The Pennsylvania State University

The Graduate School

STRUCTURAL AND MECHANISTIC ENZYMOLOGY OF NUCLEOTIDE INCORPORATION BY A-FAMILY DNA POLYMERASES

A Dissertation in

Biochemistry, Microbiology, and Molecular Biology

by

Mohammad A. AlMishwat

© 2019 Mohammad A. AlMishwat

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2019

The dissertation of Mohammad A. AlMishwat was reviewed and approved* by the following:

J. Gregory Ferry
Stanley Person Professor of Biochemistry and Molecular Biology
Dissertation Adviser
Committee Chair

Amie Boal
Associate Professor of Chemistry
Associate Professor of Biochemistry and Molecular Biology

Squire Booker
Professor of Chemistry and
Professor of Biochemistry and Molecular Biology

David Gilmour
Professor of Molecular and Cell Biology

Kenneth Keiler
Professor of Biochemistry and Molecular Biology
Associate Department Head for Graduate Education

*Signatures are on file in the Graduate School.

ABSTRACT

DNA-dependent DNA and RNA polymerases are essential enzymes to gene expression and regulation. DNA polymerases, the enzymes responsible for carrying out and regulating the faithful transmission of an organisms genetic material through the process of DNA replication have been extensively characterized, however, the exact mechanism by which a high fidelity DNA polymerase preferentially incorporates a correct nucleotide, and differentially excludes an incorrect nucleotide during DNA replication is not fully understood. Structural studies of various DNA polymerases and their complexes with DNA have provided a great deal of insight into how catalysis in nucleotide incorporation occurs, and have also provided empirical models of how fidelity is brought about and sustained in these enzymes during DNA replication.

This dissertation describes ongoing work to elucidate the mechanism of high-fidelity nucleotide incorporation during DNA replication by an A-family DNA polymerase, employing biochemical and structural studies to dissect the structural and mechanistic changes that occur during nucleotide binding and phosphodiester bond formation between the incoming nucleotide and the growing primer strand. Time-resolved X-ray crystallography is used to monitor and study in real-time, at atomic resolution, the mechanism of nucleotide incorporation *in crystallo*. This method has been successful in studying the mechanistic details of several enzymes, and we show here that it can be used to directly observe and monitor the sequential structural and mechanistic changes in an A-family DNA polymerase and its bound substrate DNA that are brought about by nucleotide binding to the DNA polymerase active site and subsequent catalysis.

TABLE OF CONTENTS

List of Tables vii

List of Figures viii

List of Abbreviations xi

Acknowledgements xiv

Structural and Mechanistic Enzymology of DNA Polymerases

Chapter 1. Introduction to DNA polymerases and DNA Replication 2

1.1 DNA Polymerases and DNA Replication 3

1.2 DNA Polymerase Structure, Function, and Organization 5

1.3 Proposed Mechanism of Nucleotide Incorporation, Fidelity, and the Fingers Domain O-helix 13

1.4 Structural Biology of A-Family DNA Polymerases 19

1.5 Time-resolved X-ray crystallography for studies in structural and mechanistic enzymology 23

1.6 Specific Aims of this Work 28

References 30

Chapter 2. Structural and Mechanistic Basis of Nucleotide Incorporation during DNA Replication by a high-fidelity, A-family Replicative DNA Polymerase 40

2.1 Abstract 41

2.2 Introduction 42

2.3 Results 45

2.3.1 *in crystallo* Single-turnover Kinetics of A-family KlenTaq DNAP 45

2.3.2 Nucleotide Binding is Spatiotemporally Separated from Metal-ion Entry 48

2.3.3 Binding and Catalytic Metal Entry is Sequential and Concerted 52

2.3.4 Pyrophosphate Release Requires O-helix Reversion to the Open Conformation 52

2.4 Discussion 55

2.4.1 Spatiotemporal Separation of Substrate and Metals – a Possible Mechanism for O-helix Closure, and Implications for Fidelity 55

2.4.2 3'-Oxyanion Activation – A Catalytic Water, Captured Transition State, and Consequences to Proton Transfer 66

2.4.3 Coupled Dynamics in Translocation..... 76

2.4.4 Two- and Three-Metal Ion Catalysis: Evolutionary Imperative or Mechanistic Contingency 81

2.4.4.1 Function-centric view to variations in metal-ion catalysis..... 83

2.4.4.2 Metal-centric view to variations in metal-ion catalysis..... 84

2.5 Materials and Methods..... 88

References 95

Chapter 3. Insights into Fidelity in A-family DNA Polymerases 104

3.1 Abstract 105

3.2 Introduction..... 106

3.3 Results..... 111

3.3.1 O-helix Conformational Change is Independent of and Unaffected by Metal-ions in Matched Ternary Complexes 111

3.3.2 Binding of Non-Mg^{2+} Divalent Metal-ions has Little Effect on Closed, Matched Ternary Complexes..... 113

3.3.3 O-helix Rapidly Reverts to the Open Conformation after Sampling the Active Site in the Presence of Mg^{2+} in Mismatched Ternary Complexes..... 113

3.3.4 O-helix Remains Closed in Chelated Mismatched Ternary Complexes 116

3.3.5 Mn^{2+}-substituted Mismatched Complexes Remain in the Closed Conformation, Poised for Catalysis 116

3.4 Discussion 119

3.4.1 Insights into the Structural and Mechanistic Determinants of Fidelity 119

3.4.2 Structural Aspects Delineated from Examining Fidelity Intrinsic to A-family DNAPs in Matched Complexes 120

3.4.3 Structural Aspects Delineated from Examining Fidelity Intrinsic to A-family DNAPs in Mismatched Complexes 120

3.4.4 Mechanisms of O-helix Reversion in Nucleotide Discrimination..... 124

3.4.4.1 Charge-driven Mechanism (Charge-repulsion)..... 124

3.4.4.2 Metal-ion-driven Mechanism (Rigidity, Steric hindrance) 124

3.4.5 Incomplete Structural Details 125

3.5 Materials and Methods 126
References 130

Chapter 4. Summary of and Future Direction in Studies on DNA Replication 132
4.1 Studies on DNA Replication and DNA Polymerases 133
References 140

Appendix A. Purification and Crystallization Screening of Extension-ready Binary and Ternary Complexes of KlenTaq DNA Polymerase 143

Appendix B. Neutron Scattering of the Ternary Complex of KlenTaq DNA Polymerase with the non-hydrolyzable nucleotide analog dGMPNPP 152

Appendix C. Crystallization of *E. coli* RNA Polymerase Holoenzyme in Complex with Promoter DNA containing an UP element 156

Appendix D. Cloning, Expression and Purification, Oligomeric States and Crystallization of MreA and Homologs 163

Appendix E. Genome-Wide Binding Patterns of the Global Transcriptional Regulator MreA in the Methanogen *Methanosarcina acetivorans* 169

LIST OF TABLES

Table 1.1 Representative DNA polymerases & their functions from the seven families.....6

Table 2.1 Crystallographic data and refinement statistics.....91

Table 2.2 Relative Metal-ion Catalytic Residue Distances (Å) 92

Table 2.3 Changes in Side-chain Position and Metal-ion Residue Distances (Å).....93

LIST OF FIGURES

Figure 1.1 Number of DNA polymerases with increasing genome complexity.....4

Figure 1.2 Domain organization of families A and B DNA polymerases.....8

Figure 1.3 Sequence of events during nucleotide incorporation.....14

Figure 1.4 Two-metal ion mechanism of DNA polymerase.....18

Figure 1.5 Crystal structure of the closed ternary complex of KlenTaq DNA polymerase.....22

Figure 2.1 Superposition of the dideoxy-terminated and extension-ready binary and ternary complexes of KlenTaq DNA polymerase.....47

Figure 2.2 Superposition of the closed dideoxy-terminated and dGMPNPP-trapped ternary complexes.....49

Figure 2.3 Time-course of events along the pathway for nucleotide incorporation by KlenTaq DNA polymerase.....50

Figure 2.4 Superposition of the open dideoxy-terminated and catalytically-competent binary complexes.....51

Figure 2.5 Superposition of the dGMPNPP-trapped and product complexes.....54

Figure 2.6 The minimal kinetic pathway and its' expansion of correct nucleotide incorporation by A-family DNA polymerases.....58

Figure 2.7 Sequence alignment of evolutionarily conserved, catalytically-relevant residues of motif B among A-family DNA polymerases.....59

Figure 2.8 Superposition of the open binary and dGMPNPP-trapped closed ternary complexes of Klentaq DNA polymerase.....61

Figure 2.9 Superposition of the open binary and closed metal-free ternary complexes of Klentaq DNA polymerase.....62

Figure 2.10 Superposition of the closed metal-free and Mg^B-coordinated ternary complexes of Klentaq DNA polymerase.....63

Figure 2.11 Superposition of the dGMPNPP-trapped and post-incorporation closed ternary complexes of Klentaq DNA polymerase.....65

Figure 2.12 Coordination of solvent molecules to Mg^A in the dideoxy-terminated ternary complex.....68

Figure 2.13 A catalytic water during product formation.....70

Figure 2.14 Catalytically-relevant distances of solvent molecules during phospho-diester bond formation.....72

Figure 2.15 Interatomic distances of Mg^B and Mg^A during and after product complex formation.....74

Figure 2.16 Structural basis for leaving group instability and effect on product complex formation.....76

Figure 2.17 Crystal packing in the single-turnover, non-translocating complex.....78

Figure 2.18 Emergence and catalytic distances of solvent molecules in dideoxy-terminated and catalytically-competent ternary complexes of KlenTaq DNA polymerase.....80

Figure 2.19 F_o-F_c omit maps of the dGMPNPP-trapped and product complexes.....82

Figure 2.20 The crystal structures of Mg^{2+}- and Mn^{2+}-substituted dGMPNPP-trapped ternary complexes.....86

Figure 3.1 Sequence conservation of motifs A and C in A-family DNA polymerases.....107

Figure 3.2 Anatomy and movement of the O-helix of the Klenow fragment of *Thermus aquaticus* DNA polymerase.....109

Figure 3.3 Conformational changes in chelated dC:dGTP ternary complexes.....112

Figure 3.4 Conformational changes in Mg^{2+}-substituted mismatched complexes....115

Figure 3.5 Conformational changes in Mg^{2+}- and Mn^{2+}-substituted mismatched complexes.....118

Figure 3.6 Minimal kinetic mechanism for incorrect nucleotide incorporation by A-family DNA polymerases.....123

Figure 4.1 Snapshots along the catalytic pathway for correct nucleotide incorporation by A-family DNAPs.....135

Figure 4.2 Expansion of the minimal kinetic pathway for correct and incorrect nucleotide incorporation by A-family DNA polymerases based on metal-ion coordination.....137

LIST OF ABBREVIATIONS

5-IdC	5I-dCTP, 5-iodo-dCTP, 5-iodo-2'-deoxycytidine-5'-triphosphate
5-IdU	5I-dUTP, 5-iodo-dUTP, 5-iodo-2'-deoxyuridine-5'-triphosphate
Pα/α-PO_4	Alpha phosphate group of an incoming nucleotide involved in phosphodiester bond formation during primer extension
AsnC	Asparagine synthase C transcription factor, part of bacterial and archaeal feast/famine regulatory proteins (FFRP's)
BC	Binary complex
BER	Base-excision repair
BF	*Bacillus* fragment
BRE	B recognition element, TFB/TFIIB recognition element upstream or downstream of the TATA-box
Bst	*Bacillus stearothermophilus*
cDNA	Complimentary DNA
ChIP-exo	Chromatin immunoprecipitation-exonuclease
ChIP-seq	Chromatin immunoprecipitation-sequencing
ChMMC	Chelated mismatched complex
Cryo-EM	Cryo-electron microscopy
ddNTP	2',3'-dideoxyribonucleoside-5'-triphosphate, chain terminating analogs
dGMP	2'-deoxyguanosine-5'-monophosphate
dGMPNPP	2'-Deoxyguanosine-5'-[(α,β)-imido]triphosphate, an αβ non-hydrolyzable analogue of dGTP
DNA	Deoxyribonucleic Acid
DNAP	DNA polymerase
DNAPγ	Human mitochondrial DNA polymerase gamma
DNAPI	DNA polymerase I
dNTP	2'-deoxyribonucleoside-5'-triphosphate
ds-	Double-stranded
E. coli	*Escherichia coli*
EMSA	Electrophoretic Mobility Shift Assay

exo	Gene encoding exonuclease
F_c	Calculated structure factor
F_o	Observed structure factor
FPLC	Fast-performance liquid chromatography
FRET	Fluorescence resonance energy transfer
HhH	Helix-hairpin-helix, DNA-binding motif
HPLC	High-pressure liquid chromatography
hTERT	Human telomerase reverse transcriptase
HTH	Helix-turn-helix, DNA-binding motif
K_{cat}	Catalytic rate constant for an enzyme-catalyzed reaction
K_d	Dissociation rate constant
KF	Klenow fragment
KlenTaq	Klenow (Large) fragment of *Thermus aquaticus* DNA polymerase I
KTQ	
Lrp	Leucine-responsive regulatory protein, a transcription factor that is part of bacterial and archaeal feast/famine regulatory proteins (FFRP's)
LrpB	*Sulfolobus solfataricus* Leucine-responsive regulatory protein
M. ace	*Methanosarcina acetivorans*
Me^A	Catalytic metal
Me^B	Binding metal
Mg^A	Catalytic Mg^{2+} metal
Mg^B	Binding Mg^{2+} metal
MMC	Mismatched complex
Mre	Methanogen regulator of energy metabolism
MWCO	Molecular weight cut-off
NAC	Nucleotide addition cycle
O3'/3'-OH	3'-Oxyanion catalytically competent of nucleophilic attack on the α-PO_4 of an incoming nucleotide
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein Data Bank

PEG	Polyethylene glycol
PEI	Polyethyleneimine
Pfu	*Pyrococcus furiosus*
PHP	Polymerase histidinol phosphatase
PIC	Pre-initiation complex
polA	Gene encoding the A-family DNA polymerase I in *E. coli*
Polβ	Human X-family DNA polymerase β
Polη	Human Y-family DNA polymerase η
PP_i/Pβγ	Inorganic pyrophosphate
RNA	Ribonucleic Acid
RNAP	RNA polymerase
RNA-seq	RNA sequencing
RT	Reverse transcriptase
RT-PCR	Reverse-Transcription Polymerase Chain Reaction
ss-	Single-stranded
Sso	*Sulfolobus solfataricus*
T7 DNAP	*Enterobacteriophage* T7 DNA polymerase
Taq	*Thermus aquaticus*
TBP	TATA-box binding protein
TERC	Telomerase RNA
TF	Transcription factor
TFB	Transcription factor B
TLS	Translesion DNA synthesis
TrmB	*Thermococcus* regulator of maltose binding
TrmBL	*Thermococcus*-like regulator of maltose binding
TSS	Transcription start site
wHTH	Winged helix-turn-helix
XFEL	X-ray free electron laser

ACKNOWLEDGEMENTS

Naming any one individual along this odyssey would only serve to diminish the contributions of all the wonderful individuals that have supported my endeavour, enriched my life, and, at times, tolerated indiscretions that were always without malice. This endeavour would have been intolerable and possibly unsuccessful had it not been for their love, support, and understanding. Those are individuals from every facet of my life that I have come to have a great deal of love and respect for as we grew closer and forged bonds that I hope will only become stronger over time and never falter.

The sense of camaraderie that I came to share with my cohort during our formative years as aspiring scientists, and even towards the end of our respective tenures, has been nothing short of awe-inspiring. My regret is that I did not invest more time and effort to express these sentiments.

Words fail me when attempting to express my gratitude for the love, understanding, patience, and encouragement that my family and loved ones have shown me time and again since the inception of my graduate studies. Truly, none of this would have been possible without them.

I am appreciative of my advisers who have afforded me the independence, sometimes to my detriment, to develop and mature as a scientist, my committee members for their guidance, support, and helpful suggestions that became the foundation of various parts of the work presented here, and past group members whose work has laid the foundation for mine, and of my peers who have contributed in many ways to my work and understanding.

I am particularly indebted to the scientists and collaborators at Oak Ridge National Laboratories, Cornell High Energy Synchrotron Source, The University of Kansas, and the Pennsylvania State University who have dedicated time and effort to facilitate the research and experiments described herein.

Structural and Mechanistic Enzymology of DNA Polymerases

Chapter 1

Introduction to DNA polymerases and DNA Replication

1.1 DNA Polymerases and DNA Replication.....3

1.2 DNA Polymerase Structure, Function, and Organization.....5

1.3 Proposed Mechanism of Nucleotide Incorporation and Models of Fidelity.....13

1.4 Structural Biology of A-Family DNA Polymerases.....19

1.5 Time-resolved X-ray crystallography in Structural Enzymology.....23

1.6 Specific Aims of this Work.....28

References.....30

1.1 DNA Polymerases and DNA Replication

Deoxyribonucleic acid (DNA) constitutes the genetic material in all domains of life (*Bacteria*, *Archaea,* and *Eukarya*), and acellular life (viruses whose genomes are composed of DNA). The genomes of any such organisms are double-stranded DNA molecules configured in a helical, antiparallel nature and comprised of four nitrogenous bases, the nucleotides. The purines, Adenine and Guanine base-pair with their orthogonal pyrimidine counterparts, Thymine and Cytosine, respectively. This base-pairing was initially suggested by Watson and Crick (X) to allow a DNA molecule to direct the synthesis of a new DNA molecule identical to the first. This process is what we now call DNA replication, and it is discussed in greater detail in the following chapters.

DNA replication is ubiquitous to all organisms whose genomes are comprised of DNA. The faithful transmission of genetic material through successive generations is an evolutionary imperative for the continuation of any species. Errors made during this process can lead to mutations that may or may not be phenotypic and detrimental. A single change in one of the nucleotides coding for a gene could prove inconsequential or could lead to altered protein products or changes in gene expression and regulation. The latter often leads to genomic instability, disruption of metabolic and gene regulatory pathways, and the prevalence of disease states; most notably Cancer and other neoplastic syndromes.

For this faithful transmission of genetic material to occur during DNA replication organisms must rely on sophisticated proteins, enzymes, that can match the speed of cell division for a particular cell type or organism whilst maintaining a high degree of fidelity. These enzymes are termed DNA polymerases, enzymes that are capable of synthesizing new DNA molecules under the direction of a parent DNA molecule (1,2).

DNA polymerases not only replicate genomic DNA but have evolved to perform much more sophisticated tasks in order to repair genomes and maintain genomic stability. As such, there is a direct correlation between genome complexity and the number and multisubunit characteristics of DNA polymerases (Fig. 1.1). The more complex the genome of an organism, the larger the number of DNA polymerases encoded within it to carry out highly specialized tasks.

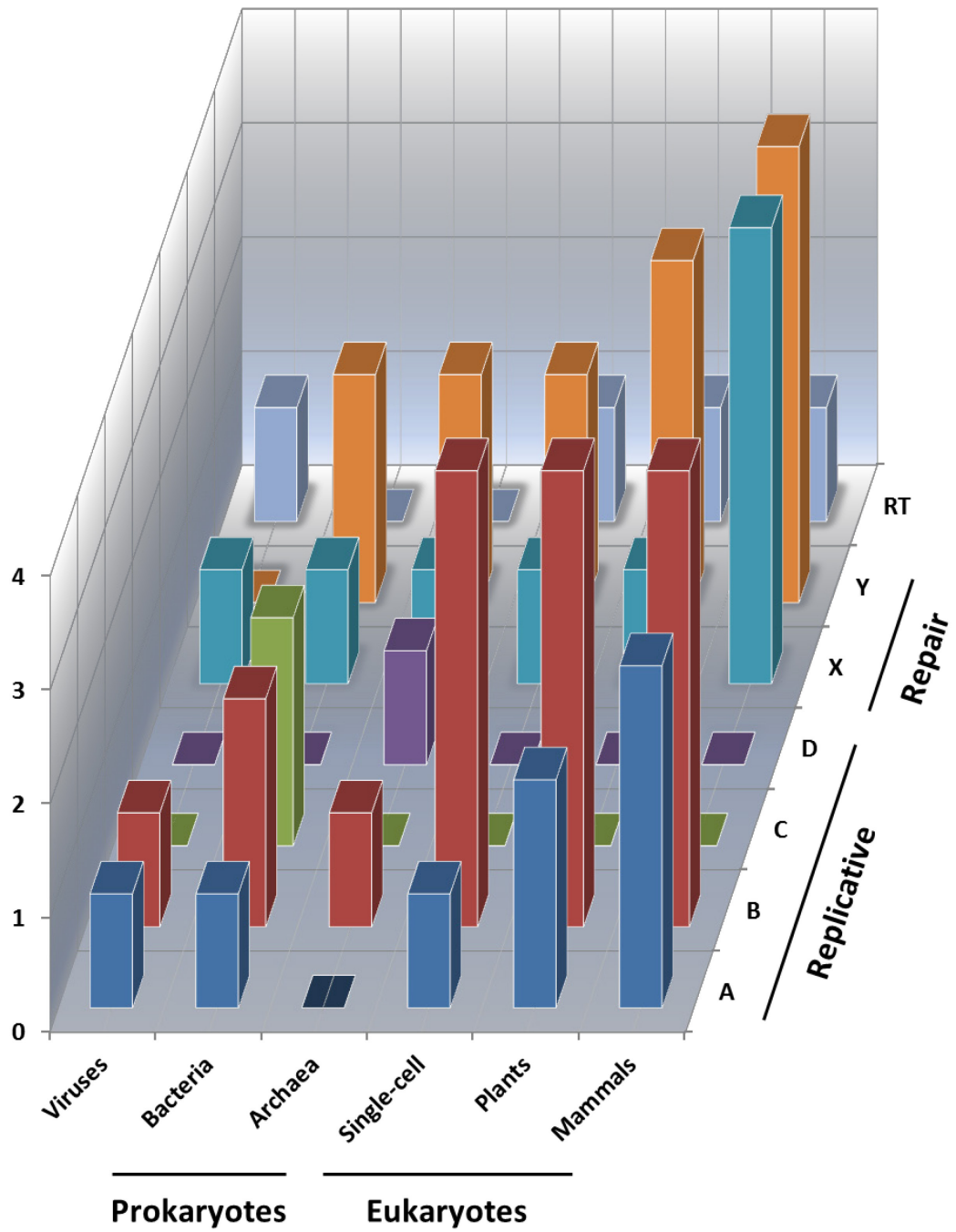


Figure 1.1 Number of DNA polymerases with increasing genome complexity. The data presented in this graph is a deconstruction of the general trend of the number, size, and family-distribution of DNA polymerases between the different domains of life (*Bacteria, Archaea, Eukaryotes*) and acellular life (Viruses), and within individual domains, with increasing genome complexity (3-5).

1.2 DNA Polymerase Structure, Function, and Organization

DNA polymerases can be classified as replicative or repair enzymes with only a few exceptions that perform both tasks (3,4) and are grouped into seven different families (A-D, X, Y, and RT) based on sequence homology and structural similarities (5) (Table 1.1). The evolution of specialized DNA polymerases is directly correlated with genome complexity, wherein simple bacteriophages contain a single replicative DNA polymerase, and higher eukaryotes contain at least 15 different DNA polymerases that participate in replication and repair.

There is a clear relationship between DNA polymerase structure and function, and common structural elements conserved across all known DNA polymerases exist (6,7). The overall structure of the polymerase domain in DNA polymerases resembles a human right hand that consists of 3 subdomains: palm, fingers, and thumb. When the DNA polymerase holoenzyme binds to duplex DNA and a 2'-deoxyribonucleoside-5'-triphosphate (dNTP) — its cognate substrate — a conformational change in the DNA polymerase and the duplex DNA leads to the assembly of an active site. The fingers domain interacts with the incoming dNTP and single-stranded DNA and forms a closure around the substrates in the active site to facilitate Watson-Crick base-pairing (8,9), and the thumb domain binds to double-stranded DNA. The palm domain contains catalytic residues that bind divalent metal ions required for the phosphoryl transfer reaction in which a 2'-deoxyribonucleoside-5'-monophosphate (dNMP) is attached to the 3'-OH group of the growing primer DNA strand and inorganic pyrophosphate (PP_i) is released.

Several DNA polymerases also contain exonuclease activities in 3'→5', 5'→3', or both directionalities through a structurally conserved exonuclease domain (5' to 3') (10) and within the polymerase domain (3' to 5') which provide such a polymerase with a secondary active-site that is in careful equilibrium with the first. The existence of exonuclease activity provides high-fidelity replication. Upon binding of duplex DNA by DNA polymerase, both active sites have been shown to initially interact sequentially with the duplex DNA in the open conformation. It is believed that once correct Watson-Crick base-pairing is verified through closure of the O-helix of the fingers subdomain on the substrates, the exonuclease active-site ceases to interact with the DNA substrate.

Table 1.1 Representative DNA polymerases & their functions from the seven families

Category	Family	DNA Polymerase	Specific Functions	Taxonomy
Replicative and Repair	A	T7 DNAP	Replication, Primer Extension	Viral
		E. coli DNAP I	Replication, repair	Bacterial
		B. stearothermophilus	Replication, repair	Bacterial
		T. aquaticus DNAP I	Replication, repair	Bacterial
		DNAP γ	Replication in mitochondria	Eukaryotic
		DNAP θ	Replication of cross-links, BER	Eukaryotic
		DNAP υ[a]	Repair	Eukaryotic
	B	T4 DNAP	Replication	Viral
		RB69 DNAP	Replication	Viral
		φ29 DNAP	Replication	Viral
		E. coli DNAP II	Repair, TLS	Bacterial
		DNAP BI	Replication	Archaeal
		DNAP BIII	Repair, base-excision (BER)	Archaeal
		DNAP α	Replication priming	Eukaryotic
		DNAP δ	Lagging strand replication, repair	Eukaryotic
		DNAP ε	Leading strand replication, repair	Eukaryotic
		DNAP ζ	Extension in translesion synthesis TLS	Eukaryotic
	C	*E. coli* DNAP III	Replication	Bacterial
		B. subtilis DNAP III	Replication	Bacterial
		T. aquaticus DNAP III	Replication	Bacterial
	D	*M. mazei* DNAP D	Replication	Archaeal
Repair	X	ASFV DNAP	Repair	Viral
		B. subtilis DNAP X	Repair	Bacterial
		T. thermophilus DNAP X	Repair	Bacterial
		T. aquaticus DNAP X	Repair	Bacterial
		M. mazei DNAP X	Repair	Archaeal
		DNAP β	Repair, base-excision	Eukaryotic
		DNAP λ	Repair, BER, DSB, TLS	Eukaryotic
		DNAP μ	Repair, non-homologous end-joining	Eukaryotic
	Y	*E. coli* DNAP IV	Repair, TLS	Bacterial
		E. coli DNAP V	Repair, TLS	Bacterial
		S. solfataricus Dpo4 DNAP	Repair, TLS	Archaeal
		DNAP η	Repair, TLS	Eukaryotic
		DNAP κ	Repair, TLS	Eukaryotic
		DNAP ι	Repair, TLS	Eukaryotic
	RT	Reverse Transcriptase	RNA-dependent DNA synthesis	Viral
		Telomerase	Replication of chromosome ends	Eukaryotic

However, once an incorrect nucleotide is incorporated, the newly synthesized DNA is released from the polymerase active-site and is bound by the exonuclease active-site for hydrolysis of the incorrect base.

A-Family DNA polymerases are perhaps the most well-studied family of replicative and repair DNA polymerases after the discovery of the prototypical DNA polymerase I from *E. coli* in 1958 by Arthur Kornberg (11). Over the years, as more and more DNA polymerases were discovered, the need to systematically characterize and classify them due to their differences resulted in a framework that utilized amino acid sequence homology and structural similarities to group these enzymes in classes distinct from each other but highly conserved within the constituents of each class.

These DNA polymerases shared common structural and functional organizations. The quintessential A-family DNA polymerase I from *E. coli* and other members of this family are organized in three functional domains: an N-terminal domain containing 5'-3' exonuclease activity, a central domain containing 3'-5' exonuclease proofreading activity, and a C-terminal domain containing the polymerase activity (12). This organization, coupled with amino acid sequence homology for these conserved motifs, is the fundamental characteristic for this family of DNA polymerases which included viruses, bacteria, and the eukaryotic mitochondrial DNA polymerase γ.

B-Family DNA polymerases, using the same scheme for A-family DNA polymerases, are homologous to *E. coli* DNA polymerase II. This second *E. coli* DNA polymerase was discovered a decade after its predecessor by genetic and biochemical characterization of a mutant strain of *E. coli* (*polA1*) that had significantly reduced levels of DNA polymerase I but was still capable of growth. This new DNA polymerase retains 3'-5' exonuclease activity, but the responsible domain is spatially organized in a different manner relative to its counterpart in A-family polymerases (Fig. 1.2) (13). Interestingly, this polymerase performs a host of functions not possible with its predecessor. Its gene expression level, or amount of protein product, is induced sevenfold in response to DNA damage (14). It was genetically mapped and identified as the *E. coli* dinA gene, one of many confirmed genes induced during the SOS response to DNA damage (15).

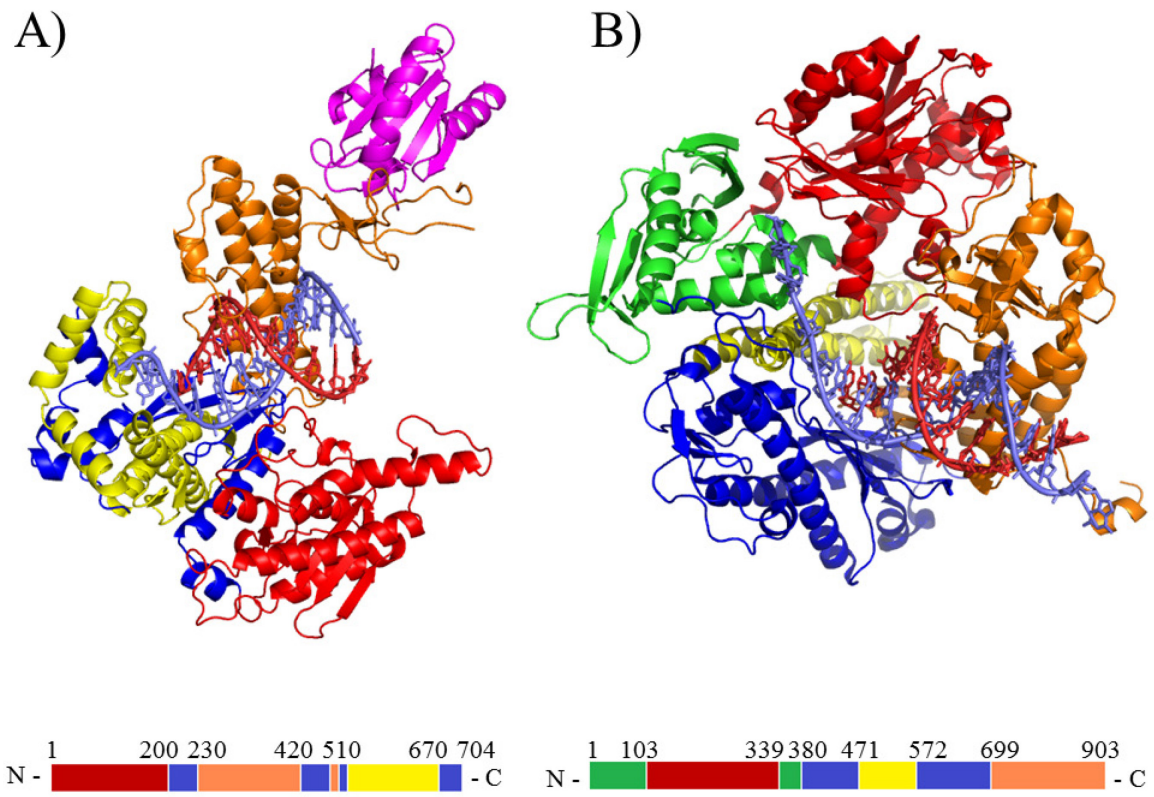


Figure 1.2 Domain organization of families A and B DNA polymerases. The figure shows the relative locations of the Exonuclease domain in prototypical A and B family DNA polymerases A) T7 and B) RB69, respectively. The exonuclease domain is shown in red, the thumb domain in orange, the fingers domain in yellow, and the palm domain in blue. The exonuclease domain in T7 DNA polymerase is located away from the polymerase active site, whereas in RB69 it is nestled between the thumb and fingers domains, providing for a drastic difference in proof-reading capabilities by both polymerases, where polymerase-exonuclease switching is more energy consuming for T7 DNA polymerase. In A) *E. coli* Thioredoxin, T7 DNA polymerase processivity factor is shown in magenta and in B) spacers in the RB69 DNA polymerase are shown in green. Primer:template duplex shown in red and blue, respectively.

Such a mechanism is often prevalent when *E. coli* cells are exposed to ultraviolet (UV) radiation. Based on this evidence and subsequent studies it was established that *E. coli* DNA polymerase II was also responsible for error-free replication restart in lesion bypass; a situation where DNA replication is blocked due to template DNA damage (16).

This family of DNA polymerases grew to encompass replicative and repair DNA polymerases from viruses, bacteria, and archaea as well as the major eukaryotic multi-subunit replicative and repair DNA polymerases α, δ, ε, and ζ which are responsible for priming of DNA replication, lagging strand replication and repair, leading strand replication and repair, and translesion DNA synthesis (TLS), respectively.

Multisubunit characteristics of DNA polymerases are not exclusive to the eukaryotic enzymes. Arthur Kornberg, along with two other labs, isolated a replicative DNA polymerase III from *E. coli polA* mutant strain that was found to be very inefficient in DNA replication *in vitro* compared to DNA polymerase I. However, a crude extract of that same mutant was much more efficient in DNA replication *in vitro*. This observation suggested that a more complex form of this DNA polymerase exists. Subsequent isolation of this complex proved that it contained ten subunits, including a clamp loader and sliding clamp — proteins that increase the processivity of this polymerase by stabilizing complexes with single- and double-stranded DNA (17). Upon further investigation, it was found that this DNA polymerase III holoenzyme was the major DNA replication complex in *E. coli* responsible for leading and lagging strand chromosomal DNA replication. No 3'-5' exonuclease activity was encoded by the catalytic subunit of *E. coli* DNA polymerase III (α subunit), it was instead encoded by an accessory subunit (ε) within the holoenzyme (18) and thus emerged C-Family DNA polymerases – multisubunit enzymes, exclusive to bacteria, which perform editing and proofreading by an accessory subunit rather than the catalytic DNA polymerase subunit.

Archaea are the only known species characterized thus far that have been shown to contain what are now called D-Family DNA polymerases (18). This DNA polymerase D is a heterodimer composed of a large DP2 and small DP1 subunits. The large subunit (DP2) is the catalytic subunit while the small (DP1) subunit acts as an accessory factor, and their interaction has been shown to be a prerequisite for optimal 3'-5' exonuclease

activity (19). Further studies have shown that the small subunit possessed an intrinsic proofreading activity and was homologous to the small accessory subunits of eukaryotic B-family DNA polymerases α, δ and ε (20).

X-Family DNA polymerases are fascinating in that they are the only class of DNA polymerases that exist in viruses, bacteria, archaea, and eukaryotes, and are exclusively repair enzymes (21). They are not limited to only one DNA repair pathway, as is the case for Y-family DNA polymerases discussed below. Because these enzymes participate in a multitude of DNA repair pathways, it appears that they share distinct homologies and organizations within a particular domain of life, but have stark variations in domain organization between various domains of life and acellular life (22). The unifying structural theme of this family of DNA polymerases is the pol-X polymerase domain. In eukaryotes, DNA polymerase β (Pol β), an X-family repair DNA polymerase, contains an N-terminal 5'-deoxyribose 5'-phosphate lyase (dRP-lyase) domain that contains a Helix-hairpin-Helix (HhH) motif which interacts with DNA downstream of a gap (abasic site, etc.) and aids in the removal of 5'-terminal abasic sugars from damaged DNA (23). Its function lies in base-excision repair (BER) of damaged DNA. Bacteria and archaea have a Pol β-like domain fused at its C-terminus to a phosphodiesterase domain, also known as polymerase histidinol phosphatase (PHP) domain (24). This domain contains an intrinsic 3'-5' exonuclease activity residing within it that allows the removal of single-stranded 3'-termini in gapped DNA. Taken together with the Pol β-like domain, bacteria and archaea are capable of not only base-excision repair but also non-homologous end joining (NHEJ) in the repair of damaged DNA by filling in gaps (25).

Excluding Pol β, all other eukaryotic X-family DNA polymerases (λ, μ, TdT, etc.) possess a BRCT domain fused to the N-terminus of a Pol β-like domain (26). The BRCT domain is the C-terminal fragment of BRCA1 (breast cancer susceptibility protein-1) that mediates protein-protein and protein-DNA interactions (27). The interactions mediated by this domain allow these eukaryotic repair polymerases to not only perform base-excision repair but also perform double-strand break (DSB) repair and immunoglobulin V(D)J recombinational repair. Only Pol λ is capable of performing translesion DNA synthesis (TLS) in addition to the other repair functions (28).

Translesion DNA synthesis (TLS), as its name implies, is a bypass strategy that allows DNA synthesis or replication across damaged sites in DNA. During DNA replication when a DNA polymerase encounters a site containing a lesion (thymine dimers, bulky base adducts) or an abasic site, events are triggered in an attempt to correct the damage to the site by base-excision, nucleotide-excision, or mismatch repair, or the more laborious homologous recombination and non-homologous end joining. In certain circumstances, the damage cannot be repaired, and the site must be bypassed in order to continue DNA replication and sustain genome stability (29). This scenario induces the TLS response which, owing to its function and the situation at hand, is a last-resort when attempting to salvage DNA replication. This response is carried out by the highly specialized, but always error-prone, Y-family DNA polymerases (30).

Y-Family DNA polymerases exclusively carry out translesion synthesis and are restricted to a last-resort approach due to their considerably lower fidelity because of the absence of 3'-5' exonuclease activity and lax structural constraints in their active sites when compared to their replicative counterparts (31). Because of the nature of this repair route, their low fidelity provides for flexibility in bypassing lesions. A structural basis for this low-fidelity was deduced from crystal structures of archaeal DNA polymerases Dbh and Dpo4 (32). It was found that, while the palm domain was nearly identical to its counterparts in other families, the fingers domain appears in the closed conformation regardless of the presence or absence of any substrate. Just as interesting was the finding that the enzyme made minimal, non-specific contacts with its substrate (33). These translesion polymerases take one of two routes, error-free, and error-prone bypass, in their function depending on the lesion encountered and the polymerase used. For example, lesions induced by UV irradiation are corrected in an error-free bypass by human pol η, but pol ι will introduce mutations at this site (34).

The last family in this enzyme class is the reverse transcriptase, or RT-family of DNA polymerases. These RNA-directed DNA polymerases exist in retroviruses, nearly all eukaryotes, and some parasites, and are quite distinct in their function (35). Retroviruses are so called due to the requirement that their RNA genomes be transformed by reverse transcription into single-stranded DNA—one prerequisite for integration into a

host genome. The most prominent of these RT's are HIV-I reverse transcriptase and human Telomerase, which is responsible for the addition of 6 nucleotide repeats to the 3'-end (the telomeres) of chromosomes, thereby slowing the progressive loss of genetic information that occurs after successive cycles of chromosome replication in cell division (36).

HIV-I reverse transcriptase (RT) is an asymmetric dimer composed of two subunits, p66 and p51, that are derived from proteolytic cleavage of precursor proteins in the virus genome. The p51 subunit is in fact derived from proteolytic cleavage of the full-length p66 subunit at its RNase H-containing C-terminus (37). This RT possesses both a DNA polymerase domain and an RNase H domain that are separated by a 100 amino acid residue connector domain in the p66 subunit of the asymmetric dimer. The RNase H domain is responsible for the removal of RNA primers from RNA:DNA hybrids that are used to prime the reverse transcription reaction for the production of complementary DNA (cDNA) to the viral genes (38). RNase H is not exclusive to reverse transcriptases and is often found in many DNA polymerases to specifically carry out primer removal (39). The DNA polymerase is similar to its counterparts in other families with respect to the three subdomains (fingers, palm, and thumb), however, the truncated p51 subunit, despite having an intact DNA polymerase domain, has these three domains spatially rearranged to prevent extensive interaction with the substrate DNA (40).

Human Telomerase (hTERT) is a ribonucleoprotein complex capable of adding 6-nucleotide repeats of 5'-TTAGGG-3' (telomeres) to the 3' ends of human chromosomes, and is composed of two molecules each of telomerase reverse transcriptase (TERT), telomerase RNA (TERC), and dyskerin protein which is required for correct processing of TERC (41). The TERC templating region contains the sequence 3'-CAAUCCCAAUC-5' which allows the telomerase to bind the first few nucleotides of the template to the last telomere sequence on the chromosome and reverse transcribe it to obtain a new telomere repeat at the 3'-end of the chromosome. It then translocates this TERC:cDNA hybrid and realigns the new 3'-end telomere to the template and repeats the process. A DNA polymerase then fills the gaps in the lagging strand (42).

1.3 Proposed Mechanism of Nucleotide Incorporation, Fidelity, and the Fingers Domain O-helix

DNA polymerase fidelity results in the accurate replication of a DNA template and is a requirement to prevent, or at least reduce, mutations in the newly synthesized DNA strand that result from misincorporation. The mechanism of nucleotide incorporation follows five steps (Fig. 1.3) (43-45). First, a binary complex is formed by the binding of DNA polymerase to its substrate double-stranded DNA. A correct incoming deoxyribonucleoside triphosphate (dNTP) then binds to the binary complex, forming a ternary complex that also includes two divalent metal ions which mediate formation of a pentacovalent coordination complex between the triphosphate moiety of the incoming nucleotide, two carboxylate-donating Asp residues, and the 3'-oxyanion of the primers 3'-terminus. A conformational change occurs in the newly formed ternary complex from open to closed conformation, followed by a chemistry step leading to phosphodiester bond formation between the α-phosphate of the incoming dNTP and the 3'-oxyanion at the 3'-end of the growing primer strand. Finally, inorganic pyrophosphate is released as a byproduct, and the polymerase reverts to the open conformation to facilitate translocation to the new templating base. At the end of this nucleotide addition cycle, post-replication proofreading occurs prior to the subsequent cycle. Fidelity of nucleotide incorporation is largely affected by the degree of discrimination during nucleotide binding, conformational change and the chemistry step (46).

The binding of an incoming nucleotide to a binary complex in itself is an intricate network of thermodynamic, kinetic and geometric parameters that must all be satisfied before a nucleotide is accepted to undergo catalysis. In general, an incoming nucleotide must form the proper hydrogen bonds with the templating base and base-stack properly with neighbouring nucleotides (47). These first two criteria are required to obtain a proper geometry of the incoming nucleotides triphosphate moiety in a catalytically-ready state (48). Proper base pairing is energetically favourable due to entropic contributions by freed water molecules and the low energetic cost of hydrogen bond formation. However, base pairing is not required for efficient dNTP incorporation for some DNA polymerases (49) but does significantly increase fidelity by serving as an initial checkpoint for correct base

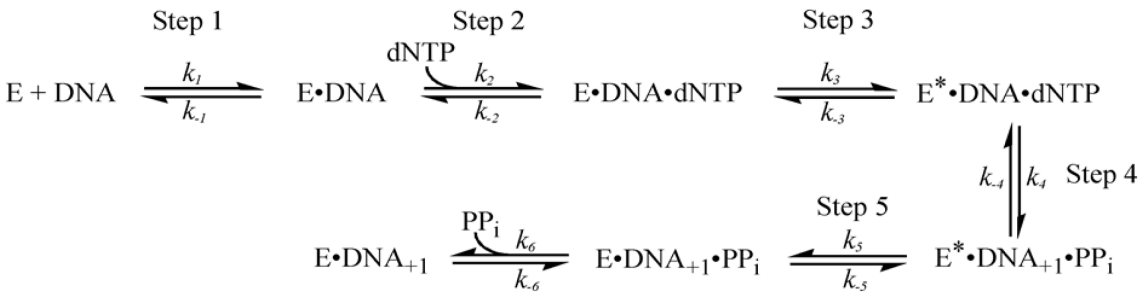


Figure 1.3 Sequence of Events During Nucleotide Incorporation. The scheme presented here is the minimal kinetic pathway for nucleotide incorporation by DNA polymerases based on current literature. Here, after formation of a binary complex (1), an incoming nucleotide enters the active site (2) and induces a conformational change to the closed ternary complex poised for catalysis (3). As chemistry occurs (4), a phosphodiester bond forms between the α-PO_4 of the incoming nucleotide and the primer 3'-end, via cleavage of the αβ-PO_4 phosphodiester bond, resulting in a newly incorporated dNMP, and PP_i. The enzyme returns to an inactive state (5) and subsequently releases PP_i to generate the binary complex once more for the next cycle of nucleotide incorporation.

pair geometry (50). Base pairing with an incorrect nucleotide would change hydrogen bonding and base stacking patterns, as is the case with Hoogsteen base-pairs, which would in turn force the incoming nucleotides triphosphate moiety to adopt an incorrect geometry that is not conducive to catalysis (51).

Enthalpic and entropic contributions to free energy change by binding of correct versus incorrect nucleotide to the polymerase active site have been shown to contribute to fidelity but are not large enough to explain the differential fidelity of various polymerases (52). Combining free energy change produced through exclusion of water with the free energy change obtained through imposition of the geometric constraints of the polymerase active site provides approximately 1-3 kcal/mol that are primarily contributed through base pair hydrogen bonds and reduction of solvation around an incoming nucleotide (53). This alone does not explain the high selectivity of several polymerases for the correct over incorrect nucleotide that in some instances was as much as ten-fold higher (54).

Base stacking in nucleotide selectivity has garnered many proponents based on evidence that it is perhaps a much more discriminatory step compared to hydrogen bonding in base pairs. Base stacking is a stabilizing force in duplex DNA and contributes heavily to helix stability (55). Studies using nucleotide analogs lacking hydrogen bonding capabilities but retaining base stacking properties show that these analogs, which have similar geometries to their natural counterparts, can be incorporated by some DNA polymerases (56), albeit with significantly reduced catalytic rates and fidelity (57), indicating that while hydrogen bonds are critical to polymerase fidelity, base stacking that provides proper geometry is much more discriminatory in the proper alignment with active site residues for catalysis (58).

Many polymerases exploit base-stacking as a significant contributor to their mechanism of nucleotide selectivity. This is a characteristic of A-family polymerases known as “base flipping” where an evolutionarily conserved Tyr residue (*E. coli* Klenow 766; *Bacillus* 714; *Taq* 671) mimics and takes the position of the 5’-nucleobase adjacent to the first templating base by “flipping” the displaced 5’-nucleobase out of the helix and mediating proper base stacking with the templating base (59). Similarly, a conserved Phe

residue in these polymerases stacks against, and stabilizes, the incoming nucleotide's nucleobase to facilitate proper geometry (60). Studies of mutant Klenow polymerases containing "non-stacking" amino acid residues in place of Tyr766 in *E.* coli showed a significant decrease in fidelity (61). This result indicated that base stacking plays a vital role in the binding and stabilization of the incoming nucleotide.

Hydrogen bonding and base stacking, taken together, serve as an initial checkpoint in the fidelity of nucleotide incorporation. However, the free energy change of binding correct versus incorrect nucleotide contributed by these interactions as stated previously does not provide a universal mechanism that adequately explains the differential fidelity of various polymerases and therefore, while appealing, is unlikely to be the rate-limiting step in nucleotide incorporation (62). Kinetic studies performed to examine the differential binding of correct and incorrect nucleotides by Klenow polymerase, T7 DNA polymerase and the Klenow fragment of *Taq* DNA polymerase (KlenTaq), have shown binding affinities on average 200 to 400 times greater for the binding of the correct nucleotide compared to the incorrect nucleotide (K_d's of 20μM and 4-8mM, respectively) (63-65), and have also shown that Klenow polymerase, compared to its counterparts, has much lower overall fidelity due to lack in initial selectivity with K_d's of 5uM and 12uM for the binding of correct and incorrect nucleotides, respectively (66).

These results prompted the proposal of two models to explain the potential mechanisms underlying the significant differences in fidelity. An "induced-fit" model posits that polymerase fidelity would be the result of conformational coupling in which the free energy change from correct nucleotide binding is used to drive a rate-limiting conformational change in the polymerase that precedes the chemistry step (67). Nucleotide selection in this model would be a two-step process in which ground state binding of the correct nucleotide, which would be more favourable compared to the incorrect one, is followed by the formation of a catalytically-competent ternary complex driven by the favourable energy produced from said ground state binding. The second "rate-limiting transition state" model posits that polymerase fidelity is attributed to a rate-limiting chemistry step in which the free energy change in the formation of transition state species of correct and incorrect nucleotides dictates whether the polymerase will

proceed in the chemistry step or revert to an open conformation and release the bound nucleotide (68). In this model, binding of the correct nucleotide is followed by a rapid conformational change and a subsequent slower chemistry step. Evidence in support of both models exists, and it appears that different polymerases, and different families, adopt different mechanisms to sustain fidelity (69,70).

The chemistry step during nucleotide incorporation (also referred to as nucleotidyl or phosphoryl transfer) is responsible for the formation of a phosphodiester bond between the incoming nucleotide and the 3'-end of the growing primer strand. During this step a pentacovalent transition state intermediate is formed by a bidentate interaction of the β- and γ-phosphates of the incoming nucleotide with a divalent (binding, B) metal ion that is also coordinated to oxyanions in the carboxylates of catalytic residues and the α-phosphate, and a divalent (catalytic, A) metal ion that coordinates oxyanions from the same catalytic residues to the α-phosphate and the 3'-oxyanion from the 3'-end of the primer (Fig. 1.4) (71).

Conforming to a two metal ion-dependent phosphoryl transfer mechanism (discussed in greater detail in the subsequent chapter), the reaction occurs through activation of the 3'-hydroxyl by the catalytic metal, Me^{A}, which then allows the resultant 3'-oxyanion to act as a nucleophile and attack the α-phosphate of the incoming nucleotide. This step causes a negative charge buildup on the αβ-bridging (leaving) oxygen, which is then stabilized by the binding metal, Me^{B}, that also chelates the βγ-phosphate (pyrophosphate) leaving group. Thus, both metals are required to stabilize the structure and charge of the pentacovalent transition state (72). It should be noted that evidence is accumulating in support of 3'-oxyanion activation using water molecules observed in several crystal structures near the 3'-OH rather than the catalytic metal (73). It should also be noted that some enzymes require three carboxylate-containing catalytic residues (Asp, Glu, or a combination), one of which is hypothesized to activate the 3'-oxyanion, whereas others specifically utilize one Glu and two Asp residues for carboxylate group coordination (74).

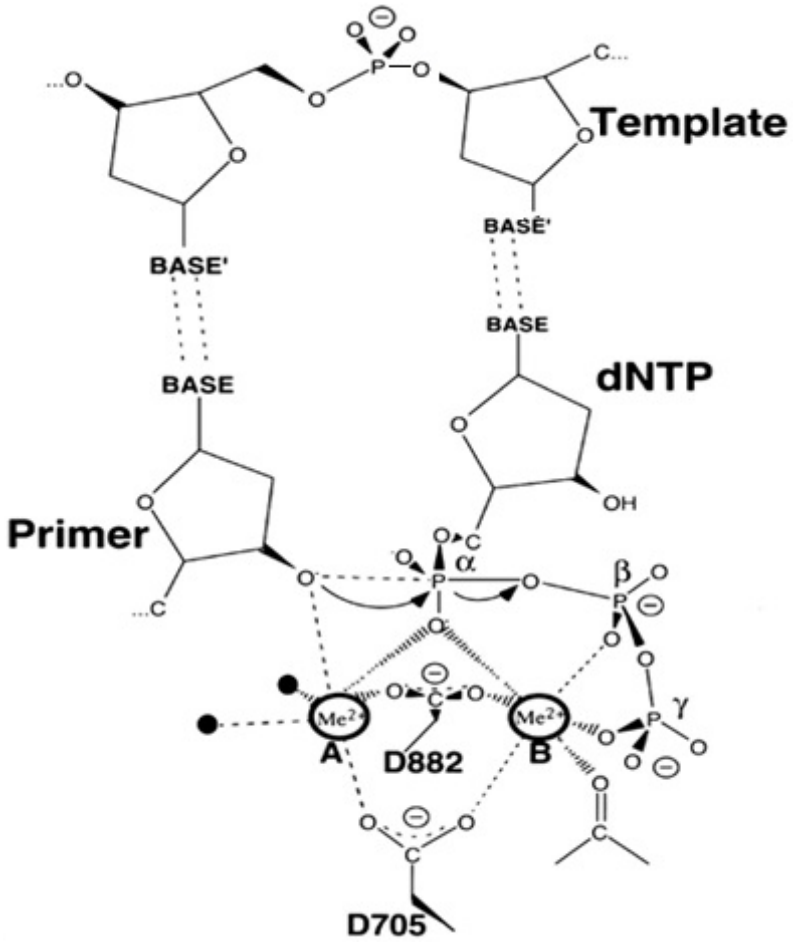


Figure 1.4 Two-metal ion mechanism of DNA polymerase. The two conserved aspartates have the *E. coli* DNA polymerase I numbers. The active site features two metal ions that stabilize the resulting pentacoordinated transition state. Metal ion A activates the primer's 3′-OH for attack on the α-phosphate of the dNTP. Metal ion B plays the dual role of stabilizing the negative charge that builds up on the leaving oxygen and chelating the β- and γ-phosphates.

Modified from Steitz *et al*. (1998) A mechanism for all polymerases. *Nature* **391**, 231-232

1.4 Structural Biology of A-Family DNA Polymerases

The first solved crystal structure of a DNA polymerase was that of the Klenow (large) fragment of *E. coli* DNA polymerase I (75), making it a structural template to which all subsequent structures could be aligned and compared. Since then, a plethora of DNA polymerases from various organisms, belonging to different families, were crystallized and their structures solved. This development led to an abundance of structural evidence that provided a better understanding of the structure-function relationship both within and between families of DNA polymerases. A biochemical understanding of the different functions several polymerases perform also provided empirical rules that aided in the successful crystallization of protein-nucleic acid complexes.

A-Family DNA polymerases were, for many years, the primary focus of structure-function relationship studies. Not surprisingly, the number of crystal structures for this family of polymerases in the Protein Data Bank (PDB) makes up approximately a third of available DNA polymerase structures. The accumulation of structures of the apoenzymes, and their complexes with DNA (binary) and nucleotides or their analogues (ternary), has shown marked structural changes that accompany the different states of the enzyme (76).

As previously described, the subdomain organization of the polymerase domains of DNA polymerases includes a palm, fingers, and thumb subdomain that serve individual and coordinated functions in DNA binding and binary complex formation, and the subsequent nucleotide (or analogue) binding and ternary complex formation leading to catalysis. The interaction and spatial arrangement of these three subdomains forms a DNA-binding crevice that allows double-stranded DNA to bind and subsequently become stabilized by a network of interactions with amino acid residues from each of these subdomains. The base of this crevice is the palm subdomain which contains three acidic residues responsible for catalysis through their carboxylates, while the upper edges of the crevice are made up of the thumb and fingers subdomains that stabilize the DNA substrate and interact with the growing primer strand, respectively.

The fingers subdomain in A-family polymerases carries an O-helix (Fig. 1.5), so designated based on the original structure of *E. coli* DNA polymerase I, which is largely responsible for the high fidelity in this family of polymerases. In the open conformation of the binary complex the O-helix of the fingers subdomain is seen rotated outwards by approximately 46°, similar to the apo form of the enzyme, and showing a conserved Tyr residue stacking against the 5'-templating base and mediating "base flipping" of the neighboring 5'-nucleobase of the template strand (77). This stacking is said to facilitate better base pairing with the incoming nucleotide.

As the incoming nucleotide binds to form a ternary complex, the O-helix rotates inwards by approximately 40° to form a complete active site that now includes a network of residues afforded by the O-helix that further interact with and stabilize the incoming nucleotide (77). Three distinct networks of amino acid motifs (A, B and C) are characteristic of the A-family, with motif B encoding the O-helix. Early evidence suggests that the movement of the O-helix, due to its extent, is likely the rate-limiting step for catalysis in this family of polymerases, as such a large conformational change is likely to be energetically costly.

