A theoretical model for self-assembly of flexible tiles

Ana Staninska
University of South Florida

Follow this and additional works at: http://scholarcommons.usf.edu/etd
Part of the American Studies Commons

Scholar Commons Citation

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
A Theoretical Model for Self-Assembly of Flexible Tiles

by

Ana Staninska

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
Department of Mathematics
College of Arts and Sciences
University of South Florida

Co-Major Professor: Nataša Jonoska, Ph.D.
Co-Major Professor: Gregory L. McColm, Ph.D.
Arunava Mukherjea, Ph.D.
Masahiko Saito, Ph.D.
Stephen Suen, Ph.D.

Date of Approval:
May 1, 2007

Keywords: DNA, spectrum, pot graph, probabilistic model, complete complexes, cyclic molecules

©Copyright 2007, Ana Staninska
Dedication

To my sister Tanja.
ACKNOWLEDGEMENTS

Although I am the sole author of this dissertation, many people have helped me reach this goal. First and foremost, I would like to express my sincere and deepest gratitude to my research supervisors, dissertation advisors, and mentors, Nataša Jonoska and Gregory McColm. Their guidance and support helped me and inspired me in my research, my studies and everyday life. Besides overseeing my research, they taught me very valuable lessons in and outside the University. I have learned so much from them, and can just hope to be nearly as good an advisor one day. Nothing in this dissertation would have been possible without them. I consider myself the luckiest student to have the best mentors possible.

I extend my thanks to all the committee members, Arunava Mukherjea, Masahiko Saito, and Stephen Suen, for creating an excellent research atmosphere and providing me with valuable comments and ideas. The time I spent in Nadrian Seeman’s laboratory at NYU was an unforgettable and very valuable experience. I am grateful to him for giving me this opportunity to work on wet lab experiments, and to Banani Chakraborty for introducing me to the world of chemistry. I would also like to thank the other students in the laboratory for being helpful and patient with a mathematician. In this occasion I would like to thank my undergraduate thesis advisor Dončo Dimoski for setting me on the right way on my mathematical education. I also wish to thank Mohamed Eddaoudi, chairman of my committee, for taking on the obligation without a qualm.

I would not have come this far if it was not for my immediate family. They have been my rock, my supporters, and they gave me encouragement when it was
most needed. There are no words with which I can express my gratitude towards my parents Krum and Ilinka, my sister Tanja, her husband Ljupco, and the center of my world, their daughters, Bojana and Kalina.

It is beyond any formal acknowledgment with which I can express my thanks to Ferenc, for his friendship, help, love, support, and understanding. Most of all for bearing these years with me and being always there for me.

I am very fortunate to have a big and supportive family. They have been giving me love and support from afar, following my progress and making my summers in Macedonia very pleasant and eventful. I am expressing my love and gratitude my aunt Snezana, uncle Gligor, aunt Zorica, uncle Lambe, aunt Buba, uncle Tome, and my cousins: Sašo, Deni, Keti, Goko, Irina, Klime, Vlatko, Dime, Vesela and Vesna, with her family, Kire, Goce and Ilina.

During the six years I have spent in Tampa, I could always rely on my friends. I met very dear and good friends whose friendship will last a lifetime. I am expressing my thanks to all of them and to my long time friends.

Finally, I express my appreciation to all my professors from the Department of Mathematics at USF, to the chair Marcus McWaters, associate chair Scott Rimbey, to Beverly, Maryann, Aya, Nancy, Barbara and Sarina for their generous help. The music of Toše Proeski had been keeping me company all these years.

The research for this dissertation was supported by the National Science Foundation grants CCF#0432009 and CCF #0523928.
# Table of Contents

List of Figures .............................................................. iii  
Abstract ................................................................. viii  
1 Introduction ............................................................ 1  
2 Introduction to the Model ............................................. 6  
   2.1 The structure of DNA ........................................... 6  
   2.2 Definition of the Model ........................................ 8  
   2.3 Pot Type Classification ....................................... 16  
3 Spectrum of a Pot ..................................................... 23  
   3.1 Definitions ...................................................... 23  
   3.2 Geometric Representation of the Spectrum .................. 25  
   3.3 Algebraic Representation of the Spectrum ................... 36  
   3.4 Maple Program .................................................. 38  
4 Graph of a Pot with DNA complexes ............................... 40  
   4.1 Introduction .................................................... 40  
   4.2 Definition of a Pot Graph ..................................... 41  
   4.3 Homomorphisms ............................................... 46  
   Minimal and Maximal Complete Complex Graphs ............... 66  
5 Probabilistic Analysis ............................................... 70  
   5.1 Introduction .................................................... 70  
   One Type Pot ...................................................... 72  
   Multi Type Pot .................................................... 77  
   5.2 Another Method for Obtaining the Expected Number of Cycles 77  
   Expectations ....................................................... 78
List of Figures

2.1 a) Schematic representation of one nucleotide. b) Schematic representation of two single stranded DNA molecules connected through Watson-Crick complementarity. .................................................. 7

2.2 Above: Watson-Crick bonding of two DNA junction molecules. Below: Junction graph that represents bonding of the two DNA junction molecules depicted on the left. .................................................. 9

2.3 a) 1-branched junction molecule, i.e., hairpin b) 2-branched junction molecule. .................................................. 10

2.4 a) Tiles $t_1, t_2, t_3$ b) Possible complexes that can be obtained by gluing the tiles in a). .................................................. 10

2.5 The complex $C = \langle T, S, J \rangle$. .................................................. 17

2.6 The complexes $C_1$ and $C_2$. .................................................. 18

2.7 a) The complex $C_C$. .................................................. 19

2.8 a) Satisfiable pot type that is not strongly satisfiable ($t_3$ and $t_4$ cannot be a part of a complete complex) b) weakly satisfiable pot type that is not satisfiable (the sticky end $c$ cannot be a part of a complete complex). .................................................. 21

2.9 a) $|P| = |H^+|$ b) $|P| < |H^+|$ c) $|P| > |H^+|$. .................................................. 21

3.1 a) The minimal complete complexes for the pot type given in Figure 2.9 c b) A complete complex that is minimal, but not extremal complete complex. .................................................. 32
3.2 a) Strongly satisfiable pot type with spectrum \( \{(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})\} \), b) Strongly satisfiable pot type with spectrum \( S(P) = \{(u, 4u-1, 2-5u) : \frac{1}{4} \leq u \leq \frac{2}{5}\} \). c) Strongly satisfiable pot type with spectrum \( \{(1-u-v, u, v) : 0 \leq u \leq 1, 0 \leq v \leq 1, u + v \leq 1\} \).

3.3 The closure of the spectrum of the pot type given in Example 3.2.20 b) is the line segment; the closure of the spectrum of the pot type given in Example 3.2.20 c) is the triangle bounded by the dotted lines along with its interior.

4.1 Pot type \( P \) and its pot graph

4.2 a) A pot type \( P = \{t_1, t_2, t_3\} \). b) The pot graph corresponding to the pot type \( P \) given in a)

4.3 Labeled graph that is not a pot graph.

4.4 Elements from the set of complete complex graphs for the pot type given in Example 4.1

4.5 A complete complex graph and a pot graph

4.6 \( t \cong t' \) via an isomorphism that preserves sticky ends. \( t \cong t'' \) via an isomorphism that does not preserves sticky ends.

4.7 The pot types \( P_1 \) and \( P_2 \) are equivalent

4.8 Two pot types \( P_1 \) and \( P_2 \) with equivalent sets of complete complex graphs

4.9 The sets of complete complex graphs for the pot types given in Figure 4.8

4.10 The sets of complex graphs \( G(P_1) \) and \( G(P_2) \) are equivalent, but \( P_1 \not\cong P_2 \)

4.11 \( G_{P_1} \cong G_{P_2} \), but \( P_1 \not\cong P_2 \)

4.12 A pot type \( P \) and its corresponding pot type \( \tilde{P} \) that don’t contain tile types with the same sticky ends.

