Analysis of the Intricacies of Substrate Recognition of High Mobility Group Proteins and Aminoacyl-tRNA Synthetases Using Non-Cognate Substrates

Douglas Van Iverson II

University of Southern Mississippi

Follow this and additional works at: https://aquila.usm.edu/dissertations

Part of the Biochemistry Commons, Evolution Commons, Medicinal-Pharmaceutical Chemistry Commons, and the Nucleic Acids, Nucleotides, and Nucleosides Commons

Recommended Citation


This Dissertation is brought to you for free and open access by The Aquila Digital Community. It has been accepted for inclusion in Dissertations by an authorized administrator of The Aquila Digital Community. For more information, please contact Joshua.Cromwell@usm.edu.
ANALYSIS OF THE INTRICACIES OF SUBSTRATE RECOGNITION OF HIGH
MOBILITY GROUP PROTEINS AND AMINOACYL-TRNA SYNTHETASES
USING NON-COGNATE SUBSTRATES

by

Douglas Van Iverson II

A Dissertation
Submitted to the Graduate School
and the Department of Chemistry and Biochemistry
at The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

Approved:

___________________________________________
Dr. Anthony Bell, Committee Chair
Assistant Professor, Chemistry and Biochemistry

___________________________________________
Dr. Gordon Cannon, Committee Member
Professor, Chemistry and Biochemistry

___________________________________________
Dr. Vijay Rangachari, Committee Member
Associate Professor, Chemistry and Biochemistry

___________________________________________
Dr. Karl Wallace, Committee Member
Associate Professor, Chemistry and Biochemistry

___________________________________________
Dr. Philip Bates, Committee Member
Assistant Professor, Chemistry and Biochemistry

___________________________________________
Dr. Karen S. Coats
Dean of the Graduate School

August 2016
COPYRIGHT BY

Douglas Van Iverson II

2016

Published by the Graduate School

THE UNIVERSITY OF SOUTHERN MISSISSIPPI
ABSTRACT

ANALYSIS OF THE INTRICACIES OF SUBSTRATE RECOGNITION OF HIGH MOBILITY GROUP PROTEINS AND AMINOACYL-TRNA SYNTHETASES USING NON-COGNATE SUBSTRATES

by Douglas Van Iverson II

August 2016

The studies presented in section 1 (Chapters I-IV) focus on the design and development of nucleic acid four-way junctions (4WJs) to target a member of the high mobility group (HMG) proteins, the proinflammatory cytokine high mobility group box 1 protein (HMGB1). In the present study, hybrid PNA-DNA 4WJs based on a model DNA 4WJ were constructed to improve the thermal stability of 4WJs while maintaining strong binding affinity toward HMGB1. An electrophoretic mobility shift assay (EMSA) was used to examine the binding affinity of an isolated DNA binding domain of HMGB1, the HMGB1 b-box (HMGB1b), toward a set of PNA-DNA hybrid 4WJs. EMSAs showed that HMGB1b recognizes single-PNA hybrid 4WJs with a similar affinity to the DNA control. Circular dichroism (CD) spectroscopy was used to examine the structure and monitor thermal transitions of hybrid PNA-DNA four-way junctions in low (0.01mM Mg²⁺) and high (2.00mM Mg²⁺) ionic strength environments. CD analysis suggests a large deviation in helical structure between DNA and PNA hybrid junctions. Blunt-ended hybrid junctions b4WJ-PNA₁ and b4WJ-PNA₃ had higher melting temperatures (Tₘs) than their full-length counter parts with ∆Tₘs of 1.55 and 5.43°C, respectively. Junction b4WJ-PNA₃ was shown capable of binding
HMGB1b with an affinity similar to that of its parent DNA junction and has a $T_m$ of 41.18°C, 1.14°C higher than its parent DNA junction J1 and well above normal body temperature, suggesting that b4WJ-PNA3 may be a viable therapeutic agent for targeting HMGB1 \textit{in vivo}.

In the study presented in section 2 (Chapters V-VIII), the amino acid (AA) selectivity of aminoacyl-tRNA synthetases (aaRSs) was examined to investigate a possible role for aaRSs in genetic code development. A radiometric assay was used to perform an exhaustive survey of the 20 natural AAs vs. the 20 aaRSs from Escherichia coli (\textit{E. coli}). This study presents an AA chronology of Early vs. Late AAs based on misactivation frequency. Findings were compared with current theories on the evolution of AA recruitment and codification. The results of this study show a correlation between misactivation frequency and the order of AA codification, suggesting that aaRSs may have played a role in the process of AA codification.
ACKNOWLEDGMENTS

Firstly, I would like to express my sincere gratitude to my advisor, Dr. Anthony Bell, for the continuous support of my Ph.D. study and related research, for his patience, understanding, motivation, and knowledge. His guidance helped make my research and writing of this dissertation possible. I could not have imagined having a better advisor and mentor for my Ph.D. study.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Gordon Cannon, Dr. Vijay Rangachari, Dr. Karl Wallace, and Dr. Philip Bates for their insightful comments and encouragement, and for the hard questions which pushed me to widen my research from various perspectives.

My sincere thanks also goes to my lab mates who provided me an endless source of assistance, queries, and humor. Without their support, it would not be possible to conduct this research.
DEDICATION

I would like to dedicate this work to my mother, Bonnie Iverson. Without whom, I would, both literally and figuratively, not be here. She, more than anyone else, has helped shape me into the man I am today. My mother has been a high school algebra teacher for over 30 years and has touched the lives of thousands of students. I could not be more proud of the great work she has accomplished shaping the minds of young adults for over a generation. I only hope she is as proud of me as I am of her.
TABLE OF CONTENTS

ABSTRACT ................................................................................................................................. ii

ACKNOWLEDGMENTS ........................................................................................................... iv

DEDICATION ........................................................................................................................... v

LIST OF TABLES ...................................................................................................................... xi

LIST OF ABBREVIATIONS ....................................................................................................... xvi

CHAPTER I – EXAMINING HYBRID PNA-DNA 4WJS AS POTENTIAL THERAPEUTIC AGENTS TO TARGET THE PROINFLAMMATORY CYTOKINE HMGB1 .............................................................................................................................. 1

1.1 Synopsis and Hypotheses ................................................................................................. 1

1.2 High Mobility Group (HMG) Proteins ........................................................................... 3

1.2.1 Classification of HMG proteins ................................................................................. 4

1.2.2 High Mobility Group B1 (HMGB1) .......................................................................... 4

1.2.3 Classical and Alternate Functions of HMGB1 ......................................................... 5

1.2.4 HMGB1 and Disease Pathogenesis ........................................................................ 10

1.2.5 Strategy to Target HMGB1 .................................................................................... 12

1.3 DNA Four-Way Junctions ............................................................................................ 13

1.4 Nonstandard Nucleic Acid - Peptide Nucleic Acid (PNA) ........................................ 19

CHAPTER II - MATERIALS AND METHODS ..................................................................... 24

2.1 Polyacrylamide Gel Electrophoresis (PAGE) .............................................................. 24
2.1.1 Four-Way Junction Formation ................................................................. 24
2.1.2 Electrophoretic Mobility Shift Assays (EMSAs) ................................. 26
2.2 Circular Dichroism (CD) Analysis of Nucleic Acids ............................... 26
   2.2.1 CD Analysis of Classic Nucleic Acids .............................................. 26
   2.2.2 CD Analysis of PNA ........................................................................ 30
   2.2.3 CD Thermal Denaturation Analysis .................................................. 31
   2.2.4 CD Spectroscopy of 4WJs ................................................................. 32
CHAPTER III - RESULTS ............................................................................... 34
   3.1 PAGE Analysis of Native and Hybrid PNA-DNA 4WJs ......................... 34
   3.2 HMGB1 Binding .................................................................................... 39
   3.3 CD Analysis of Hybrid 4WJs ................................................................. 43
   3.4 Thermal Melting Curves and Derivation of Thermodynamic Parameters
       ........................................................................................................... 47
CHAPTER IV - DISCUSSION AND CONCLUSION ........................................ 60
   4.1 Discussion ............................................................................................ 60
   4.2 Conclusion ............................................................................................ 65
CHAPTER V – A POSSIBLE LINK BETWEEN AMINO ACID SELECTIVITY
OF AMINOACYL-TRNA SYNTHETASES AND GENETIC CODE
DEVELOPMENT .......................................................................................... 68
   5.1 Synopsis and Hypotheses ..................................................................... 68
5.2 Code Origin Theories ........................................................................................................... 70

5.2.1 Connections between AA Emergence and the Early Earth ................. 70

5.2.2 Connections between AA Emergence and the Coevolution Theory (CET) ........................................................................................................................................ 72

5.2.3 Connections between AA Emergence and AA Biosynthetic Pathways ........................................................................................................................................ 74

5.2.4 Connections between AA Emergence and AA Physiochemical Properties ........................................................................................................................................ 76

5.3 Aminoacyl tRNA-Synthetases (aaRSs) ................................................................. 77

5.3.1 Aminoacyl-tRNA Synthesis ....................................................................................... 77

5.3.2 aaRSs Role in Genetic Code Development ............................................................ 79

5.3.3 Classes of aaRSs ........................................................................................................ 80

5.3.3.1 Class I aaRSs ........................................................................................................ 81

5.3.3.2 Class II aaRSs ........................................................................................................ 83

5.3.4 Fidelity of aaRSs ....................................................................................................... 87

5.4 Radiometric Amino Acid Activation Assay .............................................................. 88

CHAPTER VI - MATERIALS AND METHODS ......................................................... 91

6.1 Radiometric Assays ...................................................................................................... 91

6.2 Measurement of Activation Reaction Rates ............................................................. 94

CHAPTER VII - RESULTS .............................................................................................. 95
7.1. Analysis of AA Activation Patterns Using Class I aaRSs: ....................... 95

7.1.1 Analysis of AA Activation Patterns Using ArgRS: ............................ 95
7.1.2 Analysis of AA Activation Patterns Using CysRS: ............................ 96
7.1.3 Analysis of AA Activation Patterns Using IleRS: .............................. 96
7.1.4 Analysis of AA Activation Patterns Using LeuRS: ............................ 99
7.1.5 Analysis of AA Activation Patterns Using MetRS: ......................... 101
7.1.6 Analysis of AA Activation Patterns Using TrpRS: ............................ 102
7.1.7 Analysis of AA Activation Patterns Using TyrRS: ............................ 103
7.1.8 Analysis of AA Activation Patterns Using ValRS: ............................ 104

7.2. Analysis of AA Activation Using Class II aaRSs: .......................... 106

7.2.1 Analysis of AA Activation Using AlaRS: ....................................... 106
7.2.2 Analysis of AA Activation Using AsnRS: ....................................... 107
7.2.3 Analysis of AA Activation Using AspRS: ....................................... 108
7.2.4 Analysis of AA Activation Using GlyRS: ....................................... 109
7.2.5 Analysis of AA Activation Using HisRS: ....................................... 110
7.2.6 Analysis of AA Activation Using LysRS: ....................................... 111
7.2.7 Analysis of AA Activation Using PheRS: ....................................... 112
7.2.8 Analysis of AA Activation Using ProRS: ....................................... 113
7.2.9 Analysis of AA Activation Using SerRS: ....................................... 114
7.2.10 Analysis of AA Activation Using ThrRS: ..................................... 115
LIST OF TABLES

Table 1 4WJ $T_m$s in low (0.01mM) and high (2.00mM) Mg$^{2+}$ .............................. 50
Table 2 Hybrid junctions vs. J1 in low (0.01mM) and high (2.00mM) Mg$^{2+}$ .......... 51
Table 3 Full-length junctions vs. blunt-ended junctions in low (0.01mM) and high (2.00mM) Mg$^{2+}$ .................................................................................................................. 52
Table 4 $\Delta H_{VH}$ of Hybrid junctions vs. J1 in low (0.01mM) and high (2.00mM) Mg$^{2+}$ ........................................................................................................................................ 57
Table 5 Prebiotic AAs produced in early Earth simulations ................................. 71
Table 6 AAs identified in carbonaceous meteorites ............................................. 72
Table 7 The Aminoacyl-tRNA Synthetases of E. coli .............................................. 81
Table 8 Known identity elements of E. coli tRNAs ................................................. 86
Table 9 AA Groups used in the activation analysis ................................................. 92
Table 10 Summary of class I aaRS activation ....................................................... 117
Table 11 Summary of class II aaRS activation ...................................................... 118
Table 12 Early vs. Late AAs according to class I aaRS misactivation ............ 119
Table 13 Early vs. Late AAs according to class II aaRS misactivation ............. 119
Table 14 AA Chronologies based on early Earth/meteorite studies, Physiochemical Properties, and Biosynthetic Pathways ................................................. 121
Table 15 Consensus AA Chronology Based on Class I and II aaRS Activation 121
LIST OF ILLUSTRATIONS

Figure 1. HMGB1 a-box and b-box Structures.......................................................... 5
Figure 2. HMGB1 sequence from human ................................................................ 7
Figure 3. Relationship between HMGB1 function and posttranslational
modifications............................................................................................................. 8
Figure 4. HMGB1 extra nuclear interactions............................................................. 10
Figure 5. HMGB1 and disease .................................................................................. 11
Figure 6. Holliday junctions in genetic recombination.............................................. 14
Figure 7. Model immobile 4WJ J1 ......................................................................... 15
Figure 8. Open-X vs. stacked-X conformational isomers......................................... 16
Figure 9. X-ray crystal structures of synthetic 4WJs............................................... 17
Figure 10. 4WJ Isomerization ................................................................................ 18
Figure 11. DNA sugar phosphate backbone vs. PNA AEG backbone ................. 20
Figure 12. Schematic of DNA/PNA duplex ............................................................ 21
Figure 13. NMR/X-ray Structures of Nucleic Acid Duplexes................................. 23
Figure 14. Differences in the endo pucker of DNA deoxyribose sugar A- and B-
DNA structures....................................................................................................... 27
Figure 15. Example DNA duplex structures............................................................ 28
Figure 16. Diagram of Hybrid 4WJs........................................................................ 35
Figure 17. Nondenaturing Polyacrylamide Gel of Hybrid 4WJs............................. 36
Figure 18. Formation Controls for 4WJ-PNA\textsubscript{1} and b4WJ-PNA\textsubscript{1}.................... 37
Figure 19. Formation Controls for 4WJ-PNA\textsubscript{3} and b4WJ-PNA\textsubscript{3}..................... 38
Figure 20. Formation Controls for 4WJ-PNA\textsubscript{1,3} and b4WJ-PNA\textsubscript{1,3}...................... 38
Figure 21. EMSA of HMGB1b binding ................................................................. 42
Figure 22. CD spectra of 4WJs in low (~0.01mM) and high (2mM) Mg$^{2+}$ .......... 44
Figure 23. Apparent fraction of B- and A-form DNA ........................................ 45
Figure 24. Thermal melting curves of 4WJs in low versus high Mg$^{2+}$ .......... 49
Figure 25. First derivatives of 4WJ melting curves ............................................. 55
Figure 26. Schematic of the effects of changing $J_{twist}$ angle on HMGB1 binding. ................................................................................................................................. 62
Figure 27. Space filling 3D models of B-DNA, A-B intermediate DNA, and A-DNA ................................................................................................................................. 64
Figure 28. A schematic of the aaRS catalyzed aminoacylation reaction ............ 69
Figure 29. Codon evolution map proposed in the coevolutionary theory .......... 74
Figure 30. AA biosynthetic pathways .................................................................. 76
Figure 31. Aminoacyl-tRNA synthesis ................................................................. 78
Figure 32. Crystal structures of class I and II aminoacyl-tRNA synthetases ...... 81
Figure 33. Class I Rossmann Fold of GlnRS from E.coli ................................. 82
Figure 34. Class II Beta Fold of AspRS from E. coli .......................................... 83
Figure 35. An X-ray crystal structure of tRNA$^{Phe}$ from yeast ......................... 84
Figure 36. Possible pathways of the activation reaction ................................. 89
Figure 37. Radiometric activation assay ............................................................ 90
Figure 38. Activation scheme for alanyl-synthetase (AlaRS) ............................ 93
Figure 39. TLC results and forward reaction curves for AlaRS ....................... 93
Figure 40. Forward reaction curves of ArgRS .................................................... 95
Figure 41. Forward reaction curves of CysRS .................................................... 96
Figure 42. Forward reaction curves of IleRS with Val, Met, and Thr .................. 97
Figure 43. Forward reaction curves of IleRS with Glu, Leu, Cys, and Ala. ....... 98
Figure 44. Forward reaction curves of IleRS with Glu, Leu, Cys, Ala, and Neg control (zoomed view) without the Pos control ........................................ 98
Figure 45. Forward reaction curves of LeuRS.............................................. 99
Figure 46. Forward reaction curves of LeuRS (zoomed view) ...................... 100
Figure 47. Forward reaction curves of MetRS............................................. 101
Figure 48. Forward reaction curves of TrpRS............................................. 102
Figure 49. Forward reaction curves of TyrRS........................................... 103
Figure 50. Forward reaction curves of ValRS with Ile, Leu, Asp, Cys, and Ala 104
Figure 51. Forward reaction curves of ValRS with Ser, Glu, Thr, Met, and Gly 105
Figure 52. Forward reaction curves of AlaRS............................................. 106
Figure 53. Forward reaction curves of AsnRS........................................... 107
Figure 54. Forward reaction curves of AspRS........................................... 108
Figure 55. Forward reaction curves of GlyRS........................................... 109
Figure 56. Forward reaction curves of HisRS........................................... 110
Figure 57. Forward reaction curves of LysRS........................................... 111
Figure 58. Forward reaction curves of PheRS.......................................... 112
Figure 59. Forward reaction curves of ProRS.......................................... 113
Figure 60. Forward reaction curves of SerRS.......................................... 114
Figure 61. Forward reaction curves of ThrRS.......................................... 115
Figure 62. Forward reaction curves of ThrRS (zoomed view) .................... 116
Figure 63. AA misactivation vs. AA biosynthetic pathways ..................... 123
Figure 64. AA misactivation vs. CET ................................................................. 124
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMG</td>
<td>High mobility group protein</td>
</tr>
<tr>
<td>HMGB1</td>
<td>High mobility group box protein 1</td>
</tr>
<tr>
<td>PNA</td>
<td>Peptide nucleic acid</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>4WJ</td>
<td>Four-way junction</td>
</tr>
<tr>
<td>EMSA</td>
<td>Electrophoretic mobility shift assay</td>
</tr>
<tr>
<td>HMGB1b</td>
<td>High mobility group box protein 1 b-box</td>
</tr>
<tr>
<td>CD</td>
<td>Circular dichroism</td>
</tr>
<tr>
<td>AFa</td>
<td>Apparent fraction of A-form helical character</td>
</tr>
<tr>
<td>AFB</td>
<td>Apparent fraction of B-form helical character</td>
</tr>
<tr>
<td>Tm</td>
<td>Melting temperature</td>
</tr>
<tr>
<td>aaRS</td>
<td>Aminoacyl-tRNA synthetase</td>
</tr>
<tr>
<td>AA</td>
<td>Amino acid</td>
</tr>
<tr>
<td>ECM</td>
<td>Extra-cellular matrix</td>
</tr>
<tr>
<td>HMGA</td>
<td>High mobility group AT-hook protein</td>
</tr>
<tr>
<td>HMGN</td>
<td>High mobility group nucleotide binding protein</td>
</tr>
<tr>
<td>HTH</td>
<td>Helix-turn-helix</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------------------------------------</td>
</tr>
<tr>
<td>DAMP</td>
<td>Damage associated molecular pattern molecule</td>
</tr>
<tr>
<td>CXC</td>
<td>Cysteine-X-cysteine</td>
</tr>
<tr>
<td>CXCLR12</td>
<td>Chemokine (C-X-C Motif) Ligand 12</td>
</tr>
<tr>
<td>RAGE</td>
<td>Receptor for advanced glycation products</td>
</tr>
<tr>
<td>TLR</td>
<td>Toll like receptor</td>
</tr>
<tr>
<td>NF-κB</td>
<td>Nuclear factor kappa-light-chain-enhancer of activated B cells</td>
</tr>
<tr>
<td>CXCR4</td>
<td>Chemokine (C-X-C Motif) Receptor 4</td>
</tr>
<tr>
<td>ROS</td>
<td>Reactive oxygen Species</td>
</tr>
<tr>
<td>DN</td>
<td>Diabetic neuropathy</td>
</tr>
<tr>
<td>SLE</td>
<td>Systemic lupus erythematosus</td>
</tr>
<tr>
<td>K_d</td>
<td>Dissociation constant</td>
</tr>
<tr>
<td>HR</td>
<td>Homologous recombination</td>
</tr>
<tr>
<td>DSB</td>
<td>Double strand break</td>
</tr>
<tr>
<td>SDSA</td>
<td>Synthesis dependent strand annealing</td>
</tr>
<tr>
<td>DSBR</td>
<td>Double strand break repair</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>FRET</td>
<td>Fluorescence energy transfer</td>
</tr>
<tr>
<td>AEG</td>
<td>Aminoethylglycine</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
</tr>
<tr>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>3WJ</td>
<td>Three-way junction</td>
</tr>
<tr>
<td>Lys</td>
<td>L-lysine</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>$\Delta H_{VH}$</td>
<td>van't Hoff's enthalpy</td>
</tr>
<tr>
<td>$C$</td>
<td>Concentration</td>
</tr>
<tr>
<td>$R$</td>
<td>Ideal gas constant</td>
</tr>
<tr>
<td>$T$</td>
<td>Temperature</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Fraction of strands in junction form</td>
</tr>
<tr>
<td>$B$</td>
<td>Tetramolecular association constant</td>
</tr>
<tr>
<td></td>
<td>(full-width)</td>
</tr>
<tr>
<td>$B'$</td>
<td>Tetramolecular association constant</td>
</tr>
<tr>
<td></td>
<td>(half-width)</td>
</tr>
<tr>
<td>$P/J$</td>
<td>Protein to junction ratio</td>
</tr>
<tr>
<td>$MD$</td>
<td>Molecular dynamics</td>
</tr>
<tr>
<td>$AA$-AMP</td>
<td>Aminoacyladenylate</td>
</tr>
<tr>
<td>$aa$-tRNA</td>
<td>Aminoacyl-tRNA</td>
</tr>
<tr>
<td>$ATP$</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>$AMP$</td>
<td>Adenosine monophosphate</td>
</tr>
<tr>
<td>$PP_i$</td>
<td>Diphosphate</td>
</tr>
<tr>
<td>$CET$</td>
<td>Coevolutionary theory</td>
</tr>
<tr>
<td>$CP$</td>
<td>Connective peptide</td>
</tr>
<tr>
<td>$mnm^5s^2U$</td>
<td>5-methylamino-methyl 2-thiouridine</td>
</tr>
<tr>
<td>$t^6A$</td>
<td>Threonylcarbamoyl adenosine</td>
</tr>
<tr>
<td>Proc</td>
<td>Full Name</td>
</tr>
<tr>
<td>------</td>
<td>----------------------</td>
</tr>
<tr>
<td>TLC</td>
<td>Thin layer chromatography</td>
</tr>
<tr>
<td>Gly</td>
<td>Glycine</td>
</tr>
<tr>
<td>Ala</td>
<td>L-alanine</td>
</tr>
<tr>
<td>Val</td>
<td>L-valine</td>
</tr>
<tr>
<td>Asp</td>
<td>L-aspartic acid</td>
</tr>
<tr>
<td>Pro</td>
<td>L-proline</td>
</tr>
<tr>
<td>Ser</td>
<td>L-serine</td>
</tr>
<tr>
<td>Glu</td>
<td>L-glutamic acid</td>
</tr>
<tr>
<td>Leu</td>
<td>L-leucine</td>
</tr>
<tr>
<td>Thr</td>
<td>L-threonine</td>
</tr>
<tr>
<td>Arg</td>
<td>L-arginine</td>
</tr>
<tr>
<td>Asn</td>
<td>L-asparagine</td>
</tr>
<tr>
<td>Gln</td>
<td>L-glutamine</td>
</tr>
<tr>
<td>Ile</td>
<td>L-isoleucine</td>
</tr>
<tr>
<td>Cys</td>
<td>L-cystine</td>
</tr>
<tr>
<td>His</td>
<td>L-histidine</td>
</tr>
<tr>
<td>Phe</td>
<td>L-phenylalanine</td>
</tr>
<tr>
<td>Met</td>
<td>L-methionine</td>
</tr>
<tr>
<td>Tyr</td>
<td>L-tyrosine</td>
</tr>
<tr>
<td>Trp</td>
<td>L-tryptophan</td>
</tr>
</tbody>
</table>
CHAPTER I – EXAMINING HYBRID PNA-DNA 4WJS AS POTENTIAL THERAPEUTIC AGENTS TO TARGET THE PROINFLAMMATORY CYTOKINE HMGB1