However, Fluorescence Resonance Energy Transfer (FRET) analysis to study O-helix motion and catalysis in KlenTaq DNA polymerase suggests that closure of the O-helix is in fact not the rate-limiting step, but precedes such a step, which is believed to be the chemistry step (78). Thus, two lines of evidence now support both models for fidelity by A-family DNA polymerases and conclusive evidence to support or negate either is yet to be found.

After nucleotidyl transfer, two events must occur; inorganic pyrophosphate (PP_i) release and reversion to the open complex. It is not clear which comes first, and whether either event provides some intrinsic energy for the other to occur. Once a nucleotide is incorporated the polymerase must translocate to the next templating nucleobase but it is not yet clear whether it can do so in a sliding fashion with the O-helix in the closed conformation, or whether it needs to revert to the open conformation, and whether reversion also requires it to physically dissociate from its substrate, translocate and re-

form the binary complex. For subsequent nucleotide incorporation cycles to occur, the O-helix must first revert to the open conformation to allow for the aforementioned steps in nucleotide binding to occur prior to catalysis, but whether or not this is required for, or is a consequence of, translocation is not well understood. If this is the case then translocation, whether by sliding onto the next templating nucleobase, or physically dissociating and re-associating with its substrate DNA, must be a costly energy-driven process. The origin of such energy should be investigated along with contributions, if any, made by pyrophosphate release.

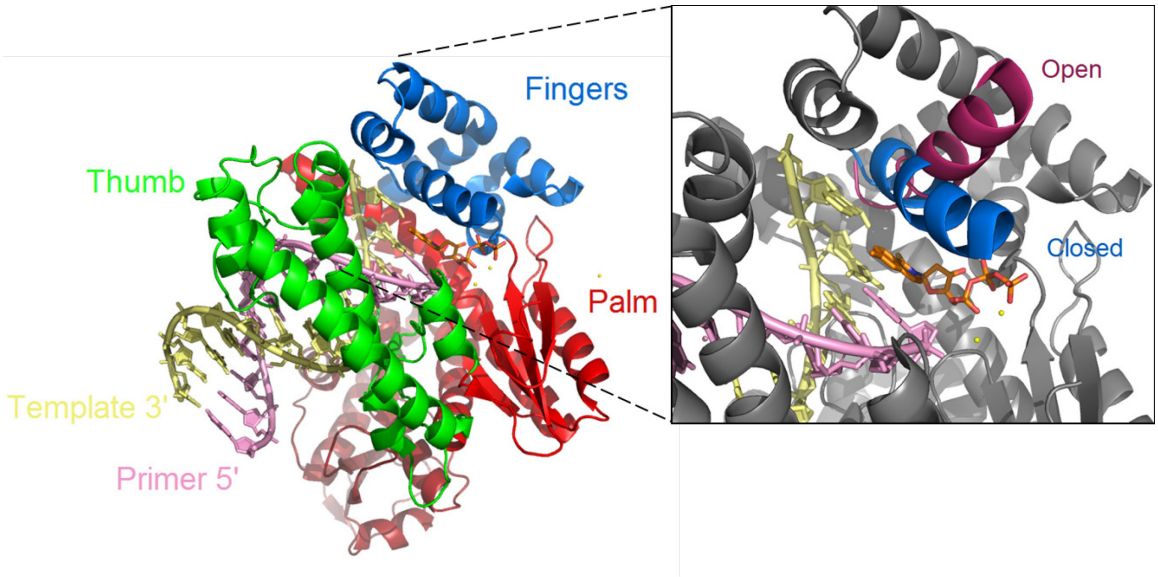


Figure 1.5 Crystal structure of the closed ternary complex of KlenTaq DNA polymerase. The subdomain organization of the polymerase domain of a prototypical A-family DNA polymerase shown in the closed conformation, bound to duplex DNA. The inlay is a close-up view of the O-helix with the open (magenta) and closed (blue) conformations superimposed. The duplex DNA is shown in yellow (template) and pink (primer).

1.5 Time-resolved X-ray crystallography for studies in structural and mechanistic enzymology

Time-resolved X-ray crystallography using Laue diffraction—diffraction using polychromatic X-rays—was established as an empirical method in the study of structural changes in proteins in the late 1970s. This was possible due to the high solvent content often found in protein crystals, which allowed the induction of significant, biologically relevant structural changes (79). The first example to employ this technique was a study on muscle twitching mediated by myosin contractions (80). It allowed researchers to study differences in the intensities of X-ray reflections as a function of time on a millisecond timescale. Data would then be complemented with spectroscopic methods and existing biochemical evidence to establish a correlation with a working model. The method, however, was impractical due to severe limitations that preclude its use as a primary tool in *in crystallo* mechanistic enzymology.

The method was limited by the requirement of the use of high-intensity, polychromatic synchrotron radiation for rapid data collection from protein crystals (81). The reason for the success of this method is also one that hinders it. Such high-intensity synchrotron radiation, while allowing rapid data collection, also causes tremendous radiation damage to the protein crystal that quickly "dies" (82). Another limitation to this approach was the localized application of these polychromatic X-rays. In this scenario, the protein crystal is stationary with no mechanical movement that allows rotation upon any one axis, which results in much more rapid decay of the protein crystal due to highly concentrated radiation damage.

The different kinetic constants for a particular protein and their correlation with structural changes often require modifications to a time-resolved Laue diffraction experiment. A typical time-resolved Laue diffraction experiment hinges on two steps: reaction initiation and monitoring of reaction progression. Both steps pose their own unique challenges and ultimately affect protein- and crystal-dependent diffraction quality.

Reaction initiation may be as simple as diffusion of a substrate into a protein crystal or may prove to be an elaborate, intricate scheme requiring specialized reagents

and equipment. The initiation must be rapid and uniform throughout the protein crystal without causing damage to protein molecules or disrupting the crystal lattice (83). Rapid initiation of the reaction is necessary to avoid convolution of the time-course of catalysis with that of initiating the reaction. Similarly, the initiation must be uniform throughout the protein crystal to avoid initial spatial gradients where a fraction of protein molecules in the crystal are undergoing catalysis while others await interaction with the substrate. Both criteria are paramount to obtaining isomorphic diffraction patterns and minimizing damage and perturbation of the crystal lattice.

Reaction initiation can be triggered by a multitude of methods. The most direct method is soaking of a protein crystal in a solution containing substrate, activator or co-factor which diffuses into the crystal to interact with protein molecules (84) — provided that a crystal is of diffraction quality and common dimensions (i.e., not excessively large and within 300μm^3 in volume) substrate diffusion can uniformly occur on a time-scale of seconds, which is substrate-dependent. It may be longer for larger crystals, solutions of high viscosity, and interactions with other agents. Such aspects depend on the nature of the protein under study. If catalysis occurs rapidly and reaction intermediates are short-lived, simple diffusion of substrate would be impractical as reaction intermediates may progress to their respective products before diffraction studies begin (85).

To overcome obstacles arising from the rate of catalysis, it is possible to exploit the physicochemical properties of the enzyme such as withholding necessary metal-ions or co-factors, manipulation of pH, salts, precipitants, and temperature and later introducing them to initiate the reaction. Several DNA polymerases have been crystallized in an inactive form by exploiting the fact that they are able to accept metal ions of similar geometry and atomic radii which were not conducive to catalysis, such as Ca^{2+} and Zn^{2+} for some polymerases, and then rapidly exchange these for the endogenous metal ion for catalysis to occur (86). Similarly, as was shown for RNase A, the first example of exploiting temperature in reaction initiation, in very low temperatures (212K) RNase A will not bind substrate or inhibitor despite its presence in the protein crystal. However, when the temperature is raised to 228K, the enzyme rapidly binds substrate to undergo catalysis (87).

With the emergence of more advanced synthetic technologies, it became possible, based on *a priori* knowledge of the interactions between substrate and protein, to synthesize substrate analogues in an inactive form that can still interact with their target proteins but cannot undergo catalysis until triggered by altering physicochemical parameters or activation by some agent. An example of this is reaction initiation in crystals of Ha-Ras p21 in complex with caged GTP (88). This GTP analogue contains a 2-nitrophenylethyl protecting group on the γ-phosphate of GTP that renders this "caged" analogue inactive until such time as the protecting group is removed. Because this GTPase activity renders short-lived reaction intermediates, it was necessary to utilize a strategy where GTP hydrolysis of this caged analogue can be triggered immediately during data collection. As such, the photolabile 2-nitrophenylethyl group on the analogue in the p21-caged GTP complex was cleaved by light pulses to release active GTP that was immediately hydrolyzed by the GTPase activity of the p21 protein. This strategy allowed the study of the structural changes accompanying GTP hydrolysis by p21 oncogene to produce GDP. It is also possible to utilize photo-excitation to activate naturally photosensitive enzymes such as heme proteins (89). In this example, complexes of carbon monoxide with heme proteins are prepared in a reversibly inactive photosensitive state, and the reaction is initiated by light pulses.

Even with such controlled reaction initiation, data collection and exposure times were lengthy. Hours and even days were necessary to obtain complete data sets. This aspect in itself proved to be another obstacle for monitoring reaction progression in protein crystals during X-ray diffraction. The assumption that a uniform saturation of protein molecules with substrate, coupled with rapid initiation, would provide an isomorphic population of protein molecules behaving similarly in a crystal lattice is sometimes invalid. Structural changes in one protein molecule do not necessarily favour or facilitate a similar structural change in neighbouring protein molecules primarily due to large solvent content that provides some flexibility in crystal packing unless crystal packing is partially maintained by symmetry-related molecules. It is more likely that protein molecules behave independently of one another (90), within the limitations and constraints imposed by crystal packing. This can lead to incomplete occupancy and less-than-uniform distribution of the substrate that was initially believed to have saturated all

protein molecules in the crystal. Thus, a polymorphic population of protein molecules with slightly different conformations may arise, and the total structure factor of any one reflection will be a weighted vector sum of each conformations structure factor based on its fractional occupancy. However, over time, this population will evolve into an isomorphic one depending on the individual proteins rate constant and the concentration of substrate diffused into the protein crystal.

Another limitation and perhaps the most important is that the methods above will likely only yield reaction intermediates and structural changes involved in catalysis while providing very little, if any, structural information for events preceding catalysis (e.g., substrate binding-induced structural changes). Thus, to increase the probability of success in any time-resolved X-ray crystallographic experiment, it is imperative that a candidate protein be crystallized and its structure solved in the open or apo, intermediate-trapped, and closed conformations. These initial static structures are often simple to obtain, provided that the protein is amenable to crystallization in these states. Such structures would simplify analysis of any discrete intermediates and structural changes that occur as the protein changes conformation.

Technologies and techniques pertinent to the method have advanced significantly to ameliorate many of the obstacles above. The advent of more powerful “home-source” X-ray generators with rotating anodes, improved beam-line synchrotron sources, microfocus beams with diameters in the 100μm range, and new photodetectors coupled to high-performance computing has led to millisecond exposure times and rapid data processing with minimal crystal damage. A complete data set of diffraction patterns for a single crystal can be obtained in mere minutes rather than days. Such short exposure times significantly reduce signal-to-noise ratio in data processing, and the possibility of reaction completion prior to complete data set collection.

Many contemporary examples of the method provide insight into the advanced techniques to manipulate the entire course of the reaction and eventually obtain a rather complete profile of all structural changes that occur in the crystal to show the mechanistic details of substrate binding, reaction intermediate formation (including short-lived

species), and product formation (91). One such example is the “Soak-trigger-freeze” or “Soak-freeze” method, which our lab uses extensively and is a fundamental part of the work presented in this manuscript. The method exploits the kinetics of an enzyme, through alterations in the physicochemical environment, by either soaking preformed crystals of the candidate enzyme in solutions with low pH, salts, and so forth, or by equilibration of the crystal in temperatures that would reduce the catalytic rate of an enzyme by several orders of magnitude. These pre-equilibrated crystals are then soaked in similar solutions containing the desired substrate and chemical components that render the physicochemical environment conducive to catalysis followed by flash-freezing, or trapping, in liquid nitrogen of individual crystals at various time-points during reaction progression. Each crystal is then diffracted and the representative structure solved by molecular replacement using the open, intermediate-trapped or closed conformation structures and a complete reaction profile is built in this manner.

1.6 Specific Aims of this Work

This work aims to investigate several questions pertaining to the structure and function of A-family DNA polymerases using a structural biology approach. In the process of doing so, a refinement of various time-resolved techniques will be investigated to address the feasibility of using X-ray crystallography in the study of structural and biochemical aspects of enzymology.

Specifically, this work will investigate the mechanism of catalysis of A-family polymerases by dissecting the structural and mechanistic changes that accompany two of the steps affecting fidelity in these polymerases: that of nucleotide binding, and the chemistry step in phosphodiester bond formation between an incoming nucleotide and the primer 3'-end. This is accomplished by obtaining X-ray crystal structures of various binary and ternary complexes that will represent "snapshots" of the catalytic reaction in its entirety. In doing so, we will also be able to examine the contribution of the fingers domain O-helix and the catalytic residues of motifs A and C to fidelity and, ultimately, how pyrophosphate release contribution, if any, affects O-helix conformational changes.

These questions will be addressed using time-resolved X-ray crystallography of the Klenow fragment of *Thermus aquaticus* DNA polymerase (KlenTaq). Given the parameters affecting the use of time-resolved X-ray crystallography, KlenTaq appears to be an ideal candidate in that its rate of catalysis at room temperature *in crystallo* is slow enough to allow direct observation, and crystallographic "snapshots," of the interactions mediated by residues in its active site with both double-stranded DNA and incoming nucleotide. For the same reasons, direct observation of the contributions of the O-helix to fidelity will be possible due to the large conformational changes in the fingers subdomain that follow nucleotide binding but precede the chemistry step. Observation at atomic level resolution of phosphodiester bond formation in conjunction with the movement of, and interactions with, the O-helix and motifs A and C residues will likely answer the question of whether fidelity is due to an "induced fit" or "rate-limiting transition state" for this family of DNA polymerases, and will shed light to the possibility that said residues mediate proton transfer to the pyrophosphate leaving group during nucleotide incorporation.

Following the same scheme, direct observation of events following phosphoryl transfer will certainly provide snapshots of any microstates, or rapid conformational changes that precede pyrophosphate release or O-helix opening, thereby providing direct evidence of whether pyrophosphate release prompts O-helix opening, or whether the O-helix opens to facilitate pyrophosphate release, and how either scenario affects, or is affected by, translocation of the enzyme along duplex DNA. The answer to this question will prompt further investigation into whether O-helix opening is an energy driven process and, potentially, allow observation of any translocation events that occur for subsequent incoming nucleotides to be incorporated.

The significance of this work is paramount to a mechanistic and structural understanding of the contribution of DNA polymerases to human health and disease. Mammalian DNA polymerases have been implicated in various neoplastic and neurodegenerative disorders (92). Mutations in the genes encoding such polymerases result in the production of aberrant forms of these enzymes that cannot sustain functional fidelity relative to their naturally occurring, wild-type counterparts, and rational drug-design has not yet yielded chemotherapeutic agents that specifically target only malfunctioning polymerases, although not for lack of interest by researchers in the field, leading to increased cytotoxicity on wild-type enzymes that results in further genomic instability. Similarly, antiviral agents directed towards viral polymerases in infections are not entirely specific to said enzymes and act on some mammalian polymerases, which leads to phenotypic side-effects similar to those of chemotherapeutic agents (93). This work will contribute a great deal to structure-based drug design that can exploit mechanistic and structural differences between the various DNA polymerases within and across different families to increase drug specificity and reduce cytotoxicity.

References

1. Gill, J., Wang, J., Millar, D. (2011). DNA polymerase activity at the single-molecule level. *Biochem. Soc. Trans.* **39**, 595-599.

2. Rossi, M., Purohit, V., Brandt, P., Bambara, R. (2006). Reconstituted Okazaki Fragment Processing Indicates Two Pathways of Primer Removal. *J. Biol. Chem.* **281**, 26051-26061.

3. Dianov, G., Lindahl, T. (1994). Reconstitution of the DNA base excision--repair pathway. *Curr. Biol.* **4**, 1069-1076.

4. Cooper, P., Hanawalt, P. (1972). Role of DNA polymerase I and the rec system in excision-repair in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **69**, 1156–1160.

5. Hubscher, U., Maga, G., Spadari, S. (2002). Eukaryotic DNA polymerases. *Annu. Rev. Biochem.* **71**, 133-163.

6. Joyce, C., Steitz, T. (1994). Function and structure relationships in DNA polymerases. *Annu. Rev. Biochem.* **63**, 777-822.

7. Steitz, T. (1999). DNA Polymerases: Structural Diversity and Common Mechanisms. *J. Biol. Chem.* **274**, 17395-17398.

8. Doublie, S., Sawaya, M., Ellenberger, T. (1999). An open and closed case for all polymerases. *Structure* **7**, R31-35.

9. Kunkel, T., Bebenek, K. (2000). DNA replication fidelity. *Annu. Rev. Biochem.* **69**, 497-529.

10. Shevelev, I., Hubscher, U. (2002). The 3'-5' exonucleases. *Nat. Rev. Mol. Cell. Biol.* **3**, 364-376.

11. Bessman, MJ., Kornberg, A., Lehman, IR., Simms, ES. (1956). Enzymatic synthesis of deoxyribonucleic acid. *Biochim. Biophys. Acta*. **21**, 197-198.

12. Patel, PH., Suzuki, M., Adman, E., Shinkai, A., Loeb, LA. (2001). Prokaryotic DNA polymerase I: evolution, structure, and "base flipping" mechanism for nucleotide selection. *J. Mol. Biol.* **308**, 823-837.

13. Joyce, CM., Grindley, ND. (1984). Method for determining whether a gene of *Escherichia coli* is essential: application to the polA gene. *J. Bacteriol.* **158**, 636-643.

14. Bonner, CA., Randall, SK., Rayssiguier, C., Radman, M., *et al.* (1988). Purification and Characterization of an Inducible *Escherichia coli* DNA Polymerase Capable of Insertion and Bypass at Abasic Lesions in DNA. *J. Biol. Chem.* **263**, 18946-18952.

15. Friedberg, EC., Siede, W., Wood, RD., Schultz, RA., Ellenberger, T. (2006). *DNA Repair and Mutagenesis, 2nd Edition*. Washington, ASM Press.

16. Sommer, S., Boudsocq, F., Devoret, R., Bailone, A. (1998). Specific RecA amino acid changes affect RecA–UmuD′C interaction. *Mol. Microbiol.* **28**, 281-291.

17. McHenry, CS. (1988). DNA polymerase III holoenzyme of *Escherichia coli*. *Annu. Rev. Biochem*. **57**, 519-550.

18. Marians, KJ. (1992). Prokaryotic DNA replication. *Annu. Rev. Biochem*. **61**, 673-719.

19. Cann, IK., *et al.* (1998). A heterodimeric DNA polymerase: evidence that members of Euryarchaeota possess a distinct DNA polymerase. *Proc. Natl. Acad. Sci. USA* **95**, 14250-14255.

20. Makiniemi, M., *et al.* (1999). A novel family of DNA-polymerase-associated B subunits. *J. Trends Biochem. Sci.* 24, 14-16.

21. Hubscher, U., Spadari, S., Villani, G., Maga, G. (2010). *DNA Polymerases. Discovery, Characterization and Functions in Cellular DNA Transactions.* World Scientific Publishing.

22. Blasius, M. Shevelev, I., Jolivet, E., Sommer, S., Hubscher, U. (2006). DNA polymerase X from *Deinococcus radiodurans* possesses a structure-modulated 3′→5′ Exonuclease activity involved in radioresistance. *Mol. Microbiol.* **60**, 165-176.

23. Ramadan, K., Shevelev, I., Hubscher, U. (2004). The DNA-polymerase-X family: controllers of DNA quality? *Nat. Rev. Mol. Cell Biol.* **5**, 1038-1043.

24. Moon, AF., Garcia-Diaz, M., Batra, VK., Beard, WA., Bebenek, K. *et al*. (2007). The X Family Portrait: Structural Insights into Biological Functions of X Family Polymerases. *DNA Repair (Amst).* **6**, 1709-1725.

25. Lee, JW., Blanco, L., Zhou, T., Garcia-Diaz, M., Bebenek, K., *et al.* (2004). Implication of DNA Polymerase λ in Alignment-based Gap Filling for Nonhomologous DNA End Joining in Human Nuclear Extracts. *J. Biol. Chem.* **279**, 805-811.

26. Garcia-Diaz, M., Bebenek, K., Gao, G., Pederson, LC., London, RE., Kunkel, TA. (2005). Structure-function studies of DNA polymerase lambda. *DNA Repair (Amst)* **4**, 1358-1367.

27. Bork, P. *et al*. (1997). A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. *FASEB J.* **11**, 68–76.

28. Uchiyama, Y., Takeuchi, R., Kodera, H., Sakaguchi, K. (2009). Distribution and roles of X-family DNA polymerases in eukaryotes. *Biochimie.* **91**, 165-170.

29. Friedberg, E., Walker, G., Siede, W. (1995). *DNA Repair and Mutagenesis. American Soc. for Microbiology Press*.

30. Fuchs, RP., Fujii, S., Wagner, J. (2004). Properties and functions of *Escherichia coli*: Pol IV and Pol V. *Adv. Protein. Chem.* **69**, 229-264.

31. Tang, M., Pham, P., Shen, X., Taylor, JS., O'Donnell, M., *et al*. (2000). Roles of *E. coli* DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis. *Nature* **404**, 1014-1018.

32. Zhou, BL., Pata, JD., Steitz, TA. (2001). Crystal Structure of a DinB Lesion Bypass DNA Polymerase Catalytic Fragment Reveals a Classic Polymerase Catalytic Domain. *Mol. Cell* **8**, 427-437.

33. Ling, H., Boudsocq, F., Woodgate, R., Yang, W. (2001). Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell* **107**, 91-102.

34. Waters, LS., Minesinger, BK., Wiltrout, ME., D'Souza, S., Woodruff, RV., Walker, GC. (2009). Eukaryotic Translesion Polymerases and Their Roles and Regulation in DNA Damage Tolerance. *Microbiol. Mol. Biol. Rev.* **73** (1): 134–54.

35. Baltimore, D. (1970). Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of RNA Tumour Viruses. *Nature* **226**, 1209-1211.

36. Gavory, G., Farrow, M., Balasubramanian, S. (2002). Minimum length requirement of the alignment domain of human telomerase RNA to sustain catalytic activity *in vitro*. *Nucleic Acids Res.* **30**, 4470–80.

37. Sarafianos, SG., Marchand, B., Das, K., Himmel, DM., Parniak, MA., Hughes, SH., Arnold, E. (2009). Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J. Mol. Biol.* **385**, 693-713.

38. Klarmann, GJ., Hawkins, ME., Le Grice, SF. (2002). Uncovering the complexities of retroviral ribonuclease H reveals its potential as a therapeutic target. *AIDS Rev* **4**, 183–94.

39. Schultz, SJ., Champoux, JJ. (2008). RNase H activity: structure, specificity, and function in reverse transcription. *Virus Res*. **134**, 86–103.

40. Kohlstaedt, LA., Wang, J., Friedman, JM., Rice, PA., Steitz, TA. (1992). Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**, 1783–1790.

41. Cohen, S., Graham, M., Lovrecz, G., Bache, N., Robinson, P., Reddel, R. (2007). Protein composition of catalytically active human telomerase from immortal cells. *Science* **315**, 1850–3.

42. Blackburn, EH. (2005). Telomeres and telomerase: their mechanisms of action and the effects of altering their functions. *FEBS Lett.* **579**, 859–62.

43. Steitz, TA. (1999). DNA Polymerases: Structural Diversity and Common Mechanisms. *J. Biol. Chem*. **274**, 17395-17398.

44. Johnson, KA. (2010). The kinetic and chemical mechanism of high-fidelity DNA polymerases. *Biochem. Biophys. Acta.* **1804**, 1041–1048.

45. Brautigam, C., Steitz, T. (1998). Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* **8**, 54-63.

46. Lee, I., Berdis, A. (2010). Non-natural nucleotides as probes for the mechanism and fidelity of DNA polymerases. *Biochem. Biophys. Acta.* **1804**, 1064-1080.

47. Kool, ET. (2001). Hydrogen bonding, base stacking, and steric effects in DNA replication. *Annu. Rev. Biophys. Biomol. Struct*. **30**, 1-22.

48. Echols, H., Goodman, MF. (1991). Fidelity Mechanisms in DNA replication. *Annu. Rev. Biochem.* **60**, 477-511.

49. Morales, JC., Kool, ET. (2000). Varied molecular interactions at the active sites of several DNA polymerases: nonpolar nucleoside isosteres as probes. *J. Am. Chem. Soc.* **122**, 1001-1007.

50. Sloane, DL., Goodman, MF., Echols, H. (1988). The fidelity of base selection by the polymerase subunit of DNA polymerase III holoenzyme. *Nucleic Acids Res.* **16**, 6465-6475.

51. Sarafianos, SG., *et al*. (1999). Touching the heart of HIV-1 drug resistance: the fingers close down on the dNTP at the polymerase active site. *J. Biol. Chem.* **6**, R137-146.

52. Johnson, KA. (1993). Conformational coupling in DNA polymerase fidelity. *Annu. Rev. Biochem.* **62**, 685-713.

53. Aboul-Ela, F., *et al.* (1985). Base-base mismatches. Thermodynamics of double helix formation for dCA3XA3G + dCT3YT3G (X, Y = A,C,G,T). *Nucleic Acids Res.* **13**, 4811-4824.

54. Loeb, LA., Kunkel, TA. (1982). Fidelity of DNA synthesis. *Annu. Rev. Biochem.* **51**, 429-457.

55. Petersheim, M., Turnr, DH. (1983). Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, CCGGUp, and ACCGGUp. *Biochemistry* **22**, 256-263.

56. Sint, H., Kool, E. (2006). Remarkable sensitivity to DNA base shape in the DNA polymerase active site. *Angew. Chem. Int. Ed. Engl.* **45**, 1974-1979.

57. Obeid, S., Baccaro, A., Welte, W., Diederichs, K., Marx, A. (2010). Structural basis for the synthesis of nucleobase modified DNA by *Thermus aquaticus* DNA polymerase. *Proc. Natl. Acad. Sci. USA* **107**, 21327-21331.

58. Bear, WA., Osheroff, WP., Prasad, R., Sawaya, MR., *et al*. (1996). Enzyme-DNA interactions required for efficient nucleotide incorporation and discrimination in human DNA polymerase beta. *J. Biol. Chem.* **271**, 12141-12144.

59. Kiefer, JR., Mao, C., Braman, JC., Beese, LS. (1998). Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature* **391**, 304-307.

60. Eom, SH., Wang, J., Steitz, TA. (1996). Structure of Taq polymerase with DNA at the polymerase active site. *Nature* **382**, 278-281.

61. Bell, JB., Eckert, KA., Joyce, CM., Kunkel, TA. (1997). Base miscoding and strand misalignment errors by mutator Klenow polymerases with amino acid substitutions at tyrosine 766 in the O helix of the fingers subdomain. *J. Biol. Chem.* **272**, 7345-7351.

62. Huang, H., Chopra, R., Verdine, GL., Harrison, SC. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **282**, 1669-1675.

63. Wong, I., Patel, SS., Johnson, KA. (1991). An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics. *Biochemistry* **30**, 526-537.

64. Kati, WM., Johnson, KA., Jerva, LF., Anderson, KS. (1992). Mechanism and fidelity of HIV reverse transcriptase. *J. Biol. Chem.* **267**, 25988-25997.

65. Eger, BT., *et al*. (1991). Mechanism of DNA replication fidelity for three mutants of DNA polymerase I: Klenow fragment KF(exo+), KF(polA5), and KF(exo-). *Biochemistry* **30**, 1441-1448.

66. Kuchta, RD., Mizrahi, V., Benkovic, PA., Johnson, KA., Benkovic, SJ. (1987). Kinetic mechanism of DNA polymerase I (Klenow). *Biochemistry* **26**, 8410-8417.

67. Arora, K., Beard, W., Wilson, S., Schlick, T. (2005). Mismatch-induced conformational distortions in polymerase beta support an induced-fit mechanism for fidelity. *Biochemistry* **44**, 13328-13341.

68. Showalter, AK., Tsai, MD. (2002). A reexamination of the nucleotide incorporation fidelity of DNA polymerases. *Biochemistry* **41**, 10571-10576.

69. Sampoli Benitez, BA., Arora, K., Schlick, T. (2006). In silico studies of the African swine fever virus DNA polymerase X support an induced-fit mechanism. *Biophys. J.* **90**, 42-56.

70. Joyce, CM., Potapova, O., DeLucia, AM., Huang, X., Basu, VP., Grindley, N. (2008). Fingers-Closing and Other Rapid Conformational Changes in DNA Polymerase I (Klenow Fragment) and Their Role in Nucleotide Selectivity. *Biochemistry* **47**, 6103-6116.

71. Sawaya, MR., Prasad, R., Wilson, SH., Kraut, J., Pelletier, H. (1997). Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: evidence for an induced fit mechanism. *Biochemistry* **36**, 11205-11215.

72. Steitz, TA. (1998). A mechanism for all polymerases. *Nature* **391**, 231-232.

73. Yang, L., Arora, K., Beard, W., Wilson, S., Schlick, T. (2004). Critical Role of Magnesium Ions in DNA Polymerase β's Closing and Active Site Assembly. *J. Am. Chem. Soc.* **126**, 8441-8453.

74. Doublie, S., Tabor, S., Long, AM., Richardson, CC., Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2-Å resolution. *Nature* **391**, 251-258.

75. Ollis, DL., Brick, P., Hamlin, R., Xuong, NG., Steitz, TA. (1985). Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature* **313**, 762-766.

76. Li, Y., Waksman, G. (2001). Crystal structures of a ddATP-, ddTTP-, ddCTP-, and ddGTP-trapped ternary complex of Klentaq1: Insights into nucleotide incorporation and selectivity. *Protein Sci.* **10**, 1225-1233.

77. Li, Y., Korolev, S., Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17**, 7514-7525.

78. Rothwell, PJ., Allen, WJ., Sisamakis, E., Kalinin, S., Felekyan, S., Widengren, J., Waksman, G., Seidel, CA. (2013). dNTP-dependent conformational transitions in the fingers subdomain of Klentaq1 DNA polymerase: insights into the role of the "nucleotide-binding" state. *J. Biol. Chem.* **288**, 13575-13591.

79. Makinen, MW., Fink, AL. (1977). Reactivity and Cryoenzymology of Enzymes in the Crystalline State. *Annu. Rev. Biophys. Bioeng.* 6, 301-343.

80. Matsubara, I., Yagi, N. (1978). A time-resolved X-ray diffraction study of muscle during twitch. *J. Physiol.* **278**, 297-307.

81. Moffat, K., Szebenyi, D., Bilderback, DH. (1984). X-ray Laue Diffraction from Protein Crystals. *Science* **223**, 1423-1425.

82. Blundell, TL., Johnson, LN. (1976). *Protein Crystallography*, *Academic Press*.

83. Moffat, K., Helliwell, J. (1989). The Laue Method and its Use in Time-Resolved Crystallography. *Topics. Curr. Chem.* **151**, 61-74.

84. Wyckoff, HW., Doscher, M., Tsernoglou, D., Inagami, T., Johnson, LN., Hardman, K., Allewell, NM., Kelley, DM., Richards, FM. (1967). Design of a diffractometer and flow cell system for X-ray analysis of crystalline proteins with applications to the crystal chemistry of ribonuclease-S. *J. Mol. Biol.* **27**, 563-578.

85. Westbrook, EM., Sigler, PB. (1984). Enzymatic function in crystals of Δ^5-3-ketosteroid isomerase. Catalytic activity and binding of competitive inhibitors. *J. Biol. Chem.* **259**, 9090-9095.

86. Freudenthal, BD., Beard, WA., Shock, D., Wilson, SH. (2013). Observing a DNA Polymerase Choose Right from Wrong. *Cell* **154**, 157-168.

87. Tilton, RF. (1992). Effects of Temperature on Protein Structure and Dynamics: X-ray Crystallographic Studies of the Protein Ribonuclease-A at Nine Different Temperatures from 98 to 320 K. *Biochemistry* **31**, 2469-2481.

88. Schlichting, I., Petsko, G., *et al*. (1990). Time-resolved X-ray crystallographic study of the conformational change in Ha-Ras p21 protein on GTP hydrolysis. *Nature* **345**, 309-315.

89. Parkhurst, LJ., Gibson, QH. (1967). The Reaction of Carbon Monoxide with Horse Hemoglobin in Solution, in Erythrocytes, and in Crystals. *J. Biol. Chem.* **242**, 5762-5770.

90. Moffat, K. (1989). Time-Resolved Macromolecular Crystallography. *Annu. Rev. Biophys. Biophys. Chem.* **18**, 309-332.

91. Gleghorn, M., Davydova, E., Basu, R., Rothman-Denes, L., Murakami, K. (2011). X-ray crystal structures elucidate the nucleotidyl transfer reaction of transcript initiation using two nucleotides. *Proc. Natl. Acad. Sci. USA* **108**, 3566-3571.

92. Loeb, L., Monnat, R. (2008). DNA polymerases and human disease. *Nat. Rev. Genet.* **9**, 594-604.

93. Berdis, A. (2008). DNA polymerases as therapeutic targets. *Biochemistry* **47**, 8253-8260.

Chapter 2

Structural and Mechanistic Basis of Nucleotide Incorporation during DNA Replication by a high-fidelity, A-family Replicative DNA Polymerase

2.1 Abstract.....41

2.2 Introduction.....42

2.3 Results.....45

2.3.1 *in crystallo* Single-turnover Kinetics of A-family KlenTaq DNAP.....45

2.3.2 Nucleotide Binding is Spatiotemporally Separated from Metal-ion Entry...48

2.3.3 Binding and Catalytic Metal Entry is Sequential and Concerted.....52

2.3.4 Pyrophosphate Release Requires O-helix Reversion to the Open Conformation.....52

2.4 Discussion.....55

2.4.1 Spatiotemporal Separation of Substrate and Metals – a Possible Mechanism for O-helix Closure, and Implications for Fidelity.....55

2.4.2 3'-Oxyanion Activation – A Catalytic Water, Captured Transition State, and Consequences to Proton Transfer.....66

2.4.3 Coupled Dynamics in Translocation.....76

2.4.4 Two- and Three-Metal Ion Catalysis: Evolutionary Imperative or Mechanistic Contingency.....81

2.4.4.1 Function-centric view to variations in metal-ion catalysis.....83

2.4.4.2 Metal-centric view to variations in metal-ion catalysis.....84

2.5 Materials and Methods.....88

References.....95

---

* Work described in this chapter will be submitted for publication under the following:

AlMishwat, MA., Ferry, JG., Murakami, KS. (2019). Structural and Mechanistic Basis of Fidelity in Nucleotide Incorporation during DNA Replication by an A-family Replicative DNA Polymerase.

2.1 Abstract

Fidelity in DNA replication facilitates the faithful transmission of genetic material through successive generations, an evolutionary imperative for the continuation of any species. The process of DNA replication has been extensively characterized, however, the exact mechanism by which a high fidelity DNA polymerase preferentially incorporates a correct nucleotide, and differentially excludes an incorrect nucleotide during DNA replication is not fully understood. Using time-resolved X-ray crystallography as a platform to investigate high-fidelity nucleotide incorporation we delineate the structural and mechanistic determinants of intermediate steps during DNA replication by an A-family replicative DNA polymerase, and show that formation of a catalytically-competent ternary complex poised for chemistry is a step-wise process wherein nucleotide-binding and metal-ligation are decoupled. This spatiotemporal separation of substrate and metals serves as the first checkpoint in high-fidelity nucleotide incorporation, as metal entry subsequent to nucleotide binding is contingent upon an optimal geometry of said nucleotide in the active site.

2.2 Introduction

DNA replication is ubiquitous to all organisms with DNA genomes. The faithful transmission of genetic material through successive generations is an evolutionary imperative for the continuation of any species. To accomplish this, organisms from different domains of life must rely on specialized DNA polymerases (DNAPs) during DNA replication that can match the speed of cell division for a particular cell type or organism whilst maintaining a high degree of fidelity. Errors made during this process can lead to genomic instability, disruption of metabolic and regulatory pathways, and the prevalence of disease states.

DNA polymerases are grouped into seven different families (A-D, X, Y, and RT) based on sequence and structural similarities (1) and are classified as replicative (A-D) or repair (X and Y) enzymes with only a few exceptions that perform both tasks (2,3).

A great deal of insight has been obtained from available crystal structures of binary complexes of DNAPs and DNA, and ternary complexes containing incoming nucleotides or their analogues. Such structures illustrate a clear structure-function relationship in DNAPs, and common structural elements conserved across all known DNAPs exist (4,5). The overall structure of the polymerase domain of a DNAP resembles a human right hand that consists of 3 subdomains: the palm, fingers, and thumb. The fingers subdomain interacts with an incoming dNTP via conserved Lys and Arg residues and the dNTP α- and βγ-phosphates, respectively, and Phe or Tyr and the dNTP deoxyribose sugar. The fingers subdomain also interacts with single-stranded template DNA via a conserved Tyr residue that, in the open binary complex stacks against the templating nucleobase and "flips" out the adjacent 5'-nucleobase and, in the closed ternary complex forms a closure around the substrates in the active site to facilitate Watson-Crick base-pairing (6,7). The thumb subdomain binds to and stabilizes double-stranded DNA. The palm subdomain contains catalytic Asp residues that ligate two divalent metal ions (Mg^{2+}) required for phosphoryl transfer. Metal binding in the palm subdomain was shown to result in an inward rotation of the catalytic Asp residues. Structures of 2',3'-dideoxynucleotide-terminated ternary complexes and ternary complexes with non-hydrolyzable dNTP analogs provided the framework for active site

architecture and reveal the coordination complex resulting from ligation of the two metal ions to the catalytic Asp residues and the triphosphate moiety of the incoming nucleotide which places both metals on either end of the scissile phosphate facilitating proper alignment of the Pα to the primer O3' (catalytic metal, Mg^A) and stabilization of the pyrophosphate (Pβγ, PP_i) leaving group (binding metal, Mg^B).

Recent studies on DNA polymerases β and η (8,9), X- and Y-family repair DNAPs, respectively, and bacteriophage N4 RNA polymerase (10) (RNAP) utilizing time-resolved X-ray crystallography have made significant contributions to our understanding of the mechanisms of DNA repair and transcription initiation. Time-resolved X-ray crystallography has facilitated, at atomic resolution, unprecedented mechanistic details in the structural enzymology of DNA and RNA polymerases that would otherwise not be possible using any other technique. Studies on DNAP β and η have facilitated kinetic investigation in a steady-state fashion that precludes observation of pre-steady-state structural and mechanistic information that would result from nucleotide binding to an open, catalytically-competent binary complex under physiologically-relevant conditions.

However, there are distinct structural and mechanistic differences between replicative and repair DNAPs (11) that warrant investigation of high-fidelity nucleotide incorporation under physiological conditions that facilitate monitoring the nucleotidyl transfer reaction *ab initio* in a catalytically-competent binary complex of a replicative DNAP. Crystal structures of repair enzymes have revealed a third catalytically-important Mg^{2+} metal ion appearing near the α-phosphate of the incoming nucleotide prior to phosphoryl transfer and remaining in the vicinity even after pyrophosphate hydrolysis where it is observed to ligate the free α-phosphate non-bridging oxygen and the αβ-bridging oxygen now part of the PPi leaving group. It is hypothesized that this interaction stabilizes the product complex by neutralizing the potential negative charge buildup that would result from PP_i release, thereby preventing DNA translocation. Another finding in these studies is a transient water molecule believed to play a role in facilitating activation of the primer O3' by accepting a proton resulting from the deprotonation of the O3' prior to nucleophilic attack onto the Pα of the incoming nucleotide.

Unlike their replicative counterparts, repair DNAPs from the X- and Y-family, due to the highly specialized roles they play in base-excision and translesion DNA synthesis repair, respectively, are error-prone and have significantly lower fidelity. The X-family DNA polymerase β (Polβ) contains an N-terminal 5'-deoxyribose 5'-phosphate lyase (dRP-lyase) domain that contains a Helix-hairpin-Helix (HhH) motif which interacts with DNA downstream of a gap (abasic site, etc.) and aids in the removal of 5'-terminal abasic sugars from damaged DNA (12). As a result, the polymerase makes more extensive contacts with the phosphate backbone of the duplex DNA and minimal contacts through its N-subdomain with the DNA minor groove. Y-family DNAPs exclusively carry out translesion synthesis to bypass damaged DNA and attempt to salvage DNA replication. Because of the nature of this repair route, their low fidelity provides for flexibility in bypassing lesions. Crystal structures of archaeal DNA polymerases Dbh and Dpo4 (13) and their mammalian counterpart Polη (14) have revealed that while the palm subdomain was nearly identical to its counterparts in other families, the fingers subdomain appears in the closed conformation regardless of the presence or absence of any substrate to which it makes minimal, non-specific contacts (15).

It is thus necessary to investigate nucleotide incorporation by a high-fidelity replicative DNAP to characterize the mechanism of nucleotide discrimination, the location of initial binding of incoming nucleotides and how this binding triggers conformational changes from open to closed state in the enzyme leading to active-site assembly and nucleotidyl transfer.

Here, using time-resolved X-ray crystallography, we report studies on the structural and mechanistic basis of nucleotide incorporation during DNA replication by a high-fidelity, replicative A-family DNA polymerase from *Thermus aquaticus*.

2.3 Results

2.3.1 *in crystallo* Single-turnover Kinetics of A-family KlenTaq DNAP

Structures of A-family DNAPs solved thus far are those of the canonical *Escherichia coli* DNAP I (Klenow Fragment), Bacteriophage T7, *Thermus aquaticus* DNAP I (KlenTaq), *Bacillus stearothermophilus* DNAP I and human mitochondrial DNAPγ. In order to investigate high-fidelity nucleotide incorporation by A-family DNAPs, we evaluated representative structures deposited to the Protein Data Bank (PDB) and systematically refined selection of candidate DNAPs by exclusion of mesophilic members that have been reported to be active at 20-25°C and those reported to be quite rapid catalytically *in crystallo*. Candidates with constrained crystal packing were also excluded from the studies presented. The *Bacillus* fragment (BF) of *B. stearothermophilus* DNAPI was excluded as previous studies have shown that, while it is capable of incorporating dNTP's within a crystal (16), the rate of incorporation *in crystallo* is quite rapid that no available catalytically competent structure is available with a complete pre-chemistry active site, i.e. a structure in which both binding and catalytic metals, incoming dNTP and primer O3' are readily visible in the resultant electron densities. Attempts at obtaining said intermediate always lacked the catalytic metal (Mg^{A}). Li *et al.* solved the crystal structure of the Klenow fragment of *Thermus aquaticus* (Klentaq1, KlenTaq) DNAP in complex with dideoxy-terminated duplex DNA and a 2',3'-dideoxyribonucleoside-5'-triphosphate (ddNTP), which corresponded to the closed ternary complex (17). By subsequently soaking the closed ternary complex crystals in mother liquor lacking the ddNTP they were able to generate the open binary complex, demonstrating a degree of flexibility in the fingers subdomain and, consequently, in the packing of complexes within the crystal that may be exploited for the purposes of these studies. However, two aspects initially precluded adopting this system and required further modifications to the platform to make it amenable to time-resolved studies. The inherent flexibility of the fingers subdomain of Klentaq1 in binary complexes prevents suitable crystal packing, making it necessary to devise a strategy that would maintain the enzyme in a stable state long enough to bring about favourable crystal contacts that would lead to diffraction-quality crystals. This strategy was compounded by the lack of

structures containing both 3'-hydroxyl groups on the primer end, and 5'-overhangs to serve as the next templating bases. Once the templating nucleobase is available for potential base-pairing with an incoming nucleotide, the O-helix continuously samples the active site, generating transient heterogeneous populations of open and closed binary and ternary complexes, respectively. Thus, it was necessary to prevent this constant flux between open and closed states in order to facilitate crystal packing (Fig. 2.1).

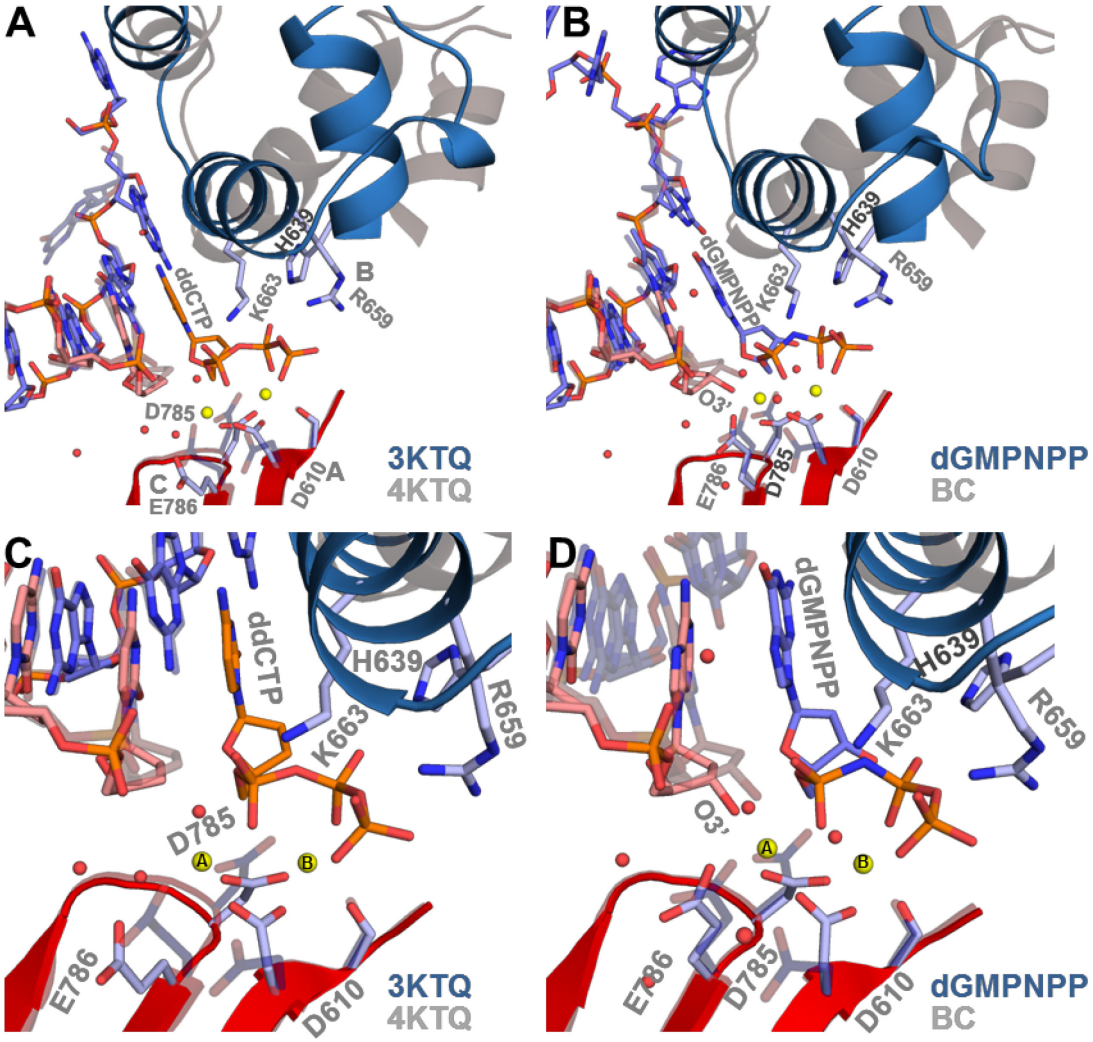


Figure 2.1 Superposition of the dideoxy-terminated and extension-ready binary and ternary complexes of KlenTaq DNA polymerase. In panel A and B the crystal structures of dideoxy-terminated (3KTQ, 4KTQ) and extension-ready (dGMPNPP, BC), respectively, binary and ternary complexes are superimposed to highlight the conformational change of the O-helices from open binary (BC, partially transparent) to closed ternary complexes upon (d)dNTP binding. Panels C and D provide a close-up view of the respective active sites to illustrate motif A and C changes consequent to (d)dNTP binding to the respective binary complexes. Motif C Glu786 shows little involvement in the absence of an O3' in dideoxy-terminated complexes and becomes engaged once said moiety becomes available for activation immediately prior to chemistry.

2.3.2 Nucleotide Binding is Spatiotemporally Separated from Metal-ion Entry

To obtain crystals of a catalytically-competent binary complex of KlenTaq in the open conformation, KlenTaq was crystallized in a closed ternary complex by introducing the α-β non-hydrolyzable dGTP analogue dGMPNPP as the next incoming nucleotide. This scheme trapped the polymerase in the closed conformation and produced crystals exhibiting atomic resolution X-ray diffraction. Subsequent soaking of the resultant ternary complex crystals in mother liquor lacking the non-hydrolyzable analogue permitted its diffusion from the active site which brought about reversion of the O-helix to the open conformation.

The ternary complex with dGMPNPP provided a static pre-chemistry intermediate structure that highlights all structural and chemical elements in their proper geometric and conformational spaces and coordination of both binding and catalytic metals. Here, relative to the open binary complex 4KTQ, the carboxylates of D610, D785 and the carbonyl oxygen of Y611 in motifs A and C are rotated inwards to ligate the catalytic and binding metals, respectively, which in turn also ligate the O3' of the primer end, Pα (Mg^A), and Pβγ (Mg^B) of dGMPNPP. In contrast to 4KTQ, Mg^B ligation to the dGMPNPP-ternary complex facilitates a 2.2Å rotation of D610 inward towards dGMPNPP during assembly of a complete active site.

While Li *et al.* reported that KlenTaq was active in their crystal system, as is evident from the conformational changes from open to closed state, and vice versa of the enzyme, it is not to be considered catalytically competent as the absence of an O3' at the primers 3'-end precludes any incorporation but only allows examination of events leading up to but excluding chemistry (Fig. 2.2). The hallmark of this structure, the presence of a O3' capable of participating in chemistry, has in fact shed light on an aspect that has long been a topic of debate regarding the involvement of solvent molecules, an evolutionarily conserved Glu residue in motif C, or both in deprotonation and activation of the O3' to generate the nucleophile for an in-line attack on Pα of the incoming dNTP, in which the conserved Glu residue would be the first proteon acceptor during acivation of the O3'. In 2',3'-dideoxy-terminated structures of KlenTaq, the carboxylate of E786 is seen at a distance of 3.8Å away from Mg^A, however, in the dGMPNPP-ternary structure,

the presence of O3', and solvent molecules in its' vicinity, brings the carboxylate of E786 approximately 1.5Å closer to Mg^A to result in a catalytically-relevant distance of 2.3Å. The direct involvement of the conserved Glu residue in motif C has been theorized and computationally predicted but never observed in crystal structures of A-family DNAPs.

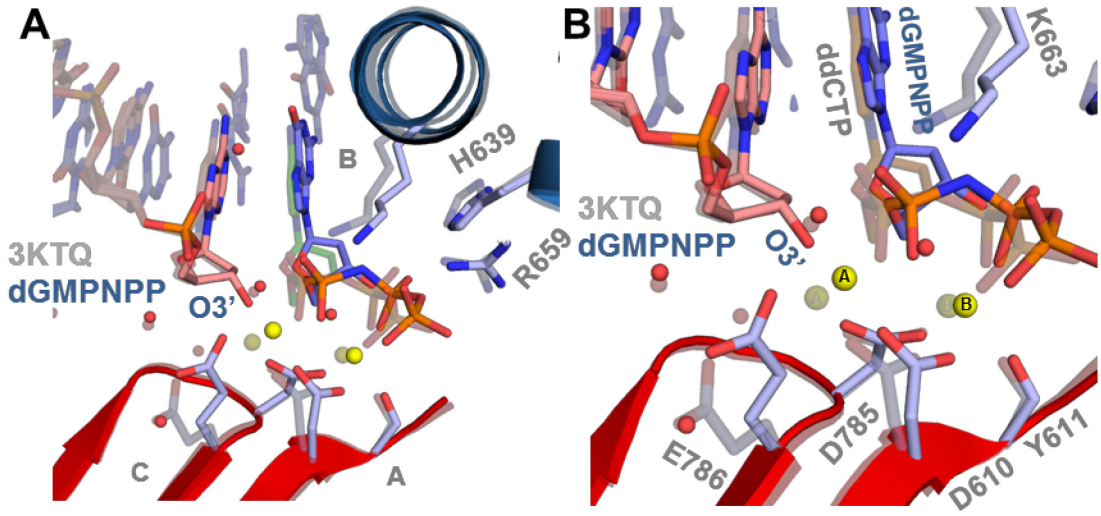


Figure 2.2 Superposition of the closed dideoxy-terminated and dGMPNPP-trapped ternary complexes. The superimposed structures illustrate both variation in the interatomic distance of the binding (Mg^B) and catalytic (Mg^A) metal-ions indicative of an inactive (3KTQ, partially transparent) or active (dGMPNPP) state of the enzyme with motif C Glu786 involvement in O3' activation.

The open, catalytically-competent binary complex was obtained after soaking out dGMPNPP but maintaining the concentration of Mg^{2+} ions. The binary complex (Fig. 2.3A) shows an O-helix in the open conformation, similar to 4KTQ, and no residual electron density for either dGMPNPP or metals. In contrast to 4KTQ, this resultant binary complex shows D610 moving 2.2Å farther away from the active site and E786 remaining in the position it occupied in the presence of Mg^A in the dGMPNPP-ternary complex (Fig. 2.4).

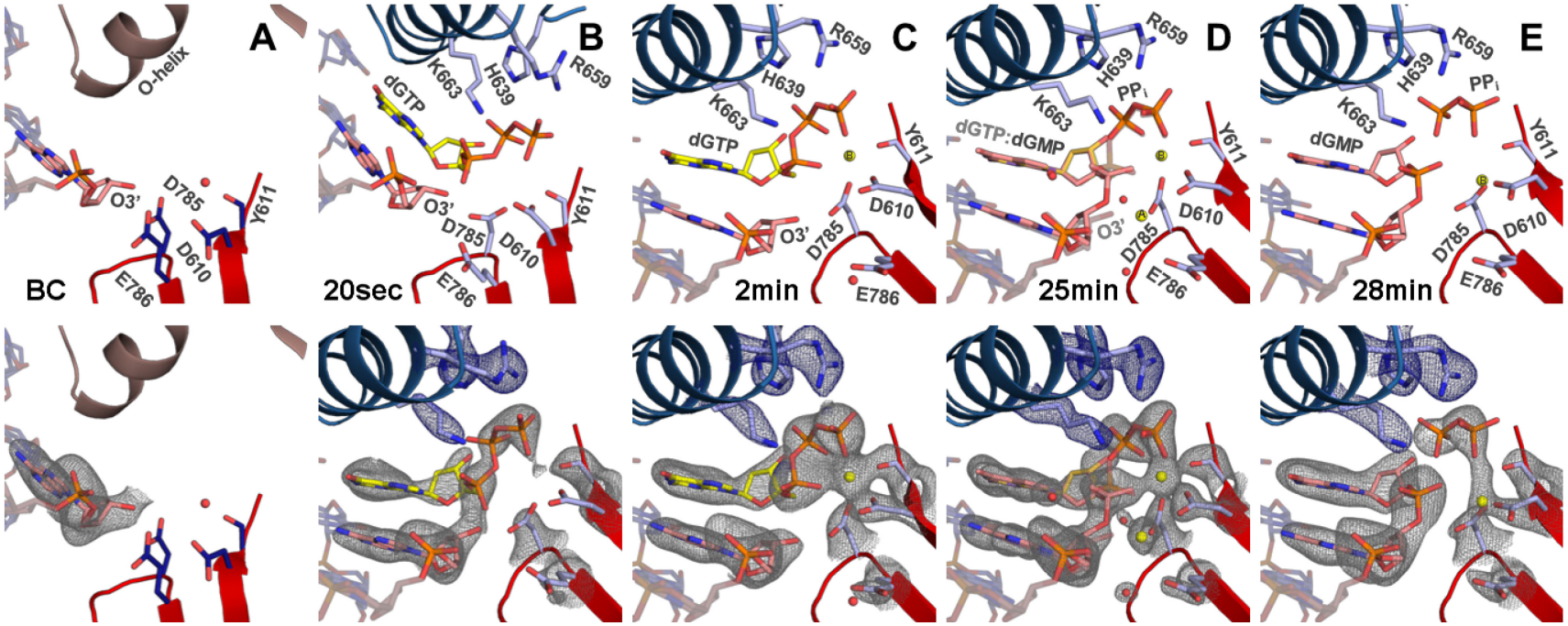


Figure 2.3 Time-course of events along the pathway for nucleotide incorporation by KlenTaq DNA polymerase. The top panel shows models of crystal structures obtained over time after introduction of the correct nucleotide to a catalytically-competent binary complex of KlenTaq DNAP. The bottom panel superimposes F_o-F_c omit maps contoured to 3σ (grey mesh) for dNTPs, metals, motif A and C residues, and $2F_o$-F_c maps (blue mesh) for motif B residues on respective structures for each complex. The binary complex (BC, panel A) shows the canonical open conformation of the O-helix in the absence of substrate. Upon dGTP binding (panel B) the O-helix undergoes conformational change to the closed complex in the absence of binding and catalytic metals. Shortly thereafter the binding metal (Mg^B) coordinates the triphosphate moiety (panel C) prior to catalytic metal entry (Mg^A). Once Mg^A enters the active site (panel D) and coordinates the remaining moieties, O3' activation occurs to generate the nucleophile to attack Pα of the incoming nucleotide and product formation (phosphodiester bond) accumulates. The resultant PP_i leaving group remains ligated to Mg^B (panel E) to stabilize the product complex before ultimately diffusing out of the active site upon dissociation of Mg^B.