4.13 Elements from the sets of complete complex graphs for the pot types \( P \) and \( \tilde{P} \) given in Figure 4.12.
4.14  a) Two complete complexes that share a sticky end in common b) A complete complex build from the complexes in a) .................................................. 68

4.15  pot types that are not similar, but with the same spectrum $S = \{(\frac{1}{2} - u, \frac{1}{2} - u, u, u) : 0 \leq u \leq \frac{1}{4}\}$ and same support $Supp = \{(1, 1, 1, 1)\} .................................................. 69

5.1  (a) Three 2-armed tiles form a triangle which represents a $K_3$ complex. (b) Three tile graphs used in a pot to assemble $K_3$. (c) Complete complexes for this pot will be cycles of length divisible by 3. Cycles $K_3$ and $C_6$ are depicted. .................................................. 85

6.1  The desired product .................................................. 91

6.2  This is a 12% native gel to check the formation of Molecule 1. A 10 nucleotide marker is in the first lane. In the second lane is Molecule 1, in the third lane is a complex consisting of the strands AS11 and AS12, in the fourth lane is a complex consisting of strands AS11 and AS13, in the fifth lane is a complex consisting of the strands AS12 and AS13, in the sixth lane is the strand AS11, in the seventh lane is the strand AS12, and in the eight lane is the strand AS13. .................................................. 94

6.3  This is a 12% native gel to check the formation of Molecule 2. A 10 nucleotide marker is in the first lane. In the second lane is Molecule 2, in the third lane is a complex consisting of the strands AS21 and AS22, in the fourth lane is a complex consisting of strands AS21 and AS23, in the fifth lane is a complex consisting of the strands AS22 and AS23, in the sixth lane is the strand AS21, in the seventh lane is the strand AS22, and in the eight lane is the strand AS23. .................................................. 95
6.4 This is a 12% native gel to check the formation of Molecule 3. A 10 nucleotide marker is in the first lane. In the second lane is Molecule 3, in the third lane is a complex consisting of the strands AS31 and AS32, in the fourth lane is a complex consisting of strands AS31 and AS33, in the fifth lane is a complex consisting of the strands AS32 and AS33, in the sixth lane is the strand AS31, in the seventh lane is the strand AS32, and in the eighth lane is the strand AS33.

6.5 (a) 10 nucleotide marker, (b) ligated molecules from the test tube T5 of 1 \( \mu M \) concentration, (c) exo treatment of the ligated product from test tube T5 of 1 \( \mu M \) concentration, (d) ligated molecules from test tube T4 of 1.5\( \mu M \), (e) exo treatment of the ligated product from test tube T4 of 1.5\( \mu M \), (f) ligated molecules from test tube T3 of 2\( \mu M \), (g) exo treatment of the ligated product from test tube T3 of 2\( \mu M \), (h) ligated molecules from test tube T2 of 2.5\( \mu M \). (i) exo treatment of the ligated product from test tube T2 of 2.5\( \mu M \).

6.6 (a) 10 nucleotide marker (b) ligated molecules from the test tube T10 of 0.1 \( \mu M \) concentration, (c) exo treatment of the ligated product from test tube T10 of 0.1 \( \mu M \) concentration, (d) ligated molecules from test tube T9 of 0.2\( \mu M \), (e) exo treatment of the ligated product from test tube T9 of 0.2\( \mu M \), (f) ligated molecules from test tube T8 of 0.3\( \mu M \), (g) exo treatment of the ligated product from test tube T8 of 0.3\( \mu M \), (h) ligated molecules from test tube T7 of 0.4\( \mu M \), (i) exo treatment of the ligated product from test tube T7 of 0.4\( \mu M \), (j) molecules from test tube T6 of 0.5\( \mu M \), (k) exo treatment of the ligated product from the test tube T6 of 0.5\( \mu M \).

6.7 (a) 10 nucleotide marker, (b) ligated molecules from the test tube T1 of 1\( \mu M \) concentration, (c) exo treatment of the ligated product from test tube T1 of 1\( \mu M \) concentration.
6.8 (a) 10 nucleotide marker, (b) ligated molecules from the test tube T2 of 0.75µM concentration, (c) exo treatment of the ligated product of 0.75µM concentration, (d) ligated molecules from test tube T3 of 0.5µM concentration, (e) exo treatment of the ligated product from test tube T3 of 0.5µM concentration.

6.9 (a) 10 nucleotide marker, (b) ligated molecules from the test tube of 3µM concentration, (c) exo treatment of the ligated product of 3µM concentration, (d) ligated molecules from the test tube of 1.5µM concentration, (e) exo treatment of the ligated product of 1.5µM concentration, (f) ligated molecules from the test tube of 1µM concentration, (g) exo treatment of the ligated product of 1µM concentration.

6.10 (a) 10 nucleotide marker, (b) ligated molecules from the test tube of 0.5µM concentration, (c) exo treatment of the ligated product of 0.5µM concentration, (d) ligated molecules from the test tube of 0.1µM concentration.

6.11 (a) 10 nucleotide marker, (b) ligated molecules from the test tube of 1µM concentration, (c) exo treatment of the ligated product of 1µM concentration, (d) ligated molecules from the test tube of 0.5µM concentration, (e) exo treatment of the ligated product of 0.5µM concentration, (f) ligated molecules from the test tube of 0.1µM concentration, (g) exo treatment of the ligated product of 0.1µM concentration.
Mathematical Models for Molecular Self-assembly
Ana Staninska

Abstract

We analyze a self-assembly model of flexible DNA tiles and develop a theoretical description of possible assembly products. The model is based on flexible branched DNA junction molecules, which are designed in laboratories and could serve for performing computation. They are also building blocks for make of even more complex molecules or structures.

The branched junction molecules are flexible with sticky ends on their arms. They are modeled with “tiles”, which are star like graphs, and “tile types”, which are functions that give information about the number of sticky ends. A complex is a structure that is obtained by gluing several tiles via their sticky ends. A complex without free sticky ends is called “complete complex”. Complete complexes are our main interest.

In most experiments, besides the desired end product, a lot of unwanted material also appears in the test tube (or pot). The idea is to use the proper proportions of tiles of different types. The set of vectors that represent these proper proportions is called the “spectrum” of the pot. We classify the types of pots according to the complexes they can admit, and we can identify the class of each pot from the spectrum and affine spaces. We show that the spectrum is a convex polytope and give an algorithm (and a MAPLE code), which calculates it, and classify the pots in PTIME.

In the second part of the dissertation, we approach molecular self-assembly from a graph theoretical point of view. We assign a star-like graph to each tile in a pot, which
induces a “pot-graph”. A pot-graph is a labeled multigraph corresponding to a given pot type, whose vertices represent tile types. The complexes can be represented by “complex-graphs”, and each such graph is mapped homomorphically into a pot-graph. Therefore, the pot-graph can be used to distinguish between pot types according to the structure of the complexes that can be assembled.

We begin the third part of the dissertation with a pot containing uniformly distributed DNA junction molecules capable of forming a cyclic graph structure, in which all possible Watson-Crick connections have already been established, and compute the expectation and the variance of the number of self-assembled cycles of any size.

We also tested our theoretical results in wet lab experiments performed at Prof. Nadrian C. Seemans laboratory at New York University. Our main concern was the probability of obtaining cyclic structures. We present the obtained results, which also helped in defining an important parameter for the theoretical model.
1 Introduction

Self-assembly can be natural or synthesized process that is understood as a spontaneous organization of simpler structures into more complex ones. Molecular self-assembly is one of the most important aspects of nanotechnology that may lead toward understanding many processes in nature and has a potential for many applications. In this dissertation, we consider several problems to advance our understanding of DNA self-assembly.

