Recording Information into DNA

by

María Elisa Tanner

Department of Mechanical Engineering and Materials Science
Duke University

Date:_______________________

Approved:

___________________________
Jonathan M. Protz, Supervisor

___________________________
Chris Dwyer

___________________________
Thomas H. LaBean

___________________________
Anne A. Lazarides

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Mechanical Engineering and Materials Science in the Graduate School of Duke University

2009
ABSTRACT

Recording Information into DNA

by

María Elisa Tanner

Department of Mechanical Engineering and Materials Science
Duke University

Date:_______________________

Approved:

___________________________

Jonathan M. Protz, Supervisor

___________________________

Chris Dwyer

___________________________

Thomas H. LaBean

___________________________

Anne A. Lazarides

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Mechanical Engineering and Materials Science in the Graduate School of Duke University

2009
Abstract

The objective of this research is to develop the concept of “genetic memory”, the storage of information into genetic material, through the demonstration of the feasibility and benefits of recording the time-history of one or more environmental state variables into genetic material. First, the amount of information that could be stored into non-coding DNA using a lossy mechanism such as regulated diffusion is determined by developing a mathematical model. Next, a mechanism through which this concept of sensing, recording, and storing information on the nanoscale could be accomplished is proposed.

In conjunction with DNA, which is the actual means of storing information, the proposed approach for the realization of genetic memory also uses thermosensitive liposomes as a means of sensing state variables and acting as a valve to transport and release the DNA in response to the appropriate stimuli. Each variant of thermally sensitive liposomes contains a unique DNA sequence that serves as an identifier. Upon release, the DNA encounters ligase, ATP, and other cofactors and binds – preserving a record of the stimulus experienced by the liposomes. Liposomes, through careful design, can be developed to have unique transition temperatures at which the permeability rate of the contents encapsulated by the liposome increases. It is only at or above this transition temperature that a liposome will become porous.
Through modeling and experimentation, the feasibility of the liposome/DNA system as a mechanism for information storage is successfully demonstrated.
## Contents

Abstract .........................................................................................................................................iv

List of Tables ..................................................................................................................................x

List of Figures ..................................................................................................................................xi

Acknowledgements .......................................................................................................................xvi

1. Introduction ............................................................................................................................... 1
   1.1 Motivation .........................................................................................................................1
   1.2 Objective and Specific Aims .............................................................................................3
   1.3 Approach for Proposed Mechanism of Genetic Memory ..................................................4
   1.4 Specific Scholarly Contribution ..........................................................................................6
   1.5 Advancing the State of the Art .........................................................................................6
   1.6 Novelty ...............................................................................................................................7
   1.7 Potential Applications .......................................................................................................7
   1.8 Summary of Chapters ........................................................................................................10

2. Literature Review .................................................................................................................... 12
   2.1 Information Theory .........................................................................................................12
      2.1.1 Lossy Data Conversion ............................................................................................13
      2.1.2 Noisy Channel Coding Theorem .............................................................................13
   2.2 Biomolecular Computing .................................................................................................15
      2.2.1 Materials of Biomolecular Computing .....................................................................17
      2.2.2 Processes of Biomolecular Computing ....................................................................18
2.2.2.1 Ligation ................................................................................................................ 18
2.2.2.2 Enzymes ..............................................................................................................21
2.2.2.3 Gel Electrophoresis ............................................................................................21
2.2.2.4 Polymerase Chain Reaction .............................................................................. 22
2.2.2.5 DNA Sequencing ................................................................................................ 22
2.3 Liposomes as a Means of Regulation ....................................................................... 23
2.3.1 Composition ............................................................................................................... 25
2.3.2 Methods of Preparation ............................................................................................ 28
2.3.2.1 Reverse-Phase Evaporation ..............................................................................28
2.3.2.2 Thin Film Method ..............................................................................................29
2.3.3 Characterization ........................................................................................................30
2.3.4 Loading and Release .................................................................................................32
2.3.5 Applications ...............................................................................................................35
3. Mathematical Model ........................................................................................................ 36
3.1 Liposome Permeability .................................................................................................38
3.2 Diffusion ..........................................................................................................................40
3.3 Reaction Kinetics ..........................................................................................................47
3.4 Monomer Incorporation .................................................................................................48
3.4.1 Symbol Definitions .................................................................................................... 49
3.4.2 Probability of Concentration ....................................................................................50
3.4.3 Relationship between Input and Signal ................................................................. 54
3.4.4 Energy Spectral Densities ........................................................................................ 57
3.4.5 Signal-to-Noise Ratio ................................................................. 60
3.4.6 Spectral Efficiency ................................................................. 62

4. Simulation ......................................................................................... 65
4.1 Liposome Permeability ............................................................... 65
4.2 Diffusion ...................................................................................... 66
4.3 Reaction Kinetics ....................................................................... 68
4.4 Monomer Incorporation ............................................................. 72
4.5 Results ....................................................................................... 73

5. Materials and Methods .............................................................. 74
5.1 Materials .................................................................................... 74
5.1.1 Liposomes ............................................................................. 74
5.1.2 DNA ..................................................................................... 75
5.1.2.1 Sticky Ends ...................................................................... 75
5.1.2.2 Nucleotides ....................................................................... 75
5.1.2.3 Single Strand DNA ......................................................... 76
5.1.2.4 Primers ............................................................................ 76
5.1.3 Ligase ................................................................................... 77
5.2 Approach .................................................................................. 78
5.2.1 Sticky Ends ........................................................................... 78
5.2.2 Nucleotides ........................................................................... 80
5.2.3 Single Strand DNA ............................................................... 81
5.2.4 Basic Experimental Procedure ............................................. 82
6. Initial Experiments and Results

   6.1 Efficiency of Sephadex Filtering
   6.2 Confirmation of Ligation
   6.3 Validation of Encapsulation
   6.4 Verification of Liposome Release

7. Final Experiments and Results

   7.1 Sticky Ends
      7.1.1 Experimental Parameters
      7.1.2 Results
   7.2 Nucleotides
   7.3 Single Strand DNA
      7.3.1 Experimental Parameters
      7.3.2 Simulation Results
      7.3.3 Data Analysis
      7.3.4 Statistical Analysis
      7.3.5 Discussion

8. Conclusions

References

Biography
List of Tables

Table 1: Variables used in the calculation of the diffusion and characteristic time scale of
the system.........................................................................................................................68

Table 2. Liposomes used experimentally and their respective transition temperatures..74

Table 3: Sticky end sequences used in experiments...............................................................75

Table 4: Single strand DNA sequences used in experiments...........................................76

Table 5: Start primers used in experiments.............................................................................77

Table 6: Sequencing primers used in experiments.................................................................77

Table 7: Experimental set-up for ligation experiments. The following amounts of single
strand DNA were used in each sample ...................................................................................88

Table 8: Concentration data from samples obtained from Nanodrop. .........................89

Table 9: For each of the iterations of the settle method of filtering of Sample B, the
volume of supernatant removed and replaced, the measured concentration of the
supernatant, and moles calculated are summarized. ............................................................92

Table 10: For each of the iterations of the settle method of filtering for Sample E, the
volume of supernatant removed and replaced, the measured concentration of the
supernatant, and moles calculated are summarized. ............................................................93

Table 11: Division of liposomes and temperatures used in verifying the release of the
encapsulated contents. .........................................................................................................94

Table 12: Strand lengths generated from the simulation for each of the temperature
profiles........................................................................................................................................114

Table 13: Range of correlations necessary to achieve a 95% or 98% confidence interval for
the realized threshold...........................................................................................................121
List of Figures

Figure 1: An example of a (a) DNA sticky end and a (b) DNA blunt end......................... 19

Figure 2. Liposomes spontaneously form from dry lipid films that are exposed to an aqueous environment. Picture from Avantilipids [47]................................................................. 23

Figure 3: Phospholipids are composed of a polar head and two fatty acid, non-polar tails (left) and frequently form lipid bilayers (right)................................................................. 26

Figure 4: Permeability rate (1/min) vs. Temperature [86]. ....................................................34

Figure 5: Block diagram of the proposed mechanism kinetics from the change in temperature to the resulting DNA sequence. ................................................................. 36

Figure 6: Two potential data strands given the liposomes’ permeability at temperatures higher than its transition temperature. For both strands, a sample graph of the temperature experienced is displayed. (left) Given a temperature, the corresponding liposome releases the assigned sequence as well as all thermally sensitive liposomes with transition temperatures less than the given temperature in a mix of units. (right) Given a temperature, only the corresponding liposome releases the assigned sequence................................................................. 39

Figure 7: The aqueous space of the liposome in which the encapsulated DNA is held is shown as the holding reservoir, the lipid bilayer is denoted as the valve, and the outside solution where the reaction would take place is labeled as the chamber. The three points of interest of the device when determining the chemical dynamics are the start and end of the valve and the chamber. ................................................................. 41

Figure 8: Bode plots for the chemical dynamics: (top) amplitude and (bottom) phase.... 43

Figure 9: Modified block diagram of diffusion dynamics that takes into account the time lag due to the sequences diffusing across the lipid bilayer......................................................... 44

Figure 10: Three DNA concentration points of interest in the device when determining the valve dynamics are the concentration inside the liposome $C_1$, the concentration at the midpoint of the lipid bilayer $C_2$, and the concentration at the “entrance” to the solution $C_3$................................................................. 45
Figure 11: Block diagram of the system being modeled. The input is fed through a channel and becomes the signal. Noise is added to the signal and the output is produced. ..............................................................................................................................................48

Figure 12: Variable definitions used in derivation.................................................................50

Figure 13: Bode plot of transfer function $G'(\omega)$ ...............................................................57

Figure 14: Plot of SNR (red line) and selected approximation of SNR (black dots)..........62

Figure 15: Comparison of the storage density (Hz/in\(^2\)) and spectral efficiency of DNA recording to other electronic devices. .................................................................64

Figure 16: Once a liposome is heated to its transition temperature, DNA is released, subsequently stored, and incorporated. These quantities are conserved. ....................66

Figure 17: Graph of the concentration of unbound DNA in solution in which two varieties of liposomes which different transition temperatures are present. The x-axis is the time of the experiment in minutes; the y-axis is the concentration of the unbound DNA in solution, and the secondary y-axis is the temperature of the solution. The concentration of the unbound DNA from the lower transition temperature liposome is shown in dark blue; the concentration of the unbound DNA from the higher transition temperature liposome is shown in magenta; and the temperature profile is shown in cyan. ...........................................................................................................................................70

Figure 18: Methodology of the simulation. When the liposome is open, the concentration of the solution and reservoir are adjusted. If while opened, the liposomes just transitioned, the reference concentrations and time are updated. If the liposomes are not open, the concentration of the solution is recalculated. ........................................71

Figure 19: Sample DNA sequence generated with the simulation. The start primer, tail, sequence for the low transition liposome, and sequence for the high transition liposome are indicated. .....................................................................................................................................73

Figure 20: Single strand DNA concept: hashed colors denote the complements to the sold segments encapsulated in the liposomes. Upon release, the DNA from the liposomes binds with the complements available in solution which preserves the time history data. ..................................................................................................................................82

Figure 21: Graph of the DNA concentration in the supernatant: concentration of the supernatant (ng/\(\mu\)l) vs. runs. Blue diamonds represent measurements from a Sephadex
column. Red squares represent measurements from the settle method using an extruded sample. Green triangles represent measurements from the settle method using a non-extruded sample.

Figure 22: Polyacrylamide gel from a single strand DNA ligation experiment. Lane 2 is a 20/100 DNA ladder and Lanes 3-10 are samples using different experiment samples. The gel demonstrates that binding occurred and that the samples with both liposome DNA present were longer than the solutions with only one liposome DNA present.

Figure 23: Expected (top) and actual (bottom) results from liposome release experiment using DPPC and DHPC. The area of the gray, shaded circles represent the maximum concentration from the three samples; the area of the black, solid circles represent the average concentration of the three samples; and the area of the white circles represent the minimum concentration from the three samples.

Figure 24: Expected (top) and actual (bottom) results from liposome release experiment using DPPC and DSPC. The area of the gray, shaded circles represent the maximum concentration from the three samples; the area of the black, solid circles represent the average concentration of the three samples; and the area of the white circles represent the minimum concentration from the three samples.

Figure 25: Procedure used in preliminary experiments. Two varieties of liposomes are heated. At 41°C, Liposome A’s transition temperature is attained and its DNA fragments are released and bound. At 54°C, Liposome B’s transition temperature is achieved and both A and B release DNA fragments which are bound.

Figure 26: Temperature profiles used in preliminary experiments. A denotes Liposome A only. B denotes Liposome B only. AB1 and AB2 denote both Liposomes A and B.

Figure 27: Preliminary data from experiments testing Liposome A (top) and B (bottom) at their respective transition temperatures. Axes clockwise from bottom: Nondimensionalized Strand Length, log (CG/AT), Nondimensionalized Time, and Temperature (°C). The blue solid line denotes the log (CG/AT) with respect to time. The red dashed lines represent the respective ratios of the sequences stored in Liposome A and B. The green dotted line represents the temperature profile for the specific experiment.

Figure 28: Preliminary data from experiments testing Liposome A and B at the lower transition temperature of A (top) and at both transition temperatures along with the two methods of adding ligase to the experiment (bottom). Axes clockwise from bottom:
Nondimensionalized Strand Length, log (CG/AT), Nondimensionalized Time, and Temperature (°C). The blue solid line denotes the log (CG/AT) with respect to time. The red dashed lines represent the respective ratios of the sequences stored in Liposome A and B. The green dotted line represents the temperature profile for the specific experiment.

Figure 29: Variations in parameters that could be used in single strand experiments...

Figure 30: Experiment matrix used in the single strand experiment describing the combinations of liposomes, DNA, ligase, and temperature profiles used.

Figure 31: Graph of the concentration of unbound DNA in solution in which two varieties of liposomes which different transition temperatures are present. The x-axis is the time of the experiment in minutes; the y-axis is the concentration of the unbound DNA in solution, and the secondary y-axis is the temperature of the solution. The concentration of the unbound DNA from the lower transition temperature liposome is shown in dark blue.; the concentration of the unbound DNA from the higher transition temperature liposome is shown in magenta; and the temperature profile is shown in cyan.

Figure 32: Graph of the concentration of unbound DNA in solution in which two varieties of liposomes which different transition temperatures are present. The x-axis is the time of the experiment in minutes; the y-axis is the concentration of the unbound DNA in solution, and the secondary y-axis is the temperature of the solution. The concentration of the unbound DNA from the lower transition temperature liposome is shown in dark blue.; the concentration of the unbound DNA from the higher transition temperature liposome is shown in magenta; and the temperature profile is shown in cyan.

Figure 33: Graph of the concentration of unbound DNA in solution in which two varieties of liposomes which different transition temperatures are present. The x-axis is the time of the experiment in minutes; the y-axis is the concentration of the unbound DNA in solution, and the secondary y-axis is the temperature of the solution. The concentration of the unbound DNA from the lower transition temperature liposome is shown in dark blue.; the concentration of the unbound DNA from the higher transition temperature liposome is shown in magenta; and the temperature profile is shown in cyan.

Figure 34: Plots of the simulation and experimental cumulative data and first harmonics for the temperature profiles of low-high and high-low-high. The correlation in A

xv
between the simulation and experimental data is shown in the upper left corner of each plot. ................................................................. 115

Figure 35: Plots of the simulation and experimental cumulative data and first harmonics for each of the tested temperature profiles. The correlation in A between the simulation and experimental data is shown in the upper left corner of each plot. .............................. 116

Figure 36: Experiment matrix used with additional columns for the correlation in A, C, G, and T between the simulation data and experimental data. Red indicates trials with expected poor correlation and blue indicates trials with expected good correlation. .... 117

Figure 37: Plot of the correlation between two experiments and their respective simulations. The dashed lines indicate the realized thresholds. For correlations above these thresholds, the model would hold true. The correlations would take the form of a 3-D Gaussian distribution which would result in most of the correlations being above the realized thresholds. ................................................................. 119

Figure 38: Using the realized threshold (red, dashed lines), the range of correlations (dotted, black lines) for which a 95% confidence interval can be established is shown in solid red. ................................................................. 121
Acknowledgements

I would like to thank my advisor, Dr. Jonathan Protz. He provided me with many helpful suggestions, important advice and constant encouragement during the course of this work. Over the last four years, he has assisted me in developing my thesis topic, aided me in the development of long derivations, offered sound advice for my future career, and simultaneously encouraged me to actively manage my research independently. I learned a great deal during my time at Duke and I am appreciative of the opportunity to work with him.

Similarly, I am very grateful to Dr. Thomas LaBean, Dr. Chris Dwyer, and Dr. Anne Lazarides for their guidance and direction. They were invaluable to me in troubleshooting, offering fresh insight, and pushing me to make my research better.

Without Kathy Parrish, who knows how many deadlines I would have missed or questions I would have never had answered. Thank you so much, Kathy! You are such an asset to the department!

I am indebted to my student colleagues and friends who offered daily assistance and entertainment. Graduate school would not have been a fraction of the fun without my dear friend, Jessica Sanders, who was always ready with a laugh or a second set of eyes to tackle a tricky problem. I could never have learned all the nuances of liposomes without Beth Vasievich – thanks for offering so much of your time to help me in lab!
Finally, and most importantly, I owe my deepest gratitude to my family. Their constant support, love, and prayers during my graduate career were greatly needed and deeply appreciated. I could not have done this without my parents who gave me every opportunity throughout my life, set excellent examples for me, and encouraged me to succeed. My mother, Lisa, offered unconditional love and support, much needed sanity checks, and was always available for proofreading. My father, Buck, offered advice and guidance in my work and was always excited to help. I have the best siblings in the world: Melissa, Melinda, Buck, and Margaret – they kept me motivated and inspired. Tom Slowe, my fiancé, empathized with my every worry and celebrated every success. I could not have done it without the love and support from each of you.

Thank you.
1. Introduction

This chapter details the motivation behind this research, the objectives and specific aims that were accomplished, an approach for the proposed mechanism of genetic memory, the specific scholarly contribution, the advancement of the state of the art, and the novelty of the research. It closes with potential applications.

1.1 Motivation

As nanotechnology continues to develop and progress, device and component sizes will continue to shrink and truly nano-sized products will be realized. In order to fully utilize these mechanical packages, new methods of sensing, recording, and storing information that are compatible on the nanoscale will need to be developed and refined.

Present approaches for the sensing and recording of environmental state variables on the nano- and sub-nano scale are limited by their reliance on electronics as a means of detecting and storing information. This restriction currently limits the practical size of such devices to the micro-scale rather than the nano-scale and imposes the requirement of providing electrical power to the devices by some means.

DNA is nature’s way of storing information. It is potentially the smallest physical way of storing a quantity of information. DNA is nanometers in size and the order of base pairs can contain vast amounts of data. One gram of DNA can store as much information as a trillion CDs [1]. In the last decade, using DNA to encode
messages directly has been studied extensively [2-6]. Researchers have proposed various methods of assigning specific sequences to individual letters or words to increase the reliability and ways of increasing the efficiency of data storage into DNA by varying the codon length. DNA has also been proposed as a method of long-term storage of information by programming a specific sequence into bacteria which can be subsequently extracted from future generations [7]. Certain species of bacteria were selected in this experiment due to their physical endurance and the researchers successfully demonstrated the concept. DNA offers many advantages over electronics for the storage of data with the quantity of information that can be stored, the potential for long-term storage, and the ability to move information undetected.

“Non-coding” DNA, or DNA that does not lead to the transcription of proteins, comprises the majority of an organism’s DNA. The role and usefulness of “junk” DNA has been debated and researched extensively over the last decades and many theories have been proposed including some stating that it plays a role in evolution, regulates the expression of certain genes, or is simply a spacer material [8-11]. These vast portions of DNA could offer a location for the storage of non-genetic data. In other words, “non-coding” DNA could be used to code non-genetic data. Once this data is encoded, it could be possible for the data to be passed onto an organism’s offspring.

Lamarckian evolution, or soft inheritance, was once a popular idea that said that an organism can pass characteristics acquired during its lifetime to its progeny. Many
researchers have attempted to demonstrate that Lamarckian evolution does exist, but generally all of these experiments have been discredited [12-15]. There has been some evidence that soft inheritance can occur in single-celled organisms but it has not been widely accepted [16]. While there appears to be no definite example of soft inheritance occurring biologically in nature, one could hypothesize that it could be engineered through the alteration of DNA in an organism using techniques described in this research.

