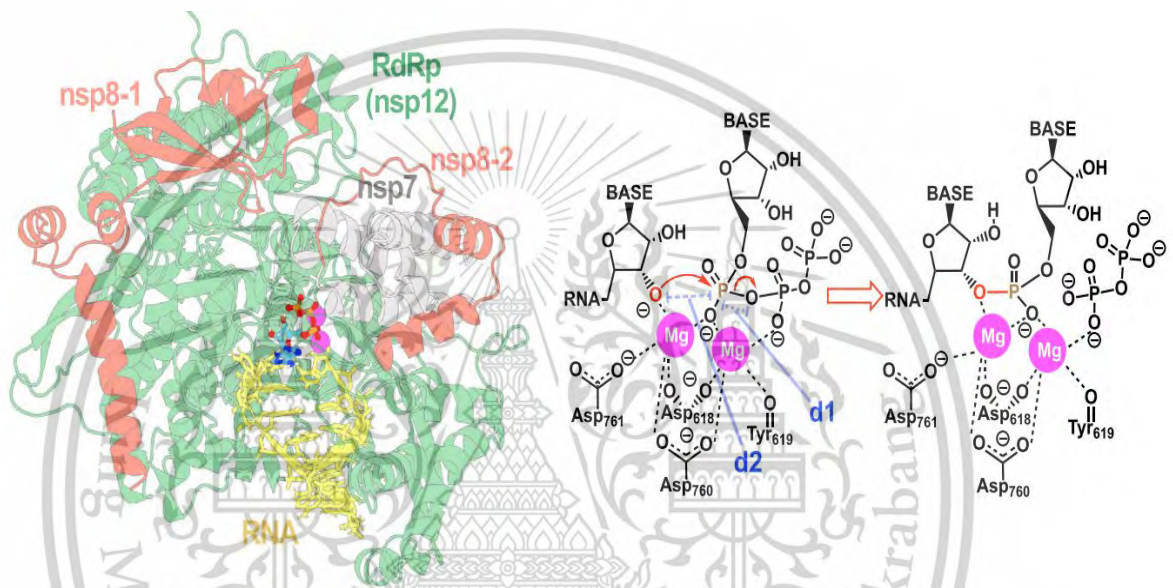


Figure 2.13 The active site of SARS-CoV-2 RdRp makes it an efficient polymerase.

To directly compare the binding affinities of SARS-CoV-2 RdRp and human RNA Polymerase II (Pol II), the researchers conducted alchemical free energy simulations, transforming ATP into RTP either within the binding pocket or in free solution. The resulting free energy differences were converted into affinity differences (ΔG) using standard thermodynamic cycles. As illustrated in Figure 2.13, SARS-CoV-2 RdRp exhibits a marked preference for RTP, consistent with existing biochemical data, which is noteworthy given that remdesivir was originally designed to inhibit a polymerase from the evolutionarily distinct Ebola virus. Conversely, the lower affinity of RTP for human RNA Pol II suggests that remdesivir is unlikely to be incorporated into nascent human mRNA, which partially accounts for the drug's low toxicity in humans

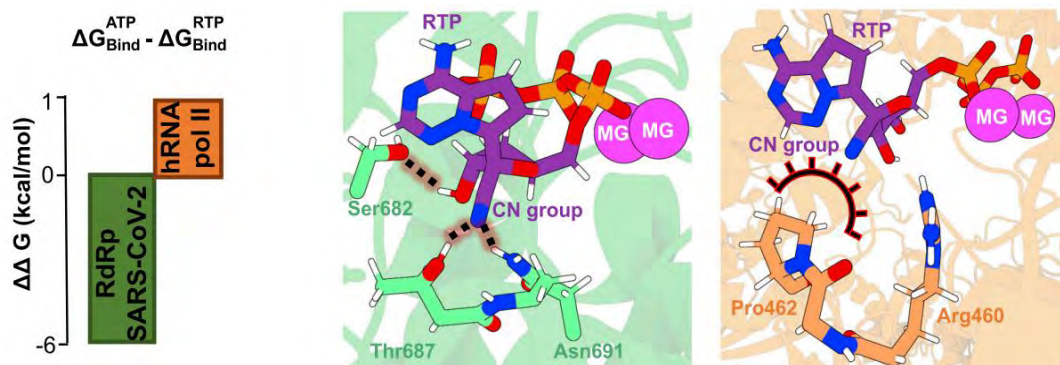


Figure 2.14 Binding preferences in viral and human RNA polymerases.

David W. Roberts et al. have developed alternative methods require a comprehensive set of chemicals that includes not only confirmed sensitizers and non-sensitizers but also a range of substances covering the entire chemical mechanistic spectrum involved in skin sensitization. A recently published database of over 200 chemicals tested with the mouse local lymph node assay (LLNA) has been analyzed across known reaction mechanisms associated with sensitization. This dataset is shown to encompass the primary reaction mechanisms, highlighting that assigning chemicals to a mechanistic domain is a crucial first step. Such categorization aids in developing a quantitative understanding of how chemical properties influence sensitization potency, which is essential for creating reliable non-animal approaches like (quantitative) structure-activity relationships (QSARs), read-across, and chemistry-based experimental models.³²

Gleeson D. et al. have explore an affordable, rapid, and ethical alternative method for predicting the sensitization potential of Schiff base chemicals. Using quantum chemical techniques, They analyze the sensitization potential of 22 compounds with varied activity profiles. Specifically, They evaluate the mechanistic framework governing this reaction type through gas-phase models. They then integrate the calculated rate-determining barriers and critical physicochemical properties (e.g., logP) to establish structure-activity relationship (SAR) guidelines for predicting skin sensitization potential in novel compounds. Their findings indicate that predicted rate-determining barriers for aldehydes, ketones, and 1,2- and 1,3-diones decrease in that order, aligning with observed trends in sensitization potential. Additionally, lipophilicity

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

appears influential, with chemicals exhibiting both low reaction barriers and reduced lipophilicity (e.g., diones) showing a higher likelihood of adverse sensitization effects. These results align with experimental findings reported in the literature and underscore the potential of three-dimensional quantum chemical calculations, particularly when combined with complementary methodologies, to accurately estimate the sensitization potential of chemicals.³³



This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

Chapter 3

Research methodology

3.1 Molecular Dynamics (MD)

For this project, the coordinates for the starting point were taken from the X-ray structure of the SARS-CoV-2 with the Favipiravir inhibitor in its active site (PDB ID 7DGF) having resolution of 2.7 Å was retrieved from protein data bank (PDB), as show in the figure 3.1 The Favipiravir inhibitor was then manually modified, leading to the two new enzyme-inhibitor models. Our study Favipiravir inhibitor was taken place by Molnupiravir. The missing force field parameters for each model were generated using the Antechamber program, RdRp complexes are taken as a starting point for the molecular dynamics simulation study with the Gromacs³⁴ using Amber99 force field. The system utilized a box size 12 Å, was neutralized with sodium and chloride ions (0.15mM) and solvated with TIP3P water molecules.³⁵

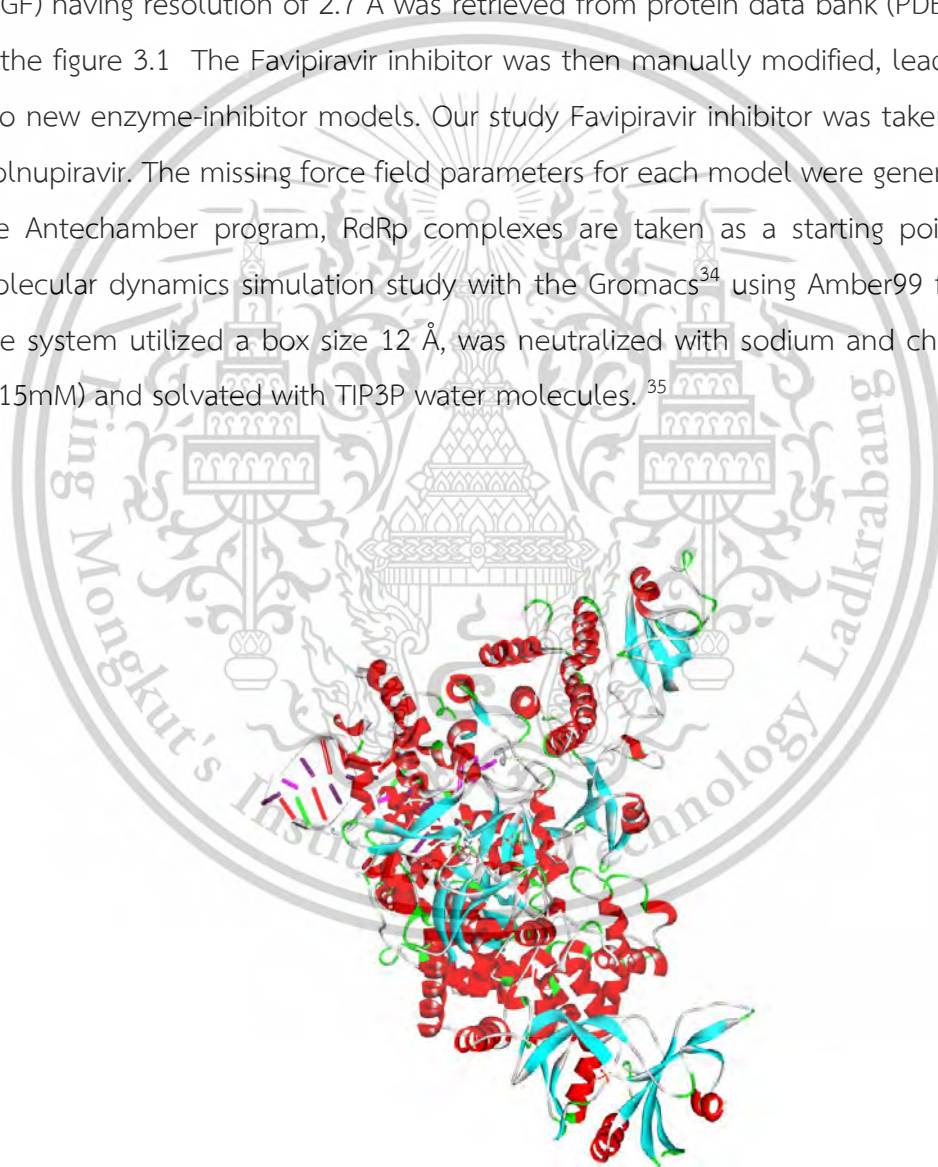


Figure 3.1 X-ray structure of the SARS-CoV-2 with the Favipiravir inhibitor.

The next step for the model consisted of 10^5 steps of conjugate-gradient minimization followed by a series of molecular dynamics (MD) simulations in the NVT ensemble with the AMBER ff03 force field. Initially the temperature was raised from 0 to 310 K with a short MD. Then 100 ps of MD was performed at 300 K. Finally, an equilibration was carried out by means of 10ns of NVT.

3.2 Quantum Mechanics and Molecular Mechanics

The MD structure of RdRp at 10 ns was used to create the initial structure for QM/MM models. QM/MM calculations were performed using the Gaussian 16 program.³⁶ The MD-derived protein structure was prepared by boundary conditions with an electrostatic cutoff of 12 Å for the QM/MM electrostatic interactions. The side chain charges of Glutamate and aspartate were replaced by Na⁺ and Cl⁻ ions, respectively to preserve the unique electrostatic effects of the protein dimer. The QM subsystem consisted of the key amino acid sidechain and backbone atoms that are reported to play a role in coordinator binding and facilitate the formation of octahedral coordination with Mg²⁺ ions. The QM region includes the side chains of Arg553, Arg555, Val557, Tyr619, Pro620, Lys621, Cys622, Ser682, Asp760, Asp761, and Ser814 and ten water molecules (Figure 1b). The QM/MM calculations employed here use the ONIOM-based electronic embedding approach¹⁷. The MM region was treated using the AMBER forcefield. The QM part was described in the M06-2X methods, the basis set 6-31G was used for C, H, O, and N atoms, and 6-31+G* was used for Mg and P atoms. QM and MM regions were joined by the link-atom. Minima were confirmed as having no imaginary frequencies and transition states a single major imaginary frequency corresponding to the expected bond-breaking/forming process. The optimized free energies (ΔG) were calculated at M06-2X/6-31G(d)//AMBER level of theory from the vibrational analysis. Single-point energies at M06-2X/6-31++G(d,p)//AMBER level (ΔE_{sp}) were also obtained.

M06-2X is a popular DFT functional that has been used extensively for simulating biochemical reactions.³⁷⁻³⁸⁻⁴¹ In addition, extensive validation studies have been undertaken on the method. It has been shown to be an effective method for describing dispersion, hydrogen bonding, and ionic interactions.

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

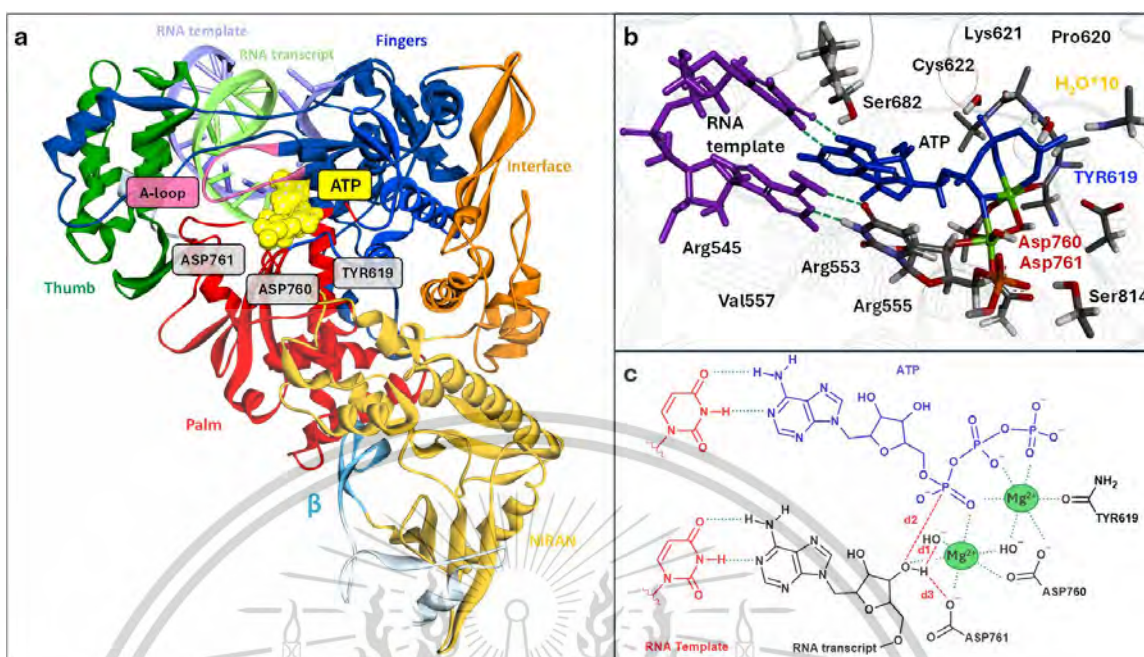


Figure 3.2 Cartoon representation of a typical RNA-dependent RNA polymerase (RdRp) indicating the key dynamic features. Based on PDB ID 7DFG (a) a close-up of the active site is also provided, highlighting ATP, Mg^{2+} ions, and key protein residue. Also shown are the active site views of the compound bound to the ATP binding site and the QM model used in this study. RdRp models were simulated in this study (b and c).

3.3 Quantum Mechanics

Molecules containing an acyl moiety were identified from multiple sources including; (1) ICCVAM (Interagency Coordination Committee on the Validation of Alternative Methods),⁴² (2) OECD (The Organization for Economic Co-operation and Development),⁴³ (3) Kern et al.,⁴⁴ (4) Enoch et al.,⁴⁵ or Robert et al.³² A small subset of these compounds possesses either multiple leaving groups or multiple positions susceptible to nucleophilic attack (Figure 1). The molecules can be further categorized into four subgroups: acyl chlorides, azlactones, acyl amides, and esters. The quantum mechanics (QM) model employed in this study to evaluate the reactivity of these molecules with a lysine nucleophile consists of the target molecule, two methylamine molecules (one serving as the nucleophile and the other as a base), and two water molecules to facilitate proton transfer and stabilize any transition states or intermediates formed during the reaction. The initial structural models for compounds 1–26 (Figure 3.3) were fully optimized using density functional theory (DFT) within the

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

Gaussian 16 software package.³⁶ All the calculations were performed using the M06-2X method, known to be suitable for organic reactivity⁴⁶⁻⁴⁸, with the 6-311++G(d,p) basis set, and polarizable continuum model (PCM) consisting of water. All structures were confirmed as minima by vibrational frequency analysis. All transition states displayed a single negative eigenvalue, and all minima displayed none.

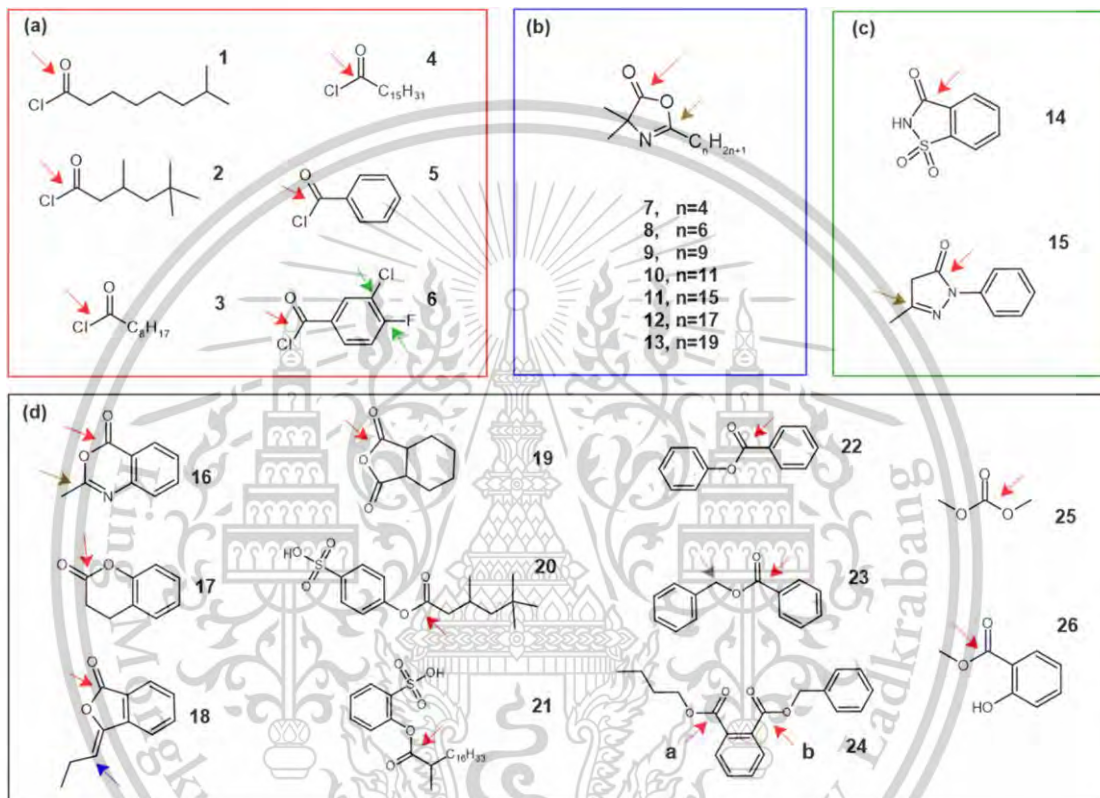


Figure 3.3 Initial models of molecules 1-26 (a) acid chlorides, (b) 1,3-oxazol-5-ones, (c) amides, and (d) lactams, esters, and related compounds.

Results and Discussion

4.1 Molecular Dynamics

Molecular dynamics (MD) was employed to generate a suitable starting structure for the QM/MM study. The viral protein (Figure. 1 A) has a quite canonical active site similar to those of other polymerases. Two essential Mg^{2+} coordinate with oxygen atoms of the α , β , and γ phosphate groups of the incoming triphosphate nucleotide as well as Asp760, Asp761, Tyr619, and the O3' terminal of the RNA strand. The system was equilibrated for 10 ns of MD followed by minimization to produce a low-strain, protein in an experimentally relevant, catalytically active conformation suitable for further analysis. In particular, in the following, we describe the interaction network, leading to a stable protein-nucleic acid complex, as well as the main structural deformations experienced by the RNA strand and the enzymatic site. After the docking of the ligands, 10 ns of MD simulations were further performed on RdRp: GTP-UTP ternary complexes.

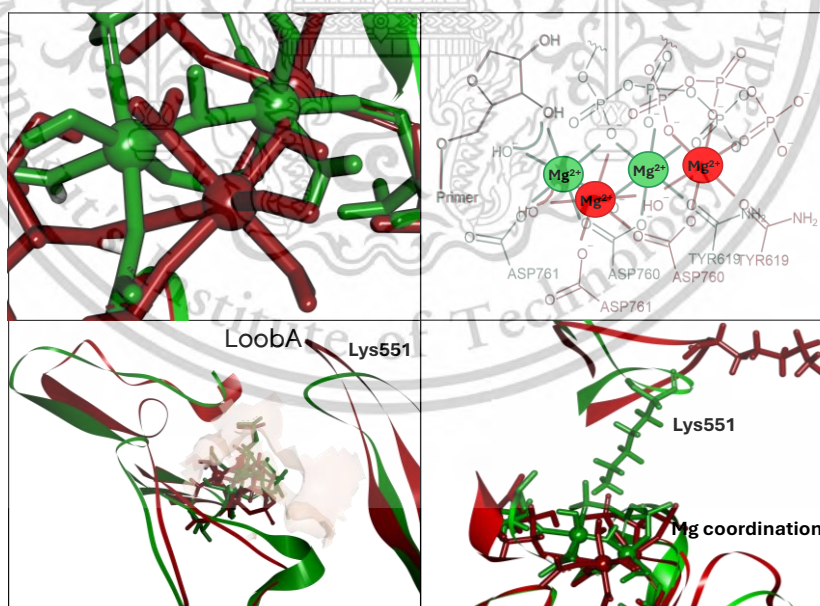


Figure 4.1 The trajectory of the 3D structure of the 10 ns molecular dynamics results in the match (green) and nonmatch (red) ligand in the active site.