In recent years, understanding self-assembly as a process and discovering new ways to use the molecule has led to many scientific advances, both experimentally and theoretically. Many nanostructures, nanomaterials, nanodevices and computational models have been developed based on the principles of self-assembly [10, 40, 60, 61].

In 1987, Tom Head in his paper [15] give an idea for using DNA for computational purposes. In 1994, Len Adleman solved small instance of the Hamiltonian Path Problem [1] using DNA molecules. Since then several different models for biomolecular computation have been developed and many instances of NP complete problems have been addressed using DNA computing [23, 30, 41]. In addition, molecular simulations of finite state automata, cellular automata and Turing machines have been designed mainly through self-assembly and enzyme restrictions [6, 42].

While there has been significant experimental progress in self-assembly, the theoretical understanding is still lagging behind. Several theoretical models for DNA self-assembly have appeared, mostly using rigid square tiles [2, 4, 26, 38, 39]. They model Wang tiles and therefore simulate a universal Turing Machine. Although there are results in the theoretical self-assembly, there is still a need for understanding the
limitations and the complexity of the process. Partial results, mainly concerning rigid tile models have already been observed; the minimal number of tile types needed to build an $N \times N$ squares is known to be $O(\log N/ \log \log N)[3, 39]$; the minimal time needed to build the squares is $O(N)$ [3]; computing the smallest tile set needed for unique self-assembly in a given shape is NP hard [2]; and an arbitrary shape can be assembled with $O(\text{Kolmogorov complexity})$ tiles with scaling [48]. A kinetic tile model was developed by Winfree, that is often used in silico simulations [54].

Also questions about the design and error correction of the rigid tile model have been addressed and answered using different approaches as: proofreading tiles [56], snaked tiles [9], and self-healing tiles [55].

Apart from the other theoretical models on self-assembly, the model that is described in this dissertation uses flexible tiles as its main building block. It was first reported in [20], and further on elaborated in [23, 24]. Problems solvable by the flexible tile model of DNA assembly are precisely the NP time problems [23]. The flexible tiles model branched junction molecules with free sticky ends (single stranded sequences) on their branches. Due to the natural Watson-Crick complementarity, complementary sticky ends of two molecules can glue together and form more complex structures.

Besides encoding NP complete problems, flexible tiles are used for building nanostructure and nanomaterials. For example in [14] a tetrahedron was built from flexible tiles, in [31] Borromean rings were constructed. Also they have been used for construction of two dimensional arrays [32], and suggested for growing a DNA fractal-like molecule [8]. Several experiments have been done using the self-assembly process of flexible tiles for DNA computing purposes [19, 20, 35, 34]. In this dissertation we use the flexible tile model for theoretical analysis of the self-assembly process and its products. This is the first attempt to have a systematic theoretical study of this model.

Due to the complexity of the problem, we concentrate on the static model and do not consider any thermodynamic properties of the solution. The static model deals
only with the input and the output of an experiment. We hope to extend this model to a dynamic version.

The Watson-Crick complementarity is represented with an involution function. Each branched junction molecule is represented as a tile of certain type, where a *tile type* is a function that gives information about the number of sticky ends on the molecule. A *pot* is a collection of tiles of various types and a *pot type* is the set of different tile types.

A computational problem can be encoded with flexible tile types in such a way that a solution to the problem exists if and only if a complete complex (a complex with no sticky ends) of certain size is formed in the sufficiently large pots of that pot type. That is one of the reasons why we concentrate on complete complexes.

Although the main inspiration for the model came from DNA computing, with this model we address issues related to the self-assembly process, such as: predicting the possible outcomes, and determining the perfect mix of tiles for experiments.

In this dissertation three different problems are investigated. They are: a *necessary condition for obtaining only complete complexes*, a *description of pot types and complexes with graphs*, and the *probability of the formation of cyclic structures*.

A *necessary condition for obtaining only complete complexes*.

Given a set of flexible branched junction molecules with sticky ends we consider the question of determining the proper stoichiometry such that all sticky ends could end up connected. The necessary condition for obtaining only complete complexes at the end of an experiment is to use the proper proportion of each type of molecules, which in general is not uniform. The set of vectors for the proper proportions is called the “spectrum”.

The pot types are classified in four classes according to possible components that assemble in complete complexes: unsatisfiable, weakly satisfiable, satisfiable, and strongly satisfiable. Through investigating subsets of affine spaces, we provide an algorithm to identify the class of a given pot type. The pot classification is PTIME computable and can be determined from the spectrum and the support of a pot type,
where the support is a binomial vector of dimension equal to the number of different sticky end types. We also give a Maple program that performs this task. This study is included in Chapter 3.

*Description of Pot Types and Complexes with Graphs.*

Since every complex has a naturally arising graph structure, we turn to graph theory to better describe the products of the self-assembly process. The graph model is used to determine what complete complexes can be assembled from a given pot type as well as compare and classify the pot types themselves. It represents an application of graph homomorphism theory.

The flexible tiles and the way they connect lend themselves naturally to graph representation. Each pot type can be represented as a labeled multigraph, called a *pot-graph*. Each vertex represents a tile type while each edge represents a connection between two tile types that have complementary sticky ends. If two tile types can connect via complementary sticky ends, for example \( h \) and \( \hat{h} \), then the tile types are represented as two vertices in the graph connected with an edge labeled \( h \).

Similarly, each complete complex is represented by a graph. This graph is called *complete complex-graph*. Every complete complex-graph is an isomorphic pre-image of a subgraph of the pot-graph. Using the pot-graph we are able to determine the types of complete complexes one pot type admits. We also characterize the pots according to the pot-graphs.

Two pot types are *equivalent* if they have equal number of tile types and equal number of sticky end types with a proper bijection between them. Since the properties of the pot types can be determined by looking at the corresponding pot graphs, it can be shown that equivalent pot types have isomorphic graphs. Pot types that have isomorphic complete complex-graphs are said to be *similar*. It can be shown that equivalent pot types are also similar, but not vice versa. This study is presented in Chapter 4.

*The Probability of the Formation of Cyclic Structures.*

If one designs a pot with tiles that could build a particular graph structure (com-
plete complex), besides the desired complete complex. Other complete complexes may appear as well. In this chapter we consider a set of tiles that can build cyclic complete complexes and perform a study of the expected numbers of cyclic complete complexes of any size.

The process of self-assembly is stochastic and as a first approximation we present a new random graph model of the products of self-assembly processes. Our model differs from other existing random graph models because it uses non-uniform probabilities for the edge appearance.

We show that the smallest cycles are more likely to appear than larger ones. The expected number of the different cycles is inversely proportional to their length and the standard deviation of the expected number of cyclic complexes is very small.

To check the theory developed, we conducted a wet lab experiment. The experiment consisted of three different 2-branched junction molecules, uniformly distributed, capable of forming a cyclic complete complex of length 3. We wanted to find the lowest concentration for which only cyclic molecules will be formed. However, even in very diluted solution, appearance of double cycles (dimers) were observed. These results are given in Chapter 6. We also used the obtained results to define an important parameter for the random graph model.

The random graph model is used to predict the stability of the experiment and calculate the expected number of cyclic molecules of certain size which appear at the end of an experiment. However, it cannot be used to calculate the expected number of complexes that are not cyclic; further adjustments would be required for that.

Some of the short and long terms goals concerning the model and directions for future research are given in Chapter 7.
The main inspiration for the model presented in this dissertation is the self-assembly of branched junction DNA molecules. Although most of the work is dedicated to DNA self-assembly, the methods and the ideas presented can be applied to other self-assembly processes as well.

We start this chapter with a short introduction on the structure of the DNA molecule, to the extent needed for this dissertation.

### 2.1 The structure of DNA

DNA molecules are very widespread and can be found in every living organism. They have very important roles in the living cell, carrying the genetic information from one generation into the next one and playing a crucial role in the synthesis and regulation of proteins.

