UC Irvine
UC Irvine Electronic Theses and Dissertations

Title

DNA Synthesis of the PhiX Genomic DNA library from Deoxynucleotide Diphosphate Substrates

Permalink

https://escholarship.org/uc/item/5v4486v5

Author

Argelagos, Allan

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
IRVINE

DNA Synthesis of the PhiX Genomic DNA library from

Deoxynucleotide Diphosphate Substrates

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Pharmacological Sciences

by

Allan Argelagos

Thesis Committee:
Professor Andrej Lupták, Chair
Professor David Mobley
Professor Amal Alachkar

2022

Table of Contents

List of Figures

Acknowledgements

Abstract

Abstract of the Thesis

by

Allan Argelagos

Master of Science in Pharmacological Sciences

University of California, Irvine 2022

Professor Andrej Lupták, Chair

DNA synthesis is an essential process necessary for the survival of living organisms. This biochemical event is important for self-maintenance and is even required to pass down genetic information to future offspring. Though this process has been well characterized and studied, very little evidence is provided for *how* this process emerged, especially in a pre-biotic world where enzymatic processes had yet to exist. In this work we investigated how the DNA polymerase selectively chose the deoxynucleotide triphosphate (dNTP) as its naturally occurring substrate to make copies of its DNA. We studied DNA synthesis with a more prebiotically relevant substrate, that being the deoxynucleotide diphosphate (dNDP). Previous work has already shown the various thermophilic and mesophilic DNA polymerase can utilize dNDPs to perform DNA synthesis with a short DNA template (~100 bps), but it has yet to be discovered if this substrate will produce full length product when provided a longer DNA template (500 bps) with a higher sequence diversity DNA library. A single triphosphate was replaced by a diphosphate substate in DNA polymerase synthesis to see whether it were possible

to amplify the 500 bps PhiX genomic DNA library. Here showed that when replacing guanosine triphosphate with guanosine diphosphate (dGDP), full length product is produced using *Taq* DNA polymerase. *Vent(-exo)* DNA polymerase can also produce full length product with dGDP, but when performing this same replacement experiment with any other diphosphate, no full-length product can be visualized for both polymerases. What this means is that these DNA polymerases have difficulty fully reading down the template strand when length of template and diversity of nucleotide context is increased with dNDP synthesis. This counteract the idea a dNDP to be prebiotically reliant, but one must consider the template context as the increase diversity in nucleotides may not have existed in pre-biotic times. Further experiments on sequence context is necessary to better understand how it is the that diphosphates can be utilized by DNA polymerases.

Introduction & Background

Role of Phosphorus in the Emergence of Life

Phosphorus is a key element vital to the proliferation and maintenance of cellular life. It is found within the cell membrane providing a *barrier* for living organisms through the phospholipid bilayer [Goñi]. Also, phosphate is constantly exchanged in cell signaling pathways that allow cells to *communicate* internally with one another [Michigami]. Lastly, phosphorus is even found to provide structure to arguably the most critical biological polymer necessary for the maintenance of life, the deoxyribonucleic acid (DNA)[Šponer]. With all the functions that phosphorus provides to life, it can be difficult to image how life could occur without the presence of this element.

At some point in early life, phosphorus became incorporated into living organism, and continues to persist as a necessity to life's maintenance. Therefore, we would like to obtain a better understanding as to how life began to incorporate phosphorus into its cellular machinery. For the purposes of this project, we will investigate how the deoxynucleotide triphosphate (dNTP) became the main phosphorus containing group involved in the preservation of genetic information via DNA synthesis.

DNA synthesis and storage of information

DNA is important for organisms to exist as it provides the code necessary for life. Cells utilize DNA as the instructions to build proteins [Koonin], and without this information, cells would not have the necessary molecular machinery to sustain themselves [McLachlan]. Therefore, it is equally important for cells to be able to effectively and accurately copy its DNA during cell division because too many mistakes will lead to a loss of genetic information eventually leading to cell death or the development of metastatic cancers [Teng X]. Each cell in the human body, for example, must perfectly copy the 6.4 billion nucleotides that DNA is composed of upon performing cell division [Nurk S], which is an incredible task considering the likelihood of incorporating an error in the genetic information. For that reason, it is worth learning about the role that the triphosphate substrate plays in DNA polymerization and DNA synthesis.

A DNA polymerase copies DNA any time a cell must undergo cell division [DePamphilis]. For a DNA polymerase to do this, there is a list of materials necessary to perform this copying of the DNA. This list includes the DNA template (the original strand of DNA) that must be copied, the DNA polymerase itself which catalyzes the copying of the DNA, magnesium as a co-factor for catalysis, and a primer strand of DNA or RNA where the polymerase may extend DNA synthesis [Smith]. The last molecule that a polymerase needs is a deoxynucleotide triphosphate (dNTP). This substrate that, once incorporated in the DNA strand, makes up the backbone of the double helix of the DNA structure [Travers A]. In a general sense, after the polymerase

2

attaches itself to the primer/template complex, it correctly selects the next nucleotide triphosphate in solution that will create correct base pairing (A matching with T and G matching with C). Therefore, DNA polymerases are essential in copying genetic information.

The aim of this work is to explore how a polymerase can perform this reaction with such high fidelity and specificity. Particularly, we would like to explore how it came to be that a DNA polymerase chose such a high energy phosphate-containing substrate (dNTP), and we would like to investigate this from an origins of life perspective. We would like to understand how DNA polymerases chose to use triphosphates in the first place.