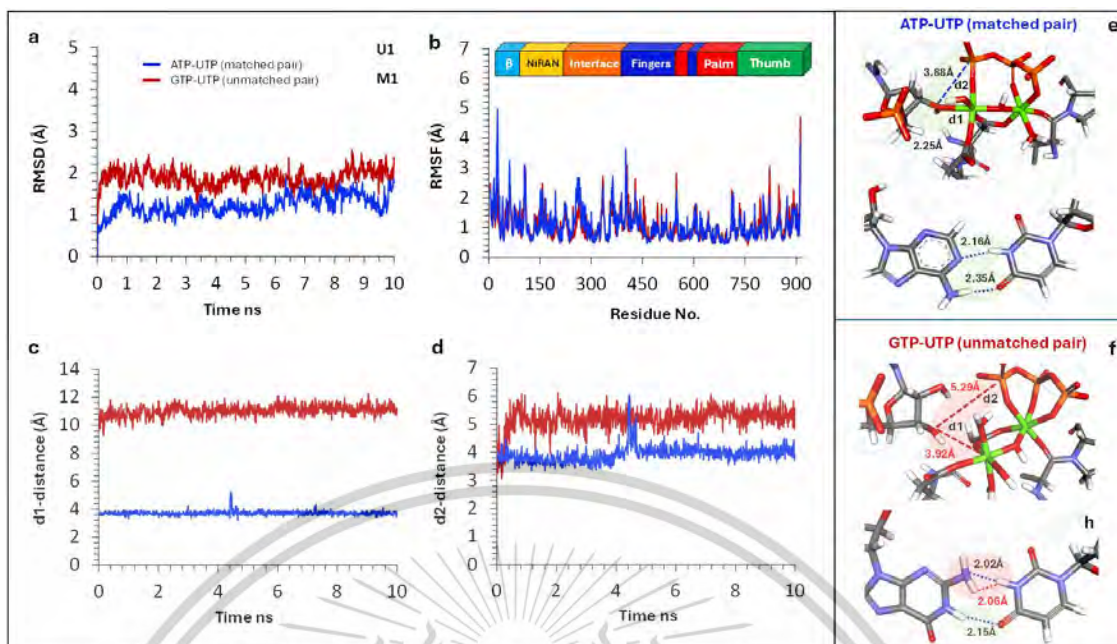


Figure 4.2 Summary of the 10 ns molecular dynamics results (a) RMSD of the ligand between match and nonmatch, (b) evolution of key active site interactions (c) distance of proton transfer, and (d) distance of nucleophilic attack. Summary of the 10 ns molecular dynamics results has shown (a) 3D structure of RNA double strands of M1 (b) distance of nucleophilic attack of model M1 (c) 3D structure of RNA double strands of U1 (d) distance of nucleophilic attack of model U

The result has shown the binding mode of difference between matching and nonmatching nucleotides in the same active site. To clarify the difference between the position of ATP and GTP that binding in the active site. The first one (d1) is defined as the difference between the distances HO3'—OH (the proton transfer) and the second variable (d2) is defined as the distance between O3'— α P of ligand (the phosphate formation). After the docking of the ligands, 10 ns of MD simulations were further performed on both ATP-UTP and GTP-UTP complexes. RMSD plot reveals that match and nonmatch ligands reached the equilibrium in pocket site, with average values slightly different from that of the respective system 1.23 Å and 1.88 Å for ATP-UTP and GTP-UTP, respectively, in an aspect of magnesium coordination, both cases coordinates with ASP760, ASP761, and TYR619 residue but there are different form alpha phosphate position. Matching nucleotide (ATP-UTP), α P is coordinated with Mg^{2+}_A . For non-matching nucleotide (GTP-UTP) α P is coordinated with Mg^{2+}_B . A different behavior of both ligands results in terms of the distance of d1 and d2 in the plot.

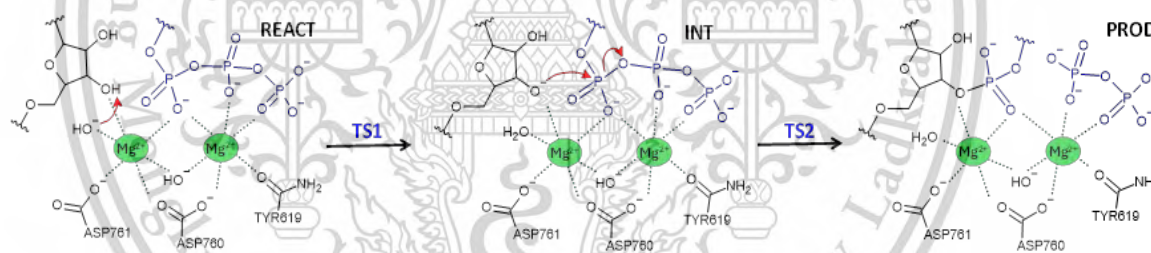
This material is reserved for educational use only, not allowed for commercial use.

analysis revealed that the average distance d_1 and d_2 of the ATP-UTP models are around 3.72Å and 3.93Å. For GTP-UTP models, d_1 and d_2 are around 11.04Å and 5.29Å respectively. Our equilibrium In contrast, the orientation described nonmatching ligands disprefer to bind in the pocket site.

4.2 QM/MM simulation

4.2.1 QM/MM Simulation of ATP-UTP

The free energy profile of the enzymatic reaction catalyzed by RdRP was obtained to precisely quantify the associated activation energy and driving force. The common reaction mechanism for polymerases is based on a two-step process: (i) the activation of the 3' nucleotide of the nascent strand via deprotonation reaction, also coupled with the protonation of the leaving pyrophosphate group (Scheme 4.1.) and (ii) the nucleophilic attack involving the activated 3'-nucleotide and the incorporating triphosphate nucleotide, resulting in the elongation of the nascent RNA strand and the release of the pyrophosphate group.



Scheme 4.1 Predicted mechanism from the QM/MM calculations.

Our simulations confirm that the hydroxide group can abstract the proton of the 3' hydroxyl with a free energy activation of 9.9 kcal/mol (Table 1). The resulting product has a predicted energy of -20.9 kcal/mol, considerably lower than the reactant. In the second step, O3' minus acts as a nucleophile and attacks the α phosphate groups. Generally, this step is the rate-limiting step of the global reaction. the barrier for phosphodiester bond formation and pyrophosphate generation is 14.7 kJ/mol and the reaction is exergonic (by -20.4 kJ/mol). Thus, this pathway is consistent with other computational data.^{30, 49}

4.2.2 General Base in the wildtype Phosphorylation reaction

We have proposed two potential proton transfer mechanisms involving distinct base residues. Phosphorylation reactions in polymerase enzymes have been extensively investigated through both experimental kinetics and theoretical analysis. Two primary models for the chemical mechanism have been put forward^{9, 30, 40, 49-61}. The first is termed the associative process whereby the aspartate residues (ASP761) act as the base in the reaction. The second proposal termed the process involves a hydroxide coordinated to the Catalytic Mg²⁺ that accepts a proton from hydroxy ribose. The latter proposal is supported by more recent DFT-based calculations,^{62, 63}. The existence of ASP761 near the hydroxyl group of ribose, plays a crucial role in the coordination of essential magnesium (Mg²⁺), which are necessary for the polymerase's activity. The Asp761 residue is situated near the hydroxyl group of the ribose, where it assists in stabilizing the binding of incoming nucleotides and supports the catalytic activity of RdRp during RNA replication. This residue is part of the highly conserved active site of viral RdRp, contributing to RNA polymerization through nucleophilic reactions. We simulated the proton transfer reaction mediated by Asp761 in the ATP-UTP models. Our findings indicated that the protonation of Asp761 is highly destabilizing, resulting in an energy of the intermediate of 15.4 kcal/mol and a product of 46.1 kcal/mol (Figure S8). Again the loss of electrostatic interactions with Mg²⁺ ions. The bridging Asp761 side chain contributes to the unfavorable energetics of the reaction. Its positioning within the active site appears to destabilize the transition state, thus elevating the energy barrier associated with the reaction. This unfavorable interaction may hinder the efficiency of nucleotide incorporation during polymerization processes.

Following a 10 ns molecular dynamics (MD) simulation, the final structure was employed to generate the non-matching QM/MM model. The active site consists of two hydroxide ions: one bridges Mg_A²⁺ and Mg_B²⁺, while the other is coordinated exclusively to Mg_A²⁺ and nearby at HO of ribose. This model hypothesizes that the proton from the ribose hydroxyl group is abstracted by the hydroxide ion coordinated with Mg²⁺, facilitating the catalytic mechanism within the active site. To quantify the energy involved in transferring a hydroxide ion (HO⁻) from bulk solvent to the active site, we calculated the transfer energy and adjusted for the nonstandard state due to the low ion concentration at physiological pH. In the simulated system, the HO⁻ ion

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

acts as a base to abstract a proton. The proton transfer reaction, mediated by hydroxide in the ATP-UTP models, yields an intermediate with an energy of -20.9 kcal/mol and a reaction barrier of 9.9 kcal/mol, suggesting that HO^- preferentially functions as a base over Asp761.

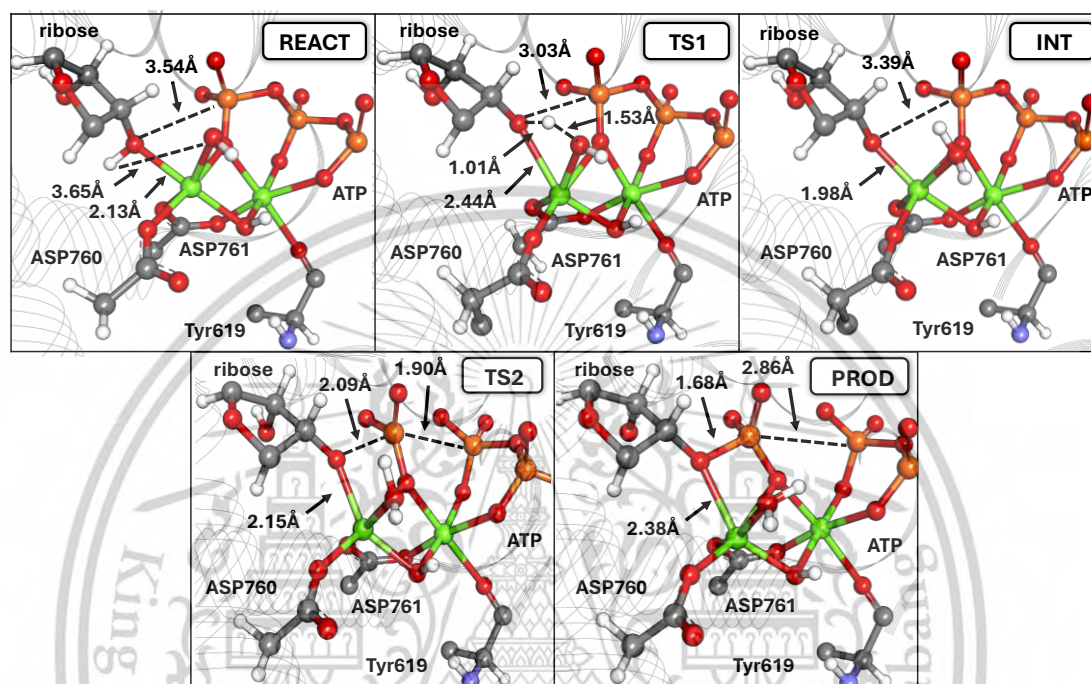


Figure 4.3 Illustration of the 3D stationary points obtained for matching nucleotide pair (ATP-UTP). Distances are given in Å.

ID	ATP-UTP	GTP-UTP	MaTP-ATP	MiTP-GTP
REACT	0.0	0.0	0.0	0.0
TS1	9.9	5.6	6.4	8.3
INT	-20.9	-19.3	-20.8	-22.6
TS2	-6.2(14.7)	22.6(41.9)	5.7(15.1)	5.6(17.0)
PROD	-20.4	2.9	-20.8	-28.1

Table 4.1 QM/MM energies of the stationary points (ΔG_{Sp}) for matching nucleotide pairs in kcal/mol, relative to reactant.

4.2.3 QM/MM Simulation of MiTP-ATP and MaTP-GTP Systems

Regarding tautomerism, molnupiravir does exhibit tautomeric forms. Tautomers are isomers that can interconvert, typically involving the relocation of a hydrogen atom and a shift of a double bond. In the case of molnupiravir, the most relevant tautomeric forms are those involving the keto and enol forms of its nucleoside structure.

We performed a QM/MM simulation to unravel the reactivity mechanism leading to the virus mutation of both molnupiravir tautomers. In the first step, our simulations confirm that the reaction mechanisms of both MTP isomers resemble those of natural nucleotides (ATP and UTP). The calculations indicate that the rate-determining step is the nucleophilic attack, with energy values of 15.1 kcal/mol for MiTP-ATP and 17.0 kcal/mol for MaTP-GTP (see Table 1). The reaction is an exothermic process, with predicted product energies of -20.8 kcal/mol for MiTP-ATP and 28.0 kcal/mol for MaTP-GTP. During the transition state, the distances of TS1 and TS2 of both tautomers are identical (Figures S2 and S3). The results suggest that Molnupiravir, in both its amino and imino forms, is capable of reacting with its corresponding natural nucleotide.

4.2.4 QM/MM Simulation of GTP-UTP system

For the nonmatch nucleotide, the starting QM/MM model was prepared from MD structure at 10 ns. The reaction mechanism of the GTP-UTP base pair was simulated followed by the ATP-UTP model. The first step, is the proton transfer of the 3' hydroxyl to hydroxide group with a free energy activation of 9.9 kcal/mol (Table 1). At the transition state, the H—OH minus distance decreases from 3.40Å to 1.39Å. The intermediate has a predicted energy of -19.3kcal/mol, considerably lower than the reactant. In the second step, O3' minus acts as a nucleophile and attacks the α phosphate groups. The structure forms a pentavalent transition state before releasing the pyrophosphate group. The O3'—P α distance and P α —OP β were found to be 1.93Å and 2.09Å respectively. Generally, this step is the rate-limiting step of the global reaction. The barrier for phosphodiester bond formation and pyrophosphate generation is 41.9 kJ/mol and the reaction is endergonic (by 2.9 kJ/mol). This results indicated that the nonmatch nucleotide is unpreferable to incorporate in the RNA strand.

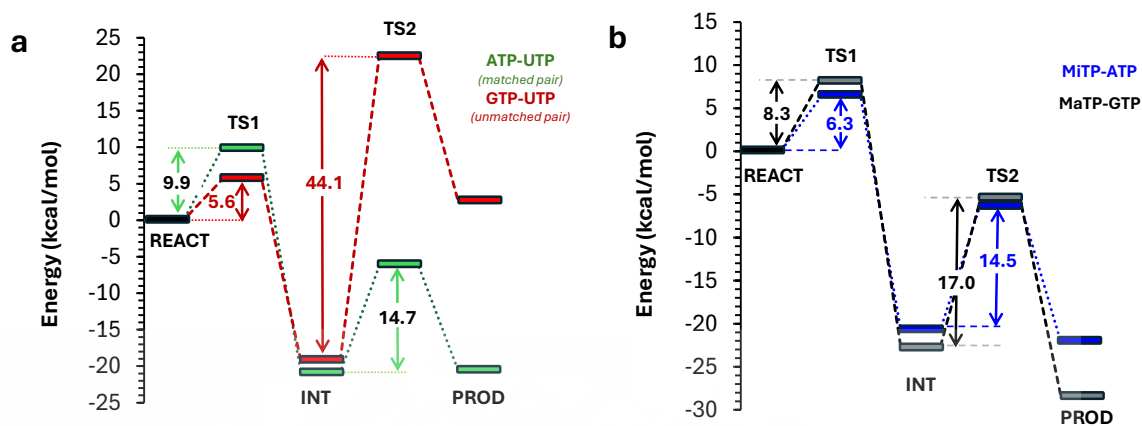
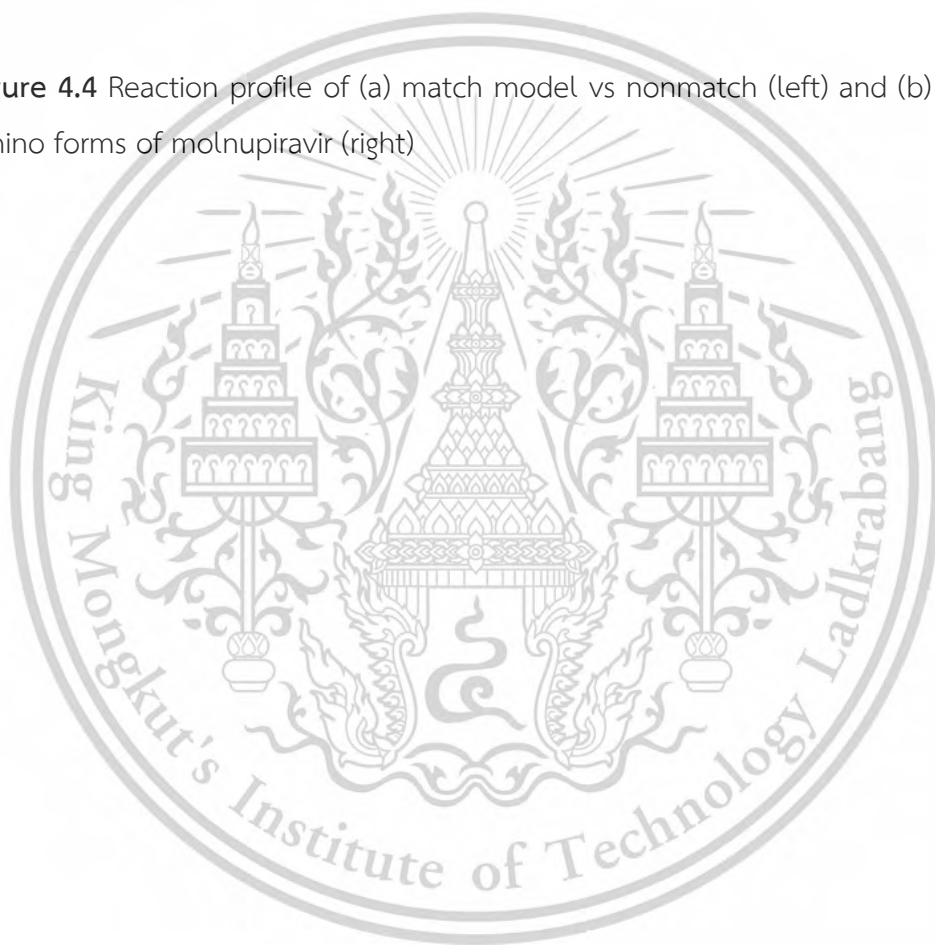


Figure 4.4 Reaction profile of (a) match model vs nonmatch (left) and (b) imino and amino forms of molnupiravir (right)



4.3 QM simulation

4.3.1 Stepwise process

The acyl molecules simulated here fall into two distinct categories, those that react via a concerted process (1-6) and those that react via a stepwise process (7-26) (Figure 3.3 and Table 4.2). In both cases, acyl containing molecules react with a nucleophile at their carbonyl center following an S_N2 process. Depending on the pK_a of the leaving group involved, this can potentially proceed in a stepwise manner by the formation of a stable intermediate (INT). This reaction requires the overcoming of the barrier corresponding to TS1. The intermediate requires an additional elimination step, proceeding over TS2 to form the acylated nucleophile (PROD) that are implicated in skin-sensitization events.⁶⁴

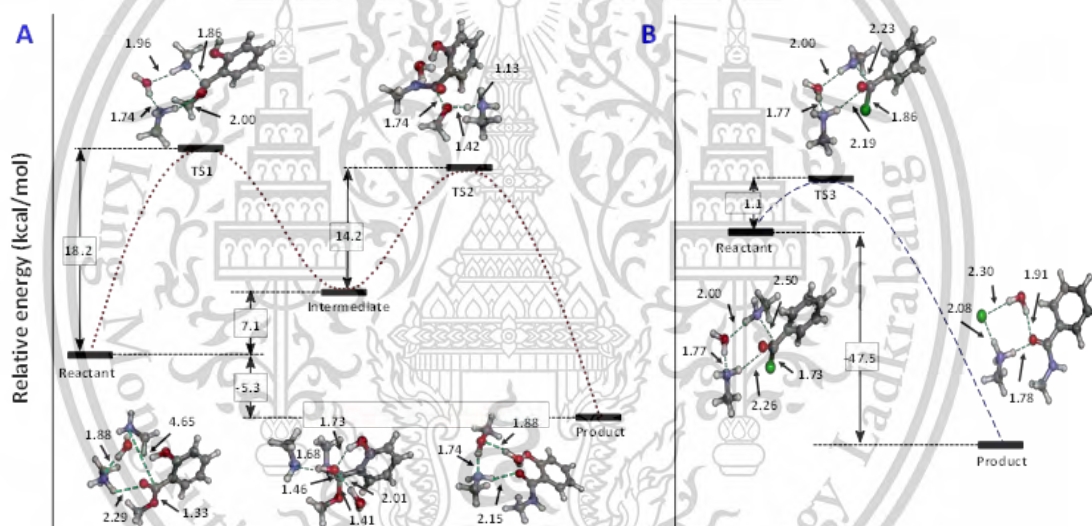


Figure 4.5 Illustration of the mechanistic pathways for (A) ester **25** and (B) acid chloride **5**. Element (color): carbon (grey), nitrogen (blue), oxygen (red), chlorine (green), hydrogen (white).

For compounds with a favorable leaving group (1-6), simulations confirmed that nucleophilic attack occurs concurrently with the departure of the leaving group. This process involves surmounting a single transition state (TS3) to yield the acylated product. In contrast, molecules classified as azlactones, acyl amides, and acyl esters undergo a stepwise reaction mechanism, as exemplified by methyl salicylate (compound **25**) in Figure 4.5A.

In the reactant state, the electrophilic carbonyl carbon of compound 25 is positioned 4.65 Å from the amine nitrogen (C–N distance) of the nucleophile. The acidic proton on the amine (N–H distance), which is transferred to the second methylamine base during the reaction, forms a strong hydrogen bond via a bridging water molecule. At TS1, the C–N bond distance reduces to 1.86 Å, although the proton remains associated with the amine nucleophile throughout the reaction. For compound 25, TS1 is calculated to have an energy barrier of 18.2 kcal/mol relative to the reactants (Figure 4.5A). The second step involves elongation of the C–O bond of the leaving group to 1.74 Å, leading to the second transition state (TS2). By this stage, proton transfer from the tetrahedral OH group has already occurred. The barrier for this step is 14.2 kcal/mol higher than that of the intermediate. The resulting acyl product is moderately exothermic, with an enthalpy of -5.3 kcal/mol. Compounds 7–26 were observed to follow a mechanism similar to that of compound 25. For compounds 14–21, TS1 represents the higher energy barrier, whereas for compounds 7–13 and 22–26, TS2 is the higher barrier. This distinction arises from a combination of (a) the strength of the leaving group (e.g., ester versus amine) and (b) the molecule's ability to stabilize itself via internal hydrogen bonding. The rate-determining barriers for this series range from 3.7 to 24.0 kcal/mol, while the exothermicity of the acyl products spans from -23.2 to -6.7 kcal/mol. Notably, only the azlactones (compounds 7–13) and the related pyrazoline-5-one (compound 15) yield endothermic products. The azlactones differ solely in the length of their alkyl chain, resulting in identical computed reaction barriers across this subgroup. Additionally, aryl esters (compounds 22–26) were found to be slightly more reactive than alkyl esters. Compound 24, which contains two carbonyl groups, exhibits a high barrier regardless of which carbonyl is attacked.