1.1 Synopsis and Hypotheses

Inflammation is part of the body’s response to injury or invasion and is a necessary part of the innate immune system. However, uncontrolled inflammation can cause damage to healthy tissue and exacerbate certain diseases and disorders. The objective of the studies presented in Chapters I through IV is to focus on the design and development of nucleic acid four-way junctions (4WJs) to target the proinflammatory cytokine, HMGB1 in an effort to reduce symptoms of inflammation. To date, there is no commercially available treatment for HMGB1-linked disorders. The rational for selecting 4WJs as ligands against HMGB1 is based on the multi-functional nature of the protein. HMGB1 is an abundant DNA-binding protein that mediates nuclear homeostasis.\(^1\)\(^-\)\(^7\) It is now clear that HMGB1 also functions in the extracellular matrix (ECM) as a cytokine that initiates unintended leukocyte production.\(^5\)\(^,\)\(^8\)\(^-\)\(^{19}\) Due to the high cellular abundance of HMGB1 (10\(^6\) copies per cell), it is perhaps not surprising that it is considered to be a biomarker and target for a variety of diseases and immune disorders such: atherosclerosis\(^8\)\(^,\)\(^{18}\)\(^,\)\(^{19}\), lupus\(^10\)\(^,\)\(^{16}\)\(^,\)\(^{18}\), rheumatoid arthritis\(^{16}\)\(^,\)\(^{19}\)\(^,\)\(^{20}\), and sepsis\(^{21}\)\(^,\)\(^{22}\). Consequently, a great deal of attention is focused toward targeting HMGB1 in efforts to reduce the harmful effects of inflammation associated with these conditions.
There is a need to develop novel therapeutics that target HMGB1 to treat chronic inflammation through pathways that attenuate damage signals without compromising the innate immune system. I hypothesize that immobile 4WJs that mimic natural intracellular ligands of HMGB1, can: i) effectively target the protein and ii) block harmful immune reactions. With a view to enhancing the stability of 4WJs, certain DNA strands of 4WJs are replaced with the nonstandard nucleic acid - peptide nucleic acid (PNA). I hypothesize that hybrid PNA-DNA 4WJs possess enhanced thermal stability compared to pure DNA 4WJs. Inflammation is often associated with fever. Therefore, any therapeutic designed to treat inflammation will have to remain viable at elevated body temperatures (>37°C). The rationale for choosing PNA as a DNA analogue is based on studies that show duplex, triplex, and quadruplex structures composed of PNA-DNA have Tₘs up to 30% higher than pure DNA structures (69°C and 54°C for 15bp PNA-DNA hybrid and DNA duplexes, respectively).²³⁻²⁶

Chapter I of the document focuses on the background of HMGB1, 4WJs, and PNA. The subsequent chapters focus on the design, development, and structure of hybrid PNA-DNA 4WJs vs. the DNA 4WJ control, J1. Next, the binding affinity of HMGB1b toward each hybrid 4WJ is evaluated. These data are followed by a more detailed investigation of the conformational stability of each hybrid 4WJ. Finally, the discussion presents hypotheses that describe the possible connections between the structure of each hybrid 4WJ and its protein recognition properties.
An electrophoretic mobility shift assay (EMSA) was used to examine the binding affinity of HMGB1 toward the multi-PNA and blunt-ended hybrid junctions. HMGB1b recognizes single-PNA hybrid 4WJs with similar affinity to the DNA control, J1. All single-PNA hybrid junctions form 4:1 protein to junction complexes at concentrations similar to that of J1. HMGB1b recognizes multi-PNA hybrids but does not form stable 4:1 protein to junction complexes, as represented by a single band in the absence of intermediate with the multi-PNA substrates. CD spectroscopy was used to examine the structure and monitor thermal transitions of hybrid PNA-DNA four-way junctions in low (0.01mM Mg\(^{2+}\)) and high (2mM Mg\(^{2+}\)) ionic strength environments.

1.2 High Mobility Group (HMG) Proteins

HMG proteins are “architectural” nuclear proteins and the most abundant class of non-histone nuclear proteins. Architectural DNA-binding proteins, unlike transcription factors (TFs), do not bind in a sequence specific manner.\(^{27-29}\) Architectural nuclear proteins bind preferentially to certain structural DNA motifs. One primary function of HMG proteins is to remodel chromatin. During this process, HMG proteins bind/bend the topology of DNA within chromatin to facilitate the binding of TFs to specific DNA sequences.\(^{1,28,30-32}\) TF binding interactions in turn initiate gene expression. Hence, HMG proteins are generally associated with gene expression vs. repression.\(^{1,3,27,33}\) HMG proteins also play an expanded role in additional nuclear processes such as genetic recombination and DNA repair.\(^{33,34}\)
1.2.1 Classification of HMG proteins

HMG proteins are split into three categories: HMG\textsubscript{A}, HMG\textsubscript{B}, and HMG\textsubscript{N}. Each protein class is distinguished by their respective DNA recognition motifs\textsuperscript{1,28,33-36}. HMG\textsubscript{A} proteins contain three binding domains called AT-hook motifs\textsuperscript{35,36}. The AT-hook sequences of HMG\textsubscript{A} proteins possess a conserved PRGRP amino acid sequence. Two of the three AT-hooks are involved in DNA recognition and binding\textsuperscript{35}. AT-hook motifs recognize the minor groove of duplex DNA and induce bending\textsuperscript{1,36}. HMG\textsubscript{N} proteins affect the post-translational modifications (i.e. acetylation and phosphorylation) of core histones and compete with linker histone H1 to regulate chromatin structure\textsuperscript{34,37}. HMG\textsubscript{N} is also active in DNA lesion repair\textsuperscript{33,34,38}. HMG\textsubscript{N} proteins possess the conserved sequence RRSARLSA that serves as the DNA-binding region\textsuperscript{38}. The HMGB family of proteins includes seven members: HMGB1, HMGB2, HMGB3, HMGB4, HMG1, HMG2, and HMG2a\textsuperscript{28}. HMGB proteins’ functional DNA-binding motifs are referred to as “box” domains. An HMG box domain is composed of three alpha helices connected by short loops to form a HTH motif\textsuperscript{1,6,28}.

1.2.2 High Mobility Group B1 (HMGB1)

HMGB1 is the most highly abundant ($10^6$ copies/cell) of HMGB proteins. HMGB1 is composed of 215 amino acids that are arranged into three subunits, two DNA-binding “box” domains and an acidic C-terminus\textsuperscript{29}. The a- and b-box subunits from HMGB1 are displayed in Fig. 1 A and B, respectively\textsuperscript{39,40}. DNA binding occurs on the concave side of the L-shaped structure\textsuperscript{39}. This hydrophobic face interacts with the minor groove of DNA. Two residues within the b-box,
Phe106 and Lys126 (shown in green in Fig. 1B), intercalate between DNA strands. This induces an approximate 90° bend in the bound DNA strand.\textsuperscript{29,39} The a-box has only a single intercalating residue, Phe38, shown in green in Fig. 1A. The single intercalating residue of the a-box produces a less dramatic bend in DNA of approximately 60°.\textsuperscript{29,40}

Figure 1. HMGB1 a-box and b-box Structures

(A) Solution NMR structure of HMGB1 A-box domain from human highlighting DNA intercalating residue Phe38 [PDB: 2RTU]\textsuperscript{40}. (B) Solution NMR structure HMGB1 B-box domain from rat highlighting DNA intercalating residues Phe106 and Lys126 [PDB: 1HME]\textsuperscript{39}.

1.2.3 Classical and Alternate Functions of HMGB1

Like other HMG proteins, HMGB1 was initially classified as a nuclear protein. In the nucleus, HMGB1 binds preferentially to bent and cruciform DNA (i.e. 4WJs) to control homeostatic functions such as chromatin remodeling and
genetic recombination and repair.\textsuperscript{2,41,42} With respect to chromatin remodeling, HMGB1 binds DNA to induce bends that alter the topology of DNA.\textsuperscript{27,43} The bent topology of DNA enhances TF binding at gene promoter regions. Moreover, HMGB1 (like HMGN proteins) can displace histones from bent DNA and relax structural constraints for transcription factor binding.\textsuperscript{3,43}

Extracellular HMGB1 was initially discovered as a 30-kDa heparin-binding protein in the developing brains of rats, and termed "amphoterin" before being identified as HMGB1.\textsuperscript{5,11,44} HMGB1 was shown to be required for neuronal outgrowth.\textsuperscript{44-46} Antibodies against HMGB1 were shown to inhibit neuronal outgrowth in cell cultures.\textsuperscript{44,46} HMGB1 is localized in the filopodia extending from the cell bodies of neurons and regulates outgrowth of axons and dendrites by promoting adhesion of the growing membrane edges.\textsuperscript{44} HMGB1 was further found to bind RAGE on the surface of rat embryonic cortical neurons, and is co-expressed with RAGE in developing rat neuroepithelium.\textsuperscript{46}

HMGB1 has since been implicated in a variety of alternate cellular processes such as: neuronal outgrowth, tumor proliferation, platelet activation, cell adhesion, and as a mediator of immune response.\textsuperscript{5,9,12,15,17,30,31,41,44,46-54} The role of HMGB1 in activation of immune response(s) has gained arguably the majority of attention in recent years. With respect to the specific role of HMGB1 in immunity, HMGB1 acts as a redox-sensitive damage-associated molecular pattern molecule (DAMP).\textsuperscript{17,18,49,53} HMGB1 can be released either actively by immune cells or passively by damaged/necrotic cells (Fig. 3A).\textsuperscript{17,48} Active
HMGB1 release can be triggered by stimulation of macrophages and monocytes. Passive release is normally associated with cellular injury.\textsuperscript{17,48,55}

As stated earlier, posttranslational modifications influence the function of HMGB1.\textsuperscript{5,9,18,48,52,56,57} The sequence for human HMGB1 is shown in Fig. 2. HMGB1 can also function in the ECM as a damage associated molecular pattern molecule (DAMP). In this capacity the oxidation state and posttranslational modifications of the protein are critical.\textsuperscript{40,48,52} The key residues that undergo modification are highlighted in Fig. 2 and include: Lys2, Lys11, Lys28, Lys29, and Lys30 (acetylation sites highlighted maroon); Cys23, Cys45, and Cys106 (oxidation sites highlighted in dark blue); Ser35, Ser39, Ser42, Ser46, and Ser53 (phosphorylation sites highlighted in grey); and Lys43 (methylation site highlighted in light blue).\textsuperscript{5,40,48}

Figure 2. HMGB1 sequence from human

The A-box shown in red text, B-box in blue text, and the acidic tail in green text. Acetylation sites (maroon), oxidation sites (dark blue), DNA-binding (green), methylation (light blue) and phosphorylation (grey).

Acetylation of lysine residues in the A-box eliminates electrostatic interaction of these residues with DNA and impedes binding.\textsuperscript{56} Acetylation of these lysine residues in isolated A-boxes abolishes binding to DNA 4WJs. Upon
acetylation, the full-length protein retains its ability to bind 4WJs through interactions with the B-box. In macrophages and monocytes, acetylation of lysine residues in the A-box precedes cytosolic translocation of HMGB1. In macrophages and monocytes, acetylation of lysine residues in the A-box precedes cytosolic translocation of HMGB1. Phosphorylation then leads active secretion to the extracellular matrix by lysosomes and secretion vesicles. In neutrophiles extra cellular transport is induced through methylation of Lys42.

Figure 3. Relationship between HMGB1 function and posttranslational modifications

The function of HMGB1 is dictated by its oxidation state (Fig. 3B). HMGB1 is found in the nucleus in its fully reduced form. In damaged cells, HMGB1 localizes within the cytosol. HMGB1 acts as a chemo-attractant in its fully reduced form, enhancing autophagy and promoting lysosomal degradation of damaged cellular components through interaction with the Beclin1-Bcl2 complex. Under oxidative conditions, Cys23 and Cys45 of the A-domain form a disulfide bond. In the extracellular space disulfide-HMGB1 acts as a proinflammatory cytokine. In this capacity, HMGB1 forms complexes with or binds directly to immune receptors (Fig. 4) such as: chemokine (C-X-C Motif) Ligand 12 (CXCL12), receptor for advanced glycation end products (RAGE), and toll-like receptor (TLR)-4 to activate the nuclear factor kappa-B (NF-κB) pathway and control leukocyte migration and activation. The complex formed between HMGB1 and CXCL12 increases the affinity of CXCL12 for its receptor CXCR4 and promotes chemotaxis. Binding of HMGB1 to RAGE and TLR4 activates the NF-κB pathway.

Fully oxidized HMGB1 loses its immune function. The oxidation of Cys106 has been shown to promote relocalization from the nucleus to the cytoplasm in apoptotic cells and promotes immune tolerance. This process is promoted by cellular reactive oxygen species (ROS). Caspase activation targets the mitochondria to produce ROS during apoptosis. Cys106 is oxidized by these mitochondrial ROS eliminating the proinflammatory activity of HMGB1. This process has been shown to be necessary for neutralizing HMGB1 as a danger signal and promoting immune tolerance of apoptotic cells.
1.2.4 HMGB1 and Disease Pathogenesis

HMGB1 is gathering a great deal of attention as a disease biomarker and therapeutic target. A growing number of studies link unintended HMGB1 signaling with pathogenesis in a number of diseases and immune disorders (Fig. 5) such as: atherosclerosis\(^8,18\), diabetes\(^62,63\), lupus\(^10,18\), rheumatoid arthritis\(^16,18,20\), and sepsis\(^21,22,53\). HMGB1 contributes to atherosclerosis by promoting vascular inflammation and lesions.\(^8\) HMGB1 is found at elevated levels in the blood plasma of patients with type 2 diabetes and is up regulated in response to high glucose levels.\(^62-64\) HMGB1 promotes the production of proinflammatory cytokines through activation of the NF-κB signaling pathway exacerbating conditions such as diabetic neuropathy (DN).\(^62\)

HMGB1 expression was shown to correlate with the severity of symptoms of systemic lupus erythematosus (SLE) in mice.\(^10\) HMGB1 was shown to
exacerbate renal pathology by enhancing macrophage inflammatory response through activation of RAGE.\textsuperscript{10} Active HMGB1 release has been observed at sites of joint inflammation in mice with arthritis. Injection of HMGB1 into the joints of healthy mice induced arthritis in these locations.\textsuperscript{16,20}

During the late phase of sepsis, HMGB1 release activates the coagulation system and results in disseminated intravascular coagulation.\textsuperscript{22} HMGB1 secreted from dying cells induces the release of additional HMGB1 and other cytokines from macrophages and other cell types.\textsuperscript{22,50} HMGB1 has also been shown to mediate acute lung injury and acute kidney injury during sepsis through activation of the NF-κB pathway.\textsuperscript{21,22}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5}
\caption{HMGB1 and disease}
\end{figure}

1.2.5 Strategy to Target HMGB1

HMGB1 provides a target for novel therapeutics to treat chronic inflammation through pathways that attenuate damage signals without compromising the innate immune system. In mouse models with diabetes, HMGB1 inhibition reduced the up regulation of pro-inflammatory cytokines in response to high glucose.\textsuperscript{62,63} In mouse models, antibodies against HMGB1 have been shown to alleviate inflammation associated with rheumatoid arthritis and prevent the progression of arthritis.\textsuperscript{20}

Cell and animal model studies have found many pharmacologically active compounds with the ability to suppress immune response by blocking HMGB1, though many of these compounds were expensive, showed cytotoxicity at required dosages, or have harsh side effects.\textsuperscript{12,20,49} Our strategy is to develop cruciform nucleic acid structures (i.e. 4WJs) composed of DNA and PNA as ligands against HMGB1. These hybrid PNA-DNA 4WJs mimic the natural DNA 4WJ ligands of HMGB1 in its role in genetic recombination. Several in vitro studies show that HMG proteins bind 4WJs with binding constants at nM range.\textsuperscript{66,67} Small (32bp – 28bp) 4WJs were developed to avoid immunogenicity associated with larger DNA sequences.\textsuperscript{15,65} 4WJs are also able to bind HMGB1 in 4:1 protein to junction complexes offering a distinct advantage over the single binding sites of duplex structures. The hybrid 4WJs are based on the model immobile DNA 4WJ, J1. J1 is composed of four asymmetric DNA strands that ensure the lattice of the molecule is immobilized or “locked”. The long-term objective is to reassign 4WJs to the ECM to target HMGB1. 4WJs are intended to
sequester HMGB1 and block unintended binding interactions between HMGB1 and immune receptors that initiate immune dysfunction. With respect to unnatural substrates, PNA are used to replace DNA sequences within the lattice of J1 to enhance the stability of the resulting hybrid DNA-PNA 4WJs.

We suspect that the binding affinity of HMGB1 toward 4WJs ($K_d \approx 10^{-160}$ nM) is strong enough to prevent HMGB1 binding to immune receptors.\textsuperscript{66,67} HMGB1 binds receptors such as RAGE and toll-like receptor 4 (TLR-4) with binding constants that range from $K_d \approx 10.0$ - $100.0$ nM).\textsuperscript{46} Inhibition of HMGB1 by hybrid 4WJs is hypothesized to alleviate inflammation by scavenging HMGB1 from the ECM. This approach is designed to remove/prevent the interaction of HMGB1 with RAGE and TLRs to effectively shut off immune responses.

1.3 DNA Four-Way Junctions

DNA 4WJs, also known as Holliday junctions, are key intermediate for genetic recombination and double strand break repair.\textsuperscript{68,69} These processes are essential to promoting genetic diversity and maintaining genomic stability. One of the most extensive networks of Holliday junctions is found in the homologous recombination (HR) steps of meiosis illustrated in Fig. 6. In the initial stage of meiotic HR, double strand breaks (DSBs) are produced at numerous chromosomal DNA sites to produce single strand 3’ overhangs. Next, DNA recombinases bind the 3’ overhangs and search for an intact homologous dsDNA sequence. The recombinases then promote strand invasion by the ssDNA ends to form displacement loops. The initial strand invasion intermediates can be processed by one of two pathways: rejoining of the broken strands by synthesis.
dependent strand annealing (SDSA) or by double strand break repair (DSBR).\textsuperscript{69,70} The SDSA pathway results in displacement of the invading strand and leads to non-crossover products. The DSBR pathway is the result of a more stable strand invasion intermediate and leads to a second end capture. This produces a double 4WJ that is eventually resolved to allow chromosome segregation.\textsuperscript{68-72}

![Figure 6: Holliday junctions in genetic recombination.](image)

Branch migration is required in genetic recombination in order to produce crossover products. However, branch migration makes it difficult to probe the structural dynamics of 4WJs. Hence, a great deal of effort was put forth to assemble locked or “immobilized” 4WJs that form stable conformational isomers.\textsuperscript{73,74} An algorithm developed by Seeman and Kallenbach, designed to minimize sequence symmetry to ensure that the junction lattice is immobilized,
was used to devise a series of immobile 4WJs. The sequence identity of each strand of J1 minimizes symmetry (to ensure that the junction lattice is immobilized), while maintaining thermodynamic stability of the junction. The algorithm is based on a set of rules that prevents: i) the repeat of four consecutive nucleotides in any two strands and ii) the presence of complimentary sequences in an arm and branch point. The sequence of the immobilized junction J1 is displayed in Fig. 7.

![Diagram of J1](image)

**Figure 7. Model immobile 4WJ J1**

Individual strands are labeled as 101, 102, 103, AND 104, respectively.

Subsequent studies using a variety of biophysical techniques such as: gel electrophoresis, chemical foot printing, enzymatic foot printing (protection assays), fluorescence resonance energy transfer (FRET), nuclear magnetic resonance (NMR), and X-ray studies reveal that the global conformation of 4WJs fluctuate between extended or "open" structures and compact "stacked" structures. The extended conformer, referred to as open-X, has four-fold
symmetry (90° rotational symmetry) with all four arms pointing away from the center (Fig. 8A). The compact or stacked-X isomers shown in Fig. 8B and C possess two-fold symmetry (180° rotational symmetry) and can populate two conformational isomers. Each stacked-X isomer contains two continuous strands and two crossover strands. The continuous strands of the stacked-X isomers have single axes throughout their respective helices. The crossover strands switch helices at the branch or strand exchange point. Studies of J1 have shown a preference for one of the isomers based on the sequences flanking the branch point.  

![Figure 8. Open-X vs. stacked-X conformational isomers](image)

(A) The open conformer showing fully extended arms. (B) and (C) display the I/II and III/IV stacked-X conformers, respectively.

The structure of 4WJJs is highly dependent upon the concentration of cations such as Na⁺ and Mg²⁺. Fig. 9 shows X-ray crystal structures of two
synthetic immobile 4WJs in the open-X (Fig. 9A)\textsuperscript{86}, and in the stacked-X conformation (Fig. 9B).\textsuperscript{87} In the absence of cations, electrostatic repulsion between phosphates causes the helical arms to extend into the open-X conformation (Fig. 9A). When the concentration of Mg\textsuperscript{2+} $\geq$ 100 $\mu$M, the charge repulsion between negatively charged phosphate backbones is greatly reduced promoting the transition from an open conformer to a stacked-X conformer.\textsuperscript{80,82,88}

Figure 9. X-ray crystal structures of synthetic 4WJs

(A) X-ray crystal structure of a DNA 4WJ in an open conformation at a resolution of 2.8 Å [PDB: 1XNS]\textsuperscript{86}. (B) X-ray crystal structure of a DNA 4WJ in a stacked-X conformation at a resolution of 1.94 Å [PDB: 4GQD]\textsuperscript{87}.

The current studies focus on the immobilized 4WJ, J1. Electrophoresis experiments confirmed that a 1:1:1:1 ratio of each strand of J1 forms a stable complex.\textsuperscript{73,74} CD and NMR analyses revealed that 4WJs (including J1) posses B-DNA helical structure.\textsuperscript{74,82,88,89} With respect to local heterogeneity within the junction lattice, J1 (Fig 10A) was shown to prefer a stacking arrangement of arms I/II - III/IV with strands 101 and 103 as the continuous strands and 102 and 104 as the crossover strands analogous to Fig. 9B.\textsuperscript{66,74,76,90} Examination of a J1 derivative, J2, with the CG and AT pairs at the branch point of arms III and IV swapped revealed a much weaker preference for the I/II - III/IV isomer.\textsuperscript{89,90} Under
conditions where J1 exists solely in the I/II - III/IV stacked-X isomeric form, J2 exists as a 5:1 mixture of the I/II - III/IV and I/IV - II/III isomers shown in Fig. 10B. Fig. 10C shows another J1 derivative, J_EPS, with AT pairs replacing the GC pairs at the ends of the junction arms. Studies utilizing FRET analysis revealed that J_EPS presented the same preference for the I/II - III/IV stacking conformation as J1. It can be concluded from these studies that the bases at the ends of the junction arms have little effect on the stacking arrangement while the sequence near the branch point is a strong factor in determining the preference for and stability of a particular structural isomer.

