COMPUTATIONAL MODELING OF THE HYDROLYSIS OF 2′-DEOXYRIBONUCLEIC ACIDS

JENNIFER L. PRZYBYLSKI
B. Sc. University of Lethbridge, 2007

A Thesis Submitted to the
Department of Chemistry and Biochemistry
of the University of Lethbridge
in Partial Fulfillment of the
Requirements for the Degree

MASTER OF SCIENCE
in Chemistry

FACULTY OF ARTS & SCIENCE

LETHBRIDGE, ALBERTA

July 2009

Abstract

The mechanism for the hydrolysis of 2′-deoxyribonucleosides is examined using computational chemistry techniques. Initially, a model capable of accurately predicting the mechanism and activation barrier for the uncatalyzed hydrolysis of 2′-deoxyuridine is designed. It is found that the smallest model includes both explicit and implicit solvation during the optimization step. Next, this hybrid solvation model is applied to four natural nucleosides, namely 2′-deoxyadenosine, 2′-deoxycytidine, 2′-deoxyguanosine and thymidine. The hybrid model correctly predicts the trend in activation Gibbs energies for the pyrimidines and purines, separately. Finally, the concepts developed during the generation of the uncatalyzed hydrolysis model are applied to the mechanism of action of a glycosylase enzyme, namely human uracil DNA glycosylase. A hybrid ONIOM approach is utilized to study the experimentally proposed two-step mechanism. Results regarding the protonation state of His148 are inconclusive, and future directions are proposed.

iv

Acknowledgements

I would first like to thank Sean, who could still find me a reason to smile when there didn't seem to be anything worth smiling about. Without your constant support I don't think I would have made it to the end.

I would like to thank Tim Horton's, without which I wouldn't have survived.

Thanks to my mother for convincing me that 'yes, it *is* worth it'.

I would like to thank my supervisor, Dr. Stacey Wetmore, for believing in me and my work, especially when it seemed like I would never get funding.

A big thank you to NSERC for funding this project.

Thank you to my friends and colleagues for convincing me to go for drinks every once and a while before my life went into atrophy.

I would like to thank my students, who reminded me why I went into chemistry in the first place.

Finally, this quote has pulled me through many tough times:

> "If someone prays for patience, you think God gives them
> patience? Or does He give them the opportunity to be patient.
> If he prayed for courage, does God give him courage? Or does
> He give them the opportunities to be courageous."
> ~God, *Evan Almighty* 2007.

Table of Contents

Chapter Two: Hydrolysis of 2′-Deoxyuridine

Chapter Three: Hydrolysis of the Natural 2′-Deoxyribonucleosides

Chapter Four: Mechanism of Action of Uracil DNA Glycosylase

List of Figures

List of Tables

List of Abbreviations

A	Adenine
AID	Activation-Induced Deaminase (EC#: 3.5.4.5)
AMBER	Assisted Model Building and Energy Refinement
AP-site	Apurinic/apyrimidinic site
B3LYP	3 Parameter Becke88 Exchange Functional with the Lee-Yang-Parr Correlation Functional
BER	Base Excision Repair
C	Cytosine
C2′H	C2′ Proton Abstraction Product
CHARMM	Chemistry at HARvard Macromolecular Mechanics
ΔE	Relative Energy
$\Delta G^{\ddagger}$	Gibbs Free Energy of Activation
dA	2′-Deoxyadenosine
dC	2′-Deoxycytidine
DFT	Density Functional Theory
dG	2′-Deoxyguanosine
DNA	Deoxyribonucleic acid
dR	2′-Deoxyribose
dT	Thymidine
dU	2′-Deoxyuridine
dUTP	2′-Deoxyuridine Triphosphate
E. coli	*Escherichia coli*
eUDG	*E. coli* Uracil DNA Glycosylase (EC#: 3.2.2.3)
G	Guanine
hOGG1	Human 8-Oxoguanine DNA Glycosylase (EC#: 3.2.2.-4.2.99.18)
hUNG2	Nuclear Human Uracil DNA N-Glycosylase (EC#: 3.2.2.3)
IEF-PCM	Integral Exchange Formalism Polarizable Continuum Model
KIE	Kinetic Isotope Effect

MC	Monte Carlo
MD	Molecular Dynamics
MM	Molecular Mechanics
NER	Nucleotide Excision Repair
NMR	Nuclear Magnetic Resonance
ONIOM	Our Own N-layered Integrated Molecular Orbital + Molecular Mechanics
PCM	Polarizable Continuum Model
PDB	Protein Data Bank
PES	Potential Energy Surface
ψU	Pseudouridine
QM	Quantum Mechanics
SMUG1	Single-Stranded Mismatch Uracil DNA Glycosylase One (EC#: 3.2.2.3)
S_N1	Unimolecular Nucleophilic Substitution
S_N2	Bimolecular Nucleophilic Substitution
T	Thymine
TCG	Temperature Correction to Gibbs Energy
TS	Transition State
TTP	Thymidine Triphosphate
U	Uracil
UDG	Uracil DNA Glycosylase (specifically referring to hUNG2)
UTP	Uridine Triphosphate
UV-B	B Range Ultra-Violet Light
ZPVE	Zero-Point Vibrational Energy

Chapter 1: Thesis Introduction

1.1 General Overview

Water, which makes up a majority of the body, is a small molecule essential to the life of every known organism. In contrast to its simplistic structure, the chemistry of water is very complicated and research involving water is broad and crosses many disciplines. The first water related reaction introduced to most chemistry and biology students is hydrolysis. The textbook definition of hydrolysis involves the cleavage of a bond by a water molecule where one product receives a hydroxyl (OH^-) group and the other product receives a proton (H^+) (Figure 1.1).[1] However, this definition leaves to the imagination the number and the order of the steps involved in these reactions. Figure 1.1 illustrates that hydrolysis requires (at a minimum) two bonds to break and two new bonds to form, and thus there are many possible pathways by which the reaction may proceed. This makes hydrolysis mechanisms difficult to predict.

$$A\text{–}B + H_2O \longrightarrow A\text{–}OH + H\text{–}B$$
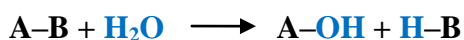
Figure 1.1 *The general definition of a hydrolysis reaction.*

Similar to water being a biologically important small molecule, a biologically important macromolecule is deoxyribonucleic acid (DNA). DNA stores the genetic material of the cell, where all information required for life is assembled in a sequence generated by four unique DNA building blocks, namely 2′-deoxyadenosine (dA), 2′-deoxycytidine (dC), 2′-deoxyguanosine (dG) and thymidine (dT) (Figure 1.2). The double-helical structure of DNA is formed by the pairing of two anti-parallel strands of nucleotides. These strands bind to each other in a manner that places a purine (dA and

dG) opposite a pyrimidine (dT and dC), such that dA is paired with dT and dG is opposite

dC. Each nucleotide is composed of a phosphate group, a 2′-deoxyribose (dR) sugar

moiety and a nucleobase, where the nucleobase is connected to the sugar moiety via an

N-glycosidic bond at N9 (purines) or N1 (pyrimidines) (Figure 1.2). In DNA, the

glycosidic bond of the nucleotides is very stable, having a decomposition half life of over

70 years,[3] and this stability contributes to the integrity of the genetic code.

Due to the prevalence of water in the cell, DNA hydrolysis reactions continuously

occur in biological systems. DNA hydrolysis can be spontaneous or require a means of

Nucleotide: R = HPO_3^- Nucleoside: R = H
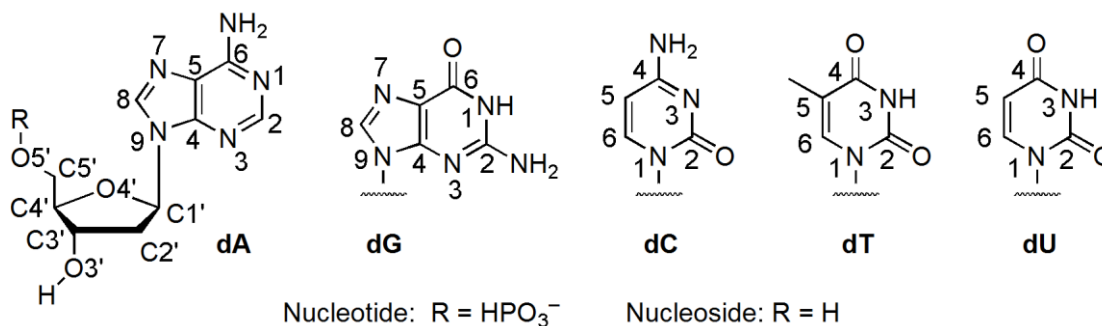
Figure 1.2   *Structures and atomic numbering of the five natural nucleic acids, namely 2'-deoxyadenosine (**dA**), 2'-deoxyguanosine (**dG**), 2'-deoxycytidine (**dC**), thymidine (**dT**) and 2'-deoxyuridine (**dU**). In nucleotide form, the 5'- end of the sugar contains a phosphate group, while in nucleoside form the 5'- end is a hydroxyl group.*

catalysis (e.g., enzymes), where some of these reactions are beneficial and others are

detrimental. An example of a frequently occurring detrimental hydrolysis reaction

involves the nucleobase cytosine (C), which can undergo hydrolysis at the C4–N4 amide

bond (Figure 1.2) to yield uracil (U) (see Section 1.2.1).[2] The conversion of C to U

causes premutagenic uracil–guanine (G) pairs, since adenine (A) will be placed opposite

uracil when the DNA helix is replicated due to the structural similarity between U and

thymine (T). An example of a situation where it is essential for water to be highly

reactive is nucleic acid deglycosylation. If a nucleobase becomes modified, such as

deamination of cytosine, there is generally a decrease in the stability of the N-glycosidic bond, which makes the damaged site susceptible to attack by nearby water molecules. In this case, water may spontaneously remove modified nucleobases, which prevents possible miscoding during replication or transcription, and thus maintains genome integrity.[4] In addition, the hydrolytic reactivity of water is often amplified by enzymes. For example, the enzyme uracil DNA glycosylase (see Section 1.2.2) enhances the effectiveness of 2′-deoxyuridine (dU) hydrolysis by a factor of 10^{12}, reducing the barrier to ~50 kJ mol^{-1}.[5]

The above examples illustrate the importance of the hydrolysis reaction with respect to nucleic acids and demonstrate the variety of roles that a single water molecule may have in nature. It is thus essential to understand not only enzymatic hydrolysis reactions, but also uncatalyzed reactions to determine what role water plays and how the hydrolysis mechanism can be manipulated. It is the underlying objective of this thesis to understand how nucleic acids are deglycosylated, in both solution and in the presence of an enzyme. The similarities and differences between these mechanisms can tell us a great deal about the reaction and the function of various constituents. For example, do the enzyme catalyzed and the uncatalyzed reactions use the same type of mechanism, and how does the enzyme stabilize the otherwise high-energy process? To this end, computational chemistry techniques will be utilized to study the hydrolytic deglycosylation of dU due to the high frequency of C deamination to U, and the abundance of experimental studies on the enzymatic repair of U lesions. The following sections will discuss the general background of dU hydrolysis, including enzymatic and non-enzymatic mechanisms.

1.2 Background

1.2.1 Uracil in DNA

Damage to DNA is not limited to attack by water molecules, and can occur by a variety of mechanisms. Some moieties directly cleave bonds in the helix and therefore, cause strand scission (for example, radical addition of low energy electrons).[6] Other types of damage introduce new bonds between nearby nucleotides, which form links that prevent replication and translational processes (e.g., thymine dimers caused by UV-B sunlight).[7] Still others modify the four naturally occurring DNA nucleobases (A, C, G and T), where these changes can involve alkylation,[8] oxidation,[9] and deamination,[2,10–11] to name a few examples. Damage-causing modifications to the nucleobases vary in their long term effects, and can be toxic and/or carcinogenic.

One of the most frequent damaging nucleobase modifications is the deamination/oxidation of cytosine to uracil (Figure 1.3), which can occur due to naturally produced intermediates such as bisulfite,[10] nitrous anhydride (N_2O_3)[11] or water.[2] As mentioned previously, this alteration leads to C:G to T:A point conversions after replication. Cytosine can also be enzymatically converted to uracil by activation-induced deaminase (AID), which can be utilized by the cell to break down nucleic acid strands.[12]

In addition to deamination of C, dU can be directly incorporated into the helix via misincorporation by DNA polymerase, due to the structural similarities to dT (Figure 1.2). The frequency of dU addition is dependent on the ratio of thymidine triphospate (TTP) to 2′-deoxyuridine triphosphate (dUTP) in the cell,[5] where dUTP is formed
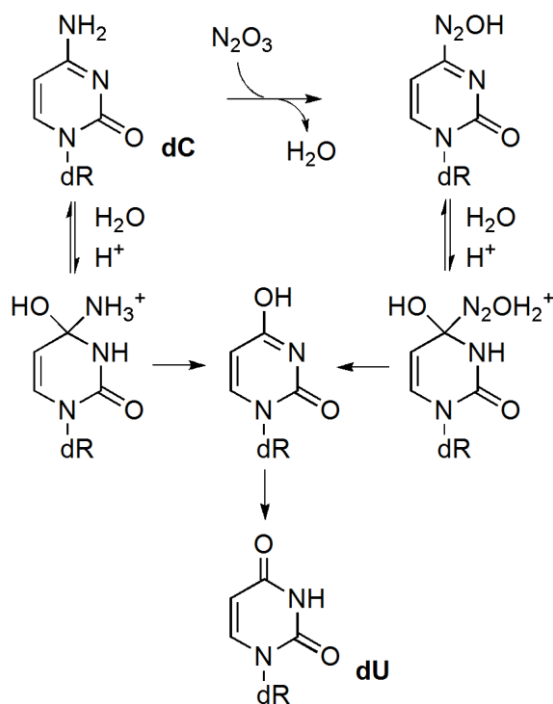
Figure 1.3 *The deamination of **dC** to **dU** via hydrolytic and nitrosative processes (adapted from reference 5).*

naturally in the pyrimidine recycling process that generates uridine triphosphate (UTP) for the production of ribonucleic acids (RNA). Combined, deamination of C and misincorporation of dU into DNA can lead to approximately 200 DNA lesions per cell per day.[13]

The deamination of C is so prevalent that one theory for the evolution of dT is the introduction of a simple way to detect C deamination.[14] The structural difference between U and T implies that any U in DNA is due to an error and should be removed, as opposed to the natural occurrence of U in RNA (instead of T). RNA is much shorter lived than DNA,[14] where this disparity has been attributed to the single-stranded nature of RNA. The lack of interactions with the Watson-Crick face of C in RNA makes C more susceptible to deamination and, additionally, it is also difficult to determine whether the appearance of a U residue is a lesion since the bases are unpaired. Thus, the RNA strand

is recycled to prevent transfer of mutations along the translation pathway, which may subsequently lead to protein malfunction.[14]

1.2.2 Repair of dU Lesions by UDG

The cell has evolved many methods to reverse the damage done to DNA. These methods include direct reversal (for example, transfer of alkyl groups from the damaged base to an enzyme), base excision repair (BER, removal of the base followed by repair of the apurinic/apyrimidinic site (AP-site), see Figure 1.4) and nucleotide excision repair (NER, removal of the damaged nucleotide with no generation of an AP-site).

Uracil lesions are most commonly repaired by a family of enzymes called uracil DNA glycosylases that initiate the BER process. In humans, hUNG2 (responsible for over 90% of dU removal) and SMUG1 are the most common enzymes for repairing nuclear dU damage.[2b,15] hUNG2 (hereafter referred to as UDG) has been extensively studied experimentally since its discovery. In fact, the first glycosylase discovered was a uracil DNA glycosylase (specifically *E. coli* UDG).[16] UDG is a 304 residue protein, with an active site that is highly conserved across prokaryotes and eukaryotes.[17] It is a mono-functional N-glycosylase, which implies that it will hydrolytically cleave the N-glycosidic bond of dU and leave an AP-site as a product for another enzyme to repair.[18] UDG has been identified to function on single- and double-stranded DNA, and can bind as a proofreading enzyme during replication to remove any dU residues misincorporated by the DNA polymerase.[19] Experimentally, it has been found that UDG can increase the hydrolysis rate of dU by $\sim 10^{12}$, and thus UDG is one of the most efficient glycosylases known.[20] UDG is very specific for dU, and will not cleave the RNA form (uridine), dC, dT or the electrostatically similar 5-fluoro-2′-deoxyuridine (Figure 1.5).[19,21] It is also
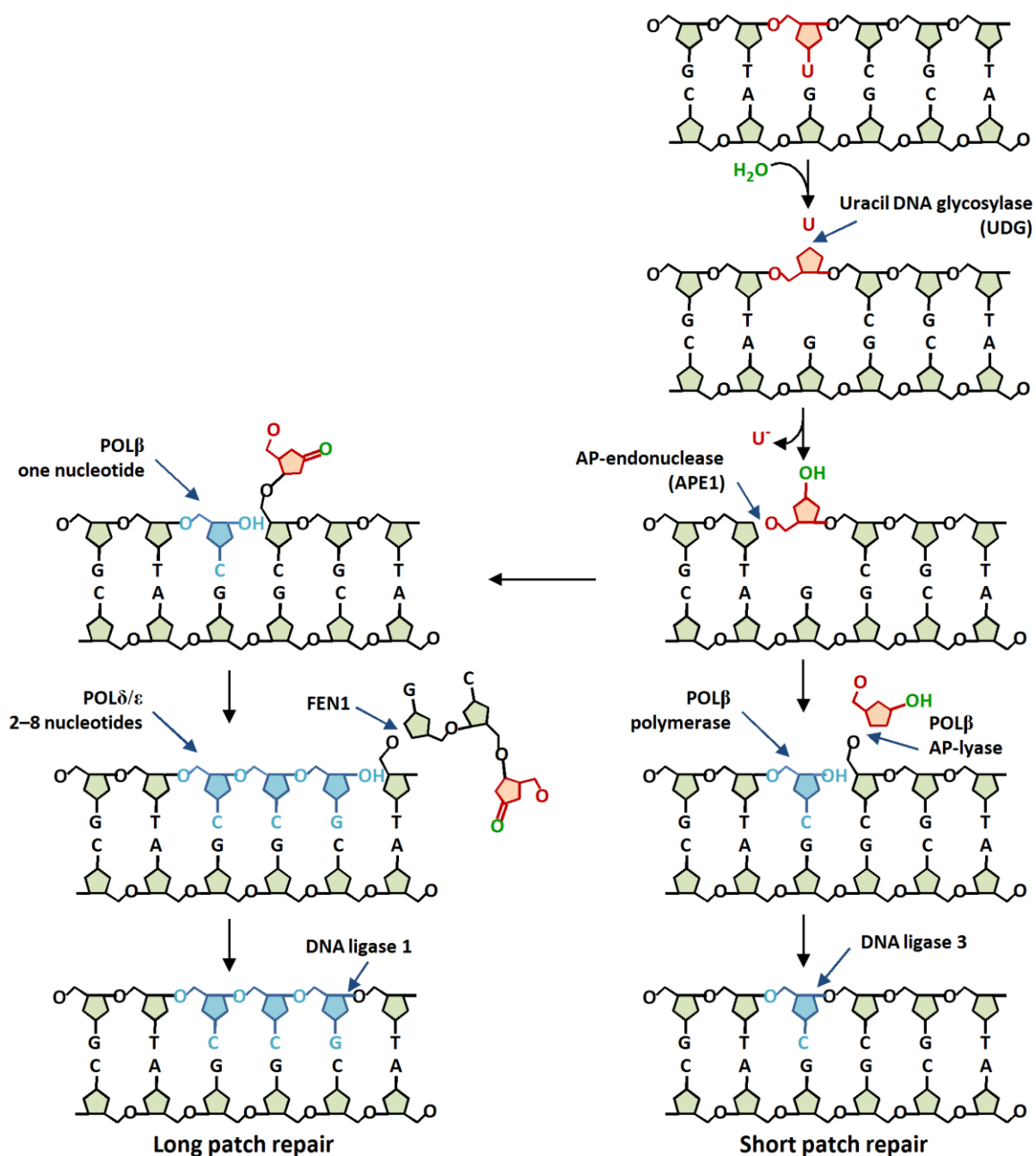
Figure 1.4 *The BER pathway utilized in humans for the removal of dU lesions. First, UDG cleaves the dU N-glycosidic bond leaving an AP-site. Next, APE1 cleaves the backbone at the 5' end of the lesion. In short-patch repair (right), the AP-site is removed by the AP-lyase function of POLβ, followed by insertion of the correct base. The nick in the backbone is then repaired by DNA ligase 3. In long-patch repair (left), POLβ inserts the replacement nucleotide, followed by extension of the chain for 2 – 8 residues by POLδ/ε. FEN1 cleaves the flap, and DNA ligase 1 closes the nick in the backbone (adapted from reference 2).*

inhibited by free U. These facts have led to UDG being one of the most widely studied

glycosylases to date and the results of this work will be discussed in detail in the
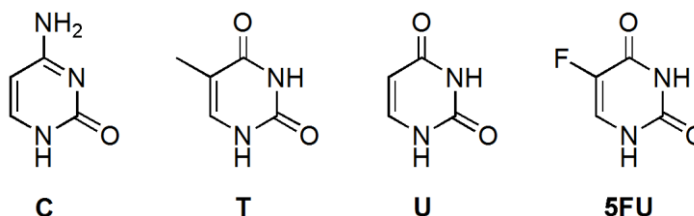
Figure 1.5 *Structural comparison of the pyrimidines, namely cytosine (**C**), thymine (**T**) and uracil (**U**), and the inhibitor 5-fluorouracil (**5FU**).*

following subsections.

1.2.2.1 UDG Proposed Mechanism

A large body of experimental work has been carried out on UDG (and eUDG) to

determine the mechanism utilized to remove uracil with such specificity and speed. The

methods implemented include X-ray crystal structure analysis,[17,22] mutational studies,[17,23]

heteronuclear NMR spectroscopy,[24] Raman spectroscopy,[25] kinetic isotope effect (KIE)

experiments,[26] and computational modeling.[27–30] The major findings in studies using each

method are summarized below.

1.2.2.1.1 Structural Analysis

A crystal structure for the unbound form of UDG (PDB code 1AKZ, 1.90 Å
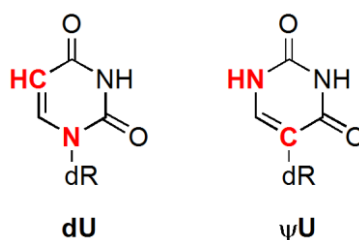
Figure 1.6 *The structure of 2'-deoxyuridine (**dU**) compared with pseudouridine (**ψU**) (atomic differences highlighted in red).*

resolution) shows a 21 Å wide groove lined with positively charged residues, which is large enough to cradle the DNA helix and bind to the negatively charged phosphate backbone.[17] This study also identified a leucine finger that protrudes into this positively charged groove of UDG. This protrusion is proposed to aid in pushing the uracil lesion out of the helix and into the enzyme active site leaving a leucine residue intercalated (i.e., inserted between the U + 1 and U − 1 nucleotides) in the DNA strand.

A Michaelis-complex analogue was first crystallized in 2000 by the Slupphaug and Krokan group (PDB code 1EMH) using the reactant mimic pseudouridine (ψU) (Figure 1.6).[22c] This study identified key active-site interactions that indicate UDG uses reactant destabilization as a catalytic method. Figure 1.7 summarizes the important active-site residues, each of which plays a unique role. Interactions with the hydrogen-
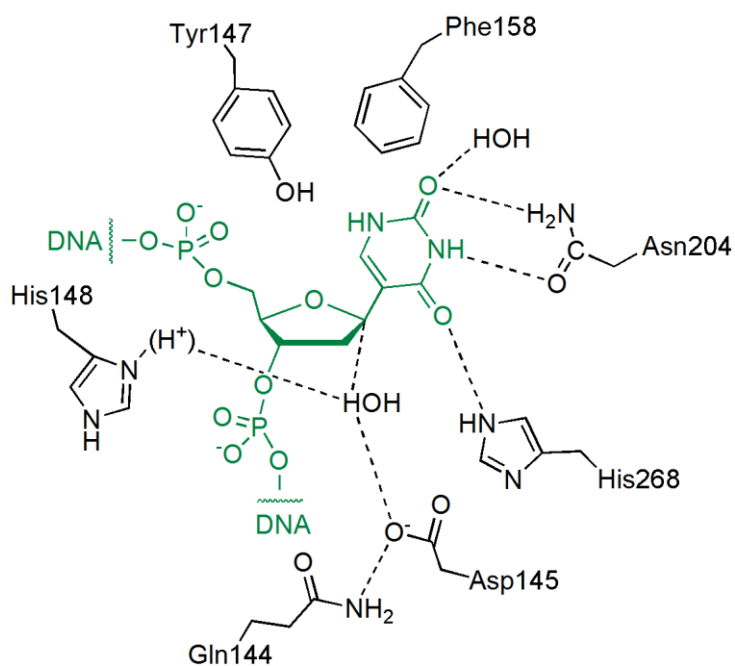
Figure 1.7　*Important active-site residues identified for UDG through crystal structure analysis. DNA is displayed in green and the enzyme and water molecules in black. Note that residue His148 may be cationic, where the second proton is in parentheses. (PDB: 1EMH)*

bonding sites of ψU by His268, Asn204[a] and a crystallographic water molecule pull the base into the active site, and have been hypothesized to aid in the flipping of uracil out of the helix. A phenylalanine residue (namely Phe158) is stacked above the uracil moiety and pushes on the π-system of pseudouracil, which leads to distortion at C1 (corresponding to N1 in dU). The distortion causes strain in the reactant complex since it breaks the aromaticity of the ring system. The non-planarity can also be explained through hyperconjugation as illustrated in Figure 1.8. A distortion at N1 of U is required to align the molecular orbitals between O4′ and O2 to allow electron transfer through the glycosidic bond. This crystal structure also pinpoints a nucleophilic water molecule, and proposes Asp145 as a catalytic residue that is positioned to strip a proton from the water molecule to activate it for nucleophilic attack.
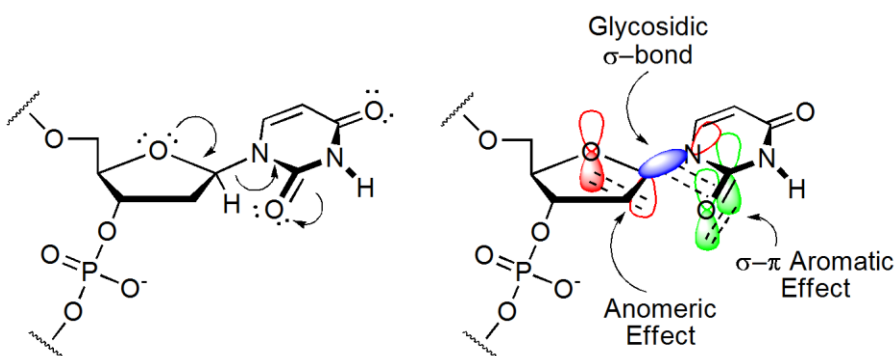
Figure 1.8 *Hyperconjugation effects that transfer electron density from O4′ of the sugar to O2 of uracil. The left diagram shows the electron motion described by this effect. The right diagram shows the molecular orbitals involved in the transfer of the electrons. The bend at N1 is required to orient the molecular orbitals correctly (adapted from reference 22d).*

A 1.90 Å crystal structure of UDG bound with a product complex (PDB code: 1SSP) suggests that UDG holds onto the AP-site until the damaged site is passed to the

---

[a] hUNG2 numbering is used throughout.

next enzyme in the repair process, AP-endonuclease (Figure 1.4). There is also evidence that uracil is expelled as an anion.[22a,b]

Inspired by KIE experiments that indicated an S_N1, dissociative mechanism (see Section 1.2.2.1.2), the Stivers group proposed a second, associative, transition state (TS) mimic (Figure 1.9) consisting of a free U anion and cationic 1-azadeoxyribose.[31] The crystal structure of the UDG:TS mimic complex (PDB code 1Q3F, 1.90 Å) indicates that the 1-azadeoxyribose contains an N1'-*exo* puckering.[22d] The authors interpreted this to indicate that the sugar distorts from the planar oxacarbenium cation intermediate to move towards the nucleophilic water, as opposed to the water migrating towards the cation.

1.2.2.1.2 Kinetic Isotope Effect Experiments

KIE experiments by Stivers and Werner found that the hydrolysis occurs in a stepwise (S_N1) fashion with a fully formed oxacarbenium cation intermediate.[26] This was the first example of a DNA glycosylase utilizing such a mechanism since other glycosylases were identified to use a bimolecular (S_N2) mechanism.[18b]

Associative Transition State

1-azadeoxyribose Mimic
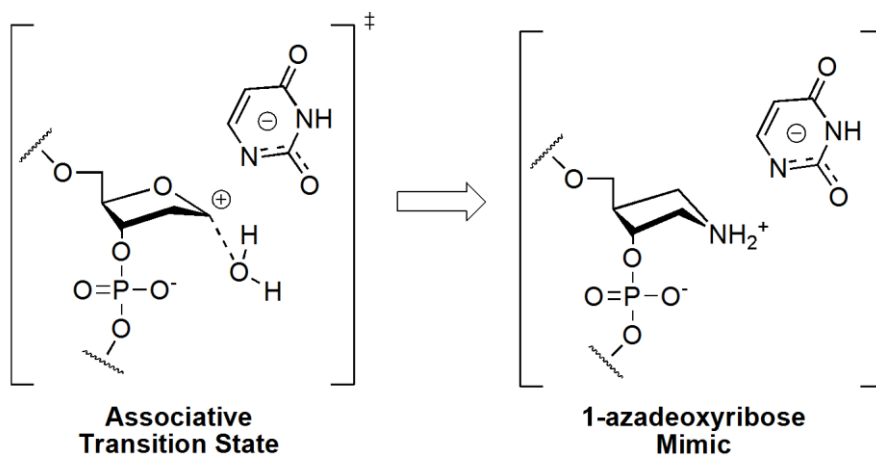
Figure 1.9 *Design of the 1-azadeoxyribose transition-state mimic to simulate the second, associative, transition state of the hydrolysis of dU by UDG (adapted from reference 22e).*

1.2.2.1.3 Raman Spectroscopy

Raman studies have shown that the uracil moiety exists in the product complex in an anionic form, where there is a negative charge delocalized over N1 and the O2 carbonyl.[27] This suggests that hydrogen bonds with active-site residues provide enough leaving group stabilization to negate the necessity of base protonation as seen for acid catalyzed hydrolysis[32] and supports propositions that neutral His268 is catalytic through a hydrogen bond to O2.[22c]

1.2.2.1.4 NMR Spectroscopy

There is debate in the literature as to the protonation state of the active-site residue His148. This residue is in a position to hydrogen bond to the U + 2 phosphate group (Figure 1.10), and hydrogen bonds to a crystallographic water molecule that bridges to the U − 1 phosphate. Furthermore, in the reactant-mimic crystal structure, His148 may hydrogen bond to the nucleophilic water molecule and thereby anchor it underneath the sugar moiety.[22c] Arguments have been made for the tautomer where $N\varepsilon_2$ is protonated, as well as the cationic form.[17,22,24,28a] A recent study by the Stivers group found NMR (^{1}H–^{15}N LR-HMQC) evidence supporting a neutral histidine in the unbound state;[24d] however, the authors indicate that the protonation state could change throughout the mechanism.
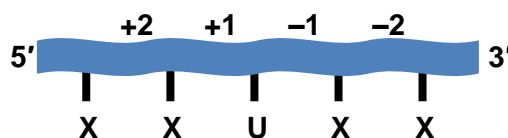
Figure 1.10 *The definition of the phosphate group numbering with respect to uracil (U) in DNA, where phosphates towards the 5'-end are positive and those towards the 3'-end are negative.*

1.2.2.1.5 Mutational Studies

A large number of mutations have been made to the UDG system, including

mutations to the enzyme, the uracil nucleotide, the base opposite the uracil nucleotide and

methylation of nearby phosphate groups.[17,23] These studies have found that the $U - 1$, $U +$

1 and $U + 2$ phosphate groups (Figure 1.10) play an integral role in stabilizing the

reaction intermediate. They also identified important residues, including the previously

mentioned His268, His148, Asp145, Asn204, and Phe158, and the additional residues of

Gln144, and Leu272. Gln144 is proposed to be important for orienting the catalytic

Asp145, and Leu272 intercalates into the DNA helix pushing dU into the enzyme active

site. It should be noted that while mutation of Asn204 affects the binding strength of the

enzyme-DNA complex, mutation of His268 leads to a large change in the specific

activity of UDG.

1.2.2.1.6 Computational Studies

In contrast to the abundance of experimental work on the subject, very little has

been accomplished computationally to elucidate the mechanism of UDG. The work that

has been done can be separated into two classes: large-model and small-model studies.

The large-model studies used low levels of theory to describe the entire UDG:DNA

complex, while the small-model studies truncated the system to allow for the use of high

levels of theory. Each approach provides different information, and therefore will be

discussed separately.

The Osman group carried out molecular dynamics (MD) simulations on the

UDG:DNA complex[27a] before the elucidation of the UDG:ψU crystal structure by Parikh

et al.[22c] Since only the product bound structure was available,[22a] many of the results from

this study no longer hold. Nevertheless, this work investigated some key concepts related to binding of the enzyme to the DNA strand and the possible overall excision mechanism. A later study by the same group discussed the catalytic Asp145 residue.[27b] Specifically, they calculated the pK_a of the residue to be 4.4 ± 0.1 in the reactant complex,[b] and thus determined that the aspartate residue would not act as a proton acceptor. Instead, they proposed that additional water molecules bridging Asp145 to the 3′-phosphate group of dU would act as the activator and generate the hydronium ion and the hydroxide required for hydrolysis (Figure 1.11). This bridged structure is not present in the reactant-mimic crystal structure of Parikh et al.,[22c] and thus this proposal has yet to be supported experimentally.
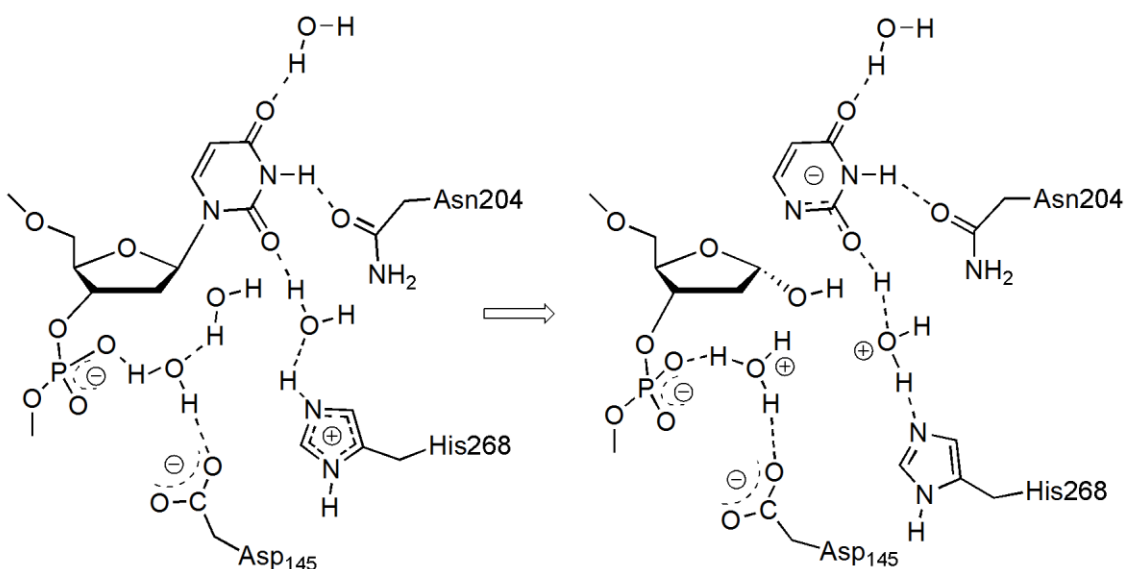
Figure 1.11 *General reaction for dU removal by UDG proposed by Osman et al. The −1 phosphate activates the nucleophilic water molecule through a bridging water molecule, and the uracil anion leaving group is stabilized by transfer of a proton from His268 through another water molecule.*

---

[b] Note that the reactant complex in this study was designed from the product complex crystal structure (PDB code 4SKN), which is considerably more open than the reactant complex.

