UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

On the origin of the canonical nucleobases : selection pressures and hydrolytic stabilities of N-glycosyl bonds

Permalink

https://escholarship.org/uc/item/0bx9p84v

Author

Rios, Andro C.

Publication Date

2012

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

On the origin of the canonical nucleobases: selection pressures and hydrolytic stabilities of *N*-glycosyl bonds

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Chemistry

by

Andro C. Rios

Committee in Charge:

Professor Yitzhak Tor, Chair
Professor Jeffrey Bada
Professor Stanley Opella
Professor Emmanuel Theodorakis
Professor Jerry Yang

2012

Copyright

Andro C. Rios, 2012

All rights reserved.

The Dissertation of Andro C. Rios is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2012

DEDICATION

To the memories of Leslie E. Orgel (1927–2007) and Stanley L. Miller (1930–2007). It is because of their work in the field of prebiotic and origin of life chemistry that I was inspired to pursue a career in chemistry. And especially to Professor Orgel, thank you for your inspiration and encouragement.

TABLE OF CONTENTS

Signature Page iii

Dedication iv

Table of Contents v

List of Figures vii

List of Tables ix

Acknowledgements x

Vita xiii

Abstract of the Dissertation xvi

Introduction 1

Chapter 1: On the origin of nucleobase composition in nucleic acids 3

The native bases of the genetic alphabet 4

The occurrence of modified bases 5

Why the selection of the native bases? 11

Considerations from prebiotic chemistry and selection pressures 12

Selection during the formation of the first informational polymers? 24

Base selection in an early RNA world? 27

A continuous process of refinement 35

References 36

Chapter 2: Refining the genetic alphabet: a late period selection pressure? 42

Introduction 43

Prebiotic chemistry and alternative bases… 43

The emergence of labile *N*-glycosyl bonds and DNA Repair 46

Greater *N*-glycosyl stability may have aided in the utility of diverse bases in the RNA world 51

Conclusion 56

Acknowledgements 56

References 57

Chapter 3: Hydrolytic stabilities of *N*-glycosyl bonds in modified, damaged and alternative nucleosides 63

Introduction 64

Results 68

Discussion 78

Conclusion 96

Materials and Methods 97

References 100

Chapter 4: Outlook 107

Appendix: Supporting Information for Chapter 3 109

LIST OF FIGURES

Figure 1.1: The native nucleobases of the genetic alphabet 4

Figure 1.2: Watson-Crick base pairing 5

Figure 1.3A: Extensive base modifications 7

Figure 1.3B: Modest base modifications 8

Figure 1.4: Alternative bases and base pairs 10

Figure 1.5: Periods of possible selection pressures 12

Figure 1.6: Examples of prebiotically relevant purines and pyrimidines 13

Figure 1.7: Reaction genealogy of three prebiotic bases 14

Figure 1.8: Factors that can affect prebiotic abundance 16

Figure 1.9: Selection pressures in the assembly of RNA nucleosides 26

Figure 1.10: Genetic fidelity pressures 29

Figure 2.1: Alternative bases in bacteriophage DNA 45

Figure 2.2: Half-lives for spontaneous damage to RNA and DNA 47

Figure 2.3: Spontaneous deglycosylation and subsequent DNA strand cleavage 49

Figure 2.4: Hypothesis diagram illustrating refinement of the genetic alphabet 55

Figure 3.1: Examples of reaction monitoring by UV-vis spectroscopy 69

Figure 3.2: Arrhenius plots of deglycosylation kinetics 72

Figure 3.3: Comparative deglycosylation kinetics 77

Figure 3.4: Alternative bases and base pairs in a four letter alphabet 93

Figure A.1: UV-Vis data and plot of deoxyadenosine kinetics at 15C 110

Figure A.2: UV-Vis data and plot of deoxyguanosine kinetics at 20C 111

Figure A.3: UV-Vis data and plot of deoxy-2,6-diaminopurine kinetics at 20C 112

Figure A.4: UV-Vis data and plot of deoxyinosine kinetics at 20C 113

Figure A.5: UV-Vis data and plot of deoxyisoguanosine kinetics at 50C..... 114
Figure A.6: UV-Vis data and plot of deoxyzebularine kinetics at 25C 115
Figure A.7: UV-Vis data and plot of deoxyxanthosine kinetics at 50C..... 116
Figure A.8: UV-Vis data and plot of adenosine kinetics at 75C..... 117
Figure A.9: UV-Vis data and plot of guanosine kinetics at 75C..... 118
Figure A.10: UV-Vis data and plot of 2,6-diaminopurineriboside kinetics at 65C..... 119
Figure A.11: UV-Vis data and plot of inosine kinetics at 65C..... 120
Figure A.12: UV-Vis data and plot of isoguanosine kinetics at 75C 121
Figure A.13: UV-Vis data and plot of zebularine kinetics at 65C..... 122
Figure A.14: UV-Vis data and plot of xanthosine kinetics at 65C..... 123
Figure A.15: Combined Eyring plots of deoxynucleosides and ribonucleosides..... 126

LIST OF TABLES

Table 1.1: Half-life values of hydrolytic deamination 19

Table 1.2: Comparison of excited-state lifetimes 23

Table 1.3: Base pair stability ranking from DNA thermal denaturation studies 33

Table 1.4: Comparison of ΔpK_a and pK_a-pH correlations 34

Table 2.1: Examples of the dichotomous occurrence of bases in RNA and DNA..... 53

Table 3.1: Nucleosides used in study 67

Table 3.2: Rate constants, half-lives and activation parameters at 37C..... 73

Table 3.3: Rate enhancements and ΔGibbs free energy of activation..... 75

Table 3.4: Entropic comparison of nucleosides kinetics..... 76

Table 3.5: Rate constants and half-lives at 50C and variable pH..... 78

Table 3.6: Comparison of pK_a values with ribonucleoside rate constants 88

Table 3.7: Comparison of pK_a values with deoxynucleoside rate constants 88

Table 3.8: Significant DNA base modifications and native letter substitutions 89

Table A.1: Compilation of rate constants for deoxynucleosides at 0.1M HCl 124

Table A.2: Compilation of rate constants for ribonucleosides at 0.1M HCl..... 125

ACKNOWLEDGEMENTS

First and foremost I would like to thank Professor Yitzhak Tor for giving me the scientific freedom and supportive work environment over the entirety of my doctoral tenure. While supportive and encouraging he was also always an essential source of constant critical feedback that undoubtedly aided in the progress of my work. Additionally, in the classroom, Professor Tor is probably one of the most eloquent, creative, intelligent, patient and engaging educators I have ever witnessed. Serving as his teaching assistant over the past five years in his organic reaction mechanisms course has been an absolutely insightful and rewarding experience in my own development of chemical education. While I expected teaching to be an important aspect of my doctoral education, I never expected how influential the experience in working with him, would be for me.

Two department members who I also would like to acknowledge are professors Ulrich Muller and Haim Weizman. Both of these individuals enriched my knowledge, experience, and overall development as a doctoral student. Professor Muller was always a source of inspiration and feedback in my endeavors into the field of origin of life and even though he was not formally on my committee, I consider him a mentor indeed. Professor Haim Weizman is another highly creative, intelligent and caring educator who influenced me tremendously. I began and ended my teaching assistantship tenure with him and along the way he has given me numerous opportunities for improving my own teaching proficiency. Also his successful endeavors in script writing and production of chemical safety videos, of which I was a cast member on two occasions, gave me an experience that I clearly was not expecting, but to which I am so grateful.

I would like to acknowledge all Tor Group members and especially previous members of the Tor group. I acknowledge Dr. Edith (Phoebe) Glazer, who gave me the opportunity to assist in her rendition of teaching Chemistry 15. I acknowledge Dr. Yun Xie, one of the most talented individuals I met during my time in graduate school, and who trained me in many areas of DNA/RNA chemistry. Dr. Andrew Dix, an individual who I could count on at any time of the day or night and who was a source of constant feedback and input to my ideas and projects, and showed me that I can always streamline things even further. I am still trying of course. And to Dr. Mary Noe, an individual who is also the holder of a diversity of skills, interests, and experiences. Her path is one that makes me realize, I haven't yet lived and gives me the reason to strive. To the current post-doctoral scientists, Dr. Renatus Sinkeldam, Dr. Dongwon Shin, and Dr. Noam Freeman, thank you all for your continued support over the years. To Rich Fair, Lisa McCoy, Ryan Weiss, Kristina Hamill, Partrycja Hopkins and Jasmine Kalsi, thank you for making my experience with the group a great one and good luck to you all in your doctoral ventures. To Sandy Yu, I am so grateful that you were a part of the kinetics project because your contribution improved the project kinetics itself.

I would like to acknowledge my entire family: Mom, Dad, Eon, Temoc, Uncle Eric, and especially to my love, Jamillah Moore and mi hijo, Thomas. Lastly, to the honorary Dr. Velma Moore who gave me tremendous support during the early days of graduate school in allowing me the flexibility to stay late in the lab and travel to important conferences. Gracias Velma!

Chapter 1 is adapted from a manuscript in preparation for an invited contribution to a thematic issue in the *Israel Journal of Chemistry*. The dissertation author is the main author for this work.

Chapter 2 is a minimally modified reprint from: Rios, A.C., and Tor, Y., *Astrobiology,* **2012**, *12,* 884 – 891. The dissertation author is the main author and researcher for this work.

Chapter 3 is adapted from a manuscript in preparation of: Rios, A.C., Yu, H.T., Tor, Y., Hydrolytic stability of *N*-glycosyl bonds in modified, alternative and damaged nucleosides: Trends and implications on the refinement of the genetic alphabet. The dissertation author is the main author and researcher for this work.

VITA

EDUCATION

2006 – 2012	Ph.D., Chemistry, University of California, San Diego
2006 – 2008	M.S., Chemistry, University of California, San Diego
2001 – 2006	B.S., Chemistry, California State University, Sacramento *Summa Cum Laude*

FELLOWSHIPS

2009 – 2010	National Science Foundation (NSF) Graduate STEM Fellowship in K–12 Education
2008	Carl Storm Underrepresented Minority Fellowship, Gordon Research Conference, Origin of Life
2007 – 2009	Molecular Biophysics Training Grant, National Institute of Health (NIH) /UCSD
2006 – 2007	Alliance for Graduate Education and the Professoriate (AGEP) Fellowship, NSF and the Office of Graduate Studies, UCSD
2004 – 2006	American Chemical Society Scholar
2002 – 2004	Ronald E. McNair Scholar, California State University, Sacramento
2001 – 2003	Hispanic Scholarship Fund Scholar

ACADEMIC AWARDS AND HONORS

2012	Teaching Assistant Excellence Award, Department of Chemistry & Biochemistry, UCSD
2012	Graduate Diversity and Outreach Award, Department of Chemistry & Biochemistry, UCSD
2010	American Chemical Society Award for Outstanding Oral Research Presentation, SACNAS National Conference
2007	Honorable Mention, NSF Graduate Research Fellowship
2006	Honorable Mention, The National Academics Ford Pre-doctoral Fellowship
2006	1st Place, Undergraduate Biological and Agricultural Sciences, 20th Annual California State University Student Research Competition, CSU Channel Islands

2006 Dean's Award for the College of Natural Sciences and Mathematics, California State University, Sacramento

TEACHING EXPERIENCE

2006 – 2012 Teaching Assistant for Department of Chemistry & Biochemistry, UCSD. Honors and regular organic chemistry (Chem 140B and 140BH, 140C and 140CH) mechanistic organic chemistry (Chem 154/254), organic chemistry labs (Chem 143AH, 143A, 143B), and chemistry for non-majors (Chem 15)

2003 – 2006 Adjunct course instructor for Science Education Equity Program, California State University, Sacramento, general chemistry (Chem 1B), organic chemistry (Chem 124)

2003 – 2005 Teaching Assistant for Department of Chemistry, California State University, Sacramento, Chemical Calculations (Chem 4)

RESEARCH EXPERIENCE AND FIELDS OF STUDY

2009 – 2012 Dissertation research, advisor: Yitzhak Tor,
Field of study: Physical organic chemistry, Bioorganic chemistry, Astrobiology

2006 – 2009 Synthesis of fluorescent nucleosides, advisor: Yitzhak Tor,
Fields of study: Synthetic organic chemistry, Bioorganic chemistry, nucleic acids

2004 – 2006 Synthesis and thermal reactivity studies of extended enediyne heterocycles, advisor: John Spence
Field of study: Synthetic organic chemistry

2003 – 2004 Method development for aerosol analysis of carbohydrate monomers, advisor: Roy Dixon,
Field of study: Analytical chemistry

SELECTED PRESENTATIONS

2012 9th Annual Yale Bouchet Conference on Diversity in Graduate Education, March 30 – April 1, Yale University, Connecticut
Session: Genetics

2012 243rd American Chemical Society National Meeting. March 25 – 29, Session: From Geochemistry to Biochemistry and the Origin of Life

2011 **Invited talk:** 43rd Western Regional American Chemical Society Meeting, November 10 – 12, Pasadena, California,
Session: Enhancing Student Success through Programmatic and Curriculum Innovation

2011	242nd American Chemical Society National Meeting, August 28 –September 1, Denver, Colorado
2010	Society for the Advancement of Hispanics/Chicanos & Native Americans in Science (SACNAS) National Conference, September 30 – October 3, Anaheim, California
2010	AAAS Annual Meeting, February 18, Session: NSF GK-12 Special Focus Meeting, San Diego, California
2008	Molecular Biophysics Training Program Monthly Speakers. May 13, Urey Room, University of California, San Diego

PUBLICATIONS

Rios, A.C., Yu, H.T., Tor, Y., Hydrolytic stability of *N*-glycosyl bonds in modified, alternative and damaged nucleosides: Trends and implications on the refinement of the genetic alphabet, *in preparation*.

Spence, J.D., **Rios, A.C**., Frost, M.A., McCutcheon, C.M., Cox, C.D., Chavez., S., Fernandez, R.P., Gherman, B.F., Syntheses, Thermal Reactivity and Computational Studies of Aryl–Fused Quinoxalenediynes: Effect of Extended Benzannelation on Bergman Cyclization Energetics, *Journal of Organic Chemistry*, **2012**, *77*, 10329-10339.

Rios, A.C., Tor, Y., Refining the genetic alphabet: A late period selection pressure? *Astrobiology*, **2012,** *12*, 884–891.

Noe, M., **Rios, A.C**., Tor, Y., Design, synthesis and spectroscopic properties of extended and fused Pyrrolo-dC and Pyrrolo-C analogs, *Organic Letters*, **2012** ,*14*, 3150 – 3153.

Rios, A.C., French, G., Introducing bond-line organic structures in high school biology: an activity that incorporates pleasant-smelling molecules. *Journal of Chemical Education*, **2011**, *88*, 954 - 959.

Rios, A.C., Tor, Y. Model Systems: How Chemical biologists study RNA. *Current Opinions in Chemical Biology*, **2009**, *13*, 660-668.

Tor, Y.; Del Valle, S.; Jaramillo, D.; Srivatson, S.; **Rios, A**.; Weizman, H. Designing New Isomorphic Fluorescent Nucleobase Analogues: The Thieno[3,2-d]pyrimidine Core.*Tetrahedron*, **2007**, *63*, 3608-3614.

ABSTRACT OF THE DISSERTATION

On the origin of the canonical nucleobases: selection pressures and hydrolytic stabilities of *N*-glycosyl bonds

by

Andro C. Rios

Doctor of Philosophy

University of California, San Diego, 2012

Professor Yitzhak Tor, Chair

The diversity and complexity of life on this planet is based on a narrow selection of organic molecules. One approach towards understanding the emergence and evolution of life is to address questions concerning the particular selection of these biochemical building blocks. The five nucleobases of DNA and RNA, also known as the genetic alphabet, are the archetypal examples of nature's selection. The origin of these canonical bases is the subject of this dissertation.

The work begins with a detailed discussion and assessment of selection pressures that could have shaped the chemical and biological evolution of the alphabetic bases. This is followed by the introduction of a new hypothetical selection pressure that could have occurred after the transition from the RNA to the DNA world.

The pressure is based on the differences in *N*-glycosidic bond stabilities between RNA and DNA and the consequences of their hydrolytic rupture. A systematic investigation on the comparative stability of *N*-glycosyl bonds (via hydrolytic deglycosylation experiments) in modified, alternative and damaged nucleosides was undertaken. Kinetics studies were performed under the same conditions to measure relative changes in *N*-glycosyl destabilization based on differences in ribo and deoxyribo-glycosidic bonds. Thermodynamic activation parameters were obtained to help explain the observed rate enhancements and relative rates of hydrolysis. A low to mild acidic pH based kinetics study was also undertaken to compare how the deoxynucleosides varied in their relative stabilities. Implications for the occurrence of modified bases in DNA, the exclusion of certain surrogate bases, and for the composition of the genetic alphabet are discussed.

.

INTRODUCTION

The question concerning the origin of life on this planet is one of the most intriguing and difficult problems in science. Given the historical nature of events surrounding this problem it is likely a question that will never be fully understood and, thus, once thought to be outside of the realm of science. However a dedicated few in the early to mid 19th century began hypothesizing and eventually conducting experiments in attempts to address questions relating to the biochemical origin of life (1) From these pioneering ideas and experiments, a new field, called prebiotic chemistry, was born and became highly influential for bringing questions surrounding the origin of life into the scientific realm (2). Investigators from disparate fields of science realized they too could pursue discipline specific questions that contributed to a broader understanding on the origin of life. Geologists, astronomers, solar physicists, geophysicists, microbiologists, evolutionary biologists, molecular biologists, ecologists, organic, analytical, inorganic, and physical chemists, mathematicians and many others, have all made contributions to shaping the field. Today the scientific quest towards understand the origin of life is now a core theme under the interdisciplinary field of science called Astrobiology.

The ideas and work presented in this dissertation stem largely from the physical organic chemistry sub-discipline, but are very much astrobiological in the broader context. Stanley Miller, a pioneer and often called “father of prebiotic chemistry” was seemingly able to contribute easily across many disciplines of chemistry, including physical organic chemistry. His wide contributions to our understanding on the stability of organic compounds and especially from his lab’s

work on nucleic acid components (sugars and nucleobases especially) remain salient examples of the utility of physical organic chemistry to our understanding of chemical evolution. This dissertation is an attempt to further contribute an understanding on specific chemical bonds that played a critical role in the evolution of life.

1. Lazcano A (2010) Historical Development of Origins Research. *Cold Spring Harbor Perspect. Biol.* 2(11):a002089.

2. Bada JL & Lazcano A (2003) Prebiotic Soup--Revisiting the Miller Experiment. *Science* 300(5620):745-746.

CHAPTER 1: On the origin of nucleobase composition in nucleic acids

Abstract

The native bases of RNA and DNA are prominent examples of the narrow selection of organic molecules upon which life is based. How did nature "decide" upon these specific heterocycles? It is presented in this chapter that the composition of nucleobases is likely a result of multiple selection pressures that occurred in chemical and early biological evolution. There is evidence that many types of heterocycles could have been present on the early Earth, but plausible selection pressures in the prebiotic epoch may have included hydrolytic and photochemical stabilities. The persistence of the fittest heterocycles in the prebiotic environment may have given some bases a selective advantage upon incorporation into the first informational polymers. Unfortunately the prebiotic formation of polymeric nucleic acids using the native bases is still a very difficult problem. Alternative views have focused on the idea that the emergence of the RNA world may have included many types of nucleobases as a way around this problem. This is supported by the extensive base modification utilized in extant RNA and the resemblance that many modifications have to observations from prebiotic chemistry. Thus selection pressures in the RNA world could have narrowed the composition of the nucleic acid bases. Two such pressures may have been related to genetic fidelity and duplex stability. In consideration of these possible pressures, the native bases along with other related heterocycles do seem to exhibit a certain level of fitness. Thus, it is likely that other selection pressures would have played a role in the refinement of alphabetic composition.

The native bases of the genetic alphabet

Just five nucleobases, also termed the genetic alphabet, are known to dominate the composition of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), (Figure 1.1). They appear simple, robust and their well known canonical or Watson-Crick base pairing properties (Figure 1.2) are so elegant that considering if Nature could have ever used any other heterocycles seems almost unlikely. Yet, we present in this chapter, and subsequent chapters, that the composition of the genetic alphabet may not have emerged in one pivotal epoch before or at the time of the origin of life. Instead, the alphabetic composition is a product of a continual process of refinement that evolved to its current state. We begin this discussion with the observation on the widespread occurrence of modified bases in the biological world (1).

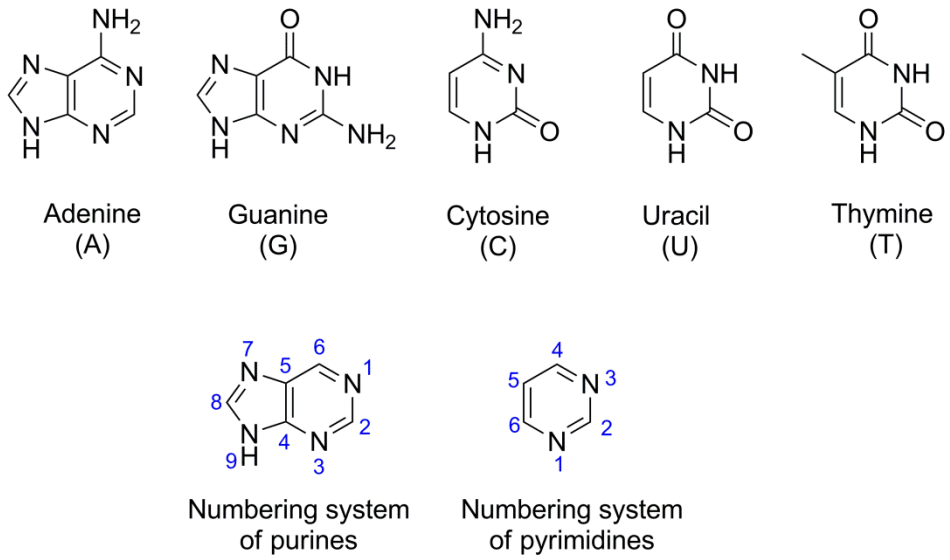


Figure 1.1: The native nucleobases of the genetic alphabet and the numbering schemes of purines and pyrimidines

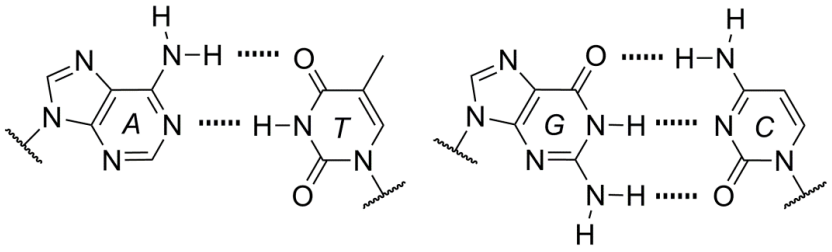


Figure 1.2: The Watson-Crick base pairing relationship of the DNA nucleobases. In RNA the base uracil (U) assumes the place of thymine (T).

The occurrence of modified bases

Nucleobase modification is a ubiquitous post-transcriptional activity found across all domains of life. These transformations are vital to cellular function since they provide the means for genetic expression, editing, proper translation, nucleic acids folding and stability and many other essential functions (1). The diversity of modifications found in nucleic acids can range from extensive alterations (Figure 1.3A) that almost disguise the base entirely, to more modest transformations (Figure 1.3B) that closely resemble the native bases and retain the canonical base pairing. When did the practice of base modification originate in biotic evolution? Were they always post-transcriptional modifications or do they have a more ancient origin (2-4)? As shown in Figure 1.3B, it is often overlooked that thymine itself is a base modification of uracil (5), called 5-methyluracil. Interestingly, of all the DNA nucleotides biosynthesized, thymidine is the only one that is made from another deoxynucleotide. Cellular biosynthesis of thymidine monophosphate (TMP) occurs via the methylation of deoxyuridine monophosphate (dUMP) and it, from the reduction of uridine monophosphate (5). The roundabout way that cells generate TMP coupled to the frequent misincorporation UMP into DNA has led to the idea that DNA, at one

time was based on AGCU (5-6). There is direct evidence of uracil based DNA in extant biochemistry; it is one of the more common native base substitutions found in bacteriophage DNA (7). More intriguing is that other modified bases have also been found to completely replace one of the native bases in bacteriophage DNA. 5-methylcytosine, 5-hydroxymethylcytosine, and 5-hydroxymethyluridine are the most prevalent examples (8-9). The occurrence of these modified bases and others that can serve as viable genetic DNA surrogates, and the idea that thymine may have been a modification that replaced uracil in early DNA evolution, has influenced how investigators think about the origin of the nucleobase composition in the extant alphabet.

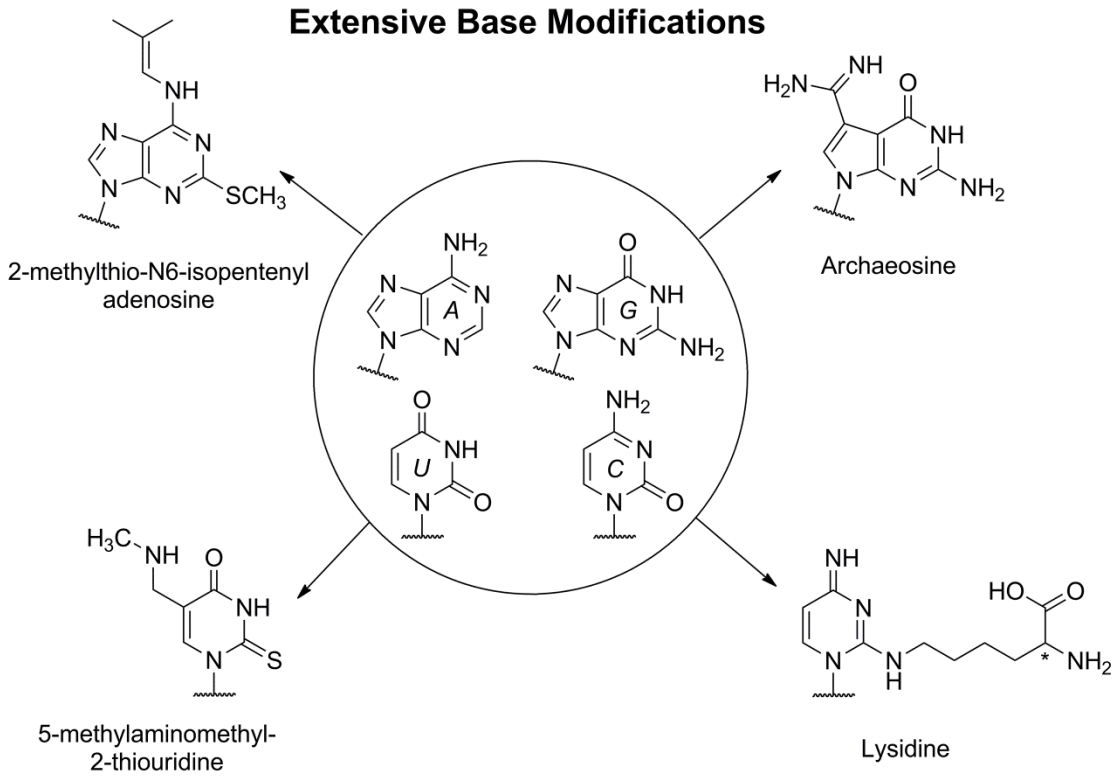


Figure 1.3A: Examples of the extensive base modification that have been identified in biological nucleic acids. Highlighted here are just four intriguing examples originating from the each of the bases of RNA. Adenine is most often alkylated on its N6 amino group, but in this case it has also been altered at the C2 position with an alkylthio group. Guanine has been heavily reworked by the removal of its heterocyclic N7 atom and replaced with a carbon and to which is added a pseudo-extended guanidine group in conjugation to the N9 nitrogen. Cytosine has been aminated at the C2 position with a lysine residue. This affects the tautomer of the amino group at the C4 position. Overall this cytosine modification actually fundamentally changes the heterocycle, as it can be now considered an alkylated 2, 4-diaminopyrimidne. Lastly, uracil, as many pyrimidines usually do, has been substituted at the C5 position. In this case uracil has been amino alkylated at the C5 position in addition to the transformation into a C2 thione.

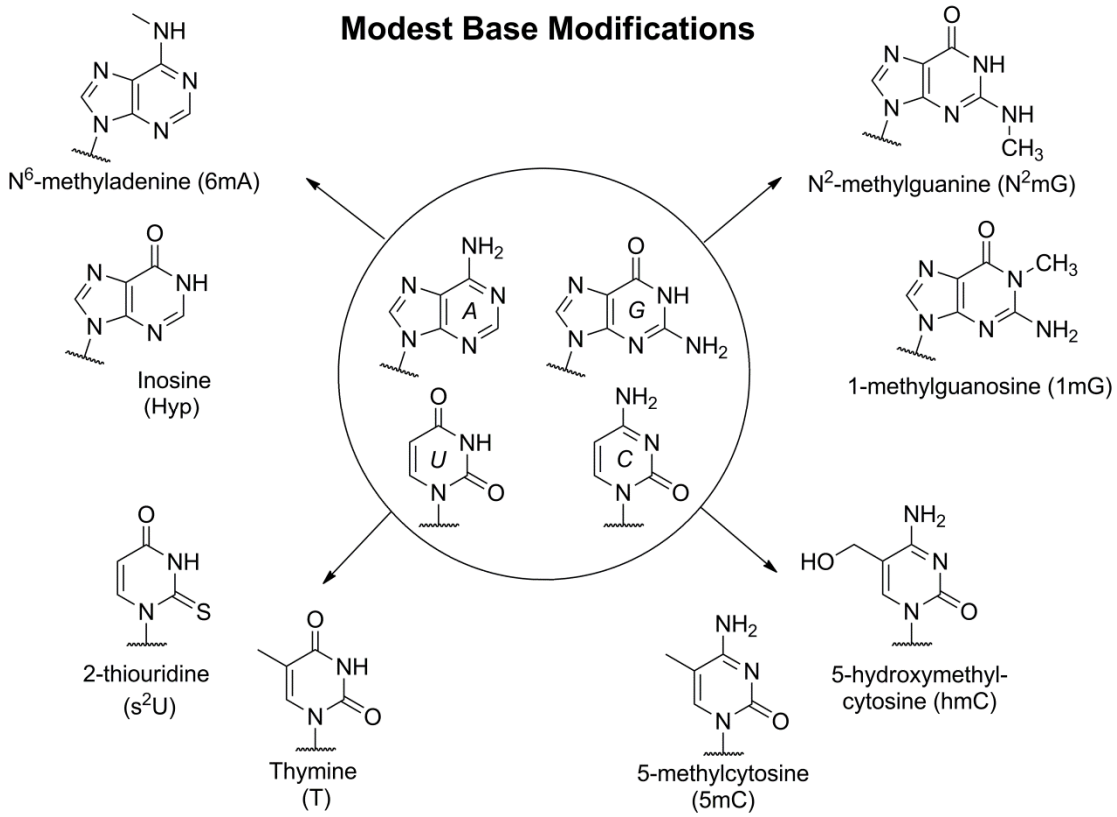


Figure 1.3B: Examples of conservative base modifications that are known to occur more often in RNA, but the pyrimidines in particular, with the exception of, s^2U, are utilized in DNA genomes. Most of the alterations are usually small methylations of the purine exocyclic amino groups (as N2 of guanine and N6 of adenine) or C5 of pyrimidines. Others include the deamination of adenine leading to hypoxanthine (called inosine in RNA) that is an important modification and used as a guanine surrogate.

A major contribution to our understanding of this evolutionary perception was made by Steven Benner in the late 1980's and early 1990's. His lab demonstrated that RNA and DNA polymerases could utilize non-native bases and base pairs and exhibit faithful replication in the presence of native bases (10-11). In short, he demonstrated that nature has the capacity to utilize an expanded genetic alphabet and that the native bases don't necessarily exhibit any special genetic property. Further intrigue came with the identity of the artificial bases used as many appeared to be just as elegant and simple in structure as the native ones. One pair in particular consisted of isomeric structures of guanine and cytosine (called isoG:isoC) that had previously been discussed by Alex Rich in 1962. He hypothesized that isoG and isoC, may have been plausible alphabetic components in early life given the prebiotic feasibility of these heterocyclic structures (12). In fact many types of bases and base pairs have since become known to be considered prebiotically feasible (a topic we discuss shortly) which has added further intrigue to questions surrounding the specific nature of the genetic alphabet. A short and modest list of alternative bases and base pairs to the genetic alphabet that will be relevant to this chapter is given in Figure 1.4. Many of the bases and alternative pairs are drawn from their occurrence in biology, demonstrated utility in synthetic biology and highlight some of the most intriguing examples for considering alternative heterocycle that may have been contenders during the course of chemical and early biological evolution of nucleic acids (13).

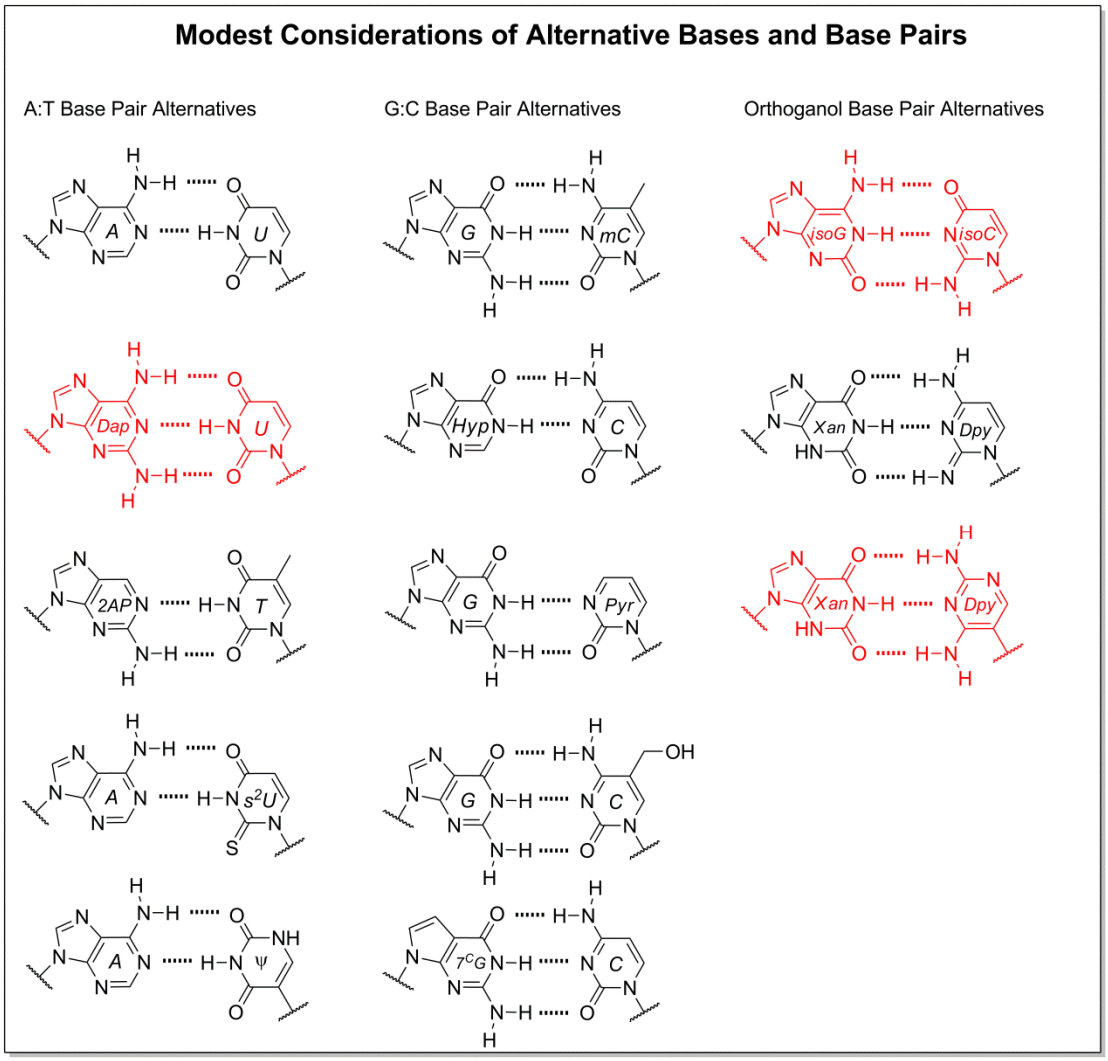


Figure 1.4: Many examples of modified or alternative bases have been considered over the years and shown to be capable of genetic function. The first two columns are bases that have been known to be reliable surrogates for adenine or thymine either in biology or from artificial incorporation. The second column contains the surrogates to guanine or cytosine. The last column represents base pairing relationships that do not occur (at least that we know) in nature. The base pairs colored red in this figure were among those first demonstrated to be enzymatically incorporated into RNA and DNA from the Benner lab.