1.2 Objective and Specific Aims

The objective of this research is the development of “genetic memory”, the storage of information into genetic material, through the in vitro demonstration of the feasibility and benefits of recording the time-history of one or more environmental state variables into genetic material. This concept could also be feasibly accomplished in vivo. First, the amount of information that could be stored into non-coding DNA using a lossy mechanism such as regulated diffusion is determined. Next, a mechanism through which this concept could be accomplished is proposed.

The specific aims include (1) developing a mathematical model to determine the amount of information that could be stored lossily into DNA, (2) creating a concept of sensing, recording, and storing information on the nanoscale, and (3) demonstrating the feasibility of such a method through experimental means.
1.3 Approach for Proposed Mechanism of Genetic Memory

In conjunction with DNA, which is the actual means of storing information, the proposed approach for the realization of genetic memory also uses thermosensitive liposomes (TSL) as a means of sensing state variables and acting as a valve to transport and release the DNA in response to the appropriate stimuli. Upon release, the DNA encounters ligase, ATP, and other cofactors and binds – preserving a record of the stimulus experienced by the liposomes. Thermally sensitive liposomes, through careful design, can be developed to have unique transition temperatures at which the permeability rate of the contents encapsulated by the liposome increases. It is only at or above this transition temperature that a liposome will become porous. Each variant of liposome contains a unique DNA sequence that serves as an identifier.

The solution contains two or more varieties of liposomes, with each variety corresponding to a unique transition temperature. Each liposome is loaded with a quantity of individual nucleotides, short DNA segments with overhanging “sticky ends”, or single stranded DNA. The DNA segments contained in each variety of liposome are unique to that variety of liposome and identical for all liposomes of that variant. The solution surrounding the liposomes contains the reagents required for assembly of nucleotides or DNA segments into a long DNA chain.

Given this design, the time history of the temperature experienced by the proposed mechanism is recorded by the device as follows: (1) when the temperature is
at its initial level, only those liposomes whose threshold temperature falls below the initial temperature will have open pores and the DNA stored in these liposomes will slowly diffuse into the immediate surroundings while the DNA stored in other liposomes is unable to diffuse; (2) the required ligase continuously assembles the available DNA segments into a long DNA chain, thereby recording information on the current temperature; (3) if the temperature is raised, new liposomes will open and their DNA sequences will also be incorporated into the chain as it is assembled, whereas if the temperature is lowered, the high-threshold liposome pores will close and the high-temperature segments will no longer be incorporated; (4) by sequencing the DNA strands constructed by the ligase over the course of the experiment, it will be possible to deduce the time history of the temperature experienced by the mixture from the statistics of the strand sequence. Multiple strands form, so the measured strands' sequence represents the statistical sequencing of an ensemble of strands.

The various aspects of this proposed mechanism are well understood and readily implemented including the process of fabricating liposomes with temperature-sensitivity or otherwise sensitive pores and the chemistry governing the assembly of sticky-end DNA segments or single strand complements into longer chains using T4 DNA ligase. The chemistry governing the assembly of individual DNA nucleotides into longer chains using TdTase has advanced substantially in recent years. The physics of the diffusion process that governs the transport of the DNA segments from the interior
of porous liposomes to the ligase is well understood. Finally, the tools needed to sequence the resulting segments are readily available and affordable.

**1.4 Specific Scholarly Contribution**

This research develops the concept of *in vitro* lossy storage of information in genetic material using regulated diffusion or a similar fully biological mechanism. To the author’s knowledge, the concept is novel and has not been previously demonstrated.

**1.5 Advancing the State of the Art**

In comparison to traditional data-storing micro- and nano-sensors, this approach makes no use of electronic systems and, therefore, requires no transistor or other electronic components whose individual size would limit the lower bound of the device size, requires no external electrical power source, and makes use of no easily-detected non-biological materials.

In comparison to existing means of DNA synthesis, this approach: (1) involves a chemical process that scales with volume rather than surface area; (2) uses inexpensive and readily-available precursor materials; (3) requires little or no expensive proprietary laboratory equipment; (4) requires little or no human intervention at any intermediate step of the DNA-writing process; and (5) should ultimately allow the high-volume fabrication of directly-written synthetic DNA at a substantially lower cost than existing approaches.
1.6 Novelty

It is the author’s understanding that to date there has been no demonstration of the sensing, recording, or storage of the time history of an environmental state variable into genetic material.

By using chemistry to perform the necessary computational step, this approach overcomes traditional limitations on sensing device size. Because a device utilizing this concept would be fabricated entirely of common biomaterials and should be incorporable into living microorganisms, this approach would eliminate the requirement to supply electrical power and enable sensor operation for indefinite durations.

Furthermore, the combination of liposomes containing encapsulated DNA and the use of different ligase has not been previously demonstrated or used to implement the concept of genetic memory.

1.7 Potential Applications

There are numerous potential applications for this research. The envisioned ultimate goal of this work is the bioengineering of a micro-organism that can function as a self-powered sensor by its being able to: (1) sense and store the time history or time series of one or more environmental state variables in genetic material for subsequent access and recall by sequencing or transcription; (2) transmit this stored information to its offspring; and, possibly, (3) act on this stored information, to include being able to edit the genetic material of its offspring on the basis of the stored information.
Release of DNA or any encapsulated materials from liposomes can be triggered by a variety of external stimuli. By incorporating additional types of liposomes, it will be straightforward to use the proposed approach to allow simultaneous sensing of temperature, light levels, pH levels, sound levels, and protein presence, and to record the levels of all of these sensed variables into the same DNA chain. The result would be a ‘multi-spectral’ sensor technology.

Additionally, vesosomes, large liposomes which can encapsulate other liposomes, could be used to create a biological container for the experiment that would replace test tubes used in other experiments to contain the reaction solution. This vesosome would have transition properties that ensure that its contents will not release during operation. The ligase are either in the encapsulated solution of the vesosome or lie in a separate small liposome with the lowest transition temperature ensuring that it will be released first during operation. The vesosome would also provide a potential mechanism for transfection via lipofection by entering cells where it can pass on its genetic memory, thus ensuring that while the “machinery” may die, the memory continues as a record inside infected cells.

Another potential application is a high-volume DNA copy machine which accomplishes its function by using the information encoded in the sequence of the original DNA strand to specify the level of one or more environmental variables such as pH, chemical concentration, or other such variables. These levels in turn regulate the
release of nucleotides from the reservoirs to yield multiple copies of a new DNA strand whose contents are a replica of the original strand.

The concept could also be used as a means of high-density archival data storage. This application would be accomplished by arraying onto a surface a single layer of reaction chambers, which would contain the mechanisms for chemical sensing and recording of environmental state information already described in this document. The reaction chambers might be vesosomes bound to the surface, voids or depressions in the material that constitutes the surface, or any other suitable structure. The relevant environmental variable could be temperature, light level, magnetic field orientation, or some other such convenient environmental state variable. This environmental state variable would be manipulated over time in order to write a time sequence of data at each point on the grid. The source of change of the environmental variable has several options such as a mobile write head which is maneuvered over the grid, the grid might be maneuvered under a static write head, or both the grid and the write head might be maneuvered as in a compact disc writer or DVD writer. The result could be, for example, a disc like a compact disc or DVD where each point illuminated by the writing laser can store one unit of information for each revolution of the disc during the write phase; resulting in a many-orders-of-magnitude increase in the storage density of the disc. The sequence of the starting length of DNA or other polymer at each grid point would be unique to that grid point. The individually written segments could be
assembled into a much longer polymer in an order prescribed by the address encoded in their starting lengths.

Chemical recording or genetic memory could be used as a means of recording the shipping and handling history of a consumer product. This application would be accomplished by coating the product with a biofilm, polymer film, or some other coating containing a mechanism for the sensing and chemical recording of the time history of one or more of the environmental state variables, to include geographical location, of the environments to which the foodstuff or consumer product was exposed.

Similarly, genetic memory could be used as a means of recording information on the past locations and experiences of an individual by applying to the individual or to the clothing of the individual a particle or particles, or a biofilm, polymer film, or some other object or coating containing a mechanism for the sensing and chemical recording of the time history of one or more of the environmental state variables, to include geographical location, of the environments to which the individual is exposed.

1.8 Summary of Chapters

The document is organized as follows. Chapter Two presents an extensive literature review on biomolecular computing, DNA, liposomes, and information theory. Chapter Three details the mathematical model that was developed to analyze the amount of information that can be stored lossily into DNA. Chapter Four summarizes a simulation used to model the mathematical model from the previous chapter. Chapter
Five describes the materials and methods used in experiments. Chapter Six outlines the initial results that were obtained from experiments designed to test specific steps in the process. Chapter Seven summarizes the results obtained from experimental work. Chapter Eight presents the conclusions.
2. Literature Review

The following sections outline topics relevant to determining how much information can be stored into DNA in a non-exact method and the methods that support the proposed approach. First, some principles of information theory will be described including the Shannon-Hartley Theorem and the transmission of information through a noisy channel. These concepts will later be applied to determine how much information can be stored into DNA in a non-coding manner. Biomolecular computing will be reviewed including a brief history of its development and progress and some of the primary materials and tools used. The use of liposomes as a means of regulating the release of an encapsulated species, including composition, methods of preparation, characterization, loading and release, and applications will be described. These methods will later be utilized in realizing a genetic memory concept.

2.1 Information Theory

Information theory is a branch of applied mathematics that involves the quantification of information. It was founded by Claude Shannon in 1948 with his seminal work “A Mathematical Theory of Communication” and has been developed to find the fundamental limits on compressing and reliably communicating data [17]. Its key measure is entropy expressed by the average number of bits needed for storage and communication [18-21]. Information theory does not consider the importance or
meaning of the message, or the quality of data, only the quantity and reliability of the message which can be determined solely by probabilities. The central paradigm of classical information theory is the engineering problem of the transmission of information over a noisy channel.

### 2.1.1 Lossy Data Conversion

Converting data between storage formats can result in data that is not exact to the original but similar enough to still be useful. This process is known as lossy data conversion.

Lossy data compression is the process of compressing data in such a way that when decompressed, the data that may be different from the original but close enough to be useful in some way. It is commonly used to compress multimedia data such as jpegs. The method suffers from generation loss: repeatedly compressing and decompressing the file will cause it to progressively lose quality. The information-theoretical foundations of lossy data compression are provided for by the rate-distortion theory.

### 2.1.2 Noisy Channel Coding Theorem

In a noiseless channel, the message sent is the same as the received signal and the message received can be used to exactly replicate the sent message. In a noisy channel, the message sent is most likely different from the received message due to errors gained
from noise during the transmission process. The received message cannot be used to exactly replicate the sent message, but information is still transmitted. The received signal is a function of the transmitted signal and noise [17]. One method of decreasing the probability of errors is to repeat the message several times but this greatly reduces the efficiency and reduces the rate of transmission.

Shannon’s noisy channel coding theorem states that reliable communication is possible over noisy channels provided that the rate of communication is below a certain threshold called the channel capacity [17, 18]. Channel capacity is the maximum possible rate of transmission. In other words, if the rate of communication is less than the channel capacity, the output of the source can be transmitted over the channel with a small frequency of errors; if the rate of communication is larger than the channel capacity, the output of the source suffers from increasing errors as the rate increases.

The Shannon-Hartley Theorem describes the channel capacity $C$, or the amount of data that can be transmitted through a noisy channel, given a signal power $S$, a specified bandwidth $B$, and noise power $N$.

$$C = B \log_2 \left( 1 + \frac{S}{N} \right)$$

This equation is valid for a channel with white noise and calculates a transmission rate as close as desired to the channel capacity [17, 19]. By increasing the bandwidth or the signal-to-noise ratio, the information rate also increases. As a result, a
trade-off between the bandwidth and signal-to-noise ratio can be used to achieve the same channel capacity.

The generalization of this result for a channel with colored noise is

\[
C = \frac{1}{2\pi} \int_{-\infty}^{\infty} \log_2 \left(1 + SNR \right) \cdot d\omega
\]

where \(SNR\) denotes the signal-to-noise ratio.

### 2.2 Biomolecular Computing

In recent history, computer engineers have worked to incorporate lessons from biology and nature into computers such as evolutionary programming, self-healing, or the ability to learn in an effort to improve their computational abilities [22]. In addition to taking lessons from nature to advance computers, computers can also be biologically based. Using molecules to accomplish computational tasks is known as biomolecular computing. The primary approaches to biomolecular computing are (1) developing molecular devices that replicate existing components of conventional computing devices and (2) developing new methods of computing that utilize the specific characteristics of molecules [5, 6].

Biomolecular computing originated with the Turing machine proposed in 1936 which is composed of three basic principles: (1) store information into a string of symbols from a fixed alphabet, (2) process these symbols in a stepwise manner, and (3) modify or add symbols according to fixed rules. It was these principles that became the
basis of electronic computers. The Turing machine, however, could just as readily be realized through biochemical means such as nucleic acids, enzymes, and ligase. In fact, biomolecular computing offers several specific advantages over electrical computing such as the ability to interface with cells, operate in a biological environment, and greater energy efficiency.

In 1994, Adleman ushered in a new era of biomolecular computing by demonstrating that it was possible to solve computational tasks using DNA [1, 23]. The following year, Lipton published a paper describing the theoretical basis for DNA computing [24]. Adleman utilized unique segments of DNA to determine the correct solution to a Hamiltonian Path Problem. Given a graph of vertices, the problem asks whether a path which visits each vertex exactly once can be found. This particular problem has been found to be computationally difficult for traditional computers and generally requires an exhaustive search of all possible solutions. Biomolecular computing is ideally suited to the task due to the possibility of massive parallelization.

Parallelization is one of the primary advantages of biomolecular computing. Some of the drawbacks to molecular computing are that as the complexity of the problem being solved increases, the number of operations required rapidly becomes cumbersome and prone to errors and the volume of DNA grows exponentially. Biomolecular computing, therefore, will probably not replace conventional computing but supplement it by offering new methods for solving certain problems [5].
Biomolecular computing continues to advance. DNA logic gates [25], DNA computing on surfaces [26-28], the use of RNA in biomolecular computing [26, 29], more complicated computing problems [30], attempts at automaton [31-33], and formal DNA computers [34] are being developed.

The following sections detail the primary materials and methods employed in biomolecular computing.

2.2.1 Materials of Biomolecular Computing

The most commonly used material in biomolecular computing is deoxyribonucleic acid which encodes the genetic information of cells. The four subunits or nucleotides are adenine (A), cytosine (C), guanine (G), and thymine (T). Adenine and guanine are purines and cytosine and thymine are pyrimidines. Purines are heterocyclic rings composed of carbon and nitrogen. Pyrimidines are heterocyclic rings and purines are heterocyclic rings combined with an imidazole ring. Nucleotides bind covalently to form single strand DNA. These strands have a 5’ end from the free 5’ phosphate group and 3’ end from the free deoxyribose hydroxyl group.

In Watson-Crick base pairing, adenine (A) binds to thymine (T) using two hydrogen bonds and cytosine (C) binds to guanine (G) with three hydrogen bonds. DNA strands with higher concentrations of C-G bonding tend to be more stable and also have higher melting temperatures. Through hybridization, single strands of DNA form
a double helix with a sugar phosphate backbone. DNA is approximately 2 nm in
diameter and 3-4 nm in length of a full turn which is between 11 to 12 base pairs.

2.2.2 Processes of Biomolecular Computing

Several processes make up the core of biomolecular computing. These tools
allow the many manipulations that biomolecular computing require such as binding
together strands, increasing the length, cutting at a specific sequence or increasing the
quantity of DNA present. Denaturing is a process in which double stranded DNA is
heated past the melting temperature of the strands which results in the DNA separating
into two strands of single strand DNA. Annealing is the reverse of denaturing and is the
process whereby single strands of DNA bind together complementarily through the
slow decrease in temperature. Additional tools such as ligation, various useful
enzymes, gel electrophoresis, polymerase chain reaction, and DNA sequencing are
detailed below.

2.2.2.1 Ligation

Ligase are used to bind strands of DNA together under specific conditions either
by closing nicks such as those that occur when sticky ends bind or through blunt end
ligation. A DNA sticky end is a single strand overhang that extends from a double
strand of DNA, while blunt end ligation is the joining of two DNA sequences with no
overhands (Figure 1).
Figure 1: An example of a (a) DNA sticky end and a (b) DNA blunt end.

The following sections detail the necessary ligase for each of these types of ligation.

### 2.2.2.1.1 T4 DNA Ligase

T4 DNA ligase is obtained from infecting *E. Coli* with bacteriophage T4 and is used to bind the 3’ hydroxyl to the 5’ phosphoryl end in three reaction steps [35]. The first step is the activation of the enzyme through the presence of ATP which results in the formation of ligase-adenylate and pyrophosphate (PPi). Next, a pyrophosphate linkage is made between the 5’ end of the DNA to be joined and the adenosine monophosphate (AMP) by the ligase transferring the adenylate group to the DNA. Finally, a phosphodiester bond is formed between the 5’ phosphate end and the 3’ hydroxyl end and the AMP is removed [35, 36].

The reaction rate of T4 DNA ligase depends on the concentration of ATP, Mg$^{2+}$, pH, and temperature. The presence of ATP in concentrations that are either too low or too high can inhibit both blunt end ligation and nick sealing ligation. Below 5 μM of ATP, ligation is reduced; at 5 mM of ATP in solution, the blunt end ligation is inhibited and increasing the concentration to 7.5 mM of ATP will inhibit the nick sealing ligation [37, 38]. Research has also shown that the presence of Mg$^{2+}$ aids in DNA ligation and
that the optimal concentration is 10 mM [38, 39]. Therefore, for ligation to occur, DNA fragments, T4 DNA ligase, ATP, Mg$^{2+}$, and water must be present in the solution.

Under the same conditions, T4 DNA ligase is twice as likely to participate in binding sticky ends, which requires nick sealing, than blunt end ligation [37]. However, with the addition of RNA ligase, the turnover number for the blunt end ligation increases significantly while not improving nick sealing performance, which brings the binding of sticky ends and blunt end ligation to be approximately equivalent [40].

<table>
<thead>
<tr>
<th>2.2.2.1.2 Terminal Deoxynucleotidyl Transferase</th>
</tr>
</thead>
</table>
| Terminal deoxynucleotidyl transferase (TdTase) is a template independent polymerase that adds mononucleotides to the 3’ end of DNA segments [41-44]. It was first isolated from the calf thymus gland in 1966 [43]. It is a single polypeptide with a molecular weight between 44 – 60 kDa [41, 44]. Single stranded DNA is its preferred initiator. Depending on the cofactor present in solution, TdTase adds nucleotides preferentially. For example, in the presence of magnesium, TdTase adds the nucleotides so that G and A are added at a greater rate than C and T. Conversely, in the presence of cobalt, TdTase adds T and C at a greater rate than A and G [41]. As a primer, TdTase requires at least a trinucleotide with a free 3’-hydroxyl and 5’-phosphate [41, 43].

Recently, TdTase has been used to create nanostructures as part of a tool set in enzymatic fabrication [42]. The height of these structures was proportional on the lateral feature size of the substrate features.
2.2.2.2 Enzymes

DNA polymerases are involved in the repair and duplication of DNA. These enzymes add single, complementary nucleotides to the 3’ end, using a primer as an initiator and a template as a guide.

Restriction enzymes recognize a specific sequence of DNA known as a restriction site. Any double strand of DNA that contains the restriction site within its sequence is cut by the enzyme at that point. The resulting cleavage generates either a sticky end or blunt end.