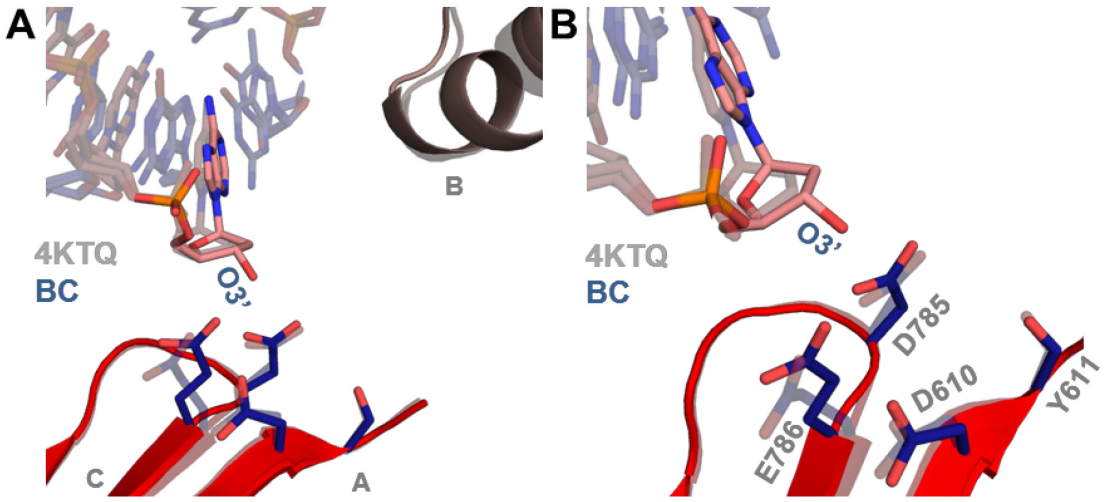


Figure 2.4 Superposition of the open dideoxy-terminated and catalytically-competent binary complexes. The presence of the O3' in the binary complex (BC), relative to 4KTQ, results in the deoxyribose sugar puckering and in motif C Glu786 in BC rotating inwards 1.0-1.5Å towards the active site presumably to stabilize the O3' along with interactions with motif C Asp785.

To investigate the structural intermediates along the catalytic pathway, crystals of the open binary complex were soaked in mother liquor containing 5mM dGTP and flash-frozen in liquid nitrogen to quench the reaction and their representative structures solved for each time point. During these time-course experiments, no significant structural changes occur prior to 20 seconds after soaking crystals in mother liquor containing dGTP. Structures solved between 20 seconds and 1 minute show a rapid conformational change of the O-helix from open to closed state, similar to 3KTQ, with a continuous electron density for the incoming nucleotide at the active site (Fig. 2.3B). In this closed conformation, relative to the open binary complex, Y671 is seen shifted outwards and no longer stacking against the templating nucleobase, with F667 stacking against the nucleobase of the incoming nucleotide and K663, H639 and R659 side-chains interacting with the Pα, Pβ and Pγ, respectively, of its' triphosphate moiety. However, neither binding (Mg^B) nor catalytic (Mg^A) metals were found ligated to the triphosphate, or carboxylates of D785 and D610, and no apparent dGTP coordination complex to mitigate any charge buildup. In fact, the side-chains of D785 and D610 remain rotated away from the triphosphate, similar to the binary complex, in an attempt to reduce the negative

charge buildup around the active site prior to metal ligation and formation of a Mg^{2+}-coordinated ternary complex. However, unlike the catalytically-competent binary complex, at 20 seconds after soaking in dGTP, the evolutionarily conserved E786 of motif C rotates 2.37Å away from the O3' as the Pα of the incoming nucleotide approaches the O3' during pre-chemistry.

2.3.3 Binding and Catalytic Metal Entry is Sequential and Concerted

Metal binding to the closed ternary complexes was found to be sequential and was accompanied by conformational changes in both motifs A and C, and in the triphosphate of the incoming nucleotide. The binding metal (Mg^B) was not observed in the electron density to be at or near the active site until 2.5-3min after dGTP was soaked into the open binary complex (Fig. 2.3C) wherein it ligated the triphosphate moiety to the carboxylates of D785 and D610 and the backbone carbonyl of Y611, thereby facilitating a 0.5Å inward rotation of D610 closer to dGTP. Consequently, this results in a geometry that can now accommodate stable ligation of the catalytic metal (Mg^A), which appears in the electron density of structures at and beyond 5.5-6min, to the carboxylate of D785, O3', and Pα of the incoming nucleotide resulting in formation of a Mg^{2+}-coordinated ternary complex.

2.3.4 Pyrophosphate Release Requires O-helix Reversion to the Open Conformation

At 25 minutes after soaking dGTP into a binary complex crystal, more than 60-70% of the resultant complex is that of the product (Fig. 2.3D), by comparison of relative occupancy and signal intensity in respective F_o-F_c omit maps for each structure. As the product complex accumulates and PP_i is hydrolyzed, Mg^A can no longer sustain its rigid coordination requirements as the newly formed phosphodiester bond precludes this from being possible and begins to evacuate the active site and is observed 0.95Å away from its optimal location for coordination chemistry (Fig. 2.5). As this occurs, E786 moves back to its original position in the open binary complex, and D785 is now 0.66Å away from its catalytic position. Pyrophosphate is now coordinated by $Mg^{B,}$ which is also coordinating the Pα non-bridging oxygen (O2A), the carboxylate of D610 and the carbonyl of Y611. Evolutionarily conserved O-helix residues K663 and R659 are also observed transiently

maintaining this complex. The interaction of O2A and the non-bridging oxygen of Pβ (O2B) in the newly hydrolyzed pyrophosphate are transiently stabilized by the conserved K663. The pyrophosphate is also concurrently stabilized by the imidazole ring of the evolutionarily conserved N-helix H639 (Pβ), and R659 (Pγ).

It is after 28 minutes subsequent to dGTP soaking into binary complex crystals that >90% of the predominant F_o-F_c density in the active site is that of the product complex wherein the density has now become discontinuous between pyrophosphate and the newly incorporated dGMP (Fig. 2.3E). In this state, the F_o-F_c density of the binding (Mg^B) metal is significantly diminished relative to the structure obtained at 25 minutes. As the binding (Mg^B) metal is released, the carboxylate of D610 is now rotated 0.6Å away from its canonical position observed in the post-catalytic complex at 25 minutes, and the pyrophosphate moves 0.45Å closer to H639. The inability of this complex to translocate, due to crystal packing constraints, occludes access to the O3' of the newly formed 3'-end which results in E786 remaining away from the active site in a conformation similar to that in the 2',3'-ddNMP-terminated ternary complex, 3KTQ.

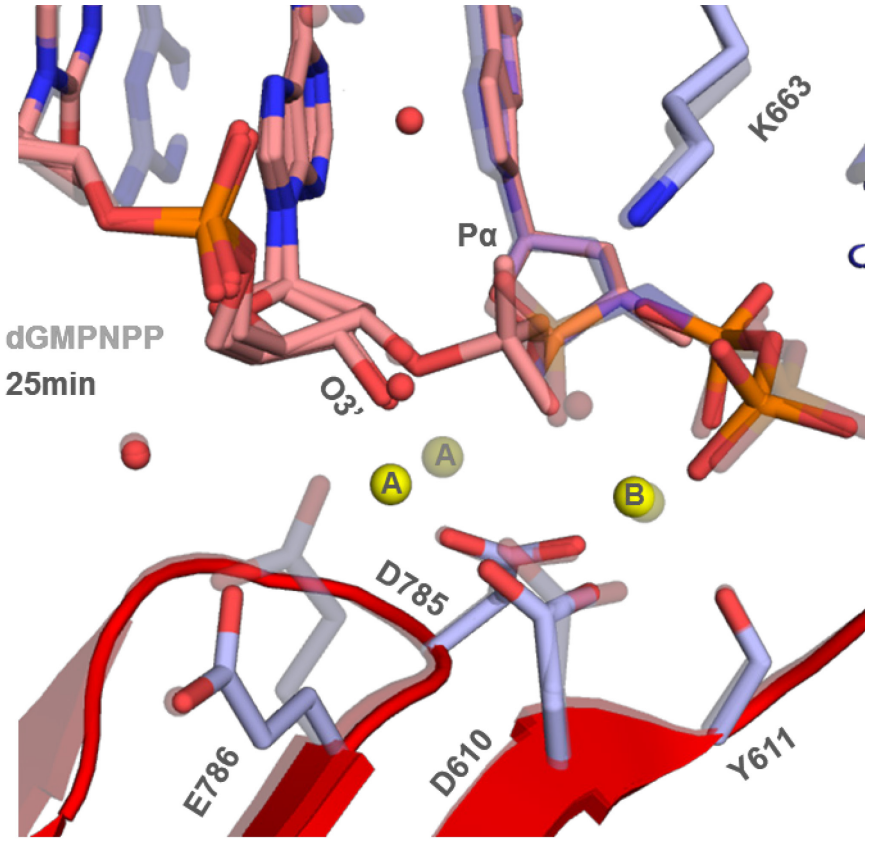


Figure 2.5 Superposition of the dGMPNPP-trapped and product complexes. The relative interatomic distance of Mg^{2+} and the distance of Mg^{A} to Pα, O3', and E786 are illustrated here to highlight the pre-chemistry (activation complex) distances, and distances as chemistry occurs. Glu786 in the dGMPNPP-trapped ternary complex (partially transparent) interacts with Mg^{A} and places it at an optimal distance of 3.5Å from Mg^{B}, the catalytic or active distance, to facilitate deprotonation and activation of the O3'. As chemistry occurs, highlighted by phosphodiester bond formation in the structure obtained at 25min, E786 is rotated 3.6Å away from Mg^{A}, destabilizing Mg^{A} and placing it at an inactive distance of 4.1Å away from Mg^{B} and 3Å away from Pα. Y611 is the carbonyl backbone of respective residue.

2.4 Discussion

in crystallo time-resolved X-ray crystallography has been successful in deducing structural and mechanistic determinants of nucleotide incorporation and its fidelity, or lack thereof, for X- and Y-family repair DNA polymerases (8,9). Attempts at deducing catalytic intermediates during nucleotide incorporation by A-family DNA polymerases, characterized as high-fidelity replicative and repair polymerases, were unsuccessful in *B. stearothermophilus* (Bst) DNAP as the rate of catalysis precluded resolution of the necessary catalytic intermediates (16). The work presented here attempts to bridge the gap in our collective understanding of the mechanistic variations between different DNA polymerase families by cultivating a platform utilizing the high-fidelity A-family DNAP from *Thermus aquaticus* to examine single-turnover kinetics *in crystallo* and extract structural and mechanistic information from captured reaction intermediates to expand the current, minimal kinetic pathway of nucleotide incorporation in this DNAP family.

2.4.1 Spatiotemporal Separation of Substrate and Metals – a Possible Mechanism for O-helix Closure, and Implications for Fidelity

Fidelity in nucleotide incorporation has been ascribed to the degree of discrimination in three distinct processes: ground-state binding of an incoming nucleotide, conformational changes subsequent to nucleotide binding, and chemistry (23). Divalent metal-ion choice, whether a consequence of polymerase preference or necessity due to availability, and concentration thereof, has been shown to influence fidelity during ground-state binding of nucleotides (24-29), manifested as metal ion-dependent variations in the equilibrium dissociation constants of nucleotides, both correct and incorrect, in B-family polymerases from the bacteriophages T4 and RB69 (25,30-32). These findings, however, were obtained from *in vitro* studies that examined the effect of divalent metal-ion choice and concentration on base selectivity in an environment where ground-state binding and metal-ion entry were coupled during reaction initiation (32) and, while valid, preclude the resolution and examination of the contributions of structural elements and metal-ion entry separately, a drawback that can be mitigated and fine-tuned in *in crystallo* studies that allow for resolution of structural features resulting

from ground-state binding and subsequent metal-ion entry and ligation illustrated in an expansion of the minimal kinetic pathway for correct nucleotide incorporation (Fig. 2.6).

It is widely held that (d)NTP's, stabilized after chemical or enzymatic synthesis in the form of cationic salts, are delivered to the polymerase active site with their complexed metals intact. However, this assumption is suspect in light of current hypotheses regarding the initial binding site of an incoming (d)NTP (18) and the findings that metal-ion entry is subsequent to and a consequence of formation of the closed ternary complex (Fig. 2.6). The evolutionarily conserved Lys and Arg residues of motif B in A-family DNA polymerases (Fig. 2.7), due to the nature of their interaction with an incoming nucleotide's triphosphate moiety, were proposed to constitute the initial binding site of an incoming nucleotide. This result was observed in crystal structures of the apo-form of KlenTaq DNAP when natural dNTP's were soaked into representative crystals (19). The apoenzyme appears in the open conformation with the dNTPs bound to the O-helix through interactions between the amines of K663, R659 and non-bridging oxygens of αγ- and γ-phosphates, respectively, of the bound nucleotide. In line with this reasoning, the charge stabilization that results from interactions between the NH_3-containing side-chains of Lys and Arg and the incoming nucleotides triphosphate would render the metal ions coordinated to such a triphosphate partially or completely unnecessary, or might result in a slight positive charge that may facilitate an alternative route to O-helix closure via attraction to motif A and C carboxylates in preparation for metal-ion entry. The O-helix in A-family polymerases is quite flexible and has been shown *in vitro* to constantly sample the active site (33), and will likely remain closed when all charge, hydrogen-bonding, and geometric constraints are met by delivery of a properly aligned dNTP to the active site. In fact, when an incoming nucleotide binds to a metal-free binary complex, the ground-state binding alone is sufficient to induce permanent O-helix conformational change, completely unaffected by the absence of metals (unpublished results/Chapter 3). Mechanistically, this is more favorable than delivery of metal ion-ligated dNTP's into the active site, since such a mechanistic scheme would commit the polymerase to chemistry regardless, albeit at varying catalytic rates, and deprive it of much of the intrinsic fidelity that it would otherwise enjoy during

ground-state nucleotide binding and subsequent O-helix conformational change to the closed ternary complex prior to metal-ion entry.

Sequential, concerted Mg^{2+} entry into the active site after the ternary complex undergoes conformational change affords A-family polymerases a higher degree of fidelity in base selectivity. This spatiotemporal separation provides the polymerase with an added checkpoint in selectivity by allowing it to present its bound nucleotide in slightly different conformations and varied distances upon O-helix closure, unencumbered by any ligated metals that may attempt to realign the bound dNTP. The resultant closed ternary complex would either allow the binding metal (Mg^B) to coordinate the triphosphate moiety of the incoming nucleotides to motif A and C carboxylates, in case of correct nucleotide binding, thereby facilitating proper substrate alignment for subsequent ligation of the catalytic metal (Mg^A), or limit proper geometry and coordination by obscuring the triphosphate in improperly aligned mismatched complexes that would preclude Mg^B coordination and ligation and prevent catalysis.

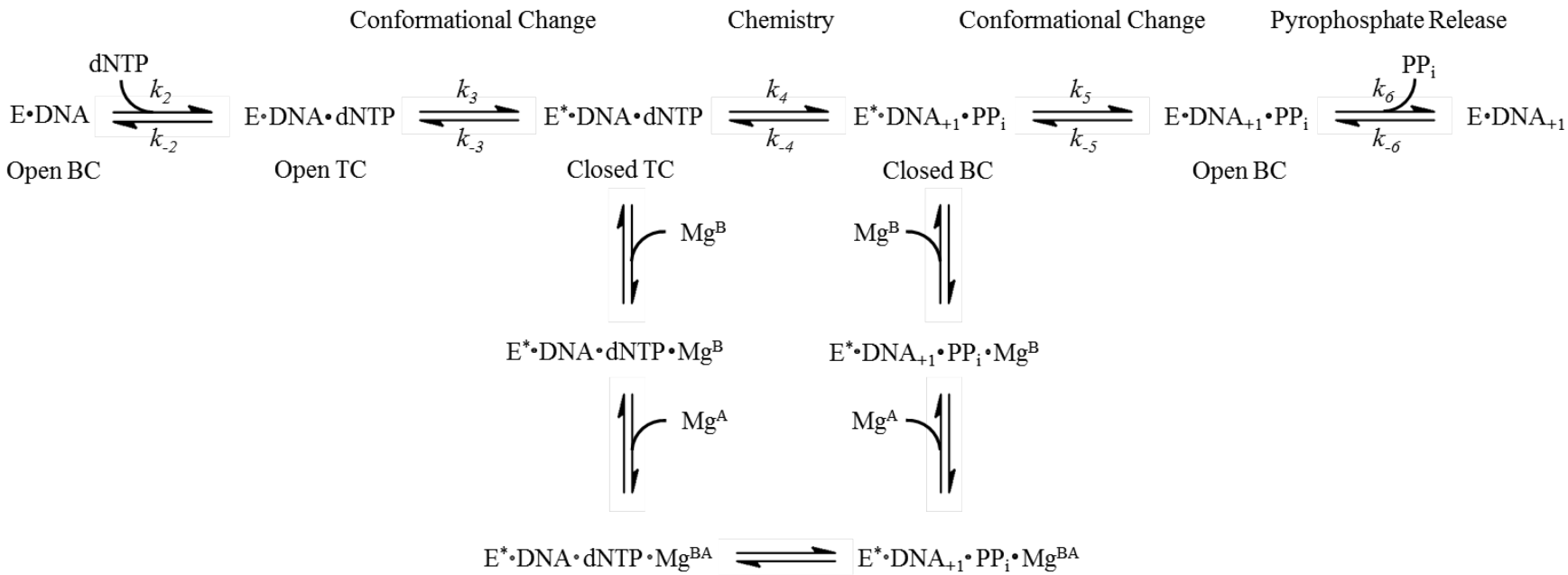


Figure 2.6 The minimal kinetic pathway and its' expansion of correct nucleotide incorporation by A-family DNA polymerases. Here, the minimal kinetic pathway for nucleotide incorporation by A-family DNA polymerases is expanded to highlight the spatiotemporal separation of substrate and metals, that places further constraints on the coordination chemistry, and sequential changes therein, than is encountered in static structures of 2',3'-ddNMP-terminated ternary complexes. E: enzyme, BC: binary complex, TC: ternary complex, PP_i: pyrophosphate.

Motif B

T. aquaticus	658	M	R	R	A	A	K	T	I	N	F	G	V	L	Y	G
E. coli	753	Q	R	R	S	A	K	A	I	N	F	G	L	I	Y	G
D. radiodurans	783	Q	R	R	A	A	K	T	V	N	F	G	V	L	Y	G
S. pneumoniae	702	D	R	R	N	A	K	A	V	N	F	G	V	V	Y	G
G. stearothermophilus	701	M	R	R	Q	A	K	A	V	N	F	G	I	V	Y	G
H. influenzae	756	Q	R	R	N	A	K	A	I	N	F	G	L	I	Y	G
M. tuberculosis	728	L	R	R	R	V	K	A	M	S	Y	G	L	A	Y	G
Bacteriophage T7	516	T	R	D	N	A	K	T	F	I	Y	G	F	L	Y	G
S. cervesiae Polγ	744	S	R	N	E	A	K	I	F	N	Y	G	R	I	Y	G
H. sapiens Polγ	942	S	R	E	H	A	K	I	F	N	Y	G	R	I	Y	G

Figure 2.7 Sequence alignment of evolutionarily conserved, catalytically-relevant residues of motif B among A-family DNA polymerases. Arginine and Lysine residues (solid blue box) in evolutionarily conserved O-helices of A-family DNA and RNA polymerases are immutable, consistently-spaced and are of the same helical length (O-helices) to facilitate proper geometry and interaction with incoming (d)NTPs.

Motif A and C conformational changes in substrate-free binary complexes and metal-free and metal-ligated ternary complexes are evident in the structures presented here following reversion of the closed ternary complex containing the non-hydrolyzable dGTP analog, dGMPNPP, to the open conformation during preparation of catalytically-competent binary complexes (Fig. 2.8) in which a conserved Asp residue in motif A, ligating the binding metal, is seen rotated 3.3Å away from its canonical orientation and distance when coordinating the binding metal, while the second conserved Asp residue in motif C, which coordinates the catalytic metal, is seen rotated only 0.9Å away from its canonical position and the conserved Glu of the same motif rotated a striking 2.4Å inwards, closer to and stabilizing the O3' in the absence of substrate and metals. The negative charge buildup upon metal-free substrate binding (Fig. 2.9) repels motif C Glu 2.4Å away from its canonical position whilst simultaneously facilitating the inwards rotation of motif A Asp by 2.8Å, still ~0.5Å short of the canonical position, in preparation for an incoming binding metal to which it would ligate. Indeed, upon Mg^B binding a modest but concerted change in motif A and C residues occurs to both coordinate the binding metal and repel further motif C Glu by ~1.0Å (Fig. 2.10), sufficiently disrupting hydrogen-bonding patterns that may lead to inadvertent activation of the 3'-oxyanion before the polymerase is able to discern whether or not the correct nucleotide is bound, with the aid of the binding metal, prior to catalytic metal (Mg^A) entry, coordination, and commitment to chemistry. This result is consistent with a structural role for the binding metal that facilitates increased selectivity of bound substrate by enforcing a rigid coordination geometry characteristic of Mg^{2+} metal-chelate complexes (34,35).

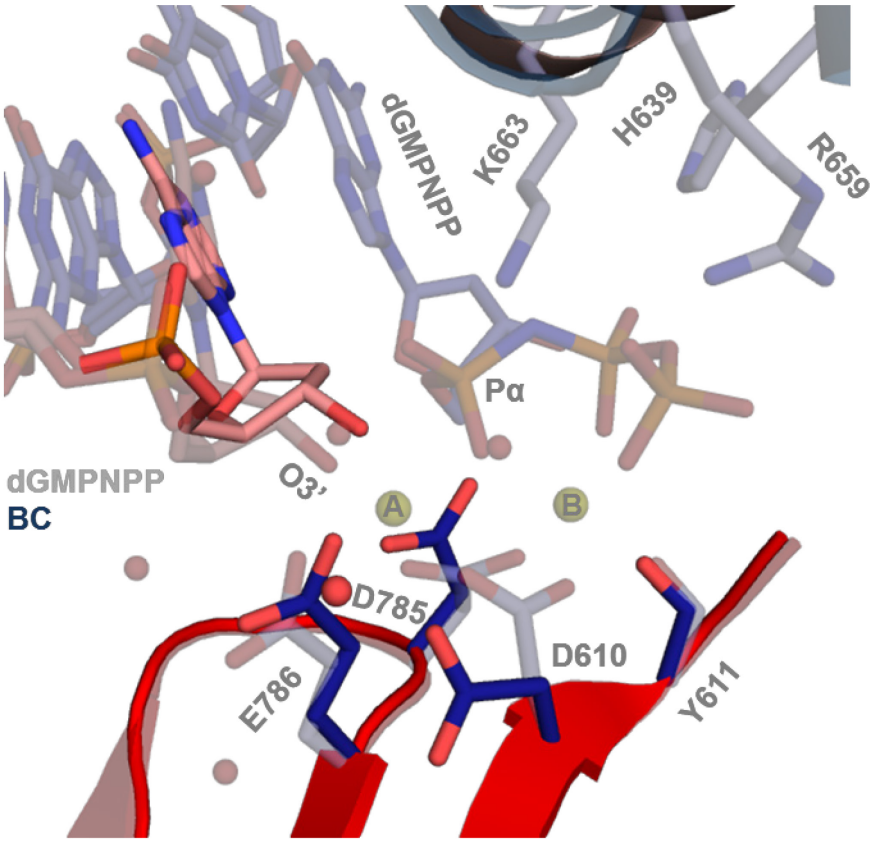


Figure 2.8 Superposition of the open binary and dGMPNPP-trapped closed ternary complexes of Klentaq DNA polymerase. Crystal structures of the dGMPNPP-trapped ternary (partially transparent) and open binary complexes are superimposed to highlight the structural changes that occur as a result of dGMPNPP diffusion out of the active site to generate the extension-ready binary complex. Motif C Glu786 shows little change relative to catalytic Asp residues (D785, D610) highlighting its role in stabilizing the O3' and Mg^A prior to chemistry.

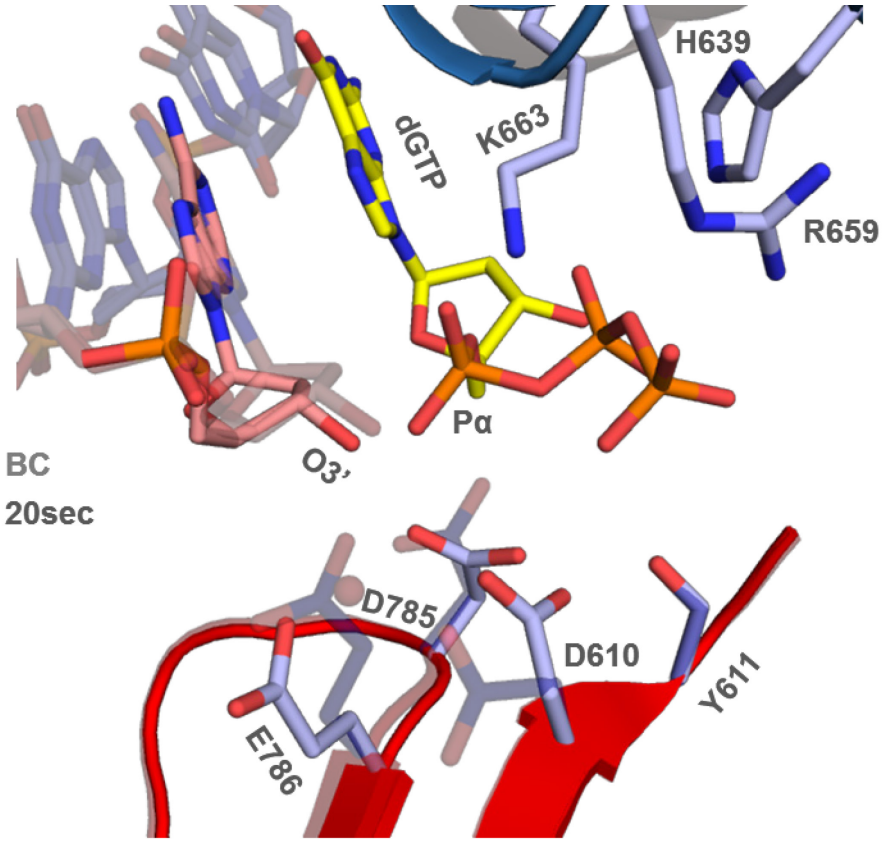


Figure 2.9 Superposition of the open binary and closed metal-free ternary complexes of Klentaq DNA polymerase. Crystal structures of the metal-free ternary and extension-ready binary complex (BC, partially transparent) are superimposed to highlight motif A and C rotations. Upon dGTP binding, motif C Glu786 rotates away from the active site to prevent premature activation of the O3' nucleophile. The metal-free ternary complex supports the spatiotemporal separation of dNTP binding and metal-ion entry.

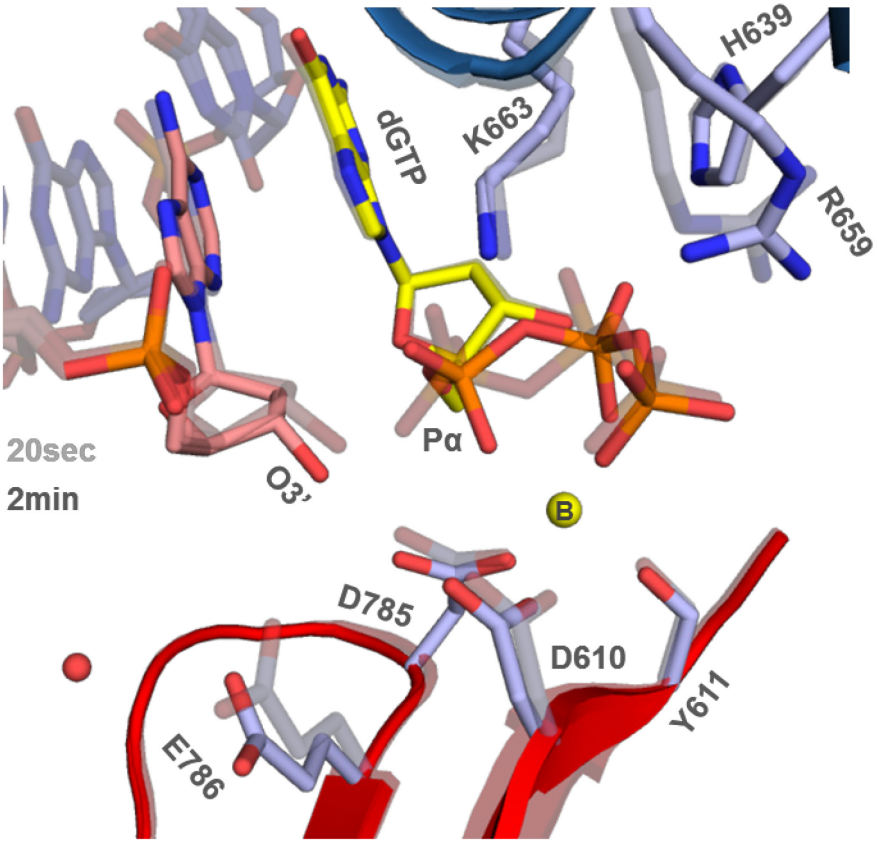


Figure 2.10 Superposition of the closed metal-free and Mg^B-coordinated ternary complexes of Klentaq DNA polymerase. Crystal structures of the metal-free (partially transparent) and Mg^B-coordinated ternary complexes are superimposed to highlight motif A and C and dNTP structural changes. Little structural change occurs in motifs A and C residues upon Mg^B binding, but dGTP undergoes conformational change to facilitate proper orientation of the triphosphate non-bridging oxygens to coordinate Mg^B, supporting a structural role for Mg^B and highlighting the sequential metal-ion entry into the active site after dNTP binding.

Catalytic metal (Mg^A) entry subsequent to Mg^B ligation is more flexible in its coordination requirements as its octahedral coordination in the canonical position for Mg^A can be satisfied by coordination of one or more water molecules that assist in nucleophile generation in the pentacovalent transition state complex. This observation is also seen in the structures presented here for the transition state complex representing the predominant species accumulating prior to phosphodiester bond formation in the non-hydrolyzable dGMPNPP complex. Octahedral coordination is satisfied via involvement of a water molecule and motif C Glu residue with all coordinating moieties within 2.0-2.4Å bond lengths. Immediately after O3' activation as phosphodiester bond formation commences E786 rotates 2.4Å away from the active site, the implications of which in O3' activation are discussed below, and destabilizing Mg^A that now shifts 1.0Å below the plane of canonical coordination resulting in bond lengths of 2.4-2.8Å with 4 of the previous 6 coordinating moieties, and weak electrostatic interaction with a solvent molecule ~4Å away (Fig. 2.11). These observations are in line with mutagenesis studies of motif C carboxylates in polymerases which result in the enzyme still retaining some activity (34,35). However, mutations in motif A completely disrupt activity (34,35). Thus, a kinetic pathway in which metal ion entry is both sequential and concerted, and decoupled from ground state substrate binding and conformational change, as shown here, is a tremendous advantage for polymerases that increases selectivity in an effort to sustain fidelity.

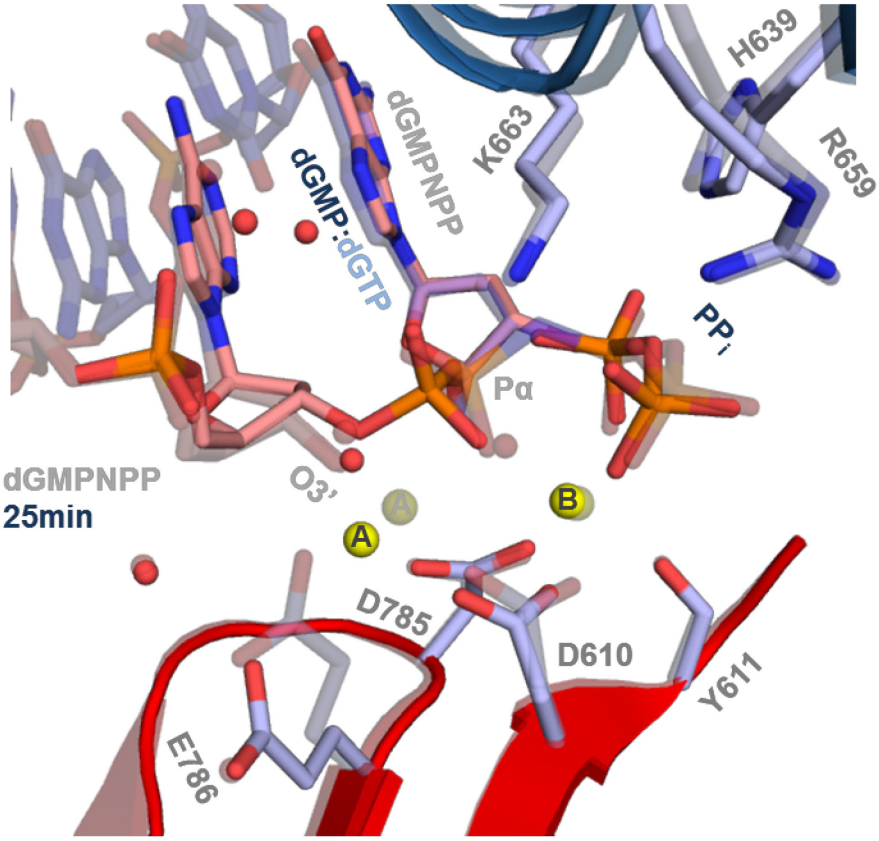


Figure 2.11 Superposition of the dGMPNPP-trapped and post-incorporation closed ternary complexes of Klentaq DNA polymerase. The crystal structures of the dGMPNPP-trapped ternary complex (partially transparent) and the product complex obtained at 25min are superimposed to highlight the structural variation that occurs as the complex transitions from pre-chemistry to chemistry and product formation. Glu786 of motif C is seen rotated outwards by 3.6Å after activation, destabilizing Mg^{A}, increasing the Mg^{BA} interatomic distance to that of the inactive state, in preparation for its release from the active site.

Prior to this work, fidelity in nucleotide incorporation by various DNAP families was proposed to follow one of two models in an attempt to explain the potential mechanisms underlying the significant differences in fidelity by different DNAPs. An "induced-fit" model posits that polymerase fidelity would be the result of conformational coupling in which the free energy change from correct nucleotide binding is used to drive a rate-limiting conformational change in the polymerase that precedes the chemistry step

(20,21). Nucleotide selection in this model would be a two-step process in which ground state binding of the correct nucleotide, which would be more favourable compared to the incorrect one, is followed by the formation of a catalytically-competent ternary complex driven by the favourable energy produced from ground state binding. The second "rate-limiting transition state" model posits that polymerase fidelity is attributed to a rate-limiting chemistry step in which the free energy change in the formation of transition state species of correct and incorrect nucleotides dictates whether the polymerase will proceed in the chemistry step or revert back to an open conformation and release the nucleotide (21). In this model, binding of the correct nucleotide is followed by a rapid conformational change and a subsequent slower chemistry step. While evidence in support of both models exists, and it has been shown that different DNAP families adopt different mechanisms for fidelity (22), it appears that fidelity of A-family DNAPs, or at least KlenTaq, can be in part attributed to a rate-limiting transition state model.

2.4.2 3'-Oxyanion Activation – A Catalytic Water, Captured Transition State, and Consequences to Proton Transfer

Two metal-ion catalysis as the mechanism for phosphoryl transfer has gained significant support since its proposal in 1993 by Steitz and Steitz. This support came from crystal structures of DNA and RNA polymerases, and restriction endonucleases, in which the ubiquitous observation of two Mg^{2+} ions coordinated by two conserved Asp residues in the active sites of these enzymes was made (36,37). The observation of two water molecules (S1 and S2) (Fig. 2.12) in dideoxy-terminated ternary complexes coordinated to the catalytic metal, Mg^{A}, fulfilling the coordination geometry requirement of the two Mg^{2+} ions, now in the active site coordinated by the triphosphate of the bound nucleotide, a backbone carbonyl oxygen, two conserved Asp carboxylates, and the approximate would-be position of the O3', suggested that the likely role of the catalytic metal, Mg^{A}, is the reduction of the pK_a of a nearby water molecule, presumably one of the two canonical water molecules coordinated by this metal, to assist in deprotonation of the O3', generating the 3'-oxyanion for nucleophilic attack, while Mg^{B} stabilizes the pentacovalent transition state. While the number of water molecules speculated to be involved at any point during the reaction was serendipitously correct, the absence of an O3' in previous 2',3'-dideoxy-terminated structures pointed to the possibility that one of

the water molecules was present only to satisfy the coordination geometry of Mg^A as the water molecule (S2) was observed in the crystal structure of the 2',3'-dideoxy-terminated ternary complex (3KTQ) to be consistently near the approximate location of a would-be O3' in an otherwise catalytically-competent counterpart. This observation provided further support that the minimal octahedral geometry of the two Mg^{2+} ions would be satisfied in a catalytically competent closed ternary complex in the presence of the two water molecules coordinated to Mg^A. The presence of an O3' in a true catalytically-competent ternary complex, however, suggests further involvement of other moieties in the microenvironment of the active site, and provides new insights with structural and mechanistic implications that are observed and described here for the first time for A-family DNA polymerases.

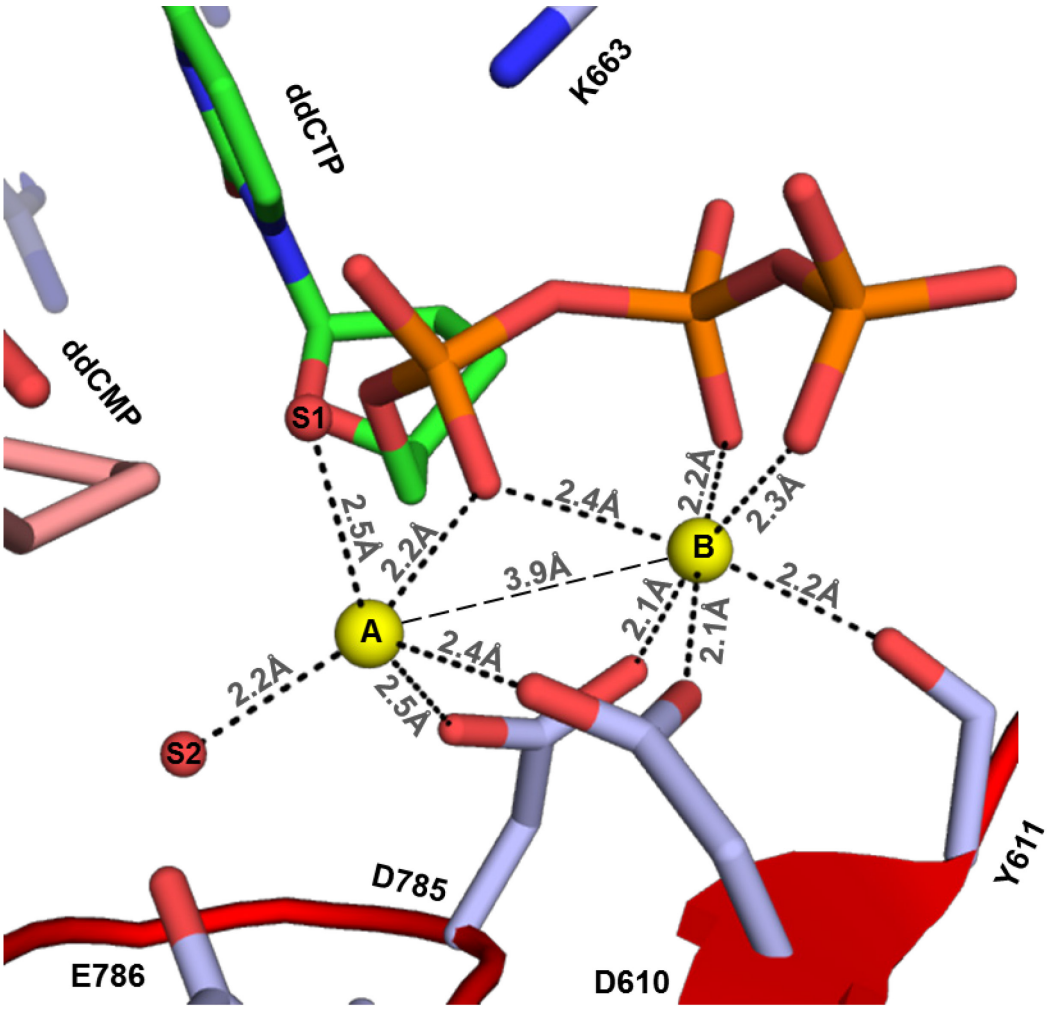


Figure 2.12 Coordination of solvent molecules to Mg^A in the dideoxy-terminated ternary complex. The crystal structure of the 2',3'-ddNMP-terminated ternary complex, 3KTQ, shows binding and catalytic metal at an interatomic distance that is unfavourable for catalysis but maintaining the proper coordination of solvent molecules to Mg^A even in the absence of the O3' to participate in coordination.

Illustrated in the structures presented herein, 3'-oxyanion activation is likely the result of a third water molecule (S3) within bonding distance of the scissile phosphate *pro*-R_p oxygen (O2A) that appears, after the inwards rotation of motif C Glu, to coordinate Mg^A at a distance of 2.2-2.3Å. This water molecule (S3) remains at that approximate position as the newly formed phosphodiester bond accumulates, observed in the structure obtained at 25 minutes (Fig. 2.13) with E786 seen rotated away from the canonical position (~2.3Å) and now no longer interacting with Mg^A due to the absence of

the second water molecule (S2). The catalytic water molecule (S3) disappears after phosphodiester bond formation, immediately before or simultaneously with Mg^A movement away from Mg^B (inactivation) and diffusion out of the active site, as observed in the structure obtained at 28 minutes. The position of E786 in the transition state makes it an ideal candidate for a general base poised to accept the first proton transferred as a result of 3'-oxyanion activation. Immediately after this transition state, and before pyrophosphate (PP_i) release, E786 reverts to its original location at a distance of ~3.7Å away from Mg^A. These findings are consistent with the proposed role of the conserved motif C Glu as a general base in the transition state that simultaneously transfers a proton to the conserved O-helix Lys (motif B) acting as general acid and seen stabilizing non-bridging oxygens of the α- and γ-phosphates (O2A and O2G).

Further support of the role of conserved motif C Glu as a general base comes from studies in which substitution of its conserved equivalent Glu883 in *E. coli* results in canonical metal-ion substrate binding and coordination but reduced catalytic rate (38,39). Engagement of motif C Glu in Mg^A coordination seen in the dGMPNPP-trapped transition state structure presented here, along with this evidence suggests that conserved motif C Glu coordination of Mg^A is likely the rate-limiting step, particularly in light of little movement in motif A and C Asp's between the stages of Mg^A entry and PP_i release.

The potential role of conserved O-helix Lys as a general acid has been hypothesized in A-family DNA and RNA polymerases, and its functionally-homologous conserved Lys in the P-helix of B-family DNA polymerases (16,17,40,41). This conserved Lys coordinates the *pro*-R_p oxygen (O2A) of the Pα in A- and B-family polymerases and has no functional equivalent in the low fidelity X- and Y-family repair polymerases β and Dpo4/η, respectively, which are discussed further below. As previously mentioned, in the pre-chemistry and chemistry structures, during and after E786 engagement, the third water molecule (S3) is seen within 3Å of the *pro*-R_p oxygen that is also coordinated by the conserved motif B Lys until PPi release (Fig. 2.13). Consistent with this line of reasoning, alterations to this conserved Lys in the O-helix of A-family DNA and RNA polymerases or the P-helix of B-family RB69 DNA polymerase severely decreased or completely abolished activity (42-44).

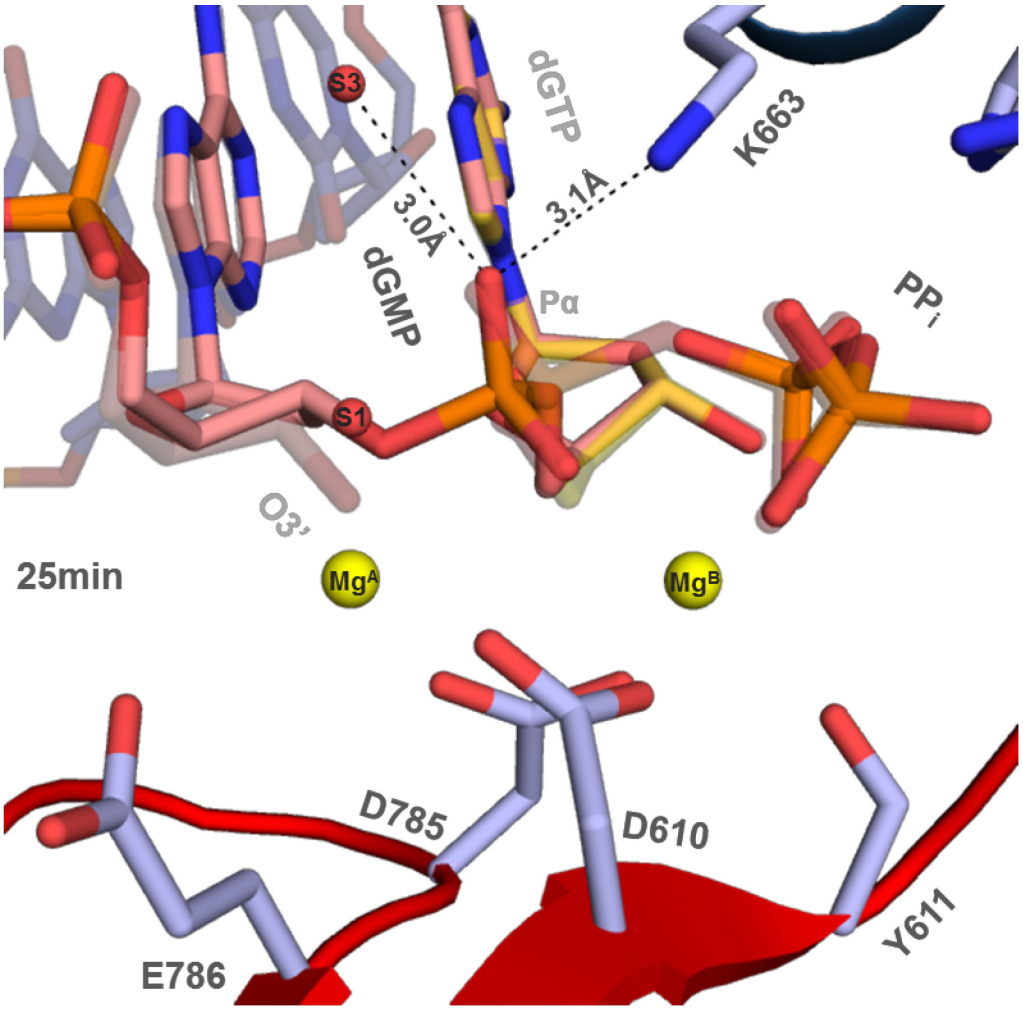


Figure 2.13 A catalytic water during product formation. Unlike the structure obtained for pre-chemistry (dGMPNPP-trapped) or the 2',3'-ddNMP-terminated ternary complexes, a water molecule appears in the structure obtained at 25 minutes (25min) within catalytic distance of Pα and remains in the vicinity until the product complex saturates the electron density in the active site, and disappears before, during, or after Mg^A diffusion out of the active site.

In line with this proposed role for conserved motif B and C Lys and Glu, respectively, acting as general acid and base for proton transfer in two metal-ion catalysis, the least speculative pathway for the order of events in reaction progression during phosphoryl transfer *in crystallo* is likely the following: 1) upon correct nucleotide binding to an open, catalytically-competent binary complex the O-helix undergoes very rapid conformational change from the open to closed states within the first 20 seconds to

form a metal-free closed substrate complex, at which point 2) the canonical binding metal (Mg^B) diffuses into the active site and is coordinated by conserved motif A and C Asp's, a neighbouring motif A backbone carbonyl oxygen, and the non-bridging oxygens of the triphosphate of the bound nucleotide. The consequent inwards rotation of motif A Asp to participate in coordinating Mg^B and 3) the proximity of this new negative charge that is yet to be chelated by Mg^A entry results in a larger negative charge at the scissile phosphate that would later be stabilized by hydrogen bonding with the catalytic, third water molecule (S3) interacting with both the *pro*-R_p oxygen of the scissile phosphate and the conserved motif B Lys. The subsequent 4) entry of Mg^A and coordination to the 3'-OH, scissile phosphate, and motif A and C Asp's and motif C Glu places Mg^A 4.1Å away from Mg^B with the two canonical water molecules, one coordinated at a distance of ~2.6Å from Mg^A (S1), and the second (S3) transiently interacting with Mg^A and motif B Lys at a distance of ~4Å and ~3.5Å, respectively, resulting in the same number of canonical water molecules completing Mg^A coordination in previous crystal structures except that they are placed on opposite sides of the scissile phosphate (S1 and S3) and only one of which (S1) directly coordinates Mg^A along with motif C Glu (Fig. 2.14 & 2.15).

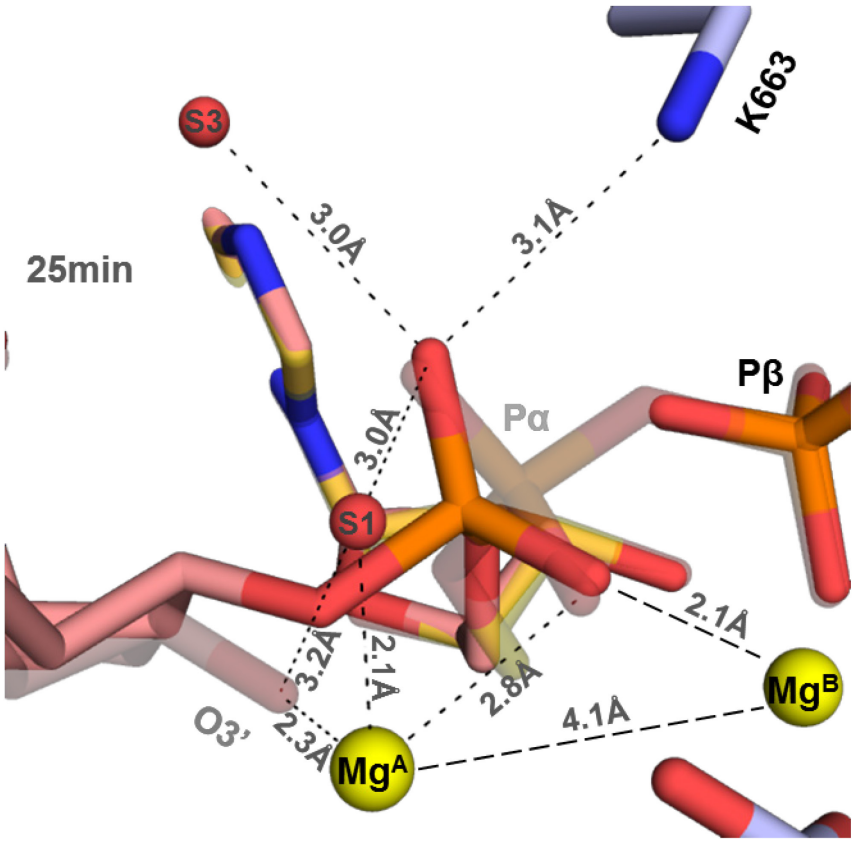


Figure 2.14 Catalytically-relevant distances of solvent molecules during phosphodiester bond formation. Both solvent molecules S1 and S3 and motif B Lys663 are 3.0-3.1Å away from the Pα non-bridging oxygen available to participate in potential proton transfer during phosphodiester bond formation.

Activation 5) of the 3'-oxyanion by S3 occurs simultaneously with direct coordination of Mg^A by conserved motif C Glu which results in a 0.6Å decrease in the interatomic distance of Mg^A and Mg^B, now at 3.5Å, thereby promoting the now energetically favourable chemistry (Fig. 2.15). Changes in the interatomic distance of binding and catalytic Mg^{2+} metal ions have been shown to be the basis for activation of catalysis (34). The canonical 4Å interatomic distance observed in crystal structures of restriction endonucleases and B-family RB69 DNAP has also been shown to decrease to 3.5Å, and to 3Å in T7 RNA polymerase, for catalysis (45). Direct coordination of Mg^A by motif C Glu allows the conserved residue to act as a general base, accepting the first proton from 3'-oxyanion activation by the activating or catalytic water (S3) that maintains

contact with the scissile phosphate *pro*-R_p oxygen and motif B Lys, making Lys an excellent general acid to facilitate proton transfer from Glu to the PP_i leaving group.

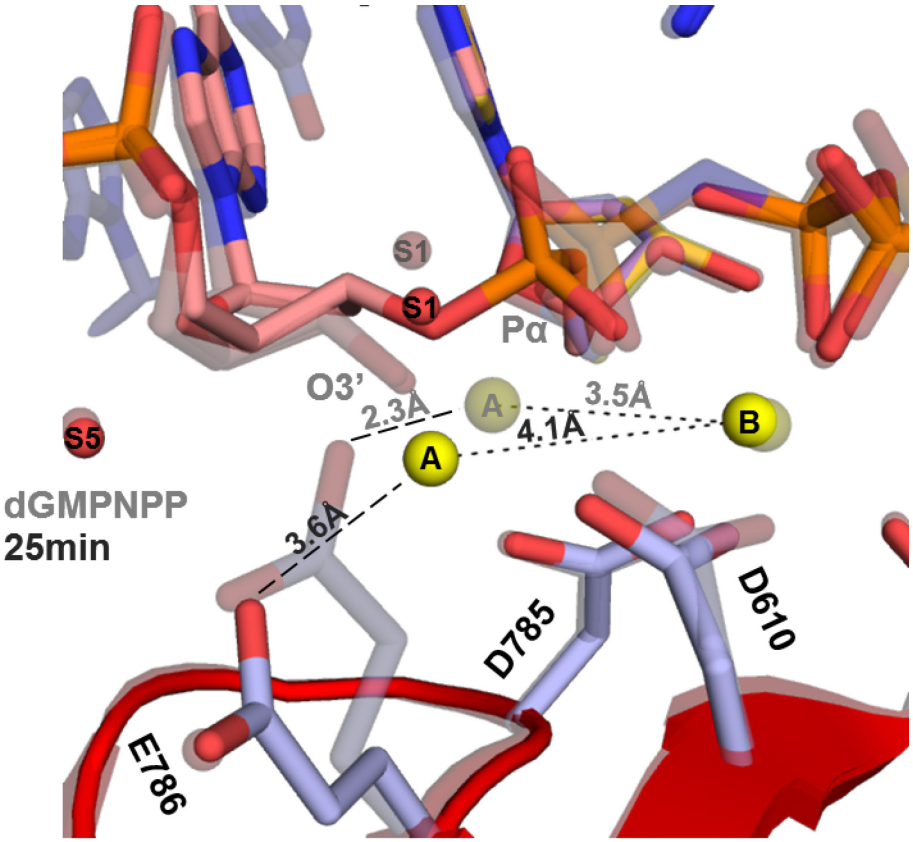


Figure 2.15 Interatomic distances of Mg^B and Mg^A during and after product complex formation. The structures of the dGMPNPP-trapped (partially transparent) and product ternary complexes are superimposed here to illustrate both the interatomic distances of Mg^{BA} and the distances of Mg^A to motif C Glu786 before and after chemistry.