Figure 10. 4WJ Isomerization

Each junction is shown in the open-X conformation at the top. The preferred stacked-X isomer is shown at the bottom. (A) J1 shown to the left with a single preferred isomer. (B) J2 shown in the middle has no clear preference for a specific stacked-X isomer. (C) J_EPS shown to the right with its preferred stacked-X isomer.
Therapeutics based on nucleic acids have shown moderate success in targeting HMGB1. One study utilized DNA beads coated with DNA oligomers to sequester HMGB1 in the intestines of mice after induction acute colitis. Oral administration of the beads was shown to sequester HMGB1 from the gut and reduce levels in serum samples. This study determined that larger DNA constructs (i.e. 4WJs) show greater affinity for HMGB1 in vitro. However, the 4WJ constructs did not bind appreciably at low pH. I hypothesize that modifying the 4WJ lattice by inserting more chemically stable nucleic acid analogs (i.e. PNA) will produce enhanced in vivo stability. PNAs offer several properties that make them ideal candidates for use as DNA analogs in 4WJs. PNAs form stable hybrid structures with DNA through Watson-Crick base pairing. The global conformations of PNA-DNA duplexes and 4WJs largely match the global conformations of corresponding DNA structures. However, PNA-DNA structures do show changes in local helical conformation compared to DNA structures.

The uncharged PNA backbone provides decreased charge repulsion between strands, increasing the thermal stability of nucleic acid structures containing PNA strands. This is evidenced by an increase in $T_m$s of PNA-DNA hybrid duplexes over their pure DNA counterparts by up to 30%. The peptide backbone of PNA also provides nuclease resistance to hybrid structures.

1.4 Nonstandard Nucleic Acid - Peptide Nucleic Acid (PNA)

PNAs are nucleic acid analogs that are composed of polyamide linkages vs. sugar/phosphate bonds. PNAs were originally studied as gene-targeting drugs that were designed to form complementary base pairs within gene
regulating regions of DNA.\textsuperscript{98,99} PNAs contain standard nucleobases, however, the sugar-phosphate backbone is replaced with an amide backbone composed of aminoethyl glycine (AEG) units. AEG was selected as the original PNA backbone based on computer models that show an AEG linkage provides a similar distance between nucleobases as a sugar phosphate backbone.\textsuperscript{98,99} Fig. 11 shows a comparison of the functional units of PNA vs. DNA. PNAs contain the same nucleobases as DNA or RNA. Therefore, PNAs form standard Watson-Crick base pairs with PNA, DNA or RNA templates.\textsuperscript{100}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure11.png}
\caption{DNA sugar phosphate backbone vs. PNA AEG backbone}
\end{figure}

(A) A sugar phosphate backbone unit of DNA. (B) An AEG backbone unit of PNA.

The polyamide backbone of PNA possesses several inherent structural and biochemical advantages vs. DNA and RNA. The polyamide backbone of PNA is not recognized by cellular nucleases making PNA more resistant to digestion than DNA and RNA.\textsuperscript{57,97} The uncharged PNA backbone provides decreased charge repulsion between strands, increasing the stability of nucleic acid structures containing PNA strands.\textsuperscript{24,25,96} Studies examining 24 - 30bp DNA and PNA-DNA duplexes showed higher T\textsubscript{m} values for the PNA-DNA hybrids.
DNA duplexes showed $T_m$s between 47.8 and 53.5°C with the corresponding PNA-DNA hybrids displaying $T_m$s between 61.8 to 68.5°C.\textsuperscript{24,26} Corresponding PNA-RNA duplexes showed even higher $T_m$s up to 72.3°C.\textsuperscript{24} Hence, hybrid PNA-DNA and PNA-RNA duplexes possess enhanced thermal stability vs. pure DNA and RNA duplexes.\textsuperscript{24,25,85,96,100,101} These features have provided the impetus to develop PNA-derived therapeutics or biotechnology reagents for many years.

Figure 12. Schematic of DNA/PNA duplex

Duplex obeys Watson-Crick hydrogen bonding rules in an antiparallel orientation.

As stated earlier, PNAs bind nucleic acids via Watson-Crick base pairing. Thermal stability studies of hybrid duplexes showed them to be more stable in an
anti-parallel orientation with the 3’ end of the DNA aligned with the N-terminus of the PNA. Fig. 12 shows a schematic of a PNA-DNA duplex oriented in an anti-parallel direction with a standard Watson-Crick hydrogen-bonding pattern.

Previous x-ray and NMR studies show that PNA duplexes can adopt several discrete structures. These conformers come as a result of rotation around the backbone amides, and transitions between them are slow, reflecting the energy barrier between conformations resulting from the coordinated motion required to avoid steric clashes. PNA-DNA duplexes, however, showed only a single conformation. These hybrid duplexes show structural alterations in both strands. PNA-DNA duplexes also show a slight perturbation of base-stacking arrangement compared to the DNA/DNA duplex. Fig. 13 shows x-ray crystal structures of DNA/DNA (A), PNA-DNA (B), and PNA/PNA (C) duplexes. The structural similarities between DNA/DNA and PNA-DNA duplex helicities are evident from Fig. 13. However, PNA-DNA duplexes show a slight increase in the rise per bp (3.5 Å) compared to B-form DNA duplexes (3.3 Å). The results of previous structural studies suggest that PNA-DNA hybrid structures adopt helical conformations that are quite similar to DNA structures. These structural similarities no doubt facilitate the binding of PNA-DNA hybrids to DNA binding proteins like HMGB1.
Figure 13. NMR/X-ray Structures of Nucleic Acid Duplexes

(A) X-ray crystal structure of a DNA/DNA duplex at a resolution of 1.99 Å [PDB: 4HLI][106]. (B) NMR solution structure of a PNA-DNA hybrid duplex [PDB: 1PDT][105]. (C) X-ray crystal structure of a PNA/PNA\textsuperscript{\textregistered} duplex at a resolution of 1.7 Å [PDB: 1PUP]. DNA strands are shown in orange as cartoons. PNA strands are shown in green as stick structures.
CHAPTER II - MATERIALS AND METHODS

2.1 Polyacrylamide Gel Electrophoresis (PAGE)

Nondenaturing PAGE has been used extensively to characterize DNA 4WJs. Junction formation is observed as a reduction in electrophoretic mobility of the junction vs. single and double-stranded DNA.\textsuperscript{35,80,81,84,107-109} This technique has also been used to observe changes in junction conformation in the presence of metal ions such as Na\textsuperscript{+} and Mg\textsuperscript{2+}.\textsuperscript{81,107,108} PAGE assays have also been used to determine whether junctions show a preference for a particular stacked isomer.\textsuperscript{80,107-109} In these studies, the junction arms were truncated by enzymatic cleavage to generate multiple products that resemble either stacked-X or open-X conformers. The resulting migration pattern(s) of the cleaved junctions are used as controls to identify the mobility of stacked-X vs. open-X isomers of uncharacterized junctions.\textsuperscript{81,108,109} These studies indicate that stacked-X junctions migrate faster than open-X. The junctions investigated here all presumably form stacked-X conformers based on the high level of Mg\textsuperscript{2+} (1 mM) in the analysis buffer.

2.1.1 Four-Way Junction Formation

The sequences for J1 are: 101*, 5'-CGCAATCCTGAGCAG-3'; 102, 5'-CGTGCTCACCAGAATCG-3'; 103, 5'-GCATTCGGACTATGGC-3' and 104, 5'-GCCATAGTGGATTGC-3'. The sequences for 4WJ-PNA\textsubscript{1} are: PNA\textsubscript{1}, H-CAATCCTGAGCA-K-NH\textsubscript{2}; 102, 5'-CGTGCTCACCAGAATCG-3'; 103*, 5'-GCATTCGGACTATGGC-3' and 104, 5'-GCCATAGTGGATTGC-3'. The sequences for b4WJ-PNA\textsubscript{1} are: PNA\textsubscript{1}, HCAATCCTGAGCA-K-NH\textsubscript{2}; b102, 5'-...
TGCTCACCGAATCG-3'; 103*, 5'-GCATTGGCTATGCG-3' and b104, 5'-GCCATAGTGATTG-3'. The sequences for 4WJ-PNA₃ are: 101*, 5'-CGCAATCCTGAGCAG-3'; 102, 5'-CGTGCTTCACCTGCGG-3'; PNA₃: H-ATTCGGAGTATG-K-NH₂ and 104, 5'-GCCATAGTGATTGCG-3'. The sequences for b4WJ-PNA₃ are: 101*, 5'-CGCAATCCTGAGCAG-3'; nb102, 5'-CGTGCTTCACCTGCGG-3'; PNA₃: H-ATTCGGAGTATG-K-NH₂ and 104, 5'-GCCATAGTGATTGCG-3'. The sequences for 4WJ-PNA₁,₃ are: PNA₁, HCAATCCTGAGCA-K-NH₂; 102, 5'-CGTGCTTCACCTGCGG-3'; PNA₃: H-ATTCGGAGTATG-K-NH₂ and 104*, 5'-GCCATAGTGATTGCG-3'. The sequences for b4WJ-PNA₁,₃ are: PNA₁, H-CAATCCTGAGCA-K-NH₂; db102, 5'-TGCTTCACCTGCGG-3'; PNA₃: H-ATTCGGAGTATG-K-NH₂ and bf104*, 5'-CATAGTGATTG-3'. Strands labeled with fluorescein are denoted with an asterisk (101*, 103*, nb104* and bf104*). Each fluorescent strand was purified via HPLC; each non-labeled DNA strand was purified via denaturing polyacrylamide gels. All DNA was purchased from Integrated DNA Technologies (IDT). Junctions were formed by mixing 25μM of a fluorescein labeled strand with 125μM of each unlabeled strand. This mixture was lyophilized and resuspended in an annealing buffer consisting of 50 mM Tris-HCl (pH 7.5) and 1.0 mM MgCl₂. The reconstituted strands are then incubated at 95°C for 2 min, followed by cooling to room temperature for 12-16 h. The 4WJs were loaded onto 15% mini-PROTEAN native polyacrylamide gels (BioRad) and run for 1-5 h at 4°C. The gel running buffer was composed of 0.5 X TBE (MgCl₂ buffer (45 mM Trisma, 45 mM boric acid, 1.0 mM EDTA and 1 mM MgCl₂), pH 7.6. The gels were subsequently
scanned with a Typhoon 9400 Phosphorimager in fluorescence mode at 530 nm using an excitation wavelength of 488 nm and a light pollution filter blocking wavelengths below 510 nm. Gels were scanned at PMT settings that ranged between 400 and 800.

2.1.2 Electrophoretic Mobility Shift Assays (EMSAs)

EMSAs are used to examine the protein recognition properties of the hybrid 4WJs toward the recombinant DNA binding protein, HMGB1b. The binding affinity of the previously uncharacterized junctions: b4WJ-PNA₁, b4WJ-PNA₃, 4WJ-PNA₁,₃ and b4WJ-PNA₁,₃ are evaluated with HMGB1b. A constant 4WJ concentration of 500nM was used with HMGB1b:4WJ molar ratios ranging from 3.2:1 to 1.6  are evaluated with HMGB1b. A constant 4WJ concentration of 500nM was used with HMGB1b:4WJ molar ratios ranging from 3.2:1 to 1.6. Each sample is composed of 0.5 X TBE (MgCl₂) and 10% (w/v) glycerol. Each sample is loaded onto 15% Mini-PROTEAN TBE precast native polyacrylamide gels (BioRad). Gels are run at 4°C in a 0.5 X TBE buffer composed of 45 mM Trisma, 45mM boric acid, 1.0mM EDTA and 1mM MgCl₂ (pH 7.6) for 6-8 h. The gels were subsequently scanned with a Typhoon 9400 Phosphorimager. Each EMSA represents at least three independent runs/junction.

2.2 Circular Dichroism (CD) Analysis of Nucleic Acids

2.2.1 CD Analysis of Classic Nucleic Acids

The CD signal of nucleic acids arises from the stacking geometries of the bases. Differences in CD spectra of A- and B-DNA arise from the respective conformations of the backbone sugar moiety. Fig. 14 shows a
diagram of the two isomers. The deoxyribose and ribose sugars of A-DNA and RNA have an endo pucker at the C3’ position while the deoxyribose sugar of B-DNA has an endo pucker at the C2’ position.\textsuperscript{110}

![Diagram of the two isomers.](image)

Figure 14. Differences in the endo pucker of DNA deoxyribose sugar A- and B-DNA structures.


A consequence of this isomerization is a shift in the helical structure. Fig. 15 displays 3D-models of A- and B-DNA. A-DNA, shown in Fig. 15A, has a much narrower major groove and wider minor groove compared to B-DNA. Fig. 15B
shows the more common B-DNA structure with a wide major groove and narrow minor groove. These changes give rise to the differing CD spectra of A- and B-DNA shown at the bottom of Fig. 14. The spectrum for B-form DNA is shown at the bottom right of Fig. 14. The B-DNA spectrum contains characteristic positive and negative bands near 280nm and 240nm, respectively. These peaks are usually comparable in shape and magnitude. Smaller positive and negative absorption bands near 220nm and 205nm, respectively, are also commonly observed. The spectrum of A-form DNA is shown at the bottom left of Fig. 14 shows a strong negative peak at 205nm and a strong positive near at 270nm.\textsuperscript{110,112,113}

Figure 15. Example DNA duplex structures

(A) B-DNA shown in red. (B) A-DNA shown in blue.

The CD spectra of duplex DNA most commonly display a B-DNA pattern. However, deviations from the classical B-DNA conformation have been
observed, especially under dehydrating conditions. Analyses of DNA three way junctions (3WJs) have shown the global conformation of the junction can affect the helical structure of the junction arms. Changes in the stacking arrangement of 3WJs have been shown to promote transition from a B- DNA conformation to a more A- DNA helical conformation. CD spectra of these junctions produced a strong negative peak near 210nm, indicative of A-DNA.

CD spectroscopy can also be used to characterize DNA 4WJs. Like DNA duplexes, DNA 4WJs typically show a strong preference for B-like helical conformations. CD spectroscopy has also been used to show that the sequences flanking the junction branch point have a large impact on the global conformation of 4WJs. Junctions with branch points capable of migrating along the arms are termed “mobile”, while those with branch points locked into place are termed “immobile”. Immobile junctions contain asymmetric sequences only capable of pairing in a single configuration. CD studies were essential to confirm PAGE analysis of mobile vs. immobile DNA 4WJs. Mobile and immobile 4WJ differ by the positions of the bases at the junction branch point. Immobile 4WJs tend to possess more GC bases at the junction branch point to help immobilize or lock the junction lattice. The CD studies show that the repositioning of GC and AT bp at the junction branch point of 4WJs produced a blue shift and decrease in intensity in the peak near 280nm indicating a shift in structure. These sequence changes presumably contribute to the loss of stability of the preferred isomerization state.
2.2.2 CD Analysis of PNA

The polyamide backbone of PNA is achiral and does not produce a CD signal. However, incorporating chiral molecules into the backbone can induce handedness of single PNA strands and duplexes.\textsuperscript{95,99,103,104,116,117} The chiral moiety can be placed at either PNA terminus \textsuperscript{95,103,104,116} or internally \textsuperscript{103,117}. Examination of a variety of AAs showed Lys to be an efficient promoter of PNA helix handedness\textsuperscript{95,103,116}, while also improving PNA solubility\textsuperscript{99}. To date, all studies indicate that, when attached to a PNA, L-AAs produce left-handed helices and D-AAs generate right-handed helices.\textsuperscript{95,103,104,116,117} The direction and magnitude of PNA handedness are based on a combination of factors such as: i) the identity of the AA side chain, ii) the stereochemistry of the AA, iii) the position of the AA within the PNA backbone and iv) the sequence of the final nucleotides attached to the AA.\textsuperscript{95,103,104,116,117} Attaching the chiral ligand to the C-terminus was also shown to more efficiently induce handedness in PNA helices.\textsuperscript{95} Lys attached to a C (C-Lys) was shown to produce a much more intense CD signal than A-Lys or T-Lys. G-Lys produced a weak signal corresponding to a structure of opposite helicity from C-, A-, and T-Lys.\textsuperscript{95}

With respect to the analysis of PNA-DNA hybrid structures, the strength of signal is based predominantly on the presence of chiral DNA strand. PNA-DNA duplexes always form right-handed helices regardless of the conformation of the unbound PNA strand.\textsuperscript{100,103,117} CD analysis showed PNAs exhibiting right-handed character to bind DNA in an anti parallel fashion, while those exhibiting left-handed character bind in a parallel fashion.\textsuperscript{103,117}
2.2.3 CD Thermal Denaturation Analysis

Ultraviolet (UV) and CD spectroscopy have been used extensively to examine thermal denaturation and structural transitions of nucleic acid structures.\textsuperscript{26,66,75,84,92,96,101,104,112,114,115,118-122} These studies yield data on the relative stabilities of nucleic acid structures, usually given as a melting temperature ($T_m$). These studies allow the determination of cooperativity in melting and the presence of two-state or gradual transition.\textsuperscript{24,26,112,113,123}

By normalizing the upper and lower baselines of the melting curves to 1 and 0, respectively, the Y-axis can be redefined to reflect the fraction of DNA/PNA strands in junction form ($\alpha$). The value of $\alpha$ directly reflects the fraction strands in junction form. Therefore, at the $T_m$, $\alpha$ is equal to 0.5. Differentiating the melting curves yield plots of $\delta \alpha / \delta T$ vs. $T$, where $\delta \alpha / \delta T$ is the change in fraction of DNA/PNA strands in junction form over change in temperature and $T$ is temperature. Calculation of van’t Hoff enthalpies ($\Delta H_{vH}$) was done using the equation below:

$$\Delta H_{vH} = 10 R T_m \left( \frac{\delta \alpha}{\delta T} \right)_{T=T_m}$$  \hspace{1cm} \text{Eq. 1}

Eq. 1 was derived from a more general form of van’t Hoff’s equation to extract thermodynamic data for tetramolecular processes such as 4WJ formation and yields van’t Hoff transition enthalpies ($\Delta H_{vH}$) at the $T_m$.\textsuperscript{24,26,121,123} Alternatively, $\Delta H_{vH}$ was calculated using the full width (Eq. 2) and half width (Eq. 3) of the differentiated melting curve at the half-height.\textsuperscript{24,123} The following two general forms of van’t Hoff’s equation are used for these calculations:
Both B and B’ are association reaction constants for polynucleic acids that depend on the molecularity (n) (i.e. the number of strands) of the process and are in units of cal/mol-K. B is used for the calculation of ΔH_{VH} using the full width (T_1 to T_2) of the melt curve first derivatives. B’ is used for the calculation of ΔH_{VH} using the half width (T_{max} to T_2) of the melt curve first derivatives. Values for B and B’ are 15.40 and 5.63 cal/mol-K, respectively, reported in previous work by Marky and Breslauer.\textsuperscript{123}

Rescaling the melting curve still yields an X value at the minima (T_{min}) corresponds to the T_m. The Y value at the minima or T_m is δα /δT_m. These values for T_m and δα /δT_m can be plugged into Eq. 1 to yield ΔH_{VH}. Alternatively, values for T_1 and T_2 taken at (δα /δT)_{0.5} can be plugged into Eq. 2 to yield ΔH_{VH}. A third method used to calculate ΔH_{VH} (Eq. 3) relies on T_m and T_2 taken at (δα /δT)_{0.5}.\textsuperscript{24,123}

2.2.4 CD Spectroscopy of 4WJs

CD spectra were recorded using Jasco J-815 spectrometer. A 2.0 μM solution of each 4WJ was prepared in CD analysis buffer: 20 mM HEPES, 30mM NH₄Cl, 200 mM KCl, 2 mM DTT, and 10% glycerol. The low Mg²⁺ CD buffer contained 100 μM MgCl₂; the high Mg²⁺ CD buffer contained 2 mM MgCl₂. All spectra were measured in a 0.1 cm path-length quartz cuvette; spectra were
recorded from 320 to 200 nm in 1.0 nm increments at 4°C. The thermal
denaturation profile of each 4WJ was monitored under the same buffer
conditions (low and high Mg\(^{2+}\)) at a fixed wavelength (maxima near 275 nm) from
4 to 100°C. The T\(_m\) values were calculated by the fitting the average value of at
least three independent scans to a sigmoidal curve equation using GraphPad\(^\circledR\).
CHAPTER III - RESULTS

3.1 PAGE Analysis of Native and Hybrid PNA-DNA 4WJs

The six hybrid PNA-DNA 4WJs being investigated for HMGB1b binding capacity are shown in Fig. 16, along with the DNA control J1. EMSAs were used to examine the protein recognition properties of the hybrid 4WJs toward the recombinant DNA binding protein, HMGB1b. Each hybrid 4WJ contains PNA strand(s) at the 101 and/or 103 positions (Fig. 10). Strands 101 and 103, representing the continuous strands of the J1 stacked-X conformer, were selected as the initial replacement strands. Future studies will focus on replacing strands 102 and 104.

PNA length was limited to 12 residues in order to mitigate problems associated with synthesis of longer PNA sequences. As a result, certain hybrids (4WJ-PNA₁, 4WJ-PNA₃, and 4WJ-PNA₁,₃) have DNA strands that are 4 nucleotides longer than the corresponding PNA strands. Our initial report showed that the presence of DNA overhangs did not prohibit the formation of 4WJ-PNA₁, 4WJ-PNA₃. However, the free motion of these DNA overhangs may result in structural instability of the hybrid junctions. Here the effects of truncating these DNA overhangs were investigated. Hybrid junctions b4WJ-PNA₁, b4WJ-PNA₃, and b4WJ-PNA₁,₃, shown at the bottom of Fig. 16, contain truncated DNA sequences in positions complementary to PNA strands. DNA sequences were truncated to match the length of PNA sequences, producing blunt-ended junction arms.
Figure 16. Diagram of Hybrid 4WJs

From top left: J1, 4WJ-PNA1, 4WJ-PNA3, 4WJ-PNA1,3, b4WJ-PNA1, b4WJ-PNA3, and b4WJ-PNA1,3

Gel electrophoresis was used to confirm junction formation. Stable 4WJ formation is evidenced by a significant reduction in electrophoretic mobility of the junction compared to single strands of DNA. Fig. 17 illustrates the differences in mobility between the 101 strand (lane 1) and junctions (lanes 2-9). J1 is loaded in lanes 2 and 9 to give a more accurate comparison of electrophoretic mobility of each hybrid vs. the control. Despite the reduction in charge from their peptide backbones, each single PNA hybrid 4WJ (Fig. 17 lanes 3-6) showed migration patterns similar to J1. The multi PNAs (Fig. 17 lanes 7 and 8), with a substantial reduction in charge, showed only a slight reduction in mobility. PAGE analysis showed the multi PNA hybrids migrate approximately 90 - 95% the distance of J1 and the single-PNA constructs. These results match those of previous studies that showed electrophoretic mobility of hybrid junctions to be a function of size and conformation, with little effect from the reduction in charge due to the PNA
Studies have shown that the difference in electrophoretic mobility of PNA-DNA triplexes was strongly dependent on the shape of the triplex. More specifically, triplexes that possessed a significant bend migrated more slowly than linear triplexes. The 4WJs examined here migrate together due to similarities in size (24-32bp) and global conformation (stacked-X). High mobility bands present in the PNA$_1$ constructs (Fig. 17 lanes 3 and 4) were an area of concern as they may be the result of contaminating structures (i.e. duplexes or triplexes) that promote the assembly of larger structures (i.e. non-4WJ tetramolecular DNA assemblies containing multiple copies of one or two strands) that mimic the electrophoretic mobilities of 4WJs. A series of controls were devised (Fig. 18-20) to test each junction for the presence of higher order contaminating structures composed of two and three strand combinations.

Figure 17. Nondenaturing Polyacrylamide Gel of Hybrid 4WJs.