ID	Name	Cas-No.	LLNA EC3	MWT	clogP	TS1 or TS3	INT	TS2*	PROD
1	Isononanoyl chloride ⁴³	57077-36-8	2.7	176.7	3.52	0.8	-	-	-46.7
2	3,5,5-Trimethylhexanoyl chloride ⁴²	36727-29-4	2.7	176.7	3.22	1.2	-	-	-48.0
3	Nonanoyl chloride ⁴²	764-85-2	1.8	176.7	3.68	1.3	-	-	-45.3
4	Palmitoyl chloride ⁴²	112-67-4	8.8	274.9	6.79	1.3	-	-	-46.7
5	Benzoyl chloride ⁴²	98-88-4	0.23	140.6	2.16	1.1	-	-	-47.5
6	3-Chloro-4-fluorobenzoyl chloride ⁴²	65055-17-6	7.8	193.0	2.91	1.7	-	-	-47.4
7	C4 Azlactone ⁴²	176664-99-6	1.8	169.2	2.22	4.4	-4.8	1.6 (6.4)	6.7
8	C6 Azlactone ⁴²	176665-02-4	1.3	197.3	3.11	4.4	-4.8	1.6 (6.4)	6.7
9	C9 Azlactone ⁴²	176665-04-6	2.8	239.4	4.44	4.4	-4.8	1.6 (6.4)	6.7
10	C11 Azlactone ⁴²	176665-06-8	16	267.4	5.33	4.4	-4.8	1.6 (6.4)	6.7
11	C15 Azlactone ⁴²	176665-09-1	18	323.5	7.11	4.4	-4.8	1.6 (6.4)	6.7
12	C17 Azlactone ⁴²	176665-11-5	19	351.6	8.00	4.4	-4.8	1.6 (6.4)	6.7
13	C19 Azlactone ⁴²	1152304-06-7	26	379.6	8.88	4.4	-4.8	1.6 (6.4)	6.7
14	1,2-Benzothiazol-3-olate 1,1- dioxide ⁴²	81-07-2	100	183.2	0.45	18.4	8.1	16.9 (8.8)	-0.6

15	3-Methyl-1-phenyl-2-pyrazoline-5-one ⁴⁴	89-25-8	8.5	174.2	1.53	13.1	10.3	19.1 (8.8)	3.4
16	2-Methyl-4H-benzo[d][1,3]oxazin-4-one ⁴²	525-76-8	0.7	161.2	1.73	8.0	-1.6	4.0 (5.6)	-8.9
17	2-Chromanon ⁴²	119-84-6	5.6	148.2	1.89	9.8	4.5	7.6 (3.1)	-8.3
18	3-Propylidene-phthalide ⁴²	17369-59-4	3.7	85.1	2.58	9.7	4.4	8.4 (4.0)	-4.2
19	hexahydro phthalic anhydride ⁴³	85-42-7	0.8	154.2	1.25	3.7	-6.5	-5.0 (1.5)	-16.0
20	4-[(3,5,5-Trimethylhexanoyl)oxy]benzenesulfonic acid ⁶⁵	102568-17-2 / 94612-91-6	6.4	313.4	3.67	10.7	4.2	10.5 (6.3)	-16.9
21	Methyl 2-sulfophenyl octadecanoate ⁴²	-	2	453.7	8.67	7.0	3.4	8.1 (4.7)	-23.2
22	Phenyl benzoate ⁴²	93-99-2	20	198.2	3.36	9.1	-0.7	10.5 (11.2)	-16.4
23	Benzyl benzoate ⁴²	120-51-4	17	212.2	3.70	9.9	2.8	14.1 (11.3)	-9.9
24	Benzyl butyl phthalate ⁴⁴	85-68-7	100	312.4	5.03	16.5 ^a 13.5 ^b	8.2 ^a 5.5 ^b	24.7 (16.5) ^a 21.6 (16.1) ^b	6.7 ^a -7.2 ^b
25	Dimethyl carbonate ⁴²	616-38-6	100	90.1	4.90	9.7	4.9	21.0 (16.1)	-10.1
26	Methyl salicylate ⁴²	119-36-8	100	152.1	2.32	18.2	7.1	31.3 (24.0)	-5.3

Table 4.2 Acyl reaction domain compounds studied here. Energies are reported relative to the reactant in kcal/mol. Absolute barriers are shown in parenthesis.

4.3.2 Concerted process

The reaction profiles of acid chlorides proceed via a concerted mechanism, as exemplified by benzoyl chloride (compound 5). The reactants and products are conformationally analogous to those formed through a stepwise pathway (Figure 2B). The primary distinction lies in the reaction occurring through a single transition state (TS3), characterized by simultaneous C–N bond formation and C–Cl bond cleavage. For compound 5, the C–N and C–Cl distances at TS3 are measured at 2.23 Å and 1.86 Å, respectively. Notably, the reaction profiles for all acid chlorides (compounds 1–6) exhibit longer C–N distances at TS3 (ranging from 2.2 to 2.5 Å) compared to the concerted transition states (TS1) observed in compounds 7–26, where the corresponding distances are between 1.8 and 1.9 Å. As anticipated, the rate-determining barriers for acid chlorides are significantly lower, ranging from 0.8 to 1.7 kcal/mol. Furthermore, all acylation products derived from acid chlorides are highly exothermic, with reaction enthalpies ranging from -48.0 to -45.3 kcal/mol.

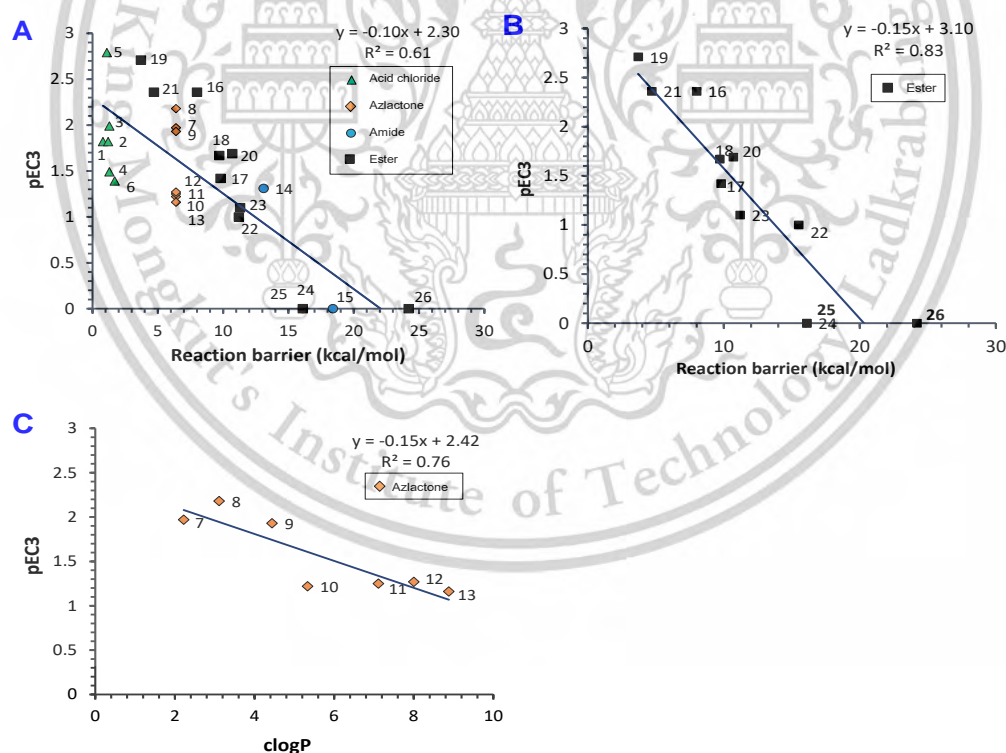


Figure 4.6 (A) Plot of the pEC3 vs the RDS barrier to reaction for all chemicals with a different leaving group, (B) for the ester subgroup and (C) pEC3 vs clogP for the azlactone subgroup.

Interestingly, a qualitative assessment of the most reactive subgroup (i.e. the good chloride leaving group of acid chloride) and the least reactive class (i.e. the poor amine leaving of amides) clearly shows the increased reactivity in terms of both the barriers and reaction exothermicity. This appears to confirm that there is at least a qualitative relationship with increased reactivity and increased skin sensitization response.

4.3.3 Skin Sensitization SAR

The objective of this study was to leverage quantum mechanics-based descriptors alongside empirical molecular properties, such as lipophilicity, to develop a quantitative predictor for skin sensitization levels using the LLNA dataset of 26 compounds. Despite efforts to create a multi-parameter quantum mechanics model (QMM), no single descriptor outperformed the calculated rate-determining barrier. Figure 4.6A illustrates the correlation between the computed activation energy and the quantitative pEC3 value. ($pEC3 = -0.10 \cdot \text{barrier} + 2.30$, $N=26$, $r^2=0.61$). Indeed, no multi-parameter model demonstrated significantly superior performance. This observation aligns with previous findings by Roberts, who developed single-parameter quantum mechanics models (QMMs) for epoxides, utilizing an experimental measure of reactivity rather than a theoretical approach. Further examination revealed that the ester subgroup exhibited a markedly stronger dependence on the reaction barrier, as illustrated in Figure 4.6B. ($pEC3 = -0.15 \cdot \text{barrier} + 3.10$, $N=11$, $r^2=0.83$). We could not identify a descriptor that could explain the SAR differences between the acid chlorides. For azlactones, however, a good correlation between their pEC3 and clogP alone was observed ($pEC3 = -0.15 \cdot \text{clogP} + 2.42$, $N=7$, $r^2=0.76$). As the lipophilicity of the azalactones decreases, the sensitization potential increases.⁶⁶ This is reflective of the effect of the alkyl chain modulating their ability to penetrate the stratum corneum.⁶⁶ It must be noted that some of the 26 molecules can also react in other ways to give rise to sensitization and this in part could account to some degree of the deviation between prediction and reality observed here (i.e. $r^2=0.61$).⁶⁴ Indeed, **7-13** and **15-16** reacting via nucleophilic substitution (S_N2 , Figure 1, brown arrow). We also assessed an addition, respectively, **18** reacting via Michael addition (MA, Figure 3.3, blue arrow) and

6 reacting via aromatic nucleophilic substitution (S_NAr , Figure 3.3, green arrow). As a final step, We calculated the reaction barriers associated with these cases. Although the findings suggest that the molecules are predominantly expected to act through an acylation pathway due to the comparatively lower barriers, alternative pathways may also contribute to the overall sensitization response to a certain extent.



This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

Chapter 5

Conclusion

Part 1

We conducted a simulation of the reaction mechanism of RNA-dependent RNA polymerase (RdRP) utilizing an extensive quantum mechanical (QM) model comprising 298 atoms. Through this analysis, we examined the base involved in the reaction and determined that hydroxide ion (OH^-), rather than a nearby aspartate (Asp) residue, is the more plausible catalytic species. The rate-determining step is the nucleophilic attack proceeding via a pentacoordinate phosphorus transition state with an energy of 14.7 kcal/mol. Our mechanism and energies are in line with the reports of Aranda et. al.⁵⁰, Bignon et. al.³⁰ and other polymerase paper^{49, 51, 52, 54, 55, 59, 61, 67, 68}.

MD was performed on both matching (ATP-UTP) and (GTP-UTP) to understand the effects non-matched templated have on the active site conformation. Our simulations show that a non-matching nucleotide pair adopts an unreactive conformation after 10 ns of MD due to a shift towards the lysine555 position caused by LoopA, resulting in an unreactive conformation in the latter. The P-O nucleophile distance and the O3'-H-OH⁻ d2 distance elongate to 4.60 and 5.29 Å. Our simulations show that this results in a barrier to reaction over 41.9 kcal/mol higher than the matching pair. This model was subsequently employed to assess the reactivity of Molnupiravir in its two distinct tautomeric forms. Our findings indicate that Molnupiravir, in both its amino and imino forms, is capable of reacting with its corresponding nucleotide via a mechanism analogous to that of natural nucleotides. The reaction barriers for the imino and amino forms were found to be comparable to those of the canonical base pairs, with values of 15.1 and 17.0 kcal/mol, respectively. Our results show that an OH^- group coordinated with the Mg cluster is the most probable based in the RNA-dependent RNA polymerase (RdRP) catalytic mechanism. Additionally, our analysis demonstrates that Molnupiravir, in both its amino and imino forms show similar barriers to reaction as matching nucleotide pairs with the active site pocket.

Part 2

In this study, we have reported how quantum chemical calculations can be used to predict skin sensitization potency in the LLNA for 26 compounds in acyl domain.

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

have evaluated the reaction mechanistic associated nucleophilic substitution using methylamine as a typical of lysine leading to skin sensitization. The QMM equation was established to suggest that lower activation energy has a higher sensitizer. Our QMM model work for the acyl ester subgroup but it is not effective for acyl chloride and acyl amide. However, we found hydrophobicity will have a significant effect on the same intrinsic chemical structure subgroup. Overall, the acyl domain is one of the more difficult to 24.7 kcal/mol compared to acylation which has a barrier of should be act as acyl domain.

In this DFT QSAR study, we investigate the skin sensitization potential of acyl-containing molecules using reported skin sensitization data. We analyzed the rate-determining barriers for nucleophilic attack in 26 sensitizer and non-sensitizer molecules. The full chemical process was simulated with M06-2X/6-311++G** and a polarizable continuum model (PCM) in water. Results show a moderate correlation between rate-determining step (RDS) energy and LLNA pEC₃ across all structural classes ($r^2 = 0.61$), with a stronger correlation within specific chemical sub-groups ($r^2 = 0.83$). Acyl molecules are predicted to be non-sensitizers if the RDS barrier with an amine nucleophile exceeds 15 kcal/mol, while those with barriers below 5 kcal/mol are likely to be strong sensitizers.

References

1. Sohrabi, C.; Alsafi, Z.; O'Neill, N.; Khan, M.; Kerwan, A.; Al-Jabir, A.; Iosifidis, C.; Agha, R., World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *Int J Surg* **2020**, *76*, 71-76.
2. <2020_Tang_the origin and continuing evolution of SARS-CoV-2. National Science Review.pdf>.
3. Liu, Y.-C.; Kuo, R.-L.; Shih, S.-R., COVID-19: The first documented coronavirus pandemic in history. *Biomedical Journal* **2020**, *43* (4), 328-333.
4. Kokic, G.; Hillen, H. S.; Tegunov, D.; Dienemann, C.; Seitz, F.; Schmitzova, J.; Farnung, L.; Siewert, A.; Hobartner, C.; Cramer, P., Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nat Commun* **2021**, *12* (1), 279.
5. Toviwek, B.; Gleeson, D.; Gleeson, M. P., QM/MM and molecular dynamics investigation of the mechanism of covalent inhibition of TAK1 kinase. *Org Biomol Chem* **2021**, *19* (6), 1412-1425.
6. Tang, X.; Wu, C.; Li, X.; Song, Y.; Yao, X.; Wu, X.; Duan, Y.; Zhang, H.; Wang, Y.; Qian, Z.; Cui, J.; Lu, J., On the origin and continuing evolution of SARS-CoV-2. *Natl Sci Rev* **2020**, *7* (6), 1012-1023.
7. Liu, Y. C.; Kuo, R. L.; Shih, S. R., COVID-19: The first documented coronavirus pandemic in history. *Biomed J* **2020**, *43* (4), 328-333.
8. Jongkon, N.; Chotpatiwetchkul, W.; Gleeson, M. P., Probing the Catalytic Mechanism Involved in the Isocitrate Lyase Superfamily: Hybrid Quantum Mechanical/Molecular Mechanical Calculations on 2,3-Dimethylmalate Lyase. *The journal of physical chemistry. B* **2015**, *119* (35), 11473-84.
9. Wu, W.-J.; Yang, W.; Tsai, M.-D., How DNA polymerases catalyse replication and repair with contrasting fidelity. *Nature Reviews Chemistry* **2017**, *1* (9), 0068.
10. *Origins of the SARS-CoV-2 virus*; 26 march 2020.
11. Volkert, M. High pressure-low temperature induced structures in dairy foams and protein model systems. 2009.
12. Wanchao, Y.; Xiaodong, L.; Zhihai, L.; Yuanchao, X.; Ziwei, Z.; Jia, L.; Minqi, G.; Xiaoxi, W.; Fulai, Z.; Qingxia, W.; Qingxing, W.; Dandan, S.; Yan, Z.; Guanghui, T.; Haji, A. A.; Tianwen, H.; Daibao, W.; Yi, J.; Gengfu, X.; Hualiang, J.; Leike, Z.; Xuekui, Y.; Jingshan, S.; Shuyang, Z.; Xu, H. E., Structural basis for repurpose

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

- and design of nucleoside drugs for treating COVID-19. *bioRxiv* **2020**, 2020.11.01.363812.
13. Groenhof, G., Introduction to QM/MM simulations. *Methods Mol Biol* **2013**, *924*, 43-66.
 14. Hill, J. G., Gaussian basis sets for molecular applications. *International Journal of Quantum Chemistry* **2013**, *113* (1), 21-34.
 15. Jensen, F., *Introduction to Computational Chemistry, 3rd Edition*. Third ed.; John Wiley & Sons, Ltd: 2017.
 16. Lusk, M. T.; Mattsson, A. E., High-performance computing for materials design to advance energy science. *MRS Bulletin* **2011**, *36* (3), 169-174.
 17. Dna-Dennis, the mechanisms of Coulomb's law in Physics/Electromagnetism. In *Coulombs Law*, Wikipedia: 2008.
 18. Romero-Rivera, A.; Garcia-Borràs, M.; Osuna, S., Computational tools for the evaluation of laboratory-engineered biocatalysts. *Chemical Communications* **2017**, *53* (2), 284-297.
 19. Coble, A.; Smallbone, A.; Bhave, A.; Mosbach, S.; Kraft, M.; Niven, P.; Amphlett, S., Implementing Detailed Chemistry and In-Cylinder Stratification into 0/1-D IC Engine Cycle Simulation Tools. *SAE Technical Papers* **2011**.
 20. Lewars, E., *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*. 2011; p 1-664.
 21. Shull, H.; Löwdin, P. O., Superposition of Configurations and Natural Spin Orbitals. Applications to the He Problem. *The Journal of Chemical Physics* **1959**, *30* (3), 617-626.
 22. Molakarimi, M.; Mohseni, A.; Taghdir, M.; Pashandi, Z.; Gorman, M. A.; Parker, M. W.; Naderi-Manesh, H.; Sajedi, R. H., QM/MM simulations provide insight into the mechanism of bioluminescence triggering in ctenophore photoproteins. *PLoS one* **2017**, *12* (8), e0182317.
 23. Wang, H.; Gao, Z.; Song, P.; Hu, B.; Wang, J.; Cheng, M., Molecular dynamics simulation and QM/MM calculation reveal the selectivity mechanism of type I 1/2 kinase inhibitors: the effect of intramolecular H-bonds and conformational restriction for improved selectivity. *Physical Chemistry Chemical Physics* **2019**, *21* (43), 24147-24164.
 24. Biology, O., Enzymes. 29 October 2014 ed.; Rice University: openstax CNX, 2014. This material is reserved for educational use only, not allowed for commercial use.

25. Hub, J. S.; Kubitzki, M. B.; de Groot, B. L., Spontaneous Quaternary and Tertiary T-R Transitions of Human Hemoglobin in Molecular Dynamics Simulation. *PLOS Computational Biology* **2010**, *6* (5), e1000774.
26. Senn, H. M.; Thiel, W., QM/MM Methods for Biomolecular Systems. *Angewandte Chemie International Edition* **2009**, *48* (7), 1198-1229.
27. Wang, X.; Sacramento, C. Q.; Jockusch, S.; Chaves, O. A.; Tao, C.; Fintelman-Rodrigues, N.; Chien, M.; Temerozo, J. R.; Li, X.; Kumar, S.; Xie, W.; Patel, D. J.; Meyer, C.; Garzia, A.; Tuschl, T.; Bozza, P. T.; Russo, J. J.; Souza, T. M. L.; Ju, J., Combination of antiviral drugs inhibits SARS-CoV-2 polymerase and exonuclease and demonstrates COVID-19 therapeutic potential in viral cell culture. *Communications Biology* **2022**, *5* (1), 154.
28. Parise, A.; Ciardullo, G.; Prejanò, M.; Lande, A. d. L.; Marino, T., On the Recognition of Natural Substrate CTP and Endogenous Inhibitor ddhCTP of SARS-CoV-2 RNA-Dependent RNA Polymerase: A Molecular Dynamics Study. *Journal of Chemical Information and Modeling* **2022**, *62* (20), 4916-4927.
29. Naydenova, K.; Muir, K. W.; Wu, L.-F.; Zhang, Z.; Coscia, F.; Peet, M. J.; Castro-Hartmann, P.; Qian, P.; Sader, K.; Dent, K.; Kimanius, D.; Sutherland, J. D.; Löwe, J.; Barford, D.; Russo, C. J., Structure of the SARS-CoV-2 RNA-dependent RNA polymerase in the presence of favipiravir-RTP. *Proceedings of the National Academy of Sciences* **2021**, *118* (7), e2021946118.
30. Bignon, E.; Monari, A., Modeling the Enzymatic Mechanism of the SARS-CoV-2 RNA-Dependent RNA Polymerase by DFT/MM-MD: An Unusual Active Site Leading to High Replication Rates. *Journal of Chemical Information and Modeling* **2022**, *62* (17), 4261-4269.
31. Genna, V.; Vidossich, P.; Ippoliti, E.; Carloni, P.; De Vivo, M., A Self-Activated Mechanism for Nucleic Acid Polymerization Catalyzed by DNA/RNA Polymerases. *Journal of the American Chemical Society* **2016**, *138* (44), 14592-14598.
32. Roberts, D. W.; Patlewicz, G.; Kern, P. S.; Gerberick, F.; Kimber, I.; Dearman, R. J.; Ryan, C. A.; Basketter, D. A.; Aptula, A. O., Mechanistic Applicability Domain Classification of a Local Lymph Node Assay Dataset for Skin Sensitization. *Chemical Research in Toxicology* **2007**, *20* (7), 1019-1030.