In 2001, Dinner et al. carried out a quantum mechanics/molecular mechanics (QM/MM) study on the UDG:ψU structure.[28a] This study used the AM1 semi-empirical method to describe the active-site region, while the remainder of the DNA strand and enzyme were described using the CHARMM molecular mechanics force field. Solvent shielding was incorporated in the model using the Poisson-Boltzmann technique (see reference 28a, and references therein, for a description of this method). A two-dimensional adiabatic potential energy surface scan was conducted by changing the N1–C1′ and the O_w–C1′ distances and relaxing the remainder of the system. The rate-limiting step was found to be dissociation of the base from the sugar moiety, which has an internal energy barrier of ~60 kJ mol^{-1} relative to the reactant. They reported that the Asp145 residue does indeed activate the nucleophilic water molecule. The authors also determined the contribution of each residue to the barrier by setting the respective charges to zero, which is equivalent to removing the electrostatic contribution from the MM portion, and removing the residue completely from the QM portion, of the calculation. This method determined that the contribution of the U + 2 phosphate group to catalysis is larger than the catalytic effect of the total enzyme. To address this disparity with experiment, Dinner and Karplus subsequently studied the effect of solvation on the role of the phosphate groups in 2006.[28b] It was determined that the discrepancy was due to the use of double-stranded DNA in the computational work compared to the single-stranded DNA utilized in the experimental study.[23a] The authors also reported that the catalytic contribution of the phosphates is masked by hydrogen bonds to solvent molecules when single-stranded DNA is considered. Finally, it was noted that if His148

is cationic, its proximity to the cationic intermediate will be anti-catalytic, and thus somewhat balance the large contribution of the phosphate groups.

The Wetmore group has carried out a variety of small model studies on specific portions of the UDG mechanism (Figure 1.12). Experimentally, it has been suggested that active-site hydrogen bonds between dU and His268 or Asn204 not only have a selectivity function, but may also contribute catalytically (Section 1.2.2.1.5).[17] To investigate this idea, the Wetmore group examined the effects of hydrogen bonds with a variety of small molecules and truncated amino acids on the N1 acidity of the U nucleobase,[29] since an increase in the acidity will reduce the glycosidic bond cleavage barrier and indicates catalysis of the reaction. It was found that hydrogen bonds can greatly affect the N1 acidity of U, where this effect is slightly dampened by implicit solvent.[c] Interestingly, it was found that the change in acidity due to imidazole (the functional group of histidine) hydrogen bound to O2 of uracil in the presence of a weak dielectric field is nearly the same as the proposed catalytic contribution of His268 calculated using data from mutational kinetic studies.[24b]
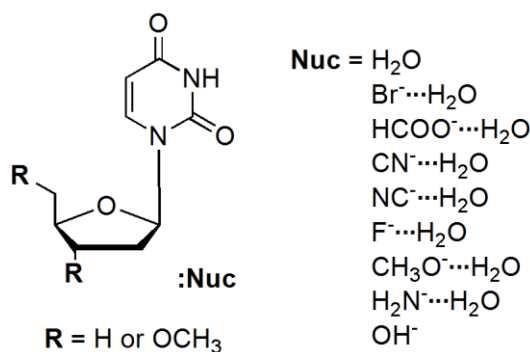
Figure 1.12 *Basic model used by the Wetmore group to study the hydrolysis of dU, including two different sugar truncations and nine different levels of water nucleophile (Nuc) activation. The nucleophiles are given in order of increasing strength.*

---

[c] Solvent effects were added post-optimization using (PCM) continuum methods with a dielectric of $\varepsilon = 4$.

1.2.2.1.7 Summary

Combined, these experimental and computational studies have determined that UDG utilizes a dissociative S_N1 mechanism to remove uracil from DNA, where short, strong hydrogen bonds between the base and active-site residues (His268 and Asn204) account for most of the catalytic activity. The uracil anion is stable enough to be expelled from the active site without (full) protonation. Most studies agree that Asp145 activates the nucleophilic water. This conclusion was drawn from the proximity of the functional group to the crystal structure water molecule, as well as, mutational studies that report reduced efficiency of UDG upon replacing Asp145 with a non-polar amino acid. However, there is no evidence indicating which step of the reaction Asp145 catalyzes, where observed reductions in catalytic efficiency may be due to the loss of a negative charge in the active site to stabilize the oxacarbenium cation intermediate rather than loss of nucleophile activation. Computational chemistry is the perfect tool for discovering which step of the reaction Asp145 is most likely catalyzing, as well as determining the protonation state of His148. This thesis will attempt to shed light on these questions.

1.2.3 Nucleoside Deglycosylation

Ideally, when studying an enzyme mechanism, the whole system should be included in the computational model; however, the Michaelis complex is simply too large to be treated with quantum mechanical techniques. Therefore, a balance must be found between accuracy and computational expense. In general, this is done by removing a large portion of the enzyme, and explicitly treating the important active-site interactions with a high level of theory.[33] Before this approach can be implemented many questions must be answered, such as: What are the 'important' active-site interactions? How much

of the enzyme is explicitly required? How much of the environment can be estimated using a continuum?

Any methodology that involves truncation or level definition requires the user to specify the relative importance of different regions of the system. Often important interactions are self-evident, such as obvious hydrogen bond contacts between the nucleophile or substrate and active-site amino acids. However, enzymes are so complicated that it is difficult to determine at the outset what groups must be treated explicitly. Therefore, it is a common practice to proceed from the most basic computational model (i.e., the substrate and nucleophile) and systematically increase the model until it resembles the enzymatic system.

This section will discuss previous literature that has examined glycosidic bond cleavage of 2′-deoxyribonucleic acids (including non-hydrolytic cleavage) using truncated models. However, these reduced systems often yield unrealistic geometries. In addition, the 'enzyme-catalyzed' models are truncated so far that they resemble non-enzymatic glycosidic bond cleavage reactions. Therefore, the next section discusses the experimental work which has been carried out on the non-enzymatic hydrolysis of 2′-deoxyribonucleic acids, where these results can be used to guide the generation of an accurate model and provide verification of results. Finally, the computational studies that specifically examined hydrolysis of the glycosidic bond will be summarized.

1.2.3.1 Computational Deglycosylation Models

There are many examples of computational enzyme studies using a highly truncated approach in the literature, where Figure 1.13 summarizes a variety of small models used to study catalytic cleavage of the N-glycosidic bond in 2′-deoxyribonucleic
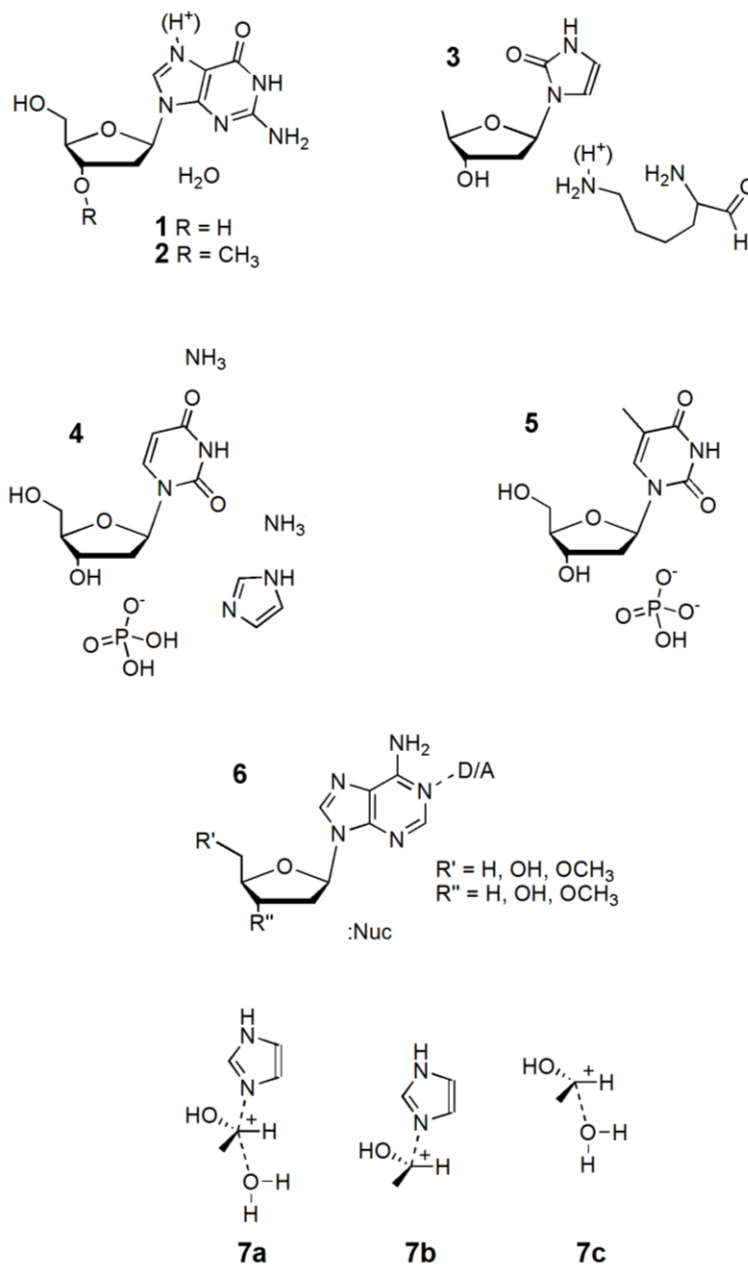
Figure 1.13 *Example computational models used to study deglycosylation of 2'-deoxyribonucleic acids.*

acids. Models $\mathbf{1}^{34}$ and $\mathbf{2}^{35}$ were used to study the acid-catalyzed hydrolysis of dG, where

model **1** did not include association of the water molecule (i.e., unimolecular

deglycosylation). The hOGG1 catalyzed cleavage of 8-oxoguanine from DNA has been

modeled using both truncated base and sugar moieties and a lysine residue as the

nucleophile (Figure 1.13, **3**).[36] The mechanism of action of thymidine phosphorylase has

been modeled in a variety of ways, where two examples are given in Figure 1.13 (**4** and

5). Model **4** contains the functional groups of three active-site amino acids (specifically, a

histidine modeled as imidazole, and two lysine residues modeled as ammonia),[37] while

these groups are absent in model **5**.[38] The Wetmore group recently expanded their small

model study of UDG (Section 1.2.2.1.6) to include the four standard DNA nucleotides,

where a variety of models were considered as shown for dA in Figure 1.13, model **6**.[30,39]

Finally, very small models are typically utilized to interpret kinetic isotope effects. For

example, ricin toxin A-chain may hydrolytically cleave the N-glycosidic bond of dA, and

possible transition states were modeled as shown in Figure 1.13, **7a-c**.[40]

From these examples, it can be seen that there is generally a lack of leaving group

stabilization and/or nucleophile activation in computational models of nucleotide

deglycosylation reactions (with the exception of **4** and **6**). In combination with the lack of

explicit charge stabilization, the reaction mechanisms were optimized in the gas phase,

which destabilizes charge separated transition states and intermediates. The inaccuracies

in the computational models have led to some anomalies. For example, some studies

report dissociative transition states that involve transfer of a proton from the sugar moiety

to the nucleobase,[30,35,36,39] where an example is shown in Figure 1.14.
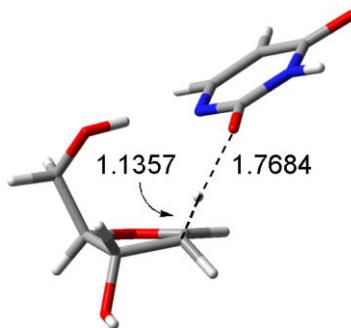
Figure 1.14 *An example of an irregular deglycosylation transition state with an elongated C–H bond in the sugar moiety and partial transfer to O2 of the uracil moiety due to incomplete computational models.*

Considering the difficulties involved in truncated enzymatic systems, it might be useful to step back and examine the non-enzymatic mechanism. A study of the uncatalyzed reaction should provide information on what types of interactions are required to allow the reaction to proceed. These results can then be used to identify the components in the enzymatic mechanism that should be treated at high or low levels of theory, or not included at all in the computational model. For example, if the uncatalyzed reaction requires protonation of the leaving group, then one or more residues in the active site that can, at a minimum, act as a donor in a strong hydrogen bond to the substrate should be included.

1.2.3.2 Experimental dR Hydrolysis

Since UDG uses a water molecule as a nucleophile, the uncatalyzed reaction corresponds to (neutral) dR hydrolysis, which has been studied experimentally for several decades. The N-glycosidic bond of nucleic acids is very stable, with a half-life of 70 to 230 years at room temperature.[3] Although the nucleoside form is more labile, experimental work has utilized either acidic conditions,[32,41,42] or high heat to induce the desired reactions[42,43] to overcome these large barriers.

Early work on the acid-catalyzed hydrolysis of pyrimidine nucleosides indicated that protonation of the nucleoside precedes deglycosylation.[32] However, it was not until the late '60s early '70s that it was determined where the protonation occurs. Shapiro and Kang[41a] (pyrimidines) and Zoltewicz et al.[41b] (purines) both found evidence of nucleobase protonation, which contrasted a previously proposed mechanism of O4′ protonation.[32] Later work determined the uncatalyzed reaction proceeds at elevated temperatures via spontaneous depurination/depyrimidination, followed by addition of a water molecule.[3,43a,43c] This two-step reaction corresponds to an S_N1 mechanism, where Schroeder et al. found the deglycosylation half-lives to be between 12 and 43 years at 37°C.[3]

In summary, experimental studies indicate that the neutral hydrolysis of 2′-deoxyribonucleosides occurs via a two-step, dissociative mechanism, where the nucleobase spontaneously dissociates before the addition of the water molecule. In the case of acid-catalyzed hydrolysis, the nucleobase strips a proton from the solvent prior to deglycosylation. Therefore, there is no evidence of the C2′-proton abstraction product seen in previous computational work. This suggests that the previous computational approach of starting with a model that only includes the substrate and nucleophile is incapable of describing the uncatalyzed reaction, and it is unknown to what degree this affects the enzyme catalyzed reaction.

1.2.3.3 Computational dR Hydrolysis

Armed with the knowledge that the hydrolysis of 2′-deoxyribonucleosides occurs in a two-step process, with possible protonation of the base in acidic environments, the deglycosylation models mentioned above can be reevaluated. Of the models discussed in

Section 1.2.3.1, only three (**1**, **2** and **6**) utilized a water molecule as the nucleophile, and results from these studies will be discussed below.

In 2002, Baik et al. examined the effect of protonation or addition of cisplatin at N7 on the depurination of dG (Figure 1.13, **1**).[34] In this study, the glycosidic bond was incrementally elongated from the reactant to the product. After each gas-phase optimization, a water molecule was added near the glycosidic bond before calculating the solvation energy. This approach was taken to avoid complications with adding bulk solvent effects. While this study does not report a C2′-proton abstraction product, a short hydrogen bond between the water molecule and C2′–H was observed during the deglycosylation process (Figure 1.15). The short hydrogen bond may be due to the sugar moiety stabilizing charge build up by transferring a proton to the water molecule rather than the nucleobase, as seen in the previous studies. Nevertheless, Baik et al. report a Gibbs energy of 126 kJ mol^{-1}, which is close to the experimental value of 127 kJ mol^{-1}.[3]
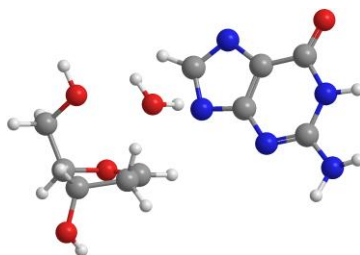
Figure 1.15   *dG deglycosylation structure reported by Baik et al. with a glycosidic bond distance elongated to 2.833 Å. The water molecule was added post-optimization.*

Rios-Font et al. later extended the Baik et al. study to include the full hydrolysis of dG (neutral and N7 protonated) using gas-phase optimizations with solvent effects included post-optimization (Scheme 1.4, model **2**).[35] This group utilized a slightly modified sugar where the 5′-OH was capped with a methyl group to prevent hydrogen

bonding, and one explicit water molecule was included as the reaction nucleophile. Transition state modeling was implemented, where nucleophilic attack was considered from a variety of orientations. In the models that led to depurination, a proton was abstracted from C2′, either by the base directly or by the nucleophilic water, which subsequently protonated the base.[34] This high energy process led to a Gibbs free energy of activation of 222 kJ mol^{-1} for the direct cleavage process, and 192 kJ mol^{-1} for the water assisted process. These values are larger than that calculated for the S_N2 hydrolysis by Millen et al. (127.4 kJ mol^{-1}),[39] the neutral depurination of dG reported by Baik et al. (126 kJ mol^{-1})[34] and experimental hydrolysis barriers determined by Schroeder et al. (127 kJ mol^{-1}).[3]

Millen et al.[30,39] (Figure 1.13, model **6**), focused on the synchronous (S_N2) hydrolysis mechanism, since the previous study on dU hydrolysis found the S_N1 mechanism to be much higher in energy.[30] This study reports the correct activation energy trend for dU, dT, dA and dC; however, dG was found to have the second highest activation energy instead of the lowest. The calculated trend corresponds to the (relative) N1 (pyrimidine) and N9 (purine) acidities previously found by the Wetmore group.[29d] However, this study incorrectly described the overall (experimental) mechanism.

1.2.3.4 Summary

Thus far, the few published computational studies of the neutral deglycosylation of 2′-deoxyribonucleosides have utilized very small models. Furthermore, all quantum mechanical studies that have been done were initially carried out in the gas phase, with solvent effects included post-optimization. A combination of gas-phase treatment and over truncation may have led some studies to report transition states and mechanisms that

are not supported by experiment. Therefore, there is still a great deal of research to be done in order to develop a computational approach capable of accurately predicting deglycosylation mechanisms.

1.3 Thesis Overview

As outlined in the present Chapter, although nucleotide hydrolysis is of great interest, the previous computational work on the deglycosylation of 2′-deoxyribonucleosides focused on very small model systems with no explicit solvent. However, previous literature on reactions involving small organic molecules indicates that some reactions (such as proton transfer,[44] tautomerization,[45] and deamination[46]) require additional explicit water molecules for accurate modeling.[47] Therefore, the initial portion of this thesis (Chapter 2) investigates the incorporation of solvent into previous dU hydrolysis models to determine an appropriate computational model. Once an appropriate hydrolysis model is developed for dU, the applicability of this method will be tested by studying the hydrolysis of the natural 2′-deoxyribonucleosides (dA, dC, dG and dT) in Chapter 3. The key concepts discovered in the preceding chapters will later be utilized to model the hydrolysis of dU by the enzyme UDG in Chapter 4. Finally, Chapter 5 will summarize key concepts discovered while studying nucleoside hydrolysis.

1.4 References

(1)   Campbell, N. A.; Reece, J. B. *Biology*, 6th ed.; Pearson Education, Inc: San Francisco, CA, 2002.

(2)   (a) Lindahl, T.; Nyberg, B. *Biochemistry* **1974**, *13*, 3405–3410; (b) Mosbaugh, D. W.; Bennett, S. E. *Prog. Nucleic Acids Res. Mol. Biol.* **1994**, *48*, 315–370.

(3)   Schroeder, G. K.; Wolfenden, R. *Biochemistry* **2007**, *46*, 13638–13647.

(4)   (a) Barsky, D.; Foloppe, N.; Ahmadia, S.; Wilson, D. M.; MacKerell, A. D. *Nucleic Acids Res.* **2000**, *28*, 2613–2626; (b) Wilson, D. M.; Barsky, D. *Mutat. Res.-DNA Repair* **2001**, *485*, 283–307.

(5)   (a) Slupphaug, G.; Eftedal, I.; Kavli, B.; Bharati, S.; Helle, N. M.; Haug, T.; Levine, D. W.; Krokan, H. E., *Biochemistry* **1995,** *34*, 128–138; (b) Sousa, M. M. L.; Krokan, H. E.; Slupphaug, G. *Mol. Asp. Med.* **2007**, *28*, 276–306.

(6)   For recent reviews see: (a) Akbari, M.; Krokan, H. E. *Mech. Ageing Dev.* **2008**, *129*, 353–365; (b) Cadet, J.; Carell, T.; Cellai, L.; Chatgilialoglu, C.; Gimisis, T.; Miranda, M.; O'Neill, P.; Ravanat, J. L.; Robert, M. *Chimia* **2008**, *62*, 742–749; (c) Caldecott, K. W. *Nat. Rev. Genet.* **2008**, *9*, 619–631.

(7)   See, for example: (a) Colis, L. C.; Raychaudhury, P.; Basu, A. K. *Biochemistry* **2008**, *47*, 8070–8079; (b) Labet, V.; Morell, C.; Grand, A.; Cadet, J.; Cimino, P.; Barone, V. *Org. Biomol. Chem.* **2008**, *6*, 3300–3305; (c) Stone, M. P.; Cho, Y. J.; Huang, H.; Kim, H. Y.; Kozekov, I. D.; Kozekova, A.; Wang, H.; Minko, I. G.; Lloyd, R. S.; Harris, T. M.; Rizzo, C. J. *Acc. Chem. Res.* **2008**, *41*, 793–804; (d) Su, D. G. T.; Kao, J. L. F.; Gross, M. L.; Taylor, J. S. A. *J. Am. Chem. Soc.* **2008**, *130*, 11328–11337.

(8)   See, for example: (a) Ye, W. J.; Sangaiah, R.; Degen, D. E.; Gold, A.; Jayaraj, K.; Koshlap, K. M.; Boysen, G.; Williams, J.; Tomer, K. B.; Ball, L. M. *Chem. Res. Toxicol.* **2006**, *19*, 506–510; (b) de-los-Santos-Alvarez, N.; de-los-Santos-Alvarez, P.; Lobo-Castanon, M. J.; Lopez, R.; Miranda-Ordieres, A. J.; Tunon-Blanco, P. *Electrochem. Comm.* **2007**, *9*, 1862–1866; (c) Martinez, G. R.; Ravanat, J. L.; Cadet, J.; Medeiros, M. H. G.; Di Mascio, P. *J. Mass Spectrom.* **2007**, *42*, 1326–1332; (d) Cadet, J.; Douki, T.; Ravanat, J. L. *Acc. Chem. Res.* **2008**, *41*, 1075–1083.

(9)   See, for example: (a) Jones, P. A.; Takai, D. *Science (Washington, D. C.)* **2001**, *293*, 1068–1070; (b) Aas, P. A.; Otterlei, M.; Falnes, P. O.; Vagbo, C. B.; Skorpen, F.; Akbari, M.; Sundheim, O.; Bjoras, M.; Slupphaug, G.; Seeberg, E.; Krokan, H. E. *Nature (London)* **2003**, *421*, 859–863; (c) Riggins, J. N.; Pratt, D. A.; Voehler, M.; Daniels, J. S.; Marnett, L. J. *J. Am. Chem. Soc.* **2004**, *126*, 10571–10581; (d) Svedruzic, Z. M.; Reich, N. O. *Biochemistry* **2004**, *43*, 11460–11473; (e) Minoshima, M.; Bando, T.; Sasaki, S.; Shinohara, K.; Shimizu, T.; Fujimoto, J.; Sugiyama, H. *J. Am. Chem. Soc.* **2007**, *129*, 5384–5390; (f) Loeppky, R. N.; Shi, J. Z. *Chem. Res. Toxicol.* **2008**, *21*, 319–329.

(10)   (a) Sono, M.; Wataya, Y.; Hayatsu, H. *J. Am. Chem. Soc.* **1973**, *95*, 4745–4749; (b) Chen, H.; Shaw, B. R. *Biochemistry* **1993**, *32*, 3535–3539,

(11)   (a) Dedon, P. C.; Tannenbaum, S. R. *Arch. Biochem. Biophys.* **2004**, *423*, 12–22; (b) Rayat, S.; Qian, M.; Glaser, R. *Chem. Res. Toxicol.* **2005**, *18*, 1211–1218.

(12) Veliz, E. A.; Easterwood, L. M.; Beal, P. A. *J. Am. Chem. Soc.* **2003**, *125*, 10867–10876.

(13) Kavli, B.; Otterlei, M.; Slupphaug, G.; Krokan, H. E. *DNA Repair* **2007**, *6*, 505–516.

(14) Lindahl, T. *Nature (London)* **1993**, *362*, 709–715.

(15) (a) Lucaccioni, A.; Pavlov, Y. I.; Achilli, A.; Babudri, N. *Curr. Genet.* **2007**, *52*, 239–245; (b) Pettersen, H. S.; Sundheim, O.; Gilljam, K. M.; Slupphaug, G.; Krokan, H. E.; Kavli, B. *Nucleic Acids Res.* **2007**, *35*, 3879–3892.

(16) Lindahl, T. *Proc. Natl. Acad. Sci. U.S.A.* **1974**, *71*, 3649–3653.

(17) Mol, C. D.; Arvai, A. S.; Slupphaug, G.; Kavli, B.; Alseth, I.; Krokan, H. E.; Tainer, J. A. *Cell (Cambridge, Mass.)* **1995**, *80*, 869–878.

(18) (a) Krokan, H. E.; Standal, R.; Slupphaug, G. *Biochem. J.* **1997**, *325*, 1–16; (b) Stivers, J. T.; Jiang, Y. L. *Chem. Rev.* **2003**, *103*, 2729–2759; (c) Berti, P. J.; McCann, J. A. B. *Chem. Rev.* **2006**, *106*, 506–555.

(19) (a) Otterlei, M.; Warbrick, E.; Nagelhus, T. A.; Haug, T.; Slupphaug, G.; Akbari, M.; Aas, P. A.; Steinsbekk, K.; Bakke, O.; Krokan, H. E. *EMBO J.* **1999**, *18*, 3834–3844; (b) Liu, P. F.; Burdzy, A.; Sowers, L. C. *Chem. Res. Toxicol.* **2002**, *15*, 1001–1009; (c) Duraffour, S.; Ishchenko, A. A.; Saparbaev, M.; Crance, J. M.; Garin, D. *Biochemistry* **2007**, *46*, 11874–11881.

(20) Stivers, J. T.; Pankiewicz, K. W., Watanabe, K. A. *Biochemistry* **1999**, *38*, 952–963.

(21) Fischer, J. A.; Muller-Weeks, S.; Caradonna, S. J. *Cancer Res.* **2006**, *66*, 8829–8837.

(22) (a) Mol, C. D.; Arvai, A. S.; Slupphaug, G.; Kavli, B.; Alseth, I.; Krokan, H. E.; Tainer, J. A. *Cell (Cambridge, Mass.)* **1995**, *80*, 869–878; (b) Slupphaug, G.; Mol, C. D.; Kavli, B.; Arvai, A. S.; Krokan, H. E.; Tainer, J. A. *Nature (London)* **1996**, *384*, 87–92; (c) Parikh, S. S.; Mol, C. D.; Slupphaug, G.; Bharati, S.; Krokan, H. E.; Tainer, J. A. *EMBO J.* **1998**, *17*, 5214–5226; (d) Parikh, S. S.; Walcher, G.; Jones, G. D.; Slupphaug, G.; Krokan, H. E.; Blackburn, G. M.; Tainer, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5083–5088; (e) Bianchet, M. A.; Seiple, L. A.; Jiang, Y. L.; Ichikawa, Y.; Amzel, L. M.; Stivers, J. T. *Biochemistry* **2003**, *42*, 12455–12460.

(23) (a) Drohat, A. C.; Jagadeesh, J.; Ferguson, E.; Stivers, J. T. *Biochemistry* **1999**, *38*, 11866–11875; (b) Jiang, Y. L.; Kwon, K.; Stivers, J. T. *J. Bio. Chem.* **2001**, *276*, 42347–42354; (c) Jiang, Y. L.; Stivers, J. T. *Biochemistry* **2001**, *40*, 7710–7719; (d) Jiang, Y. L.; Stivers, J. T. *Biochemistry* **2002**, *41*, 11236–11247; (e)

Jiang, Y. L.; Ichikawa, Y.; Song, F.; Stivers, J. T. *Biochemistry* **2003**, *42*, 1922–1929.

(24)   (a) Drohat, A. C.; Xiao, G. Y.; Tordova, M.; Jagadeesh, J.; Pankiewicz, K. W.; Watanabe, K. A.; Gilliland, G. L.; Stivers, J. T. *Biochemistry* **1999**, *38*, 11876–11886; (b) Drohat, A. C.; Stivers, J. T. *Biochemistry* **2000**, *39*, 11865–11875; (c) Drohat, A. C.; Stivers, J. T. *J. Am. Chem. Soc.* **2000**, *122*, 1840–1841; (d) Parker, J. B.; Stivers, J. T. *Biochemistry* **2008**, *47*, 8614–8622.

(25)   Dong, J.; Drohat, A. C.; Stivers, J. T.; Pankiewicz, K. W.; Carey, P. R. *Biochemistry* **2000**, *39*, 13241–13250.

(26)   Werner, R. M.; Stivers, J. T. *Biochemistry* **2000**, *39*, 14054–14064.

(27)   (a) Luo, N.; Mehler, E.; Osman, R. *Biochemistry* **1999**, *38*, 9209–9220; (b) Osman, R.; Fuxreiter, M.; Luo, N. *Comput. Chem.* **2000**, *24*, 331–339.

(28)   (a) Dinner, A. R.; Blackburn, G. M.; Karplus, M. *Nature (London)* **2001**, *413*, 752–755; (b) Ma, A.; Hu, J.; Karplus, M.; Dinner, A. R. *Biochemistry* **2006**, *45*, 13687–13696.

(29)   (a) Di Laudo, M.; Whittleton, S. R.; Wetmore, S. D. *J. Phys. Chem. A* **2003**, *107*, 10406–10413; (b) Whittleton, S. R.; Hunter, K. C.; Wetmore, S. D. *J. Phys. Chem. A* **2004**, *108*, 7709–7718; (c) Hunter, K. C.; Millen, A. L.; Wetmore, S. D. *J. Phys. Chem. B* **2007**, *111*, 1858–1871; (d) Hunter, K. C.; Wetmore, S. D. *J. Phys. Chem. A* **2007**, *111*, 1933–1942.

(30)   Millen, A. L.; Archibald, L. A. B.; Hunter, K. C.; Wetmore, S. D. *J. Phys. Chem. B* **2007**, *111*, 3800–3812.

(31)   Jiang, Y. L.; Cao, C.; Stivers, J. T.; Song, F.; Ichikawa, Y. *Bioorg. Chem.* **2004**, *32*, 244–262.

(32)   Garrett, E. R.; Seydel, J. K.; Sharpen, A. J. *J. Org. Chem.* **1966**, *31*, 2219–2227.

(33)   (a) Chen, S. L.; Fang, W. H.; Himo, F. *Theor. Chem. Acc.* **2008**, *120*, 515–522; (b) Ramos, M. J.; Fernandes, P. A. *Acc. Chem. Res.* **2008**, *41*, 689–698; (c) Schwabe, T.; Grimme, S. *Acc. Chem. Res.* **2008**, *41*, 569–579.

(34)   Baik, M. H.; Friesner, R. A.; Lippard, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 4495–4503.

(35)   Rios-Font, R.; Rodriguez-Santiago, L.; Bertran, J.; Sodupe, M. *J. Phys. Chem. B* **2007**, *111*, 6071–6077.

(36)   Schyman, P.; Danielsson, J.; Pinak, M.; Laaksonen, A. *J. Phys. Chem. A* **2005**, *109*, 1713–1719.

(37)   Rick, S. W.; Abashkin, Y. G.; Hilderbrandt, R. L.; Burt, S. K. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 242–252.

(38)   Mendieta, J.; Martin-Santamaria, S.; Priego, E. M.; Balzarini, J.; Camarasa, M. J.; Perez-Perez, M. J.; Gago, F. *Biochemistry* **2004**, *43*, 405–414.

(39)   Millen, A.L.; Wetmore, S.D. *Can. J. Chem.* **2009**, *87*, 850–863.

(40)   Chen, X. Y.; Berti, P. J.; Schramm, V. L. *J. Am. Chem. Soc.* **2000**, *122*, 6527–6534.

(41)   (a) Shapiro, R.; Kang, S. *Biochemistry* **1969**, *8*, 1806–1810; (b) Zoltewicz, J. A.; Clark, D. F.; Sharpless, T. W.; Grahe, G. *J. Am. Chem. Soc.* **1970**, *92*, 1741–1750.

(42)   Shapiro, R.; Danzig, M. *Biochemistry* **1972**, *11*, 23–29.

(43)   (a) Lindahl, T.; Karlstro.O. *Biochemistry* **1973**, *12*, 5151–5154; (b) Lindahl, T.; Nyberg, B. *Biochemistry* **1974**, *13*, 3405–3410; (c) Ratsep, P. C.; Pless, R. C. *J. Org. Chem.* **1988**, *53*, 3241–3246.

(44)   See, for example, (a) Kohanoff, J.; Koval, S.; Estrin, D. A.; Laria, D.; Abashkin, Y. *J. Chem. Phys.* **2000**, *112*, 9498–9508; (b) Liang, W. C.; Li, H. R.; Hu, X. B.; Han, S. J. *J. Phys. Chem. A* **2004**, *108*, 10219–10224.

(45)   See, for example, (a) Kim, Y.; Lim, S.; Kim, H. J.; Kim, Y. *J. Phys. Chem. A* **1999**, *103*, 617–624; (b) Li, P.; Bu, Y. X. *J. Phys. Chem. B* **2004**, *108*, 18088–18097; (c) Li, Q. G.; Xue, Y.; Yan, G. S. *J. Mol. Struct. (THEOCHEM)* **2008**, *868*, 55–64.

(46)   (a) Manojkumar, T. K.; Suh, S. B.; Oh, K. S.; Cho, S. J.; Cui, C. Z.; Zhang, X.; Kim, K. S. *J. Org. Chem.* **2005**, *70*, 2651–2659; (b) Almatarneh, M. H.; Flinn, C. G.; Poirier, R. A.; Sokalski, W. A. *J. Phys. Chem. A* **2006**, *110*, 8227–8234; (c) Almerindo, G. I.; Pliego Jr., J. R. *J. Braz. Chem. Soc.* **2007**, *18*, 696–702; (d) Wu, Y.; Xue, Y.; Xie, D. Q.; Kim, C. K.; Yan, G. S. *J. Phys. Chem. B* **2007**, *111*, 2357–2364; (e) Wu, Y.; Jin, L.; Xue, Y.; Xie, D. Q.; Kim, C. K.; Guo, Y.; Sen Yan, G. *J. Comput. Chem.* **2008**, *29*, 1222–1232; (f) Gao, J. Y.; Zeng, Y.; Zhang, C. H.; Xue, Y. *J. Phys. Chem. A* **2009**, *113*, 325–331.

(47)   (a) Kallies, B.; Mitzner, R. *J. Mol. Model.* **1998**, *4*, 183–196; (b) Zhan, C. G.; Landry, D. W.; Ornstein, R. L. *J. Am. Chem. Soc.* **2000**, *122*, 2621–2627; (c) Tautermann, C. S.; Voegele, A. F.; Loerting, T.; Kohl, I.; Hallbrucker, A.; Mayer, E.; Liedl, K. R. *Chem. Euro. J.* **2002**, *8*, 66–73; (d) Zhang, L. D.; Xie, D. Q.; Xu, D. G.; Guo, H. *Chem. Commun.* **2007**, 1638–1640; (e) Zeng, Y.; Xue, Y.; Yan, G. *J. Phys. Chem. B* **2008**, *112*, 10659–10667.