Why the Native bases and not the others?

The base pairs shown in Figure 1.4 appear to be just as plausible as the native bases. But as organic chemists know all too well, what looks feasible on paper doesn't always translate in practice. Even small changes to structure can have large consequences to the intermolecular, intramolecular and macromolecular "chemical physiology" of nucleic acids (14). However, since nature uses modified bases, often to a high degree (even completely replacing a native letter), it does suggest that not all bases or base pairs would be precluded from an alphabetic role. Numerous and insightful investigations have been pursued to identify plausible selection pressures that might have occurred to narrow our understanding of this evolutionary process. Whether these pressures have been found to be plausible or not, along the way, the community has benefited from the generation of detailed understanding about the bases within the genetic alphabet (13, 15). What follows are contributions from select areas of investigation that have enhanced our understanding for the "fitness" of the native bases in comparison to those that are largely highlighted in Figure 1.4. The pressures we will consider are segmented into periods or events (Figure 1.5) that are currently viewed as essential to the chemical and biological evolution of the genetic alphabet. These pressures are not necessarily stationed in one hypothetical period, since it is quite likely that similar pressures could arise in other periods/events. In the interest of this chapter though, we have associated the relevance of the pressures to a particular period that is most significant to our current state of knowledge.



Figure 1.5: Periods of possible selection pressures. Evolutionary arrow from the formation of the Earth ~4.5 billion years ago and prebiotic chemistry to LUCA based life and extant biology. The timeline is marked by major "events" that could have been crucial to hosting multiple selection pressures that shaped the composition of genetic alphabet.

Consideration from prebiotic chemistry and selection pressures

Results from simulated prebiotic chemistry experiments conducted over the past fifty years (16-20) and the on-going analysis of meteorites(21-22) provides evidence that not only the native bases were likely present only the early earth, but so were many others(Figure 1.6) Furthermore a diverse population doesn't have to originate from separate formation scenarios. Purines and pyrimidines are known to undergo chemical reactions of prebiotic relevance that can enrich the population diversity of nucleobases. Hydrolytic deamination is probably one for the most ubiquitous reaction pathways that increases the population and as shown in Figure 1.7, the entire genetic alphabet and close relatives can be generated from just three simple 'molecular ancestors'. Oxidation, via Fenton chemistry (Fe^{2+} and H_2O_2) to the C8/C2 position of the purines and C5 position of pyrimidines is another way that modified or core bases can result. Outside of deamination, electrophilic aromatic substitution to the C5 position of the pyrimidines is probably the second most important way that modified pyrimidines could have populated the landscape (23). The pathway has also been hypothesized to have opened up the possibility for methylation of uracil in a prebiotic reaction (24-25). Many more prebiotic scenarios for

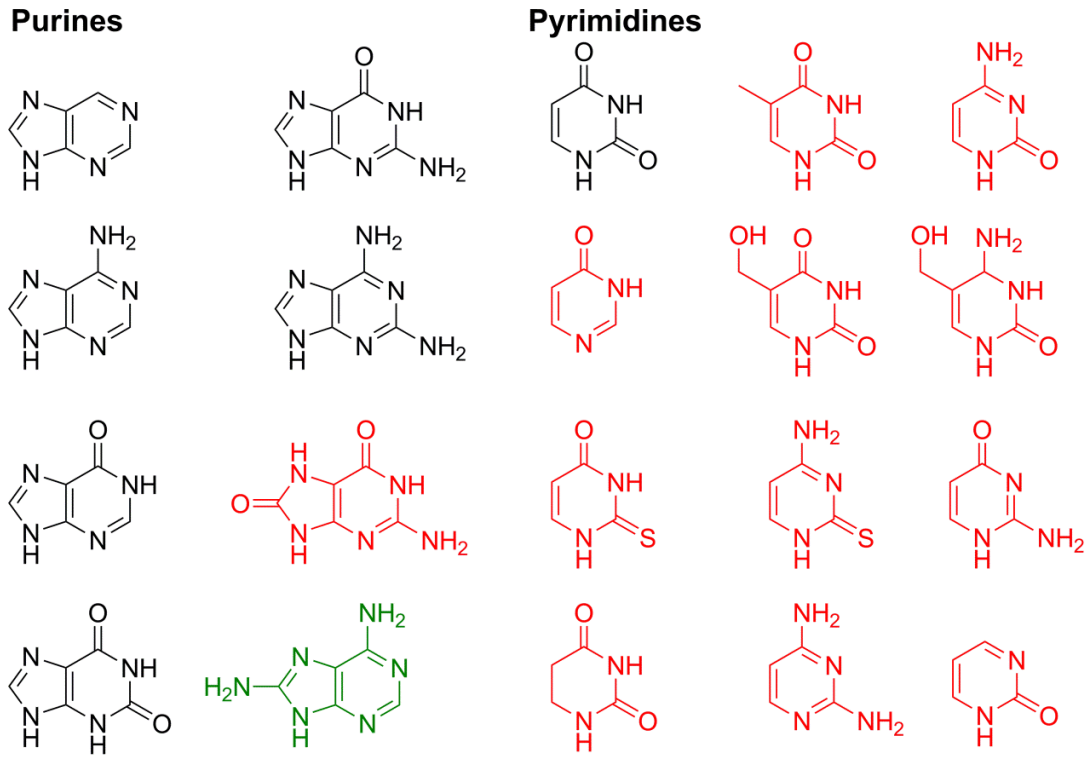


Figure 1.6: Numerous heterocycles have been indentified in prebiotic chemistry simulation experiments and in meteorites. Shown here are some of the relevant purine and pyrimidine nucleobases to our discussion, many more heterocycles have been identified. The structures shown in black are identified in both meteorites and observed in prebiotic chemistry experiments. The structures in red have only been observed in prebiotic chemistry experiments. The structure in green has been identified recently in meteorites.

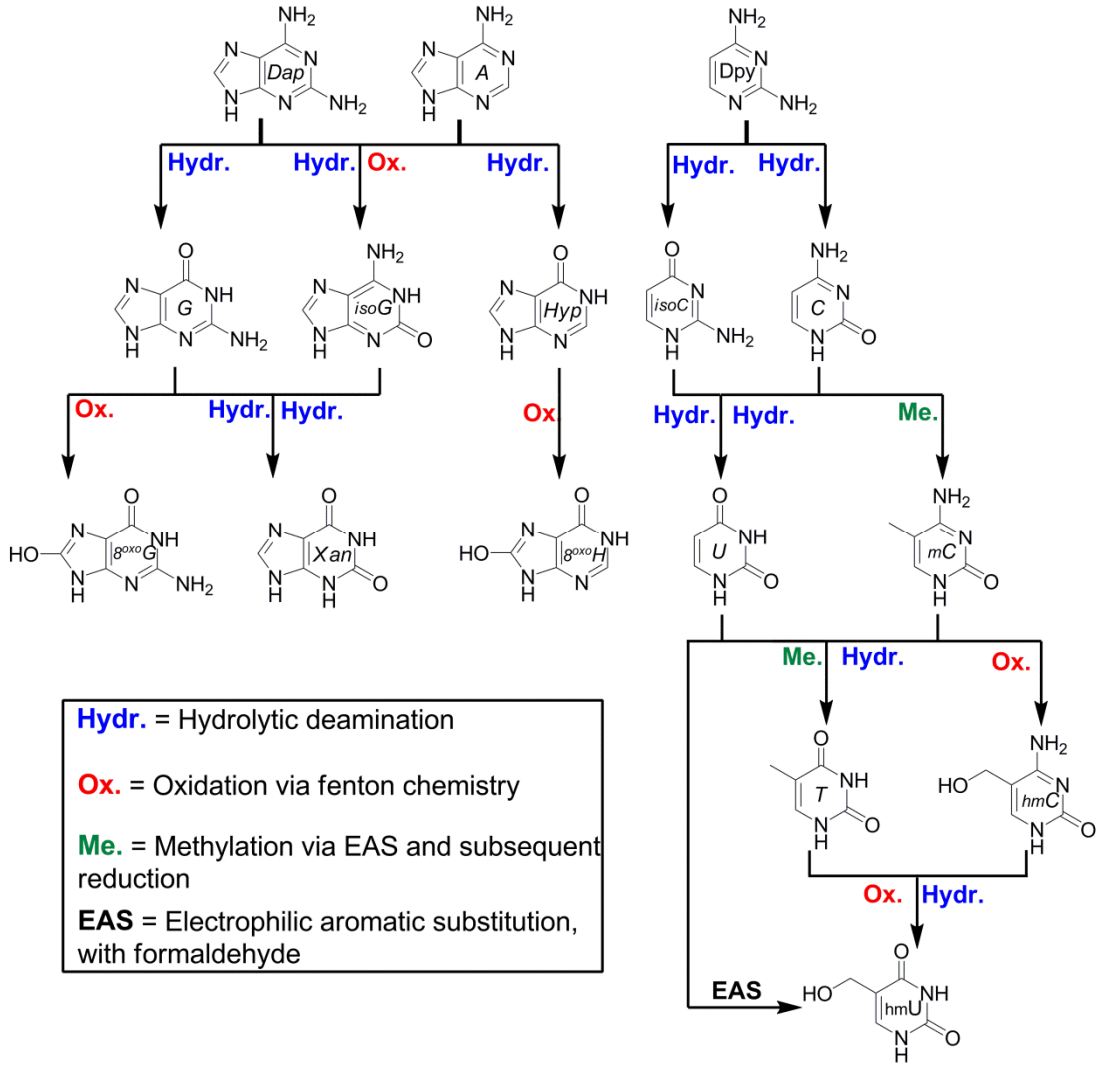


Figure 1.7: Reaction genealogy of purines and a pyrimidine from three simple amino substituted heterocycles. 2, 6-diaminopurine (Dap), Adenine (A), and 2, 4-diaminopyrimidine (Dpy) all considered prebiotically plausible heterocycles, can in principle give rise to many different types of nucleobases, including the native alphabet and its surrogates. The reactivities shown here are all well known and occur even in contemporary biochemistry.

generation of modified bases either from the native bases or independently, have been observed but not shown here (26-29). Clearly, the considerations of prebiotic formation scenarios don't appear to provide much in way of selection with the bases of interest in this chapter. However we haven't taken into account what is considered by many to be the more important factor regarding prebiotic abundance. Stability. Every organic molecule has a threshold to how much it can endure given the set of conditions in which it is subjected. In the context of prebiotic chemistry, the persistence of a nucleobase in a particular hypothetical environment can be considered as a balance between synthetic production(or deposition), side reactions(productive or deterrent) and degradation (Figure 1.8) Measuring the comparative stability(usually in the form of kinetics experiments) of the native bases to others when confronted with reactions that might challenge their persistence has thus been a useful way for understanding the relative fitness of the native bases. There is a limited record of what the early earth was like (30-32) and so it is often difficult to identify the most important environmental conditions that could have contributed selection pressures (Hot/cold temperature, pH, etc). But it does seem that at some point two environmental factors would have likely challenged the persistence of nucleobase on the early Earth; hydrolysis and UV irradiation. We discuss both of these and assess what is known from the literature.

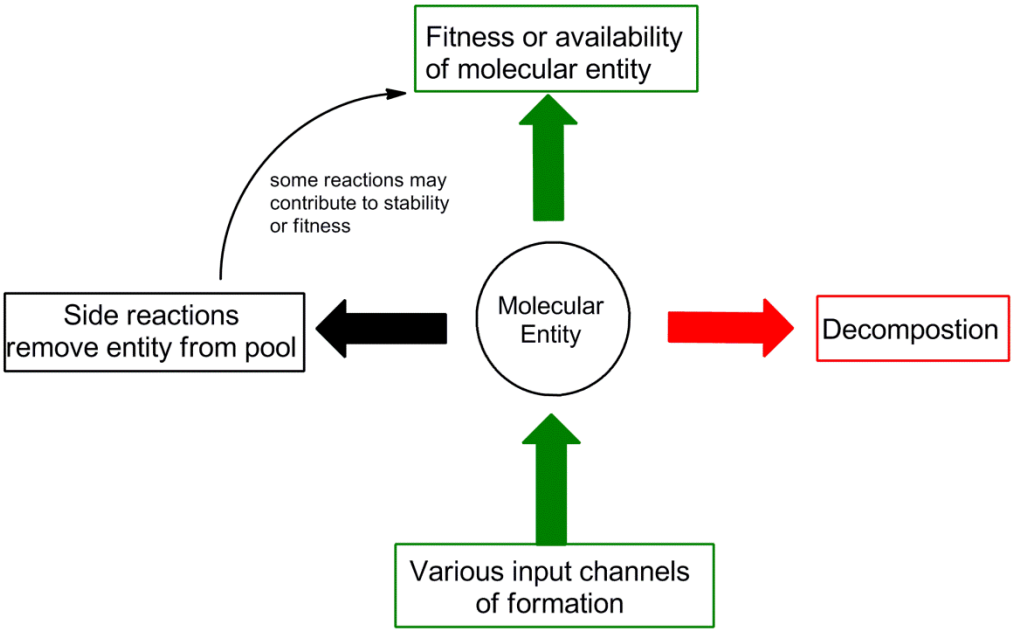


Figure 1.8: Factors that could influence the plausible prebiotic abundance of nucleobases.

Hydrolytic degradation of the bases

The amino groups on the bases in the native alphabet (Figure 1.2) and the ones shown in Figure 1.4, are essential to the fidelity of genetic information. Spontaneous deamination reactions that can occur to the native bases are thus highly deleterious and can lead to genetic mutations (33). It would seem reasonable to hypothesize that the bases used by nature would have been selected to exhibit some of the highest stability against these spontaneous deamination reactions in comparison to alternatives. In a prebiotic scenario though, it could also be possible that the native bases exhibited the most robust heterocyclic stability in an aqueous environment against the action of deamination and ring degradation. The greater persistence in this environment would have given the native bases an advantage over others and possibly facilitating their selective incorporation into the first primitive genetic polymers. Much work has been done on the stability of the bases under a variety of conditions (26, 34-36), but the contributions that are of most relevance to this discussion were from the laboratories of Stanley Miller (26, 37-38). They conducted a comparative stability study using native and alternative bases under the same conditions. From their study a ranking of the deamination and ring opening reaction rates could be made which provided clues to the relative stability of the native heterocycles. Table 1.1 provides the ranking of the bases studied by Miller and Levy with half-life values as the measure of stability. The longer the half-life value indicates the higher stability.

With regards to hydrolytic deamination reactions, it is clear that native purines (A and G) are among the most stable. But the longer half-live value reported for 2,6-diamiopuine deamination reaction provides evidence the native purines are not THE

most stable under these condition. In ring opening or general degradation, it is quite obvious that thymine and uracil are significantly more stable than any of the deaminated purines. The ring opening half-life value of 5,6-dihydrouracil under similar conditions, is included from another Miller report (37) to give perspective on the stability differences between aromatic and non-aromatic species. However, the true quandary has always remained with cytosine as it is the one native base that is highly susceptible to deamination. Upon comparison to the half-life values of other pyrimidine structures in the table there does not appear to be a decent alternative that would dramatically enhance its stability. The occurrence of cytosine in the native alphabet doesn't necessarily remove hydrolytic stability as a plausible selection pressure since it is apparent that the other native bases, are quite robust by comparison to the others. The hypothesized abundance of cytosine on the prebiotic earth in spite of its hydrolytic instability could have resulted from a larger contribution of formation pathways that balanced out the decomposition one (Figure 1.8) One such formation pathway could simply rely on the degradation of other heterocycles. The deamination of the base 2,4-diaminopyrimidine(Figure 1.7 and Table 1.2) is one such example that could lead to a slower continuous release of cytosine(and isocytosine) into the prebiotic environment. In an earlier study, Miller reported that the degradation rate of 2,4-diaminopyrimidine(occurring from deamination reactions) under these conditions, is largely attributed to the C4 deamination reaction that generates cytosine (26).

Table 1.1: Half-Life values of hydrolytic deamination and ring degradation/opening at pH 7 and 100^{0}C

Deamination	Half-life value	Ring degradation or opening	Half-life value
2,6- diaminopurine	2.1 yr	Thymine	56 yr
Adenine	1 yr	Uracil	12 yr
Guanine	0.8 yr	Xanthine	146 days
2,4-Diaminopyrimdine	42 days	Hypoxanthine	12 days
N^4-Methycytosine	38 days	5,6-Dihydrouracil	9.1 hours
Isocytosine	21 days		
Isoguanine	20 days		
Cytosine	19 days		
5-Hydroxymethylcytosine	13 days		
2-Thiocytosine	11 days		
5-methycytosine	9 days		

Data obtained from (26, 37-38)

Photochemical considerations:

The action of UV light on prebiotic mixtures has been historically considered a driver of prebiotic chemistry, but also as a selection pressure for the most stable molecules (18, 39). The nucleobases are strong ultra-violet chromophores with a combined absorbance range ~230 – 280 nm, and molar absorptivity range of their λ_{max} ~8,000 – 15, 000 M^{-1} cm^{-1} in water. As such, the native bases are not inert to photochemical damage. One of the most significant photochemical reactions that can occur is the photodimerization of uracil or thymine residues in RNA and DNA (40). However as free bases they exhibit outstanding photostabilities in comparison to other heterocycles (41). The last 15 years has been a time of increased detailed investigations to probing the photophysical properties of the native bases (41). What has emerged is an understanding that upon radiative excitation (via a $\pi\pi^*$ transition), the bases exhibit efficient non-radiative decay mechanisms back to the ground state, resulting in exceptionally fast excited-state lifetimes(τ) (41). It is this attribute that explains the extremely low fluorescent and many photochemical quantum yields observed in the native bases (41). The ultrafast excited state lifetimes shared by the native bases have prompted investigators to hypothesize that photostability in the prebiotic environment was a viable selection pressure and the native bases, the clear winners (42). Assessing the plausibility of this selection pressure has only recently been possible given the published investigations on excited state lifetimes of alternative and modified nucleobases. Table 1.2 lists these contributions and most are of the studies conducted in aqueous solutions. A few reports though, because they contain relevant heterocycles are included from experimental work in the gas phase. The values of τ =0.1 – 7 picosecs (ps) reported for the native base under neutral pH are indeed fast, but it is apparent they are not the only ones. The

derivatives of xanthine that were studied implicate xanthine with a fast lifetime. But it is hypoxanthine that is now considered to exhibit the fastest known lifetime yet (43) More intriguing is how relatively simple structural changes can create drastic difference to the decay duration of a heterocycle. The isomer of adenine, 2-aminopurine(2AP), (see figure 1.4), a highly fluorescent base (QY= 0.68) with a τ excited state lifetime of τ =11.8 nanosecs (compared to 0.18 picosecs for adenine) is one of the largest photophysical alterations ever reported. The other adenine analog, 2,6-diaminopurine is also known for its fluorescent properties(QY ~0.01) was recently reported in gas phase experiments, to have a long life time ($\boldsymbol{\tau}$ =6300 ps). Much smaller differences but still considered of significance to this community are the changes to the native pyrimidines upon methylation at the C5 position or modification to amino substituents. Whereas uracil and thymine are close in lifetime values ($\boldsymbol{\tau}$ = 1.9ps (U) and 2.8ps(T), the analogous cytosine($\boldsymbol{\tau}$ =1ps) and 5-methylcytosine($\boldsymbol{\tau}$ =7ps display a larger difference.

The native bases in their neutral state exhibit fast decay times, but a change in pH can increase these values quite substantially as a result to protonation or deprotonation of the heterocycle. These pH studies are especially relevant for consideration to a prebiotic selection pressure because there appears to be much flexibility as to what the pH of the early oceans may have been (44). The base guanine (in this case as guanosine) under acidic condition (pH 1.5) has a dramatic lifetime change from τ =0.16 to 191 ps upon protonation. Interestingly the alternative base hypoxanthine was reported to not exhibit any significant lifetime alteration, and 5-mC actually exhibited a decrease from τ =7 to 2 ps. Alkaline pH though can have an

effect, cytosine is less drastic (from τ =1 to 13ps) but, 5mC increases its value significantly from τ =7 to 250 ps.

Many of the non-native bases listed in Table 1.2 with longer lifetime values are found in RNA as base modifications(mC, ac^4C, Hyp) or in the case of 2,4-diaminopyrimidine, it is the core heterocycle formed in Lysidine, (Figure 1.3A). The occurrence of these "photophysically unfit" bases in nature may seem to contradict the selection pressure posed here. As free bases, photostability may have been an advantage in the prebiotic environment. But in terms of the first genetic polymers or more importantly the first biotic systems, the inclusion of alternative bases with heightened photochemical activity may actually have offered an advantage. Recent publications by Cynthia Burrows and co-workers have contributed a fresh angle to the importance of these photophysically interactive bases in the context of early RNA genomes. Her work with 8-oxoG (see Figure 1.7), a base widely known as a damaged oxidative lesion in DNA has been shown to be quite successful in its photo-induced role as a repair catalyst that reverses the photodimerization products of thymine and uracil residues within RNA and DNA polymers (45) This activity has been attributed to the enhanced photophysical and redox activity of 8-oxoG over G (46). The excited state lifetimes of 8-oxoG have not been reported, but that of 5-hydroxyuracil has (τ =1800 ps) and this is also a base that Burrows has mentioned to that might also exhibit similar repair properties given is redox activity (46). Conversely, her attempts to employ xanthine, since it had a lower redox potential than G, did not perform any photo induced repair activity. Based on the ultra short lifetime data obtained for xanthine derivatives (Table 1.2) it would seem reasonable to exclude this base from any favorable photochemical activity.

Table 1.2: Comparison of excited state lifetime values of native and alternative nucleobases in aqueous solutions[a]

Nucleobase	Relevance to Biochemistry	Excited state Lifetimes(τ)/ picosec	Reference
Uncharged bases (neutral pH 6.8 – 7.3)			
Hypoxathanine (Hyp)	RNA Mod.	0.13 ± 0.3	(43)
Adenine(A)	Native Letter	0.18 ±0.3	(47)
Guanine/(G)	Native letter	0.16[b]	(48)
Xanthine (Xan)	RNA precur.	0.28 - .50[c]	(43)
Cytosine (C)	Native Letter	1.0 ± 0.2, 2.9, 12	(49-50)
Uracil (U)	Native Letter	1.9 ±0.1, 24 ± 0.2	(50)
Thymine (T)	Native Letter	2.8 ±0.1, 30 ± 13	(50)
5-Methylcytosine (mC)	DNA/RNA Mod.	7.2 ± 0.4	(49)
N^4-actetycytosine (ac^4C)	RNA Mod.(3 Domains)	280 ± 30	(49)
2,4-diaminopyrimidine	Core heterocycle of Lysidine , Fig 1.3A	10–1000[a] (gas)	(51)
5-Hydroxyuracil(5HoU)	RNA Mod	1800[a]	(52)
2,6-diaminopurine(Dap)	Rare aturally occuring adenine surrogate in DNA	6300 ± 400[a] (gas state	(51)
2-aminopurine(2AP)	Adenine surrogate but not found in nature	11,800	(53)
Cation charged bases(pH 0 – 2) via protonatation			
Guanine		191± 4[b]	(54)
Hypoxanthine		< 0.2	(48)
5-Methylcytosine		2.57 ± 0.22	(49)
Anion charged bases (pH 13)			
Hypoxanthine (pH 10)		19	(48)
Cytosine		13.3± 0.4	(49)
5-Methylcytosine		250 ±30	(49)

[a] gas phase experimental studies . [b] due to solubility problems of Guanine the community has relied on the value obtained from deoxyguanosine and guanosine to model/approxiamte nucleobase values G. [c] expected value range based on the derivaties in study

Selection during the formation of the first informational polymers

One of the more enigmatic and difficult problems confronting the prebiotic chemistry community is identifying how the monomers of RNA, or pre-RNA, or even non-related polymeric components selectively formed and self-assembled out of the presumed random prebiotic mixtures(55-57). It is however in this transition (Informational polymer assembly, Figure 1.5) where significant selection processes must have occurred not only for the base composition, but for the other components of nucleic acids (or nucleic acid precursors) (44, 58) and others. Focusing on just a narrow view of RNA precursors, the linking of the nucleobase to a ribose sugar is one such pressure. There are multiple ways that a nucleobase can be attached to ribose via an *N*-glycosidic bond, but in nature only one position (the N9 of purines, or N1 of pyrimidines) is used as the glycosidic bond. Achieving regio and stereo-specific selectivity of this reaction under simulated prebiotic conditions has plagued the community ever since Leslie Orgel and others began working on this problem (17-18, 55) (Figure 1.9A). The problem with prebiotic glycosylation has consequently lead John Sutherland and co-workers to circumvent this reaction by pursuing the route of building the nucleobase heterocycles and sugars together (59-60) He has achieved success in his prebiotic experiments and most intriguingly demonstrated preferential formation of the native ribonucleotides(61). His routes may be criticized for its “external” or directed nature of synthetic utility, in addition to abandoning the utility of the presumed abundance of nucleobases that populated the early earth environment. Additionally, the selective formation of native ribonucleotides does not preclude the base itself from further modification. Even still, Sutherland’s contributions are highly creative and do employ simple prebiotic precursors. His unique approach is also

reminiscent of Orgel's Second Rule, which states that "Evolution is cleverer than you are".

In another creative manner other investigators have realized that while the native bases themselves are too difficult to form glycosidic bonds, more success comes from the use of modified or alternative purine and pyrimidine heterocycles. Hypoxanthine was one of the first non-native bases from Orgel's pioneering studies in the 1970's (Figure 1.9B), shown to exhibit some of the highest yields of ribonucleoside formation and the major product was the "correct" anomer (17). More recently, Nick Hud and co-workers were the first to demonstrate that a non-native pyrimidine could undergo glycosylation reactions with D-ribose (Figure 1.9C), in some of the highest yields to date, by using the analog to cytosine, 2-pyrimidinone, giving the zebularine ribonucleoside (62). Previous work from the Miller lab pursued the use of non-pyrimidine or non-purine base heterocycles with much better success and possible pre-RNA world surrogates (28). The easier formation of regio and stereo specific glycosidic bonds between non-native bases might have played a role in the assembly to the first informational polymers (55). The eventual modification to generate the native bases during the course of evolution could have then given rise to the extant alphabet.

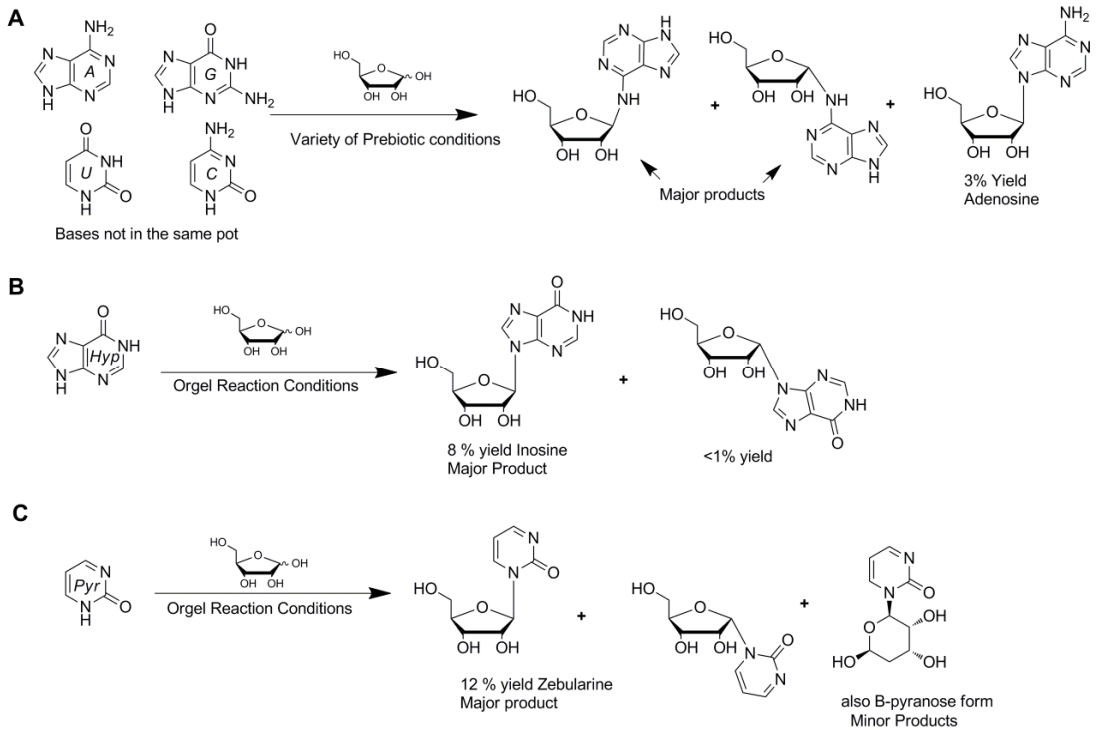


Figure 1.9: Experiments conducted to generate nucleosides under prebiotic conditions using the nucleobases and D-ribose. **A.** Leslie Orgel and coworkers demonstrated that of all the native bases, only adenine formed the relevant glycosidic bond to produce adenosine, but still with only low yields (17). **B.** Using a concentrated solution of magnesium sulfate,magnesium chloride(modeling seawater) and excess of ribose, hypoxanthine has been observed to produce the correct regio and stereochemical product. **C.** Under similar conditions, the Hud labs were the first to report a successful prebiotic glycosylation reaction using an alternative pyrimidine base.

Base Selection in an Early RNA world?

The RNA world is an assumed period near or at the origin of life where RNA was the genomic and catalytic center of early cells (63). The native bases while exceptionally good in their genetic capabilities, don't appear to be particularly useful at catalytic functional roles in comparison to the side chains of amino acids. However it is the observations of base modification in extant RNA that has encouraged many to ponder that the RNA world may have utilized an expanded set of heterocycles (64). Many of the modifications found in extant RNA (also see Figure 1.3A and 1.3B) do have striking similarities to the functional groups of the amino acids (23, 65). While this may seem to compound the problem of narrowing the alphabetic composition there are still plausible pressures that could have occurred in this epoch. This is especially true as RNA began to partition its role (genomic vs. catalytic), or relinquishing it catalytic involvement during the recruitment of proteins. Two such pressures based on the reliability of genetic information, and another with contributions to the stability of duplex polymers, are both inherently related to individual bases.

Genetic fidelity considerations:

Genetic fidelity is likely to be among the most important functions of a base when part of a living system (however primitive). Spontaneous deamination reactions that challenge the integrity of the base pairing face as previously discussed is certainly a viable pressure in the RNA world. But even with the integrity of the heterocycle intact, not all bases are equal in maintaining faithful coding properties. The well known example comes from the base isoguanine (isoG). Isoguanine, when incorporated into nucleic acids can base pair with its natural complement isoC (Figure

1.10A), but also with uracil (Figure 1.10B). This is due to the propensity of isoG to tautomerize to its enol form which enables the alternative pairing (66). Furthermore, in the presence of polymerases it was observed that isoG can also direct the incorporation of U and similarly the opposite scenario. This problem, along with their deamination susceptibilities (Table 1.1), has been previously hypothesized as one reason for why the isoG:isoC pair probably never became a part of the genetic alphabet(67). However the native alphabet is not entirely free of promiscuous pairing either. Guanine also recognizes uracil but with the use of a wobble base pair (Figure 1.10C) The G•U wobble pair is actually quite ubiquitous in extant biology and considered to be one of the most important non-canonical pairing interactions used by RNA for its functional diversity (68). Because of the inherent stability and pervasive nature of this interaction it might be pondered as to how life in the RNA world maintained genetic fidelity without the assistance of evolved polymerases to ensure proper replication? One possibility comes from clues in extant biochemistry where life ingeniously found ways around this 'problem' from the utility of modified bases such 2-thiouracil and its derivatives. It has been reported that these thiolated modifications occur most often in the wobble position of tRNA when the base pairing of A-U needs to be ensured for proper translation (1, 69-70). The switch from oxygen to sulfur in the C2 position of uracil creates a highly specific base pair with adenine because it removes the possibility for a wobble G-U interaction given the position of the sulfur on the modified uracil (71-73) (Figure 1.10D). The occurrence of 2-thiouracil and derivatives in extant biology provides insight as to how alternative bases could have been advantageous for replication fidelity in the early RNA world. However, the evolution of proteinaceous polymerase in the presence of the native bases has also largely guaranteed proper fidelity during replication.

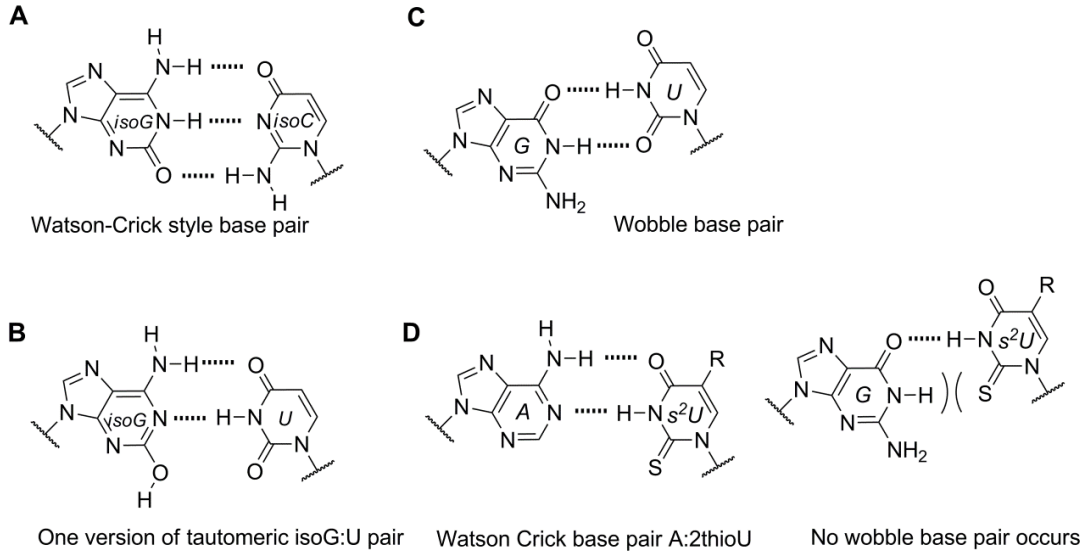


Figure 1.10: Genetic fidelity pressures A. The isoG:isoC base pair that utilizes the dominant tautomer. **B**. A minor tautomeric form of isoG allows for a base pairing with U. **C.** Guanine can also base pair with uracil, utilizing a wobble pair. **D.** The 2-thioU base is highly specific for adenine since it can't wobble pair with guanine.

Duplex stability considerations:

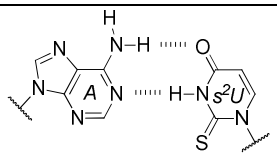
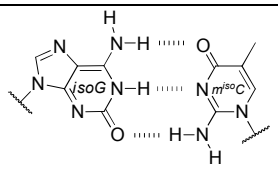
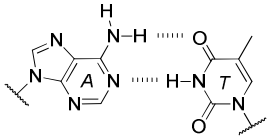
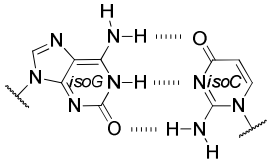
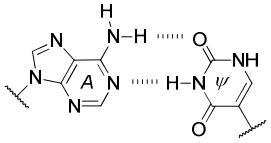
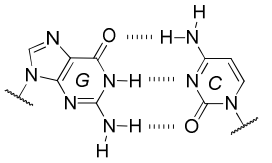
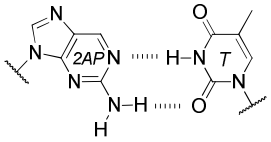
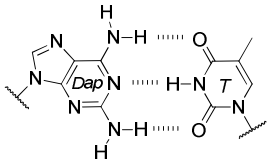
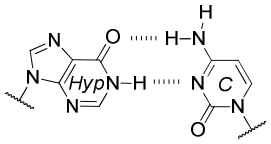
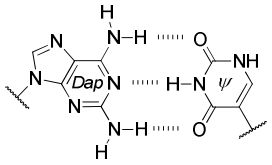
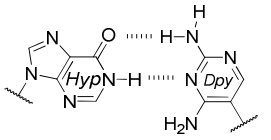
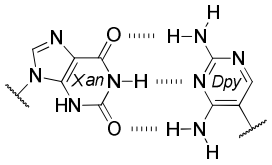
From observations of DNA damage in extant biology, it would seem that a strong selection pressure for cells to maintain double stranded nucleic acids would have occurred early in biotic evolution. More than providing two copies of genetic information, duplex systems are highly advantageous in protecting genetic material from any chemical or physical challenge. Just about every type of damage that can occur to DNA (deamination, deglycosylation, oxidation, and alkylation and even some photochemical reactions) are known to be reduced when nucleic acids are kept in duplex systems(74) And it is thought that most of the damage to DNA does indeed occur when portions of the genome are exposed or separated from the duplex, as it does during the many processes of normal cellular function(33). Many factors contribute to duplex stability (15, 75) such as an anionic sugar-phosphate backbone, water hydration, and metal ions, etc. But two important contributions directly related the bases are the hydrophobic interaction of π- π aromatic stacking and the base pairing strength (14). Of these two contributions, the base pairing strength seems to be inherently related to the specific nature of the heterocycle. Assuming the same sugar-phosphate backbone, the question could reasonably be asked; do the native bases contribute the highest stabilities of RNA and DNA duplexes? Building upon his work of utilizing non-natural bases, Benner and co-workers in 2003 undertook a systematic study of measuring how alternative bases and their pairing relationships affected the temperature of thermal denaturation experiments of short sequences (14 oligonucleotides in length) of duplex DNAs (76). The differences in melting temperatures (called Tm's) which are dependent on the nature of the base pair under investigation, provided a way to probe the relative stability of base pair strengths.