In previous work from the Lupták Lab, it has already been discovered that triphosphates may not be the only substrate polymerases recognize and utilize when copying DNA [Burke]. In 2018, it has been discovered that several mesothermic and thermal DNA polymerases such as *Taq*, *Vent(-exo)*, and *Pfu* can utilize and recognize a lower energy substrate known as a deoxynucleotide diphosphate (dNDP). This is extraordinary, considering the substrate specificity of most naturally occurring enzymes [Juárez-Vázquez]. Usually, enzymes don't readily accept any other substrate than the preferred starting material, but in this case, dNDPs can be utilized by several DNA polymerases. The difference between a dNTP and dNDP is that there is one less phosphate group on the dNDP, making it a similar, yet lower energy substrate (Figure 1). Therefore, our hypothesis is that there is an amino acid sequence within the DNA polymerase that is conserved from prebiotic life that still allow the polymerase to

recognized diphosphates, even though currently a triphosphate is preferred. Put in another way, our hypothesis states a dNDP is a chemical precursor of the dNTP because of the structural conservation of polymerase from early life that allows it to use dNDPs.

To test this hypothesis, it then becomes necessary to simulate *in vitro* early earth conditions. This is done by utilizing a dNDP to test polymerase activity for DNA synthesis. If polymerases can utilize dNDP in a similar way to dNTPs, then it would provide support for the hypothesis. From previous research [Burke], we know that dNDPs can be recognized and utilized by various DNA polymerase, but we do not know whether the same polymerase fidelity exists when utilizing the different substrate for longer templated sequences with high diversity in sequence context. In short, we will be performing a long DNA with a much larger diversity in sequence context, while using the dNDP and comparing polymerization with a dNTP.

Figure 1. Structural relationship between a deoxynucleotide diphosphate (dNDP) and deoxynucleotide triphosphate (dNTP)

Highlighted are the key structural differences between a deoxynucleotide diphosphate (left) and the deoxynucleotide triphosphate (right). Both are composed of the nucleobase, deoxyribose sugar, and the acidic phosphate group. The major difference between the two is that the dNDP lacks a phosphate group that the dNTP contains. Though this difference exists, several DNA polymerases can still recognize and utilize a dNDP, meaning that the extra phosphate group is not necessary for polymerization.

dNDP   vs   dNTP
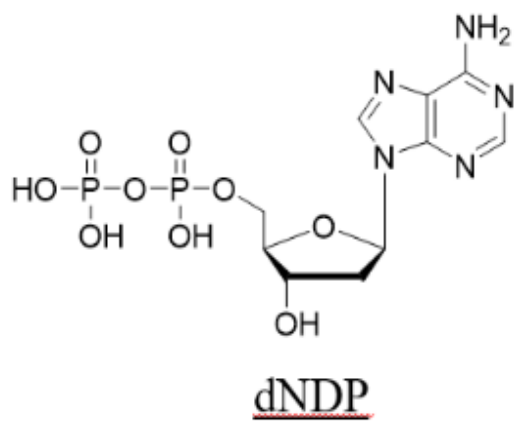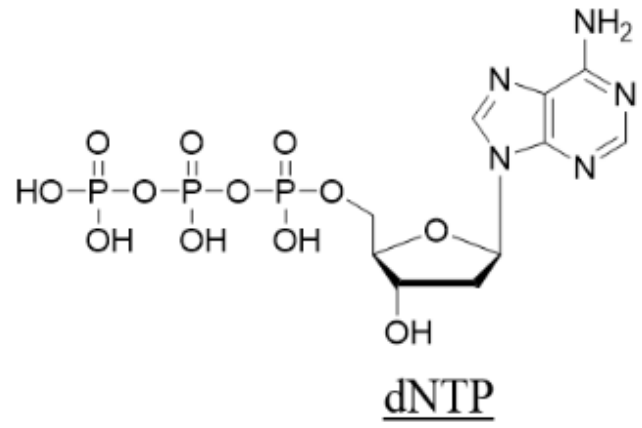
Created using ChemDraw ®

Results and Discussions

Quantification of Mismatch Pairs in DNA Polymerase Fidelity

Even though it is generally known that a DNA polymerase can copy DNA with high accuracy (fidelity), very little evidence has been provided to understand the mechanism by which the polymerase is able to distinguish between choosing the correct nucleotide triphosphate when participating in DNA synthesis. One approach to understanding this mechanism of correct nucleotide incorporation would be to look at error rates of incorporation [Song Y-S]. The fidelity of a polymerase can be quantified by synthesizing a strand of DNA and counting the number of mistakes (mismatches) that occur while the polymerase reads the template strand and incorporates nucleotides onto the primer. In this case, we will synthesize DNA with dNDPs and dNTPs in separate experiments to compare the number of mistakes made when synthesizing DNA between the two conditions.

Two problems that happen when looking at the frequency of mistakes after DNA synthesis. First, though a DNA polymerase can continue to elongate a DNA primer with multiple mistakes, if too many mismatches occur, the DNA polymerase will no longer continue to DNA synthesis [Bertram]. We don't know if there is a bias as to which template sequences may prevent DNA synthesis if we decided to use the dNDPs. This will prevent the full length of the template from being copied creating a problem for downstream analysis. Also, the frequency of mistakes changes depending on the sequence content of the template [de Paz AM, Cybulski TR]. If there are too many G/C

7

nucleotides in a row in the template sequence, for example, there is a higher likelihood for the polymerase to make more mistakes as it incorporates the substrate into the primer strand [Bertram]. This happens not just because of the exchanged substrate but because of the nucleotide order within the template. We would like to understand DNA synthesis when a dNDP is used, but we would also like to avoid the two previously mentioned problems.

To address these two issues simultaneously, it becomes important to synthesize various DNA templates with multiple different sequences of DNA with high nucleotide diversity all at once. By increasing the nucleotide diversity within the template strand, we will be able to synthesize various combinations of nucleotide series. This would allow us to test multiple sequence contexts while also allowing us to analyze certain sequence motifs that may prevent full-length product from being formed when using a dNDP substrate. Therefore, an adapter ligated genomic DNA library is used to provide multiple DNA template strands that call all be copied after just one reaction. Specifically, we will be using the PhiX genomic DNA library to accomplish the synthesis of a variety of DNA templates simultaneously.