Once protonation of the PP_i leaving group occurs, and its subsequent movement ~0.5Å closer to conserved N-helix His, motif C Glu simultaneously moves back to its canonical inactive position in polymerases thereby destabilizing Mg^A, which now moves back to its canonical inactive 4Å distance from Mg^B along with dissipation of the activating water (S3). As Mg^A becomes further destabilized by the loss of another coordinating partner after phosphodiester bond formation, unable to complete its octahedral geometry, it diffuses out of the active site, leaving Mg^B-coordinated PP_i, with its γ-PO4 non-bridging oxygen still hydrogen bonded to the newly incorporated phosphate backbone non-bridging oxygen via motif B Lys (Fig. 2.18C and D), which ultimately dissociates Mg^B resulting in its diffusion out of the active site and destabilizing

PP_i due to the negative charge buildup before its diffusion out of the active site as well. Whether or not the negative charge build-up on the PP_i leaving group after Mg^B dissociation is the likely mechanism by which the O-helix reverts to the open conformation after nucleotide incorporation is uncertain. It is certainly appealing, however, and fits the current dogma that the still Arg- and His-coordinated PP_i would accumulate a negative charge upon Mg^B dissociation that results in electrostatic repulsion by active site carboxylates and backbone carbonyl oxygens that forces the O-helix to revert to the open conformation and release the pyrophosphate. Attempts to address this are presented elsewhere in this chapter.

It should be noted, however, that if Lys is the penultimate proton acceptor or simply a proton shuttle (PP_i being the final proton acceptor) then an interpretation is warranted regarding the marked difference in catalytic efficiency between A-family DNA and RNA polymerases. If protonation of the PP_i leaving group is the final step in two metal-ion catalysis prior to PP_i release then involvement of residues from the N-helix in A-family DNA and RNA polymerases, which interact with a β-PO_4 non-bridging oxygen, and water molecules and conserved O-helix Arg coordinating γ-PO_4 non-bridging oxygens, should not significantly affect the rate of polymerization (k_{pol}) . This is, however, not the case as A-family DNA polymerases have a conserved His in the N-helix that directly interacts with a non-bridging oxygen of the β-PO_4 throughout all reaction intermediates, including PPi release, while A-family RNA polymerases have a conserved Tyr (T7- and N4 RNAP Y571 and Y612, respectively) at that position which results in a reaction rate 10-times slower relative to A-family DNA polymerases (Fig. 2.16) (10). Since the N-helix His can participate as both a general acid and general base, the implication of a Tyr substitution resulting in a markedly decreased rate warrants an examination as to the possibility of the N-helix His serving as general acid. In fact, mutations of this conserved Tyr to Ser in A-family RNA polymerases, in an effort to disrupt this interaction but maintain the polar character of the side chain, results in loss of activity (46). An alternative explanation is that the conserved His residue is able to better stabilize the PP_i leaving group relative to its Tyr counterpart which may destabilize the PP_i by repulsion resulting in slower PP_i release thereby decreasing the rate of formation

of transition state species during catalysis and effectively destabilizing the product complex and shifting the equilibrium in favour of the substrate complex.

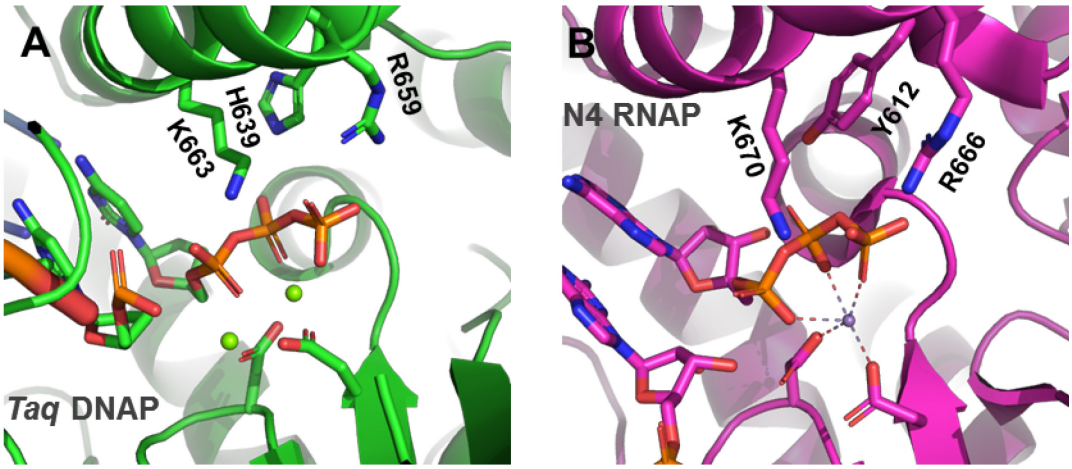


Figure 2.16 Structural basis for leaving group instability and effect on product complex formation. The crystal structures rendered here are of A) *Taq* and B) N4 phage DNA and RNA polymerases, respectively, and show the evolutionarily conserved motif B (O-helix) Lys and Arg residues, and the variant N-helix His (DNAP) or Tyr (RNAP), that interact with the triphosphate moiety of the incoming nucleotide.

2.4.3 Coupled Dynamics in Translocation

Translocation by DNA polymerases subsequent to nucleotide incorporation is a rapid process *in vivo* and *in vitro* whose kinetic parameters are not entirely defined as it is difficult to capture and resolve a translocating complex. As such, translocation was not considered a potential mechanistic candidate for the rate-limiting step in nucleotide incorporation. Multiple models for translocation have emerged driven primarily by the understanding that a negative charge build-up consequent to phosphodiester bond formation results in charge repulsion as the newly incorporated scissile phosphate now lies in close proximity to motif A and C Asp carboxylates (47,48). Following the catalytic pathway delineated thus far in previous sections, the catalytic metal (Mg^A) is first destabilized by loss of two coordinating partners through the outwards rotation of conserved motif C Glu, and loss of 3'-oxyanion upon phosphodiester bond formation, which results in an unsatisfied coordination geometry leading to release of Mg^A, and buildup of a slight negative charge as the new phosphate clashes with the conserved

Asp's that are also coordinating Mg^B-coordinated pyrophosphate. The presence of water molecules in the vicinity of the Mg^B-coordinated PP_i aids in slightly mitigating this negative charge build-up. However, once Mg^B exits the active site, it leaves behind a more significant negative charge buildup from the PP_i still coordinated to O-helix and N-helix Arg and His, respectively, and this new net negative charge has been postulated to be sufficient to drive reversion of the O-helix to the open conformation (47,48).

Interestingly, in the presence of metal-free PP_i in the active site after nucleotide incorporation, or even after pyrophosphate release, the O-helix remains in the closed conformation (Fig. 2.3E/Fig. 2.17) in the structures presented here for *in crystallo* single turnover kinetics. This platform was designed to prevent or slow down translocation through conservation of crystal packing which forces the 5'-end of the duplex DNA to stack against aromatic residues in symmetry-related molecules at an approximate distance equal to that observed for adjacent base pairs, thereby preventing translocation under these conditions. Models for translocation along the duplex DNA onto the next templating nucleobase have long suggested that pyrophosphate release, O-helix opening, or both would provide the intrinsic energy necessary to drive translocation. Contrary to this, the structures presented here provide the first direct evidence that suggests that translocation in A-family polymerases is necessary for O-helix opening. This observation of coupled dynamics between translocation and O-helix opening indicates that the charge buildup at the active site subsequent to pyrophosphate release alone is not sufficient to drive O-helix opening but that translocation must occur prior to, or simultaneously with O-helix reversion to the open conformation.

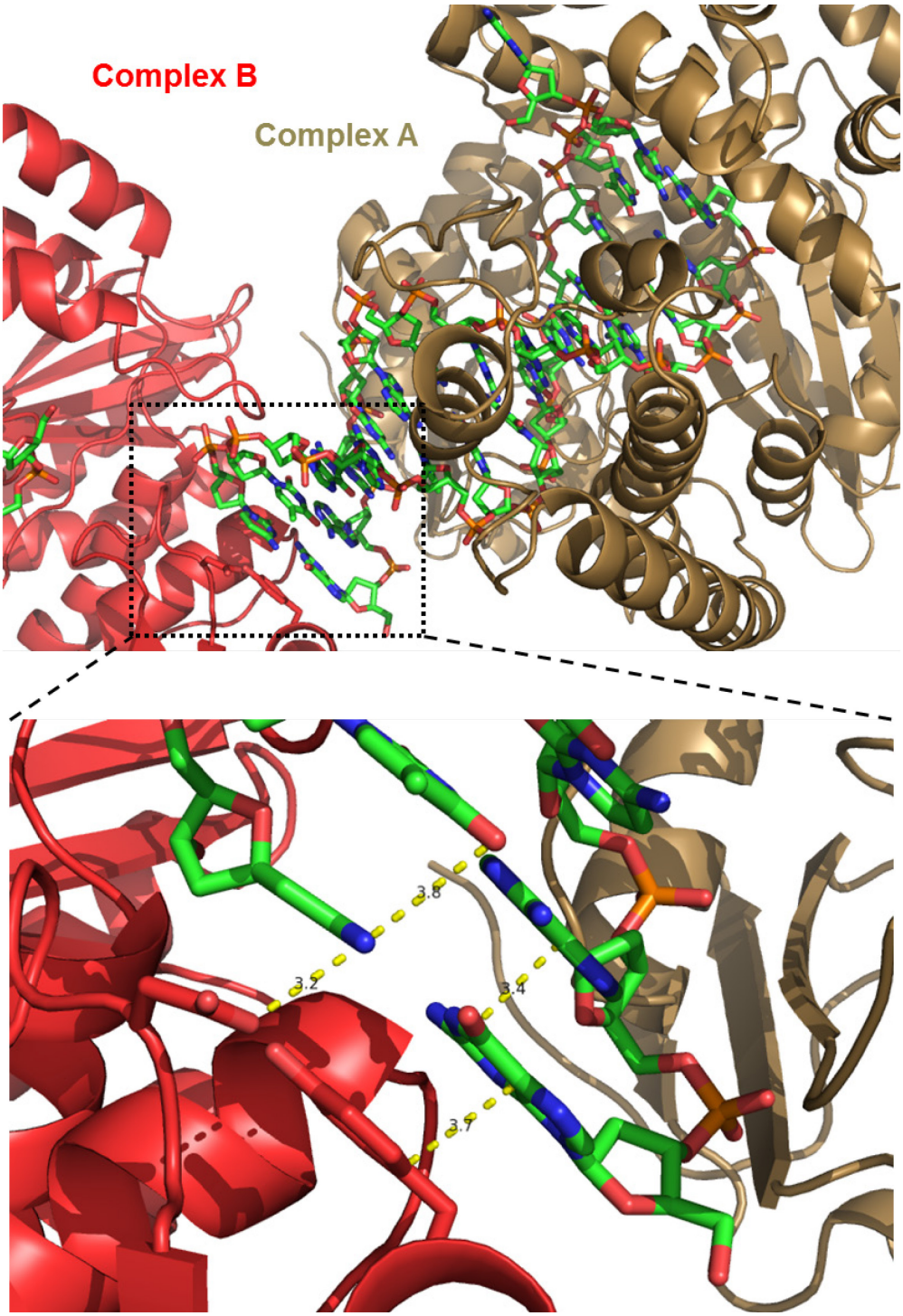


Figure 2.17 Crystal packing in the single-turnover, non-translocating complex. Duplex DNA in the ternary complex is constrained by contact with, and stacking against Y339 and D320 from symmetry-related molecules necessary for crystal packing and affecting crystallization of the dGMPNPP-trapped ternary complex. The observed distance is approximately similar to the base stacking distance, 3.6Å, providing a possible rationale for the inability to observe translocation *in crystallo.*

It should be noted that very recent structural studies of *Escherichia coli* DNA polymerase IV, a Y-family DNAP, showed in X-ray crystal structures that PP_i is not immediately released after Mg^B exits the active site but that it first is hydrolyzed to two phosphate ions through pyrophosphorolysis (49). The authors suggest that nucleotide incorporation and PP_i release are likely only slightly favourable and that pyrophosphate hydrolysis is an added step to the conventional minimalist kinetic scheme for reaction progression that is necessary, even in the absence of other enzymes such as pyrophosphatases, to generate a larger negative free energy change that renders the reaction more favourable. The authors further show using *in vitro* primer extension assays by other DNA polymerase families (A, B, and RT) using βγ-non-hydrolyzable analogues of the correct incoming nucleotide, wherein said nucleotide is incorporated, but the resultant released PP_i group cannot be hydrolyzed to monophosphates, that all polymerases tested failed to incorporate the modified nucleotides and conclude that PP_i hydrolysis is necessary for DNA synthesis and that it may be a conserved feature among DNAP families. The authors, however, did not examine the effect of this phenomenon on translocation and reversion to an open complex. We have not observed the same species in X-ray crystal structures obtained at later time points, but given that our platform is amenable to resolving PP_i at said time points it would be a worthwhile investigation into this possibility. Our results, however, do not indicate that the released pyrophosphate undergoes pyrophosphorolysis prior to evacuating the active site as even after PP_i release, or monophosphates, if that can be confirmed in our structures, the O-helix remains in the closed conformation in the absence of a translocation event (Fig. 2.18D).

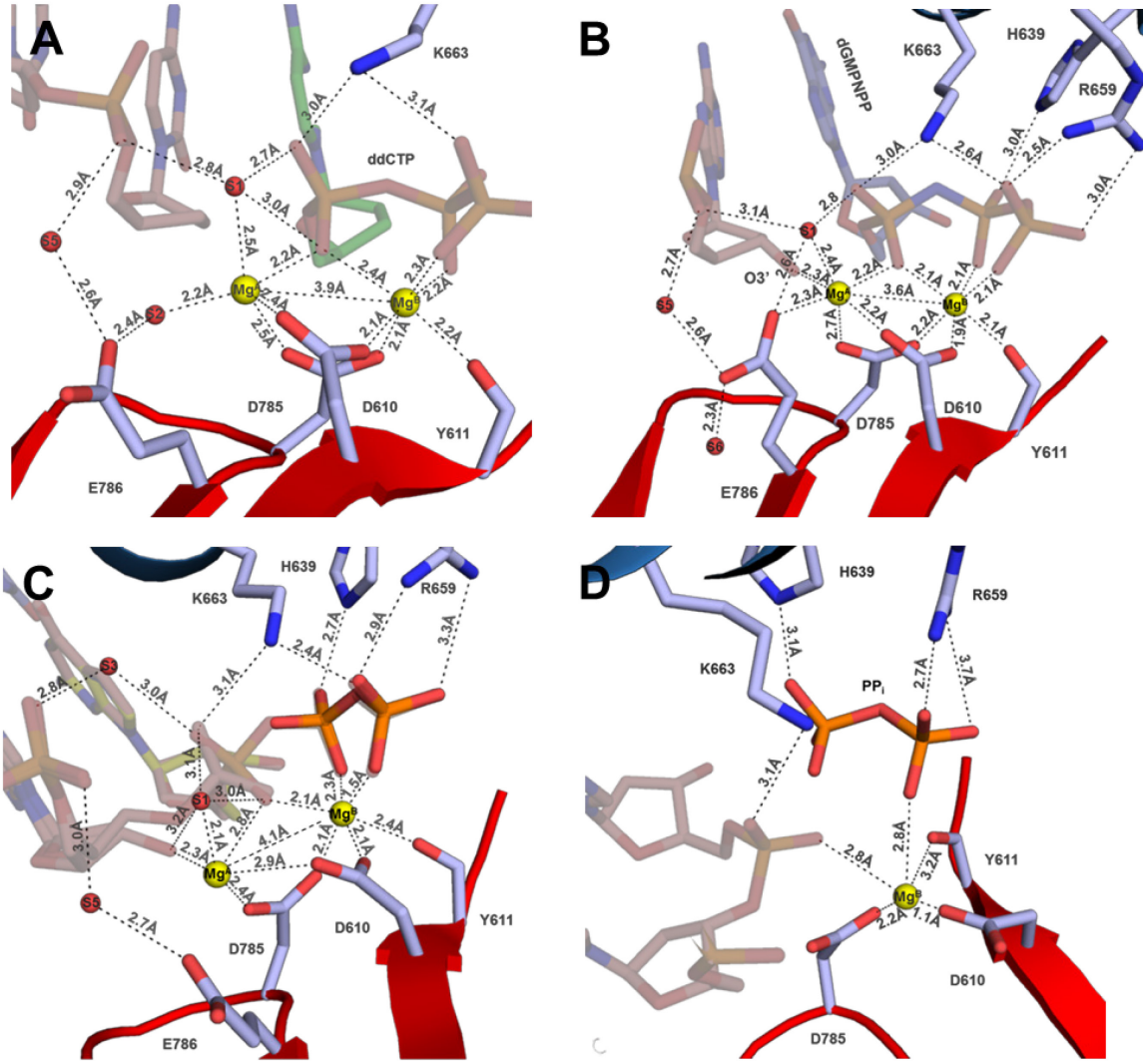


Figure 2.18 Emergence and catalytic distances of solvent molecules in dideoxy-terminated and catalytically-competent ternary complexes of KlenTaq DNA polymerase. In panels A, B, and C the interaction of active site residues, substrate, metal ions, and solvent molecules are conserved, however, the catalytic water (S3) only appears during phosphodiester bond formation (C) and after O3' activation by the transiently interacting Glu786 (B). In the Mg^B-coordinated PP_i complex (D), no solvent molecules are observed in the active site.

2.4.4 Two- and Three-Metal Ion Catalysis: Evolutionary Imperative or Mechanistic Contingency

The mechanism for all phosphoryl transfer reactions in nucleic acids has been shown to be bimolecular nucleophilic substitution (S_N2) (50-52) in which a nucleophile, as described above, is activated via deprotonation for an in-line attack on the phosphate group in question resulting in phosphodiester bond cleavage or formation, and, in case of formation, pyrophosphate protonation and release, which occur simultaneously via a pentacoordinated transition state. Two-metal ion catalysis, while generally accepted, is not without its detractors. Historically, the minimal kinetic scheme for nucleic acid enzymes required two metal-ions, barring catalytically-relevant waters and side-chains. This is supported by several decades of crystal structures of the ternary complexes of DNA and RNA polymerases, substrate-bound and apo forms of type II restriction endonucleases in which the metal-ions ligate the penultimate backbone phosphate oxygens, DNA transposases, and group I ribozymes that consistently show that proper geometry of coordination and formation of a pentacovalent complex can be achieved using only two Mg^{2+} metal ions (53-58), resulting in a kinetic scheme where one Mg^{2+} stabilizes the pentacovalent intermediate (Mg^B) while another (Mg^A) reduced the pK_a of a nucleophile, in this case the O3', via deprotonation of a coordinated water molecule whose hydroxide now deprotonates the O3' to generate the 3'-oxyanion nucleophile that will subsequently attack the scissile phosphate. Proton transfer during deprotonation of the nucleophile and protonation of the leaving group in the pentacovalent coordination complex relies on the transient participation of nearby water molecules, active site residues, or both, to act as general acids or bases, to donate or accept protons during catalysis, as we've provided evidence to support in A-family polymerases (Fig. 2.19).

Our initial contention was that the divide between two- and three-metal ion catalysis lies in the replicative or repair nature of polymerases, respectively, which may very well prove to be the case. This, however, may be a simplistic and naïve view as it disregards the overabundant structural and functional variety within a family in the repair enzymes in favour of an abstract classification driven by functional similarities and sequence conservation. The overarching observations regarding the structural and functional diversity of DNA polymerases and the emergence of a third metal ion in the

active sites of some but not all DNA polymerases during catalysis suggest that this may be due to either the evolutionary imperative of repair families with regards to their active site architecture, or a mechanistic contingency explained by the ability of various polymerases to utilize different divalent metal ions in different environments.

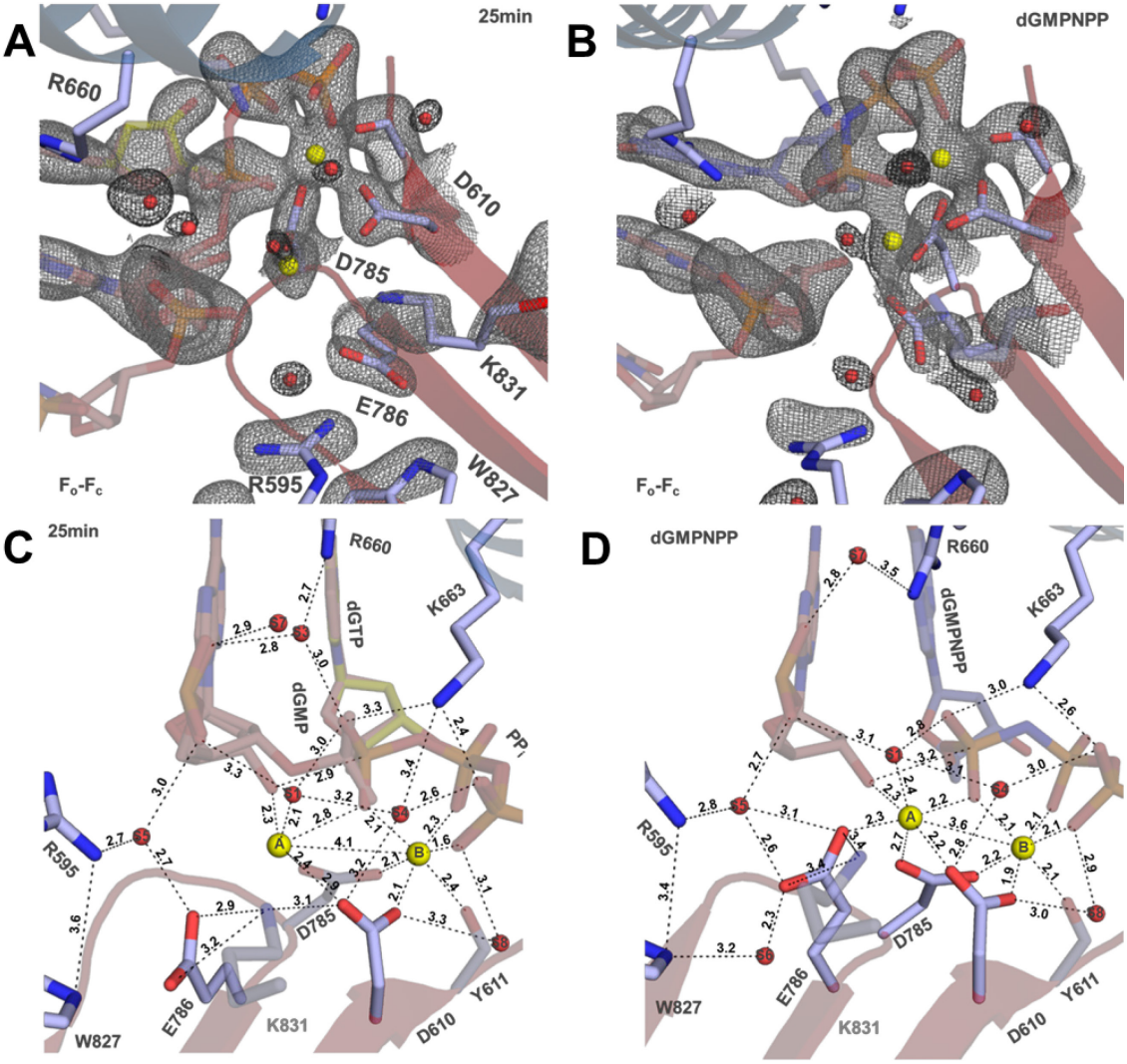


Figure 2.19 F_o-F_c omit maps of the dGMPNPP-trapped and product complexes.

F_o-F_c omit maps (grey mesh), contoured to 3σ, of the structure of the product complex obtained at 25min (A) and the dGMPNPP-trapped ternary complex (B) highlighting the number, catalytically-relevant distances (C and D, respectively), and coordination of solvent molecules in their respective active sites. The number and distances of solvent molecules is largely conserved except for the catalytic water (S3) that only appears during phosphodiester bond formation.

2.4.4.1 Function-centric view to variations in metal-ion catalysis

A function-centric view attempts to ascribe the emergence of a third metal ion to the evolutionary functional imperative of X- and Y-family polymerases which necessitates a dramatic decrease in fidelity brought about by reduced selectivity in a flexible and forgiving active site. The desire to maintain cellular integrity is one marked by a compromise in selectivity, to a much larger extent in Y- than in X-family polymerases, in an attempt to prevent cell cycle arrest. In line with the evidence presented here in support of two Mg^{2+} metal-ion catalysis in A-family polymerases, the lack of conserved basic residues coordinating the scissile phosphate, a Lys in A- and B-family polymerases, and stabilizing the transition state in X- and Y-family DNA polymerases, while biologically advantageous and functionally necessary owing to their respective roles as repair polymerases, suggests the existence of elaborate mechanisms that confer some degree of fidelity or selectivity, no matter how small, and catalytic efficiency. The absence of this prototypical basic residue in repair polymerases would require stabilization of the transition state by other moieties that are either absent or unnecessary during catalysis in A- and B-family polymerases.

The lack of a basic residue in place of the conserved A- and B-family Lys results in less geometric constraints during nucleotide incorporation in X- and Y-family polymerases. This, however, requires metals, waters, other side-chains, or an interplay of all three moieties to complete coordination, stabilize the transition state, and facilitate proton transfer to the PP_i leaving group during chemistry. It is undoubtedly functionally advantageous to mechanistically utilize small moieties like waters and metals since they allow for tremendous flexibility in incorporation as coordination, variations in bond lengths, from ~1.5-4Å, and variations in bonding partners, e.g., waters appearing and disappearing during different stages in catalysis, means such polymerases can develop a substantially higher tolerance for lesions, bulky adducts, and misincorporations, and, consequently, a larger functional repertoire to aid in the various repair functions.

In the crystal structures of both human DNA polymerases β and η, X- and Y-family polymerases, respectively, the catalytic, or activating, transient water molecule which we also describe here can be observed bridging the O3' and the scissile phosphate

before catalysis and, unlike A-family polymerases where it bridges the *pro*-R_p oxygen to the non-bridging oxygen of the primers' phosphate backbone (Fig. 2.18), remains shortly thereafter as the enzyme-product complex becomes more abundant, presumably to stabilize the product complex by neutralizing the negative charge buildup and, more importantly in this discussion, to coordinate a third Mg^{2+} metal ion appearing after the new phosphodiester bond begins to form and occupying the coordinate space reserved for the conserved Lys in A- and B-family polymerases. The presence of this third metal after product formation suggests it plays a role in either stabilizing the product complex, facilitating or participating in proton transfer, or both (8,9).

2.4.4.2 Metal-centric view to variations in metal-ion catalysis

However, it has recently been shown in the crystal structure of an archaeal B-family polymerase from *Thermococcus sp. 9°N-7* that it can contain two or three metal ions, one or two Mg^{2+} and one Mn^{2+}, coordinating active site residues and bound substrates (59), which warrants examination with respect to the divide we proposed for replicative and repair polymerases in the utilization of two or three Mg^{2+}, respectively, based on active site architecture alone. This examination adopts a metal-centric view as a mechanistic contingency in non-ideal catalytic environments. In crystal structures of A-family polymerases there is often one more metal ion on the opposite face of the O-helix or much farther upstream along the duplex DNA that in either case is not close enough to the active site to be involved in chemistry, but might, plausibly, become involved depending on catalytic conditions.

The observation that various DNA polymerases within and across different families are able to utilize non-ideal divalent metal ions (Mn^{2+}, Ni^{2+}, Co^{2+}, Cd^{2+}, Ca^{2+}, Zn^{2+}) to varying degrees, in some instances with catalytically significant variations in rate, substrate selectivity, and coordination numbers and bond lengths, suggests a necessity for active site flexibility (60-62), which may also include the transient presence of additional waters, metals, or both to satisfy the requirements for catalysis. Similarly, the pH-dependent variations in reaction rates during catalysis by different polymerases suggests that choice and number of metals, their consequent changes in coordination, and available solvent molecules may point to a mechanistic contingency in which the

polymerase is able to carry out its functions in non-ideal environments with reasonable fidelity (63-68). This can be demonstrated by contrasting Mg^{2+}- and Mn^{2+}-coordinated, catalytically-competent, ternary complexes of KlenTaq and the non-hydrolyzable analog of dGTP, dGMPNPP, in which the coordination of Mn^{2+} precludes involvement and potential activation by motif C Glu786 and instead is activated or at least coordinated by the terminal Lys831 to facilitate potential proton transfer (Fig. 2.20). A-family DNAPs from *T. aquaticus* and *B. stearothermophilus* possess this terminal Lys, however, *E. coli* and bacteriophage T7 have His residues at that position, and mutations of this His in *E. coli* abolish activity.

This mechanistic contingency is observed in archaeal B-family DNA polymerases that appear to possess a degree of flexibility in their active sites that is unlike anything seen in other families. Because of their environmental stresses, shifting metabolic functions and biochemical pathways, including utilization of alternative metal ions to facilitate catalysis with minor aberrations, depending on the metal in question, archaeal B-family polymerases are able to maintain function in very harsh environments. Unlike non-archaeal B-family polymerases, and indeed other polymerase families, this group can utilize Zn^{2+}, Cd^{2+}, and even Ca^{2+} for catalysis (65,69), where historically these divalent metal ions did not support catalysis in other polymerase families (68-70). Most DNA polymerases are inactive when Ca^{2+} is the coordinating metal but others, like the archaeal B- and X-families (Dpo4), are not, and those active with Ca^{2+} are also active in other, smaller metals, like Mg^{2+}, Mn^{2+}, or Zn^{2+}, which also have the same octahedral coordination geometry constraints, but with varying flexibility in bond lengths, suggesting some degree of flexibility in the active site that can accommodate and maintain proper coordination of such disparate metal ions in non-ideal catalytic environments. Whereas RNA polymerases differentially exhibit varying rates of transcription depending on whether the coordinating metal-ion is Mg^{2+} or Mn^{2+}, with little effect on fidelity, the latter significantly reduces the fidelity of replicative DNA polymerases.

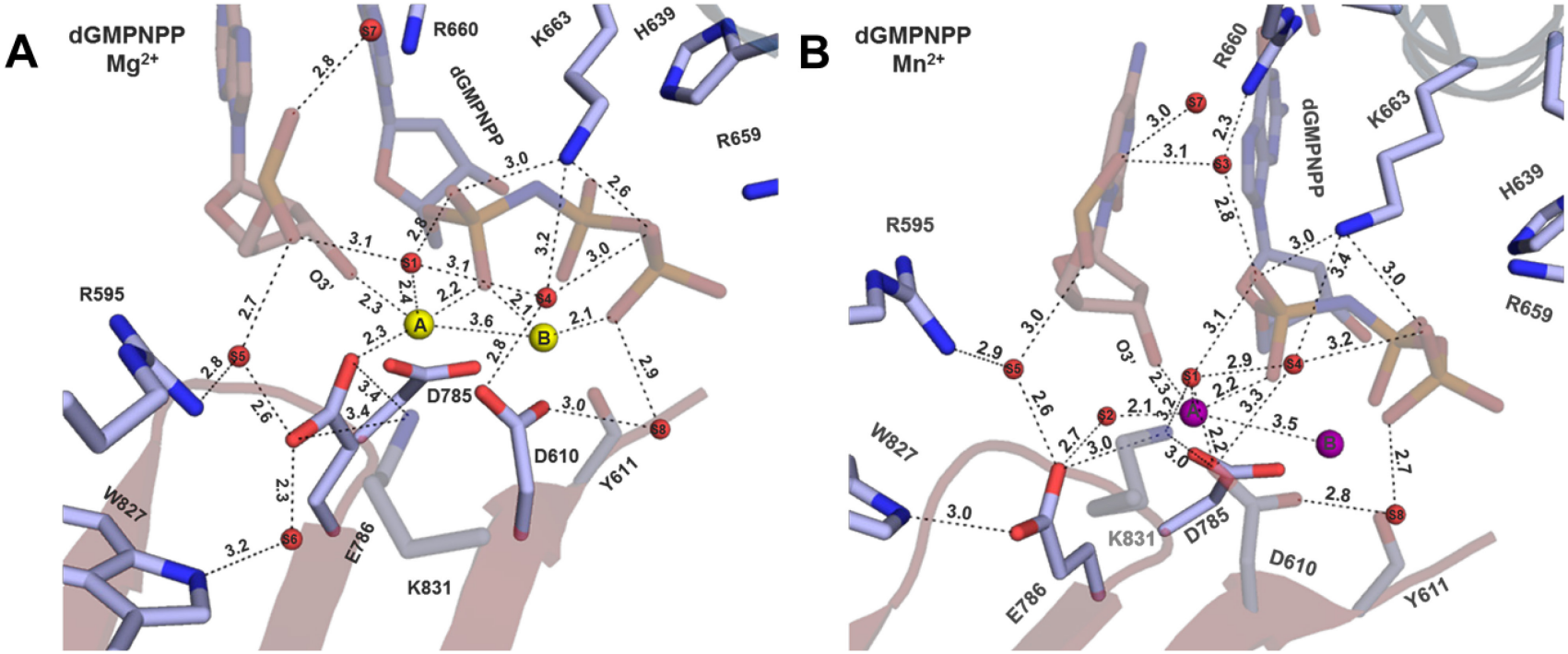


Figure 2.20 The crystal structures of Mg^{2+}- and Mn^{2+}-substituted dGMPNPP-trapped ternary complexes. The use of suboptimal metal ions drastically alters the distance of S1 to Me^{A} thereby requiring a potentially different strategy to activate and facilitate nucleotidyl transfer. Interestingly, here, the use of Mn^{2+} in place of Mg,$^{2+}$ perhaps due to coordination distances, facilitates the emergence of the catalytic water, S3. Water molecule, S1, is seen directly interacting with K831 concurrently with Mn^{A}, suggesting a potential alternative (contingency) mechanism for activation and nucleotidyl transfer. This may explain the basis of misincorporation using suboptimal metal ions. Mutations of this functionally-conserved Lys (His in some A-family DNAPs) nearly completely abolish the activity of A-family DNA polymerases. Structure in panel B) is of PDB ID: 6FBI

Higher eukaryotes, however, have evolved to possess several redundancy mechanisms for both their replicative and repair polymerases with many belonging to each but the C- and D- (bacterial and archaeal) families of polymerases, and containing as many as 3-5 members within a family, each of which with very specialized, sometimes overlapping, functions and minute but functionally significant differences from other members within the same family (71-73). This functional redundancy is one form of a mechanistic contingency that also allows for promiscuity in choice of metal-ion, Mg^{2+} or Mn^{2+}, for vertebral X-family repair polymerases (12,74,75). In fact, X-family polymerases λ and μ display preference for Mn^{2+} over Mg^{2+} (76,78) which in turn affords these polymerases a higher tolerance for incorporation of incorrect nucleotides and those containing bulky adducts.

Along those same lines of convergent evolution, *Archaea* are the only domain of life to have evolved two different DNA polymerase families, B- and D-family, independently to perform the same function in DNA replication, albeit with significantly varied degrees of fidelity (79-81). The major leading and lagging strand replicative polymerase in archaea, *polD*, is the functional equivalent of B- and C-family polymerases in other domains of life. Conversely, the existence of an A-family polymerase in all but the domain *Archaea*, that serves both replicative and repair functions, suggests that the archaeal B-family polymerase must be its functional equivalent owing to it retaining both replicative and repair functions that are seen only in A-family polymerases (82).

Similarly, a B-family polymerase from an archaeon found in deep-sea brines of the Red Sea with thermophilic and halophilic adaptations was shown to have the ability to not only utilize Mg^{2+} or Mn^{2+}, similar to eukaryotic X-family polymerases, but also utilize Zn^{2+} as a catalytic metal-ion with no loss in activity relative to Mg^{2+} (83). Thus, it appears that archaeal B-family polymerases may represent an interesting hybrid between A- and B-family polymerases that are found in all other domains of life and may offer insight into a mechanistic contingency wherein a binary reaction mechanism allows the utilization of either two or three metal-ions depending on the polymerases metal-ion preference, metal-ion availability and coordination preference, and active site geometry.

2.5 Materials and Methods

2.5.1 Materials

Oligonucleotides used in crystallization were PAGE and HPLC purified from Integrated DNA Technologies (Carlsbad, CA USA). Deoxyribonucleoside triphosphates (dNTP's) were obtained from New England Biolabs (Ipswich, MA USA). 2',3'-Dideoxyribonucleoside triphosphate (ddNTP) analogues were obtained from GE Healthcare Bio-Sciences Corporation (Piscataway, NJ USA). The non-hydrolyzable analogue of dGTP, dGMpNHpp was purchased from Jena Bioscience (Jena, Germany). Crystallization screens and reagents were obtained from Hampton Research (Aliso Viejo, CA). All other reagents were of the highest grade available.

2.5.2 Protein Expression and Purification

The coding sequence for the Klenow fragment of *Thermus aquaticus* (KlenTaq) DNA polymerase (residues 281-832) was amplified from *Thermus aquaticus* genomic DNA and cloned into a phage T7-derived expression vector to facilitate production of recombinant KlenTaq as previously described (Barnes, *et al.*). Briefly, cell pellets were lysed by sonication followed by heat denaturation and ultracentrifugation. The resultant material was precipitated with polyethyleneimine (PEI), and subsequently purified by affinity chromatography on Heparin-Sepharose, anion-exchange chromatography on Q-Sepharose, and size-exclusion chromatography on Superdex 75, yielding >99% pure KlenTaq.

2.5.3 Crystallization of dGMPNPP-Trapped Ternary Complex

KlenTaq DNAP ternary complex was formed by mixing 7.5mg/mL KTQ with preformed duplex DNA (Primer 5'-GACCACGGCGCA-3', Template 3'-CTGGTGCCGCGTCAAA-5'), and dGMPNPP in a molar ratio of 1:1.5:10. Crystals of the ternary complex were obtained by vapor diffusion against a reservoir buffer containing 0.1M Tris, pH 8, 0.2M $Mg(OAc)_2$, 0.2M $(NH_4)OAc$, 10mM TCEP and 17-20% PEG 8,000 or 18-25% PEG 3,350. Crystals were cryoprotected successively in mother liquor containing 10mM $Mg(OAc)_2$ and increasing concentrations of PEG 8,000 or PEG 3,350 up to 30% and 35%, respectively.

2.5.4 Derivation of the Open Binary Complexes

Crystals of the open, catalytically-competent binary complex crystals were obtained by soaking crystals in mother liquor lacking dGMPNPP to remove the non-hydrolyzable analogue and facilitate reversion to the enzyme's open state. Resultant crystals were cryoprotected successively in mother liquor containing 10mM $Mg(OAc)_2$, 0.1M HEPES, pH 7, and increasing concentrations of PEG 8,000 or PEG 3,350 up to 30% and 35%, respectively.

2.5.5 Time-course of Nucleotide Incorporation

Catalytically-competent binary complex crystals, now at physiological conditions, were soaked in cryoprotectant containing 5mM dGTP for 30-second intervals up to 5 minutes, 1-minute intervals from 6- to 30 minutes, 10-minute intervals from 30- to 60 minutes, and 30-minute intervals from 1- to 4 hours. For each time point crystals were flash-frozen in liquid nitrogen prior to data collection.

2.5.6 Data Collection and Processing

Diffraction data were collected at the Macromolecular Diffraction facility of the Cornell High Energy Synchrotron Source (MacCHESS) using both A-1 and F-1 beamlines (Cornell University, Ithaca, NY USA), and processed using HKL2000 (84). Further data processing was carried out using the PHENIX software suite (85).

2.5.7 Structure Determination and Refinement

The phase for the open binary complex structure was solved by molecular replacement using PHASER (86) with a modified model of the open, dideoxy-terminated binary complex of KlenTaq (PDB ID: 4KTQ) which lacks residues 630-680 that encode a portion of the fingers domain containing the O-helix. COOT (87) was used to rebuild the deleted region in the F_o-F_c map of the resultant model followed by replacement of the template 5'- and primer 3'-nucleobases. In a similar fashion, the phase for the dGMPNPP-trapped ternary complex was solved by molecular replacement with a modified model of the closed dideoxy-terminated ternary complex of KlenTaq (PDB ID: 3KTQ) lacking the incoming dNTP and Mg^A/Mg^B. The primer 3'-end, dGMPNPP, Mg^A

and Mg^B were modelled into the resultant F_o-F_c map. The final structures of the open binary complex and closed ternary complex lacking dGMPNPP and metals served as search models in molecular replacement for structure determination of subsequent time-course structures.

Structures of ternary complexes up to 5 minutes were solved by molecular replacement using the model for the open binary complex to determine the time-point at which the O-helix converts to the closed conformation followed by modelling of dGTP and Me^B/Me^A into the F_o-F_c when pertinent. Structures of subsequent time points were solved by molecular replacement using the model for the closed ternary complex and subsequent modelling of dGTP, or dGMP and PP_i, and Me^B/Me^A into the F_o-F_c when pertinent. Resultant structures were refined using PHENIX.

Accession Numbers

Coordinates and structure factors have been deposited to the Protein Data Bank. The accession codes and refinement statistics are provided in Table 2.1.

Complex	dGMPNPP	Open Binary	Metal-free dGTP	Mg^{B}:dGTP	Mg^{BA}:dGMP:PP_i	dGMP:PP_i
Data Collection						
PDB ID	TBD	TBD	TBD	TBD	TBD	TBD
Space group	$P3_121$	$P3_121$	$P3_121$	$P3_121$	$P3_121$	$P3_121$
Unit Cell Dimensions						
a, b, c (Å)	108.3 108.3 90.6	114.7 114.7 91.2	115.1 115.1 91.7	113.5 113.5 91.5	108.6 108.6 90.6	114.5 114.5 91.5
α, β, γ (°)	90, 90,120	90, 90,120	90, 90,120	90, 90,120	90, 90,120	90, 90,120
Resolution (Å)	29.56-2.145	26.38-2.47	29.25-2.3	28.87-2.453	29.64-2.0	29-2.65
I/σI	8.48 (1.14)	13.63 (1.69)	12.54 (1.77)	9.11 (1.31)	9.80 (1.30)	10.92 (1.28)
Completeness (%)	98.60 (86.70)	92.22 (85.56)	100.00 (100.00)	100.00 (100.00)	99.86 (98.63)	99.94 (99.80)
Refinement						
No. of reflections	33533	23292	31578	25374	42209	20464
R_{work}	0.1827	0.213	0.1985	0.1887	0.1745	0.1891
R_{free}	0.2447	0.2699	0.2484	0.2494	0.2244	0.2586
B factors						
Protein/DNA	43.9	66.7	68.9	57.6	37	57.7
dGMP(N/O)PP	44.3	-	69.2	55.3	35.7	59.6
Rmsds						
Bond lengths (Å)	0.012	0.011	0.011	0.011	0.01	0.009
Bond angles (°)	1.39	1.67	1.35	1.36	1.26	1.29

Table 2.1 Crystallographic data and structure refinement.

	Motif C	(d)dNTP (PO_4)				Motif A	
		(Mg^A/Mg^B)					
	E786:Mg^A	D785	α-	β-	γ-	D610	Y611:Mg^B
dGMPNPP	2.3	2.9/2.3	2.1/2.2	-/2.3	-/2.0	2.3/2.1	2.1
3KTQ	3.8	2.5/2.1	2.2/2.4	-/2.2	-/2.3	2.4/2.1	2.2
E786:S2:Mg^A	2.2						
			dGMP(PO_4)	PP_i			
Mg^{BA}:dGMP:PP_i	3.7	2.4/2.1	2.8/2.2	-/2.1	-/2.0	2.6/2.1	2.2
	E786					D610	
3KTQ/4KTQ	(+2.1)					(-2.2)	
dGMPNPP/Open	(+2.4)					(-2.2)	

Table 2.2 Relative Metal-ion Catalytic Residues Distances (Å). relative to reference dideoxy-terminated structures for open and closed complexes, and for intermediate structures from open complex to pyrophosphate release.

	Motif C		dNTP (PO_4)			Motif A	
			(Mg^A/Mg^B)				
	E786	D785	α-	β-	γ-	D610	Y611:Mg^B
dGMPNPP	2.3(Mg^A)	2.9/2.3	2.1/2.2	-/2.3	-/2.0	2.3/2.1	2.1
Open	(+0.9)	(-0.9)	-	-	-	(-3.3)	-
Metal-free dGTP	(-2.4)	(+0.9)	-	-	-	(+2.8)	-
Mg^B:dGTP	(-0.9)	-/2.2	-/2.3	-/2.5	-/2.3	-/2.4	1.9
Mg^{BA}:dGMP:PP_i	3.7	2.4/2.1	2.8/2.2	-/2.1	-/2.0	2.6/2.1	2.2
dGMP:PP_i	NC	(-0.7)/-	-	-	-	(-1.0)	-

Table 2.3 Side-chain Position and Metal-ion Distances (Å) Relative to Preceding Structure.

Acknowledgements

We thank the staff and scientists at the Macromolecular Diffraction facility of the Cornell High Energy Synchrotron Source, and the Huck Institute for the Life Sciences X-ray Crystallography facility at the Pennsylvania State University for technical assistance and support. This work was supported by NIH grant GM087350, the Stanley Person Graduate Fellowship in Biochemistry and Molecular Biology, and Kuwait University. Figures were prepared using PyMOL.

References

1. Hubscher, U., Maga, G., Spadari, S. (2002). Eukaryotic DNA polymerases. *Annu. Rev. Biochem*. **71**, 133-163.

2. Dianov, G., Lindahl, T. (1994). Reconstitution of the DNA base excision--repair pathway. *Curr. Biol*. **4**, 1069-1076.

3. Cooper, P., Hanawalt, P. (1972). Role of DNA polymerase I and the rec system in excision-repair in *Escherichia coli*. *Proc. Natl. Acad. Sci*. USA **69**, 1156–1160.

4. Joyce, C., Steitz, T. (1994). Function and structure relationships in DNA polymerases. *Annu. Rev. Biochem*. **63**, 777-822.

5. Steitz, T. (1999). DNA polymerases: structural diversity and common mechanisms. *J. Biol. Chem*. **274**, 17395-17398.

6. Doublie, S., Sawaya, M., Ellenberger, T. (1999). An open and closed case for all polymerases. *Structure* **7**, R31-35.

7. Kunkel, T., Bebenek, K. (2000). DNA replication fidelity. *Annu. Rev. Biochem*. **69**, 497-529.

8. Freudenthal, B., Beard, W., Shock, D., Wilson, S. (2013). Observing a DNA polymerase choose right from wrong. *Cell* **154**, 157-168.

9. Nakamura, T., Zhao, Y., Yamagata, Y., Hua, Y., Yang, W. (2012). Watching DNA polymerase η make a phosphodiester bond. *Nature* **487**, 196-201.

10. Basu, R., Murakami, K. (2013). Watching the bacteriophage N4 RNA polymerase transcription by time-dependent soak-trigger-freeze X-ray crystallography. *J. Biol. Chem*. **288**, 3305-3311.

11. Brautigam, C., Steitz, T. (1998). Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol*. **8**, 54-63.

12. Ramadan, K., Shevelev, I., Hubscher, U. (2004). The DNA-polymerase-X family: controllers of DNA quality. *Nat. Rev. Mol. Cell Biol.* **5**, 1038-1043.

13. Zhou, B., Pata, J., Steitz, T. (2001). Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain. *Mol. Cell* **8**, 427-437.

14. Waters, L., Minesinger, B., Wiltrout, M., D'Souza, S., Woodruff, R., Walker, G. (2009). Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiol. Mol. Biol. Rev*. **73** (1): 134–54.

15. Ling, H., Boudsocq, F., Woodgate, R., Yang, W. (2001). Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell* **107**, 91-102.

16. Kiefer, J., Mao, C., Braman, J., Beese, L. (1998). Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature* **391**, 305-307.

17. Li, Y., Korolev, S., Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J*. **17**, 7514-7525.

18. Rothwell, P., Allen, W., Sisamakis, E., Kalinin, S., Felekyan, S., Widengren, J., Waksman, G., Seidel, C. (2013). dNTP-dependent conformational transitions in the fingers subdomain of Klentaq1 DNA polymerase: insights into the role of the "nucleotide-binding" state. *J. Biol. Chem*. **288**, 13575-13591.

19. Li, Y., Waksman, G. (2001). Crystal structures of a ddATP-, ddTTP-, ddCTP-, and ddGTP-trapped ternary complex of Klentaq1: Insights into nucleotide incorporation and selectivity. *Protein Sci*. **10**, 1225-1233.

20. Arora, K., Beard, W., Wilson, S., Schlick, T. (2005). Mismatch-induced conformational distortions in polymerase beta support an induced-fit mechanism for fidelity. *Biochemistry* **44**, 13328-13341.

21. Showalter, A., Tsai, M. (2002). A reexamination of the nucleotide incorporation fidelity of DNA polymerases. *Biochemistry* **41**, 10571-10576.

22. Joyce, C., Potapova, O., DeLucia, A., Huang, X., Basu, V., Grindley, N. (2008). Fingers-closing and other rapid conformational changes in DNA Polymerase I (Klenow Fragment) and their role in nucleotide selectivity. *Biochemistry* **47**, 6103-6116.

23. Lee, I., Berdis, A. (2010). Non-natural nucleotides as probes for the mechanism and fidelity of DNA polymerases. *Biochem. Biophys. Acta.* **1804**, 1064-1080.

24. Miyaki, M., Murata, I., Osabe, M., Ono, T. (1977). Effect of metal cations on misincorporation by *E. coli* DNA polymerases. *Biochem. Biophys. Res. Commun.* **77**, 854–860.

25. Goodman, M., Keener, S., Guidotti, S., Branscomb, E. (1983). On the enzymatic basis for mutagenesis by manganese. *J. Biol. Chem.* **258**, 3469–3475.

26. Chin, Y., Snow, E., Cohen, M., Christie, N. (1994). The effect of divalent nickel (Ni^2) on *in vitro* DNA replication by DNA polymerase α. *Cancer Res.* **54**, 2337–2341.

27. Bock, C., Katz, A., Markham, G., Glusker, J. (1999). Manganese as a replacement for magnesium and zinc: functional comparison of the divalent ions. *J. Am. Chem. Soc.* **121**, 7360–7372.

28. Dube, D., Loeb, L. (1975). Manganese as a mutagenic agent during *in vitro* DNA synthesis. *Biochem. Biophys. Res. Commun*. **67**, 1041–1046.

29. Jin, Y., Clark, A., Slebos, R., Al-Refai, H., Taylor, J., Kunkel, T., Resnick, M., Gordenin, D. (2003). Cadmium is a mutagen that acts by inhibiting mismatch repair. *Nat. Genet.* **34**, 326 –329.

30. Hays, H., Berdis, A. (2002). Manganese substantially alters the dynamics of translesion DNA synthesis. *Biochemistry* **41**, 4771– 4778.

31. Lee, H. R., Wang, M., Konigsberg, W. (2009). The reopening rate of the fingers domain is a determinant of base selectivity for RB69 DNA polymerase. *Biochemistry* **48**, 2087–2098.

32. Vashishtha, A., Konigsberg, W. (2016). Effect of different divalent cations on the kinetics and fidelity of RB69 DNA polymerase. *Biochemistry* **55**, 2661–2670.

33. Dahlberg, M., Benkovic, S. (1991). Kinetic mechanism of DNA polymerase I (Klenow fragment): identification of a second conformational change and evaluation of the internal equilibrium constant. *Biochemistry* **30**, 4835–4843.

34. Nowotny, M., Gaidamakov, S., Crouch, R., Yang, W. (2005). Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **121**, 1005–1016.

35. Derbyshire, V., Grindley, N., Joyce, C. (1991). The 30-50 exonuclease of DNA polymerase I of *Escherichia coli*: contribution of each amino acid at the active site to the reaction. *EMBO J*. **10**, 17–24.

36. Beese, L., Steitz, T. (1991). Structural basis for the 30-50 exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J*. **10**, 25–33.

37. Kim, E., Wyckoff, H. (1991). Reaction mechanism of alkaline phosphatase based on crystal structures. Two-metal ion catalysis. *J. Mol. Biol*. **218**, 449–464.

38. Gangurde, R., Modak, M. (2002). Participation of active-site carboxylates of *Escherichia coli* DNA polymerase I (Klenow fragment) in the formation of a prepolymerase ternary complex. *Biochemistry* **41**, 14552–14559.

39. Polesky, A., Dahlberg, M., Benkovic, S., Grindley, N., Joyce, C. (1992). Side chains involved in catalysis of the polymerase reaction of DNA polymerase I from *Escherichia coli*. *J. Biol. Chem.* **267**, 8417–8428.

40. Doublie, S., Tabor, S., Long, A., Richardson, C., Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2Å resolution. *Nature* **391**, 251–258.

41. Franklin, M., Wang, J., Steitz, T. (2001). Structure of the replicating complex of a Pol α family DNA polymerase. *Cell* **105**, 657–667.

42. Osumi-Davis, P., de Aguilera, M., Woody, R., Woody, A. (1992). Asp537, Asp812 are essential and Lys631, His811 are catalytically significant in bacteriophage T7 RNA polymerase activity. *J Mol Biol* **226**, 37-45.

43. Yang, G., Franklin, M., Li, J., Lin, T., Konigsberg, W. (2002). Correlation of the kinetics of finger domain mutants in RB69 DNA polymerase with its structure. *Biochemistry* **41**, 2526–2534.

44. Pandey, V., Kaushik, N., Modak, M. (1994). Role of lysine 758 of *Escherichia coli* DNA polymerase I as assessed by site-directed mutagenesis. *J Biol Chem* **269**, 13259–13265.

45. Yin, Y., Steitz, T. (2004). The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell* **116**, 393–404.

46. Tunitskaya, V., Kochetkov, S. (2002). Structural-functional analysis of bacteriophage T7 RNA polymerase. *Biochemistry (Moscow)* **67**, 1124-1135.

47. Sarafianos, S., Clark, A., Das, K., Tuske, S., Birktoft, J., Ilankumaran, P., Ramesha, A., Sayer, J., Jerina, D., Boyer, P., Hughes, S., Arnold, E. (2002). Structures of HIV-1 reverse transcriptase with pre- and post-translocation AZTMP-terminated DNA. *EMBO J.* **21**, 6614-6624.

48. Patel, P., Jacobo-Molina, A., Ding, J., Tantillo, C., Clark, A., Raag, R., Nanni, R., Hughes, S., Arnold, E. (1995). Insights into DNA polymerization mechanisms from structure and function analysis of HIV-1 reverse transcriptase. *Biochemistry* **34**, 5351-5363.

49. Kottur, J., Nair, D. (2018). Pyrophosphate hydrolysis is an intrinsic and critical step of the DNA synthesis reaction. *Nucleic Acids Res.* **46**, 5875-5885.

50. Brody, R., Frey, P. (1981). Unambiguous determination of the stereochemistry of nucleotidyl transfer catalyzed by DNA polymerase I from *Escherichia coli*. *Biochemistry* **20**, 1245–1252.

51. Burgers, P., Eckstein, F. (1979). Stereochemistry of internucleotide bond formation by polynucleotide phosphorylase from *Micrococcus luteus*. *Biochemistry* **18**, 450–454.

52. Mizuuchi, K., Nobbs, T., Halford, S., Adzuma, K., Qin, J. (1999). A new method for determining the stereochemistry of DNA cleavage reactions: application to the *SfiI* and *HpaII* restriction endonucleases and to the MuA transposase. *Biochemistry* **38**, 4640–4648.

53. Steitz, T., Steitz, J. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6498-502.

54. Westover, K., Bushnell, D., Kornberg, R. (2004). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* **119**, 481–489.

55. Stahley, M., Strobel, S. (2005). Structural evidence for a two metal-ion mechanism of group I intron splicing. *Science* **309**, 1587–1590.

56. Steiniger-White, M., Rayment, I., Reznikoff, W. (2004). Structure/function insights into Tn5 transposition. *Curr. Opin. Struct. Biol.* **14**, 50–57.

57. Chevalier, B., Sussman, D., Otis, C., Noel, A., Turmel, M., Lemieux, C., Stephens, K., Monnat, R., Stoddard, B. (2004). Metal-dependent DNA cleavage mechanism of the I-CreI LAGLIDADG homing endonuclease. *Biochemistry* **43**, 14015–14026.

58. Horton, N., Perona, J. (2004). DNA cleavage by EcoRV endonuclease: two metal ions in three metal ion binding sites. *Biochemistry* **43**, 6841–6857.