Lane 1) 101, 2) J1, 3) 4WJ-PNA$_1$, 4) b4WJ-PNA$_1$, 5) 4WJ-PNA$_3$, 6) b4WJ-PNA$_3$, 7) 4WJ-PNA$_{1,3}$, 8) b4WJ:PNA$_{1,3}$ and 9) J1.
The data in Fig. 18 - 20 display the electrophoretic mobility patterns of six hybrid PNA–DNA 4WJs vs. i) J1 and ii) different combinations of potentially contaminating strands of DNA. Fig. 18 shows 4WJ-PNA\textsubscript{1} and b4WJ-PNA\textsubscript{1}. Fig. 19 shows 4WJ-PNA\textsubscript{3} and b4WJ-PNA\textsubscript{3}. Fig. 20 shows 4WJ-PNA\textsubscript{1,3} and b4WJ-PNA\textsubscript{1,3}. The contaminating strands are composed of single strands used to form J1. The data are based on the protocol used by Seeman and Kallenbach to characterize, J1.\textsuperscript{73} The data provides a direct method to compare the mobility of immobile hybrid PNA–DNA junctions vs. a DNA control junction (J1) and potentially contaminating multi-PNAs. In each gel (Fig. 18 - 20), the single strand control (101) is loaded in lane 1; the different single strand combinations are loaded in lanes 2–11; the DNA control, J1 is loaded in lane 12, and the hybrid PNA–DNA junctions are loaded in lanes 13 and 14.

Figure 18. Formation Controls for 4WJ-PNA\textsubscript{1} and b4WJ-PNA\textsubscript{1}

Lanes 1-14: 1) 101*, 2) 101*-102, 3) 101*-103, 4) 101*-104, 5) 102-103*, 6) 102-104*, 7) 103-104*, 8) 101*-102-103, 9) 101*-102-104, 10) 101*-103-104, 11) 102-103-104*, 12) J1(101*), 13) 4WJ-PNA\textsubscript{1}(103*), and 14) b4WJ-PNA\textsubscript{1}(103*).
Figure 19. Formation Controls for 4WJ-PNA\textsubscript{3} and b4WJ-PNA\textsubscript{3}

Lanes 1-14: lanes 1-14: 1) 101\textsuperscript{*}, 2) 101\textsuperscript{*}-102, 3) 101\textsuperscript{*}-103, 4) 101\textsuperscript{*}-104, 5) 102-103\textsuperscript{*}, 6) 102-104\textsuperscript{*}, 7) 103-104\textsuperscript{*}, 8) 101\textsuperscript{*}-102-103, 9) 101\textsuperscript{*}-102-104, 10) 101\textsuperscript{*}-103-104, 11) 102-103 104\textsuperscript{*}, 12) J1(101\textsuperscript{*}), 13) 4WJ-PNA\textsubscript{3}(101\textsuperscript{*}), and 14) b4WJ-PNA\textsubscript{3}(101\textsuperscript{*}).

Figure 20. Formation Controls for 4WJ-PNA\textsubscript{1,3} and b4WJ-PNA\textsubscript{1,3}

Lanes 1-15: lanes 1-14: 1) 101\textsuperscript{*}, 2) 101\textsuperscript{*}-102, 3) 101\textsuperscript{*}-103, 4) 101\textsuperscript{*}-104, 5) 102-103\textsuperscript{*}, 6) 102-104\textsuperscript{*}, 7) 103-104\textsuperscript{*}, 8) 101\textsuperscript{*}-102-103, 9) 101\textsuperscript{*}-102-104, 10) 101\textsuperscript{*}-103-104, 11) 102-103 104\textsuperscript{*}, 12) J1 (101\textsuperscript{*}), 13) 4WJ-PNA\textsubscript{1,3}(104\textsuperscript{*}), and 14) b4WJ-PNA\textsubscript{1,3}(bf104\textsuperscript{*}).
The electrophoretic mobilities of J1 and the hybrid junction bands (Fig. 18 - 20 lanes 12-14) are significantly reduced compared to bands seen with the double and triple DNA strand combinations (Fig. 18 - 20 lanes 2 - 11). Double and triple strand combinations in lanes 2, 7, 8, 9, and 10 from each analysis representing combinations of 101-102, 103-104, 101-102-103, 101-102-104, and 101-103-104, respectively, display bands that run slower than the 101 strand control (lane 1) but faster than the junction bands in lanes 12 - 14. These bands likely represent duplex or possibly triplex structures. Fig. 18 - 20 show that none of the double and triple strand combinations shown in lanes 2 - 11 form species that run with the same mobility as bands representing the 4WJs in lanes 12 - 14. These results display indirectly that the duplex and triplex structures do not aggregate to form contaminating structures that run with the same electrophoretic mobility as immobilized 4WJs. Instead, it seems that immobilized junctions merely dissociate to a certain extent during gel analysis. High mobility bands seen in lanes 2 - 9 of Fig 17 then represent single strand and duplex and/or triplex structures and low mobility bands represent 4WJs.

3.2 HMGB1 Binding

EMSAs were used to evaluate the binding affinity of b4WJ-PNA1, b4WJ-PNA3, 4WJ-PNA1,3 and b4WJ-PNA1,3 toward the recombinant DNA binding protein, HMGB1b. Fluorescence binding and analytical ultracentrifugation analysis from previous studies have shown that HMGB1b binds J1 with a stoichiometry of four to one (4:1), presumably favoring the open configuration. Previous EMSA studies were used to measure the binding constants of 4WJ-
PNA₁ and 4WJ-PNA₃ toward HMGB1b. Both 4WJ-PNA₁ and 4WJ-PNA₃ formed 4:1 complexes that migrate as single bands without intermediate binding species. Previous studies suggest that hybrid junction 4WJ-PNA₃ binds HMGB1b with a nearly identical affinity to J₁ (Kₐ ≈ 6.0 μM). Previous studies suggest that the binding affinity of HMGB1b for 4WJ-PNA₁ was only slightly less than that of J₁ (Kₐ ≈ 16.0 μM). The current study examines the HMGB1b binding properties of four additional hybrid junctions, b4WJ-PNA₁, b4WJ-PNA₃, 4WJ-PNA₁,₃, and b4WJ-PNA₁,₃. The DNA overhangs have been truncated from 4WJ-PNA₁ and 4WJ-PNA₃ to produce the blunt-ended junctions b4WJ-PNA₁ and b4WJ-PNA₃, respectively. The multi-PNA junctions 4WJ-PNA₁,₃ and b4WJ-PNA₁,₃ each contain two PNA strands. The b4WJ-PNA₁,₃ hybrid is the smallest of the 4WJs examined, with truncated DNA sequences at the end of each arm. These new hybrids were developed in an effort to increase the thermal and nuclease stability of 4WJs. However, these changes to the DNA strand sizes and junction PNA content may affect HMGB1 binding affinity. EMSAs were used to examine the binding affinity of b4WJ-PNA₁, b4WJ-PNA₃, 4WJ-PNA₁,₃ and b4WJ-PNA₁,₃ toward HMGB1b in order to understand the effects of removing DNA overhangs and adding an additional PNA strand on the protein recognition properties of hybrid 4WJs toward the recombinant DNA binding protein, HMGB1b.

The EMSA data is displayed in Fig. 21. Panels A-E represent: J₁, b4WJ-PNA₁, b4WJ-PNA₃, 4WJ-PNA₁,₃ and b4WJ-PNA₁,₃. The protein and junction are expressed in terms of molar ratio of protein to junction (P/J). Lane 1 corresponds to 0.05 mM of 4WJ, lanes 2-8 represent each 4WJ incubated with each protein at
P/J ratios of 3.2:1, 6.4:1, 12.8:1, 25.6:1, 47.8:1, 63.4:1 and 79:1. Fig. 21A shows initial binding of HMGB1b to J1 in lane 3 at a ratio of 6.4:1. Complete binding of HMGB1b to J1 is achieved by lane 4 at a P/J ratio of 12.8:1. Fig. 21B shows HMGB1b binding to b4WJ-PNA1 with affinity similar to J1. Initial binding interactions with b4WJ-PNA1 again take place at a P/J ratio of 6.4:1 (lane 3). Complete binding of HMGB1b to b4WJ-PNA1 is seen in lane 4 at a P/J ratio of 12.8:1. In Fig. 21C HMGB1b was shown to bind b4WJ-PNA3 with a slightly lower affinity than J1 and b4WJ-PNA1. Initial binding of b4WJ-PNA3 takes place at a P/J ratio of 12.8:1 (lane 4), and complete binding of HMGB1b to b4WJ-PNA3 is achieved at a P/J ratio of 25.6:1 (lane 5). The multi-PNA hybrids, 4WJ-PNA1,3 and b4WJ-PNA1,3 are shown in Fig. 21D and E. HMGB1b showed similar binding patterns with 4WJ-PNA1,3 and b4WJ-PNA1,3. Both 4WJ-PNA1,3 and b4WJ-PNA1,3 EMSAs show the appearance of broadened/smeared bands with increasing HMGB1b concentration, indicative of the formation of unstable protein/junction complexes that dissociate under experimental conditions. Initial HMGB1b binding is evidenced by the appearance of smeared bands at P/J ratios of 12.8:1 for 4WJ-PNA1,3 (Fig. 21D, lane 5) and 24:1 for b4WJ-PNA1,3 (Fig. 21E, lane 6). The smeared bands continue to broaden as the P/J molar ratio increases out to 79:1 (lane 8) indicating an increase in junction binding. However, complete binding of 4WJ-PNA1,3 and b4WJ-PNA1,3 is never achieved.

Previous binding studies show that HMGB1b has a similar binding affinity for J1 and hybrid junctions 4WJ-PNA1 and 4WJ-PNA3. The EMSA results of the current study suggest that the blunt-ended hybrid junctions b4WJ-PNA1 and
b4WJ-PNA$_3$ also bind HMGB1b with a similar affinity to J1. This result suggests that truncating the DNA overhangs did not adversely affect the HMGB1b binding capacity of hybrid 4WJs, and demonstrates the therapeutic potential of these new constructs as ligands for HMGB1. The multi-PNA hybrid junctions, 4WJ-PNA$_1$,$_3$ and b4WJ-PNA$_1$,$_3$ do not form stable complexes and are not expected to be viable therapeutic ligands for HMGB1.

Figure 21. EMSA of HMGB1b binding

A) J1 and novel hybrid junctions  B) b4WJ-PNA$_1$, C) b4WJ-PNA$_3$, D) 4WJ-PNA$_1$,$_3$ and E) b4WJ-PNA$_1$,$_3$. Lanes 1-8 of each panel contain protein/DNA ratios of 1) 0:1, 2) 3.2:1, 3) 6.4:1, 4) 12.8:1, 5) 25.6:1, 6) 47.8:1, 7) 63.4:1 and 8) 79:1, respectively.
3.3 CD Analysis of Hybrid 4WJs

B-form and A-form DNA possess different CD spectra. The spectra of B-form DNA show positive bands at 280nm and negative bands at 250nm that are roughly symmetrical. A-form DNA show a characteristic negative band at ~205nm and a strong positive band between 260 and 270nm. Fig. 22A and B show comparisons of CD spectra of J1 and the hybrid 4WJs in low (0.01mM) and high (2mM) concentrations of Mg$^{2+}$ ion. The 2mM Mg$^{2+}$ ion concentration was chosen to correspond with the high end of the standard ECM Mg$^{2+}$ ion concentration range of 1mM – 2mM.$^{125}$ CD spectra of the hybrid 4WJs are quite different from that of J1. Spectra of the hybrid 4WJs show positive bands between 265 and 278nm are blue shifted and lower in intensity compared to those of J1. These positive bands displayed a lack of symmetry with negative bands centered at 250nm in all hybrid 4WJ spectra. Hybrid junctions also showed an intense peak near 205nm under both sets of conditions. The lack of symmetry in the bands between 250 and 280nm and the presence of the peaks at 205nm are highly indicative of a shift toward more A-form conformation. These shifts were more pronounced in the case of the multi PNA constructs.
Figure 22. CD spectra of 4WJs in low (~0.01mM) and high (2mM) Mg$^{2+}$

CD spectra highlighting A- and B-form signals for J1 vs. single-PNA and multi-PNA 4WJ in low and high Mg$^{2+}$. Panel A) represents each single- and multi-PNA hybrid vs. J1 in low Mg$^{2+}$. Panel B) represents each single- and multi-PNA hybrid vs. J1 in high Mg$^{2+}$.

CD spectra can be used to effectively examine the relative contribution of A and B form helical character to the overall spectra. This is done by examining signals at 205 and 250nm. The ratio between the magnitudes of the two peaks yields the apparent fraction of A and B form helices.$^{78,94,114,115}$ Fig. 23 shows the apparent fraction of A (AF$_A$) and B (AF$_B$) form helices determined for each junction in high and low Mg$^{2+}$. The DNA control J1 displayed AF$_B$ values
exceeding 0.85 in both low and high Mg\(^{2+}\) denoting its predominantly B-form character.

Hybrid junctions all showed a propensity for higher fractions (0.5-0.95) of A-form helices in the absence of Mg\(^{2+}\). Single PNA b4WJs showed a propensity for greater B-form character than their full-length counterparts. All single PNA hybrids displayed considerable structural sensitivity to Mg\(^{2+}\) concentration. In the presence of 2mM Mg\(^{2+}\) all single PNA hybrids shifted to higher AF\(_B\) with the greatest shift seen for b4WJ-PNA\(_3\) (0.5 to 0.9 AF\(_B\)). The multi PNA hybrids possessed the highest AF\(_A\). The structure of the multi PNA hybrids was largely unaffected by Mg\(^{2+}\) concentration, retaining <0.9 AF\(_A\) under both sets of conditions. The signals between 265 nm and 280 nm are excluded because they are common to both A- and B-form conformers. The hatched bars represent helical fractions calculated in low Mg\(^{2+}\); the solid bars represent fractions calculated in high Mg\(^{2+}\).

Figure 23. Apparent fraction of B- and A-form DNA

(A) B-form DNA and (B) A-form DNA of each junction in low and high Mg\(^{2+}\).
The $F_B$ value of each single-PNA construct increases in high Mg$^{2+}$ (solid bars); three of four hybrids possess $AF_B$ values > 0.5. The secondary structure of the PNA$_3$ derived hybrids proved more sensitive to Mg$^{2+}$ than PNA$_1$ derived hybrids, again illustrating the sequence dependence of hybrid structure and stability. 4WJ-PNA$_3$ and b4WJ-PNA$_3$ display an approximate 2-fold increase in B-form content in high Mg$^{2+}$, with the $AF_B$ value for b4WJ-PNA$_3$ being nearly identical to J1. The large $AF_A$ values (> 0.9) for the multi-PNA hybrids reflect the strong A-form CD signature displayed in Fig. 22. Moreover, the structure of the multi-PNA junctions showed little sensitivity to Mg$^{2+}$ concentration in the range examined (0.01mM – 2.00mM).

The CD spectra of the hybrid junctions indicate that the helicities of the single-PNA hybrids more closely resemble those of J1 vs. the multi-PNA hybrids. The CD wavelength scans suggest that secondary structure of the single-PNA hybrids is largely composed of helices that possess A- and B-form characteristics. Moreover, the CD spectra clearly show that ionic strength strongly influences the secondary structure of the single-PNA hybrid junctions. The CD spectra of the single-PNA hybrid junctions display a prominent increase in B-form helices upon shifting the ionic strength of the buffer from 0.01mM to 2.00mM. Again, these spectra more closely resemble that of the DNA control, J1. The CD spectra of the multi-PNA hybrid 4WJs deviate more significantly from J1. The multi-PNA hybrids to have larger A-form helical character vs. single-PNA hybrids and J1. Moreover, the secondary structure (A-form content) of the multi-PNA hybrids are not sensitive to changes in ionic strength. One plausible
explanation for the reduced level of ionic strength sensitivity may be the 
increased hydrophobicity from the additional polyamide PNA peptide back-bone.

3.4 Thermal Melting Curves and Derivation of Thermodynamic Parameters

To examine the contribution of the PNA strands to thermal stability and 

further explore the link between ionic strength and junction structure, the effects 
of Mg\(^{2+}\) concentration, end blunting, and PNA strand content on the 

conformational stability of hybrid 4WJs were examined. Thermal denaturation of 

hybrid 4WJs was monitored by CD spectroscopy. Denaturation was examined by 

monitoring the CD signal at the maxima near 280nm as each junction was 

heated from 4°C to 100°C. Denaturation curves were generated by fitting the 

data with a Boltzmann sigmoidal function, and are based on at least three 

independent assays.

Denaturation curves are shown in Fig. 24A - G. The data has been 

normalized from 0 - 1 to reflect the fraction of intact junction, with 1 being fully 

intact and 0 being fully denatured. Junctions were examined at a concentration of 

2.0μM. The data was fit with a Boltzmann sigmoidal function using GraphPad\(^{\circ}\). 

Each line represents at least three independent melts. The denaturation curves 

for J1 in low (0.01mM) and high (2.00mM) Mg\(^{2+}\) ion concentrations, shown in Fig 

24A, have distinct S-shapes. These melting profiles are indicative of cooperative 

two-state transitions from folded junction to unfolded strands.\(^{123}\) J1 denaturation 

curves show estimated onsets of melting near 20°C, evidenced by a rapid 

decrease in slope of the denaturation curve. The slopes of the J1 curves begin to 

rapidly level off near 55°C, indicating J1 is almost fully denatured at this
temperature. Full-length single-PNA junctions, 4WJ-PNA\textsubscript{1} and 4WJ-PNA\textsubscript{3} (Fig. 24B and D, respectively), show a much earlier onset of melting (<10°C) than the other hybrid junctions (20°C - 25°C) in both low (0.01mM) and high (2.00mM) Mg\textsuperscript{2+} ion concentrations. 4WJ-PNA\textsubscript{1} and 4WJ-PNA\textsubscript{3} denaturation curves also have more gradual slopes than those of the other junctions, giving them comparatively larger melting ranges (≈10°C - 60°C). This may indicate more gradual melting transitions or secondary pre-melting transitions for the full-length single-PNA junctions.\textsuperscript{123} The blunt-ended single-PNA junctions, b4WJ-PNA\textsubscript{1} and b4WJ-PNA\textsubscript{3} (Fig. 24C and E, respectively) appear to display clear cooperative transitions between folded and unfolded states in low (0.01mM) and high (2.00mM) Mg\textsuperscript{2+} ion concentrations. Both multi-PNA hybrids, 4WJ-PNA\textsubscript{1,3} and b4WJ-PNA\textsubscript{1,3} (Fig. 24F and G, respectively) also seem to display cooperative two-state transitions in low (0.01mM) and high (2.00mM) Mg\textsuperscript{2+} ion concentrations. Both 4WJ-PNA\textsubscript{1,3} and b4WJ-PNA\textsubscript{1,3} showed particularly narrow melting ranges (approximately 25°C to 45°C) and sharp transitions compared to the other junctions.

The melt curves show that the blunt-ended single-PNA hybrids and multi-PNA hybrids undergo sharper and more cooperative melting transitions compared to the full-length single-PNA hybrids, suggesting an increase in thermal stability. The earlier onset of melting and the more gradual slopes suggest the full-length single-PNA hybrids exhibit some form of premelting behavior, possibly due to the increased freedom of motion of the DNA overhangs.
Figure 24. Thermal melting curves of 4WJs in low versus high Mg\textsuperscript{2+}

(A) J1, (B) 4WJ-PNA\textsubscript{1}, (C) b4WJ-PNA\textsubscript{1}, (D) 4WJ-PNA\textsubscript{3}, (E) b4WJ-PNA\textsubscript{3}, and (G) b4WJ-PNA\textsubscript{1,3}. High Mg\textsuperscript{2+} (2.00mM) conditions are shown with open circles. Low Mg\textsuperscript{2+} (0.01mM) conditions are shown with solid circles. The curves shown are Boltzmann sigmoid fits of the data. All junctions were examined at 2μM concentrations from 4°C to 100°C. The data shown are based on at least three independent assays.
Melting curves were used to generate $T_m$ values for J1 and the hybrid junctions. Table 1 shows the $T_m$ values for each junction in low (0.01mM) vs. high (2.00mM) Mg$^{2+}$ ion concentrations and $\Delta T_m$ from low to high Mg$^{2+}$ concentration. The $T_m$s for the single-PNA junctions and b4WJ-PNA$_{1,3}$ are higher in the presence of 2.00mM Mg$^{2+}$ compared to 0.01mM Mg$^{2+}$, suggesting that Mg$^{2+}$ has a stabilizing effect on these junctions. For these junctions $\Delta T_m$s from low to high Mg$^{2+}$ conditions range from 1.10°C (b4WJ-PNA$_3$) to 4.28°C (4WJ-PNA$_3$). The multi-PNA junction 4WJ-PNA$_{1,3}$ showed the opposite affect in the presence of 2.00mM Mg$^{2+}$ showing a slight decrease in $T_m$ ($\Delta T_m = -0.55$) compared to 0.01mM Mg$^{2+}$. For the majority of the junctions examined, the $\Delta T_m$s for low vs. high Mg$^{2+}$ suggest that hybrid 4WJs are stabilized by the presence of 2.00mM Mg$^{2+}$ ions. This stabilization effect may be attributed to an increase in the stabilization of the solvation shell surrounding the junction lattice.$^{104,126}$

Table 1

4WJ $T_m$s in low (0.01mM) and high (2.00mM) Mg$^{2+}$

<table>
<thead>
<tr>
<th>Junction</th>
<th>J1</th>
<th>4WJ-PNA$_1$</th>
<th>b4WJ-PNA$_1$</th>
<th>4WJ-PNA$_3$</th>
<th>b4WJ-PNA$_3$</th>
<th>4WJ-PNA$_{1,3}$</th>
<th>b4WJ-PNA$_{1,3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_m$ (°C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low Mg$^{2+}$</td>
<td>$39.47 \pm 1.03$</td>
<td>$32.53 \pm 1.34$</td>
<td>$32.83 \pm 0.86$</td>
<td>$31.47 \pm 2.25$</td>
<td>$40.08 \pm 0.87$</td>
<td>$35.60 \pm 0.78$</td>
<td>$30.94 \pm 0.63$</td>
</tr>
<tr>
<td>high Mg$^{2+}$</td>
<td>$40.04 \pm 1.04$</td>
<td>$34.00 \pm 0.87$</td>
<td>$35.55 \pm 1.05$</td>
<td>$35.75 \pm 1.83$</td>
<td>$41.18 \pm 1.21$</td>
<td>$35.05 \pm 0.40$</td>
<td>$32.95 \pm 0.75$</td>
</tr>
<tr>
<td>$\Delta T_m$ (°C)</td>
<td>$0.57$</td>
<td>$1.47$</td>
<td>$2.72$</td>
<td>$4.28$</td>
<td>$1.10$</td>
<td>$-0.55$</td>
<td>$2.01$</td>
</tr>
</tbody>
</table>

$\Delta T_m$ values show differences in $T_m$s of each junction in low (0.01mM) vs. high (2mM) Mg$^{2+}$. 

50
The ΔT_m values for each hybrid vs. J1 in low (0.01mM) and high (2.00mM) Mg^{2+} ion concentrations are listed in Table 2. The ΔT_m values show the majority of the hybrid junctions to be slightly less thermally stable than J1 at both high and low Mg^{2+} concentrations. Only b4WJ-PNA_3 showed an increase in T_m versus J1 (ΔT_m = 1.14°C in 0.01mM Mg^{2+} ion concentration and ΔT_m = 1.14°C in 2.00mM Mg^{2+} ion concentration). The ΔT_m value for each hybrid vs. J1 in 0.01mM Mg^{2+} range from -8.53°C (b4WJ-PNA_{1,3}) to 0.61°C (b4WJ-PNA_3). Upon shifting to 2mM Mg^{2+}, the ΔT_m values of the hybrids vs. J1 range from -7.09°C (b4WJ-PNA_{1,3}) to 1.14°C (b4WJ-PNA_3). This further illustrates the Mg^{2+} sensitivity of hybrid 4WJs. The increase in T_m of b4WJ-PNA_3 compared to J1 indicates that addition of PNA strands to the junction lattice is a viable method for increasing the thermal stability of 4WJs.