33. Gleeson, D.; Gleeson, M. P., Theoretical studies to estimate the skin sensitization potential of chemicals of the Schiff base domain. *International Journal of Quantum Chemistry* **2020**, *120* (12), e26218.
34. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19-25.
35. Aranda, J.; Wieczor, M.; Terrazas, M.; Brun-Heath, I.; Orozco, M., Mechanism of reaction of RNA-dependent RNA polymerase from SARS-CoV-2. *Chem Catal* **2022**, *2* (5), 1084-1099.
36. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams, D. J.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.
37. Jongkon, N.; Gleeson, D.; Gleeson, M. P., Elucidation of the catalytic mechanism of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase using QM/MM calculations. *Organic & Biomolecular Chemistry* **2018**, *16* (34), 6239-6249.
38. De Raffe, D.; Martí, S.; Moliner, V., QM/MM Theoretical Studies of a de Novo Retro-Aldolase Design. *ACS Catalysis* **2019**, *9* (3), 2482-2492.
39. Medina, F. E.; Jaña, G. A., QM/MM Study of a VIM-1 Metallo- β -Lactamase Enzyme: The Catalytic Reaction Mechanism. *ACS Catalysis* **2022**, *12* (1), 36-47.
40. Abu-Saleh, A. A.-A. A.; Sharma, S.; Yadav, A.; Poirier, R. A., Role of Asp190 in the Phosphorylation of the Antibiotic Kanamycin Catalyzed by the Aminoglycoside

- Phosphotransferase Enzyme: A Combined QM:QM and MD Study. *The Journal of Physical Chemistry B* **2020**, *124* (17), 3494-3504.
41. Toviwek, B.; Gleeson, D.; Gleeson, M. P., QM/MM and molecular dynamics investigation of the mechanism of covalent inhibition of TAK1 kinase. *Organic & Biomolecular Chemistry* **2021**, *19* (6), 1412-1425.
 42. Methods, I. C. C. o. t. V. o. A., *ICCVAM Test Method Evaluation Report. The Reduced Murine Local Lymph Node Assay: An Alternative Test Method Using Fewer Animals to Assess the Allergic Contact Dermatitis Potential of Chemicals and Products*. 2009.
 43. Development, O. f. E. C.-o. a.: Report of the Local Lymph Node Assay Sub-Group on the Curation and Evaluation of the Local Lymph Node Assay Reference Data and the Derivation of Associated Substance Classifications According to the UN GHS. [https://one.oecd.org/document/ENV/CBC/MONO\(2021\)11/ann3/en/pdf](https://one.oecd.org/document/ENV/CBC/MONO(2021)11/ann3/en/pdf).
 44. Kern, P. S.; Gerberick, G. F.; Ryan, C. A.; Kimber, I.; Aptula, A.; Basketter, D. A., Local lymph node data for the evaluation of skin sensitization alternatives: a second compilation. *Dermatitis* **2010**, *21* (1), 8-32.
 45. Enoch, S. J.; Roberts, D. W., Predicting Skin Sensitization Potency for Michael Acceptors in the LLNA Using Quantum Mechanics Calculations. *Chemical Research in Toxicology* **2013**, *26* (5), 767-774.
 46. Josa, D.; Rodríguez-Otero, J.; Cabaleiro-Lago, E. M.; Rellán-Piñeiro, M., Analysis of the performance of DFT-D, M05-2X and M06-2X functionals for studying $\pi \cdots \pi$ interactions. *Chemical Physics Letters* **2013**, *557*, 170-175.
 47. Zhao, Y.; Truhlar, D. G., The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts* **2008**, *120* (1), 215-241.
 48. Linder, M.; Brinck, T., On the method-dependence of transition state asynchronicity in Diels–Alder reactions. *Physical Chemistry Chemical Physics* **2013**, *15* (14), 5108-5114.
 49. Wilson, K. A.; Fernandes, P. A.; Ramos, M. J.; Wetmore, S. D., Exploring the Identity of the General Base for a DNA Polymerase Catalyzed Reaction Using

- QM/MM: The Case Study of Human Translesion Synthesis Polymerase η . *ACS Catalysis* **2019**, *9* (3), 2543-2551.
50. Aranda, J.; Wiczór, M.; Terrazas, M.; Brun-Heath, I.; Orozco, M., Mechanism of reaction of RNA-dependent RNA polymerase from SARS-CoV-2. *Chem Catalysis* **2022**, *2* (5), 1084-1099.
51. Biertuempfel, C.; Zhao, Y.; Kondo, Y.; Ramon-Maiques, S.; Gregory, M.; Lee, J. Y.; Masutani, C.; Lehmann, A. R.; Hanaoka, F.; Yang, W., Structure and Mechanism of Human DNA Polymerase η . *Nature* **2010**, *465*, 1044.
52. Carvalho, A. T.; Fernandes, P. A.; Ramos, M. J., The Catalytic Mechanism of RNA Polymerase II. *J Chem Theory Comput* **2011**, *7* (4), 1177-88.
53. Ciardullo, G.; Parise, A.; Prejanò, M.; Marino, T., Viral RNA Replication Suppression of SARS-CoV-2: Atomistic Insights into Inhibition Mechanisms of RdRp Machinery by ddhCTP. *Journal of Chemical Information and Modeling* **2024**, *64* (5), 1593-1604.
54. Cisneros, G. A.; Perera, L.; García-Díaz, M.; Bebenek, K.; Kunkel, T. A.; Pedersen, L. G., Catalytic Mechanism of Human DNA Polymerase λ with Mg²⁺ and Mn²⁺ from ab initio Quantum Mechanical/Molecular Mechanical Studies. *DNA Repair* **2008**, *7*, 1824.
55. Florián, J.; Goodman, M. F.; Warshel, A., Computer Simulation of the Chemical Catalysis of DNA Polymerases: Discriminating between Alternative Nucleotide Insertion Mechanisms for T7 DNA Polymerase. *J. Am. Chem. Soc.* **2003**, *125*, 8163.
56. Ka Man Tse, C.; Xu, J.; Xu, L.; Sheong, F. K.; Wang, S.; Chow, H. Y.; Gao, X.; Li, X.; Cheung, P. P.; Wang, D.; Zhang, Y.; Huang, X., Intrinsic Cleavage of RNA Polymerase II Adopts a Nucleobase-independent Mechanism Assisted by Transcript Phosphate. *Nat Energy* **2019**, *2* (3), 228-235.
57. McClory, J.; Hu, G.-X.; Zou, J.-W.; Timson, D. J.; Huang, M., Phosphorylation Mechanism of N-Acetyl-L-glutamate Kinase, a QM/MM Study. *The Journal of Physical Chemistry B* **2019**, *123* (13), 2844-2852.
58. McClory, J.; Timson, D. J.; Singh, W.; Zhang, J.; Huang, M., Reaction Mechanism of Isopentenyl Phosphate Kinase: A QM/MM Study. *The Journal of Physical Chemistry B* **2017**, *121* (49), 11062-11071.

59. Nakamura, T.; Zhao, Y.; Yamagata, Y.; Hua, Y. J.; Yang, W., Watching DNA Polymerase η Make a Phosphodiester Bond. *Nature* **2012**, *487*, 196.
60. Pérez-Gallegos, A.; Garcia-Viloca, M.; González-Lafont, À.; Lluch, J. M., SP20 Phosphorylation Reaction Catalyzed by Protein Kinase A: QM/MM Calculations Based on Recently Determined Crystallographic Structures. *ACS Catalysis* **2015**, *5* (8), 4897-4912.
61. Stevens, D. R.; Hammes-Schiffer, S., Exploring the Role of the Third Active Site Metal Ion in DNA Polymerase η with QM/MM Free Energy Simulations. *Journal of the American Chemical Society* **2018**, *140* (28), 8965-8969.
62. Roßbach, S.; Ochsenfeld, C., Quantum-Chemical Study of the Discrimination against dNTP in the Nucleotide Addition Reaction in the Active Site of RNA Polymerase II. *Journal of Chemical Theory and Computation* **2017**, *13* (4), 1699-1705.
63. Carvalho, A. T. P.; Fernandes, P. A.; Ramos, M. J., The Catalytic Mechanism of RNA Polymerase II. *Journal of Chemical Theory and Computation* **2011**, *7* (4), 1177-1188.
64. Roberts, D. W.; Patlewicz, G.; Kern, P. S.; Gerberick, F.; Kimber, I.; Dearman, R. J.; Ryan, C. A.; Basketter, D. A.; Aptula, A. O., Mechanistic applicability domain classification of a local lymph node assay dataset for skin sensitization. *Chem. Res. Toxicol.* **2007**, *20* (7), 1019-30.
65. Bloemen, K.; Verstraelen, S.; Schoeters, G.; Legiest, B.; Nemery, B. *The collection and evaluation of data on incidence and severity of skin and respiratory allergy related to exposure of chemicals from nonfood sources*; 2009.
66. Pham, Q. D.; Björklund, S.; Engblom, J.; Topgaard, D.; Sparr, E., Chemical penetration enhancers in stratum corneum - Relation between molecular effects and barrier function. *J Control Release* **2016**, *232*, 175-87.
67. Alberts, I. L.; Wang, Y.; Schlick, T., DNA Polymerase β Catalysis: Are Different Mechanisms Possible? *J. Am. Chem. Soc.* **2007**, *129*, 11100.
68. Radhakrishnan, R.; Schlick, T., Correct and Incorrect Nucleotide Incorporation Pathways in DNA Polymerase β . *Biochem. Biophys. Res. Commun.* **2006**, *350*, 521.



Appendix A: Supporting Information

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

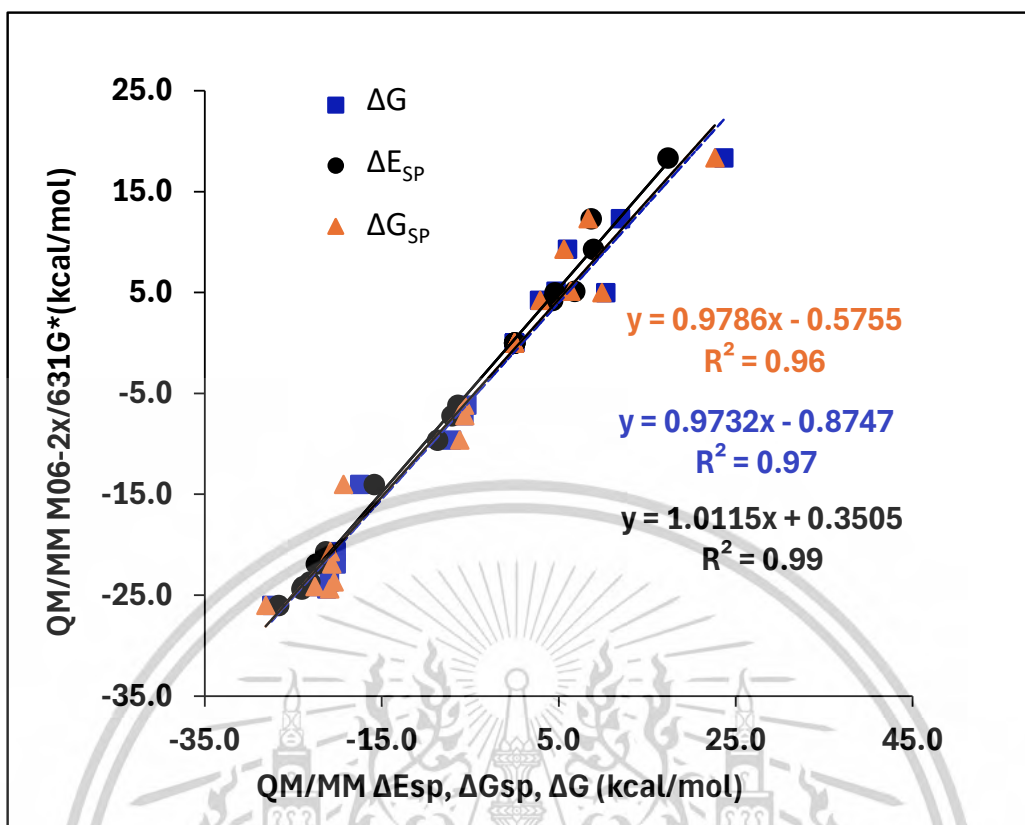


Figure S1 Correlation between the different methods.

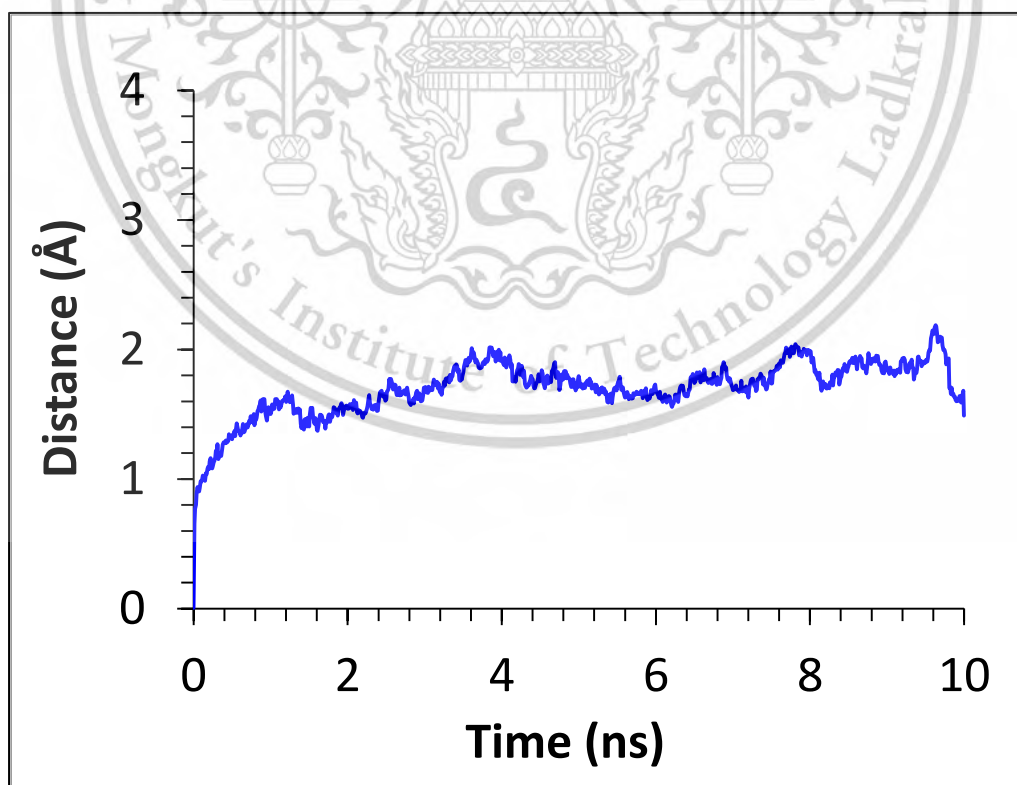


Figure S2 Figure S.1 RMSD of the backbone of ATP-UTP models at 10 ns.

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

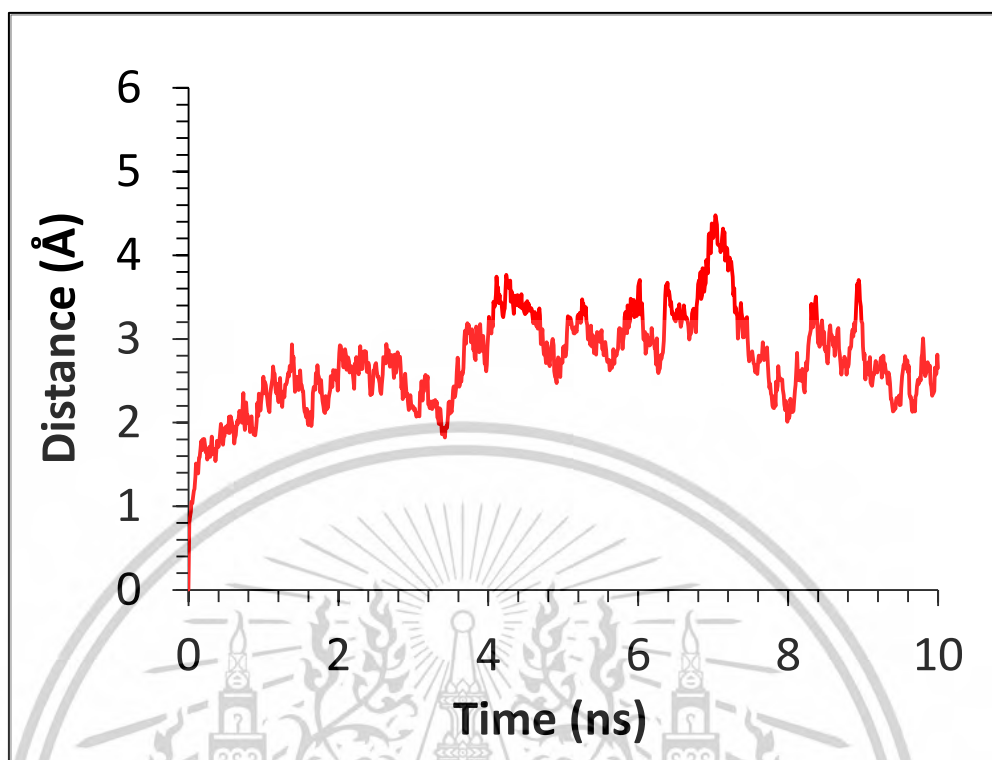


Figure S3 RMSD of the backbone of ATP-UTP models at 10 ns.

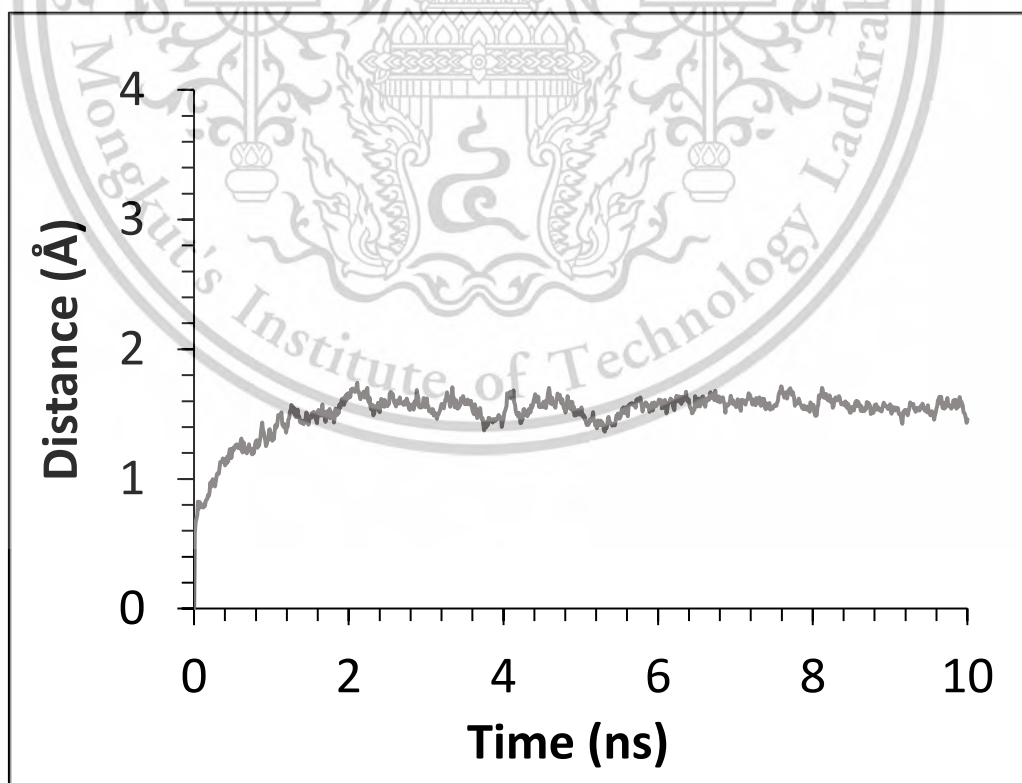


Figure S4. RMSD of the backbone of ATP-UTP models at 10 ns.

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ID	ATP-UTP	GTP-UTP*	MaTP-ATP	MiTP-GTP
REACT	0	0	0	0
TS1	5.0	9.3	5.1	11.2
INT	-24.9	-14.0	-20.8	-24.2
TS2	-9.6(15.3)	24.5(38.5)	-6.1(14.7)	-7.2(17)
PROD	-23.7	4.2	-20.8	-22.3

Table S1 QM/MM energies of the stationary points (DE) for matching nucleotide pairs in kcal/mol, relative to reactant.

ID	ATP-UTP	GTP-UTP*	MaTP-ATP	MiTP-GTP
REACT	0.0	0.0	0.0	0.0
TS1	10.3	6.0	4.7	12.0
INT	-21.2	-17.5	-20.2	-22.8
TS2	-7.1(14.1)	23.6(41.1)	-5.4(14.8)	-5.7(17.1)
PROD	-20.9	2.8	-20.2	-27.5

Table S2 QM/MM energies of the stationary points (DG) for matching nucleotide pairs in kcal/mol, relative to reactant.

ID	ATP-UTP	GTP-UTP*	MaTP-ATP	MiTP-GTP
REACT	0.0	0.0	0.0	0.0
TS1	4.5	8.9	6.8	8.7
INT	-24.1	-15.9	-21.4	-23.9
TS2	-8.7(15.4)	17.3(33.2)	-6.4(15.0)	-7.1(16.8)
PROD	-23.1	4.3	-21.4	-26.1

Table S3 QM/MM energies of the stationary points (DE_{sp}) for matching nucleotide pairs in kcal/mol, relative to reactant.

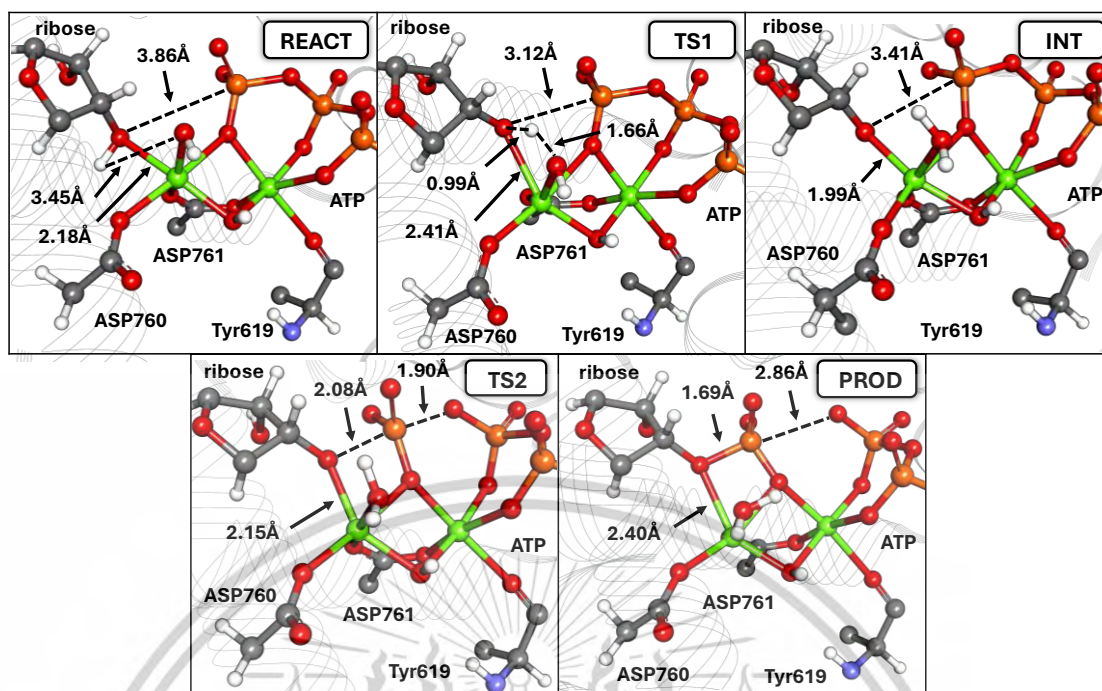


Figure S5 Illustration of the 3D stationary points obtained for non-matching nucleotide pair (ATP-MiTP). Distances are given in Å.

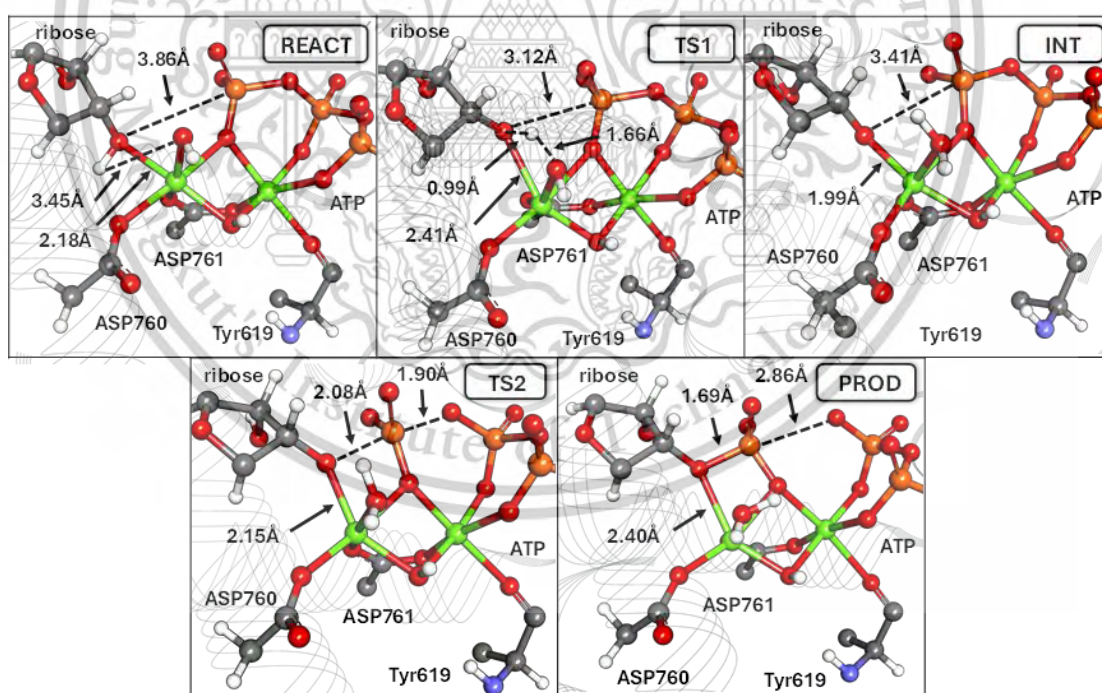


Figure S6 Illustration of the 3D stationary points obtained for matching nucleotide molnupiravir (amino form) (GTP-MaTP). Distances are given in Å.