A genomic DNA library takes the entire genetic information of a certain species, randomly sheers the full-length genome to a certain average size (200-500 nucleotides), then flanks the end of this worked up DNA with constant primer regions through a ligation process [Head SR]. In this case, we are using the genome of PhiX, which is a bacteriophage only composed of about five thousand three hundred nucleotides [Sanger F]. Working with a smaller genome makes the process of counting the correct

incorporation of nucleotides a lot easier. Since the genomic DNA is flanked by constant

primer regions, we can use the same forward and reverse DNA primer to amplify the

DNA and choose the specific substrate reaction conditions to see if we can get full

length product. After the DNA is verified to form full length product, it then becomes

possible to submits our samples to for high-throughput sequencing and follow up

bioinformatic analysis to quantify the number of mistakes made from either dNDP

synthesis of dNTP synthesis. Copying the DNA using dNTP would be our control in this

case. We will be synthesizing DNA using both dNTP and dNDPs in separate reactions

to see how many mistakes are being made by the DNA polymerase.

Before we perform this, we need first to verify whether the dNDP conditions can

amplify the PhiX genome. The PhiX library is much larger in length (~500 nucleotides)

compared to DNA templates used previously for dNDP synthesis (~130 nucleotides)

[Burke]. Also, this is the first time a DNA library is used to copy DNA with dNDPs.

Previously a single simple template was used with one type of sequence context, and it

has been yet to be established if it is possible to copy a DNA library using dNDPs. For

this reason, it then becomes necessary to test this process.

We decided to test long DNA synthesis by dNDPs by first performing a single

diphosphate replacement to perform DNA amplification via PCR. Instead of replacing all

four triphosphates (dGTP, dATP, dCTP, dTTP) with diphosphates, we systematically

replaced one triphosphate with a diphosphate to determine if we could obtain full length

product (i.e. **dGDP**, dATP, dCTP, dTTP). Because diphosphates incorporated at about

20 times slower rate than triphosphates [Burke] we also increased the time of reaction

9

to ninety seconds compared to the normal thirty-second time point when using triphosphates. First, replaced the guanine triphosphates with a guanine diphosphate, but we maintained typical PCR conditions for the remaining triphosphates (Figure 2). Each diphosphate was purified via ion exchange chromatography as previously described [Burke]. Here performed the PCR amplification with *Taq* DNA polymerase as it is the most well studied DNA polymerase within *in vitro* studies.

With the starting template concentration of 10 pM, it is possible to see that after just 20 cycles of PCR, full length product is formed using the dGDP replacement for 2 mM and 1 mM. When looking at the 0.2 mM dGDP condition, we do not see full length product meaning that we need a higher concentration of dGDP than 0.2 mM in order to obtain full length product. Three controls are present within this gel. Lane 5 shows normal PCR conditions with all four triphosphates present, and under these conditions it actually takes about 12 cycles to see full length product. This is consistent with previous work [Burke] as diphosphates are known to have a slower incorporation rate when using all four diphosphates to synthesize DNA, thus requiring more cycles of PCR. The last two controls are negative control as neither of them contains the template strand. Lane six (Figure 2) mimics lane two conditions but with no template present, while lane seven mimics lane five again without the template. These controls are added to ensure the polymerase is working properly, because if we were to see full length produce in these lanes, then all experimental conditions would be invalid. Overall, this gel in Figure 2 shows that replacing one triphosphate substrate with a diphosphate, we are able to amplify a genomic DNA library that is about 500 bps in length.

We extended this same procedure with a dCDP substrate again using *Taq* polymerase (Figure 3). The same reaction conditions were performed in which the highest concentration of dCDP was 2 mM as previously performed utilizing the dGDP. From the gel it is clear to see that no full-length product is formed under the dCDP condition. The positive control still produces full-length product, meaning that *Taq* did not producing full-length product under the dCDP with the number of PCR cycles and reaction time given. This same experiment was repeated with dTDP and finally dADP. Unfortunately, neither produce full length product as well (Figure 3).

This means that *Taq* has difficulty utilizing the dCDP, dTDP, and dADP substrates when amplifying this genomic library. Possible explanations for this difficulty in amplification may be that the library is either is too large for the polymerase to complete the amplification of the full length product, or that the reaction time of elongation of ninety seconds is not enough time for the polymerase to complete the amplification of the template. For this reason, we submitted samples for sequencing utilizing the dGDP substrate using three different concentrations (2 mM, 1 mM, 0. 5 mM). Though the 0.2 mM dGDP condition did not produce a visible product, the location by which the full-length library should have been when using 0.5 mM was still cut from the gel and sent in for analysis and sequence verification. As a positive control for sequencing, the dNTP amplification of the genomic library with *Taq* conditions was also submitted.