Table 2

*Hybrid junctions vs. J1 in low (0.01mM) and high (2.00mM) Mg^{2+}*

<table>
<thead>
<tr>
<th>Junction</th>
<th>J1</th>
<th>4WJ-PNA_1</th>
<th>b4WJ-PNA_1</th>
<th>4WJ-PNA_3</th>
<th>b4WJ-PNA_3</th>
<th>4WJ-PNA_{1,3}</th>
<th>b4WJ-PNA_{1,3}</th>
</tr>
</thead>
<tbody>
<tr>
<td>T_m (°C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low Mg^{2+}</td>
<td>39.47 ±1.03</td>
<td>32.53 ±1.34</td>
<td>32.83 ±0.86</td>
<td>31.47 ±2.25</td>
<td>40.08 ±0.87</td>
<td>35.60 ±0.78</td>
<td>30.94 ±0.63</td>
</tr>
<tr>
<td>ΔT_m (°C)</td>
<td></td>
<td>-6.94</td>
<td>-6.64</td>
<td>-8.04</td>
<td>0.61</td>
<td>-3.87</td>
<td>-8.53</td>
</tr>
</tbody>
</table>

| T_m (°C) |    |           |            |           |            |              |              |
| high Mg^{2+} | 40.04 ±1.04 | 34.00 ±0.87 | 35.55 ±1.05 | 35.75 ±1.83 | 41.18 ±1.21 | 35.05 ±0.40 | 32.95 ±0.75   |
| ΔT_m (°C) |    | -6.04     | -4.49      | -4.29     | 1.14       | -4.99        | -7.09        |

ΔT_m values show differences in T_m's of each hybrid vs. J1.
Table 3

Full-length junctions vs. blunt-ended junctions in low (0.01mM) and high (2.00mM) Mg$^{2+}$

<table>
<thead>
<tr>
<th>Junction</th>
<th>4WJ-PNA$_1$</th>
<th>b4WJ-PNA$_1$</th>
<th>4WJ-PNA$_3$</th>
<th>b4WJ-PNA$_3$</th>
<th>4WJ-PNA$_1,3$</th>
<th>b4WJ-PNA$_1,3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_m$ (°C) low Mg$^{2+}$</td>
<td>32.53 ±1.34</td>
<td>32.83 ±0.86</td>
<td>31.47 ±2.25</td>
<td>40.08 ±0.87</td>
<td>35.60 ±0.78</td>
<td>30.94 ±0.63</td>
</tr>
<tr>
<td>$\Delta T_m$ (°C) blunt-full</td>
<td>-</td>
<td>0.30</td>
<td>-</td>
<td>8.61</td>
<td>-</td>
<td>-4.66</td>
</tr>
<tr>
<td>$T_m$ (°C) high Mg$^{2+}$</td>
<td>34.00 ±0.87</td>
<td>35.55 ±1.05</td>
<td>35.75 ±1.83</td>
<td>41.18 ±1.21</td>
<td>35.05 ±0.40</td>
<td>32.95 ±0.75</td>
</tr>
<tr>
<td>$\Delta T_m$ (°C) blunt-full</td>
<td>-</td>
<td>1.55</td>
<td>-</td>
<td>5.43</td>
<td>-</td>
<td>-2.10</td>
</tr>
</tbody>
</table>

$\Delta T_m$ values show differences in $T_m$s of each full-length junction containing overhangs vs. blunt-ended junctions.

The freedom of motion of the DNA overhangs present in full-length junctions 4WJ-PNA$_1$, 4WJ-PNA$_3$, and 4WJ-PNA$_1,3$ may result in structural instability of these hybrid junctions. To evaluate the effects of removing DNA overhangs, $T_m$ values of full-length junctions were compared to blunt-ended junctions. The resulting $\Delta T_m$s are listed in Table 3. Truncating the overhangs increased the $T_m$s of the single-PNA hybrids in both low (0.01mM) and high (2.00mM) Mg$^{2+}$. The junction b4WJ-PNA$_3$ showed the highest $\Delta T_m$s in both low ($\Delta T_m = 8.61{\degree}C$) and high Mg$^{2+}$ ($\Delta T_m = 5.43{\degree}C$). The results for the single-PNA hybrids suggest that removal of DNA overhangs produces a stabilizing effect on these hybrid junctions. However, a reduction in $T_m$ was observed between the full-length multi-PNA hybrid and its blunt-ended counter part. For the multi-PNA hybrids, the thermal unfolding data shows the $T_m$ value for b4WJ-PNA$_1,3$ is -
4.66°C lower than 4WJ-PNA1,3 in low Mg²⁺ and -2.10°C lower in high Mg²⁺. The reduction in Tₘ shown for 4WJ-PNA1,3 vs. b4WJ-PNA1,3 may be a result of different forces stabilizing the tertiary structure (i.e. stacked-X conformation). The increased hydrophobicity imparted by the PNA strands may lead to a stabilization from the hydrophobic effect. The DNA overhangs of 4WJ-PNA1,3 may act as a negatively charged shell around a more hydrophobic core, ordering the solvent cage around 4WJ-PNA1,3 resulting in an increased Tₘ compared to b4WJ-PNA1,3.

In order to further examine the temperature-dependent melting behavior of 4WJs, the melting curves from Fig. 24 were analyzed for van’t Hoff transition enthalpies (ΔHᵥH). The values of ΔHᵥH are a function of the equilibrium constant (Kₑq) and calculated from the rate of change in fraction of DNA/PNA strands in junction form at the Tₘs. The melting curves were normalized to produce upper and lower baselines of 1 and 0, respectively. The Y-axis then reflects the fraction of DNA/PNA strands in junction form as α. Plots of δα/δT vs. T (Fig. 25) were produced from the first derivatives of the Boltzmann sigmoidal functions used to fit the normalized melting data. The smooth S-shaped melt curves (Fig. 24) displayed by most of the hybrid 4WJs indicate a two-state transition between folded and unfolded states. However, the unfolding of the full-length single-PNA constructs was slow, possibly indicating a gradual transition. The presence of multiple peaks or shoulders in the first derivatives of the melting curves would indicate a non-cooperative gradual transition. Fig. 25 shows the first derivatives of the melting curves have only single peaks giving a clear indication that all
hybrid 4WJs undergo a cooperative two state transition upon melting.\textsuperscript{24,26,112,113,123}
Figure 25. First derivatives of 4WJ melting curves

The curves show the first derivatives of the Boltzmann sigmoidal functions used to fit the thermal denaturation data. The change in fraction of free strands as a function of time ($\delta\alpha / \delta T$) is shown on the Y-axis. Minima represent junction T_m's. Each junction is shown in low vs. high Mg$^{2+}$. High Mg$^{2+}$ (2mM) conditions are shown with a solid line. Low Mg$^{2+}$ (100μM) conditions are shown with a dashed line.
Calculation of $\Delta H_{VH}$ from the first derivatives of the melting curves was done using the three previously discussed methods. Eq. 1, derived from the general form of van’t Hoff’s equation, is listed as method 1. The Y values at the maxima of the differentiated melting curves are $\delta \alpha / \delta T_m$. Eq. 2 was used to calculate $\Delta H_{VH}$ from the full width of the differentiated melting curve at half height, and is listed as method 2. Values for $T_1$ (lower value) and $T_2$ (upper value) were taken at $(\delta \alpha / \delta T)_{0.5}$ from the differentiated melting curves. Eq. 3 was used to calculate $\Delta H_{VH}$ from the half width of the differentiated melting curve at half height, and is listed as method 3. The association reaction constants for tetramolecular nucleic acid interactions used for the calculation of $\Delta H_{VH}$ at the full (B) and half-widths (B’) were reported in previous work by Marky and Breslauer.\textsuperscript{123} In this case a molecularity of $n=4$ yields values of 15.40 for B and 5.63 for B’.

Comparisons of the $\Delta H_{VH}$ values for J1 vs. each hybrid junction are shown in Table 4 for both low and high Mg\textsuperscript{2+} ion concentrations. Table 4 shows that, at a 2.00mM Mg\textsuperscript{2+} ion concentration, $\Delta H_{VH}$ values of 4WJ-PNA\textsubscript{1} and 4WJ-PNA\textsubscript{3}, decrease by approximately 1.4 to 5.4 kcal/mol compared to J1. These results are in line with the $T_m$ results, suggesting that 4WJ-PNA\textsubscript{1} and 4WJ-PNA\textsubscript{3} are less stable than J1 at a 2.00mM Mg\textsuperscript{2+} ion concentration. At a 0.01mM Mg\textsuperscript{2+} ion concentration $\Delta H_{VH}$ values for 4WJ-PNA\textsubscript{1} and 4WJ-PNA\textsubscript{3} decrease by 0.5 and 14.8 kcal/mol, respectively, again suggesting a decrease in stability compared to J1 in line with the $T_m$ results. At a 0.01 mM Mg\textsuperscript{2+} ion concentration b4WJ-PNA\textsubscript{3} appears slightly more stable than J1, showing a 3.3 kcal/mol increase in $\Delta H_{VH}$
compared to J1. However, at a 2.00 mM Mg$^{2+}$ ion concentration b4WJ-PNA$_3$ showed a 4.7 kcal/mol decrease in $\Delta H_{VH}$ compared to J1. This is in contrast to the T$_m$ comparison between b4WJ-PNA$_3$ and J1, which showed b4WJ-PNA$_3$ to have a higher T$_m$ than J1 at a 2.00mM Mg$^{2+}$ ion concentration and a lower T$_m$ than J1 at 0.01mM Mg$^{2+}$ ion concentration.

Table 4

$\Delta H_{VH}$ of Hybrid junctions vs. J1 in low (0.01mM) and high (2.00mM) Mg$^{2+}$

<table>
<thead>
<tr>
<th>Junction</th>
<th>J1</th>
<th>4WJ-PNA$_1$</th>
<th>b4WJ-PNA$_1$</th>
<th>4WJ-PNA$_3$</th>
<th>b4WJ-PNA$_3$</th>
<th>4WJ-PNA$_1,3$</th>
<th>b4WJ-PNA$_1,3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta H_{VH}$ (kcal/mol)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low Mg$^{2+}$</td>
<td>42.5 ±6.3</td>
<td>42.0 ±5.7</td>
<td>57.7 ±8.6</td>
<td>27.7 ±3.3</td>
<td>45.8 ±8.6</td>
<td>61.7 ±9.4</td>
<td>53.3 ±7.5</td>
</tr>
<tr>
<td>$\Delta H_{VH}$ (kcal/mol)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hybrid - J1</td>
<td>-</td>
<td>-0.5</td>
<td>15.2</td>
<td>-14.8</td>
<td>3.3</td>
<td>19.2</td>
<td>10.8</td>
</tr>
<tr>
<td>$\Delta H_{VH}$ (kcal/mol)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>high Mg$^{2+}$</td>
<td>46.6 ±6.7</td>
<td>41.2 ±5.6</td>
<td>64.3 ±9.3</td>
<td>32.3 ±4.7</td>
<td>41.9 ±6.3</td>
<td>72.2 ±11.5</td>
<td>61.9 ±9.3</td>
</tr>
<tr>
<td>$\Delta H_{VH}$ (kcal/mol)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hybrid - J1</td>
<td>-</td>
<td>-5.4</td>
<td>17.7</td>
<td>-1.43</td>
<td>-4.7</td>
<td>25.6</td>
<td>15.3</td>
</tr>
</tbody>
</table>

Values are shown for each junction at low (0.01mM) and high (2.00mM) Mg$^{2+}$ ion concentrations.

The reduction in $\Delta H_{VH}$ shown for the full-length single-PNA junctions 4WJ-PNA$_1$ and 4WJ-PNA$_3$ compared to J1 may be attributed to destabilizing effects from the loss of four DNA base pairs and the increased freedom of motion of the unpaired DNA overhangs. The PNA$_3$ constructs 4WJ-PNA$_3$ and b4WJ-PNA$_3$ may have lower $\Delta H_{VH}$ values than J1 due not only to loss of four base pairs, but also to the presence of the three AT pairs at the ends of junction arm III. The AT pairs at the ends of 4WJ-PNA$_3$ and b4WJ-PNA$_3$ are less stable than the GC pairs at
the ends of J1 and may lead to premature onset of melting and lower $\Delta H_{VH}$ values.

Table 4 shows that in both low (0.01mM) and high (2.00mM) Mg$^{2+}$ ion concentrations the $\Delta H_{VH}$ values of b4WJ-PNA$_1$, 4WJ-PNA$_{1,3}$, and b4WJ-PNA$_{1,3}$ increase by 10 - 25 kcal/mol compared to J1. This is in contrast to a decrease of approximately 4°C - 9°C in the $T_m$s of b4WJ-PNA$_1$, 4WJ-PNA$_{1,3}$, and b4WJ-PNA$_{1,3}$ compared to J1. The decrease in $T_m$ values coupled with an increase in $\Delta H_{VH}$ values compared to J1 suggests that the b4WJ-PNA$_1$ and the multi-PNA hybrids may be more stable than J1 in the lower range of temperatures examined (4°C - 25°C). This is evidenced by the relatively flat upper portion of the melt curves between 4°C and 25°C, and the sharp drop in CD signal near the $T_m$s for b4WJ-PNA$_1$, 4WJ-PNA$_{1,3}$, and b4WJ-PNA$_{1,3}$ (Fig. 24C, F, and G, respectively). The melt profile of J1 (Fig. 24A) displays an initial drop in CD signal at a slightly lower temperature ($\approx$20°C) in comparison to b4WJ-PNA$_1$, 4WJ-PNA$_{1,3}$, and b4WJ-PNA$_{1,3}$ ($\approx$25°C), and displays a much more gradual loss of signal.

More favorable base-stacking interactions from the reduced charge repulsion imparted by the PNA$_1$ strand coupled with the stabilizing effect of removing DNA overhangs may lead to a higher $\Delta H_{VH}$ for b4WJ-PNA$_1$ compared to J1. The increase values of $\Delta H_{VH}$ shown for the multi-PNA hybrids compared to J1 may again be due to more favorable base-stacking interactions. The increased flexibility of PNA and the lack of charge repulsion between the DNA and PNA strands may allow the multi-PNA hybrids to maximize base-stacking interactions of all four arms, producing a stabilizing effect on the helical structure.
of these junctions at temperatures below 25°C. Alternatively, the multi-PNA hybrids may possess a more stable stacked-X structure than the other junctions at temperatures below 25°C that must first unfold before denaturation can occur. Above 25°C the stabilizing effect of favorable base-stacking interactions or stacked-X structure may be overcome by the destabilizing effects of loss of eight base pairs compared to J1 and a more disordered solvent cage from the increased hydrophobicity imparted by the PNA strands. These combined effects may lead to the higher $\Delta H_{VH}$ and lower $T_m$s observed for 4WJ-PNA$_{1,3}$ and b4WJ-PNA$_{1,3}$ compared to J1.
CHAPTER IV - DISCUSSION AND CONCLUSION

4.1 Discussion

The goal of this study was to examine hybrid PNA-DNA 4WJs as possible therapeutics to target HMGB1. EMSAs, used to probe binding interactions of HMGB1b with hybrid 4WJs, suggest that HMGB1b recognizes single-PNA hybrid 4WJs, b4WJ-PNA1 and b4WJ-PNA3, with similar affinity to the DNA control, J1 (Fig. 21B and C) supporting the hypothesis that the DNA-binding cytokine, HMGB1, recognizes unnatural nucleic acid substrates. CD, used to examine structural differences between hybrid PNA-DNA junctions and pure DNA junctions, showed that single-PNA hybrid junctions have a more B-form helical character than the multi-PNA hybrids. Together, the EMSA and CD structural analysis results suggest that the helical structure of the hybrid 4WJs may be a key factor influencing HMGB1b binding affinity. Thermal melting behavior was examined via CD to gauge the relative stabilities of hybrid PNA-DNA junctions compared to pure DNA junctions. The single-PNA hybrid b4WJ-PNA3 showed increased T_m compared to J1. The EMSA and thermal analysis results suggest that the single-PNA hybrid b4WJ-PNA3 poses a similar HMGB1b binding affinity and enhanced thermal stability compared to its DNA parent structure J1. This suggests that b4WJ-PNA3 may be a suitable therapeutic ligand for HMGB1.

There are several plausible explanations to explain why HMGB1b recognizes certain hybrid 4WJs more readily. Two of the main possibilities are linked to: i) the rotation angle between adjacent helices in the junction lattice (\( J_{\text{twist}} \)) and ii) topology (dimensions) of the minor groove of each 4WJ. Previous
studies utilizing X-ray crystallography, electrophoresis, FRET and atomic force microscopy have shown that the rotation angle ($J_{\text{twist}}$) between the two adjacent helices of a J1 in the stacked-X form is between $56.5^\circ - 60^\circ$.

These $J_{\text{twist}}$ dimensions presumably facilitate protein recognition by placing the junction branch point in a favorable conformation for HMGB1b binding. We suspect that the single-PNA hybrids have a similar $J_{\text{twist}}$ value to J1 that promotes protein binding. The multi-PNA, however, possess PNA strands that represent ~50% of the junction lattice. In this case, the flexible PNA backbones of the multi-PNA hybrids enable the junction lattice to assume $J_{\text{twist}}$ values that are either $> or < 56.5^\circ - 60^\circ$. As a result, the multi-PNA junctions may adopt unfavorable conformations that destabilize HMGB1b binding interactions. As a result, discreet 4:1 complexes do not form. The schematic below (Fig. 26) shows a model of the potential binding interactions of HMGB1b with single-PNA (Fig. 26A) and multi-PNA hybrids (Fig. 26B). I suspect that the $J_{\text{twist}}$ angles between the junction arms of the single-PNA hybrids (Fig. 26A) more closely match the $60^\circ$ angles of the DNA control, J1 than do the multi-PNA hybrids. The $J_{\text{twist}}$ angles of the multi-PNA hybrids (Fig. 26B) may deviate to produce more acute or obtuse angles that are unfavorable for HMGB1b binding. The reduction in binding affinity of HMGB1b toward multi-PNA hybrids seen in the EMSAs (Fig. 21F and G) may be due to reduced access to the 4WJ binding sites or additional energy differences between unbound and bound 4WJ conformations as a result of unfavorable $J_{\text{twist}}$ angles as illustrated in Fig 26B. Another possible reason for the reduced affinity observed between HMGB1b and the multi-PNA hybrids may
be the lack of electrostatic interactions between HMGB1b and both PNA backbones of these junctions. Additionally, electrostatic interactions between positively charged residues along the binding face of HMGB1b and junctions are lost from replacement of two sugar/phosphate backbones with uncharged PNAs for the multi-PNA 4WJs vs. one for the single-PNA junctions. This 50% reduction in backbone charge of the multi-PNA junctions is a likely contributor to the reduced affinity shown for HMGB1b in the EMSAs.

Figure 26. Schematic of the effects of changing Jtwist angle on HMGB1 binding.

With respect to changes in the topology of 4WJs, it is well known that the minor groove of A-DNA is broader than B-form DNA.\textsuperscript{112,126,128} HMGB1b is binds within the minor groove of B-DNA. Presumably, the narrow minor groove of B-DNA promotes enhanced binding contacts with HMGB1b. Conversely, the hybrid 4WJs with large fraction of A-form helices (i.e. the multi-PNA junctions) possess
broader minor groove dimensions that may reduce contacts with HMGB1b. NMR studies of hybrid PNA-DNA duplexes by Eriksson et al. showed these substrates to possess characteristics of both A- and B-form DNA. In this study, the multi-PNA hybrid junctions have an approximately 9-fold larger A-form signal at 205 nM compared to J1. Although CD does not provide high-resolution data, one may surmise that the minor groove dimensions of the helices measured via CD have similarities to the known 3-D structures of B- and A-form DNA duplexes. If this is the case, the B-form helices found in J1 have a narrow, shallow minor groove topology, while the helices of the hybrid junctions have broader, shallower minor grooves intermediate between the topology of A- and B-DNA. Fig. 27 displays a model of minor groove of B-, B/A, and A-DNA helices. It is clear that B-DNA helices (left panel) have narrow minor groove that may enable proteins (presumably HMGB1b) to fit firmly into the groove. The intermediate structure B/A possesses a broader minor groove. I hypothesize that single-PNA structures have a minor groove topology that is more similar to the intermediate B/A model. Hence, HMGB1b fits firmly into this minor groove and forms discreet 4:1 complexes similar to J1. This viewpoint is strengthened by NMR data of PNA-DNA duplexes; these hybrid PNA-DNA duplexes possess structural features intermediate between B- and A-DNA (as shown in Fig. 13). With respect to the multi-PNA hybrids, the CD data shows that these constructs are composed largely of A-DNA helices. Hence, we hypothesize the multi-PNA 4WJs have a broader minor groove that is more similar to the A-DNA structure shown in Fig.
27 (right panel). If this is the case, HMB1b does not fit firmly into the minor groove.

Figure 27. Space filling 3D models of B-DNA, A-B intermediate DNA, and A-DNA.

With respect to the thermostability of hybrid 4WJs vs. J1, thermal melting curves (Fig. 24) show that high Mg$^{2+}$ enhances the stability of five of the six hybrid junctions examined (Table 1). The single-PNA hybrids showed a noticeable correlation between ionic strength of the solvent (H$_2$O), B-form helical content, and $T_m$. The single-PNA constructs favor B-form helices and possess enhanced $T_m$ values in high Mg$^{2+}$ compared to the multi-PNA junctions. This correlation is most evident for b4WJ-PNA$_3$. In the presence of high Mg$^{2+}$ ion concentration, b4WJ-PNA$_3$ possesses higher B-form structure (AF$_B$) and $T_m$ values than the DNA control, J1. This effect may again be due to an increase in solvent accessibility of Mg$^{2+}$ for the major grooves of the single-PNA junctions reducing charge repulsion between DNA strands and the forming of a more organized solvent cage around the junctions.$^{126,129,130}$ Support for this hypothesis

64
is provided by molecular dynamics (MD) simulations data that show B-DNA helices are preferred in high activity solvent(s) due to favorable solvation free energy values.\textsuperscript{126,130} The corresponding solvation free energy values are considerably smaller for A-DNA helices.\textsuperscript{126} The addition of uncharged PNA strands to the 4WJs may provide a more hydrophobic environment near their complementary DNA strands, reducing the solvation free energies. This results in hybrid PNA-DNA 4WJs having more A-form like helices compared to J1, as evidenced by the CD results (Fig. 22). In high ionic strength environments (i.e. 2mM Mg\textsuperscript{2+}), counter ion interactions with the DNA phosphate backbone may be strengthened as a result of an increased extent of counter ion condensation in the more hydrophobic environment produced by the PNA strands. This is evidenced by the increase in T\textsubscript{m} for most junctions in 2mM Mg\textsuperscript{2+} ion concentrations (Table 1).

4.2 Conclusion

Both single-PNA junctions b4WJ-PNA\textsubscript{1} and b4WJ-PNA\textsubscript{3} bind HMGB1b with similar affinity to pure DNA 4WJs. Replacement of a single J1 strand with a PNA did not seem to perturb HMGB1 binding affinity. While the single PNA constructs b4WJ-PNA\textsubscript{1} and b4WJ-PNA\textsubscript{3} were shown to bind HMGB1 with similar affinities to that of the parent DNA 4WJ, J1, the multi-PNA junctions showed a marked decrease in binding affinity and did not form stable complexes. This may be a result of the increased freedom of movement imparted by lack of electrostatic repulsion between the DNA strands and the uncharged PNA backbone. HMGB1b presumably binds each junction across the branch point.
The increased freedom of motion may cause a deviation in the canonical *Jtwist* angle at the branch point and/or a deviation in helical structure. This deviation in *Jtwist* angle may limit access of HMGB1b to the junction branch point or decreasing the stability of protein/junction complexes formed between HMGB1b and the multi-PNA hybrids.

The strand replacement and end-blunting strategy showed success in increasing the thermal stability of b4WJ-PNA3 vs. J1 under conditions used to mimic average ECM Mg\(^{2+}\) ion concentration (i.e. 2.00mM). At a Mg\(^{2+}\) ion concentration of 2.00mM, junction b4WJ-PNA3 showed an increase in T_m vs. J1 of 1.14°C and a large increase in T_m vs. 4WJ-PNA3 of 8.61°C. At a 2.00mM Mg\(^{2+}\) ion concentration, single-PNA hybrid junction b4WJ-PNA3, with a T_m of 41.18°C, was shown to remain stable well above normal body temperature of 37°C and slightly above high fever temperatures of 40°C.