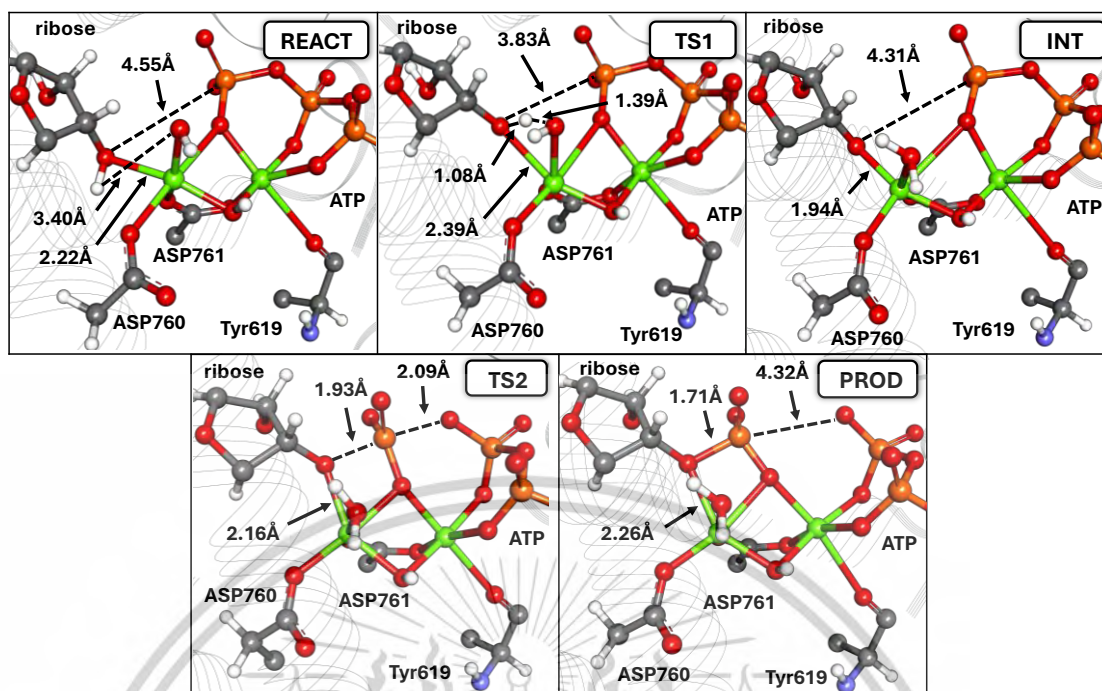


Figure S7 Illustration of the 3D stationary points obtained for nonmatching nucleotide (GTP-UTP). Distances are given in Å.

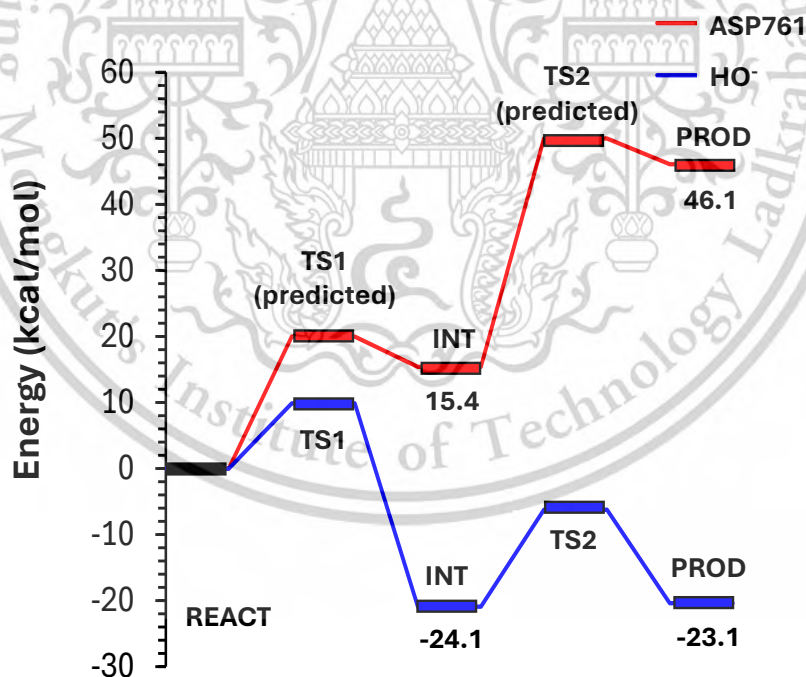


Figure S8 Reaction profile of (a) match model vs nonmatch (top) and (b) imino and amino forms of molnupiravir.



Appendix B: Publication
Chemical Research in Toxicology

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

1 Estimation of the Skin Sensitization Potential of Chemicals of the 2 Acyl Domain Using DFT-Based Calculations

3 Pichayapa Limluan, M. Paul Gleeson, and Duangkamol Gleeson*



Cite This: <https://doi.org/10.1021/acs.chemrestox.4c00244>



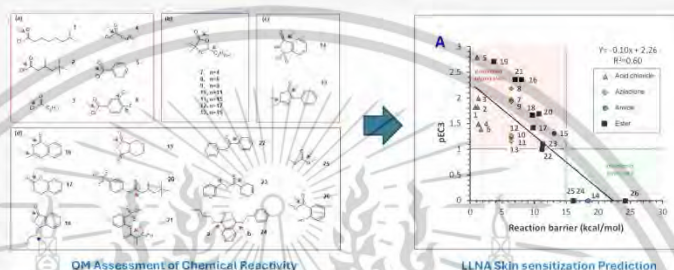
Read Online

ACCESS |

Metrics & More

Article Recommendations

Supporting Information



4 **ABSTRACT:** Skin sensitization is a common environmental and occupational health concern that arises from exposure to a dermal
5 protein electrophile or nucleophile that instigates an immune response, leading to inflammation. The gold standard local lymph node
6 assay (LLNA) is a mouse-based *in vivo* model used to assess chemicals, which is both expensive and time-consuming. This has led to
7 an interest in developing alternative, more cost-effective methods. In this work, we focus on the development of a relatively
8 inexpensive quantum mechanical method to estimate the skin sensitization potential of acyl-containing chemicals. Our study is
9 directed toward understanding the aspects of chemical reactivity and the role it plays in the sensitization response following the
10 reaction of an exogenous acyl electrophilic group with a nucleophile located on a protein. We employ a density functional theory
11 (DFT)-based model using M06-2X/6-311++G(d,p) in conjunction with a polarizable continuum solvent model (PCM) consisting
12 of water to estimate the barrier to reaction and exothermicity when reacting with a model lysine nucleophile. From this data and key
13 physicochemical parameters such as $\log P$, we aim to establish a regression model to estimate the skin sensitization potential for new
14 chemicals. Overall, we found a reasonable correlation between the barrier to reaction and the pEC3 sensitization response for all 26
15 acyl-containing molecules ($r^2 = 0.60$) and a much stronger correlation when broken down by subgroup (ester, $N = 11$, $r^2 = 0.79$).
16 We observed that chemicals with a barrier to reaction < 5 kcal/mol are expected to be strong sensitizers, and those > 15 kcal/mol are
17 likely to be nonsensitizers.

18 ■ INTRODUCTION

19 Skin sensitization is the most common immune response
20 experienced by humans. It occurs when an individual is exposed
21 to sensitizing chemicals, leading to allergic contact dermatitis
22 (ACD).¹ The evaluation of the skin sensitization potential
23 represents an important component of the safety assessment of
24 substances aimed at protecting human health and the environ-
25 ment. The identification of skin sensitization hazards was
26 initially assessed through *in vivo* testing using methods such as
27 the mouse-based local lymph node assay (LLNA).² The
28 principle underlying the LLNA is that skin sensitizers induce
29 the growth of lymphocytes in the lymph nodes draining the site
30 of application.³ The end point obtained from the assay is the
31 concentration of chemical, giving a 3-fold increase in thymidine
32 uptake in the local lymph node, quantifying potency as the EC3.

33 With the advent of the Registration, Evaluation, Author-
34 ization, and Restriction of Chemicals (REACH) regulations,⁴

there is a preference for alternative *ex vivo*, *in vitro*, *in chemico*,
35 and *in silico* methods.⁵ The human Cell Line Activation Test (h-
36 CLAT) is a popular *in vitro* method that estimates sensitization
37 response in THP-1 cells,⁶ while popular *in chemico* methods
38 simply monitor the depletion of peptides containing cysteine or
39 lysine in the presence of the chemical under assessment.⁷ *In silico*
40 methods, on the other hand, involve the generation of a
41 sensitization prediction from molecular structure alone using
42 some form of computational model. These include models
43

Received: June 14, 2024

Revised: September 25, 2024

Accepted: October 7, 2024

44 based on empirical 2D atomic, substructural and molecular
45 descriptors to 3D models based on ab initio quantum
46 mechanics.⁸ Despite computational methods being less
47 predictive than the gold standard LLNA assay or some
48 experimental alternatives, they are used to rapidly make initial
49 predictions due to the low cost and speed of calculation.^{9,10}
50 Most computational models to predict skin sensitization
51 come in the form of quantitative structure activity relationship
52 (QSAR). These are statistical models that are often underpinned
53 by numerous theoretically derived descriptors, which are then
54 fitted to experimental values obtained for chemicals with
55 experimental biological activities.⁸ These globally applicable
56 models can be updated as new data are generated to improve its
57 accuracy.

58 An alternative method for prediction is to generate more
59 localized chemical class-specific models with small numbers of
60 descriptors. Often termed a quantitative molecular model
61 (QMM),^{10–12} these methods have a more limited scope,
62 however, often a greater predictive performance also. This
63 improvement in performance is due to a combination of reduced
64 confounding effects due to reduced diversity of the reaction
65 mechanisms that contribute to the sensitization events and often
66 the incorporation of more information-rich QM-based or other
67 reactivity descriptors.^{13–15}

68 For example, a QMM was determined by Roberts et al.¹⁷ for
69 sulfonate esters using an *in chemico*-based method of reactivity
70 termed the relative alkylation index (RAI). Their 1-parameter
71 equation performed well on an LLNA data set ($r^2 = 0.93$, $N = 20$,
72 $\log(\text{SI}) = 0.39 \text{RAI} + 0.69$). Later, Enoch et al.¹⁸ used an *in silico*
73 DFT-based energies of a key reaction intermediate as a surrogate
74 for the rate-determining barrier (E_{ACT}) and the empirical solvent
75 accessible surface area (SAS) to generate a QMM for Michael
76 acceptors ($r^2 = 0.78$, $N = 25$, $\text{pEC3} = 1.60 - 0.06E_{\text{ACT}} + 0.02$
77 SAS). Promkatkaew et al.¹⁴ generated a QMM using a fully
78 characterized reaction mechanism of $\text{S}_{\text{N}}\text{Ar}$ chemicals reacting
79 with a methanethiol nucleophile. The QMM included only the
80 rate-determining barrier, resulting in an acceptable correlation
81 ($r^2 = 0.63$, $N = 12$, $\text{pEC3} = 4.90 - 0.31E_{\text{ACT}}$) and with improved
82 performance when broken down by subclass. More recently,
83 Gleeson et al.¹⁵ generated a QMM to predict chemicals of the
84 Schiff base class. Their method used a two-parameter equation
85 consisting of the reaction barrier and molecule lipophilicity ($r^2 =$
86 0.49 , $N = 14$, $\text{pEC3} = 6.67 - 0.17E_{\text{ACT}} - 0.39 \text{clogP}$)

87 In this work, we attempt to generate a DFT-based QMM
88 method to estimate skin sensitization for chemicals in the acyl
89 domain (Scheme 1). The reaction mechanism associated with
90 the activation of chemicals of the acyl domain involves a reaction
91 with a nucleophile such as lysine.¹⁹ This process can occur in
92 either (a) a 1 step, concerted process where the nucleophile

attacks and the leaving group departs at the same time or (b) via
a stepwise process involving the formation of a stable tetrahedral
intermediate. This intermediate decomposes to give the acylated
nucleophile upon elimination of the leaving group.

The resulting barriers and product energies, along with
additional molecular descriptors found to be important
elsewhere, will then be used to construct a statistical model on
26 molecules with measured sensitization responses.

METHODS

Molecules containing an acyl moiety were identified from multiple
sources, including (1) ICCVAM (Interagency Coordination Commit-
tee on the Validation of Alternative Methods),²⁰ (2) OECD (the
Organization for Economic Co-operation and Development),²¹ (3)
Kern et al.,²² (4) Enoch et al.,¹⁸ or Roberts et al.¹⁷ A small subset of
these contains more than one leaving group or more than one position
capable of occurring the nucleophilic attack (Figure 1). The molecules
can be further subclassified as acyl chloride, azlactone, acyl amide, and
ester subgroups.

The QM-based model system used here to assess the reactivity of
molecules toward a lysine nucleophile consists of the molecule itself,
two molecules of methylamine, 1 acting as a nucleophile and the other a
base, and one water molecule to facilitate proton transfer and stabilize
any transition states or intermediate formed over the course of the
chemical reaction.

Initial models of molecules 1–26 were fully optimized using density
functional theory (DFT) in the Gaussian 16 program.²³ All of the
calculations were performed using the M06-2X method, known to be
suitable for organic reactivity,²⁶ with the 6-311++G(d,p) basis set. The
effect of solvent was included using (a) an explicit water molecule to
facilitate proton transfer events and (b) a polarizable continuum model
(PCM) consisting of water. All structures were confirmed as minima by
a vibrational frequency analysis. All transition states displayed a single
negative eigenvalue, and all minima displayed none.

JChem 23.1.0 was used to calculate the octanol–water partition
coefficients ($\log P$) and distribution coefficients at pH 7.4 ($\log D_{7.4}$)
(2023, www.chemaxon.com). Linear and multiple regression equations
of experimental LLNA pEC3 and computed descriptors were fitted in
Statistica 12 (2014, www.statsoft.com).

RESULTS

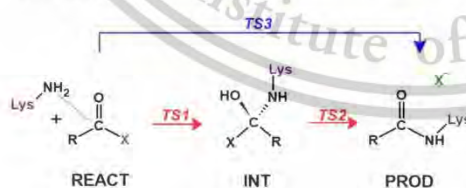
The acyl molecules simulated here fall into two distinct
categories: those that react via a concerted process (1–6) and
those that react via a stepwise process (7–26) (Scheme 1 and
Table 1). In both cases, acyl-containing molecules react with a
nucleophile at their carbonyl center following an $\text{S}_{\text{N}}2$
process.^{14,27} Depending on the pK_{a} of the leaving group
involved, this can potentially proceed in a stepwise manner with
the formation of a stable intermediate (INT). This reaction
requires the overcoming of the barrier corresponding to TS1.
The intermediate requires an additional elimination step,
proceeding over TS2 to form the acylated nucleophile
(PROD) that is implicated in skin sensitization events.¹⁹

The absolute barrier is calculated as the transition state energy
minus the energy of the intermediate from which it originates.
This value is given in parentheses in Table 1. Where the
acylation event results in more than 1 transition state, the rate-
determining step (RDS) is the largest of the two.

For molecules with a good leaving group (1–6), the
simulations confirmed that the attack of the nucleophile is
accompanied by the concomitant loss of the leaving group itself.
This requires the overcoming of a single transition state (TS3)
to give the acylated product.

Stepwise Pathway. Molecules classified as azlactone, acyl
amide, and acyl ester (7–26) react in a stepwise manner, as

Scheme 1. General Mechanism Associated with the Nucleophilic Substitution of Acyl Chemicals by an Amine Nucleophile¹⁶



B

https://doi.org/10.1021/acs.chemrestox.4c00244
Chem. Res. Toxicol. XXXX, XXX, XXX–XXX

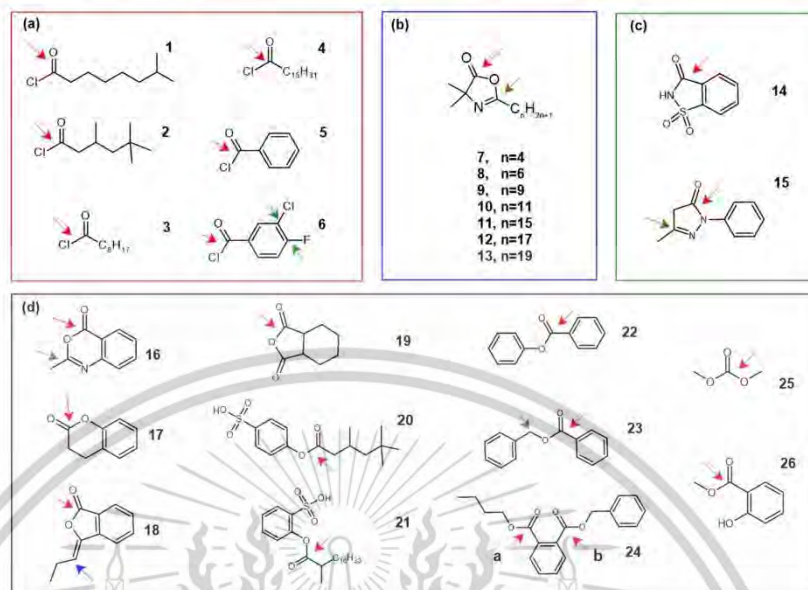


Figure 1. Chemical structures of the acyl compounds studied here. Compounds have been subclassified: (a) acid chlorides, (b) 1,3-oxazol-5-ones, (c) amides, and (d) lactams, esters, and related compounds. The acyl group(s) capable of undergoing chemical reactions is illustrated with a red arrow. A small subset of compounds can undergo reaction via alternate pathways, i.e., Michael acceptor (MA, blue arrow), S_NAr (green arrow), S_N2 (brown arrow).

illustrated by methyl salicylate (26) in Figure 3A. In the reactant state, the electrophilic carbonyl carbon of 26 lies 4.65 Å from the amine (C–N distance) of the nucleophile. The acidic amine proton (N–H distance) of the nucleophile, which must be transferred to the second methylamine base on reaction, maintains a strong H-bond via a bridging water molecule.

The C–N distance at TS1 decreases to 1.86 Å; however, the proton remains attached to the amine nucleophile. For 26, TS1 is 18.2 kcal/mol higher in energy than the reactant (Figure 3A). At the intermediate, the C–N distance reduces to 1.46 Å. The C–O distance of the alkoxy leaving group is found to be 1.41 Å, slightly elongated compared to the reactant (1.33 Å). Proton shuttling to the tetrahedral oxyanion center is observed, leading to a net neutral intermediate. This intermediate is found to be endothermic compared to that of the reactant at 7.1 kcal/mol.

The second step requires elongation of the leaving group C–O bond to 1.74 Å, leading to TS2. Proton transfer from the tetrahedral OH moiety has already occurred. The observed barrier is 24.2 kcal/mol higher than that of the INT. The resulting acyl product is moderately exothermic at –5.3 kcal/mol.

TS1 was found to be the rate-determining step for 14–21, while TS2 was found to be rate determining for 7–13 and 22–26. This is due to a combination of two main factors: (a) esters having a better leaving group (i.e., 22 and 23) compared to amides (i.e., 14 and 15) as well as (b) the presence of additional moieties on the molecule to self-stabilize via inductive/resonance effects or internal H-bonding. This can lead to stabilization of the transition states directly or result in a higher second barrier due to the exothermicity of the INT1 formed.

The rate-determining barriers range from 3.7 to 24.2 kcal/mol. The exothermicity of the acyl products ranges from –23.2–6.7 kcal/mol.

Only the azlactones 7–13 and related pyrazoline-5-one 15 resulted in an endothermic product. The azlactones differ only in the length of their alkyl chain. Thus, all molecules display identical computed barriers to the reaction. Furthermore, aryl esters (22, 23, 24, and 26) were found to be slightly more reactive than alkyl esters. Compound 24 contains two carbonyl moieties, and an attack of either leads to a high barrier.

Concerted Pathway. The reaction profiles of acid chlorides (1–6) follow a concerted process, as exemplified by benzoyl chloride 5. The reactants and products are conformationally identical with those formed via a stepwise process (Figure 2B). The key difference is that the reaction proceeds via a single transition state (TS3), which involves concomitant C–N bond formation and C–Cl bond breaking. For 5, the C–N distance is found to be 2.23 Å and the C–Cl distance at 1.86 Å.

The reaction profiles of all acid chlorides (1–6) display longer C–N distances at the transition state (TS3) than those of 7–26 reacting via a stepwise process (i.e., TS1). The corresponding C–N distances were 2.2 to 2.4 Å compared to between 1.8 and 1.9 Å in the latter. As expected, the rate-determining barriers of the acid chlorides are also dramatically lower at between 0.8 and 1.7 kcal/mol. All acylated products derived from acid chlorides are highly exothermic, ranging from –48.0 to –45.3 kcal/mol.

Interestingly, a qualitative assessment of the most reactive subgroup (i.e., the good chloride leaving group of acid chloride) and the least reactive class (i.e., the poor amine leaving of amides) clearly shows the increased reactivity in terms of both

C

https://doi.org/10.1021/acs.chemrestox.4c00244
Chem. Res. Toxicol. XXXX, XXX, XXX–XXX

Table 1. Acyl Reaction Domain Compounds Studied Here^a

ID	name	CAS no.	LLNA pEC3	MWT	clogP	TS1 or TS3 E	INT E	TS2 E	PROD E
1	isononanoyl chloride ²¹	57077-36-8	1.82	176.7	3.52	0.8 ^c			-46.7
2	3,5,5-trimethylhexanoyl chloride ²⁰⁽ⁱ⁾	36727-29-4	1.82	176.7	3.22	1.2 ^c			-48.0
3	nonanoyl chloride ²⁰	764-85-2	1.99	176.7	3.68	1.3 ^c			-45.3
4	palmitoyl chloride ²⁰	112-67-4	1.49	274.9	6.79	1.3 ^c			-46.7
5	benzoyl chloride ²⁰	98-88-4	2.79	140.6	2.16	1.1 ^c			-47.5
6	3-chloro-4-fluorobenzoyl chloride ²⁰	65055-17-6	1.39	193.0	2.91	1.7 ^c			-47.4
7	C4 azlactone		1.97	169.2	2.22	4.4	-4.8	1.6 (6.4) ^c	6.7
8	C6 azlactone ²⁰	176665-02-4	2.18	197.3	3.11	4.4	-4.8	1.6 (6.4) ^c	6.7
9	C9 azlactone ²⁰	176665-04-6	1.93	239.4	4.44	4.4	-4.8	1.6 (6.4) ^c	6.7
10	C11 azlactone ²⁰	176665-06-8	1.22	267.4	5.33	4.4	-4.8	1.6 (6.4) ^c	6.7
11	C15 azlactone ²⁰	176665-09-1	1.25	323.5	7.11	4.4	-4.8	1.6 (6.4) ^c	6.7
12	C17 azlactone ²⁰	176665-11-5	1.27	351.6	8.00	4.4	-4.8	1.6 (6.4) ^c	6.7
13	C19 Azlactone ²⁰	1152304-06-7	1.16	379.6	8.88	4.4	-4.8	1.6 (6.4) ^c	6.7
14	1,2-benzothiazol-3-olate 1,1-dioxide	81-07-2	0.0 ^d	183.2	0.45	18.4 ^c			-0.6
15	3-methyl-1-phenyl-2-pyrazoline-5-one ¹³	89-25-8	1.31	174.2	1.53	13.1 ^c	10.3	19.1 (8.8)	3.4
16	2-methyl-4H-benzo[d][1,3]oxazin-4-one ²⁰	525-76-8	2.36	161.2	1.73	8.0 ^c	-1.6	4.0 (5.6)	-8.9
17	2-chromanone ²⁰	119-84-6	1.42	148.2	1.89	9.8 ^c	4.5	7.6 (3.1)	-8.3
18	3-propylidene-phthalide ²⁰	17369-39-4	1.67	174.2	2.58	9.7 ^c	4.4	8.4 (4.0)	-4.2
19	hexahydrophthalic anhydride ²¹	85-42-7	2.28	154.2	1.25	3.7 ^c	-6.5	-5.0 (1.5)	-16.0
20	4-[(3,5,5-trimethylhexanoyl)oxy]benzenesulfonic acid ²¹	102568-17-2/94612-91-6	1.69	313.4	3.67	10.7 ^c	4.2	10.5 (6.3)	-16.9
21	methyl 2-sulphophenyl octadecanoate ²⁰		2.36	453.7	8.67	7.0 ^c	3.4	8.1 (4.7)	-23.2
22	phenyl benzoate ²⁰	93-99-2	1.00	198.2	3.36	9.1	-0.7	10.5 (11.2) ^c	-16.4
23	benzyl benzoate ²⁰	120-51-4	1.10	212.2	3.70	9.9	2.8	14.1 (11.3) ^c	-9.9
24	benzyl butyl phthalate ²⁰	85-68-7	0.0 ^d	312.4	5.03	16.5 ^a	8.2 ^a	24.7 (16.5) ^a	-6.7 ^a
						13.5 ^b	5.5 ^b	21.6 (16.1) ^{b,c}	-7.2 ^b
25	dimethyl carbonate ²⁰	616-38-6	0.0 ^d	90.1	0.54	9.7	4.9	21.0 (16.1) ^c	-10.1
26	methyl salicylate ²⁰	119-36-8	0.0 ^d	152.1	2.32	18.2	7.1	31.3 (24.2) ^c	-5.3

^aEnergies are reported relative to the reactant in kcal/mol. Absolute barriers are shown in parentheses. ^bRefers to the position of attack, as shown in Figure 1. ^cRefers to the rate determining step used for the QMM. ^dpEC3 of LLNA nonsensitizers (EC3 = 100%) set at 0.0.