Chapter 2: Hydrolysis of 2′-Deoxyuridine[d]

2.1 Introduction

As discussed in Section 1.1.3, much of the previous small model, computational

work on the uncatalyzed hydrolysis of dU,[1,2] and other nucleotides,[3-5] has been unable to

correctly predict a two-step, dissociative, mechanism as determined by experiment.[6-8]

Thus far, the calculations have been carried out in the gas phase, with (bulk) solvent

effects generally included post-optimization as a correction to the energy.[9] Furthermore,

to the best of my knowledge, only one study has correctly predicted the experimental
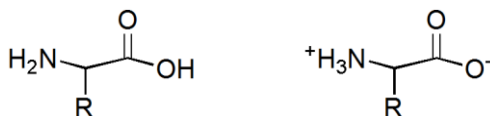
Figure 2.1   *Comparison of the tautomeric forms of an amino acid, where R is the functional group defining the amino acid.*

depurination barrier using this approach.[10]

Computational studies on the stability of amino acid tautomers indicate that the

charge separated zwitterion form (Figure 2.1, right) is only stable when implicit solvent

effects are included during the optimization step.[11] With respect to dU hydrolysis, this

implies that the neutral S_N2 transition state would be more stable than the charge

separated S_N1 intermediate when optimized in the absence of implicit solvent.[12] In

addition to the issue of charge separation, a body of literature on the hydrolysis of small

organic molecules, as well as tautomerization of nucleic acids, indicates that one or more

water molecules may participate in the reaction.[10,13] Therefore, inclusion of several

discrete water molecules in computational models is necessary to allow for proton

transfer, which also aids charge stabilization.

This chapter uses these ideas to generate a hydrolysis model that is capable of

correctly predicting the mechanism and rate-limiting barrier for the (neutral) hydrolysis

of dU. Since many computational studies utilize reduced systems,[14] initially, a model that

only includes the nucleoside will be considered to determine the effect of such a

truncation. The above studies suggest that using both implicit and explicit approaches

may stabilize the charge separated intermediate (and TS) involved in dU hydrolysis.

Therefore, explicit solvent molecules will be incorporated into the model, where this

work will identify the appropriate number of water molecules to include. Understanding

the effect of implicit and explicit solvent on the mechanism of dU hydrolysis will

contribute to the generation of more realistic, and accurate models for future studies of

the hydrolysis of other DNA nucleosides and nucleotides, as well as hydrolysis reactions

catalyzed by DNA repair (or other) enzymes.

2.2 Computational Details

Previous work on the depurination of dG by Baik et al. indicates that

computational difficulties with modeling N-glycosidic bond dissociation may arise since

the transition state region of the potential energy surface (PES) is very flat.[10] This

obstacle can potentially be overcome by slowly following the reaction coordinate using

constrained optimizations, rather than optimize the transition state directly.[10]

To ensure all possible reaction routes were considered, and to limit mechanistic

bias, the hydrolysis of dU was systematically examined in the present work using fixed

coordinate optimizations to generate detailed reaction PES. All geometry optimizations

were carried out using the B3LYP density functional method with the 6-31+G(d) basis set. Test calculations with a larger basis set (6-31+G(d,p)) displayed a change in reported energies of less than 1.5 kJ mol^{-1}, except for structures with multiple proton transfers, where the change was approximately 3 kJ mol^{-1}. This method and basis set combination has been recently used in PES scans for the hydrolysis of phosphodiesters,[15] and B3LYP has also been most commonly used to study nucleic acid glycosidic bond cleavage,[1,3–5] which will allow direct comparison to the current results. Implicit solvation effects were incorporated using the integral-equation formalism polarizable continuum model (IEF-PCM, referred to as PCM throughout) with a dielectric constant of $\varepsilon = 78.39$ for water at 298 K.[e] In order to better understand the effect of the solvation approach on the reaction PES, two techniques for incorporating bulk solvation effects were utilized: (1) solvent-phase (PCM) single-point (energy) calculations on gas-phase geometries, and (2) full solvent-phase (PCM) optimizations. In both cases, the same level of theory as implemented in the gas-phase optimizations was employed. All calculations were carried out using Gaussian 03.[16] Additional computational details for the different sized systems are provided below.

2.2.1 Unimolecular Cleavage

The unimolecular cleavage of dU was modeled using a one-dimensional PES scan. First, a variety of dU conformations were optimized in the gas phase by altering C5′- and C3′-hydroxyl orientations, as well as the sugar puckering. The lowest energy geometry (the reactant) has a glycosidic bond length of 1.477 Å, and C2′-*endo* sugar puckering, with both of the hydroxyl groups directed away from the base. The *anti*

---

[e] All PCM calculations were performed using the following keywords: *radii*=UAHF, *Rmin*=0.5, and *ofac*=0.8.

conformation of dU was utilized since it can be compared to the UDG mechanism and the *syn* and *anti* conformers are close in energy (not shown). Next, the glycosidic bond length was constrained to 1.5 Å and dU was re-optimized. Subsequently, the glycosidic bond was elongated or compressed by 0.1 Å and held fixed during the new optimization while the remainder of the system was allowed to relax. Thus, each new structure is dependent on the adjacent point on the surface. This process was repeated to generate the PES in a step-by-step manner with glycosidic bond lengths spanning 1.2 – 3.5 Å.

2.2.2 Hydrolysis

The hydrolysis of dU was investigated using two-dimensional PES scans and models containing one or three discrete water molecules. Initially, Monte Carlo (MC) searches were performed to determine the most likely location of the water molecule(s). Specifically, an initial geometry was generated by inspection, and MC calculations were run in the gas phase using the default routine and the AMBER[17,f] molecular mechanics method (parm99ff) as implemented in HyperChem.[18] The MC calculations were initially performed at a variety of temperatures between 100 and 400 K to determine the flexibility of the system containing one discrete water molecule with respect to the location of the water nucleophile. It was found that a simulation temperature of 300 K gives a broad sampling of possible nucleophile locations, and thus this temperature was utilized in all MC calculations. The most stable geometries, which typically span an approximate 5 kJ mol^{-1} range with respect to the (AMBER) global minimum, were subsequently optimized using B3LYP/6-31+G(d). In general, all geometries optimized to the same minimum, which was used as the initial geometry to construct the PES.

---

[f] Charges for the system were taken from a B3LYP/6-31+G(d) optimized initial geometry.

Starting from the optimum dU-water complexes identified using MC searches, PES scans were carried out by optimizing the reactant complex with the glycosidic bond length constrained to 1.5 Å and the nucleophile-sugar distance constrained to 3.4 Å.[g] Each subsequent point on the PES was generated from a preceding point on the reaction coordinate. In other words, the dissociative portion of the surface was modeled by slowly elongating the glycosidic bond and the concerted portion of the surface was generated by alternating whether the $N1–C1'$ or $O_w–C1'$ distance was elongated or compressed. A detailed description of how each point was generated can be found in Appendix A (Tables A1–A3). Over the course of the scan, the glycosidic bond was varied from approximately 1.3 – 4.3 Å and the nucleophile distance was altered between 1.2 and 3.8 Å, where both bonds were varied in 0.2 Å increments. Using the one water model as a test case, the resolution of the scan was increased to 0.1 Å increments near possible stationary points; however, this did not significantly alter the relative energies, and had no effect on the geometries (not shown). Indeed, another study that utilized a similar methodology found that further decreasing the increment size to 0.05 Å does not significantly affect the results.[15] The 0.2 Å increment leads to a total of approximately 200 (doubly) constrained optimizations to generate the PES for each model. Since each PES is a projection of a many dimensional surface onto a three dimensional surface, it is possible for several points on the true surface to correspond to one point on the calculated three-dimensional surface. Therefore, each structure was carefully compared to all

---

[g] When the unconstrained reactant with one explicit water molecule is fully optimized, the $C1'\cdots O_w$ distance is 3.452 Å. Therefore, the nucleophile distance was constrained to 3.4 Å for the initial point on the potential energy surface.

neighboring points to ensure consistency. Additional details and explanations of the models chosen will be presented in the appropriate Results and Discussion section.

2.2.3 Gibbs Energy Profiles

The Gibbs energy surfaces were calculated in the standard fashion.[19] Specifically, the surfaces generated from gas-phase optimized geometries were calculated by adding the scaled (0.9806) zero-point vibrational energy (ZPVE), thermal correction and (unscaled) entropy term (TCG) obtained from the gas-phase frequency calculation to the solvent-phase single-point energy (Equation 2.1). Similarly, the surfaces generated from the solvent-phase optimized geometries were calculated by adding the (scaled) zero-point energy, thermal correction and (unscaled) entropy term calculated in water to the solvent-phase (optimization) energy for each structure on the surface.

[2.1] $$G^{aq} = E_{aq}^{gas} + ZPC^{gas} + TCG^{gas}$$

It should be noted that the entropy of the system is calculated using the harmonic oscillator approximation for vibrational motion. Therefore, the Gibbs energy contains errors due to low vibrational frequencies, as well as the anharmonicity of a given mode. In terms of the present study, the very low barrier, and thus low vibrational frequency, associated with rotation of a water molecule about a hydrogen bond means that each additional explicit water molecule will introduce additional errors in the calculated Gibbs energy.

2.3 Results and Discussion

2.3.1 Unimolecular Glycosidic Bond Cleavage in dU

Experimental studies suggest that the hydrolysis of dU proceeds via a dissociative S_N1 pathway with a rate-limiting step of glycosidic bond cleavage to produce a fully formed uracil anion and oxacarbenium cation.[6–8] This indicates that the simplest computational model that may reproduce the experimental barrier for dU hydrolysis will correspond to the first, dissociative step of the S_N1 mechanism. Thus, the glycosidic bond cleavage in dU was initially modeled in the absence of the (water) nucleophile (i.e., unimolecular cleavage).

Figure 2.2A plots the relative energy (ΔE) of the nucleoside model as a function of the glycosidic bond length and reveals that the gas-phase surface is unstable. Specifically, once the glycosidic bond is elongated to approximately 3 Å, a sharp energy reduction occurs from roughly 160 to 90 kJ mol^{-1}. This energy drop is associated with a drastic geometrical change, where the uracil anion abstracts a proton from C2′ of the sugar moiety (Figure 2.2A, inset). Despite the lack of experimental evidence for a C2'-proton (C2′H) abstraction product, previous computational studies of glycosidic bond cleavage reactions in dU,[1] as well as other DNA nucleosides,[5] which implement full (unconstrained) TS optimizations, report similar findings. Indeed, the dU dissociation barrier estimated from the energy plateau preceding the geometrical change (~160 kJ mol^{-1}) is comparable to the previously published barrier (155 kJ mol^{-1}) based on full optimizations.[1h]

---

[h] It should be noted that differences between this thesis and previous studies also arise since the previous barriers were calculated at the B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) level of theory and include (0.9806) scaled zero-point vibrational energy corrections.
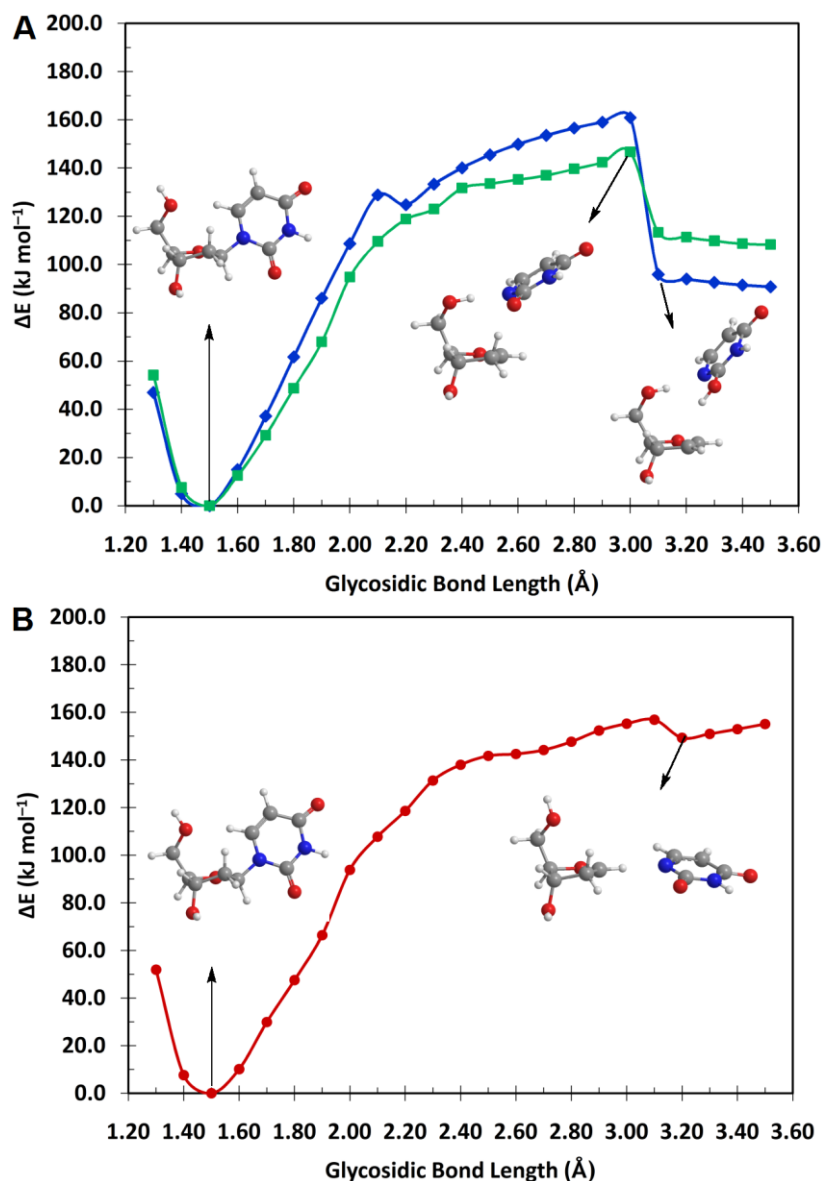
Figure 2.2   *B3LYP/6-31+G(d) potential energy surface for dU unimolecular glycosidic bond cleavage calculated in the gas phase (**A**, blue diamonds), in the solvent phase (water) using gas-phase geometries (**A**, green squares), and in the solvent phase (water) using solvent-phase geometries (**B**). Representative structures along the reaction are inset.*

Due to structural discrepancies between the calculated gas-phase unimolecular

pathway and experimental data, it is clear that gas-phase optimizations are inadequate for

this system even though they are often used in the literature. Indeed, the C2′H-abstraction

discussed above likely arises due to the inability of gas-phase calculations to model the

charge separation present in the expected product. This shortcoming of gas-phase calculations may also contribute to the previous incorrect computational prediction that dU hydrolysis occurs via an S_N2, rather than an S_N1, pathway.[1] Therefore, the next step to improve the computational model is to accommodate charge separation by including bulk solvent effects.

Bulk solvent effects can be incorporated into a computational model in two ways: 1) by performing (implicit) solvent-phase (water) single-point calculations on the gas-phase geometries; and 2) by performing (implicit) solvent-phase optimizations. The first method will determine the *energetic* effects of the (bulk) solvent and the second method will determine the *structural* effects of the (bulk) solvent. Figure 2.2 compares the potential energy surfaces for dU deglycosylation calculated upon inclusion of bulk solvent effects using the two techniques. Although solvent-phase single-point calculations stabilize the gas-phase activation energy by about 15 kJ mol^{-1} (Figure 2.2A), problems with the gas-phase geometry (e.g., C2′H-abstraction in the product complex) can clearly not be rectified. In contrast, solvent-phase optimizations do not lead to C2′H-abstraction in the product complex. Instead, the reaction proceeds to a charge-separated product, where N1 of the uracil anion hydrogen bonds with C1′ of the cationic sugar moiety (Figure 2.2B, inset).

The barrier for the solvent-phase optimized reaction is estimated from the energy plateau to be ~155 kJ mol^{-1}, which is similar to the gas-phase estimated barrier (~160 kJ mol^{-1}). Interestingly, this activation energy (Figure 2.2B) falls between the gas-phase activation energy and that calculated using solvent-phase single-point calculations on gas-phase geometries up to the discontinuity (Figure 2.2A). Thus, even though solvent-

phase optimizations may not significantly change the calculated barrier, it is imperative to consider solvent-phase optimizations to yield realistic structures along the reaction coordinate. It is also noted that the continuous increase in the solvent-phase potential energy as the glycosidic bond is elongated (Figure 2.2B) is likely the reason previous studies have highlighted difficulties modeling nucleoside (or nucleotide) deglycosylation reactions using full optimizations.[9,11]

Since the evaluation of Gibbs energies should permit a direct comparison of calculated and experimental barriers, the Gibbs energy profile for dU deglycosylation in water was determined using both the gas and solvent-phase optimized geometries (Figure 2.3). Although solvent-phase single-point calculations on gas-phase geometries over-stabilize the gas-phase potential energy when compared to solvent-phase optimizations (by ~10 kJ mol^{-1}), the same is not seen for the Gibbs energies. In this case, the gas-phase structures yield a flat Gibbs energy plateau at approximately 120 kJ mol^{-1} and the true solvent-phase Gibbs energy surface contains a similarly broad maximum near 115 kJ mol^{-1}. Furthermore, considering the size of the computational model, which includes the nucleoside in the absence of a discrete (water) nucleophile, and despite difficulties associated with the accurate calculation of Gibbs energies, which are underestimated due to over estimations of the entropy term,[20] the calculated $\Delta G^{\ddagger}$ values are close to the most recently reported experimental value (127 kJ mol^{-1}).[8] Thus, the calculations reported here support experimental suggestions that the hydrolysis of dU follows a dissociative, S_N1, mechanism.[6-8] However, while both solvation methods yield similar energies, the gas-phase optimizations breakdown at extended glycosidic bond lengths. This suggests that previous works including solvent effects through energy corrections to gas-phase
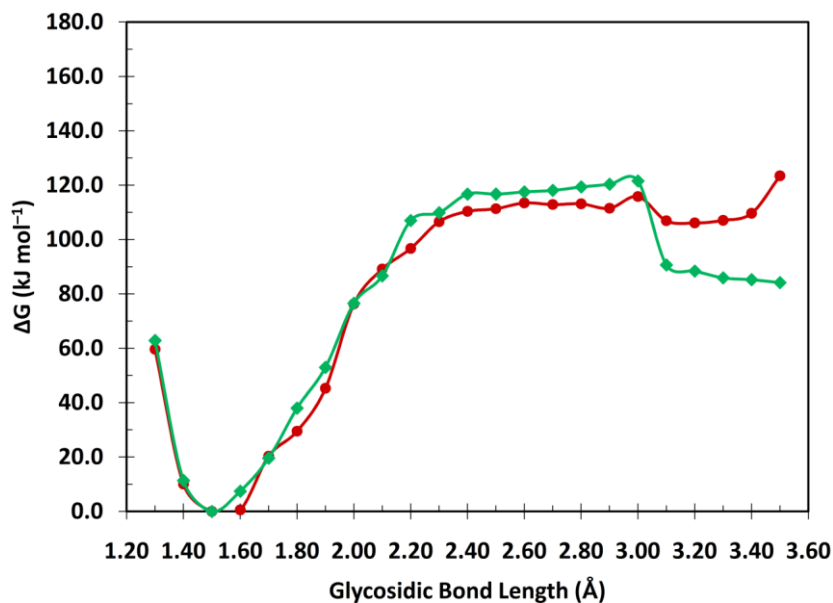
Figure 2.3 *PCM-B3LYP/6-31+G(d) Gibbs energy surface for dU unimolecular glycosidic bond cleavage calculated in (bulk) water using gas-phase geometries (green diamonds) and solvent-phase geometries (red circles).*

optimizations are likely not accurately modeling nucleoside deglycosylation reactions.

2.3.2 dU Hydrolysis Modeled with One Discrete Water Molecule

Although studying the unimolecular glycosidic bond cleavage in dU gives valuable information about the first, dissociative, step of the hydrolysis reaction, this model does not afford a complete picture of the reaction mechanism. Therefore, the next step in systematically increasing the computational model is to include a discrete water molecule as the reaction nucleophile. While the previous section revealed the importance of including the bulk solvent in optimizations of reaction intermediates for the unimolecular bond cleavage, it is not clear how the presence of a discrete nucleophile will affect this conclusion. Furthermore, one of the goals of the present study is to determine both the *energetic* and *structural* effects of bulk solvent. Hence, the potential energy surface for hydrolysis using a model that includes one discrete water nucleophile

was initially characterized in the gas phase, and subsequently bulk solvent effects were incorporated implicitly in either single-point calculations or optimizations.

Figure 2.4A displays the three-dimensional PES for the gas-phase hydrolysis of dU as a contour plot where the x-axis represents the distance between N1 of uracil and C1′ of the sugar (i.e., the glycosidic bond length), the y-axis represents the distance between the oxygen in the nucleophile and C1′ of the sugar (i.e., the nucleophile distance) and the relative energy is plotted in colour. The top-left corner of the surface with glycosidic bond lengths between 1.3 and 2.1 Å and nucleophile distances between 1.2 and 1.8 Å was not modeled since these geometrical constraints yield highly compressed structures and preliminary calculations (not shown) indicate that this region is much higher in energy than the rest of the surface. Representative structures throughout the reaction coordinate are displayed in the figures and were selected based on relative energy and the number of imaginary frequencies. Although structures reported as TS are not always true first-order saddle points, the largest (and generally only) imaginary frequency corresponds to the desired motion.

The reactant is the lowest energy point on the PES (bottom left corner), where the water nucleophile hydrogen bonds to both the C3′-hydroxyl of the sugar and O2 of U (R, Figure 2.4). A dissociative S_N1 transition state (denoted as TS_D in Figure 2.4C) corresponding to U dissociating from the sugar moiety can be found, where the water nucleophile follows the uracil anion away from the sugar due to a strong O_w–H⋯O2 hydrogen bond. This interaction prevents the uracil moiety, as well as the lone pairs on the nucleophile, from efficiently stabilizing the positive charge developing on the sugar. Consequently, as the glycosidic bond is further lengthened, geometrical instabilities and
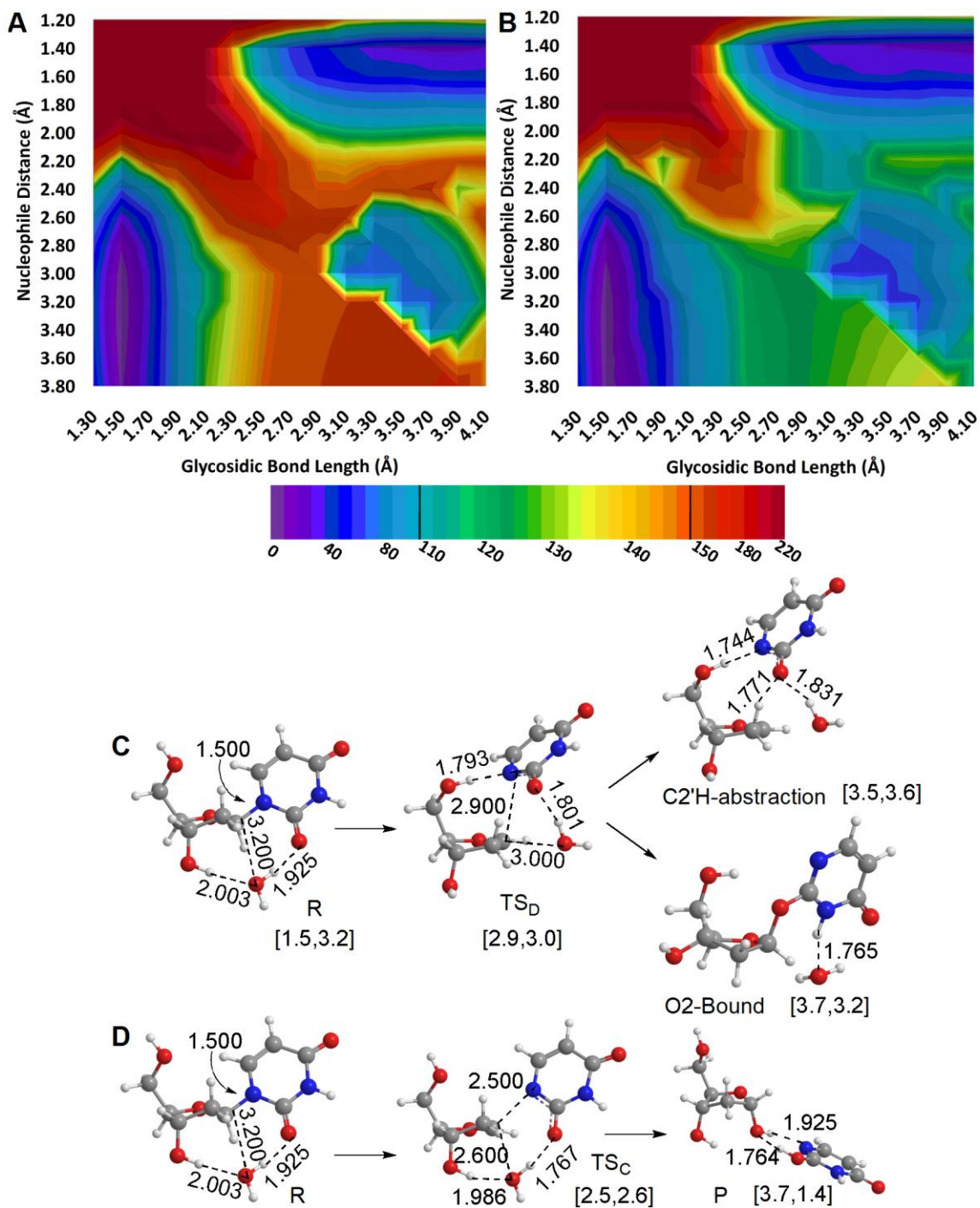
Figure 2.4 *Potential energy surface for the hydrolysis of dU with one explicit water molecule in the gas phase (A) and in the solvent phase (water) using gas-phase geometries (B). Representative structures that exhibit S_N1 (C) and S_N2 (D) character are shown. (Relative energies in kJ mol^{-1}, colour scale provided in legend, select B3LYP/6-31+G(d) distances (Å) provided including glycosidic bond length and nucleophile distance represented as [C1'···N1,C1'···Ow].)*

discontinuities (unrealistic strains on the system that alleviate themselves through spontaneous geometrical rearrangement) arise in the gas-phase PES (bottom right quadrant Figure 2.4A). Specifically, the red region (glycosidic bond lengths ranging between 2.9 and 3.9 Å) corresponds to the previously discussed C2′H-abstraction dissociation product, while the blue energy well (glycosidic bond lengths greater than 2.9 Å and nucleophile distances ranging between 2.6 and 3.8 Å) corresponds to a reaction intermediate where O2 of uracil is fully bound to C1' of the sugar moiety (O2-Bound, Figure 2.4C). Interestingly, the discontinuous region occurs at a glycosidic bond length (> 2.9 Å) similar to that discussed for the unimolecular gas-phase reaction (Figure 2.2A), and likely arises due to the instability of charge-separated species in the gas phase. Despite the collapse of the gas-phase PES, these calculations suggest that the dissociation barrier will be greater than 160 kJ mol^{-1}, which is very similar to the value calculated for unimolecular cleavage.

Since both S_N1 intermediates discussed above results from a collapse of the PES, neither structure (O2-bound nor C2′H-abstraction) directly leads to a hydrolysis product. However, a continuous *concerted* pathway (Figure 2.4D) can be followed from the reactant to the product, which occurs in the top right purple well on the PES. In the product complex, the uracil anion abstracts a proton from the original (water) nucleophile to form the O2 protonated uracil tautomer, which hydrogen bonds to the new C1′-hydroxyl group using both N1 and the new O2 proton (P, Figure 2.4D). The S_N2 TS (TS$_C$, Figure 2.4D) leads to an approximate barrier of 175 kJ mol^{-1}. Therefore, even though only the S_N2 pathway can be properly characterized on the gas-phase optimized surface, the path of slowest ascent corresponds to S_N1 dissociation.

Interestingly, this gas-phase PES is consistent with previously published findings by the Wetmore group obtained using full optimizations and a sugar model with methoxyl groups replacing the C3'- and C5'-hydroxyl groups.[1] Specifically, the S_N2 TS and barrier (~175 kJ mol^{-1}) found in the present work are similar to those previously reported (165.4 kJ mol^{-1}).[1,i] Furthermore, the product complexes identified in both studies are nearly identical, and the reaction energies differ by less than 4 kJ mol^{-1}.

The above discussion reveals that the two major problems with the gas-phase hydrolysis surface remain in the computational model including a discrete (water) nucleophile. First, there are instabilities in the dissociative intermediate region of the surface which prevent full characterization of the experimentally determined S_N1 pathway. Second, the calculated rate-limiting energy barrier cannot be confirmed by direct comparison to experiments, which were measured in solution. Therefore, bulk solvent effects were incorporated as discussed for the unimolecular mechanism.

Initially, solvent-phase single-point calculations were performed on gas-phase geometries to evaluate the energetic effect of the bulk solvent and to allow the calculation of the Gibbs energy surface for comparison to experiment. In contrast with the gas-phase surface (Figure 2.4A), the dissociative energy barrier obtained from solvent-phase single-point calculations on gas-phase geometries (Figure 2.4B) has decreased by over 30 kJ mol^{-1} to 125 kJ mol^{-1}, which is much greater stabilization than seen for the unimolecular mechanism (approximately 15 kJ mol^{-1}, Figure 2.2A). When the corresponding Gibbs energy surface is calculated (Figure 2.5), the dissociative $\Delta G^{\ddagger}$ is approximately 100 kJ

---

[i] It should be noted that differences between this thesis and previous studies also arise since the previous barriers were calculated at the B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) level of theory and include (0.9806) scaled zero-point vibrational energy corrections.
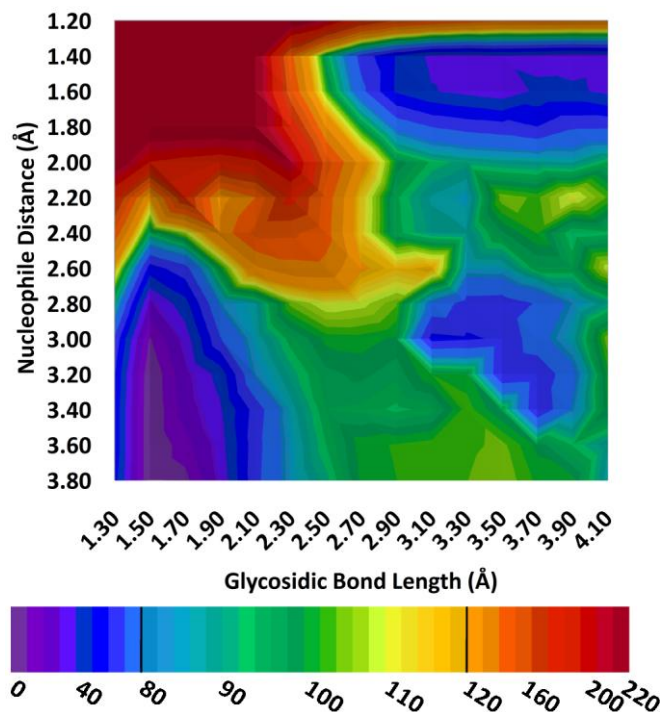
Figure 2.5 *PCM-B3LYP/6-31+G(d) Gibbs energy surface (kJ mol^{-1}, colour scale provided in legend) for dU hydrolysis modeled in bulk water with one explicit water molecule using gas-phase geometries.*

mol^{-1}, which is roughly 20 kJ mol^{-1} lower than that calculated for unimolecular cleavage (120 kJ mol^{-1}, Figure 2.3) and almost 30 kJ mol^{-1} lower than the experimental value (127 kJ mol^{-1}).[8] Most importantly, it must be emphasized that while solvent-phase single-point calculations on gas-phase geometries are able to provide more experimentally relevant energies, they do not address difficulties with the structures of reaction intermediates (i.e., the C2′H-abstraction and O2-bound states). Thus, although many groups justify gas-phase reaction mechanisms by comparing experimental barriers to calculated barriers adjusted using solvent-phase single-point calculations, this study shows that caution should be taken when interpreting mechanistic conclusions drawn from gas-phase geometries, especially for structures involving charge separation.

Subsequently, solvent-phase optimizations were performed to evaluate the structural effect of the bulk solvent in hopes to remove instabilities on the surface, and

thereby yield a better description of the dissociative S_N1 pathway. The resulting PES (Figure 2.6A) is quite different from the surfaces generated using gas-phase geometries (Figures 2.4A and 2.4B). On the solvent-phase surface, the reactant is still the lowest energy point (R, Figure 2.6), and resembles the gas-phase optimized structure (R, Figure 2.4). However, the path of slowest ascent now leads to a dissociative intermediate, where the uracil anion is involved in a C1′–H···O2 hydrogen bond (I, Figure 2.6C). The corresponding dissociative barrier is approximately 130 kJ mol^{-1}, where the TS (TS$_D$, Figure 2.6C) shows partial dissociation of the glycosidic bond and formation of the new (C1′–H···O2) hydrogen bond observed in the intermediate. In contrast to the gas-phase optimized geometry (Figure 2.4C), the water nucleophile stays below C1′ of the sugar in the transition state, and thus helps stabilize the oxacarbenium cation being formed. A high energy associative TS can also be found (TS$_A$, Figure 2.6C) with a barrier of 5 kJ mol^{-1} relative to the intermediate (138 kJ mol^{-1} relative to the reactant). However, the path of slowest ascent (yellow band, Figure 2.6A) does not point towards the associative TS, but rather suggests further migration of the uracil anion away from the sugar moiety with no association of the water nucleophile.
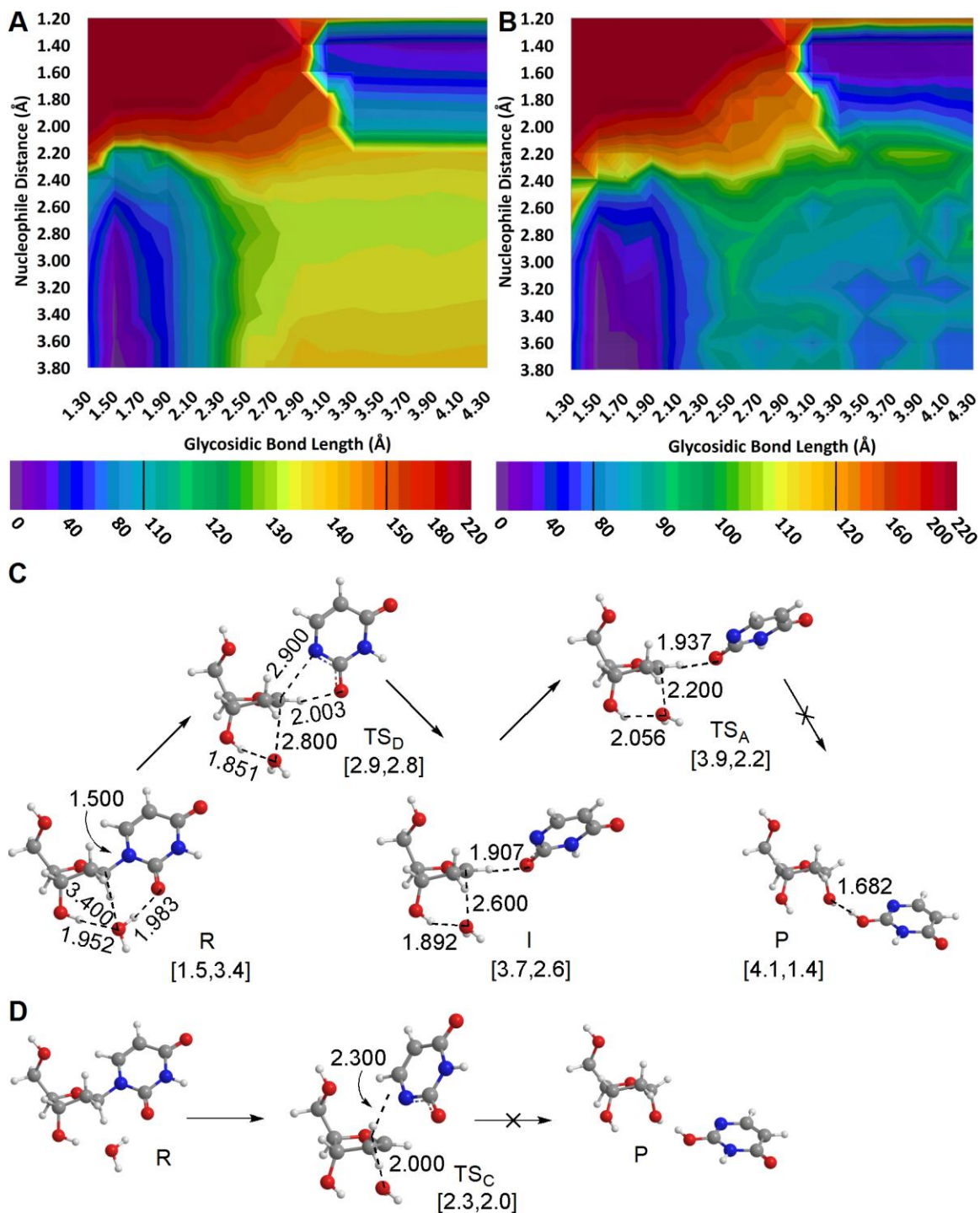
Figure 2.6 *The potential (**A**) and Gibbs (**B**) energy surfaces for the hydrolysis of dU with one explicit water molecule using solvent-phase geometries, where representative structures determined from the PES that exhibit S_N1 (**C**) or S_N2 (**D**) character are shown. (kJ mol^{-1}, colour scale provided in legend, and PCM-B3LYP/6-31+G(d) distances (Å) provided including glycosidic bond length and nucleophile distance represented as [C1'···N1,C1'··O$_w$].)*

It is interesting to note that the dissociative pathway on the solvent-phase optimized PES is very flat (Figure 2.6A), where the intermediate occurs in a mere 2 kJ mol^{-1} well on the surface. However, the corresponding Gibbs energy surface (Figure 2.6B) suggests that there is a dissociative intermediate in a well with a depth of over 15 kJ mol^{-1} (bottom right corner). This suggests that, in addition to allowing for comparison to experiment, the Gibbs energy surface can be used to aid the identification of important points that may be difficult to locate on the PES. A similar conclusion was drawn in a previous study of the glycosidic bond dissociation in dG due to platination.[9] It is also of note that the intermediate identified on the PES appears at a nucleophile distance of 2.6 Å, while the intermediate on the Gibbs energy surface occurs at a larger nucleophile distance (3.2 Å), which implies less association of the nucleophile. Thus, consideration of the Gibbs surface can also change the structure of intermediates along the reaction coordinate.