Table 1.3 lists a selection of the findings from Benner's work in addition to one other report (77) that relates to the alternative bases discussed in this chapter (see Figure 1.4). Two values are typically listed for each column containing base pairs as a result to the differences observed when the base is oriented on the two strands (A:T versus T:A, for example). The higher reported Tm value results from a greater relative stability from the associated base pair. In this short list of closely related bases and base pairs it is observed that the native bases do contribute some of the stronger base pair strengths, but generally not by much, and they are not the highest either. This observation has actually led Benner to the conclusion that in general the bases of nucleic acids are quite interchangeable and can be manipulated to a high degree (75). Some of the interesting examples here are the bases or the base pairs that are more stable than the native bases. Most striking is the A:s^{2}U interaction, one that occurs in nature, because even with just two hydrogen bonds it still approaches the strength of a G:C pair. The isoG:5-methylisoC is also intriguing since it is probably one of the strongest reported for bases that appear to be prebiotically plausible. Other studies measuring the original isoG:isoC pair also found that it was either at or above the base pair strength of the G:C interaction(66, 78). The data shown for the Hyp:Dpy, or the Xan:Dpy base pairs are considerably lower in Tm's and may provide some indication that not all of the bases could have been viable contenders. Some of the results from Table 1.3 can be incorporated into the hypothesis recently reported from the lab of Ramanarayanan Krishnamurthy, where he is investigating a selection pressure related to the pK_a values of the bases and how they affect the base pairing strength and overall duplex stability (79). His observations reside with two correlations. The first is that optimum base pair strength is observed when the difference between the pK_a values of the hydrogen bonding face, ΔpK_a, are ≥ 5 units.

This appears to be rooted in the principle that greater base pairing strength comes from more polarized intermolecular interactions. The second correlation is that duplex stability is greater when the pK_a values of individual bases are > 2 units away from the pH of the aqueous medium which he denoted as the pK_a–pH correlation. This lies with the assumption that bases with pK_a values closer to the pH of the aqueous medium (<2 units) are likely to be ionized to a greater extent. This internal charge may then disrupt the hydrophobic region of base pairing and stacking within the duplexes. Table 1.4 lists the values using his empirical correlations with some of experimental data from Table1.3. While not completely congruent, there is a general trend, and it should be noted that Krishnamurthy was considering, in his article, bases that exhibited more exotic alterations to the heterocyclic structure (79). However the entries of Xan:Dpy and Hyp:Dpy base pairs in Table 1.4 did display the lowest melting temps which is in agreement to the calculated ΔpK_a and pK_a–pH values.

So was base pair strength a viable selection pressure? We know from studies of thermophiles and hyperthermophiles that duplex stability is indeed important given the correlation of higher GC rich residues that are associated with the genomes of these organisms residing at high temperatures. Similarly, there may have been base pairs like the Xan:Dpy example, that posed a distinct disadvantage even at modest temperatures, and thus not utilized(79). But in comparison to some of the other base pairs listed in Table 1.3, the native bases don't appear to exhibit any clear advantage.

Table 1.3: Base Pair Stability Ranking from Relative Thermal Denaturation at pH 7.9,[a]

2 H-Bond Base Pair	Melting Temp (^{0}C)	3-H Bond Base Pair	Melting Temp (^{0}C)
	Predicted to be ~58.7[a]		61.5, 63.3^o
	52.9, 55.9		Expected to rank here (66, 78)
	54.4		58.5, 59.5
	52.1, 54.7		56.9, 58.4
	52.8, 54.6		56.7
	47.5		47.5

Data from: (76) [a] Data from this study (77) which compared 7mer oligos but of RNA duplexes monitoring one substitution. The s^2U-A base pair substantially increased the melting temp to 57.4^{o}C in comparison to 51.6 ^{o}C for a U-A pair.

Table 1.4: Comparison of some data from Table 1.3 to pK_a correlations made by R. Krishnamurthy (79)

Base pair	Melting Temp	ΔpK_a	pK_a–pH for each base pH was conducted at 7.9
3-H bond			
isoG:isoC	61.5, 63.3[o]	~4.8	1(isoG), 3.7(isoC)
G:C	58.5, 59.5	~5.1	1.8(G), 3.5(C)
Dap:T	56.9, 58.4	~4.6	2.4(Dap), 1.9(T)
Dap:Psi	56.7	~4.5	2.4(Dap), 1.8(Psi)
Xan:Dpy[*]	47.5	~2	~1(Xan), ~1.6(Dpy)
2H Bond			
A:s^{2}U	58.7[a]	~5.1	~4.2(A), ~0.9 (s^{2}U)
A:T	52.9, 55.9	~6.1	4.2(A), 1.8 (T)
A:Psi	54.4	~6	4.2(A),1.8(Psi)
Hyp:C	52.8, 54.6	~4.5	~1(Hyp), 3.5(C)
Hyp:Dpy	47.5	~1.8	1(Hyp), 1.6(Dpy

*pH of experiment was measured at 5.4

A continuous process of refinement

From the topics included in this chapter, it should be readily apparent that many types of selection pressures could have occurred that shaped the composition of the nucleic acid bases. But it is also likely that not any single pressure (at least from these topics introduced) could have designated the native bases as clear winners. It is more appropriate to view these pressures on the nature of the genetic alphabet as part of a continual process of refinement. We share the view that many forms of bases became part of the first genetic polymers which contributed to the viability of the RNA world (58). As early life evolved so did the nature of the bases since new pressures also emerged. The remainder of this dissertation will be based on the exploration and investigation of one such pressure that confronted cells at the dawn of one of the greatest transitions for early life, the transition from genomic RNA to DNA.

Acknowledgements

Chapter 1 is adapted from a manuscript in preparation for an invited contribution to a thematic issue in the *Israel Journal of Chemistry*. The dissertation author is the main author for this work.

References

1. Carell T*, et al.* (2012) Structure and Function of Noncanonical Nucleobases. *Angew. Chem. Int. Ed. Engl.* 51(29):7110-7131.

2. Martı*, et al.* (1998) On the Function of Modified Nucleosides in the RNA World. *J. Theor. Biol.* 194(4):485-490.

3. Forterre P & Grosjean H (2009) Chapter 19. The Interplay between RNA and DNA Modifications:Back to the RNA World. *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution,* Molecular Biology Intelligence United Grosjean H (Landes Bioscience, Austin, Texas), pp 259 - 274.

4. Cermakian N & Cedergren R (1998) Chapter 29. Modified Nucleosides Always Were: an Evolutionary Model. *Modification and Editing of RNA*, eds Grosjean H & Benne R (ASM Press, Washington DC), pp 535 - 541.

5. Berg JM, Tymoczko JL, & Stryer L (2007) *Biochemistry* (W.H. Freeman, New York) 6th Ed p

6. Poole A, Penny D, & Sjöberg B-M (2001) Confounded cytosine! Tinkering and the evolution of DNA. *Nat Rev Mol Cell Biol* 2(2):147-151.

7. Warren RAJ (1980) Modified Bases in Bacteriophage DNAs. *Annu. Rev. Microbiol.* 34(1):137-158.

8. Rae PMM & Steele RE (1978) Modified bases in the DNAs of unicellular eukaryotes: an examination of distributions and possible roles, with emphasis on hydroxymethyluracil in dinoflagellates. *Biosystems* 10(1-2):37-53.

9. Gommers-Ampt J & Borst P (1995) Hypermodified bases in DNA. *The FASEB Journal* 9(11):1034-1042.

10. Piccirilli JA, Benner SA, Krauch T, & Moroney SE (1990) Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature* 343(6253):33-37.

11. Switzer C, Moroney SE, & Benner SA (1989) Enzymatic incorporation of a new base pair into DNA and RNA. *J. Am. Chem. Soc.* 111(21):8322-8323.

12. Rich A (1962) On the problems of evolution and biochemical information transfer. *Horizons in Biochemistry* eds Kasha M & Pullman B (Academic Press, New York), pp 103 - 126.

13. Eschenmoser A (1999) Chemical Etiology of Nucleic Acid Structure. *Science* 284(5423):2118-2124.

14. Saenger W (1984) *Principles of nucleic acid structure* (Springer-Verlag, New York) pp 556.

15. Benner SA & Sismour AM (2005) Synthetic biology. *Nat Rev Genet* 6(7):533-543.

16. Borquez E, Cleaves HJ, Lazcano A, & Miller SL (2005) An Investigation of Prebiotic Purine Synthesis from the Hydrolysis of HCN Polymers. *Origins Life Evol. Biospheres* 35(2):79-90.

17. Orgel LE (2004) Prebiotic Chemistry and the Origin of the RNA World. *Crit. Rev. Biochem. Mol. Biol.* 39(2):99-123.

18. Miller SL & Orgel LE (1974) *The origins of life on the earth* (Prentice-Hall, Englewood Cliffs, N.J.,) pp x, 229 p.

19. Cleaves HJ, Nelson K, & Miller S (2006) The prebiotic synthesis of pyrimidines in frozen solution. *Naturwissenschaften* 93(5):228-231.

20. Barks HL*, et al.* (2010) Guanine, Adenine, and Hypoxanthine Production in UV-Irradiated Formamide Solutions: Relaxation of the Requirements for Prebiotic Purine Nucleobase Formation. *ChemBioChem* 11(9):1240-1243.

21. Botta O & Bada JL (2002) Extraterrestrial Organic Compounds in Meteorites. *Surv. Geophys.* 23(5):411-467.

22. Callahan MP*, et al.* (2011) Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases. *Proc. Natl. Acad. Sci. U.S.A.* 108(34):13995-13998.

23. Robertson M & Miller S (1995) Prebiotic synthesis of 5-substituted uracils: a bridge between the RNA world and the DNA-protein world. *Science* 268(5211):702-705.

24. Choughuley ASU, Subbaraman AS, Kazi ZA, & Chadha MS (1977) A possible prebiotic synthesis of thymine: Uracil-formaldehyde-formic acid reaction. *Biosystems* 9(2-3):73-80.

25. Saladino R*, et al.* (2003) One-Pot TiO2-Catalyzed Synthesis of Nucleic Bases and Acyclonucleosides from Formamide: Implications for the Origin of Life. *ChemBioChem* 4(6):514-521.

26. Robertson MP, Levy M, & Miller SL (1996) Prebiotic synthesis of diaminopyrimidine and thiocytosine. *J. Mol. Evol.* 43(6):543-550.

27. Levy M & Miller SL (1999) The Prebiotic Synthesis of Modified Purines and Their Potential Role in the RNA World. *J. Mol. Evol.* 48(6):631-637.

28. Kolb VM, Dworkin JP, & Miller SL (1994) Alternative bases in the RNA world - the prebiotic synthesis of urazole and its ribosides. *J. Mol. Evol.* 38(6):549-557.

29. Menor-Salvan C & Marin-Yaseli MR (2012) Prebiotic chemistry in eutectic solutions at the water-ice matrix. *Chemical Society Reviews* 41(16):5404-5415.

30. Sleep NH (2010) The Hadean-Archaean Environment. *Cold Spring Harbor Perspect. Biol.* 2(6).

31. Zahnle K, Schaefer L, & Fegley B (2010) Earth's Earliest Atmospheres. *Cold Spring Harbor Perspect. Biol.* 2(10).

32. Nisbet EG & Sleep NH (2001) The habitat and nature of early life. *Nature* 409(6823):1083-1091.

33. Gates KS (2009) An Overview of Chemical Processes That Damage Cellular DNA: Spontaneous Hydrolysis, Alkylation, and Reactions with Radicals. *Chem. Res. Toxicol.* 22(11):1747-1760.

34. Garrett ER & Tsau J (1972) SOLVOLYSES OF CYTOSINE AND CYTIDINE. *J. Pharm. Sci.* 61(7):1052.

35. Shapiro R (1995) The prebiotic role of adenine: A critical analysis. *Origins Life Evol. Biospheres* 25(1):83-98.

36. Shapiro R (1999) Prebiotic cytosine synthesis: A critical analysis and implications for the origin of life. *Proc. Natl. Acad. Sci. U.S.A.* 96(8):4396-4401.

37. House CH & Miller SL (1996) Hydrolysis of Dihydrouridine and Related Compounds. *Biochemistry* 35(1):315-320.

38. Levy M & Miller SL (1998) The stability of the RNA bases: Implications for the origin of life. *Proc. Natl. Acad. Sci. U.S.A.* 95(14):7933-7938.

39. Sagan C (1973) Ultraviolet selection pressure on the earliest organisms. *J. Theor. Biol.* 39(1):195-200.

40. Friedberg EC (2003) DNA damage and repair. *Nature* 421(6921):436-440.

41. Middleton CT, *et al.* (2009) DNA Excited-State Dynamics: From Single Bases to the Double Helix. *Annu. Rev. Phys. Chem.* 60(1):217-239.

42. Gustavsson T, Improta R, & Markovitsi D (2010) DNA/RNA: Building Blocks of Life Under UV Irradiation. *The Journal of Physical Chemistry Letters* 1(13):2025-2030.

43. Chen JQ & Kohler B (2012) Ultrafast nonradiative decay by hypoxanthine and several methylxanthines in aqueous and acetonitrile solution. *Phys. Chem. Chem. Phys.* 14(30):10677-10682.

44. Kua J & Bada J (2011) Primordial Ocean Chemistry and its Compatibility with the RNA World. *Origins Life Evol. Biospheres* 41(6):553-558.

45. Nguyen KV & Burrows CJ (2011) A Prebiotic Role for 8-Oxoguanosine as a Flavin Mimic in Pyrimidine Dimer Photorepair. *J. Am. Chem. Soc.* 133(37):14586-14589.

46. Nguyen KV & Burrows CJ (2012) Whence Flavins? Redox-Active Ribonucleotides Link Metabolism and Genome Repair to the RNA World. *Acc. Chem. Res.*

47. Cohen B, Hare PM, & Kohler B (2003) Ultrafast Excited-State Dynamics of Adenine and Monomethylated Adenines in Solution:Implications for the Nonradiative Decay Mechanism. *J. Am. Chem. Soc.* 125(44):13594-13601.

48. Villabona-Monsalve JP, Noria R, Matsika S, & Peon J (2012) On the Accessibility to Conical Intersections in Purines: Hypoxanthine and its Singly Protonated and Deprotonated Forms. *J. Am. Chem. Soc.* 134(18):7820-7829.

49. Malone RJ, Miller AM, & Kohler B (2003) Singlet Excited-state Lifetimes of Cytosine Derivatives Measured by Femtosecond Transient Absorption. *Photochemistry and Photobiology* 77(2):158-164.

50. Hare PM, Crespo-Hernández CE, & Kohler B (2007) Internal conversion to the electronic ground state occurs via two distinct pathways for pyrimidine bases in aqueous solution. *Proc. Natl. Acad. Sci. U.S.A.* 104(2):435-440.

51. Gengeliczki Z*, et al.* (2010) Effect of substituents on the excited-state dynamics of the modified DNA bases 2,4-diaminopyrimidine and 2,6-diaminopurine. *Phys. Chem. Chem. Phys.* 12(20):5375-5388.

52. Nachtigallova D*, et al.* (2010) The effect of C5 substitution on the photochemistry of uracil. *Phys. Chem. Chem. Phys.* 12(19):4924-4933.

53. Neely RK, Magennis SW, Dryden DTF, & Jones AC (2004) Evidence of Tautomerism in 2-Aminopurine from Fluorescence Lifetime Measurements. *The Journal of Physical Chemistry B* 108(45):17606-17610.

54. Pecourt J-ML, Peon J, & Kohler B (2001) DNA Excited-State Dynamics: Ultrafast Internal Conversion and Vibrational Cooling in a Series of Nucleosides. *J. Am. Chem. Soc.* 123(42):10370-10378.

55. Bean Heather D, Lynn David G, & Hud Nicholas V (2009) Self-Assembly and the Origin of the First RNA-Like Polymers. *Chemical Evolution II: From the*

Origins of Life to Modern Society, ACS Symposium Series, (American Chemical Society, Washington DC), pp 109-132.

56. Benner SA, Kim H-J, & Yang Z (2012) Setting the Stage: The History, Chemistry, and Geobiology behind RNA. *Cold Spring Harbor Perspect. Biol.* 4(1).

57. Robertson MP & Joyce GF (2012) The Origins of the RNA World. *Cold Spring Harbor Perspect. Biol.* 4(5).

58. Engelhart AE & Hud NV (2010) Primitive Genetic Polymers. *Cold Spring Harbor Perspect. Biol.* 2(12):a002196.

59. Powner MW, Gerland B, & Sutherland JD (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 459(7244):239-242.

60. Powner MW, Sutherland JD, & Szostak JW (2010) Chemoselective Multicomponent One-Pot Assembly of Purine Precursors in Water. *J. Am. Chem. Soc.* 132(46):16677-16688.

61. Powner MW, Sutherland JD, & Szostak JW (2011) The Origins of Nucleotides. *Synlett* 2011:1956- 1964 .

62. Bean HD*, et al.* (2007) Formation of a β-Pyrimidine Nucleoside by a Free Pyrimidine Base and Ribose in a Plausible Prebiotic Reaction. *J. Am. Chem. Soc.* 129(31):9556-9557.

63. Joyce GF (1989) RNA evolution and the origins of life. *Nature* 338(6212):217-224.

64. Benner SA, Burgstaller P, Battersby TR, & Jurczyk S (1999) Chapter 6. Did the RNA World Exploit an Expanded Genetic Alphabet? *The RNA World*, eds Raymond F Gesteland, Cech TR, & Atkins JF (Cold Spring Harbor Laboratory Press, Cold Spring Harbor), Vol 2, pp 163-181.

65. Rozenski J, Crain PF, & McCloskey JA (1999) The RNA Modification Database: 1999 update. *Nucleic Acids Res.* 27(1):196-197.

66. Roberts C, Bandaru R, & Switzer C (1997) Theoretical and Experimental Study of Isoguanine and Isocytosine: Base Pairing in an Expanded Genetic System. *J. Am. Chem. Soc.* 119(20):4640-4649.

67. Switzer CY, Moroney SE, & Benner SA (1993) Enzymic recognition of the base pair between isocytidine and isoguanosine. *Biochemistry* 32(39):10489-10496.

68. Varani G & McClain WH (2000) The GU wobble base pair - A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep.* 1(1):18-23.

69. Yarian C*, et al.* (2002) Accurate translation of the genetic code depends on tRNA modified nucleosides. *J. Biol. Chem.* 277(19):16391-16395.

70. Ajitkumar P & Cherayil JD (1988) Thionucleosides in transfer ribonucleic acid-diversity, structure, biosynthesis, and function. *Microbiological Reviews* 52(1):103-113.

71. Donohue J (1969) ON N-HS hydrogen bonds. *Journal of Molecular Biology* 45(2):231.

72. Siegfried NA, Kierzek R, & Bevilacqua PC (2010) Role of Unsatisfied Hydrogen Bond Acceptors in RNA Energetics and Specificity. *J. Am. Chem. Soc.* 132(15):5342-.

73. Sismour AM & Benner SA (2005) The use of thymidine analogs to improve the replication of an extra DNA base pair: a synthetic biological system. *Nucleic Acids Res.* 33(17):5640-5646.

74. Friedberg EC*, et al.* (2006) *DNA Repair and Mutagenesis, 2nd Edition* (ASM Press, Washington, D.C.) 2nd ed Ed.

75. Benner SA (2004) Understanding Nucleic Acids Using Synthetic Chemistry. *Acc. Chem. Res.* 37(10):784-797.

76. Geyer CR, Battersby TR, & Benner SA (2003) Nucleobase Pairing in Expanded Watson-Crick-like Genetic Information Systems. *Structure (London, England : 1993)* 11(12):1485-1498.

77. Testa SM, Disney MD, Turner DH, & Kierzek R (1999) Thermodynamics of RNA-RNA duplexes with 2-or 4-thiouridines: Implications for antisense design and targeting a group I intron. *Biochemistry* 38(50):16655-16662.

78. Chen X, Kierzek R, & Turner DH (2001) Stability and Structure of RNA Duplexes Containing Isoguanosine and Isocytidine. *J. Am. Chem. Soc.* 123(7):1267-1274.

79. Krishnamurthy R (2012) Role of pKa of Nucleobases in the Origins of Chemical Evolution. *Acc. Chem. Res.*

CHAPTER 2: Refining the Genetic Alphabet: A Late-Period Selection Pressure?

Abstract

The transition from genomic RNA to DNA in primitive cells may have created a selection pressure that refined the genetic alphabet, resulting from the global weakening to the *N*-glycosyl bonds. Hydrolytic rupture of these bonds, termed deglycosylation, leaves an abasic site that is the single greatest threat to the stability and integrity of genomic DNA. The rates of deglycosylation are highly dependent on the identity of the nucleobases. Modifications made to the bases such as deamination, oxidation, and alkylation can further increase deglycosylation reaction rates suggesting that the native bases provide optimum *N*-glycosyl bond stability. To protect their genomes, cells have evolved highly specific DNA repair enzymes known as glycosylases that detect and remove these damaged bases. However, the occurrence of modified bases in RNA is deliberate. The dichotomous behavior that cells exhibit toward base modifications may have originated in the RNA world. Modified bases would have been advantageous for the functional and structural repertoire of RNA catalysis. Yet in an early DNA world the utility of these heterocycles was greatly diminished and their presence posed a distinct liability to the stability of cells' genomes. A natural selection for bases exhibiting the greatest resistance to deglycosylation would have ensured the viability of early DNA life along with the recruitment of DNA repair.

Introduction: A selection pressure in the early DNA world

Deoxyribonucleic acid (DNA) consists of four letters that comprise the genetic alphabet: adenine (*A*), thymine (*T*), guanine (*G*) and cytosine (*C*). What were the underlying mechanisms of natural selection that favored these specific nucleobases during the course of biochemical evolution? This question has occupied investigators for decades given its fundamental relation to understanding the origin and evolution of nucleic acids (1-4). Here we hypothesize that a refinement of the genetic alphabet could have taken place after the transition from genomic RNA to DNA in primitive cells. We describe a selection pressure that emerged in the early DNA world and is based on the difference in *N*-glycosyl bond stabilities between RNA and DNA. Stronger *N*-glycosyl bonds may have contributed to the utility of a wide variety of bases employed in RNA for diverse structural and functional capabilities. But with the invention of DNA a global weakening to these bonds emerged and their hydrolysis became the governing threat to genomic stability. Because the stability of *N*-glycosyl bonds is directly related to the identity of the linked heterocycles, a selection pressure for nucleobases most resistant to deglycosylation would have been imposed, resulting in the refinement of the genetic alphabet.

Prebiotic chemistry and alternative bases

Studies in prebiotic chemistry routinely suggest that the native bases, while synthetically accessible and common were likely to have been present in a mixture with numerous other nucleobases (5-7). Reactions such as deamination, aromatic substitution, oxidation and alkylation of exocyclic amines are examples of modifications that readily occur with purines and pyrimidines under prebiotic conditions (8-15). Analyses of carbonaceous meteorites also suggest that both the

extraterrestrial and early terrestrial environment may have been diversely populated with related heterocycles (16-18). Furthermore, contributions illustrating that alternative bases and base pairs can both replace and expand the genetic alphabet continue to underscore the question as to why nature selected the native letters (19-22). The discovery of bacteriophages that employ modified bases completely replacing one of the native letters (Figure 2.1) is testament to the utility of modified letters in a functional DNA alphabet. Pyrimidine derivatives appear to be the most common modifications (23), but there is at least one case where 2,6-diaminopurine (Dap) was found to completely replace adenine (24). Hypoxanthine (Hyp) has not yet been identified in a functional DNA genome but it is known to serve as a reliable guanine surrogate in RNA editing (25) and in DNA (26), and thus in principle could be employed to some extent in viral DNA.

Selection pressures favoring the native bases such as increased photochemical stability (27-28), decreased susceptibility to tautomerization (29-30), and greater comparative stability against decomposition (12), in comparison to other accessible heterocycles, have been discussed. Consequently, many investigations have operated under the assumption that the selection of the native bases could have been made during the prebiotic epoch (13, 31-32), pre-RNA (33-35) and/or the RNA world (36-37). However, with the emergence of DNA, another opportunity for base selection or refinement seems plausible.

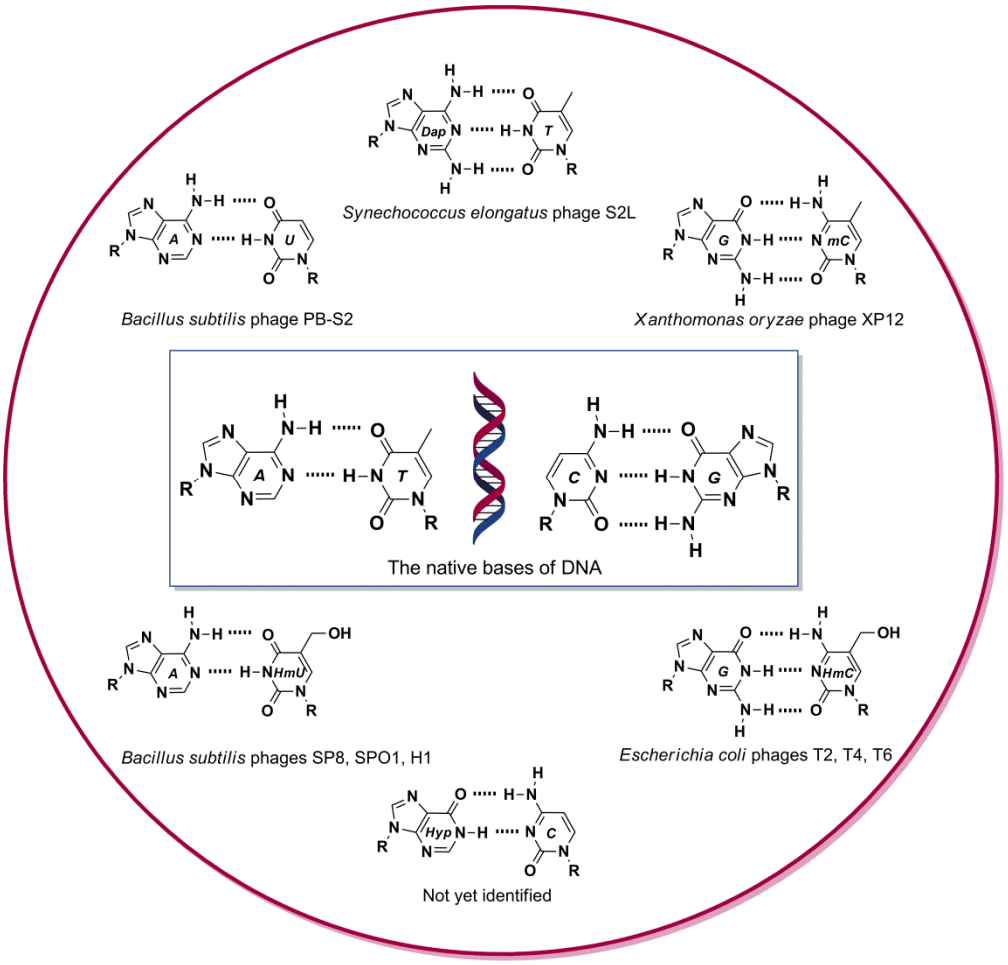


Figure 2.1: Examples of bacteriophages that employ a modified base (HmU=5-hydroxymethyluracil, HmC=5-hydroxymethylcytosine, mC=5-methylcytosine. Dap=2,6-diaminopurine, Hyp=Hypoxanthine) completely replacing one of the native bases in their genomes (23). All of the modified bases shown here are also considered prebiotically relevant heterocycles and the purines have recently been identified in meteorites (17). It is interesting that while hypoxanthine (commonly known as inosine in nucleic acids) is routinely employed as a guanine letter in RNA, has not been identified in bacteriophage functional DNA genomes.

The emergence of labile *N*-glycosyl bonds and DNA repair

The transition from genomic RNA to DNA is widely accepted to be a result of a selection pressure for early forms of life to overcome the kinetic instability of the 3', 5'-phosphodiester bond in ribonucleotides (Figure 2.2) (38-39). The removal of the 2'-OH group in the deoxyribonucleotides has weakened, however, the *N*-glycosyl bonds (40). DNA is vulnerable to a specific type of hydrolytic damage called deglycosylation which involves the loss of a nucleobase via rupture of the *N*-glycosidic bond. Unlike RNA, where the 3',5'-phosphodiester bonds are subjected to transesterification reactions by the 2'-OH group, the loss of genetic information and backbone stability in DNA is dependent on the specific identity of the bases (41). Among these, the purines (A/G) deglycosylate under physiological conditions more frequently than the pyrimidines (C/ T), (Figure 2.2), and heat, divalent metals or a low pH can enhance these reactions even further (40). While RNA can also suffer from depurination, the occurrence of these reactions takes place at significantly reduced rates and low pH (42) (Figure 2.2). Importantly, even slight modifications made to the DNA bases, such as deamination, methylation, or oxidation, typically result with *increased* deglycosylation rates, suggesting that the native bases may provide optimum stability of their *N*-glycosyl bonds (41, 43).

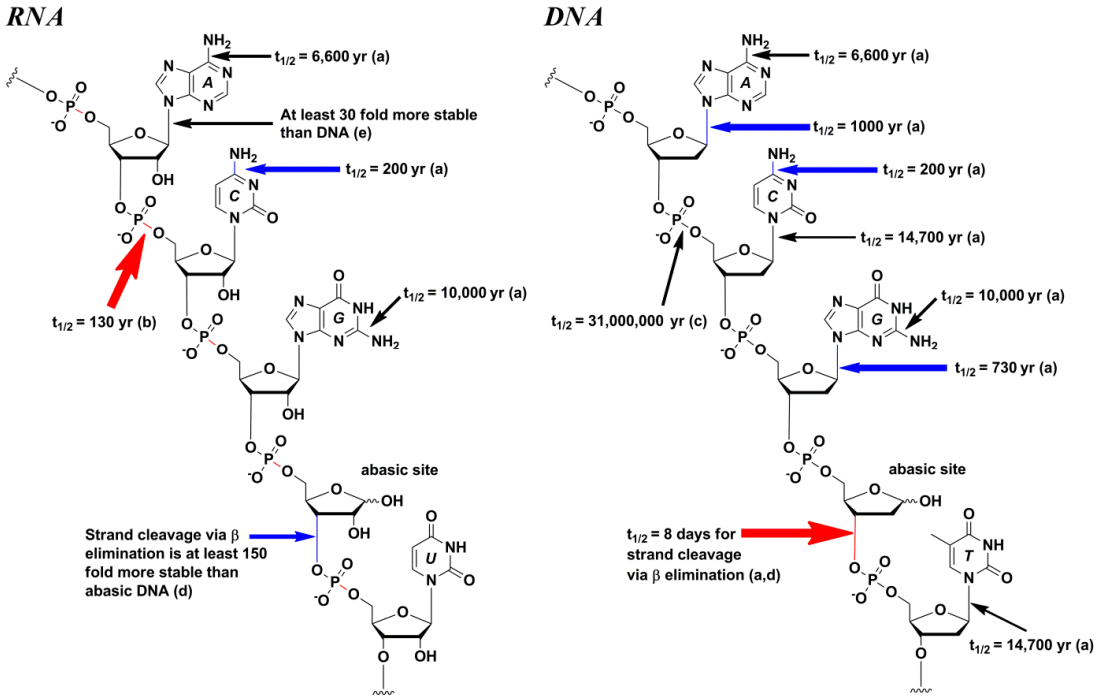


Figure 2.2: Half-life values for spontaneous damage to RNA and DNA. Both polymers are highly susceptible to spontaneous hydrolysis/chemistry as indicated by the color (red: most vulnerable, blue: vulnerable, and black: minimally vulnerable) and arrow width of experimentally determined kinetics for single stranded polymers at neutral pH and extrapolated to 25 – 37°C. While only one 5'-phosphodiester bond is indicated by an arrow, note that any other RNA backbone linkage is vulnerable to cleavage. Phosphodiesters in DNA appear to be essentially stable within the lifetime of any organism. However, it is the formation of abasic sites resulting from deglycosylation that exposes the Achilles Heel of DNA. Abasic sites in DNA, being hemi-acetals, are in equilibrium with their open chain aldehydes (about 1%) and are prone to β elimination reactions and strand cleavage. *Experimental conditions:* **(a)** pH 7.4 , 37°C (40-41, 44); (**b)** pH 7, 23°C (39) ; **(c)** pH 7, 25°C (45-46); **(d)** comparison of abasic stability determined under alkaline pH, 37°C (47); **(e)** comparison of adenosine *N*-glycosyl stability determined at pH7, extrapolated to 25°C(48).

The product of deglycosylation is called an apurinic or apyrimidinic site (AP), also known as an abasic site (Figure 2.2) (40). The formation of an AP site is the single greatest threat to the integrity and stability of DNA (40-41). AP sites are both powerful mutagenic lesions (49) and cytotoxic species given their reactive nature, which could lead to strand cleavage (Figure 2.3) (50-52). For RNA however, abasic sites are less reactive than the DNA counterparts since RNA AP sites maintain enhanced stability against degradation (Figure 2.2) (47). It is important to note that while DNA is extremely resistant to direct phosphodiester bond cleavage (45, 53), it can readily suffer from the same problem that plagues RNA upon deglycosylation (Figure 2.3) (54-55).

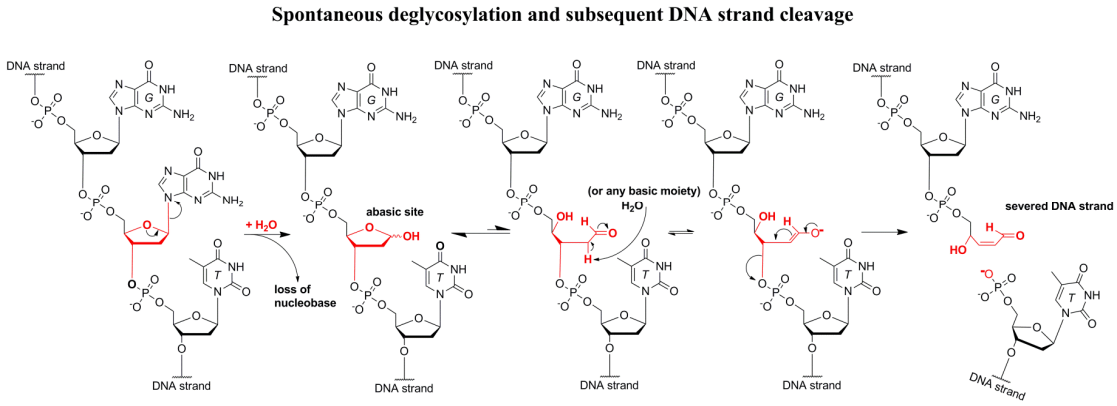


Figure 2.3: General mechanism of hydrolytic rupture of an N-glycosyl bond and the subsequent reactivity of the abasic site.