2.2.2.3 Gel Electrophoresis

Gel electrophoresis is a sorting tool; it separates molecules by size. The rate of migration of a molecule in an electric field depends on its shape and charge. Since DNA molecules carry a negative charge, when placed in an electric field they tend to migrate to the positive pole. Gels are frequently made of agarose or polyacrylamide which acts as a resistive medium so that the larger the molecule or the longer the strand, the slower it moves through the gel. In order to visualize the result, a stain is applied – typically ethidium bromide – and the gel is viewed under ultraviolet light. The brightness of a band depends on the amount of DNA of a certain length present in the sample. Larger concentrations of DNA absorb more dye and therefore appear brighter.
2.2.2.4 Polymerase Chain Reaction

Polymerase chain reaction (PCR) is used to quickly amplify the amount of DNA in a given solution. A cycle of PCR consists of (1) heating the solution to denature the double stranded DNA, (2) adding primers which are the complementary to an area of the single strand and delineate the region of duplication, (3) and adding polymerase to the reaction. The polymerase will successively add complementary nucleotides to the single strands to make double stranded DNA. Each cycle of the reaction doubles the quantity of each strand giving an exponential growth in the number of strands.

2.2.2.5 DNA Sequencing

DNA sequencing is the process by which the order of nucleotides in a DNA strand is determined. The chain termination method of DNA sequencing is similar to PCR. It requires a DNA template, a primer, DNA polymerase, nucleotides, and buffer. The nucleotides are a mixture of deoxynucleotides (dNTPs) and dideoxynucleotides (ddNTPs). ddNTPs are missing the 3’ hydroxyl group and once added to the strand, the chain is terminated. The DNA is typically separated into four sequencing reactions, in which all four dNTPs are present but only one ddNTP. At the end of the reaction, there are DNA fragments of varying length. The fragments are then denatured and separated by size using gel electrophoresis – one lane for each of the four sequencing reactions. A band in a given lane corresponds to a reaction that terminated at that particular position. By reading the bands and using their relative positions, the DNA sequence is found.
2.3 Liposomes as a Means of Regulation

In the proposed approach for genetic memory, liposomes are a critical component. A liposome is a lipid bilayer that encapsulates an aqueous core. The protective coating of every living cell is a lipid bilayer, which has many functions including transport across the membrane, recognition, and signaling. Additionally, lipid membranes are found in eukaryotic cells in the form of organelles, which perform the majority of tasks within the cells including energy production, protein translation and folding, storage, and DNA maintenance. Lipids self-assemble in an aqueous environment into vesicles termed liposomes and were first described in 1965 [45, 46]. This formation is depicted in Figure 2.

Figure 2. Liposomes spontaneously form from dry lipid films that are exposed to an aqueous environment. Picture from Avantilipids [47].
Liposomes fall into three general categories: multilamellar vesicles (MLV), small unilamellar vesicles (SUV) under 1000 Å in diameter, and large unilamellar vesicles (LUV) [45]. Liposomes are inherently temperature sensitive – below a certain transition temperature they are solid, while above the transition temperature they are porous. This transition temperature is dependent on the acyl chain length – a longer chain generally means a higher transition temperature [48]. “Custom” transition temperatures can be achieved by mixing lipids with different transition temperatures.

Liposomes can be utilized to carry hydrophobic and hydrophilic molecules. Hydrophobic molecules can be stored in the bilayer and hydrophilic molecules can be stored in the aqueous core. This ability allows a wide range of materials to be transported within liposomes including chemotherapeutic agents, MRI contrast agents, or in this current application, DNA.

The liposomal delivery system requires the production of uniform, stable liposomes that are loaded with a specific payload which can be released at the appropriate stimulus. Many obstacles can arise and impede the liposomal delivery system at any of these steps. Scientific reproducibility and mass production in industry demand a standardized product; therefore the size of the liposomes must be uniform between batches. This requirement necessitates the characterization of liposomes and methods for achieving a homogeneous product. Drugs and ions can present barriers to loading high concentrations of drugs into the aqueous core of the liposome. The
stability of liposomes, *in vivo* and *in vitro*, must be ensured and qualified, and if inadequate, the composition of the liposome or protocol for loading needs to be altered. Finally, the release of the encapsulated solution must be achieved consistently and reliably.

The specificity of release is well characterized. Release can be triggered by an external stimulus, such as light [49], pH of the microenvironment [50-52], ultrasound [53, 54], enzymes [55, 56], or temperature [57-59]. Active targeting includes labeling the outside of the liposome with any receptor over-expressed in the delivery tissue of interest, such as folate receptors, integrin surface receptors, or with the use of antibodies [60-68].

Temperature-sensitive liposomes (TSL), the focus of this research, have the ability to release its contents quickly at specific temperatures based on the composition of the liposome.

**2.3.1 Composition**

Liposomes can be composed of a wide range of lipids or mixtures of lipids, but are frequently composed of phospholipids which are a major component of biological membranes. Phospholipids are amphipathic: they have a polar head containing glycerol and a phosphate group and two nonpolar fatty acid tails as shown in Figure 3 [69]. Because of this property, phospholipids form lipid bilayers – the polar heads and non-polar heads align which results in hydrophilic and hydrophobic regions. For
liposomes, this means that the exterior and enclosed interior spaces are hydrophilic, while the center of the lipid bilayer is hydrophobic.

Figure 3: Phospholipids are composed of a polar head and two fatty acid, non-polar tails (left) and frequently form lipid bilayers (right).

Thermally sensitive liposomes are often composed of several different lipids in order to give the liposome different properties. Most liposomes use a neutrally charged 16-carbon phospholipid, 1,2-Dipalmitoyl-sn-Glycero-3-Phosphocholine (DPPC) as the majority of the membrane [49]. DPPC has a transition temperature between 41.0 and 41.5°C [70-72]. To change the transition temperature, other lipids with different transition temperatures are included. The simplest changes come from including other phosphatidylcholines of different carbon chain length such as 1,2-Dimyristoyl-sn-Glycero-3-Phosphocholine (DMPC), a 14-carbon phospholipid with a transition temperature of 34°C, and 1,2-Dimyristoyl-sn-Glycero-3-Phosphocholine (DSPC), an 18-carbon phospholipid with a transition temperature of 54°C. By mixing the lipids in
different molar amounts, the transition temperature can be shifted between the main transition temperatures of each of the two lipids alone.

The inclusion of lysolipid in the composition has been shown to dramatically increase the peak of the transition temperature and release rate [73]. Lysolipids have only one fatty acid chain and therefore can create micelles with a higher curvature and lower concentration required. The critical micelle content for the lysolipid used is about 3 μM, above which, the lysolipid will separate from the bilayer and create micelles. Two different lysolipids that are commonly used are: 1-Palmitoyl-2-Hydroxy-\textit{sn}-Glycero-3-Phosphocholine (MPPC), a 16-carbon unsaturated fatty acid and 1-Stearoyl-2-Hydroxy-\textit{sn}-Glycero-3-Phosphocholine (MSPC), an 18-carbon unsaturated fatty acid. Both of these lysolipids also lower the transition temperature when included with DPPC [74].

Cholesterol is generally not included in the formulation of thermosensitive liposomes because at cholesterol concentrations greater than 25-30 mole percent, the enthalpy required for the transition to take place is reduced or even eliminated such that no transition or increase in permeability is observed [45, 75, 76]. An effort was made to optimize the amount of cholesterol included in most conventional liposomes because of the benefit of additional membrane stability [77-83]. This benefit is particularly appealing in drug delivery applications because cholesterol can decrease the permeability of the bilayer to small polar molecules and prevent proteins from penetrating [45].
2.3.2 Methods of Preparation

There are several methods for producing liposomes, including the reverse-phase evaporation method [48, 70, 71] or thin film lipid extrusion method [45, 84, 85]. These methods produce MLV liposomes which through the use of sonication or extrusion can be modified into SLV liposomes.

2.3.2.1 Reverse-Phase Evaporation

Reverse-phase evaporation requires a mixture of phospholipid, cholesterol, organic solvent, and aqueous buffer. Sonication can be used at this point to produce a homogenous emulsion. The organic solvent is removed using a rotary evaporator under reduced pressure and partial vacuum to produce a gel-like phase which will subsequently and spontaneously form liposomes in aqueous solution. Filtration can then be used to remove any excess solvent or unencapsulated material. The primary advantages of this method is that it produces liposomes with a large interior aqueous space, provides a significantly higher encapsulation efficiency, and enables the encapsulation of larger macromolecules than other methods of preparation [45, 48]. The method is not particularly well suited to thermally sensitive liposomes because of the high concentration of cholesterol required. It is the inclusion of cholesterol that produces the increased encapsulated volume, but it is also this inclusion that leads to liposomes that are exceedingly stable and no longer retain temperature sensitivity [45, 48].
2.3.2.2 Thin Film Method

The thin film method is detailed as follows. Lipids are dissolved and stocked in chloroform in free lipid form. The desired mixture of lipids is added directly to a clean round-bottom flask and dried to a thin film using a Rotoevaporator with recommended manufacturer settings. The flask is covered with aluminum foil and placed in a desiccator to which a vacuum is pulled for 30 minutes. The desiccator is subsequently sealed off and the vacuum removed. The thin film sits for 6 hours or overnight in the desiccator which ensures that the film is completely dry.

Next, the lipid film is hydrated with an aqueous solution that forms multilamellar liposomes between 100-800 nm [86]. The water bath to be used for hydration should be turned to ~9°C above the transition temperature or above the transition temperature of the highest melting component added to the mixture. The round bottom flask is taken out of the desiccator and the DNA buffer solution is added to achieve the desired concentration of lipid. A clean Teflon stir bar is gently swirled in the flask at the appropriate temperature bath for 10-15 minutes to hydrate the lipids. Hydration time has been shown to be important to the amount of encapsulation achieved – the longer the hydration time, the more encapsulated material and the larger the liposomes produced [45].

Liposome extrusion is essential to getting unilamellar, same-sized particles, and a crucial control to any further liposome study. Sonication can also be used alone to make
similar sized liposomes or become a step before extrusion to reduce particle size slightly and make extrusion easier [87].

The extruder is assembled with a drain disk and two polycarbonate filters. The size of these filters will determine the size of liposome yielded. For example, using two polycarbonate 100nm filters are necessary to achieve liposomes of ~100 nm. A recirculating water bath is attached to the thermobarrel of the extruder to keep the temperature of the liposomes above the transition temperature during extrusion. If the temperature falls to the transition temperature or below, the liposomes will clog the filters of the extruder. First, the buffer solution used to hydrate the thin film is used to wet the filters; then the hydrated lipid solution is extruded at least 5 times through the thermobarrel using nitrogen pressure, regulated at 150-200 psig. After the last extrusion, the liposomes are stored at 4 ºC for at least 10-15 minutes to ensure the liposomes anneal into the gel phase.

**2.3.3 Characterization**

Depending on the application, certain characteristics of liposomes need to be measured or defined. These characteristics include size, charge, transition temperature, permeability, and stability.

The determination of the liposomes’ size distribution is important in biological applications. In particular, ensuring that the size is standardized between batches of liposomes is important for reproducing results and commercial products. When
lipoisomes are formed, there is typically a heterogeneous mixture of sizes. Using extrusion and gel filtration or sonication, a homogeneous mixture can be produced. For a homogenous mixture of liposomes, light scattering techniques, analytical ultracentrifugation, and NMR spectroscopy are methods of determining the size [45].

Determining the phase transition temperature is essential to characterizing thermally sensitive liposomes so that the temperature at which the encapsulated material will be released is known. The phase transition of thermosensitive liposomes is most commonly measured with differential scanning calorimetry (DSC) [86]. In DSC, the sample of interest and a reference, which is typically water or buffer, are slowly heated and the temperature and energy required to increase the temperature are recorded. Throughout this heating, the sample is at the same temperature as the reference, but may require more or less heat than the reference to maintain this temperature. Phase transitions can be endothermic such as melting from a solid to a liquid or exothermic such as crystallization. When a sample is undergoing a phase transition, the heat required by the sample to keep it at the same temperature as the reference is different than the heat used by the reference to maintain the temperature. It is this difference that indicates that the sample is undergoing a phase transition.

In 1973, it was discovered that phospholipid bilayers heated through their phase transitions had increased permeability to small molecules at that temperature [88]. Subsequently, assays were developed to measure the permeability of liposomes to ions
as a function of temperature [86]. Liposomes are prepared with 1.0 mol% NBD-PE, which absorbs ultraviolet light. The NBD labeled formulation and unlabeled formulation are equilibrated at temperature in a UV spectrometer and dithionite is added. The permeability is recorded as the outside of the liposomes is quenched first from the NBD on the outer membrane, and the permeability through the membrane is directly proportional to the absorbance quenching. From an exponential model fit of the internal quenching, the permeability rate constant can be found.

2.3.4 Loading and Release

Loading the desired encapsulation species into the liposomes can be achieved in numerous ways, either actively, passively, or by utilizing equilibrium on the inside and outside of the liposome. Active loading uses a pH gradient between the inside and the outside of the liposome to draw amphipathic drugs into the aqueous core of the liposome. This method has been used to maximize the loading of chemotherapeutic drugs, such as Doxorubicin (Dox), into the liposome with a citrate buffer [57, 89]. By using a buffer with low pH, a pH gradient is created as the external solution is taken away and replaced with a buffer near pH 7 [72, 79]. When the Dox is added to the solution, it is pulled into the liposome and protonated, which prevents the drug from escaping prior to triggered release.

Passive loading is achieved by hydrating the dried lipid thin film with the solution to be encapsulated. As the lipids swell in the solution, the dissolved solute to
be encapsulated is surrounded by lipid and becomes part of the aqueous interior. The amount of encapsulation attained using this method can be as low as 10-20% of the total solute [45]. After extrusion, the external solution is removed using gel filtration columns, and the desired solution is sequestered in the liposomes.

Equilibrium loading uses preformed liposomes heated to their transition temperature in a bathing solution that is to be encapsulated. The increased permeability at this temperature allows the interior and exterior solutions to reach equilibrium. The liposomes are cooled below their transition temperature, the elevated permeability no longer exists and the external solution can be removed by gel filtration leaving the intact liposomes with the solution of interest encapsulated.

The release of thermosensitive liposomes occurs at the transition temperature due to leaky interfacial membranes [58, 90]. Below the transition temperature, the liposome forms a stable lamellar crystalline Lα state; as the temperature is increased, the liposome transitions to a lamellar gel or ordered bilayer Lβ' state [91]. In lysolipid-containing thermosensitive liposomes (LTSL), the individual gel phase plates are separated by grain boundaries that occur in the Lβ phase of phospholipid membranes when the temperature is decreased from above to below the transition temperature [92]. The lysolipid enhances the permeable defect structures, either by desorbing rapidly from boundary regions or by stabilizing porous defects [86]. In LTSL, phase transition
has been recorded to increase by 20 fold compared to DPPC liposomes [57]. This increase is also seen in Figure 4.

![Permeability rate (1/min) vs. Temperature](image)

**Figure 4: Permeability rate (1/min) vs. Temperature [86].**

In addition to demonstrating that LTSL have a greater permeability compared to liposomes composed with DPPC only, this figure shows that immediately following the transition temperature for the liposome, the permeability decreases [86]. For DPPC, this decrease is maintained until the temperature is raised about 45°C, at which point, the permeability again increases. In the LTSL compositions shown in the figure, the permeability peaks at the transition temperature, dramatically decreases at temperatures slightly above the transition temperature and then does not regain the same permeability until temperatures significantly higher than the transition temperature.
2.3.5 Applications

For the past 40 years, liposomes have been researched extensively as drug delivery devices, encapsulating a solution of interest with the goal of delivering a therapeutic payload *in vivo* [57, 58, 89, 93, 94]. These liposomes are delivered to the tumor site which is subsequently heated slightly above the body’s temperature. This temperature increase results in the liposomes releasing their encapsulated drug payloads and delivering concentrated levels of drug directly to the tumor site. This method of treatment is particularly advantageous in cancer treatments by delivering concentrated drug locally, which prevents systematic exposure of the entire body to the drug and allows for a higher concentration of the drug to be used than would otherwise be feasible. Another delivery mechanism utilizes the fact that tumors have a lower pH than the ambient environment. pH sensitive liposomes have also been tested as vehicles of drug delivery [52].

Currently marketed liposomes lack active targeting and rely on passive diffusion and nonspecific degradation of the liposome [55]. As a result, active targeting and triggered release have been pursued to maximize payload delivery to the area of interest.
3. Mathematical Model

One of the primary objectives of this research is to develop a mathematical model to determine the amount of information that can be stored in a lossy manner into a DNA strand. The mechanism selected to realize this concept of genetic memory is the use of variants of thermally sensitive liposomes to encapsulate unique sequences of DNA. When a liposome reaches its transition temperature, it becomes porous and releases its encapsulated payload. These released DNA sequences diffuse across the lipid bilayer and into the solution. As the DNA sequences bind with the starter strand through ligation, the strand contains information about the temperatures experienced and a relative time scale. Figure 5 shows the block diagram of the dynamics of the system from the change in temperature to the resulting DNA sequence.

Figure 5: Block diagram of the proposed mechanism kinetics from the change in temperature to the resulting DNA sequence.

Depending on the temperature, the liposome permeability will change as a result of the transformation in porosity experienced by the lipid structure. The change in porosity results in the release of the DNA or the diffusion of the DNA across the lipid bilayer and into the solution. As this diffusion occurs, the concentration of that
particular species of DNA changes in the solution which further alters the reaction kinetics and impacts the expected arrival rate of a given sequence of DNA at the end of the recording strand. Given the expected arrival rate of a particular sequence of DNA at the end of the strand, the monomer incorporation, which species is added to the strand, will vary. It is at this step that the noise of the system is introduced. At the conclusion of the process, the output is the DNA sequencer data from which a record of the temperature is preserved.

Once information is stored into DNA through the type or order of base pairs added to a strand, the data can be recalled or retrieved through DNA sequencing. In order to sequence the information, a sequencing primer must be used which requires knowledge about the tail of the DNA created.

From the sequencer, the order of arrival to the chain is known, but not the order from which it left the liposomes. Also, the average length of the subchains is given. A sliding window of varying width can be used to analyze the DNA. The width of the window is specified in space but not in time. In other words, the location of a nucleotide within the strand is known but not the time that it was added to the strand. The link is deduced through the length of the strand. Despite the fact that an exact replica of time cannot be made in space, information can be stored into DNA.

The following sections will detail each of these mechanisms further and their contribution to the mathematical model.
3.1 Liposome Permeability

The liposome permeability given a specific temperature is determined through experiment and characterized using differential scanning calorimetry.

Given that the proposed mechanism of genetic memory uses thermally sensitive liposomes releasing DNA which then diffuses through the solution while binding, the order of arrival in the chain is not necessarily the order of departure from the liposome. Each of the varieties of thermally sensitive liposomes will release at a specific transition temperature. Above the transition temperature, there are two possibilities for what occurs: the permeability of the liposomes can decrease if the composition contains lysolipid or the liposomes can remain porous if their composition does not contain lysolipid. Either of these possibilities makes it possible to record information about the temperature into DNA. Figure 6 demonstrates what the potential DNA strands could look like with either of these options.
Figure 6: Two potential data strands given the liposomes’ permeability at temperatures higher than its transition temperature. For both strands, a sample graph of the temperature experienced is displayed. (left) Given a temperature, the corresponding liposome releases the assigned sequence as well as all thermally sensitive liposomes with transition temperatures less than the given temperature in a mix of units. (right) Given a temperature, only the corresponding liposome releases the assigned sequence.

One option, demonstrated on the right of Figure 6, is that at temperatures higher than the specific temperature, the permeability of the thermally sensitive liposome declines and the contents would stop being released. This option would mean that the liposome would only release DNA at its transition temperature. However, at temperatures much higher than the transition temperature, the liposome becomes highly permeable, effectively releasing its entire contents. These temperature profiles have been characterized for many liposome compositions and would have to be carefully considered during the data analysis phase [73]. For example, at a specific temperature, the assigned thermally sensitive liposome would release. Liposomes that have transition temperatures lower than the current temperature would stop releasing, but thermally sensitive liposomes with transition temperatures that are much lower than the current temperature would dump the encapsulated load.
Another option, demonstrated on the left of Figure 6, is that at temperatures higher than the specific temperature, the liposome is still porous and releasing its encapsulated contents. This option would mean that the liposome with transition temperatures at or below the current temperature would be releasing DNA.

Under either option, the units released at each of the temperatures will bind randomly. The data analysis has to be designed to take into account the multiple units released by greater temperatures. One simplifying condition is that in the solution there will be a starting sequence or primer onto which all subsequent units will bind. This starting sequence would be designed in such a way as to allow binding in only one direction, thus preventing the units from binding randomly on either side which would greatly complicate the data analysis.