Contrary to the gas-phase model where the S_N2 (red) region extends to a nucleophile distance of ~2.8 Å (Figure 2.4A), the S_N2 region only reaches a nucleophile distance of ~2.4 Å in bulk water and therefore a majority of the solvent surface (Figure 2.6A) is indicative of an S_N1 mechanism. Nevertheless, a concerted, S_N2, transition state can be identified in the solvent phase (TS$_C$, Figure 2.6D), which leads to an approximate activation energy of 175 kJ mol^{-1}. Compared with the gas-phase structure (TS$_C$, Figure 2.4D), this S_N2 TS is less strained, where the strong O$_w$–H⋯O2 hydrogen bond seen in the gas phase is not present in the solvent-phase optimized geometry. Instead, the water nucleophile initiates attack on C1′ from directly below the sugar moiety with no explicit

activation by uracil, and the uracil anion begins to depart from the cationic sugar with no explicit charge stabilization.

Similar to the gas-phase product (P, Figure 2.4D), the solvent-phase product (P, Figure 2.6) shows transfer of a proton from the nucleophilic water to O2 of the uracil anion, where the O2 protonated uracil tautomer interacts with the C1′-hydroxyl group through an O2–H···O_w hydrogen bond. However, unlike the gas-phase optimized product, this structure is not smoothly connected to either the S_N1 or the S_N2 transition states on the solvent-phase surface. In the case of the S_N1 pathway, structures with nucleophile distances < 2.0 Å immediately change from those resembling the dissociative intermediate, which contain a strong C1′–H···O2 hydrogen bond, to structures where O2 of the uracil anion abstracts a proton from the water nucleophile. In the case of the S_N2 pathway, the continuous activation of water by uracil observed on the gas-phase concerted surface (TS_C, Figure 2.4D) is not seen on the solvent-phase surface (TS_C, Figure 2.6D). Instead, bulk solvent effects stabilize the charges formed in the transition state, but are insufficient to stabilize the fully formed uracil anion in the product, and therefore proton abstraction from water by the uracil anion spontaneously occurs beyond a glycosidic bond distance of 3.1 Å.

The solvent-phase optimized surface reveals that simply including bulk solvent and one discrete (water) nucleophile does not correct all instabilities in the gas-phase surface and in fact introduces new difficulties. Furthermore, the Gibbs energy surface reveals that the dissociative $\Delta G^{\ddagger}$ is approximately 93 kJ mol^{-1} (Figure 2.6B), which is even smaller than the value calculated using solvent-phase single-point calculations on gas-phase geometries (~100 kJ mol^{-1}) and therefore severely underestimates the

experimental value. Consequently, solvent-phase hydrolysis in a model including one discrete water molecule both fails to describe the full hydrolysis reaction and to predict the experimental $\Delta G^{\ddagger}$.

Previous literature on enzymatic hydrolysis reactions indicates that nucleophile activation is extremely important.[21–24] Therefore, a possible reason for the instabilities on the solvent-phase optimized surface may be the lack of adequate nucleophile activation. Indeed, evidence to support this statement is seen near the concerted transition state, where further association of the nucleophile leads to high energy, unstable structures and causes a collapse of the surface to structures containing an O_w–H···O2 hydrogen bond that subsequently activates the nucleophile. In parallel to the lack of nucleophile activation, the current model system does not provide complete stabilization of the leaving group (uracil) through, for example, discrete hydrogen-bonding interactions as seen in the active site of uracil DNA glycosylase.[21,25–27] While the bulk (implicit) solvent partially stabilizes charge separation and thereby prevents full proton abstraction by O2 in the uracil anion in the dissociation step, it cannot fully stabilize the leaving group. Therefore, even when implicit solvent is included during optimizations, additional stabilization of the uracil anion occurs in the product in the form of a strong C1′–H···O2 hydrogen bond (~1.9 Å) to the sugar cation or abstraction of a proton from the water nucleophile (which also acts as nucleophile activation). This suggests that an extended model system should include discrete hydrogen-bonding interactions to both activate the nucleophile and stabilize the departing base anion.

2.3.3 dU Hydrolysis Modeled with Three Explicit Water Molecules

The previous section suggests that a dU hydrolysis model containing one explicit water molecule is not large enough to accurately describe a continuous reaction surface. Specifically, during solvent-phase optimizations, instabilities were caused by the lack of both discrete nucleophile activation and leaving group stabilization. Therefore, initially one additional water molecule was added to the model between the nucleophilic water and O2 of uracil due to previous suggestions that some hydrolysis reactions require several explicit water molecules to permit a proton shuffle that activates the nucleophile and stabilizes the leaving group.[10,13] However, the lowest energy reactant for this model will lead to the same strain in the TS discussed for the model with one explicit water molecule, where the hydrogen bond angles are far from linear and the distances are not ideal. Therefore, a model formed by adding two discrete water molecules between the nucleophile and O2 of uracil in the original system was considered, which yields a realistic reactant geometry with decreased strain (R, Figure 2.7). Finally, since all previous models reveal the importance of including bulk solvent in the optimization routine, only solvent-phase structures were considered for this extended system.

The solvent-phase optimized PES for dU hydrolysis modeled using three discrete water molecules (Figures 2.7A) reveals that this system does not yield any of the inconsistencies discussed for the other models. Specifically, the S_N1 reaction progresses smoothly from the reactant (R, Figure 2.7C) to a reaction intermediate through an approximate 120 kJ mol^{-1} dissociative transition state (TS$_D$, Figure 2.7C). In the intermediate (I, Figure 2.7C), the anionic base has fully departed from the sugar and there is slight association of the nucleophile. In the second step of the reaction, the nucleophile
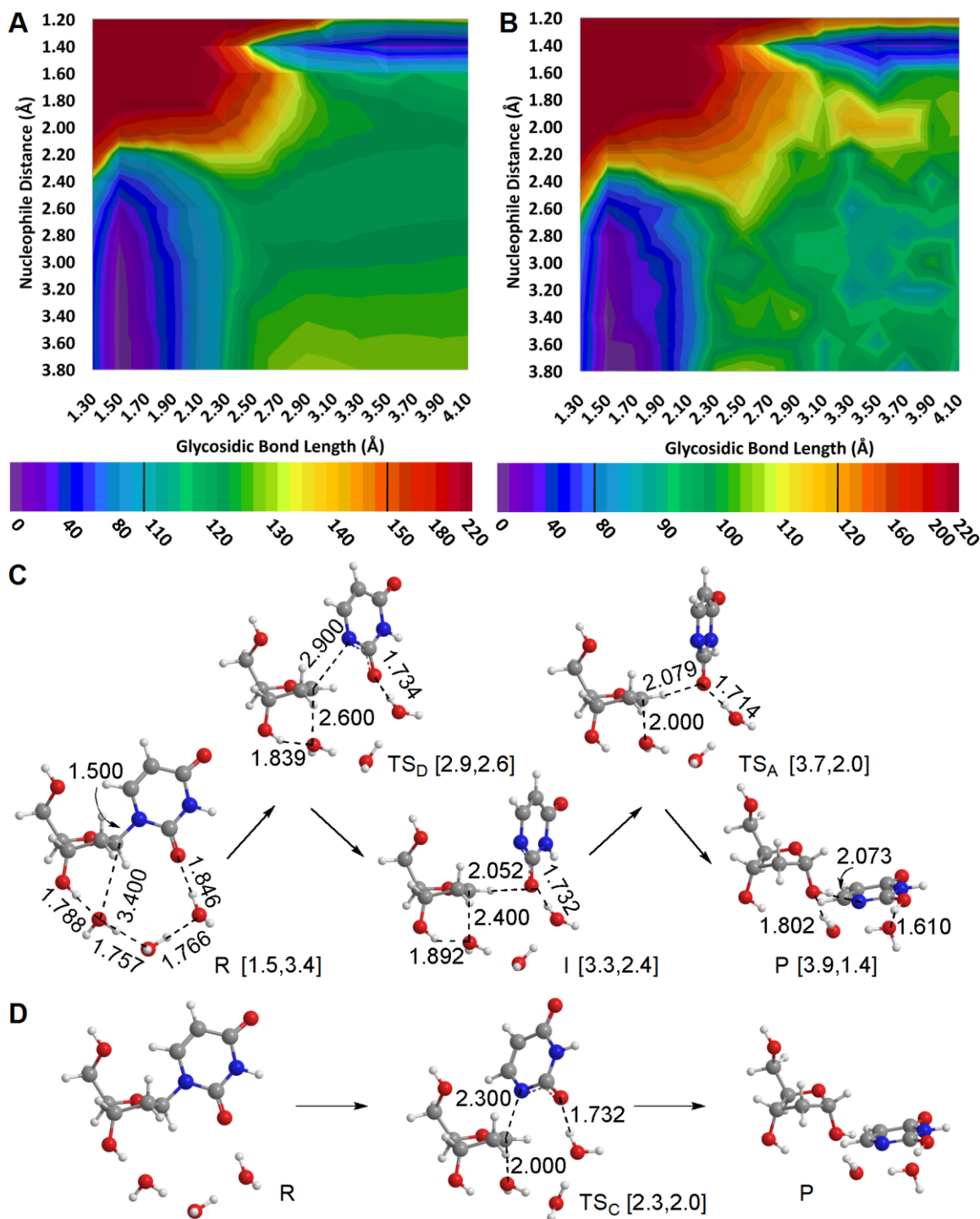
Figure 2.7 *The potential (**A**) and Gibbs (**B**) energy surfaces for the hydrolysis of dU with three explicit water molecules, where representative structures determined from the PES that exhibit S_N1 (**C**) or S_N2 (**D**) character are shown. (kJ mol^{-1}, colour scale provided in legend; PCM-B3LYP/6-31+G(d) distances (Å) provided including glycosidic bond length and nucleophile distance represented as [C1'···N1,C1'··O_w].)*

adds to C1' of the sugar through a ~3 kJ mol^{-1} barrier relative to the intermediate (123 kJ mol^{-1} relative to the reactant), where the TS involves a C1'–H···O2 hydrogen bond between the sugar cation and uracil anion (TS$_A$, Figure 2.7C). Finally, in the product, the dissociated base is located near the new hydroxyl group at C1', where the hydroxyl proton is hydrogen bonded to N1 of the uracil moiety (P, Figure 2.7C). The overall reaction energy is approximately 30 kJ mol^{-1}.

Throughout the scan, the O···O distances between water molecules remain nearly constant. However, structures with protons in between the water oxygen atoms can be found on the constrained surface, and full proton transfer from the nucleophile to O2 of the uracil anion occurs in the product. This proton transfer occurs in a smooth two-step process, where first a proton is transferred from the nucleophile to the nearest water molecule, and then full transfer to the uracil anion through the two additional waters occurs when the nucleophile is within 1.6 Å of C1'. This asynchronous proton transfer is characteristic of a water-assisted reaction mechanism.[12,13]

As found for the model containing one water molecule, an S$_N$2 transition state can be identified for the model containing three water molecules (TS$_C$, Figure 2.7D), which has a corresponding barrier of ~160 kJ mol^{-1}. Interestingly, this barrier falls between the S$_N$2 barriers previously reported for dU hydrolysis in the presence of one water molecule (165.4 kJ mol^{-1}) and dU hydrolysis by partially activated water (119.3 kJ mol^{-1} for the weakest activator studied).[1,j] Nevertheless, these new calculations suggest that the S$_N$1 pathway is almost 40 kJ mol^{-1} lower in energy than the S$_N$2 mechanism and therefore is

---

[j] It should be noted that differences between this thesis and previous studies also arise since the previous barriers were calculated at the B3LYP/6-311+G(2df,p)//B3LYP/6-31G(d) level of theory and include (0.9806) scaled zero-point vibrational energy corrections.

the most likely pathway. This is the first time that a computational model has matched the experimentally predicted nucleoside hydrolysis mechanism for dU.

As discussed previously, calculating the corresponding Gibbs energies will aid the identification of important reaction intermediates and allow direct comparison to experiment. Indeed, similar to the model with one discrete water molecule, the Gibbs energy surface for the model with three explicit water molecules (Figure 2.7B) shows a much more defined intermediate region than the PES (Figure 2.7A), which is almost 15 kJ mol^{-1} below the transition state region (compared with less than 5 kJ mol^{-1} for the potential energy surface). The calculated dissociative $\Delta G^{\ddagger}$ is ~100 kJ mol^{-1}, the intermediate falls at ~90 kJ mol^{-1}, the associative $\Delta G^{\ddagger}$ is ~25 kJ mol^{-1} relative to the intermediate (115 kJ mol^{-1} relative to the reactant) and the product relative Gibbs energy is ~35 kJ mol^{-1}.

A comparison of the calculated $\Delta G^{\ddagger}$ for the S_N2 pathway (140 kJ mol^{-1}) to that of the S_N1 pathway (100 kJ mol^{-1}) supports the previous conclusion that the reaction will proceed via an S_N1 mechanism. Thus, even though the calculated $\Delta G^{\ddagger}$ for the rate-determining step (100 kJ mol^{-1}) is smaller than the experimentally predicted value (127 kJ mol^{-1}), which could be due to known barrier underestimation by B3LYP[27] and intrinsic errors in the calculation of the entropy correction,[20] these results provide the first computational support for the experimentally predicted mechanism. Furthermore, it is interesting to note that the S_N1 Gibbs energy barrier calculated using solvent-phase optimizations on a model with *three* explicit water molecules is the same as that calculated using solvent-phase single-point calculations on gas-phase geometries for a model containing only *one* explicit water molecule. This reemphasizes the fact that even

though a model may yield acceptable energies, it may not correctly map the reaction coordinate, and therefore great care must be taken when drawing conclusions from potentially incomplete models.

2.4 Conclusions

The present chapter investigated the hydrolysis of 2′-deoxyuridine using a variety of computational models. Initially, the unimolecular glycosidic bond cleavage of dU was considered, which corresponds to the rate-limiting step in the proposed hydrolysis mechanism. The (Gibbs energy) barriers for unimolecular cleavage calculated in the presence of bulk (implicit) solvent relatively closely match the experimentally predicted hydrolysis barrier. Although this energetic agreement is obtained regardless of whether bulk solvent effects are incorporated during or after geometry optimizations, gas-phase geometries are unrealistic. It should be emphasized here that many computational works in the literature justify gas-phase optimized reaction mechanisms by comparing solvent-phase single-point energies to experimental values. The similarity in the gas-phase and solvent-phase optimized PES reported here indicate that, while the energies may be comparable, the geometries may be dramatically different, and care should be taken when interpreting gas-phase optimized reaction schemes involving dissociative mechanisms.

Next, (one or three) explicit water molecules were incorporated into the model. Careful analysis of the corresponding PES suggests that explicit solvent molecules that both activate the (water) nucleophile and stabilize the (uracil anion) leaving group, in addition to implicit solvent during optimization, must be incorporated into larger models in order to remove all instabilities. Indeed, a hydrolysis model that includes three discrete water molecules is the only model considered that yields a reaction surface with no

inconsistencies or instabilities along both the dissociative S_N1 and concerted S_N2 pathways. This work suggests that the reason for incorrect predictions reported in the literature in the past are due to the implementation of gas-phase optimizations.

In summary, this chapter emphasizes the importance of including both implicit and explicit solvation in computational models when studying the hydrolysis of DNA nucleosides. Furthermore, these findings indicate that studies involving hydrolysis of other nucleosides or nucleotides should not be limited to one discrete water molecule (acting as the nucleophile) in the computational system, which has been the standard practice to date for these systems. The model should also account for (bulk) solvent effects during the optimization routines, where previous studies of DNA hydrolysis have generally accounted for bulk solvent effects only through single-point calculations.

Although it is difficult to directly transfer these findings to complicated biological processes, this Chapter provides clues about designing appropriate computational models for the enzymatic hydrolysis of dU by uracil DNA glycosylase and other related enzymes. Specifically, since the current calculations reveal that both nucleophile activation and leaving group stabilization are crucial for nucleoside hydrolysis, computational models of hydrolysis by enzymes must explicitly include molecules to fulfill both of these roles and describe these molecules at the quantum mechanical level when hybrid quantum mechanics/molecular mechanics techniques are implemented. Finally, the results in this Chapter also have more general implications for the accurate modeling of any reaction that involves charge separation along the reaction coordinate.

The hypotheses presented above regarding how the results in the current Chapter can be translated to other systems will be further probed in the next two Chapters. In

Chapter 3, the hydrolysis model will be applied to the remaining 2′-deoxyribonucleosides

to test the general applicability of the model. Subsequently, the key interactions identified

when studying the dU hydrolysis mechanism will be utilized in Chapter 4 to design a

reduced UDG model, for studying the enzymatic deglycosylation of dU.

2.5 References

(1)   Millen, A. L.; Archibald, L. A. B.; Hunter, K. C.; Wetmore, S. D. *J. Phys. Chem. B* **2007**, *111*, 3800–3812.

(2)   Millen, A.L.; Wetmore, S.D. *Can. J. Chem.* **2009**, *87*, 850–863.

(3)   (a) Chen, X. Y.; Berti, P. J.; Schramm, V. L. *J. Am. Chem. Soc.* **2000**, *122*, 6527–6534; (b) McCann, J. A. B.; Berti, P. J. *J. Am. Chem. Soc.* **2007**, *129*, 7055–7064; (c) McCann, J. A. B.; Berti, P. J. *J. Am. Chem. Soc.* **2008**, *130*, 5789–5797.

(4)   (a) Birck, M. R.; Schramm, V. L. *J. Am. Chem. Soc.* **2004**, *126*, 2447–2453; (b) Osakabe, T.; Fujii, Y.; Hata, M.; Tsuda, M.; Neya, S.; Hoshino, T. *Chem-Bio. Infomatics J.* **2004**, *4*, 73–92.

(5)   (a) Rios-Font, R.; Rodriguez-Santiago, L.; Bertran, J.; Sodupe, M. *J. Phys. Chem. B* **2007**, *111*, 6071–6077; (b) Schyman, P.; Danielsson, J.; Pinak, M.; Laaksonen, A. *J. Phys. Chem. A* **2005**, *109*, 1713–1719.

(6)   Shapiro, R.; Kang, S. *Biochemistry* **1969**, *8*, 1806–1810.

(7)   (a) Garrett, E. R.; Seydel, J. K.; Sharpen, A. J. *J. Org. Chem.* **1966**, *31*, 2219–2227; (b) Lindahl, T.; Karlstro.O. *Biochemistry* **1973**, *12*, 5151–5154.

(8)   Schroeder, G. K.; Wolfenden, R. *Biochemistry* **2007**, *46*, 13638–13647.

(9)   (a) Rejnek, J.; Hanus, M.; Labelac, M.; Ryjacek, F.; Hobza, P. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2006–2017; (b) Aikens, C. M.; Gordon, M. S. *J. Am. Chem. Soc.* **2006**, *128*, 12835–12850; (c) Kim, S.; Schaefer, H. F. *J. Chem. Phys.* **2006**, *125*, 144305; (d) Bachrach, S. M. *J. Phys. Chem. A* **2008**, *112*, 3722–3730;

(10)   Baik, M. H.; Friesner, R. A.; Lippard, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 4495–4503.

(11)   (a) Chuchev, K.; BelBruno, J. J. *J. Mol. Struct. (Theochem)* **2008**, *850*, 111–120; (b) Michaux, C.; Wouters, J.; Perpète, E. A.; Jacquemin, D. *J. Phys. Chem. B* **2008**, *112*, 7702–7705.

(12)   (a) Manojkumar, T. K.; Suh, S. B.; Oh, K. S.; Cho, S. J.; Cui, C. Z.; Zhang, X.; Kim, K. S. *J. Org. Chem.* **2005**, *70*, 2651–2659; (b) Almatarneh, M. H.; Flinn, C. G.; Poirier, R. A.; Sokalski, W. A. *J. Phys. Chem. A* **2006**, *110*, 8227–8234; (c) Almerindo, G. I.; Pliego Jr., J. R. *J. Braz. Chem. Soc.* **2007**, *18*, 696–702; (d) Wu, Y.; Xue, Y.; Xie, D. Q.; Kim, C. K.; Yan, G. S. *J. Phys. Chem. B* **2007**, *111*, 2357–2364; (e) Wu, Y.; Jin, L.; Xue, Y.; Xie, D. Q.; Kim, C. K.; Guo, Y.; Sen Yan, G. *J. Comput. Chem.* **2008**, *29*, 1222–1232; (f) Gao, J. Y.; Zeng, Y.; Zhang, C. H.; Xue, Y. *J. Phys. Chem. A* **2009**, *113*, 325–331.

(13)   (a) Kallies, B.; Mitzner, R. *J. Mol. Model.* **1998**, *4*, 183–196; (b) Zhan, C. G.; Landry, D. W.; Ornstein, R. L. *J. Am. Chem. Soc.* **2000**, *122*, 2621–2627; (c) Tautermann, C. S.; Voegele, A. F.; Loerting, T.; Kohl, I.; Hallbrucker, A.; Mayer, E.; Liedl, K. R. *Chem. Euro. J.* **2002**, *8*, 66–73; (d) Zhang, L. D.; Xie, D. Q.; Xu, D. G.; Guo, H. *Chem. Commun.* **2007**, 1638–1640; (e) Zeng, Y.; Xue, Y.; Yan, G. *J. Phys. Chem. B* **2008**, *112*, 10659–10667.

(14)   See, for example, (a) Schramm, V. L. *Arch. Biochem. Biophys.* **2005**, *433*, 13–26; (b) Chen, S. L.; Fang, W. H.; Himo, F. *Theor. Chem. Acc.* **2008**, *120*, 515–522; (c) Hopmann, K. H.; Himo, F. *J. Chem. Theory Comput.* **2008**, *4*, 1129–1137; (d) Ramos, M. J.; Fernandes, P. A. *Acc. Chem. Res.* **2008**, *41*, 689–698.

(15)   Rosta, E.; Kamerlin, S. C. L.; Warshel, A. *Biochemistry* **2008**, *47*, 3725–3735.

(16)   Gaussian 03, Revision D.02, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc., Wallingford CT, **2004**.

(17)   Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

(18)   HyperChem™ Professional 7.5; Hypercube, Inc.; 1115 NW 4th Street, Gainesville, Florida 32601, USA.

(19)   (a) Brown, T. N.; Mora-Diez, N. *J. Phys. Chem. B* **2006**, *110*, 9270–9279, (b) Brown, T. N.; Mora-Diez, N. *J. Phys. Chem. B* **2006**, *110*, 20546–20554.

(20)   Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*; Wiley: Chichester, West Sussex, England, **2004**.

(21)   (a) Liu, P. F.; Burdzy, A.; Sowers, L. C. *Chem. Res. Toxicol.* **2002**, *15*, 1001–1009; (b) Pettersen, H. S.; Sundheim, O.; Gilljam, K. M.; Slupphaug, G.; Krokan, H. E.; Kavli, B. *Nucleic Acids Res.* **2007**, *35*, 3879–3892.

(22)   (a) Werner, R. M.; Stivers, J. T. *Biochemistry* **2000**, *39*, 14054–14064; (b) Bianchet, M. A.; Seiple, L. A.; Jiang, Y. L.; Ichikawa, Y.; Amzel, L. M.; Stivers, J. T. *Biochemistry* **2003**, *42*, 12455–12460; (c) Jiang, Y. L.; Cao, C.; Stivers, J. T.; Song, F.; Ichikawa, Y. *Bioorg. Chem.* **2004**, *32*, 244–262.

(23)   For a review of the importance of water activation, see: Anderson, J. A.; Hopkins, B. W.; Chapman, J. L.; Tschumper, G. S. *J. Mol. Struct. (THEOCHEM)* **2006**, *771*, 65–71.

(24)   See, for example, (a) Krokan, H. E.; Standal, R.; Slupphaug, G. *Biochem. J.* **1997**, *325*, 1–16; (b) Xu, D. G.; Guo, H. *FEBS Lett.* **2005**, *579*, 4249–4253; (c) Deng, H.; Botting, C. H.; Hamilton, J. T. G.; Russell, R. J. M.; O'Hagan, D. *Angew. Chem.-Int. Ed.* **2008**, *47*, 5357–5361; (d) Hadler, K. S.; Tanifum, E. A.; Yip, S. H. C.; Mitic, N.; Guddat, L. W.; Jackson, C. J.; Gahan, L. R.; Nguyen, K.; Carr, P. D.; Ollis, D. L.; Hengge, A. C.; Larrabee, J. A.; Schenk, G. *J. Am. Chem. Soc.* **2008**, *130*, 14129–14138; (e) Zha, M. W.; Guo, Q.; Zhang, Y. C.; Yu, B.; Ou, Y.; Zhong, C.; Ding, J. P. *J. Mol. Biol.* **2008**, *379*, 568–578.

(25)   (a) Stivers, J. T.; Jiang, Y. L. *Chem. Rev.* **2003**, *103*, 2729–2759; (b) Berti, P. J.; McCann, J. A. B. *Chem. Rev.* **2006**, *106*, 506–555.

(26)   For an excellent review on uracil in DNA, see: Sousa, M. M. L.; Krokan, H. E.; Slupphaug, G. *Mol. Asp. Med.* **2007**, *28*, 276–306.

(27)   (a) Nilsen, H.; Stamp, G.; Andersen, S.; Hrivnak, G.; Krokan, H. E.; Lindahl, T.; Barnes, D. E. *Oncogene* **2003**, *22*, 5381–5386; (b) Studebaker, A. W.; Ariza, M. E.; Williams, M. V. *Biochem. Biophys. Res. Commun.* **2005**, *334*, 509–515.

(28)   Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Phys. Chem. A* **2007**, *111*, 10439–10452.

Chapter 3: Hydrolysis of the Natural 2′-Deoxyribonucleosides

3.1 Introduction

Apart from depurination of dG,[1] very few studies in the literature have focused on the hydrolysis of natural dR nucleosides. Furthermore, the experimental trend in the Gibbs energy of activation with respect to the natural dR nucleosides (dG < dU ~ dT < dA < dC)[2] has yet to be reproduced computationally. For example, Millen and Wetmore[3] carried out a gas-phase DFT study on the hydrolysis of the natural dR nucleotides, where solvent effects were accounted for via PCM (water) single-point energy corrections. This study predicted the correct activation energy trend for four of the standard dR nucleotides (dU ~ dT < dA < dC), but dG was found to have the second highest activation energy instead of the lowest. Nevertheless, this trend in deglycosylation barriers matches the trend in the calculated N1 (pyrimidine) and N9 (purine) gas-phase acidities.[4] In addition, the Millen and Wetmore study only examined the concerted, S_N2, mechanism for hydrolytic deglycosylation since the initial dU hydrolysis study predicted a large S_N1 barrier.[5] Therefore, this study reports an incorrect mechanism compared with the experimentally predicted S_N1 mechanism for all 2′-deoxyribonucleosides.[2] It should be noted that this is another example of gas-phase calculations agreeing with experimental energies, but predicting completely different mechanisms.

Since the newly derived model for dU hydrolysis presented in Chapter 2 for the first time successfully predicts the correct hydrolysis mechanism for dU, it is of interest to determine whether the same theoretical methodology can rectify the discrepancies between computational and experimental results for the remaining dR nucleosides. Therefore, the two-fold approach developed in Chapter 2, which includes hybrid

solvation and two-dimensional PES scans, will be applied to the four canonical dR

nucleosides (dA, dC, dG and dT). This will not only act as a test for the broader

applicability of the model developed in Chapter 2, but will aid in uniting theory and

experiment, since neither the energetics nor mechanism for the hydrolytic

deglycosylation of the 2′-deoxyribonucleosides has yet to be correctly predicted by

calculations.

3.2 Computational Details

In order to accurately test the applicability of the hydrolysis model developed in

Chapter 2, the PCM-B3LYP/6-31+G(d) solvent-phase reaction PES for the hydrolysis of

the four canonical dR nucleosides were generated in the same fashion as discussed for dU

in Section 2.2. Appendix B contains a detailed description of how each surface was

generated. The Gibbs energy surfaces were calculated as discussed in Section 2.2.3.[6] All

calculations were carried out with Gaussian 03,[7] except for the Monte Carlo simulations,

which were performed using HyperChem 7.5.[8] The reported product complexes contain

non-standard nucleobase tautomers. The tautomerization of these product complexes to

the global minimum has been investigated elsewhere, and thus is not discussed here.[9] The

reactant complexes for the pyrimidines, dC and dT, were prepared by simply mutating

the uracil base in the dU reactant to either cytosine or thymine and relaxing the resultant

structure to a minimum. However, the reactant complexes for the purines were generated

separately, and therefore will be described below.

The imidazole ring of dA and dG increases the distance compared with dU that

the discrete hydrogen-bonding network must span between the nucleophilic water and the

nucleobase to properly describe nucleophile activation and leaving group stabilization.

For example, the distance between anomeric carbon (C1′) and the hydrogen acceptor (O2 (dU) or N3 (dA)) increases from 2.802 Å in dU to 3.127 Å in dA (Figure 3.1). These changes suggest an additional water molecule is needed in the computational model. Therefore, dA and dG models including four explicit water molecules, which form a hydrogen-bonding network between O3′ in the sugar moiety and N3 of the purine, were generated. C2′-*endo* sugar puckering was utilized to be consistent with the pyrimidine systems. The *anti* orientation of the base (the Watson-Crick binding face pointed away from the sugar) in the nucleoside was utilized to allow for the hydrogen-bonding network to N3. In addition, solvent-phase studies have shown that the *anti* conformation is lower in energy in solution,[10] and therefore the *syn* conformer was not examined. As for the dU study, the water molecules were initially added by inspection, and the 3′ and 5′ hydroxyl groups were positioned similar to the dU reactant. Subsequently, Monte Carlo simulations were carried out, where the geometries selected for further optimization fell within a 10 kJ mol^{-1} range (instead of a 5 kJ mol^{-1} range as implemented for dU). Additional dA and dG structures were optimized at the DFT level compared with the number considered for dU since the extra water molecule introduces additional degrees of freedom. After optimization with B3LYP, the structure with the lowest energy and largest grouping of similar structures (i.e., the largest number of AMBER structures that optimized to a particular B3LYP structure) was selected as the starting reactant for the PES scans.
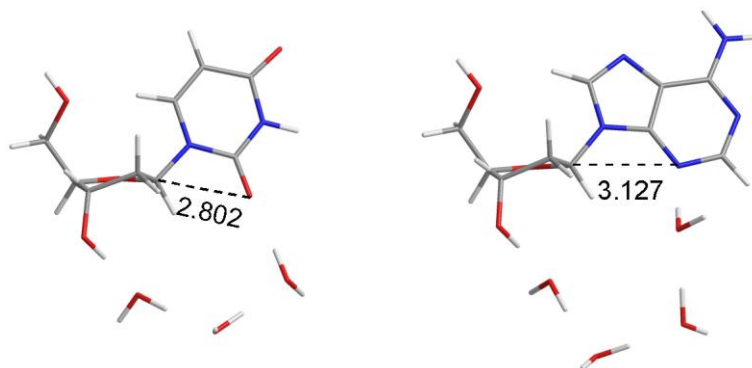
Figure 3.1 *Structural differences between the pyrimidines and the purines that affect the size of the discrete water hydrogen-bonding network included in the computational model.*

3.3 Results and Discussion

Since different computational models were implemented, hydrolysis of the natural pyrimidines, dC and dT, and purines, dA and dG, will be discussed separately below.

3.3.1 Pyrimidines

Due to their analogous structures, the canonical pyrimidines behaved similarly to dU. In all cases, the potential and Gibbs energy surfaces indicate that the stepwise, S_N1, mechanism is lower in energy than the concerted, S_N2, mechanism, which is in agreement with empirical evidence for the first time. Therefore only the S_N1 mechanism will be discussed explicitly.

The three pyrimidines adopt similar structures throughout most of the S_N1 reaction coordinate. In particular, dissociation of the base is stabilized by a hydrogen bond between C1′–H and O2, and there is full proton transfer to O2 from a water molecule once the nucleophile is within 1.8 Å of C1′ (Figure 3.2). The proton transfer through the hydrogen-bond network occurs in two steps. Initially, an O–H bond in the nucleophilic water is elongated when the nucleophilic distance is near 1.8 Å. When the attacking distance decreases to 1.6 Å, transfer through the remaining water molecules

occurs to yield either partial or full transfer to O2 of the pyrimidine anion (depending on the N1–C1′ distance). At a nucleophilic distance of 1.4 Å, full transfer has occurred, and the remaining hydroxyl adds to C1′ of the sugar moiety. Other interesting features of the dC and dT reaction surfaces will be discussed below.
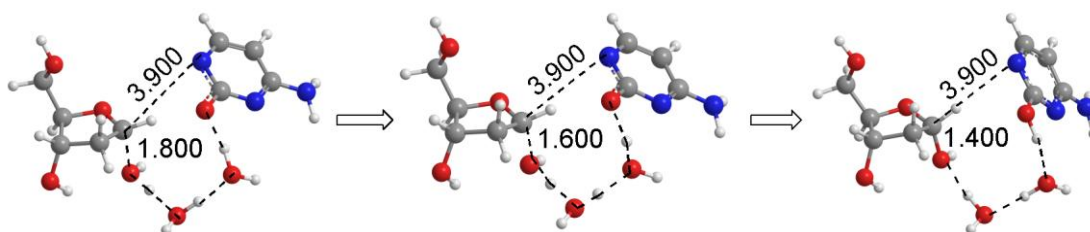
Figure 3.2   *An example (dC) of the proton transfer from the nucleophilic water molecule to O2 of the pyrimidines as the nucleophilic distance is reduced. Selected PCM-B3LYP/6-31+G(d) distances displayed in Å.*

3.3.1.1 Hydrolysis of dC

The potential and Gibbs energy surfaces for the hydrolysis of dC with three explicit water molecules are shown in Figures 3.3 A and B, respectively. It can be seen that these surfaces are similar in shape to the dU surfaces (Figures 2.9 and 2.10); however, the reported relative energies are higher for dC. Specifically, the path of slowest ascent leads to a dissociative transition state (TS$_D$, Figure 3.3C) with a barrier of ~140 kJ mol^{-1} (~110 kJ mol^{-1} on the Gibbs surface), which is approximately 20 kJ mol^{-1} larger than the calculated $\Delta E^{\ddagger}$, and 10 kJ mol^{-1} larger than $\Delta G^{\ddagger}$, reported for dU in Chapter 2. Beyond the dissociative barrier, the reaction progresses to a shallow intermediate (I, Figure 3.3C) and, similar to dU, passes through a small associative barrier before reaching the product complex (P, Figure 3.3C).