The generation of abasic sites would have posed a formidable barrier to the persistence of DNA life had it not been for the recruitment of repair proteins (44, 56). Modern cells devote substantial resources to the surveillance and maintenance of their genome which includes AP sites and damaged bases (57). The base excision repair (BER) pathway is a preeminent defense mechanism used by cells which utilizes a variety of enzymes to detect and remove lesions and others to repair AP sites (57). The need for cells to recruit enzymes that repair AP sites could have been the primary pressure that helps explain the origin of this highly sophisticated pathway. This activity has recently been discovered to also exist within the enzymatic capabilities of a DNA polymerase, hinting at a possible early pressure for maintaining genomic stability combined with proper replication (58). Yet, particularly advantageous to the evolution of BER was the recruitment of a class of enzymes known as glycosylases that specifically undertake the task of finding and removing damaged bases as a form of preventive measure to ensure the ‘health’ of a cells’ genome (59) Primitive glycosylases may have evolved to differentiate bases simply by the relative ease of glycosidic bond excision. This feature, being able to differentiate between normal and damaged bases due to differences in N-glycosyl stability has been observed even in modern glycosylases (59-61). It is proposed however, that many extant glycosylases use other methods for the detection and removal of lesions (62). But in an early DNA world using less sophisticated forms of BER, glycosylases that exploit the difference in glycosidic bond stability would seem to be the simplest.

Greater N-glycosyl stability may have aided in the utility of diverse bases in the RNA world

Intriguingly, many of the damaged DNA bases so diligently removed by glycosylases are the exact modifications created by proteins in tRNA, rRNA and mRNA (Table 2.1). Cells seem to exploit the enhanced stability of N-glycosyl bonds in RNA. Modified bases such as alkylated purines and pyrimidines (e.g. 7-methylguanine, 3-methylcytosine) that are unstable lesions in DNA are found to reside in RNAs (Table 2.1) (41, 63). Hypoxanthine, having a weaker glycosidic bond compared to A or G and being a particularly potent mutagenic lesion in DNA (43) is a ubiquitous modification in RNA, which is employed as a reliable guanine surrogate in RNA editing (Figure 2.1 and Table 2.21) (25). Even uracil, the base excluded from the DNA alphabet, is known to exhibit higher deglycosylation rates under neutral pH in comparison to thymine (64). Not all modifications found in RNA are necessarily excluded from DNA. 5-methylcytosine and 5-hydroxymethylcytosine are two examples of nucleobases that are utilized in DNA (Table 2.1, and Figure 2.1) but present interesting peculiarities. These bases along with the parent cytosine are notorious for their rapid rates of spontaneous deamination in comparison to the other letters (Figure 2.2) (12). Yet cytosine is the one base that nature has selected to exploit and its modifications can make up a substantial presence in DNA (65-66). From the viewpoint of genetic fidelity, deamination reactions of C, mC and hmC are the most problematic, but with regards to DNA stability, cytosine retains one of the strongest glycosidic bonds (Figure 2.2). It was further shown in a detailed study using deoxynucleosides to measure spontaneous deglycosylation rates, that cytosine contributes the strongest N-glycosyl bond above all of the native bases(43). It could be that one explanation for nature's particular selection of modifying the C5 position

of cytosine in DNA is because the global impact to N-glycosyl stability is comparatively minimal. Many other modifications highlighted in Table 2.1 (D, HoU, 1mA, N6mA, isoG), while employed in RNA, are necessarily removed from DNA given their inability to maintain or exhibit genetic fidelity or function. Lacking, however, are thorough investigations of relative deglycosylation rates of these and other modified bases compared to the native letters that could provide a quantitative perspective as to how various modifications also affect the N-glycosyl bonds.

Table 2.1: Examples of the dichotomous occurrence of bases in RNA and DNA

Heterocycle	Identified RNA occurrence	Identified DNA occurrence
Hypoxanthine (Hyp)	All three domains in tRNA; used in eukarya mRNA; product of RNA editing; found in eukarya rRNA	Potent mutagenic lesion that arises from adenosine deamination; is a glycosylase substrate.
Uracil (U)	Native to all RNA, is also generated from C in mRNA as part of the eukarya RNA editing process	Potent mutagenic lesion that arises from cytosine deamination and misincorporation by DNA polymerase; is a glycosylase substrate.
5-methyluracil (T)	All three domains in tRNA, also found in rRNA of domain bacteria	Native to DNA (thymine) but also can arise from deamination of 5-methylcytosine and is a glycosylase substrate.
5-methylcytosine (mC)	tRNA in domain archaea, eukarya rRNA all three domains	Used as a epigenetic marker in domain eukarya and completely replaces C in some bacteriophage DNA (see Figure 1)
5,6-Dihydrouracil (D)	All three domains in tRNA, and rRNA in the domain bacteria	An oxidative–deaminated lesion of cytosine; is a glycosylase substrate.
Isoguanine (isoG)	Only known as a naturally occurring ribonucleoside	A lesion resulting from oxidative damage to adenine; is a glycosylase substrate
5-Hydroxyuracil (HoU)	tRNA in domain bacteria, eukarya	An oxidative–deaminated lesion of cytosine; is a glycosylase substrate.
5-hydroxymethyl cytosine (HmC)	Generated in rRNA of domain eukarya	Associated with epigenetic pathways, and completely replaces C in some bacteriophage DNA (see Figure 1)
7-methylguanine (7mG)	All three domains in tRNA, rRNA; used in mRNA as the 5' cap	Lesion resulting from alkylation; is a glycosylase substrate
1-methyladenine (1mA)	All three domains in tRNA and in rRNA of the domain bacteria	Lesion resulting from alkylation; is a glycosylase Substrate.
N^6-methyladenine (N^6mA)	All three domains in rRNA, and in tRNA of archaea and eukarya	Lesion resulting from alkylation; is a glycosylase substrate
3-methylcytosine (3mC)	tRNA of domain bacteria and eukarya	Lesion resulting from alkylation; is a glycosylase substrate

Table References: (59, 63, 67-69)

What is the evolutionary origin of the diversity seen in RNA bases? Although all these heterocycles result from posttranscriptional modifications, their structures resemble side chains of amino acids (11, 70-71). The greater stability of RNA *N*-glycosyl bonds may have thus been an advantageous feature in the RNA world (Figure 2.4). The utility of modified and exotic bases could have expanded the repertoire of catalytically competent RNA oligomers (11, 70, 72-75) without the consequences of rapid deglycosylation. However, with the emergence of DNA and the greater utility of proteins, the functionalized bases would have become obsolete and a detriment to the survival of life (Figure 2.4). While the selection of the DNA bases may have been aided by those exhibiting the greatest glycosidic bond stability, the eventual refinement of RNA bases would have largely mirrored the selection process based on energetic costs to the cell. With the increasing takeover by proteins only the most essential bases for structural and functional roles in RNA would have continued to persist. Modified bases identified in contemporary RNA may thus actually have been part of the original larger family of diverse bases utilized in the RNA world (74).

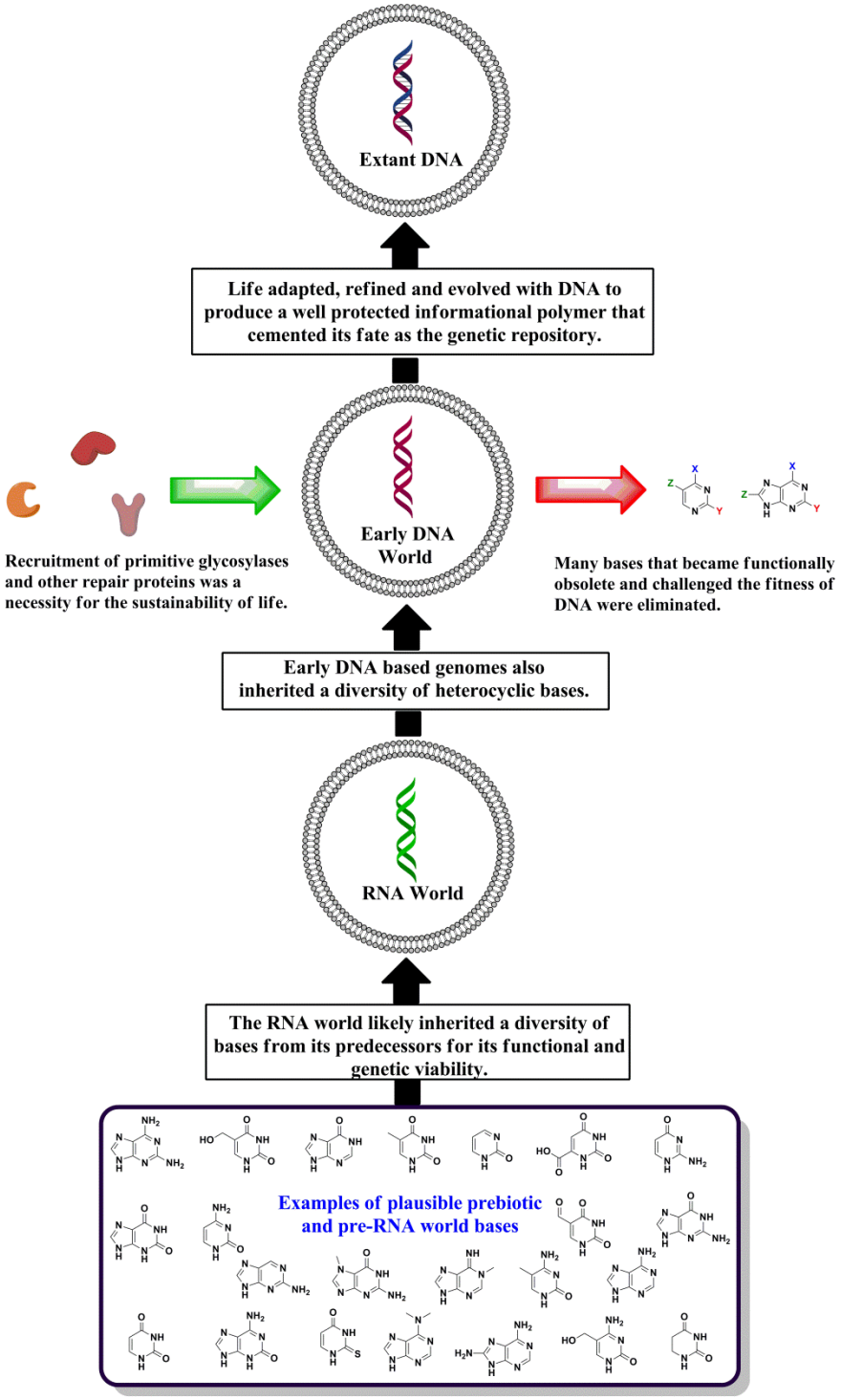


Figure 2.4: Hypothesis diagram illustrating a refinement of the genetic alphabet.

Conclusion

The selection of the native bases did not occur in any one hypothetical period. It is more likely that a continuous process of refinement directed their selection throughout prebiotic and early biotic epochs. As suggested here, the fundamental change to the differences in hydrolytic susceptibility between RNA and DNA may have contributed to this refinement process. While life in the RNA world may have been challenged by the nature of the sugar moiety and its impact on backbone stability, in an early DNA world the governing pressure came from the identity of the attached nucleobase.

In this sense, the arrival of DNA should not be considered just a later modification of RNA, rather, it is a unique biopolymer in its own right that challenged life to adapt to its specific chemical vulnerabilities, to further refine the genetic alphabet and evolve repair pathways that allowed for the ubiquity of DNA as we know it.

Acknowledgements

We are grateful to the NIH for support (via grant number GM 069773), Dr. Ulrich Muller and the reviewers for their helpful comments to our manuscript.

Chapter 2 is a minimally modified reprint from: Rios, A.C., and Tor, Y., *Astrobiology,* **2012**, *12,* 884 – 891. The dissertation author is the main author and researcher for this work.

References

1. Rich A (1962) On the problems of evolution and biochemical information transfer. *Horizons in Biochemistry* eds Kasha M & Pullman B (Academic Press, New York), pp 103 - 126.

2. Westheimer F (1987) Why nature chose phosphates. *Science* 235(4793):1173-1178.

3. Eschenmoser A (1999) Chemical Etiology of Nucleic Acid Structure. *Science* 284(5423):2118-2124.

4. Szathmary E (2003) Why are there four letters in the genetic alphabet? *Nat Rev Genet* 4(12):995-1001.

5. Benner SA, Kim H-J, Kim M-J, & Ricardo A (2010) Planetary Organic Chemistry and the Origins of Biomolecules. *Cold Spring Harbor Perspect. Biol.* 2(7):a003467.

6. Borquez E, Cleaves HJ, Lazcano A, & Miller SL (2005) An Investigation of Prebiotic Purine Synthesis from the Hydrolysis of HCN Polymers. *Origins Life Evol. Biospheres* 35(2):79-90.

7. Orgel LE (2004) Prebiotic Chemistry and the Origin of the RNA World. *Crit. Rev. Biochem. Mol. Biol.* 39(2):99-123.

8. Shapiro R (1999) Prebiotic cytosine synthesis: A critical analysis and implications for the origin of life. *Proc. Natl. Acad. Sci. U.S.A.* 96(8):4396-4401.

9. Barks HL*, et al.* (2010) Guanine, Adenine, and Hypoxanthine Production in UV-Irradiated Formamide Solutions: Relaxation of the Requirements for Prebiotic Purine Nucleobase Formation. *ChemBioChem* 11(9):1240-1243.

10. Robertson MP & Miller SL (1995) An efficient prebiotic synthesis of cytosine and uracil. *Nature* 375(6534):772-774.

11. Robertson M & Miller S (1995) Prebiotic synthesis of 5-substituted uracils: a bridge between the RNA world and the DNA-protein world. *Science* 268(5211):702-705.

12. Levy M & Miller SL (1998) The stability of the RNA bases: Implications for the origin of life. *Proc. Natl. Acad. Sci. U.S.A.* 95(14):7933-7938.

13. Powner MW, Gerland B, & Sutherland JD (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 459(7244):239-242.

14. Shapiro R (1995) The prebiotic role of adenine: A critical analysis. *Origins Life Evol. Biospheres* 25(1):83-98.

15. Siegel JS & Tor Y (2005) Genetic alphabetic order: what came before A? *Org. Biomol. Chem.* 3(9):1591-1592.

16. Botta O & Bada JL (2002) Extraterrestrial Organic Compounds in Meteorites. *Surv. Geophys.* 23(5):411-467.

17. Callahan MP*, et al.* (2011) Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases. *Proc. Natl. Acad. Sci. U.S.A.* 108(34):13995-13998.

18. Martins Z*, et al.* (2008) Extraterrestrial nucleobases in the Murchison meteorite. *Earth Planet. Sci. Lett.* 270(1-2):130-136.

19. Benner SA (2004) Understanding Nucleic Acids Using Synthetic Chemistry. *Acc. Chem. Res.* 37(10):784-797.

20. Piccirilli JA, Benner SA, Krauch T, & Moroney SE (1990) Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature* 343(6253):33-37.

21. Benner SA & Sismour AM (2005) Synthetic biology. *Nat Rev Genet* 6(7):533-543.

22. Chiba J & Inouye M (2010) Exotic DNAs Made of Nonnatural Bases and Natural Phosphodiester Bonds. *Chem. Biodiversity* 7(2):259-282.

23. Warren RAJ (1980) Modified Bases in Bacteriophage DNAs. *Annu. Rev. Microbiol.* 34(1):137-158.

24. Kirnos MD, Khudyakov IY, Alexandrushkina NI, & Vanyushin BF (1977) 2-Aminoadenine is an adenine substituting for a base in S-2L cyanophage DNA. *Nature* 270(5635):369-370.

25. Nishikura K (2010) Functions and Regulation of RNA Editing by ADAR Deaminases. *Annu. Rev. Biochem.* 79(1):321-349.

26. Budke B & Kuzminov A (2006) Hypoxanthine Incorporation Is Nonmutagenic in Escherichia coli. *J. Bacteriol.* 188(18):6553-6560.

27. Serrano-Andres L & Merchan M (2009) Are the five natural DNA/RNA base monomers a good choice from natural selection? A photochemical perspective. *J. Photochem. Photobiol. C-Photochem. Rev.* 10(1):21-32.

28. Abo-Riziq A*, et al.* (2005) Photochemical selectivity in guanine–cytosine base-pair structures. *Proc. Natl. Acad. Sci. U.S.A.* 102(1):20-23.

29. Roberts C, Bandaru R, & Switzer C (1997) Theoretical and Experimental Study of Isoguanine and Isocytosine: Base Pairing in an Expanded Genetic System. *J. Am. Chem. Soc.* 119(20):4640-4649.

30. Robinson H*, et al.* (1998) 2'-Deoxyisoguanosine Adopts More than One Tautomer To Form Base Pairs with Thymidine Observed by High-Resolution Crystal Structure Analysis. *Biochemistry* 37(31):10897-10905.

31. Sutherland JD (2010) Ribonucleotides. *Cold Spring Harbor Perspect. Biol.* 2(4):a005439.

32. Powner MW, Sutherland JD, & Szostak JW (2010) Chemoselective Multicomponent One-Pot Assembly of Purine Precursors in Water. *J. Am. Chem. Soc.* 132(46):16677-16688.

33. Bean Heather D, Lynn David G, & Hud Nicholas V (2009) Self-Assembly and the Origin of the First RNA-Like Polymers. *Chemical Evolution II: From the Origins of Life to Modern Society,* ACS Symposium Series, (American Chemical Society, Washington DC), pp 109-132.

34. Joyce GF (2002) The antiquity of RNA-based evolution. *Nature* 418(6894):214-221.

35. Engelhart AE & Hud NV (2010) Primitive Genetic Polymers. *Cold Spring Harbor Perspect. Biol.* 2(12):a002196.

36. Joyce GF (1989) RNA evolution and the origins of life. *Nature* 338(6212):217-224.

37. Bean HD*, et al.* (2007) Formation of a β-Pyrimidine Nucleoside by a Free Pyrimidine Base and Ribose in a Plausible Prebiotic Reaction. *J. Am. Chem. Soc.* 129(31):9556-9557.

38. Lazcano A, Guerrero R, Margulis L, & Oró J (1988) The evolutionary transition from RNA to DNA in early cells. *J. Mol. Evol.* 27(4):283-290.

39. Li Y & Breaker RR (1999) Kinetics of RNA Degradation by Specific Base Catalysis of Transesterification Involving the 2'-Hydroxyl Group. *J. Am. Chem. Soc.* 121(23):5364-5372.

40. Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* 362(6422):709-715.

41. Gates KS (2009) An Overview of Chemical Processes That Damage Cellular DNA: Spontaneous Hydrolysis, Alkylation, and Reactions with Radicals. *Chem. Res. Toxicol.* 22(11):1747-1760.

42. Kochetkov NK & Budovskii EI (1972) *Organic chemistry of nucleic acids* (Plenum Press, New York,) pp 425 - 428.

43. Schroeder GK & Wolfenden R (2007) Rates of Spontaneous Disintegration of DNA and the Rate Enhancements Produced by DNA Glycosylases and Deaminases†. *Biochemistry* 46(47):13638-13647.

44. Friedberg EC*, et al.* (2006) *DNA Repair and Mutagenesis, 2nd Edition* (ASM Press, Washington, D.C.) 2nd ed Ed.

45. Schroeder GK, Lad C, Wyman P, Williams NH, & Wolfenden R (2006) The time required for water attack at the phosphorus atom of simple phosphodiesters and of DNA. *Proc. Natl. Acad. Sci. U.S.A.* 103(11):4052-4055.

46. Wolfenden R (2011) Benchmark Reaction Rates, the Stability of Biological Molecules in Water, and the Evolution of Catalytic Power in Enzymes. *Annu. Rev. Biochem.* 80(1):645-667.

47. Küpfer PA & Leumann CJ (2006) The chemical stability of abasic RNA compared to abasic DNA. *Nucleic Acids Res.* 35(1):58-68.

48. Stockbridge RB, Schroeder GK, & Wolfenden R (2010) The rate of spontaneous cleavage of the glycosidic bond of adenosine. *Bioorganic Chem.* 38(4-6):224-228 .

49. Loeb LA & Preston BD (1986) Mutagenesis by Apurinic/Apyrimidinic Sites. *Annu. Rev. Genet.* 20(1):201-230.

50. Shapiro R (1981) Damage to DNA caused by hydrolysis *Chromosome Damage and Repair*, eds Seeberg E & Kleppe K (Plenum Press, New York), pp 3-18.

51. Boiteux S & Guillet M (2004) Abasic sites in DNA: repair and biological consequences in Saccharomyces cerevisiae. *DNA Repair* 3(1):1-12.

52. Lhomme J, Constant J-F, & Demeunynck M (1999) Abasic DNA structure, reactivity, and recognition. *Biopolymers* 52(2):65-83.

53. Williams NH, Takasaki B, Wall M, & Chin J (1999) Structure and nuclease activity of simple dinuclear metal complexes: Quantitative dissection of the role of metal ions. *Acc. Chem. Res.* 32(6):485-493.

54. Eigner J, Boedtker H, & Michaels G (1961) Thermal degradation of nucleic acids. *Biochim. Biophys. Acta* 51(1):165-168.

55. Sugiyama H*, et al.* (1994) Chemistry of Thermal Degradation of Abasic Sites in DNA. Mechanistic Investigation on Thermal DNA Strand Cleavage of Alkylated DNA. *Chem. Res. Toxicol.* 7(5):673-683.

56. Jensen RA (1976) Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* 30:409-425.

57. Lindahl T & Wood RD (1999) Quality Control by DNA Repair. *Science* 286(5446):1897-1905.

58. Banos B, Villar L, Salas M, & de Vega M (2010) Intrinsic apurinic/apyrimidinic (AP) endonuclease activity enables Bacillus subtilis DNA polymerase X to recognize, incise, and further repair abasic sites. *Proc. Natl. Acad. Sci. U.S.A.* 107(45):19219-19224.

59. O'Brien PJ (2006) Catalytic Promiscuity and the Divergent Evolution of DNA Repair Enzymes. *Chem. Rev.* 106(2):720-752.

60. Bennett MT*, et al.* (2006) Specificity of Human Thymine DNA Glycosylase Depends on N-Glycosidic Bond Stability. *J. Am. Chem. Soc.* 128(38):12510-12519.

61. O'Brien PJ & Ellenberger T (2004) The Escherichia coli 3-Methyladenine DNA Glycosylase AlkA Has a Remarkably Versatile Active Site. *J. Biol. Chem.* 279(26):26876-26884.

62. Friedman JI & Stivers JT (2010) Detection of Damaged DNA Bases by DNA Glycosylase Enzymes. *Biochemistry* 49(24):4957-4967.

63. Limbach PA, Crain PF, & McCloskey JA (1994) Summary: the modified nucleosides of RNA. *Nucleic Acids Res.* 22(12):2183-2196.

64. Shapiro R & Kang S (1969) Uncatalyzed hydrolysis of deoxyuridine, thymidine, and 5-bromodeoxyuridine. *Biochemistry* 8(5):1806-1810.

65. Poole A, Penny D, & Sjöberg B-M (2001) Confounded cytosine! Tinkering and the evolution of DNA. *Nat Rev Mol Cell Biol* 2(2):147-151.

66. Nabel CS, Manning SA, & Kohli RM (2011) The Curious Chemical Biology of Cytosine: Deamination, Methylation,and Oxidation as Modulators of Genomic Potential. *ACS Chem. Biol.* 7(1):20-30.

67. Fuhrman F, Fuhrman G, Nachman R, & Mosher H (1981) Isoguanosine: isolation from an animal. *Science* 212(4494):557-558.

68. Ushijima Y*, et al.* (2005) A functional analysis of the DNA glycosylase activity of mouse MUTYH protein excising 2-hydroxyadenine opposite guanine in DNA. *Nucleic Acids Res.* 33(2):672-682.

69. Rozenski J, Crain PF, & McCloskey JA (1999) The RNA Modification Database: 1999 update. *Nucleic Acids Res.* 27(1):196-197.

70. Levy M & Miller SL (1999) The Prebiotic Synthesis of Modified Purines and Their Potential Role in the RNA World. *J. Mol. Evol.* 48(6):631-637.

71. Lazcano A (1994) *The transition from nonliving to living* (Columbia University Press, New York) pp x, 630 p.

72. Nguyen KV & Burrows CJ (2011) A Prebiotic Role for 8-Oxoguanosine as a Flavin Mimic in Pyrimidine Dimer Photorepair. *J. Am. Chem. Soc.* 133(37):14586-14589.

73. Benner SA, Burgstaller P, Battersby TR, & Jurczyk S (1999) Chapter 6. Did the RNA World Exploit an Expanded Genetic Alphabet? *The RNA World*, eds Raymond F Gesteland, Cech TR, & Atkins JF (Cold Spring Harbor Laboratory Press, Cold Spring Harbor), Vol 2, pp 163-181.

74. Cermakian N & Cedergren R (1998) Chapter 29. Modified Nucleosides Always Were: an Evolutionary Model. *Modification and Editing of RNA*, eds Grosjean H & Benne R (ASM Press, Washington DC), pp 535 - 541.

75. Forterre P & Grosjean H (2009) Chapter 19. The Interplay between RNA and DNA Modifications:Back to the RNA World. *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution,* Molecular Biology Intelligence United Grosjean H (Landes Bioscience, Austin, Texas), pp 259 - 274.

CHAPTER 3: Hydrolytic stabilities of *N*-glycosyl bonds in modified, alternative and damaged nucleosides: Trends and implications on the refinement of the genetic alphabet

---

Abstract

A systematic investigation of relative *N*-glycosyl stabilities of ribonucleosides and deoxyribonucleosides under acidic conditions (pH 1 – 3) was conducted. Nucleosides were chosen based on the relevance of the heterocycles to prebiotic chemistry, RNA and DNA modifications and DNA damage. Rate constants were measured as a function of temperature (at pH 1) to produce Arrhenius and Eyring plots for extrapolation to 37^{o}C and determination of thermodynamic activation parameters. Rate enhancements (k_{deoxy}/k_{ribos}) based on the difference in reaction rates of deoxyribo and ribo-glycosidic bonds were found to vary under the same conditions. This was verified by the calculated difference in Gibbs free energy of activation ($\Delta\Delta G^{\ddagger}$), which was found to range from 3 – 4.5 kcal/mol. No noticeable correlation to pK_a values was noticed for the purine nucleosides that could explain the reaction rates under acidic conditions. While the enthalpic parameter of activation ($\Delta H^{\ddagger}$) was always the dominant contributing parameter to $\Delta G^{\ddagger,}$,it was found that the entropic contribution ($\Delta S^{\ddagger}$) affected the overall hierarchy in relative reaction rates. Across the pH range of 1 – 3 it was found that the native purine deoxynucleosides exhibited the slowest reaction rates and were similar to one other in their pH dependent reactivity. Only the 2,6-diaminopurine deoxynucleoside approached the slower reactivity of native purines as the pH was increased to 3. Along with literature data, the implications for extant base modification and relevance to the composition of the genetic alphabet are discussed.

Introduction

Nature employs diverse nucleobases in nucleic acids. While these alternative heterocycles can be viewed as modifications to the canonical bases (1-2), their origin may also reflect prebiotic or early biotic processes (3-6). The functionally-limited bases of the contemporary genetic alphabet present a chemical problem to the assumed catalytic diversity that existed within functional RNAs according to the RNA world hypothesis. This predicament has been countered with a relaxation on the assumption of a four base alphabet (7). An expanded set of nucleobases included in the RNA world is supported with observations from simulated prebiotic experiments (4, 8-10) identification of alternative bases in meteorites (11-12) and the known capacity for functional RNAs to be enzymatically modified with non-native bases(13-14). Recent contributions have added intrigue to this concept with the demonstration that even 'damaged bases' could have endowed RNA with beneficial properties in the RNA world (15). It seems reasonable to suggest that a diverse population of nucleobases during the prebiotic epoch had influenced the emergence of primordial genetic polymers and the origin of life (16-18) This suggestion might appear to exacerbate questions relating to the eventual selection of the genetic alphabet, but selection pressures may have existed during and after the era of the RNA world as life made the transition to the early DNA world (19).

We have recently hypothesized that fundamental differences in *N*-glycosidic bond stabilities between RNA and DNA could have been beneficial in the RNA world but created a selection pressure in the early DNA world (20). As *N*-glycosyl stabilities are inherently linked to the nature of the nucleobase, this pressure may explain why some bases are widely distributed in RNA but excluded from DNA, and consequently helped refine the specific identity of the alphabetic letters. Intriguingly, the *N*-

glycosidic bond as it relates to the origin of nucleic acids and the genetic alphabet has largely been studied in the context of its formation from a prebiotic “bottom up” approach (21-25). Conversely, considerations of *N*-glycosyl stabilities and their rupture have been largely and traditionally studied in the context of DNA (and RNA) damage, mutagenesis and DNA repair(26-28). We submit, however, that the study of relative *N*-glycosyl stabilities from the perspective of biochemical evolution could provide insight as to the specific composition of bases found in the nucleic acids of extant biology.

In this contribution we systematically study the hydrolytic stability of *N*-glycosyl bonds in selected nucleosides (Table 3.1), which contain heterocycles that are relevant to prebiotic chemistry and are of significance in DNA damage as well as in RNA and DNA modification. We compare these results to the hydrolytic stability of the native purine and pyrimidine nucleosides. Deglycosylation kinetics were determined for ribo- and deoxyribo-nucleosides under acidic conditions to address two fundamental questions: a) Does the stability difference between a ribo-glycosidic and a deoxyribo-glycosidic bond remain the same or does it vary depending on the nature of the attached nucleobase?, and, b) how do the native nucleosides, with a particular focus on the deoxynucleosides, compare in their global *N*-glycosyl stability to other related modified or alternative nucleosides. Experimentally, the temperature dependence of deglycosylation rates was measured under acidic conditions (0.1M HCl) for all nucleosides in order to obtain rate extrapolations, enhancements and thermodynamic activation parameters at 37^{o}C from the corresponding Arrhenius and Eyring curves. Additionally, comparative hydrolysis kinetics at 50^{o}C between pH 1 and pH 3, was determined for the deoxynucleosides. Low pH was chosen for this study mainly to facilitate the comparison of ribonucleoside and deoxyribonucleoside

kinetics under the same experimental conditions and for comparison to previous studies. It is worth noting that given the abundance of data obtained over the years, little attention has surprisingly been given to the role of glycosidic bond stability in the evolution of nucleobase repertoire and the genetic alphabet (23, 25). We demonstrate that relative *N*-glycosyl stability in ribonucleosides compared to deoxyribonucleosides under the same conditions can change as a result to the differences in Gibbs free energy of activation ($\Delta\Delta G^{\ddagger}$), We show that while modified nucleosides, even with modest changes to the heterocycles can exhibit wider differences in their glycosidic bond stabilities, the native purines exhibit similar *N*-glycosyl stabilities to each other and deglycosylate with rates that are the slowest. The implications of these stability differences in the context of RNA and DNA base modification and damage, as well as their potential role as a selection criterion refining the composition of the genetic alphabet are also discussed.

Table 3.1: Nucleosides used in study

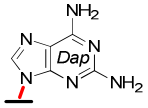
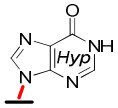
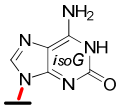
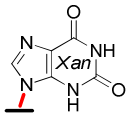
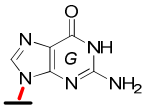
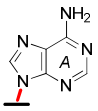
Heterocycle	Prebiotic relevance	Occurrence in RNA Biology	Occurrence in DNA Biology	Ref.
	Identified in Meteorites, Low abundance in prebiotic experiments	Not yet identified	Replaces adenine(100%) in dsDNA of Cyanophage S-2L	(11, 36-39)
	Identified in Meteorites; High abundance in prebiotic experiments, related to A via deamination	Ubiquitous across All domains; tRNA, rRNA, and in mRNA editing of Biosynthetic Precursor to purine ribonucleotides	Occurs as a lesion from spontaneous deamination to Adenine	(11, 39-44)
	Not yet found in meteorites. Low abundance in prebiotic experiments related to Dap via deamination	Occurs as a ribonucleoside in isolated organisms, not yet identified in polymeric RNA	Occurs as a lesion from oxidative damage to Adenine	(39, 45-46)
	Found in meteorites Moderate abundance in prebiotic simulated experiments and related to Dap, isoG and G via deamination	Biosynthetic precursor to Guanosine monophosphate Not yet identified in polymeric RNA	Occurs as a lesion from spontaneous deamination to guanine	(11, 43-44)
	Identified in prebiotic experiments, demonstrated to be first successful prebiotic glycosylation of a pyrimidine with ribose	Not yet identified	Not yet identified	(10, 23)

Table 3.1: cont.

	Found in meteorites moderate abundances in simulated prebiotic experiments, related to Dap via deamination	Alphabet letter	Alphabet letter	(9, 11, 38-39, 43)
	Found in Meteorites. High abundances in prebiotic experiments	Alphabet letter	Alphabet letter	(11, 39, 43)

Results

Rates of acid-catalyzed deglycosylation reactions were studied as functions of pH and temperature and followed spectrophotometrically using the alkaline quench method. Selected examples of spectral changes that occur over the time course of the reaction for 2,6-diaminopurine-2'-deoxyriboside, 2'-deoxyisoguanosine, and 2'-deoxyinosine are shown in Figure 3.1A–C, respectively. Also shown in Figure 3.1 are the resulting pseudo-first order kinetics curves from the corresponding plots of a single-wavelength absorbance versus time, which acid-catalyzed deglycosylation reactions are known to exhibit (33) Additional example of UV-Vis spectra of nucleoside deglycosylation reactions and kinetics plots are included in Appendix A.

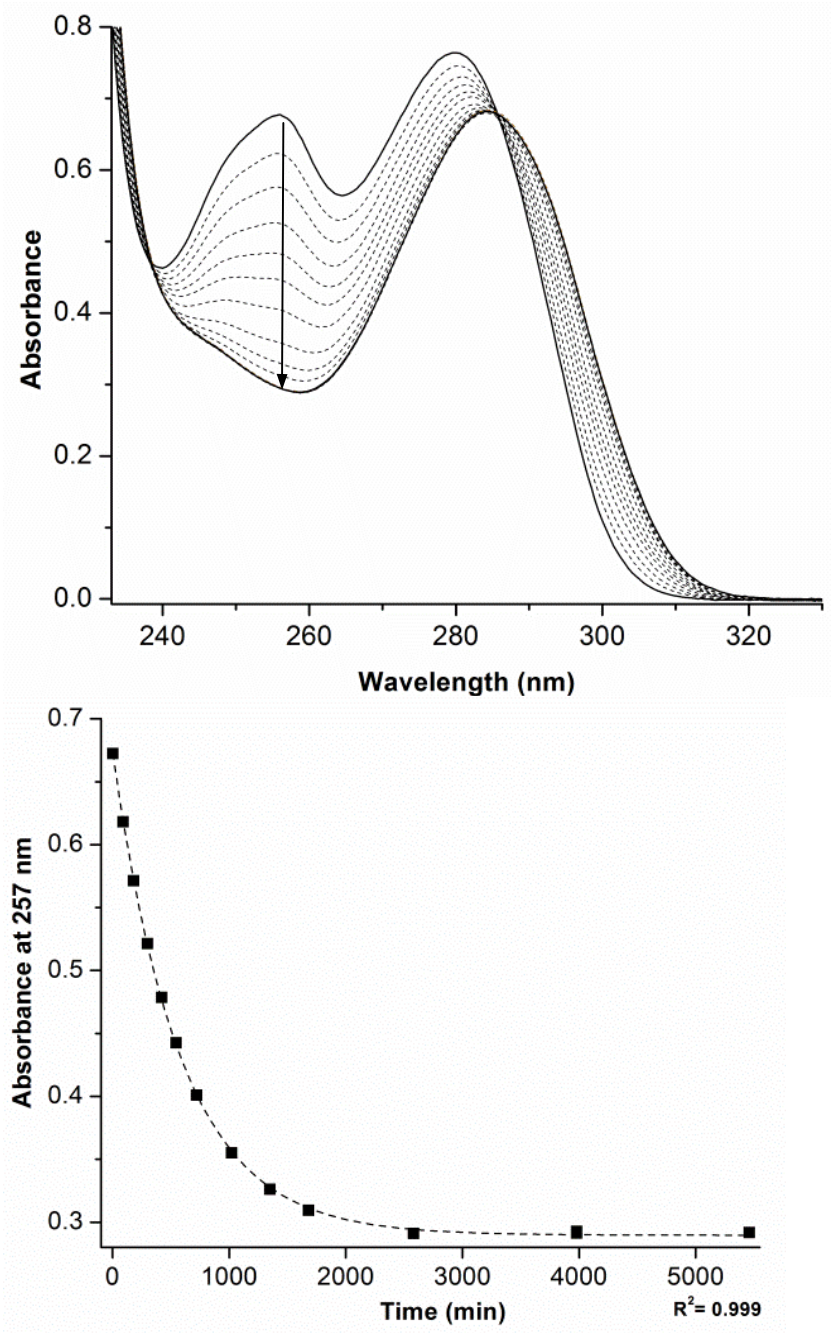


Figure 3.1: Examples of reaction monitoring by UV-Vis spectroscopy using the alkaline quench method (see text and Materials and Methods) for three deoxynucleosides during the course of a deglycosylation reaction at pH 3 and 50^{o}C. In all cases distinct hypochromicity coupled to bathchromic shifts were observed during the course of the reactions. The λ with the greatest absorption difference (indicated with arrows) were chosen for each of the corresponding plots of absorbance vs. time to determine rate constants. **A.** 2,6-Diaminopurine-2'-deoxyriboside →2,6-diaminopurine spectral overlay (top) and absorbance, λ = 257 nm vs. time plot (bottom).