59. Kropp, H., Betz, K., Wirth, J., Diederichs, K., Marx, A. (2017). Crystal structures of ternary complexes of archaeal B-family DNA polymerases. *PLoS One* **12**(12), e0188005.

60. Xia, S., Wang, M., Blaha, G., Konigsberg, W., Wang, J. (2011). Structural insights into complete metal ion coordination from ternary complexes of B family RB69 DNA polymerase. *Biochemistry* **50**, 9114–9124.

61. Pelletier, H., Sawaya, M., Wolfle, W., Wilson, S., Kraut, J. (1996). A structural basis for metal ion mutagenicity and nucleotide selectivity in human DNA polymerase β. *Biochemistry* **35**, 12762–12777.

62. Harding, M. (2006). Small revisions to predicted distances around metal sites in proteins. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 678–682.

63. Snow, E., Xu, L., Kinney, P. (1993). Effects of nickel ions on polymerase activity and fidelity during DNA replication *in vitro*. *Chem. Biol. Interact* **88**, 155–173.

64. Sirover, M., Dube, D., Loeb, L. (1979). On the fidelity of DNA replication: metal activation of *Escherichia coli* DNA polymerase I. *J. Biol. Chem.* **254**, 107–111.

65. Sirover, M., Loeb, L. (1976). Infidelity of DNA synthesis *in vitro*: screening for potential metal mutagens or carcinogens. *Science* **194**, 1434–1436.

66. Sirover, M., Loeb, L. (1976). Metal-induced infidelity during DNA synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **73**, 2331–2335.

67. Sirover, M., Loeb, L. (1976). Metal activation of DNA synthesis. *Biochem. Biophys. Res. Commun.* **70**, 812–817.

68. Seal, G., Shearman, C., Loeb, L. (1979). On the fidelity of DNA replication: studies with human placenta DNA polymerases. *J. Biol. Chem.* **254**, 5229–5237.

69. Irimia, A., Zang, H., Loukachevitch, L., Eoff, R. , Guengerich, F., Egli, M. (2006). Calcium is a cofactor of polymerization but inhibits pyrophosphorolysis by the *Sulfolobus solfataricus* DNA polymerase Dpo4. *Biochemistry* **45**, 5949–5956.

70. Pelletier, H., Sawaya, M., Wolfle, W., Wilson, S., Kraut, J. (1996). A structural basis for metal ion mutagenicity and nucleotide selectivity in human DNA polymerase β. *Biochemistry* **35**, 12762–12777.

71. Benedict, C., Gilfillan, S., Thai, T., Kearney, J. (2000). Terminal deoxynucleotidyl transferase and repertoire development. *Immunol. Rev.* **175**, 150-7.

72. Bertocci, B., De Smet, A., Weill, J., Reynaud, C. (2006). Nonoverlapping functions of DNA polymerases Mu, Lambda, and Terminal deoxynucleotidyltransferase during immunoglobulin V(D)J recombination *in vivo*. *Immunity* **25**, 31-41.

73. Di Santo, R., Maga, G. (2006). Human terminal deoxynucleotidyl transferases as novel targets for anticancer chemotherapy. *Curr. Med. Chem.* **13**, 2353-68.

74. Moon, A., Garcia-Diaz, M., Batra, V., Beard, W., Bebenek, K. Kunkel, T., Wilson, S., Pedersen, L. (2007). The X family portrait: structural insights into biological functions of X family polymerases. *DNA Repair (Amst.)* **6**, 1709-25.

75. Beard, W., Wilson, S. (2006). Structure and mechanism of DNA polymerase β. *Chem. Rev.* **106**, 361-82.

76. Blanca, G., Shevelev, I., Ramadan, K., Villani, G., Spadari, S. Hubscher, U., Maga, G. (2003). Human DNA polymerase λ diverged in evolution from DNA polymerase β toward specific Mn^{++} dependence: a kinetic and thermodynamic study. *Biochemistry* **42**, 7467-76.

77. Ruiz, J., Dominguez, O., Lain de Lera, T., Garcia-Diaz, M., Bernard, A., Blanco, L. (2001). DNA polymerase mu, a candidate hypermutase? *Philos. Trans, R Soc Lond B Biol Sci.* **356**, 99-109.

78. Zhang, Y., Wu, X., Yuan, F., Xie, Z. (2001). Highly frequent frameshift DNA synthesis by human DNA polymerase μ. *Mol. Cell. Biol.* **21**, 7995-8006.

79. Daimon, K., Ishino, S., Imai, N., Nagumo, S., Yamagami, T., Matsukawa, H., Ishino. Y. (2018). Two family B DNA polymerases from *Aeropyrum pernix*, based on revised translational frames. *Front Mol Biosci.* **5**:37.

80. Abellon-Ruiz, J., Walderon, K., Connolly, B. (2016). *Archaeoglobus fulgidus* DNA polymerase D- A zinc-binding protein inhibited by hypoxanthine and uracil. *J Mol Biol*. **428**(14), 2805-13.

81. Huber, C., Marx, A. (2017). Variants of sequence family B *Thermococcus kodakaraensis* DNA polymerase with increased mismatch extension selectivity. *PLoS One* **12**(8), e0183623.

82. Kornberg, A., Baker, T. (1992). *DNA Replication, 2nd edition*. New York: W.H. Freeman..

83. Takahashi, M., Takahashi, E., Joudeh, L., Marini, M., Gobind Das, G., Elshenawy, M., Akal, A., Sakashita, K., Alam, I., Tehseen, M., Sobhy, M., Stingl, U., Merzaban, J., Di Fabrizio, E., Hamdan, S. (2018). Dynamic structure mediates halophilic adaptation of a DNA polymerase from the deep-sea brines of the Red Sea. *FASEB J.* **32**, 3346-3360.

84. Otwinowski, Z., Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*. **276**, 307–326.

85. Adams, P., Afonine, P., Bunkóczi, G., Chen, V., Davis, I., Echols, N., Headd, J., Hung, L., Kapral, G., Grosse-Kunstleve, R., McCoy, A., Moriarty, N., Oeffner, R., Read, R., Richardson, D., Richardson, J., Terwilliger, T., Zwart, P. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221.

86. McCoy, A., Grosse-Kunstleve, R., Adams, P., Winn, M., Storoni, L., Read, R. (2007). Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674.

87. Emsley, P., Lohkamp, B., Scott, W., Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501.

Chapter 3

Insights into Fidelity in A-family DNA Polymerases

3.1 Abstract.....105

3.2 Introduction.....106

3.3 Results.....111

3.3.1 O-helix Conformational Change is Independent of and Unaffected by Metal-ions in Matched Ternary Complexes.....111

3.3.2 Binding of Non-Mg^{2+} Divalent Metal-ions has Little Effect on Closed, Matched Ternary Complexes.....113

3.3.3 O-helix Rapidly Reverts to the Open Conformation after Sampling the Active Site in the Presence of Mg^{2+} in Mismatched Ternary Complexes.....113

3.3.4 O-helix Remains Closed in Chelated Mismatched Ternary Complexes....116

3.3.5 Mn^{2+}-substituted Mismatched Complexes Remain in the Closed Conformation, Poised for Catalysis.....116

3.4 Discussion.....119

3.4.1 Insights into the Structural and Mechanistic Determinants of Fidelity.....119

3.4.2 Structural Aspects Delineated from Examining Fidelity Intrinsic to A-family DNAPs in Matched Complexes.....120

3.4.3 Structural Aspects Delineated from Examining Fidelity Intrinsic to A-family DNAPs in Mismatched Complexes.....120

3.4.4 Mechanisms of O-helix Reversion in Nucleotide Discrimination.....124

3.4.4.1 Charge-driven Mechanism.....124

3.4.4.2 Metal-ion-driven Mechanism.....124

3.4.5 Incomplete Structural Details.....125

3.5 Materials and Methods.....126

References.....130

3.1 Abstract

DNA-dependent DNA and RNA polymerases sustain fidelity in nucleotide incorporation via intricate interplay between polymerase active-site geometry, physicochemical variables within the active-site environment, and nature of incoming substrate. The exact mechanism by which a DNA polymerase sustains fidelity – that is, to preferentially incorporate a correct nucleotide, and differentially exclude an incorrect nucleotide during DNA replication – is not fully understood. We provide here a series of time-resolved X-ray crystallographic structures of a catalytically-competent binary complex of the high-fidelity A-family DNA polymerase from *Thermus aquaticus* and incoming nucleotides, correct or incorrect, in the presence or absence of various metal ions. We show that fidelity is a two-stage, two-component, synergistic process that relies on both structural elements within the polymerase active-site and nature of metal-ions destined for chemistry to confer onto the polymerase the ability to accept or reject an incoming nucleotide with varying degrees of discrimination. Results from this work strongly favour a rate-limiting transition state model governing fidelity in A-family DNA polymerases.

3.2 Introduction

DNA polymerase (DNAP) fidelity results in the accurate replication of a DNA template and is a requirement to prevent, or at least reduce, mutations in the newly synthesized DNA strand that result from misincorporation. The proposed mechanism of nucleotide incorporation follows five steps (1-3). First, a binary complex is formed by the binding of DNA polymerase to its substrate double-stranded DNA. A correct incoming deoxyribonucleoside triphosphate (dNTP) binds to the binary complex, forming a ternary complex that also includes two divalent metal ions (Mg^{2+}) which mediate formation of a pentacovalent coordination complex between the triphosphate moiety of the incoming nucleotide, two carboxylate-donating Asp residues from the evolutionarily conserved motifs A and C (Fig. 3.1), and the 3'-oxyanion (O3') of the primers 3'-terminus. A conformational change occurs in the newly formed ternary complex from open to closed conformation, followed by a chemistry step leading to phosphodiester bond formation between the α-phosphate of the incoming dNTP and the 3'-oxyanion at the 3'-end of the growing primer strand. Finally, inorganic pyrophosphate is released as a byproduct, and the polymerase reverts to the open conformation to facilitate translocation onto the next templating nucleobase. Fidelity of nucleotide incorporation is primarily affected by the degree of discrimination during nucleotide binding, conformational change, and chemistry (4).

	Motif C						Motif A				
T. aquaticus	783	V	H	D	E	L	609	L	D	Y	S
E. coli	880	V	H	D	E	L	704	A	D	Y	S
D. radiodurans	908	V	H	D	E	L	734	A	D	Y	S
S. pneumoniae	829	V	H	D	E	I	652	S	D	Y	S
G. stearothermophilus	828	V	H	D	E	L	652	A	D	Y	S
H. influenzae	882	V	H	D	E	L	707	A	D	Y	S
M. tuberculosis	651	V	H	D	E	L	679	A	D	Y	S
Bacteriophage T7	855	V	H	D	E	I	473	I	D	A	S
S. cervesiae Polγ	890	I	H	D	E	I	692	A	D	V	D
H. sapiens Polγ	1133	I	H	D	E	V	889	A	D	V	D

Figure 3.1 Sequence conservation of motifs A and C in A-family DNA polymerases. Sequence alignment of conserved motifs A and C in A-family DNA and RNA polymerases responsible for coordinating and chelating both binding and catalytic metal-ions Mg^B and Mg^A, respectively, during O3' activation and phosphoryl transfer. Residues whose side-chains directly coordinate metal-ions are in solid red boxes, and those in which backbone carbonyl oxygens coordinate metal-ions are in dashed red boxes.

Two models are proposed to explain the potential mechanisms underlying the significant differences in fidelity by DNAPs. An "induced-fit" model posits that polymerase fidelity would be the result of conformational coupling in which the free energy change from correct nucleotide binding is used to drive a rate-limiting conformational change in the polymerase that precedes the chemistry step (5). Nucleotide selection in this model would be a two-step process in which ground state binding of the correct nucleotide, which would be more favourable compared to the incorrect one, is followed by the formation of a catalytically-competent ternary complex driven by the favourable energy produced from ground state binding. The second "rate-limiting transition state" model posits that polymerase fidelity is attributed to a rate-limiting chemistry step in which the free energy change in the formation of transition state species of correct and incorrect nucleotides dictates whether the polymerase will proceed to chemistry or revert to an open conformation and release the bound nucleotide (6). In this model, binding of the correct nucleotide is followed by a rapid conformational change and a subsequent slower chemistry step. Evidence in support of both models exists, and it appears that different polymerases, and different families, adopt different mechanisms to sustain fidelity (7,8).

The fingers subdomain in A-family polymerases carries an O-helix (Fig. 3.2), so designated based on the original structure of *E. coli* DNA polymerase I, which is largely responsible for the high fidelity in this family of polymerases. In the open conformation of the binary complex the O-helix of the fingers domain is seen rotated outwards by approximately 46°, similar to the apo form of the enzyme, and showing a conserved Tyr residue stacking against the 5'-templating base and mediating "base flipping" of the neighboring 5'-nucleobase of the template strand (9). This stacking is said to facilitate better base pairing with the incoming nucleotide.

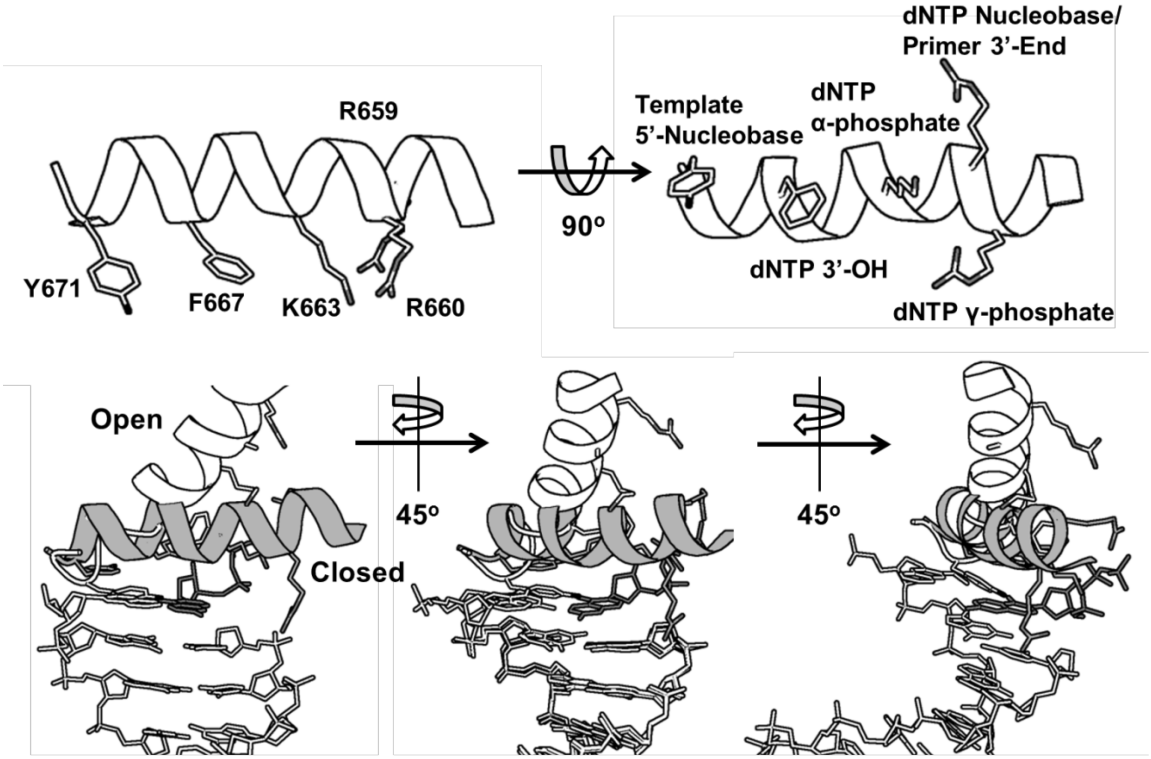


Figure 3.2 Anatomy and movement of the O-helix of the Klenow Fragment of *Thermus aquaticus* DNA polymerase. The top panel shows evolutionarily conserved, catalytically relevant side-chains in the O-helix of the fingers subdomain of KlenTaq DNA polymerase and the moieties with which they interact in the closed conformation of a matched ternary complex. The bottom panel shows the relevant downwards and inwards rotation of the O-helix during closure in a matched ternary complex. Structures rendered using PyMOL.

As the incoming nucleotide binds to form a ternary complex, the O-helix rotates inwards by approximately 40° to form a complete active site that now includes a network of residues afforded by the O-helix that further interact with, and stabilize, the incoming nucleotide (9). Three distinct networks of amino acid motifs (A, B, and C) are characteristic of the *polA* family, with motif B encoding the O-helix. Evidence suggests that the movement of the O-helix, due to its extent, is likely the rate-limiting step for catalysis in this family of polymerases, as such a large conformational change is likely to be energetically costly. However, previous studies using Fluorescence Resonance Energy Transfer (FRET) analysis to study catalysis in *KlenTaq* DNA polymerase suggests that closure of the O-helix is in fact not the rate-limiting step, but precedes such a step, which is believed to be the chemistry step (10). Thus, two lines of evidence now support both models for fidelity and conclusive evidence to support or negate either is yet to be found.

Here we utilize a time-resolved X-ray crystallographic approach to examine the structural and mechanistic determinants of fidelity in correct and incorrect ternary complexes of the high-fidelity A-family DNA polymerase from *Thermus aquaticus* in the presence or absence of various metal ions.

3.3 Results

3.3.1 O-helix Conformational Change is Independent of and Unaffected by Metal-ions in Matched Ternary Complexes

To examine the structural and mechanistic determinants of fidelity in A-family DNA polymerases (DNAPs) the exonuclease-deficient Klenow (large) fragment of *Thermus aquaticus* DNAP was crystallized in a dC:dGMPNPP-trapped closed ternary complex (TC), as in Chapter 2 and the resultant crystals were subsequently soaked in cryoprotectant lacking the αβ non-hydrolyzable dGTP analogue, dGMPNPP, and several crystals were used to collect x-ray diffraction data to monitor dGMPNPP diffusion and consequent reversion to the open binary complex (BC). These crystals are robust and allow for a host of physicochemical manipulations with little effect on crystal quality. We attempted to exploit this aspect by first stripping BC crystals of all divalent metal-ions through successive washes in mother liquor containing cryoprotectant and increasing concentrations of EDTA (5-25mM). To ensure that the open binary complex crystals were completely chelated, crystals were soaked in mother liquor as before but containing 5mM dGTP, the correct dNTP ($dNTP_c$), and the resultant ternary complexes were monitored up to 60 minutes for chemistry, densities of $Mg^{B/A}$, pyrophosphate (PP_i), and phosphodiester bond formation between dGTP and the 3'-oxyanion (O3').

In line with our previous findings of spatiotemporal separation of substrate entry and metal ligation, nucleotide binding in chelated matched complexes (ChMC's) was sufficient to trigger O-helix conformational change to the closed ternary complex in the absence of metals. In fact, the O-helix remains closed even at 60 minutes beyond diffusion of dGTP into the active site – which is well beyond the amount of time necessary for chemistry to occur in an otherwise catalytically-competent ternary complex – with no density for metal-ions or a newly formed phosphodiester bond (Fig. 3.3).

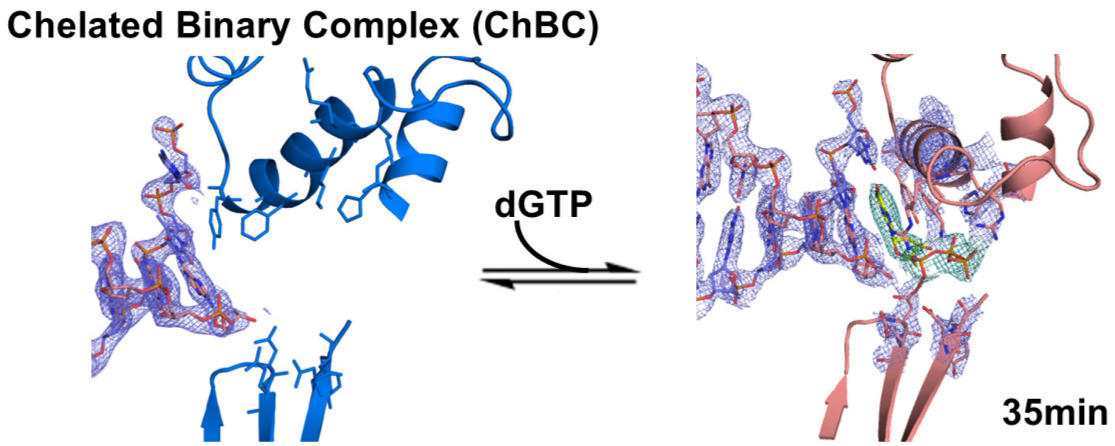


Figure 3.3 Conformational changes in chelated dC:dGTP ternary complexes.
F_o-F_c map of chelated binary and ternary dC and dC:dGTP complexes, respectively. The chelated dC:dGTP ternary complex illustrates that nucleotide binding alone is sufficient to drive conformational change to the closed ternary complex in the absence of metal-ions and is maintained in the closed conformation far beyond the point at which chemistry would occur, and further reinforces the spatiotemporal separation of nucleotide binding and metal-ion entry and subsequent ligation. Structures rendered using PyMOL.

3.3.2 Binding of Non-Mg^{2+} Divalent Metal-ions has Little Effect on Closed, Matched Ternary Complexes

We then attempted to examine the effect of incorrect metal-ions on O-helix conformation by introduction of a bulky, constrained metal-ion (Ca^{2+}) into the chelated matched ternary complex and found that, despite slight distortion in the O-helix, it remains in the closed conformation with Ca^{2+} occupying the active site and ligating dGTP to motifs A and C. Thus, in matched complexes, the absence of metals, or the presence of incorrect metals, alone is insufficient in altering the pathway from ground-state $dNTP_c$-binding to closed ternary complex.

3.3.3 O-helix Rapidly Reverts to the Open Conformation after Sampling the Active Site in the Presence of Mg^{2+} in Mismatched Ternary Complexes

The aforementioned findings that metals contribute little to the physicochemical environment in matched ternary complexes, due to proper alignment and geometry of the incoming $dNTP_c$ after proper base-stacking against its neighbouring nucleobase and hydrogen-bonding with its orthogonal partner, suggests that metals may play a structural role in fidelity that is more pronounced in the presence of an incorrect dNTP ($dNTP_{ic}$) that has been shown in crystal structures to be misaligned in mismatched complexes. To investigate this possible role that metal-ions may play in nucleotide discrimination, we introduced incorrect nucleotides to Mg^{2+}-substituted binary complexes, which will result in dC-mismatched complexes containing dATP, dTTP, 5-iodo-dCTP, or 5-iodo-dUTP, and monitored O-helix conformational changes upon nucleotide binding in a similar fashion to monitoring nucleotide binding in matched complexes.

In all instances, at 30 seconds, the incorrect nucleotides appear in the active site with clear densities in the F_o-F_c difference maps using the open binary complex model with an intermediate conformation of the O-helix between the open and closed state (Fig. 3.4), similar to the mechanistic and structural trajectory seen for the matched complex. Following the reaction to 1 minute, the structures and densities for the mismatched complexes are nearly identical to the matched complexes, with the exception of slight misalignment of the incorrect nucleotide in the active site, showing a clearly defined closed O-helix and absence of any metals ligating the triphosphate moieties to motifs A

or C. However, once the reaction is allowed to proceed to 2 minutes –the point in time at which the binding metal, Mg^B, enters the active site and ligates motif C to the triphosphate moieties in matched complexes– all mismatched complexes are seen reverted to the open ternary complex with residual, but dissipating, density for the incorrect nucleotides at the active site, suggesting an attempt, upon Mg^B ligation, at forced geometric rearrangement or realignment of the incorrect nucleotides in the active site, resulting in steric clashes and ultimately forcing the O-helix to revert to the open conformation to release the incorrect nucleotide.

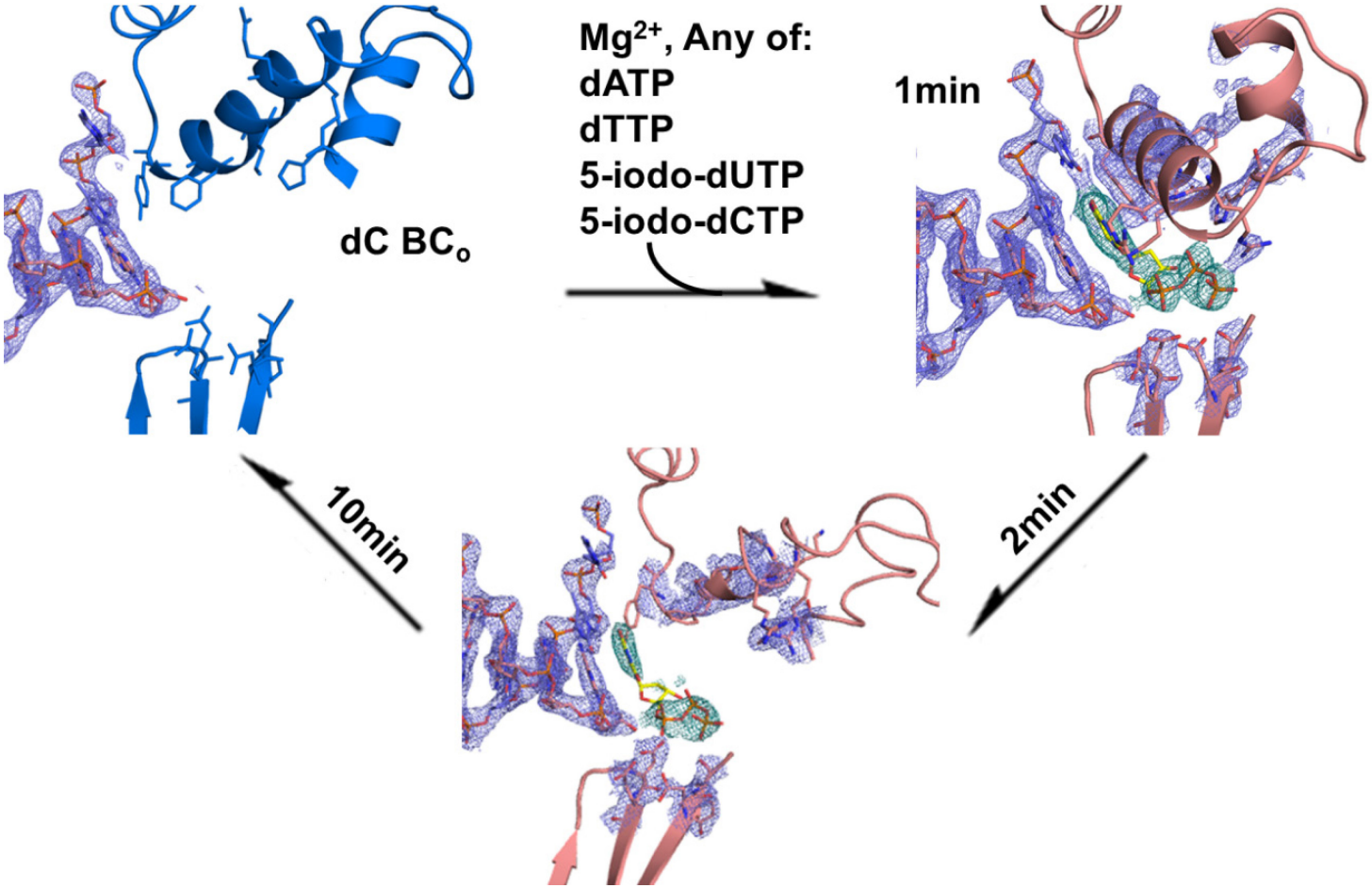


Figure 3.4 Conformational changes in Mg^{2+}-substituted mismatched complexes. F_o-F_c maps obtained after introduction of dATP, dTTP, 5-iodo-dUTP, or 5-iodo-dCTP to Mg^{2+}-substituted binary complexes with a dC-templating base results in similar conformational changes to metal-less ternary complexes obtained prior to Mg^B entry. The resultant ternary complexes undergo conformational change to the closed complex after 1-minute post-$dNTP_{ic}$ soaks, then revert to the open ternary complex by 2-minutes to facilitate release of the bound $dNTP_{ic}$, which diffuses out of the active site and is no longer visible in F_o-F_c maps obtained at 10-minutes. Structures rendered using PyMOL.

3.3.4 O-helix Remains Closed in Chelated Mismatched Ternary Complexes

It became clear from the previous results that metal ions, and choice thereof, play a significant role in fidelity through concerted structural interactions with $dNTP_{ic}$ in mismatched ternary complexes which consequently force steric hindrance with the O-helix that results in reversion to the open ternary complex to release the incorrect nucleotide. To support an attempt at forced re-alignment of the incorrect nucleotide in the active site that results in reversion to the open ternary complex we examined mismatched complexes in the absence of metals by introducing incorrect nucleotides to chelated binary complex crystals.

In these mismatched complex crystals, the absence of metals results in stable closed ternary complexes, identical to the 1 minute closed mismatched complex, that do not revert to the open conformation even after 30-60 minutes post-introduction of an incorrect nucleotide, suggesting that the misalignment of the incorrect nucleotide can be tolerated in the active site until a suitable divalent metal-ion ligates the nucleotides triphosphate moiety and attempts to realign it, resulting in forced structural clashes that destabilize the closed ternary complex and revert the O-helix to the open conformation for release of the incorrect nucleotide.

3.3.5 Mn^{2+}-substituted Mismatched Complexes Remain in the Closed Conformation, Poised for Catalysis

Thus far, our observations support a structural role for Mg^B, as was hypothesized in the past, but go further to support the claim that Mg^B also participates in orienting the incoming nucleotide to facilitate proper pre-catalytic geometry, base-stacking, and hydrogen-bonding with a rigid bidentate coordination. The rigidity of Mg^{2+} and its consequent coordination number requirement during ligation may be the driving force behind attempts at forced realignment of the incorrect nucleotide in the active site that ultimately clashes with the O-helix and results in reversion to the open complex for release of said nucleotide.

If this is indeed true then maintaining the metal-ion characteristics of Mg^{2+} with its more flexible counterpart, Mn^{2+}, may prevent the steric clashes that occur during

attempts at realignment as Mn^{2+} adopts different oxidation states, coordination numbers, and geometry to facilitate suboptimal, but stable, Mn^{2+}-ligated ternary complexes poised for catalysis. Use of non-ideal divalent metal-ions such as Mn^{2+} is historically established and experimentally verified as the pathway giving rise to "mutator" polymerases that are notorious for their lack of fidelity.

We sought to examine this with one of the mismatched complexes above and chose a dC:dATP mismatch to study the effects that Mn^{2+} may have in these complexes relative to its Mg^{2+}-substituted counterpart. This mismatch complex preserves the geometric constraints of the active site that would otherwise be present if the correct incoming nucleotide, dGTP, where to enter the active site. Chelated binary complex crystals were soaked in mother liquor containing 10mM $MnCl_2$ and 5mM dATP and incubated as previously described. Unlike its Mg^{2+}-substituted counterpart, this aberrant mismatched complex remained in the closed conformation up to 15 minutes (Fig. 3.5), the extent to which the reaction was monitored, with a clear density for dATP in the active site. This observation further supports claims of a structural role for Me^B in alignment and orientation of incoming nucleotides in the active site, and that a steric, structural clash of the incorrect nucleotide against the O-helix, brought about by rigid Mg^{2+} ligation and coordination, is a likely component in the pathway and mechanism by which A-family DNAPs sustain fidelity.

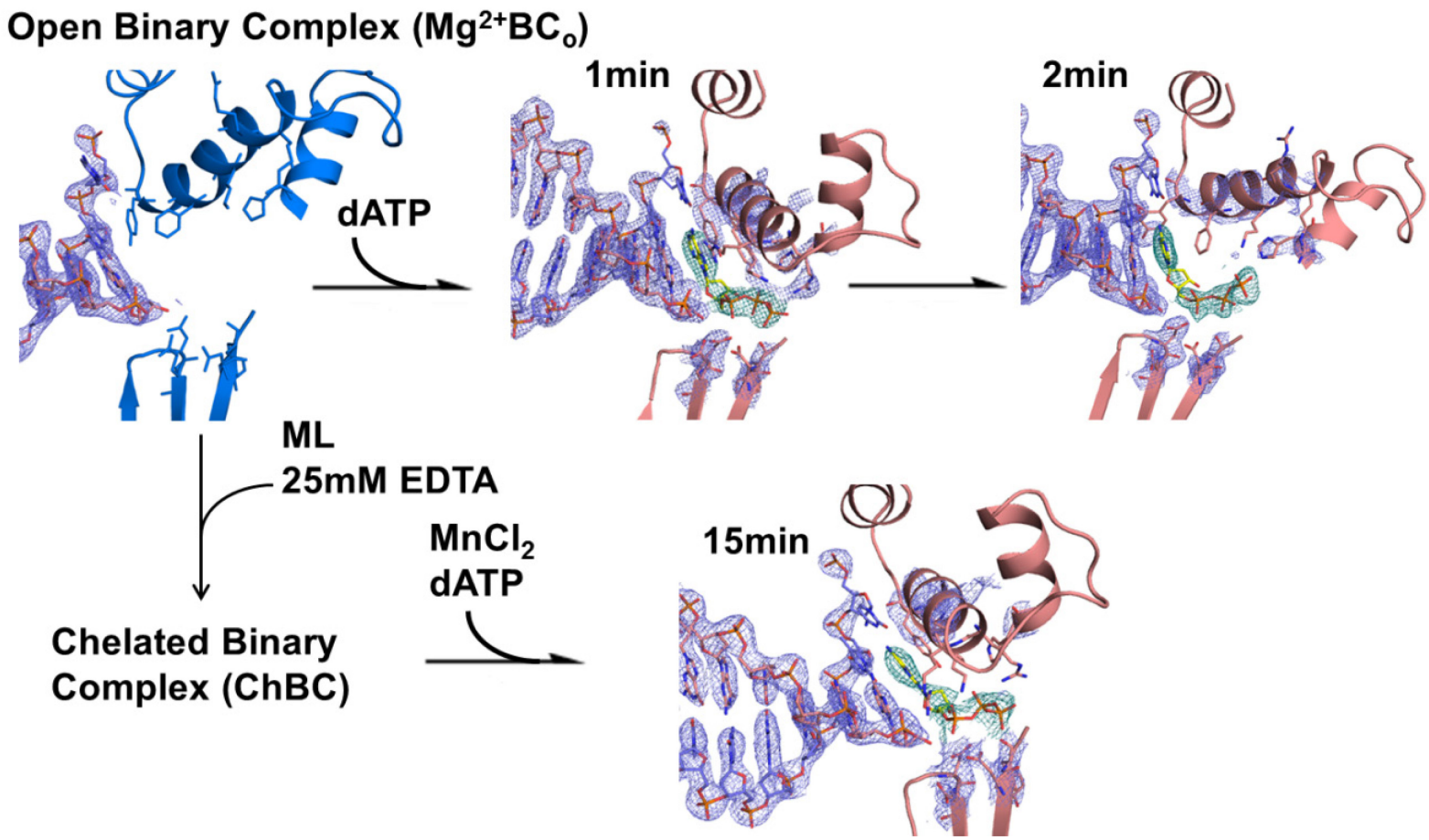


Figure 3.5 Conformational changes in Mg^{2+}- and Mn^{2+}-substituted mismatched complexes. F_o-F_c maps of Mg^{2+}- and Mn^{2+}-substituted binary complexes and the consequent conformational changes after introduction of dATP opposite a dC-templating base. Unlike their Mg^{2+}-substituted, Mn^{2+}-substituted mismatched complexes remain in the closed conformation possibly poised for chemistry beyond the length of time at which Mg^{2+}-substituted mismatched complexes revert to the open conformation. Structures rendered using PyMOL.

3.4 Discussion

3.4.1 Insights into the Structural and Mechanistic Determinants of Fidelity

By utilizing a robust platform for *in crystallo* single-turnover kinetics in A-family DNAPs, a plethora of physicochemical properties can be altered and fine-tuned to examine various aspects of nucleotide incorporation and determinants of fidelity intrinsic to this family of polymerases. This study has contributed a great deal of insight into how fidelity is sustained in A-family DNAPs by facilitating the dissection of the role of individual components within a ternary complex during catalysis.

The previously established, and generally accepted, catalytic pathway of nucleotide incorporation, as it pertains to A-family DNAPs, is devoid of information derived from time-resolved structural data that provides evidence in support of spatiotemporal separation of ground-state nucleotide binding and conformational change, and metal entry to the active site and sequential ligation to the nucleotides triphosphate moiety, motifs A and C and the binding (Mg^B) and catalytic (Mg^A) metal-ions. These aspects are evident in results obtained from this work and results from previous studies described in Chapter 2.

The results presented here suggest that fidelity is a two-stage, two-component process, one governed by the O-helix and the other by metal entry and choice thereof. The discrete intermediate states of the O-helix, during its closing, act as a first-line fidelity check-point to sample the active site and differentiate between correct and incorrect nucleotide through base-pairing, base-stacking, and alignment geometry of the bound nucleotide. These states can be seen in structures of matched complexes obtained under 1 minute. In mismatched complexes the O-helix undergoes conformational change to close down in a slightly distorted state, or does not close down entirely, and remains in one of the discrete microstates, distorted, or ajar conformation, and is stabilized by charge separation and chelation between R659 and K663 of motif B (O-helix), and the triphosphate moiety of the incorrect nucleotide, until either the correct metal binds to chelate or disrupt charge separation by attempting to realign the incorrect nucleotide, thereby forcing the O-helix to revert to the open conformation, or the incorrect metal

binds and maintains the O-helix in the distorted state to allow the polymerase to proceed to chemistry.

3.4.2 Structural Aspects Delineated from Examining Fidelity Intrinsic to A-family DNAPs in Matched Complexes

The patterns deduced from the results presented here alter the established abstract view of the catalytic pathway of nucleotide incorporation and show that, in matched complexes (MC's), the absence of metals, or the presence of incorrect metals (Ca^{2+}, Mn^{2+}) alone is insufficient to drive O-helix conformational changes. In fact, in chelated MC's (ChMC), soaking the next incoming dNTP, which in this study facilitates a dC:dGTP ternary complex, into chelated binary complexes was sufficient to drive O-helix closure, suggesting that base-pairing, neighbouring nucleobase-stacking (π:π interactions), and shape-complementarity play only a partial, but significant role in fidelity as primary check-points. In these chelated matched ternary complexes, the O-helix remains closed, despite the absence of any density representative of metal-ions, even after 35-60 minutes –which is well beyond the time-point at which chemistry has been observed in our structures –waiting for the correct metals to ligate motifs A and C, and the triphosphate moiety of the incoming nucleotide prior to chemistry.

3.4.3 Structural Aspects Delineated from Examining Fidelity Intrinsic to A-family DNAPs in Mismatched Complexes

In all mismatched complexes (MMC) with the correct metal-ion (Mg^{2+}, MgMMC) at one minute post-introduction of an incorrect nucleotide, the O-helix samples the active-site by closing down on the incorrect dNTP before rapidly reverting to the open conformation after 2 minutes of reaction progression –the point in time at which the binding metal-ion (Mg^{B}) would saturate the active site and ligate the βγ-PO_4 to motifs A and C in matched complexes– to release the incorrect nucleotide. Interestingly, these MMC's show a unique O-helix conformation close, but not identical, to a previously-described "ajar" microstate. In all MgMMC's the O-helix is seen pushed outwards from the prototypical correct ternary complex by ~0.5-1.0A. This should not be confused with the outwards and upwards conformation that follows the normal trajectory from open to closed state, but is instead akin to a pre- and post-translocated step, such that the

movement of the O-helix is lateral rather than vertical, suggesting some variation on "backtracking" that has been reported in some DNA polymerases and postulated to be implicated in fidelity and proofreading. In these structures G668 (motif B, O-helix) acts as a flexible hinge to allow for a unique conformation of F667, of the same motif, where F667 no longer stacks against the deoxyribose and 3'-OH of the incoming dNTP but is instead flipped ~90° upwards and base-stacks against Y696, which provides similar π-π interactions and stacking to those that the deoxyribose provides, thereby stabilizing this new ring-stacking and preventing F667 from interacting with the deoxyribose of the incorrect dNTP and destabilizing this MMC by a cascade effect as its rotation propagates along the O-helix. As a result, in all MgMMC's R660 which normally interacts with the nucleobase of the incoming dNTP is seen extended away from the nucleobase and no longer interacting in any way with the incorrect dNTP, further destabilizing these MgMMC's to promote reversion of the O-helix to the open conformation and consequently release the incorrect nucleotide. This is contrasted with other A-family DNAPs from Enterobacteriophage T7, and yeast and human mitochondrial DNAPγ, which have a Tyr residue in place of their functional equivalent, F667 in KlenTaq, that alters the polymerases ability to discriminate between 2'-ribose and deoxyribose sugars of the incoming (d)NTP, in line with the conclusions drawn above.

Metal-ion preferences are clearly larger contributors to fidelity in mismatched complexes than they are in their matched counterparts. An examination of the contribution of metals to fidelity in MMC's was described above in which chelated binary complex crystals were soaked in mother liquor that included $MnCl_2$ and dATP, resulting in a dC:dATP mismatch. Remarkably, the O-helix transitions to the closed state, similar to MC's and remains closed for extended periods of time with a geometry that appears to be poised to undergo chemistry (Fig. 3.6). The reaction was not monitored to completion in this complex due to constraints on synchrotron beam-time availability. However, it has been shown that bacteriophage T7 DNAP, and certainly many more polymerases from other families, more readily allow misincorporation when Manganese (Mn^{2+}) is used to facilitate chemistry. The reasoning postulated for this is due to the degree of rigidity in coordination geometry for each metal ion. For Mg^{2+}-coordinated complexes, octahedral geometry is enforced prior to chemistry, whereas Mn^{2+} is more

flexible, capable of accommodating square planar, tetrahedral and octahedral coordination. If the incorrect dNTP is rotated in the active site in such a way as to obscure proper alignment of the triphosphate moiety with respect to catalytic residues in MnMMC's, chemistry would nonetheless occur resulting in misincorporation.

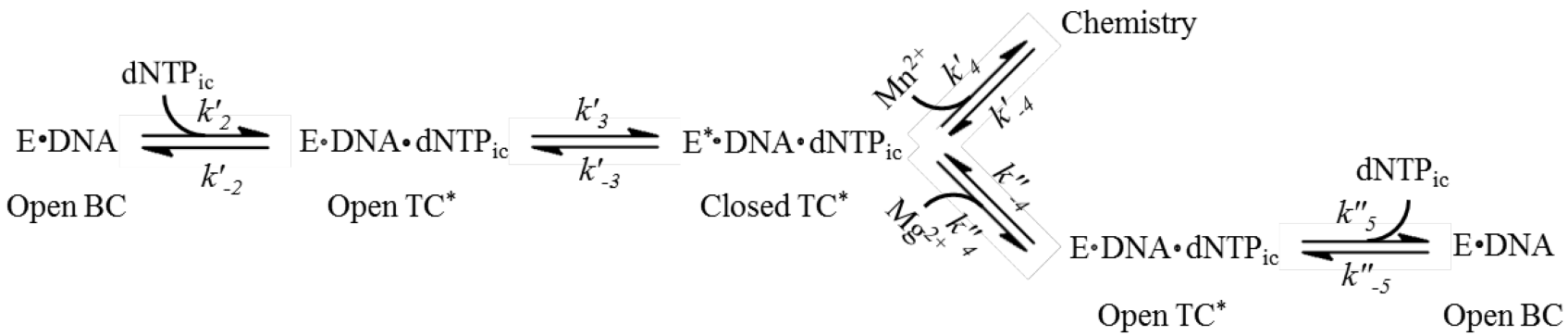


Figure 3.6 Minimal kinetic mechanism for incorrect nucleotide incorporation by A-family DNA polymerases. Here the minimal kinetic mechanism for incorrect nucleotide incorporation is expanded to delineate the effect of metal-ions on misincorporation. E: enzyme, BC: binary complex, TC: ternary complex.

3.4.4 Mechanisms of O-helix Reversion in Nucleotide Discrimination

3.4.4.1 Charge-driven Mechanism (Charge-repulsion)

As previously discussed, although no distinct density for the binding metal-ion, Mg^B, is observed in MgMMC's, it appears that one possible mechanism by which an incorrect nucleotide in these complexes induces conformational change in the O-helix is due to the rigidity of the octahedral coordination of Mg^{2+}, wherein upon its ligation, or attempts thereof, a forced rotation of the incorrect nucleotide causes its 3'-OH to clash with F667 of motif B (O-helix), forcing F667 to rotate upwards and stack against Y696. The rotation propagates throughout the O-helix, from G668 to R660, resulting in R660 also being rotated away from the nucleobase of the incorrect nucleotide which further destabilizes the MgMMC until the O-helix opens rendering the incorrect nucleotide solvent-exposed with little chelation of the negative charge now built-up on the triphosphate moiety, which results in charge repulsion with motifs A and C carboxylates, further instability of the incorrect dNTP and ultimate release from the active site.

3.4.4.2 Metal-ion-driven Mechanism (Rigidity, Steric hindrance)

An alternative mechanism, and perhaps more likely, given the structural features of MnMMC of dC:dATP, is that the O-helix will close down regardless and that its conformation will depend on whether the correct or incorrect nucleotide is in the active site. If the correct nucleotide is bound then no distortion occurs in the O-helix conformation and ligation of Me^B (Mg^{2+} or Mn^{2+}) stabilizes the nascent complex for entry and subsequent ligation of Me^A, allowing polymerization to continue, and if the incorrect nucleotide is bound, the O-helix alters its conformation to accommodate the rotated incorrect nucleotide and prevent a steric clash in the closed state until a metal-ion comes into the active site, thus delegating nucleotide discrimination to the nature of the ligating metal-ions and resulting in a fidelity mechanism that is now driven by choice of ligating metal and not the O-helix. In this scenario, Mg^B cannot ligate the triphosphate moiety of the incorrect nucleotide due to its rotation away from motif A and the inherent geometric constraints of Mg^{2+}, allowing for a negative charge buildup in the triphosphate moiety of the incorrect nucleotide and motif A and C carboxylates which ultimately forces the incorrect dNTP to push upwards against, and open the O-helix to release the

incorrect dNTP, or the more flexible Mn^{2+} will bind and stabilize the MMC in the closed state until either Mg^{2+} competes with it and causes the O-helix to open, or chemistry occurs and results in misincorporation.

Since the MnMMC, dC:dATP, was able to maintain a closed O-helix with the incorrect conformation, owing to the coordination flexibility of Mn^{2+}, it stands to reason that fidelity in MgMMC's is primarily governed by the presence or absence of Mg^{B} binding. However, in MnMMC's whatever measure of fidelity that is sustained is likely due to Watson-Crick geometry, base-pairing, and base-stacking.

3.4.5 Incomplete Structural Details

Without ChMMC to further delineate the role of metals in stabilizing correct or incorrect complexes it is difficult to conclude which model accurately identifies structural determinants for fidelity. Does the distorted O-helix remain closed until "metal-sensing" is accomplished, or does it eventually revert to the open conformation without the aid of metals? Similarly, an examination of scaffolds with variant templating nucleobases is necessary to examine whether fidelity is universal, regardless of whether the templating nucleobase and incoming nucleotide are purines or pyrimidines, and that slight deviation in the binding pocket due to transitions or transversions will not affect the mechanism by which the polymerase sustains fidelity.

With regards to MnMMC, the fate of this complex must be determined to ascertain the extent of incorrect metal involvement in misincorporation and fidelity. If chemistry in MnMMC occurs then it is likely that some structural determinant along the catalytic pathway, beyond what has been described here, will be different between matched and mismatched complexes.

3.5 Materials and Methods

3.5.1 Materials

Oligonucleotides used in crystallization were PAGE and HPLC purified from Integrated DNA Technologies (Carlsbad, CA USA). Deoxyribonucleoside triphosphates (dNTP's) were obtained from New England Biolabs (Ipswich, MA USA). 2',3'-Dideoxyribonucleoside triphosphate (ddNTP) analogues were obtained from GE Healthcare Bio-Sciences Corporation (Piscataway, NJ USA). The non-hydrolyzable analogue of dGTP, dGMpNHpp was purchased from Jena Bioscience (Jena, Germany). 5-iodo- dUTP and 5-iodo-dCTP were purchased from TriLink BioTechnologies (San Diego, CA USA). Crystallization screens and reagents were obtained from Hampton Research (Aliso Viejo, CA). All other reagents were of the highest grade available.

3.5.2 Protein Expression and Purification

The coding sequence for the Klenow fragment of *Thermus aquaticus* (KlenTaq) DNA polymerase (residues 281-832) was amplified from *Thermus aquaticus* genomic DNA and cloned into a phage T7-derived expression vector to facilitate production of recombinant KlenTaq as previously described (Barnes, *et al.*). Briefly, cell pellets were lysed by sonication followed by heat denaturation and ultracentrifugation. The resultant material was precipitated with polyethyleneimine (PEI), and subsequently purified by fractionation on Heparin-Sepharose, Q-Sepharose, and Superdex 75, yielding >99% pure KlenTaq.

3.5.3 Crystallization of dGMPNPP-Trapped Ternary Complex

KlenTaq DNAP ternary complex was formed by mixing 7.5mg/mL KTQ with preformed duplex DNA (Primer 5'-GACCACGGCGCA-3', Template 3'-CTGGTGCCGCGTCAAA-5'), and dGMPNPP in a molar ratio of 1:1.5:10. Crystals of the ternary complex were obtained by vapor diffusion against a reservoir buffer containing 0.1M Tris, pH 8, 0.2M $Mg(OAc)_2$, 0.2M $(NH_4)OAc$, 10mM TCEP and 17-20% PEG 8,000 or 18-25% PEG 3,350. Crystals were cryoprotected successively in mother liquor containing 10mM $Mg(OAc)_2$ and increasing concentrations of PEG 8,000 or PEG 3,350 up to 30% and 35%, respectively.

3.5.4 Derivation of the Open Binary Complexes

Crystals of the open, catalytically-competent binary complex crystals were obtained by soaking crystals in mother liquor lacking dGMPNPP to remove the non-hydrolyzable analogue and facilitate reversion to the enzyme's open state. Resultant crystals were cryoprotected successively in mother liquor containing 10mM $Mg(OAc)_2$, 0.1M HEPES, pH 7, and increasing concentrations of PEG 8,000 or PEG 3,350 up to 30% and 35%, respectively.

3.5.5 Time-course of Nucleotide Binding

Catalytically-competent binary complex crystals, now at physiological conditions, were soaked in cryoprotectant containing 5mM of either dGTP, dATP, dTTP, 5-iodo-dUTP or 5-iodo-dCTP for 30 second intervals up to 5 minutes, 1 minute intervals from 6- to 30 minutes, 10 minute intervals from 30- to 60 minutes, and 30 minute intervals from 1- to 4 hours. Chelated, or Metal-less complexes were washed in mother liquor containing 5-25mM EDTA, followed by soaking in cryoprotectant and subsequent soaking in the same cryoprotectant containing 5mM of the aforementioned dNTPs and analogues. For each time point crystals were flash-frozen in liquid nitrogen prior to data collection.

3.5.6 Data Collection and Processing

Diffraction data were collected at the Macromolecular Diffraction facility of the Cornell High Energy Synchrotron Source (MacCHESS) using both A-1 and F-1 beamlines (Cornell University, Ithaca, NY USA), and processed using HKL2000 (11). Further data processing was carried out using the PHENIX software suite (12).

3.5.7 Structure Determination and Refinement

The phase for the open binary complex structure was solved by molecular replacement using PHASER (13) with a modified model of the open, dideoxy-terminated binary complex of KlenTaq (PDB ID: 4KTQ) which lacks residues 630-680 that encode a portion of the fingers domain containing the O-helix. COOT (14) was used to rebuild the deleted region in the F_o-F_c map of the resultant model followed by replacement of the

template 5'- and primer 3'-nucleobases. In a similar fashion, the phase for the dGMPNPP-trapped ternary complex was solved by molecular replacement with a modified model of the closed dideoxy-terminated ternary complex of KlenTaq (PDB ID: 3KTQ) lacking the incoming dNTP and Mg^A/Mg^B. The primer 3'-end, dGMPNPP, $Mg^{A,}$ and Mg^B were modelled into the resultant F_o-F_c map. The final structures of the open binary complex and closed ternary complex lacking dGMPNPP and metals served as search models in molecular replacement for structure determination of subsequent time-course structures.

Structures of ternary complexes up to 5 minutes were solved by molecular replacement using the model for the open binary complex to determine the time-point at which the O-helix converts to the closed conformation followed by modelling of dNTPs or analogues and Me^B/Me^A into the F_o-F_c when pertinent. Structures of subsequent time points were solved by molecular replacement using the model for the closed ternary complex and subsequent modelling of dNTPs and analogues, and Me^B/Me^A into the F_o-F_c when pertinent. Resultant structures were refined using PHENIX (12).

Acknowledgements

We thank the staff and scientists at the Macromolecular Diffraction facility of the Cornell High Energy Synchrotron Source, and the Huck Institute for the Life Sciences X-ray Crystallography facility at the Pennsylvania State University for technical assistance and support. This work was supported by NIH grant GM087350, the Stanley Person Graduate Fellowship in Biochemistry and Molecular Biology, and Kuwait University. Figures were prepared using PyMOL.

References

1. Steitz, TA. (1999). DNA Polymerases: Structural Diversity and Common Mechanisms. *J. Biol. Chem.* **274**, 17395-17398.

2. Johnson, KA. (2010). The kinetic and chemical mechanism of high-fidelity DNA polymerases. *Biochem. Biophys. Acta.* **1804**, 1041–1048.

3. Brautigam, C., Steitz, TA. (1998). Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* **8**, 54-63.

4. Lee, I., Berdis, A. (2010). Non-natural nucleotides as probes for the mechanism and fidelity of DNA polymerases. *Biochem. Biophys. Acta.* **1804**, 1064-1080.

5. Arora, K., Beard, W., Wilson, S., Schlick, T. (2005). Mismatch-induced conformational distortions in polymerase beta support an induced-fit mechanism for fidelity. *Biochemistry* **44**, 13328-13341.

6. Showalter, AK., Tsai, MD. (2002). A reexamination of the nucleotide incorporation fidelity of DNA polymerases. *Biochemistry* **41**, 10571-10576.

7. Sampoli Benitez, BA., Arora, K., Schlick, T. (2006). In silico studies of the African swine fever virus DNA polymerase X support an induced-fit mechanism. *Biophys. J.* **90**, 42-56.

8. Joyce, CM., Potapova, O., DeLucia, AM., Huang, X., Basu, VP., Grindley, N. (2008). Fingers-Closing and Other Rapid Conformational Changes in DNA Polymerase I (Klenow Fragment) and Their Role in Nucleotide Selectivity. *Biochemistry* **47**, 6103-6116.

9. Li, Y., Korolev, S., Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17**, 7514-7525.

10. Rothwell, PJ., Allen, WJ., Sisamakis, E., Kalinin, S., Felekyan, S., Widengren, J.,

Waksman, G., Seidel, CA. (2013). dNTP-dependent conformational transitions in the fingers subdomain of Klentaq1 DNA polymerase: insights into the role of the "nucleotide-binding" state. *J. Biol. Chem.* **288**, 13575-13591.

11. Otwinowski, Z., Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*. **276**, 307–326.

12. Adams, P., Afonine, P., Bunkóczi, G., Chen, V., Davis, I., Echols, N., Headd, J., Hung, L., Kapral, G., Grosse-Kunstleve, R., McCoy, A., Moriarty, N., Oeffner, R., Read, R., Richardson, D., Richardson, J., Terwilliger, T., Zwart, P. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221.

13. McCoy, A., Grosse-Kunstleve, R., Adams, P., Winn, M., Storoni, L., Read, R. (2007). Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674.

14. Emsley, P., Lohkamp, B., Scott, W., Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501.

Chapter 4

Summary of and Future Direction in Studies on DNA Replication

4.1 Studies on DNA Replication and DNA Polymerases

The overarching theme of this work is the desire, obsessive-compulsive or not, to visually monitor, at atomic resolution, dissect and examine enzyme-catalyzed reactions. Specifically, the objective of this work is the unambiguous deconstruction of the entire catalytic pathway of nucleotide incorporation by a high-fidelity A-family DNA polymerase to address a host of unanswered questions pertaining to catalysis and fidelity.

With growing evidence in support of multiple models of nucleotide incorporation fidelity (1,2), and metal ion-dependent mechanistic variations (3,4), it was unclear how discrimination and selectivity are brought about in A-family DNA polymerases. Due to the lack of direct evidence, the initial binding site for the incoming nucleotide was yet to be determined and was likely to provide definitive evidence in support of either an induced-fit or a transition state model for fidelity (1,2). The determination of initial binding site was also posited to provide a structural basis for triggering O-helix conformational changes prior to nucleotidyl transfer (5).