The PESs report larger ΔE values for dC than for dU since an electron withdrawing (C4) carbonyl in U is replaced with an electron donating amine group in C, which makes C a poorer leaving group than U. This effect is evident upon analysis of the
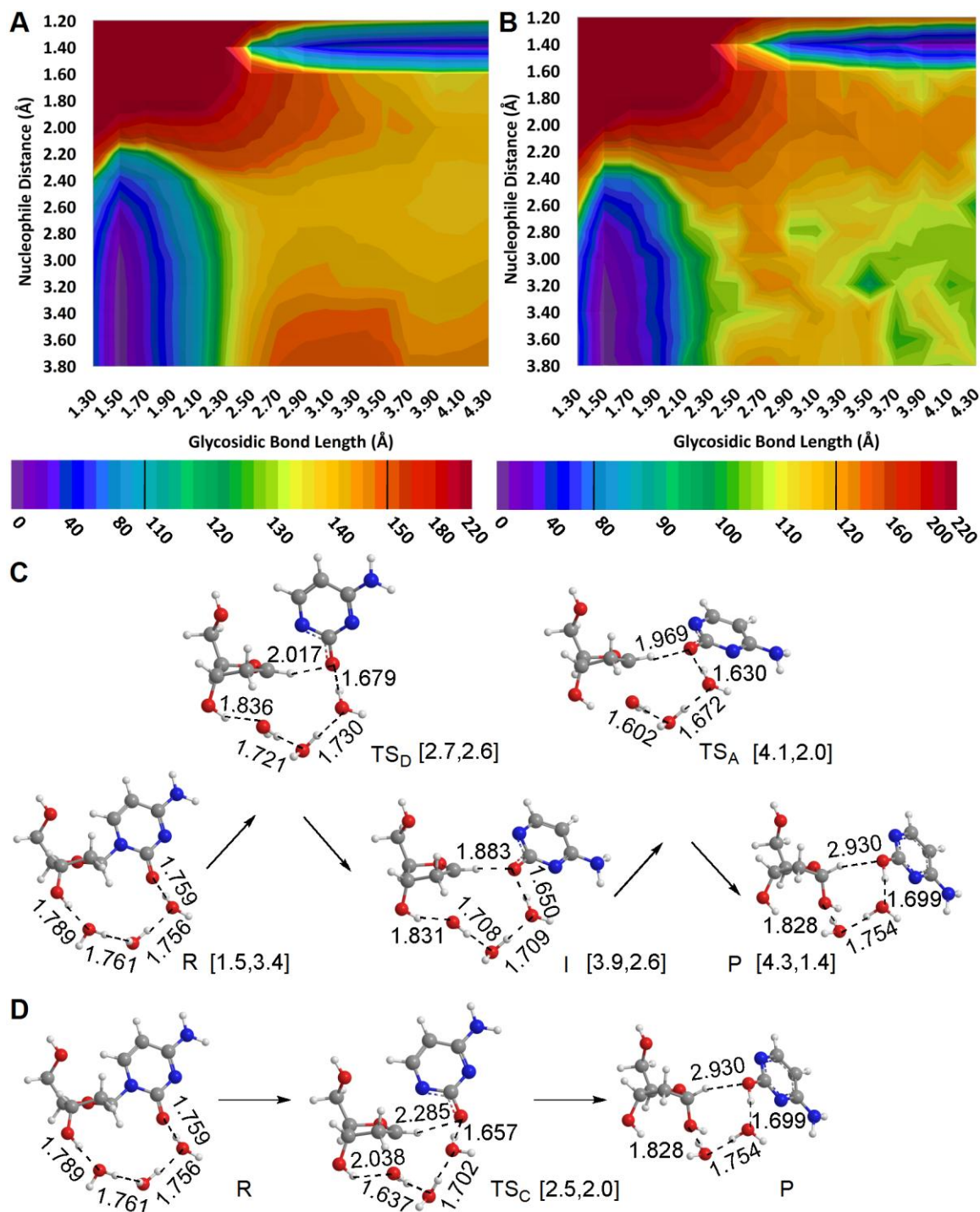
Figure 3.3 *The potential (**A**) and Gibbs (**B**) energy surfaces for the hydrolysis of dC using a model with three explicit water molecules, where representative structures determined from the PES that exhibit S_N1 (**C**) or S_N2 (**D**) character are shown. (Relative energies in kJ mol⁻¹, colour scale displayed in legend, and select PCM-B3LYP/6-31+G(d) distances (Å) provided including glycosidic bond length and nucleophile distance represented as [C1'···N1, C1'···O_w].)*

calculated N1 proton affinities of the two nucleobases (1391.0 and 1444.5 kJ mol^{-1} for U

and C, respectively).[4] With the exception of the rate-limiting barrier, the dC surfaces and

corresponding structures agree with those for dU. Experimental studies on the hydrolysis

of dR nucleosides have determined dC to be more stable than dU, and therefore these

results coincide with previous findings.[2] In addition to these energetic differences, the dC

hydrolysis product involves a C1′–H···O2 hydrogen bond, rather than an O1′–H···N1

hydrogen bond with the new hydroxyl group at C1′ as found in the dU product. This may

be due to O2 interacting with the sugar moiety rather than N1 in the associative TS, since

the O2-protonated cytosine tautomer is more stable than the N1-protonated tautomer.[11]

3.3.1.2 Hydrolysis of dT

The PES for the hydrolysis of thymidine (Figure 3.4A) using a model with three

explicit water molecules closely resembles the analogous dU and dC hydrolysis surfaces

(Figures 2.9 and 3.3A). Specifically, the path of slowest ascent corresponds to a stepwise,

S_N1, mechanism, which contains a very shallow intermediate region that is more easily

identified on the Gibbs energy surface (Figure 3.4B). The barrier for the dissociation step

is calculated to be ~129 kJ mol^{-1} (~95 kJ mol^{-1} on the Gibbs surface), which is very close

to the energetics reported in Chapter 2 for dU hydrolysis ($\Delta E^{\ddagger}$ = 120 kJ mol^{-1}, $\Delta G^{\ddagger}$ = 100

kJ mol^{-1}).[5] This result agrees with experimental evidence that dU and dT deglycosylate at

nearly the same rate, and that both are more labile than dC.[2,12] The relative $\Delta E^{\ddagger}$ for the

three pyrimidines (dU < dT < dC) follow with their calculated relative N1 acidities.[13]

The reason for the slight increase in activation energy for hydrolysis of dT

compared to dU can be understood by examining the structures along the reaction

coordinate (Figures 3.4C and D). Unlike dU, as the glycosidic bond length increases, N1

67
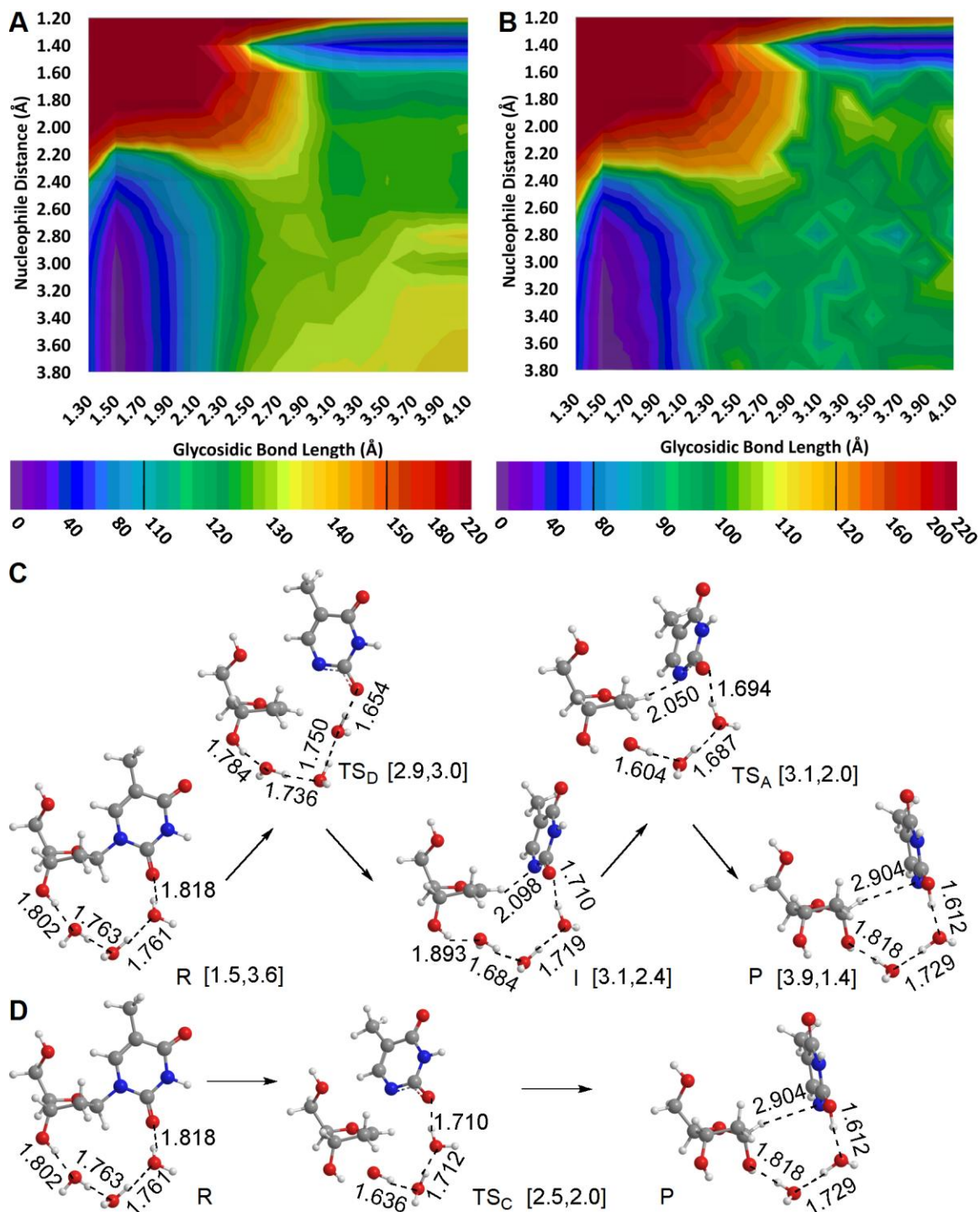
Figure 3.4 *The potential (**A**) and Gibbs (**B**) energy surfaces for the hydrolysis of dT using a model with three explicit water molecules, where representative structures determined from the PES that exhibit S_N1 (**C**) or S_N2 (**D**) character are shown. (Relative energies in kJ mol^{-1}, colour scale displayed in legend, and select PCM-B3LYP/6-31+G(d) distances (Å) provided including glycosidic bond length and nucleophile distance represented as [C1'···N1, C1'···O$_w$].)*

migrates over the C1′–H bond to form a C1′–H···N1 interaction (TS$_D$, I, Figure 3.4C). Although a similar base–sugar interaction occurs in the purine models as discussed in Section 3.3.2, this configuration does not afford as much charge stabilization as seen for dU. Therefore, the association step of the mechanism occurs at a shorter glycosidic bond distance for dT than dU (TS$_A$, Figure 3.4), as well as dC. Excluding these two changes, the remainder of the mechanism and surface is the same as discussed for dU.

In summary, the model designed in Chapter 2 for the hydrolysis of the dU nucleoside correctly predicts the deglycosylation mechanism for all three standard pyrimidines. Furthermore, this model predicts the experimental free energy trend for the pyrimidines. Since previous literature was unable to predict the mechanism, this shows the strength of the new model presented in this thesis.

3.3.2 Purines

Parallel to the pyrimidines, the path of slowest ascent for the purines corresponds to a dissociative, S$_N$1, mechanism. Therefore, this thesis presents the first computational model to agree with experimental findings.[2] Along the S$_N$1 pathway the purines migrate across the C1′–H bond, which replaces the N9–C1′ glycosidic bond with a C1′–H···N9 hydrogen bond (Figure 3.5, middle). Subsequently, as the base anion moves farther away from the sugar moiety, the water molecule hydrogen bound to N3 moves closer to C1′ and generates a bridge between C1′–H and N3 (Figures 3.5, right and 3.6, TS$_C$). This bridge permits the base to rotate about the O–H···N3 hydrogen bond and thereby allows N9 to interact with the C1′-hydroxyl group in the product (see, for example, Figure 3.6, P). Previous studies on dG deglycosylation have found similar water-bridged structures.[1]

In addition, at extended glycosidic and nucleophilic distances, water bridged structures

can be identified for the pyrimidines.
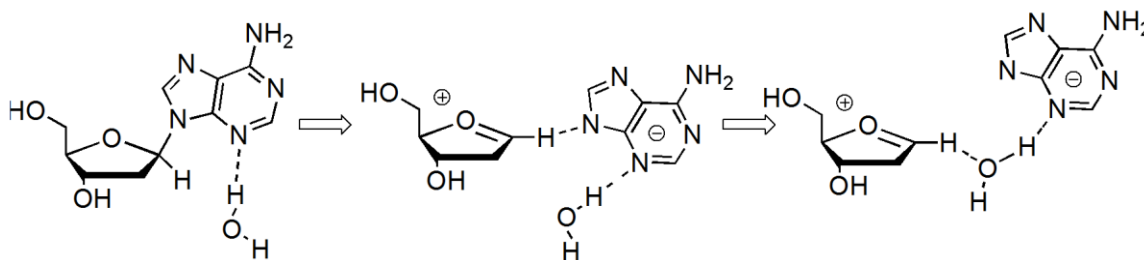
Figure 3.5   *A schematic depicting how the water molecule near N3 of the purines moves toward the sugar upon elongation of the glycosidic bond to yield a bridged structure. (The remaining three explicit water molecules are not shown.)*

Despite the difference outlined above, the remainder of the reaction mechanism is

very similar between the purines and the pyrimidines. Specifically, dissociation of the

base leads to an intermediate with a planar oxacarbenium cation moiety containing a

slightly elongated C4′–O4′ bond length, and a hydrogen bond between C1′–H and the

base anion. The associative transition state occurs before proton transfer from the

nucleophilic water molecule to the base (O2 (pyrimidines) or N3 (purines)) has occurred,

where the transfer is barrierless. In the product complex, the base anion interacts with the

C1′-hydroxyl group to complete the hydrogen-bond network. Energetics for dA and dG

hydrolysis will be discussed in the following two subsections.

3.3.2.1 Hydrolysis of dA

From the PES (Figure 3.6A), the activation energy for the dissociation of dA is

calculated to be ~134 kJ mol^{-1}, while the relevant $\Delta G^{\ddagger}$ is ~115 kJ mol^{-1}. The intermediate

(I, Figure 3.6C) is found in a shallow well and is ~6 kJ mol^{-1} more stable than the

dissociative TS. The associative TS (TS$_A$, Figure 3.6C) occurs at an energy of ~5 kJ

mol^{-1} relative to the intermediate, or 133 kJ mol^{-1} relative to the reactant. As before, the
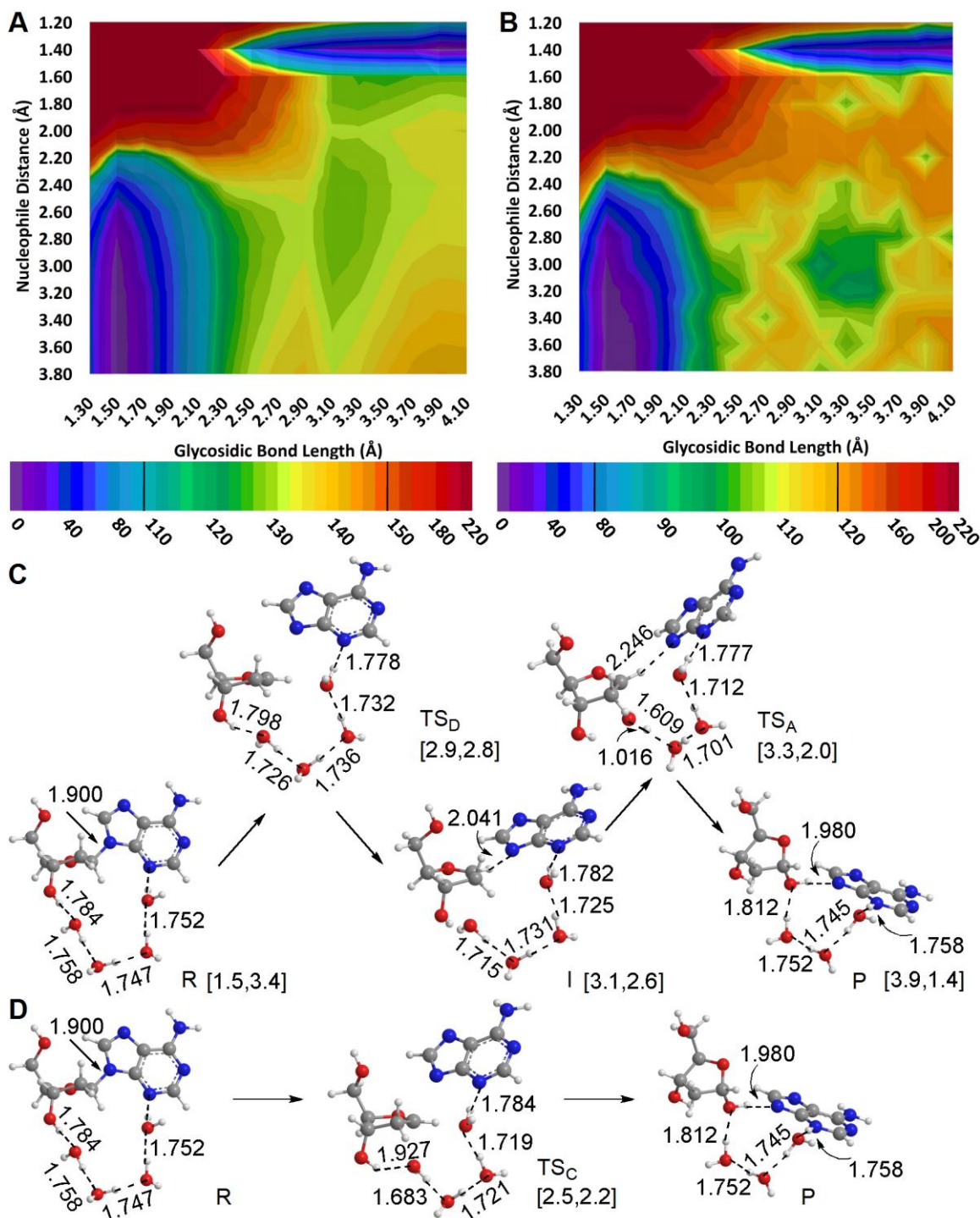
Figure 3.6   *The potential (**A**) and Gibbs (**B**) energy surfaces for the hydrolysis of dA using a model with four explicit water molecules, where representative structures determined from the PES that exhibit S_N1 (**C**) or S_N2 (**D**) character are shown. (Relative energies in kJ mol^{-1}, colour scale displayed in legend, and select PCM-B3LYP/6-31+G(d) distances (Å) provided including glycosidic bond length and nucleophile distance represented as [C1'···N1, C1'···Ow].)*

calculated Gibbs barrier is similar to the value reported by Schroeder et al. (~129 kJ mol^{-1}).[2] However, even though the calculated Gibbs energy of activation is near the experimental value, the calculated barrier for dA hydrolysis is too large compared to that calculated for dC. Therefore, the calculated trend across the pyrimidines and dA does not match the trend seen experimentally. In contrast, the corresponding trend in the potential energies is correct, which implies that the discrepancy in the Gibbs energy may arise because the additional water molecule in the dA hydrolysis model introduces errors in the thermal and entropy corrections.

3.3.2.2 Hydrolysis of dG

The reaction potential energy surface for the hydrolysis of dG contains the same shape as the surface for dA hydrolysis. The most obvious difference is the height of dissociative barrier (Figure 3.7A). The calculated activation energy for dG ($\Delta E^{\ddagger}$ = ~138 kJ mol^{-1}) is larger than that of dA; however, the Gibbs energy surface reports the opposite trend ($\Delta G^{\ddagger}$ = ~108 kJ mol^{-1} for dG, compared to ~115 kJ mol^{-1} for dA). The trend in the calculated Gibbs barriers result is in agreement with experiment, which indicates dG deglycosylation is faster than dA under neutral conditions.[2,12] However, although the free energy trend among the purines is correct, experiment finds dG to hydrolyze faster than any other 2′-deoxyribonucleoside, while calculations determine dG to be the third easiest nucleoside to hydrolyze.

A potential reason for the large dG deglycosylation barrier relative to other nucleosides can be seen in Figure 3.7. In the reactant (R) and dissociative transition state (TS$_D$), the hydrogen-bond distance between a water molecule and N3 is quite large (> 1.8 Å) compared to a standard O–H···N hydrogen bond involving water. The long hydrogen
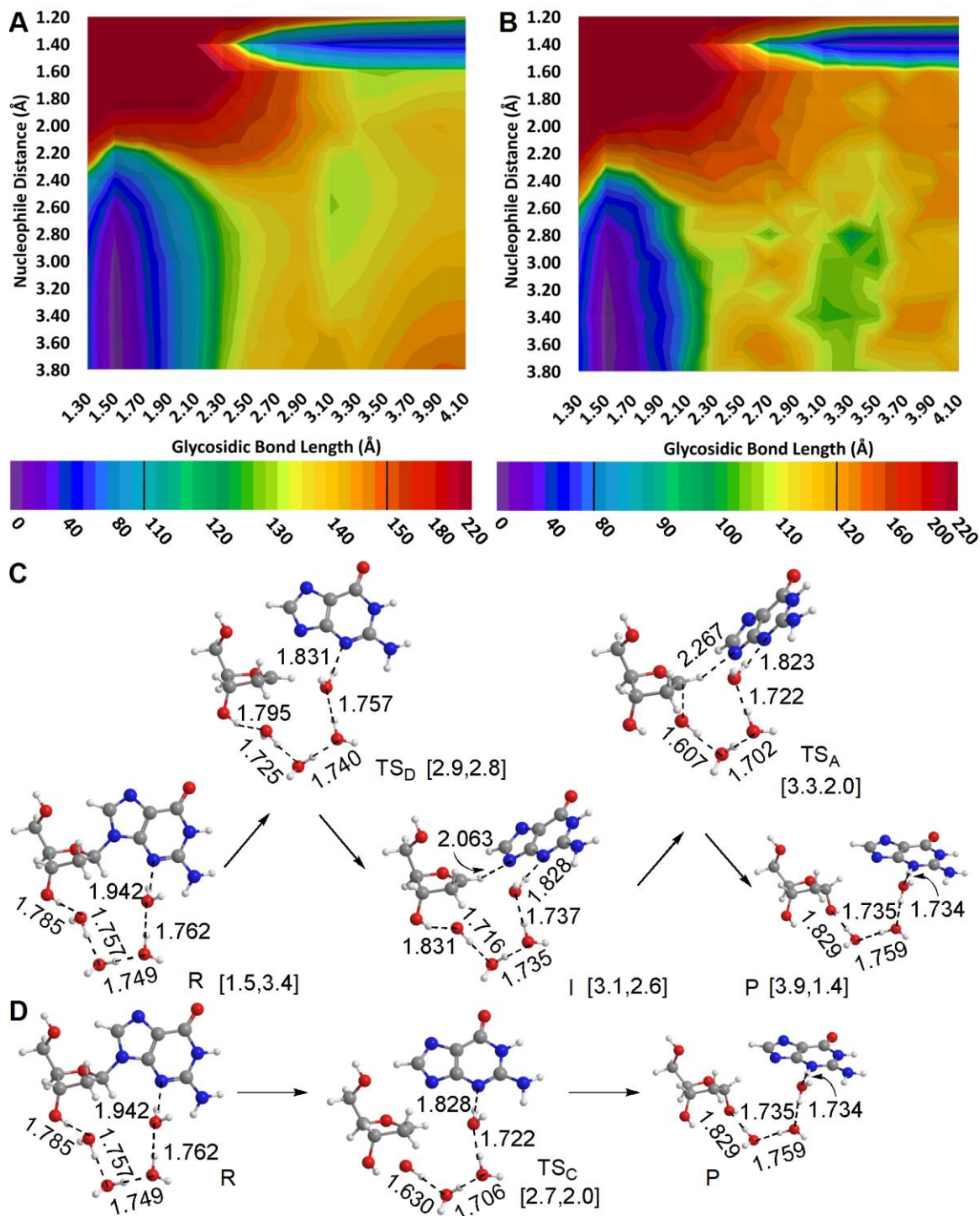
Figure 3.7 *The potential (**A**) and Gibbs (**B**) energy surfaces for the hydrolysis of dG using a model with four explicit water molecules, where representative structures determined from the PES that exhibit S_N1 (**C**) or S_N2 (**D**) character are shown. (Relative energies in kJ mol^{-1}, colour scale displayed in legend, and select PCM-B3LYP/6-31+G(d) distances (Å) provided including glycosidic bond length and nucleophile distance represented as [C1'⋯N1, C1'⋯O$_w$].)*

bond in the dG structures decreases the amount of charge stabilization to the nucleobase anion in the dissociation step, and therefore increases the barrier height compared to other 2′-deoxyribonucleosides. The weak interaction at N3 is due to the relatively low proton affinity of this site, where protonation at N7 is calculated to be much more favourable.[4]

Although the model for dR hydrolysis designed in Chapter 2 for dU cannot correctly predict the trend in experimental hydrolysis barriers across all nucleosides, it can correctly predict the mechanism of purine nucleoside hydrolysis. In addition, the calculated trend in purine hydrolysis free energy barriers matches that predicted from experiment.

3.4 Conclusions

In this Chapter, the model for nucleoside hydrolysis developed in Chapter 2 for dU was applied to the remaining (dC, dT, dA and dG) 2′-deoxyribonucleosides. In all cases, it was found, that the hydrolysis occurs via a stepwise, S_N1, mechanism, where the rate-limiting step is dissociation of the nucleobase from the sugar moiety. This is the first time that the correct reaction mechanism has been computationally predicted for the (neutral) hydrolysis of the 2′-deoxyribonucleosides.

A previous gas-phase study by Millen and Wetmore found the trend in the calculated S_N2 ($\Delta E^{\ddagger}$) barriers to be dU ~ dT < dA < dG < dC,[3] which corresponds with the calculated $\Delta E^{\ddagger}$ trend reported in this thesis. The similar trend calculated for the gas-phase S_N2 and solvent-phase S_N1 potential energies once again highlights one of the central ideas of this thesis. Specifically, gas-phase calculated energies may agree with experiment, but the associated structures, and therefore mechanistic predictions, may be

very different. Therefore, this thesis shows that, in the case of dissociative mechanisms, solvent effects (implicit and explicit) should be included during optimizations.

The trend in the calculated $\Delta G^{\ddagger}$ for the nucleosides approximated from the dissociative portion of the Gibbs energy surface (dT < dU < dG < dC < dA) is slightly different from the trend in the calculated $\Delta E^{\ddagger}$. Furthermore, the model utilized in this study correctly predicts the experimental deglycosylation trends for the pyrimidines and the purines, separately. However, the free energies of activation for hydrolysis of the purines are shifted slightly higher compared with experiment. Since the trend in the deglycosylation rate of the standard dR nucleosides changes upon moving from the PES to the free energy surface, the discrepancy between calculations and experiment may come from the calculated thermal and entropy corrections. Furthermore, additional errors may be due to the additional degrees of freedom associated with the extra water molecule in the purine model compared to the pyrimidine model. In addition, the large calculated barrier for dG hydrolysis may be due to the small proton affinity of the N3 site which makes a hydrogen bonding network to N3 unfavourable.[14]

Based on the results presented in this Chapter, it can be proposed that the two-fold model for nucleoside hydrolysis utilized in this thesis is applicable to other 2′-deoxyribonucleosides, and may be applied to other bases, including damaged nucleosides. In the case of pyrimidine nucleosides, the hydrogen-bond network should bridge the nucleophilic water to O2, while purine nucleosides should bridge the water to N3 or possibly N7 depending on their relative proton affinities. In all cases, implicit solvent must also be included in the optimization routine and the Gibbs energy surface should be considered.

3.5 References

(1)   (a) Baik, M. H.; Friesner, R. A.; Lippard, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 4495–4503; (b) Rios-Font, R.; Rodriguez-Santiago, L.; Bertran, J.; Sodupe, M. *J. Phys. Chem. B* **2007**, *111*, 6071–6077.

(2)   Schroeder, G. K.; Wolfenden, R. *Biochemistry* **2007**, *46*, 13638–13647.

(3)   Millen, A. L.; Wetmore, S. D. *Can. J. Chem.* **2009**, *87*, 850–863.

(4)   Chandra, A. K.; Nguyen, M. T.; Uchimaru, T.; Zeegers-Huyskens, T. *J. Phys. Chem. A*, **1999**, *103*, 8853–8860.

(5)   Millen, A. L.; Archibald, L. A. B.; Hunter, K. C.; Wetmore, S. D. *J. Phys. Chem. B* **2007**, *111*, 3800-3812.

(6)   Przybylski, J. L.; Wetmore, S. D. *J. Phys. Chem. B* **2009**, *113*, 6533–6542.

(7)   Gaussian 03, Revision D.02, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc., Wallingford CT, **2004**.

(8)   HyperChem™ Professional 7.5; Hypercube, Inc.; 1115 NW 4th Street, Gainesville, Florida 32601, USA.

(9)   (a) Gorb, L.; Leszczynski, J. *Int. J. Quantum Chem.* **1998**, *70*, 855–862; (b) Gu, J.; Leszczynski, J. *J. Phys. Chem. A* **1999**, *103*, 2744–2750; (c) Ahn, D.-S.; Lee, S.; Kim, B. *Chem. Phys. Lett.* **2004**, *390*, 384–388; (d) Hu, X.; Li, H.; Liang, W.; Han, S. *J. Phys. Chem. B* **2005**, *109*, 5935–5944.

(10)   See, for example,: Foloppe, N.; Nilsson, L. *J. Phys. Chem. B* **2005**, *109*, 9119–9131.

(11)   (a) Sambrano, J. R.; de Souza, A. R.; Queralt, J. J.; Andrés, J. *Chem. Phys. Lett.* **2000**, *317*, 437–443; (b) Trygubenko, S. A.; Bogdan, T. V.; Rueda, M.; Orozco,

M.; Luque, F. J.; Šponer, J.; Slaviček, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4192–4203; (c) Alparone, A.; Millefiori, A.; Millefiori, S. *Chem. Phys.* **2005**, *312*, 261–274.

(12)   (a) Garrett, E. R.; Seydel, J. K.; Sharpen, A. J. *J. Org. Chem.* **1966**, *31*, 2219–2227; (b) Shapiro, R.; Kang, S. *Biochemistry* **1969**, *8*, 1809–1810; (c) Shapiro, R.; Danzig, M. *Biochemistry* **1972**, *11*, 23–5154.

(13)   (a) Di Luado, M.; Whittleton, S. R.; Wetmore, S. D. *J. Phys. Chem. A* **2003**, *107*, 10406–10413; (b) Chandra, A. K.; Michalska, D.; Wysokiñsky, R.; Zeegers-Huyskens, T. *J. Phys. Chem. A* **2004**, *108*, 9593–9600; (c) Whittleton, S. R.; Hunter, K. C.; Wetmore, S. D. *J. Phys. Chem. A* **2004**, *108*, 7709–7718; (d) Rejnek, J.; Hanus, M.; Kabeláč, M.; Ryjáček, F.; Hobza, P. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2006–2017; (e) Jiao, D.; Wang, H.; Zhang, Y.; Tang, Y. *Can. J. Chem.* **2009**, *87*, 406–415.

(14)   Ratsep, P. C.; Pless, R. C. *J. Org. Chem.* **1988**, *53*, 3241–3246.

Chapter 4: Mechanism of Action of Uracil DNA Glycosylase

4.1 Introduction

As discussed in Chapter 1, there have been very few computational studies on the UDG catalyzed deglycosylation of dU,[1-3] and disparity exists not only between experiment and theory, but also amongst the theoretical work. For example, the protonation state of His148 has recently been addressed in the literature,[4] but no definitive conclusion as to whether this residue acquires a charge has been drawn. Therefore, there are two main goals of this chapter: 1) use the conclusions of Chapter 2 and the ONIOM[5] methodology to design a model that is capable of studying the mechanism of UDG action; and 2) use this model to determine the role and protonation state of His148.

The first step to accomplishing these goals involves determining the most simplistic representation of the enzyme-substrate complex capable of reproducing the currently accepted steps in the mechanism and catalytic efficiency of UDG. An example of a concept that must be considered in developing the model is the amount of truncation that may be used to approximate the system. Since a large portion of the enzyme is not directly involved in the reaction, but has other biological roles (i.e., binding sites for proteins that form complexes with UDG),[6] the computational model can likely be reduced to a small portion of the total complex. However, a difficult question is how much of the complex can be removed while maintaining the functionality of the enzyme.

Chapter 2 provides important clues for designing a truncated computational model

of UDG. It was found that the minimal model for dU hydrolysis includes proton transfer

from the nucleophile to the uracil anion, which provides leaving group stabilization and

water nucleophile activation. Therefore, a good starting point for a minimal enzyme

model is including groups that offer the same chemistry. In the case of leaving group

stabilization, this amounts to including a crystallographic water near O4, an asparagine

residue (Asn204) near O4 and N3, and a histidine residue (His268) hydrogen bound to

O2 of uracil (Figure 4.1A). Water nucleophile activation is less trivial since there are two

residues that can accept a hydrogen bond (namely Asp145 and Pro146), and a residue

(His148) that can either accept or donate a hydrogen bond depending on the protonation
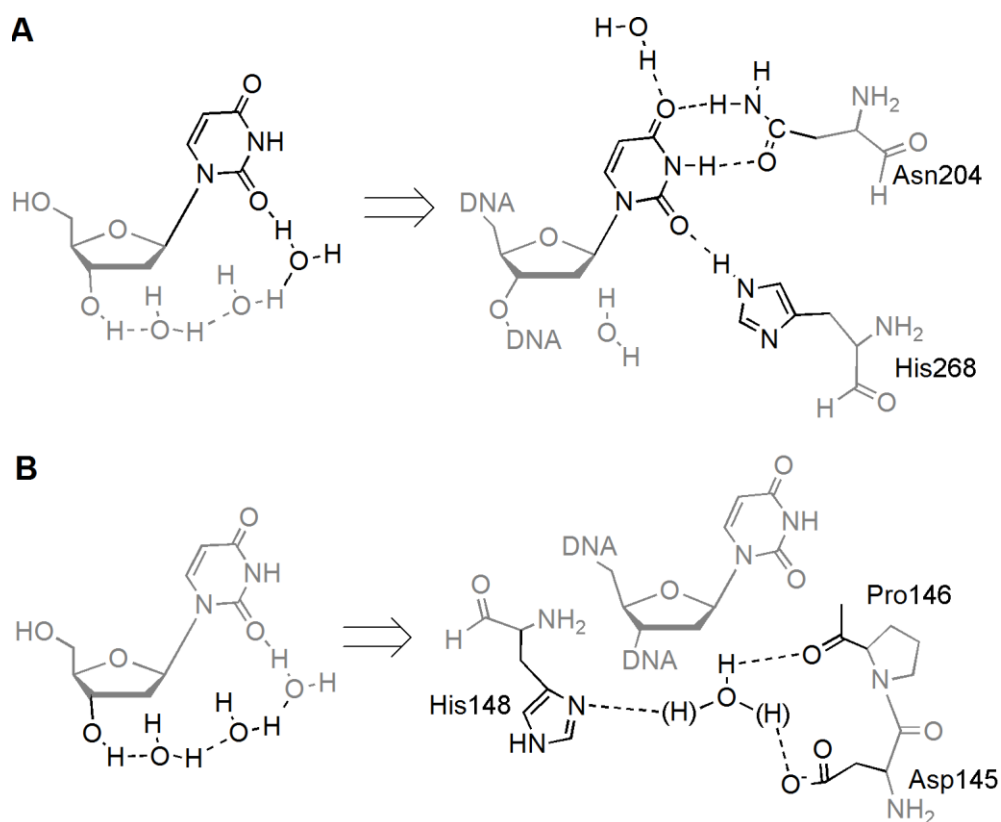
Figure 4.1 *Active-site residues in UDG that provide leaving group stabilization (**A**) and nucleophile activation (**B**) required by the uncatalyzed hydrolysis model.*

state (Figure 4.1B). Therefore, all of these residues must be included in the model, and a variety of protonation states and hydrogen-bonding orientations should be examined. The uncatalyzed dU hydrolysis also requires solvent effects to be included during the optimization step. Therefore, an enzyme model should account for the environment, either through explicit inclusion of residues or implicit 'solvation' approaches.