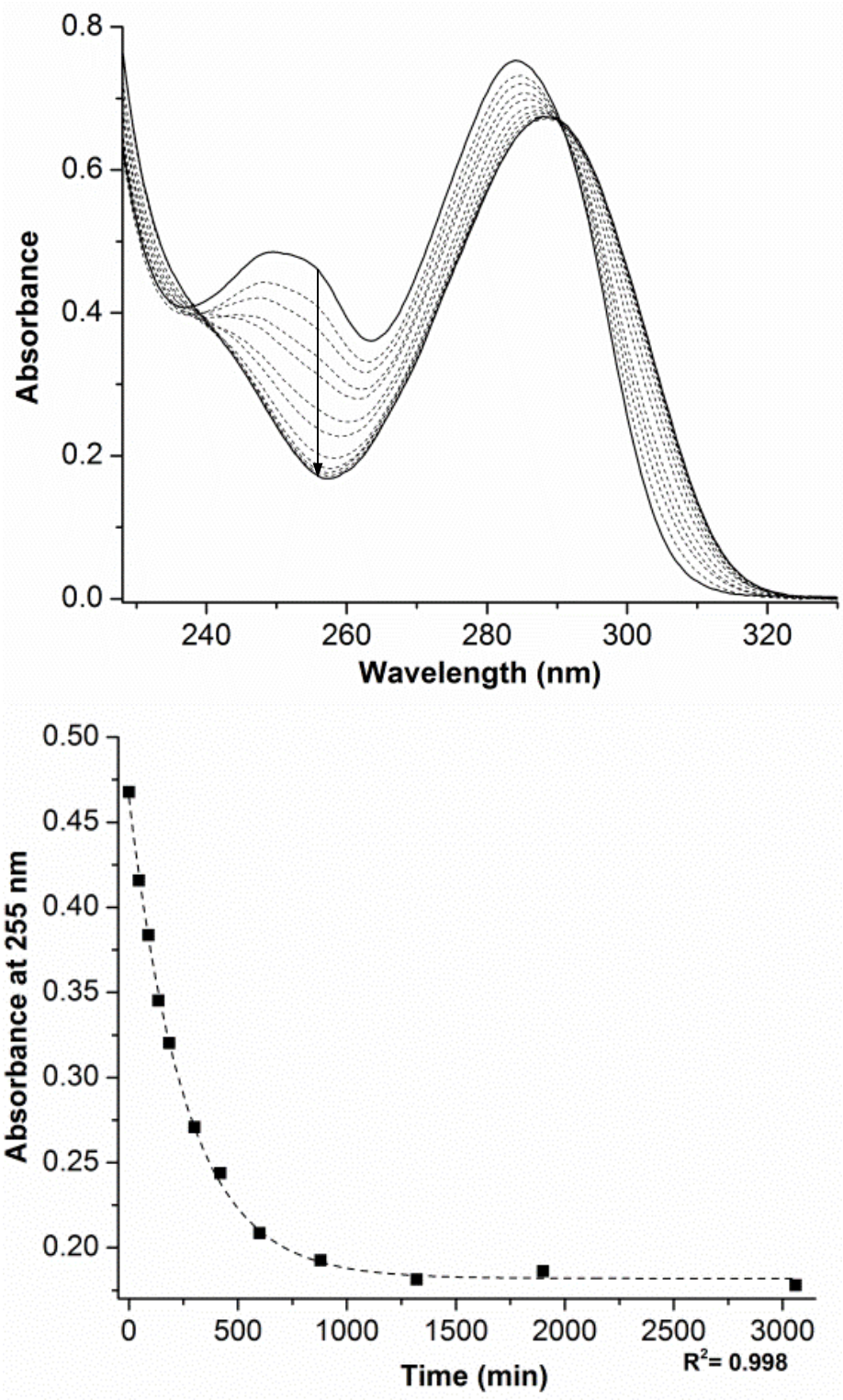
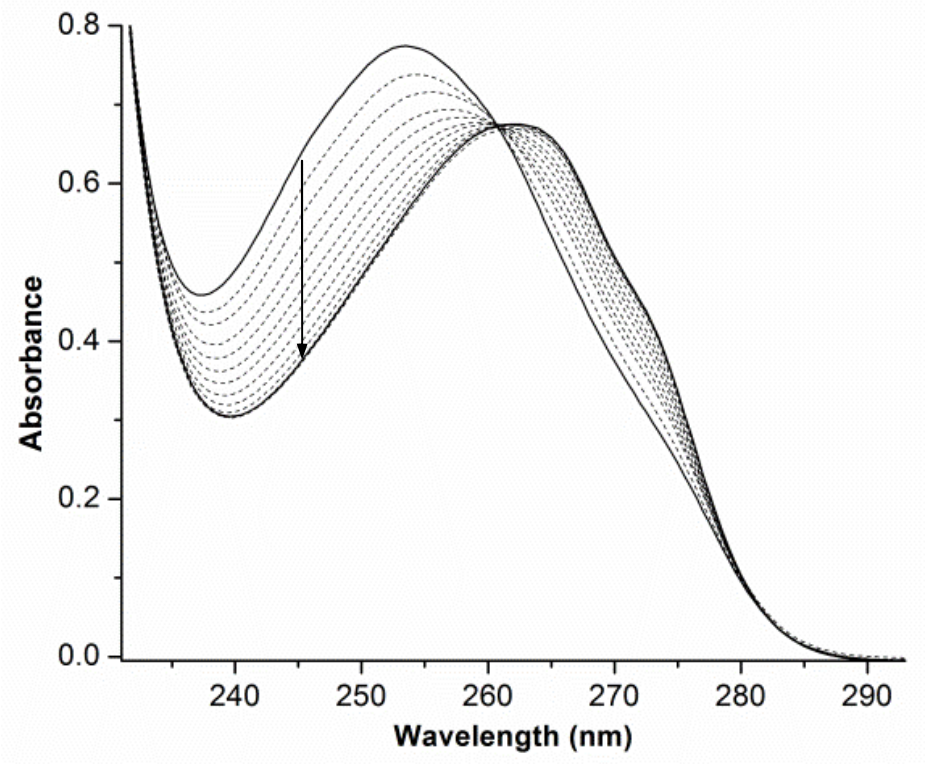


Figure 3.1, **Cont. B**. 2'-deoxyisoguanosine→isoguanine spectral overlay (top) and absorbance, λ = 255 nm vs. time plot (bottom).



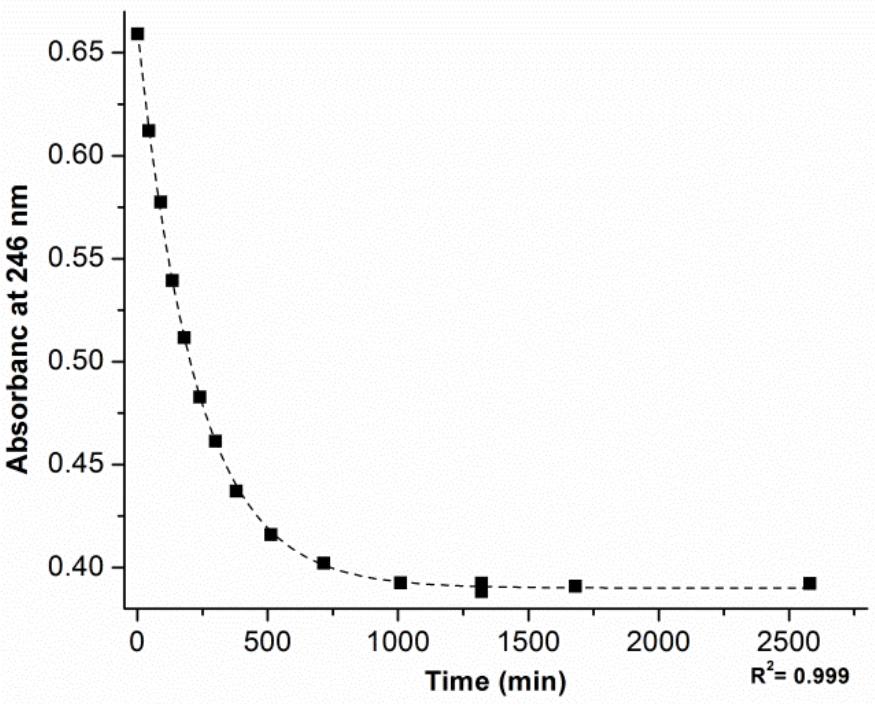


Figure 3.1, **Cont. C.** 2'-deoxyinosine→hypoxanthine spectral overlay (top) and absorbance, λ = 246 nm vs. time plot (bottom)

Figure 3.2 shows the compiled Arrhenius curves, which contain the extrapolated rate constants to 37°C for all the ribonucleosides and all but two of the deoxynucleosides (see caption Figure 3.2). These data along with the corresponding thermodynamic parameters of activation at 37°C, obtained from Eyring plots (see Appendix A) , are summarized in Table 3.2 and arranged by nucleoside with the slowest deglycosylation rate listed first and the fastest listed last.

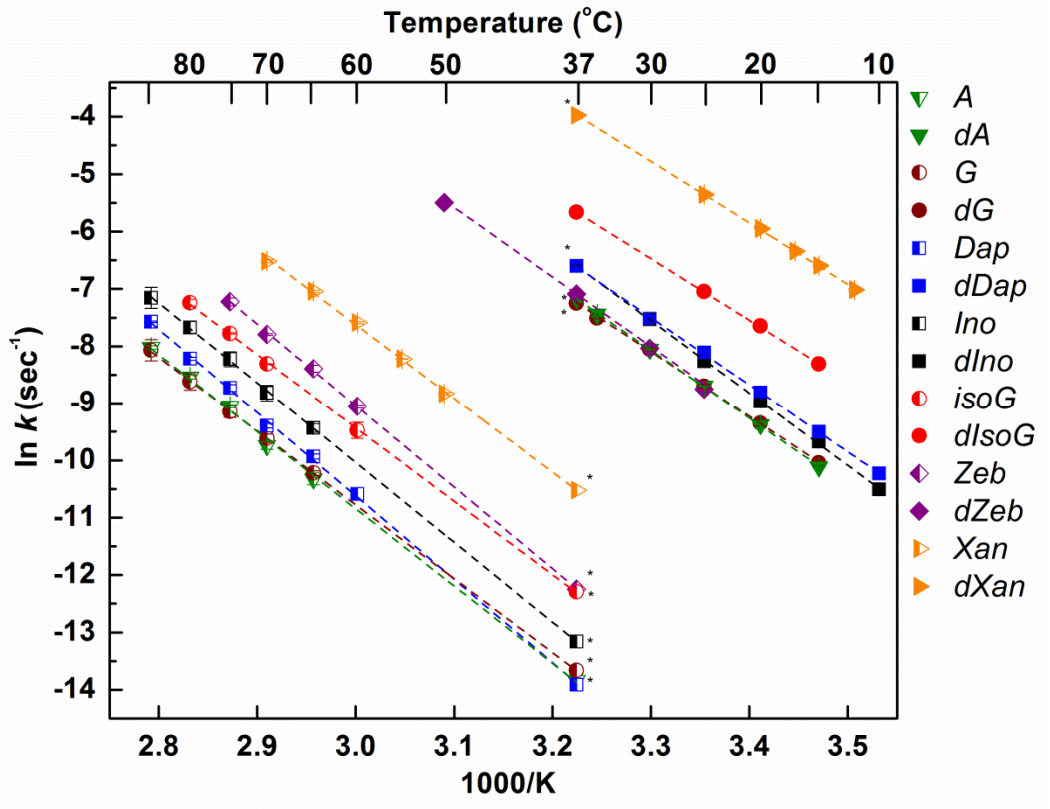


Figure 3.2: Arrhenius plots of deglycosylation kinetics, at pH 1, of ribonucleosides and deoxyribonucleosides. Extrapolated data of ribonucleosides and four deoxyribonucleosides to 37 °C are indicated with an asterisk. The nucleosides disoG and dZeb are the only two that have reaction rates experimentally determined (duplicate and triplicate trials respectively) at 37°C.

Table 3.2: Rate constants, half-lives, and thermodynamic activation parameters for acid catalyzed (0.1M HCl) deglycosylation reactions at 37°C.

Nucleoside	Rate, $10^6 k$, s^{-1}	$t_{1/2}$	$\Delta G^{\ddagger}$ kcal/mol (kJ/mol)	$\Delta H^{\ddagger}$ kcal/mol	$\Delta S^{\ddagger}$ cal/mol•K
2,6-Diaminopurine riboside (Dap)	0.91 ± 0.03	8.8 days	26.75 ±0.02 (111.9)	28.4 ± 0.8	5.4 ± 0.2
Adenosine (A)	0.98 ± 0.05	8.2 days	26.71 ±0.03 (111.7)	26.1 ± 0.9	-1.87 ± 0.1
Guanosine (G)	1.17 ± 0.04	6.9 days	26.60 ±0.02 (111.3)	24.9 ± 0.6	-5.3 ± 0.2
Inosine (Ino)	1.92 ±0.02	4.2 days	26.29 ±0.01 (110.0)	27.1 ± 0.2	2.6 ± 0.1
Isoguanosine (IsoG)	4.58 ±0.04	1.7 days	25.76 ±0.01 (107.7)	24.8 ± 0.2	-2.9 ± 0.1
Zebularine (Zeb)	4.81 ±0.05	1.7 days	25.73 ±0.01 (107.6)	27.8 ± 0.3	6.8 ± 0.1
Xanthosine (Xan)	26.9±0.5	7.1 h	24.67 ± 0.01 (103.2)	25.2 ± 0.4	1.8 ± 0.1
2'-Deoxyguanosine (dG)	712±19	16 mins	22.65 ±0.02 (94.7)	21.7 ± 0.5	-3.2 ± 0.1
2'-Deoxyadenosine (dA)	774 ±12	15 mins	22.59 ±0.01 (94.5)	22.9 ± 0.3	1.1± 0.1
2'-Deoxyzebularine (dZeb)	835± 14	13.8 mins	22.55 ± 0.01 (94.3)	24.1 ± 0.9	5.0 ± 0.1
2,6-Diaminopurine-2'-deoxyriboside (dDap)	1360 ± 16	8.4 mins	22.25 ±0.01 (93.1)	22.9 ± 0.2	2.1± 0.1
2'-Deoxyinosine (dIno)	1400 ±36	8.5 mins	22.25 ±0.02 (93.1)	24.4 ± 0.5	7.0 ± 0.2

Table 3.2: cont.

2'-Deoxyisoguanosine (disoG)	3470 ±70	3.32 mins	21.67 ±0.01 (90.7)	20.7 ± 0.1	-2.9 ± 0.1
2'-Deoxyxanthosine (dXan)	18780 ±111	37 s	20.63 ±0.01 (86.3)	20.9 ± 0.3	1.3 ±0.1

The difference in the Gibbs free energy of activation, $\Delta\Delta G^{\ddagger}$, for acid catalyzed *N*-glycosyl cleavage at 37^{o}C between a ribose and deoxyribose nucleoside for each of the bases studied, along with the corresponding rate enhancements ($k_{N\text{-deoxyribose}}/k_{N\text{-ribose}}$) are highlighted in Table 3.3. We include both kcal/mol and kJ/mol for $\Delta G^{\ddagger}$ (Table 3.2) and $\Delta\Delta G^{\ddagger}$ values (Table 3.3), but the latter unit is a more appropriate metric for comparing smaller changes in $\Delta\Delta G^{\ddagger}$ values.

Table 3.3: Rate enhancements ($k_{N\text{-deoxyribose}} / k_{N\text{-ribose}}$) and change in Gibbs free energy of activation ($\Delta\Delta G^{\ddagger}$) corresponding to differences in reaction rates between ribose and deoxyribose *N*-glycosyl hydrolysis at pH 1 and 37°C.

Nucleobase	$k_{N\text{-deoxyribose}}/k_{N\text{-ribose}}$	$\Delta\Delta G^{\ddagger}$ kcal/mol (kJ/mol)
2,6-Diaminopurine (Dap→dDap)	1428	4.50 (18.8)
Adenine (A→dA)	790	4.12 (17.2)
Isoguanine (IsoG→dIsoG)	754	4.09 (17.1)
Hypoxanthine (Ino→dIno)	737	4.04 (16.8)
Xanthine (Xan→dXan)	731	4.04 (16.9)
Guanine (G→dG)	592	4.01 (16.6)
2-Pyrimidinone (Zeb→ dZeb)	174	3.18 (13.3)

The enthalpic activation parameter ($\Delta H^{\ddagger}$) is the dominant contributor to the Gibbs free energy of activation but the hierarchy of the determined $\Delta G^{\ddagger}$ values (Table 3.2), and thus the overall ranking of deglycosylation rates appear to be a result of the contribution from the entropic activation parameter ($\Delta S^{\ddagger}$). A list of $\Delta S^{\ddagger}$ values along with $\Delta H^{\ddagger}$ and $\Delta G^{\ddagger}$ arranged by related nucleosides is shown in Table 3.4.

Table 3.4: Entropic comparison of nucleosides, and their $\Delta H^{\ddagger}$ and $\Delta G^{\ddagger}$ values for *N*-glycosyl hydrolysis at pH 1 and 37^{O}C.

Ribonucleoside/ Deoxynucleoside	$\Delta S^{\ddagger}$ cal/mol•K	$\Delta H^{\ddagger}$ kcal/mol	$\Delta G^{\ddagger}$ kcal/mol
Dap	5.4 ± 0.2	28.4± 0.8	26.75± 0.02
dDap	2.1 ± 0.1	22.9± 0.2	22.25 ± 0.01
A	-1.8 ± 0.1	26.1± 0.9	26.71 ±0.03
dA	1.1 ± 0.1	22.9± 0.3	22.59 ± 0.01
G	-5.3 ± 0.2	24.9± 0.6	26.60 ±0.02
dG	-3.2 ± 0.1	21.7± 0.5	22.59 ± 0.01
Ino	2.6 ± 0.1	27.1± 0.2	26.29 ±0.01
dIno	7.0 ± 0.2	24.4± 0.5	22.25 ± 0.02
IsoG	-2.9 ± 0.1	24.8± 0.2	25.76 ±0.01
disoG	-2.9 ± 0.1	20.7± 0.1	21.67 ± 0.01
Zeb	6.8 ± 0.1	27.8± 0.3	25.73 ±0.01
dZeb	5.0 ± 0.1	24.1± 0.9	22.55 ± 0.01
Xan	1.8 ± 0.1	25.2± 0.4	24.67 ± 0.01
dXan	1.3 ± 0.1	20.9 ± 0.3	20.63 ± 0.01

Figure 3.3 graphically shows the results of deglycosylation kinetics of all the deoxynucleosides included in this study, determined at 50^{o}C and pH ranging from 1 to 3. A tabular form of the same data obtained is given in Table 3.5. This specific temperature was selected as a compromise, to facilitate kinetic measurements as the reactions slow down with increasing pH, while at the same time represent conditions which closely mirror the relative reactivity at 37^{o}C

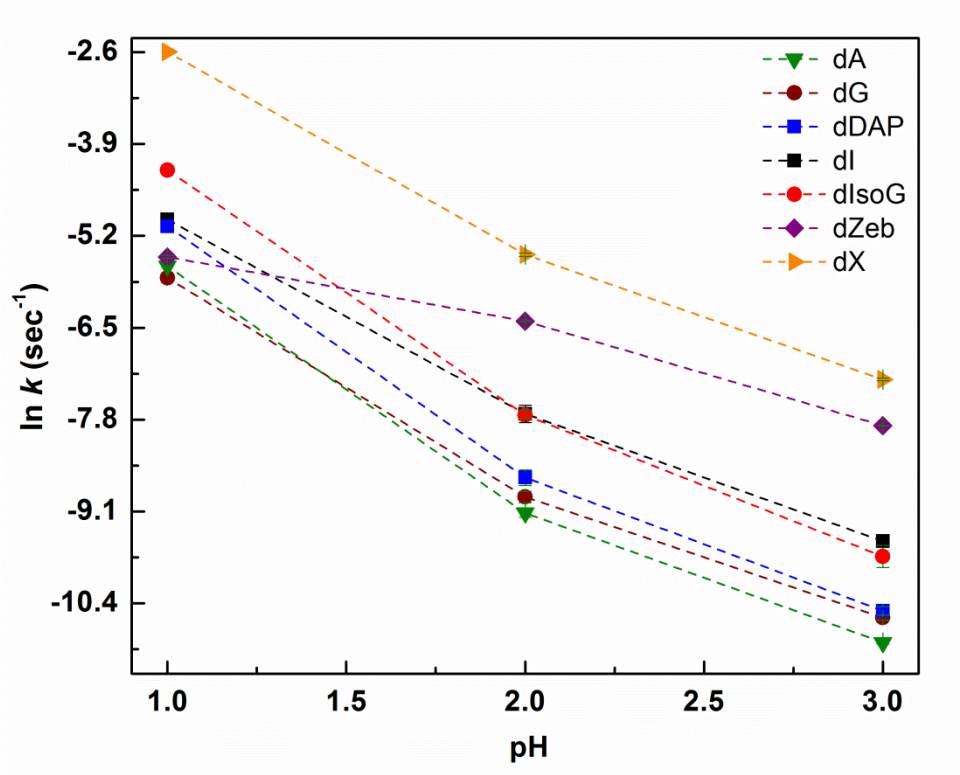


Figure 3.3: Comparative deglycosylation rates at 50°C of 2'-deoxynucleosides as a function of pH from 1 – 3. The average values for all deoxynucleosides at pH 2 and 3 were determined from direct measurements (in triplicate). At pH 1, only the value of dZeb was determined directly at 50°C (in triplicate), the other values are determined from extrapolation of the Arrhenius plots (Figure 3.2).

Table 3.5: Rate constants and $t_{1/2}$ values of deglycosylation reactions for deoxyribonucleosides at 50^{o}C

	pH 1		pH 2		pH 3	
Nucleoside	$10^6 k$, s^{-1}	$t_{1/2}$	$10^6 k$, s^{-1}	$t_{1/2}$	$10^6 k$, s^{-1}	$t_{1/2}$
dA	3600 ±54	3.2 m	109 ± 6	1.8 h	17.4 ±0.3	11 h
dG	3040 ±81	3.7m	137 ±12	1.4 h	24.8 ±0.9	7.7 h
dDap	6330 ±77	1.8 m	181 ±19	1.1h	27 ±1	7.0 h
dIno	6950 ±185	1.6 m	447 ±53	25 m	73 ±2	2.6 h
disoG	14020 ±70	0.8 m	436 ±25	26 m	63 ±2	3 h
dZeb	4100 ±68	2.8 m	1650 ±25	7 m	377 ±2	31 m
dXan	74690 ±444	9.3 s	4230 ±97	2.7 m	720 ±12	16 m

Discussion

The well known stability of riboglycosidic bonds makes it practically impossible, in terms of time scales, to study deglycosylation kinetics at neutral pH and temps < 100C. Depending on the nature of the heterocycle, choosing to carry out these ribonucleoside reactions at neutral pH and necessarily higher temperatures (usually > 120C) could complicate the kinetics of monitoring *N*-glycosyl cleavage by accessing additional reaction pathways that lead to heterocyclic degradation (28-30). Even with our method, the ribonucleoside kinetics still had to be conducted at higher temperatures (up to 85^{o}C) to extrapolate rate constants down to 37^{o}C for direct comparison to the deoxynucleosides. In terms of biological relevance, depurination events in RNA and DNA are assumed to take place much faster than depyrimidination because of the greater susceptibility of the native purines towards acid catalysis. This provided the idea to study the alternative/modified

deoxynucleosides across the low to mid pH range. The existence of a rich history of nucleoside hydrolysis data (28, 31-35) was important to this work as it provided initial clues and inspired us to begin this systematic investigation.

It is evident both from inspection of Table 3.2 or Figure 3.2, the ribose moiety imparts significant stabilization to the *N*-glycosyl bonds for heterocycles compared to the deoxyribose series. Even for xanthosine which displayed the fastest deglycosylation rate (k =2.7 x10^{-5} s^{-1}) of the ribonucleoside series, was still markedly slower in its reaction rate in comparison to the most stable *N*-glycosyl bonds of the deoxyribonucleosides. Deglycosylation of xanthosine proceeds a rate that is only 3–4 % to that of deoxyguanosine (k = 7.1 x10^{-4} s^{-1}) or deoxyadenosine (k = 7.4 x10^{-4} s^{-1}) at pH 1, 37^{o}C.

The isomeric structures of isoguanine and guanine have distinctions between their $\Delta\Delta G^{\ddagger}$ values. Isoguanine, with a $\Delta\Delta G^{\ddagger}$ = 17.1 kj/mol and rate enhancement of 754 fold, had values larger than those of guanine, $\Delta\Delta G^{\ddagger}$ =16.6 kj/mol and 592 fold. However, with the comparison of parent bases (A, G and isoG) and their deamination products, small differences in $\Delta\Delta G^{\ddagger}$ values were mostly observed. Adenine with $\Delta\Delta G^{\ddagger}$ = 17.2 kj/mol, and a rate enhancement of 790 fold, was only a bit larger compared to hypoxanthine, $\Delta\Delta G^{\ddagger}$ = 16.8 kj/mol and 737 fold. Xanthine with a $\Delta\Delta G^{\ddagger}$ = 16.9 kj/mol and rate-enhancement of 731 fold was flanked between either of its parent isomers, isoguanine and guanine. The purine heterocycle, 2,6-diaminopurine, was markedly distinct. It displayed the biggest difference in the change in Gibbs free energy of activation, $\Delta\Delta G^{\ddagger}$ = 18.8 kj/mol and corresponding rate enhancement of 1428 fold. This was substantially greater than any of its possible deamination products such as guanine, isoguanine, or xanthine. The pyrimidine analog to cytosine, 2-pyrimidinone,

exhibited the smallest value, $\Delta\Delta G^{\ddagger}$ = 13.3 kJ/mol and a corresponding rate enhancement of 175 fold.

The enthalpic activation values determined for A (26.1 kcal/mol) and dA (22.9 kcal/mol) were lower than either of the $\Delta H^{\ddagger}$ values determined for the deamination products, Ino(27.1 kcal/mol) and dIno(24.4 kcal/mol) respectively. The overall lower values of $\Delta G^{\ddagger}$ observed for Ino/dIno though can be attributed to the more favorable entropic contribution exhibited by Ino (2.6 cal/mol•K) and dIno (7.0 cal/mol•K) in comparison to the A (-1.8 cal/mol•K) and dA (1.1cal/mol•K). The enthalpic values determined for G (24.9 kcal/mol) and dG (21.7 kcal/mol) were very close to the related deamination products of Xan (25.2 kcal/mol) and dXan (20.9 kcal/mol). The differences in $\Delta G^{\ddagger}$ values between G/dG and Xan/dXan is a result of the negative entropic values G (-5.3 cal/mol•K) and dG (-3.2 cal/mol•K) compared to the positive entropic contribution determined for Xan (1.8 cal/mol•K) and dXan (1.3 cal/mol•K). The isomer isoG was found to have a $\Delta H^{\ddagger}$ value (24.8 kcal/mol) close to G but a less negative entropic contribution $\Delta S^{\ddagger}$= -2.9 cal/mol•K by comparison, resulting in an overall lower $\Delta G^{\ddagger}$ value. Deoxyisoguanosine did not display a change in the $\Delta S^{\ddagger}$ value in comparison to its ribonucleoside version, but in comparison to dG, the $\Delta H^{\ddagger}$ value of disoG(20.7 kcal/mol) was smaller. The ribonucleoside Dap, was found to exhibit the largest $\Delta H^{\ddagger}$ value (28.4 kcal/mol) of the entire series, and a relatively large positive $\Delta S^{\ddagger}$, 5.4 cal/mol•K. The deoxynucleoside version, dDap had modest $\Delta H^{\ddagger}$ (22.9 kcal/mol) and $\Delta S^{\ddagger}$ values (2.1 cal/mol•K) in comparison to the other deoxynucleosides. Both zebularine and deoxyzebularine exhibited large enthalpic activation parameters Zeb = 27.8 kcal/mol and dZeb=24.1 kcal/mol in comparison to the other ribo and deoxyribonucleosides comparatively. But in the case of Zeb, its lower $\Delta G^{\ddagger}$ parameter can be attributed to its larger positive entropic contribution ($\Delta S^{\ddagger}$

= 6.8 cal/mol•K), which was the highest of the ribonucleosides. For dZeb, the $\Delta S^{\ddagger}$ value of 5.0 cal/mol•K, lowered the overall $\Delta G^{\ddagger}$ parameter to match those of dA and dG.

Across the pH ranges shown in Figure 3.3, dA and dG exhibited the slowest reaction rates (Table 3.5) and the other deoxynucleosides are compared to them. Deoxyinosine was observed to increase its 2–3 fold reaction rate over dG and dA at pH 1 to over 3 – 4 times the native nucleosides at pH 2 and 3. Deoxyisoguanosine began with a 3.9 – 4.6x faster deglycosylation rate over dA and dG, respectively at pH 1, but dropped to about 2.5x the rate of dG and 3.7x the rate of dA at pH 3. Deoxyxanthosine, maintained the largest reaction rates across this pH range. At pH 1 it exhibited a 20 – 24 fold faster reaction rate over dA and dG, and the difference becomes larger at pH 3 where it was nearly 30 fold faster than dG and even 41 times the rate of dA. 2,6-diamonopurine-deoxyriboside displayed reaction rates that approached those of the native nucleosides as the pH was increased to 2 and 3. Lastly, the pyrimidine nucleoside, dZeb, exhibited the smallest change in reaction rate as the pH was increased. Its rate was close to the native purine nucleosides at pH 1 but eventually became 15–20 times the deglycosylation rates at pH 3.

The transition from ribo-glycosidic to deoxy-ribo glycosidic bonds

RNA in contemporary biology across all domains is known to utilize many types of modified and even exotic bases (2, 47) We have previously hypothesized that the higher stability of glycosidic bonds associated in RNA due to the ribose moiety could have been an advantageous feature for the base diversity in an RNA world(20). The transition to DNA (19) and the consequence of global weakening to *N*-glycosidic bonds (48) may have facilitated a further refinement of bases employed in

DNA. The exerted pressure was the need to eliminate the most problematic *N*-glycosyl bonds that gave rise to the increased presence of abasic sites. A subtle idea we are exploring is that relative *N*-glycosyl stabilities of heterocycles/bases may differ in moving from a ribo-glycosidic bond to a deoxyribose one. The variation in rate enhancements resulting from the $\Delta\Delta G^{\ddagger}$ values is our first detailed/systematic look at this phenomenon. While some similarities in $\Delta\Delta G^{\ddagger}$ were noticed in the deamination products(Ino, Xan) to the bases A, G, isoG the comparison instead of the native bases to surrogates, such as the A→dA to Dap→dDap rate enhancement or G→dG to Ino→dIno (Table 3.3) highlights much larger differences. It is unknown how a C→dC rate enhancement compares to our determined value for the Zeb→dZeb transition, given the historic difficulty in obtaining kinetics data of cytidine and deoxycytidine under the same conditions (34, 49-50). This is largely attributed to the comparatively fast deamination reaction of cytidine leading to uridine in temps< 100^{o}C, which inherently obstructs accurate experimental measurements of cytidine deglycosylation rates. It is hoped that this will be an area aided by future computational studies(51) Mostly importantly, rate enhancements of this comparison are expected to differ depending on pH and temperature and remain unique to the identity of the heterocycle whether purines or pyrimidines. In a study where we directly measured the deglycosylation rates of xanthosine and deoxyxanthosine at pH 2, 37^{o}C we observed a rate enhancement of 1680, which is a significant change from the observed rate enhancement of 731 measured at pH 1 and 37^{o}C (Table 3.2). The base 5-hydroxyuracil, a tRNA modification, but also a DNA oxidative-deamination lesion of cytosine was observed under acidic conditions (1 M HCl) to exhibit a rate enhancement of around 370 in comparing deglycosylation rate of 5-hydroxyuridine and 5-hydroxy-2'-deoxyuridine at 60^{o}C(52).Wolfenden has recently shown that under

neutral pH and extrapolation to 25C, the weakening of the glycosidic bond in moving from adenosine to deoxyadenosine exhibited only a 30 fold difference in reaction rate, (35) another significant change from the 790 fold we observed under different conditions(Table 3). Comparison to what may be known at the same temperature under similar pH for a U→dU transition is an interesting relation. Under neutral pH, limiting the possibilities of acid catalysis, dU is actually known to exhibit faster deglycosylation rates than deoxyadenosine (28, 53). This can be attributed in part to the better leaving group ability of the uracil monoanion (pK_a 9.5) over the monoanion of adenine (pK_a 9.8). By contrast, the ribo-glycosidic bond of uridine appears to be much more stable in comparison to adenosine. Detailed kinetics data is lacking in this area, but a recent report seems to implicate the greater stability of uridine over adenosine and other ribonucleosides studied (23). Furthermore, Miller and Orgel published a preliminary study(54) that measured the deglycosylation rates of U under neutral pH and extrapolated down to 25^{o}C and compared to the Shapiro and Kang study (32) of dU under similar conditions. The determined half-lives of hydrolysis for U ($t_{1/2}$ = 1.2x 10^5 yrs) compared to dU ($t_{1/2}$ = 365 years) gives a very approximate rate enhancement of around 300. This appears to be consistent with the observation that uridine while exhibiting a stronger ribo-glycosidic bond undergoes a greater reduction in stability when attached to the deoxyribose moiety in comparison to the A → dA transition. Still, inconsistencies between different experimental methods need to be eliminated by undertaking a comprehensive study measuring deglycosylation kinetics of U/dU and A/dA under the same conditions. In general, the observed variations in rate enhancements underscore how stability differences between a ribo-glycosidic and 2'-deoxyribo-glycosidic bond is not uniform, but instead dependent on the specific nature of the attached heterocycle. From a perspective of biochemical

evolution, the variation of rate enhancements could mean that the transition from genomic RNA to DNA created yet another shuffling of relative glycosidic bond strength and in the context of an early DNA world, a selection for the strongest bonds became necessary.