There is the possibility that “subchains” would form. If chemical units came into contact with each other before reaching the main chain, it would be possible for them to bind, forming a subchain. However, even with the presence of subchains, the chains could be evaluated in a similar manner as described because there should be a probabilistic distribution of lengths and unit content.

### 3.2 Diffusion

When a temperature is experienced that changes the liposome’s permeability, the bilayer becomes porous, allowing the sequences to diffuse across the bilayer and into the
reaction solution. The diffusion model described here considers the dynamics of the sequences diffusing across the lipid bilayer which is modeled as a valve.

To determine the chemical dynamics and resulting transfer function, a lumped sum approximation was made. This approximation is justified because the Biot number, $Bi$, is significantly less than 1. In Figure 7, the aqueous space of the liposome in which the encapsulated DNA is held is shown as the holding reservoir, the lipid bilayer is denoted as the valve, and the outside solution where the reaction would take place is labeled as the chamber. The three points of interest are the start of the valve, the end of the valve, and the chamber.

The mass flow rates, $m_i$, between the different points are

$$
m_{1-2} = \frac{MD_{1-2} A_{1-2}}{L_{1-2}} (C_1 - C_2)
$$

$$
m_{2-3} = m_{1-2}
$$

$$
m_{3-\infty} = -k[C_3 - 0]MV_3 + \dot{m}_{2-3} = \frac{\partial C_1}{\partial t} MV_3
$$
where $M$ is the molecular weight, $D_{AB}$ is the diffusion coefficient of DNA in water, $A_{1-2}$ is the cross-sectional area of the valve, $L_{1-2}$ is the length of the valve, $V_3$ is the volume of the mixing chamber, $C_1$, $C_2$, and $C_3$ are the concentrations at the start and end of the valve and the mixing chamber respectively, and $k$ is the reaction rate. Assumptions about the concentrations are made. $C_1$ is considered a constant and $C_2 = C_3$. Applying these assumptions and combining the mass flow rates

$$\frac{\partial C_3}{\partial t} V_3 M = -k C_3 M V_3 + \frac{MD_{AB} A_{1-2}}{L_{1-2}} (C_1 - C_2)$$

$$\frac{\partial C_3}{\partial t} = -k C_3 + \frac{D_{AB} A_{1-2}}{V_3 L_{1-2}} (C_1 - C_3)$$

results in the steady background which needs to be linearized for perturbations

$$\frac{\partial C_3'}{\partial t} = -k C_3' + \frac{D_{AB}}{V_3 L_{1-2}} (A_{1-2}' C_1' - \bar{A}_{1-2} C_3') .$$

Rearranging this equation and taking the Laplace Transform yields

$$\frac{C_3'}{\bar{C}_1} s + \left( k + \frac{D_{AB} \bar{A}_{1-2}}{V_3 L_{1-2}} \right) \frac{C_3'}{\bar{C}_1} = \frac{D_{AB} A_{1-2}'}{V_3 L_{1-2} A_{ref}}$$

which can be further rearranged to give

$$\frac{C_3' / \bar{C}_1}{A_{1-2}' / A_{ref}} = \frac{D_{AB} A_{ref}}{V_3 L_{1-2} A_{ref}} .$$

(32)
This equation is the ratio of the percent change of concentration to cross-sectional area and the transfer function for the chemical dynamics. Bode plots of the transfer function were generated (Figure 8).

\[
K = \frac{D_{AB}A_{ref}}{kV_3L_{i-2} + D_{AB}A_{i-2}}
\]

Figure 8: Bode plots for the chemical dynamics: (top) amplitude and (bottom) phase.

The gain for the chemical dynamics is
and the bandwidth is

\[
\frac{1}{\tau} = \frac{kV_3L_{t-2} + D_{4b} \bar{A}_{t-2}}{V_3L_{t-2}}.
\]

The first approximation of the model neglected the valve dynamics of the device. A more accurate model takes into account these dynamics. A more detailed block diagram of this particular step is depicted in Figure 9. The mechanical dynamics with transfer function, \( G_0 \), have an input of the temperature and output of the valve cross-sectional area. The valve dynamics require two transfer functions, \( G_1 \) and \( G_2 \), with inputs of the valve cross-sectional area and concentration of the reaction solution, which is obtained as feedback, and output of the mass flow rate out of the valve into the solution. The chemical dynamics use this flow rate as the input and output the solution concentration for transfer function, \( G_3 \).

![Figure 9: Modified block diagram of diffusion dynamics that takes into account the time lag due to the sequences diffusing across the lipid bilayer.](image)

The three DNA concentration points of interest are defined differently from the first approximation and are roughly depicted in Figure 10. \( C_1 \) is defined as the concentration in the liposome and assumed constant; \( C_2 \) is the concentration at the midpoint of the lipid bilayer and \( C_3 \) is the concentration at the “entrance” to the solution.
Figure 10: Three DNA concentration points of interest in the device when determining the valve dynamics are the concentration inside the liposome $C_1$, the concentration at the midpoint of the lipid bilayer $C_2$, and the concentration at the “entrance” to the solution $C_3$.

The mass flow into the valve or lipid bilayer, $\dot{m}_{in}$, where $D_{AB}$ is the diffusion coefficient of DNA in water, $A_{1-2}$ is the cross-sectional area of the valve, $L_{1-2}$ is the length of the valve, is

$$\dot{m}_{in} = (C_1 - C_2) \frac{A_{1-2}}{L_{1-2}/2} D_{AB} = (C_1 - (C_2 + C_2')) \frac{(A_{1-2} + A_{1-2}')}{L_{1-2}/2} D_{AB}$$

which has also been linearized for perturbations resulting from the change in permeability and resulting change in cross-sectional area of the valve, $A_{1-2}$, due to changes in the liposomes’ porosity. The mass flow into the valve can be decomposed into steady, $\bar{m}_{in}$, and unsteady components, $\dot{m}_{in}'$,

$$\bar{m}_{in} = (C_1 - C_2) \frac{A_{1-2}}{L_{1-2}/2} D_{AB}$$

$$\dot{m}_{in}' = (C_1 - C_2) \frac{A_{1-2}}{L_{1-2}/2} D_{AB} - C_2' \frac{A_{1-2}}{L_{1-2}/2} D_{AB}$$

(1)

The unsteady mass flow out of the valve, $\dot{m}_{out}'$, is
\[ \dot{m}^\prime = (C'_2 - C'_3) \frac{\overline{A}_{1-2}}{L_{1-2} / 2} D^\prime_{AB}, \]

which can be rearranged to yield

\[ C'_2 = \frac{\dot{m}^\prime_{out} + \frac{C'_3 \overline{A}_{1-2} D^\prime_{AB}}{L_{1-2} / 2}}{\frac{\overline{A}_{1-2} D_{AB}}{L_{1-2} / 2}} = \frac{L_{1-2} \dot{m}^\prime_{out}}{2 \overline{A}_{1-2} D_{AB}} + C'_3 \quad (2) \]

Establishing the relationship

\[ V_{1-2} \frac{\partial C'_2}{\partial t} = \dot{m}^\prime_{in} - \dot{m}^\prime_{out}, \]

and substituting in the appropriate relationship in Equation (1) results in

\[ V_{1-2} \frac{\partial C'_2}{\partial t} = (C_1 - \overline{C}_2) \frac{A'_{1-2}}{L_{1-2} / 2} D_{AB} - C'_2 \frac{\overline{A}_{1-2}}{L_{1-2} / 2} D_{AB} - \dot{m}^\prime_{out} \]

which can be rearranged to

\[ \dot{m}^\prime_{out} = -V_{1-2} \frac{\partial C'_2}{\partial t} - C'_2 \frac{\overline{A}_{1-2}}{L_{1-2} / 2} D_{AB} + (C_1 - \overline{C}_2) \frac{A'_{1-2}}{L_{1-2} / 2} D_{AB} \quad (3) \]

Substituting Equation (2) into Equation (3) and simplifying results in

\[ \frac{V_{1-2} L_{1-2}}{2 \overline{A}_{1-2} D_{AB}} \frac{\partial \dot{m}^\prime_{out}}{\partial t} + \dot{m}^\prime_{out} \frac{s^2}{2} = -V_{1-2} \frac{\partial C'_3}{\partial t} - \frac{2 \overline{A}_{1-2} D_{AB} C'_3}{L_{1-2}} + \frac{2 A'_{1-2} D_{AB}}{L_{1-2}} (C_1 - \overline{C}_2) \quad (4) \]

Taking the Laplace Transform of Equation (4) gives

\[ \dot{m}^\prime_{out} \left( \frac{V_{1-2} L_{1-2}}{2 \overline{A}_{1-2} D_{AB}} s + s^2 \right) = -C'_3 \left( \frac{2 \overline{A}_{1-2} D_{AB}}{L_{1-2}} + V_{1-2} s \right) + \frac{2 A'_{1-2} D_{AB}}{L_{1-2}} (C_1 - \overline{C}_2) \]

which can be used to find the two transfer functions of the valve dynamics, \( G_i \) and \( G_s \).
\[
\frac{\dot{m}_{\text{out}}'}{A_{1-2}'} = \frac{4D_{AB}^2A_{1-2}'}{2A_{1-2}L_{1-2}D_{AB}s^2 + V_{1-2}L_{1-2}s} (C_1 - C_2')
\]

\[
\frac{\dot{m}_{\text{out}}'}{C_3'} = \frac{-2A_{1-2}D_{AB}'}{2A_{1-2}L_{1-2}D_{AB}s^2 + V_{1-2}L_{1-2}s} \left(2A_{1-2}D_{AB} + V_{1-2}L_{1-2}s\right)
\]

The chemistry dynamics are defined as

\[
V_3 \frac{\partial C_3'}{\partial t} = \dot{m}_{\text{out}}' - kV_3C_3'
\]

which says that the mass in the solution is the mass flow out of the valve and into the solution minus the mass consumed in the binding reaction. The transfer function, \(G_3\), is

\[
\frac{C_3'}{\dot{m}_{\text{out}}'} = \frac{1}{V_3(s + k)}
\]

### 3.3 Reaction Kinetics

Ligases are used to join sequences of DNA together. T4 DNA ligase can be used in double strand DNA to either seal nicks or join together blunt ends. In the proposed methods, this ligase is involved in nick sealing reactions. TdTase can be used in the incorporation of nucleotides into a strand.

The reaction kinetics of ligation are driven by the concentration of the solution sequences, the ligase and its necessary co-factors, and the influx of liposome DNA sequences into the solution. T4 DNA ligase from a rapid ligation kit can complete roughly \(1E7\) reactions/\(\mu\)g in 1 minute [95].
3.4 Monomer Incorporation

Given the expected arrival rate of a particular sequence of DNA at the end of the strand, the monomer incorporation, which species is added to the strand, will vary. At this step, the noise of the system is introduced. Only the intrinsic noise from the incorporation is modeled. Other sources of potential noise throughout the process are the result of modeling error. Figure 11 is a more detailed view of the block diagram presented in Figure 5. At the conclusion of the process, the output is the DNA sequencer data from which a record of the temperature is preserved.

![Block diagram of the system being modeled](image)

From the sequencer data, the average length of the formed DNA subchains in the solution and the order of arrival to the chain are known, but not the order from which
the sequences left the liposomes. In other words, the location of a nucleotide within the strand is known but not the time that it was added to the strand. The link is deduced through the length of the strand.

3.4.1 Symbol Definitions

The following are the definitions for the variables used in the following model:

- $c$: the concentration in the window, equals $k/n$
- $i$: the number of nucleotides backwards from the sequencing primer
- $j$: the species of nucleotide (A, C, G, T)
- $k$: the number of the $j$-th species contained within the window
- $L$: the length of the strand
- $\bar{L}$: the expected strand length
- $n$: width of the window
- $x$: $i/L$
- $t$: forward time
- $T$: total time of the experiment
- $\tau$: nondimensional backwards time
- $\lambda^j$: the expected arrival rate of $j$ at time $t$ which is assumed constant over the window $(t+\Delta t)$
- $\lambda$: the expected arrival rate of any species at time $t$ which is assumed constant over the window $(t+\Delta t)$
\( \theta^{-1} \) the average arrival rate of any species over \( T \), equals \( \bar{L} / T \)

\( p \) the relative arrival rate of the \( j \)-th species, equals \( \lambda^j / \lambda \)

These variables are pictured in Figure 12.

![Figure 12: Variable definitions used in derivation](image)

3.4.2 Probability of Concentration

A sliding window of varying width can be used to analyze the DNA. The width of the window is specified in space through a number of nucleotides but not in time. Within each window, the concentration of a specific species of DNA is determined.

Given that the window is \( n \) units wide, the probability that it takes until time \( \Delta \tau \) to write the \( n \)-th unit to the window is given by an Erlang distribution with arrival rate \( \lambda \):

\[
P(\Delta \tau \mid n) = \frac{(\lambda)^n \cdot \Delta \tau^{(n-1)} \cdot e^{-\lambda \Delta \tau}}{(n-1)!} d\Delta \tau
\]

An Erlang distribution measures the amount of time that occurs between events or until a certain event occurs. In this scenario, the specified event is the writing of the \( n \)-th unit to a window. This distribution defines the window in length and time.
Approximating this probability density function with a Gaussian distribution yields:

\[
P(\Delta \tau \mid n) = \frac{\lambda}{\sqrt{2\pi \sqrt{n}}} e^{-\frac{1}{2} \left( \frac{\Delta \tau - n}{\sqrt{n}} \right)^2} d\Delta \tau \quad (5)
\]

Within the specified window, the probability that a total of \( k \) units of the \( j \)-th species are added to the strand during the interval of time length \( \Delta \tau \) is given by the following Poisson distribution with arrival rate \( \lambda^{(j)} \):

\[
P(A) \equiv P(N^{(j)}(\Delta \tau) = k) = \frac{e^{-\lambda^{(j)}\Delta \tau} \left( \lambda^{(j)} \Delta \tau \right)^k}{k!}
\]

A Poisson distribution describes the number of events that occur during a certain window of time. In this scenario, the events are the arrivals of the \( j \)-th species. This distribution describes that within the window that was previously defined, \( k \) units of the \( j \)-th species arrived. It is an accurate approximation to model these reaction kinetics as a Poisson distribution [96].

The probability that a total of \( n \) units are added to the strand during the interval of time length \( \Delta \tau \) is given by the following Poisson distribution with arrival rate \( \lambda \) :

\[
P(B) \equiv P(N(\Delta \tau) = n) = \frac{e^{-\lambda \Delta \tau} \left( \lambda \Delta \tau \right)^n}{n!}
\]

The probability that a total of \( n \) units are added to the strand during the interval \( \Delta \tau \) given that a total of \( k \) units of the \( j \)-th species are added to the strand during that same interval, which is the same as the probability that a total of \( n - k \) units of species
other than the \( j \)-th species are added to the strand during the interval \( \Delta \tau \), is given approximately by the following Poisson distribution:

\[
P(B \mid A) = P(N^0(\Delta \tau) = n - k) = \frac{e^{-\lambda_j}(\lambda_j \cdot \Delta \tau)^{n-k}}{(n-k)!}
\]

Combining these results using Bayes Rule and recognizing that \( N^0(\Delta \tau) = N_n^{(j)} \) if \( \Delta N(\Delta \tau) = n \) and \( c = k / n \) if \( N_n^{(j)} = k \) gives the probability that the concentration of the \( j \)-th species contained in the window given that the \( n \)-th unit of the strand was written at time \( \tau \):

\[
P(A \mid B) = P(N_n^{(j)} = k \mid N(\Delta \tau) = n) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}
\]

where \( p(\tau) = \lambda_j / \lambda \). Approximating this result with a Gaussian distribution, yields:

\[
P(N_n^{(j)} = k \mid \Delta N(\Delta \tau) = n) \approx \frac{1}{\sqrt{2\pi np(1-p)}} \cdot e^{-\frac{\left( k - np \right)^2}{2np(1-p)}} dk
\]

After some manipulation, this may be rewritten as

\[
P(c \mid n = N(\Delta \tau)) \approx \frac{\sqrt{n}}{\sqrt{2\pi np(1-p)}} \cdot e^{\frac{n\left(c - np\right)^2}{2np(1-p)}} dc / n
\]

Defining \( dc \equiv dk / n \) and \( c \equiv k / n \) then results in

\[
P(c \mid n = N(\Delta \tau)) \approx \frac{\sqrt{n}}{\sqrt{2\pi np(1-p)}} \cdot e^{\frac{n\left(c - np\right)^2}{2np(1-p)}} dc
\]  \hspace{1cm} (6)

This equation is the probability of the concentration given the window is \( n \) units long and those units were added during the time interval \( \Delta \tau \).
By combining Equation (5) which defines a window of length \( n \) written in the time interval \( \Delta \tau \) and Equation (6) which describes the probability of the concentration given the window is \( n \) units long, the probability of the concentration given a time is described as follows:

\[
P(c | t) = \sum_{\Delta \tau} P(\Delta \tau | n) \cdot P(c | n, \Delta \tau)
\]

\[
P(c | t) = \int_{\Delta \tau} \frac{\lambda}{\sqrt{2\pi n}} e^{-\frac{1}{2n} \left( \frac{t}{\sqrt{n}} \right)^2} d\Delta \tau \cdot \frac{\sqrt{n}}{\sqrt{2\pi \sqrt{p \cdot (1-p)}}} \cdot e^{-\frac{n}{2} \left( \frac{c-p}{\sqrt{p(1-p)}} \right)^2} dc
\]

Evaluating the probability distribution function yields:

\[
P(c | t) = \frac{\sqrt{n}}{\sqrt{2\pi \sqrt{p \cdot (1-p)}}} e^{-\frac{n}{2} \left( \frac{c-p}{\sqrt{p(1-p)}} \right)^2} dc
\] (7)

Given that the tail is \( i \) units long, the probability that the tail was written after time \( t \), which is the same as the probability that it took until time \( t \) to write the \((L-i)\)th unit of the head of the strand is given by the following Erlang distribution:

\[
P(t | x) = \frac{(\theta^{-1})^{L-i} \cdot i^{(L-i-1)} \cdot e^{-i \theta^{-1}}}{(L-i-1)!} dt
\]

Approximating the probability density function with a Gaussian and defining \( x \equiv i / L \) and \( L = \theta^{-1} T \) yields:

\[
P(t | x) \approx \frac{\theta^{-1}}{\sqrt{2\pi \sqrt{L-i}}} e^{-\frac{1}{2} \left( \frac{i}{\sqrt{L-i}} \right)^2} dt = \frac{\sqrt{L}}{\sqrt{2\pi \sqrt{1-x}}} e^{-\frac{1}{2} \left( \frac{1-i/T}{\sqrt{L^{-1}(1-x)}} \right)^2} \frac{dt}{T}
\] (8)
Combining Equation (7) which is the probability of concentration at a given time and Equation (8) which is the probability of a given time given a location along the strand in $x$ yields:

$$P(c \mid x) = \int_t P(c \mid t) \cdot P(t \mid x) \cdot dt \cdot dc$$

$$P(c \mid x) = \int_t \frac{\sqrt{n}}{\sqrt{2\pi} \sqrt{p \cdot (1 - p)}} e^{-\frac{n}{2} \left(\frac{c - \mu}{\sqrt{p \cdot (1 - p)}}\right)^2} \cdot \phi \cdot dt \cdot dc$$

(9)

where $\phi = \frac{(\theta^{-1})^{i-1} \cdot (L-i-1)^{1-i} \cdot e^{-i\theta^{-1}}}{(L-i-1)!} \approx \frac{\sqrt{L}}{\sqrt{2\pi} \sqrt{1-x}} \cdot e^{-\frac{1}{2} \left(\frac{1}{L} \cdot \frac{x}{1-x}\right)^2}$

This equation describes the probability of the concentration given a location along the strand in $x$.