At this point we tested another polymerase to identify whether it can amplify the genomic library. We chose utilized *Vent(-exo)* DNA polymerase. *Vent(-exo)* is a variant

11

Figure 2. Amplification of PhiX genomic DNA library with dGDP and Taq Polymerase

We are testing to see if replacing a single triphosphate with a diphosphate will allow us to amplify the PhiX genomic DNA library. These are typical PCR reaction conditions including 1x commercial *Taq* buffer, 200 μM triphosphate (dTTP,dCTP,dATP), 1 μM forward and revere primer primer, 10 pM template, and 2 mM MgCl$_2$. We tested a concentration gradient of the replaced triphosphates (dGDP). Reaction conditions of each lane are provided bellow. No template controls (NTC) are provided for both di and triphosphate conditions. From the gel, it would appear we can obtain full length product of the PhiX library as indicated by the 500 bps ladder in lane one when using dGDP at both 2 mM and 1 mM dGDP concentration. Below this concentration, it is not possible to see the full-length product with these number of PCR cycles. This was a 2% agarose gel visualized with ethidium staining.
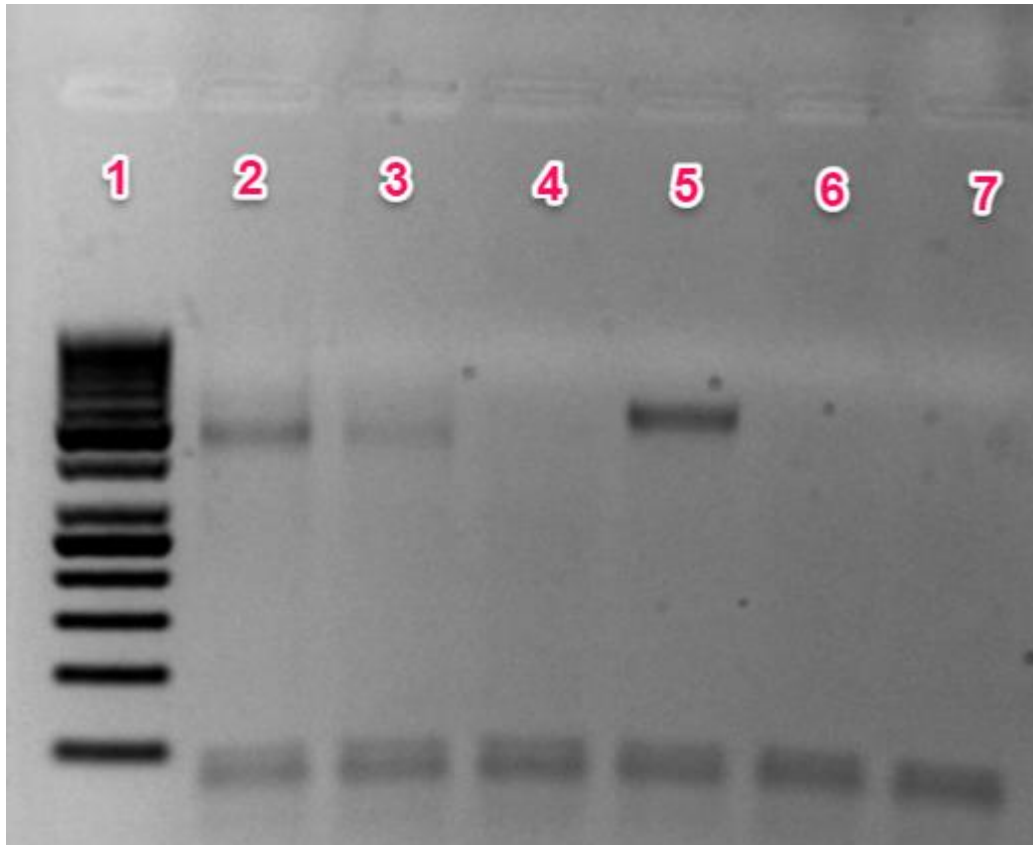
1. 50 Base Pair Ladder
2. 2 mM dGDP, 0.2 mM dNTP(-dGTP)
3. 1 mM dGDP, 0.2 mM dNTP(-dGTP)
4. 0.2 mM dGDP, 0.2 mM dNTP(-dGTP)
5. Positive Control [dNTPs, no dGDP] 12 Cycles
6. NTC [dNTPs, no dGTP, add dGDP at highest concentration (2 mM rxn)]
7. NTC [dNTPs]

of *Vent* polymerase but is mutated to disable the 3' to 5' exonuclease domain that many DNA polymerases contain [Kong H]. This exonuclease domain is an innate proof-reading capability that Vent usually contains. When it is not mutated, the polymerase can go backward and correct the mistakes it makes if it incorrectly incorporates nucleotides [Kong H]. More mistakes tend to can happen when utilizing a new substrate, as this is a prime example of how mutations can purposefully be introduced if performing some type of mutagenesis [McCullum EO]. Hence, if *Vent(-exo)* makes many mistakes when copying DNA using a dNDP, it then becomes more accessible to track as the exonuclease domain of *Vent(-exo)* will not be able to correct the mistakes introduced. But before we can get to this step of sequence analysis, again we need to see if *Vent(-exo)* can produce full length produce with a genomic DNA library (Figure 5).

From the previous gel, it can be understood that *Vent(-exo)* requires a higher concentration of dGDP compared to *Taq* polymerase. To get the same product formation, we need to drastically increase the amount of dGDP utilized in solution, probably because the dGDP substrate has a lower affinity to *Vent(-exo)* polymerase. This same single nucleotide replacement experiment was again repeated with the other diphosphates, but only dGDP is able to make fully product. From both different cases of the polymerases used, it can be concluded that only dGDP can effectively be utilized to synthesize the PhiX genomic library. This is consistent with the literature, as up to this point, there has only been one diphosphate present that can be utilized by a human DNA polymerase known as β human DNA polymerase [Varela FA]. This paper traps the dGDP substrate in the catalytic core of β human DNA polymerase right before catalysis

within a crystal structure. This paper only utilizes dGDP, meaning that perhaps DNA polymerase has a better tolerance in recognizing and incorporating dGDP compared to the other diphosphate nucleotides. Though it is not explicitly mentioned in the paper as to why the other diphosphates weren't included in their analysis, it may be likely that the DNA polymerase could not utilize the other diphosphates as effectively.