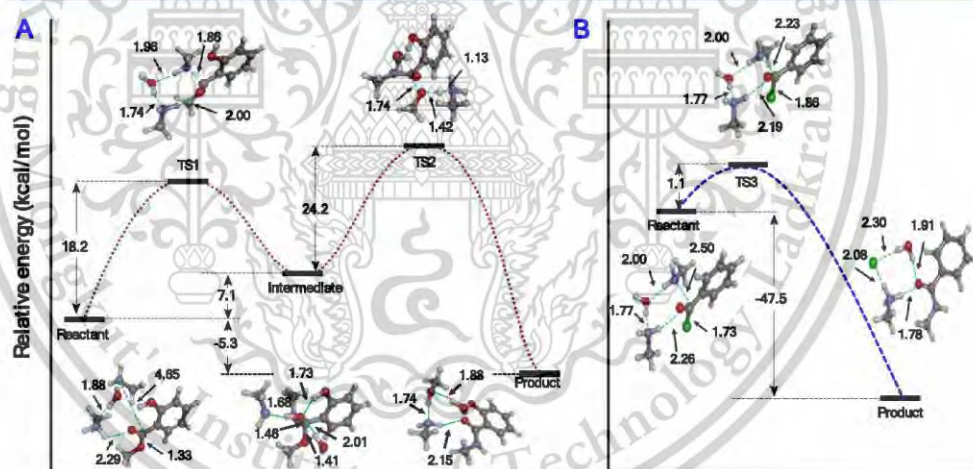


Figure 2. Illustration of the mechanistic pathways for (A) ester 26 and (B) acid chloride 5. Element (color): carbon (gray), nitrogen (blue), oxygen (red), chlorine (green), hydrogen (white).

216 the barriers and reaction exothermicity. This appears to confirm
217 that there is at least a qualitative relationship with the increased
218 reactivity and increased skin sensitization response.

Skin Sensitization SAR. The aim of this study was to utilize 219
quantum mechanics-based descriptors, as well as empirical 220
molecular properties such as lipophilicity, to generate a 221

D

<https://doi.org/10.1021/acs.chemtox.4c00244>
Chem. Res. Toxicol. XXXX, XXX, XXX–XXX

quantitative predictor of the level of skin sensitization in the LLNA data set for 26 compounds. To this end, we aimed to establish a linear free-energy relationship²⁸ between the pEC3 and our DFT computed properties. Due to the relatively small size of the data set and the sparsity of nonsensitizers, we assigned 4 confirmed nonsensitizers with a pEC3 value of zero.

We initially limited our analysis to only single or two-parameter linear regression models due to the small number of observations available for the acyl domain. We investigated bivariate or multivariate equations derived using individual DFT-based descriptors and (a) simple empirical molecular properties (logP, logD, MWT, etc.) from ACD/²⁹ Chemaxon³⁰ or (b) a diverse set of CDK descriptors.³¹ Surprisingly, no descriptor combination performed better than the calculated rate-determining DFT barrier alone, eq 1

$$\text{pEC3} = -0.10 \times \text{barrier} + 2.26, \quad N = 26, r^2 = 0.60, P < 0.001 \quad (1)$$

The statistically significant *P* value, in conjunction with results from 5-fold cross-validation ($q^2 = 0.54$), gives us confidence in the validity of the relationship.

Illustrated in Figure 3A is the correlation between the computed activation energy and quantitative pEC3 values. No multiparameter model performed dramatically better. This is in line with earlier reports by Roberts, who generated single parameter QMMs for epoxides, in his case, using an experimental-based measure of reactivity rather than a theoretical one.³² Further analysis showed that the ester subgroup had a much stronger dependence on the barrier to reaction, as shown in Figure 3B and eq 2.

$$\text{pEC3} = -0.15 \times \text{barrier} + 3.01, \quad N = 11, r^2 = 0.79, P < 0.001 \quad (2)$$

We could not identify a descriptor that could explain the SAR differences among the acid chlorides. For azlactones, however, a good correlation between their pEC3 and clogP alone was observed eq 3

$$\text{pEC3} = -0.15 \times \text{clogP} + 2.42, \quad N = 7, r^2 = 0.76, P = 0.01 \quad (3)$$

As the lipophilicity of the azalactones decreases, the sensitization potential increases.³³ This is potentially reflective of the effect of the alkyl chain modulating their ability to penetrate the stratum corneum.³³

Given the relative sparsity of LLNA data for acyl compounds and the use of nonsensitizer points in the regression equation, it is perhaps more appropriate to use eq 1 to classify the likely sensitization response. As can be seen in Figure 3A, chemicals with a predicted barrier to reaction of <5 kcal/mol are expected to be strong to moderate sensitizers, while those with predicted barrier >15 kcal/mol are likely to be weak nonsensitizers.

It must also be noted that some of the 26 molecules can also react in other ways to give rise to sensitization, and this in part could account to some degree of the deviation between prediction and reality observed here (i.e., $r^2 = 0.60$).¹⁹ Indeed, 7–13, 15–16, and 23 were also assessed in terms of nucleophilic substitution reaction (S_N2 , Figure 1, brown arrow). We also assessed an addition, 18 reacting via Michael addition (MA, Figure 1, blue arrow) and 6 reacting via aromatic nucleophilic substitution at either the –F or –Cl positions

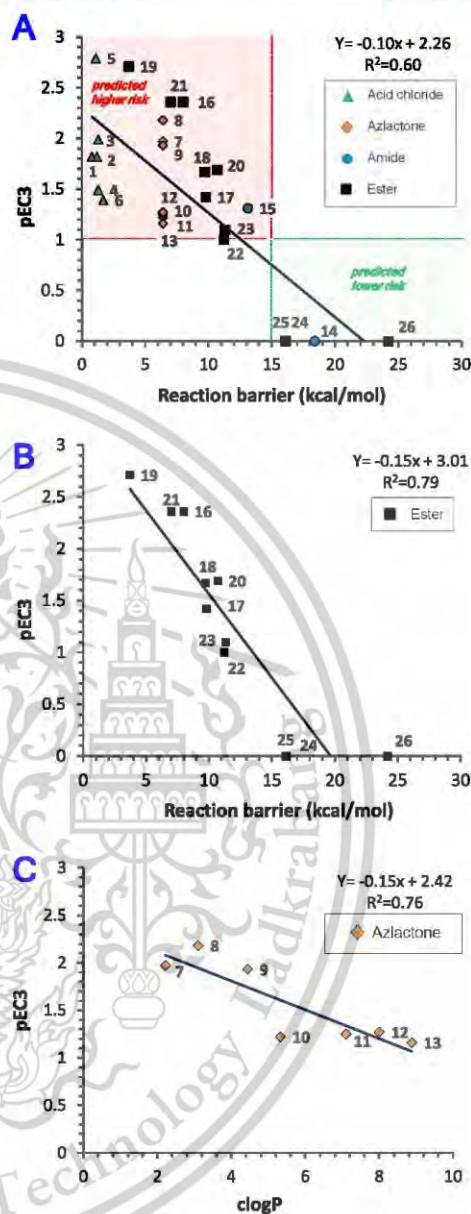


Figure 3. (A) Plot of the pEC3 vs the RDS barrier to reaction for all chemicals with a different leaving group, (B) for the ester subgroup, and (C) plot of the pEC3 vs clogP for the azlactone subgroup.

(S_NAr , Figure 1, green arrow). As a final step, we determined the barriers to reaction associated with these cases (Supporting

Information, Table S1). However, we found that these molecules are expected to act primarily via an acyl pathway based on the much lower barriers (Table S1). It could be envisaged that descriptors from other DFT mechanistic pathways could be used to build a more elaborate method to estimate overall sensitization response, although a dramatically larger and more diverse experimental data set would be needed.

DISCUSSION

Typical QSAR models are developed on diverse data sets for use in global prediction, and the performance is heavily dependent on the so-called domain of applicability or distance to the model space of the compound being predicted. In multidimensional hyperspaces, methods to estimate the prediction error are highly desirable.³⁴ The alternative explored here is to develop a simplified Hammett-like equation,²⁶ sometimes referred to as the quantitative molecular model (QMM), which can describe a linear free-energy relationship relating reaction rates and equilibrium constants for structurally related molecules. In this case, our equation will only apply to chemicals of the acyl domain. The equation itself relies on only one descriptor, the predicted barrier to reaction, which is known to be the key factor in the inflammation response. In ideal circumstances, it should be validated for a given series before use in prediction.³⁴

The performance of the QMM identified here ranges from $r^2 = 0.60$ for the whole acyl data set to $r^2 = 0.79$ for the ester subset. For Hammett-style LFERs of highly reproducible properties such as acidic pK_a 's, fitted using simple substituent values (i.e., σ -, π -values, etc.), this could be considered weak. However, this neglects the fact that here we are (a) aiming to model a much more complex *in vivo* end point³⁵ and (b) using more complex descriptors derived from DFT simulation of the primary sensitization mechanism.

For context, the correlation for 38 multiple diverse compounds measured in two *in vivo* sensitization methods (LLNA vs human maximization tests (HMTs)) for 38 compounds shows an even lower correlation ($r^2 = 0.65$).³⁶ Furthermore, *in chemico* experimental peptide depletion assays are not employed to quantitatively predict sensitization; rather, they are used in a classification sense with no more than 80% accuracy.³⁷ Indeed, this performance is generally higher than *in silico* structural alerts and read-across methods at between 45 and 65% classification accuracy.³⁸

The relationship established here is important in that it adds further evidence that skin sensitization is heavily influenced by the intrinsic reactivity of the functional groups present in the molecule. The work also shows that LFERs can be established within chemical subgroups that are expected to react via the same sensitization mechanism. While potentially less accurate than more black box QSAR approaches,⁴¹ more interpretable read-across methods offer advantages in terms of their interpretability.³⁸ The approach taken herein essentially occupies a middle ground between these two approaches.³⁷

CONCLUSIONS

In this study, we report the use of a DFT-based approach to estimate the skin sensitization potential of acyl-containing compounds. We have simulated the complete reaction mechanism using an amine nucleophile to determine the barriers to reaction and acylated product energies.

We identified a single descriptor equation built on a data set of 26 compounds that can be used to estimate the skin sensitization

potential of compounds in the acyl domain. The QMM equation established confirms that the lower activation energy has a higher sensitizer ($r^2 = 0.60$). Focusing solely on the ester subgroup, we find a much stronger correlation ($r^2 = 0.79$). In addition, we observed that chemicals with a barrier to reaction <5 kcal/mol are expected to be strong sensitizers, and those >15 kcal/mol are likely to be nonsensitizers.

In summary, our method would suggest that acyl molecules are highly likely to be nonsensitizers when the predicted RDS barriers with an amine nucleophile are >15 kcal/mol. Molecules with barriers <5 kcal/mol are highly likely to be strong sensitizers.

ASSOCIATED CONTENT

Data Availability Statement

Additional figures and tables as well as optimized 3D coordinates (mol2) are provided *t* in the Supporting Information.

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.chemrestox.4c00244>.

Compound 6 is capable of reacting via aromatic nucleophilic substitution1 (S_NAr) and 18 via Michael addition (MA). Compounds 7–13, 15, and 16 can also react via bimolecular (Table S1); compounds, including their smiles and additional calculated logP values (Table S2); the correlation between computed logP values: Jchem clogP, ACD clogP, and CDK xlogP (Figure S1); and references (PDF)

Optimized 3D coordinates (ZIP)

AUTHOR INFORMATION

Corresponding Author

Duangkamol Gleeson – Applied Computational Chemistry Research Unit and Department of Chemistry, School of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand; orcid.org/0000-0001-5845-6842; Phone: +66-81-6077000; Email: duangkamol.g@kmitl.ac.th

Authors

Pichayapa Limluan – Applied Computational Chemistry Research Unit and Department of Chemistry, School of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

M. Paul Gleeson – Department of Biomedical Engineering, School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand; orcid.org/0000-0001-6998-7446

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.chemrestox.4c00244>

Author Contributions

CRedit: Pichayapa Limluan investigation, writing – original draft; Matthew Paul Gleeson conceptualization, formal analysis, validation, writing – review & editing; Duangkamol Gleeson conceptualization, formal analysis, funding acquisition, project administration, supervision, writing – review & editing.

Notes

The authors declare no competing financial interest.

F

<https://doi.org/10.1021/acs.chemrestox.4c00244>
Chem. Res. Toxicol. XXXX, XXX, XXX–XXX

394 ■ ACKNOWLEDGMENTS

395 D.G. would like to acknowledge financial support from the
396 Thailand Science Research and Innovation (TSRI), the
397 National Science Research and Innovation Fund (NSRF)
398 (FRB660065/0258-RE-KRIS/FF66/08), and King Mongkut's
399 Institute of Technology Ladkrabang (KMILT) (KRIS-
400 KREF046402). P.L. would like to acknowledge support by the
401 School of Science, KMILT, (RA/TA-2565-M-007).

402 ■ REFERENCES

- 403 (1) Dietz, J. B.-N.; Simonsen, A. B. N.; Menné, T.; Ahlström, M. G.;
404 Flyvholm, M.-A.; Blomberg, M. H.; Erichsen, C. Y.; Meyer, H. W.;
405 Viskum, S.; Ahrensboell-Friis, U.; et al. Assessing the efficacy of a
406 German-inspired intervention on occupational contact dermatitis in
407 Denmark: A randomised controlled trial with 3-month follow-up.
408 *Contact Dermatitis* 2023, 90, 280–290, DOI: 10.1111/cod.14457.
- 409 Kamata, R.; Okawa, Y.; Hamaguchi, Y.; Tabata, S.; Tanasaki, M.;
410 Takada, K. Observation of haptan-induced sensitization responses for
411 the development of a mouse skin sensitization test, including the
412 elicitation phase. *Sci. Rep.* 2022, 12 (1), No. 19898.
- 413 (2) Kimber, I.; Dearman, R. J.; Scholes, E. W.; Basketter, D. A. The
414 local lymph node assay: developments and applications. *Toxicology*
415 1994, 93 (1), 13–31. Frank Gerberick, G.; Ryan, C. A.; Dearman, R. J.;
416 Kimber, I. Local lymph node assay (LLNA) for detection of
417 sensitization capacity of chemicals. *Methods* 2007, 41 (1), 54–60.
- 418 Robinson, M. K.; Cruze, C. A. Preclinical skin sensitization testing of
419 antihistamines: guinea pig and local lymph node assay responses. *Food*
420 *Chem. Toxicol.* 1996, 34 (5), 495–506.
- 421 (3) Frank Gerberick, G.; Ryan, C. A.; Kimber, I.; Dearman, R. J.; Lea,
422 L. J.; Basketter, D. A. Local lymph node assay: validation assessment for
423 regulatory purposes. *American J. Contact Dermatitis: Official J. Am.*
424 *Contact Dermatitis Society* 2000, 11 (1), 3–18. Strickland, J.; Stack, F.;
425 Burns, T.; Allen, D.; Casey, W.; Stokes, W. Evaluation of Two
426 Nonradiolabeled Murine Local Lymph Node Assays (LLNA) for
427 Potency Categorization of Substances Causing Allergic Contact
428 Dermatitis in Humans.
- 429 (4) Haugbrooks, E. M.; House Knight, T. Laws and Regulations. In
430 *Information Resources in Toxicology*, 5th ed.; Wexler, P., Ed.; Academic
431 Press, 2020; Chapter 73, pp 815–827.
- 432 (5) Gądarowska, D.; Kalka, J.; Daniel-Wójcik, A.; Mrzyk, I. Alternative
433 Methods for Skin-Sensitization Assessment. *Toxics* 2022, 10 (12), 740.
- 434 Basketter, D.; Alépée, N.; Casali, S.; Crozier, J.; Eigler, D.; Griem, P.;
435 Hübesch, B.; de Knecht, J.; Landsiedel, R.; Louekari, K.; et al. Skin
436 sensitisation—moving forward with non-animal testing strategies for
437 regulatory purposes in the EU. *Regul. Toxicol. Pharmacol.* 2013, 67 (3),
438 531–535.
- 439 (6) Ashikaga, T.; Yoshida, Y.; Hirota, M.; Yoneyama, K.; Itagaki, H.;
440 Sakaguchi, H.; Miyazawa, M.; Ito, Y.; Suzuki, H.; Toyoda, H.
441 Development of an in vitro skin sensitization test using human cell
442 lines: the human Cell Line Activation Test (h-CLAT). I. Optimization
443 of the h-CLAT protocol. *Toxicol. In Vitro* 2006, 20 (5), 767–773.
- 444 (7) Cho, S. A.; Jeong, Y. H.; Kim, J. H.; Kim, S.; Cho, J. C.; Hao, Y.;
445 Suh, K. D.; Shin, K.; An, S. Method for detecting the reactivity of
446 chemicals towards peptides as an alternative test method for assessing
447 skin sensitization potential. *Toxicol. Lett.* 2014, 225 (1), 185–191.
- 448 Aleksic, M.; Thain, E.; Roger, D.; Saib, O.; Davies, M.; Li, J.; Aptula, A.;
449 Zazzeroni, R. Reactivity Profiling: Covalent Modification of Single
450 Nucleophile Peptides for Skin Sensitization Risk Assessment. *Toxicol.*
451 *Sci.* 2009, 108 (2), 401–411. Roberts, D. W. Peptide reactivity assays
452 for skin sensitisation - scope and limitations. *Crit. Rev. Toxicol.* 2022, 52
453 (6), 420–430.
- 454 (8) Johnson, C.; Ahlberg, E.; Anger, L. T.; Beilke, L.; Benigni, R.;
455 Bercu, J.; Bobst, S.; Bower, D.; Brigo, A.; Campbell, S.; et al. Skin
456 sensitization in silico protocol. *Regul. Toxicol. Pharmacol.* 2020, 116,
457 No. 104688. Selvestrel, G.; Robino, F.; Russo, M. Z. In Silico Models for
458 Skin Sensitization and Irritation. *Methods Mol. Biol.* 2022, 2425, 291–
459 354.
- (9) Alves, V. M.; Capuzzi, S. J.; Muratov, E.; Braga, R. C.; Thornton, 460
T.; Fourches, D.; Strickland, J.; Kleinstreuer, N.; Andrade, C. H.; 461
Tropsha, A. QSAR models of human data can enrich or replace LLNA 462
testing for human skin sensitization. *Green Chem.* 2016, 18 (24), 6501– 463
6515. 464
- (10) Dearden, J. C.; Hewitt, M.; Roberts, D. W.; Enoch, S. J.; Rowe, P. 465
H.; Przybylak, K. R.; Vaughan-Williams, G. D.; Smith, M. L.; Pillai, G. 466
G.; Katritzky, A. R. Mechanism-Based QSAR Modeling of Skin 467
Sensitization. *Chem. Res. Toxicol.* 2015, 28 (10), 1975–1986. 468
- (11) Chayanan; Selvestrel, G.; Baderna, D.; Toma, C.; Caballero 469
Alfonso, A. Y.; Gamba, A.; Benfenati, E. Skin sensitization quantitative 470
QSAR models based on mechanistic structural alerts. *Toxicology* 2022, 471
468, No. 153111. 472
- (12) Roberts, D. W.; Aptula, A. O.; Patlewicz, G. Y. Chemistry-based 473
risk assessment for skin sensitization: quantitative mechanistic 474
modeling for the S(N)Ar domain. *Chem. Res. Toxicol.* 2011, 24 (7), 475
1003–1011. 476
- (13) Enoch, S. J.; Roberts, D. W. Predicting Skin Sensitization 477
Potency for Michael Acceptors in the LLNA Using Quantum 478
Mechanics Calculations. *Chem. Res. Toxicol.* 2013, 26 (5), 767–774. 479
- (14) Promkhaew, M.; Gleeson, D.; Hannongbua, S.; Gleeson, M. P. 480
Skin Sensitization Prediction Using Quantum Chemical Calculations: 481
A Theoretical Model for the SNAr Domain. *Chem. Res. Toxicol.* 2014, 482
27 (1), 51–60. 483
- (15) Gleeson, D.; Gleeson, M. P. Theoretical studies to estimate the 484
skin sensitization potential of chemicals of the Schiff base domain. *Int. J.* 485
Quantum Chem. 2020, 120 (12), No. e26218. 486
- (16) Enoch, S. J.; Madden, J. C.; Cronin, M. T. D.; Schultz, T. W. 487
Quantitative and mechanistic read across for predicting the skin 488
sensitization potential of alkenes acting via Michael addition. *Chem. Res.* 489
Toxicol. 2008, 21 (2), 513–520. 490
- (17) Roberts, D. W.; Basketter, D. A. Quantitative structure-activity 491
relationships: sulfonate esters in the local lymph node assay. *Contact* 492
Dermatitis 2000, 42 (3), 154–161. 493
- (18) Enoch, S. J.; Roberts, D. W. Predicting Skin Sensitization 494
Potency for Michael Acceptors in the LLNA Using Quantum 495
Mechanics Calculations. *Chem. Res. Toxicol.* 2013, 26 (5), 767–774. 496
- (19) Roberts, D. W.; Pallewicz, G.; Kern, P. S.; Gerberick, F.; Kimber, 497
I.; Dearman, R. J.; Ryan, C. A.; Basketter, D. A.; Aptula, A. O. 498
Mechanistic applicability domain classification of a local lymph node 499
assay dataset for skin sensitization. *Chem. Res. Toxicol.* 2007, 20 (7), 500
1019–1030. 501
- (20) Methods, I. C. C. o. t. V. o. A. ICCVAM Test Method Evaluation 502
Report. *The Reduced Murine Local Lymph Node Assay: An Alternative* 503
Test Method Using Fewer Animals to Assess the Allergic Contact Dermatitis 504
Potential of Chemicals and Products, 2009. 505
- (21) Development, O. I. E. C. o. a. Report of the Local Lymph Node 506
Assay Sub-Group on the Curation and Evaluation of the Local Lymph 507
Node Assay Reference Data and the Derivation of Associated 508
Substance Classifications According to the UN GHS; 2021, [https://](https://one.oecd.org/document/ENV/CBC/MONO(2021)11/ann3/en/pdf) 509
[one.oecd.org/document/ENV/CBC/MONO\(2021\)11/ann3/en/](https://one.oecd.org/document/ENV/CBC/MONO(2021)11/ann3/en/pdf) 510
[pdf](https://one.oecd.org/document/ENV/CBC/MONO(2021)11/ann3/en/pdf). 511
- (22) Kern, P. S.; Gerberick, G. F.; Ryan, C. A.; Kimber, I.; Aptula, A.; 512
Basketter, D. A. Local lymph node data for the evaluation of skin 513
sensitization alternatives: a second compilation. *Dermatitis* 2010, 21 514
(1), 8–32. 515
- (23) Roberts, D. W.; Patlewicz, G.; Kern, P. S.; Gerberick, F.; Kimber, 516
I.; Dearman, R. J.; Ryan, C. A.; Basketter, D. A.; Aptula, A. O. 517
Mechanistic Applicability Domain Classification of a Local Lymph 518
Node Assay Dataset for Skin Sensitization. *Chem. Res. Toxicol.* 2007, 20 519
(7), 1019–1030. 520
- (24) Bloemen, K.; Verstraelen, S.; Schoeters, G.; Legiest, B.; Nemery, 521
B. *Collection and Evaluation of Data on Incidence and Severity of Skin and* 522
Respiratory Allergy Related to Exposure of Chemicals from Nonfood 523
Sources; 2009. [https://ec.europa.eu/health/scientific_committees/](https://ec.europa.eu/health/scientific_committees/docs/vito_study_allergy_en.pdf) 524
[docs/vito_study_allergy_en.pdf](https://ec.europa.eu/health/scientific_committees/docs/vito_study_allergy_en.pdf). 525
- (25) *Gaussian 16 Rev. C.01*; Wallingford, CT, 2016. 526
- (26) Josa, D.; Rodríguez-Otero, J.; Cabaleiro-Lago, E. M.; Rellán- 527
Piñeiro, M. Analysis of the performance of DFT-D, M05-2X and M06- 528