This chapter will use a variety of computational approaches (including full optimizations and constrained PES scans) and models (including two protonation states of His148) to study the hydrolytic deglycosylation of dU catalyzed by UDG. Structures calculated using full optimizations will be initially discussed, since these are used as the basis for the constrained optimizations. Second, (constrained) PES scans will be used to gain insight into the mechanism of UDG.

4.2 Computational Details

4.2.1 ONIOM Methodology

The above discussion leads to a computational UDG model composed of a minimum of five amino acids, an active-site water molecule, the substrate and (water) nucleophile, as well as further residues that simulate the protein environment. Due to the prohibitively large size of the system under investigation, the entire enzyme-substrate complex could not be treated with a quantum mechanical level of theory. Therefore, Morokuma's ONIOM methodology was utilized.[5] In this method, the total (real) system is divided into multiple layers. In all models used in this thesis, the system was divided into two layers. In a two-layer ONIOM calculation, the smaller portion is referred to as the 'model' region and is treated with a high level of theory, typically a QM method, while the real system is treated with a lower level of theory. The low level may be

another quantum mechanical or semi-empirical method, or it may be molecular mechanics.

Figure 4.2 provides an example of the definition of a two-layer system. The whole nucleoside is considered the 'real' system and is treated with the low-level method. The 'model' system is composed of the uracil moiety, and will be considered at both the low and the high levels of theory. In this example, the layer division bisects the N-glycosidic bond and a nitrogen (N1) is left in the 'model' region that contains a bond to nothing, which is referred to as a dangling bond. In the model, the dangling bond is capped with a proton (indicated with an asterisk in the figure). Since an N–C bond is longer than an N–H bond, the new bond length is scaled relative to the 'real' bond length using a default scaling factor. During an optimization, the angles and dihedrals involving the capping proton are held fixed for the 'model' system. Thus, all information regarding the position of the proton comes from the 'real' system, which prevents the model system from introducing erroneous hydrogen bonds or migration of truncated groups. For example, the functional groups of a protein may be included in the 'model' system, while the backbone is only present in the 'real' system. By keeping the capping protons fixed, the functional
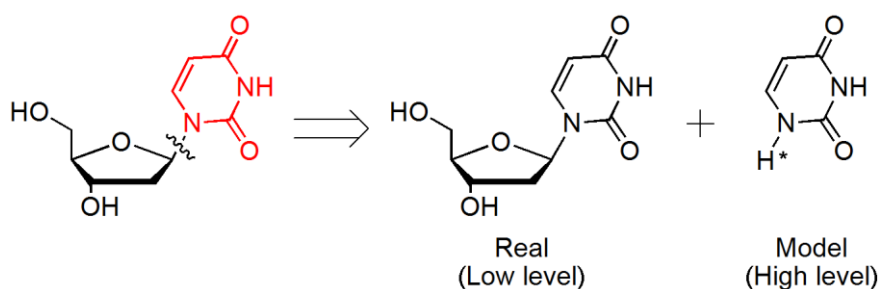
Figure 4.2 *An example of how the ONIOM method divides a system into a real and model component, where each system is treated at a different level of theory. The dangling bond that is created by removing the sugar moiety is capped with a proton (indicated by an asterisk).*

groups cannot migrate away from the backbone and/or each other. Furthermore, the model system can 'see' the real system through electrostatic embedding.

The energy of the system (E_{real}^{ONIOM}) is calculated using the results from three different calculations: low level of theory on the real system (E_{real}^{low}), low level of theory on the model system (E_{model}^{low}), and high level of theory on the model system (E_{model}^{high}). To determine the total energy of the system the high-level calculation is added to the difference between the two low-level layers (Equation 1), which removes double counting of the model region. In other words, the ONIOM calculation adds a low-level correction due to environment to the high-level, model energy, which includes polarization and steric effects during the optimization.

$$E_{real}^{ONIOM} = E_{model}^{high} + \left(E_{real}^{low} - E_{model}^{low}\right)$$ [1]

Gaussian 03 Version D.02 allows for the calculation of ONIOM energies, geometry optimizations and harmonic frequencies.[7] However, it does not allow for the incorporation of bulk solvent effects through a self-consistent reaction field (SCRF) method such as PCM. Therefore, the effect of the environment must be incorporated through explicit inclusion of the surrounding residues rather than a continuum method as utilized in the previous chapters, as discussed in Section 4.2.2.

Recent literature by Morokuma[8] and others[9] has found the B3LYP method combines well with semi-empirical methods such as PM3.[10] Therefore, all ONIOM calculations utilized B3LYP/6-31+G(d) as the high level of theory, to be consistent with the method implemented in Chapters 2 and 3, and PM3 as the low level of theory. The low-level region was further divided into constrained and unconstrained regions. Crystal structure overlays of the reactant, intermediate and product mimic containing complexes

indicate that many of the surrounding residues move very little throughout the reaction and therefore these residues were held fixed. Constraining the outer portion of the model simulates the steric constraints provided by the bulk of the enzyme, and prevents the system from artificially swelling in size to relieve electrostatic stress. In addition, holding Phe158 fixed prevents the introduction of errors due to inadequate treatment of the π–π interaction between the uracil moiety and Phe158 by PM3.

4.2.2 Basic Model Generation

Each model was built from a crystal structure, either 1EMH (reactant mimic),[12] 1Q3F (intermediate mimic)[13] or 1EMJ (product mimic),[12] where the specific residues incorporated are further discussed below. All models were designed using the same methodology (Figure 4.3). First, the coordinates for the residues (and substrate mimic) included in the model were taken from crystal structure PDB files. A hydrogen atom was
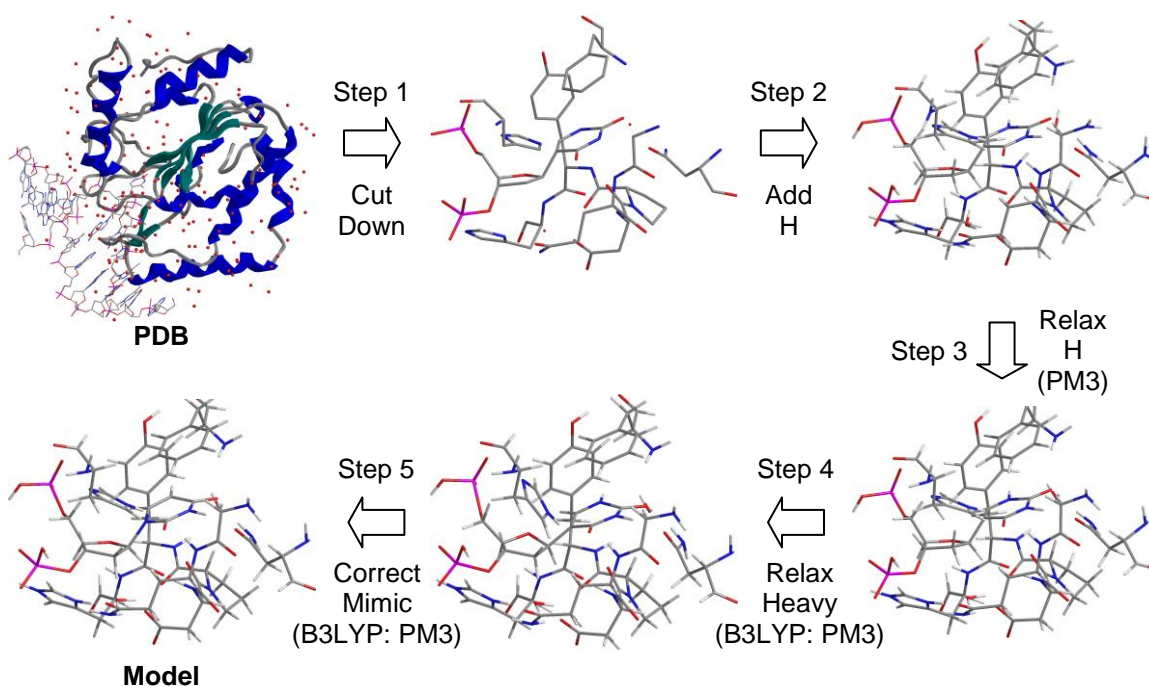
Figure 4.3 *The methodology utilized to generate a computational enzyme model from an X-ray crystal structure (1EMH shown).*

placed at all truncation points and the dihedrals of the truncated atoms were held constant

to maintain the protein and DNA backbones. Second, the remaining hydrogen atoms were

manually added to the residues, the hydrogen atoms were positioned to allow for

hydrogen bonding between nearby atoms where possible. For example, the hydroxyl

group of Tyr147 is near the backbone carbonyl of Phe158, and therefore all models

include a hydrogen bond between these two residues. Both phosphate groups were

assigned a negative charge, the Asp145 was assumed to be anionic, His268 was assumed

to be neutral and protonated at $N\varepsilon_2$, and His148 was either protonated at $N\varepsilon_2$ (neutral

His148 model) or protonated at both $N\varepsilon_2$ and $N\delta_1$ (cationic His148 model). The uracil

moiety was treated as an anion at extended glycosidic bond distances. In the third step,

the locations of the hydrogen atoms were optimized using PM3, while the position of all

heavy atoms was kept frozen. The terminal backbone dihedral angles were also

constrained to prevent the capping protons from migrating. These structural relaxations

were also carried out at the HF/3-21G level of theory; however, there was no measurable

difference between the two methods and therefore the semi-empirical (PM3) method was

used throughout to reduce the computational cost. Fourth, the system was allowed to

relax to the ONIOM model using B3LYP/6-31+G(d):PM3 (high level:low level). Finally,

the mimic was replaced with the true substrate and the entire system was relaxed.

The general description of the active site utilized to study the UDG mechanism is

depicted in Figure 4.4. As discussed in Section 4.1, the enzyme model contains discrete

active-site residues for leaving group stabilization (Asn204, His268, and a water

molecule) and nucleophile activation (His148, Asp145, and Pro146), as well as to

account for environmental effects (Gly143, Gln144, Tyr147, Phe158, and U + 1, U − 1
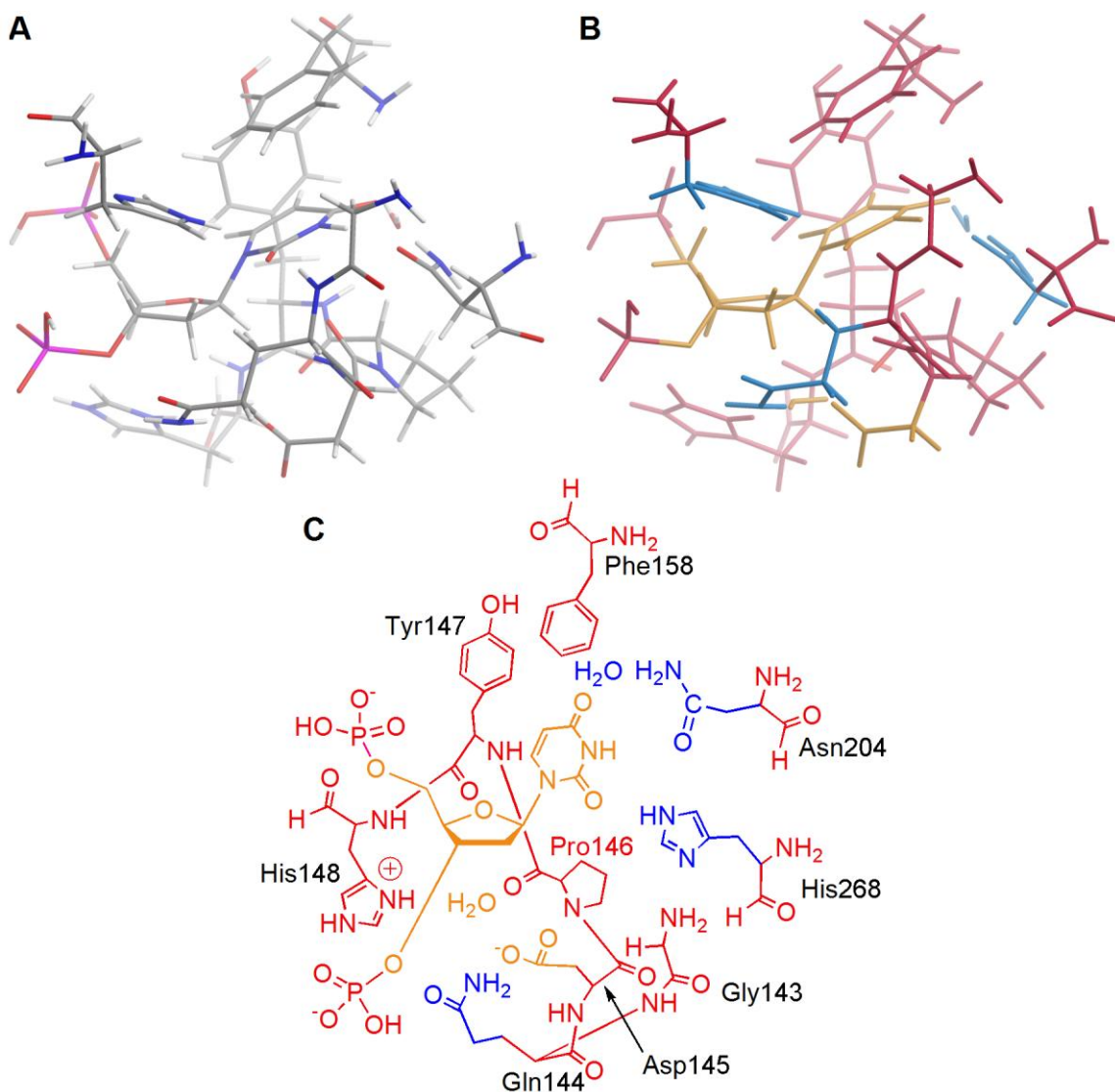
Figure 4.4 *A: The 3D geometry of the UDG active site, where the heavy-atom locations are taken from the 1EMH crystal structure. **B**: ONIOM level definitions. **C**: A 2D representation of the model. The high-level (B3LYP/6-31+G(d)) residues are indicated in orange, and the low-level (PM3) atoms are indicated in blue (unconstrained) and red (constrained). His148 was modeled as either neutral or cationic (shown).*

phosphates).[k] While the majority of the model was constrained throughout the

optimizations, the R-groups of the stabilizing residues (Asn204, His268, and Gln144)

(Figure 4.4, blue) were unconstrained to maintain hydrogen bonds with the substrate and

---

[k] While it has been debated that the +2 phosphate contributes to catalysis, it does not directly interact with any residues in the unconstrained region and therefore, was not included in the model.

nucleophile throughout the reaction coordinate. In addition, the Gln144 R-group was left unconstrained since overlays of the 1EMH, 1Q3F and 1EMJ crystal structures, as well as analysis of their respective B-values, indicates that there is a large degree of thermal motion associated with this residue. Since the protonation state of His148 is unknown, both the neutral and cationic states were modeled, which yields a system charge of –3 or –2, respectively.

4.2.3 Reaction Modeling

4.2.3.1. Relaxed Reaction Coordinate

Initially, the mechanism of action of UDG was modeled by separately optimizing each of the stationary points along the reaction coordinate. Each minimum was modeled from the nearest crystal structure. Specifically, the reactant was modeled from the 1EMH (reactant mimic) crystal structure, the intermediate was modeled from the 1Q3F (intermediate mimic) structure, and the product was modeled from the 1EMJ (product mimic) crystal structure. While overlays of the three crystal structures indicate that there is very little movement of the active-site residues throughout the reaction, each minimum (R, I and P) was generated from a different crystal structure to obtain a better estimate of the location of the fixed residues. Section 4.2.2 described in detail how each model was generated.

All ONIOM optimizations utilized the keyword *opt*=(*maxstep*=15), which aids in convergence by reducing the G03 default step-size in the optimization routine by half. Harmonic frequencies were calculated at the optimization level of theory to ensure that minima contained only real vibrational frequencies in the unconstrained region. Single-point (energy) calculations were carried out at the B3LYP/6-311+G(2df,p):PM3 level of

theory. Scaled (0.9806) zero-point vibrational corrections were included in the reported high-level energies.

4.2.3.2 Constrained Reaction Coordinate

The unimolecular cleavage studied in Chapter 2 indicated that the deglycosylation transition state is very flat. Indeed, it was found that the only way to reliably locate a transition state for hydrolysis was to follow the reaction coordinate in an incremental fashion. In addition, recent literature indicates that the current releases of the Gaussian suite cannot optimize protein transition states.[14] Therefore, the transition states for the S_N1 hydrolysis of dU by UDG were derived using reaction (PES) scans from the fully-optimized minima.

The reaction was followed in the forward direction from the closest optimized (minimum) structure. The glycosidic and nucleophilic distances were systematically varied in 0.1 Å increments, and the remainder of the system was allowed to relax (within the constraints of the model).

4.3 Results and Discussion

The model under investigation was designed to study the widely accepted mechanism for UDG action (Figure 4.5), whereby a water nucleophile is activated by an anionic aspartate residue, and the uracil anion product is stabilized through a short and strong hydrogen bond with His268 at O2. The deglycosylation mechanism is proposed to occur in a stepwise fashion, with dissociation of the nucleobase followed by addition of the water molecule to C1′ of the sugar. Therefore, there are five chemically relevant structures along the reaction pathway (denoted as RC, TS_D, IC, TS_A, and PC). Initially,

the three reaction minima will be discussed (Section 4.3.1). Then, the constrained minima

and related transition states will be examined (Section 4.3.2).
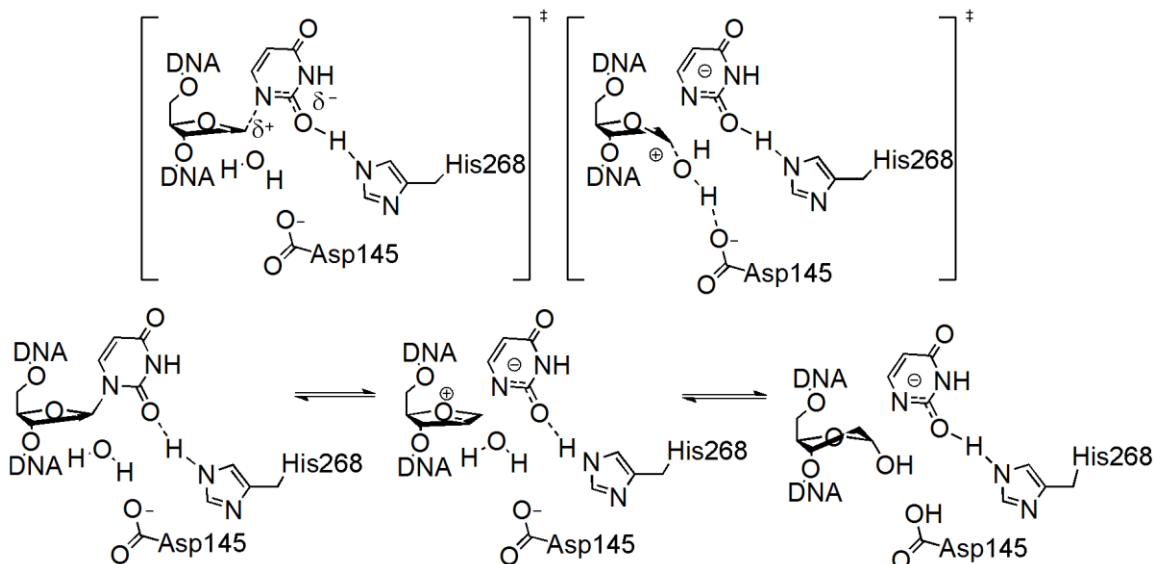
Figure 4.5 *Proposed mechanism for the hydrolysis of dU by UDG. The water nucleophile is activated by Asp145, and the uracil anion product is stabilized by a short and strong hydrogen bond to the neutral His268 residue.*

4.3.1 Relaxed Reaction Pathway

As discussed above, the UDG mechanism of action whereby the nucleophilic

water molecule is activated by a nearby aspartate residue was modeled using full

optimizations. While the transition states for the reaction cannot be determined using this

method, valuable thermodynamic information can be derived from the optimized minima.

In addition, the optimized minima can be compared to previous computational and

experimental studies, which may aid in determining the protonation state of His148.

4.3.1.1 Cationic His148 Model

As mentioned in the Introduction, the reactant geometry was designed from the

1EMH crystal structure,[12] which contains a non-planar base moiety at C1 (N1 in uracil).

Mol et al. proposed that the non-planarity arises due to reactant destabilization where the

enzyme promotes hyperconjugation and decreases glycosidic bond stability (Section 1.2.2.1.1).[12] It has also been suggested that the distortion may be due to the mimic, and therefore may not appear in 2′-deoxyuridine. Figure 4.6 compares the crystal-structure geometry and the ONIOM optimized geometry for the reactant complex. It can be seen that significant distortion remains at N1 after relaxation of the model. Furthermore, when N1 contains an sp^3-like geometry, the resonance of the pyrimidine ring is broken causing a slight boat conformation (i.e., O4 tilts down to the sugar side of the base). Indeed, optimization of a reactant starting with the uracil moiety arranged to be planar or starting from the intermediate mimic crystal structure (1Q3F) also leads to this distortion. Further consideration of the intermediate structure reveals that the bend in the uracil moiety may be caused by the active site exhibiting preferential binding of the intermediate since the uracil anion is parallel to the phenylalanine side-chain in the intermediate.
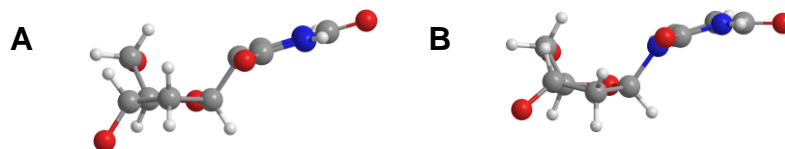
Figure 4.6 *Distortion in the reactant for the cationic his148 model in the ψU, 1EMH crystal structure (**A**) and in the relaxed optimization (**B**).*

The reactant, intermediate and product complexes for the cationic His148 model are outlined in Figure 4.7. In the reactant, the water nucleophile is held under the sugar moiety through three hydrogen bonds to Asp145, Pro146 and His148 (Figures 4.7 B and C, left). The Watson-Crick binding face of uracil participates in hydrogen bonds to Asn204, His268 and a water molecule. It is interesting to note that the base is accepting three hydrogen bonds and donating only one proton, which stabilizes the negative charge accumulating on the base during glycosidic-bond cleavage. The sugar moiety has a very

slight C2′-*exo* pucker, which is similar to the planar geometry in the crystal structure. The

sugar puckering may be due to the restraints placed upon the backbone through hydrogen

bonds between the U + 1 and U − 1 phosphate groups and nearby serine residues (not

shown), as well as from dU being flipped out of the helix.

In the intermediate complex (Figure 4.6 middle column), the water nucleophile

has not moved appreciably; however, the hydrogen bond to Asp145 has shortened (from
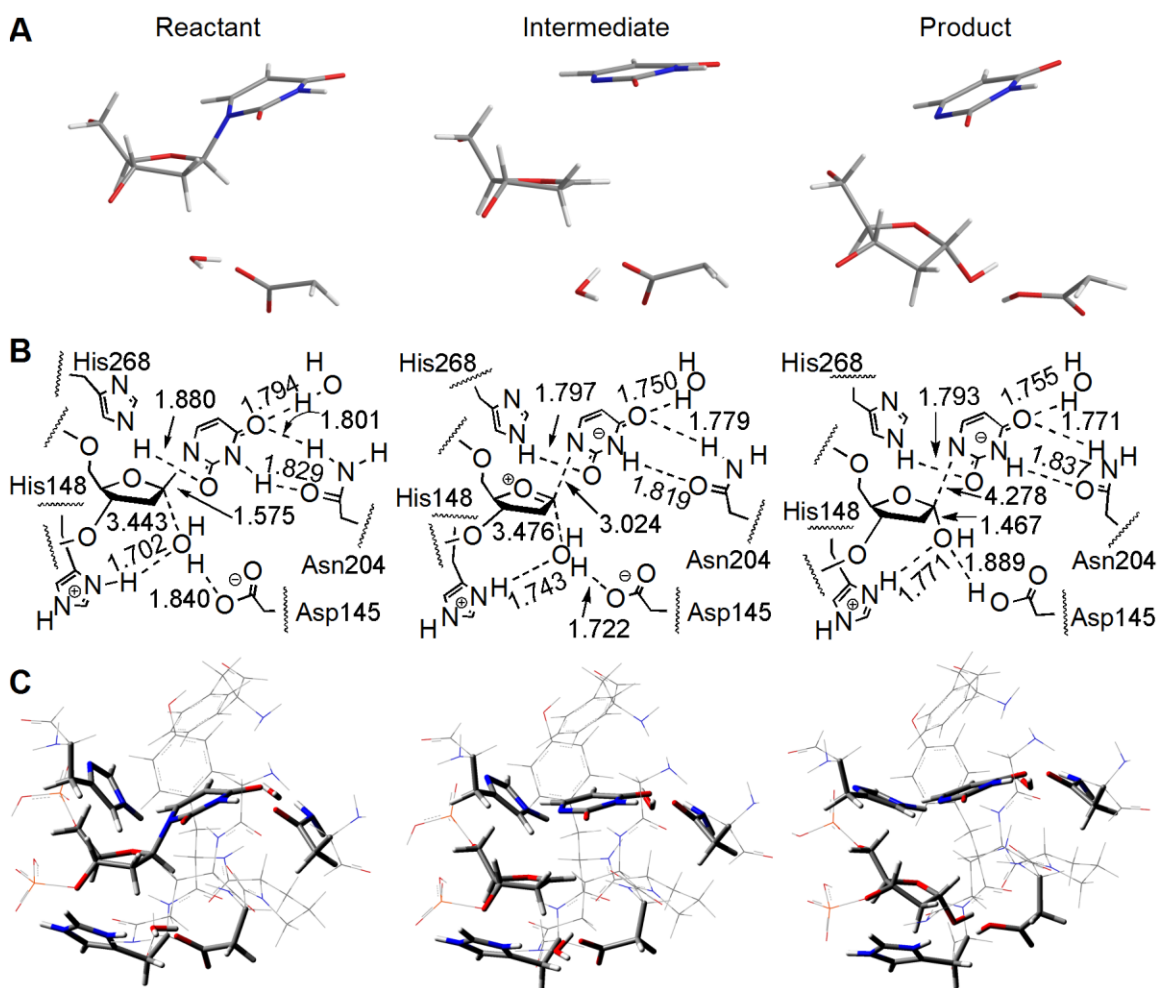
Figure 4.7 *The reactant, intermediate and product structures for the mechanism of UDG action with cationic His148 obtained through full optimizations using B3LYP/6-31+G(d):PM3. **A**: a close up of the DFT region; **B**: a 2D representation with important distances highlighted (Å); **C**: the full ONIOM model with all residues identified in **B** emphasized in tubing.*

1.840 to 1.722 Å). The uracil anion has migrated away from the oxacarbenium cation, where the hydrogen bonds to all donating residues have strengthened substantially. The sugar moiety is planar, and does not show any of the C1′-*exo* character displayed by the parent 1-azadeoxyribose mimic in the 1Q3F crystal structure.[13] However, this TS mimic is inherently sp^3 hybridized and therefore is expected to contain some puckering. Furthermore, the enzyme holds the backbone phosphates quite tightly, which may force the mimic to display N1′-*exo* pucker and therefore this structure may not carryover to the deoxyribose moiety. The oxacarbenium cation is stabilized by an electrostatic sandwich between the uracil anion, the aspartate anion, and two nearby phosphate groups, which is only slightly masked by the cationic His148 residue (Figure 4.7C, middle) leading to a structure that is 52.4 kJ mol^{-1} less stable than the reactant geometry (Table 4.1). Therefore, the activation energy for this mechanism is > 52.4 kJ mol^{-1}. This can be compared to the activation energy of 62.1 kJ mol^{-1}, and an intermediate complex energy of 31.0 kJ mol^{-1}, calculated by Dinner et al. using QM/MM on the entire 1EMH,[2a] as well as, the experimental activation energy for a UDG clone (~50.6 kJ mol^{-1}).[15] These comparisons indicate that this model is close to accurately predicting the barrier, but destabilizes the intermediate.

Table 4.1 *Relative energies of the minima obtained for the mechanism of UDG action with cationic His148 using full optimizations.*

	Reactant	Intermediate	Product
B3LYP/6-31+G(d):PM3	0.0	13.4	−70.5
B3LYP/6-311+G(2df,p):PM3*	0.0	52.4	−49.3

* Calculated as a single-point energy on B3LYP/6-31+G(d):PM3 optimized geometries, including (0.9806) scaled zero-point vibrational correction.

In the product complex, the aspartate residue has stripped a proton from the water nucleophile, and the resulting hydroxyl adds to C1′. The sugar moiety contains a C3′-

endo pucker, which is one of the standard forms for 2′-deoxyribose. The uracil anion migrates to 4.278 Å away from C1′ of the sugar, and is stabilized by hydrogen bonds. Although it has been proposed that the hydrogen bond between O2 and $N\varepsilon_2H$ of His268 falls into the short and strong hydrogen bond (SSHB) classification, the (O⋯N) heavy atom distance is 2.753 Å, which is larger than accepted SSHB distance of < 2.65 Å.[16] However, this interaction is between the values reported for the neutral and anionic uracil⋯imidazole hydrogen-bond distances reported by Hunter et al.[17] which indicates that the protein environment is slightly masking this interaction. The product complex is highly exothermic, and is 49.3 kJ mol⁻¹ more stable than the reactant (101.7 kJ mol⁻¹ more stable than the intermediate).

4.3.1.2 Neutral His148 Model

The UDG model with neutral His148 yields reaction minima that are structurally similar to those discussed for the cationic histidine model (Figure 4.8). In the ONIOM optimized reactant (Figure 4.8, left), the glycosidic bond length is 1.609 Å, which is notably longer (0.034 Å) than reported for the His148 cationic model (Figure 4.7, left). This indicates that the barrier to glycosidic bond cleavage is likely significantly lower than discussed for the previous model. Since the hydrogen-bond distances between the active-site residues and the Watson-Crick binding face of uracil are all within 0.02 Å of those for the His148 cationic model, the increase in the glycosidic bond length is due to the sugar moiety pulling away from the base, rather than vice versa. The only other major difference between the two models is the location of the nucleophilic water. In the cationic model, the water molecule was held under the sugar moiety via a hydrogen bond to the $N\delta_1$ proton of His148; however, this interaction cannot occur in the neutral model
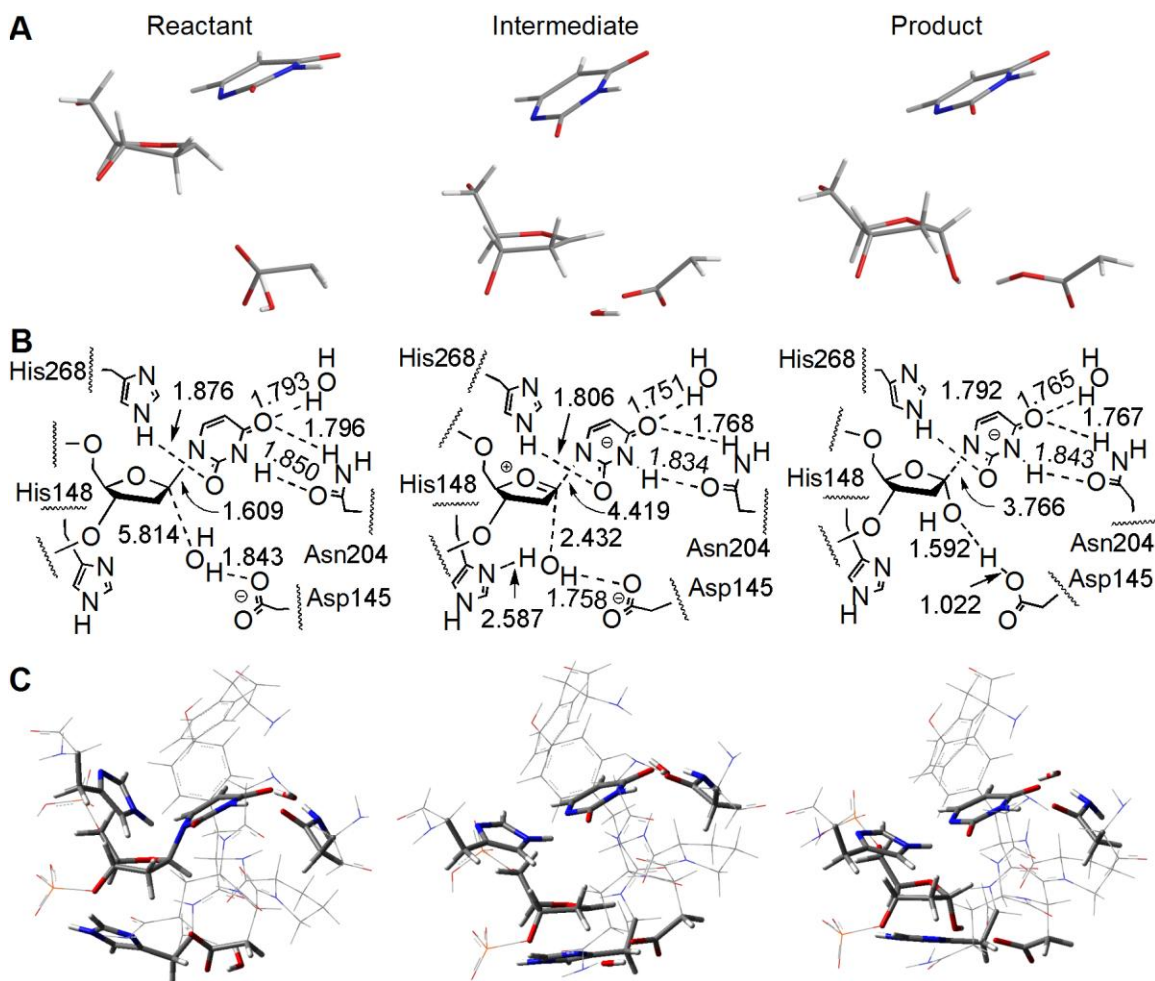
Figure 4.8 *The reactant, intermediate and product structures for the mechanism of UDG action with neutral His148 obtained through full optimizations using B3LYP/6-31+G(d):PM3. **A**: a close up of the DFT region; **B**: a 2D representation with important distances highlighted (Å); **C**: the full ONIOM model with residues identified in **B** emphasized in tubing.*

if the water molecule is interacting with Asp145 and Pro146. Therefore, the water

nucleophile migrates away from C1′ of the sugar by almost 6 Å.