Conclusion

This work has shown that *Taq* and *Vent(-exo)* DNA polymerases can both extend the PhiX genomic DNA library to fully length product when a dGDP substrate replaces dGTP in conducting PCR. When performing this same experiment with the other diphosphate substrates, it has been shown that no observable. Overall, this implies these two DNA polymerases do not utilize the diphosphate substrate in the same way that is uses the triphosphates. When only dGDP can produce full length product, it becomes difficult to argue for the presence of diphosphates as an evolutionary precursor to the prebiotic world. Yet, one must consider that the genomic DNA library contains a sequence context that may not have existed during life's beginnings. It may be better to perform these same experiments with a couple of sequences at a time to control for the variability produced from amplifying multiple sequence simultaneously by the diphosphate substrate.

Figure 3. Amplification of PhiX genomic DNA library with dCDP and Taq Polymerase

When using dCDP to amplify DNA with *Taq* polymerase, we see that we do not observe full length product of the PhiX genomic library. These were the same reaction conditions utilized when amplifying dGDP, but full-length product is not formed when using dCDP. To make sure the polymerase still had activity, the positive control (lane 5) utilizing all dNTPs showed full length product. These were again 20 cycles of PCR under all concentration gradient conditions of the dCDP.
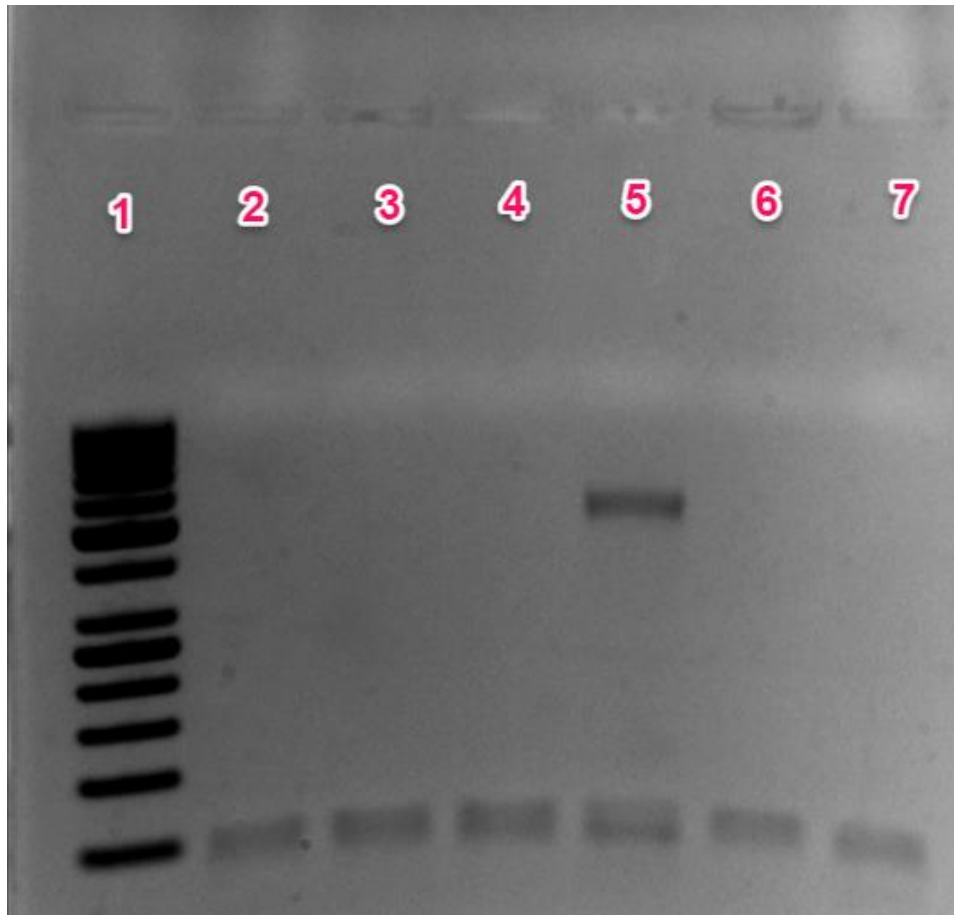
1. Ladder
2. 2 mM dCDP, 0.2 mM dNTP(-dCTP)
3. 1 mM dCDP, 0.2 mM dNTP(-dCTP)
4. 0.2 mM dCDP, 0.2 mM dNTP(-dCTP)
5. Positive Control [dNTPs, no dCDP] (12 Cycles)
6. NTC [dNTPs, no dCTP, add dCDP at highest concentration (2 mM rxn)]
7. NTC [dNTPs]