The results of this study suggest that hybrid PNA-DNA 4WJs warrant further investigation as possible therapeutic agents against unwanted HMGB1 cytokine activity. Junction b4WJ-PNA3 was shown capable of binding HMGB1b with an affinity similar to that of its parent DNA junction and has a T_m well above normal body temperature. This study suggests that b4WJ-PNA3 may successfully target HMGB1 *in vivo* and that b4WJ-PNA3 may remain thermally stable in the body even at elevated temperatures often associated with inflammation.

Future plans are to examine the consequences of replacing strands 102 and 104. In the stacked-X conformation strands 101 and 103 represent the continuous strands. Replacing one or both of the crossover strands may have
much different effects on local and global conformation as well as junction stability. Additionally, studies are planned to examine the serum stability and nuclease resistance of hybrid 4WJs. The formulation of junctions containing chimeric strands composed of both PNA and DNA is being examined as a possible strategy for improving junction stability while maintaining substrate viability.
5.1 Synopsis and Hypotheses

The genetic code is a cornerstone of molecular and cell biology that has fascinated scientists for decades. Despite having a clear grasp of the basic features of the code, several key questions about its origin remain unsolved. Among these questions, the chronology of amino acid (AA) emergence and codification are largely unresolved. The objective of Chapters V through VIII is to use experimental data to provide evidence to support current genetic code origin theories. Aminoacyl-tRNA synthetases (aaRSs), the enzymes responsible for production of aminoacyl-tRNAs (aa-tRNAs) in the first step of protein translation, are used to perform a selection survey of the 20 natural AAs. Several previous studies have suggested that aaRSs are among the oldest families of proteins.\(^\text{131-134}\) It has also been suggested that ancestral aaRSs arose during a period of an incomplete genetic code (i.e. fewer than 20 AAs) and initially had little AA specificity.\(^\text{131,134}\) These suppositions imply that ancestral aaRSs initially developed specificity toward an incomplete or primitive set of AAs. I hypothesize that extant aaRSs may retain a selection preference for the substrates of their ancestors and that the pool of AAs selected by multiple aaRSs represent an early set AAs used in primitive translation.

In this study, a radiometric assay is used to probe for an AA activation imprint (i.e. a selection preference) that is inherent to certain AAs. Using experimental techniques focused on elucidating connections between AAs and
extant aaRSs may provide novel insight into the process of genetic code development. By examining AA selection in the form of misactivation profiles we will provide empirical evidence to support a chronology of AA codification. Fig. 28 shows the two-step synthesis of aa-tRNAs. In Fig. 28 step 1 an AA and a molecule of ATP react to form an aminoacyl-adenylate (AA-AMP) with release of a diphosphate (PPi). In Fig 28 step 2 the AA moiety of the AA-AMP is transferred to a tRNA to form an aa-tRNA. This study focuses solely upon the activation reaction shown in step 1 of Fig. 28 and examines the AA discriminating ability of each aaRS. The omission of tRNAs provided an opportunity to examine if certain AAs are structurally and/or chemically predisposed to reacting with aaRSs. Moreover, these investigations revealed potential biosynthetic imprints between AAs and aaRSs in the form of selection preferences, that exist in the absence of tRNAs, and suggest a relationship between aaRS specificity and AA biosynthesis and codification. These findings were compared to genetic code origin studies examining AA abundance in early Earth simulations\textsuperscript{135-138} and meteorites\textsuperscript{139-142}, physiochemical properties of AAs\textsuperscript{143-145}, and AA biosynthetic pathways\textsuperscript{146-149}.

**Step 1:** Activation

\[
\text{AA} + \text{ATP} + \text{AARS} \rightarrow \text{AARS-\text{AA-AMP} + PPi}
\]

**Step 2:** Charging

\[
\text{AARS-\text{AA-AMP} + tRNA} \rightarrow \text{AARS + aa-tRNA + AMP}
\]

**Figure 28.** A schematic of the aaRS catalyzed aminoacylation reaction

Step 1 shows the activation of an AA to produce AA-AMP. Step 2 shows the charging of the AA moiety of the AA-AMP to a tRNA to produce an aa-tRNA.
5.2 Code Origin Theories

5.2.1 Connections between AA Emergence and the Early Earth

The original simulated primitive earth experiments performed by Miller and Urey reportedly produced a variety of organic compounds including AAs such as: Ala, Gly, Val, Asp, and Glu.\textsuperscript{136} The Miller-Urey experiments have been rigorously scrutinized and modified to yield similar results under a wide range of possible conditions.\textsuperscript{135,137,138,150-152} The original experiments used electric discharges on a mixture of hydrogen (H\textsubscript{2}), H\textsubscript{2}O, methane (CH\textsubscript{4}), and ammonia (NH\textsubscript{3}) to produce organic compounds.\textsuperscript{136,150} Subsequent experiments incorporated other gases such as carbon dioxide (CO\textsubscript{2}), nitrogen (N\textsubscript{2})\textsuperscript{137,150}, and hydrogen sulfide (H\textsubscript{2}S)\textsuperscript{151} and an aspirator to inject steam\textsuperscript{138}. These simulated primitive earth experiments have produced a variety of organic compounds, including simple AAs. The combined results of four such analyses are shown in Table 5.\textsuperscript{135,137,138,150}
Table 5

Prebiotic AAs produced in early Earth simulations

<table>
<thead>
<tr>
<th>Amino Acid identified in Early Earth Simulations</th>
<th>Relative Abundance (Gly = 100)</th>
<th>Amino Acid identified in Early Earth Simulations</th>
<th>Relative Abundance (Gly = 100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>α Glycine</td>
<td>100</td>
<td>α N-isopropyl glycine</td>
<td>0.5</td>
</tr>
<tr>
<td>α Alanine</td>
<td>180</td>
<td>α N-ethyl-β-alanine</td>
<td>0.5</td>
</tr>
<tr>
<td>α Amino-α-butyric acid</td>
<td>61</td>
<td>α Isotamine</td>
<td>0.8</td>
</tr>
<tr>
<td>α Hydroxy-α-amino butyric acid</td>
<td>17</td>
<td>α Isovaline</td>
<td>0.99</td>
</tr>
<tr>
<td>α Valine</td>
<td>14</td>
<td>α Amino-n-heptanoic acid</td>
<td>0.3</td>
</tr>
<tr>
<td>α Aspartic acid</td>
<td>12.3</td>
<td>α-Proline</td>
<td>0.3</td>
</tr>
<tr>
<td>α α,β-diaminobutyric acid</td>
<td>7.7</td>
<td>α Threonine</td>
<td>0.2</td>
</tr>
<tr>
<td>α Amino-α-butyric acid</td>
<td>7.6</td>
<td>α Allothreonine</td>
<td>0.2</td>
</tr>
<tr>
<td>α N-ethyl glycine</td>
<td>6.8</td>
<td>β-Amino-n-butyric acid</td>
<td>0.1</td>
</tr>
<tr>
<td>Valine</td>
<td>4.4</td>
<td>α Amino-α-butyric acid</td>
<td>0.1</td>
</tr>
<tr>
<td>β-Valine</td>
<td>4.3</td>
<td>N-ethyl alanine</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>N-methyl alanine</td>
<td>3.4</td>
<td>β-Hydroxy aspartic acid</td>
<td>0.01</td>
</tr>
<tr>
<td>Leucine</td>
<td>2.6</td>
<td>β-Pipolic acid</td>
<td>0.01</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>1.7</td>
<td>α-Ethionine</td>
<td>0.01</td>
</tr>
<tr>
<td>α,β-diaminopropionic acid</td>
<td>1.5</td>
<td>α Ornithine</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Alloisoleucine</td>
<td>1.2</td>
<td>α Lysine</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Serine</td>
<td>1.1</td>
<td>α T-Leucine</td>
<td>&lt;0.005</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>1.1</td>
<td>β 2-Methyl serine</td>
<td>0.0016</td>
</tr>
<tr>
<td>Norleucine</td>
<td>1.4</td>
<td>α Methionine sulfide</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Isoine</td>
<td>2</td>
<td>α-Amino adipic acid</td>
<td>0.00038</td>
</tr>
<tr>
<td>α Methyl α-α-mono-α-butyric acid (isoaline)</td>
<td>1</td>
<td>α Homoserine</td>
<td>0.00034</td>
</tr>
<tr>
<td>α N-methyl-β-alanine</td>
<td>1.0</td>
<td>β 2-methyl glutamic acid</td>
<td>0.00024</td>
</tr>
<tr>
<td>γ Aminobutyric acid</td>
<td>0.9</td>
<td>α Phenylalanine</td>
<td>0.0005</td>
</tr>
<tr>
<td>Homocysteine</td>
<td>0.5</td>
<td>α Methionine sulfone</td>
<td>&lt;0.0002</td>
</tr>
<tr>
<td>α N-propyl glycine</td>
<td>0.5</td>
<td>α 5-Methylcysteine</td>
<td>&lt;0.0002</td>
</tr>
</tbody>
</table>

AAs are listed according to the highest relative abundance compared to Gly (Gly = 100)^135,137,138,150.

Many of the same organic compounds produced in early Earth simulations have also been found in studies of carbonaceous meteorite compositions. Studies examining the compositions of carbonaceous meteorites have shown them to be rich sources of organic compounds, including a large number of AAs. The Murchison^139,141^, Orgueil^140^, and Murray^142^ are well studied carbonaceous meteorites shown to contain many of the same canonical proteinacious AAs seen in the early Earth simulations including Gly, Ala, Val, Pro, Ser, Asp, and Glu. A list of AAs found in examinations of these meteorites is shown in Table 6.139-142

Early Earth simulations and meteorite examinations provide evidence for the existence of prebiotic AA sources from which a primitive translation system could be built upon. The identification of proteinacious AAs in early Earth simulations and meteorite examinations provides evidence for the early availability of these
AAs, but does not identify them as early coded AAs. However, studies examining genetic code origins often point to many of the AAs identified in these early Earth simulations and meteorite examinations as early coded AAs.147-149,153-162

Table 6

**AAs identified in carbonaceous meteorites**

<table>
<thead>
<tr>
<th>Amino Acid Identified in Carbonaceous Meteorites</th>
<th>1,2-methyl aspartic acid</th>
<th>3-amino-2-ethylpropanoic acid</th>
<th>Leucine</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-Amino-2-ethyl-3-methylbutanoic acid</td>
<td>3-amino-2-methylbutanoic acid</td>
<td>Norleucine</td>
<td></td>
</tr>
<tr>
<td>2-amino-2-ethylbutyric acid</td>
<td>3-amino-2,2-dimethylpropanoic acid</td>
<td>Norvaline</td>
<td></td>
</tr>
<tr>
<td>2-amino-2-ethylpentanoic acid</td>
<td>3-aminopentanoic acid</td>
<td>Proline</td>
<td></td>
</tr>
<tr>
<td>2-amino-2-methylhexanoic acid</td>
<td>3,3 methyl aspartic acid</td>
<td>Pseudoleucine</td>
<td></td>
</tr>
<tr>
<td>2-amino-2,3-dimethylbutyric acid</td>
<td>4-amino-2-methylbutanoic acid</td>
<td>Valine</td>
<td></td>
</tr>
<tr>
<td>2-amino-2,3-dimethylpentanoic acid</td>
<td>4-amino-3-methylbutanoic acid</td>
<td>Serine</td>
<td></td>
</tr>
<tr>
<td>2-amino-2,3,3-trimethylbutanoic acid</td>
<td>4-aminopentanoic acid</td>
<td>Threonine</td>
<td></td>
</tr>
<tr>
<td>2-amino-2,4-dimethylpentanoic acid</td>
<td>4-aminopentanoic acid</td>
<td>Valine</td>
<td></td>
</tr>
<tr>
<td>2-amino-3-ethylpentanoic acid</td>
<td>4-aminopentanoic acid</td>
<td>Serine</td>
<td></td>
</tr>
<tr>
<td>2-amino-3-methylhexanoic acid</td>
<td>4-aminopentanoic acid</td>
<td>Threonine</td>
<td></td>
</tr>
<tr>
<td>2-amino-3,3-dimethylpentanoic acid</td>
<td>4-aminopentanoic acid</td>
<td>Valine</td>
<td></td>
</tr>
<tr>
<td>2-amino-3,4-dimethylpentanoic acid</td>
<td>4-aminopentanoic acid</td>
<td>Valine</td>
<td></td>
</tr>
<tr>
<td>2-amino-4,4-dimethylpentanoic acid</td>
<td>4-aminopentanoic acid</td>
<td>Valine</td>
<td></td>
</tr>
<tr>
<td>2-amino-5-methylhexanoic acid</td>
<td>4-aminopentanoic acid</td>
<td>Valine</td>
<td></td>
</tr>
<tr>
<td>2-aminoheptanoic acid</td>
<td>4-aminopentanoic acid</td>
<td>Valine</td>
<td></td>
</tr>
<tr>
<td>2-methyl glutamic acid</td>
<td>4-aminopentanoic acid</td>
<td>Valine</td>
<td></td>
</tr>
<tr>
<td>2-methyl norvaline</td>
<td>4-aminopentanoic acid</td>
<td>Valine</td>
<td></td>
</tr>
</tbody>
</table>

All studies employed mass spec coupled GC to examine material from the inner portion of fragments from each meteorite. Amino acids denoted with an a or b were also found in the Orgueil and Murray meteorites, respectively 139-142.

5.2.2 Connections between AA Emergence and the Coevolution Theory (CET)

There are a many theories to explain the establishment of the genetic code, some of the most notable are: the coevolution theory (CET)147-149,153,155, the stereochemical hypothesis157,163,164 and error minimization hypothesis 158,159,165. These theories are not mutually exclusive and each supposition likely played a role in the formation of the genetic code.149,153,154

The CET predicts that the structure of the genetic code is an imprint of these biosynthetic pathways and the precursor product relationships between AAs.147 The CET assumes that the genetic code was initially comprised of a
smaller set of precursor AAs (phase 1 AAs) encoded by proto-tRNAs. As primitive biosynthetic pathways became more complex, the evolution of precursor AAs to product AAs (phase 2 AAs) was facilitated by these proto-tRNAs. Genetic code structure then becomes a consequence of precursor AAs conceding codons to product AAs. Support for this theory lays in the extant precursor/product relationships of AAs and the similarity of their codons. Figure Fig. 29 shows a proposed evolutionary map of the genetic code development with arrows denoting extant AA precursor/product relationships (single-headed arrows) and interconversion relationships (double-headed arrows). Codons shown in solid boxes correspond to the extant use in the genetic code. Codons in the dotted boxes for Glu and Asp were presumably conceded to product Gln and Asn in the later stages of codon assignment. The codons of each precursor/product pair are contiguous, differing by only a single nucleotide. This mechanism of codon assignment proposed in the CET assumes that primordial AAs each occupied continuous regions of the genetic code and that codons from this region were conceded to product AAs consistent with the observed contiguity of precursor AA and product AA codons. From the biosynthetic and codon relationships shown in Fig. 29, a set of central AAs emerges from which all others seemingly evolved and allows the generation of a basic chronology of AA codification. This set of seven precursor AAs comprised of Glu, Asp, Ala, Ser, Gly, Phe, and Val is presumed by the CET to be the earliest set of coded AAs. These AAs are found in relatively high abundance in early Earth and meteorite studies giving support to their “early” status.
5.2.3 Connections between AA Emergence and AA Biosynthetic Pathways

The investigation of AA biosynthetic pathways has been used to make inferences of the relative ages of AAs and aaRSs. It has been proposed that class II aaRSs and their respective AAs are, as a group, more ancient than class I aaRSs and AAs. This proposal was based on the AA precursor/product relationships and distances from the glycolysis pathway, pentose pathway, and citric acid cycle. Fig. 30 shows a schematic of the AA biosynthetic pathways showing the AAs and their respective codons. AAs specific to class I aaRSs are
shown in white and AAs specific to class II aaRSs are shown in grey. While it is true that a higher number of class II specific AAs occupy interior nodes of the diagram, several of the class I AAs occupy positions very close to the glycolysis pathway and citric acid cycle. Also, the class I AA Glu is an important precursor in the biosynthesis of Pro, a class II AA. Though the primary goal of this previous AA biosynthetic pathway study was to examine the relative ages of the aaRS classes, it also provided insight into the chronology of AAs. Fig. 30 illustrates the extant precursor/product relationships between AAs and shows their proximity to glycolysis, the pentose phosphate pathway, and the citric acid cycle. The extant machinery involved in these biosynthetic pathways is undoubtedly quite different from that found in early metabolisms. However, the precursor/product relationships of the AAs may be an artifact of an early metabolic system. The precursor product relationships shown in Fig. 30 were used to produce a three stage chronology of AA codification with Ala, Val, Gly, Ser, Asp, and Glu as more ancient; Asn, Thr, Gln, and Pro at an intermediary stage; and Leu, Phe, Tyr, Trp, His, Cys, Met, Lys, Ile, and Arg as more recent.146
5.2.4 Connections between AA Emergence and AA Physiochemical Properties

Studies examining physiochemical properties of AAs have revealed AA stability, polarity, and complexity as factors that correlate strongly with previously discussed code origin theories.\textsuperscript{143-145} One study compared the free energies of the 20 natural AAs to sets of corresponding isomers.\textsuperscript{166} Comparing the free energies of formation of the 20 natural AAs to one another showed a set of particularly stable AAs (Gly, Ala, Ser, Pro, Val, Thr, Asp, and Glu) that largely correspond to the set considered ancient by early Earth simulations, meteorite studies, and the CET. Larger AAs His, Phe, Arg, Tyr, and Trp were shown to be
the least stable in terms of free energy of formation.\textsuperscript{166} Another study examined a set of 40 criteria and hypotheses associated with genetic code development.\textsuperscript{144} This study identified thermal stability of codon-anticodon interactions, codon complementarities, abiotic origin, and AA size/complexity as major factors that correlate with AA age. This study was used to produce a chronological ranking of AAs from earliest to late as follows Gly/Ala, Val/Asp, Pro, Ser, Glu/Leu, Thr, Arg, Asn, Lys, Gln, Ile, Cys, His, Phe, Met, Tyr, Trp.\textsuperscript{144}

All of the theories discussed here assume a gradual progression of genetic code development. During this period, codons would not have been specific but instead coded for a set of related AAs producing a non-degenerate but functional code. The presence of a primitive code, even an ambiguous code with a smaller set of AAs, suggests the presence of primitive proteins.\textsuperscript{167,168} The near ubiquitous nature of aaRSs throughout the three kingdoms of life suggests they are very ancient. If aaRS ancestors existed during the period of this primitive code, it is plausible that aaRSs may have played a role in AA codification. It is conceivable that extant aaRS retain recognition features from this era that will be evident in the selection preference for the prebiotic AAs identified in the early Earth simulations and meteorite studies.

5.3 Aminoacyl tRNA-Synthetases (aaRSs)

5.3.1 Aminoacyl-tRNA Synthesis

AaRSs are highly specific enzymes that are sentinels of the genetic code. Researchers have defined a clear two-step mechanism to describe the role of aaRSs in the selection and subsequent attachment of AAs onto their cognate
tRNA. The reaction mechanisms for activation and tRNA charging are displayed in Fig. 31. In the first step (Fig 31A), an aaRS transfers AMP (from ATP) onto an AA to form a highly energetic mixed anhydride AA-AMP. Next, the aaRSs attaches the adenylate to the 3' terminal adenosine of a tRNA. The class II aaRSs attach the AA directly to the 3' hydroxy (Fig 31C). Class I aaRSs initially attach AAs to the 2' hydroxy (Fig. 31B), followed by transesterification to the 3' hydroxyl (Fig 31D). *In vitro* studies have shown that aaRSs possess a discrimination factor of 200 – 10,000 during activation; the stringency factor increases to > 10,000 during charging.\textsuperscript{169,170}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure31.png}
\caption{Aminoacyl-tRNA synthesis}
\end{figure}
Two step reaction: (A) Aminoacyl-adenylate formation (B) Transfer of the activated amino acid to the tRNA 3'-CCA terminus 2'-OH by Class I synthetase. C) Transfer of the activated amino acid to the tRNA 3'-CCA terminus 3'-OH by Class II. D) 2'-aminoacyl esters undergo a transesterification to the 3'-OH of tRNA.

5.3.2 aaRSs Role in Genetic Code Development

It has been suggested that ancestral aaRSs were initially tasked with enhancing charging rates in primitive translation with little specificity toward a particular amino acid. Under this assumption aaRS development could have followed two possible scenarios: i) aaRSs developed AA specificity in the presence the genetic code in its current form ii) aaRS specificity coevolved with the genetic code. The first scenario implies that a fully-fledged genetic code existed in an RNA based translation system prior to the advent of peptides that assisted in RNA charging (proto-aaRSs). The second scenario implies the existence of a primitive genetic code, similar to the operational codes proposed by Schimmel et. al. and Rodin et. al., capable of producing functional peptides. Further more, it implies that aaRSs were present during the period of code expansion and/or ambiguity reduction. Under this second scenario, aaRSs may have played a role in determining the structure of the genetic code. This may be observable in the interactions between extant aaRSs and AAs, specifically in the preferential selection (i.e. misactivation) of more ancient AAs by extant aaRSs. Here, AA misactivation profiles are compared with AA chronologies based on: AA abundance in early Earth simulations, AA biosynthetic pathways, and physiochemical properties of AAs and codons/anticodons. I hypothesize that extant aaRS will show a selection preference for the prebiotic AAs identified in the early Earth simulations and
meteorite studies, and for the precursor AAs identified in the AA biosynthesis study and CET.

5.3.3 Classes of aaRSs

The structures of aaRSs are separated by the different dinucleotide binding domains of their active sites.\(^{134,175-178}\) The aaRSs from \textit{E. coli} are listed in Table 7; each sub-class is color-coded. The majority of bacterial class I aaRSs are monomeric with the exception of TrpRS and TyrRS; these synthetases are dimers composed of identical subunits (\(\alpha_2\) dimers).\(^{177,179,180}\) The majority of bacterial class II synthetases are \(\alpha_2\) dimers, with the exception of GlyRS, PheRS, and AlaRS. GlyRS and PheRS are \(\alpha_2\beta_2\) tetramers while AlaRS is an \(\alpha_4\) tetramer.\(^{179,181,182}\) Fig. 32A and B display the crystal structures of the class I synthetase GlnRS and class II synthetase AspRS from \textit{E. coli}. The monomeric structure of GlnRS\(^{183}\) and dimeric structure of AspRS\(^{184}\) are typical of each class of synthetase.
Table 7

The Aminoacyl-tRNA Synthetases of E. coli

<table>
<thead>
<tr>
<th>Class I aaRS</th>
<th>Sub-class</th>
<th>Structure</th>
<th>Class II aaRS</th>
<th>Sub-class</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>ArgRS</td>
<td>la</td>
<td></td>
<td>AlaRS</td>
<td>IIa</td>
<td>4</td>
</tr>
<tr>
<td>CysRS</td>
<td>la</td>
<td></td>
<td>ProRS</td>
<td>IIa</td>
<td>2</td>
</tr>
<tr>
<td>IleRS</td>
<td>la</td>
<td></td>
<td>SerRS</td>
<td>IIa</td>
<td>2</td>
</tr>
<tr>
<td>LeuRS</td>
<td>la</td>
<td></td>
<td>ThrRS</td>
<td>IIa</td>
<td>2</td>
</tr>
<tr>
<td>MetRS</td>
<td>la</td>
<td></td>
<td>HisRS</td>
<td>IIa</td>
<td>2</td>
</tr>
<tr>
<td>ValRS</td>
<td>la</td>
<td></td>
<td>GlyRS</td>
<td>IIb</td>
<td>2 2</td>
</tr>
<tr>
<td>GluRS</td>
<td>lb</td>
<td></td>
<td>AsnRS</td>
<td>IIb</td>
<td>2</td>
</tr>
<tr>
<td>GlnRS</td>
<td>lb</td>
<td></td>
<td>AspRS</td>
<td>IIb</td>
<td>2</td>
</tr>
<tr>
<td>TyrRS</td>
<td>lc</td>
<td>2</td>
<td>LysRS</td>
<td>IIb</td>
<td>2</td>
</tr>
<tr>
<td>TrpRS</td>
<td>lc</td>
<td>2</td>
<td>PheRS</td>
<td>IIc</td>
<td>2 2</td>
</tr>
</tbody>
</table>

Figure 32. Crystal structures of class I and II aminoacyl-tRNA synthetases

A) X-ray crystal structure of the monomeric Class I synthetase GlnRS at a resolution of 2.6 Å [PDB: 1NYL]183. B) X-ray crystal structure of the Class II synthetase AspRS α2 dimer at a resolution of 2.7 Å [PDB: 1EQR]184.