It was also unclear, during nucleotidyl transfer, whether O3' oxyanion activation, for an inline attack on the α-PO_4 of the incoming nucleotide, is brought about by the evolutionarily conserved Glu residue in motif C (6) or whether such deprotonation relies on water molecules in the vicinity of the active site (7). After nucleotidyl transfer, three events must occur; pyrophosphate release, O-helix opening, and translocation. Kinetic and FRET (5) studies were incapable of discerning whether pyrophosphate release occurs prior to, simultaneously with, or subsequent to O-helix reversion due to the rapid and transient nature of the two events, and thus could not accurately discern whether either event provides intrinsic energy for the other to occur.

Once a nucleotide is incorporated the polymerase must translocate to the next templating nucleobase, however, it was not entirely clear whether it could do so in a sliding fashion with the O-helix in the closed conformation, or whether it needed to revert to the open conformation first, and whether reversion also required it to physically dissociate from its substrate, translocate and re-form the binary complex. One would assume that for subsequent nucleotide incorporation, the O-helix must first revert to the

open conformation to allow for the aforementioned steps in nucleotide binding to occur prior to catalysis, but whether or not this is required for translocation is not well understood. If this is the case then translocation, whether by sliding onto the next templating nucleobase, or physically dissociating and re-associating with its substrate DNA, must be a costly energy-driven process due to the mechanical nature of this step and the origin of such chemical energy and its transduction into mechanical energy should be investigated along with contributions, if any, made by pyrophosphate release.

To that end, we developed and cultivated a time-resolved X-ray crystallographic platform utilizing the large (Klenow) fragment of *Thermus aquaticus* (KlenTaq) DNA polymerase, retaining its high-fidelity proof-reading polymerase domain but lacking the N-terminal exonuclease domain (8), to interrogate the catalytic pathway in an effort to delineate the structural and mechanistic determinants of nucleotide incorporation and its fidelity.

Using this platform we examined single-turnover kinetics of nucleotide incorporation, beginning with a catalytically-competent binary complex in which the correct incoming nucleotide was soaked, and followed the reaction to post-chemistry and provided evidence that formation of a catalytically-competent ternary complex poised for chemistry in A-family DNA polymerases utilizes two rather than three metal ions (Mg^{2+}) and is a step-wise process wherein nucleotide-binding and metal-ligation are decoupled (Fig. 4.1). This spatiotemporal separation of substrate and metals serves as the first checkpoint in high-fidelity nucleotide incorporation, as metal entry subsequent to nucleotide binding is contingent upon an optimal geometry of said nucleotide in the active site.

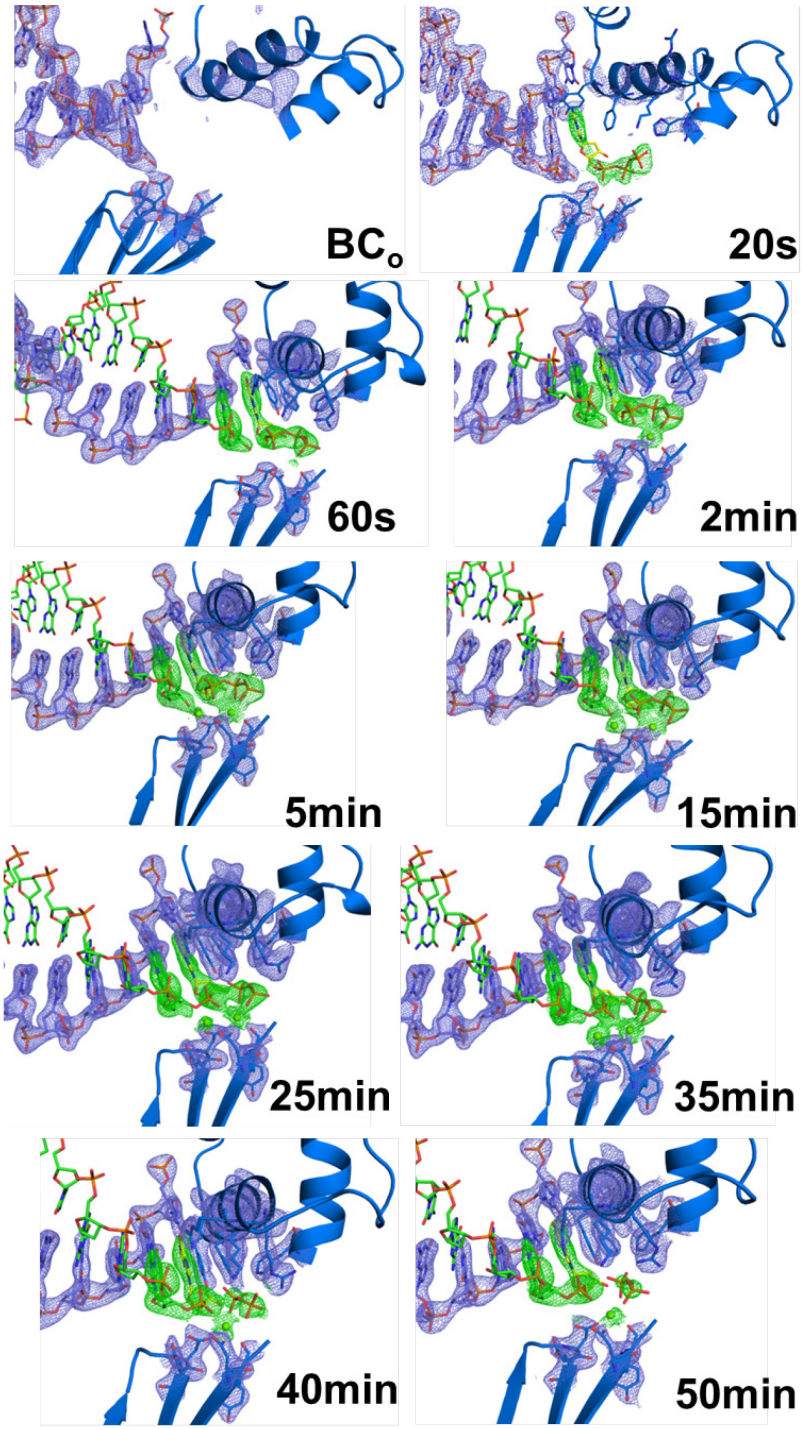


Figure 4.1 Snapshots along the catalytic pathway for correct nucleotide incorporation by A-family DNAPs. The panels show reaction intermediates after soaking the correct nucleotide into a Mg^{2+}-substituted binary complex, highlighting the spatiotemporal separation of substrate and metals and sequential concerted metal-ion entry and exit prior to PP_i release. $2F_o$-F_c density contoured to 1σ is shown as blue mesh and F_o-F_c density, contoured to 3σ is shown as green mesh covering the primer 3'-end, incoming nucleotide and metal-ions.

The aforementioned observation of sequential metal entry warranted an examination of the role of metal ions in the fidelity of nucleotide incorporation and their interplay with the structural stability of ground-state and pre-catalytic complexes. To do so, we examined time-courses of correct and incorrect nucleotide incorporation, or attempts thereof, in catalytically-competent binary complexes containing the correct (Mg^{2+}), suboptimal (Mn^{2+}), or no metal ions, and provide evidence that the O-helix undergoes conformational change to the closed ternary complex upon binding of either correct or incorrect nucleotide, forming stable metal-free ternary complexes, with a slightly distorted O-helix in the case of incorrect nucleotide binding, that relies on incoming metal-ions to reconcile whether it should commit to chemistry, in the case of correct nucleotide binding and correct or suboptimal metal-ion entry, or incorrect nucleotide binding and suboptimal metal-ion entry, albeit at a much lower rate, or revert back to the open conformation and release the incoming nucleotide, in the case of incorrect nucleotide binding and correct metal-ion entry (Fig. 4.2).

Taken together, our observations support a two metal-ion (Mg^{2+}) catalytic mechanism for A-family DNA polymerases that exhibits a hybrid, metal ion-dependent model for fidelity, wherein utilization of Mg^{2+} favours an "induced fit" model, and utilization of Mn^{2+} conforms to a "rate-limiting transition state" model for fidelity (1,2) in which said fidelity is sustained via free-energy change derived from the formation of correct or incorrect transition state species –pre-chemistry complexes– that dictate whether the polymerase will proceed to chemistry or revert to an open conformation and release the nucleotide.

Because the platform was cultivated to facilitate single turnover kinetics, owing to packing of molecules to facilitate favourable crystal contacts, the aforementioned unaddressed questions regarding pyrophosphate release, O-helix reversion and translocation could not be individually examined but did provide interesting details regarding coupled dynamics in O-helix reversion and polymerase translocation, and in pyrophosphate release and O-helix reversion. We show that after pyrophosphate release the O-helix remains in the closed conformation with pyrophosphate remaining in the electron density at that location for an extended period of time.

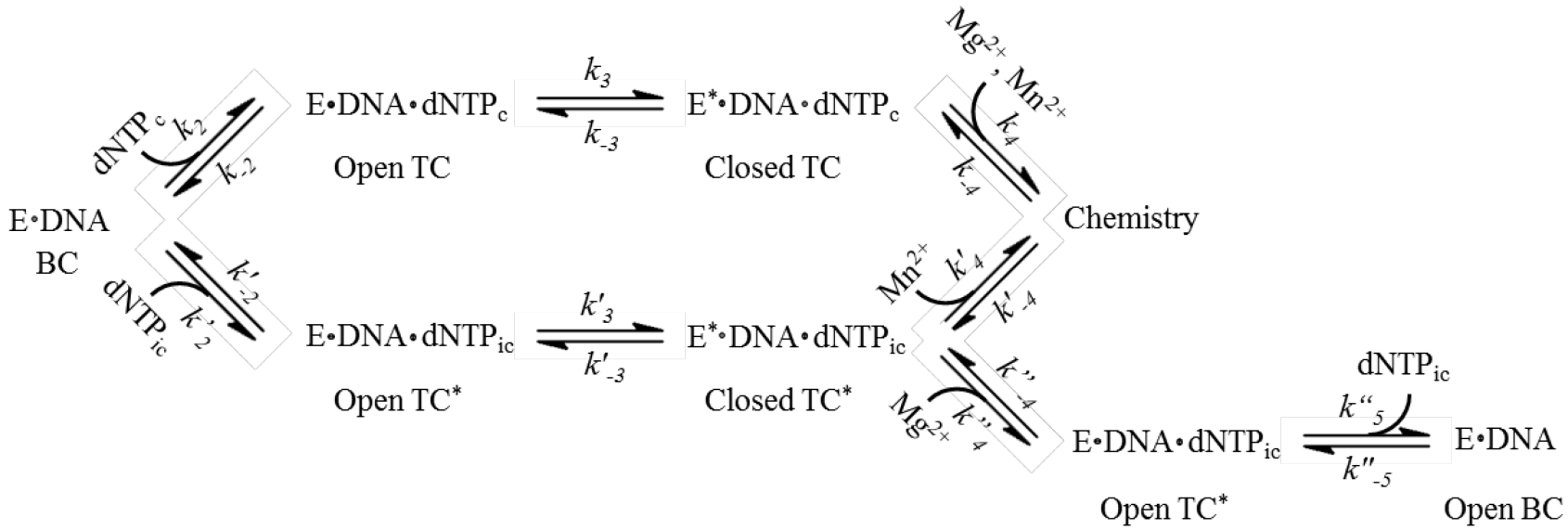


Figure 4.2 Expansion of the minimal kinetic pathway for correct and incorrect nucleotide incorporation by A-family DNA polymerases based on metal-ion coordination. The kinetic scheme presented herein is an expansion of the minimal kinetic scheme for nucleotide incorporation by A-family DNAPs that takes into account metal-ion choice during differential nucleotide binding and metal ion-dependent incorporation or misincorporation. Here, an incorrect nucleotide binding to a chelated binary complex (BC) will allow the enzyme to undergo the necessary conformational changes, albeit with slight structural variations relative to the correct nucleotide binding, and awaits metal-ion entry. If the optimal metal-ion, Mg^{2+}, is used, due to its rigid geometric constraints, the enzyme (E) is able to differentiate between correct and incorrect nucleotide and reverts to the open conformation to release the incorrect nucleotide. However, a suboptimal metal-ion, such as Mn^{2+}, due to its flexible coordination, is able to occupy the position of Me^{B} and offsetting the variation in distances brought about by a rotated triphosphate moiety for the incorrect nucleotide, and potentially allowing the enzyme to proceed to Mn^{A} entry and coordination and subsequent chemistry. TC: ternary complex.

Models for translocation along the duplex DNA onto the next templating nucleobase have long suggested that pyrophosphate release, O-helix opening, or both would provide the intrinsic energy necessary to drive translocation. Contrary to this, however, the structures presented here provide direct evidence that translocation is necessary for O-helix opening. This observation of coupled dynamics between translocation and O-helix opening indicates that the charge buildup at the active site subsequent to pyrophosphate release alone is insufficient to drive O-helix opening but that translocation must occur prior to, or simultaneously with O-helix opening, and that pyrophosphate release, in a physiologically and mechanistically relevant manner, likely requires reversion of the O-helix to the open conformation.

While our platform served this endeavour quite well, it is not without its limitations with regards to studies in structural and mechanistic enzymology. As described in previous chapters, the methodology is highly protein-dependent. It depends entirely on the propensity of the target protein to crystallize under conditions that would permit the necessary conformational and catalytic changes without disrupting the crystal lattice (9-11). Crystal packing, i.e., protein and nucleic acid complex interactions with neighbouring complexes in the asymmetric unit of the unit cell characteristic of the representative space group in which the complex crystallizes, must be maintained in order for this technique to succeed (9-11). This requirement immediately excludes large complexes or macromolecular assemblies that undergo rather large conformational changes or protein or subunit exchange prior to, during, or after catalysis.

The aforementioned drawback limits or entirely precludes future studies necessary to understand multisubunit nucleic acid polymerases and polymerases that require accessory factors (helicases, clamps, etc.) to perform their respective functions. As a result of this drawback, other techniques at or near atomic resolution in the solution-phase, or single molecule techniques are necessary to continue these studies into catalysis and fidelity by nucleic acid polymerases.

Recent advances in hardware (detectors, etc.) and software used in conducting cryo-electron microscopy (Cryo-EM) studies have facilitated atomic resolution structural examination of macromolecular complexes and assemblies (12,13) that were otherwise

difficult to study using crystallography due to either size, propensity to crystallize or large conformational changes along their catalytic pathways that would otherwise disrupt lattice contacts and render the undertaking futile (12,13). Cryo-EM was successful in attempts to obtain the three-dimensional structure of the multisubunit replicative DNA polymerase III holoenzyme from *E. coli* in complex with the DNA sliding clamp β, the proofreading exonuclease ε, and the clamp loader subunit at 6.7Å (14); a feat that is impossible, at least given current technology, to have been fruitful using X-ray crystallography. Similarly, a plethora of Cryo-EM structures have been deposited in recent years for multisubunit prokaryotic and eukaryotic RNA polymerases, in apo- and holoenzyme forms, and in complexes with other transcription factors and antibiotics, at resolutions nearing 3Å (15,16).

For proteins and complexes amenable to crystallization and utilization of a time-resolved approach, advances in X-ray free electron lasers (XFELs), now becoming more and more common at synchrotrons, can facilitate true single-crystal time-resolved crystallographic studies at physiological conditions on a femtosecond scale (17-19), reducing radiation damage and possible perturbation of the protein molecules within a crystal that may ultimately convolute the catalytic pathway under study, and eschewing the need to utilize different crystals of the same complex for the same study, which inevitably results in variations in resolution, crystal size, mosaicity, substrate or structural element occupancy, and most importantly, uncertainty as to whether crystal to crystal variations are somehow convoluting the time-course or order of events or intermediates of the catalytic pathway under examination (20).

References

1. Arora, K., Beard, W., Wilson, S., Schlick, T. (2005). Mismatch-induced conformational distortions in polymerase beta support an induced-fit mechanism for fidelity. *Biochemistry* **44**, 13328-13341.

2. Showalter, AK., Tsai, MD. (2002). A reexamination of the nucleotide incorporation fidelity of DNA polymerases. *Biochemistry* **41**, 10571-10576.

3. Freudenthal, B., Beard, W., Shock, D., Wilson, S. (2013). Observing a DNA polymerase choose right from wrong. *Cell* **154**, 157-168.

4. Nakamura, T., Zhao, Y., Yamagata, Y., Hua, Y., Yang, W. (2012). Watching DNA polymerase η make a phosphodiester bond. *Nature* **487**, 196-201.

5. Rothwell, PJ., Allen, WJ., Sisamakis, E., Kalinin, S., Felekyan, S., Widengren, J., Waksman, G., Seidel, CA. (2013). dNTP-dependent conformational transitions in the fingers subdomain of Klentaq1 DNA polymerase: insights into the role of the "nucleotide-binding" state. *J. Biol. Chem.* **288**, 13575-13591.

6. Doublie, S., Tabor, S., Long, AM., Richardson, CC., Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2-Å resolution. *Nature* **391**, 251-258.

7. Yang, L., Arora, K., Beard, W., Wilson, S., Schlick, T. (2004). Critical Role of Magnesium Ions in DNA Polymerase β's Closing and Active Site Assembly. *J. Am. Chem. Soc.* **126**, 8441-8453.

8. Li, Y., Korolev, S., Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17**, 7514-7525.

9. Moffat, K., Helliwell, J. (1989). The Laue Method and its Use in Time-Resolved Crystallography. *Topics. Curr. Chem.* **151**, 61-74.

10. Wyckoff, HW., Doscher, M., Tsernoglou, D., Inagami, T., Johnson, LN., Hardman,

K., Allewell, NM., Kelley, DM., Richards, FM. (1967). Design of a diffractometer and flow cell system for X-ray analysis of crystalline proteins with applications to the crystal chemistry of ribonuclease-S. *J. Mol. Biol.* **27**, 563-578.

11. Westbrook, EM., Sigler, PB. (1984). Enzymatic function in crystals of delta 5-3-ketosteroid isomerase. Catalytic activity and binding of competitive inhibitors. *J. Biol. Chem.* **259**, 9090-9095.

12. Fernandez-Leiro, R., Conrad, J., Scheres, S., Lamers, M. (2015). cryo-EM structures of the *E. coli* replicative DNA polymerase reveal its dynamic interactions with the DNA sliding clamp, exonuclease and *t*. *eLife* **4**, e11134.

13. Vonck, J., Mills, D. (2017). Advances in high-resolution cryo-EM of oligomeric enzymes. *Curr. Opin. Struc. Biol.* **46**, 48-54.

14. Fernandez-Leiro, R., Conrad, J., Scheres, SH., Lamers, MH. (2015). cryo-EM structures of the *E. coli* replicative DNA polymerase reveal its dynamic interactions with the DNA sliding clamp, exonuclease and τ. *eLife* **4**, e11134.

15. Narayanan, A., Vago, FS., Li, K., Qayyum, MZ., Yernool, D., Jiang, W., Murakami, KS. (2018). Cryo-EM structure of *Escherichia coli* σ^{70} RNA polymerase and promoter DNA complex revealed a role of σ non-conserved region during the open complex formation. *J. Biol. Chem.* **293**, 7367-7375.

16. Liu, X., Farnung, L., Wigge, C., Cramer, P. (2018). Cryo-EM structure of a mammalian RNA polymerase II elongation complex inhibited by α-amanitin. *J. Biol. Chem.* **293**, 7189-7194.

17. Johansson, L., Stauch, B., Ishchenko, A., Cherezov, V. (2017). A bright future for serial femtosecond crystallography with XFELs. *Cell* **42**, 749–762.

18. Mizohata, E., Nakane, T., Fukuda, Y., Nango, E., Iwata, S. (2018). Serial femtosecond crystallography at the SACLA: breakthrough to dynamic structural biology. *Biophys. Rev.* **10**, 209-218.

19. Spence, J. (2017). XFELs for structure and dynamics in biology. *IUCr J.* **4**, 322-339.

20. Moffat, K. (1989). Time-Resolved Macromolecular Crystallography. *Annu. Rev. Biophys. Biophys. Chem.* **18**, 309-332.

Appendix A

Purification and Crystallization Screening of Extension-ready Binary and Ternary Complexes of KlenTaq DNA Polymerase

Purification of KlenTaq DNA polymerase

Performing structural studies with KlenTaq DNA polymerase required production of large quantities of highly pure protein. Using the native source would not be in-line with these requirements and the necessity for truncated or mutated variants precluded this approach. We sought to clone the Klenow fragment of Taq DNA polymerase first reported by Barnes, *et al*. This construct was an N-terminal truncated variant retaining amino acid residues 281-832 and included a short peptide (MGKRKST) at the N-terminus that was shown to increase solubility and expression levels during recombinant protein expression in *E. coli*. This construct was also the first KlenTaq variant that was crystallized by Waksman *et al*., and was shown to be highly amenable to crystallization under a host of conditions.

Purification of KlenTaq DNA polymerase performed using a combination of heat-treatment, precipitation and chromatographic separation (Fig. A1). Heat-treatment proved to be a viable first step in the isolation of KlenTaq, as is the case with purification of most thermostable proteins. The heat denatures most other thermolabile proteins but for short periods of time *in vitro* leaves KlenTaq intact and, depending on protein and salt concentration, even still complexed with DNA. However, while more than 90% of endogenous cellular proteins are often denatured and removed during this step, approximately 30-60% of the target polymerase activity is also lost when heated at 80°C for 45 minutes. Some protection is afforded by the fact that KlenTaq is still bound to DNA during this step which precluded performing Polyethyleneimine (PEI) precipitation prior to heat treatment, as KlenTaq exhibits biphasic DNA binding that is both protein and salt concentration-dependent. It is also necessary to agitate or mix the lysate during this heating step at regular intervals to distribute heat evenly throughout the solution and prevent any localization that may result in excessive heating and denaturation.

PEI precipitation served two functions, the removal of endogenous DNA- and RNA-binding proteins and the removal of DNA bound to target KlenTaq DNA polymerase. The positively charged PEI polymer interacts with, and subsequently binds the negatively charged phosphate backbones of nucleic acids and precipitates them, and

their bound proteins, out of solution. It is empirically determined, batch by batch, the required percentage of PEI to be used to precipitate proteins and nucleic acids, however in most cases 0.3% was sufficient to remove DNA from the preparation. Variations in the recovery often exist due to the strength of the binding interaction between KlenTaq and substrate DNA. To circumvent further loses in recovery of KlenTaq DNAP, the lysate is equilibrated to 0.5M NaCl to disrupt any interactions between KlenTaq and DNA. PEI precipitation is subsequently performed by drop-wise addition to the resultant lysate at $4^{o}C$ until the desired 0.3% is reached. A white precipitate gradually forms as the PEI binds to nucleic acids and precipitates them out of solution. After a1-hour incubation with mixing at $4^{o}C$, the homogenous lysate is centrifuged at 30,000xg for 45minutes at $4^{o}C$, which results in precipitation of heavy PEI-complexed molecules.

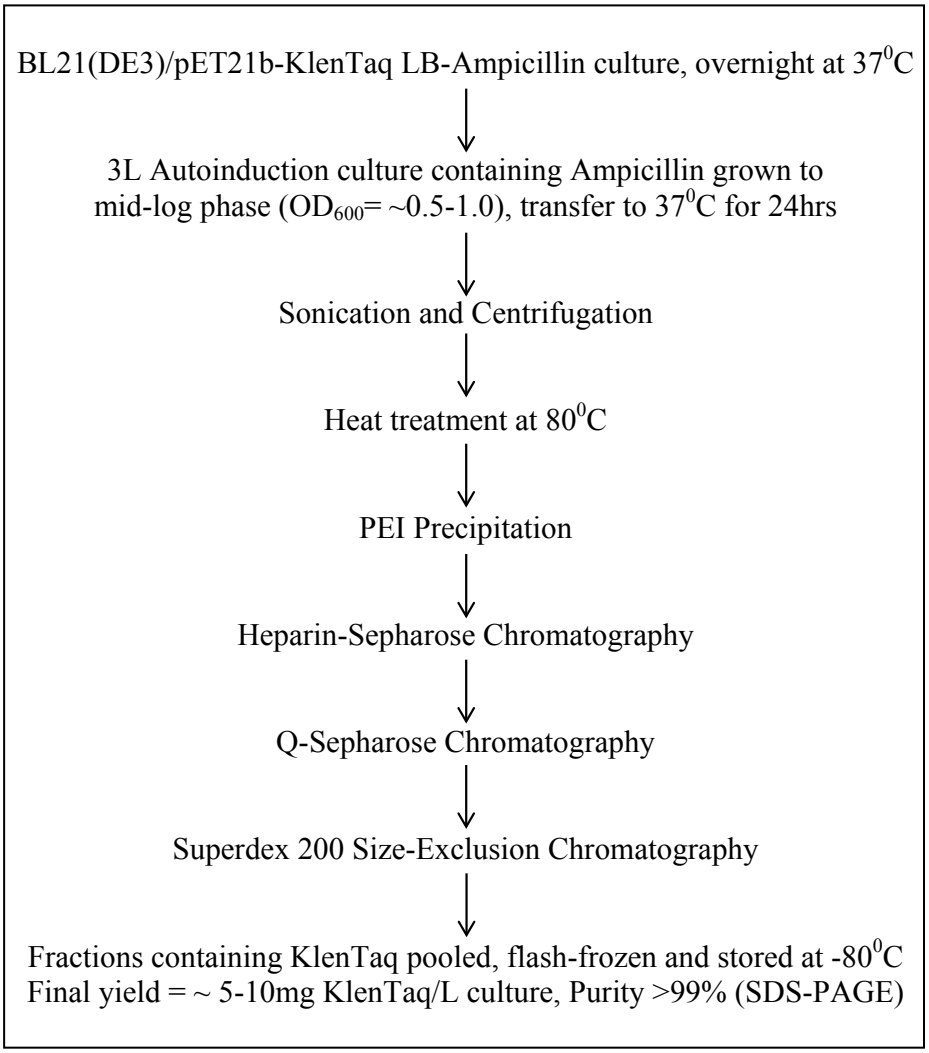


Figure A1. Purification scheme for KlenTaq DNA polymerase.

It is often the case, as we and others have observed, that the lysate extracted after centrifugation usually contains residual PEI that significantly interferes with subsequent chromatographic separations. More so with chromatography on Heparin-substituted agarose or sepharose matrices. Strategies we have often used to remove residual PEI after centrifugation included subsequent centrifugation steps, PEI adsorption to the weak cation-exchanger Bio-Rex 70, and in some cases the omission of the PEI precipitation step and replacement with chromatography on Q-sepharose anion-exchange for removal of DNA. Anion-exchange on Q-sepharose for this purpose is costly and as such is not quite common but proved to be efficient, more reproducible relative to other methods, and inconsequential to downstream processing.

After removal of residual PEI, we often obtain approximately 40% recovery of KlenTaq DNA polymerase relative to the starting material. At this point, the resultant lysate is diluted to a conductivity of 100mM NaCl to facilitate adsorption onto Heparin-sepharose as the first chromatographic step. More than 90% of applied KlenTaq DNA polymerase is retained on the column. However, no binding studies were conducted to explore how this retention is affected by the amount loaded and the binding capacity of Heparin in mg/mL for KlenTaq. After subsequent washing at 200mM NaCl, most other cellular proteins are eluted. KlenTaq DNA polymerase elutes at approximately 250mM NaCl during a linear gradient of 200-600mM NaCl (Fig. A2).

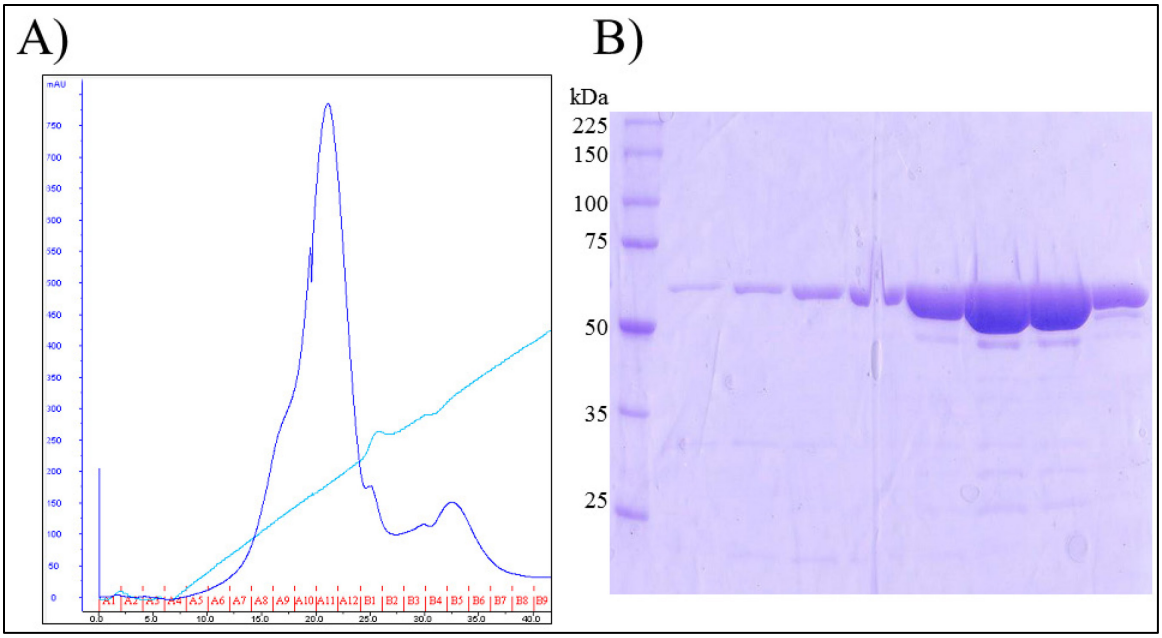


Figure A2. Elution Profile of KlenTaq DNA Polymerase on Heparin-Sepharose.
A) Chromatogram of elution profile of KlenTaq from Heparin-Sepharose. The major peak eluting at approximately 240mM NaCl is that of KlenTaq. B) Analysis of peak fractions on 10% SDS-PAGE showing more than 95% purity for KlenTaq.

The eluted fractions of KlenTaq represent more than 95% purity and are readily crystallizable due to the proteins high propensity for crystallization, but crystallization is not consistent and rapid degradation occurs if concentrated (~30mg/mL) aliquots are stored at 4°C for more than several weeks. Subsequent mass spectrometric identification of protein species in the preparation put suspicions of protease contamination at rest but have, however, shown that the only other existing protein was that of *E. coli* GTP cyclohydrolase. Thus, a subsequent chromatographic step was warranted. Moreover, seeing as how residual DNA may affect later crystallization, the approach used utilized strong anion-exchange chromatography on Q-sepharose that would serve to potentially separate monomers of GTP cyclohydrolase while also binding any residual DNA for removal.

The pooled KlenTaq fractions from Heparin were diluted to a conductivity of 20mM NaCl and applied to a Q-sepharose column. Proteins were eluted with a linear gradient from 20-300mM NaCl, and KlenTaq was found to elute at approximately 80mM NaCl. KlenTaq fractions were pooled, concentrated, and further purified by size-exclusion chromatography on Superdex 200 (Fig. A3). Here, KlenTaq was found to elute beginning at 65mL, approximately one-half of the column volume.

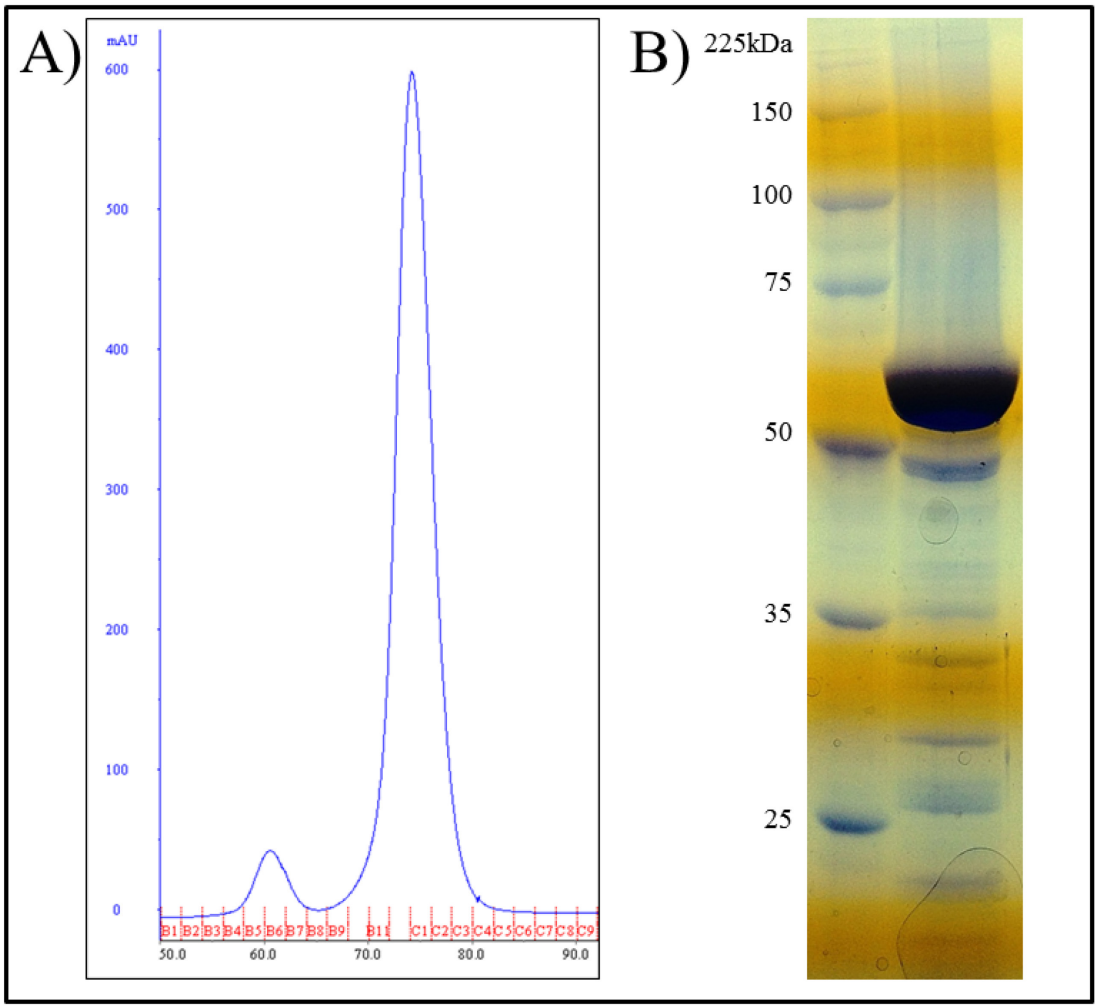


Figure A3. Elution Profile of KlenTaq DNA Polymerase on Superdex 200.

A) Chromatogram of KlenTaq elution profile from Superdex 200 size-exclusion chromatography. The major peak eluting at approximately 65mL (½ CV, 120mL) represents KlenTaq. B) Analysis of the major Superdex 200 peak fraction. The discreet doublet-bands below the major band are artefacts of excessive heating to denature the sample. Verified by MS/MS.

Crystallization Screening of Extension-ready Binary and Ternary Complexes

The successful observation of structural changes in our initial dideoxy-terminated system prompted us to further pursue this goal by modifying the system to allow the observation of a complete cycle of nucleotide addition. To do so, this required crystallizing KlenTaq in an "Extension-ready" complex; meaning a 3'-hydroxyl on the primers' 3'-end was required for further work to study the mechanism of catalysis, or the chemistry step, upon completion of nucleotide binding. No known crystal structures of KlenTaq exist in this state. Available structures are often dideoxy-terminated to lock the O-helix in the closed conformation and facilitate crystal packing for crystallization as the O-helix actively samples the active site in an otherwise catalytically-competent complex.

A KlenTaq-directed crystallization screen was designed to include novel conditions derived from all published structures of KlenTaq. Multiple strategies were undertaken to screen for new conditions. Screens were set up to evaluate crystal growth using conventional scaffolds. i.e., the CG scaffold used in KTQ_{CGddG}, or a 1-base extended scaffold to preclude a translocation event (A-TC) (Fig. A4).

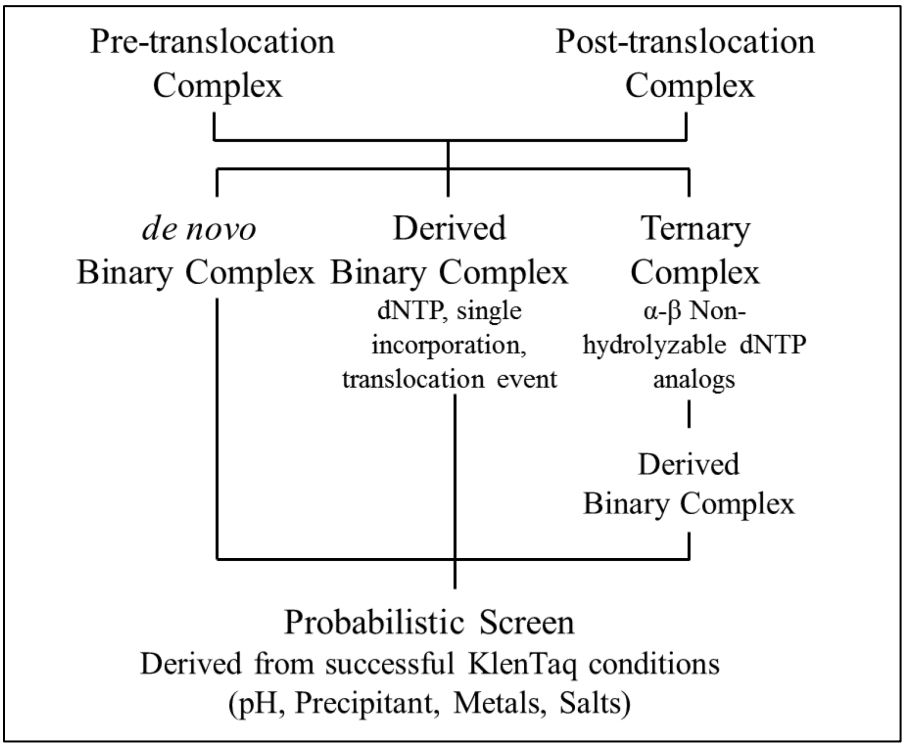


Figure A4. Algorithmic, mechanism-driven screen design for crystallization of KlenTaq DNAP.

Conditions also utilized various divalent metals and nucleotide analogues to either trap the complex in an inactive state, or lock the O-helix in the closed conformation. Also, screens were initially set up with microseeds of the dideoxycytidine-terminated ternary complex to facilitate rapid screening (Fig. A5).

The CG scaffold produced crystals in the pH range of 6.5 to 8.5 when Ca^{2+} was substituted for Mg^{2+} in the presence of dGTP, and as a control, at pH 7.5 using the same metal ion, ddGTP and dCTP. The 1-base extended A-TC scaffold, in which the "A" denotes the extended nucleotide and "TC" denotes the available templating nucleobases, was capable of producing binary crystals in the presence of only Mg^{2+}, but these crystals did not show promise in our preliminary results.

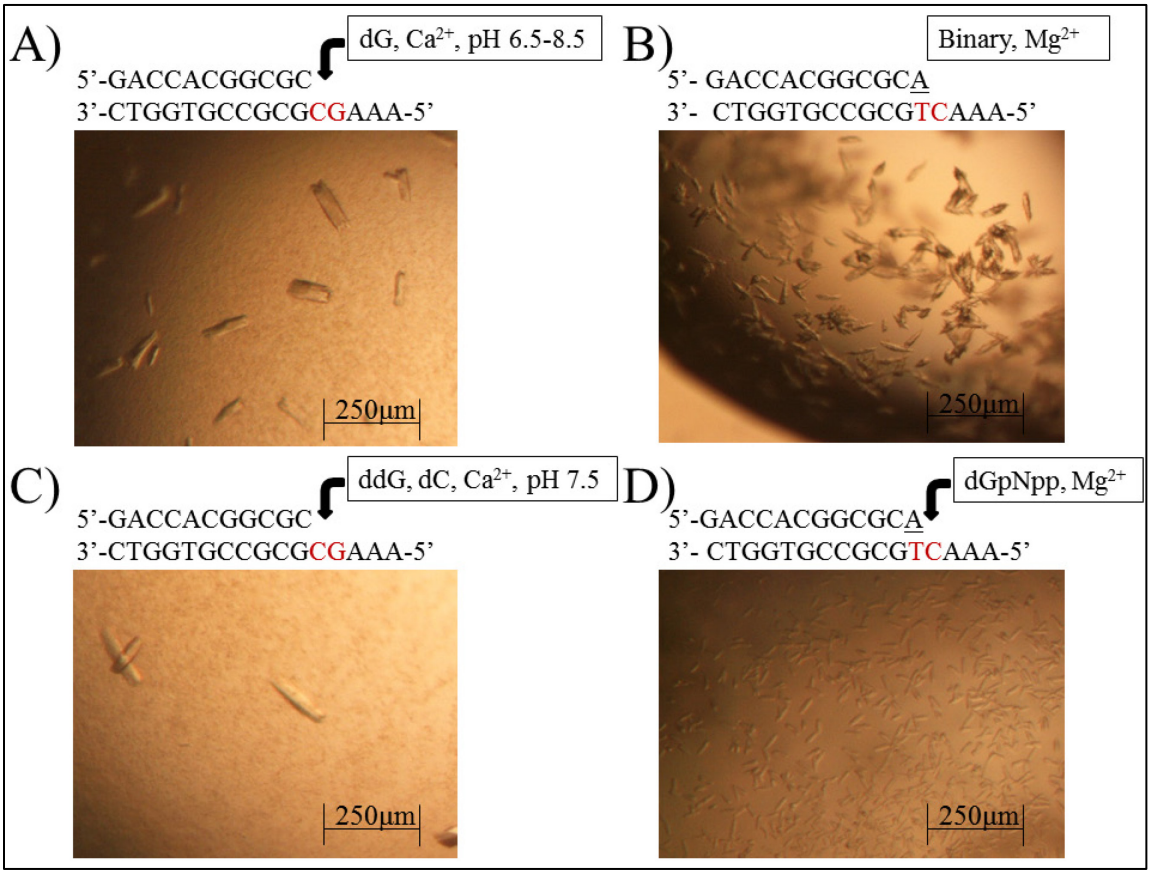


Figure A5. Screening for "Extension-Ready" Crystals. The panel shows crystals obtained for the various conditions producing extension-ready complexes. In A) & C) crystals obtained for a binary and ternary complex using the "CG" scaffold in an inactive state, respectively. In B) & D) a 1-base extended "A-TC" scaffold is used to produce a binary and ternary complex, respectively.

Using the A-TC scaffold, it was possible to obtain crystals using the αβ-non-hydrolyzable dGTP analogue dGMPNPP when NH_4OAc was used in conjunction with $Mg(OAc)_2$ in the pH ranges of 6.5 to 9.5. Crystals appeared within 1 hour. One would assume that this is due to microseeding during screen setup, but this was not the case when using other scaffolds under the same or different conditions. While small in size during initial screening, the number of nuclei, and the mechanism utilized to obtain these crystals, prompted us to further optimize this condition for our work.

The best condition obtained during screening for dGMPNPP ternary complex was further optimized using an Additive Screen from Hampton Research. Interestingly, when this secondary screen was scored (Fig. A6), the best conditions were obtained using organics and additives often used during PCR-based amplification of DNA sequences. This pattern was expected as some of these additives have been reported to improve PCR-based amplification of DNA sequences by stabilizing the DNA polymerase, DNA, or the interactions between them.

Diffraction data were collected for dGMPNPP ternary complex crystals, and the structures show a clear density for the structural elements characteristic of the closed complex with a distinct density for the 3'-oxyanion of the primers 3'-end.

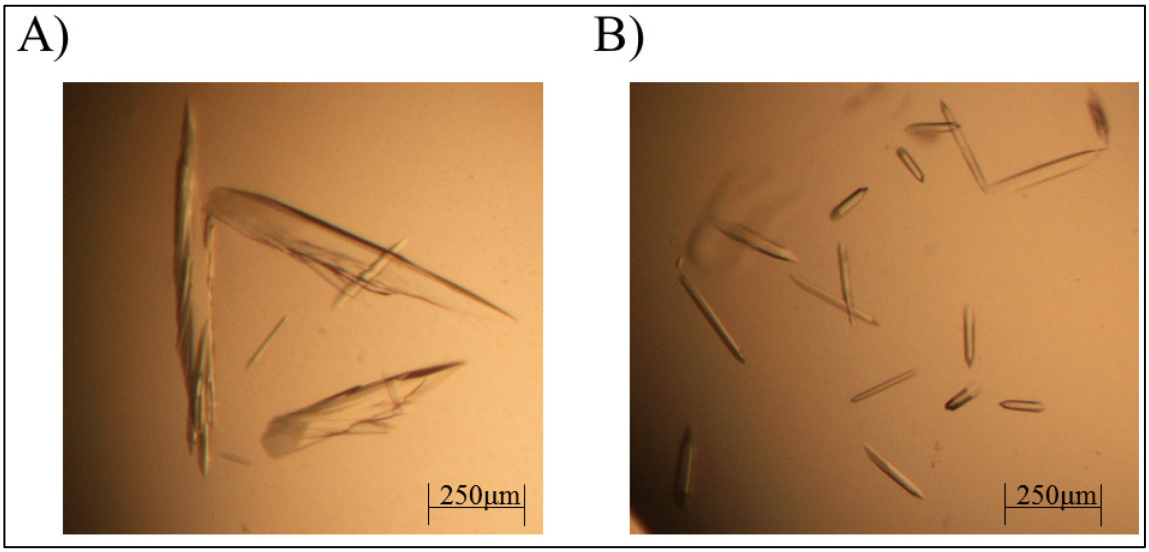


Figure A6. Optimized Conditions for Extension-ready dGMPNPP Ternary Complex. A) crystals obtained using conditions above, and B) crystals obtained with 5mM TCEP.

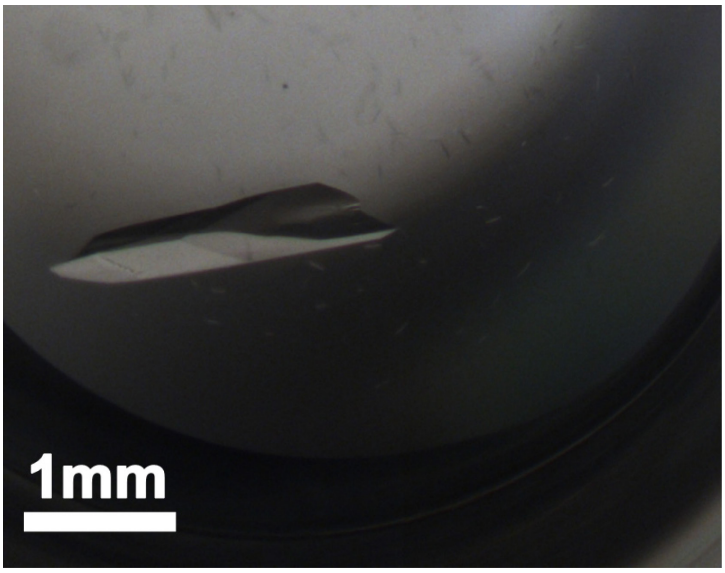
Appendix B

Neutron Scattering of the Ternary Complex of KlenTaq DNA Polymerase with the non-hydrolyzable nucleotide analogue dGMPNPP

High Flux Isotope Reactor (HFIR)
Oak Ridge National Laboratory

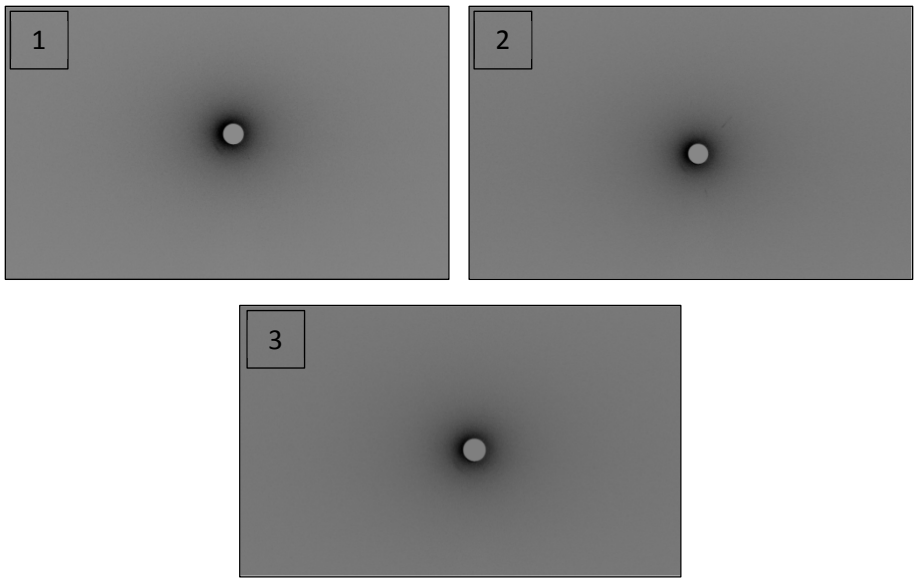
Experiment Summary

Large crystals of dGMPNPP-trapped ternary complex of KTQ281:AAG were obtained by sitting drop vapor diffusion crystallization (10-20uL drops) against a 500uL reservoir containing 11-22% PEG 3,350, 0.2M $NH_4(OAc)$, 0.2M $Mg(OAc)_2$, 0.1M Tris-Cl, pH 8.5 (RT), and 10mM TCEP. Six large crystals with approximate dimensions (1-2mm) X (0.3-0.6mm) X (0.2-0.4mm) yielding minimum and maximum cubic volumes of 0.6 and 0.48 mm^3 were transferred 1.0 mm glass capillaries containing mother liquor. Excess mother liquor was removed by capillary wicking action followed by introduction of 90% deuterated solvent plugs. To reduce fluctuations in pH, all mother liquor components except for 0.1M Tris-Cl, pH 8.5 (RT), were prepared in D_2O yielding a 90% deuterated solvent. Deuteration was allowed to proceed for three weeks via vapor diffusion at 18^oC. Crystals were then transferred to fresh capillaries filled with deuterated solvent and packed into padded 15mL conical tubes which were then sealed and prepared for transport to ORNL.

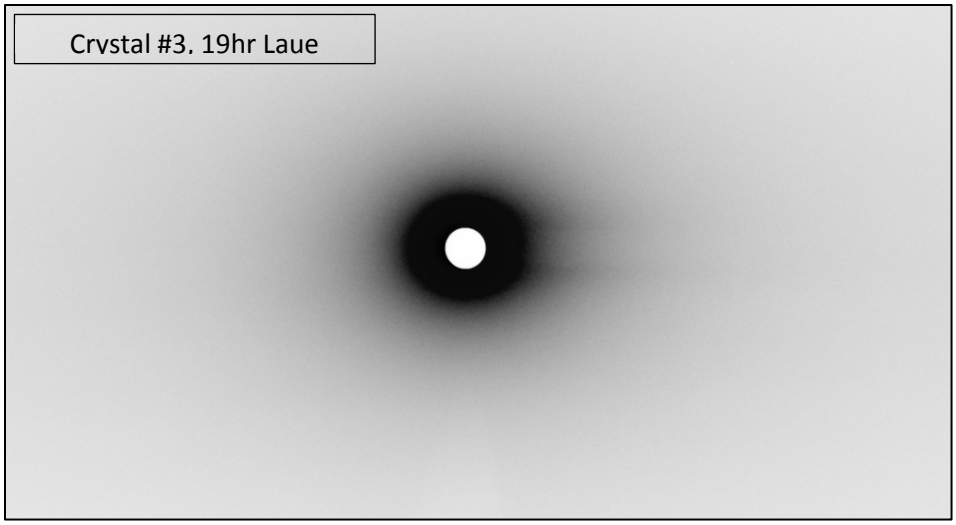


Transportation proved somewhat challenging and resulted in the loss of one of six capillaries (broken) and destruction of 2/6 large crystals (shattered into a shower of smaller crystals). The three remaining crystals also broke, but large enough fragments were quickly remounted upon arrival to preserve sample integrity and prepare for diffraction experiments. Resulting crystals could not have been larger than $0.1mm^{3,}$ and anecdotal evidence from beamline scientists suggests that the aforementioned crystal size is a minimum for obtaining reasonable neutron diffraction. Thus, the experiment was allowed to proceed but had a low probability of success.

Three crystals were exposed to a Laue (polychromatic, whole spectrum) neutron beam for 1 hour each with a 200mm fixed crystal-detector distance sufficient for a diffraction limit of 10-2.78Å on the IMAGINE single-crystal diffractometer with a neutron image plate in the High Flux Isotope Reactor (HFIR) at ORNL. The choice of reactor (HFIR) vs. spallation (SNS/MaNDi) neutron sources is limited by unit cell dimensions and crystal volume. The aforementioned crystallization condition yields "locked" ternary complexes of KTQ_{281}:AAG:dGMPNPP in the $P3_121$ space group with unit cell dimensions of a=110Å, b=110Å, c=90Å and angles of $\alpha=90^o$, $\beta=90^o$, $\gamma=120^o$ and a unit cell volume of 943102 $Å^3$. The unit cell contained 6 ASU's with one complex/ASU.



The third crystal showed weak, low-resolution Laue diffraction at 1 hour and was chosen to test an exposure time of 19 hours for a single frame. The 19hr exposure did not produce any diffraction (below) and was entirely overshadowed by background intensity accumulated during this long exposure.



Appendix C

Crystallization of *E. coli* RNA Polymerase Holoenzyme in Complex with Promoter DNA containing an UP element

Structural studies of the protein-DNA interactions occurring between the αCTD of *E. coli* RNAP and UP-element-containing promoter DNA

Escherichia coli RNA polymerase (RNAP) is a multisubunit complex composed of five subunits ($\alpha_2\beta\beta'\omega$) representing the 'core' enzyme which is capable of transcription elongation but not initiation. For promoter recognition during transcription initiation in *E. coli*, the core RNAP must form a holoenzyme with one of seven different σ factors, each responsible for recognizing a specific subset of promoters during various growth conditions. In the predominant σ^{70} holoenzyme, promoter recognition is established by the binding of up to 3 regulatory elements by σ^{70}. The indispensable -10 element is recognized by σ region 2.3-2.4 whereas the -35 element is recognized by σ region 4.2. An infrequent 'extended -10' element is recognized by σ region 3. Still, some bacterial promoters, often directing transcription of vital sequences for bacterial growth, also contain UP elements; sequences upstream of the -35 element that enhance transcription by as much as 170-fold by binding of αCTD and recruiting the remainder of the RNAP holoenzyme.

Over the years, however, it was found that these UP elements differ in their capacity to stimulate transcription in various promoters and extensive genetic and biochemical studies of bacterial growth in mutants containing sequence- and/or distance-variants, relative to the -35 element, has shown that UP elements positioned in the -43 region had the highest stimulatory effect. These UP elements can contain both distal and proximal subsites, again relative to the -35 element, and in such an instance are occupied each by one of the αCTD protomers. Due to the established protein-DNA mode of interaction that αCTD exhibits on UP elements, and protein-protein interactions with Class I activator proteins, exemplified by actions on *rrnBP*1 and *lac* promoters, respectively, it was suggested, based on models built using the crystal structures of *E. coli* αCTD-DNA and *Taq* RNAP holoenzyme-promoter DNA complexes, that the proximity of αCTD's occupying the proximal subsite in the -43 region to the -35 element occupied by σ^{70}R4.2 may facilitate interactions between αCTD and σ^{70}R4.2 that structurally explain variations in UP element-dependent stimulation of transcription and "remodelling" of the RNAP holoenzyme.

Despite the availability of crystal structures of both *Thermus thermophilus* and *Thermus aquaticus* holoenzymes in complex with promoter DNA, no direct evidence for this interaction between αCTD and σ^{70}R4.2 exists as αCTD in the aforementioned complexes is not resolved alone, bound to DNA, or interacting with the equivalent of σ^{70}R4.2 due to the flexible linker tethering αCTD to αNTD and thus to the remainder of the RNAP holoenzyme in those species. Biochemical evidence complementing this structural model of the hypothesized interaction was provided by work showing that αCTD can be crosslinked to free unbound σ^{70} that was mapped to surface-exposed residues D259 and E261 in αCTD and R603 on σ^{70}. When alanine-scanning mutagenesis and DNase I sensitivity were employed in the study of these residues, it was found that the resulting promoters had a marked deficiency in stimulation and structural alterations to the DNA-bending induced by the proposed interaction, respectively.