Figure 4. Amplification of PhiX genomic DNA library dTDP/dADP and Taq polymerase

The last two diphosphates were tested in these last PCR reactions. Neither dTDP (Left) nor dADP (right) could produce full length product when performing 20 cycles of PCR. Positive controls are present (lane 5) to make sure the polymerase could still work with dNTPs. To test whether increasing the number of cycles would increase the likelihood of producing more product, the number of PCR cycles for dCDP, dTDP, and dADP conditions were all increased having experienced 28 cycles of PCR, yet not full-length product can be seen on the gel.
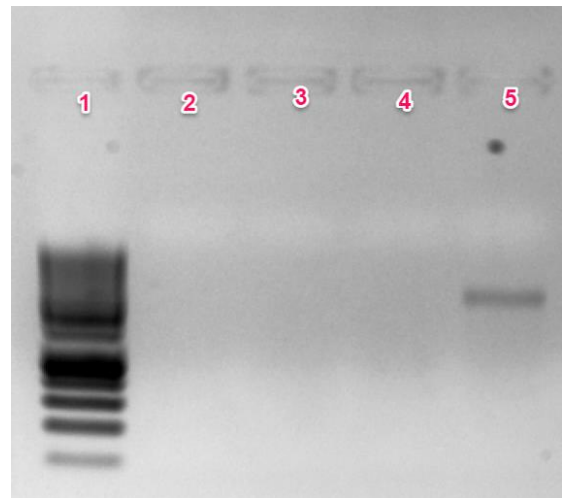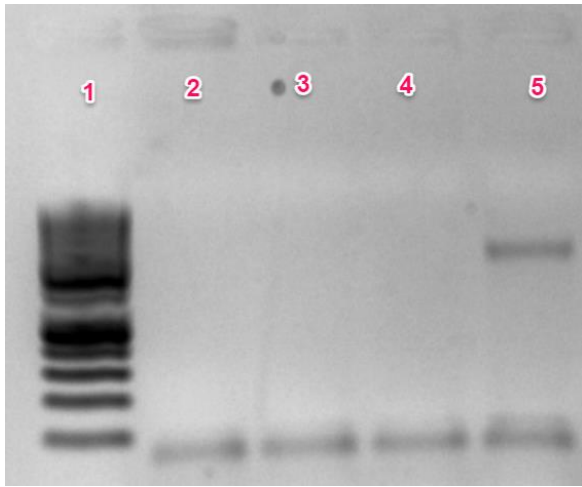
1. Ladder(50bps)
2. 2 mM dTDP/dADP, 0.2 mM dNTP(-dTDP/dADP)
3. 1 mM dTDP/dADP, 0.2 mM dNTP(-dTDP/dADP)
4. 0.2 mM dTDP/dADP, 0.2 mM dNTP(-dTDP/dADP)
5. Positive Control [dNTPs, no dTDP/dADP] (12 Cycles with same thermocycler conditions)

mutagenesis [McCullum EO]. Hence, if *Vent(-exo)* makes many mistakes when copying

DNA using a dNDP, it then becomes more accessible to track as the exonuclease

domain of *Vent(-exo)* will not be able to correct the mistakes introduced. But before we

can get to this step of sequence analysis, again we need to see if *Vent(-exo)* can

produce full length produce with a genomic DNA library (Figure 5).

From the previous gel, it can be understood that *Vent(-exo)* requires a higher

concentration of dGDP compared to *Taq* polymerase. To obtain same product

formation, we need to drastically increase the amount of dGDP utilized in solution,

probably because the dGDP substrate has a lower affinity to *Vent(-exo)* polymerase.

This same single nucleotide replacement experiment was again repeated with the other

diphosphates, but only dGDP is able to make fully product. From both different cases of

the polymerases used, it can be concluded that only dGDP can effectively be utilized to

synthesize the PhiX genomic library. This is consistent with the literature, because up to

point, there has only been one diphosphate present that can be utilized by a human

DNA polymerase known as β human DNA polymerase [Varela FA]. This paper traps the

dGDP substrate in the catalytic core of β human DNA polymerase right before catalysis

within a crystal structure. This paper only utilizes dGDP, meaning that perhaps DNA

polymerase has a better tolerance in recognizing and incorporating dGDP compared to

the other diphosphate nucleotides. Though it is not explicitly mentioned in the paper as

to why the other diphosphates weren't included in their analysis, it may be likely that the

DNA polymerase could not utilize the other diphosphates as effectively.

Figure 5. Amplification of PhiX genomic DNA library with Vent(-exo) and dGDP

Vent(-exo) polymerase is utilized to amplify the PhiX genomic DNA library under a

dGDP substrate gradient. A major difference that occurs here is that dGDP has to be

utilized at a higher concentration (5 mM) in order to be utilized by *Vent(-exo).* All other

conditions remain the same, except a higher amount of Mg2+ was utilized in reaction

the (4 mM). It took 20 cycles of PCR to get full form product to show up except for the

dNTP positive control (lane 5), which only took 12 cycles.
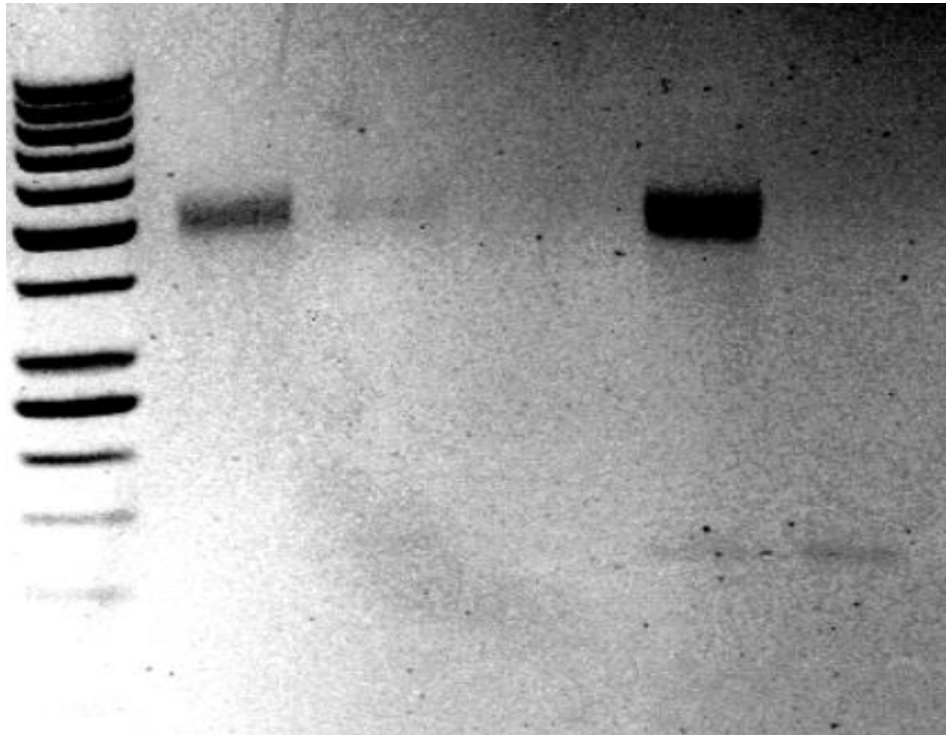
1. Ladder
2. 5 mM dGDP
3. 3 mM dGDP
4. 2 mM dGDP
5. dNTP
6. dNTP, NTC