The native purines and pyrimidines, particularly in DNA, may exhibit optimum N-glycosyl stability in the presence of acid catalysis

The ease of hydrolytic rupture of native purine *N*-glycosyl bonds, in comparison to the native pyrimidines, under physiological conditions in DNA is thought to occur as a result to the greater susceptibility of acid or metal ion catalyzed deglycosylation exhibited by the purines (55). Evidence to support this comes from the known relative pK_a values, the sites of protonation and the observed pH dependence on deglycosylation rates that are more accessible for the purine (33, 53). Acid catalysis lowers the enthalpic barrier ($\Delta H^{\ddagger}$) of deglycosylation because it increases the leaving group ability of the purine from the sugar moiety and hence destabilizes the *N*-glycosyl bond. Thus it might be expected that a more basic nucleobase (higher pK_a value) in related purines, would exhibit a greater reduction in the $\Delta H^{\ddagger}$ values and a faster reaction rate. However an interesting observation from this study is that differences in deglycosylation rates under acidic conditions did not appear to be closely correlated to the pK_a values or even to the relative magnitude of the enthalpic barrier. Tables 3.6 and 3.7 list the known pK_a values (56) of relevance under these conditions for the heterocycles used in this study, along with the rate constants, *k*, and enthalpy activation parameters, for the ribonucleoside and deoxyribonucleoside deglycosylation reactions, respectively. While the pK_a values for each of the heterocycles are known to be lower when connected to the

ribose/deoxyribose moiety, not completely accurate nucleoside pK_a values seem to be available in the literature. Thus we use the heterocycle values only to serve as a systematic guide for the relative basicity of the corresponding nucleosides. It is apparent upon inspection of Table 3.6 and 3.7 that rate constants don't reflect the basicity of the listed heterocycles. Both in the ribo and deoxyribonucleosides series, Xan and dXan, contain the least basic heterocycles, but are the nucleosides that exhibit the fastest reaction rates. A look to the $\Delta H^{\ddagger}$ values for Xan (Table 3.6) and dXan (Table 3.7), which are among the lowest, provide only a partial explanation for their apparently weak *N*-glycosyl bonds. Comparisons of Xan ($\Delta H^{\ddagger}$ = 25.2 kcal/mol) to isoG ($\Delta H^{\ddagger}$ = 24.8 kcal/mol) and dXan ($\Delta H^{\ddagger}$ = 20.9 kcal/mol) to disoG ($\Delta H^{\ddagger}$ = 20.7 kcal/mol) indicate that even with similar enthalpic values (within error) in each pair, the Xan and dXan nucleosides still exhibit deglycosylation kinetics that are about 6 fold faster. In the ribonucleoside series (Table 3.6), Dap, contains the heterocycle with the highest pK_a, but appeared to exhibit a particularly strong ribo-glycosidic bond, having a $\Delta H^{\ddagger}$ = 28.4 kcal/mol, and a slow reaction rate. Its native counterpart, adenosine, containing a less basic heterocycle than Dap, was determined to exhibit the same reaction rate under these conditions, even though it had a distinctly lower enthalpy of activation, $\Delta H^{\ddagger}$ = 26.4 kcal/mol. However in the deoxyribonucleosides series (Table 3.7), both dDap and dA displayed the same enthalpic values (both at $\Delta H^{\ddagger}$ = 22.9 kcal/mol) but exhibiting different reaction rates, with dDap being 2 fold faster over dA. Other comparisons to mention include isoG/disoG and Ino/dIno to the native nucleosides G/dG. From the literature data isoG and disoG appear to contain heterocycles with a pK_a value higher than G/dG, and while the rates reflect a trend (isoG is about 4x faster in deglycosylation over G), the enthalpic values of the ribonucleoside isomers are determined to be just about the same (24.8 – 24.9

kcal/mol). Yet in the deoxynucleosides series, the values adjust even more, resulting in an enthalpic barrier of disoG to be about a full 1 kcal/mol unit lower than that of dG and exhibits a 5 fold greater reaction rate than dG. Inosine, a genetic surrogate for guanosine, contains a heterocycle that is much less basic in comparison (Table 3.6) and correspondingly found to exhibit a $\Delta H^{\ddagger}$ value which is a full 2 kcal/mol greater than G. However, even with this enthalpic barrier difference, the reaction rate of inosine was not slower, but actually 1.6x faster than guanosine. A larger difference was observed in comparing the deoxyribonucleosides versions. Deoxyinosine had a 2.7 kcal/mol greater enthalpic barrier over dG, but yet still resulted in a deglycosylation rate that was two-fold faster than dG.

It might be concluded that pyrimidines, as a class, are less susceptible than purines to acid catalysis under the same conditions (33, 52, 55), but not all pyrimidines are preordained for such stability. The alternative pyrimidine nucleosides, Zeb and dZeb have previously been shown for their susceptibility towards deglycosylation (57) (25). A comparison of the enthalpic parameters from a previous study on deoxycytidine hydrolysis (34) under similar conditions to the dZeb values determined here, hints at the origin of the extremely fast reaction rates. Even with a higher heterocyclic pK_a value, cytosine = 4.4 compared to 2-pyrimidinone = 2.1, the enthalpic barrier for dC under these conditions is calculated to be 37.2 kcal/mol, which is an staggering 13 kcal/mol greater activation barrier than the value of dZeb ($\Delta H^{\ddagger}$ =24.1 kcal/mol) and still 9 kcal/mol larger that the ribonucleoside, Zeb ($\Delta H^{\ddagger}$ =27.1 kcal/mol). The corresponding deglycosylation rate of dC at pH 1 and 37^{o}C is, k =1.7x10^{-8} s^{-1} which means that deoxycytidine deglycosylates at only 0.002 % the rate of dZeb and 0.4 % the rate of Zeb. Yet even more drastic changes can occur to the relative stability of a pyrimidine *N*-glycosyl bond from a simple isomerization of the

cytosine face. It was previously reported under the same acidic conditions (0.1M HCl) (58)studied here, but at 40^{o}C, that disoC deglycosylates with a $t_{1/2}$ = 3.5 mins. By comparison, dZeb, deglycosylates at 40^{o}C with a $t_{1/2}$= 9.4 mins and the determined values for dC at 40^{o}C, gave a $t_{1/2}$= 260 days. Thus, disoC deglycosylates about 3 fold faster than dZeb and over 100, 000 times that of its native isomer. Even with considering error in differences between experimental method to ours or that reported for disoC, it is quite remarkable how such a simple pyrimidine analog to cytosine or the isomer, disoC, both that can be considered close relatives to the native pyrimidine can suffer a dramatic destabilization in the *N*-glycosyl bonds. It is especially note worthy to point out that these non-native pyrimidines exhibit deglycosylation susceptibilities that match the purine nucleosides.

While relevant pK_a values and enthalpic barriers didn't correlate well with the observed rate constants under acidic conditions, as previously mentioned the overall hierarchy of reaction rates for Table 3.2 was fine tuned by the entropic parameter. One feature that does stand out from these data and those obtained in Figure 3.3, is that the native purines and expectantly the native pyrimidine nucleosides appear to exhibit the highest level of *N*-glycosyl stability that is possible given their particular chemical susceptibilities. This is a feature that may have been more important in the early DNA world and may have implications for base modification and even possibly a selection pressure for the particular bases used in the genetic alphabet.

Table 3.6: Comparison of nucleobase pK_a values with rate constants and enthalpic activation parameters for *N*-glycosyl hydrolysis of ribo-glycosidic bonds at pH 1 and 37^{O}C.

Ribonucleoside	pK_a of nucleobase	Rate, 10^6k, s^{-1}	$\Delta H^{\ddagger}$ kcal/mol
Dap	5.09	0.91	28.4± 0.8
isoG	4.51	4.6	24.8± 0.2
A	4.15	0.98	26.1± 0.9
G	3.3	1.2	24.9± 0.6
Zeb	2.2 – 3.1	4.8	27.8± 0.3
Ino	1.98	1.9	27.1± 0.2
Xan	0 – 0.8	27	25.2± 0.4

Ref for pK_a values: (56)

Table 3.7: Comparison of nucleobase pK_a values with rate constants and enthalpic activation parameters for *N*-glycosyl hydrolysis of 2'-deoxyribo-glycosidic bonds at pH 1 and 37^{O}C.

2'-Deoxynucleoside	pK_a of nucleobase	Rate, 10^6k, s^{-1}	$\Delta H^{\ddagger}$ kcal/mol
dDap	5.09	1300	22.9 ±0.2
disoG	4.51	3470	20.7± 0.1
dA	4.15	774	22.9± 0.3
dG	3.3	710	21.7± 0.5
dZeb	2.2 – 3.1	835	24.1± 0.9
dIno	1.98	1400	24.4± 0.5
dXan	0 – 0.8	19730	20.9 ± 0.3

Ref for pK_a values: (56)

Specific DNA modification found in the biological world may carry a common theme in relation to N-glycosyl stability

Modification of bases in DNA, by comparison to RNA is known to be quite limited (2). When DNA modifications are found in nature, it appears that pyrimidines are the outstanding sites of modification (Table 3.8) (59-60).

Table 3.8: Significant DNA base modifications and native letter substitutions

Nucleobase	Modifications originating from the base	Substitutions of a native letter	Ref.
Cytosine	5-methylcytosine (mC), 5-hydroxymethyl cytosine (hmC) N^4-methylcytosine (m^4C) Uracil,(U) Thymine (T), 5-hydroxymethyluracil (hmU) 5-formylcytosine (fC) 5-carboxylcytosine (caC),	Hmc, mC	(2, 59-61)
Thymine	none	hmU, U	(59-60)
Adenine	N^6-methyladenine(m^6A),	2,6-Diaminopurine (Dap)	(2, 59-60)
Guanine	None yet identified	None yet identified	(59-60)

Cytosine however is the salient example for DNA base modification, and the question as to why this particular heterocycle has remained nature's choice for unique alterations has been elegantly put forward (61). In our view, the selection of cytosine would be a logical choice given its high *N*-glycosidic bond stability under neutral pH in DNA (20, 53). The specific modifications made to cytosine along with

experimentally or predicted *N*-glycosyl stabilities reflect this hypothesis. 5-methylcytosine (mC) and 5-hydroxylmethylcytosine, which are useful signaling species in genes and even found to replace C entirely in some bacteriophage genomes,(60) are expected to exhibit similar *N*-glycosyl stabilities(34, 62-63) (Table 3.8) When modified cytosines need to be removed, in the case of epigenetic processes, further alterations are performed to generate fC (Table 3.8) or caC, both of which have been predicted to destabilize the *N*-glycosyl bond enough for effective removal by glycosylases (63-64). While cytosine may seem to present far reaching opportunities in its capacity for useful DNA modification (61), there may be limits with respect to *N*-glycosyl stability. Removal of the C4 amino group with subsequent reduction, generating 2'-deoxyzebularine, or conducting an isomerization reaction to deoxyisocytidine would transform one of the most stable *N*-glycosidic bonds in DNA into one of the weakest. It appears that life may have found a perfect balance in relative *N*-glycosyl stability with the modification of cytosine in DNA

When a purine is modified in DNA, adenine appears to be the choice used by cells, and almost exclusively on its N6 position (with a few on reports on the C2 position)(47) but the extent of modification has been observed so far to be only 10% compared to the 100% substitution of the pyrimidine modifications that have been identified.(59) However, there does exist one intriguing report on the utility of a modified base completely replacing adenine and that is occurrence of 2,6-diaminopurine in the duplex DNA of the Cyanophage S2-L (36, 65). From the argument of glycosidic bond stability as one angle of this specific and unique occurrence, Dap, was observed to exhibit one of the slowest deglycosylation rates (especially at lower temperatures) over the other modified purines (Table 3.2). Notably, as the pH was increased in the deoxynucleosides, the stability of dDap

appeared to approach that of the native purines (Figure 3.3). Given the higher pK_a value of 10.77 for Dap in comparison to adenine, pK_a = 9.8 (56) as a qualitative assessment of its leaving group ability in the form of a monoanion, it is predicted that Dap will likely exhibit even slower reaction rates than deoxyadenosine at neutral pH. The absence of deoxyinosine employed in DNA as a guanine surrogate is another curious observation. In the ribo-version, inosine occupies a pervasive role in extant biology, having occurrences in tRNA, rRNA and even mRNA. It is also the precursor purine nucleotide generated first, known as inosine monophosphate that is biosynthesized and modified into adenosine and guanosine monophosphate (43). In eukaryotic cells, the post-transcriptional deamination of adenosine to produce inosine as a reliable guanosine analog is utilized in the phenomenon known as RNA editing (40, 66). By comparison to native purine deoxynucleosides, the *N*-glycosyl stability of deoxyinosine is modestly weaker (up to 3 fold faster reaction rate over G as pH is increased (Figure 3.3, Table 3.5) and at neutral pH (28). In RNA, even much larger differences in *N*-glycosyl stability attributed to more exotic base modification may matter little (20), but is this stability difference in DNA, significant? Cells have evolved repair pathways to maintain the integrity of its genetic material (67) which provides convincing evidence that the problem of spontaneous deglycosylation (55) was a significant threat. Substituting an already susceptible residue (deoxyguanosine) for one that exhibits an even weaker *N*-glycosyl bond (deoxyinosine) might have been previously carried out in the early DNA world. But a limit to how much a cell could endure with respect to the energetic cost of its resources may have prevented this guanine surrogate from ever really taking hold in DNA.

Selection Pressures for the Genetic Alphabet

The occurrence of modified bases that are utilized in RNA and DNA by cells, especially as they are employed as genetic surrogates, naturally invokes the question "how did nature select upon the specific bases of the genetic alphabet?" The heterocycles employed in this study are known to the field of prebiotic chemistry (Table 3.1) and many of them have been the subjects to questions surrounding the origin, evolution and size of the genetic alphabet (68-70). Different types of alternative bases and base pairs have been considered and even shown to be enzymatically incorporated into polymeric RNA and DNA (13, 71-72), but in the context of this study we consider more modest/conservative alternatives. We assume that a strong and weak base pairing relationship (two hydrogen bonds and three hydrogen bonds) in addition to a purine: pyrimidine system was an essential attribute for the origin and early biochemical evolution of nucleic acids (73-74).

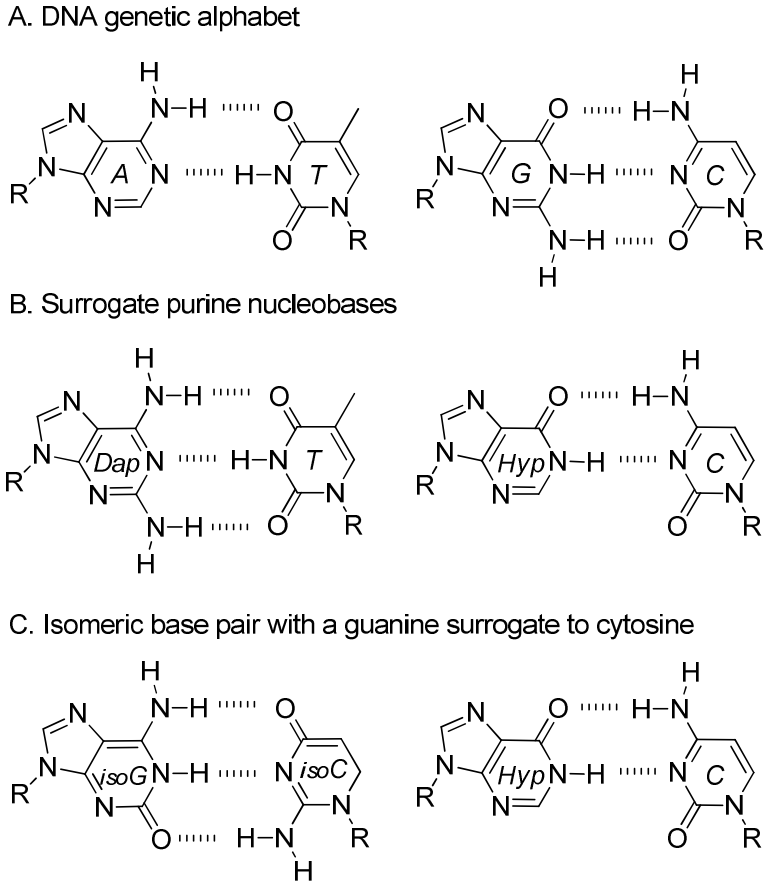


Figure 3.4: Alternative bases and base pairs in a four letter alphabet that maintain a two and three hydrogen bonding pattern. **A.** Native bases. **B**. Switching Adenine with Dap and Guanine with Hyp, maintains the same letter capabilities **C.** Employing hypoxanthine as a guanine surrogate along with the orthogonal base pair of isoG and isoC.

Two such examples of alternative alphabets to the native genetic alphabet are shown in Figure 3.4. Replacing A with Dap, and G with Hyp, maintains overall 2:3 hydrogen bonding pattern, but now in an inverse Watson-Crick fashion (Figure 3.4B). Similar substitutions such as these either employing one or two surrogates, have been utilized in nucleic acids chemistry (75-78) So why not this alphabet? Previous studies have indicated that the global base pairing strength of each pair in this surrogate alphabet doesn't appear to obtain the same level of stability in a DNA duplex as the native ones(71, 79-80). While fully utilizing three hydrogen bonds, the

Dap:T interaction has been observed to exhibit, on average, a base pair strength that is in between the native strong G:C and the weak, A:T. The Hyp:C base pair with its two hydrogen bonds is also reported to be slightly less than the native A:T system (71). The origin to the differences in base pair strength for these and other more exotic base pairing systems in relation to their heterocyclic pK_a values has been recently discussed as yet another pressure during the course of nucleic acids evolution and base selection (81). The consideration of this alphabet though is intriguing because as we have shown here, dDap may indeed contribute a stable glycosidic bond (Figure 3.3) as the pH approaches neutrality. The only noticeable weaker link would originate from Hyp (deoxyinosine). The natural occurrence of Dap in bacteriophage DNA exists in the presence of a three: three hydrogen bond system (Dap:T, and G:C), which makes it hard to assess if nature finds the utility of a stronger base pairing duplex or the integrity of its *N*-glycosyl bonds more important, or both. In either case, it may also be that the rare occurrence of Dap in the biological world may simply be attributed to its lower prebiotic abundance or higher synthetic complexity that isn't compensated by a significant advantage over the utility of adenine.

If base pair strength in duplex stability was a significant pressure, then consideration of an isoG:isoC pair (72, 82) along with Hyp:C might be considered a more plausible alternative alphabet (Figure 3.4C). The three hydrogen bond pair of isoG:isoC has been reported to exhibit the same base pair stability as a G:C pair (83), and with a Hyp:C pair that is very close to an A:T pair, this alphabet would seem to closely approach the duplex stability of the native system. Furthermore, this hypothetical version might even offer an advantage over the native alphabet. Without the A:T (or A:U) base pair, the alphabet in Figure 3.4C, would benefit from an almost complete (though not entire) elimination of genetic ambiguity upon spontaneous

deamination events. In the native alphabet, the problem of cytosine/5-methylcytosine deamination leads to the uracil/thymine lesion and can be highly mutagenic (a GC→TA transition mutation) since the lesion itself is an alphabetic letter. Nature has evolved solutions around it but it is a bit more complicated and not without its problems (84). For the hypothetical alphabet of Figure 3.4C, deamination events of C and isoC leading both to uracil as a non-native letter, would, presumably, be an immediately recognizable lesion and thus efficiently removed. One reason for why others have hypothesized against the plausibility of a naturally occurring isoG:isoC base pair is because of the fidelity problems of isoG, attributed to its significant (about 10% in hydrophobic regions) tautomeric enol form that codes quite reliably for thymine (14, 83, 85), However in the case of this alphabet (Figure 3.4C) the problem would become almost irrelevant, since the absence of adenine and thymine as genetic letters would inherently prevent the generation of isoG:isoC →A:T transition mutations. Not every problem would be solved since there would still remain the likelihood that a uracil interloper not excised by repair could promote a Hyp:C→isoG:isoC transition mutation via a U:isoG pairing from the enol form of isoG. However, the most convincing arguments against this alphabet would seem to come from the problems of deamination and deglycosylation rates in comparison to the native system. Cytosine is clearly the most problematic base with regards to spontaneous deamination in the native alphabet and adding two more bases (isoG and isoC) with rates that are comparable or even greater than cytosine would seem to be especially challenging for the repair resources in the hypothetical alphabet (14, 30). Probably even more significant is the consideration that each of the bases with the exclusion of cytosine, (isoG, isoC, and Hyp), exhibits a weaker *N*-glycosyl bond in comparison to the native bases (Figure 3.30) (58). Additionally, the resulting

deamination product from isoG, leading to xanthine would further destabilize the glycosidic bond and likely increase depurination under mild acid (Figure 3.3) or possibly metal ion catalysis (28, 86).

Conclusion

The specific differences in hydrolytic susceptibility between polymeric RNA and DNA and their components have been known for a long time. In the context of chemical and biochemical evolution, much has been focused on the emergence of pre-RNA polymers and the transition to the assumed RNA world (87). Much less focus has been put on the transition between the RNA and early DNA world (19, 88). Only fairly recently have reports been made addressing the evolutionary implications of hydrolytic susceptibilities differences between RNA and DNA under the same experimental conditions (89-90). However, not much is known about the role and significance of non-canonical bases in the RNA and early DNA worlds, especially given the knowledge that these two polymers employ base modifications in contemporary biochemistry (2, 47). Comparative hydrolytic stabilities of alternative bases are available in the literature (18, 29-30, 91-92) and greatly provide a perspective on the relative hydrolytic fitness of the native bases, but their presence in components of RNA and DNA needs to be given a detailed investigation. As we have shown here, even simple modifications can have a large impact on reaction kinetics of *N*-glycosyl rupture. It seems plausible to hypothesize that the stability of the *N*-glycosyl linkage in DNA would have presented a potent pressure in the early DNA world (20). While the bases, particularly the purines, in the native genetic alphabet aren't perfect (48), they might still have presented early life with the optimum level of *N*-glycosyl stability for safeguarding their genetic material.

Material and Methods

Nucleosides and nucleobases were purchased from commercial sources and used without further purification. The native and many modified nucleosides and bases were obtained from Sigma-Aldrich, Acros, or MP Biochemicals. 2'-deoxyisoguanosine, was purchased from Berry & Associates (Dexter, Michigan). 2'-deoxyxanthosine was purchased from Carbosynth LLC (San Diego, CA) Isoguanosine, isoguanine, 2,6-diaminopurine-2'-deoxyriboside, and zebularine were purchased from Toronto Research Chemicals (North York, Ontario). 2'-deoxyzebularine was purchased from Trilink Biotechnologies (San Diego, California). For experiments conducted under acidic conditions (pH 1 – 3) and temperature ranges (10 – 85^{o}C), a Shimadzu UV-2450, UV-Vis Spectrophotometer was used to monitor kinetics in real-time or after an alkaline quench (as described below) and monitored at a single-wavelength.

2'-Deoxyribonucleosides:

Kinetics reactions for deoxynucleosides at pH 1 (0.1M HCl), were initiated by using 10mL volumetric flasks containing 5mL of 0.2M HCl, 1 mL of stock nucleoside solution (10^{-4} M) and diluted with deionized water. Sample analysis was conducted by the alkaline quench method, adapted from Garrett and Mehta (93). All solutions, vials and flasks were equilibrated at the corresponding reaction temperature for 30 mins and then rapidly mixed in the volumetric flasks followed by transfer to 5mL reaction vials and placed in a temperature controlled (±0.1 ^{o}C) water/ethylene glycol bath. At designated time points, 250 µL aliquots of the reaction mixture were removed and quenched in a UV-Vis cuvette with 250 µL of cold 1M NaOH solution. Full absorbance spectra were recorded and rates were determined by monitoring a

decrease in absorbance at the largest change of λ between the nucleoside and base (Appendix A) Kinetic experiments were conducted in duplicate or triplicate. Comparison of the alkaline spectra of nucleosides and the corresponding nucleobases (obtained from commercial sources) were conducted to verify that the reaction spectra were indeed monitoring deglycosylation kinetics.

For three nucleosides (deoxyisoguanosine, deoxyzebularine and deoxyxanthosine), kinetics were monitored directly by the UV-Vis spectrophotometer under the reactions conditions. The same set up was employed (pre-equilibration and mixing) but once mixed, a portion of the solution was transferred to a pre-equilibrated cuvette and completely filled with the reaction mixture and closed with a stopper. This was quickly placed in a cell holder that was specially designed to be controlled by a similar temperature controlled ethylene-glycol/water bath (±0.1 ^{o}C) and kinetics were obtained by monitoring decrease in absorbance at one wavelength.

For reactions conducted at pH 2 and 3, (50 ^{o}C), buffers of 0.1M sodium phosphate (pH 2-2.1), 3.-3.1 were used as dilution media and the experiments were conducted using the alkaline quench method and reaction progress was monitored spectrophotometrically.

Ribonucleosides:

Given their inherent stability even at pH 1, the ribonucleoside reaction mixtures were prepared by using 10mL volumetric flasks containing 5mL of 0.2M HCl, 0.5 - 1 mL of stock nucleoside solution (10^{-4} M) and diluted with deionized water at room temperature. Then aliquots of 600 µL were used to completely fill (0.6mL reaction vial (borosilicate glass, ChemGlass # CV-1510-0740) and crimp sealed (ChemGlass # CV-3450-0008). Vials (about 10-13) were placed in a dry bath (Grant

Instruments # QBH2) and modified with sand and oil to ensure equal and consistent heating at the desired temperature. In all cases, for the given temperatures chosen the time it took for vials to reach reaction temperature (<1 min) had no noticeable impact on the reaction rate. At designated time points, a reaction vial was removed from the heat bath and submerged in ice-water. Upon cooling, 250 µL aliquots of the reaction mixture were removed and quenched in a UV-Vis cuvette with 250 µL of 1M NaOH solution. Full absorbance spectra were recorded and rates were determined by monitoring a decrease in absorbance at the largest change of λ between the nucleoside and base (Appendix A). Kinetic experiments were conducted in duplicate and some cases triplicate.

Acknowledgements

Chapter 3 is adapted from a manuscript in preparation of: Rios, A.C., Yu, H.T., Tor, Y., Hydrolytic stability of *N*-glycosyl bonds in modified, alternative and damaged nucleosides: Trends and implications on the refinement of the genetic alphabet. The dissertation author is the main author and researcher for this work.

References

1. Rozenski J, Crain PF, & McCloskey JA (1999) The RNA Modification Database: 1999 update. *Nucleic Acids Res.* 27(1):196-197.

2. Carell T*, et al.* (2012) Structure and Function of Noncanonical Nucleobases. *Angew. Chem. Int. Ed. Engl.* 51(29):7110-7131.

3. Levy M & Miller SL (1999) The Prebiotic Synthesis of Modified Purines and Their Potential Role in the RNA World. *J. Mol. Evol.* 48(6):631-637.

4. Robertson M & Miller S (1995) Prebiotic synthesis of 5-substituted uracils: a bridge between the RNA world and the DNA-protein world. *Science* 268(5211):702-705.

5. Cermakian N & Cedergren R (1998) Chapter 29. Modified Nucleosides Always Were: an Evolutionary Model. *Modification and Editing of RNA*, eds Grosjean H & Benne R (ASM Press, Washington DC), pp 535 - 541.

6. Joyce GF (2002) The antiquity of RNA-based evolution. *Nature* 418(6894):214-221.

7. Benner SA, Burgstaller P, Battersby TR, & Jurczyk S (1999) Chapter 6. Did the RNA World Exploit an Expanded Genetic Alphabet? *The RNA World*, eds Raymond F Gesteland, Cech TR, & Atkins JF (Cold Spring Harbor Laboratory Press, Cold Spring Harbor), Vol 2, pp 163-181.

8. Menor-Salván C, Ruiz-Bermejo DM, Guzmán MI, Osuna-Esteban S, & Veintemillas-Verdaguer S (2009) Synthesis of Pyrimidines and Triazines in Ice: Implications for the Prebiotic Chemistry of Nucleobases. *Chemistry – A European Journal* 15(17):4411-4418.

9. Barks HL*, et al.* (2010) Guanine, Adenine, and Hypoxanthine Production in UV-Irradiated Formamide Solutions: Relaxation of the Requirements for Prebiotic Purine Nucleobase Formation. *ChemBioChem* 11(9):1240-1243.

10. Nuevo M, Milam SN, & Sandford SA (2012) Nucleobases and Prebiotic Molecules in Organic Residues Produced from the Ultraviolet Photo-Irradiation of Pyrimidine in NH3 and H2O+NH3 Ices. *Astrobiology* 12(4):295-314.

11. Callahan MP*, et al.* (2011) Carbonaceous meteorites contain a wide range of extraterrestrial nucleobases. *Proc. Natl. Acad. Sci. U.S.A.* 108(34):13995-13998.

12. Botta O & Bada JL (2002) Extraterrestrial Organic Compounds in Meteorites. *Surv. Geophys.* 23(5):411-467*.*

13. Piccirilli JA, Benner SA, Krauch T, & Moroney SE (1990) Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature* 343(6253):33-37.

14. Switzer CY, Moroney SE, & Benner SA (1993) Enzymic recognition of the base pair between isocytidine and isoguanosine. *Biochemistry* 32(39):10489-10496.

15. Nguyen KV & Burrows CJ (2011) A Prebiotic Role for 8-Oxoguanosine as a Flavin Mimic in Pyrimidine Dimer Photorepair. *J. Am. Chem. Soc.* 133(37):14586-14589.

16. Engelhart AE & Hud NV (2010) Primitive Genetic Polymers. *Cold Spring Harbor Perspect. Biol.* 2(12):a002196.

17. Joyce GF (1989) RNA evolution and the origins of life. *Nature* 338(6212):217-224.

18. Robertson MP, Levy M, & Miller SL (1996) Prebiotic synthesis of diaminopyrimidine and thiocytosine. *J. Mol. Evol.* 43(6):543-550.

19. Lazcano A, Guerrero R, Margulis L, & Oró J (1988) The evolutionary transition from RNA to DNA in early cells. *J. Mol. Evol.* 27(4):283-290.

20. Rios AC & Tor Y (2012) Refining the Genetic Alphabet: A Late-Period Selection Pressure? *Astrobiology* 12(9):884-891.

21. Orgel LE (2004) Prebiotic Chemistry and the Origin of the RNA World. *Crit. Rev. Biochem. Mol. Biol.* 39(2):99-123.

22. Sutherland JD (2010) Ribonucleotides. *Cold Spring Harbor Perspect. Biol.* 2(4):a005439.

23. Bean HD*, et al.* (2007) Formation of a β-Pyrimidine Nucleoside by a Free Pyrimidine Base and Ribose in a Plausible Prebiotic Reaction. *J. Am. Chem. Soc.* 129(31):9556-9557.

24. Kolb VM, Dworkin JP, & Miller SL (1994) Alternative Bases in the RNA world – the prebiotic synthesis of urazole and its ribosides. *J. Mol. Evol.* 38(6):549-557.

25. Sheng Y, Bean HD, Mamajanov I, Hud NV, & Leszczynski J (2009) Comprehensive Investigation of the Energetics of Pyrimidine Nucleoside Formation in a Model Prebiotic Reaction. *J. Am. Chem. Soc.* 131(44):16088-16095.

26. Friedberg EC*, et al.* (2006) *DNA Repair and Mutagenesis, 2nd Edition* (ASM Press, Washington, D.C.) 2nd ed pp 3-69.

27. Shapiro R (1981) Damage to DNA caused by hydrolysis *Chromosome Damage and Repair*, eds Seeberg E & Kleppe K (Plenum Press, New York), pp 3-18.

28. Schroeder GK & Wolfenden R (2007) Rates of Spontaneous Disintegration of DNA and the Rate Enhancements Produced by DNA Glycosylases and Deaminases. *Biochemistry* 46(47):13638-13647.

29. House CH & Miller SL (1996) Hydrolysis of Dihydrouridine and Related Compounds. *Biochemistry* 35(1):315-320.

30. Levy M & Miller SL (1998) The stability of the RNA bases: Implications for the origin of life. *Proc. Natl. Acad. Sci. U.S.A.* 95(14):7933-7938.

31. Seela F & Gabler B (1994) Facile syntheis of 2'-deoxyisoguanosine and realted 2',3'-dideoxyribonucleosides. *Helvetica Chimica Acta* 77(3):622-630 .

32. Shapiro R & Kang S (1969) Uncatalyzed hydrolysis of deoxyuridine, thymidine, and 5-bromodeoxyuridine. *Biochemistry* 8(5):1806-1810.

33. Kochetkov NK & Budovskii EI (1972) *Organic chemistry of nucleic acids* (Plenum Press, New York,) pp 425-448.

34. Shapiro R & Danzig M (1972) Acidic hydrolysis of deoxycytidine and deoxyuridine derivatives. General mechanism of deoxyribonucleoside hydrolysis. *Biochemistry* 11(1):23-29.

35. Stockbridge RB, Schroeder GK, & Wolfenden R (2010) The rate of spontaneous cleavage of the glycosidic bond of adenosine. *Bioorganic Chem.* 38(4-6):224-228

36. Kirnos MD, Khudyakov IY, Alexandrushkina NI, & Vanyushin BF (1977) 2-Aminoadenine is an adenine substituting for a base in S-2L cyanophage DNA. *Nature* 270(5635):369-370.

37. Borquez E, Cleaves HJ, Lazcano A, & Miller SL (2005) An Investigation of Prebiotic Purine Synthesis from the Hydrolysis of HCN Polymers. *Origins Life Evol. Biospheres* 35(2):79-90.

38. Levy M, Miller SL, & Oró J (1999) Production of Guanine from NH4CN Polymerizations. *J. Mol. Evol.* 49(2):165-168.

39. Cleaves HJ & Lazcano A (2009) The Origin of Biomolecules. *Chemical Evolution II: From the Origins of Life to Modern Society,* ACS Symposium Series, (American Chemical Society, Washington DC), pp 17-43.

40. Nishikura K (2010) Functions and Regulation of RNA Editing by ADAR Deaminases. *Annu. Rev. Biochem.* 79(1):321-349.

41. Botta O & Bada JL (2002) Extraterrestrial Organic Compounds in Meteorites. *Surv. Geophys.* 23(5):411-467.

42. Limbach PA, Crain PF, & McCloskey JA (1994) Summary: the modified nucleosides of RNA. *Nucleic Acids Res.* 22(12):2183-2196.

43. Berg JM, Tymoczko JL, & Stryer L (2007) *Biochemistry* (W.H. Freeman, New York) 6th Ed pp 709-731.

44. Gates KS (2009) An Overview of Chemical Processes That Damage Cellular DNA: Spontaneous Hydrolysis, Alkylation, and Reactions with Radicals. *Chem. Res. Toxicol.* 22(11):1747-1760.

45. Fuhrman F, Fuhrman G, Nachman R, & Mosher H (1981) Isoguanosine: isolation from an animal. *Science* 212(4494):557-558.

46. Cheng Q, Gu J, Compaan KR, & Schaefer HF (2012) Isoguanine Formation from Adenine. *Chemistry – A European Journal* 18(16):4877-4886.

47. Grosjean H (2009) *DNA and RNA modification enzymes : structure, mechanism, function and evolution* (Landes Bioscience, Austin, Tex.) p 653 p.

48. Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature* 362(6422):709-715.

49. Lönnberg H & Käppi R (1985) Competition between the hydrolysis and deamination of cytidine and its 5-substituted derivatives in aqueous acid1. *Nucleic Acids Res.* 13(7):2451-2456.

50. Garrett ER & Tsau J (1972) Solvolyses of cytosine and cytidine. *J. Pharm. Sci.* 61(7):1052.

51. Przybylski JL & Wetmore SD (2009) Modeling the Dissociative Hydrolysis of the Natural DNA Nucleosides. *The Journal of Physical Chemistry B* 114(2):1104-1113.

52. Garrett ER, Seydel JK, & Sharpen AJ (1966) Acid-catalyzed solvolysis of pyrimidine nucleosides. *Journal of Organic Chemistry* 31(7):2219.

53. Berti PJ & McCann JAB (2006) Toward a Detailed Understanding of Base Excision Repair Enzymes: Transition State and Mechanistic Analyses of N-Glycoside Hydrolysis and N-Glycoside Transfer. *Chem. Rev.* 106(2):506-555.

54. Miller SL & Orgel LE (1974) *The origins of life on the earth* (Prentice-Hall, Englewood Cliffs, N.J.,) pp x, 229 p.

55. Gates KS (2009) An Overview of Chemical Processes That Damage Cellular DNA: Spontaneous Hydrolysis, Alkylation, and Reactions with Radicals. *Chem. Res. Toxicol.* 22(11):1747-1760.

56. Dawson RMC (1986) *Data for biochemical research* (Clarendon Press, Oxford) 3rd Ed pp xii, 580 p.

57. Iocono JA, Gildea B, & McLaughlin LW (1990) Mil-acid hydrolysis of 2-pyrimidinone containing DNA fragments generates apurininc/apyrimidinic sites. *Tetrahedron Letters* 31(2):175-178.

58. Seela F & He Y (2000) 2 '-deoxyuridine and 2 '-deoxyisocytidine as constituents of DNA with parallel chain orientation: The stabilization of the iC(d)center dot Gd base pair by the 5-methyl group. *Helvetica Chimica Acta* 83(9):2527-2540.

59. Gommers-Ampt J & Borst P (1995) Hypermodified bases in DNA. *The FASEB Journal* 9(11):1034-1042.

60. Warren RAJ (1980) Modified Bases in Bacteriophage DNAs. *Annu. Rev. Microbiol.* 34(1):137-158.

61. Nabel CS, Manning SA, & Kohli RM (2011) The Curious Chemical Biology of Cytosine: Deamination, Methylation,and Oxidation as Modulators of Genomic Potential. *ACS Chem. Biol.* 7(1):20-30.

62. Bennett MT*, et al.* (2006) Specificity of Human Thymine DNA Glycosylase Depends on N-Glycosidic Bond Stability. *J. Am. Chem. Soc.* 128(38):12510-12519.

63. Maiti A & Drohat AC (2011) Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine. *J. Biol. Chem.* 286(41):35334-35338.

64. Williams RT & Wang Y (2012) A Density Functional Theory Study on the Kinetics and Thermodynamics of N-Glycosidic Bond Cleavage in 5-Substituted 2′-Deoxycytidines. *Biochemistry* 51(32):6458-6462.