Structural evidence demonstrating this interaction is of paramount importance to understanding the mechanism by which this interaction is brought about but also to its contribution in potential remodelling of RNAP upon DNA binding through any structurally induced changes that result from such an interaction. While the mechanism of such an interaction and its implications for transcription initiation are important, it is my goal to extend this work to developing a structural platform for direct observation of transcription by *E. coli* RNA polymerase *in crystallo* which will allow significant progress in structure-based drug design and provide structural information as to the mechanisms of inhibition by various antibiotics.

To investigate the structural basis of protein-protein and protein-DNA interactions that occur between αCTD, σ^{70} and promoter DNA regulatory elements we used synthetic oligonucleotide fork junction scaffolds that form stable complexes of the RNA polymerase-σ^{70} holoenzyme. These synthetic fork junction scaffolds are a hybrid of the wild-type sequence of the *rrnB* P1 promoter containing the consensus UP element with a stabilizing synthetic 5'-blunt end consisting of multiple GC pairs and synthetic downstream -35 and -10 regulatory elements with varying length spacers (16-17 bases) followed by downstream (dw)DNA to +1 of the transcription start site. The rationale behind this design, at least in this early stage, is to preclude RNA polymerase from

undergoing the isomerization and open-bubble formation steps required during transcription initiation. Structural studies of isomerization and open-bubble formation are also of interest and will likely be investigated later on. Thus, these fork junction scaffolds will serve to explore holoenzyme binding to UP-element-containing promoter DNA without inducing significant structural changes prior to crystallization. The resultant model would provide valuable information as to the mechanism by which αCTD and σ^{70} interact with each other and with DNA to induce any structural changes leading to isomerization and open-bubble formation which, again, may be explored in subsequent work using extended dwDNA scaffolds.

Recombinant, *E. coli* RNA polymerase holoenzyme, *in vivo* reconstituted using a bacterial strain carrying multiple expression vectors, was purified to homogeneity (Fig. C1) and complexed at 10mg/mL in a 1:1.5 molar ratio with a synthetic UP-element-containing scaffold that readily forms a fork junction (Fig. C2), and subjected to crystallization screening utilizing commercially available screening kits. An initial hit provided small crystals when the complex was equilibrated via vapor diffusion against 50mM Na-Cacodylate, pH=6, 15mM $Mg(OAc)_2$, and 1.7M Ammonium Sulfate (Fig. C3). The condition was subsequently optimized using a commercially available additive screen and yielded three conditions that improved crystal size (Fig. C4). Crystals optimized against polyvinylpyrrolidone K were flash-frozen in liquid nitrogen and subjected to X-ray diffraction studies (Fig. C5A) at the Cornell High Energy Synchrotron Source (CHESS) but were of poor quality and yielded diffraction to ~8Å (Fig. C5B).

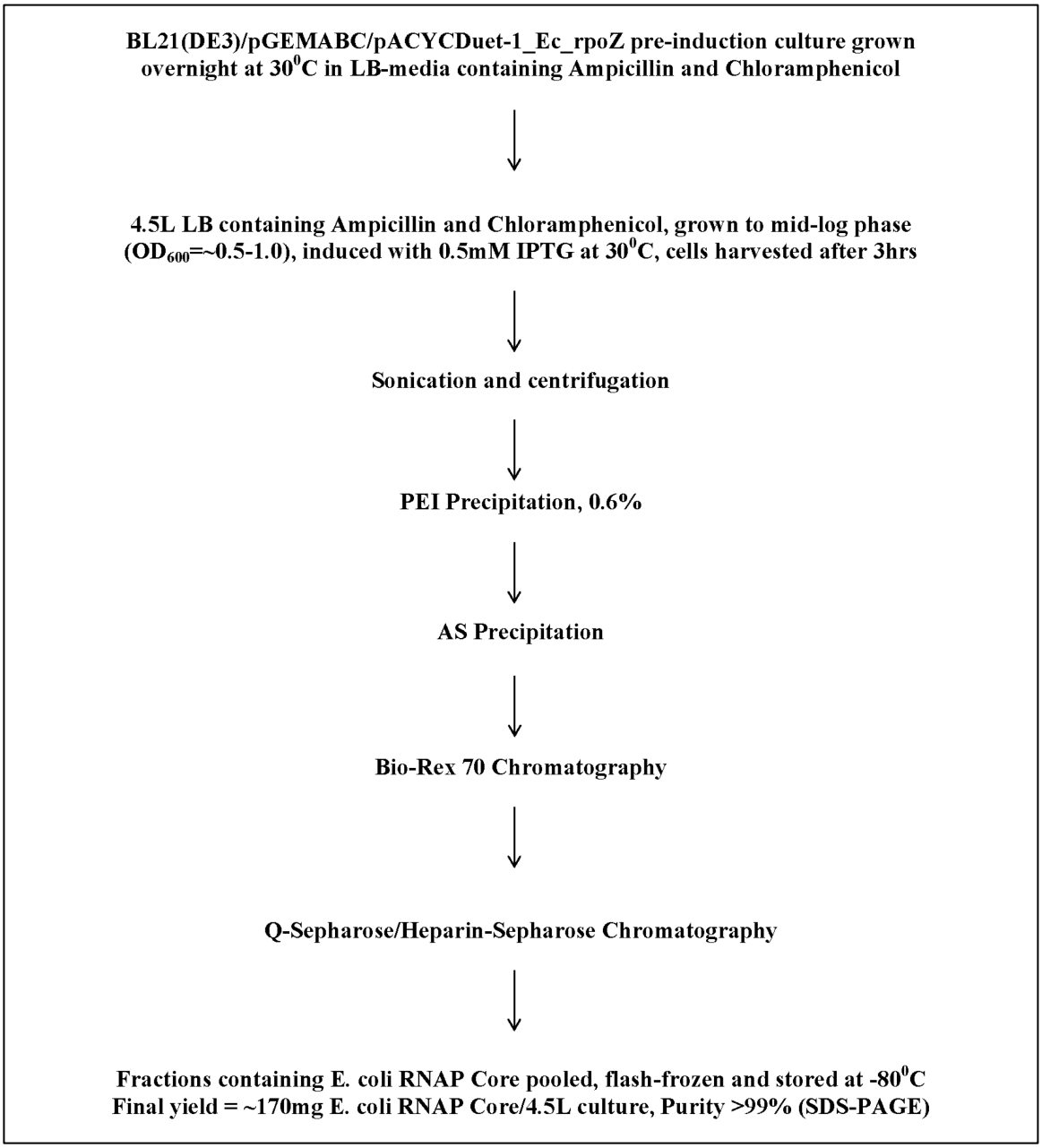


Figure C1. Purification scheme and yield of recombinant *E. coli* RNA polymerase.

```
5′-GGCCGTCAGAAAATTATTTTAAATTTCCTCTTGACAAAAGTGTTAAATTGTGCTATAAT
3′-CCGGCAGTCTTTTAATAAAATTTAAAGGAGAACTGTTTTCACAATTTAACACGA
                UP-element              -35                    -10
```

Figure C2. Synthetic scaffold DNA containing an UP-element and fork junction.

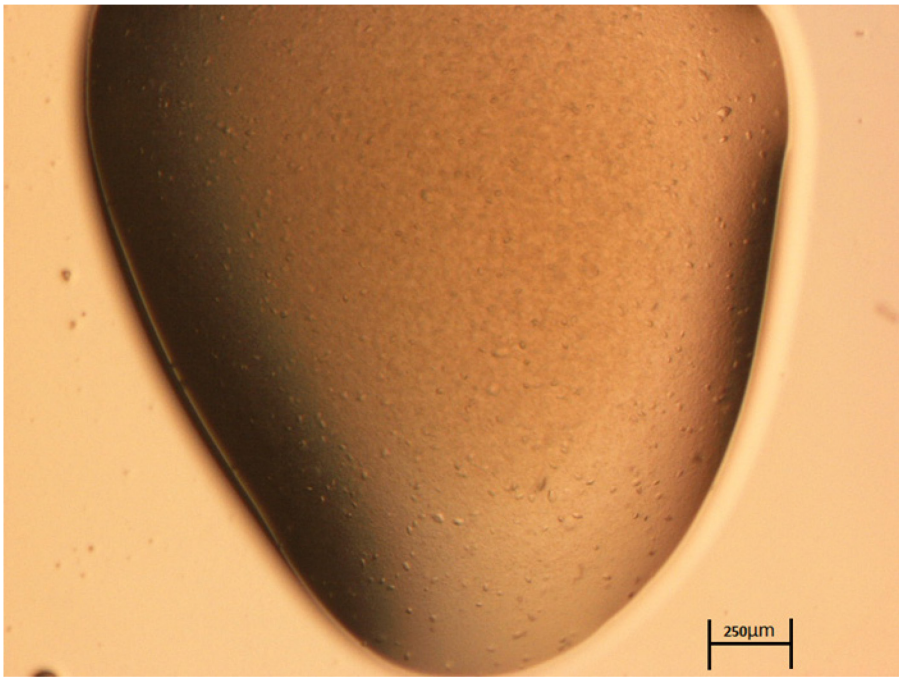


Figure C3. Initial crystallization hit for *E. coli* RNA polymerase holoenzyme in complex with UP element-containing fork junction promoter DNA.

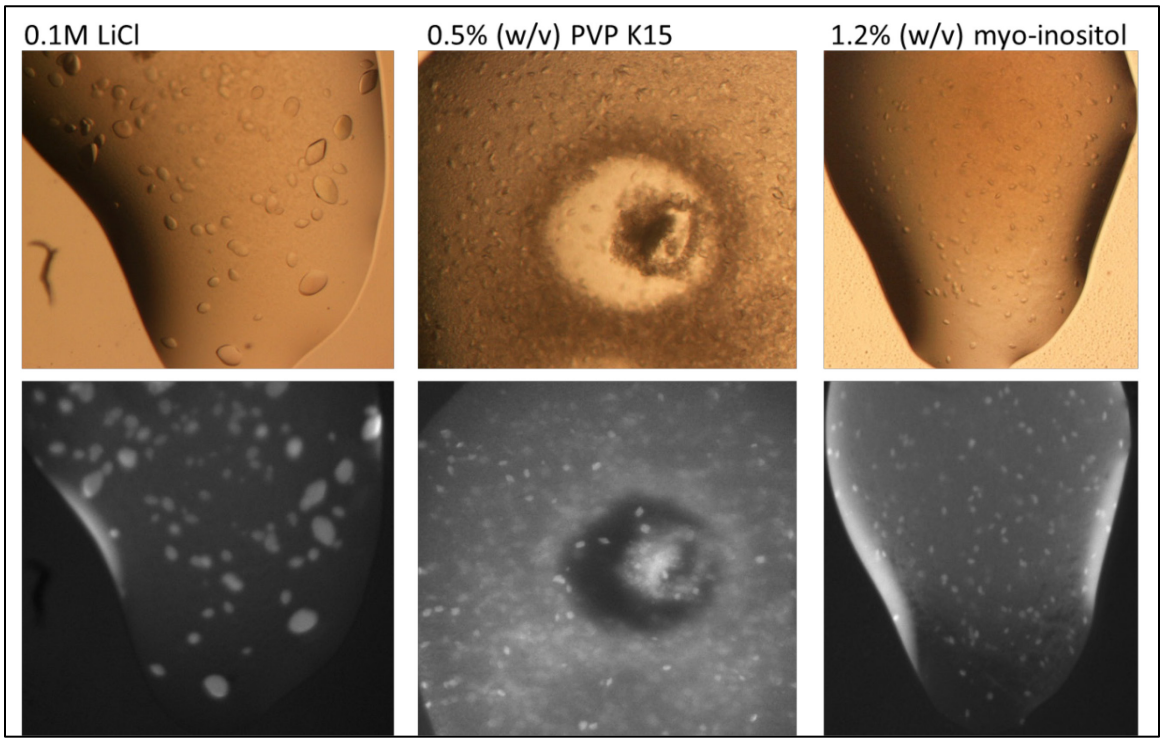


Figure C4. Additive screening to improve initial hit condition. The original condition from Fig. C3 was optimized by addition and titration of LiCl, PVP K15 or myo-inositol.

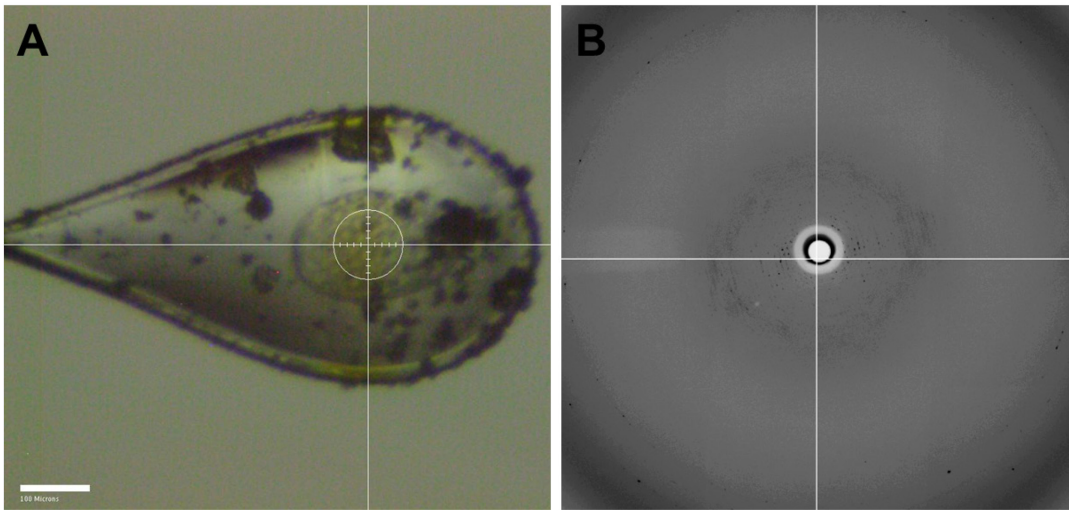


Figure C5. *E. coli* RNA polymerase holoenzyme in complex with UP element-containing fork junction promoter DNA. A) Snapshot of loop-mounted crystal in the X-ray beam path, B) Diffraction pattern obtained for crystal optimized using PVP-K15.

Appendix D

Cloning, Expression and Purification, Oligomeric States and Crystallization of MreA and Homologs

Preparation of MreABCD

MreABCD were expressed and purified as C-terminal fusions to SUMO (Smt3) to both facilitate solubility and prevent or at least decrease aggregation to aid in the characterization of native and artificial homo-oligomers. Expression was carried out at 18°C for 20 hours after addition of 0.25mM IPTG in Rosetta 2(DE3), a lysogenized BL-strain derivative carrying an accessory plasmid encoding rare tRNA codons to enhance the solubility of recombinant proteins. Cells were harvested by centrifugation for 30 minutes at 7,000xg and 4°C. Pellets were frozen at -80°C for a period of 2 hours followed by rapid thawing to room temperature to weaken the integrity of the cell walls. The Thawed pellets were resuspended in a millilitre volume of lysis buffer equivalent to twice the wet weight of the cell pellet. The lysis buffer contained 40mM Tris-Cl, pH=8.0 at 4°C, 300mM NaCl, 5% glycerol and 5mM β-Mercaptoethanol. The suspension was sonicated on ice in a Misonix S3000 sonicator at a power level of 6.5 for a total period of 3 minutes with 1-second pulses and 3 seconds rest.

The resultant extract was centrifuged for 30 minutes at 30,000xg and 4°C. The supernatant was collected, applied to a Ni-NTA column and washed extensively in lysis buffer containing 1M NaCl to dissociate any sheared and still-bound DNA that may impede analysis. The columns were then washed with lysis buffer (300mM NaCl) to baseline A_{280}, and then again with lysis buffer containing up to 50mM Imidazole to baseline A_{280}. Finally, columns were eluted with step-gradients of 250- and 500mM Imidazole. Collected fractions were analyzed on a 4-12% Bis-Tris polyacrylamide gel in MES running buffer (Fig. D1). Fractions were pooled and dialyzed overnight against lysis buffer containing 150mM NaCl and 100U of Ulp-1 protease to cleave the fusions and yield the respective Mre's in native form. After overnight dialysis, samples were once again applied to Ni-NTA columns to remove Ulp-1 protease, uncleaved Mre's, and protein impurities that bound strongly to the metal chelate matrix prior to dialysis. Mature MreABCD were analyzed on the same 4-12% Bis-Tris polyacrylamide gel in MES running buffer (Fig. D1) before being flash-frozen in liquid nitrogen and stored at -80°C.

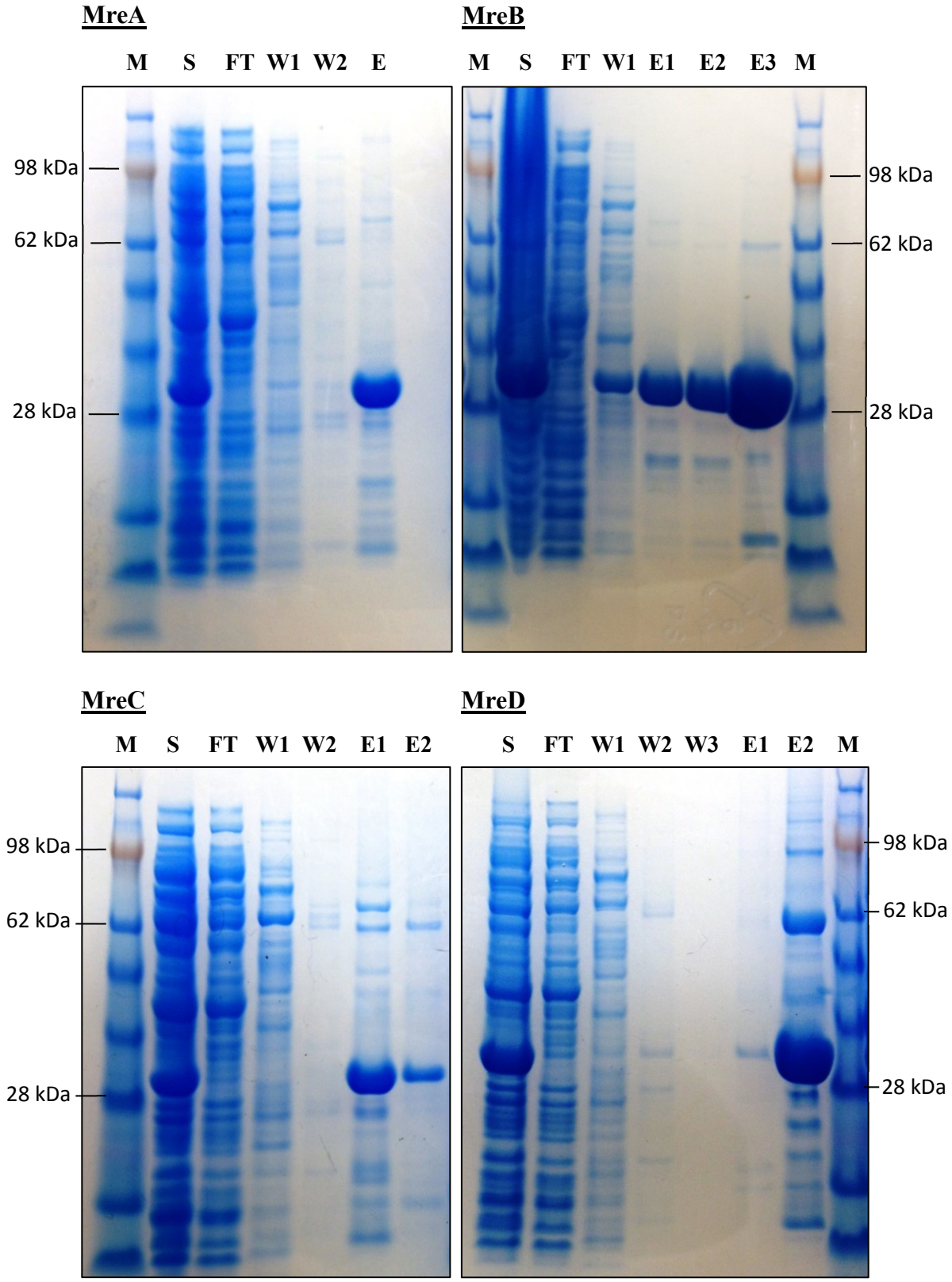


Figure D1. SDS-PAGE analysis of fractions obtained during purification of Mre homologs.

Preliminary Analysis of MreABCD

To investigate the oligomeric state of MreABCD, 500uL samples of 5mg/mL MreA, C or D, and 20mg/mL of MreB, which was much more soluble relative to other homologues, were analyzed on an analytical Superdex 75 10/300 GL size-exclusion column (GE Healthcare) with a bed volume of approx. 24mL and a fractionation range of M_r 3,000-75,000. Initial predictions from bioinformatics analyses and structural modelling suggested that Mre's would predominantly be homodimers with a M_r of ~30,000-35,000. In subsequent work, a similar column with a broader fractionation range (Superdex 200) of M_r 10,000-650,000 will be used to better approximate the apparent molecular weight of the higher-order homo-oligomers. As it is, Superdex 75 has an exclusion limit of M_r 100,000, which approximates the presence of octamers.

As can be seen from the chromatograms of each homologue, monomeric species do not exist within the population but rather the predominant species in MreA, C and D are homo-octamers (~90kDa) followed by homodimers (~32kDa) for MreC and MreD. MreA, under these conditions, does not appear to exist as a homodimer but rather a homotetramer. It should be noted that MreC and MreD show a distinct, but minor, middle peak with retention indicative of a tetrameric form, which may or may not be artifactual in nature (due to protein aggregation, irreversible denaturation by domain swapping), and requires further analysis. MreA, on the other hand, shows a spectrum of species retained at intermediate weights between the homotetramer and homo-octamer. Again, these may be artifactual and must be re-evaluated. MreB represents the most interesting homologue with various implications regarding mechanism. Even at such high saturations in concentration, it exists predominantly in the homotetrameric form with remarkable stability.

As can be seen from accompanying gel images, when samples of respective peaks from each chromatogram were boiled and analyzed by denaturing SDS-PAGE, all higher molecular weight species dissociated to the monomeric forms, indicating that these peaks, regardless of size are all composed of the monomers from each homologue.

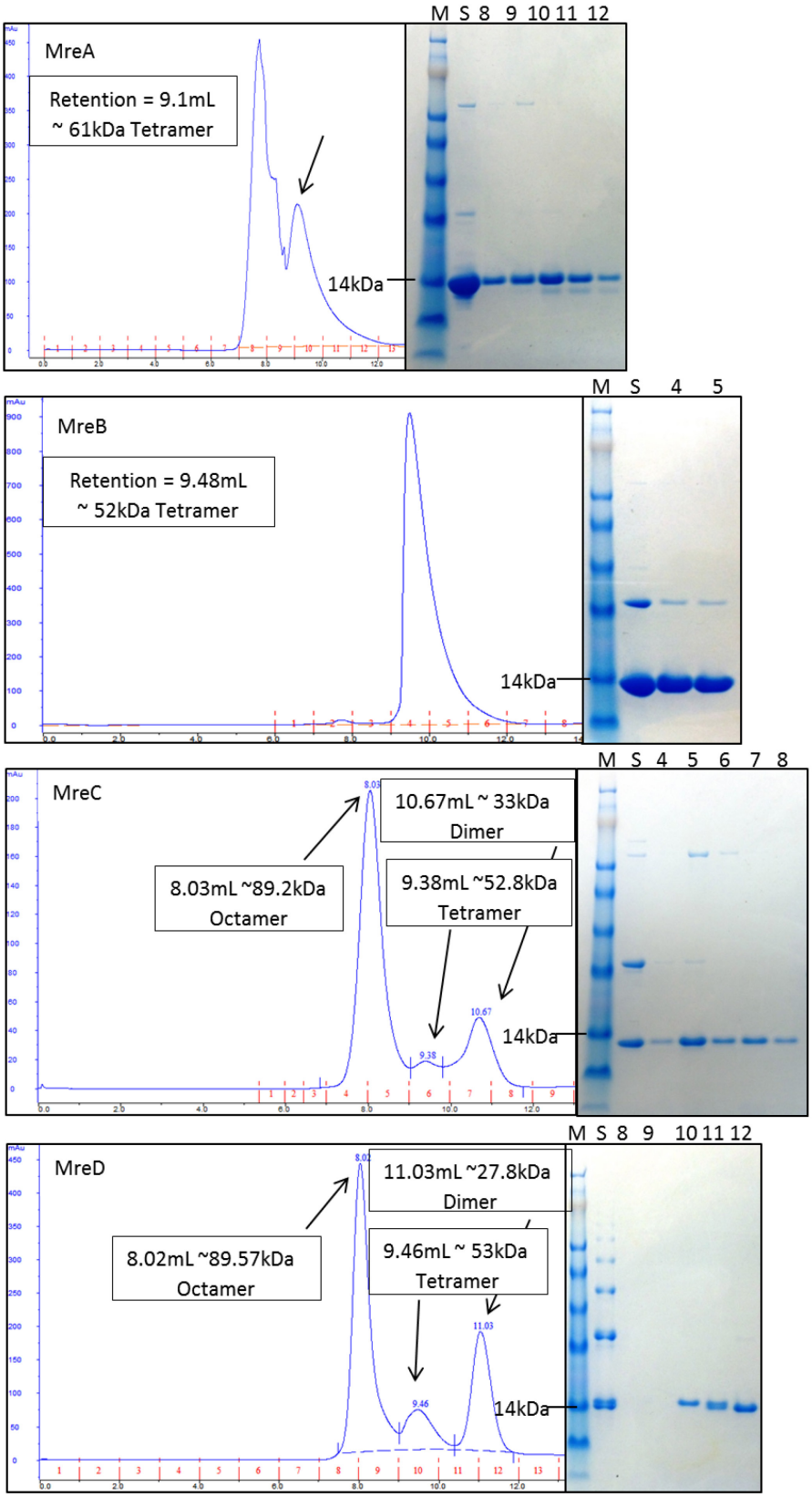


Figure D2. Oligomeric states of Mre homologs.

Crystallization of MreABCD

Recombinant MreA, -B, -C and –D were purified to homogeneity and subjected to crystallization screening utilizing commercially available screening kits. An initial hit provided large crystals of MreD (Fig. D3) when the drop was equilibrated via vapor diffusion against 50mM Tris-Cl, pH=7.5, 40mM $Mg(OAc)_2$, 20% (v/v) MPD. Crystals of MreD were flash-frozen in liquid nitrogen and subjected to X-ray diffraction studies at the Cornell High Energy Synchrotron Source (CHESS) but yielded no diffraction.

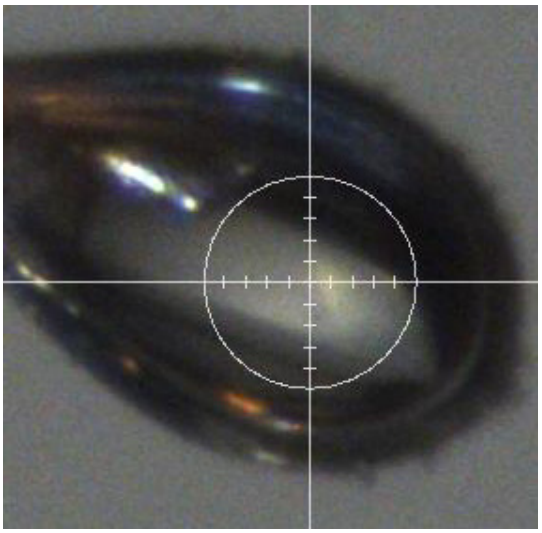


Figure D3. Loop-mounted crystal of MreD in the path of the X-ray beam.

Appendix E

Genome-Wide Binding Patterns of the Global Transcriptional Regulator MreA in the Methanogen *Methanosarcina acetivorans*

E.1 Transcription Initiation in Archaea

Archaea possess a unique chimeric landscape for transcription. Whereas archaeal TFs more closely resemble prokaryotic counterparts (1), the mechanism by which transcription is achieved, in fact, more closely resembles eukaryotic organisms (2,3). Archaea contain basal transcription machinery, akin to eukaryotic counterparts, wherein a pre-initiation complex (PIC) must first be formed prior to recruitment of RNA polymerase (RNAP) (2,3). Archaeal RNAPs also more closely resemble eukaryotic RNAP(II) in subunit composition –8-13 subunits in thus far characterized archaeal genomes– and complexity.

Archaeal genomes are organized in a similar fashion to prokaryotic genomes with dense operonic organization and short intergenic spacing (4). A basal archaeal promoter, in mapped and characterized genomes, consists of two regulatory elements upstream of the transcription start site (TSS), a TATA-box centred at position -26/-27 with respect to the TSS, and a factor B response element (BRE) located immediately upstream of the TATA-box (5).

The core bacterial RNAP is composed of 5 subunits ($\alpha_2\beta\beta'\omega$) which can elongate transcription products, requires one class of TFs, σ-subunits, to facilitate DNA-binding to promoter elements and transcription initiation. In contrast, archaeal basal transcription machinery includes (at least) two or more TFs to initiate transcription. Transcription initiation in archaea is hierarchical with the two general TFs, TBP and TFB, alone or in concert with other general or specific TFs bind to regulatory regions upstream of the TSS to form the pre-initiation complex (PIC) prior to recruitment of RNAP to initiate transcription (6).

PIC formation is also hierarchical and sequential. TBP must first bind to the TATA-box prior to TFB interactions with and recruitment to the BRE. TFB recognizes the DNA-bound TBP and makes specific protein-protein interactions with TBP while simultaneously recognizing the BRE sequence immediately upstream of the TATA-box to determine the orientation of the PIC (7). It is at this point that RNAP is recruited to the PIC in proper positioning and orientation to initiate transcription at the TSS.

E.2 Archaeal Transcription Factors (TFs)

Prokaryotic transcription factors (TFs) are most commonly single-component systems that contain the necessary domains for activity. This organization in prokaryotic TFs provides for a minimum of two domains, a DNA-binding domain and a sensory domain responsible for stimulus-response (8). Stimuli can be small effector molecules (ligands available as metabolites or organic compounds taken up by the cell), protein-protein interactions, temperature fluctuations, and cellular redox state. Once the condition for a specific stimulus is met, the stimulus-response domain, through a variety of mechanisms, conveys the output in a manner that alters DNA-binding properties, thereby altering transcription levels and ultimately gene expression.

Characterized archaeal TFs contain a significant proportion of homologues to bacterial activators and repressors. However, a more striking point to make regarding most of the sequenced archaeal genomes thus far is the fact that the number of transcription factors, predicted or experimentally determined, is significantly lower per genome size relative to bacterial and eukaryotic counterparts of similar genome size (9). Evolutionarily, more complex genomes require much more sophisticated machinery to regulate gene expression. Where ~10% of bacterial genomes encode for TFs, archaeal genomes of similar sizes contain nearly half that percentage despite the increase in genome complexity (10). Research in the field has pointed to multiple hypotheses to explain these disparate values. It is, of course, feasible that the novelty of archaea and what little is known thus far could point to the fact that we have yet to delineate archaeal genome complexity fully and the plethora of potential mechanisms archaea can employ to diversify their regulatory repertoire. It may also be the case that archaea do in fact contain a smaller proportion of TFs but are able to regulate their physiology by alternative strategies that exploit the eukaryote-like basal transcription machinery. i.e., alternative protein-protein interaction profiles with TBP, TFB, and RNAP with or without interactions with accessory proteins can bring about very different responses to gene expression, as will be discussed below.

Nearly half of the archaeal transcription factors characterized thus far are very small in size (100-200 residues), even compared to their bacterial homologues, which

suggests that domain organization in archaeal TFs is unlike their bacterial homologues. In fact, when sequences of characterized archaeal TFs are aligned against their bacterial homologues, nearly 75% of the archaeal TFs are single-domain TFs (11)—that is, they only contain the DNA-binding domain. This finding suggests that a stimulus-response (ligand-binding) domain may achieve its functional role by participating in protein-protein interactions with these small, single-domain TFs, similar to eukaryotic regulation by TFs. Taken together, research seems to point to a combinatorial interplay of TFs, wherein a TF of a particular class, identifying only a subset of promoters, can interact with different ligand-binding domain-containing proteins under different physiological conditions to bring about activation or repression of the particular subset of promoters to which the TF can bind. Thus, the possible mechanisms by which archaea can regulate transcription may be in the form of homo- or hetero-oligomeric complex formation in response to small-molecule effectors or changes in the environment. Still, the most striking feature in some archaeal TFs is regulation of transcription by alternating oligomeric states (12).

The Leucine-responsive regulatory protein (Lrp)/Asparagine synthase C transcription factor (AsnC) family of TFs is both global and specific regulators of transcription with activation and repression being brought about by oligomeric state (13). The average length of TFs in the AsnC family is 160 residues which can form dimers, tetramers, and octamers to facilitate various responses. Even more interesting is the *Thermococcus* regulator of maltose binding (TrmB) family of TFs characterized from *Thermococccus litoralis* (*Tli*) and *Pyrococcus furiosus* (*Pfu*). Both TFs exist predominantly in two oligomeric states, tetramers and octamers, depending on the stimulus present, and facilitate occlusion by structural hindrance of the binding of TBP and TFB (14). TrmB from *Pfu* exists as a tetramer and converts to an octamer upon binding of an inducer, maltose or maltotriose to alter its DNA-binding properties to activate or repress transcription (15).

E.3 Structural Biology of Archaeal TF:DNA Complexes and Interactions

The structural basis for TF function lies in its sequence-specific interaction with DNA. Since the single-domain archaeal TFs, which only contain the DNA-binding domain, are more bacterial in nature than eukaryotic, they recognize and bind specific DNA sequences in a similar fashion to their bacterial counterparts. DNA sequence motifs almost invariably have a semi-palindromic nature, which reflects the two-fold symmetry of the binding proteins caused by homodimerization (16).

Structural stability in TF:DNA binding depends on binding motifs that can range in size from 8-24bp (17), with typical binding motifs having sizes between 11- and 17bp with less sequence-specificity in the centre of the motif. The non-specific interaction in the centre of the sequence motif is stabilized by the flanking sequence-specific interactions. The bulk of the sequence-specific interactions occurs in the major groove of DNA by the symmetry-determining half-sites (because of the semi-palindromic nature of the sequence and the helical architecture of B-form DNA). This interaction with the major groove of the target DNA sequence occurs by means of a recognition helix in Helix-turn-Helix (HTH) motif-containing TFs (Fig. E.1) (18). The alignment of the two half-sites mentioned above in a binding site within the major groove of DNA has been demonstrated to be highly constrained (19), primarily due to the rigidity of the recognition helix and its sequence-specificity. Research has shown that a 1-2bp insertion in the centre of the binding site (the non-specific sequence) for the recognition helix causes a significant decrease in binding affinity (19). However, in some cases, a 1bp deletion completely disrupts DNA binding, which may be indicative of limited conformational flexibility of the transcription factor, the recognition helix, or a combination of effects (19).

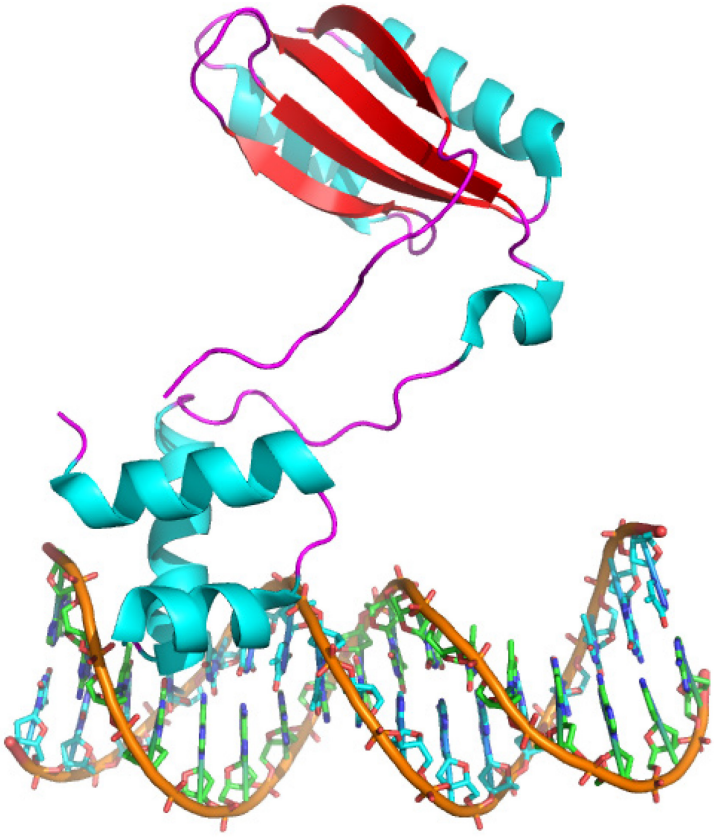
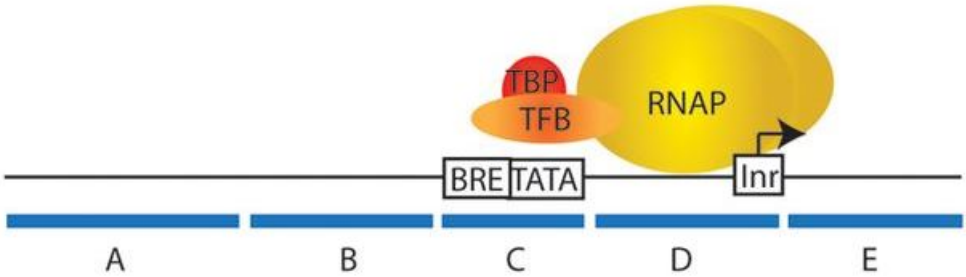


Figure E.1 *Pyrococcus* OT3 putative wHTH transcriptional regulator in complex with DNA. Motif stabilized by hydrophobic core at the interface of the 3 helices which helps to present helix 3 for sequence-specific DNA recognition. The DNA recognition helix inserts into the major groove of duplex DNA allowing side chains to make base-specific and sugar-phosphate backbone contacts. PDB ID: 2E1C. Figure created using PyMOL.

Most TFs recognize an array of binding sites throughout the genome to facilitate global regulation. Some TFs even have multiple binding sites within a single intergenic sequence to modulate activity levels. In such instances, a primary binding site facilitates regulation and auxiliary binding sites assist in maintaining or increasing occupancy of the primary site by the TFs in question (20). However, for such activity to occur, there are spatial constraints on the placement of multiple binding sites which are governed by the DNA helical structure in a manner that allows binding of TFs to the same face of the helix. Centre-to-centre distances corresponding to two or three full helical turns between binding sites are most commonly found (21). Thus, altering this spacing with merely a half-helical turn can have a profound effect on DNA binding, affinity, and the resultant regulatory effect.

The primary determinant of whether a TF acts as an activator or repressor is most often the location of the binding site with respect to the promoter in question (Fig. E.2). Activators tend to bind upstream of the promoter, and repressors overlapping with, or downstream of the promoter (22). In archaea, repressors have been found to function in two ways. Repressors either bind at a site overlapping the BRE and TATA-box thereby precluding access to these sites by TBP and TFB, or they may bind downstream of the TATA-box and, while not occluding TBP and TFB binding, prevent recruitment of RNAP to the TBP-TFB-DNA pre-initiation complex (23). Other mechanisms of repression exist, and some TFs can repress transcription initiation by a combination of the two mechanisms above.

Activators often bind at a single or primary site located immediately upstream of or overlapping, the BRE. For 15-17bp sites, this binding positions the TF on the same helical face as the basal TFs (TBP, TFB) resulting in a centre-to-centre distance of approximately two helical turns between the TF binding site and the TATA-box. Interestingly, there are cases in which TFs activate a promoter from a site farther upstream with binding progressing cooperatively towards the promoter in a concentration-dependent manner (24).



Type	Mechanism	TF binding				
		A	B	C	D	E
Activation	Recruitment of TBP	(+)	+	-	(+)	-
Activation	Recruitment of TFB	(+)	+	-	(+)	-
Repression	Inhibition of TBP/TFB binding	-	-	+	-	-
Repression	Displacement of TBP/TFB	-	(+)	-	-	-
Repression	Inhibition of RNAP recruitment	-	-	-	+	(+)
Repression	Roadblock	-	-	-	-	(+)

Figure E.2 Archaeal TF binding location effect on regulation.

Overview of molecular mechanisms of archaeal transcription regulation and binding locations for which the corresponding mechanism has been observed, ranging from far upstream of promoter (A), close upstream of promoter (B), overlapping core promoter (C), downstream of promoter (D) to downstream of TSS (E). The symbol + indicates that the mechanism has been demonstrated, the symbol (+) indicates that the mechanism has been postulated without conclusive proof.

Peeters *et al*. (2013) Cis-regulatory logic in archaeal transcription. *Biochem. Soc. Trans.* **41**, 326–331

In contrast to repressors, activators are generally associated with promoters having large deviations from the consensus BRE sequence, TATA-box sequence, or both with activation *in vitro* only occurring in the presence of minimal concentrations of basal TFs (25). This phenomenon points to a concentration-dependent mechanism of action that is mediated by protein-protein interactions between activators and TBP, TFB or both.

To facilitate global regulation, some TFs can act as autoregulators to either activate or repress transcription in the previously discussed concentration-dependent manner. The relevance here lies in the fact that such a function was characterized in TrmB-like (TrmBL2) TFs to which the Mre's belong. HTH-motif containing TFs, such as LrpB from *Sulfolobus solfataricus* which belongs to the TrmB family, can have opposite effects on the same promoter in a concentration-dependent manner (Fig. E.3) (26). At very low concentrations, the *Sso* LrpB binds to three different locations upstream of the promoter, with the two flanking sites having high binding affinity, and the middle site having a low binding affinity, such that full occupancy of the middle site is only achieved when the two outer sites are fully occupied and participate in protein-protein interactions with the middle-site to stabilize it (27). This cooperativity in binding tends to increase with increasing concentrations of LrpB and results in deformation of the DNA containing that binding region which ultimately represses transcription from that promoter.

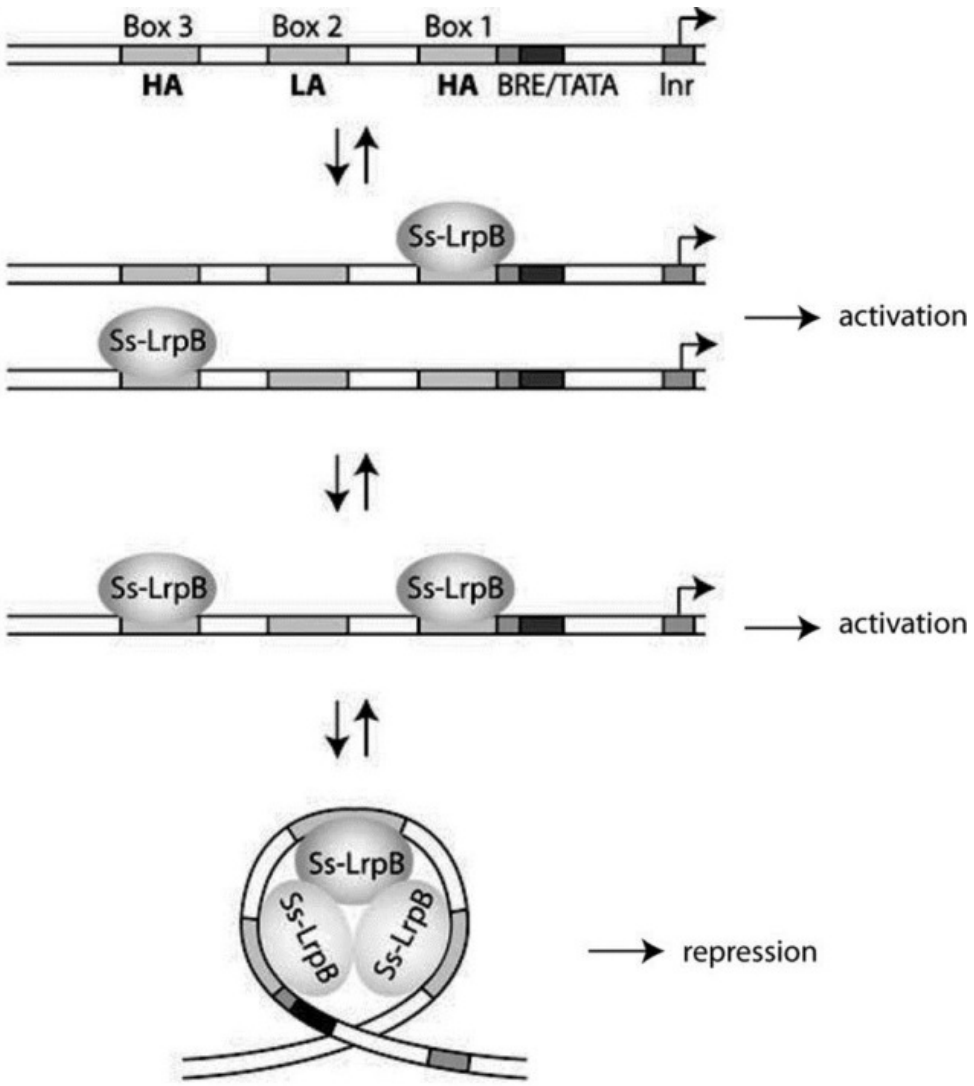


Figure E.3 Autoregulation by the *S. solfataricus* TF Ss-LrpB.

A hypothetical model of the autoregulatory process of Ss-LrpB. Binding sites of Ss-LrpB (Box1, Box2 and Box3) and the main promoter elements BRE, TATA box and initiator (Inr), containing the first transcribed nucleotide, are indicated (HA, high affinity; LA, low affinity). The scheme displays the major Ss-LrpB-DNA-binding states that exist in a dynamic equilibrium, and the assumed regulatory output (activation or repression) for each of these binding states.

Peeters *et al*. (2013) Cis-regulatory logic in archaeal transcription. *Biochem. Soc. Trans.* **41**, 326–331

E.4 *Methanosarcina acetivorans* Methanogen Regulator of Energy Conversion

Methanosarcina acetivorans is a recently discovered acetotrophic methanogen whose genome, one of the largest in archaea, has been sequenced. Unlike other methanogens which utilize a single pathway for methanogenesis, the genus *Methanosarcina* can utilize all three known pathways for methanogenesis using a variety of one-carbon compounds, acetate or carbon dioxide (28). These methanogens are remarkably versatile organisms with a diverse physiology regulated by basal transcription machinery that contains transcription factor TFIIB (TFB), three putative TATA-box binding proteins (TBP's) (29) and a polymerase (Pol) II-like RNA polymerase, and bring about the regulation of their physiology via an interplay between the aforementioned components and other general and specific transcription factors.

Proteomic studies of differentially grown *M. acetivorans* in which acetate or methanol was utilized as substrate for methanogenesis identified a significant number of proteins elevated, and others depressed, in response to the substrate utilized, indicating differential involvement in regulating the acetotrophic and methylotrophic pathways (30). One such hypothetical protein, MA3302, was upregulated 250-fold when the organism utilizes acetate as growth substrate, with sequence identity to the N-terminal DNA-binding domain of *Thermococcus* regulators of maltose binding (TrmB) (31) but lacking the C-terminal sugar sensory domain making it homologous to TrmBL2 transcription factors (32). Using genetic and transcriptomic analyses, MA3302 was found to be elevated during growth on acetate and simultaneously upregulated genes unique to the acetotrophic pathway and downregulated genes unique to the methylotrophic pathway (30). A Δ*mreA* strain showed a marked decrease in transcripts of genes involved in the acetotrophic pathway and increase in transcripts of genes involved in the methylotrophic pathway, further supporting a simultaneous role in both activation and repression, respectively, of transcription of genes unique to the acetotrophic or methylotrophic pathways (33). This duality resulted in MA3302 becoming designated Methanogen regulator of energy-conversion (MreA).

Homologs of MreA from the Leucine-responsive regulatory protein (Lrp), and Asparagine synthase C transcription factor (AsnC) families of transcription factors,

similar to the TrmB-family are global and specific regulators of transcription with activation and repression being brought about by altered oligomeric states (Fig. E.4) (34). Many of the aforementioned TFs exist predominantly in two oligomeric states, tetramers and octamers, and facilitate activation or repression by adopting an oligomeric state that facilitates or interferes with TBP1-, TFB-binding, or both (35).

While MreA was previously shown to have sequence-specific binding to promoter DNA (33) sequence identity, and potential variations thereof based on growth substrate were not known. Here we use ChIP-exo (36,37) to map the genome-wide binding patterns of MreA and TBP1 and show that, in one instance, MreA binds to TATA-boxes simultaneously with the major TATA-box binding protein in *M. acetivorans* (TBP1) and results in displacement of TBP1 to the opposite face of the TATA-box.

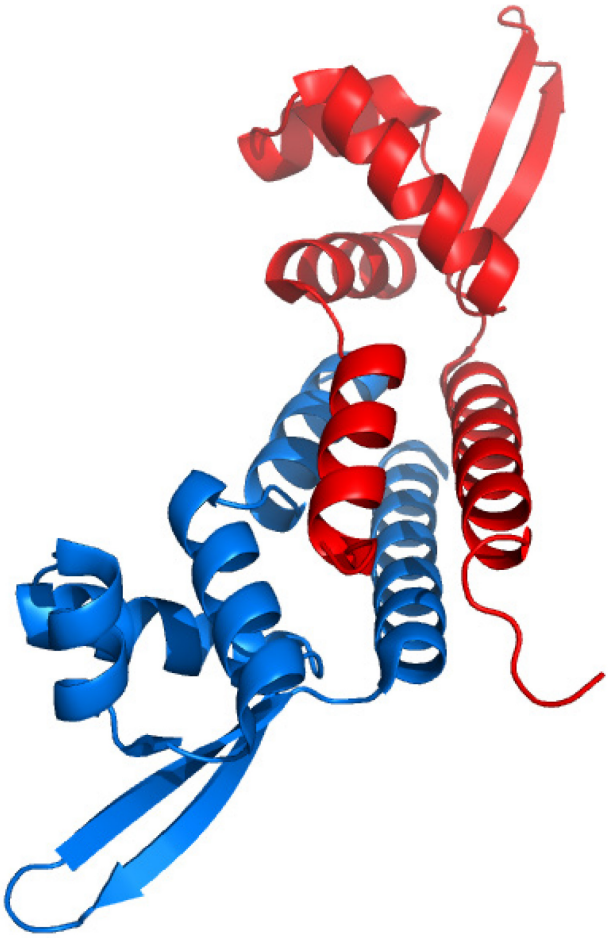


Figure E.4 *Archaeoglobus fulgidus* putative HTH transcription regulator. A TrmB-like transcriptional regulator homologous to *Methanosarcina acetivorans* MreA in a dimer configuration showing each monomer in different colours. PDB ID: 1SFX. Figure created using PyMOL.

References

1. Bell, S.D. (2005). Archaeal transcriptional regulation – variation on a bacterial theme? *Trends Microbiol.* **13**, 262–265.

2. Geiduschek, E.P. and Ouhammouch, M. (2005) Archaeal transcription and its regulators. *Mol. Microbiol*. **56**, 1397–1407.

3. Grohmann, D. and Werner, F. (2011). Recent advances in the understanding of archaeal transcription. *Curr. Opin. Microbiol*. **14**, 318–324.

4. Price, M.N., Arkin, A.P., Alm, E.J. (2006). The life-cycle of operons. *PLoS Genet*. 2, e96.

5. Soppa, J. (1999). Normalized nucleotide frequencies allow the definition of archaeal promoter elements for different archaeal groups and reveal base-specific TFB contacts upstream of the TATA box. *Mol. Microbiol*. **31**, 1589–1592.

6. Hirata, A. and Murakami, K.S. (2009). Archaeal RNA polymerase. *Curr. Opin. Struct. Biol.* **19**, 1–8.

7. Bell, S.D., Kosa, P.L., Sigler, P.B. and Jackson, S.P. (1999). Orientation of the transcription preinitiation complex in archaea. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13662–13667.

8. Pe'rez-Rueda, E. and Janga, S.C. (2010). Identification and genomic analysis of transcription factors in archaeal genomes exemplifies their functional architecture and evolutionary origin. *Mol. Biol. Evol*. **27**, 1449–1459.

9. Perez-Rueda, E., Collado-Vides, J. (2001). Common history at the origin of the position function correlation in transcriptional regulators in archaea and bacteria. *J. Mol. Evol.* **53**, 172–179.

10. Perez-Rueda, E., Collado-Vides, J. (2000). The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res.* **28**, 1838–1847.

11. Aravind, L., Koonin, E.V. (1999). DNA-binding proteins and evolution of

transcription regulation in the archaea. *Nucleic Acids Res.* **27**, 4658–4670.

12. Koike, H., Ishijima, S.A., Clowney, L., Suzuki, M. (2004). The archaeal feast/famine regulatory protein: potential roles of its assembly forms for regulating transcription. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2840–2845.

13. Krug, M., Lee, S.J., Diederichs, K., Boos, W., Welte, W. (2006). Crystal structure of the sugar binding domain of the archaeal transcriptional regulator TrmB. *J. Biol. Chem.* **281**, 10976–10982.

14. Lee, S.J., Engelmann, A., Horlacher, R., Qu, Q., Vierke, G., Hebbeln, C., Thomm, M., Boos, W. (2003). TrmB, a sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter from the hyperthermophilic archaeon *Thermococcus litoralis*. *J. Biol. Chem.* **278**, 983–990.

15. Lee, S.J., Moulakakis, C., Koning, S.M., Hausner, W., Thomm, M., Boos, W. (2005). TrmB, a sugar sensing regulator of ABC transporter genes in *Pyrococcus furiosus* exhibits dual promoter specificity and is controlled by different inducers. *Mol. Microbiol.* **57**, 1797–1807.

16. Peng, N., Xia, Q., Chen, Z., Liang, Y.X. and She, Q. (2009). An upstream activation element exerting differential transcriptional activation on an archaeal promoter. *Mol. Microbiol.* **74**, 928–939.

17. Brouns, S.J.J., Walther, J., Snijders, A.P.L., van de Werken, H.J.G., Willemen, H.L.D.M., Worm, P., de Vos, M.G., Andersson, A., Lundgren, M., Mazon, H.F. et al. (2006). Identification of the missing links in prokaryotic pentose oxidation pathways: evidence for enzyme recruitment. *J. Biol. Chem.* **281**, 27378–27388.

18. Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M. and Iyer, L.M. (2005). The many faces of the helix–turn–helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.* **29**, 231–262.

19. Peeters, E., Wartel, C., Maes, D. and Charlier, D. (2007). Analysis of the DNA-binding sequence specificity of the archaeal transcriptional regulator Ss-LrpB from

Sulfolobus solfataricus by systematic mutagenesis and high resolution contact probing. *Nucleic Acids Res.* **35**, 623–633.

20. Ochs, S.M., Thumann, S., Richau, R., Weirauch, M.T., Lowe, T.M., Thomm, M. and Hausner, W. (2012). Activation of archaeal transcription mediated by recruitment of transcription factor B. *J. Biol. Chem.* **287**, 18863–18871.

21. Ouhammouch, M., Langham, G.E., Hausner, W., Simpson, A.J., El-Sayed, N.M.A. and Geiduschek, E.P. (2005). Promoter architecture and response to a positive regulator of archaeal transcription. *Mol. Microbiol.* **56**, 625–637.

22. Bell, S.D., Cairns, S.S., Robson, R.L. and Jackson, S.P. (1999). Transcriptional regulation of an archaeal operon *in vivo* and *in vitro*. *Mol. Cell* **4**, 971–982.

23. Ouhammouch, M., Dewhurst, R.E., Hausner, W., Thomm, M. and Geiduschek, E.P. (2003). Activation of archaeal transcription by recruitment of the TATA-binding protein. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5097–5102.

24. Peeters, E., Albers, S.-V., Vassart, A., Driessen, A.J.M. and Charlier, D. (2009). Ss-LrpB, a transcriptional regulator from *Sulfolobus solfataricus*, regulates a gene cluster with a pyruvate ferredoxin oxidoreductase encoding operon and permease genes. *Mol. Microbiol.* **71**, 972–988.

25. Peng, N., Ao, X., Liang, Y.X. and She, Q. (2011). Archaeal promoter architecture and mechanism of gene activation. *Biochem. Soc. Trans.* **39**, 99–103.