DNA is an abbreviation for Deoxyribonucleic Acid; it is a polymer consisting of sequences of monomers, called deoxyribonucleotides or shortly, nucleotides. Each nucleotide consists of three components: a sugar, a phosphate group and a nitrogenous base (See Fig 2.1).

The sugar component, called Deoxyribose, consists of five carbon atoms numbered 1’ through 5’. The phosphate group is attached to the 5’ carbon, while the base is attached to the 1’ carbon. There is also a hydroxyl group (OH) attached to the 3’ carbon of the sugar molecule. When the 5’ phosphate group of one nucleotide joins with the 3’ hydroxyl group of another nucleotide, they form a **strong covalent bond**, also known as phosphodiester bond, which is hard to break. This connection gives
orientation to the molecule, because one linear strand will have a free phosphate group
hanging on one side and a free hydroxyl group on the other side. Usually when we
depict a strand of DNA, we draw an arrow from the 5’ to the 3’ end of the molecule,
since the molecule can be extended by adding nucleotides on its 3’ end.

Nucleotides differ by their bases. There are four bases and they are divided into
two groups: Purines and Pyrimidines. Adenine and guanine, or A and G for short,
are purins, and cytosine and thymine, or C and T for short, are pyrimidines.

The base of one nucleotide can join with the base of another nucleotide in a
certain way, forming a weak hydrogen bond. Adenine can bond with thymine, and
cytosine with guanine; no other base connections are possible (A-T pairing involves
the formation of two hydrogen bonds, while the C-G pairing involves the formation
of three hydrogen bonds between the two nucleotides. So the C-G bond is stronger
than A-T; (however, this fact does not play a role in our investigation). We say that
A is complementary to T and C is complementary to G. This principle of pairing is
called Watson-Crick complementarity (named after James D. Watson and Francis H.
C. Crick).

Figure 2.1: a) Schematic representation of one nucleotide. b) Schematic representation of
two single stranded DNA molecules connected through Watson-Crick complementarity.

Two single stranded DNA molecules with complementary sequences on their bases
connect through Watson-Crick connection in an anti-parallel fashion. This means that the free 5’ end of one strand is on the same side as the 3’ end of the other strand and the two strands bond to each other through their bases, forming a double helix molecule. The double helix is organized so that the sugar-phosphate bond is on the outer side, while the bases are on the inner side. In this presentation, Watson-Crick complementarity will be the main building tool for assembly.

Although the DNA molecule is mostly known as a linear helix structure, it can be constructed into more complex form. Examples include branched junction molecules, double [29] and triple crossover molecules [28] (called DX and TX molecules), etc.

2.2 Definition of the Model

This dissertation explores the theoretical aspects concerning DNA self-assembly. It is based on a theoretical model motivated by the weak hydrogen bonding of the DNA molecules. The model can be adjusted for investigating the self-assembly of DNA tiles and self-assembly of other structures. The thermodynamic properties of the molecules in the test tube (pot) are not included in the description of the assembly process. Also, a relatively uniform melting temperature for the sticky ends is assumed.

The main building blocks for the model are inspired by the branched junction DNA molecules. These are synthesized starlike molecules [44], that have flexible arms with sticky ends (see Figure 2.2 (a) to the left). Each arm has two parts: a body and a sticky end extending from the body. The body part is a double stranded DNA molecule, while the sticky end part is a single stranded DNA molecule. When single stranded parts of two arms with complementary sticky ends hybridize, they glue the molecules by the sticky ends, forming a more complex structure.

For simplicity we ignore some of the technicalities (like the sequences of the molecules, model the sequence of sticky ends with a symbol, representing the complexes as a double stranded molecules) of the self-assembled complexes and represent them as labeled graphs. For example, Figure 2.2 (a) represents the three- and the four -branched junction molecule glued together, depicted in a way close to reality,
Figure 2.2: Above: Watson-Crick bonding of two DNA junction molecules. Below: Junction graph that represents bonding of the two DNA junction molecules depicted on the left.

while Figure 2.2 (b) is a graph representation of three and four branched junction molecules before and after gluing occurred. In Figure 2.2 (b) the graph on the left represents a four-branched junction molecule with four sticky ends labeled $a, b, c,$ and $d$, while the graph in the middle represents a three branched junction molecule with sticky ends $\hat{a}, \hat{b},$ and $e$. The sticky ends $\hat{a}$ and $a$ are complementary to each other, and the sticky ends $\hat{b}$ and $b$ are complementary to each other. In the gluing process, the complementary sticky ends connect, hence the complex that is a product of the gluing of those two graphs (represented in the Figure 2.2 on the right has only sticky ends $c, d,$ and $e$.

A 1-branched junction molecule is a hairpin structure with only one sticky end, a 2-branched junction molecule is a double helix with two sticky ends, one at each end (see Figure 2.3). In general $n$-branched junction molecule will have $n$ sticky ends.
In order for all connections to be possible, the branches of the junction molecules need to be flexible. The flexibility is obtained by adding bulged T’s on the junction sequences, like in [21, 22] and [35].

Two junction molecules can connect in many different ways. For example the three tiles given in Figure 2.4 (a) can connect in four different ways, as depicted in Figure 2.4 (b).

Figure 2.3: a) 1-branched junction molecule, i.e., hairpin b) 2-branched junction molecule.

Figure 2.4: a) Tiles $t_1$, $t_2$, $t_3$ b) Possible complexes that can be obtained by gluing the tiles in a).
In our model, we consider a test tube, or so called (which we can call a pot, with DNA branched junction molecules (tiles) in it. Using appropriate chemical protocols, the complementary parts of the DNA molecules in the test tube hybridize and form more complex structures. We want to perform a study on the assembly process and on the possible outcomes. For that, a formal mathematical definition of the components and of the process is needed.

We start with the sticky ends. If two sticky ends have the same sequences of nucleotides, we say that they are of the same sticky end type and we denote by $H$ the set of sticky end types. Each sticky end in the test tube is of a certain sticky end type and since there are only finitely many sticky end types, $H$ is finite. A sticky end of type $h$ is a copy of the sticky end type $h$ and has the same sequence on nucleotides as the sticky end type $h$.

The Watson-Crick complementarity is modeled with a function $\theta : H \to H$ which is an involution, i.e., $\theta(\theta(h)) = h$ for all $h \in H$. We call $\theta(h) \in H$ the complementary sticky end type to $h$ such that sticky ends of types $h$ and types $\theta(h)$ bond. If $\theta(h) = h'$ then the sticky ends $h$ and $h'$ can connect. For each $h \in H$ we assume that $\theta(h) \neq h = \theta(\theta(h))$. Thus $H$ can be partitioned into two sets, $H^+$ and $H^-$ such that if $h$ is an element of $H^+$ then $\theta(h)$ is an element of $H^-$. To ease the notation we write $(H, \theta)$ for the set of sticky ends $H$ for which $\theta$ represents the complementary function.

We often simplify the notation by writing $\hat{h}$ for $\theta(h)$ and we fix $H$. We use notation $[n] = \{1, 2, 3 \ldots n\}$ and $\mathbb{N} = \{0, 1, 2, \ldots\}$ in what follows.

**Definition 2.2.1.** A tile type over $(H, \theta)$ is a function $t : H \to \mathbb{N}$. A tile of type $t$ has $t(h)$ sticky ends of type $h$. The degree of a tile type $t$ is $d = d(t) = \sum_{h \in H} t(h)$.

We pose a restriction on the tile types, so that it is not possible for a tile type to have sticky ends of type $h$ and $\hat{h}$ for some $h \in H^+$ at the same time, i.e., it is not possible for a given tile type $t$, $t(h) > 0$ and $t(\hat{h}) > 0$.