### 3.4.3 Relationship between Input and Signal

The definition of the signal is

$$s(x) \equiv E[c(x)] - \int_x E[c(x)] \cdot dx$$

(10)

The signal power is defined as

$$S \equiv \int_x s(x)^2 \cdot dx$$

Defining $p = p' + \bar{p}$, allows for the following relationship

$$\bar{p} \equiv \frac{1}{T} \int_t p(t) \cdot dt = \int_x E[c(x)] \cdot dx$$
which can be substituted back into Equation (10) to obtain the definition of the signal from the block diagram in Figure 11:

\[ s(x) = E[c(x)] - \bar{p} \]

Calculating the expectation or mean of \( c(x) \), the following definition is used:

\[ E[c(x)] \equiv \sum_c P(c \mid x) \cdot c \]

Substituting Equation (9), rearranging the terms, and evaluating the integrals yields:

\[ E[c(x)] = \int_{t} \phi \cdot p \cdot \frac{dt}{T} \quad (11) \]

Using this result, the signal can be rewritten as

\[ s(x) = \int_{t} p' \cdot \phi \cdot \frac{dt}{T} \quad (12) \]

In order to solve the signal, \( p'(t) \) needs to be defined. Therefore,

\[ p'(t) = \sum_{n \neq 0} \hat{p}_n \cdot e^{\frac{2\pi nt}{T}} \]

Applying the substitutions \( \tau \equiv 1 - \frac{t}{T} \) and \( \omega_n \equiv 2\pi n \) results in:

\[ p'(t) = \sum_{n \neq 0} \hat{p}_n \cdot e^{-j\omega_n \tau} \cdot e^{j\omega_n t} \]

which can be further manipulated into

\[ p'(t) = \sum_{n \neq 0} \hat{p}_n \cdot e^{-j\omega_n (x-x)} \cdot e^{j\omega_n (1-x)} \]
This equation is substituted back into Equation (12), the terms are rearranged, and the integral is evaluated, which yields

\[ s(x) = \sum_{n \neq 0} \hat{p}_n \cdot e^{j\omega_n (1-x)} \cdot e^{-\frac{\omega_n^2 (1-x)}{2L}} \quad (13) \]

Defining the following relationship,

\[ s(x) = \sum_{n \neq 0} \hat{s}_n \cdot e^{j\omega_n x} = \sum_{n \neq 0} \hat{s}_n^* \cdot e^{j\omega_n (1-x)} \]

it can be rearranged to yield

\[ \hat{s}_n^* = \int_{1-x=0}^{1} s(x) \cdot e^{-j\omega_n (1-x)} d(1-x) \]

Substituting Equation (13) and evaluating the integral results in

\[ \hat{s}_n^* = \hat{p}_n \cdot \frac{2L}{\omega_n^2} \left( 1 - e^{-\frac{\omega_n^2}{2L}} \right) \quad (14) \]

which is manipulated into the transfer function relating the input to the signal where for the convenience of notation \( \omega \) is introduced for \( \omega_n \)

\[ G^*(\omega) \equiv \frac{\hat{s}_n^*}{\hat{p}_n} = \frac{2L}{\omega^2} \left( 1 - e^{-\frac{\omega^2}{2L}} \right) \]

Bode plots of the transfer function were generated by defining \( s = \frac{j\omega}{2L} \) and are shown in Figure 13.
Figure 13: Bode plot of transfer function $G'(\omega)$

The Bode plot indicates a cut-off frequency ($\omega_c$) of 1 and that the system acts as a low pass filter.

### 3.4.4 Energy Spectral Densities

The energy spectral density relates how the energy is distributed across frequencies. The signal energy spectral density is defined as

$$\Phi_s(\omega) \equiv \frac{dS}{d\omega}$$

where $S$ is the signal power which is defined as

$$S \equiv \sum_n |\hat{s}_n|^2 = \sum_n |\hat{s}_n +|^2$$
Substituting Equation (14) which related the input to the signal into the definition of the signal power shown in Equation (16) results in

\[ S = \sum_{n} \hat{p}_n \cdot \frac{2L}{\omega^2} \cdot \left(1 - e^{-\frac{\omega^2}{2L}}\right)^2 \]  \hspace{1cm} (17)

Defining

\[ \Phi_p(\omega) \equiv \frac{dP}{d\omega} \]

and applying Parseval’s Theorem, the following is obtained

\[ P \equiv \int p^2 \cdot \frac{dt}{T} = \sum_{n} \hat{p}_n^2 \approx \int \Phi_p(\omega) \cdot d\omega \]  \hspace{1cm} (18)

This relation is substituted into Equation (17)

\[ S = \int \Phi_p(\omega) \cdot \left[\frac{2L}{\omega^2} \cdot \left(1 - e^{-\frac{\omega^2}{2L}}\right)^2\right] \cdot d\omega \]  \hspace{1cm} (19)

Utilizing the definition in Equation (15) and substituting Equation (19), the signal energy spectral density is found

\[ \Phi_s(\omega) = \Phi_p(\omega) \cdot \left[\frac{2L}{\omega^2} \cdot \left(1 - e^{-\frac{\omega^2}{2L}}\right)^2\right] \]  \hspace{1cm} (20)

The noise energy spectral density is

\[ \Phi_N(\omega) \equiv \frac{dN}{d\omega} \]  \hspace{1cm} (21)

where \( N \) is the noise power defined as
\[ N \equiv \int_x n(x)^2 \cdot dx \quad (22) \]

where \( n(x) \equiv c(x) - E[c(x)] \). The expected noise power is

\[ E[N] = E\left[ \int_x n(x)^2 \cdot dx \right] \]

Because the noise is ergodic, the following assumption can be made:

\[ E[N] = \int_x E[n(x)^2] \cdot dx \quad (23) \]

To find \( n(x)^2 \), the following manipulation is performed

\[ n(x)^2 = E\left[ (c(x) - E[c(x)])^2 \right] = E[c(x)^2] - (E[c(x)])^2 \]

Substituting Equations (9) and (11), rearranging the terms, and evaluating the integral with respect to concentration yields

\[ n(x)^2 = \int \phi \cdot \left( 1 - \frac{1}{n} \right) p^2 \cdot \frac{dt}{T} + \frac{1}{n} \int \phi \cdot p \cdot \frac{dt}{T} - \left( \int \phi \cdot p \cdot \frac{dt}{T} \right)^2 \]

where \( n \) is the window width from Equation (9). This equation is substituted back into Equation (23) to find the expected noise power.

\[ E[N] = \int_x \left( \int \phi \cdot \left( 1 - \frac{1}{n} \right) p^2 \cdot \frac{dt}{T} + \frac{1}{n} \int \phi \cdot p \cdot \frac{dt}{T} - \left( \int \phi \cdot p \cdot \frac{dt}{T} \right)^2 \right) \cdot dx \quad (25) \]

The following substitution is made

\[ p = p' + \bar{p} \]

into Equation (24), which is reduced to
\[ N = \left(1 - \frac{1}{n}\right) \int p^2 \cdot \frac{dt}{T} + \frac{1}{n} \bar{p} - \frac{1}{n} \bar{p}^2 - \int \left( \frac{\phi \cdot p'}{T} \right)^2 \cdot dx \tag{26} \]

Equation (18) is substituted into Equation (26) yielding

\[ N = \left(1 - \frac{1}{n}\right) \int \Phi_p(\omega) \cdot d\omega + \frac{1}{n} \bar{p} - \frac{1}{n} \bar{p}^2 - \int \Phi_p(\omega) \cdot \left[ \frac{2L}{\omega^2} \cdot \left(1 - e^{\frac{\omega^2}{2L}}\right) \right]^2 \cdot d\omega \]

This equation is substituted into the noise energy spectral density definition in Equation (21) to produce

\[ \Phi_N(\omega) = \frac{d\left(1 - \frac{1}{n}\right) \int \Phi_p(\omega) \cdot d\omega + \frac{1}{n} \bar{p} - \frac{1}{n} \bar{p}^2 - \int \Phi_p(\omega) \cdot \left[ \frac{2L}{\omega^2} \cdot \left(1 - e^{\frac{\omega^2}{2L}}\right) \right]^2 \cdot d\omega}{d\omega} \]

which is manipulated and can be further reduced by recognizing that \( \frac{\partial p}{\partial \omega} = 0 \) into

\[ \Phi_N(\omega) = \left(1 - \frac{1}{n}\right) \Phi_p(\omega) - \Phi_p(\omega) \cdot \left[ \frac{2L}{\omega^2} \cdot \left(1 - e^{\frac{\omega^2}{2L}}\right) \right]^2 \] \tag{27} \]

### 3.4.5 Signal-to-Noise Ratio

The signal-to-noise ratio is defined as

\[ SNR(\omega) = \frac{\Phi_s(\omega)}{\Phi_N(\omega)} \]

And substituting Equations (19) and (26) into this definition, produces
\[ SNR(\omega) = \frac{\Phi_p(\omega) \cdot \left[ \frac{2L}{\omega^2} \cdot \left( 1 - e^{-\frac{\omega^2}{2L}} \right) \right]^2}{\left( 1 - \frac{1}{n} \right) \cdot \Phi_p(\omega) - \Phi_p(\omega) \cdot \left[ \frac{2L}{\omega^2} \cdot \left( 1 - e^{-\frac{\omega^2}{2L}} \right) \right]^2} \]

which is reduced to

\[ SNR(\omega) = \frac{\left[ \frac{2L}{\omega^2} \cdot \left( 1 - e^{-\frac{\omega^2}{2L}} \right) \right]^2}{\left( 1 - \frac{1}{n} \right) - \left[ \frac{2L}{\omega^2} \cdot \left( 1 - e^{-\frac{\omega^2}{2L}} \right) \right]^2} \]

(28)

Rearranging the ratio and for \( n \gg 1 \), in this case for \( n \) larger than 8 nucleotides in width, the following assumption can be made

\[ SNR(\omega) = \frac{1}{\left( \frac{\omega^2}{\omega^2} - \frac{2L}{\omega^2} \cdot \left( 1 - e^{-\frac{\omega^2}{2L}} \right) \right)^2} - 1 \]

For ease of notation, if the following substitutions are made:

\[ z = \frac{\omega^2}{2L} \]
\[ dz = \frac{\omega}{L} \, d\omega \]
\[ d\omega = \sqrt{\frac{2}{z}} \cdot \frac{dz}{\sqrt{z}} \]

which yields
Expanding this equation into a Taylor Series, it can be approximated by

$$SNR \approx \frac{1}{z}$$  \hspace{1cm} (29)

This approximation works well and is depicted in Figure 14.

![Figure 14: Plot of SNR (red line) and selected approximation of SNR (black dots)](image)

### 3.4.6 Spectral Efficiency

By the Shannon-Hartley Theorem, the channel capacity is defined as

$$C = \frac{1}{2\pi} \int_{-\infty}^{\infty} \log_2 \left(1 + SNR\right) \cdot d\omega$$
Applying the approximation found for the SNR in Equation (29) and evaluating the integral yields

\[ C = \frac{1}{\ln(2)} \sqrt{\frac{L}{2}} \]

The bandwidth can be found from the Nyquist frequency

\[ f_B = \frac{L}{2T} \]

where \( L \) is the number of samples equal to the strand length and \( T \) is the length of the writing period or experiment.

In nondimensional frequency, \( \omega_B = 2\pi \cdot f_B \cdot T = \pi \cdot L \) and \( B = \frac{\omega_B}{2\pi} = \frac{L}{2} \) and \( C \) can be reduced to

\[ C \approx 1.44\sqrt{B} \approx 1.02\sqrt{L} \]

which can be used to find the spectral efficiency

\[ \frac{C}{B} \approx 1.44 \approx \frac{2.04}{\sqrt{B}} \]

The following figure is a plot of the spectral efficiency and storage density of this method of DNA recording compared to other electronic devices such as commercial hard drives, CDs, and cell phones.
Figure 15: Comparison of the storage density (Hz/in²) and spectral efficiency of DNA recording to other electronic devices.

DNA recording is less spectrally efficient than other electronic devices but offers superior storage density. Vastly more information can be stored in a given volume of DNA.
4. Simulation

The simulation described in the following sections models the different steps proposed in the block diagram shown in Figure 5. Based on the temperature, the liposome permeability is used to determine the porosity. The porosity is used to calculate the diffusion and subsequent concentration in solution. The concentration drives the reaction kinetics and the expected arrival rate at the end of the strand. Given the arrival rate, the incorporation can be established.

There are two basic questions that the simulation needs to answer: in a given time step, (1) did a DNA segment arrive and if so, (2) which species arrived?

In the following sections, the same variable definitions defined in Chapter 3 are used. The liposomes used are DPPC with a transition temperature of 41°C and DSPC with a transition temperature of 54°C.

4.1 Liposome Permeability

The temperature profile for the simulation was varied to mimic the various profiles used experimentally. Below the transition temperature of the liposome, the liposome is considered closed; at or above the transition temperature, the liposome is considered open. The liposomes are initially loaded with DNA to a concentration of $C_{rev,0}$. 
4.2 Diffusion

When the liposomes are at their transition temperature, they will release their encapsulated payloads of DNA into the solution. Once released, the DNA is stored in solution and some portion of that DNA will be incorporated into strands, shown in Figure 16.

The change in the amount of DNA in the solution, $N_{soln}$, is a relationship between the amount of DNA released, $\dot{N}_{rel}$, and the amount of DNA that is incorporated into strands, $\dot{N}_{inc}$

$$\frac{\partial N_{soln}}{\partial t} = (\dot{N}_{rel} - \dot{N}_{inc})$$

This relationship can also be expressed as

$$\frac{\partial C_{soln}}{\partial t} \cdot V_{int} = \dot{N}_{rel} - r \cdot V_{int} \quad (30)$$
where \( C_{\text{soln}} \) is the concentration of a species in solution, \( V_{\text{int}} \) is the interstitial volume between the liposomes in solution, and \( r \) is the reaction rate. The interstitial volume is defined as

\[
V_{\text{int}} \equiv \left( 1 - \frac{V_{\text{lip}}}{V_{\text{soln}}} \right) \cdot V_{\text{soln}}
\]

where \( V_{\text{soln}} \) is the volume of the solution and \( V_{\text{lip}} \) is the volume occupied by the liposomes. The change in the concentration within of the liposome, \( C_{\text{resv}} \), once the liposome becomes porous is approximated by

\[
\frac{\partial C_{\text{resv}}}{\partial t} \approx -C_{\text{resv}} \cdot \frac{\varepsilon_{\text{lip}} \cdot D_{AB} \cdot \pi \cdot d^2}{t}
\]

(31)

where \( \varepsilon_{\text{lip}} \) is the efficiency of the release, \( D_{AB} \) is the diffusion constant for the encapsulated DNA, \( d \) is the diameter of the liposome, and \( t \) is the thickness of the lipid bilayer. The nondimensionalization of this approximation is

\[
\frac{\partial C_{\text{resv}}}{\partial t} \approx \frac{C_{\text{resv}}}{C_{\text{resv}_0}} \cdot \frac{\varepsilon_{\text{lip}} \cdot D_{AB}}{t \cdot d}
\]

(32)

where \( C_{\text{resv}_0} \) is the starting concentration within the liposome. From Equation (32), the relationship for the characteristic time scale is found

\[
\tau_{\text{char}} = \frac{1}{\varepsilon_{\text{lip}} / 100} \cdot \frac{d \cdot t}{6 \cdot D_{AB}}
\]

The fixed variables used to calculate this part of the simulation are summarized in Table 1.
Table 1: Variables used in the calculation of the diffusion and characteristic time scale of the system

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diffusion Constant ((D_{AB}))</td>
<td>0.57E-6 cm²/s</td>
</tr>
<tr>
<td>Liposome Bilayer Thickness ((t))</td>
<td>5 nm</td>
</tr>
<tr>
<td>Liposome Diameter ((d))</td>
<td>8 µm</td>
</tr>
<tr>
<td>Efficiency ((\varepsilon_{lip}))</td>
<td>1%</td>
</tr>
</tbody>
</table>

Equation (31) can be solved to determine the concentration inside the liposome reservoir, \(C_{\text{resv}}\), when the liposomes are above their transition temperature

\[
C_{\text{resv}} = C_{\text{resv, ref}} e^{-\frac{(t-t_{\text{ref}})}{\tau_{\text{char}}}}
\]

where \(C_{\text{resv, ref}}\) and \(t_{\text{ref}}\) are reference variables for the reservoir concentration and time respectively. The reference variables are used in the simulation and are updated when the liposomes undergo a transition.

4.3 Reaction Kinetics

The reaction kinetics of the ligation are modeled as a first order chemical reaction

\[
r_A = k \cdot C_{A, \text{soln}}
\]

where \(r_A\) is the rate of consumption of the species \(A\) with units of moles/s/L, \(k\) is the rate constant with units of \(s^{-1}\), and \(C_A\) is the concentration of the unbound species \(A\) in the solution with units of moles/L.

By nondimensionalizing the concentrations from Equation (30), and applying the reaction kinetics, the following equation is obtained
\[
\frac{\partial}{\partial t} \left( \frac{C_{\text{soln}}}{C_{\text{soln}\_\text{max}}} \right) = -k \cdot \frac{C_{\text{soln}}}{C_{\text{soln}\_\text{max}}} - \frac{\partial}{\partial t} \left( \frac{C_{\text{resv}}}{C_{\text{resv}\_0}} \right)
\]  

(32)

where \( C_{\text{soln}\_\text{max}} \) is the maximum concentration of the solution based on the concentration of the inside of the liposomes. When the temperature of the solution is below the transition temperature of the liposomes and no DNA is being released, this equation simplifies to

\[
\frac{\partial}{\partial t} \left( \frac{C_{\text{soln}}}{C_{\text{soln}\_\text{max}}} \right) = -k \cdot \frac{C_{\text{soln}}}{C_{\text{soln}\_\text{max}}}
\]

which can be solved to produce

\[
C_{\text{soln}} = C_{\text{soln}\_\text{ref}} \cdot e^{-k(t-t_{\text{ref}})}
\]

This equation describes the decrease in concentration of the unbound DNA in solution when the liposomes are closed due to the incorporation into the strand through the ligase. However, when the liposome is open, Equation (32) does not reduce further. Instead, it can be manipulated and solved to produce

\[
C_{\text{soln}} = \left( C_{\text{soln}\_\text{ref}} \cdot \frac{f \cdot C_{\text{resv}\_\text{ref}}}{(k \cdot \tau_{\text{char}} - 1)} \right) \cdot e^{-k(t-t_{\text{ref}})} + \frac{f}{(k \cdot \tau_{\text{char}} - 1)} \cdot C_{\text{resv}\_\text{low}}
\]

where \( f \) the ratio of the liposome volume to the solution volume. This equation describes the concentration of the unbound DNA in solution. The concentration increases as the liposomes release their encapsulated DNA, some of which is being bound into the strand. The liposomes are initially releasing DNA faster than the DNA is bound. A graph demonstrating this is seen in Figure 17.
Figure 17: Graph of the concentration of unbound DNA in solution in which two varieties of liposomes with different transition temperatures are present. The x-axis is the time of the experiment in minutes; the y-axis is the concentration of the unbound DNA in solution, and the secondary y-axis is the temperature of the solution. The concentration of the unbound DNA from the lower transition temperature liposome is shown in dark blue; the concentration of the unbound DNA from the higher transition temperature liposome is shown in magenta; and the temperature profile is shown in cyan.