5.3.3.1 Class I aaRSs. Class I synthetases contain a Rossmann fold nucleotide binding domain within their active sites.179,180,182 A Rossmann fold is composed of two alternating β-α-β-α-β structural motifs that form a parallel β-
sheet. In aaRSs the fold is slightly altered substituting two $\alpha$-$\beta$-$\alpha$-$\beta$-$\alpha$ motifs instead. The Rossmann fold domains of most class I synthetases contain two characteristic motifs (KMSKS and HIGH). The KMSKS and HIGH motifs are named for their amino acid sequences. The KMSKS and HIGH motifs have been implicated in the stabilization of the aminoacyl adenylate. Certain class I synthetases contain one or two large connective peptide (CP) insertion domains (CP1 and CP2) that bisect the Rossmann fold domains, thought to participate in editing function. Fig. 33 shows the Rossmann fold of GlnRS from E.coli with the HIGH and KMSKS motifs are highlighted in green and red. The CP1 insertion domain of GlnRS is highlighted in purple.

Figure 33. Class I Rossmann Fold of GlnRS from E.coli

X-ray crystal structure of the class I Rossmann Fold of GlnRS from E.coli at a resolution of 2.6 Å. The HIGH motif is shown in green, the MSK motif is shown in red, and the CP1 insertion domain is shown in purple [PDB: 1NYL].
5.3.3.2 Class II aaRSs. The structure of class II aaRSs is more varied; these enzymes contain an atypical anti-parallel \( \beta \)-fold nucleotide-binding domain and three conserved structural motifs.\(^{181,182,188,189}\) Fig. 34 displays these structural features for AspRS from E. coli.\(^{184}\) Motifs 1, 2, and 3 are highlighted in red, green, and blue. Each motif has been shown to facilitate the assembly of the dimer interface and ATP binding site.\(^{181,182,188,189}\) Motif 1 (red) is composed of a long \( \alpha \)-helix and a short \( \beta \)-strand; motif 2 (green) comprises the first two strands of the anti-parallel \( \beta \)-sheet nucleotide-binding domain and motif 3 (blue) contains the central strand of the anti-parallel \( \beta \)-sheet followed by a short helix.

Figure 34. Class II Beta Fold of AspRS from *E. coli*

X-ray crystal structure of the class II Beta Fold of AspRS from *E. coli* at a resolution of 2.7 Å. Motif 1 is shown in red, Motif 2 is shown in green, and Motif 3 is shown in blue [PDB: 1EQR]\(^{184}\)
Fig. 35 shows a 3 dimensional view of the L-shaped structure $\text{tRNA}^{\text{Phe}}$ from yeast.\textsuperscript{190} Evidence from crystal structures containing tRNA shows a possible correlation in the way each class of aaRS approaches the tRNA acceptor stem.\textsuperscript{191,192} Class I aaRSs approach from the minor groove side leaving the variable loop (Fig. 35 shown in orange) facing away from the binding site and toward the solvent. Class II aaRSs, on the other hand, approach the major groove side resulting in the variable loop facing toward the binding pocket.

Figure 35. An X-ray crystal structure of $\text{tRNA}^{\text{Phe}}$ from yeast

An X-ray crystal structure of $\text{tRNA}^{\text{Phe}}$ from yeast at a resolution of 1.93 Å. The acceptor stem is shown in red. The T loop is shown in green. The D loop is shown in purple. The variable loop is shown in orange. The anticodon loop is shown in blue with the $G_m$AA anticodon in yellow. [PDB: 1EHZ]\textsuperscript{190}.

While the amino acid and nucleotide binding strategies are fairly well conserved among each class of aaRS, tRNA binding relies on a more
complicated set of tRNA structural and chemical features termed identity elements that seem to be unique to each aaRS.\textsuperscript{177,191-194} Each tRNA contains a specific set of identity elements.\textsuperscript{192} Identity elements are a set of tRNA bases that are required for recognition by its corresponding aaRS. The primary tRNA identity elements for all 20 aaRSs from various species have been established using studies involving site-directed mutagenesis of tRNAs.\textsuperscript{191-193} A list of the identity elements found in \textit{E. coli} tRNAs is shown in Table 8.\textsuperscript{193} The most common identity elements are found in the acceptor stem and anticodon loop (blue and yellow, Fig. 35), but nucleotide residues in the D stem (purple Fig. 35), T loop (green, Fig. 35), and variable arm have also been shown to be strong determinants. Several modified bases present in various tRNAs at different locations have also been established as vital to cognate tRNA recognition.\textsuperscript{191,193,194}
Table 8

Knowledge identity elements of E. coli tRNAs

<table>
<thead>
<tr>
<th>Identity Elements of E. coli tRNAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>tRNA</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>Acceptor Stem</td>
</tr>
<tr>
<td>Val</td>
</tr>
<tr>
<td>Ile</td>
</tr>
<tr>
<td>U12:A23, C29:G41</td>
</tr>
<tr>
<td>Leu</td>
</tr>
<tr>
<td>Met</td>
</tr>
<tr>
<td>Cys</td>
</tr>
<tr>
<td>G15..G48, A13..A22</td>
</tr>
<tr>
<td>Tyr</td>
</tr>
<tr>
<td>Trp</td>
</tr>
<tr>
<td>Glu</td>
</tr>
<tr>
<td>U11:A24, U13:G22-A46, Δ47</td>
</tr>
<tr>
<td>Gln</td>
</tr>
<tr>
<td>G10</td>
</tr>
<tr>
<td>Arg</td>
</tr>
<tr>
<td>His</td>
</tr>
<tr>
<td>Asp</td>
</tr>
<tr>
<td>G10</td>
</tr>
<tr>
<td>Lys</td>
</tr>
<tr>
<td>Asn</td>
</tr>
<tr>
<td>Phe</td>
</tr>
<tr>
<td>U20, G44, U45, U59, U60</td>
</tr>
<tr>
<td>Ala</td>
</tr>
</tbody>
</table>

X:Y denotes standard Watson-Crick base pairing, X:.Y denotes non-Watson-Crick base pairing, X/Y denotes that either base can occupy this position, Δ denotes the absence of a residue, L denotes Lysidine, R denotes a purine, Y denotes a pyrimidine, t^6A is threonylcarbamoyl adenosine, mmn^5s^2U is 5-methylamino-methyl 2-thiouridine.

184,192,193,195,196
5.3.4 Fidelity of aaRSs

The aaRSs are capable of selecting their cognate amino acid from a cellular pool of many chemically/structurally similar molecules with high efficiency, producing mischarging rates of <10^{-5} errors per charging event. Although initial selectivity plays a major role in excluding non-cognate amino acids, in most cases initial screenings events do not prevent production of misactivated adenylates. Editing functions that hydrolyze misactivated or misacylated products supplement substrate selectivity.

The double sieve model is widely accepted as the model to describe the high fidelity of AA activation and tRNA charging. The model proposes the existence of secondary active sites capable of expelling non-cognate substrates that subvert the primary active site. Many aaRSs can expel misactivated AAs before they are charged onto tRNAs – a process referred to as pre-transfer editing. Many aaRSs are also capable of expelling misacylated tRNAs – a process referred to as post-transfer editing. Pre-transfer editing can act in a tRNA independent or tRNA dependent manner to expel misactivated AAs. Post-transfer editing mechanisms require tRNA to initiate editing or removal of misacylated AAs from non-cognate tRNAs.

As stated earlier, the initial selectivity of AAs by aaRSs in the activation step of aa-tRNA synthesis is the focus of this study. This allows us to determine if certain AAs are structurally or chemically predisposed to reacting with aaRSs. AAs that are selected by the largest number of aaRSs presumably possess a biological imprint that suggests this pool of AAs is more ancient AAs. Finally, we
compare are AA lineage to those from: early Earth simulations, meteorite composition, the coevolution theory, and AA biosynthetic pathways.

5.4 Radiometric Amino Acid Activation Assay

Several radiometric assays have been developed that provide the sensitivity and reliability required for an accurate assessment of enzymatic amino acid activation and tRNA aminoacylation. Assays that use $[^3H]$ or $[^{14}C]$ labeled amino acids and $[^{32}P]$ labeled tRNAs have shown great success in measuring steady state and pre-steady state kinetics of AA activation and tRNA charging. However, these techniques have drawbacks. Assays that rely on the radiolabeled amino acid or tRNA to be captured on filter paper or charcoal are prone to error from product loss on the filters or during washing. Moreover, laborious washing steps are required, resulting in low throughput. Determining the fraction of activated or aminoacylated product also requires knowing the aliquot size and quantum efficiency of the scintillation counter on the filter medium. Moreover, prohibitive levels of $[^3H]$ or $[^{14}C]$ labeled amino acids are required to investigate aaRSs that possess low $K_M$ values (high binding affinity) toward their cognate AA.

More recently, sensitive and straightforward assays have been developed that use $[\alpha-^{32}P]ATP$ as a label to monitor activation. Fig. 36 shows the potential pathways of the radiolabel during activation (shown in red). In Step 1 the ATP is esterified to an amino acid to produce an AA-AM$[^{32}P]$. Step 2 shows tRNA independent pre-transfer editing of the resulting AA-AMP to yield the free amino acid and AMP. Using $[\alpha-^{32}P]ATP$ allows direct monitoring of ATP, AMP,
and AA-AMP being consumed or generated in a typical reaction. Products are separated using thin layer chromatography (TLC), and can be viewed and measured with the aid of phosphor imaging tools. Compared to other radiometric methods this process requires fewer steps, employs a much more efficient means of product separation, and can also been scaled up using a 96 well-plate format for higher throughput.

Figure 36. Possible pathways of the activation reaction

Step 1 shows activation of an AA to form AA-AMP. Step 2 illustrates tRNA independent pre-transfer editing resulting in the AA-AMP.

An illustration of the TLC separation pattern is presented in Fig. 37. The top spot corresponds to AA-AMP, the middle spot corresponds to AMP, and the bottom spot corresponds to unreacted ATP. The presence of activation is characterized by the appearance of AA-AMP with signal intensity above that of a negative control containing no AA. As shown in Fig. 37 (inset), AA-AMP concentration can be calculated by dividing the AA-AMP signal by the total signal to find the percentage of ATP converted to AA-AMP (%AA-AMP) and multiplying...
this percentage by the amount of ATP in the assay. This alleviates the need for many of the calculations required when using a scintillation counter (i.e. quantum efficiency calculations and decay corrections).

\[
\text{% AA-AMP} = \frac{\text{AA-AMP}}{\text{(AA-AMP + AMP + ATP)}} \times 100.
\]

\[
\text{AA-AMP nM} = \% \text{ AA-AMP} \times \text{ATP}_{\text{Tot}} \text{(nM)}
\]

Figure 37. Radiometric activation assay

Top Left) Basic activation reaction showing the alpha-[32P] radio label. Bottom Left) Equation used to calculate %AA-AMP. Right) TLC image of activation showing separation of AA-AMP, AMP, and ATP.
CHAPTER VI - MATERIALS AND METHODS

6.1 Radiometric Assays

6.1.1. Strategy to Identification of Preferentially Activated AAs

Each aaRS was initially examined with its cognate AA as a positive control and with no AA as a negative control. In order to increase throughput, the $\alpha$-[P$^{32}$]ATP assay was initially used to measure the activation of L-AAs in subsets or groups. Each group was incubated with an individual aaRSs. Subsets showing AA-AMP production were investigated further. Each AA from subsets showing activation was assayed individually to determine which AAs are activated. Each AA subset was grouped according to structural and chemical similarity as shown in Table 9.

6.1.2. AA Classification:

Groups were classified based upon the chemical composition of their respective R-groups (Table 9). Group one was composed of aliphatic R-groups (Ala, Gly, Ile, Leu and Val); group two contained amine and carboxamide R-groups (Arg, Asn, Gln and Lys); group three contained carboxyl R-groups (Asp and Glu); group four contained hydroxyl R-groups (Ser and Thr); group five contained aromatic R-groups (Phe and Trp); group six contained sulfhydryl/methyl sulfide R-groups (Met & Cys); group seven contained cyclic R-groups (His and Pro). Upon incubation with the aaRS, groups that generated a measurable signal were investigated further. The identity of the activated AA within each group was determined by incubating individual AAs with the aaRS of interest.
Fig. 38 displays the activation scheme using alanyl-tRNA synthetase (AlaRS). Groups 1 and 4 produced an activation signal (i.e. an increase in the TLC AA-AMP spot) and proceeded to the next stage in the analysis. The remaining subgroups (2, 3, and 5 – 7 denoted with an x in Fig. 38) showed no activation (i.e. no AA-AMP) and were not investigated further. Next, analysis of the individual AAs within subgroups 1 and 4 (step 4) revealed that Gly and Ser are preferentially selected. The TLC images for the positive control, negative control, Gly, and Ser assays are displayed in Fig. 39A. Again the AA-AMP is represented as the fastest migrating species, followed by the AMP, and
unreacted ATP does not migrate appreciably. The data shown is based on a minimum of three independent assays.

Figure 38. Activation scheme for alanyl-synthetase (AlaRS)

Step 1 is a positive control to ensure enzyme activity. Step 2 is a set of group assays. Step 3 displays the individual AA assays. An X denotes no misactivation. Gly and Ser (shown in red) were selected by AlaRS.

Figure 39. TLC results and forward reaction curves for AlaRS

A) TLC images from AlaRS Pos control (Ala), Neg control (no AA), and individual AA assays for Gly and Ser. Time increases from left to right with lanes 1 – 11 representing 0s, 0.08m, 0.17min, 0.33min, 0.67min, 1.00min, 2.00min, 4.00min, 8.00min, 12.00min, and 20.00min, respectively. B) Forward reaction curve for AlaRS individual AA assays showing the Pos control in blue, Gly in grey, Ser in green, and the Neg control in red.
6.2 Measurement of Activation Reaction Rates

Fig. 40 – 51 and Fig. 52 – 61 show the forward reaction curves for the class I and class II aaRSs, respectively. Recombinant histidine tagged aaRSs were expressed in *E. coli* and purified using a nickel affinity column according to protocols previously reported by Josephson *et. al.*\(^{212}\) Positive (Pos) and negative (Neg) controls are shown for each aaRS along with any misactivated AAs. Tables to the right of each graph display AA-AMP concentrations at 20 min, the relative initial rates, and the R\(^2\) values for each fit. Each data set was fit using GraphPad\(^\circledR\) software, according to its behavior, by one of Eq. 4 – 6. Eq. 4 is a simple linear regression where the AA-AMP concentration (C) is equal to the rate constant (k) multiplied by time (t). Eq. 5 is an exponential decay function used to describe a one-phase association model, where C\(_{\text{Max}}\) is the maximum AA-AMP concentration, k is the rate constant, and t is time. Eq. 6 is a product inhibition model used to fit data displaying reversible Michealis –Menton kinetics, where C\(_{\text{Max}}\) is the maximum AA-AMP concentration, t is time, k\(_F\) is the forward rate constant and k\(_R\) is the reverse rate constant. Relative initial rates are calculated from the first derivative of each forward reaction curve. Each graph represents the average of at least three replicate assays.

\[
C = kt
\]
\[
C = C_{\text{Max}} \times (1 - e^{kt})
\]
\[
C = \frac{C_{\text{Max}} \times t}{k_F + t(1 + \frac{t}{k_R})}
\]

Eq. 4 Linear Model

Eq. 5 One-phase Association Model

Eq. 6 Product Inhibition Model
CHAPTER VII - RESULTS

7.1. Analysis of AA Activation Patterns Using Class I aaRSs:

7.1.1 Analysis of AA Activation Patterns Using ArgRS:

The forward reaction curves for ArgRS misactivation are shown in Fig. 40. ArgRS showed no activity in the presence of its cognate AA, Arg, but showed weak misactivation with several noncognate AAs. The presence of tRNA\textsuperscript{Arg} is known to be required for activation of Arg\textsuperscript{213}, but this is apparently not so for misactivation of Ala, Ile, Ser, and Thr. ArgRS produced 0.24 – 0.45 nM adenylate concentrations with Ala, Ile, Ser, and Thr. Ala produced a biphasic activation profile and had the highest initial rate at 0.026 nM/min.

![Figure 40. Forward reaction curves of ArgRS](image)

<table>
<thead>
<tr>
<th>[AA-AMP] @ 20min (nM)</th>
<th>Initial Rate (nM/min)</th>
<th>R\textsuperscript{2}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pos</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ala</td>
<td>0.45</td>
<td>0.026 0.65</td>
</tr>
<tr>
<td>Ile</td>
<td>0.24</td>
<td>0.005 0.50</td>
</tr>
<tr>
<td>Ser</td>
<td>0.31</td>
<td>0.010 0.65</td>
</tr>
<tr>
<td>Thr</td>
<td>0.37</td>
<td>0.007 0.51</td>
</tr>
<tr>
<td>Neg</td>
<td>0.06</td>
<td>-</td>
</tr>
</tbody>
</table>
7.1.2 Analysis of AA Activation Patterns Using CysRS:

CysRS showed no activity toward any of the noncognate substrates examined. Fig. 41 shows the forward reaction curve of the positive control for CysRS. CysRS produced relatively low concentrations of cognate adenylate (1.04 nM) and displayed a slow initial rate of 0.21 nM/min.

7.1.3 Analysis of AA Activation Patterns Using IleRS:

The forward reaction curves for IleRS are shown in Fig. 42 - 44. IleRS activated the second largest number of noncognate substrates of any aaRS examined. Val, Met, Thr, Glu, Leu, Cys, and Ala were all misactivated by IleRS. Val, Met, and Thr are shown in Fig. 42. IleRS showed remarkably strong activity toward Val, exceeding Ile in adenylate production at 20 min. Val, however, showed a 3-fold reduction in initial rate versus Ile. Met and Thr also produced relatively high levels of adenylate ( > 2 nM) at 20 min with rates approximately 14

![Graph showing forward reaction curves of CysRS and IleRS reaction rates.](image-url)
fold less than with Ile. Fig. 43 shows Glu, Leu, Cys, and Ala with the Pos control in order to illustrate the weak activation displayed with these substrates. Fig. 44 shows a zoomed in view of Glu, Leu, Cys, and Ala reaction curves, and clearly shows activation above the level of the Neg control.

Figure 42. Forward reaction curves of IleRS with Val, Met, and Thr

<table>
<thead>
<tr>
<th></th>
<th>[AA-AMP] @ 20min (nM)</th>
<th>Relative Initial Rate (nM/min)</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pos</td>
<td>5.29</td>
<td>1.60</td>
<td>0.93</td>
</tr>
<tr>
<td>Val</td>
<td>7.09</td>
<td>0.528</td>
<td>0.99</td>
</tr>
<tr>
<td>Met</td>
<td>2.19</td>
<td>0.111</td>
<td>0.92</td>
</tr>
<tr>
<td>Thr</td>
<td>2.13</td>
<td>0.118</td>
<td>0.83</td>
</tr>
<tr>
<td>Neg</td>
<td>0.18</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 43. Forward reaction curves of IleRS with Glu, Leu, Cys, and Ala.

Figure 44. Forward reaction curves of IleRS with Glu, Leu, Cys, Ala, and Neg control (zoomed view) without the Pos control.
7.1.4 Analysis of AA Activation Patterns Using LeuRS:

Forward reaction curves for LeuRS are shown in Fig. 45 and 46. LeuRS showed activity toward several noncognate AAs, however, for many of these this activity was quite weak compared to the cognate AA Leu (Fig. 45). The zoomed view shown in Fig. 46 shows LeuRS misactivated Cys, Ile, Phe, Val, Ala, and Asp. Cys showed the highest level of misactivation, producing 2.23 nM of Cys-AMP at 20 min at an initial rate of 0.229 nM/min. This represents a 6.5 fold reduction in adenylate concentration and a 37 fold reduction in initial rate versus Leu.

![Figure 45. Forward reaction curves of LeuRS](image)

<table>
<thead>
<tr>
<th></th>
<th>[AA-AMP] @ 20min (nM)</th>
<th>Relative Initial Rate (nM/min)</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pos</td>
<td>14.16</td>
<td>8.420</td>
<td>0.88</td>
</tr>
<tr>
<td>Cys</td>
<td>2.23</td>
<td>0.229</td>
<td>0.76</td>
</tr>
<tr>
<td>Ile</td>
<td>1.66</td>
<td>0.162</td>
<td>0.62</td>
</tr>
<tr>
<td>Phe</td>
<td>1.03</td>
<td>0.053</td>
<td>0.70</td>
</tr>
<tr>
<td>Val</td>
<td>0.89</td>
<td>0.148</td>
<td>0.64</td>
</tr>
<tr>
<td>Ala</td>
<td>0.58</td>
<td>0.082</td>
<td>0.42</td>
</tr>
<tr>
<td>Asp</td>
<td>0.57</td>
<td>0.047</td>
<td>0.63</td>
</tr>
<tr>
<td>Neg</td>
<td>0.38</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 45. Forward reaction curves of LeuRS
Figure 46. Forward reaction curves of LeuRS (zoomed view)
7.1.5 Analysis of AA Activation Patterns Using MetRS:

MetRS was the most active aaRS to show no activity toward any of the noncognate substrates examined. Fig. 47 shows the forward reaction curve of the positive control for MetRS. MetRS produced relatively high concentrations of cognate adenylate (14.16 nM) and displayed one of the fastest initial rates at 8.42 nM/min.

![Graph of MetRS reaction](image)

Figure 47. Forward reaction curves of MetRS
7.1.6 Analysis of AA Activation Patterns Using TrpRS:

TrpRS showed no activity toward any of the noncognate substrates examined. Fig. 48 shows the forward reaction curve of the positive control for TrpRS.

![Figure 48. Forward reaction curves of TrpRS](image)

<table>
<thead>
<tr>
<th>[AA-AMP] @ 20min (nM)</th>
<th>Relative Initial Rate (nM/min)</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pos</td>
<td>3.74</td>
<td>5.94</td>
</tr>
<tr>
<td>Neg</td>
<td>0.09</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 48. Forward reaction curves of TrpRS
7.1.7 Analysis of AA Activation Patterns Using TyrRS:

The forward reaction curves for TyrRS are shown in Fig 49. TyrRS showed noncognate activity toward Cys and Phe. Both Cys and Phe produced adenylate levels at 20 min comparable to those seen with Tyr, but with drastically different initial rates. Cys displayed an initial rate of 5.72 nM/min, or about half that of Tyr. Phe, on the other hand, displayed an initial rate 17 fold lower versus Tyr.