Informally, a tile represents a type of branched junction molecules. Formally a tile is a star-like graph with (see Figure 2.8) with one central vertex of degree $d(t)$,
and $d(t)$ vertices of degree one labeled with sticky ends. If a tile $t$ is of type $t$ then for each sticky end $h$ of type $h \in H$, $t(h) = t(h) = t(h)$, meaning that the tile $t$ has exactly $t(h)$ sticky ends of type $h$. If the sticky end $h$ is on the tile $t$, we will write $t(h) = 1$, otherwise $t(h) = 0$. A sticky end $h$ can only be on one tile, so if $t_1(h) = 1$ and $t_2(h) = 1$, then $t_1 = t_2$. In a pot with DNA molecules there are many copies of a given type of junction molecules, and hence we can assume potentially an infinite supply of tiles of each type.

For example, Figure 2.4 (a) shows examples of three tiles with degrees 3, 5, and 3, respectively. The central vertex is represented with a black circle and the sticky ends are indicated with different colors and shapes. Those tile types are formally defined as follows:

$$
\begin{align*}
& t_1(a) = 1 \\
& t_1(b) = 1 \\
& t_1(c) = 1 \\
& \tilde{t}_2(\tilde{a}) = 1 \\
& \tilde{t}_2(b) = 1 \\
& \tilde{t}_2(c) = 1 \\
& t_2(d) = 1 \\
& t_3(a) = 1 \\
& t_3(b) = 1 \\
& t_3(c) = 1 \\
& t_3(d) = 1
\end{align*}
$$

**Definition 2.2.2.** A **pot type** over $(H, \theta)$ is a set $P$ of tile types over $(H, \theta)$ such that for any $h \in H$ and $t \in P$, if $t(h) > 0$ then there exists $t' \in P$ such that $t'(h) > 0$. We write $P(H, \theta)$ for a pot type over $(H, \theta)$.

(For example the pot type of Figure 2.4 is $P = \{t_1, t_2, t_3\}$.) A **pot** $P$ over $P$ is a collection of tiles from types in $P$. We model a test tube and its content with a pot $P$. Hence we work with a pot $P$ of type $P$, where $P$ contains infinite supply of distinct tiles of each tile type.

**Definition 2.2.3.** Let $T$ be a set of tiles and $S$ a set of sticky ends. The function $\text{type} : T \cup S \to P \cup H$ is called a **type function** and is defined as:

- for every $h \in S$, $\text{type}(h) = h$, if the sticky end $h$ is of type $h$.
- for every $t \in T$, $\text{type}(t) = t$, if the tile $t$ is of type $t$.

Next we give a formal mathematical definition of the gluing process and formation of complexes.
Definition 2.2.4. A complex over a pot type $P$ is a triple $C = \langle T, S, J \rangle$, where $T$ is a set of tiles with tile types in $P$, $S$ is a set of sticky ends with sticky end types in $H$, and $J$ is a set of unordered pairs $c = \{ (t, h), (t', h') \}$ satisfying the following properties:

a) for each $c = \{ (t, h), (t', h') \} \in J$, $t, t' \in T$, $h, h' \in S$, such that $\text{type}(h) = h, \text{type}(h') = \hat{h}$, and $t(h), t'(\hat{h}) > 0$, (c indicates the connection between two complementary sticky ends) and

b) for each $h \in H$, either $\sum_{t \in T} t(h) = \sum_{h \in S, \text{type}(h)=h} |\{ c : (t, h) \in c \}|$ or $\sum_{t \in T} t(\hat{h}) = \sum_{h' \in S, \text{type}(h')=\hat{h}} |\{ c : (t, h') \in c \}|$, (this prevents a complex from having complementary sticky ends),

c) the sum of the cardinalities $\sum_{h \in S, \text{type}(h)=h} |\{ c : (t, h) \in c \}| \leq t(h)$ for each $t \in T$ and $h \in H$ (this prevents the tile from making more connections than it has sticky ends),

d) $|\{ h \in S : \text{type}(h) = h \}| = \sum_{t \in T} t(h)$ (the total number of sticky ends of type $h$ in $S$ equals the total number of sticky ends of type $h$ on the tiles in $T$).

A sticky end $h$ can be only on one tile, so if $t_1(h) = 1$ and $t_2(h) = 1$, then $t_1 = t_2$.

A complex $C = \langle T, S, J \rangle$ schematically can be presented as a graph $G(C) = (V, E)$ defined in the following way: $V = V_T \cup V_S$, $E = E_T \cup E_H$, where

- $V_T = \{ t : t \in T \}$
- $V_S = \{ h \in S : h \in S, \text{ for } t \in T, t(h) = 1, (t, h) \notin c \text{ for any } c \in J \}$, deg($t$) = d($t$) and deg($h$) = 1
- $E_T = \{ \{ t, t' \} : \text{ there exists } c \in J, c = \{ (t, h), (t', h') \} \text{ for some } h, h' \in S \}$
- $E_H = \{ \{ t, h \} : t(h) = 1, (t, h) \notin c \text{ for every } c \in J \}$. 

13
From the definition of the graph for a given complex it follows that $|V| = |T| + |S| - 2|J|$ and $|E| = |J| + (|S| - 2|J|) = |S| - |J|$.

If a complex $C = \langle T, S, J \rangle$ has a sticky end $h \in S$ such that, for the tile $t \in T$ satisfying $t(h) = 1$, $(t, h) \notin c$, for every $c \in J$, then that sticky end is called a **free sticky end**.

We assume that the assembly process occurs in an extremely diluted solution, so that when two complexes meet, all of their complementary free sticky ends join up and that there are no complementary free sticky ends left. This is the part where the flexibility is needed. If they were rigid, this assumption would have not been possible. However, DNA junction molecules with an addition of bulged T’s on their junction sequences can be made to be flexible. Thus, a complex $C = \langle T, S, J \rangle$ represents one of the outcomes of gluing all possible sticky ends on the tiles in $T$.

**Definition 2.2.5.** The type of a complex $C = \langle T, S, J \rangle$ is the function $\text{type}(C) : H \rightarrow \mathbb{N}$ defined by

$$\text{type}(C)(h) = \sum_{t \in T} t(h) - \sum_{h \in S \atop \text{type}(h) = h} |\{c : (t, h) \in c\}|.$$

Informally, a complex type records the number and the types of the sticky ends that are free.

Note: Tile is also a complex $t = \langle \{t\}, S, \emptyset \rangle$ and a tile type is also a complex type. Both the tile type and the complex type keep the information about the sticky ends and not about the underlying graph structure. Therefore, we define a **structure type**. Structure types are equivalence classes: two complexes are of the same structure type if there is a graph isomorphism from one to the other that preserves the tile types, sticky end types and edges.

For example, the complexes in Figure 2.4 (b) are all of the same type ($C(a) = 1$, $C(c) = 1$, $C(\hat{e}) = 1$, and $C(k) = 0$, for $k \in H = \{a, c, \hat{e}\}$), but all of them are of different structure types.

More generally, two complexes $C_1 = \langle T_1, S_1, J_1 \rangle$ and $C_2 = \langle T_2, S_2, J_2 \rangle$ can be glued by their complementary sticky ends to form a bigger complex $C = \langle T, S, J \rangle$. As for
the tiles, complexes may glue in several different ways, but all sticky ends that can connect must be indeed connected.

Formally we define the gluing process in the following way.

**Definition 2.2.6.** We say that $C = \langle T, S, J \rangle$ is obtained by gluing complexes $C_1 = \langle T_1, S_1, J_1 \rangle$ and $C_2 = \langle T_2, S_2, J_2 \rangle$ if

$$T = T_1 \cup T_2, \quad S = S_1 \cup S_2, \quad \text{and} \quad J = J_1 \cup J_2 \cup \Delta J,$$

where $\Delta J$ is a set of unordered pairs $c = \{(t_1, h_1), (t_2, h_2)\}$ satisfying the following properties:

- **a)** for each $c = \{(t_1, h_1), (t_2, h_2)\} \in \Delta J$, $t_1 \in T_1$, $t_2 \in T_2$, $h_1 \in S_1$ and $h_2 \in S_2$ are free sticky ends such that $\theta(\text{type}(h_1)) = \text{type}(h_2)$ (c indicates the connection between two tiles from both complexes).