In the figure, two varieties of liposomes with different transition temperatures are in solution. The temperature of the solution is increased and 30 seconds into the experiment, the transition temperature for the first liposome is reached which begins the release of its encapsulated DNA. It continues to be permeable for over 4 minutes until the solution is cooled to below its transition temperature. The higher transition temperature liposome begins to release 2 minutes into the experiment and releases for 1 minute.
The methodology of the simulation is depicted in Figure 18. If the temperature of the solution is above the liposome’s transition temperature, the liposome is open. When the liposome is open, the concentration of the solution and reservoir are updated. If the liposome also just transitioned, meaning that in that time step it either just opened or just closed, the reference concentrations and time are also updated. If the liposome is not open, only the concentration of the solution is calculated to account for any binding that may be occurring.

\[
C_{\text{coli}} = C_{\text{coli,ref}} \cdot e^{-\frac{r}{k}}
\]

\[
C_{\text{coli}} = \left( C_{\text{coli,ref}} - \frac{f \cdot C_{\text{res,ref}}}{(k \cdot \tau_{\text{clus}} - 1)} \right) \cdot e^{-\frac{r}{k \cdot \tau_{\text{clus}}}} + \frac{f}{(k \cdot \tau_{\text{clus}} - 1)} \cdot C_{\text{res,ref}}
\]

\[
C_{\text{res}} = C_{\text{res,ref}} \cdot e^{-\frac{r}{k}}
\]

Figure 18: Methodology of the simulation. When the liposome is open, the concentration of the solution and reservoir are adjusted. If while opened, the liposomes just transitioned, the reference concentrations and time are updated. If the liposomes are not open, the concentration of the solution is recalculated.
4.4 Monomer Incorporation

The arrival rate of the species from the low transition temperature liposome is

\[ \lambda_{\text{low}} = \frac{k \cdot C_{\text{low soln}}}{C_{\text{primer}}} \]

Similarly, the arrival rate of the species from the high transition temperature liposome is

\[ \lambda_{\text{high}} = \frac{k \cdot C_{\text{high soln}}}{C_{\text{primer}}} \]

In the case where only these two species are present in solution, the arrival rate of any species is

\[ \lambda(t) = \lambda_{\text{low}} + \lambda_{\text{high}} = \frac{k}{C_{\text{primer}}} \left( C_{\text{low soln}} + C_{\text{high soln}} \right) \]

The ratio of the arrival rates, \( p(t)_{\text{low}} \) and \( p(t)_{\text{high}} \) are

\[ p(t)_{\text{low}} = \frac{\lambda_{\text{low}}}{\lambda(t)} = \frac{C_{\text{low soln}}}{C_{\text{low soln}} + C_{\text{high soln}}} \]

\[ p(t)_{\text{high}} = \frac{\lambda_{\text{high}}}{\lambda(t)} = \frac{C_{\text{high soln}}}{C_{\text{low soln}} + C_{\text{high soln}}} \]

The probability of an arrival is

\[ P(\text{arrival}) = \frac{\lambda(t) \cdot \Delta \tau \cdot \lambda_{\text{max}}}{\lambda_{\text{max}} \cdot T} \]

and the probability of a certain species arriving given that an arrival occurred is dependent on the ratio of the arrival rates.
4.5 Results

Five sample DNA data strands were generated using the equations developed in this chapter. The lengths of the strands were 320, 260, 200, 200, and 300 base pairs. A sample data strand is shown in Figure 19 with the start primer, tail, sequence for the low transition liposome, and sequence for the high transition liposome indicated.

![DNA sequence](image)

Figure 19: Sample DNA sequence generated with the simulation. The start primer, tail, sequence for the low transition liposome, and sequence for the high transition liposome are indicated.

To mimic sequencing, the strands were aligned at the tails and compared. At each location, the nucleotide that occurred with the greatest frequency was called and stored. The result was an “average” DNA strand.
5. Materials and Methods

The following sections detail the overall approach of the experiments, the materials and methods that were used in experimental procedures, and preliminary results.

5.1 Materials

The materials required in these experiments were liposomes, DNA, and ligase. Each of these is detailed in the following sections.

5.1.1 Liposomes

Liposomes composed of pure 1,2-Dipalmitoyl-sn-Glycero-3-Phosphocholine (DPPC), 1,2-di-O-Hexadecyl-sn-Glycero-3-Phosphocholine (DHPC), and 1,2-Dimyristoyl-sn-Glycero-3-Phosphocholine (DSPC) were used in experiments. Table 2 contains their respective transition temperatures.

<table>
<thead>
<tr>
<th>Liposome</th>
<th>Transition Temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPPC</td>
<td>41</td>
</tr>
<tr>
<td>DHPC</td>
<td>48</td>
</tr>
<tr>
<td>DSPC</td>
<td>54</td>
</tr>
</tbody>
</table>
Lipids of each variant are obtained through AvantiLipids in powder form. They are stocked in chloroform and stored below 0°C. The liposomes are prepared using the thin film extrusion method.

5.1.2 DNA

The following categories of DNA were used in experiments: sticky ends, nucleotides, and single strand DNA.

5.1.2.1 Sticky Ends

DNA sticky ends were procured from IDT in the form of two single strand segments. These segments were combined and annealed and were stocked in either PBS or DNase/RNase free, DI water at 0°C.

The specific sequences used in the experiments are summarized in Table 3.

<table>
<thead>
<tr>
<th>DNA Sequence</th>
<th>Melting Temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAG GCC TTC AAA CAA GGC CTT G TTT G</td>
<td>57.7</td>
</tr>
<tr>
<td>AAG GCC TTC GGG CAA GGC CTT G CCC G</td>
<td>64.5</td>
</tr>
</tbody>
</table>

5.1.2.2 Nucleotides

Nucleotides were obtained from New England Biolabs. They were stored at 0°C.
5.1.2.3 Single Strand DNA

Single strand DNA was procured from IDT. They were stocked in either PBS or DNase/RNase free, DI water at 0°C.

The specific sequences used in the experiments are summarized in Table 4.

**Table 4: Single strand DNA sequences used in experiments**

<table>
<thead>
<tr>
<th>Use</th>
<th>DNA Sequence (5’ to 3’)</th>
<th>Melting Temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liposome 1</td>
<td>TTT TCT CCC AGT CCA AAG CG</td>
<td>55.5</td>
</tr>
<tr>
<td>Liposome 2</td>
<td>TCT TAC GGG TGT GTA TGG CT</td>
<td>55.7</td>
</tr>
<tr>
<td>Complement</td>
<td>TGG GAG AAA ACG CTT TGG AC</td>
<td>55.7</td>
</tr>
<tr>
<td>Complement</td>
<td>ACC CGT AAG AAG CCA TAC AC</td>
<td>54.5</td>
</tr>
<tr>
<td>Complement</td>
<td>TGG GAG AAA AAG CCA TAC AC</td>
<td>52.7</td>
</tr>
<tr>
<td>Complement</td>
<td>ACC CGT AAG ACG CTT TGG AC</td>
<td>57.4</td>
</tr>
<tr>
<td>Liposome 1</td>
<td>AGT GGG TGA AGA GCC GTC AAG AGT</td>
<td>63.8</td>
</tr>
<tr>
<td>Liposome 2</td>
<td>GCT CAT AGA GGC GTC AGG AGA TGG</td>
<td>60.7</td>
</tr>
<tr>
<td>Complement</td>
<td>TCT TCA CCC ACT ACT CCT GAC CGC</td>
<td>62.1</td>
</tr>
<tr>
<td>Complement</td>
<td>GCC TCT ATG AGC CCA TCT CCT GAC</td>
<td>60.7</td>
</tr>
<tr>
<td>Complement</td>
<td>TCT TCA CCC ACT CCA TCT CCT GAC</td>
<td>59.9</td>
</tr>
<tr>
<td>Complement</td>
<td>GCC TCT ATG AGC ACT CCT GAC CGC</td>
<td>63.2</td>
</tr>
</tbody>
</table>

5.1.2.4 Primers

DNA primers were procured from IDT. They were stocked in either PBS or DNase/RNase free, DI water at 0°C.

The specific sequences used as start primers in the experiments are summarized in Table 5.
Table 5: Start primers used in experiments

<table>
<thead>
<tr>
<th>Experiment</th>
<th>DNA Sequence (5’ to 3’)</th>
<th>Melting Temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sticky Ends</td>
<td>CGC TAT CTC GAG AAG GCC TT</td>
<td>56.6</td>
</tr>
<tr>
<td>Nucleotides</td>
<td>TTT TTT TTT TTT TTT TTT TTT TTT</td>
<td>37.3</td>
</tr>
<tr>
<td>Nucleotides</td>
<td>AAT TAA CCC TCA CTA AAG GG</td>
<td>49.3</td>
</tr>
<tr>
<td>Single Strand</td>
<td>TGG GAG AAA ATT TTT TTT TC</td>
<td>44.8</td>
</tr>
<tr>
<td>Single Strand</td>
<td>ACC CGT AAG ATT TTT TTT TC</td>
<td>46.6</td>
</tr>
<tr>
<td>Single Strand</td>
<td>GCC TCT ATG AGC TTT TTT TTT C</td>
<td>50.4</td>
</tr>
<tr>
<td>Single Strand</td>
<td>TCT TCA CCC ACT TTT TTT TTT C</td>
<td>50.0</td>
</tr>
<tr>
<td>Single Strand</td>
<td>GAA AAA AAA A</td>
<td></td>
</tr>
</tbody>
</table>

The specific sequences used as sequencing primers in the experiments are summarized in Table 6.

Table 6: Sequencing primers used in experiments

<table>
<thead>
<tr>
<th>Experiment</th>
<th>DNA Sequence (5’ to 3’)</th>
<th>Melting Temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleotides</td>
<td>TTT TTT TTT TTT TTT TTT TTT TTT</td>
<td>42.6</td>
</tr>
<tr>
<td>Nucleotides</td>
<td>TTT TTT TTT TTT TTT TTT TTT TTT</td>
<td>42.3</td>
</tr>
<tr>
<td>Nucleotides</td>
<td>TTT TTT TTT TTT TTT TTT TTT TTT</td>
<td>41.2</td>
</tr>
<tr>
<td>Single Strand</td>
<td>TTT TCT CCC ATT TTT TTT TG</td>
<td>45.1</td>
</tr>
<tr>
<td>Single Strand</td>
<td>TCT TAC GGG TTT TTT TTT TG</td>
<td>47.0</td>
</tr>
<tr>
<td>Single Strand</td>
<td>AGT GGG TGA AGA TTT TTT TTT G</td>
<td>50.1</td>
</tr>
<tr>
<td>Single Strand</td>
<td>GCT CAT AGA GGC TTT TTT TTT G</td>
<td>50.7</td>
</tr>
<tr>
<td>Single Strand</td>
<td>CAA AAA AAA A</td>
<td></td>
</tr>
</tbody>
</table>

5.1.3 Ligase

T4 DNA Ligase was obtained from Promega, USB, or New England Biolabs.

TdTase was obtained from Promega. Ligase was stored at 0°C.
5.2 Approach

Nucleic acids were selected as the chemical recording unit. This selection was done for a number of reasons. First, unique sequences can be designed for each thermally sensitive liposome that can be later used in the data analysis to identify the corresponding state variable value. Secondly, these sequences can bind without being affected by steric effects. Finally, the resulting double helix of DNA is compact in size, approximately 2 – 3 nm in diameter and 3 – 4 nm long for a full turn which is between 11 and 12 base pairs.

Prior to release, the nucleotides or sequences are encapsulated in thermosensitive liposomes. Each nucleotide or unique sequence is stored in a different variety of liposome which results in each nucleotide or unique sequence being associated with a unique transition temperature at which it will be released. Once released, there are two options for binding in solution. The first option is to add primer and ligase directly to the liposome solution and the binding takes place throughout the temperature profile as the liposomes release their contents and samples are taken and analyzed. The other option is to take samples of the solution during the temperature profile and to add primer and ligase to each sample.

5.2.1 Sticky Ends

A DNA sticky end is a single strand overhang that extends from a double strand of DNA. Using this characteristic, each thermally sensitive liposome has a specific DNA
sequence with an overhang on each side, which represents a given temperature or acceleration. As the device undergoes temperature changes or accelerations, the corresponding liposomes release the analogous sequences. As long as the sticky ends are complementary, the sequences bind via the overhangs. The feasibility of this option depends on several things including the reliability and efficiency of the DNA binding between complementary sticky ends, the ability to extract out the data, and the sequence design of the overhangs.

The sequence design is particularly important. If sequences are designed so that sequence A can only bind to sequence B and sequence B can only bind to sequence A and C, etc. when the progression AAABBCCCCCD is completed, the resulting strands would look like: ABCD and ABC with the additional sequences not binding to anything. A better sequence design is to have all the DNA sequences have the same overhang sequences but a unique duplex strand middle section that will serve as the identifier of a specific thermally sensitive liposome. Therefore, all of the sequences concatenate with each other with equal probability. The middle identifying segment will be five base pairs in length to ensure sufficient variety in sequences and the sticky ends will be eight nucleotides in length which will ensure stable binding between the base pairs at room temperature. This sticky end length makes each segment twenty-one nucleotides long.

T4 DNA Ligase is used in conjunction with the sticky end DNA. It is present in solution or added to samples removed from solution. However, it was determined that
the chosen sticky end sequences certainly aided in binding upon release, there was a significant chance of “clumping” within the liposome. It would be possible to create an environment within the liposome that would inhibit this premature binding while maintaining an environment conducive to binding outside of the liposomes.

5.2.2 Nucleotides

TdTase is another option for binding together the unique payloads of thermally-sensitive liposomes. Instead of using longer sequences of DNA which could potentially be too large for the liposome pores, this polymerase allows for the use of nucleotides which can be released easily. Each of the thermally sensitive liposomes varieties encapsulates a different nucleotide, which limits the resolution. This limitation is fine for a proof of concept experiment but further work needs to be done to devise a system where better resolution could be attained. The other major limitation is that depending on the cofactor present in solution, TdTase adds nucleotides preferentially. For example, in the presence of magnesium, TdTase adds the nucleotides so that G and A are added at a greater rate than C and T. Conversely, in the presence of cobalt, TdTase adds T and C at a greater rate than A and G [41].

The combination of nucleotides with TdTase has been attempted extensively. However, despite repeated efforts, published results using TdTase under similar conditions could not be successfully replicated.
5.2.3 Single Strand DNA

In order to avoid premature binding of the DNA within the liposomes prior to release, a single strand DNA concept was developed. In this concept, shown in Figure 20, each variety of thermally sensitive liposomes encapsulates a unique 20 to 24 nucleotide long single strand of DNA. The length of the oligonucleotides is as short as possible while also ensuring that the melting temperature of the strand would be high enough to withstand the elevated liposome transition temperatures. This balance is done primarily through the GC content. The sequence design also ensures that there are no internal structures that prevent the sequences from properly binding with each other.

In the surrounding solution, in addition to ligase and cofactors, the complement of each of these liposome strands is present, as well as the linkage complement. The linkage complement takes the last half of one liposome strand complement and the first half of another liposome strand complement. As depicted in the figure, solid blue and red denote the unique DNA strands and hashed blue and red denote the complements. In addition to the exact complement of red and blue – linkage complements must also be available. In other words, if a “red” DNA and a “blue” DNA are released, a complement link must be present in order to record the order.
Figure 20: Single strand DNA concept: hashed colors denote the complements to the sold segments encapsulated in the liposomes. Upon release, the DNA from the liposomes binds with the complements available in solution which preserves the time history data.

Also present in solution is a start sequence. The start sequence is either 20 or 22 nucleotides long. At one of the start is a double strand stretch of 10 bases that is the same across all the start sequences. The remainder of the start sequence is half of the complement of one of the liposome sequences. The start sequence can be selected to be the complement to any of the liposomes with the original DNA present in solution.

This DNA scheme uses T4 DNA ligase present in solution. It is also possible to add the ligase to samples removed from the solution during experiments.

**5.2.4 Basic Experimental Procedure**

Experiments are initially preformed using only DNA and ligase to confirm binding and determine optimal experimental conditions. Subsequently, experiments are
conducted to include liposomes with the DNA and ligase. Experiments that focus on using just DNA involve varying the concentration of the DNA, primer, and ligase. The solutions are incubated at the recommended temperature for the ligase and then heated to deactivate the ligase. Then, the solutions are filtered using Sephadex and either prepared for sequencing or run on a 10% polyacrylamide gel. Sephadex filtering is accomplished by adding 1 mL of hydrated Sepahdex-50 to a spin column that is placed inside of a 1.5 mL centrifuge tube. It is briefly centrifuged to remove the extra solvent and then centrifuged at 13,000g for 2-3 minutes to remove the remaining extra solvent. The sample is added down the center of the Sephadex column and centrifuged at 13,000g for another 2-3 minutes. To prepare for sequencing, known tails must be added to the DNA in order to be able to add successful sequencing primers. Tails are accomplished either through adding a single nucleotide with TdTase or through the use of a designed “end cap” with T4 DNA ligase. After the tails are added, the ligase is again deactivated and the solution is filtered with Sephadex to remove extraneous tails.

The preparation of liposomes involves the drying of a lipid film, hydration with the desired encapsulation solution, preparation of Sephadex columns, the extrusion of the liposomes to achieve a homogenous distribution, and the filtering of the solution through the Sephadex columns to remove any extraneous DNA from the external solution. Another method for removing the unencapsulated DNA is a settling method. The settling method is accomplished by repeatedly allowing the liposomes to settle at
the bottom of the tube, carefully removing the supernatant and adding back an equal volume of DNase/RNase free water. Experiments that involve liposomes, ligase, and DNA are heated cycled through various temperature profiles which have built-in incubation periods. At the conclusion of the temperature profile, the ligase is deactivated and the solution is purified to remove the lipid material, unbound DNA, ligase, and buffer. The samples are then either prepared for sequencing or run on a 10% polyacrylamide gel.
6. Initial Experiments and Results

This chapter describes the initial experiments that were conducted to measure the efficiency of the various filtering methods, confirm that ligation occurred between the selected DNA sequences that were to be encapsulated into the liposomes, validate and quantify that the DNA could be encapsulated into the liposomes, and verify that the liposomes’ permeability at various temperatures. The results from each of these experiments are also provided.

6.1 Efficiency of Sephadex Filtering

The efficiency of Sephadex filtering was tested. Sephadex filtering was accomplished by placing a 5 mL syringe barrel into a 15 mL centrifuge tube. A ball of glass wool is placed at the bottom of the syringe barrel. Sephadex-50 solution was slowly added until the syringe barrel is filled – extra solvent is continually caught in the centrifuge tube and removed. The column was briefly centrifuged to remove the extra solvent and then centrifuged at 2000 rpm for 10 minutes to remove the remaining extra solvent. The sample was added down the center of the Sephadex column and centrifuged at 2000 rpm for another 10 minutes. In this experiment, after the sample was first collected, another volume of DNase/RNase free water was added to the column and the column as spun at 2000 rpm for another 10 minutes two additional times.
The settling method was accomplished by repeatedly allowing the liposomes to settle at the bottom of the tube, carefully removing the supernatant and adding back an equal volume of DNase/RNase free water.

Figure 21 is a graph of the concentration of the supernatant (ng/μl) versus runs. Blue diamonds represent measurements from a Sephadex column. Red squares represent measurements from the settle method using an extruded sample. Green triangles represent measurements from the settle method using a non-extruded sample.

![Graph of DNA concentration in supernatant](image)

**Figure 21:** Graph of the DNA concentration in the supernatant: concentration of the supernatant (ng/μl) vs. runs. Blue diamonds represent measurements from a Sephadex column. Red squares represent measurements from the settle method using an extruded sample. Green triangles represent measurements from the settle method using a non-extruded sample.

This data indicates that the Sephadex column removes as much of the unencapsulated DNA in the solution as the first iteration of the settle method of filtering. Both of these methods remove approximately half of the DNA found in the supernatant.
6.2 Confirmation of Ligation

In order to confirm binding of the selected DNA and ligase, several experimental trials were undertaken. The combination of nucleotides with TdTase was attempted extensively. The concentration of the start primers was varied from 1 μM to 5 μM. The concentration of the nucleotides was varied between 1mM and 100 mM. The amount of TdTase was also varied. The results from these trials were run on polyacrylamide gels under various settings: 10-12% polyacrylamide, 200 – 500V, and with bromophenol blue as the loading dye and a version of the dye that contained no visible dye to ensure that the primers were not running on top of the dye which could obscure the subsequent ethidium bromide stain. However, despite repeated efforts, published results using TdTase under similar conditions could not be successfully replicated.