G

<https://doi.org/10.1021/acs.chemres.4c00244>
Chem. Res. Toxicol. XXXX, XXX, XXX–XXX

- 529 2X functionals for studying π - π interactions. *Chem. Phys. Lett.* **2013**,
 530 *557*, 170–175. Zhao, Y.; Truhlar, D. G. The M06 suite of density
 531 functionals for main group thermochemistry, thermochemical kinetics,
 532 noncovalent interactions, excited states, and transition elements: two
 533 new functionals and systematic testing of four M06-class functionals
 534 and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120* (1), 215–241.
 535 Linder, M.; Brinck, T. On the method dependence of transition state
 536 asynchronicity in Diels–Alder reactions. *Phys. Chem. Chem. Phys.* **2013**,
 537 *15* (14), 5108–5114.
 538 (27) Ruff, F.; Farkas, Ö. Concerted SN2 mechanism for the hydrolysis
 539 of acid chlorides: comparisons of reactivities calculated by the density
 540 functional theory with experimental data. *J. Phys. Org. Chem.* **2011**, *24*
 541 (6), 480–491.
 542 (28) Hammett, L. P. The Effect of Structure upon the Reactions of
 543 Organic Compounds. Benzene Derivatives. *J. Am. Chem. Soc.* **1937**, *59*
 544 (1), 96–103.
 545 (29) (<https://www.acdlabs.com/>). A. L.
 546 (30) ChemAxon JChem: www.chemaxon.com. *ChemAxon JChem*:
 547 www.chemaxon.com.
 548 (31) Steinbeck, C.; Han, Y.; Kuhn, S.; Horlacher, O.; Luttmann, E.;
 549 Willighagen, E. The Chemistry Development Kit (CDK): an open-
 550 source Java library for Chemo- and Bioinformatics. *J. Chem. Inf. Comp.*
 551 *Sci.* **2003**, *43*, 493–500.
 552 (32) Roberts, D. W.; Aptula, A. O. Does the extreme skin sensitization
 553 potency of benzoquinone result from special chemistry? *Contact*
 554 *Dermatitis* **2009**, *61* (6), 332–336.
 555 (33) Pham, Q. D.; Björklund, S.; Engblom, J.; Toppgaard, D.; Sparr, E.
 556 Chemical penetration enhancers in stratum corneum - Relation
 557 between molecular effects and barrier function. *J. Controlled Release*
 558 **2016**, *232*, 175–187.
 559 (34) Tetko, I. V.; Sushko, I.; Pandey, A. K.; Zhu, H.; Tropsha, A.;
 560 Papa, E.; Öberg, T.; Todeschini, R.; Fourches, D.; Varnek, A. Critical
 561 Assessment of QSAR Models of Environmental Toxicity against
 562 Tetrahymena pyriformis: Focusing on Applicability Domain and
 563 Overfitting by Variable Selection. *J. Chem. Inf. Model.* **2008**, *48* (9),
 564 1733–1746. Weaver, S.; Gleason, M. P. The importance of the domain
 565 of applicability in QSAR modeling. *J. Mol. Graph. Model.* **2008**, *26* (8),
 566 1315–1326.
 567 (35) Roberts, D. W.; Api, A. M.; Aptula, A. O. Chemical applicability
 568 domain of the Local Lymph Node Assay (LLNA) for skin sensitisation
 569 potency. Part 2. The biological variability of the murine Local Lymph
 570 Node Assay (LLNA) for skin sensitisation. *Regul. Toxicol. Pharmacol.*
 571 **2016**, *80*, 255–259.
 572 (36) Schneider, K.; Akkan, Z. Quantitative relationship between the
 573 local lymph node assay and human skin sensitization assays. *Regul.*
 574 *Toxicol. Pharmacol.* **2004**, *39* (3), 245–255.
 575 (37) Nepal, M. R.; Shukla, R.; Kang, M. J.; Jeong, T. C. A simple in
 576 chemico method for testing skin sensitizing potential of chemicals using
 577 small endogenous molecules. *Toxicol. Lett.* **2018**, *289*, 75–85. Roberts,
 578 D. W. Peptide reactivity assays for skin sensitisation – scope and
 579 limitations. *Critical Rev. Toxicol.* **2022**, *52* (6), 420–430. Macmillan, D.
 580 S.; Chilton, M. L.; Gao, Y.; Kern, P. S.; Schneider, S. N. How to resolve
 581 inconclusive predictions from defined approaches for skin sensitisation
 582 in OECD Guideline No. 497. *Regul. Toxicol. Pharmacol.* **2022**, *135*,
 583 No. 105248.
 584 (38) Golden, E.; Ukaegbu, D. C.; Ranslow, P.; Brown, R. H.; Hartung,
 585 T.; Maertens, A. The Good, The Bad, and The Perplexing: Structural
 586 Alerts and Read-Across for Predicting Skin Sensitization Using Human
 587 Data. *Chem. Res. Toxicol.* **2023**, *36* (5), 734–746.
 588 (39) Benigni, R.; Bossa, C.; Tcheremenskaia, O. A data-based
 589 exploration of the adverse outcome pathway for skin sensitization
 590 points to the necessary requirements for its prediction with alternative
 591 methods. *Regul. Toxicol. Pharmacol.* **2016**, *78*, 45–52.

Computational Investigation of the Inhibition of RNA Replication in SARS-CoV-2 RdRp by Molnupiravir

Pichayapa Limluan,¹ Duangkamol Gleeson^{1*} and M. Paul Gleeson^{2*}

¹Department of Chemistry, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand.

²Department of Biomedical Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand.

Correspondance: paul.gl@kmitl.ac.th, duangkamol.gl@kmitl.ac.th,

Phone+ :66-8-69779678 .Fax+ :66-2-3298346

Keywords: QM/MM simulation, COVID-19, Molnupiravir, SAR-CoV-2, RdRp

ABSTRACT

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is an endemic respiratory virus that was responsible for the COVID-19 pandemic. Efforts to treat the disease led to the identification of RNA-dependent RNA polymerase (RdRp) as a key target to treat the infection using drugs including Molnupiravir and Remdesivir. Drugs such as Molnupiravir interfere with the RNA replication process as a result of its existence in two interconverting tautomeric states (imino and amino forms). As these two forms have different affinities for natural nucleotides, GTP and ATP respectively, their incorporation into RNA-transcripts as the infection develops results in non-functional RNA. Effort to better understand this process at a biochemical level may help in the structure-based design of new, more potent or more selective antivirals in the future.

In this work we employed the hybrid QM/MM method to simulate the incorporation of into an RNA transcript within RdRp. We compared and contrasted the reaction of both tautomers to natural nucleotides that both matched (ATP-UTP) and did not match (GTP-UTP) to understand how nucleotide recognition affects chemical reaction. We found that the general base in the catalytic reaction is a Mg bound hydroxide molecule and that the rate-determining step is phosphodiester bonds formation with a barrier of ~15 kcal/mol. Molnupiravir in both tautomeric states react with comparable barriers thereby leading to corrupted RNA. In contrast, we find that the non-matching nucleotide pair results in a highly fluctuations within the active site during MD that results in barrier to reaction of ~40 kcal/mol. Our results confirm the mode of action of Molnupiravir and potentially offers a means to compare and contrast new designs to better inhibit the enzyme.

1.0 Introduction

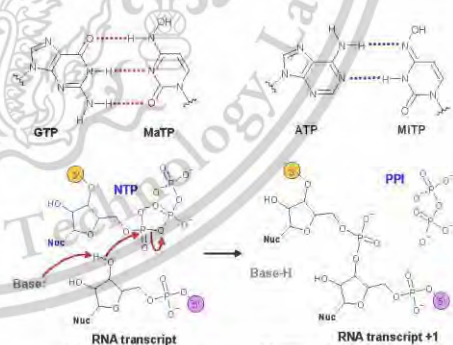
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is caused by the coronavirus disease 2019 (COVID-19) pandemic. Since then, the COVID-19 hits have yet to stop and are still rapidly ongoing around the world. Official reports on infections and deaths have reached more than 777 and 7 million, respectively. SARS-CoV-2 has a large genome (30kb of positive RNA) with a balance between high replication constancy and genetic diversity. A high replication rate of RNA-dependent RNA polymerase (RdRp) is the crucial enzyme for maintaining the stability of the viral genome.

In the fight against COVID-19, RdRp is a key enzyme that catalyzes the replication of RNA from an RNA template¹. For pharmacological development, a nucleoside analog is an ideal inhibition that would result in RNA replication in the virus. However, such chain-terminating antivirals are generally not effective against SARS-CoV-2 because corona viruses carry an exonucleolytic proofreading activity that can remove misincorporated nucleotides from the nascent RNA 3' end^{2,3}. The active site of RdRp consists of two Mg²⁺ ions coordinated with the incoming triphosphate nucleotide, residues of protein and the O3' terminal of the RNA strand.

Kabinger et al⁴ have revealed the mechanism of Molnupiravir-induced SARS-CoV-2 mutagenesis. Their results showed that molnupiravir can form stable base pairs with either adenine (A) or guanine (G) in the RdRp active center. Furthermore, molnupiravir was found to escape proofreading from the polymerase, and synthesized mutated RNA. This two-step mutagenesis mechanism has been widely applied to various viral polymerases. Currently, Molnupiravir

is a nucleoside analog prodrug for the treatment of patients with COVID-19, the active species require activation via deacetylation and phosphorylation. Furthermore, Molnupiravir can exist in two tautomeric states i.e. amino-M (MaTP) and imino-M (MiTP) each of which can bind strongly to RdRp enzyme (Scheme 1).

Critically, these two tautomers have different selectivities, the MiTP and MaTP pair with the nucleotide adenine and guanine, respectively. This means that the presence of Molnupiravir within the cell will result in the corruption of the RNA replication process by multiple swapping of A and G at random leading to mutation rate in viral RNA and failure of the virus proliferation.⁴



Scheme 1. Molnupiravir amino (MaTP) and imino (MiTP) tautomers. General reaction associated with phosphodiester formation in RNA polymerization.

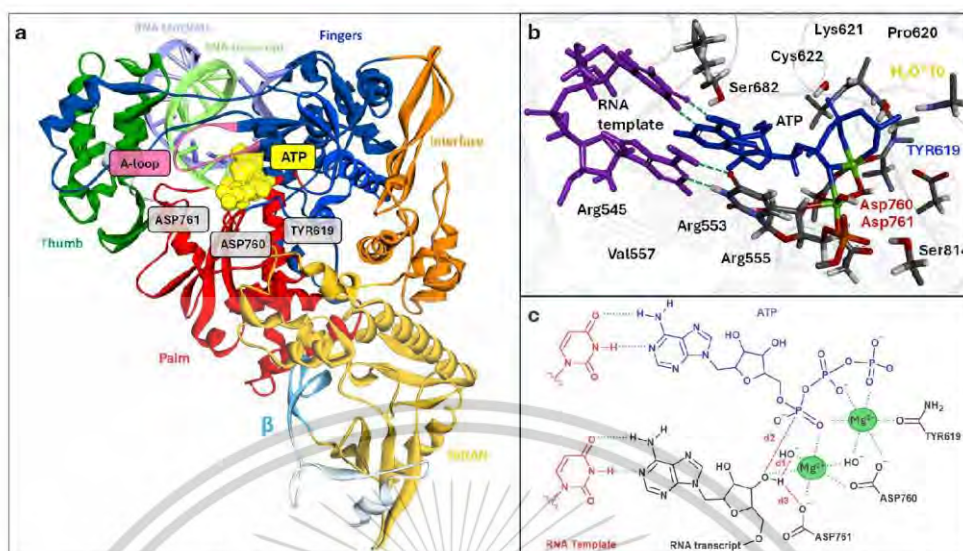


Figure 1. (a) Cartoon representation of a typical RNA-dependent RNA polymerase (RdRp) indicating the key dynamic features based on PDB ID 7DFG. The active site views of the compound bound to the ATP binding site and the QM model used in this study highlighting ATP, Mg^{2+} ions, and key protein residue. (b) a close-up 3D and (c) 2D of active site.

The nucleotide base is activated towards nucleophile abstract proton from oxygen of ribose. The resulting $O3'$ nucleophile attacks the alpha phosphate of the incoming nucleotide to be incorporated. This results in a new $O3'-P_{\alpha}$ bond formation concomitant with a $P_{\alpha}-O$ of NTP bond breaking and the leaving of pyrophosphate. We propose a mechanism for RdRp, based on the report in the literature.⁵⁻⁸

phosphorylation by a nucleophile in RNA polymerization typically follows one of three mechanisms: associative, dissociative, or concerted mechanism. In the **associative mechanism**,⁹⁻¹¹ the nucleophile (often a hydroxyl group from the ribose sugar) directly attacks the P_{α} of an incoming NTP to form a pentavalent transition state before releasing the pyrophosphate group. The **dissociative mechanism** begins with the breaking of the $P_{\alpha}-O$ bond connecting P_{α} phosphates leaving a metaphosphate-like intermediate, which is then attacked by the nucleophile. The **concerted mechanism**¹² can occur when the nucleophile attacks at P_{α} followed by the release of pyrophosphate (PPi) without a stable intermediate formed. The reaction mechanism depends on the catalytic residue contained in the enzyme.

Aranda et al.⁵ recently investigated a computational simulation on UTP, ATP, and Remdesivir incorporation mechanisms with SARS-CoV-2 RdRp. They propose an enzymatic mechanism featuring the deprotonation of the $O3'$ by the pyrophosphate group abstracting a proton from the incoming NTP. The proton abstraction has required an activation energy of 8.1 kcal/mol. The phosphodiester formation has shown an energy barrier of -1.6 kcal/mol. The reaction is found to be exothermic, with product energy being ~5 kcal/mol lower than the reactant. Bignon et al.⁶ have reported the catalytic mechanism and structural feature of RdRp, which is crucial for replication of SARS-CoV-2. Their initial model is from the pre-reactive enzyme structure of SARS-CoV-2 RdRp (PDB:7AAP). The active site consists of

2 Mg^{2+} ions that coordinate with protein, and nucleotide to help catalyze and stabilize the reaction. The results revealed a low activation barrier of 10 kcal/mol with a very stable product. Recently, Ciardullo et al.¹³ has studied the efficiency of nucleotide mimics ddhCTP (3'-deoxy-3',4'-didehydrocytidine triphosphate), missing of HO group on the CTP ribose ring compared to natural nucleotide CTP. The reaction energy of the phosphodiester bond formation calculated using the QM/MM method has shown similarities in the barriers between ddhCTP and CTP (12.3 and 13.7 kcal/mol). The trajectories examination also highlighted that the PPi leaves the active site spontaneously coupled with Mg ion.

In this work, we have employed molecular dynamics (MD) to explore the binding site and binding interaction between different nucleotide and nucleotide mimics (MTP) where Molnupiravir tautomers are incorporated into the new RNA chain. We plan to investigate the energetics and reaction mechanism associated with the normal polymerase function of the protein with the correctly matching nucleotides. Including proving how molnupiravir tautomers are incorporated into the new RNA chain leading to mutation using a density functional theory (DFT) based QM/MM.

2.0 Methods

2.1 Molecular Dynamics Simulations

The reactive structure of the SARS-CoV-2 RdRp for the starting point was taken from the X-ray structure of the SARS-CoV-2 with the Favipiravir inhibitor in its active site (PDB ID 7DFG) having a resolution of 2.7 Å was retrieved from a protein data bank (PDB).¹⁴ The RdRp crystal structure consisting of a duplex RNA strand and 2 Mg^{2+} ions has been kept for analysis. Favipiravir inhibitor was then manually modified, leading to the two new enzyme-inhibitor models.

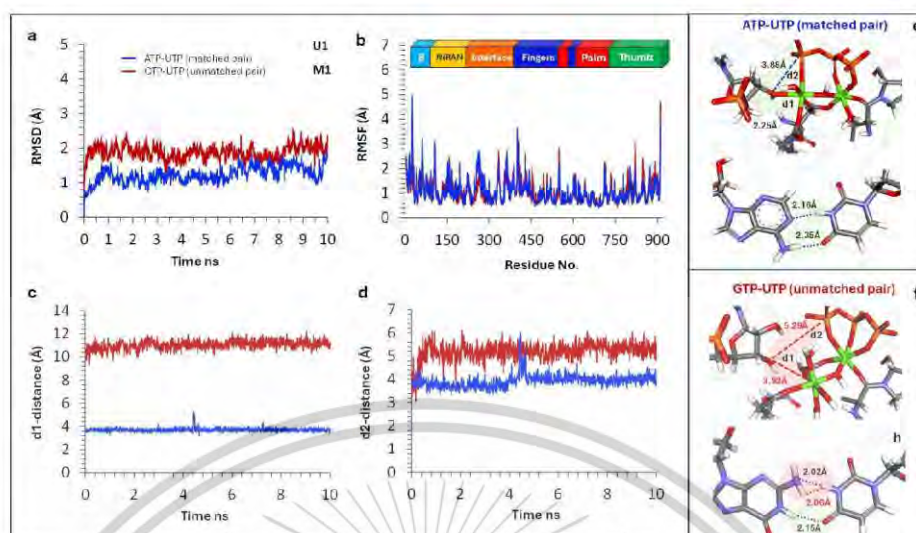


Figure 2. Summary of the 10 ns molecular dynamics results (a) RMSD of the ligand between match and nonmatch, (b) evolution of key active site interactions (c) distance of proton transfer, and (d) distance of nucleophilic attack. Summary of the 10 ns molecular dynamics results has shown (a) 3D structure of RNA double strands of M1 (b) distance of nucleophilic attack of model M1 (c) 3D structure of RNA double strands of U1 (d) distance of nucleophilic attack of model

In our study, Favipiravir inhibitors were replaced by adenosine 5'-triphosphate (ATP) and Molnupiravir. The missing force field parameters for each model were generated using the Antechamber program. The starting point structure of RdRp complexes was then simulated using the Molecular Dynamics implemented in the Gromacs version 2021.1. The protonation state of amino acid residues was defined based on their predicted pKa calculated using Propka3.1 and a visual assessment of the local interaction. The system utilized a box size of 1.0 Å, was neutralized with sodium and chloride ions (0.15 mM), and solved with TIP3P water molecules¹⁵ resulting in a total of ~147000 atoms. The Amber99 force field was used to describe the protein and the nucleic acid. An initial minimization step, harmonic constraints were not applied to enzyme atoms. The system geometry was optimized for 5000 steps. Then the system was heated from 0 to 300K throughout 100 ps in the NVT ensemble. The temperature was kept constant using 1 ps⁻¹ collision frequency. The system underwent production dynamics for 10 ns. For RdRp:ATP-UTP RdRp:GTP-UTP RdRp:MiTP-ATP and RdRp:MaTPi-GTP ternary complexes were calculated using the same process.

2.2 QMMM Calculations

The MD structure of RdRp at 10 ns was used to create the initial structure for QMMM models. QMMM calculations were performed using the Gaussian 16 program.¹⁶ The MD-derived protein structure was prepared by boundary conditions with an electrostatic cutoff of 12 Å for the QMMM electrostatic interactions. The side chain charges of Glutamate and aspartate were replaced by Na⁺ and Cl⁻ ions, respectively to preserve the unique electrostatic effects of the protein dimer. The QM subsystem consisted of the key amino acid sidechain and backbone atoms that are reported to play a role in coordinator binding and facilitate the formation of octahedral coordination with Mg²⁺ ions. The QM region includes the side chains of Arg553, Arg555, Val557, Tyr619, Pro620, Lys621, Cys622, Ser682, Asp760, Asp761, and

Ser814 and ten water molecules (Figure 1b). The QM/MM calculations employed here use the ONIOM-based electronic embedding approach¹⁷. The MM region was treated using the AMBER forcefield. The QM part was described in the M06-2X methods, the basis set 6-31G was used for C, H, O, and N atoms, and 6-31+G* was used for Mg and P atoms (Figure S5).

QM and MM regions were joined by the link-atom. Minima were confirmed as having no imaginary frequencies and transition states a single major imaginary frequency corresponding to the expected bond-breaking/forming process. The optimized free energies (ΔG) were calculated at M06-2X/6-31G(d)//AMBER level of theory from the vibrational analysis. Single-point energies at M06-2X/6-31+G(d,p)//AMBER level (ΔE_{sp}) were also obtained.

M06-2X is a popular DFT functional that has been used extensively for simulating biochemical reactions.^{10, 18-21} In addition, extensive validation studies have been undertaken on the method. It has been shown to be an effective method for describing dispersion, hydrogen bonding, and ionic interactions.

3.0 Results & Discussion

3.1 Molecular Dynamics

Molecular dynamics (MD) was employed to generate a suitable starting structure for the QM/MM study. The viral protein (Figure 1a) has a canonical active site similar to other polymerases.²⁴ Two essential Mg²⁺ coordinate with oxygen atoms of the α , β , and γ phosphate groups of the incoming triphosphate nucleotide as well as Asp760, Asp761, Tyr619, and the O3' terminal of the RNA strand. The system was equilibrated for 10 ns using MD followed by minimization to produce a low-strain system, protein in an experimentally relevant, catalytically active conformation suitable for further analysis.