- **b)** for each $h \in H$, $\text{type}(C)(h) = \max\{\text{type}(C_1)(h) + \text{type}(C_2)(h) - \text{type}(C_1)(\hat{h}) - \text{type}(C_2)(\hat{h}), 0\}$ (for each $h \in H$, as many free sticky ends of type $h$ as possible are joined).

**Definition 2.2.7.** A complex $C$ is called complete if it has no free sticky ends, i.e., for all sticky ends $h$, $\text{type}(C)(h) = 0$.

If in a pot some complexes have free sticky ends, that means they are still free to glue with other complexes. So if we want to design a certain graph structure with tiles, we would like the final outcome to be a complete complex. Also, a DNA computing problem can be encoded in the tiles, and a solution to the problem will be a complete complex of certain size. Hence, our main interest are, in particular, complete complexes.

For a pot type $P$ we denote by $C(P)$ the set of all complete complexes that can be obtained by tiles of tile types in $P$. 

15
2.3 Pot Type Classification

We classify the pots in four classes: “unsatisfiable”, “weakly satisfiable”, “satisfiable” and “strongly satisfiable” according to possible assemblies of complete complexes.

**Definition 2.3.1.** A complex $C = \langle T, S, J \rangle$ (represented by a graph $G(C) = (V, E) = (V_T \cup V_S, E_T \cup E_S)$) is **embedded** in a complex $C' = \langle T', S', J' \rangle$ (represented by a graph $G' = (V', E') = (V'_T \cup V'_S, E'_T \cup E'_S)$) if $T \subseteq T'$, $S \subseteq S'$, $J \subseteq J'$, and there exists a function $\varphi : V \to V'$. The function $\varphi$ is defined as follows: for $t \in V_T$, $\varphi(t) = t$; for $h \in V_S$,

$$\varphi(h) = \begin{cases} 
  h & \text{for } t \in T \text{ such that } t(h) = 1, (t, h) \notin c, \text{ for any } c \in J' \\
  t' & \text{for } t \in T \text{ such that } t(h) = 1, \text{ there exists } c \in J', (t, h) \in c.
\end{cases}$$

Naturally, a tile $t$ is **embedded** in a complex $C = \langle T, S, J \rangle$ if $t \in T$.

Now we classify the pot types.

**Definition 2.3.2.** A pot type $P$ is **weakly satisfiable** if it admits a complete complex, i.e., $C(P) \neq \emptyset$. Otherwise it is **unsatisfiable**.

A pot type $P$ is **satisfiable** if, for each $h \in H$, there is a complete complex $C \in C(P)$ of the pot containing at least one sticky end of type $h$.

A pot type $P$ is **strongly satisfiable** if every complex that can be generated by $P$ can be embedded into a complete complex of $P$.

Since we are only interested in complete complexes, we would like to obtain only complete complexes as the products of the assembly process. Therefore, strong satisfiability is the notion of most immediate interest in our study.

**Definition 2.3.3.** A complex $C = \langle T, S, J \rangle$ is called **k-tile complex** if $|T| = k$.

**Lemma 2.3.4.** A pot type $P$ is strongly satisfiable if and only if for every tile type $t \in P$, there exists a complete complex $C = \langle T, S, J \rangle \in C(P)$ with $t \in T$ such that $\text{type}(t) = t$. 
Proof. One implication of the lemma is trivial; if $P$ is strongly satisfiable, since every tile is a complex, it can be embedded into a complete complex.

The converse is obtained by mathematical induction on the number of tiles in a complex. If a complex consists of only one tile, i.e., if $C = \langle \{t\}, S, \emptyset \rangle$, then the complex $C$ itself is the tile $t$. By the assumption that any tile can be embedded in a complete complex, it follows that $t$ can be embedded in a complete complex.

Assume the statement holds for $k$-tile complexes; we claim it holds for $(k+1)$-tile.

Let $C = \langle T, S, J \rangle$ be a $(k+1)$-tile complex with $T = \{t_0, t_1, \ldots, t_k\}$. If $C$ is complete then the theorem is proved. Assume it is not, i.e., for some $h \in H$

$$\text{type}(C)(h) = \sum_{i=0}^{k} t_i(h) - \sum_{h' \in S \text{ type}(h') = h} |\{c : (t_i, h) \in c\}| > 0.$$ 

![Figure 2.5: The complex $C = \langle T, S, J \rangle$](image)

Consider the complex $C' = \langle T', S', J' \rangle$ with $T' = \{t_1, t_2, \ldots, t_k\}$ and $J' = J - \{c \in J : (t_0, h) \in c\}$, and the tile $t_0 = \langle \{t_0, S_0, \emptyset\} \rangle$. The complex obtained by gluing the $k$-tile complex $C'$ and the tile $t_0$ is, in fact, $C$.

CASE 1: The complex $C'$ is not complete.

By the inductive hypothesis, $C'$ and $t_0$ can join with complexes $\widehat{C}' = \langle \widehat{T}', \widehat{S}', \widehat{J}' \rangle$ ($\widehat{T}' \neq \emptyset$ since $C'$ is not complete) and $\widehat{C}_{t_0} = \langle \widehat{T}_0, \widehat{S}_0, \widehat{J}_0 \rangle$ ($\widehat{T}_0 \neq \emptyset$), respectively, to form complete complexes $\widehat{C}'$ and $\widehat{C}_{t_0}$ have sticky ends that are complementary to $C'$ and $t_0$ respectively, i.e., $\widehat{S}' = \{\widehat{h}' : \theta(\text{type}(\widehat{h}')) = \text{type}(h') \text{ for } h' \in S'\}$ and
\( \hat{S}_0 = \{ \hat{h}_0 : \theta(type(\hat{h}_0)) = type(h_0) \text{ for } h_0 \in S_0 \} \).

The complexes formed by gluing \( C' \) and \( \hat{C}' \) are complete, and the complexes obtained by gluing \( \hat{C}_{t_0} \) and \( t_0 \) are also complete (as we saw before the gluing process is not unique, so we can glue two complexes in many different ways, each way obtaining a new complex). Let choose one complex from the set of complexes obtained by gluing \( C' \) and \( \hat{C}' \), say \( C_1 = \langle T_1, S_1, J_1 \rangle \). Also, let choose an other complex from the set of complexes obtained by gluing \( \hat{C}_{t_0} \) and \( t_0 \), \( C_2 = \langle T_2, S_2, J_2 \rangle \). The complexes \( C_1 = \langle T_1, S_1, J_1 \rangle \) and \( C_2 = \langle T_2, S_2, J_2 \rangle \) are complete complexes, i.e., \( type(C_1)(h) = 0 \) and \( type(C_2)(h) = 0 \) for every \( h \in H \).

Note that for every \( h \in H \), since \( C_1 \) is a complete complex, in \( S_1 \) there are equal number of sticky ends of type \( h \) and \( \hat{h} \). Same thing hold for \( S_2 \).

\[ \begin{align*}
\hat{S}_0 &= \{ \hat{h}_0 : \theta(type(\hat{h}_0)) = type(h_0) \text{ for } h_0 \in S_0 \}.
\end{align*} \]

The complexes formed by gluing \( C' \) and \( \hat{C}' \) are complete, and the complexes obtained by gluing \( \hat{C}_{t_0} \) and \( t_0 \) are also complete (as we saw before the gluing process is not unique, so we can glue two complexes in many different ways, each way obtaining a new complex). Let choose one complex from the set of complexes obtained by gluing \( C' \) and \( \hat{C}' \), say \( C_1 = \langle T_1, S_1, J_1 \rangle \). Also, let choose an other complex from the set of complexes obtained by gluing \( \hat{C}_{t_0} \) and \( t_0 \), \( C_2 = \langle T_2, S_2, J_2 \rangle \). The complexes \( C_1 = \langle T_1, S_1, J_1 \rangle \) and \( C_2 = \langle T_2, S_2, J_2 \rangle \) are complete complexes, i.e., \( type(C_1)(h) = 0 \) and \( type(C_2)(h) = 0 \) for every \( h \in H \).