The combination of the single strand DNA with T4 DNA ligase was also completed. In each sample, 1 μL of a 150 μM solution of “start” was added. Half the samples had “Start 4” and the other half had “Start 5”. The samples that had Start 4 contained either 150 μM of liposome DNA “4” or 150 μM of both liposome DNA “4” and “5”. The samples that had Start 5 contained either 150 μM liposome DNA “5” or 150 μM of both liposome DNA “4” and “5”. All the samples contained 150 μM of each of the complements to the liposome DNA. The experiment sample set-ups are summarized in Table 7.
Table 7: Experimental set-up for ligation experiments. The following amounts of single strand DNA were used in each sample

<table>
<thead>
<tr>
<th>Bucket DNA</th>
<th>Liposome DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>μL</td>
</tr>
<tr>
<td>A</td>
<td></td>
</tr>
<tr>
<td>Start 4</td>
<td>1</td>
</tr>
<tr>
<td>4 bar</td>
<td>1.8</td>
</tr>
<tr>
<td>5 bar</td>
<td>1.8</td>
</tr>
<tr>
<td>4bar/5bar</td>
<td>1.8</td>
</tr>
<tr>
<td>5bar/4bar</td>
<td>1.8</td>
</tr>
<tr>
<td>B</td>
<td></td>
</tr>
<tr>
<td>Start 5</td>
<td>1</td>
</tr>
<tr>
<td>4 bar</td>
<td>1.8</td>
</tr>
<tr>
<td>5 bar</td>
<td>1.8</td>
</tr>
<tr>
<td>4bar/5bar</td>
<td>1.8</td>
</tr>
<tr>
<td>5bar/4bar</td>
<td>1.8</td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
<tr>
<td>Start 4</td>
<td>1</td>
</tr>
<tr>
<td>4 bar</td>
<td>1.5</td>
</tr>
<tr>
<td>5 bar</td>
<td>1.5</td>
</tr>
<tr>
<td>4bar/5bar</td>
<td>1.5</td>
</tr>
<tr>
<td>5bar/4bar</td>
<td>1.5</td>
</tr>
<tr>
<td>D</td>
<td></td>
</tr>
<tr>
<td>Start 5</td>
<td>1</td>
</tr>
<tr>
<td>4 bar</td>
<td>1.5</td>
</tr>
<tr>
<td>5 bar</td>
<td>1.5</td>
</tr>
<tr>
<td>4bar/5bar</td>
<td>1.5</td>
</tr>
<tr>
<td>5bar/4bar</td>
<td>1.5</td>
</tr>
</tbody>
</table>

The T4 DNA ligase used were from Promega and USB – one of each sample for each ligase. The Bucket DNA and Liposome DNA were mixed in solution and the ligase added. As per manufacturer’s instructions, the solutions were incubated for 10 minutes at room temperature and then the ligase were heat inactivated at 65°C for 10 minutes. The samples were then filtered using Sephadex-50 spin columns to remove any DNA sequences that were not bound from the solution. The concentrations of the solutions were measured on a Nanodrop and the results are summarized in Table 8.
Table 8: Concentration data from samples obtained from Nanodrop.

<table>
<thead>
<tr>
<th>Sample</th>
<th>ng/μL</th>
<th>260/280</th>
<th>260/230</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (Promega)</td>
<td>20.5</td>
<td>2.00</td>
<td>1.67</td>
</tr>
<tr>
<td>B (Promega)</td>
<td>97.9</td>
<td>2.04</td>
<td>2.04</td>
</tr>
<tr>
<td>C (Promega)</td>
<td>88.1</td>
<td>2.02</td>
<td>1.97</td>
</tr>
<tr>
<td>D (Promega)</td>
<td>147.9</td>
<td>2.00</td>
<td>1.99</td>
</tr>
<tr>
<td>A (USB)</td>
<td>2.5</td>
<td>2.09</td>
<td>0.84</td>
</tr>
<tr>
<td>B (USB)</td>
<td>147.8</td>
<td>2.01</td>
<td>2.12</td>
</tr>
<tr>
<td>C (USB)</td>
<td>168.8</td>
<td>1.94</td>
<td>2.10</td>
</tr>
<tr>
<td>D (USB)</td>
<td>32.6</td>
<td>1.84</td>
<td>1.90</td>
</tr>
</tbody>
</table>

The samples were then run on a 10% polyacrylamide gel and the results are shown in Figure 22. Lane 2 is a 20/100 DNA ladder; Lanes 3-6 are samples A-D ligated with the Promega kit; Lanes 7-10 are samples A-D ligated with the USB kit.
Figure 22: Polyacrylamide gel from a single strand DNA ligation experiment. Lane 2 is a 20/100 DNA ladder and Lanes 3-10 are samples using different experiment samples. The gel demonstrates that binding occurred and that the samples with both liposome DNA present were longer than the solutions with only one liposome DNA present.

All the lanes show a band at 20bp which would indicate that the unbound sequences were not removed with the Sephadex column. The bands at 30bp would indicate that the liposome DNA bound with a start. Bands at 40bp would indicate that the liposome DNA bound with one of the complement sequences. Bands at 50bp could indicate that the liposome DNA bound with a start and a complement sequence. The smears above 50bp could indicate that the liposome DNA and complement sequences
bound into longer strands. Lanes 5, 6, and 9 which contain either Samples C and D demonstrate the longest sequences which is encouraging and an expected result.

6.3 Validation of Encapsulation

The purpose of these experiments was to prove and quantify the amount of DNA that is encapsulated in the liposomes after hydration.

A 50 mg DPPC film was dried and subsequently hydrated with 2 mL of 150 μM single strand DNA solution or 3.45E-7 mol. After hydration, 1.78 mL were recovered and distributed among three samples: 0.20 mL went into sample A, 0.38 mL went into sample B and 1.2 mL went into sample C. The supernatants of samples A and B were measured to be 1095.9 ng/μl or 3.6E-8 mol and 1046 ng/μl or 6.55E-8 mol respectively. Sample C went on to be extruded using a 1 μm filter.

Sample B was filtered using a settling method in which the liposomes were repeatedly allowed to settle at the bottom of the tube, the supernatant was carefully removed, and an equal volume of DNase/RNase free water was added back. Table 9 summarizes the removed and replaced volume of supernatant, the measured concentration, and resulting moles of DNA found in the supernatant. The total number of moles of DNA from these ten runs sums to 6.90E-8 moles.
Table 9: For each of the iterations of the settle method of filtering of Sample B, the volume of supernatant removed and replaced, the measured concentration of the supernatant, and moles calculated are summarized.

<table>
<thead>
<tr>
<th>Run</th>
<th>Volume</th>
<th>ng/μl</th>
<th>Moles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>1046</td>
<td>1.17E-8</td>
</tr>
<tr>
<td>2</td>
<td>150</td>
<td>829.3</td>
<td>2.04E-8</td>
</tr>
<tr>
<td>3</td>
<td>200</td>
<td>492.8</td>
<td>1.61E-8</td>
</tr>
<tr>
<td>4</td>
<td>200</td>
<td>225.2</td>
<td>7.37E-9</td>
</tr>
<tr>
<td>5</td>
<td>200</td>
<td>119.7</td>
<td>3.92E-9</td>
</tr>
<tr>
<td>6</td>
<td>250</td>
<td>53.4</td>
<td>2.19E-9</td>
</tr>
<tr>
<td>7</td>
<td>250</td>
<td>19.6</td>
<td>8.02E-10</td>
</tr>
<tr>
<td>8</td>
<td>250</td>
<td>8.5</td>
<td>3.48E-10</td>
</tr>
<tr>
<td>9</td>
<td>300</td>
<td>8.9</td>
<td>4.37E-10</td>
</tr>
<tr>
<td>10</td>
<td>300</td>
<td>6.1</td>
<td>3.00E-10</td>
</tr>
</tbody>
</table>

After extrusion, 0.8 mL of solution was recovered and the supernatant was measured to be 656 ng/μl or 2.13E-8 mol. The extruded solution was distributed among two samples: 0.6 mL went into Sample D and 0.2 mL went into Sample E. Sample D was filtered using a Sephadex column and Sample E was filtered using the settling method described above. Table 10 summarizes the removed and replaced volume of supernatant, the measured concentration, and resulting moles of DNA found in the supernatant. The total number of moles of DNA from these ten runs sums to 1.66E-8 moles.
Table 10: For each of the iterations of the settle method of filtering for Sample E, the volume of supernatant removed and replaced, the measured concentration of the supernatant, and moles calculated are summarized.

<table>
<thead>
<tr>
<th>Run</th>
<th>Volume</th>
<th>ng/μl</th>
<th>Moles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>656.1</td>
<td>5.37E-9</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>444.4</td>
<td>2.91E-9</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>289</td>
<td>2.37E-9</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>227.5</td>
<td>1.49E-9</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>149.7</td>
<td>9.80E-10</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>159.9</td>
<td>1.31E-9</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>98.8</td>
<td>8.09E-10</td>
</tr>
<tr>
<td>8</td>
<td>50</td>
<td>80.1</td>
<td>6.56E-10</td>
</tr>
<tr>
<td>9</td>
<td>50</td>
<td>44.4</td>
<td>3.63E-10</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>39.3</td>
<td>3.22E-10</td>
</tr>
</tbody>
</table>

Between the two measured samples, the total amount of DNA not encapsulated in the liposomes was 8.56E-8 moles. Extrapolating this out to the initial total volume, 2.95E-7 moles were not encapsulated or 14.6% of the total DNA was encapsulated into the liposomes.

Further confirmation of encapsulation was obtained. After filtering Samples B and E, the samples were heated to 41°C for 20 minutes. Afterwards, the supernatant concentration was measured at 193.1 ng/μl or 1.16E-8 moles for Sample B and 158.3 ng/μl or 4.17E-9 moles for Sample E – a significant increase which would indicate that the liposomes did in fact encapsulate DNA which was subsequently released upon heating to the transition temperature. The two samples released a total of 1.58E-8 moles of DNA into the solution which represents 4.6% of the initial total DNA and 31.6% of the initial encapsulated DNA.
6.4 Verification of Liposome Release

The purpose of these experiments was to verify the liposomes’ permeability and subsequent release of its encapsulated contents at specific temperatures. The results would provide an approximation of an expected output at temperatures below, at, and above the transition temperatures of the liposomes.

For each pair of liposomes selected, three temperatures were used: (1) a temperature below both of the transition temperatures; (2) the transition temperature of the liposome with the lower transition temperature, and (3) the transition temperature of the liposome with the higher transition temperature, summarized in Table 11. This arrangement allows for (1) no liposomes to release, (2) one liposome to release, and (3) either one or two liposomes to release and demonstrates the ability to differentiate between these scenarios. Three samples for each trial were used.

Table 11: Division of liposomes and temperatures used in verifying the release of the encapsulated contents.

<table>
<thead>
<tr>
<th>Temperature</th>
<th>DPPC v. DHPC</th>
<th>DPPC v. DSPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>34°C</td>
<td>DPPC</td>
<td>28°C</td>
</tr>
<tr>
<td></td>
<td>DHPC</td>
<td>DPPC</td>
</tr>
<tr>
<td></td>
<td>DPPC/DHPC</td>
<td>DPPC/DSPC</td>
</tr>
<tr>
<td>41°C</td>
<td>DPPC</td>
<td>41°C</td>
</tr>
<tr>
<td></td>
<td>DHPC</td>
<td>DPPC</td>
</tr>
<tr>
<td></td>
<td>DPPC/DHPC</td>
<td>DPPC/DSPC</td>
</tr>
<tr>
<td>48°C</td>
<td>DPPC</td>
<td>54°C</td>
</tr>
<tr>
<td></td>
<td>DHPC</td>
<td>DPPC</td>
</tr>
<tr>
<td></td>
<td>DPPC/DHPC</td>
<td>DPPC/DSPC</td>
</tr>
</tbody>
</table>
A 50 mg film of each liposome, DPPC, DHPC, and DSPC, film was dried and subsequently hydrated with 2 mL of 150 μM single strand DNA solution. After hydration, each of these solutions was extruded and filtered using the previously described settle filtering method. The concentrations of the supernatant were measured to verify that the unencapsulated DNA had been removed from the solution. The resulting solutions were each divided into the required samples. Solutions with one liposome contained 60 μL and solutions containing two liposomes contained 120 μL. At each of the temperatures, nine samples or three samples of each liposome combination being tested were heated. After being heated, the samples were placed in the refrigerator and allowed to settle in order to measure the concentration of the supernatant.

Bubble charts are used to summarize the results. Figure 23 is the expected and actual results for DPPC and DHPC, while Figure 24 is the expected and actual results for DPPC and DSPC. The temperature of the water bath is shown in the x-axis while the different liposome combinations are shown on the y-axis along with their respective transition temperatures. The area of the bubble demonstrates the concentration of DNA in the supernatant. The black circle denotes the average concentration of the three samples. The shaded bubble is the concentration from the sample with the highest concentration while the white circle is the concentration from the sample with the lowest concentration. Ideally, the samples should be the same concentration or should show
only small variations due to being from the same “master” liposome preparation and having undergone the same temperature profile at the same time.

At the low temperature, none of the liposomes are expected to release and the supernatant concentration should remain the same as the initial post-filtration level. On the graph, all of the liposome combinations should appear equivalent in concentration and be very low. At the medium temperature, the lower transition temperature liposome, DPPC, should release while the higher transition temperature liposome, either DHPC or DSPC, should not release. On the graph, the low transition temperature liposomes should have a significantly larger circle demonstrating the release and subsequent elevation in supernatant concentration, the high transition temperature liposomes should appear the same as the cold temperature data, and the samples containing both types of liposomes should be half the size as the pure low sample due to twice the volume being present. At the high temperature, the high transition temperature liposome, either DHPC or DSPC, should begin to release and the low transition temperature liposome will either continue to release if the liposome composition does not include lysolipid or stop releasing if the liposome composition does include lysolipid. On the graph, the low transition temperature liposome should appear the same or larger than the bubble for the same liposome composition at the medium temperature depending upon the presence of the lysolipid. The high transition temperature liposome should be significantly larger than it was at the low or medium
temperature but approximately the same size as the low transition temperature liposome at the medium temperature. The samples containing both types of liposomes should be approximately as large as the corresponding bubbles for the single liposomes at their respective transition temperatures.

Figure 23: Expected (top) and actual (bottom) results from liposome release experiment using DPPC and DHPC. The area of the gray, shaded circles represent the maximum concentration from the three samples; the area of the black, solid circles represent the average concentration of the three samples; and the area of the white circles represent the minimum concentration from the three samples.
The actual results from the experiment using DPPC and DHPC show good agreement with the expected results with the same, projected trends seen in the figure. This agreement indicates that below the transition temperature of the liposomes there is little to no release of the encapsulated DNA which allows for distinct regions of release at various temperatures when there are combinations of liposomes present in solution.

The same trends can be expected from the liposome release experiment containing DPPC and DSPC and are shown in Figure 24.
Figure 24: Expected (top) and actual (bottom) results from liposome release experiment using DPPC and DSPC. The area of the gray, shaded circles represent the maximum concentration from the three samples; the area of the black, solid circles represent the average concentration of the three samples; and the area of the white circles represent the minimum concentration from the three samples.

Again, the actual results from the experiment using DPPC and DSPC show good agreement with the expected results with the same, projected trends seen in the figure.
7. Final Experiments and Results

The following sections detail the results obtained from experiments conducted using sticky end DNA, nucleotides, and single strand DNA.

7.1 Sticky Ends

Experiments with the sticky ends were conducted.

7.1.1 Experimental Parameters

The overall experimental procedure is depicted in Figure 25. Two varieties of liposomes were used. Liposome A is composed of pure DPPC using the Thin-Film and Extrusion method previously described. Liposome B is similarly composed of pure DSPC. Each liposome variety has a unique transition temperature due to its composition. Liposome A releases at approximately 41°C and Liposome B releases at approximately 54°C. The liposomes are passively loaded with a unique DNA sticky end fragment. Liposome A has a sequence ratio of log(CG/AT) of -0.041 and B’s sequence ratio is 0.212. These values are used during data analysis to determine the order of the fragments.
Figure 25: Procedure used in preliminary experiments. Two varieties of liposomes are heated. At 41°C, Liposome A’s transition temperature is attained and its DNA fragments are released and bound. At 54°C, Liposome B’s transition temperature is achieved and both A and B release DNA fragments which are bound.

Four temperature profiles were used, as shown in Figure 26, in order to test the binding of Liposome A and B alone and in combination. In combination, testing was completed to ensure that Liposome B did not release below its transition temperature.
Figure 26: Temperature profiles used in preliminary experiments. A denotes Liposome A only. B denotes Liposome B only. AB1 and AB2 denote both Liposomes A and B.

The ligase was added at one of two stages. Some experiments were run with the ligase and necessary reagents present in solution and 5 µL samples were removed at the appointed times, purified, and sent for sequencing. Other experiments were run without ligase present in solution. In these experiments, ligase was added to the samples taken which were subsequently purified and sent for sequencing. Multiple samples were taken at each of the appointed time intervals.

7.1.2 Results

Data analysis consisted of taking the sequence data and using a window of varying size and location to determine the logarithm of the ratio of C + G to A + T.

The results from the preliminary experiments testing Liposome A and B separately at their respective transition temperatures are shown in Figure 27.
Figure 27: Preliminary data from experiments testing Liposome A (top) and B (bottom) at their respective transition temperatures. Axes clockwise from bottom: Nondimensionalized Strand Length, log (CG/AT), Nondimensionalized Time, and Temperature (°C). The blue solid line denotes the log (CG/AT) with respect to time. The red dashed lines represent the respective ratios of the sequences stored in Liposome A and B. The green dotted line represents the temperature profile for the specific experiment.

In these first experiments, Liposome A was heated to 50°C for 5 minutes and then cooled to room temperature immediately by placing the sample on ice. The resulting data strand has a log (CG/AT) value that hovers close to the value of Liposome
A of -0.041. Similarly, Liposome B was heated to 60°C for 5 minutes and then cooled.

The data fell into two categories. The data shown in the lower left of Figure 27 starts high and decreases throughout the time period while the data shown in the lower right of the figure keeps a relatively constant value close to the value of Liposome B of 0.212.

The results from the preliminary experiments testing the combination of Liposome A and B with two different temperature profiles are shown in Figure 28.
Figure 28: Preliminary data from experiments testing Liposome A and B at the lower transition temperature of A (top) and at both transition temperatures along with the two methods of adding ligase to the experiment (bottom). Axes clockwise from bottom: Nondimensionalized Strand Length, log (CG/AT), Nondimensionalized Time, and Temperature (°C). The blue solid line denotes the log (CG/AT) with respect to time. The red dashed lines represent the respective ratios of the sequences stored in Liposome A and B. The green dotted line represents the temperature profile for the specific experiment.

In these experiments, both liposomes were present in solution. The data shown at the top of Figure 28 is from a temperature profile where the temperature was held at 50°C allowing Liposome A to release but not Liposome B, in theory. The resulting data
strand appears to support this. The data shown at the bottom of Figure 28 is from experiments where the temperature profile started at 50°C and was then increased to 60°C. The data on the lower left is from an experiment where the ligase was added to the sample after the sample was taken from the solution. The data on the lower right is from an experiment where the ligase was present in solution throughout the duration of the experiment. The results from these experiments are particularly interesting. Both share similar data profiles and start as a relatively constant lower ratio value and then increase to the upper ratio value after the temperature has been increased. The hypothesis is that the upper value would also remain constant for longer if the experiment was continued longer and not immediately quenched, stopping the experiment and recording.

7.2 Nucleotides

The combination of nucleotides with TdTase was attempted extensively. However, despite repeated efforts, published results using TdTase under similar conditions could not be successfully replicated.

7.3 Single Strand DNA

After the initial experiments described in Chapter 6 were conducted and the various steps of the experiment were verified, a full experiment using the single strand
DNA concept was attempted. The following sections describe the experimental conditions, the results obtained, and a comparison to the simulation.

### 7.3.1 Experimental Parameters

There are many parameters in the experiment that can be varied and as a result there are seemingly endless variations for the experiments. Figure 29 is a tree that indicates some of the scope of parameters that could be varied in the experiment, including whether one or more liposomes is present in solution, which DNA is encapsulated within the liposome, whether the ligase is either present in solution throughout the entirety of the experiment or samples are removed from the solution and added to a ligation bucket, and the temperature profiles that could be selected.

From these options, Figure 30 is a table of the combinations of the parameters that were selected. Thirty-six different combinations were used and a total of ninety samples were prepared according to the methods described previously. Once the liposomes were hydrated, extruded, and the extraneous, unencapsulated DNA was removed through filtration, the samples were made. For each liposome present in solution, 50 μL was used.
Figure 29: Variations in parameters that could be used in single strand experiments
**Figure 30:** Experiment matrix used in the single strand experiment describing the combinations of liposomes, DNA, ligase, and temperature profiles used.