26. Peeters, E., Thia-Toong, T.-L., Gigot, D., Maes, D. and Charlier, D. (2004). Ss-LrpB, a novel Lrp-like regulator of *Sulfolobus solfataricus* P2, binds cooperatively to three conserved targets in its own control region. *Mol. Microbiol.* **54**, 321–336.

27. Peeters, E., Willaert, R., Maes, D. and Charlier, D. (2006). Ss-LrpB from *Sulfolobus solfataricus* condenses about 100 base pairs of its own operator DNA into globular nucleoprotein complexes. *J. Biol. Chem.* **281**, 11721–11728.

28. Hovey, R., Lentes, S., Ehrenreich, A., Salmon, K., Saba, K., Gottschalk, G., Gunsalus, R.P., Deppenmeier, U. (2005). DNA microarray analysis of *Methanosarcina*

mazei Go1 reveals adaptation to different methanogenic substrates. *Mol. Genet. Genomics* **273**, 225–239.

29. Reichlen, M.J., Murakami, K.S., and Ferry, J.G. (2010). Functional analysis of the three TATA binding protein homologs in *Methanosarcina acetivorans*. *J. Bacteriol.* **192**, 1511-1517.

30. Li, L., Li, Q., Rohlin, L., Kim, U., Salmon, K., Rejtar, T., Gunsalus, R.P., Karger, B.L., Ferry, J.G. (2007). Quantitative proteomic and microarray analysis of the archaeon *Methanosarcina acetivorans* grown with acetate versus methanol. *J. Proteome Res.* **6**, 759 –771.

31. Lee, S.J., Moulakakis, C., Koning, S.M., Hausner, W., Thomm, M., Boos, W. (2005). TrmB, a sugar sensing regulator of ABC transporter genes in Pyrococcus furiosus exhibits dual promoter specificity and is controlled by different inducers. *Mol. Microbiol.* **57**, 1797–1807.

32. Maruyama, H., Shin, M., Oda, T., Matsumi, R., Ohniwa, R.L., Itoh, T., Shirahige, K., Imanaka, T., Atomi H., Yoshimura, S.H., Takeyasu, K. (2011). Histone and TK0471/TrmBL2 form a novel heterogeneous genome architecture in the hyperthermophilic archaeon *Thermococcus kodakarensis*. *Mol. Biol. Cell* **22**, 386 –398.

33. Reichlen, M.J., Vepachedu, V.R., Murakami, K.S., Ferry, J.G. (2012). MreA Functions in the Global Regulation of Methanogenic Pathways in *Methanosarcina acetivorans*. *mBio* **3**, e00189-12.

34. Krug, M., Lee, S.J., Diederichs, K., Boos, W., Welte, W. (2006). Crystal structure of the sugar binding domain of the archaeal transcriptional regulator TrmB. *J. Biol. Chem.* **281**, 10976–10982.

35. Lee, S.J., Engelmann, A., Horlacher, R., Qu, Q., Vierke, G., Hebbeln, C., Thomm, M., Boos, W. (2003). TrmB, a sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter from the hyperthermophilic archaeon *Thermococcus litoralis*. *J. Biol. Chem.* **278**, 983–990.

36. Rhee, H.S., Pugh, B.F. (2012). ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. *Curr Protoc Mol Biol.* **Chapter 21**.

37. Rossi, M.J., Lai, W.K.M., Pugh, B.F. (2018). Genome-wide determinants of sequence-specific DNA binding of general regulatory factors. *Genome Res.* **28**, 497-508.

38. Peeters, E., Peixeiro, N., Sezonov, G. (2013). Cis-regulatory logic in archaeal transcription. *Biochem. Soc. Trans.* **41**, 326–331.

E.5 Abstract

Transcription in archaea is a unique hybrid between eukaryotic transcriptional regulation and prokaryotic functional components. However, the number of transcriptional regulators in archaea, per genome size, is far smaller relative to their prokaryotic homologues, resulting in the need for intricate strategies to regulate transcription. One such strategy is utilized by the methanogen regulators of energy-conversion (MreABCD) in *Methanosarcina acetivorans* to regulate transcription of genes involved in acetotrophic and methylotrophic pathways in methanogenesis. Using ChIP-exo facilitated mapping of genome-wide binding patterns we show that MreA, in one aspect, differentially binds TATA-boxes across the genome of *M. acetivorans*.

E.6 Introduction

Archaea possess a unique chimeric landscape for transcription. Whereas archaeal transcription factors (TFs) more closely resemble prokaryotic counterparts (1), the mechanism by which transcription is achieved in fact more closely resembles eukaryotic organisms (2,3). Archaeal genomes are organized in a similar fashion to prokaryotic genomes with dense operonic organization and short intergenic spacing (4). A basal archaeal promoter consists of two regulatory elements upstream of a transcription start site (TSS), a TATA-box centred at position -26/-27 with respect to the TSS, recognized by one or more TATA-box binding proteins (TBP's), and a factor B recognition element (BRE) located immediately upstream of the TATA-box (5).

Characterized archaeal transcription factors contain a significant proportion of homologues to bacterial activators and repressors. However, a more striking point to make is the fact that the number of transcription factors is significantly lower per genome size relative to bacterial and eukaryotic counterparts with similar genome sizes (6). Despite this, archaea are able to regulate their complex physiology by alternative strategies that exploit the eukaryote-like basal transcription machinery, i.e., alternative protein-protein interaction profiles with TBP, TFB, and RNAP with or without interactions with accessory proteins can bring about very different responses to gene expression.

Nearly half of characterized archaeal transcription factors are very small in size (100-200 residues) compared to their bacterial homologues and 75% of which only contain a DNA-binding domain (7), suggesting that a stimulus-response domain may be formed through protein-protein interactions with other single-domain TFs, expanding the repertoire of regulation to include homo- and hetero-oligomeric complex formation in response to stimulus (8).

Methanosarcina acetivorans is a recently discovered acetotrophic methanogen whose genome, one of the largest in archaea, has been sequenced. Unlike other methanogens which utilize a single pathway for methanogenesis, the genus *Methanosarcina* can utilize all three known pathways for methanogenesis using a variety

of one-carbon compounds, acetate or carbon dioxide (9). These methanogens are remarkably versatile organisms with a diverse physiology regulated by basal transcription machinery that contains transcription factor TFIIB (TFB), three putative TATA-box binding proteins (TBP's) (10) and a polymerase (Pol) II-like RNA polymerase, and bring about the regulation of their physiology via interplay between the aforementioned components and other general and specific transcription factors.

Proteomic studies of differentially grown *M. acetivorans* in which acetate or methanol was utilized as substrates for methanogenesis identified a significant number of proteins elevated in response to the substrate utilized, indicating differential involvement in regulating the acetotrophic and methylotrophic pathways (11). One such hypothetical protein, MA3302, was upregulated 250-fold with sequence identity to the N-terminal DNA-binding domain of *Thermococcus* regulators of maltose binding (TrmB) (12) but lacking the C-terminal sugar sensory domain making it homologous to TrmBL2 transcription factors (13). Using genetic and transcriptomic analyses, MA3302 was found to be elevated during growth on acetate and simultaneously upregulated genes unique to the acetotrophic pathway and downregulated genes unique to the methylotrophic pathway (11). A Δ*mreA* strain showed a marked decrease in expression of genes involved in the acetotrophic pathway and increase in genes involved in the methylotrophic pathway, indicating both a role in activation and repression, respectively, of transcription of genes unique to the acetotrophic or methylotrophic pathways (14). This duality resulted in MA3302 becoming designated Methanogen regulator of energy-conversion (MreA).

Homologs of MreA from the Leucine-responsive regulatory protein (Lrp), and Asparagine synthase C transcription factor (AsnC) families of transcription factors, similar to the TrmB-family are global and specific regulators of transcription with activation and repression being brought about by altered oligomeric states (15). Many of the aforementioned TFs exist predominantly in two oligomeric states, tetramers and octamers, and facilitate activation or repression by adopting an oligomeric state that facilitates or interferes with TBP1-, TFB-binding, or both (16).

While MreA was previously shown to have sequence-specific binding to promoter DNA (14), sequence identity, and potential variations thereof based on growth substrate

were not known. Here we use ChIP-exo (17, 18) to map the genome-wide binding patterns of MreA and show that MreA differentially binds to TATA-boxes to affect transcriptional regulation of subsets of genes unique to the acetotrophic or methylotrophic pathways of methanogenesis in *M. acetivorans*.

E.7 Results and Discussion

The observation from proteomic and transcriptomic studies indicating MreA simultaneously activates transcription of genes unique to the acetotrophic pathway, represses transcription of genes unique to the methylotrophic pathway, and that a mreA knockout strain resulted in the opposite effects, prompted an examination of the sequence-specificity, if any, to promoter regions of genes up-regulated in the acetotrophic pathway. Previous work examined sequence-specific binding of MreA to promoter regions.

Purified recombinant MreA, heterologously expressed in *Escherichia coli*, was used in electrophoretic mobility shift assays (EMSA) to evaluate sequence-specific binding to DNA fragments spanning the promoter regions of *mreA*, *pta*, and *fpo*, to examine MreA involvement, if present, in potential autoregulation, regulation of genes unique to the acetotrophic pathway, and genes unique to the methylotrophic pathway, respectively (14). Distinct band shifts were observed for the aforementioned promoter regions but not for the nonspecific competitor, poly(dI-dC)/poly(dI-dC). These studies also showed significant smearing in some instances, and hints of multiple distinct bands in others, suggesting a concentration-dependent binding along the promoter region, similar to the pattern observed for the TrmBL2 homolog in *Thermococccus kodakaraensis*, or binding at multiple locations cooperatively, similar to the homolog in *Sulfolobus solfataricus*, LrpB, or a variable, possibly hybrid mechanism that depends on promoter region architecture (sequence, spacing, etc.).

These results presented a challenge when it came to attempts at deducing the specific sequence, or sequences, to which MreA binds. Conventional ChIP-seq methodologies would not have been of sufficient resolution to deduce the binding sequence, particularly if the binding sites were within close proximity, and varied in number, sequence, or both between regulatory regions. ChIP-seq experiments relying on sonication of crosslinked protein-DNA complexes generate libraries with DNA lengths on average ~10-40x that of the actual binding site, introducing various systematic biases in the analyses which need to be reduced using various control experiments. Clearly, such large libraries would convolute attempts at deducing binding sequences of a potential

transcription factor that adopts multiple oligomeric states, or binds multiple locations in close proximity, or along a stretch of the regulatory regions in question in a filamentous, or fibrous manner, similar to the MreA homolog from *T. kodakaraensis* that behaves as a architectural protein involved in genome organization.

E.7.1 MreA Binds to TATA-boxes Across the Genome of *M. acetivorans*

A variation on conventional ChIP-seq, ChIP-exo (17,18) can resolve DNA sequences protected by the transcription factor in question by employing the bacteriophage Lambda (λ) exonuclease to digest the 5'-ends of DNA sequences beyond the resultant complexes from mechanical shearing via sonication. The result of this further processing allows cleavage at or very near the site of transcription factor (TF):DNA cross-linking, barring any TF:DNA boundary artefacts.

Cultures of *M. acetivorans* were grown in and adapted to high-salt media providing either trimethylamine, acetate, or methanol as substrate available for utilization during growth, resulting in activation of either the acetotrophic or methylotrophic pathways which previous studies show are regulated in part by MreA. Cultures were passaged for 50 generations to ensure synchronicity and grown in the final passage to mid-log phase (OD_{600}=~0.6). Approximately 10^7 cells were subsequently cross-linked with 1% (v/v) Formaldehyde, and processed according to Rhee, *et al*. (17,18). Because of the size of the *M. acetivorans* genome, with an estimated ~5000 open reading frames, and the likely fractional occupancy of MreA across the genome, it was sufficient to sequence the resultant libraries on an Illumina NextSeq500 platform with optimal coverage. Data sets were analyzed using MultiGPS and default parameters (20).

Analysis of datasets generated using *M. acetivorans* cultured in media providing acetate as growth substrate results in MreA being mapped predominantly to TATA-boxes across the organism's genome (Fig. E.5) and generating a sequence motif suggesting the cross-link site, and therefore the boundary, of MreA binding to promoter DNA encompasses the canonical sequence normally occupied by TBP1 in *M. acetivorans*. However, from previous proteomic and biochemical studies, it was known that MreA dually regulates both acetotrophic and methylotrophic pathways by activation and

repression, respectively, of genes unique to said pathways. It is also the case that bacterial transcription factors, other than TBP, that occupy TATA-boxes act as repressors as they preclude TBP binding to said TATA-boxes. Thus, it was our initial suspicion that MreA may identify and bind regions upstream or downstream of TATA-boxes under different growth conditions to regulate transcription of subsets of genes within the genome.

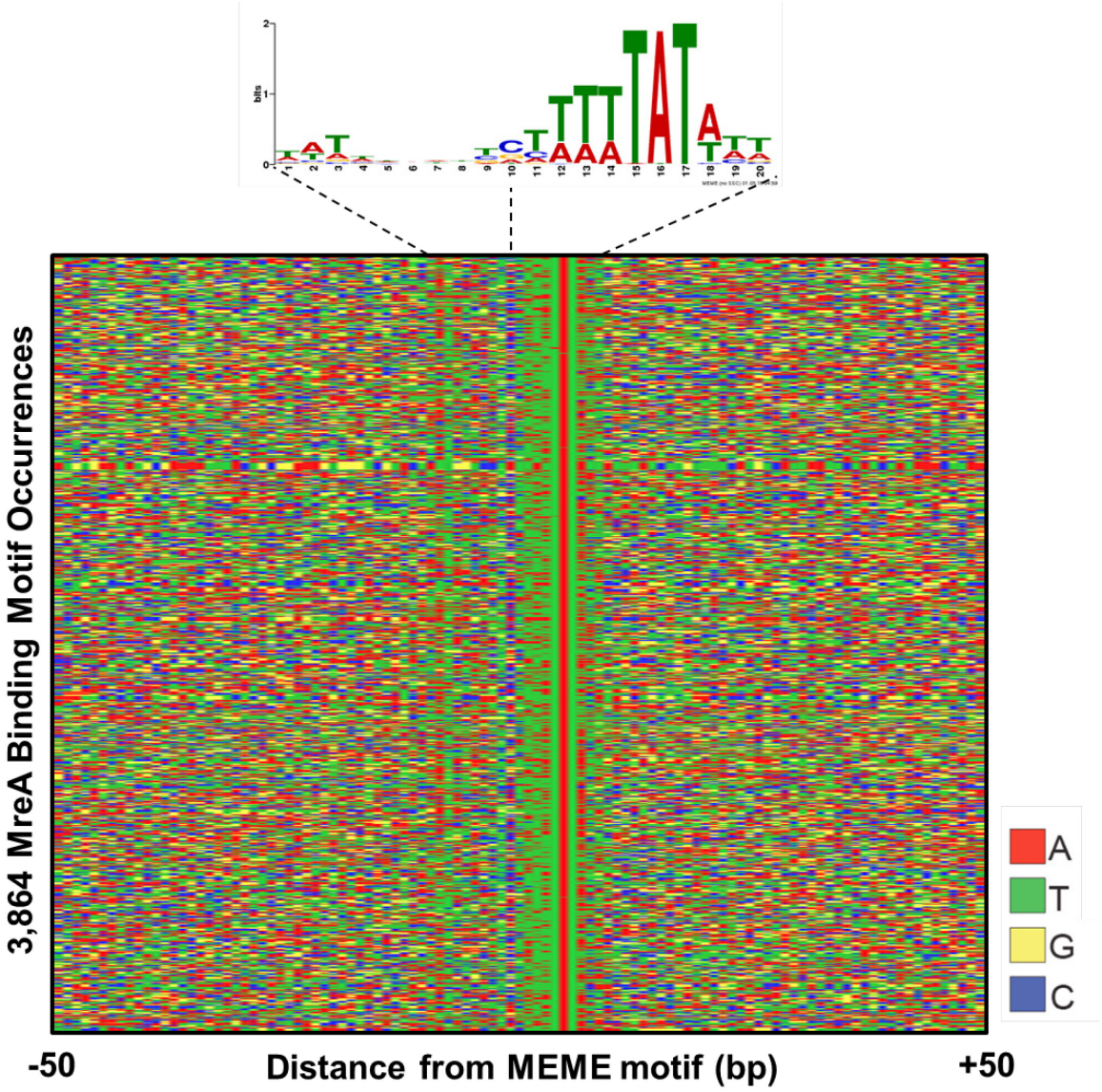


Figure E.5 Genome-wide binding of MreA *in vivo* in acetate-utilizing *M. acetivorans*. The top panel shows the resultant MEME motif and a four-colour plot (bottom) representation of 100-bp sequences centred on the motif midpoint. Each row represents a motif occurrence that passed a FIMO threshold of 1e-4.

E.7.2 Differential Binding Specificity of MreA to Promoter Regions

The finding that MreA binds to TATA-boxes made it necessary to examine sequence specificity under different growth conditions in an effort to understand, and provide rationale for, the global regulatory role of MreA within the *M. acetivorans* genome. At first, we utilized ChIP-exo to examine genome-wide binding patterns of MreA under conditions during which the organism would be cultured on high-salt (HS) media or media containing Trimethylamine (TMA) as a precursor in the methylotrophic pathway of methanogenesis, with the expectation that limiting acetate would perhaps alter the sequence specificity, in one of several mechanisms, of MreA.

However, initial results from the conditions above provided insignificant variation in both sequence specificity and number of binding events of MreA (Fig. E.6) within the genome of *M. acetivorans*.

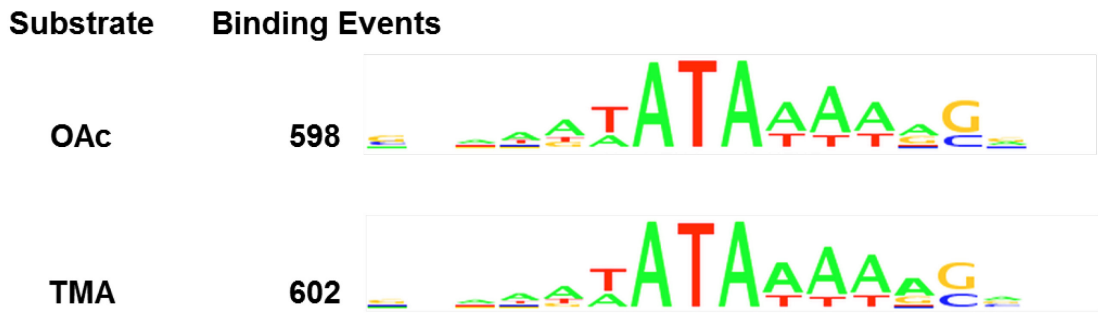


Figure E.6 Differential MreA binding *in vivo* in trimethylamine- or acetate-utilizing *M. acetivorans*. The number of binding events and resultant consensus sequence MEME logos reported from the top 500 peak-pairs from ChIP-exo of MreA in trimethylamine- (TMA) or acetate-utilizing (OAc) *M. acetivorans*. The height of each letter corresponds to its frequency at that position.

The methylotrophic pathway can utilize a number of one-carbon compounds in methanogenesis using a variety of methyltransferases and methyl-coenzyme M that each recognize and act on one of the one-carbon compounds (tri-, di-, and monomethylamine,

methanol, etc.) utilized along the methylotrophic pathway. With this in mind, we examined possible variation in sequence specificity when the organism utilized methanol instead of one of the methylamines during growth, and find that under these conditions the resultant binding sites and consequent motif (Fig. E.7) place MreA at a 12bp offset in the motif relative to the motif obtained for MreA during the organism's growth on acetate.

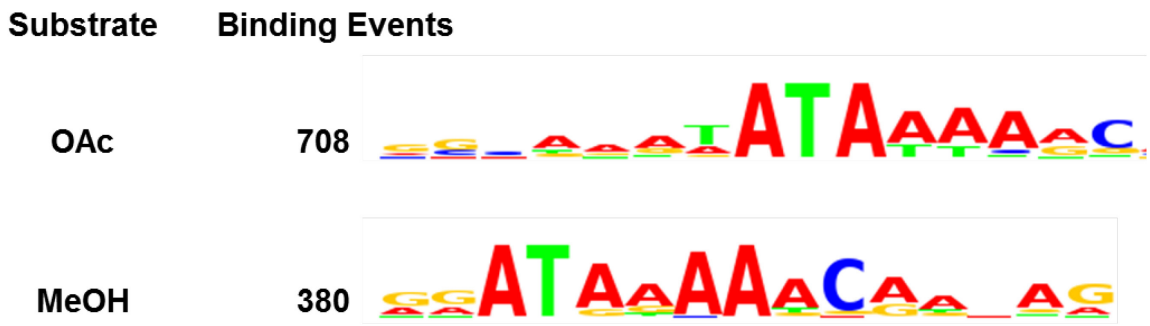


Figure E.7 Differential MreA binding *in vivo* in methanol- or acetate-utilizing *M. acetivorans*. The number of binding events and resultant consensus sequence MEME logos reported from the top 500 peak-pairs from ChIP-exo of MreA in methanol- (MeOH) or acetate-utilizing (OAc) *M. acetivorans*. The height of each letter corresponds to its frequency at that position.

The relative offset in motifs obtained for MreA binding specificity during growth on acetate and methanol, still within proximity of each other, suggests the existence of a mechanism by which MreA can alter its binding about a TATA-box to either recruit or occlude TBP1 to bring about activation or repression of different subsets of genes unique to either pathway, but its TrmB-like homologs exhibit a variety of different properties that facilitate their function by altering their oligomeric states, protein-protein interactions, chemical environment, or a combination of these attributes.

Investigating the possible mechanism by which MreA alters binding specificity would require mapping both MreA binding to promoter regions alone, in any of its' oligomeric forms, mapping possible protein-protein interactions between MreA and TBP1, TFB, or other as yet unidentified transcription factors. Examining sequence specificity *in vitro* would provide a baseline and proper control in attempts to investigate

the possible role oligomeric states may play in altering sequence specificity. To that end, genome-wide MreA binding *in vitro* in the absence of any other protein, cofactor, or metabolite, resulted in a sequence motif that indicates MreA binding at different locations around the TATA-box including a near-perfect occupancy of the 3'-end of BRE box (Fig. E.8).

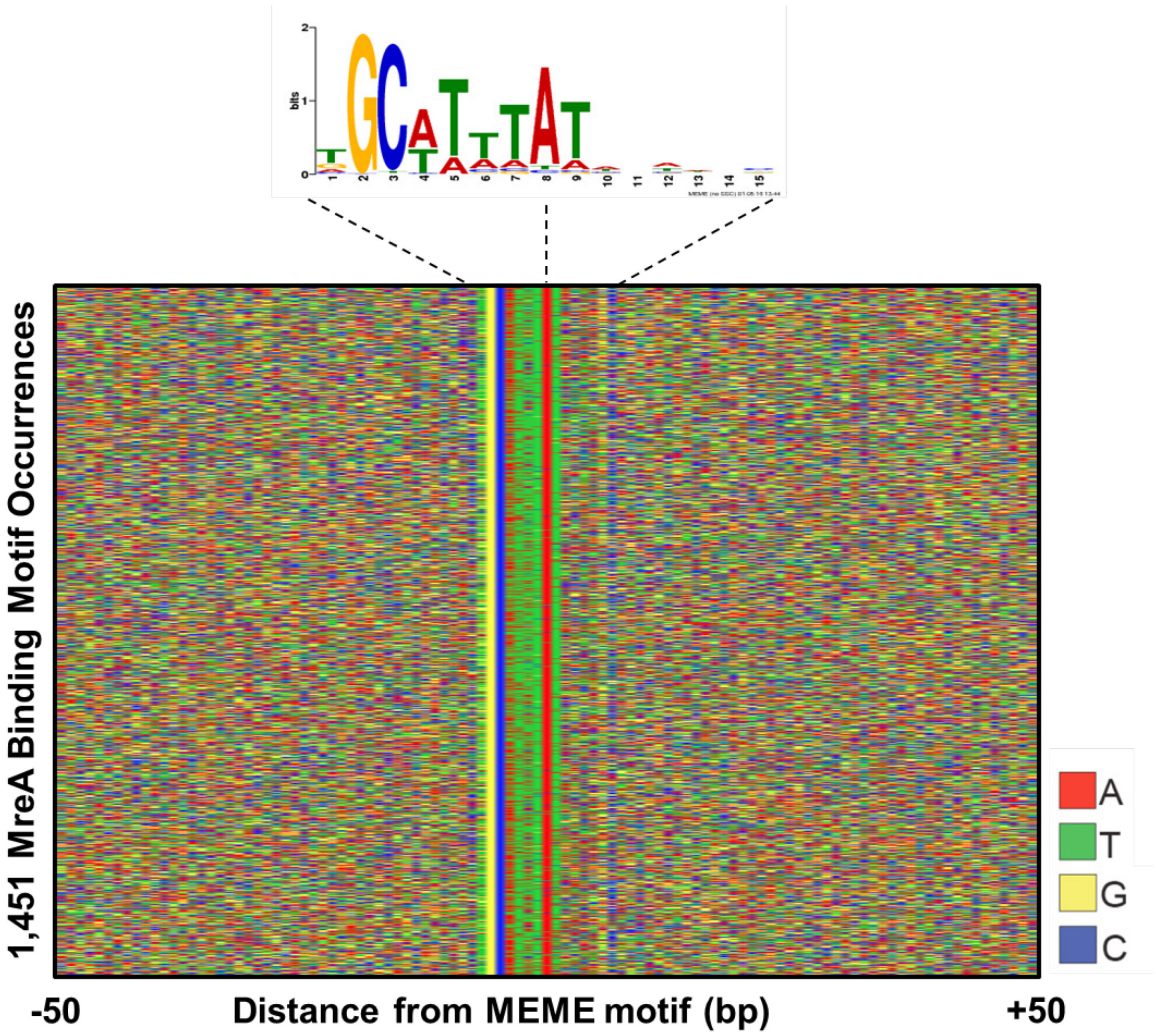


Figure E.8 Genome-wide binding of MreA *in vitro*. The top panel shows the resultant MEME motif and a four-colour plot (bottom) representation of 100-bp sequences centred on the motif midpoint. Each row represents a motif occurrence that passed a FIMO threshold of 1e-4.

E.7.3 Higher Order Architecture Mediates TATA-box Recognition

Most TrmBL2 homologs of MreA affect activation or repression predominantly via altered oligomeric states (dimers to tetramers or octamers) to structurally hinder or entirely occlude TBP from binding to TATA-boxes and recruiting RNA polymerase to form the pre-initiation complex (PIC). Other TrmBL2 homologs bring about their intended function by a concentration-dependent manner through binding to adjacent high-affinity sites separated by a low-affinity site. Once the target transcription factor saturates the high-affinity sites, the low-affinity sites become occupied until, in a filamentous manner, the binding of said factor will either deform the DNA or occupy multiple regulatory sites along the promoter region to prevent or facilitate binding of other factors.

The number of crosslink sites, that is the TF:DNA boundary formed by the protective footprint of the bound factor, can be deduced from ChIP-exo by examining the locations and spacing of said sites. We examined datasets from *in vitro* experiments, and *in vivo* experiments utilizing methanol or acetate as growth substrate (Fig. E.9), by generating composite plots, with a zero midpoint representative of the halfway point of deduced TATA-boxes, from *in vitro* and *in vivo* data and show that *in vitro* two strong peaks, one on either strand, are apparent with minimal baseline noise, corresponding to two crosslink sites representative of a single footprint. This is consistent with the possibility that higher order oligomeric states or protein-protein interactions, which would correspond to more than two crosslink sites, are required to facilitate altered binding specificity. The *in vivo* data shows a greater number of both evenly and unevenly spaced crosslink sites indicative of a binding pattern that either relies on the aforementioned higher order oligomeric states, or protein-protein interactions occurring to facilitate the binding of one factor or the other, resulting in a greater number of crosslink sites.

In *in vitro* biochemical experiments, including analytical equilibrium gel filtration (data not shown), purified MreA and its homologs appear in different oligomeric states, and predominantly as tetramers or octamers. The presence of residual DNA appears to have stabilized octameric complexes.

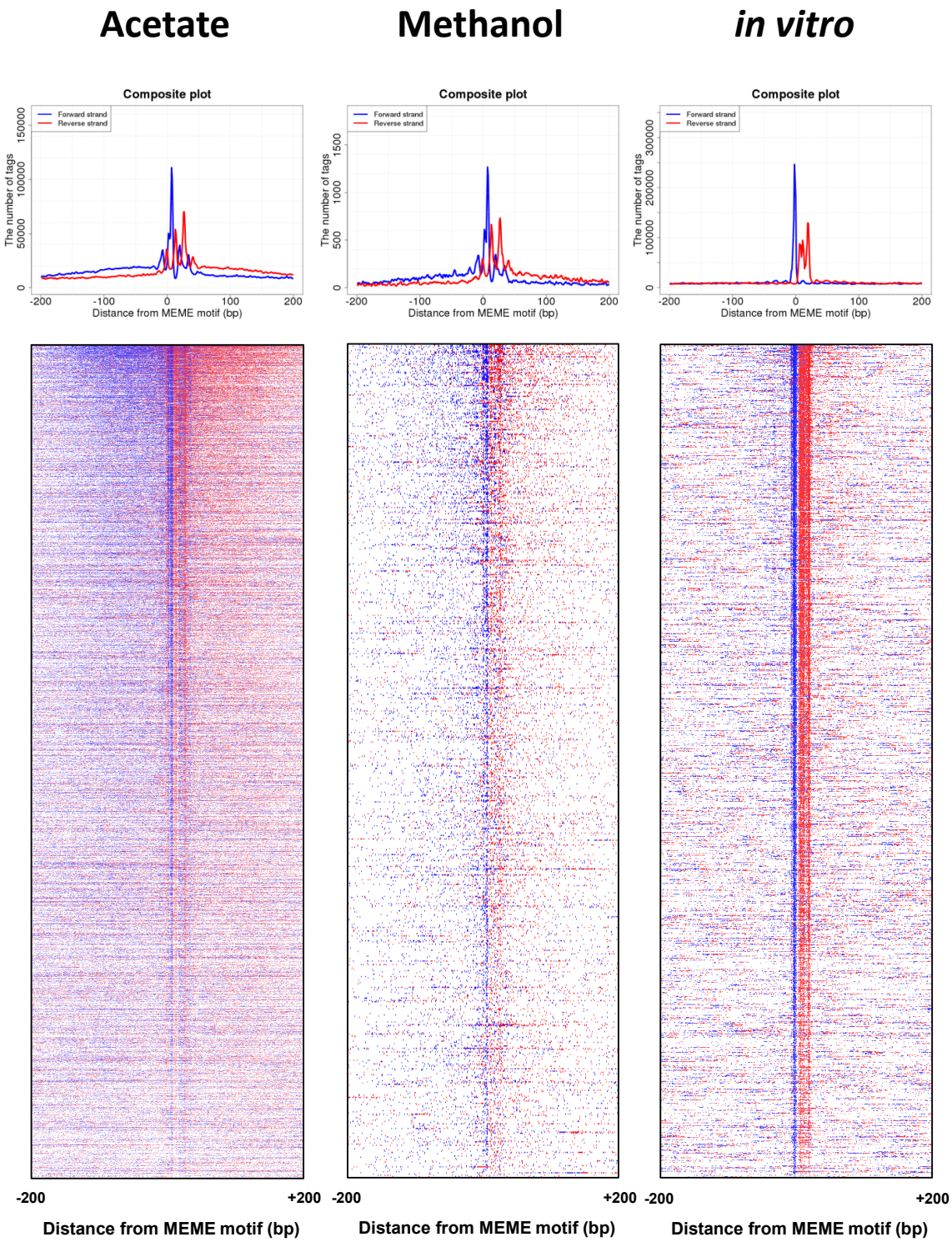


Figure E.9 Genome-wide *in vivo* and *in vitro* binding of MreA. Composite plots and heat maps of MreA binding analyses examining the number and spacing of potential crosslink sites using ChIP-exo *in vivo*, utilizing methanol (MeOH) or acetate (OAc), and *in vitro*. Distances are from TATA-box motif reference point. Blue indicates tag 5′ ends located on the same strand as the motif, whereas red are located on the opposite strand.

E.8 Materials and Methods

E.8.1 Materials

Oligonucleotides used in cloning, EMSA or crystallization were PAGE and/or HPLC purified from Integrated DNA Technologies (Carlsbad, CA USA). Deoxyribonucleoside triphosphates (dNTP's) were obtained from New England Biolabs (Ipswich, MA USA). Protein purification matrices were purchased from GE Healthcare Bio-Sciences Corporation (Piscataway, NJ USA). EMSA gels and consumables were purchased from Thermo Fisher. Rabbit polyclonal antibodies raised against MreA or TBP1 were obtained from Covance Inc. (Denver, PA USA). Crystallization screens and reagents were obtained from Hampton Research (Aliso Viejo, CA). All other reagents were of the highest grade available.

E.8.2 Cell Growth

Growth and harvesting of *M. acetivorans* C2A (DSM 800) in single cell morphology (21) cultured on high-salt (HS) broth medium supplemented with acetate or methanol was previously described (22). Growth was monitored by measuring the optical density (turbidity) at 600nm.

E.8.3 Adaptation to Different Methanogenic Substrates

Starter cultures (ca. 10mL) of methanol-adapted early stationary phase (OD_{600}= ~1.0) *M. acetivorans* were pelleted by centrifugation anaerobically at 4,000 x g, at room temperature. Cell pellets were subsequently washed 3 times with 10mL media containing the target substrate (acetate, TMA, etc.) and used to inoculate 100mL media containing the target substrate.

E.8.4 Genomic DNA Isolation

Cells were pelleted by centrifugation for 10 minutes at 4,000 x g and lysed by incubation for 30 minutes at 37°C in TE buffer (10mM Tris-Cl, pH = 7.5 at RT, 1mM EDTA, pH = 8.0) supplemented with 400mM NaCl, 50mM $MgCl_2$, 0.3% (w/v) SDS and 5ug/mL RNase A. Genomic DNA was purified from cell lysate using the PureGene genomic DNA isolation kit following the manufacturers recommended protocols.

E.8.5 Protein Expression and Purification

M. acetivorans genes MA3302, MA1671, MA3129 and MA3130 encoding MreA, -B, -C, or –D, respectively, and genes MA4331, MA0179, MA0278, and MA0610 encoding TBP1, -2, -3, or TFB, respectively, were PCR-amplified from genomic DNA using oligonucleotides listed in Table E.1, with 5'-extensions homologous to the upstream (Fwd) and downstream (Rev) sequences of the multiple cloning site (MCS) of phage T7-derived expression vector pSUMO for Vaccinia virus DNA polymerase-catalyzed homologous recombination cloning (In-Fusion®). Target genes are cloned as C-terminal fusions to His_6-tagged Smt3 to facilitate increased solubility, purification utilizing immobilized metal-affinity chromatography (IMAC) and subsequent cleavage of fusion proteins with Ulp1 protease to yield native target proteins.

E.8.6 ChIP-exo

ChIP-exo was performed as previously described (17,18) with 10 million cells grown in conditions described above and then treated with 1% formaldehyde prior to lysis and sonication. Polyclonal anti-MreA or anti-TBP1 antibody was conjugated to Protein G MagSepharose beads (GE Healthcare) and subsequently used for ChIP. ChIP-exo libraries were amplified by 18 PCR cycles and then size-selected (200-500 bp) by gel excision.

E.8.7 DNA Sequencing

ChIP-exo libraries were sequenced on the Illumina NextSeq 500, producing (40 x 36 bp) paired-end reads. Reads passing Q30 were aligned to the NC003552 reference genome of *M. acetivorans* using BWA MEM with default parameters (19). Aligned reads were then filtered to a minimum mapping quality of Q5 and deduplication was performed using the Picard MarkDuplicates function with default parameters (picard).

E.8.8 Data Analysis

Datasets generated from DNA sequencing were analyzed using MultiGPS v0.74 using default software parameters according to Mahony, *et al.* (20).

Gene	Locus	Name	Sequence (5'-3')
mreA	MA3302	pSUMO_MacMreA_F	CAGATTGGAGGTATGGTAGATGAAAGTGGTTTAG
		pSUMO_MacMreA_R	GATCCGGTCTCATTACTTTGTGGGTGTTAAAGGCAC
mreB	MA1671	pSUMO_MacMreB_F	CAGATTGGAGGTATGACCGACGATTCTCCAGTTA
		pSUMO_MacMreB_R	GATCCGGTCTCATTAATTTTGCATTATAATTTCCAGTTCC
mreC	MA3129	pSUMO_MacMreC_F	CAGATTGGAGGTATGATAGTTGTGGCAGATAATG
		pSUMO_MacMreC_R	GATCCGGTCTCATCAGGAAAGGCTCTTGAGACG
mreD	MA3130	pSUMO_MacMreD_F	CAGATTGGAGGTATGGTAGATGAAAGTGGTTTAG
		pSUMO_MacMreD_R	GATCCGGTCTCATCAGGCAATGTGCTTCAGGC
tbp1	MA4331	pSUMO_MacTBP-1_F	CAGATTGGAGGTATGAGCGAATCTAGCATTAA
		pSUMO_MacTBP-1_R	GATCCGGTCTCATTATAAAAGTCCCATGTTATCAAGCTGC
tbp2	MA0179	pSUMO_MacTBP-2_F	CAGATTGGAGGTATGGAACCCACAATAACTATAGAGA
		pSUMO_MacTBP-2_R	GATCCGGTCTCATTAATAAAGAAGTCCTAAGTTGTCCA
tbp3	MA0278	pSUMO_MacTBP-3_F	CAGATTGGAGGTATGGAATCCACAATAACTATAGAGAACGT
		pSUMO_MacTBP-3_R	GATCCGGTCTCATTAACAGAGAAGTCCTAAGTTGTCGA
tfb	MA0610	pSUMO_MacTFB_F	CAGATTGGAGGTATGGTAGAAGTCGAAAGAG
		pSUMO_MacTFB_R	GATCCGGTCTCATTAGAGGATAATCTCTATATCCAGT

Table E.1 Oligonucleotides used in cloning target genes from *M. acetivorans* genomic DNA. The sequences underlined are 12-bp extensions that are homologous to the upstream and downstream regions of the multiple cloning site (MCS) of expression vector pSUMO MCS.

E.9 Authors Contribution

Starting material and processing, and data generation between the years of 2016 and 2018, and consequent analyses have been performed by the author, Mohammad A. AlMishwat. Raw sequencing data were processed by Dr. Will Lai of the Pugh laboratory at the Pennsylvania State University. Starting material and processing between the years of 2014 and 2016 have been performed by the author, and consequent analyses were performed by co-first author of the manuscript currently in preparation, Dr. Bongsoo Park, then of the laboratory and under the direction of B. Franklin Pugh at the Pennsylvania State University. All samples since the authors undertaking of the work were sequenced on an Illumina NextSeq500 platform. All other samples prior to this were sequenced on an Illumina HiSeq2000 platform

The first replicate of the *in vitro* sample for ChIP-exo in which purified genomic DNA and recombinant MreA were mixed together, crosslinked, and submitted for downstream processing in early 2014, was generated by Dr. Ninad Walavalkar and processed by Ms. Nina Farrell, then at the Pennsylvania State University in the laboratories and under the directions of J. Gregory Ferry and B. Franklin Pugh, respectively.

The first *in vivo* sample for ChIP-exo of cells cultured on acetate was prepared by Mr. Eric Malmberg in 2013, then of the laboratory and under the direction of J. Gregory Ferry at the Pennsylvania State University. In all instances, DNA library preparation and subsequent sequencing were performed by the Pugh laboratory at the Pennsylvania State University.

The author did not contribute to computational work pertaining to DNA shape analysis reported in one version of the resultant manuscript. A biological framework for the structural and physiological aspects of regulation of transcription in Archaea, including TF structural biology and archaeal regulatory regions, first referenced in proposals and experimental design generated by the author and disseminated to co-authors, was provided by the author to Dr. Bongsoo Park to familiarize and assist in the accurate interpretation of results from computational analyses.

References

1. Bell, S.D. (2005). Archaeal transcriptional regulation – variation on a bacterial theme? *Trends Microbiol.* **13**, 262–265.

2. Geiduschek, E.P. and Ouhammouch, M. (2005). Archaeal transcription and its regulators. *Mol. Microbiol*. **56**, 1397–1407.

3. Grohmann, D. and Werner, F. (2011). Recent advances in the understanding of archaeal transcription. *Curr. Opin. Microbiol*. **14**, 318–324.

4. Price, M.N., Arkin, A.P., Alm, E.J. (2006). The life-cycle of operons. *PLoS Genet*. 2, e96.

5. Soppa, J. (1999). Normalized nucleotide frequencies allow the definition of archaeal promoter elements for different archaeal groups and reveal base-specific TFB contacts upstream of the TATA box. *Mol. Microbiol*. **31**, 1589–1592.

6. Perez-Rueda, E., Collado-Vides, J. (2001). Common history at the origin of the position function correlation in transcriptional regulators in archaea and bacteria. *J. Mol. Evol.* **53**, 172–179.

7. Aravind, L., Koonin, E.V. (1999). DNA-binding proteins and evolution of transcription regulation in the archaea. *Nucleic Acids Res.* **27**, 4658–4670.

8. Koike, H., Ishijima, S.A., Clowney, L., Suzuki, M. (2004). The archaeal feast/famine regulatory protein: potential roles of its assembly forms for regulating transcription. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2840–2845.

9. Hovey, R., Lentes, S., Ehrenreich, A., Salmon, K., Saba, K., Gottschalk, G., Gunsalus, R.P., Deppenmeier, U. (2005). DNA microarray analysis of *Methanosarcina mazei Go1* reveals adaptation to different methanogenic substrates. *Mol. Genet. Genomics* **273**, 225–239.

10. Reichlen, M.J., Murakami, K.S., and Ferry, J.G. (2010). Functional analysis of the

three TATA binding protein homologs in *Methanosarcina acetivorans*. *J. Bacteriol.* **192**, 1511-1517.

11. Li, L., Li, Q., Rohlin, L., Kim, U., Salmon, K., Rejtar, T., Gunsalus, R.P., Karger, B.L., Ferry, J.G. (2007). Quantitative proteomic and microarray analysis of the archaeon *Methanosarcina acetivorans* grown with acetate versus methanol. *J. Proteome Res.* **6**, 759 –771.

12. Lee, S.J., Moulakakis, C., Koning, S.M., Hausner, W., Thomm, M., Boos, W. (2005). TrmB, a sugar sensing regulator of ABC transporter genes in Pyrococcus furiosus exhibits dual promoter specificity and is controlled by different inducers. *Mol. Microbiol.* **57**, 1797–1807.

13. Maruyama, H., Shin, M., Oda, T., Matsumi, R., Ohniwa, R.L., Itoh, T., Shirahige, K., Imanaka, T., Atomi H., Yoshimura, S.H., Takeyasu, K. (2011). Histone and TK0471/TrmBL2 form a novel heterogeneous genome architecture in the hyperthermophilic archaeon *Thermococcus kodakarensis*. *Mol. Biol. Cell* **22**, 386 –398.

14. Reichlen, M.J., Vepachedu, V.R., Murakami, K.S., Ferry, J.G. (2012). MreA Functions in the Global Regulation of Methanogenic Pathways in *Methanosarcina acetivorans*. *mBio* **3**, e00189-12.

15. Krug, M., Lee, S.J., Diederichs, K., Boos, W., Welte, W. (2006). Crystal structure of the sugar binding domain of the archaeal transcriptional regulator TrmB. *J. Biol. Chem.* **281**, 10976–10982.

16. Lee, S.J., Engelmann, A., Horlacher, R., Qu, Q., Vierke, G., Hebbeln, C., Thomm, M., Boos, W. (2003). TrmB, a sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter from the hyperthermophilic archaeon *Thermococcus litoralis*. *J. Biol. Chem.* **278**, 983–990.

17. Rhee, H.S., Pugh, B.F. (2012). ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. *Curr Protoc Mol Biol.* **Chapter 21**.

18. Rossi, M.J., Lai, W.K.M., Pugh, B.F. (2018). Genome-wide determinants of sequence-specific DNA binding of general regulatory factors. *Genome Res.* **28**, 497-508.

19. Li, J., Liu, Y., Rhee, H.S., Ghosh, S.K., Bai, L., Pugh, B.F., Gilmour, D.S. (2013). Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing. *Mol. Cell* **50**, 711-22.

20. Mahony, S., Edwards, M.D., Mazzoni, E.O., Sherwood, R.I., Kakumanu, A., Morrison, C.A., Wichterle, H., Gifford, D.K. (2014). An integrated model of multiple-condition ChIP-seq data reveals predeterminants of Cdx2 binding. *PLoS Comput. Biol.* **10**(3), e1003501.

21. Sowers, K. R., Boone, J.E., Gunsalus, R.P. (1993). Disaggregation of *Methanosarcina spp* and Growth as Single Cells at Elevated Osmolarity. *Appl. Environ. Microbiol.* **59**, 3832-3839.

22. Lessner, D.J., Li, Q., Li, T., Rejtar, V.P., Reichlen, M., Hill, K., Moran, J.J., Karger, B.L., Ferry, J.G. (2006). An unconventional pathway for reduction of CO_2 to methane in CO-grown *Methanosarcina acetivorans* revealed by proteomics. *Proc Natl Acad Sci U S A.* **103**, 17921-6.

E.10 Studies on Novel Transcriptional Regulation in Archaea – Summary

This part of the work represents a foray, brief as it was, into the field of genomics in an effort to characterize genome-wide binding patterns of the archaeal transcriptional regulator MreA from the methanogen *Methanosarcina acetivorans* and provide insight into its mode of action. *Methanosarcina acetivorans* is a recently discovered marine methanogen responsible for two-thirds of the world's methane production through acetate fermentation (1). Very little is known about transcriptional regulation by this regulator in *M. acetivorans* –in fact, only one other publication originating from our group examines MreA in any capacity– and one key observation was sequence-specific binding of MreA to promoter regions within the genome of *M. acetivorans*.

Global transcriptional regulation in *Methanosarcina acetivorans* is affected by a number of transcriptional regulators. Within this category of transcriptional regulators exists the four homologs of Methanogen regulators of energy conversion (MreABCD) (2). Previous studies in our laboratory have shown that MreA is a global regulator of methanogenic pathways in *M. acetivorans* with differential expression levels during growth conditions utilizing substrates activating the acetotrophic or methylotrophic pathways through sequence-specific binding to promoter regions (3). Transcriptional profiling of wild-type *M. acetivorans* and mutants lacking functional MreA has identified 280 genes, including genes unique to the acetotrophic or methylotrophic pathways, with altered expression levels (4).

To investigate the role of MreA and its homologs (MreBCD) in the global regulatory network of *M. acetivorans* structural and functional studies were performed on a genome-wide scale. ChIP-exo analyses (5) probing the binding sites of MreA, MreD and mutants thereof have revealed that MreA acts as a dual regulator activating transcription of a specific subset of genes involved in the acetotrophic pathway and repressing transcription of another subset of genes involved in the methylotrophic pathway for methane production by altered binding to promoter regions through an as yet undetermined mechanism (3). Our biochemical studies thus far have shown that MreACD, but not MreB, exist as dimers, tetramers and more predominantly octamers, which may be the mechanism by which altered binding interactions with other

transcriptional regulators and specific operons is brought about. ChIP-exo studies have also shown that mutants lacking *mreA* displayed nearly 300-fold up-regulation of MreD suggesting that either MreA regulates expression of MreD or/and MreD is part of a redundancy mechanism to regulate transcription in the absence of functional MreA (3).

ChIP-exo analyses have also shown that the promoter-specific binding regions to which MreA binds are in fact TATA-boxes. However, this was convoluted somewhat as growth on either substrate, acetotrophic or methylotrophic, resulted in largely the same binding sequence motif, so it was not readily clear how transcription was differentially regulated in *M. acetivorans* in response to changes in the metabolic pathway being utilized for methane production, but it was likely that MreA binding to TATA-boxes affected TBP1 binding to those same TATA-boxes. Thus, differential regulation of transcription must occur at least in part via binding of MreA to TATA-boxes alone or simultaneously with TBP1.

Conventional ChIP-seq methodology would not have been able to provide the base-pair resolution that was necessary to obtain our findings that MreA functions, at least in part, by binding to TATA-boxes and interacting, either directly or indirectly, with TBP1. Given the sequence context of the TATA-boxes, with slight variations in sequence and spacing in archaea relative to other domains of life, MreA appears to bind to said sequences and either hinder or facilitate TBP1 binding by occluding its binding to TATA-boxes or exposing TATA-box DNA to facilitate more interaction with TBP1. This finding is unique to MreA and has not been reported in its TrmB homologs in other archaea, and certainly not in other domains of life.

As was previously alluded to, little information is available exploring transcriptional regulation in *Methanosarcina acetivorans*. In fact, as it directly pertains to MreABCD, only a single publication is produced from the work of a previous member of our group, making studies of any kind to characterize the structure, function, biochemistry, or involvement in *M. acetivorans* physiology, of Mre's novel and indispensable. The findings above serve only as a starting point for a great deal of further investigation into to a very long list of questions regarding the function of MreA and its homologs within the context of the physiology of *M. acetivorans*.

One such question that we began to investigate stems from the observation that MreB monomers but not MreA, C or D monomers, appear to form only tetramers. Within the context of TrmB-like transcriptional regulators, which lack the C-terminal sugar sensory domains of TrmB TFs (6), these oligomeric states can either be an artefact, e.g., due to degradation, or the conditions in which the crude characterization was performed, or they may be consistent with their *Pfu* (7) and *Sso* (8) counterparts that bring about activation or repression via alternative higher-order oligomeric states, presumably producing steric hindrance to occlude recruitment of general TFs. If the latter holds true then, again, given the lack of sensory domains, the mechanism by which induction of conversion of dimers to tetramers, and *vice versa*, of MreACD, must either be concentration dependent, as is the case in their *Sso* counterpart, the leucine-responsive regulatory protein LrpB (9), binding sequence-dependent, or brought about via interaction with other regulatory proteins.

Similarly, our group has shown a growth substrate-dependent up- and down-regulation of genes encoding either the acetotrophic or methylotrophic pathways in both wild-type and knockout strains of each of the four homologs which cannot be readily explained only in the context of pathway-dependent participation of one or more of the homologs, but can be addressed through subsequent ChIP-exo experiments and inclusion of antisera targeting other homologs and TBP's. Such an undertaking would facilitate parsing or binning of binding sequences representing entire subsets of the approximately 4500 genes encoded by the *M. acetivorans* genome based on target homolog and potential interactions with any basal transcription factors when the organism is utilizing substrates activating its acetotrophic or methylotrophic pathways.

Lastly, as was mentioned above, data generated from ChIP-exo experiments presented here suggests that MreA binding to TATA-boxes, with a strong resultant sequence motif, hinders TBP1 binding and is likely the mechanism by which MreA represses transcription of the relevant downstream genes. This result is in line with preliminary cross-linking experiments that show pull-downs of MreA complexes to contain TBP1, indicating that protein-protein interactions exist between MreA and TBP1, either directly or via proximity due to simultaneous occupancy of TATA-boxes. In either

case, occlusion of a subset of TATA-boxes by MreA binding must interfere with TBP1's ability to interact with and recruit TFB to the BRE (10), to bend the TATA-box DNA (11), and must also have a structural basis that would be very interesting to investigate given the novelty of MreA's actions.

References

1. Hovey, R., Lentes, S., Ehrenreich, A., Salmon, K., Saba, K., Gottschalk, G., Gunsalus, R.P., Deppenmeier, U. (2005). DNA microarray analysis of *Methanosarcina mazei Go1* reveals adaptation to different methanogenic substrates. *Mol. Genet. Genomics* **273**, 225–239.

2. Reichlen, M.J., Murakami, K.S., and Ferry, J.G. (2010). Functional analysis of the three TATA binding protein homologs in *Methanosarcina acetivorans*. *J. Bacteriol.* **192**, 1511-1517.

3. Reichlen, M.J., Vepachedu, V.R., Murakami, K.S., Ferry, J.G. (2012). MreA Functions in the Global Regulation of Methanogenic Pathways in *Methanosarcina acetivorans*. *mBio* **3**, e00189-12.

4. Li, L., Li, Q., Rohlin, L., Kim, U., Salmon, K., Rejtar, T., Gunsalus, R.P., Karger, B.L., Ferry, J.G. (2007). Quantitative proteomic and microarray analysis of the archaeon *Methanosarcina acetivorans* grown with acetate versus methanol. *J. Proteome Res.* **6**, 759 –771.

5. Rhee, H.S., Pugh, B.F. (2012). ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. *Curr Protoc Mol Biol.* **Chapter 21**.

6. Lee, S.J., Engelmann, A., Horlacher, R., Qu, Q., Vierke, G., Hebbeln, C., Thomm, M., Boos, W. (2003). TrmB, a sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter from the hyperthermophilic archaeon Thermococcus litoralis. *J. Biol. Chem.* **278**, 983–990.

7. Lee, S.J., Moulakakis, C., Koning, S.M., Hausner, W., Thomm, M., Boos, W. (2005). TrmB, a sugar sensing regulator of ABC transporter genes in Pyrococcus furiosus exhibits dual promoter specificity and is controlled by different inducers. *Mol. Microbiol.* **57**, 1797–1807.

8. Peeters, E., Thia-Toong, T.-L., Gigot, D., Maes, D. and Charlier, D. (2004). Ss-LrpB, a

novel Lrp-like regulator of *Sulfolobus solfataricus* P2, binds cooperatively to three conserved targets in its own control region. *Mol. Microbiol.* **54**, 321–336.

9. Peeters, E., Willaert, R., Maes, D. and Charlier, D. (2006). Ss-LrpB from *Sulfolobus solfataricus* condenses about 100 base pairs of its own operator DNA into globular nucleoprotein complexes. *J. Biol. Chem.* **281**, 11721–11728.

10. Bell, S.D., Kosa, P.L., Sigler, P.B. and Jackson, S.P. (1999). Orientation of the transcription preinitiation complex in archaea. *Proc. Natl. Acad. Sci. U.S.A.* 96, 13662–13667.

11. Peeters, E., Wartel, C., Maes, D. and Charlier, D. (2007). Analysis of the DNA-binding sequence specificity of the archaeal transcriptional regulator Ss-LrpB from *Sulfolobus solfataricus* by systematic mutagenesis and high resolution contact probing. *Nucleic Acids Res.* **35**, 623–633.

VITA

Mohammad A. AlMishwat

Education

The Pennsylvania State University 12/2019
Ph.D., Biochemistry, Microbiology, and Molecular Biology

The Pennsylvania State University 08/2013
M.Sc., Biochemistry, Microbiology, and Molecular Biology

The University of Kansas 08/2009
B.Sc., Genetics, Molecular Biosciences
B.Sc., Biochemistry, Molecular Biosciences

Honors and Awards

Kuwait University Graduate Scholar in Biochemistry, 2011 – Present
Person Graduate Fellowship in Biochemistry and Molecular Biology, 2013 – 2017
National Science Foundation Nanotechnology Scholarship, 2007
Research Experience Program, Univ. of Kansas Dept. of Molecular Biosciences 08/2009
International Scholar, Ministry of Higher Education of Kuwait, 2004 - 2009
Academic Merit, Embassy of the State of Kuwait, 2004 – 2009

Publications

AlMishwat, MA. (2013). *Time-resolved X-ray crystallography of DNA polymerases: A platform for studies in structural and mechanistic enzymology* (Master's thesis). Retrieved from the Pennsylvania State University's electronic theses and dissertations (eTD) database.

Mohammad Almishwat, Katsuhiko Murakami. (2014). Structural Enzymology of Nucleotide Incorporation by DNA Polymerases. Oak Ridge National Laboratory (Oak Ridge, TN) (**Poster**)

AlMishwat, MA., Ferry, JG., Murakami, KS. (2019). Structural and Mechanistic Basis of Fidelity in Nucleotide Incorporation during DNA Replication by an A-family Replicative DNA Polymerase. (Manuscript in preparation)

AlMishwat, MA., Ferry, JG., Murakami, KS. (2019). Genome-wide binding of the *Methanosarcina acetivorans* global regulator MreA revealed through ChIP-exo and biochemical characterization support a structural basis for regulatory function. (Manuscript in preparation)

AlMishwat, MA. (2019). *Structural and mechanistic enzymology of nucleotide incorporation by A-family DNA polymerases* (Doctoral dissertation).