In contrast to the His148 cationic model, which has an endothermic intermediate

complex, the neutral histidine model generates an intermediate with an energy of -28.9

kJ mol^{-1} relative to the reactant complex (Table 4.2). This model contains a more stable

intermediate than the cationic histidine model due to the absence of the His148 cation–

oxacarbenium cation repulsion. The neutral histidine intermediate contains a glycosidic

bond that is longer and a nucleophile distance that is shorter than the equivalent cationic

structure (compare Figure 4.8B middle to Figure 4.7B middle). This implies that the

intermediate occurs later in the reaction scheme for the neutral model. The asparagine

hydrogen bonds to the uracil anion are also slightly tighter for the neutral model, and the

His268 interaction with O2 is slightly looser, which indicates that the uracil anion is

rotated compared to the His148 cationic model. As with the previous model, the

oxacarbenium cation is mostly planar with a minor C3′-exo pucker.

Table 4.2 *Relative energies of the minima obtained for the mechanism of UDG action with neutral His148 using full optimizations.*

	Reactant	Intermediate	Product
B3LYP/6-31+G(d):PM3	0.0	−10.7	−109.3
B3LYP/6-311+G(2df,p):PM3*	0.0	−28.9	−276.4

* Calculated as a single-point energy on B3LYP/6-31+G(d):PM3 optimized geometries, including (0.9806) scaled zero-point vibrational correction.

The neutral model also yields a highly exothermic product complex, which boasts

a relative energy of −276.4 kJ mol^{-1} with respect to the reactant (Table 4.2). The large

stabilization can be partially attributed to the separation of the negative charges present in

the active site. A reduction in electrostatic strain results from transferring negative charge

from the aspartate residue to the uracil moiety during the hydrolysis reaction. In the

reactant, there are four negatively charged residues held close together. When uracil is

released further into the active-site pocket, the charge is spread over a larger volume.

This release of electrostatic strain is referred to in the literature as 'auto-catalysis'.[2a,4,13]

Although the same argument can be made for the cationic His148 model, the cationic

charge of the histidine will partially mask the effect. Another interesting feature of the

product complex is the position of the transferred proton. Figure 4.8 (B, right) notes the

distances between the donor (water) and acceptor (Asp145) residues. It can be seen that

the acceptor distance is slightly elongated compared to a standard O–H bond (> 1 Å), while the donor distance is quite short (< 1.6 Å). The O–H⋯O1′ hydrogen bond is short enough to fall into the short-and-strong category, which implies that full transfer to the aspartate has not occurred.

In summary, the neutral His148 model yields more stable complexes than the cationic His148 model, which agrees with NMR experiments that found His148 to be neutral in the open state.[4] The neutral His148 model also generates a reactant with less uracil distortion than the cationic His148 model. The structures derived using both models support an 'auto-catalytic' mechanism, whereby the highly concentrated negative charge in the active site is relieved by transferring the charge to uracil, which migrates away from the phosphate backbone. Despite the significance of these results, comparisons of the reaction minima only give thermodynamic information. To obtain complete mechanistic information, the next section characterizes the transition states.

4.3.2 Constrained Reaction Pathway

Since reaction pathway scans successfully predicted the dU hydrolysis mechanism for the first time (Chapter 2), this approach may be useful for characterizing the transition states for the mechanism of UDG action. Therefore, the minima found in the previous section were used as starting points to study the dissociation and association steps of the reaction. This methodology has also been very recently utilized by other groups studying enzyme mechanisms with ONIOM.[18] The results for the two steps in the reaction will be discussed separately below.

4.3.2.1 Dissociation of Uracil

The first, dissociative, step of the UDG mechanism was modeled from the previously optimized reactants (Figures 4.7 and 4.8, left). The cationic His148 model leads to a dissociative transition state with a glycosidic bond length of 2.000 Å (Figure 4.9A). This distance is significantly shorter than found for the uncatalyzed transition state (~2.9 Å) (Section 2.3). From Figure 4.9, it can be seen that the uracil moiety still contains a very slight boat conformation, which is similar to the reactant geometry. The TS also contains hydrogen-bond lengths to the uracil moiety that are between those reported for the reactant and intermediate. A similar transition state is found for the neutral His148 model (Figure 4.9B). The major difference between the two geometries is the additional
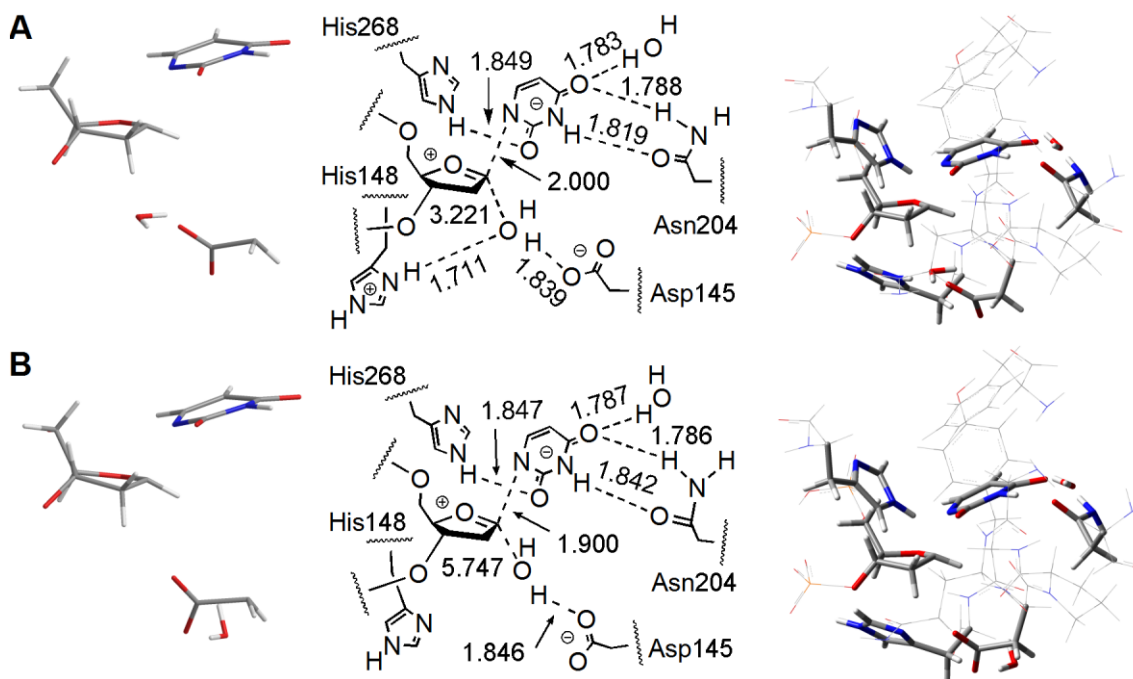
Figure 4.9   *The dissociative transition states found using incremental ONIOM (B3LYP/6-31+G(d):PM3) calculations from the optimized reactants for the cationic (**A**) and neutral (**B**) His148 models. From left to right the frames correspond to: the DFT region; a 2D representation with important distances (Å); and the full model, with important residues emphasized in tubing.*

hydrogen bond to the nucleophilic water molecule in the cationic His148 model, which changes the orientation of the nucleophile.

The energy profiles for the two models are summarized in Figure 4.10. The neutral His148 model displays barrierless dissociation at the B3LYP/6-311+G(2df,p):PM3 level of theory.[1] The cationic His148 model has a larger dissociation barrier due to the positively-charged histidine residue producing a repulsive interaction with the newly formed oxacarbenium cation. This same repulsion is
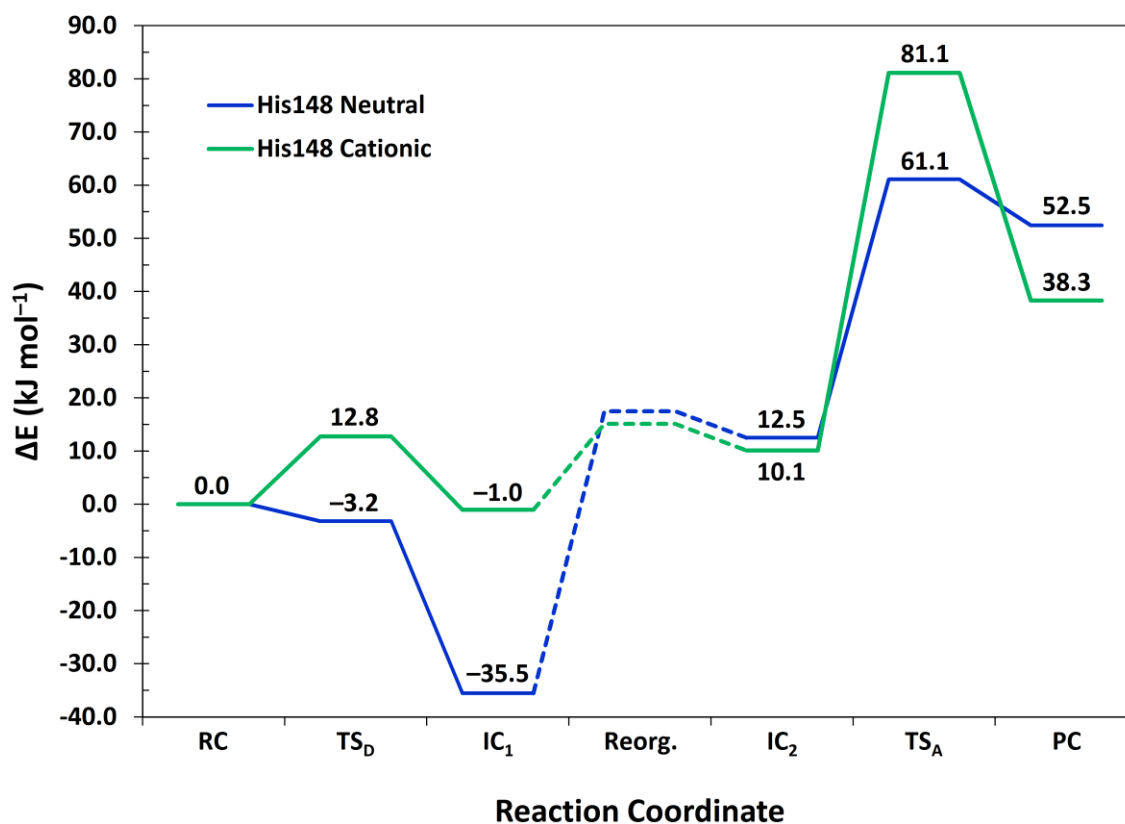
Figure 4.10 *Reaction coordinate for the hydrolysis of dU by UDG using PES scans beginning from the reactant and intermediate structures from Section 4.3.1. The reorganization energy between the scan intermediate (IC_1) and the previous intermediate (IC_2) is unknown and shown as a dashed line. Reported energies were calculated at the B3LYP/6-311+G(2df,p):PM3// B3LYP/6-31+G(d):PM3 level including scaled (0.9806) ZPC.*

---

[1] At the optimization level, the barrier for dissociation of uracil using the neutral his148 model is 4.4 kJ mol^{-1}. When the ZPC is incorporated the barrier is reduced to -2.3 kJ mol^{-1}. Therefore, the barrierless TS is due to the zero-point energy, and not to the level of theory of the single-point (energy) calculation.

responsible for the large difference in intermediate stability (IC_1, Figure 4.10), where the neutral His148 model is 34.5 kJ mol^{-1} more stable than the intermediate complex derived from the cationic His148 model. This is similar to the difference in IC energy found using relaxed optimizations as reported in the previous section.

4.3.2.2 Association of Nucleophile

The second step of the mechanism of UDG action was modeled from the previously optimized intermediate complexes (Figures 4.7 and 4.8, middle). These structures are similar to the intermediates found by following the dissociative TS; however, some reorganization of the residues occurs. Although overlays of the parent crystal structures for the reactant and intermediate complexes (1EMH and 1Q3F, respectively) indicates that the degree of motion is small, the energy of the intermediate complexes (IC_2, Figure 4.10) are $11 - 47$ kJ mol^{-1} above those found from the dissociation scan (IC_1, Figure 4.10). Therefore, the reaction scheme includes a transition state that corresponds to reorganization of the active site (Figure 4.10, dotted lines), which is a common practice for enzyme mechanisms that incorporate reorganization.[19] While the barriers between the two IC are unknown, it appears that the cationic His148 model requires less reorganization than the neutral His148 model. This may be because the cationic His148 residue holds the nucleophilic water molecule beneath the sugar residue.

The associative TS for the neutral His148 model is 20 kJ mol^{-1} more stable than the cationic histidine model, which is close to the proposed 'anti-catalytic' effect of the residue.[2] This can be partially explained by the cationic histidine residue masking the negative charge on the aspartate residue, which increases the barrier for proton transfer

from the nucleophilic water molecule. The barriers (relative to the IC_2 intermediate) for the associative step are 48.6 and 71.0 kJ mol^{-1} for the neutral and cationic His148 models, respectively. These values are near the experimental activation energy of ~50.6 kJ mol^{-1}.[15] However, the calculated barriers imply that the rate-limiting step corresponds to association and, more specifically, activation of the nucleophile. This contradicts KIE studies that find the rate-limiting step to be dissociation.[20]
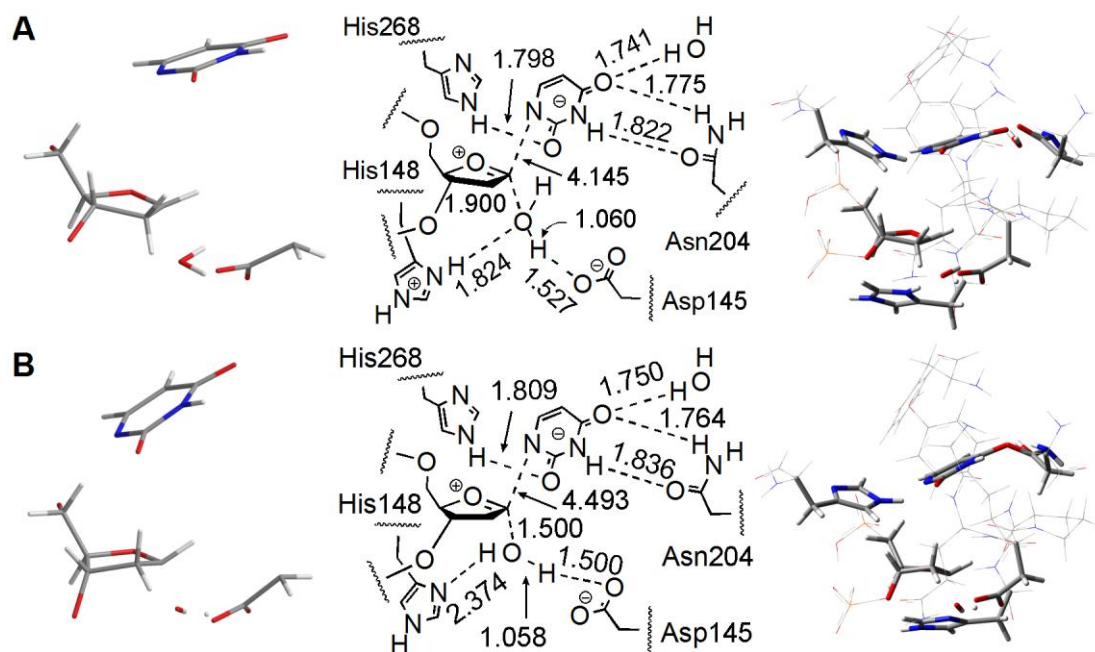

Figure 4.11 *The associative transition states found using incremental ONIOM (B3LYP/6-31+G(d):PM3) calculations from the optimized intermediates for cationic (**A**) and neutral (**B**) His148 models. From left to right the frames correspond to: the DFT region; a 2D representation with important distances (Å); and the full model, with important residues emphasized in tubing.*

Figure 4.11 summarizes the associative transition states found in this study. Similar to the corresponding transition states for the uncatalyzed hydrolysis of dU (Chapter 2), there is partial proton transfer from the nucleophilic water molecule to the nearby aspartate residue. However, while the cationic model displays an Ow–C1′ distance near that found for the uncatalyzed hydrolysis (~1.9 Å), the neutral model

contains a very short nucleophile distance (~1.5 Å). The very late transition state for the neutral His148 model implies that this reaction requires a great deal of force to occur. In addition, the secondary proton on the nucleophile rotates away from the proline backbone (not shown) to weakly interact with His148. This result is interesting, since it implies that formation of a hydrogen bond to the histidine nitrogen is more favourable than to the proline backbone carbonyl.

The product complexes found by following the reaction coordinate from the association TS display reverse relative stability compared with the fully-optimized products (Section 4.3.1). In this case, the cationic His148 model is more stable than the neutral His148 model due to the strong electrostatic interactions between the cationic imidazole moiety and the nearby phosphate groups. In the relaxed neutral histidine product, the phosphate groups are slightly further apart, which relieves the electrostatic stress. In addition, the sugar moiety relaxes from a slight C2′-*endo*/C1′-*exo* twist pucker in the constrained product complexes (not shown) to a much more stable C2′-*endo* conformation in the relaxed product complex (Figure 4.8, right).

In summary, the neutral His148 model leads to lower barriers than a model with cationic His148. However, the cationic model requires much less reorganization of the active site between the two characterized IC. These results are inconclusive and suggest that more work needs to be done regarding the protonation state of His148.

4.3 Conclusions

This chapter focused on developing a working model for the mechanism of UDG action using a variety of approaches for modeling both the enzyme and the reaction. High-resolution crystal structures for UDG were used as starting points for the active-site

geometry along the reaction coordinate. A two-layer ONIOM model, which combined DFT (B3LYP) and semi-empirical (PM3) methods, was utilized to study the reaction. Constrained one-dimensional reaction PES scans were performed from the fully-optimized minima to characterize dissociative and associative transition states. This methodology is able to locate appropriate transition structures on relatively flat surfaces, and is able to reproduce previously determined amino-acid energy contributions.

By comparing models where His148 was cationic and neutral, it was found that the cationic form of the residue destabilizes the intermediate by ~30 kJ mol^{-1}, which is similar to the proposed anti-catalytic effect of the residue reported by Dinner et al.[2a] Additionally, the cationic model shows a more pronounced N1 distortion in the reactant than the neutral model, which decreases the barrier to deglycosylation via hyperconjugation with the sugar moiety. However, the neutral His148 model generates more energetically stable structures than the cationic His148 model since the repulsive oxacarbenium cation – histidine cation interaction is not present. Combined, these results imply that His148 might be neutral during the dissociation step and therefore, supports experimental NMR evidence.[4] Although the results presented in this thesis are inclusive as to the protonation state of the His148 residue, they provide clues for generating a new model as discussed in Chapter 5.

4.4 References

(1)   (a) Luo, N.; Mehler, E.; Osman, R. *Biochemistry* **1999**, *38*, 9209–9220; (b) Osman, R.; Fuxreiter, M.; Luo, N. *J. Comput. Chem.* **2000**, *24*, 331–339.

(2)   (a) Dinner, A. R.; Blackburn, G. M.; Karplus, M. *Nature (London)* **2001**, *413*, 752–755; (b) Ma, A.; Hu, J.; Karplus, M.; Dinner, A. R. *Biochemistry* **2006**, *45*, 13687–13696.

(3)   Millen, A. L.; Archibald, L. A. B.; Hunter, K. C.; Wetmore, S. D. *J. Phys. Chem. B* **2007**, *111*, 3800–3812.

(4)   Parker, J. B.; Stivers, J. T. *Biochemistry* **2008**, *47*, 8614–8622.

(5)   (a) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170–1179; (b) Humbel, S.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 1959–1967; (c) Svensson, M.; Humbel, S.; Morokuma, K.; *J. Chem. Phys.* **1996**, *105*, 3654–3661.

(6)   Otterlei, M.; Warbrick, E.; Nagelhus, T. A.; Slupphaug, G.; Akbari, M.; Aas, P. A.; Steinsbekk, K.; Bakke, O.; Krokan, H. E. *EMBO J.* **1999**, *18*, 3834–3844.

(7)   Gaussian 03, Revision D.02, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc., Wallingford CT, **2004**.

(8)   Vreven, T.; Morokuma, K. *J. Comput. Chem.* **2000**, *21*, 1419–1432.

(9)   Sklenak, S.; Yao, L.; Cukier, R. I.; Yan, H. *J. Am. Chem. Soc.* **2004**, *126*, 14879–14889.

(10)   (a) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209–220; (b) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 221–264.

(11)   McNamara, J. P.; Hillier, I. H. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2362–2370.

(12)   Parikh, S. S.; Walcher, G.; Jones, G. D.; Slupphaug, G.; Krokan, H. E.; Blackburn, G. M.; Tainer, J. A. *Proc. Natl. Acad. Sciences U.S.A* **2000**, *97*, 5083–5088.

(13)   Bianchet, M. A.; Seiple, L. A.; Jiang, Y. L.; Ichikawa, Y.; Amzel, L. M.; Stivers, J. T. *Biochemistry* **2003**, *42*, 12455–12460.

(14)   Lundberg, M.; Kawatsu, T.; Vreven, T.; Frisch, M. J.; Morokuma, K. *J. Chem. Theory Comput.* **2009**, *5*, 222–234.

(15)   Slupphaug, G.; Eftedal, I.; Kavli, B.; Bharati, S.; Helle, N. M.; Haug, T.; Levine, D. W.; Krokan, H. E. *Biochemistry* **1995**, *34*, 128–138.

(16)   Miño G.; Contreras, R. *J. Phys. Chem. A* **2009**, *113*, 5769–5772.

(17)   Hunter, K. C.; Millen, A. L.; Wetmore, S. D. *J. Phys. Chem. B* **2007**, *111*, 1858–1871.

(18)   (a) Inoue, T.; Shiota, Y.; Yoshizawa, K. *J. Am. Chem. Soc.* **2008**, *130*, 16890–16897; (b) Li, X.; Chung, L. W.; Paneth, P.; Morokuma, K. *J. Am. Chem. Soc.* **2009**, *131*, 5115–5125.

(19)   See, for example; McCann, J. A. B.; Berti, P. J. *J. Am. Chem. Soc.* **2008**, *130* 5789–5797.

(20)   Werner, R. M.; Stivers, J. T. *Biochemistry* **2000**, *39*, 14054–14064.

Chapter 5: Conclusions and Future Work

5.1 Global Summary

This thesis set out to study how deoxyribonucleic acids undergo either uncatalyzed or enzyme catalyzed hydrolytic deglycosylation. Computational chemistry techniques, and specifically hybrid density functional theory, were utilized to examine a variety of mechanisms and model designs. In the case of uncatalyzed hydrolysis, particular attention was paid to the level of solvation required to accurately predict both experimental activation barriers and proposed mechanism. Once the uncatalyzed hydrolysis was examined, key concepts derived from the model were applied to the mechanism of action of the N-glycosylase UDG. While there is a large body of work on UDG, there are still many unanswered questions. For example, one of the goals of this thesis was to gain a better understanding of the role of His148 and its protonation state.

In this chapter, the basic conclusions found through these studies will be summarized, with emphasis on how this work can be applied to other systems of interest. In addition, future work will be proposed to fill in the missing pieces of the puzzles that were unearthed during the course of this work.

5.2 Nucleoside Hydrolysis

5.2.1 Conclusions

This study initially utilized the hydrolysis of 2′-deoxyuridine to develop a model capable of effectively describing the reaction mechanism. In general, previous computational literature on the deglycosylation of nucleic acids used a model containing only the nucleoside and the nucleophile, with no treatment of the surrounding environment in the geometry optimization steps. Instead, the most common practice is to

find the geometry using gas-phase methods and incorporate solvent effects as a correction to the energy. However, this technique is naïve, since it assumes that the surrounding environment only has an energetic effect and no structural effect on the reaction mechanism. In many cases, this assumption may hold; however, in the case of reactions involving charge separation, such as bond dissociations, it is no longer valid.

An in-depth analysis of the effect of (both explicit and implicit) solvent led to a model for dU hydrolysis that includes three explicit water molecules, which form a hydrogen-bonding network between the sugar moiety and uracil, as well as inclusion of implicit solvent effects during the optimization routine. This model follows the deglycosylation reaction pathway without suffering from errors due to structural instabilities. An interesting result from this study is that the activation energy calculated by including solvent effects in the optimization or as a correction to the gas-phase energy can be the same, even though the structures are dramatically different. This implies that verifying a computational model by comparing solvent-corrected energies to experiment may not be enough to determine the structural accuracy of the model.

The hybrid solvation model was applied to four natural 2′-deoxyribonucleosides (dA, dC, dG, and dT) and it was found that the methodology utilized to study dU can be carried over to similar systems, as long as care is taken regarding the positioning of the hydrogen-bond network. For example, the trend in the Gibbs energies of activation was correctly calculated for the pyrimidines and purines separately. However, the reported dissociation barriers for the purines are too large compared to the pyrimidines.

5.2.2 Future Work

To correct the large deglycosylation barrier for dG, a future model should include a hydrogen-bond network of water molecules that connects the 3′-hydroxyl of the sugar to N7 of guanine. This would increase the stability of the guanine leaving group since the N7 site has a higher proton affinity than the N3 site.[1]

In addition to studying the hydrolysis of natural 2′-deoxyribonucleosides, the method developed in this thesis for studying hydrolysis of N-glycosidic bonds could be applied to damaged DNA nucleobases such as those resulting from oxidation or alkylation damage. This method could also be extended to the nucleotide form of the nucleic acids, which could act as a truncated model of DNA strands. This medium-sized computational model could be used to bridge the gap between small-model, high-level studies and large-model, low-level studies currently existing in the literature. Specifically, on one hand the model will be small enough to use quantum mechanical methods, but, on the other hand, it still incorporates explicit solvation.

5.3 Mechanism of UDG Action

5.3.1 Conclusions

The mechanism of UDG action was modeled using the two-layer ONIOM methodology, where His148 was either neutral or cationic. The reactant, intermediate and product complexes were optimized from crystal structure orientations, which were subsequently utilized to generate one-dimensional reaction coordinate scans and thereby identify the corresponding transition states. Through this study, it was found that the neutral His148 model is more energetically stable than the cationic model, which likely occurs since the close proximity of cationic His148 to the oxacarbenium cation

intermediate is highly anti-catalytic. Furthermore, the second, associative step of the mechanism was found to be rate limiting, with a barrier of 48.6 and 71.0 kJ mol^{-1} for the neutral and cationic His148 models, respectively. The calculated rate-limiting barrier is near that found experimentally,[2] which, in contrast to previous propositions, indicates that the rate-limiting step may be association of the nucleophile rather than dissociation of the base.

Interestingly, the reaction scans isolated a reactant complex with N1 distortion, as seen previously in the 1EMH crystal structure,[3] and the structure calculated using QM/MM.[4] The distortion was present regardless of starting structure, indicating that this may not be due to the crystallization method, but may be the result of reactant destabilization implemented by the enzyme to facilitate catalysis.

Despite the successes listed above, the product complex and association transition state were difficult to model and, in some cases, full proton transfer from the nucleophilic water molecule to the Asp145 residue did not occur. This indicates that more research is required to determine whether the error is from the model or the mechanism and, more importantly, how to rectify the problem.

5.3.2 Future Work

Analysis of the ONIOM energy for the constrained reaction pathways studied in Chapter 4 (Figure 5.1) provides valuable information regarding the effect of the protein environment. The high-level contribution to the dissociation energy (Figure 5.1A) behaves similarly to the uncatalyzed dU hydrolysis study in Chapter 2. For example, the energy continuously increases, reaching a plateau near 130 kJ mol^{-1}. The protein environment contribution to the dissociation energy is entirely downhill, indicating that

the active site selectively binds to the intermediate geometry. Combined, the two

components give rise to an energy profile with a low energy TS and stable intermediate.
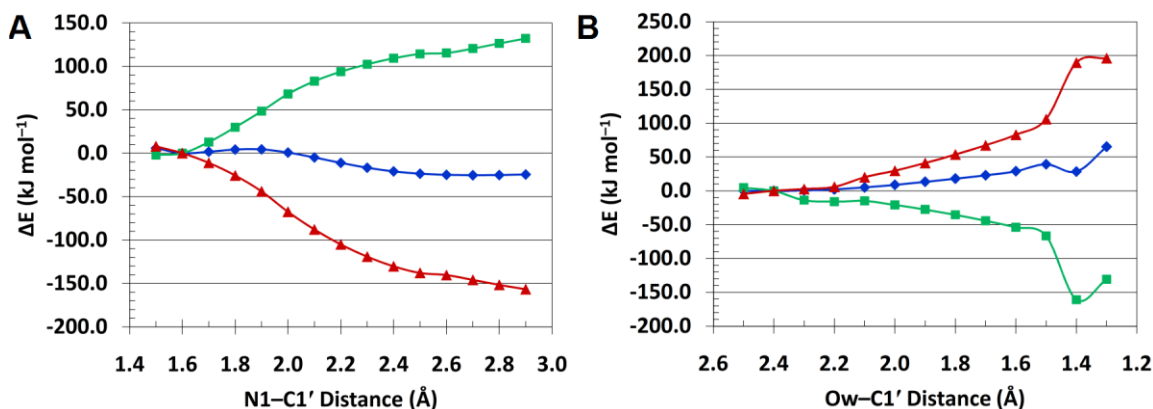
Figure 5.1 *The ONIOM component energies for the constrained reaction scans (**A**: dissociation, **B**: association) for the neutral His148 model. The ONIOM energy (blue diamonds) is the sum of the model energy at DFT (green squares) and the protein environment (real – model) energy at the low level (red triangles).*

The reverse is seen when the ONIOM component energies for the association step

are considered (Figure 5.1B), which indicates that the B3LYP model characterizes

barrierless addition of the water nucleophile to the oxacarbenium cation and leads to a

highly exothermic product. This follows with the uncatalyzed mechanism that located a

very stable product separated from the intermediate by a low barrier. However, the

protein environment component is highly endothermic, and contains a sharp increase in

energy when proton transfer from the water molecule to the Asp145 residue occurs.

Therefore, there is a feature in the low-level portion of the model that is highly repulsive

upon neutralization of the oxacarbenium cation. A possible explanation for the large

repulsion is that the pK_a of the aspartate residue is too low for the group to accept a

proton from the nucleophilic water molecule. Another possible explanation involves the

role of the phosphate groups. In the intermediate, the charge on the nearby phosphates is

partially shielded by the oxacarbenium cation; however, in the product, this shielding is removed. Therefore, the protein region of the model is more repulsive in the product than in the intermediate. Combined, these analyses imply that the Asp145 residue may not be the responsible for activating the nucleophilic water molecule. Therefore, future work should reexamine the roles of other active-site residues to determine whether the model or the mechanism is causing the highly repulsive low-energy component.

5.3.2.1 Discussion

The currently accepted UDG mechanism is based almost entirely on mutational analysis of UDG that accompanied the 1AKZ crystal structure.[5] In that study, mutating the Asp145 residue to asparagine or glutamate was found to reduce the activity of eUDG to less than 1% of the specific activity of the wild type enzyme. The authors interpreted this to imply that the aspartate residue is important for catalysis. Later groups further interpreted that the Asp145 residue was responsible for activating the nucleophile.[6] Indeed, the product complex crystal structure (1SSP) clearly shows that the aspartate residue is near the sugar and close enough to hydrogen bond to the C1′-hydroxyl.[3]

There is, however, a body of evidence indicating that the 'catalytic' residue is a histidine (possibly His148). First, the optimal pH for UDG is between 7.7 and 8.0, which is indicative of a neutral imidazole acting as a general base.[2] In fact, the 1AKZ paper proposed a mechanism whereby an active site histidine acts as either a general base or as the nucleophile.[5] The latter mechanism was later discarded once the enzyme was determined to be monofunctional. In the late 90's, the Stivers group completed pK_a studies on UDG to determine the pK_a of the catalytic residue.[6b] They found the catalytic residue to have a pK_a of 6.2, which they assigned to Asp145. However, the accepted

solution pK_a's of histidine and aspartate are 6.04 and 3.90, respectively. Although the Stivers group attributed the 2 pH unit change in pK_a of Asp145 to be due to environmental effects, Luo *et al*. calculated the apparent pK_a of Asp145 to be 4.2 in the enzyme environment.[7]
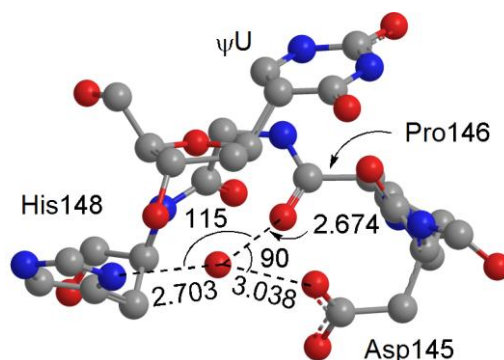
Figure 5.2 *Important distances and angles in the 1EMH crystal structure involving the nucleophilic water molecule. Distances are given in Å, and angles in degrees.*

Second, analysis of the 1EMH reactant mimic indicates that the water nucleophile may be hydrogen bonding to His148 and Pro146, rather than Asp145.[8] Figure 5.2 shows a close up of the active site, where one can see that the distances from the water to His148 and the backbone of Pro146 are ~2.7 Å, while the distance to Asp145 is ~3.0 Å. In addition to the distances, the angle between the histidine and proline residues is 115°, while the angle between the aspartate and proline is ~90°. Therefore, the tetrahedral shape of the water molecule fits better when it is not directly interacting with the aspartate. In the case of the hydrogen bond between His148 and the water nucleophile, either component could be the proton donor. However, a recent NMR study by Stivers indicates that the histidine residue is neutral in the unbound state, and therefore water is most likely the proton donor in this interaction.[9]

Third, His148 is donating a hydrogen bond to a nearby phosphate group and crystal structure water molecule, which increases its proton affinity and makes it a good activator. In contrast, Asp145 is positioned to accept a hydrogen bond from Gln144, which would decrease its proton affinity.

Collectively, the above discussion suggests that His148 could be activating the nucleophilic water molecule. Investigating this scheme would answer both the question regarding the protonation state of His148 and the catalytic role of Asp145.