![Forward reaction curves of TyrRS](image)

**Figure 49.** Forward reaction curves of TyrRS
7.1.8 Analysis of AA Activation Patterns Using ValRS:

ValRS (Fig. 50 and 51) was the most active aaRS of all those examined, both in terms of cognate and noncognate activation. ValRS was also the most promiscuous aaRS, showing activation toward a total of 10 AAs (Ile, Leu, Asp, Glu, Ala, Ser, Glu, Thr, Met, and Gly). ValRS produced the highest concentration of cognate adenylate of any aaRS at 25.79 nM and displayed the highest initial rate of cognate activation at 88.64 nM/min. Fig. 50 shows ValRS misactivated five non-cognate AAs (Ile, Leu, Asp, Glu, and Ala) to a higher degree, in terms of adenylate production, than their respective aaRSs. One non-cognate substrate, Ile, produced adenylate concentrations 20% higher than with Val. The initial activation rate with Ile, however, was 6 fold less compared to Val. Ser, Glu, Thr, and Met (Fig. 51) showed weaker, but still substantial activation.

Figure 50. Forward reaction curves of ValRS with Ile, Leu, Asp, Cys, and Ala
Figure 51. Forward reaction curves of ValRS with Ser, Glu, Thr, Met, and Gly

<table>
<thead>
<tr>
<th></th>
<th>[AA-AMP] @ 20min (nM)</th>
<th>Relative Initial Rate (nM/min)</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pos</td>
<td>25.79</td>
<td>88.639</td>
<td>0.83</td>
</tr>
<tr>
<td>Ser</td>
<td>11.02</td>
<td>0.625</td>
<td>0.89</td>
</tr>
<tr>
<td>Glu</td>
<td>9.74</td>
<td>0.560</td>
<td>0.85</td>
</tr>
<tr>
<td>Thr</td>
<td>9.11</td>
<td>17.991</td>
<td>0.88</td>
</tr>
<tr>
<td>Met</td>
<td>6.14</td>
<td>0.326</td>
<td>0.77</td>
</tr>
<tr>
<td>Gly</td>
<td>4.92</td>
<td>0.758</td>
<td>0.84</td>
</tr>
<tr>
<td>Neg</td>
<td>3.39</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
7.2. Analysis of AA Activation Using Class II aaRSs:

7.2.1 Analysis of AA Activation Using AlaRS:

Fig. 52 shows the forward reaction curves and accompanying data table for AlaRS. AlaRS showed strong activity toward Gly and Ser. Both noncognate substrates produced adenylate concentrations comparable to the cognate Ala at 20 min. Initial rates, however, were 6 – 9 fold lower with Ser and Gly, respectively, versus Ala.

<table>
<thead>
<tr>
<th></th>
<th>[AA-AMP] @ 20min (nM)</th>
<th>Relative Initial Rate (nM/min)</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pos</td>
<td>10.24</td>
<td>6.64</td>
<td>0.96</td>
</tr>
<tr>
<td>Gly</td>
<td>8.99</td>
<td>0.76</td>
<td>0.99</td>
</tr>
<tr>
<td>Ser</td>
<td>8.97</td>
<td>1.03</td>
<td>0.98</td>
</tr>
<tr>
<td>Neg</td>
<td>0.27</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 52. Forward reaction curves of AlaRS
7.2.2 Analysis of AA Activation Using AsnRS:

Fig. 53 shows the forward reaction curves for AsnRS misactivation. AsnRS was shown to misactivate the two acidic AAs, Asp, and Glu. Asp produced just under half that of the Pos control with an initial rate 18-fold less than the Pos control. AsnRS activated Glu much less efficiently, producing an AA-AMP concentration 10-fold less than the Pos control at an 80-fold lower initial rate.

```
<table>
<thead>
<tr>
<th></th>
<th>[AA-AMP] @ 20min (nM)</th>
<th>Relative Initial Rate (nM/min)</th>
<th>R^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pos</td>
<td>7.36</td>
<td>8.86</td>
<td>0.86</td>
</tr>
<tr>
<td>Asp</td>
<td>2.81</td>
<td>0.48</td>
<td>0.96</td>
</tr>
<tr>
<td>Glu</td>
<td>0.77</td>
<td>0.15</td>
<td>0.85</td>
</tr>
<tr>
<td>Neg</td>
<td>0.11</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
```

Figure 53. Forward reaction curves of AsnRS
7.2.3 Analysis of AA Activation Using AspRS:

AspRS forward reaction curves are shown in Fig. 54. AspRS displayed no
cognate activation with Asp (Pos control). AspRS did, however, misactivate Ala,
Gly, and Ser. Misactivated adenylate concentrations at 20 min were relatively low
(1.0 – 0.3nM). Ala showed the highest AA-AMP concentration at 1.09nM with an
initial rate of 2.41 nM/min.

![AspRS](image)

<table>
<thead>
<tr>
<th></th>
<th>[AA-AMP] @ 20min (nM)</th>
<th>Relative Initial Rate (nM/min)</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pos</td>
<td>0.09</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ala</td>
<td>1.09</td>
<td>2.41</td>
<td>0.90</td>
</tr>
<tr>
<td>Gly</td>
<td>0.38</td>
<td>0.11</td>
<td>0.86</td>
</tr>
<tr>
<td>Ser</td>
<td>0.31</td>
<td>0.32</td>
<td>0.86</td>
</tr>
<tr>
<td>Neg</td>
<td>0.09</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 54. Forward reaction curves of AspRS
7.2.4 Analysis of AA Activation Using GlyRS:

GlyRS showed a surprising level of substrate flexibility, though it was one of the least active aaRSs, producing only 0.31 nM Gly-AMP at 20 min. Fig. 55 shows Ala and Met misactivation produced nearly 50% higher adenylate concentrations at 20 min versus Gly, and Ser was misactivated to nearly the same degree as Gly. Initial rates for GlyRS were relatively low with Ala producing the highest initial rate at 0.55 nM/min.

![GlyRS Forward Reaction Curves](image)

**Figure 55.** Forward reaction curves of GlyRS

<table>
<thead>
<tr>
<th>[AA-AMP] @ 20min (nM)</th>
<th>Relative Initial Rate (nM/min)</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pos</td>
<td>0.31</td>
<td>0.11</td>
</tr>
<tr>
<td>Ala</td>
<td>0.48</td>
<td>0.55</td>
</tr>
<tr>
<td>Ser</td>
<td>0.29</td>
<td>0.05</td>
</tr>
<tr>
<td>Met</td>
<td>0.45</td>
<td>0.07</td>
</tr>
<tr>
<td>Neg</td>
<td>0.05</td>
<td>-</td>
</tr>
</tbody>
</table>
7.2.5 Analysis of AA Activation Using HisRS:

HisRS was the only class II aaRS to show no activity toward any of the noncognate substrates examined. Fig. 56 shows the forward reaction curve of the positive control for HisRS. HisRS produced very low concentrations of cognate adenylate (0.48 nM) but at a significant initial rate of 6.64 nM/min.

Figure 56. Forward reaction curves of HisRS
7.2.6 Analysis of AA Activation Using LysRS:

LysRS forward reaction curves are shown in Fig. 57. LysRS was one of the most active of the class II aaRSs with the highest initial rate of cognate activation of any class II aaRS. LysRS produced a biphasic activation profile with Lys with an initial rate exceeding 20 nM/min. LysRS showed noncognate activity toward Ile, Cys, and Met. All produced relatively high adenylate concentrations at 20 min, however, at rates 20 – 55 fold lower rates versus Lys.

**LysRS**

<table>
<thead>
<tr>
<th></th>
<th>Relative Initial Rate (nM/min)</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pos</td>
<td>21.68</td>
<td>0.80</td>
</tr>
<tr>
<td>Ile</td>
<td>1.70</td>
<td>0.93</td>
</tr>
<tr>
<td>Cys</td>
<td>1.74</td>
<td>0.98</td>
</tr>
<tr>
<td>Met</td>
<td>0.38</td>
<td>0.98</td>
</tr>
<tr>
<td>Neg</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 57. Forward reaction curves of LysRS
7.2.7 Analysis of AA Activation Using PheRS:

PheRS showed moderate activity toward Tyr. The forward reaction curves for PheRS are shown in Fig. 58. Tyr produced only 3 fold less adenylate concentration at 20 min, but displayed an initial rate 11 fold lower versus Phe.

![Figure 58. Forward reaction curves of PheRS](image)

<table>
<thead>
<tr>
<th>[AA-AMP] @ 20min (nM)</th>
<th>Relative Initial Rate (nM/min)</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pos</td>
<td>3.18</td>
<td>2.07</td>
</tr>
<tr>
<td>Tyr</td>
<td>1.18</td>
<td>0.18</td>
</tr>
<tr>
<td>Neg</td>
<td>0.18</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 58. Forward reaction curves of PheRS
7.2.8 Analysis of AA Activation Using ProRS:

The forward reaction curves for ProRS are shown in Fig 59. ProRS was shown to misactivate Ala. Ala misactivation produced an almost 10 fold reduction in both adenylate concentration at 20 min and in initial rate versus the cognate substrate Pro.

![ProRS Reaction Curves](image)

**ProRS**

<table>
<thead>
<tr>
<th></th>
<th>[AA-AMP] @ 20min (nM)</th>
<th>Relative Initial Rate (nM/min)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pos</strong></td>
<td>2.35</td>
<td>3.74</td>
<td>0.88</td>
</tr>
<tr>
<td><strong>Ala</strong></td>
<td>0.35</td>
<td>0.39</td>
<td>0.62</td>
</tr>
<tr>
<td><strong>Neg</strong></td>
<td>0.11</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 59. Forward reaction curves of ProRS
7.2.9 Analysis of AA Activation Using SerRS:

SerRS was shown to misactivate Cys. The forward reaction curves showing activation profiles for the Ser (pos) and Cys are shown in Fig. 60. Initial activation rates show Cys was activated 50 fold less efficiently than the cognate Ser.

![Figure 60. Forward reaction curves of SerRS](image-url)

<table>
<thead>
<tr>
<th></th>
<th>[AA-AMP] @ 20min (nM)</th>
<th>Relative Initial Rate (nM/min)</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pos</td>
<td>6.50</td>
<td>5.39</td>
<td>0.86</td>
</tr>
<tr>
<td>Cys</td>
<td>1.67</td>
<td>0.11</td>
<td>0.86</td>
</tr>
<tr>
<td>Neg</td>
<td>0.16</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
7.2.10 Analysis of AA Activation Using ThrRS:

ThrRS (Fig. 61 and 62) showed activity toward the highest number of noncognate substrates of any class II aaRS. ThrRS misactivated Ala, Gly, Ile, Ser, and Cys. Ala was misactivated 3 – 4 fold less efficiently than Thr both in terms of adenylate produced at 20 min and initial rate. Gly, Ile, Ser, and Cys were activated much less efficiently, producing adenylate concentrations 10 fold less than with Thr and at initial rates 20 – 100 fold slower.

Figure 61. Forward reaction curves of ThrRS
Figure 62. Forward reaction curves of ThrRS (zoomed view)

7.3 Analysis of AA Misactivation

Table 10 shows a summary of class I aaRS activation with aaRSs listed horizontally and AAs listed vertically. Both are arranged in order of highest number of misactivations. An X denotes activation of an AA by an aaRS. Five class I aaRSs (ArgRS, IleRS, LeuRS, TyrRS, and ValRS) misactivated at least one non-cognate aaRS. An additional three (CysRS, MetRS, and TrpRS) showed only cognate activation. Two class I aaRSs (GlnRS and GluRS) produced no activation (calibration curves for these two aaRSs are not shown). While only five of ten class I aaRSs showed noncognate activity, several of these misactivated a large number of AAs with varying side chain functionalities. A subset of closely related class Ia aaRSs comprised of ArgRS, IleRS, LeuRS, and ValRS, showed
a great deal of substrate flexibility, misactivating aliphatic, hydroxyl containing, acidic, sulfur containing, and aromatic AAs.

Table 10

**Summary of class I aaRS activation**

<table>
<thead>
<tr>
<th>Class I aaRSs</th>
<th>ValRS</th>
<th>IleRS</th>
<th>LeuRS</th>
<th>ArgRS</th>
<th>TyrRS</th>
<th>TrpRS</th>
<th>CysRS</th>
<th>MetRS</th>
<th>GluRS</th>
<th>GlnRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cys</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Trp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>His</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 11 shows a summary of class II aaRS activation with aaRSs listed horizontally and AAs listed vertically. Again, both are arranged in order of highest number of misactivations with an X denoting activation. Nine class II aaRSs (AlaRS, AsnRS, AspRS, GlyRS, LysRS, PheRS, ProRS, SerRS, and ThrRS) misactivated at least one non-cognate AA. HisRS showed only cognate activation. Class II aaRSs showed non-cognate activation to nine AAs (Ala, Ser, Gly, Cys, Ile, Met, Asp, Glu, and Tyr), and were also shown to preferentially select aliphatic, hydroxyl containing, acidic, sulfur containing, and aromatic AAs.
Table 11

Summary of class II aaRS activation

<table>
<thead>
<tr>
<th>Class II aaRSs</th>
<th>ThrRS</th>
<th>GlyRS</th>
<th>LysRS</th>
<th>AlaRS</th>
<th>AspRS</th>
<th>AsnRS</th>
<th>PheRS</th>
<th>ProRS</th>
<th>SerRS</th>
<th>HisRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Ile</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Met</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>His</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Phe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Lys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Pro</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Arg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trp</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tables 12 and 13 show AAs selected during class I and II aaRS assays listed as “Early” with those rejected listed as “Late”. These tables are based on the hypothesis that early AAs are more highly selected by extant aaRSs for misactivation. For both data sets, the list of early AAs contains a near equal number of AAs specific to both classes of aaRSs. Aliphatic AAs (Ala, Ile, Leu, and Val), hydroxyl containing (Ser and Thr), sulfur containing (Cys and Met), and acidic AAs (Asp and Glu) were selected to a high degree and placed in the Early column. The aromatic AA Phe was misactivated by a single class I aaRS, TyrRS,
and tentatively placed in the Early column. All amide and carboxamide containing AAs (Arg, Asn, Gln, and Lys) are placed in the Late column.

Table 12

*Early vs. Late AAs according to class I aaRS misactivation*

<table>
<thead>
<tr>
<th>Class I Chronology</th>
<th>Early</th>
<th>Late</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cys</td>
<td></td>
<td>Tyr</td>
</tr>
<tr>
<td>Ala</td>
<td></td>
<td>Trp</td>
</tr>
<tr>
<td>Ile</td>
<td></td>
<td>His</td>
</tr>
<tr>
<td>Thr</td>
<td></td>
<td>Lys</td>
</tr>
<tr>
<td>Val</td>
<td></td>
<td>Pro</td>
</tr>
<tr>
<td>Leu</td>
<td></td>
<td>Arg</td>
</tr>
<tr>
<td>Met</td>
<td></td>
<td>Asn</td>
</tr>
<tr>
<td>Asp</td>
<td></td>
<td>Gln</td>
</tr>
<tr>
<td>Glu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phe</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 13

*Early vs. Late AAs according to class II aaRS misactivation*

<table>
<thead>
<tr>
<th>Class II Chronology</th>
<th>Early</th>
<th>Late</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td></td>
<td>Thr</td>
</tr>
<tr>
<td>Ser</td>
<td></td>
<td>Leu</td>
</tr>
<tr>
<td>Gly</td>
<td></td>
<td>Val</td>
</tr>
<tr>
<td>Cys</td>
<td></td>
<td>Phe</td>
</tr>
<tr>
<td>Ile</td>
<td></td>
<td>His</td>
</tr>
<tr>
<td>Met</td>
<td></td>
<td>Lys</td>
</tr>
<tr>
<td>Asp</td>
<td></td>
<td>Pro</td>
</tr>
<tr>
<td>Glu</td>
<td></td>
<td>Trp</td>
</tr>
<tr>
<td>Tyr</td>
<td></td>
<td>Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Asn</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gln</td>
</tr>
</tbody>
</table>
CHAPTER VIII - DISCUSSION AND CONCLUSION

8.1 Discussion

Tables 14 A-C show Early vs. Late AAs based on three of the theories discussed earlier. AAs sharing the same box are equal in age according to each theory. Table 14A is derived from a consensus of proteinacious AAs found in early Earth simulations.\textsuperscript{137,138,150} Table 14B is derived from the codon-ceding scheme proposed in the CET.\textsuperscript{147} Table 14B shows the results from a consensus study examining a set of criteria thought to be associated with AA age.\textsuperscript{144} This study determined an AA chronology based primarily on thermodynamic stability and cost of processivity. In this study, AAs that are most thermally stable and least costly to process were considered to be early AAs. Table 15 shows a consensus table derived from Tables 12 and 13 from the current study. AAs activated by at least two noncognate aaRSs are considered early. AAs showing only a single instance of misactivation along with those showing no misactivation are considered late. There is a consensus between all chronologies for a majority of AAs. Ala, Ser, Gly, Asp, Glu, Thr, Leu, and Val are considered early by all methods. Likewise, Arg, Asn, His, Lys, Tyr, and Trp are considered late by each method. Data for Cys, Gln, Ile, Met, Phe, and Pro do not agree between methods.
Table 14 AA

*Chronologies based on early Earth/meteorite studies, Physiochemical Properties, and Biosynthetic Pathways*

Table 15

*Consensus AA Chronology Based on Class I and II aaRS Activation*

To further illustrate the compliance of misactivation profiles with the CET and AA biosynthetic pathways, Fig. 63 and 64 were generated showing the misactivation overlaid onto a schematic of extant biosynthetic pathways (Fig. 63)
and the codon-conceding scheme proposed by Wong in the CET (Fig. 64). When the AA misactivation data from the current study is overlaid onto the AA biosynthetic pathways all of the most highly misactivated AAs (early AAs) comprise all of the inner nodes. These are shown in dashed red boxes in Fig. 63. AAs showing little or no misactivation are all termination points in the pathways, shown in solid blue boxes in Fig. 63. In the biosynthetic pathways shown here all intermediate precursor AAs are from the group of highly misactivated AAs (Early AAs). AAs showing little or no misactivation (Late AAs) are not precursors of other AAs. This illustrates the selection preference of extant aaRSs for AAs that are precursors of other AAs and closer to central metabolism. Assuming that precursor AAs and AAs closer to central metabolism are more ancient suggests that extant aaRSs may have retained some activity toward their ancestral substrates (i.e. early AAs) and may suggest a role for ancestral aaRSs in primitive biosynthesis of AAs or their intermediates.
Figure 63. AA misactivation vs. AA biosynthetic pathways

AA biosynthetic pathways with AA misactivation results overlaid onto it. Highly misactivated or Early AAs are shown in dashed red boxes. AAs showing little or no misactivation are shown in solid blue boxes. From "Amino acid biogenesis, evolution of the genetic code and aminoacyl-tRNA synthetases," by Klipcan et. al. 2004. J. Theor. Biol. 228, 389-396. Copyright 2004 by Elsevier. Reprinted with permission of the author.

Fig. 64 was generated, showing the codon-conceding scheme proposed by Wong in the CET with the AA misactivation results overlaid onto it. The figure shows the proposed evolution of AA codification as codons are conceded from older AAs to newer, AAs. In this case the group of highly misactivated AAs (Early AAs) again comprise all of the inner nodes. AAs considered Late according to misactivation data are terminal points according to this scheme. The correlation between the AA selectivity of aaRSs and the codon concession scheme shown in Fig. 64 suggests that aaRS may have played a role in AA codification. If extant
aaRSs have indeed retained activity toward a set of early AAs, this suggests that development of aaRS specificity may be linked to the codon concession process.

Figure 64. AA misactivation vs. CET

Codon ceding scheme proposed in the CET with AA misactivation results overlaid onto it. Highly misactivated or Early AAs are shown in dashed red boxes. AAs showing little or no misactivation are shown in solid blue boxes. From “A co-evolution theory of the genetic code,” by Wong 1975, Proc. Nat. Acad. Sci. 72, 1909-1912. Copyright 1975 by Proc. Nat. Acad. Sci. Reprinted with permission of the author. There seems to be little relationship between aaRS class and selection of AAs for misactivation. Overall, AAs are not selected to a higher degree by aaRSs of their respective classes. Six class I specific AAs (Cys, Ile, Val, Leu, Glu, and Met) were misactivated by class I aaRSs. Additionally, five class II specific AAs (Ala, Asp, Ser, Thr, and Phe) were activated by class I aaRSs. Only four of ten class II specific AAs (Ala, Ser, Gly, and Asp) showed misactivation by the class II aaRSs. Class II aaRSs showed noncognate activity toward five class I specific
AAs (Cys, Ile, Met, Glu, and Tyr). The class II aaRSs proved to be more promiscuous as a group than the class I aaRSs. Nine of the ten class II aaRSs showed activity toward at least one noncognate substrate compared to only five of ten class I aaRSs. However, a subset of class I aaRSs (ArgRS, IleRS, LeuRS, and ValRS) showed a high degree of substrate flexibility, and activated a higher number of AAs. Both classes showed a propensity for excluding many of the same AAs. Arg, Asn, Lys, and Gln, AAs containing amide or carboxamide side chains, were completely rejected for misactivation.

8.2 Conclusion

The data presented in this study suggests that extant aaRSs have a selection preference for AAs thought to be among the most ancient. AA misactivation profiles shown in Fig. 40-62 and summarized in Tables 11-12 correlate with the many of the current theories on the evolution of AA recruitment and codification, including the CET (Fig. 64). This points to the prospect that aaRSs may have played a role in the process of AA codification during some phase of genetic code development. This proposition depends heavily on the state of the genetic code during the transition from ribozyme-based translation to protein-based translation when the ancestors of aaRSs first took up the role of tRNA charging.

It has been suggested that ancestral aaRSs were initially tasked with enhancing charging rates in primitive translation with little specificity toward a particular amino acid. The diversification of aaRS into AA specific enzymes could have followed two possible scenarios: i) aaRSs initially developed AA
specificity in the presence the genetic code in its current form or ii) aaRSs initially
developed in the presence of an incomplete set of AAs and AA specificity
coevolved with the genetic code. The first scenario implies that a fully-fledged
genetic code existed in an RNA based translation system prior to the advent of
peptides that assisted in RNA charging (proto-aaRSs). The second scenario
implies the existence of a primitive genetic code similar to the operational codes
proposed by Schimmel and Rodin\textsuperscript{167,168,173,174}, capable of producing functional
peptides. Furthermore, it implies that aaRSs were present during a period of
code expansion and ambiguity reduction.

The results of this study suggest the second scenario to be more likely
and support the supposition that ancestral aaRSs were initially tasked with
enhancing charging rates in primitive translation with little specificity toward a
particular amino acid. Many extant aaRSs display noncognate activity toward
AAs thought to be among the earliest. This activity may be a vestige of ancestral
aaRS specificity toward this group of AAs. If this is indeed the case, it indicates
that ancestral aaRSs arose during a period of an early operational code with
fewer than 20 AAs (i.e. the set identified as early AAs) and were present during
later stages of genetic code development.

Further, the selection (or misactivation) profiles of extant aaRSs suggest
that interactions between ancestral aaRSs and AAs may have played a role in
AA recruitment and in determining the structure of the genetic code. This
scenario is evidenced by the synthesis of Gln-tRNA\textsuperscript{Gln} observed in archea and
certain bacteria that lack GlnRS\textsuperscript{132}. In these organisms GluRS forms an
intermediate of Gln-tRNA$^{\text{Gln}}$ by mischarging tRNA$^{\text{Gln}}$ with Glu to form Glu-tRNA$^{\text{Gln}}$. The Glu moiety is then converted to Gln by an amidotransferase. In this manner GluRS participates in the synthesis of an intermediate of Gln. The ability of GluRS to charge both tRNA$^{\text{Glu}}$ and tRNA$^{\text{Gln}}$ with Glu also provides a mechanism for codon concession that is linked to aaRS specificity.

In a manner similar to GluRS, ancestral aaRSs may have participated in the synthesis of new AAs and conceded codons to them, thereby directing the recruitment and coding of AAs. The AA selection profiles from this study show that extant aaRSs preferentially misactivate AAs thought to be more ancient, suggesting that aaRSs may still hold a vestige of their ancestral activity toward an early set of AAs. This apparent preference for an ancient set of AAs suggests that aaRSs originally developed specificity for this ancient set of AAs and points to the likelihood that aaRSs played a part in AA synthesis and codification during a period of expansion and ambiguity reduction of the genetic code.
REFERENCES