<table>
<thead>
<tr>
<th>Samples</th>
<th>Liposome</th>
<th>DNA</th>
<th>Liposome</th>
<th>DNA</th>
<th>Primer</th>
<th>Ligase</th>
<th>Temp Profile</th>
<th>Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>DPPC</td>
<td>4</td>
<td>DSPC</td>
<td>5</td>
<td>Start 4</td>
<td>Soln</td>
<td>Low</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>DPPC</td>
<td>4</td>
<td>DSPC</td>
<td>5</td>
<td>Start 4</td>
<td>Sample</td>
<td>Low</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>DPPC</td>
<td>4</td>
<td>Start 4</td>
<td>Soln</td>
<td>Low</td>
<td>5</td>
<td>Inc</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DPPC</td>
<td>4</td>
<td>Start 4</td>
<td>Sample</td>
<td>Low</td>
<td>5</td>
<td>Inc</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DPPC</td>
<td>5</td>
<td>DSPC</td>
<td>4</td>
<td>Start 5</td>
<td>Soln</td>
<td>Low</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>DPPC</td>
<td>5</td>
<td>DSPC</td>
<td>4</td>
<td>Start 5</td>
<td>Sample</td>
<td>Low</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>DPPC</td>
<td>5</td>
<td>Start 5</td>
<td>Soln</td>
<td>Low</td>
<td>5</td>
<td>Inc</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DPPC</td>
<td>5</td>
<td>Start 5</td>
<td>Sample</td>
<td>Low</td>
<td>5</td>
<td>Inc</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DSPC</td>
<td>4</td>
<td>Start 4</td>
<td>Soln</td>
<td>Low</td>
<td>5</td>
<td>Inc</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DSPC</td>
<td>4</td>
<td>Start 4</td>
<td>Sample</td>
<td>Low</td>
<td>5</td>
<td>Inc</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DSPC</td>
<td>5</td>
<td>Start 5</td>
<td>Soln</td>
<td>Low</td>
<td>5</td>
<td>Inc</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DSPC</td>
<td>5</td>
<td>Start 5</td>
<td>Sample</td>
<td>Low</td>
<td>5</td>
<td>Inc</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DSPC</td>
<td>4</td>
<td>DSPC</td>
<td>5</td>
<td>Start 4</td>
<td>Soln</td>
<td>Low</td>
<td>2.5</td>
</tr>
<tr>
<td>2</td>
<td>DSPC</td>
<td>4</td>
<td>DSPC</td>
<td>5</td>
<td>Start 4</td>
<td>Sample</td>
<td>Low</td>
<td>2.5</td>
</tr>
<tr>
<td>3</td>
<td>DPPC</td>
<td>4</td>
<td>DSPC</td>
<td>4</td>
<td>Start 5</td>
<td>Soln</td>
<td>Low</td>
<td>2.5</td>
</tr>
<tr>
<td>2</td>
<td>DPPC</td>
<td>4</td>
<td>DSPC</td>
<td>4</td>
<td>Start 5</td>
<td>Sample</td>
<td>Low</td>
<td>2.5</td>
</tr>
<tr>
<td>3</td>
<td>DPPC</td>
<td>5</td>
<td>Start 5</td>
<td>Soln</td>
<td>Low</td>
<td>2.5</td>
<td>Inc</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DPPC</td>
<td>5</td>
<td>Start 5</td>
<td>Sample</td>
<td>Low</td>
<td>2.5</td>
<td>Inc</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DSPC</td>
<td>4</td>
<td>Start 4</td>
<td>Soln</td>
<td>Low</td>
<td>2.5</td>
<td>Inc</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DSPC</td>
<td>4</td>
<td>Start 4</td>
<td>Sample</td>
<td>Low</td>
<td>2.5</td>
<td>Inc</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DSPC</td>
<td>5</td>
<td>Start 5</td>
<td>Soln</td>
<td>Low</td>
<td>2.5</td>
<td>Inc</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DSPC</td>
<td>5</td>
<td>Start 5</td>
<td>Sample</td>
<td>Low</td>
<td>2.5</td>
<td>Inc</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DSPC</td>
<td>4</td>
<td>DSPC</td>
<td>5</td>
<td>Start 4</td>
<td>Soln</td>
<td>High Low</td>
<td>1 Inc</td>
</tr>
<tr>
<td>2</td>
<td>DSPC</td>
<td>4</td>
<td>DSPC</td>
<td>5</td>
<td>Start 4</td>
<td>Sample</td>
<td>High Low</td>
<td>1 Inc</td>
</tr>
<tr>
<td>3</td>
<td>DPPC</td>
<td>4</td>
<td>Start 4</td>
<td>Soln</td>
<td>High Low</td>
<td>1 Inc</td>
<td>3 Inc 1 Inc</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DPPC</td>
<td>4</td>
<td>Start 4</td>
<td>Sample</td>
<td>High Low</td>
<td>1 Inc</td>
<td>3 Inc 1 Inc</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DPPC</td>
<td>5</td>
<td>Start 5</td>
<td>Soln</td>
<td>High Low</td>
<td>1 Inc</td>
<td>3 Inc 1 Inc</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DPPC</td>
<td>5</td>
<td>Start 5</td>
<td>Sample</td>
<td>High Low</td>
<td>1 Inc</td>
<td>3 Inc 1 Inc</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DSPC</td>
<td>4</td>
<td>Start 4</td>
<td>Soln</td>
<td>High Low</td>
<td>1 Inc</td>
<td>3 Inc 1 Inc</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DSPC</td>
<td>4</td>
<td>Start 4</td>
<td>Sample</td>
<td>High Low</td>
<td>1 Inc</td>
<td>3 Inc 1 Inc</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DSPC</td>
<td>5</td>
<td>Start 5</td>
<td>Soln</td>
<td>High Low</td>
<td>1 Inc</td>
<td>3 Inc 1 Inc</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DSPC</td>
<td>5</td>
<td>Start 5</td>
<td>Sample</td>
<td>High Low</td>
<td>1 Inc</td>
<td>3 Inc 1 Inc</td>
<td></td>
</tr>
</tbody>
</table>
For the samples in which the ligase was present in solution throughout the temperature profile, the start sequence, complements, liposomes, ligase, and buffer were added. These samples underwent the temperature profile described including the incubation periods which were 10 minutes at room temperature. At the end of the final incubation period, the end sequence was added and the sample was incubated again. At this point, the ligases were deactivated and the samples were sent for purification to remove the remaining liposome material, unbound DNA, ligase, and buffer.

For the samples in which the ligase would not be present in solution throughout the temperature profile, only the liposomes were added and separate vials were prepared in which the start sequence, complements, ligase, and buffer were added. At the beginning of each incubation period, a 5 μL sample was removed and added to the vial with the ligase. Both the original sample and the ligation vial were incubated for 10 minutes before the original sample resumed the temperature profile described. At the end of the final incubation period, the end sequence was added and the sample was incubated again. At this point, the ligases were deactivated and the samples were sent for purification to remove the remaining liposome material, unbound DNA, ligase, and buffer.

After purification, the samples were sent for sequencing and the sequence of the strand was obtained. Due to the fact that the length and composition of the strands
varied, the sequence obtained was the most frequently occurring nucleotide at each location.

**7.3.2 Simulation Results**

The simulation was used to mimic the same experimental conditions. The following graphs summarize the three temperature profiles used and the resulting concentration of the unbound DNA in solution from the low and high transition temperature liposomes.

![Graph of concentration of DNA in solution](image)

**Figure 31:** Graph of the concentration of unbound DNA in solution in which two varieties of liposomes which different transition temperatures are present. The x-axis is the time of the experiment in minutes; the y-axis is the concentration of the unbound DNA in solution, and the secondary y-axis is the temperature of the solution. The concentration of the unbound DNA from the lower transition temperature liposome is shown in dark blue; the concentration of the unbound DNA from the higher transition temperature liposome is shown in magenta; and the temperature profile is shown in cyan.

The first temperature profile goes 42°C for five minutes and then returns to room temperature for an incubation period to allow the ligase their optimal binding.
temperature. As expected, only the low transition temperature liposome releases unbound DNA into solution.

Figure 32: Graph of the concentration of unbound DNA in solution in which two varieties of liposomes which different transition temperatures are present. The x-axis is the time of the experiment in minutes; the y-axis is the concentration of the unbound DNA in solution, and the secondary y-axis is the temperature of the solution. The concentration of the unbound DNA from the lower transition temperature liposome is shown in dark blue; the concentration of the unbound DNA from the higher transition temperature liposome is shown in magenta; and the temperature profile is shown in cyan.

The second temperature profile goes 42°C for 2.5 minutes and then returns to room temperature for an incubation period to allow the ligase their optimal binding temperature. Next, it goes to 55°C for 2.5 minutes and then again returns to room temperature for an incubation period. As expected, only the low transition temperature liposome releases unbound DNA into solution for the first temperature increase, but both liposomes release in the second temperature increase with the lower transition
temperature liposomes releasing less DNA compared to the first temperature period because the amount of DNA inside these liposomes has decreased.

![Graph of DNA concentration and temperature profile](image)

**Figure 33:** Graph of the concentration of unbound DNA in solution in which two varieties of liposomes which different transition temperatures are present. The x-axis is the time of the experiment in minutes; the y-axis is the concentration of the unbound DNA in solution, and the secondary y-axis is the temperature of the solution. The concentration of the unbound DNA from the lower transition temperature liposome is shown in dark blue; the concentration of the unbound DNA from the higher transition temperature liposome is shown in magenta; and the temperature profile is shown in cyan.

The third temperature profile increases to 55°C for one minute, returns to room temperature for an incubation period, goes to 42°C for three minutes, returns to room temperature, and finally increases to 55°C for one minute before returning to room temperature for the final incubation period to allow the ligase their optimal binding temperature.

For each of the temperature profiles, five strands were generated in the simulation. The lengths of these strands are summarized in Table 12.
Table 12: Strand lengths generated from the simulation for each of the temperature profiles

<table>
<thead>
<tr>
<th></th>
<th>Strand 1</th>
<th>Strand 2</th>
<th>Strand 3</th>
<th>Strand 4</th>
<th>Strand 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>211</td>
<td>251</td>
<td>151</td>
<td>211</td>
<td>131</td>
</tr>
<tr>
<td>Low High</td>
<td>171</td>
<td>271</td>
<td>271</td>
<td>391</td>
<td>271</td>
</tr>
<tr>
<td>High Low</td>
<td>91</td>
<td>151</td>
<td>231</td>
<td>271</td>
<td>171</td>
</tr>
<tr>
<td>High Low</td>
<td>51</td>
<td>151</td>
<td>131</td>
<td>151</td>
<td>131</td>
</tr>
<tr>
<td>High</td>
<td>131</td>
<td>211</td>
<td>171</td>
<td>391</td>
<td>291</td>
</tr>
<tr>
<td>High Low</td>
<td>91</td>
<td>151</td>
<td>131</td>
<td>251</td>
<td>211</td>
</tr>
<tr>
<td>High Low</td>
<td>131</td>
<td>71</td>
<td>71</td>
<td>151</td>
<td>111</td>
</tr>
</tbody>
</table>

7.3.3 Data Analysis

The simulation and experimental sequence data was imported into Matlab. The cumulative content of each nucleotide along the strand was found and stored into a vector. The cumulative data vector was detrended. The Fourier coefficients for the first harmonic were found for the data and used to generate sines and consines from which the correlation of the phase shift was calculated. Both the simulation and experimental data was rescaled by the amplitude. In the following figures, red denotes experimental data and blue is representative of simulation data. The cumulative data and the first harmonics are shown on the plots. Figure 34 shows plots of the experiment and simulation cumulative data and first harmonics for the low-high and high-low-high temperature profiles. The numbers shown in the upper left corners are the correlation in
A between the experiment and simulation. The left and middle plot of the low-high are from trials in which the ligase was present in solution during the temperature cycle while the plot on the right is from the trial in which the ligase was separate from the main solution and small samples were removed during the temperature and added to the ligase. Both of the plots for the high-low-high were from trials in which the ligase was present in solution.

Figure 34: Plots of the simulation and experimental cumulative data and first harmonics for the temperature profiles of low-high and high-low-high. The correlation in A between the simulation and experimental data is shown in the upper left corner of each plot.

Figure 35 shows plots of the experiment and simulation cumulative data and first harmonics for all three temperature profiles. The numbers shown in the upper left
corners are the correlation in A between the experiment and simulation. All of the plots are from trials in which the ligase was present in solution.

**Figure 35:** Plots of the simulation and experimental cumulative data and first harmonics for each of the tested temperature profiles. The correlation in A between the simulation and experimental data is shown in the upper left corner of each plot.

The experiment matrix is shown in Figure 36 with additional columns for the correlation in A, C, G, and T between the experiment and simulation data for each of the trials. Red indicates trials where poor correlation is expected which would indicate that information was not stored. Blue indicates trials where good correlation is expected which would indicate that information was stored.
### 7.3.4 Statistical Analysis

In order to demonstrate that the experiments succeeded in recording information instead of producing colored noise, a statistical analysis was conducted. The correlation
between two random variables is a relationship between the covariance and standard deviation. For two variables defined as

\[ x = A_x \cdot \cos(2mx + \phi_x) \]
\[ y = A_y \cdot \cos(2my + \phi_y) \]

the correlation is defined as

\[ \rho = \frac{\int \mathbf{x} \cdot \mathbf{y}}{\sigma_{xx} \cdot \sigma_{yy}} \]

Substituting the variable definitions, the correlation can be manipulated into the form

\[ \rho = \cos(\Delta \phi) \]

In Figure 37, two experiments have been correlated with their respective simulations and the realized thresholds are indicated with the dashed lines. For correlations above the realized thresholds, the model would hold true. The correlations would be distributed as a 3-dimensional Gaussian distribution.
Figure 37: Plot of the correlation between two experiments and their respective simulations. The dashed lines indicate the realized thresholds. For correlations above these thresholds, the model would hold true. The correlations would take the form of a 3-D Gaussian distribution which would result in most of the correlations being above the realized thresholds.

For uniformly distributed \((\phi_x - \phi_y)\), the probability that a correlation is greater than \(\rho\) is

\[
\Delta \phi = \frac{\arccos \rho}{\pi}
\]

The probability that the correlation is greater than the calculated correlation from the data is equal to the probability that the phase shift is less than the experimental phase shift which equals

\[
P(\rho \geq \rho_A) = P\left(\left|\Delta \phi\right| \leq |\Delta \phi_A|\right) = \Delta \phi_A = \frac{\arccos \rho_A}{\pi}
\]

Applying this equation to the correlations obtained from the low high temperature profiles seen in Figure 34, the following results are obtained
\[ \rho_A = 0.9993 \Rightarrow P(\rho \geq \rho_A) = 0.01191 \]
\[ \rho_A = 0.9953 \Rightarrow P(\rho \geq \rho_A) = 0.03087 \]
\[ \rho_A = 0.9558 \Rightarrow P(\rho \geq \rho_A) = 0.09499 \]

The probability of getting the three above simultaneously is obtained by multiplying these probabilities together to yield

\[ 0.01191 \cdot 0.03087 \cdot 0.09499 = 3.5 \times 10^{-5} = 1 - 0.99997 \]

Therefore, based on the obtained correlations from the low-high temperature profile trials, it can be stated with 99.997% confidence that information was recorded.

To compute the correlation threshold needed for 95% confidence using three samples:

\[ \rho^3 = 1 - 0.95 \]
\[ \rho = 0.3684 \]
\[ \Delta \phi = \frac{\arccos \rho}{\pi} = 0.3799 \]

Similarly, the correlation threshold needed for 95% confidence using one sample is 0.05; for five samples it is 0.55.

Figure 38 is a plot indicating that for a realized threshold from the correlation between the experimental and simulation data, there is a range of correlations that can be determined for which a confidence interval of 95% can be determined. The correlation ranges for a 95% and 98% confidence interval for the three trials from the low-high temperature profile and two trials from the high-low-high temperature profile are summarized in Table 13.
Figure 38: Using the realized threshold (red, dashed lines), the range of correlations (dotted, black lines) for which a 95% confidence interval can be established is shown in solid red.

Table 13: Range of correlations necessary to achieve a 95% or 98% confidence interval for the realized threshold.

<table>
<thead>
<tr>
<th></th>
<th>95% confidence interval</th>
<th>98% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min correlation</td>
<td>Max correlation</td>
</tr>
<tr>
<td>LH1</td>
<td>0.9991</td>
<td>0.9994</td>
</tr>
<tr>
<td>LH2</td>
<td>0.9943</td>
<td>0.9962</td>
</tr>
<tr>
<td>LH3</td>
<td>0.9454</td>
<td>0.9643</td>
</tr>
<tr>
<td>HLH1</td>
<td>0.0710</td>
<td>0.2657</td>
</tr>
<tr>
<td>HLH2</td>
<td>0.7721</td>
<td>0.8416</td>
</tr>
</tbody>
</table>

7.3.5 Discussion

Good correlation between the experiment and simulation in certain trials such as those where both liposomes were present for the low-high and high-low-high temperature profiles would indicate that information was in fact recorded. Out of the five trials where those conditions were tested, four of the cases produced correlations...
higher than 0.8097. One of the high-low-high trials produced lower correlation than expected but this could be attributed to a form of experimental error where the trial was not set up correctly with the various DNA sequences, cofactors, and ligase, the ligase in the trial were deactivated, or the sample did not undergo the subsequent tailing and purification steps successfully. Given the results obtained, it can be concluded that information was in fact successfully recorded using the mechanism proposed.

In addition to the noise introduced in the system at the incorporation step, the reading mechanism of sequencing provides new noise due to the fact that the strands in the sample vary in length and content. To improve reading of the noisy strand, one method would be to obtain the fractional read instead of or in addition to the call data from the sequencer. The fractional read would provide the relative content at each position of the strands.

Other alternatives to sequencing include the use of cloning or sequencing through a nanopore. In the cloning option, one strand is vectored into each *E. coli* colony. The colonies are plated and treated with antibiotic. The colonies that contain a DNA vector proliferate while those without the DNA vector die. Once the colonies with the DNA are identified, the DNA is isolated and can be amplified using PCR. This method would result in a sample of DNA that contains DNA that is consistent in length and content and would yield cleaner sequencing results. By repeating this process many times, the sample size can be increased to produce more data.
8. Conclusions

The objectives of this research were to develop “genetic memory”, the storage of information into genetic material, through the demonstration of the feasibility and benefits of recording the time-history of one or more environmental state variables into genetic material. First, the amount of information that could be stored into non-coding DNA using a lossy mechanism such as regulated diffusion was determined. Next, a mechanism through which this concept could be accomplished was proposed and tested.

The approach selected involved the combination of thermally sensitive liposomes and DNA to sense and record information about environmental state variables, in this case temperature. Each variant of thermally sensitive liposomes has a distinctive transition temperature and encapsulates a unique sequence of DNA either in the form of sticky ends, nucleotides, or single strand. At the transition temperature, the liposome becomes porous and releases its encapsulated DNA which subsequently enters the solution and binds together, storing the information.

A comprehensive mathematical model was developed that characterizes the amount of information that can be stored into DNA using a non-exact approach or lossy data conversion. This model resolves issues in the proposed approach for sensing and recording; the order of arrival to the chain is not necessarily the order from which it left the liposomes and the location of a nucleotide within the strand is known but not the
time that it was added to the strand. The link is deduced through the length of the strand. Despite the fact that an exact replica of time cannot be made in space, information can be stored into DNA. The model demonstrates that while noisy, substantial information can be stored. From the model, a simulation was developed to be used as a comparison for experiments.

To determine the feasibility of the proposed approach, experiments testing ligation of the proposed DNA sequence designs, encapsulation of the DNA into liposomes, and subsequent release were conducted. The results are very promising and indicate that information was in fact stored using the proposed mechanism and also demonstrate areas where future work can be conducted in order to optimize the process.
References


[59] K. Iga, Y. Ogawa, and H. Toguchi, "Rates of systemic degradation and reticuloendothelial system (RES) uptake of thermosensitive liposome


Biography

María Elisa Tanner was born in Fort Carson, Colorado in 1983. She earned a Bachelors of Science in Mechanical Engineering from the Massachusetts Institute of Technology in 2004 and a Masters of Science in Mechanical Engineering and Materials Science from Duke University in 2006.