Note that for every \( h \in H \), since \( C_1 \) is a complete complex, in \( S_1 \) there are equal number of sticky ends of type \( h \) and \( \hat{h} \). Same thing hold for \( S_2 \).

Consider the complex, obtained by gluing \( \hat{C}' \) and \( \hat{C}_{t_0} \), say the complex \( \hat{C} = \langle \hat{T}, \hat{S}, \hat{J} \rangle \), with \( \hat{T} = \hat{T}' \cup \hat{T}_0 \) and \( \hat{J} = \hat{J}' \cup \hat{J}_0 \cup \Delta \hat{J} \), where \( \Delta \hat{J} \) is the set of unordered pairs \( \{(p, h_p), (q, h_q)\} \), \( p \in \hat{T}' \) and \( q \in \hat{T}_0 \), \( type(h_p) = h \) for some \( h \in H \), and \( type(h_q) = \hat{h} \).

We claim that every complex formed by gluing \( C' \) and \( \hat{C} \) is a complete complex. Consider a complex \( C_C = \langle T_C, S_C, J_C \rangle \) formed by gluing \( C \) and \( \hat{C} \). By the definition of gluing we have \( T_C = T \cup \hat{T}, S_C = S \cup \hat{S}, \) and \( J_C = J \cup \hat{J} \cup \Delta J \), where \( \Delta J \) is a set of unordered pairs \( \{(r, h_1), (s, h_2)\} \) such that a free sticky end of \( h_1 \), with \( type(h_1) = h_1 \) for some \( h_1 \in H \), from tile \( r \in T \) connects to a free sticky end \( h_2 \) of type \( \hat{h}_1 \) from tile \( s \in \hat{T} \).

From the previous observation, \( T_C = T \cup \hat{T} = T \cup \hat{T}' \cup \hat{T}_0 = (\{t_1, \ldots, t_k\} \cup \hat{T}') \cup \hat{T}_0 \). 

\[ \hat{T}_0 = \{ \hat{h}_0 : \theta(type(\hat{h}_0)) = type(h_0) \} \text{ for } h_0 \in S_0 \} \]
(\{t_0\} \cup \hat{T}_0) = T_1 \cup T_2. \text{ Also, } S_C = S \cup \hat{S} = S' \cup S_0 \cup \hat{S}' \cup \hat{S}_0 = S_1 \cup S_2.

From the definition of the complex it follows that for every \(h \in H\), \(\{|h \in S_C : \text{type}(h) = h\}| = \sum_{t \in T_C} t(h)\) and \(\{|\hat{h} \in S_C : \text{type}(\hat{h}) = \hat{h}\}| = \sum_{\hat{t} \in \hat{T}_C} t(\hat{h})\). Since in \(S_C\), for every \(h \in H\), there are as many sticky ends of type \(h\) as of type \(\hat{h}\) it follows that \(|\{h \in S_C : \text{type}(h) = h\}| = |\{\hat{h} \in S_C : \text{type}(\hat{h}) = \hat{h}\}| = \sum_{h \in S_C, \text{type}(h) = h} |\{c \in J_c : c = \{(t, h), (t', h')\}, \text{type}(h) = h, \text{type}(h') = \hat{h}, t, t' \in T_C, h, h' \in S_C\}|\). Hence,

\[
\text{type}(C_C)(h) = \sum_{t \in T_C} t(h) - \sum_{h \in S_C, \text{type}(h) = h} |\{c \in J_c : (t, h) \in c\}|
= \sum_{t \in T_C} t(h) - |\{h \in S_C : \text{type}(h) = h\}|
= \sum_{t \in T_C} t(h) - \sum_{t \in T_C} t(h) = 0,
\]

i.e., \(C_C\) is a complete complex.

![Figure 2.7: a) The complex \(C_C\)](image)

**STEP 2:** The complex \(C'\) is complete.

Since \(P\) is strongly satisfiable, there exists a complex \(\hat{C}_{t_0} = \langle \hat{T}_0, \hat{S}_0, \hat{J}_0 \rangle\) such that a complex obtained by gluing \(t_0\) and \(\hat{C}_{t_0}\) is complete. Consider a complex obtained by gluing the complexes \(C\) and \(\hat{C}_{t_0}\), say a complex \(C^* = \langle T^*, S^*, J^* \rangle\). From the definition for gluing of two complexes, \(T^* = T \cup \hat{T}_0\), \(S^* = S \cup \hat{S}_{t_0}\), and \(J^* = J \cup \hat{J}_0 \cup \Delta J\), where \(\Delta J = \{c = \{(t_0, h), (t, h')\} : h \in S_0 - S, t \in \hat{T}_0, h' \in \hat{S}_0, \theta(\text{type}(h')) = \text{type}(h)\}\).
From the definition of the complex it follows that for every $h \in H$, $|\{h \in S^* : \text{type}(h) = h\}| = \sum_{t \in T^*} t(h)$ and $|\{\tilde{h} \in S^* : \text{type}(\tilde{h}) = \tilde{h}\}| = \sum_{t \in T^*} t(\tilde{h})$. Since in $S^*$, for every $h \in H$, there are as many sticky ends of type $h$ as of type $\tilde{h}$ it follows that $|\{h \in S^* : \text{type}(h) = h\}| = |\{\tilde{h} \in S^* : \text{type}(\tilde{h}) = \tilde{h}\}| = \sum_{h \in S^* \atop \text{type}(h) = h} |\{c \in J_c : c = \{(t, h), (t', h')\}, \text{type}(h) = h, \text{type}(h') = \tilde{h}, t, t' \in T^*, h, h' \in S^*\}|$.

The complex $C^*$ is complete since for every $h \in H$:

$$\text{type}(C^*)(h) = \sum_{t \in T^*} t(h) - \sum_{h \in S^* \atop \text{type}(h) = h} |\{c \in J^* : (t, h) \in c\}|$$

$$= \sum_{t \in T^*} t(h) - |\{h \in S^* : \text{type}(h) = h\}|$$

$$= \sum_{t \in T^*} t(h) - \sum_{t \in T^*} t(h) = 0,$$

It is straightforward to see that all strongly satisfiable pot types are also satisfiable and that all satisfiable pot types are weakly satisfiable. But the converse is not necessary true. Figure 2.8 a) shows a pot type that is satisfiable, but not strongly satisfiable, since tiles of the tile types $t_3$ and $t_4$ can never be embedded into a complete complex. The pot type of Figure 2.8 b) is an example of a pot type that is weakly satisfiable, but not satisfiable since the sticky end type $c$ can never be a part of any finite complete complex.
Figure 2.8: a) Satisfiable pot type that is not strongly satisfiable (t₃ and t₄ cannot be a part of a complete complex) b) weakly satisfiable pot type that is not satisfiable (the sticky end c cannot be a part of a complete complex).

Note that the number of sticky end types does not depend on the number of tile types. The pot types in all three examples in Fig 2.9 are strongly satisfiable; in the first example the number of tile types and sticky end types are equal; in the second one the number of tile types are less than free sticky end types; and in the third example the number of tile types is greater than sticky end types.

Figure 2.9: a) |P| = |H⁺|  b) |P| < |H⁺|  c) |P| > |H⁺|.

For the rest of the dissertation we reserve $m = |P|$ ( $P = \{t₁, t₂, \ldots, tₘ\}$), and $2n = |H|$ ( $H = \{h₁, \ldots, hₙ, \hat{h}_₁, \ldots, \hat{h}_ₙ\}$).