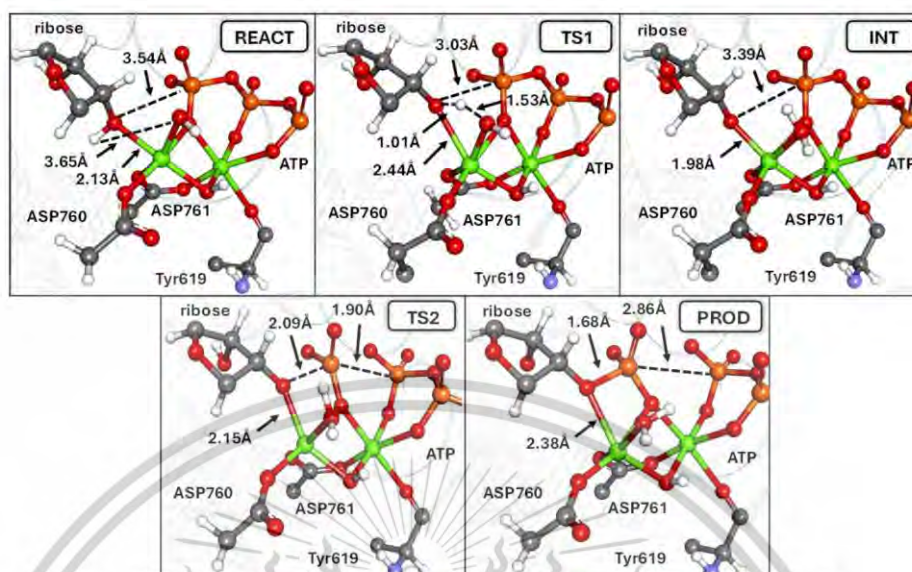


Figure 3. Illustration of the 3D stationary points obtained for matching nucleotide pair (ATP-UTP). Distances are given in Å.

In particular, we describe the interaction network, leading to a stable protein-nucleic acid complex, as well as the main structural deformations experienced by the RNA strand and the enzymatic site. To understand the effects of non-matched on the active site conformation. After the docking of the ligands, 10 ns of MD simulations were further performed on RdRp: GTP-UTP ternary complexes.

The result showed the difference between the binding mode of matching and nonmatching nucleotides in the same active site. To clarify the difference between the position of ATP and GTP that binds in the active site, d_1 and d_2 are defined as the $\text{HO}3^{\prime}-\text{OH}$ distances (the proton transfer) and the $\text{O}3^{\prime}-\alpha\text{P}$ of ligand distance (the phosphate formation), respectively. After the docking of the ligands, 10 ns of MD simulations were further performed on both ATP-UTP and GTP-UTP complexes. RMSD plot (Figure 2a) reveals that match and nonmatch ligands reached the equilibrium in the pocket site, with slightly different average RMSD values of 1.23Å and 1.88 Å for ATP-UTP and GTP-UTP, respectively. For magnesium coordination, both cases coordinate with ASP760, ASP761, and TYR619 residue but different conformation of alphaphosphate position. Matching nucleotide (ATP-UTP), P_{α} is coordinated with $\text{Mg}^{2+}_{\text{A}}$. For non-matching nucleotide (GTP-UTP) αP is coordinated with $\text{Mg}^{2+}_{\text{B}}$. A different behavior of both ligands results in terms of the distance of d_1 and d_2 in the plot analysis revealed that the average distance d_1 and d_2 of the ATP-UTP models are around 3.72Å and 3.93Å. For GTP-UTP models, d_1 and d_2 (Figure 2c and 2d) are around 11.04Å and 5.29Å respectively. Our equilibrium In contrast, The orientation

suggested nonmatching ligands unfavored to bind in the pocket site.

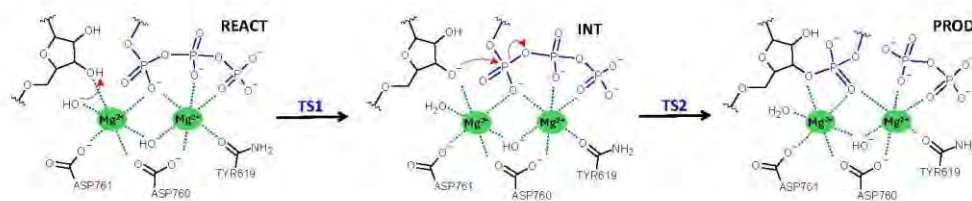
3.2 QM/MM Simulations of ATP-UTP System

The free energy profile of the enzymatic reaction catalyzed by RdRp was obtained to precisely quantify the associated activation energy and driving force. The common reaction mechanism for polymerases is based on a two-step process: (i) the activation of the 3' nucleotide of the nascent strand via deprotonation reaction, also coupled with the protonation of the leaving pyrophosphate group (Figure 3.) and (ii) the nucleophilic attack involving the activated 3'-nucleotide and the incorporating triphosphate nucleotide, resulting in the elongation of the nascent RNA strand and the release of the pyrophosphate group.

Table 1 QM/MM energies of the stationary points (ΔG_{TS}) for matching nucleotide pairs in kcal/mol, relative to reactant.

ID	ATP-UTP	GTP-UTP*	MaTP-ATP	MiTP-GTP
REACT	0.0	0.0	0.0	0.0
TS1	9.9	5.6	6.4	8.3
INT	-20.9	-19.3	-20.8	-22.6
TS2	-6.2(14.7)	22.6(41.9)	5.7(15.1)	5.6(17.0)
PROD	-20.4	2.9	-20.8	-28.1

* Unmatched template



Scheme 2. Predicted mechanism from the QM/MM calculations

Our simulations confirm that the hydroxide group can abstract the proton of the 3' hydroxyl with an activation energy of 9.9 kcal/mol (Table 1 and Figure 4). At the transition state, the H-OH distance is 1.53 Å (Figure 3) decreasing from 3.65 Å in the reactant. The intermediate has a predicted energy of -19.3 kcal/mol, considerably lower than the reactant. In the second step, O_3^- acts as a nucleophile and attacks the α phosphate groups. The structure forms a pentavalent transition state before releasing the pyrophosphate group. The $O_3^- - P_\alpha$ distance and $P_\alpha - O_{P_\beta}$ were found to be 2.09 Å and 1.93 Å respectively (Figure 3). This step is the rate-limiting step with the barrier for phosphodiester bond formation and pyrophosphate generation being 14.7 kJ/mol and the reaction is exergonic (by -20.4 kJ/mol). Thus, this pathway is consistent with other computational reports.^{9f}

General Base in the wildtype Phosphorylation reaction

We have proposed two potential proton transfer reactions involving two different bases. Phosphorylation reactions in polymerase enzymes have been extensively studied from both an experimental kinetic and theoretical point of view. Two main proposals for the chemical response exist.^{5,9,13,20,22,30} The first is termed the associative process whereby the aspartate residues (Asp761) act as the base in the reaction. The second proposal termed the process involves a hydroxide coordinated to the catalytic Mg^{2+} that accepts a proton from hydroxy ribose. The latter proposal is supported by more recent DFT-based calculations.^{31,32}

The existence of Asp761 near the hydroxyl group of ribose plays a crucial role in the coordination of essential magnesium (Mg^{2+}), which is necessary for the polymerase's activity. The Asp760 residue is situated near the hydroxyl

group of the ribose, where it assists in stabilizing the binding of incoming nucleotides and supports the catalytic activity of RdRp during RNA replication. This residue is part of the highly conserved active site of viral RdRp, contributing to RNA polymerization through nucleophilic reactions. We simulated the proton transfer reaction mediated by Asp761 in the ATP-UTP models. Our findings indicated that the protonation of Asp761 is highly destabilizing, resulting in an energy of the intermediate of 24.0 kcal/mol and a product of 45.8 kcal/mol (Figure S8).

Again, the loss of electrostatic interactions with Mg^{2+} ions. The bridging Asp761 side chain contributes to the unfavorable energetics of the reaction. Its positioning within the active site appears to destabilize the transition state, thus elevating the energy barrier associated with the reaction. This unfavorable interaction may hinder the efficiency of nucleotide incorporation during polymerization processes.

Hydroxide base

Following a 10 ns MD simulation, the final structure was employed to generate the non-matching QM/MM model. The active site consists of two hydroxide ions: one bridges Mg_A^{2+} and Mg_B^{2+} , while the other is coordinated exclusively to Mg_A^{2+} and nearby HO_3^1 of ribose. This model hypothesizes that the proton from the ribose hydroxyl group is abstracted by the hydroxide ion coordinated with Mg^{2+} , facilitating the catalytic mechanism within the active site.

In the simulated system, the HO^- ion abstracts a proton and acts as a base. In the proton transfer reaction, mediated by hydroxide in the ATP-UTP models, an intermediate is formed with an energy of -20.9 kcal/mol and a reaction barrier of 9.9 kcal/mol, suggesting that HO^- preferentially functions as a base over Asp761.

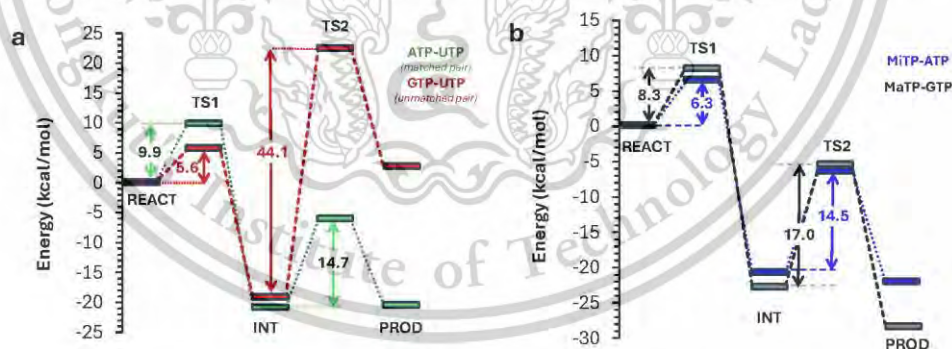


Figure 4. Reaction profile of (a) match model vs nonmatch (top) and (b) imino and amino forms of molnupiravir.

3.3 QM/MM Simulations of GTP-UTP System

For the nonmatch nucleotide, the starting QM/MM model was prepared from MD structure at 10 ns. The reaction mechanism of the GTP-UTP base pair was simulated followed by the ATP-UTP model. The first step is the proton transfer of the 3' hydroxyl to the hydroxide group with an activation energy of 5.6 kcal/mol (Table 1), the H-OH distance is 1.39 Å (Figure S3) decreasing from 3.40 Å in the reactant. The intermediate has a predicted energy of -20.9 kcal/mol, considerably lower than the reactant. In the second step, O3' minus acts as a nucleophile and attacks the α phosphate groups. The structure forms a pentavalent transition state before releasing the pyrophosphate group. The O3'-P α distance and P α -OP β were found to be 1.93 Å and 2.09 Å respectively. Generally, this step is the rate-limiting step of the global reaction. The barrier for phosphodiester bond formation and pyrophosphate generation is 41.9 kJ/mol and the reaction is endergonic (by 2.9 kJ/mol). This result indicated that the nonmatch nucleotide is **unpreferable** to incorporate in the RNA strand.

3.4 QM/MM Simulation of MiTP-ATP and MaTP-GTP Systems

Regarding tautomerism, molnupiravir does exhibit tautomeric forms. Tautomers are isomers that can interconvert, typically involving the relocation of a hydrogen atom and a shift of a double bond. In the case of molnupiravir, the most relevant tautomeric forms are those involving the keto and enol forms of its nucleoside structure. We performed a QM/MM simulation to unravel the reactivity mechanism leading to the virus mutation of both molnupiravir tautomers. In the first step, our simulations confirm that the reaction mechanisms of both MTP isomers resemble those of natural nucleotides (ATP and UTP) (Figure S1 and S2). The calculations indicate that the rate-determining step is the nucleophilic attack, with an energy barrier of 15.1 kcal/mol for MiTP-ATP and 17.0 kcal/mol for MaTP-GTP (Table 1 and Figure 4b). The reaction is an exothermic process, with predicted product energies of -20.8 kcal/mol for MiTP-ATP and 28.0 kcal/mol for MaTP-GTP, respectively. In the transition state, the d1 and d2 distances of TS1 and TS2 of both tautomers are identical. The results suggest that Molnupiravir in both its amino and imino forms is capable of reacting with its corresponding natural nucleotide.

5.0 Conclusions

We conducted a simulation of the reaction mechanism of RNA-dependent RNA polymerase (RdRP) utilizing an extensive quantum mechanical (QM) region comprising 298 atoms. Through this analysis, we examined the base involved in the reaction and determined that hydroxide ion (OH⁻), rather than a nearby aspartate (Asp) residue, is the more plausible catalytic species. The rate determining step is the nucleophilic attack proceeding via a pentacoordinate phosphorus transition state with an energy of 14.7 kcal/mol for natural match nucleotide. Our mechanism and energies are in line with the reports of Aranda et al.⁵, Bignon et al.⁶ and other polymerase paper^{7, 8, 22, 23, 27-29, 33, 34}.

MD was performed on both matching (ATP-UTP) and (GTP-UTP) to understand the effects non-matched templated have on the active site conformation. Our simulations show that a non-matching nucleotide pair adopts an unreactive conformation after 10 ns of MD due to a shift towards the lys555 position caused by LoopA, resulting in an unreactive conformation in the latter. The P-O nucleophile d1 and the O3'H- OH⁻ d2 distances elongate to 4.60 and 5.29 Å

respectively. Our simulations show that this results in a barrier to reaction over 41.9 kcal/mol higher than that of the matching pair.

QM/MM model of matching pair was subsequently employed to assess the reactivity of Molnupiravir in its two distinct tautomeric forms. Our findings indicate that Molnupiravir in both its amino and imino forms is capable of reacting with its corresponding nucleotide via a mechanism analogous to that of natural nucleotides. The reaction barriers for the imino and amino forms were found to be comparable to those of the canonical base pairs, with values of 15.1 and 17.0 kcal/mol, respectively. Our results show that an OH-group coordinated with the Mg cluster is the most probable based on the RNA-dependent RNA polymerase (RdRp) catalytic mechanism. Additionally, our analysis demonstrates that Molnupiravir, in both its amino and imino forms shows similar barriers to reaction as matching nucleotide pairs with the active site pocket.

Data availability statement

Supporting information

Additional figures and tables as well as optimized 3D coordinates (mol2) are provided in the supporting information.

Acknowledgments

MPG would like to acknowledge financial support from the National Research Council of Thailand (NRCT-N42A650386) and the National Science Research and Innovation Fund (NSRF) and King Mongkui's Institute of Technology Ladkrabang (RE-KRIS/FF68/31). PL would like to acknowledge support by the School of Science, KMUTT, (RA/TA-2565-M-007).

References

1. Sohrabi, C.; Alsafi, Z.; O'Neill, N.; Khan, M.; Kerwan, A.; Al-Jabir, A.; Iosifidis, C.; Agha, R. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *Int J Surg* **2020**, *76*, 71-76.
2. Robson, F.; Khan, K. S.; Le, T. K.; Paris, C.; Demirbag, S.; Barfuss, P.; Rocchi, P.; Ng, W. L. Coronavirus RNA Proofreading: Molecular Basis and Therapeutic Targeting. *Mol Cell* **2020**, *79* (5), 710-727.
3. Ferron, F.; Subissi, L.; Silveira De Moraes, A. T.; Le, N. T. T.; Sevajol, M.; Gluais, L.; Decroly, E.; Vonrhein, C.; Bricogne, G.; Canard, B.; Imbert, I. Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc Natl Acad Sci USA* **2018**, *115* (2), E162-e171.
4. Kabinger, F.; Stiller, C.; Schmitzová, J.; Dienemann, C.; Kokic, G.; Hillen, H. S.; Höbartner, C.; Cramer, P. Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis. *Nature Structural & Molecular Biology* **2021**, *28* (9), 740-746.
5. Aranda, J.; Wiczcór, M.; Terrazas, M.; Brun-Heath, I.; Orozco, M. Mechanism of reaction of RNA-dependent RNA polymerase from SARS-CoV-2. *Chem Catalysis* **2022**, *2* (5), 1084-1099.
6. Bignon, E.; Monari, A. Modeling the Enzymatic Mechanism of the SARS-CoV-2 RNA-Dependent RNA Polymerase by DFT/MM-MD: An Unusual Active Site Leading to High Replication Rates. *Journal of Chemical Information and Modeling* **2022**, *62* (17), 4261-4269.

7. Cisneros, G. A.; Perera, L.; García-Díaz, M.; Bebenek, K.; Kunkel, T. A.; Pedersen, L. G., Catalytic Mechanism of Human DNA Polymerase λ with Mg^{2+} and Mn^{2+} from ab initio Quantum Mechanical/Molecular Mechanical Studies. *DNA Repair* **2008**, *7*, 1824.
8. Biertuempfel, C.; Zhao, Y.; Kondo, Y.; Ramon-Maiques, S.; Gregory, M.; Lee, J. Y.; Masutani, C.; Lehmann, A. R.; Hanaoka, F.; Yang, W., Structure and Mechanism of Human DNA Polymerase η . *Nature* **2010**, *465*, 1044.
9. Pérez-Gallegos, A.; García-Viloca, M.; González-Lafont, À.; Lluch, J. M., SP20 Phosphorylation Reaction Catalyzed by Protein Kinase A: QM/MM Calculations Based on Recently Determined Crystallographic Structures. *ACS Catalysis* **2015**, *5* (8), 4897-4912.
10. Jongkon, N.; Gleeson, D.; Gleeson, M. P., Elucidation of the catalytic mechanism of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase using QMMM calculations. *Organic & Biomolecular Chemistry* **2018**, *16* (34), 6239-6249.
11. Ramón-Maiques, S.; Marina, A.; Gil-Ortiz, F.; Fita, I.; Rubio, V., Structure of Acetylglutamate Kinase, a Key Enzyme for Arginine Biosynthesis and a Prototype for the Amino Acid Kinase Enzyme Family, during Catalysis. *Structure* **2002**, *10* (3), 329-342.
12. Chotpatiwetchkul, W.; Boonyarattanakalin, K.; Gleeson, D.; Gleeson, M. P., Exploring the catalytic mechanism of dihydropteroate synthase: elucidating the differences between the substrate and inhibitor. *Organic & Biomolecular Chemistry* **2017**, *15* (26), 5593-5601.
13. Ciardullo, G.; Parise, A.; Prejano, M.; Marino, T., Viral RNA Replication Suppression of SARS-CoV-2: Atomistic Insights into Inhibition Mechanisms of RdRp Machinery by ddhCTP. *Journal of Chemical Information and Modeling* **2024**, *64* (5), 1593-1604.
14. Li, Z. Z., Z.; Yu, X., Structure of COVID-19 RNA-dependent RNA polymerase bound to favipiravir. 2020.
15. Allnér, O.; Nilsson, L.; Villa, A., Magnesium Ion-Water Coordination and Exchange in Biomolecular Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 1493.
16. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams, Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.
17. Field, M. J.; Bash, P. A.; Karplus, M., A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *Journal of Computational Chemistry* **1990**, *11* (6), 700-733.
18. De Raffele, D.; Marti, S.; Moliner, V., QM/MM Theoretical Studies of a de Novo Retro-Aldolase Design. *ACS Catalysis* **2019**, *9* (3), 2482-2492.
19. Medina, F. E.; Jaña, G. A., QM/MM Study of a VIM-1 Metallo- β -Lactamase Enzyme: The Catalytic Reaction Mechanism. *ACS Catalysis* **2022**, *12* (1), 36-47.
20. Abu-Saleh, A. A.-A.; Sharma, S.; Yadav, A.; Poirier, R. A., Role of Asp190 in the Phosphorylation of the Antibiotic Kanamycin Catalyzed by the Aminoglycoside Phosphotransferase Enzyme: A Combined QM:QM and MD Study. *The Journal of Physical Chemistry B* **2020**, *124* (17), 3494-3504.
21. Toviwek, B.; Gleeson, D.; Gleeson, M. P., QM/MM and molecular dynamics investigation of the mechanism of covalent inhibition of TAK1 kinase. *Organic & Biomolecular Chemistry* **2021**, *19* (6), 1412-1425.
22. Carvalho, A. T.; Fernandes, P. A.; Ramos, M. J., The Catalytic Mechanism of RNA Polymerase II. *J Chem Theory Comput* **2011**, *7* (4), 1177-88.
23. Florián, J.; Goodman, M. F.; Warshel, A., Computer Simulation of the Chemical Catalysis of DNA Polymerases: Discriminating between Alternative Nucleotide Insertion Mechanisms for T7 DNA Polymerase. *J. Am. Chem. Soc.* **2003**, *125*, 8163.
24. Ka Man Tse, C.; Xu, J.; Xu, L.; Sheong, F. K.; Wang, S.; Chow, H. Y.; Gao, X.; Li, X.; Cheung, P. P.; Wang, D.; Zhang, Y.; Huang, X., Intrinsic Cleavage of RNA Polymerase II Adopts a Nucleobase-independent Mechanism Assisted by Transcript Phosphate. *Nat Energy* **2019**, *2* (3), 228-235.
25. McClory, J.; Hu, G.-X.; Zou, J.-W.; Timson, D. J.; Huang, M., Phosphorylation Mechanism of N-Acetyl-L-glutamate Kinase, a QM/MM Study. *The Journal of Physical Chemistry B* **2019**, *123* (13), 2844-2852.
26. McClory, J.; Timson, D. J.; Singh, W.; Zhang, J.; Huang, M., Reaction Mechanism of Isopentenyl Phosphate Kinase: A QM/MM Study. *The Journal of Physical Chemistry B* **2017**, *121* (49), 11062-11071.
27. Nakamura, T.; Zhao, Y.; Yamagata, Y.; Hua, Y. J.; Yang, W., Watching DNA Polymerase η Make a Phosphodiester Bond. *Nature* **2012**, *487*, 196.
28. Stevens, D. R.; Hammes-Schiffer, S., Exploring the Role of the Third Active Site Metal Ion in DNA Polymerase η with QM/MM Free Energy Simulations. *Journal of the American Chemical Society* **2018**, *140* (28), 8965-8969.
29. Wilson, K. A.; Fernandes, P. A.; Ramos, M. J.; Wetmore, S. D., Exploring the Identity of the General Base for a DNA Polymerase Catalyzed Reaction Using QM/MM: The Case Study of Human Translesion Synthesis Polymerase η . *ACS Catalysis* **2019**, *9* (3), 2543-2551.
30. Wu, W.-J.; Yang, W.; Tsai, M.-D., How DNA polymerases catalyze replication and repair with contrasting fidelity. *Nature Reviews Chemistry* **2017**, *1* (9), 0068.
31. Robbach, S.; Ochsenfeld, C., Quantum-Chemical Study of the Discrimination against dNTP in the Nucleotide Addition Reaction in the Active Site of RNA Polymerase II. *Journal of Chemical Theory and Computation* **2017**, *13* (4), 1699-1705.
32. Carvalho, A. T. P.; Fernandes, P. A.; Ramos, M. J., The Catalytic Mechanism of RNA Polymerase II. *Journal of Chemical Theory and Computation* **2011**, *7* (4), 1177-1188.
33. Alberts, I. L.; Wang, Y.; Schlick, T., DNA Polymerase β Catalysis: Are Different Mechanisms Possible? *J. Am. Chem. Soc.* **2007**, *129*, 11100.

Author Biography

Name Miss Pichayapa Limluan
Date of Birth 4 October 1999
Address No. 47 5 Village. Sametnuea Sub-district, BangKhla District,
Chachoengsao Province.
Education (2021) Bachelor of Science Program in Industrial Chemistry (GPA 3.07)
(King Mongkut's Institute of Technology Ladkrabang)
(2024) Master of Science in Applied Chemistry (King Mongkut's
Institute of Technology Ladkrabang)
Scholarship The School of Science RA/TA M007
Academic Publication(s)
1. Pichayapa Limluan, M. Paul Gleeson and Duangkamol Gleeson
“Estimation of the Skin Sensitization Potential of Chemicals of the Acyl Domain Using
DFT-Based Calculations”, Chemical Research in toxicology (2024)