5.3.2.2 Proposed Mechanism

The ideas presented in the previous section suggest that one could design a mechanism of UDG action whereby the nucleophilic water molecule is activated by the neutral His148 rather than Asp145. In the new mechanism, the dissociation of the uracil moiety is catalyzed by the aspartate residue. Specifically, the negatively charged Asp145 will stabilize the charge buildup on the sugar since the intermediate will contain the standard $(- + -)$ electrostatic sandwich used by many enzymes. From Chapter 2, it is clear that the charge-separated products (namely the oxacarbenium cation and uracil anion) are highly unstable without charge stabilization and therefore, the intermediate would be too short-lived to be observed experimentally. The next step, association of the water nucleophile, is catalyzed by the neutral His148 residue. By deprotonating the nucleophilic water, histidine gains a positive charge, which will also increase the hydrogen-bond strength to nearby phosphate groups. The increase in binding strength to the phosphate backbone agrees with experiments showing that UDG binds strongly to abasic sites.[10]
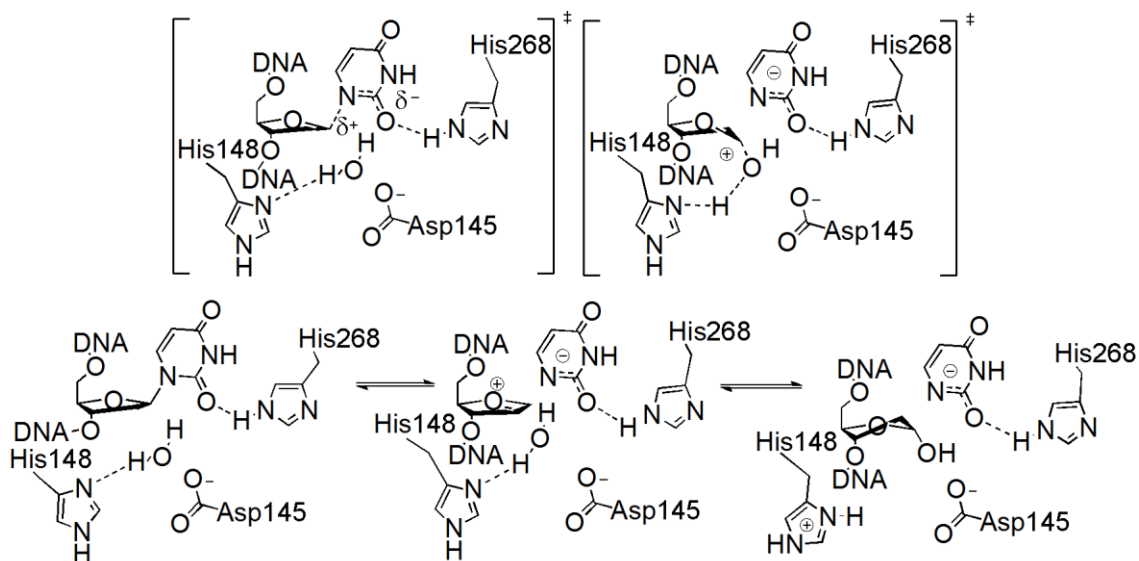
Figure 5.3 *The newly proposed UDG mechanism where the initial step, uracil dissociation, is catalyzed by the electrostatic sandwich created by Asp145, and the second step, association of the water nucleophile, is catalyzed by a neutral His148.*

This proposal is in line with mutational studies indicating that mutating Asp145 decreases the activity to < 1%, while mutating his148 decreases the activity to 57% of the specific activity of the wild type enzyme (eUDG). Specifically, mutations that remove the anionic character of the Asp145 residue will not catalyze the first, dissociative, step, and therefore no reaction will occur. However, once dissociation has occurred, the association step can take place without an enzyme being present since solvent molecules can accept the proton from the nucleophile, similar to the results for the three water model of dU hydrolysis in Chapter 2. Therefore, mutations involving His148 will not completely remove the catalytic activity of the enzyme, since the largest barrier corresponding to dissociation of uracil will generally not be affected by the mutation. Thus, the new mechanism proposes that Asp145 is the major catalytic residue not because it activates the nucleophile, but because it stabilizes the dissociation product. Therefore, the

mechanism outlined in Figure 5.3 is fully consistent with experimental data that indicates

the importance of Asp145, and that suggests the activating residue is a histidine.

5.3.2.3 Proposed UDG Studies

The above mechanism could be studied in the same fashion as outlined in Chapter

4. However, it would also be of interest to generate a three-dimensional reaction surface

as presented in Chapters 2 and 3. This would allow not only the examination of any other

possible mechanisms, but also a comparison to the surface generated by Dinner et al.

using QM/MM.[4] In addition, this model could be examined with a 3-layer ONIOM

methodology. A 3-layer model could include the remainder of the enzyme and nucleic

acid, and treat these extremities with molecular mechanics. Ideally, the new model would

include both His148 and Asp145 in the high-level region. This would remove bias

introduced by using a model that does not allow proton transfer across the level divisions.

In addition, to properly model the acidity of His148, an additional phosphate group (U +

2) must be incorporated into the quantum mechanical levels that hydrogen bonds to $N\varepsilon_2$,

and thereby increases the $N\delta_1$ proton affinity. In order to reduce the size of the system,

the U + 1 nucleobase in the crystal structure (T in 1EMH), which does not directly

interact with the remainder of the model, may be truncated to an amino group.

A secondary study could also take the above large model and make point

mutations which have previously been made in the literature. Examination of the effect of

these mutations, both structurally and mechanistically, could give additional, qualitative,

discussion to the experimentally observed changes in binding affinity and specific

activity of the mutants compared to wild-type UDG. Novel mutations could also be

carried out to possibly determine experiments that could verify the newly proposed

mechanism. For example, mutating His148 to an asparagine residue might provide interesting insight into the role of this residue.

5.3.2.4 Extension to Other Systems

In addition to reexamining the mechanism of UDG action, the methodology utilized in Chapter 4 could be extended to other glycosylases. For example, the stepwise search for transition states could be applied to other monofunctional glycosylases such as MutY, which removes alkylated adenine residues, or bifunctional enzymes such as hOGG1, which removes C8-oxidized guanine. It would also be interesting to carry out a study whereby the Asp145 residue of UDG is mutated to a lysine, to determine if it is possible to convert UDG from a monofunctional to a bifunctional glycosylase through a point mutation.

5.4 Concluding Remarks

In conclusion, research presented in this thesis was able to generate a working model for nucleoside hydrolysis, and shed light on deceptive energies obtained by adding solvent energy corrections to gas-phase geometries. This is a very important finding since many studies use gas-phase structures to obtain energy barriers (which may be coincidentally close to experiment) and to draw mechanistic conclusions. The enzymatic study presented in this thesis utilized the concepts derived from the small hydrolysis models to guide the construction of a medium-sized enzyme model for the study of uracil-DNA glycosylase. During the course of this mechanistic study, a great deal of insight into the use of the ONIOM methodology to understand the mechanism of action of DNA glycosylases was gained, which can be applied to other systems of interest. Finally, studying the accepted mechanism of UDG action uncovered an interesting

direction for further research involving a new mechanistic proposal that may resolve

discrepancies that have developed in the literature.

5.5 References

(1)   Chandra, A. K.; Nguyen, M. T.; Uchimaru, T.; Zeegers-Huyskens, T. *J. Phys. Chem. A*, **1999**, *103*, 8853–8860.

(2)   Slupphaug, G.; Eftedal, I.; Kavli, B.; Bharati, S.; Helle, N. M.; Haug, T.; Levine, D. W.; Krokan, H. E. *Biochemistry* **1995**, *34*, 128–138.

(3)   Parikh, S. S.; Mol, C. D.; Slupphaug, G.; Bharati, S.; Krokan, H. E.; Tainer, J. A. *EMBO J.* **1998**, *17*, 5214–5226.

(4)   Dinner, A. R.; Blackburn, G. M.; Karplus, M. *Nature (London)* **2001**, *413*, 752–755.

(5)   Mol, C. D.; Arvai, A. S.; Slupphaug, G.; Kavli, B.; Alseth, I.; Krokan, H. E.; Tainer, J. A. *Cell (Cambridge, Mass.)* **1995**, *80*, 869–878.

(6)   See, for example; (a) Ma, A.; Hu, J.; Karplus, M.; Dinner, A. R. *Biochemistry* **2006**, *45*, 13687–13696; (b) Drohat, A. C.; Jagadeesh, J.; Ferguson, E.; Stivers, J. T. *Biochemistry* **1999**, *38*, 11866–11875.

(7)   Luo, N.; Mehler, E.; Osman, R. *Biochemistry* **1999**, *38*, 9209–9220.

(8)   Parikh, S. S.; Walcher, G.; Jones, G. D.; Slupphaug, G.; Krokan, H. E.; Blackburn, G. M.; Tainer, J. A. *Proc. Natl. Acad. Sciences U.S.A.* **2000**, *97*, 5083–5088.

(9)   Parker, J. B.; Stivers, J. T. *Biochemistry* **2008**, *47*, 8614–8622.

(10)   Dianov, G.; Price, A.; Lindahl, T. *Mol. Cell. Biol.* **1992**, *12*, 1605–1612.

Appendix A

Reaction Profile Generation for dU

Each point on the hydrolysis surfaces was generated from a neighboring point. The parent point for each new geometry is listed in the following three tables.

Legend: The cell contents specifies from which direction the new structure was derived, for example, the structure corresponding to a glycosidic and nucleophilic distance of 1.7 and 3.4 Å, respectively, was generated from the structure to its left (glycosidic bond length of 1.5 Å).

Left: L	Below: B	Initial: I	
Right: R	Above: A		

Table A1 Parent structures for the gas-phase optimized hydrolysis model with one discrete water molecule.

Glycosidic Bond Length (Å)

	1.3	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1
1.2						B	B	L	L	B	B	B	B	B	L
1.4						B	B	L	L	B	B	B	B	B	L
1.6						B	B	B	B	B	B	B	B	B	L
1.8						B	B	B	B	B	B	B	B	B	L
2.0	B	B	B	B	B	B	B	B	B	B	B	B	B	B	L
2.2	B	B	B	L	B	B	B	B	B	L	B	B	L	L	L
2.4	B	B	B	L	L	L	B	L	B	L	L	L	L	B	L
2.6	B	B	B	L	L	L	L	L	B	L	L	L	B	L	L
2.8	B	B	B	L	L	L	L	L	B	L	L	L	B	L	L
3.0	B	B	B	L	L	B	L	L	L	L	B	B	L	L	L
3.2	B	B	B	L	L	B	L	L	L	L	L	L	L	L	L
3.4	R	I	L	L	L	L	L	L	L	L	L	L	L	L	L
3.6	A	A	A	L	L	L	L	L	L	L	L	L	L	L	L
3.8	A	A	A	L	L	L	L	L	L	L	L	L	L	L	L

Nucleophile-C1' Distance (Å)

Table A2   Parent structures for the solvent-phase (water) optimized hydrolysis model with one discrete water molecule.

Glycosidic Bond Length (Å)

Nucleophile-C1' Distance (Å)	1.3	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1	4.3
1.2						B	B	B	R	B	R	B	B	B	B	B
1.4						B	B	B	B	B	R	B	L	L	L	L
1.6						B	B	B	B	R	R	B	L	L	L	L
1.8						B	L	L	B	B	A	A	L	L	L	L
2.0	R	R	B	L	B	L	L	L	B	B	A	A	L	L	L	L
2.2	R	R	B	L	B	L	L	A	B	B	B	B	B	B	B	L
2.4	B	B	B	L	B	L	L	B	B	B	B	B	B	B	B	L
2.6	R	B	B	R	B	L	L	B	B	B	B	B	B	B	B	L
2.8	B	B	L	B	B	A	R	B	B	B	B	B	L	B	B	L
3.0	R	B	L	B	L	L	L	A	B	B	L	L	L	B	B	L
3.2	B	B	B	B	B	B	L	L	L	L	L	L	L	L	L	L
3.4	R	**I**	L	L	L	L	L	L	L	L	L	L	L	L	L	L
3.6	R	A	A	A	A	A	L	A	B	A	A	A	L	L	L	L
3.8	A	L	R	A	A	A	A	A	L	A	A	A	A	A	A	L

Table A3 Parent structures for the solvent-phase (water) optimized hydrolysis model with three discrete water molecules.

Glycosidic Bond Length (Å)

	1.3	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1
1.2						B	L	L	B	B	L	L	L	L	B
1.4						B	L	L	B	B	B	R	B	L	L
1.6						B	L	L	B	B	B	B	L	B	B
1.8						B	L	L	B	B	B	B	B	L	B
2.0	B	B	B	L	B	B	R	B	B	B	B	B	B	R	B
2.2	B	B	B	L	B	L	L	B	B	B	B	B	B	B	B
2.4	B	B	B	L	B	B	L	B	B	B	B	B	B	B	B
2.6	B	B	B	B	B	B	L	L	B	B	B	B	B	B	L
2.8	B	B	L	L	B	B	L	L	B	B	B	B	B	B	L
3.0	B	B	B	A	B	B	L	L	B	B	B	L	B	L	L
3.2	B	B	L	L	B	B	B	L	B	B	L	L	L	L	L
3.4	R	I	L	R	B	L	L	L	L	L	L	L	L	A	L
3.6	A	A	A	B	L	L	A	A	A	A	L	L	L	L	L
3.8	A	A	L	L	L	L	A	L	A	A	A	A	A	A	L

Nucleophile-C1' Distance (Å)

Table A4 Cartesian coordinates for the constrained starting structure (Initial, Tables A1-A3) for each model utilized.

A. Gas-phase optimized uni-molecular cleavage model.

O	-2.09676200	2.27906900	0.48241300
C	-2.86827100	1.74243400	-0.59843200
C	-2.61244300	0.24645700	-0.69983600
O	-1.25601200	0.01385900	-1.12233600
C	-0.57400800	-0.82784900	-0.20444400
N	0.83662900	-0.33362700	-0.07855400
C	1.08130400	1.01049400	0.11313900
C	2.33078100	1.52179500	0.21145700
C	3.48580900	0.64517400	0.10928600
O	4.66348100	0.97449800	0.17937200
N	3.13750900	-0.70674500	-0.08783300
C	1.87413000	-1.26463700	-0.18169700
O	1.69257800	-2.46590800	-0.33119900
C	-2.80938800	-0.54192100	0.61205700
C	-1.37446700	-0.76155900	1.10229800
O	-3.46988100	-1.75993800	0.25450100
H	-3.94279800	1.90608800	-0.42829700
H	-2.58504600	2.21495900	-1.54849700
H	-3.28522200	-0.16624900	-1.46122600
H	-0.48851800	-1.84259800	-0.59407200
H	0.18937400	1.62598000	0.16793400
H	2.50084300	2.58042400	0.35905000
H	3.90705800	-1.36503000	-0.16035700
H	-3.41200000	0.02009300	1.33581400
H	-1.05459400	0.09830400	1.69962400
H	-1.25819100	-1.66948600	1.70247800
H	-3.58751900	-2.30413200	1.04917700
H	-2.27027200	3.22928400	0.55809900

B. Solvent-phase (water) optimized uni-molecular cleavage model.

O	-1.95786200	2.30398000	0.43242900
C	-2.77914600	1.77685700	-0.61678400
C	-2.60597200	0.26972000	-0.70169100
O	-1.25904700	-0.05425300	-1.14109800
C	-0.59455600	-0.87522000	-0.18479300
N	0.81781600	-0.38347200	-0.06909100
C	1.03572400	0.96108100	0.13989200
C	2.27560100	1.50210700	0.23686100
C	3.43806700	0.65171100	0.11770500
O	4.61977100	1.01183600	0.18622900
N	3.12873100	-0.69611000	-0.09341200
C	1.87612100	-1.27697500	-0.19090900
O	1.73920800	-2.48776200	-0.36351800
C	-2.82171700	-0.49733900	0.61517300
C	-1.39690600	-0.75593600	1.11622700
O	-3.51213400	-1.71407500	0.28613100
H	-3.84201100	1.99286400	-0.42343900
H	-2.50225500	2.22405700	-1.58337600
H	-3.31020600	-0.10752000	-1.46832100
H	-0.55069100	-1.91843200	-0.55454400
H	0.13196800	1.55985600	0.21146300
H	2.41404000	2.56616900	0.40022900
H	3.91890700	-1.34499500	-0.18039100
H	-3.41548500	0.08489900	1.33518800
H	-1.05118100	0.09921600	1.71370000

```
H                        -1.32388500   -1.66472000    1.73039000
H                        -3.63376700   -2.22799000    1.11518000
H                        -2.06404000    3.27981900    0.44996900
```

C. Gas-phase optimized hydrolysis model with one explicit water molecule.

```
H                         0.54301700    1.30013800   -0.86817600
H                         1.11906600    1.33344800    1.54279300
H                         0.99505800   -0.44878800    1.61235100
H                         3.21179900   -0.45687600   -1.48450000
H                         3.92301800   -2.32060700   -0.10790400
H                         2.62524100   -2.83875800   -1.21936600
H                         2.28628000   -3.53966000    1.02223800
O                         2.05203700   -2.63183700    0.77826700
C                         2.85123200   -2.22365100   -0.33693900
O                         1.18304300   -0.64522600   -1.10608400
C                         0.53251000    0.34796900   -0.33933100
H                        -0.50345600   -2.04109100   -0.14518100
H                        -2.91317900   -2.72230000    0.11423600
H                        -3.81583300    1.40277700    0.04224200
C                        -3.65107500   -0.66618600    0.08424800
O                        -4.85607600   -0.84432100    0.19868400
N                        -3.13770100    0.64891100   -0.01216600
C                        -1.82223200    1.03744800   -0.12730400
O                        -1.48858400    2.22643700   -0.15169600
N                        -0.91787200   -0.01113100   -0.20727000
C                        -1.31740800   -1.32672900   -0.09906200
C                        -2.61598800   -1.68440700    0.03883500
C                         2.55882800   -0.76737500   -0.66083800
C                         2.72618100    0.23579900    0.51245800
C                         1.28471500    0.40485900    0.99201700
H                         2.64003500    2.17677900    0.08740400
O                         3.29519800    1.45233100    0.03779000
H                         3.38373500   -0.17077200    1.28835400
O                         1.10126200    3.45793100    0.15529600
H                         0.16091000    3.19765700    0.05160900
H                         1.15769600    4.39472300   -0.08261100
```

D. Solvent-phase (water) optimized hydrolysis model with one explicit water molecule.

```
H                         0.52421200    1.16270100   -1.13703000
H                         1.03649000    1.43502700    1.36739500
H                         0.94287600   -0.33232000    1.56380100
H                         3.18946100   -0.54462600   -1.51070800
H                         4.03735900   -2.22040300   -0.01900600
H                         2.72326700   -2.91816300   -1.00655400
H                         2.54032200   -3.46117100    1.27471500
O                         2.24093700   -2.56862500    0.99639200
C                         2.94933500   -2.21029400   -0.19392300
O                         1.16569200   -0.80293000   -1.07342200
C                         0.50258600    0.28191500   -0.45216700
H                        -0.60356200   -2.10326700   -0.38877900
H                        -3.01339400   -2.72809100   -0.06451200
H                        -3.79182300    1.42637700    0.20524700
C                        -3.67542900   -0.64824800    0.09405400
O                        -4.89135200   -0.80570400    0.25222100
N                        -3.13725100    0.64671500    0.06699200
C                        -1.81514500    0.99957000   -0.08666200
O                        -1.44243800    2.17890000   -0.04894200
N                        -0.94742900   -0.06234600   -0.28204900
C                        -1.38470700   -1.36406100   -0.25113600
```

```
C                    -2.69038100   -1.69193000   -0.07630100
C                     2.56276600   -0.81117600   -0.63769100
C                     2.67991800    0.29633600    0.43881600
C                     1.23054500    0.47706300    0.87881300
H                     2.57705800    2.23463400   -0.02748000
O                     3.21134200    1.49640500   -0.15031600
H                     3.34101800   -0.01357300    1.26005800
O                     1.15576900    3.54553800    0.24216600
H                     0.24691400    3.20523000    0.10635400
H                     1.16602900    4.44334600   -0.14572300
```

E. Solvent-phase (water) optimized hydrolysis model with three explicit water molecules.

```
H                     0.83901600    0.57419100   -0.35600700
H                     1.26505400   -0.16380200    1.91076100
H                     0.00366300   -1.41885800    1.79899200
H                     2.16128300   -2.46986700   -1.31065300
H                     1.29600000   -4.53123600   -0.41813000
H                     0.16906300   -3.88251000   -1.64403300
H                    -0.94553200   -4.58454900    0.29261900
O                    -0.48080400   -3.72077600    0.33467700
C                     0.57598500   -3.72574800   -0.63345900
O                     0.44459000   -1.32924800   -1.05060100
C                     0.30196100   -0.32574000   -0.05837600
H                    -1.73195700   -1.91669400    0.11111500
H                    -4.18391800   -1.40271000    0.14980200
H                    -3.08204900    2.67692000   -0.10192900
C                    -3.87854000    0.76045200    0.01567200
O                    -5.04840700    1.15900700    0.01739800
N                    -2.82659400    1.68343600   -0.05948300
C                    -1.47423400    1.41804100   -0.06853800
O                    -0.63628900    2.32368300   -0.11844300
N                    -1.14295700    0.07421700   -0.01098300
C                    -2.11409900   -0.90019000    0.07024000
C                    -3.43980400   -0.61496000    0.08715100
C                     1.31898200   -2.40010000   -0.59600900
C                     1.84923400   -1.97054000    0.79554900
C                     0.82052700   -0.92475300    1.25456900
H                     3.23932900   -0.67696900    0.16682700
O                     3.18654700   -1.47528100    0.74684100
H                     1.88200700   -2.82517800    1.48193200
O                     3.41622300    0.73267100   -0.91927100
H                     4.15324400    0.61045800   -1.54915800
H                     3.56301700    1.62019200   -0.50156600
O                     1.24585800    4.41443000    0.10686100
H                     0.56632700    3.71963800   -0.01192700
H                     1.14818300    5.01236900   -0.66111500
O                     3.72765200    3.21992300    0.20698800
H                     3.99492600    3.18940600    1.14655300
H                     2.84616800    3.67158500    0.19278300
```

Appendix B

Reaction Profile Generation for Natural Nucleoside Hydrolysis

Each point on the hydrolysis surfaces was generated from a neighboring point. The parent point for each new geometry is listed in the following three tables.

Table B1 Parent structures for the hydrolysis of dA.

Glycosidic Bond Length (Å)

	1.3	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1
1.2						B	B	B	B	R	B	B	L	B	L
1.4						B	B	B	B	R	B	B	L	B	B
1.6						B	B	B	B	B	B	B	B	B	L
1.8						B	B	L	L	B	B	B	L	B	B
2.0	B	B	B	B	B	B	B	L	L	B	B	B	B	B	B
2.2	B	B	B	L	B	B	B	L	L	B	B	B	B	B	B
2.4	B	B	B	L	B	B	L	L	L	B	B	B	B	B	L
2.6	B	B	L	L	L	L	L	L	L	L	L	L	L	L	L
2.8	B	B	L	L	L	L	L	L	L	L	L	L	L	L	L
3.0	R	B	L	L	L	L	L	L	L	L	L	L	L	L	L
3.2	R	B	L	L	L	B	L	A	L	L	L	L	L	L	L
3.4	R	I	L	L	L	L	A	A	A	A	A	L	A	L	L
3.6	R	A	A	A	A	L	A	A	A	A	A	L	A	A	L
3.8	A	A	A	A	A	A	A	A	A	A	A	A	A	A	L

Nucleophile-C1' Distance (Å)

Legend: The cell contents specifies from which direction the new structure was derived, for example, the structure corresponding to a glycosidic and nucleophilic distance of 1.7 and 3.4 Å, respectively, was generated from the structure to its left (glycosidic bond length of 1.5 Å).

Left:	L	Below:	B	Initial:	I
Right:	R	Above:	A		

Table B2   Parent structures for the hydrolysis of dC.

Glycosidic Bond Length (Å)

	1.3	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1	4.3
1.2							B	B	B	B	B	B	L	L	L	L
1.4						B	B	B	B	B	B	B	B	B	L	L
1.6						B	B	B	B	B	B	B	B	B	L	B
1.8						B	L	B	B	B	B	B	B	B	B	B
2.0	B	B	B	B	L	L	L	B	B	B	B	B	B	B	B	B
2.2	B	B	B	L	L	L	A	B	B	B	B	B	B	B	B	B
2.4	B	B	B	L	B	A	R	B	B	L	L	B	L	B	B	B
2.6	B	B	B	B	L	L	L	B	B	L	L	B	L	B	B	L
2.8	B	B	L	L	L	B	B	B	B	B	L	L	B	B	L	L
3.0	B	B	L	B	B	L	L	L	L	L	L	L	B	L	L	L
3.2	B	B	L	B	L	L	A	L	L	L	L	L	L	L	L	L
3.4	R	**I**	L	L	L	L	L	L	L	L	L	L	A	L	A	A
3.6	A	A	L	A	A	A	L	L	L	L	L	L	A	A	A	A
3.8	A	A	L	A	A	A	L	L	L	L	L	L	A	A	A	A

Nucleophile-C1' Distance (Å)

Legend: The cell contents specifies from which direction the new structure was derived, for example, the structure corresponding to a glycosidic and nucleophilic distance of 1.7 and 3.4 Å, respectively, was generated from the structure to its left (glycosidic bond length of 1.5 Å).

Left:	L	Below:	B	Initial:	I
Right:	R	Above:	A		

Table B3   Parent structures for the hydrolysis of dG.

Glycosidic Bond Length (Å)

	1.3	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1
1.2							B	B	B	B	L	L	B	B	B
1.4						R	B	B	B	B	L	L	B	B	B
1.6						B	B	B	B	L	B	B	L	B	L
1.8						B	B	B	B	L	B	B	L	B	L
2.0	B	B	B	B	B	B	L	L	B	L	B	L	B	B	L
2.2	B	B	B	B	B	L	L	L	L	B	B	L	B	B	L
2.4	B	B	L	L	L	L	L	L	A	B	L	L	B	B	B
2.6	B	B	L	L	L	L	L	L	L	L	L	L	B	B	B
2.8	B	B	L	A	B	A	R	A	L	L	L	L	L	L	L
3.0	R	B	B	B	B	B	L	L	A	A	L	L	L	L	L
3.2	B	B	L	B	L	L	B	L	L	A	L	L	L	L	L
3.4	R	**I**	L	L	L	L	L	L	L	A	A	A	L	L	L
3.6	R	A	L	L	L	L	A	A	A	L	A	A	A	A	L
3.8	R	A	L	L	L	A	A	A	A	L	A	A	A	A	A

Nucleophile-C1' Distance (Å)

Legend: The cell contents specifies from which direction the new structure was derived, for example, the structure corresponding to a glycosidic and nucleophilic distance of 1.7 and 3.4 Å, respectively, was generated from the structure to its left (glycosidic bond length of 1.5 Å).

Left:	L	Below:	B	Initial:	I
Right:	R	Above:	A		

Table B4   Parent structures for the hydrolysis of dT.

Glycosidic Bond Length (Å)

	1.3	1.5	1.7	1.9	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5	3.7	3.9	4.1
1.2						B	B	B	B	B	B	B	R	B	L
1.4						B	B	B	B	B	B	L	R	B	L
1.6						B	B	B	B	B	B	L	B	B	B
1.8						B	L	B	B	B	B	B	L	B	L
2.0	B	B	B	L	L	L	B	B	B	B	B	B	R	B	L
2.2	B	B	B	L	L	L	L	B	B	B	B	B	B	B	B
2.4	B	B	B	A	B	B	B	L	L	L	B	L	L	L	L
2.6	B	B	B	L	B	B	L	L	L	L	L	L	L	L	L
2.8	B	B	L	R	B	B	L	L	L	L	L	L	L	L	L
3.0	B	B	L	R	B	B	L	B	B	L	L	L	A	L	L
3.2	R	B	L	L	L	L	L	L	L	L	L	L	L	L	L
3.4	R	**I**	L	L	L	L	L	A	L	L	L	A	A	L	L
3.6	R	A	A	L	L	L	L	A	A	A	A	A	A	A	A
3.8	R	A	L	L	L	L	L	A	A	A	A	A	A	A	A

(Row axis label: Nucleophile-C1' Distance (Å))

Legend: The cell contents specifies from which direction the new structure was derived, for example, the structure corresponding to a glycosidic and nucleophilic distance of 1.7 and 3.4 Å, respectively, was generated from the structure to its left (glycosidic bond length of 1.5 Å).

Left:	L	Below:	B	Initial:	I
Right:	R	Above:	A		

Table B5 Cartesian coordinates for the constrained starting structure (Initial, Tables B1-B4) for each nucleoside.

A. 2′-Deoxyadenosine Reactant.

H	0.76947600	0.25938200	-0.29428400
H	1.41704100	-0.37063900	1.95785200
H	0.47219300	-1.88013300	1.85612600
H	2.75803600	-2.48268900	-1.27003800
H	2.32517800	-4.64358800	-0.32037200
H	1.06702700	-4.25571200	-1.52847000
H	0.17391000	-5.12695900	0.46233300
O	0.43964400	-4.18200500	0.46200400
C	1.45609400	-3.99995600	-0.53095700
O	0.86572800	-1.67599700	-1.00298300
C	0.48111300	-0.75164000	0.00309700
C	1.92864300	-2.55388900	-0.53990800
C	2.38118100	-1.99292500	0.82859100
C	1.14655500	-1.21070600	1.30534000
H	3.44035300	-0.48486600	0.05434700
O	3.56027400	-1.19405600	0.73270900
H	2.63208700	-2.80234200	1.52464800
O	3.27268500	0.76698800	-1.20557700
H	3.59135100	0.45527100	-2.07501900
H	3.72629700	1.63446200	-1.04687600
O	0.54813000	2.98325800	1.35743200
H	-0.13169200	2.55299300	0.78301800
H	0.04616000	3.40166000	2.08460100
O	4.43124700	3.21817800	-0.75741300
H	5.15569800	3.19818300	-0.10181600
H	3.72246300	3.78796900	-0.35898300
N	-1.01756200	-0.73188700	0.06295300
C	-1.83742300	-1.83907200	0.12148400
C	-1.85078300	0.35707500	-0.06179000
H	-1.41249600	-2.83171100	0.20804600
N	-3.12036900	-1.54782700	0.04485300
C	-3.14848600	-0.16772000	-0.07271700
N	-1.53177500	1.66517900	-0.14439100
C	-4.21045000	0.75604800	-0.20745700
C	-2.61718800	2.44086000	-0.27294200
N	-5.50521100	0.38933500	-0.22764500
N	-3.90434300	2.07275000	-0.31425900
H	-2.43073600	3.51169600	-0.34777700
H	-6.23022800	1.09183500	-0.36910200
H	-5.77644500	-0.59165200	-0.19410500
O	2.40607200	4.71226000	0.32321800
H	1.97089000	5.29566700	-0.32897200
H	1.70247600	4.09085800	0.64269900

B. 2′-Deoxycytidine Reactant.

```
H                -0.76230600     0.67101100     0.60173000
H                -1.37179400     0.24740300    -1.71563400
H                -0.32582900    -1.19225900    -1.77837200
H                -2.50935400    -2.21378800     1.31224800
H                -2.03042000    -4.27849700     0.18133800
H                -0.75304300    -3.95058800     1.38701600
H                 0.13034500    -4.61499800    -0.67875700
O                -0.18388800    -3.68791700    -0.60638600
C                -1.17758400    -3.62497100     0.42517100
O                -0.63971400    -1.33995700     1.07406900
C                -0.38017600    -0.25920900     0.18457100
H                 1.37018300    -2.08880400    -0.34933000
C                 3.79886600     0.26091900     0.00778900
N                 3.02764600     1.31688200     0.29707800
C                 1.67912600     1.18862200     0.35567100
O                 0.93096400     2.15411700     0.63119300
N                 1.10583000    -0.07253100     0.10128000
C                 1.89451600    -1.15093600    -0.18819900
C                 3.24761200    -1.03661800    -0.24973800
C                -1.69293700    -2.20148700     0.56417200
C                -2.19968100    -1.54820000    -0.74644600
C                -1.03420200    -0.63243600    -1.15186100
H                -3.33371900    -0.09881400     0.04178900
O                -3.43973100    -0.86250700    -0.57656200
H                -2.39406400    -2.30895300    -1.51200400
O                -3.23956200     1.32135200     1.12570900
H                -4.03720100     1.39295000     1.68576100
H                -3.18361200     2.17976300     0.63060700
O                -0.38574600     4.13291300    -0.74218100
H                 0.12568700     3.43257300    -0.26962400
H                -0.06891800     4.98184900    -0.37522200
O                -3.05384000     3.69828000    -0.25259400
H                -3.49369100     3.63821500    -1.12330500
H                -2.09796100     3.87787900    -0.44814900
H                 3.87896500    -1.89124200    -0.47639700
N                 5.13009300     0.45012200    -0.03737300
H                 5.76818600    -0.31912400    -0.23618800
H                 5.52246600     1.36977100     0.16283500
```

C. 2′-Deoxyguanosine Reactant.

```
H        0.77620000    0.32988600   -0.38744000
H        1.75778400    0.02876500    1.80549000
H        1.29378100   -1.69255100    1.83128200
H        3.41380800   -1.75509000   -1.47103200
H        3.70906700   -3.89960500   -0.45951000
H        2.28323600   -3.95257900   -1.53519200
H        1.87147600   -4.95365000    0.54346700
O        1.85757700   -3.97274200    0.51007100
C        2.67289900   -3.54773300   -0.58866600
O        1.39538700   -1.51527700   -1.06600600
C        0.82571700   -0.70641200   -0.04626300
C        2.69773600   -2.02892000   -0.67226900
C        3.06658500   -1.29832000    0.64010900
C        1.69669700   -0.88686300    1.20329800
H        3.54841600    0.45039200   -0.20241400
O        3.94988900   -0.19762800    0.42732800
H        3.59613500   -1.96805900    1.32832800
N       -0.60063600   -1.12883700    0.14628700
C       -1.03677500   -2.41349700    0.42009600
C       -1.72962700   -0.37293600   -0.05564100
H       -0.32526000   -3.21195600    0.58923400
N       -2.34616100   -2.52521800    0.41508300
C       -2.80210800   -1.25084500    0.11780700
N       -1.77815100    0.95336500   -0.34741700
C       -4.13350000   -0.75818300   -0.04940100
C       -3.01338100    1.40899600   -0.52423700
N       -4.12530100    0.61524300   -0.39030800
O       -5.21151000   -1.35810700    0.05347000
N       -3.21784500    2.72239600   -0.79469800
H       -2.42061300    3.24731200   -1.15636700
H       -4.13210200    3.02207200   -1.13878400
H       -5.05034600    1.03512700   -0.54159200
O        2.89498900    1.66146700   -1.33903600
H        3.27543100    1.55991000   -2.23329400
H        3.08218200    2.59812800   -1.07130800
O        3.32732000    4.25861000   -0.55254800
H        4.10712000    4.35518200    0.02862900
H        2.54767600    4.53607900   -0.00467000
O        1.13664400    4.95963100    0.93707700
H        0.56548300    5.61239400    0.48652800
H        0.56792100    4.16440500    1.09376500
O       -0.31484500    2.69071500    1.48620000
H       -0.74424700    2.13660900    0.79346700
H       -0.99555300    2.81351300    2.17734400
```

D. Thymidine Reactant.

```
H                -1.20070400    -0.40805900    -0.39420500
H                -1.44748800     0.38219500     1.88492600
H                 0.04615100     1.35067000     1.76860700
H                -1.82573700     2.86846100    -1.32276200
H                -0.55110200     4.67672200    -0.39594300
H                 0.43680900     3.81219200    -1.60800900
H                 1.63487300     4.24503600     0.35577900
O                 0.99562900     3.50038000     0.38037400
C                -0.01528100     3.73738400    -0.60728800
O                -0.40693600     1.37269400    -1.07375400
C                -0.48108300     0.35078000    -0.08978500
H                 1.84001000     1.45697600     0.14692100
H                 2.17795500    -3.30032300    -0.23853500
C                 3.35845300    -1.60638500    -0.03347300
O                 4.41458600    -2.24833000    -0.03665900
N                 2.13720200    -2.27791600    -0.15608400
C                 0.87334500    -1.73057000    -0.17007100
O                -0.13825300    -2.43399400    -0.27857100
N                 0.84168100    -0.35550000    -0.05150000
C                 2.00266400     0.38503800     0.07218000
C                 3.24580300    -0.16090100     0.08643000
C                -1.02731700     2.60232100    -0.60319300
C                -1.64907000     2.27444900     0.77620000
C                -0.85810500     1.03738600     1.22877400
H                -3.26773800     1.29245800     0.12543900
O                -3.05916500     2.05976000     0.71023600
H                -1.51818900     3.11011600     1.47383100
O                -3.74401700    -0.07214100    -0.94675600
H                -4.43782600     0.17025400    -1.59045600
H                -4.05928800    -0.90570500    -0.51111800
O                -2.33002200    -4.07757600     0.31981000
H                -1.53571400    -3.55049100     0.09150800
H                -2.37087800    -4.79387600    -0.34479900
O                -4.52762200    -2.41890000     0.24349700
H                -4.82864700    -2.29932800     1.16550400
H                -3.75494000    -3.03714800     0.29129000
C                 4.49871300     0.66022400     0.21672600
H                 5.15857300     0.51233300    -0.64751900
H                 4.25971300     1.72597600     0.28971300
H                 5.06964800     0.37196300     1.10849500
```