65. Khudyakov IY, Kirnos MD, Alexandrushkina NI, & Vanyushin BF (1978) Cyanophage S-2L contains DNA with 2,6-diaminopurine substituted for adenine. *Virology* 88(1):8-18.

66. Bass BL (2002) RNA EDITING BY ADENOSINE DEAMINASES THAT ACT ON RNA. *Annu. Rev. Biochem.* 71(1):817-846.

67. Lindahl T & Wood RD (1999) Quality Control by DNA Repair. *Science* 286(5446):1897-1905.

68. Szathmary E (2003) Why are there four letters in the genetic alphabet? *Nat Rev Genet* 4(12):995-1001.

69. Benner SA, Ricardo A, & Carrigan MA (2004) Is there a common chemical model for life in the universe? *Current Opinion in Chemical Biology* 8(6):672-689.

70. Benner SA & Sismour AM (2005) Synthetic biology. *Nat Rev Genet* 6(7):533-543.

71. Benner SA (2004) Understanding Nucleic Acids Using Synthetic Chemistry. *Acc. Chem. Res.* 37(10):784-797.

72. Switzer C, Moroney SE, & Benner SA (1989) Enzymatic incorporation of a new base pair into DNA and RNA. *J. Am. Chem. Soc.* 111(21):8322-8323.

73. Hoshika S, Chen F, Leal NA, & Benner SA (2010) Artificial Genetic Systems: Self-Avoiding DNA in PCR and Multiplexed PCR. *Angew. Chem. Int. Ed. Engl.* 49(32):5554-5557.

74. Eschenmoser A (1999) Chemical Etiology of Nucleic Acid Structure. *Science* 284(5423):2118-2124.

75. Bailly C & Waring MJ (1998) The use of diaminopurine to investigate structural properties of nucleic acids and molecular recognition between ligands and DNA. *Nucleic Acids Res.* 26(19):4309-4314.

76. Geyer CR, Battersby TR, & Benner SA (2003) Nucleobase Pairing in Expanded Watson-Crick-like Genetic Information Systems. *Structure (London, England : 1993)* 11(12):1485-1498.

77. Suspène R*, et al.* (2008) Inversing the natural hydrogen bonding rule to selectively amplify GC-rich ADAR-edited RNAs. *Nucleic Acids Res.* 36(12):e72.

78. Budke B & Kuzminov A (2006) Hypoxanthine Incorporation Is Nonmutagenic in Escherichia coli. *J. Bacteriol.* 188(18):6553-6560.

79. Martin FH, Castro MM, Aboulela F, & Tinoco I (1985) Base-pairing involving deoxyinosine- implications for probe design. *Nucleic Acids Res.* 13(24):8927-8938.

80. Cheong C, Tinoco I, & Chollet A (1988) Thermodynamic studies of base pairing involving 2,6-diaminopuring. *Nucleic Acids Res.* 16(11):5115-5122.

81. Krishnamurthy R (2012) Role of pKa of Nucleobases in the Origins of Chemical Evolution. *Acc. Chem. Res.*

82. Rich A (1962) On the problems of evolution and biochemical information transfer. *Horizons in Biochemistry* eds Kasha M & Pullman B (Academic Press, New York), pp 103 - 126.

83. Roberts C, Bandaru R, & Switzer C (1997) Theoretical and Experimental Study of Isoguanine and Isocytosine: Base Pairing in an Expanded Genetic System. *J. Am. Chem. Soc.* 119(20):4640-4649.

84. Poole A, Penny D, & Sjöberg B-M (2001) Confounded cytosine! Tinkering and the evolution of DNA. *Nat Rev Mol Cell Biol* 2(2):147-151.

85. Robinson H*, et al.* (1998) 2‘-Deoxyisoguanosine Adopts More than One Tautomer To Form Base Pairs with Thymidine Observed by High-Resolution Crystal Structure Analysis†. *Biochemistry* 37(31):10897-10905.

86. Wuenschell GE, O'Connor TR, & Termini J (2003) Stability, Miscoding Potential, and Repair of 2‘-Deoxyxanthosine in DNA: Implications for Nitric Oxide-Induced Mutagenesis†. *Biochemistry* 42(12):3608-3616.

87. Kua J & Bada J (2011) Primordial Ocean Chemistry and its Compatibility with the RNA World. *Origins Life Evol. Biospheres* 41(6):553-558.

88. Dworkin JP, Lazcano A, & Miller SL (2003) The roads to and from the RNA world. *J. Theor. Biol.* 222(1):127-134.

89. Saladino R*, et al.* (2005) Origin of Informational Polymers. *J. Biol. Chem.* 280(42):35658-35669.

90. Saladino R, Crestini C, Ciciriello F, Di Mauro E, & Costanzo G (2006) Origin of Informational Polymers. *J. Biol. Chem.* 281(9):5790-5796.

91. Shapiro R (1995) The prebiotic role of adenine: A critical analysis. *Origins Life Evol. Biospheres* 25(1):83-98.

92. Shapiro R (1999) Prebiotic cytosine synthesis: A critical analysis and implications for the origin of life. *Proc. Natl. Acad. Sci. U.S.A.* 96(8):4396-4401.

93. Garrett ER & Mehta PJ (1972) Solvolysis of adenine nucleosides. I. Effects of sugar and adenine substituents on acid solvolyses. *J. Am. Chem. Soc.* 94(24):8532-8541.

CHAPTER 4: Outlook

The ideas and work presented in this dissertation established a new area of investigation to help explain the composition of the genetic alphabet. In this study it was observed that the native purines in comparison to related nucleosides, exhibited the slowest deglycosylation rates under acidic conditions. Further investigations that include other damaged and alternative purines nucleosides along with a pH dependent study needs to be conducted in order to obtain a global perspective on *N*-glycosyl stability. This is especially imperative when considering the pyrimidine nucleosides, a group that has received even less attention with respect to alternative and modified heterocycles. As demonstrated in this dissertation, the view that pyrimidines, as a class, exhibit robust *N*-glycosyl stabilities under acidic conditions may only be limited to heterocycles that are typically associated with the native and biologically relevant modified pyrimidines. The deoxynucleosides of isocytosine and 2-pyrimidinone were two cases that highlighted how relatively modest changes to the heterocycle can dramatically alter deglycosylation kinetics. It is especially noteworthy to point out that the *N*-glycosyl stabilities of these pyrimidines were even more labile than many of the purines deoxynucleosides. A pH and temperature dependent study on the disoC nucleoside to extract activation parameters might provide insight to the dramatic differences that exist between the native dC and disoC deglycosylation rates. The ubiquitous 2-thiouracil modification found in RNA (discussed in Chapter 1), has not been identified to occur as a DNA modification. Based on its lower heterocyclic pK_a value, 2-thio-deoxyuridine may exhibit a weaker glycosidic bond in

comparison to the native pyrimidine deoxynucleosides. Work is already under way to explore this hypothesis.

Some of the most important long-term studies to be conducted are in measuring the changes in *N*-glycosyl stability as a heterocycle moves from a ribo→deoxyribose nucleoside. The variation in rate enhancements observed in this work highlight the need to continue investigations under different conditions to track the stability changes. A detailed understanding for how the hydrolytic stabilities of these pivotal linkages differ in model nucleosides will provide a first approach to understanding the pressures that may have plagued early life after the transition from genomic RNA to DNA.

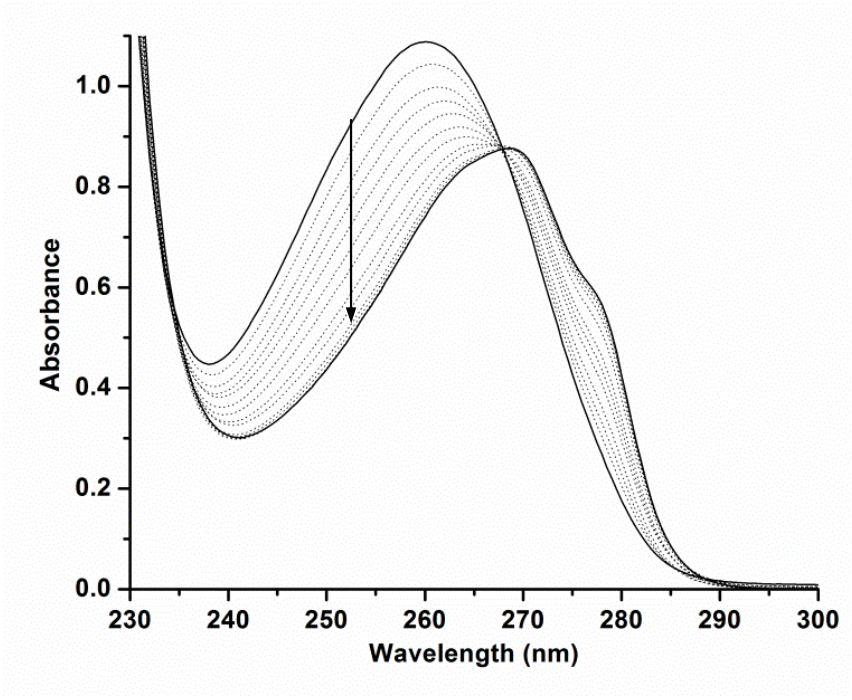
Appendix: Supporting Information for Chapter 3

Supporting Information

- **A.1** Examples of UV-spectral data and Absorbance vs. Time plots from kinetics experiments conducted at 0.1M HCl*
    - **Figure A.1**: 2'-Deoxyadenosine at 15C
    - **Figure A.2**: 2'-Deoxyguanosine at 20C
    - **Figure A.3**: 2'-Deoxy-2,6-diaminopurineriboside at 20C
    - **Figure A.4:** 2'-Deoxyinosine at 20C
    - **Figure A.5:** 2'-Deoxyisoguanosine at 50C *(pH 3)
    - **Figure A.6:** 2'-Deoxyzebularine at 25C
    - **Figure A.7:** 2'-Deoxyxanthosine at 50C(pH 3)
    - **Figure A.8:** Adenosine at 75C
    - **Figure A.9:** Guanosine at 65C
    - **Figure A.10:** 2,6-Diaminopurineriboside at 65C
    - **Figure A.11:** Inosine at 65C
    - **Figure A.12:** Isoguanosine at 75C
    - **Figure A.13:** Zebularine at 65C
    - **Figure A.14:** Xanthosine at 65C

- **A.2** Tables of rate constants determined at pH 1
    - **Table A.1:** Compilation of determined rate constants for deoxynucleosides 0.1 M HCl
    - **Table A.2:** Compilation or determined rate constants for ribonucleosides at 0.1MHCl
- **A.3** Combined Eyring plots of deoxynucleosides and ribonucleosides at 0.1 M HCl
    - **Figure A.15:** Eyring Plot

A1.1 Representative UV-spectral data from kinetics experiments conducted at 0.1M HCl but monitored using alkaline quench



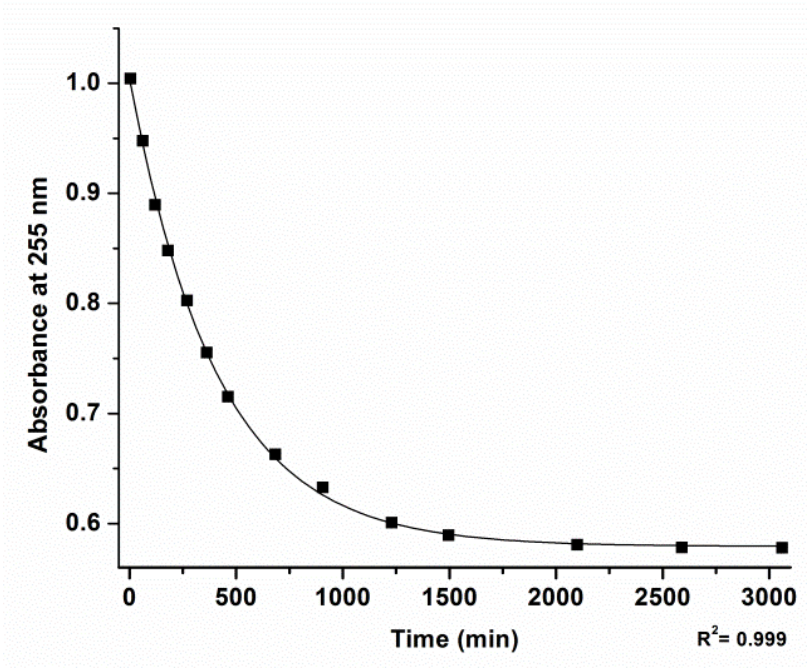
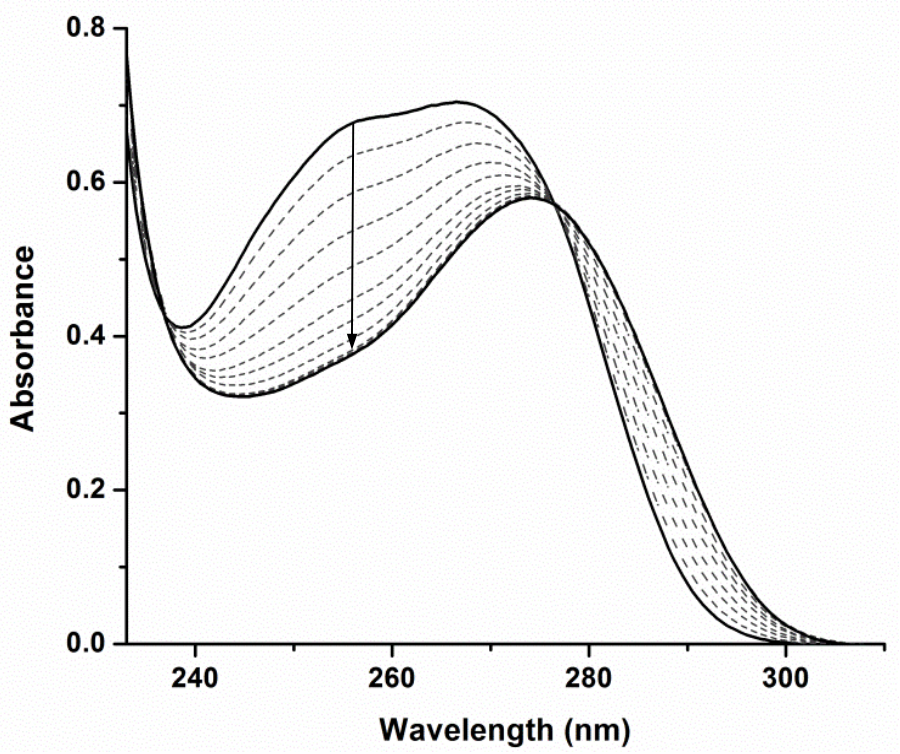


Figure A.1: (Top) 2'-deoxyadenosine→adenine deglycosylation spectral overlay at 15°C. (Bottom) Corresponding absorbance, λ = 255 nm vs. time plot at 15C.



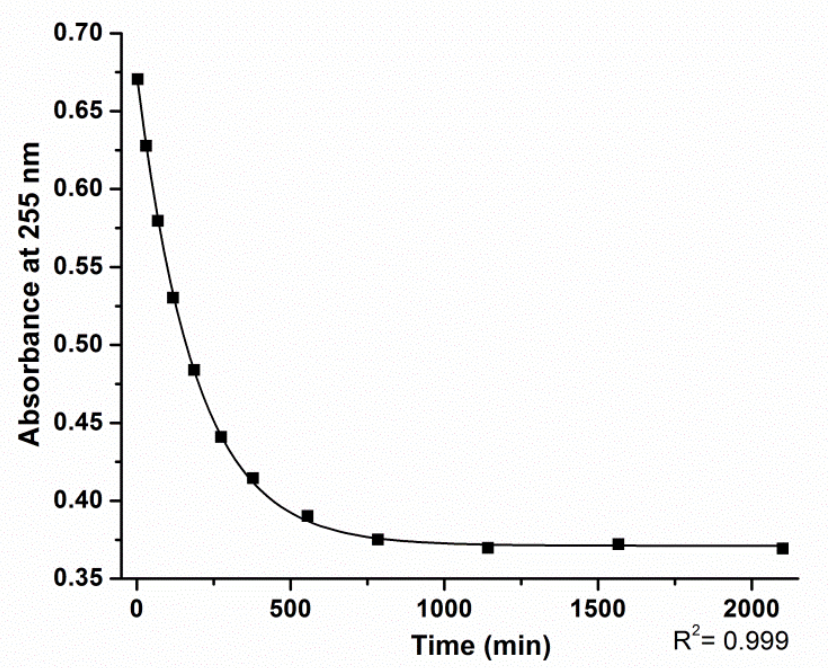
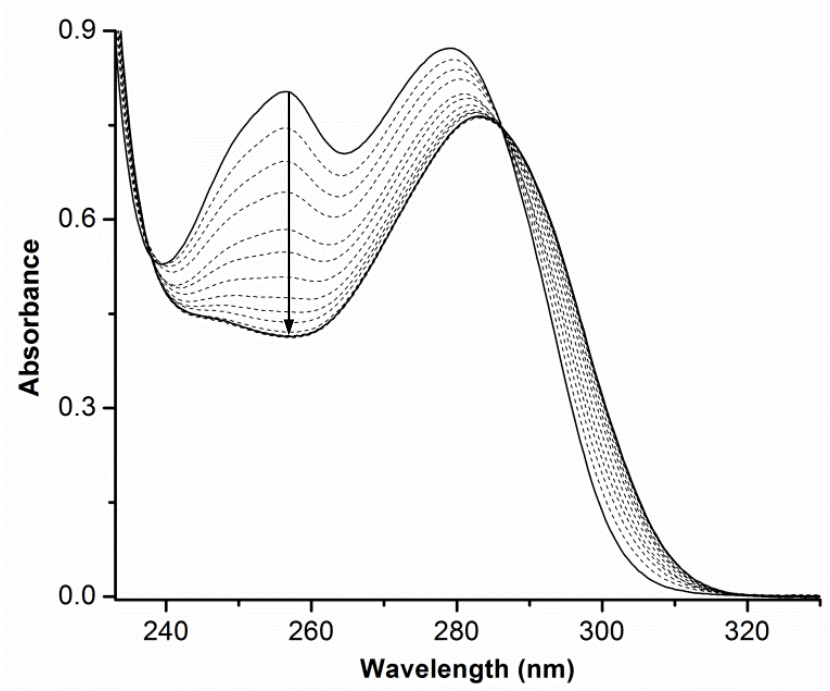


Figure A.2: (Top) 2'-deoxyguanosine→guanine deglycosylation spectral overlay at 20°C. (Bottom) Corresponding absorbance, λ = 255 nm vs. time plot at 20°C.



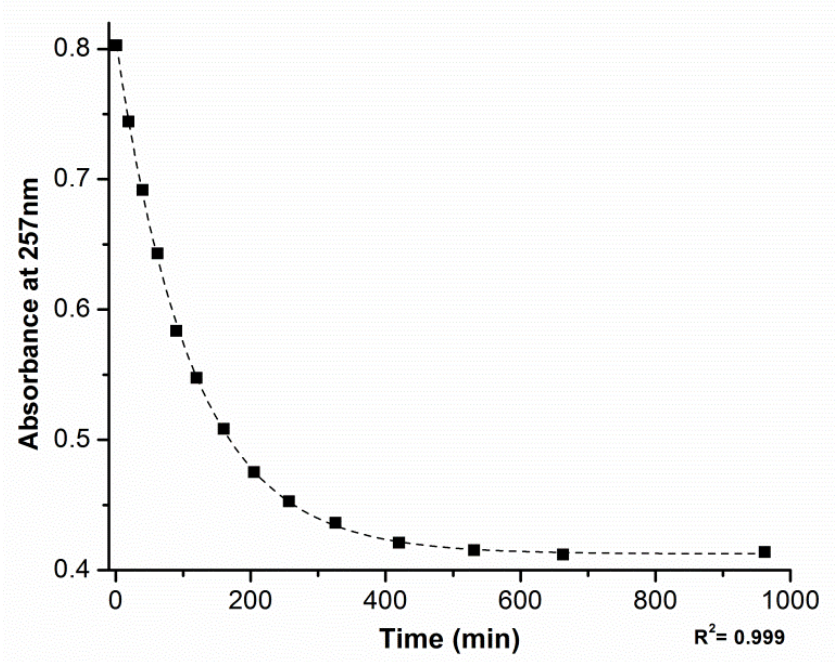


Figure A.3: (Top) 2'-deoxy-2,6-diaminopurine →2,6- diaminopurine deglycosylation spectral overlay at 20°C. (Bottom) Corresponding absorbance, λ = 257 nm vs. time plot at 20°C.

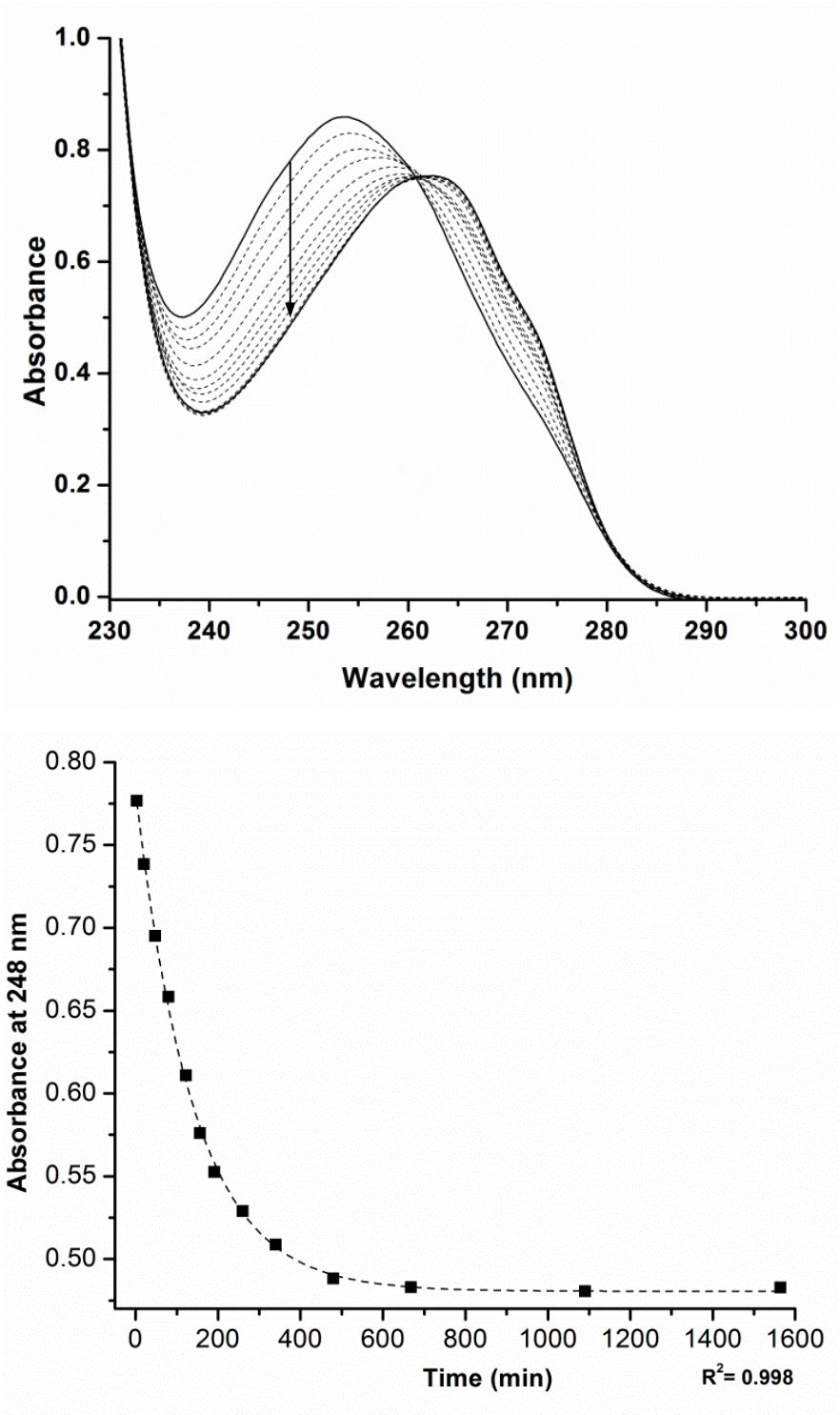


Figure A.4: (Top) 2'-deoxyinosine→hypoxanthine deglycosylation spectral overlay at 20°C. (Bottom) Corresponding absorbance, λ = 248 nm vs. time plot at 20°C.

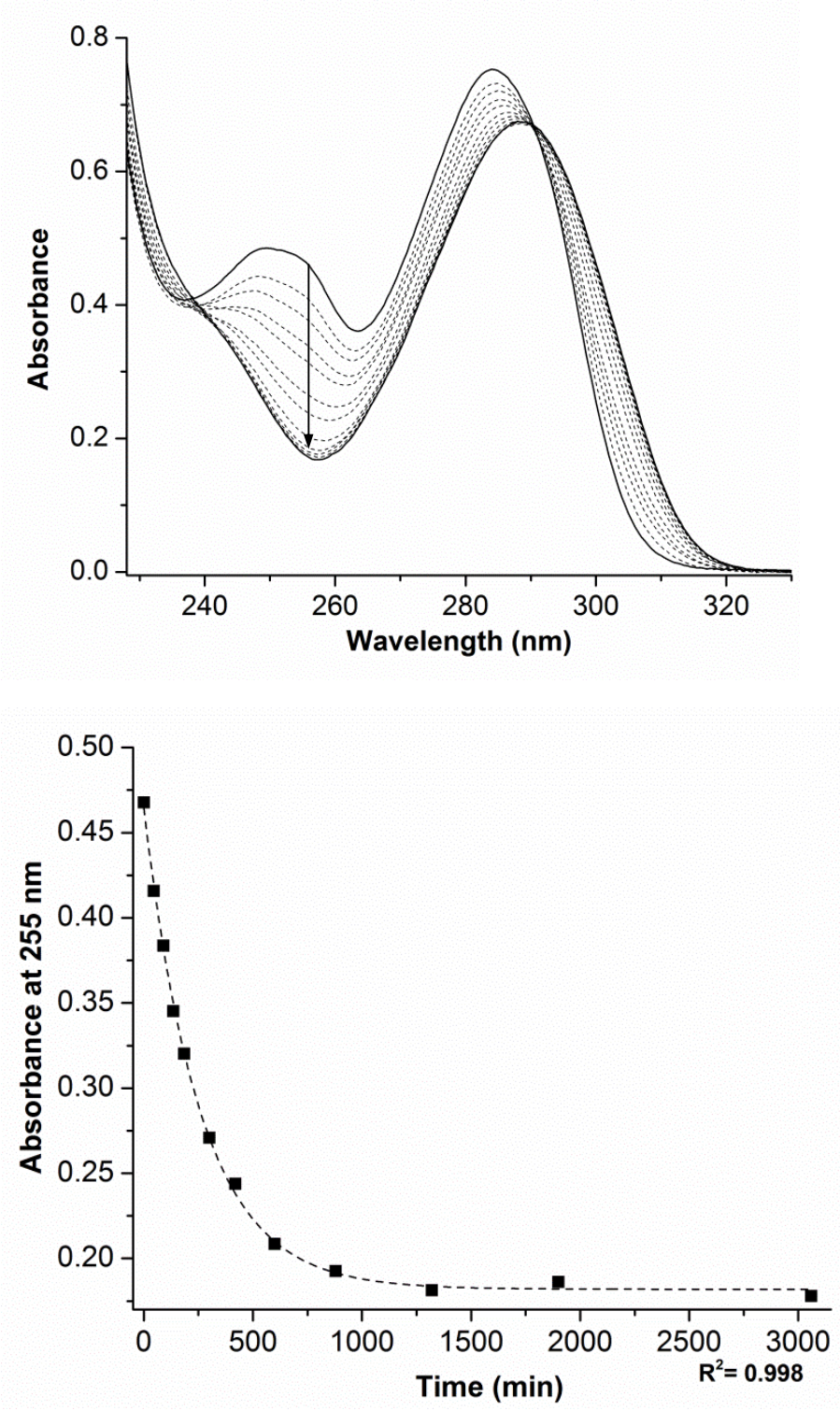


Figure A.5: (Top) 2'-deoxyisoguanosine→isoguanine deglycosylation spectral overlay at 50°C, pH 3. (Bottom) Corresponding absorbance, λ = 255 nm vs. time plot at 50°C.

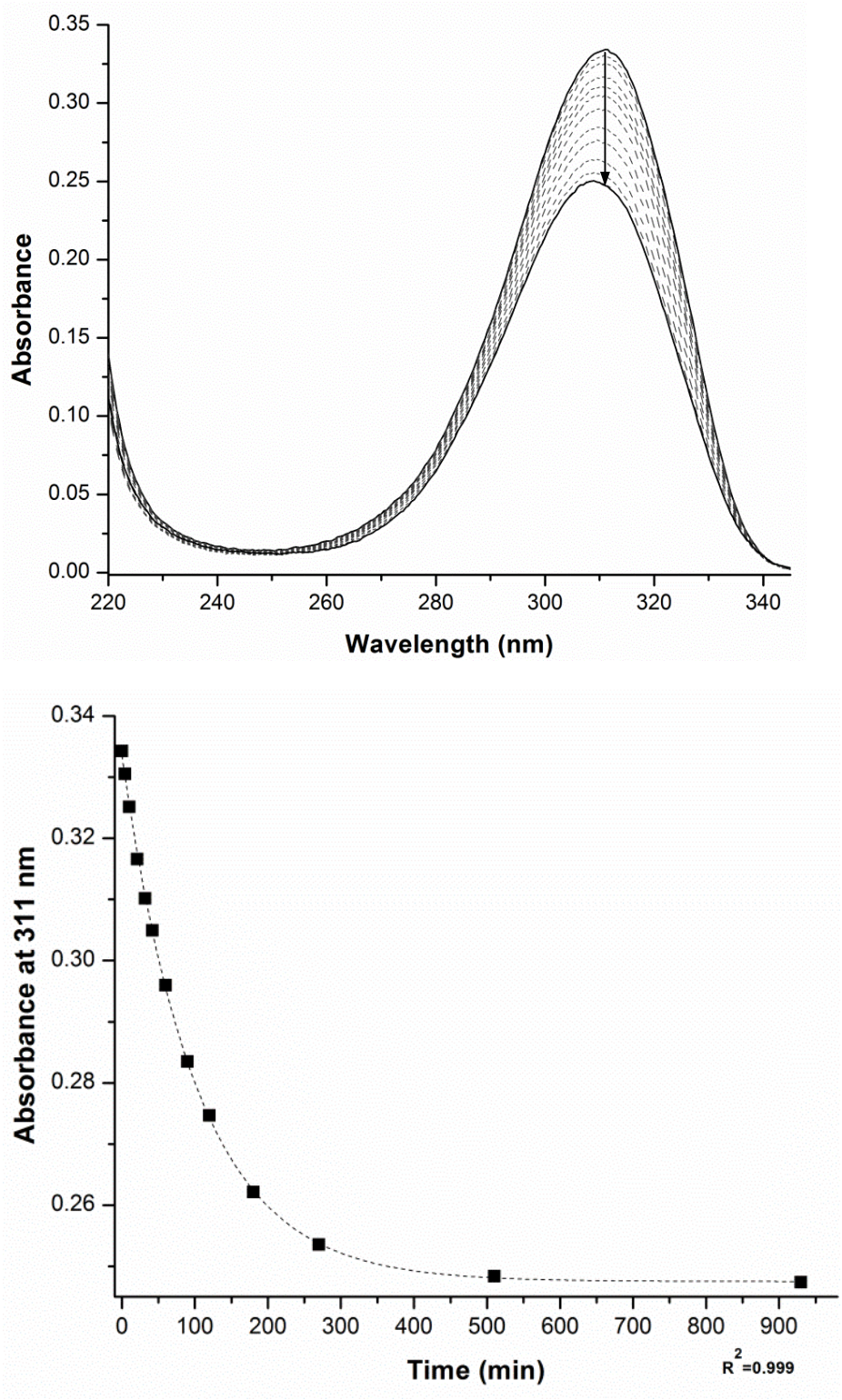
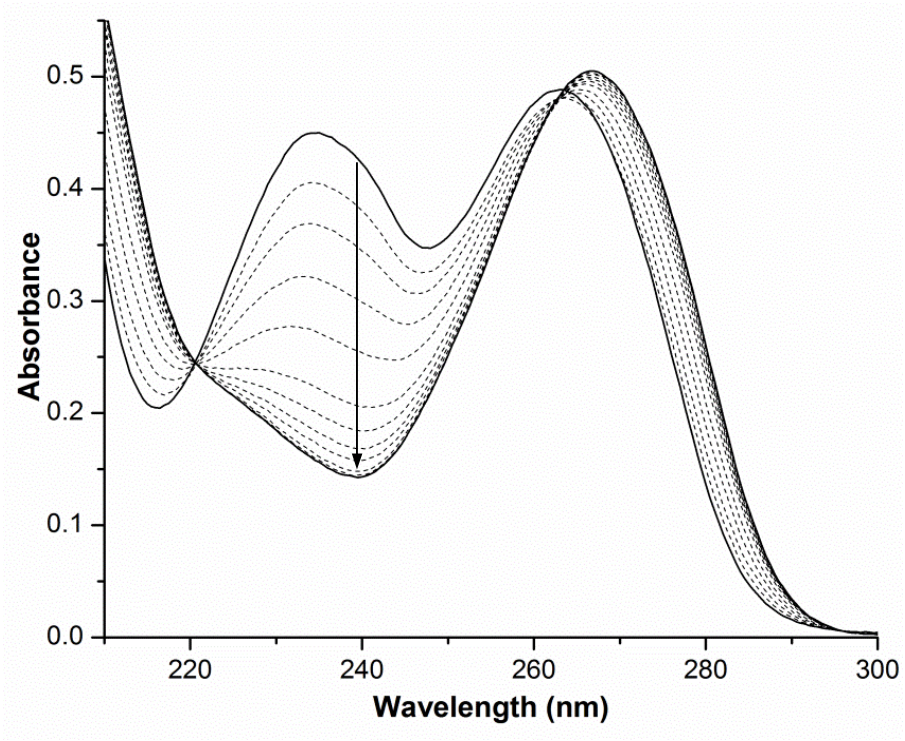


Figure A.6: (Top) 2'-deoxyzebularine→2-pyrimidinone deglycosylation spectral overlay at 25°C. (Bottom) Corresponding absorbance, λ = 311 nm vs. time plot at 25°C.



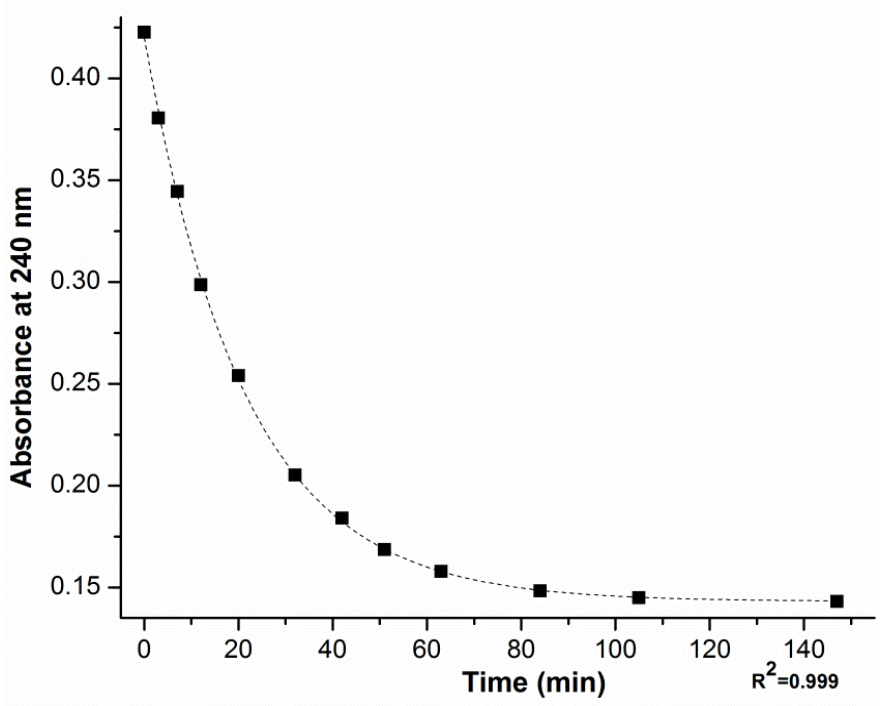


Figure A.7: (Top) 2'-deoxyxanthosine→xanthine deglycosylation spectral overlay at 50^{o}C, pH 3. (Bottom) Corresponding absorbance, λ = 240 nm vs. time plot at 50^{o}C.