Material and Methods

Diphosphate Purification

The nucleotide diphosphates were purchased from Sigma-Aldrich as a solid

sodium salt. Each diphosphate is then separately resuspended in about 3 to 5 mL of

double distilled water until fully dissolved and purified via ion exchange chromatography

as previously described [Burke]. The samples are then dried by vacuum filtration and

again resuspended in 1.5 mL of double distilled water. This process of drying and

resuspension is performed three times total for each diphosphate to remove any

remaining volatile salts present. Lastly, the diphosphates are individually resuspended

in 25 μL of nuclease free water so that they can be adjusted to neutral pH via the

addition 1 M sodium hydroxide (NaOH). The amount of NaOH added varied depended

on which diphosphate was in use.

From here, each diphosphate's concentration is measured by taking 2 μL of the

stock diphosphate solution and diluting in 8 μL of 10 mM Tris-HCl buffer, pH 7 @ 22 °C

to be measured by absorbance on the spectrophotometer. A serial dilution was

performed in the 10 mM Tris-HCl and a stock concentration was back calculated based

on absorbance values.

PhiX Genomic DNA Library

The DNA template chosen to be amplified is known as the PhiX Control v3 library

which was purchased from Illumina. This is an adapter-ligated genomic DNA library

meaning that each strand of DNA has consistent forward and reverse primer binding

23

regions in which only one set of primers are required for the initiation of DNA synthesis. The variation in nucleotide context come from the region between the primer regions as each strand of DNA has different portions of the PhiX genome. On average, the length of the DNA to be synthesized is about 500 base pairs long, which indicates fully length product upon PCR amplification. The sequences of the forward and reverse primers used to perform the amplification are provided below:

Forward Primer

5' AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T 3'

Reverse Primer

5' GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC TGA CCA ATC TCG TAT GCC GTC TTC TGC TTG 3'

Single Triphosphate Replacement Experiment

Amplification conditions for the two tested polymerases are similar in their set up. The PhiX DNA, forward/reverse primers, and polymerase are introduced to a PCR tube. One of the four triphosphates will be replaced by a diphosphatase, as such the appropriate diphosphate must be introduced to substitute the missing triphosphate (i.e. if a dGTP is removed, it is replaced by a dGDP). Stock solutions of the individual triphosphates were purchases from Sigma-Aldrich to select the appropriate combination of nucleotides in solution for each reaction condition. All these materials are present with their respective polymerase buffer solution make up a total volume of 10 μL before

being placed into the thermocycler. A positive control of all triphosphates present was also used to ensure proper polymerase. Unless otherwise state, twenty cycles of PCR per performed under the specified reaction conditions below. Each gel was run on a 2 w/v % agrose gel in 0.5 X TBE buffer.

Thermocycler Conditions for All Experiments

Stage	Time (sec)	Temperature(°C)
Initiation	30	95
Annealing	30	65
Elongation	90	72

Taq Reaction Conditions

Component	Stock Concentration	Volume (μL)	Reaction Concentration
Taq DNA Polymerase	0.5 U/μL	1	0.05 U/μL
Taq Buffer	10x	1	1x
dNTP(-dXTP)**	2 mM	1	0.2 mM
Forward Primer	10 μM	1	1 μM
Reverse Primer	10 μM	1	1 μM
Template	100 pM	1	10 pM
MgCl$_2$	20 mM	1	2 mM
dXDP**	10 mM	Varies	2, 1, 0.1 mM

Nuclease Free H2O	-	Varies	-
Total Volume	-	10	-

** The nucleotides introduced come with equal molar amounts of magnesium ions

*1X Standard Taq Reaction Buffer****

10 mM Tris-HCl

50 mM KCl

(pH 8.3 @ 25°C)

*** Buffer made in-house; the commercially available product was not used to prevent polymerase enhancers from altering results

Vent(-exo) Reaction Conditions

Component	Stock Concentration	Volume (μL)	Reaction Concentration
Vent(-exo) Polymerase	0.75 U/μL	1	0.075 U/μL
Vent(-exo) Buffer	10x	1	1x
dNTP(-dXTP)**	2 mM	1	0.2 mM
Forward Primer	10 μM	1	1 μM
Reverse Primer	10 μM	1	1 μM
Template	100 pM	1	10 pM
MgSO$_4$	40 mM	1	4 mM

dXDP**	20 mM	Varies	5, 3, & 2 mM
Nuclease Free H2O	-	Varies	-
Total Volume	-	10	-

** The nucleotides introduced come with equal molar amounts of magnesium ions

*1X Standard Vent(-exo) Reaction Buffer****

20 mM Tris-HCl

10 mM $(NH_4)_2SO_4$

10 mM KCl

0.1% Triton® X-100

(pH 8.8 @ 25°C)

*** Buffer made in-house

Works Referred

(1) Bertram, J. G., Oertell, K., Petruska, J., & Goodman, M. F. (2009). DNA polymerase fidelity: Comparing direct competition of right and wrong dntp substrates with steady state and pre-steady state kinetics. *Biochemistry*, *49*(1), 20–28. https://doi.org/10.1021/bi901653g

(2) Burke, C. R., & Lupták, A. (2018). DNA synthesis from diphosphate substrates by DNA polymerases. *Proceedings of the National Academy of Sciences*, *115*(5), 980–985. https://doi.org/10.1073/pnas.1712193115

(3) DePamphilis, M. L. (2000). Review: Nuclear structure and DNA replication. *Journal of Structural Biology*, *129*(2-3), 186–197. https://doi.org/10.1006/jsbi.2000.4219

(4) de Paz, A. M., Cybulski, T. R., Marblestone, A. H., Zamft, B. M., Church, G. M., Boyden, E. S., Kording, K. P., & Tyo, K. E. J. (2018). High-resolution mapping of DNA polymerase fidelity using nucleotide imbalances and next-generation sequencing. *Nucleic Acids Research*, *46*(13). https://doi.org/10.1093/nar/gky296

(5) Goñi, F. M. (2014). The basic structure and dynamics of cell membranes: An update of the singer–nicolson model. *Biochimica Et Biophysica Acta (BBA) - Biomembranes*, *1838*(6), 1467–1476. https://doi.org/10.1016/j.bbamem.2014.01.006

(6) Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh,

F., Salomon, D. R., & Ordoukhanian, P. (2014). Library construction for next-generation sequencing: Overviews and challenges. *BioTechniques*, *56*(2), 61–77. https://doi.org/10.2144/000114133

(7) Hickey, R. J., & Malkas, L. H. (1997). Mammalian cell DNA replication. *Critical Reviews™ in Eukaryotic Gene Expression*, *7*(1-2), 125–157. https://doi.org/10.1615/critreveukargeneexpr.v7.i1-2.80

(8) Juárez-Vázquez, A. L., Edirisinghe, J. N., Verduzco-Castro, E. A., Michalska, K., Wu, C., Noda-García, L., Babnigg, G., Endres, M., Medina-Ruíz, S., Santoyo-Flores, J., Carrillo-Tripp, M., Ton-That, H., Joachimiak, A., Henry, C. S., & Barona-Gómez, F. (2017). Evolution of substrate specificity in a retained enzyme driven by Gene Loss. *ELife*, *6*. https://doi.org/10.7554/elife.22679

(9) Kong, H., Kucera, R. B., & Jack, W. E. (1993). Characterization of a DNA polymerase from the hyperthermophile Archaea Thermococcus litoralis. vent DNA polymerase, steady state kinetics, thermal stability, processivity, strand displacement, and exonuclease activities. *Journal of Biological Chemistry*, *268*(3), 1965–1975. https://doi.org/10.1016/s0021-9258(18)53949-1

(10)   Koonin, E. V. (2015). Why the central dogma: On the nature of the Great Biological Exclusion Principle. *Biology Direct*, *10*(1). https://doi.org/10.1186/s13062-015-0084-3

(11)   McCullum, E. O., Williams, B. A., Zhang, J., & Chaput, J. C. (2010). Random mutagenesis by error-prone PCR. *Methods in Molecular Biology*, 103–109. https://doi.org/10.1007/978-1-60761-652-8_7

(12)   McLachlan, A. D. (1972). Protein structure and function. *Annual Review of Physical Chemistry*, *23*(1), 165–192. https://doi.org/10.1146/annurev.pc.23.100172.001121

(13)   Michigami, T., Kawai, M., Yamazaki, M., & Ozono, K. (2018). Phosphate as a signaling molecule and its sensing mechanism. *Physiological Reviews*, *98*(4), 2317–2348. https://doi.org/10.1152/physrev.00022.2017

(14)   Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., Vollger, M. R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S. J., Diekhans, M., Logsdon, G. A., Alonge, M., Antonarakis, S. E., Borchers, M., Bouffard, G. G., Brooks, S. Y., … Phillippy, A. M. (2022). The complete sequence of a human genome. *Science*, *376*(6588), 44–53. https://doi.org/10.1126/science.abj6987

(15)   Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., Hutchison, C. A., Slocombe, P. M., & Smith, M. (1977). Nucleotide sequence of bacteriophage ΦX174 DNA. *Nature*, *265*(5596), 687–695. https://doi.org/10.1038/265687a0

(16)   Smith, Douglas W., Schaller, Heinze., & Bonheffer, Friedrich J. (1970). DNA synthesis in vitro. *Nature*, *226*(5247), 711–713. https://doi.org/10.1038/226711a0

(17)   Song, Y.-S., Shu, Y.-G., Zhou, X., Ou-Yang, Z.-C., & Li, M. (2016). Proofreading of DNA polymerase: A new kinetic model with higher-order terminal effects. *Journal of Physics: Condensed Matter*, *29*(2), 025101.

https://doi.org/10.1088/0953-8984/29/2/025101

(18)   Teng, X., & Hardwick, J. M. (2015). Cell death in genome evolution. *Seminars in Cell & Developmental Biology*, *39*, 3–11. https://doi.org/10.1016/j.semcdb.2015.02.014

(19)   Travers, A., & Muskhelishvili, G. (2015). DNA structure and function. *FEBS Journal*, *282*(12), 2279–2295. https://doi.org/10.1111/febs.13307

(20)   Varela, F. A., & Freudenthal, B. D. (2021). Mechanism of deoxyguanosine diphosphate insertion by human DNA polymerase β. *Biochemistry*, *60*(5), 373–380. https://doi.org/10.1021/acs.biochem.0c00847

(21)   Varela, F. A., & Freudenthal, B. D. (2021). Mechanism of deoxyguanosine diphosphate insertion by human DNA polymerase β. *Biochemistry*, *60*(5), 373–380. https://doi.org/10.1021/acs.biochem.0c00847

(22)   Šponer, J., Mládek, A., Šponer, J. E., Svozil, D., Zgarbová, M., Banáš, P., Jurečka, P., & Otyepka, M. (2012). The DNA and RNA sugar–phosphate backbone emerges as the key player. an overview of quantum-chemical, Structural Biology and Simulation Studies. *Physical Chemistry Chemical Physics*, *14*(44), 15257. https://doi.org/10.1039/c2cp41987d