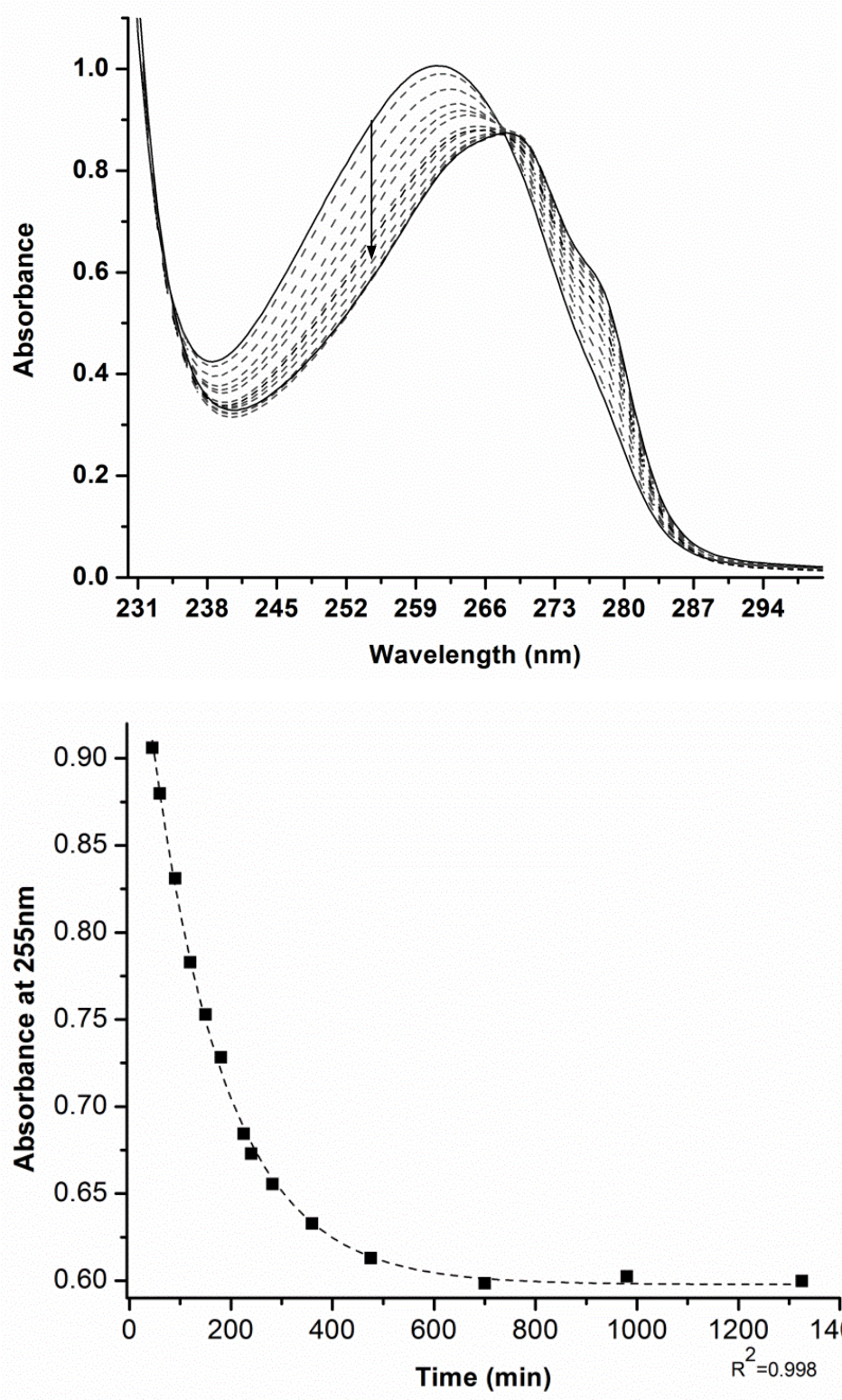
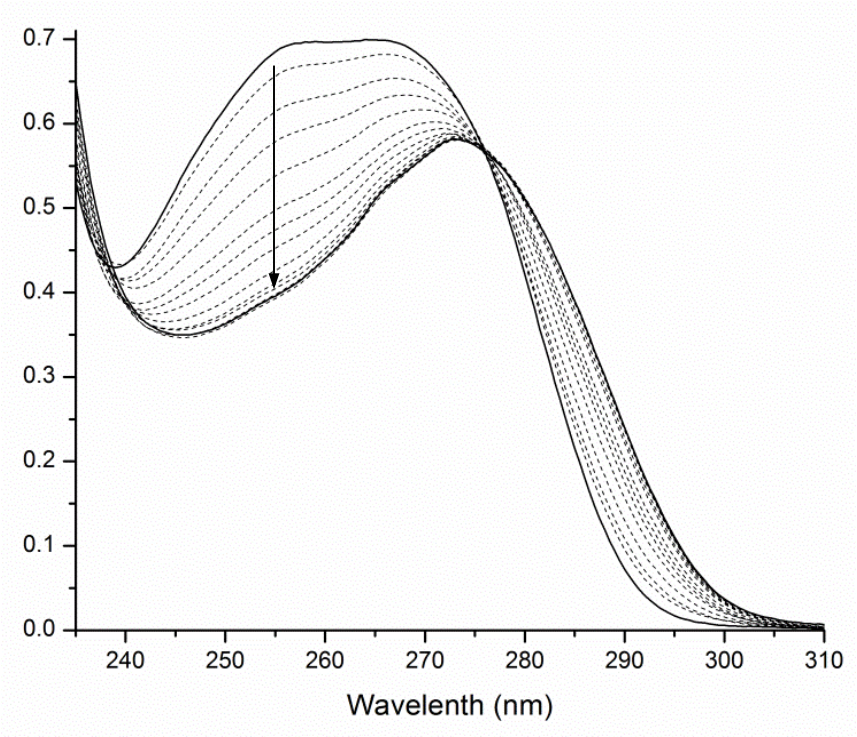


Figure A.8: (Top) Adenosine→adenine deglycosylation spectral overlay at 75°C. (Bottom) Corresponding absorbance, λ = 255 nm vs. time plot at 75°C



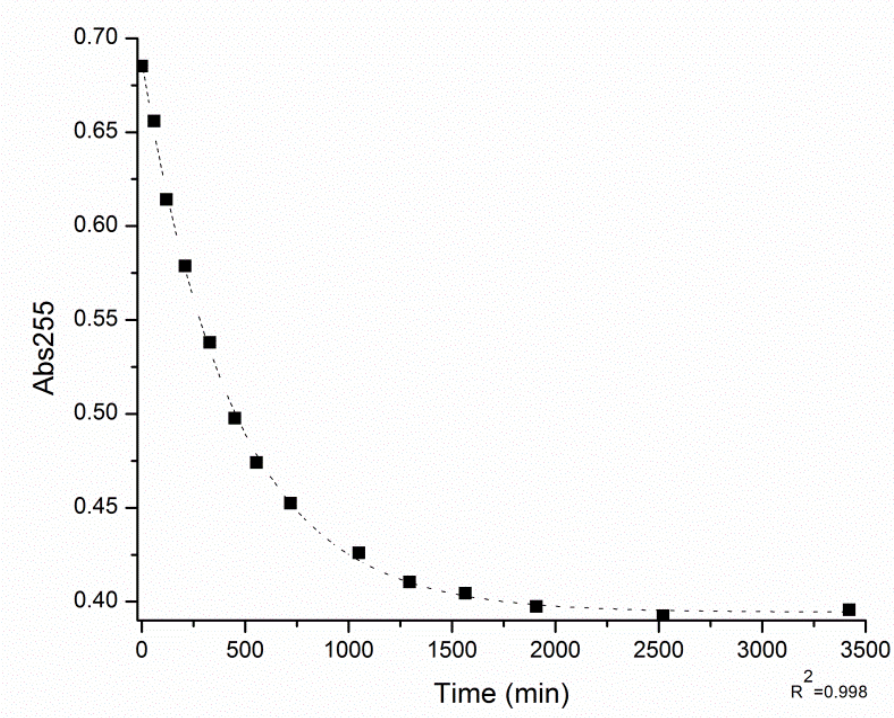
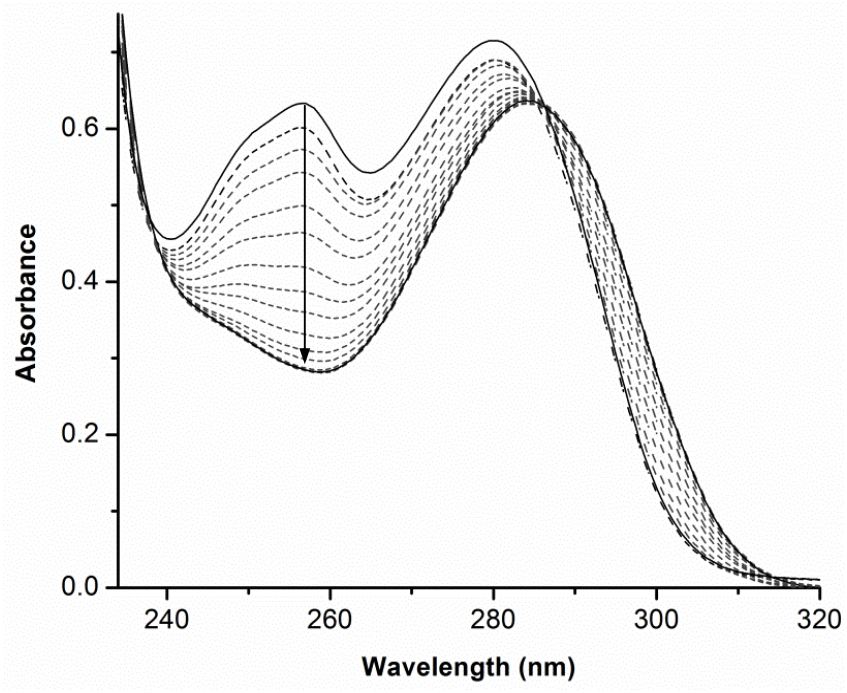


Figure A.9: (Top) Guanosine→guanine deglycosylation spectral overlay at 65°C. (Bottom) Corresponding absorbance, λ = 255 nm vs. time plot at 65°C.



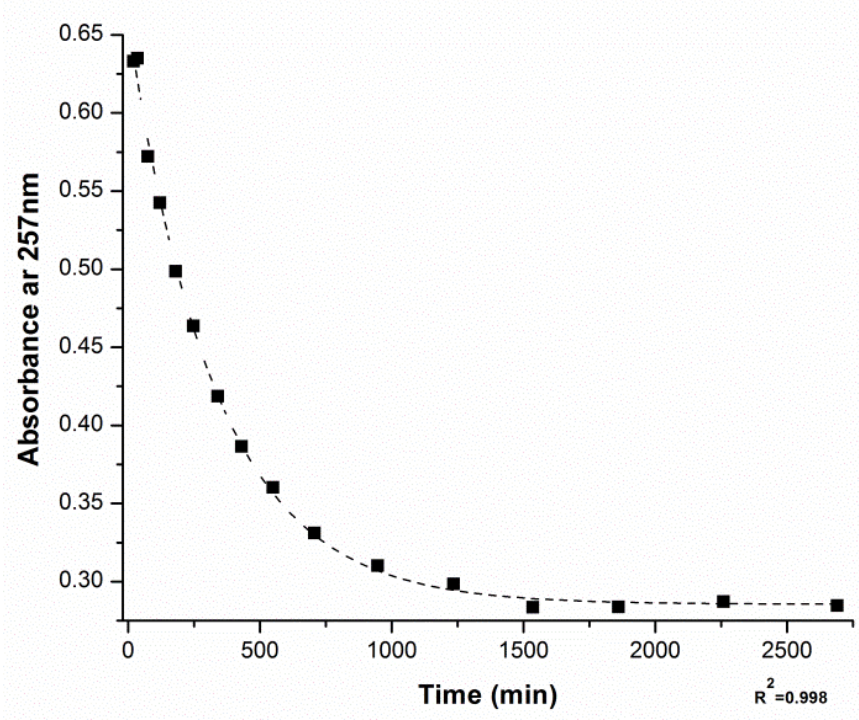
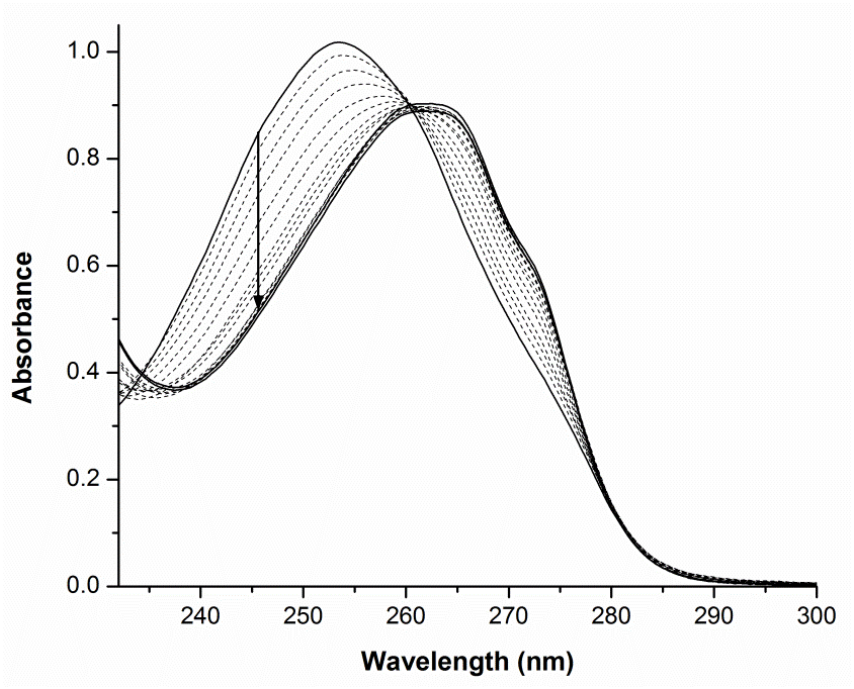


Figure A.10: (Top) 2,6-diaminopurineriboside →2,6- diaminopurine deglycosylation spectral overlay at 65°C. (Bottom) Corresponding absorbance, λ = 257 nm vs. time plot at 65°C**.**



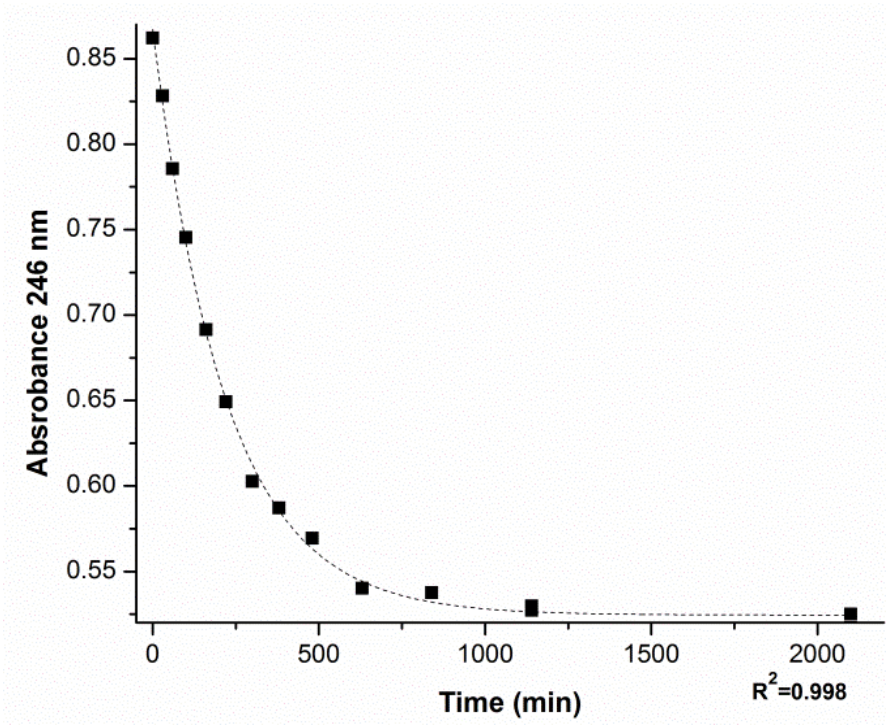
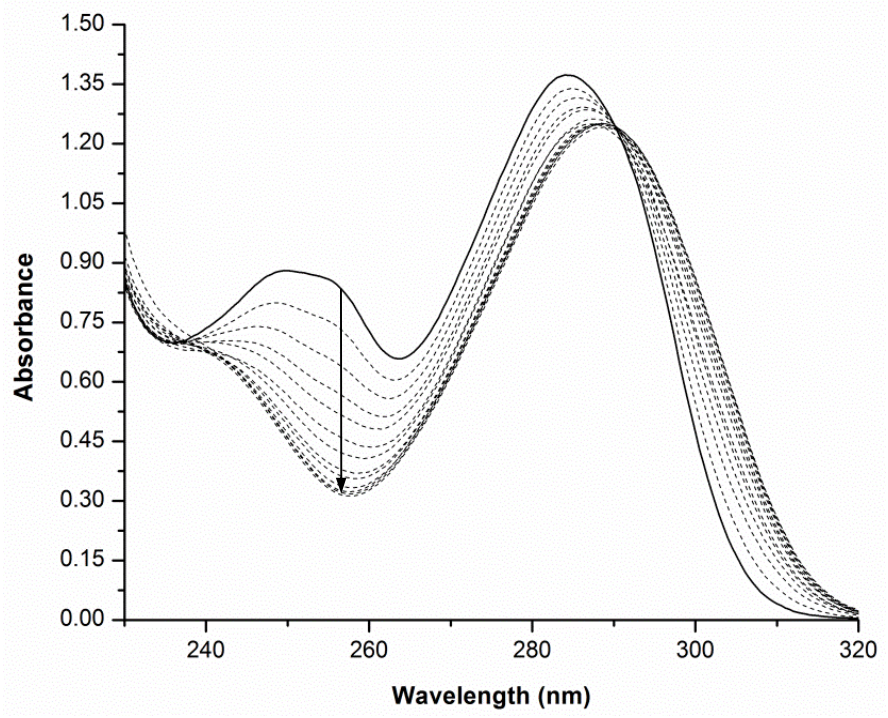


Figure A.11: (Top) Inosine→hypoxanthine deglycosylation spectral overlay at 65^{o}C. (Bottom) Corresponding absorbance, λ = 248 nm vs. time plot at 65^{o}C.



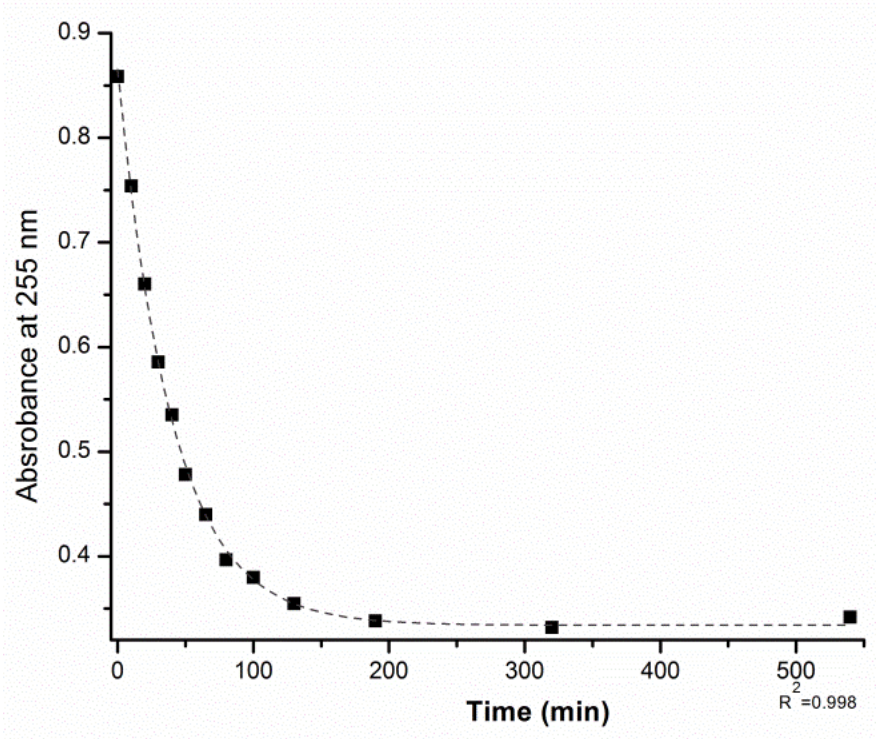
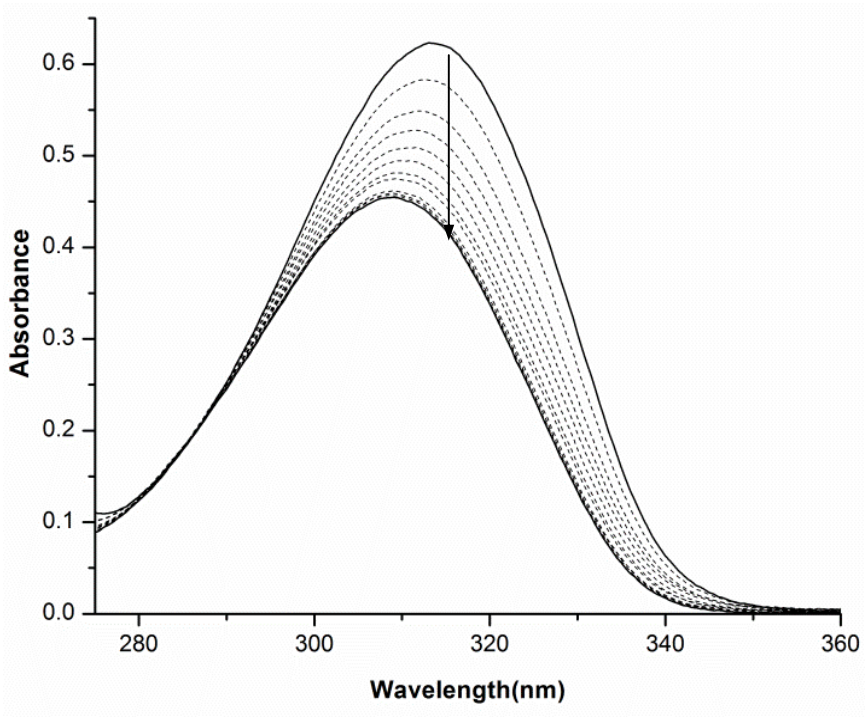


Figure A.12: (Top) Isoguanosine→isoguanine deglycosylation spectral overlay at 75^{o}C. (Bottom) Corresponding absorbance, λ = 255 nm vs. time plot at 75^{o}C.



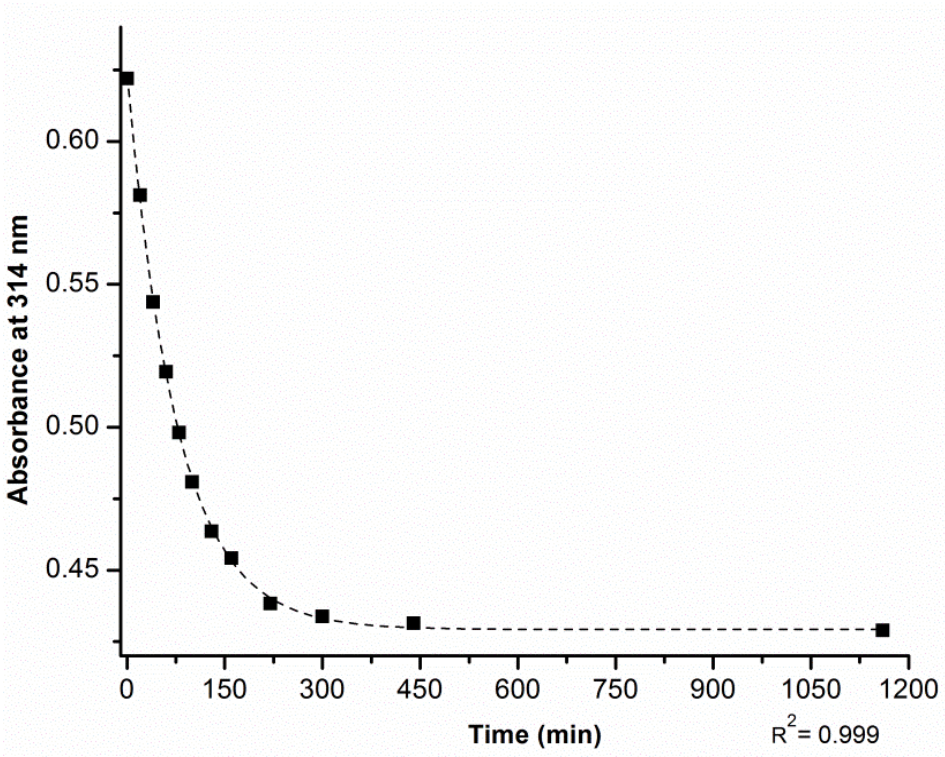
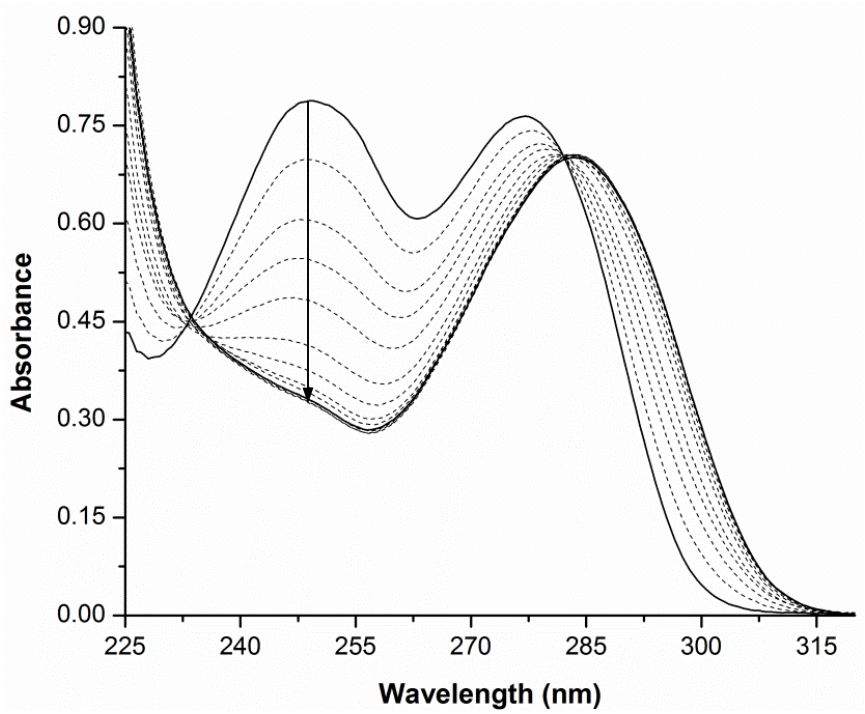


Figure A.13: (Top) Zebularine→2-pyrimidinone deglycosylation spectral overlay at 65^{o}C. (Bottom) Corresponding absorbance, λ = 314 nm vs. time plot at 65^{o}C.



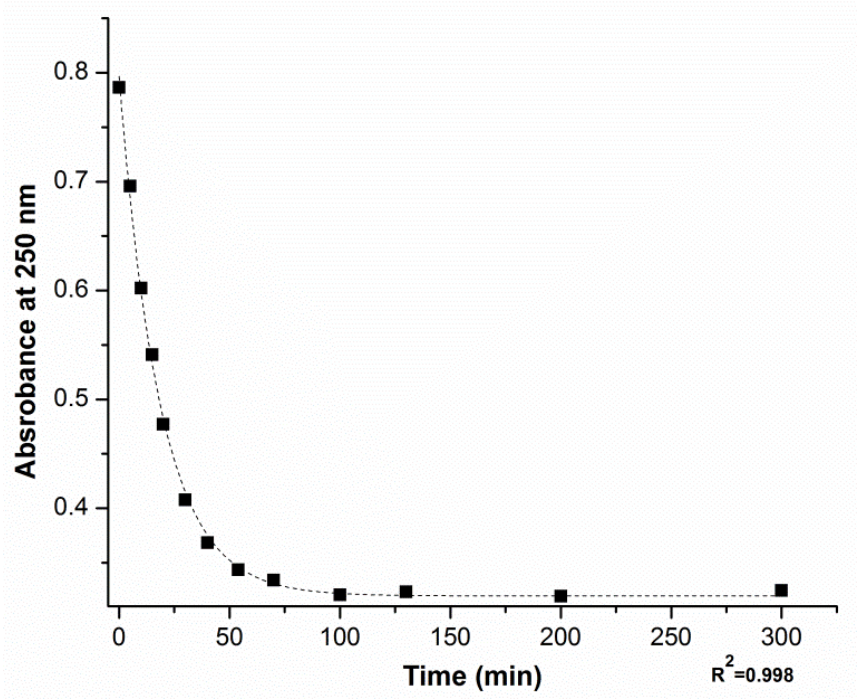


Figure A.14: (Top) Xanthosine→xanthine deglycosylation spectral overlay at 65^{o}C. (Bottom) Corresponding absorbance, λ = 250 nm vs. time plot at 65^{o}C.

A.2 Tables of rate constants determined at pH 1

Table A.1: Compilation of determined rate constants for deoxyribonucleosides used for all Arrhenius and Eyring plots at 0.1M HCl

Nucleoside	*k* (s^{-1}) at 10°C	*k* (s^{-1}) at 15°C	*k* (s^{-1}) at 20°C	*k* (s^{-1}) at 25°C	*k* (s^{-1}) at 30°C	*k* (s^{-1}) at 35°C
dA		4.06E-5, 4E-5	8.766E-5 8.18E-5	1.66E-4 1.67E-4	3.173E-4 3.19E-4	5.72E-4 6.01E-4
dG		4.38E-5 4.533E-5 4.1E-5	8.85E-5 9.083E-5 8.166E-5	1.625E-4 1.718E-4 1.663E-4	3.25E-4, 3.151E-4	5.5483E-4, 5.395E-4, 5.601E-4
dDap	3.5666E-5, 3.66667E-5	7.3833E-5, 7.75E-5 7.3E-5	1.485E-4, 1.4816E-4, 1.505E-4	3E-4, 2.946E-4 3.08E-4	5.64E-4 5.173E-4	
dIno	2.78333E-5 2.66667E-5	6.2E-5 6.366E-5	1.33E-4 1.18667E-4 1.305E-4	2.668E-4 2.605E-4 2.5E-4	5.715E-4 4.7883E-4 5.661E-4	
disoG		2.43E-4 2.49333E-4	4.59167E-4 4.96167E-4	8.875E-4 8.523E-4		***k* (s^{-1}) at 37°C** 0.00342 0.00352
dZeb			***k* (s^{-1}) at 25°C** 1.61667E-4 1.56E-4	***k* (s^{-1}) at 30°C** 3.35E-4 3.14E-4	***k* (s^{-1}) at 37°C** 8.345E-4 8.36E-4	***k* (s^{-1}) at 50°C** 0.00405 0.00415
dXan	***k* (s^{-1}) at 12°C** 9.14833E-4 8.81167E-4	***k* (s^{-1}) at 15°C** 0.00137 0.00136	***k* (s^{-1}) at 17°C** 0.00176 0.00177	***k* (s^{-1}) at 20°C** 0.00259 0.00263	***k* (s^{-1}) at 25°C** 0.00472 0.0047	

Table A.2: Compilation of determined rate constants of ribonucleosides used for all Arrhenius and Eyring plots at 0.1M HCl

Nucleoside	*k* (s^{-1}) at 60°C	*k* (s^{-1}) at 65°C	*k* (s^{-1}) at 70°C	*k* (s^{-1}) at 75°C	*k* (s^{-1}) at 80°C	*k* (s^{-1}) at 85°C
A		3.533E-5 3.05E-5	6.16667E-5 5.6E-5	1.1516E-4 1.1966E-4	2.016E-4 1.913E-4	3.21667 E-4 3.35667 E-4
G		3.75E-5 3.533E-5	7.0333E-5 6.3166E-5	1.1466E-4 9.8833E-5	1.65E-4 1.671E-4 2.133E-4	2.52833 E-4 3.48167 E-4 3.51667 E-4
Dap		5.01667E-5 4.71667E-5	8.06667E-5 8.51667E-5	1.6883E-4 1.6533E-4 1.5116E-4	2.758E-4 2.711E-4 2.631E-4	5.11333 E-4 5.195E-4
Ino		8.41667E-5 7.56667E-5	1.63167E-4 1.34333E-4	2.9033E-4 2.31E-4 2.865E-4	5.25E-4 4.49E-4 4.303E-4	9.62E-4 7.19167 E-4 6.895E-4
IsoG	8.46667E-5 6.95E-5		2.46833E-4 2.45333E-4	4.165E-4 4.245E-4	6.87E-4 7.476E-4	
Zeb	1.14667E-4 1.18833E-4	2.17333E-4 2.37E-4	4.07167E-4 4.15667E-4	6.9366E-4 7.655E-4		
Xan	***k* (s^{-1}) at 50°C** 1.43833E-4 1.48333E-4	***k* (s^{-1}) at 55°C** 2.69E-4 2.6983E-4	***k* (s^{-1}) at 60°C** 5.05333E-4 5.14833E-4 5.005E-4	***k* (s^{-1}) at 65°C** 8.69E-4 8.935E-4 8.63E-4	***k* (s^{-1}) at 70°C** 0.00145 0.00152 0.00146	

A.3 Combined Eyring plots of deoxynucleosides and ribonucleosides at 0.1 M HCl

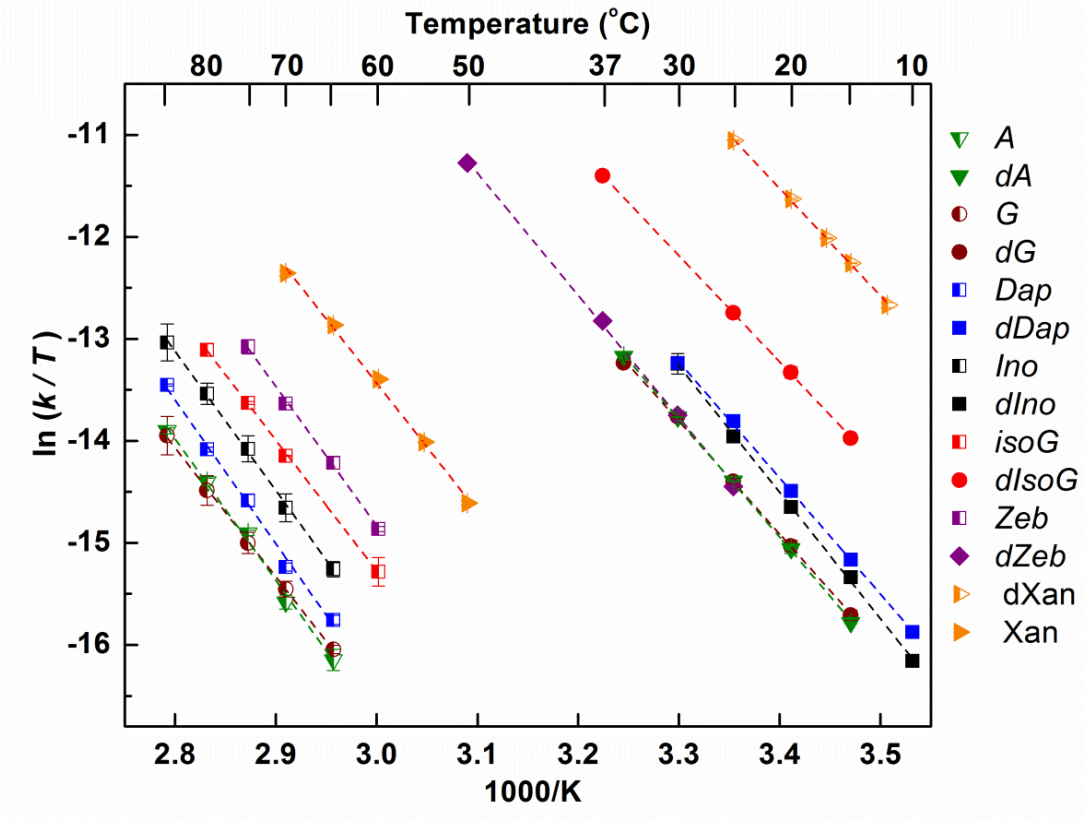


Figure A.15: Combined Eyring plots of deoxynucleosides and ribonucleosides