To each complex type $C$ we associate a vector $zₖ = (z₁, z₂, \ldots, zₙ)$ from $\mathbb{Z}^n$ such that
\( z_i : H^+ \to \mathbb{Z} \)

\[ z_i = \text{type}(C)(h_i) - \text{type}(C)(\widehat{h_i}). \]

We assume that the pot is diluted and the thermodynamic conditions are such that all sticky ends that can connect would be able to. In this sense \( z_C \) gives information about the remaining free sticky end types on the complex \( C \).

Therefore, either \( \text{type}(C)(h_i) > 0 \) or \( \text{type}(C)(\widehat{h_i}) > 0 \), but not both. If \( \text{type}(C)(h_i) > 0 \), then \( z_i > 0 \), and if \( \text{type}(C)(\widehat{h_i}) > 0 \), then \( z_i < 0 \).

Since a tile is a complex, no tile has complementary sticky ends. So, for every \( t \in P \) and every \( h \in H \), if \( t(h) > 0 \) then \( t(\widehat{h}) = 0 \). In this case, for every tile \( t \), we associate a vector \( z_t = (z_t(h_1), z_t(h_2), \ldots, z_t(h_n)) \) from \( \mathbb{Z}^n \) such that

\[
z_t(h_i) = \begin{cases} 
  t(h_i) & \text{if } t(h_i) > 0 \\
  -t(h_i) & \text{if } t(\widehat{h_i}) > 0 \\
  0 & \text{otherwise}.
\end{cases}
\]

NOTE: The method and the theory described here is correct even if tiles with complementary sticky ends are accepted.
3 Spectrum of a Pot

3.1 Definitions

From reports on DNA assemblies we know that when one runs an experiment, the desired complexes are not the only things that shows up in the pot; there may be a lot of incomplete complexes. They increase the error rate and the cost of the experiment.

If the stoichiometry in the test tube is bad, i.e., an improper ratio of each of the molecules is used, then under any conditions incomplete complexes will be present. In this section we propose a method which theoretically (ignoring all dynamic considerations such as those in [27]) eliminates the presence of incomplete complexes assuming that assembly occurs in ideal conditions in a well mixed diluted pot. The sets of vectors of the ratio of the molecules is called the spectrum of the pot, and we give an algorithm for calculating it using the Jordan-Elimination method. The closure of a spectrum of a pot in Euclidian space is a simplex, whose vertices correspond to connected complete complexes. Also, in this section through investigating subsets of affine spaces we give a method for identifying the class of a given pot type.

Definition 3.1.1. The spectrum of $P$ is the set $S$ of all vectors $r = (r_t : t \in P)$ such that:

1) For each $t$, $r_t \geq 0$ and

$$\sum_{t \in P} r_t = 1,$$  \hspace{1cm} (3.1.1)

2) for each $h$,

$$\sum_{t \in P} r_t t(h) = \sum_{t \in P} r_t \hat{t}(h),$$ \hspace{1cm} (3.1.2)
i.e., for each $h \in H$ there are as many sticky ends of type $\hat{h}$ as there are of type $h$.

Using the vector $z_t(h) = t(h) - t(\hat{h})$, associated to tile $t$ when it is considered as a complex, the second part of the definition can be rewritten in the following form

\[ \sum_{t \in P} r_t(t(h) - t(\hat{h})) = 0, \]
\[ \sum_{t \in P} r_t z_t(h) = 0. \]

An obvious observation from the definition of the spectrum, is that the spectrum can be represented as an intersection of the hyperplane $H_1 = \{ x \in \mathbb{R}^m : \sum_{i=1}^m x_i = 1 \}$ with the kernel of the linear transformations. Also, note that the spectrum $\mathcal{S}(P) \subseteq [0, 1]^{[P]}$.

If a pot has a mixture of tiles whose proportions correspond to a vector in the spectrum, then in perfect conditions only complete complexes need be expected. If the used proportion of the molecules is not in the spectrum, there are no conditions under which at the end of the experiment only complete complexes will be present in the test tube. Note that a pot type admits a complete complex if and only if its spectrum is nonempty.

For a vector $r = (r_1, r_2, \ldots, r_m) \in \mathcal{S}(P)$, $r_i \in [0, 1]$ for $i \in [m]$ and $\sum_{i=1}^m r_i = 1$. Therefore, the vector $r$ can be considered as a vector of probabilities for tiles to be on a complete complex, i.e., $r_j$ can be considered as the probability that a randomly selected tile is of type $t_j$.

**Example 3.1.2.** The spectrum of the pot type given in Figures 2.8 (a), (b) each containing three tile types, is the solution of the following systems of equations for
\[ r_1, r_2, r_3 \geq 0. \]

(a) \[ r_1 + r_2 + r_3 + r_4 = 1 \]
(b) \[ r_1 + r_2 + r_3 + r_4 = 1 \]
\[ r_1 - r_2 + 2r_3 = 0 \]
\[ r_1 - r_2 + r_3 + r_4 = 0 \]
\[ r_1 - r_2 + 2r_4 = 0 \]
\[ r_1 - r_2 = 0 \]
\[ r_1 - r_2 + 2r_4 = 0 \]
\[ r_1 - r_2 = 0 \]
\[ r_3 - r_4 = 0 \].

Both systems have the same solution, i.e., the spectrum of both pot types is \( S = \{ (\frac{1}{2}, \frac{1}{2}, 0, 0) \} \), but the first pot type is satisfiable, while the second is only weakly satisfiable. These two examples show that spectrums cannot be used to distinguish between a weakly satisfiable pot type and strongly satisfiable pot type.

In the above example for no sticky ends to remain free, i.e., only complete complexes to be assembled, the spectrum points out that one needs to use equal number of molecules of the first two types, and no use of any molecules from the other two types.

Use of proportion of tile types from the spectrum is necessary for eliminating the incomplete complexes at the end of an experiment, but it does not give information about the type of complete complexes. We can only assume that in a very diluted solution, the smallest complexes will be most favorable.

There are finite number of tiles in a given pot type, so the proportion of each tile is a rational number. For the practical purposes we consider \( S(P) \subseteq \mathbb{Q}^m \), \( \mathbb{Q} \) being the set of rational numbers.

### 3.2 Geometric Representation of the Spectrum

First, we give some definitions from Linear Programming that will be used. [11, 33] gives a good introduction to the subject.

**Definition 3.2.1.** A **polyhedron** in \( \mathbb{R}^n \) is the set \( \{ \mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b} \} \), where \( A \in \mathbb{R}^{m \times n} \) is a matrix and \( \mathbf{b} \in \mathbb{R}^m \) is a vector. A bounded polyhedron is called a **polytope**.
Definition 3.2.2. Let $a_1, a_2, \ldots, a_n$ be scalars and $b$ be a point in $\mathbb{R}$. A hyperplane is the set of all points $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ satisfying $\sum_{i=1}^{n} a_i x_i = b$.

Definition 3.2.3. Let $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ be a nonempty polyhedron. If $c$ is a nonempty vector for which $\delta = \max\{cx : x \in P\}$ is finite, then $\{x : cx = \delta\}$ is called a supporting hyperplane of $P$. A face of $P$ is $P$ itself or the intersection of $P$ with a supporting hyperplane of $P$. A point $v$ for which $\{v\}$ is a face is called a vertex of $P$.

Definition 3.2.4. A convex combination of finite number of points $a_1, a_2, \ldots, a_n$ is

$$\sum_{i=1}^{n} \lambda_i a_i, \text{ where } \sum_{i=1}^{n} \lambda_i = 1$$

and $\lambda_i \geq 0$ for $i = 1, 2, \ldots, n$.

Definition 3.2.5. A convex hull of a set of points $S$ is the set of all convex combinations of the points from $S$. An extreme point of $S$ is a point that cannot be written as a convex combination of two other points from $S$.

Definition 3.2.6. Consider the system of equations $Ax = b$, $x \geq 0$, where $x$ is an $n$-vector, $b \in \mathbb{R}^m$, $A$ is an $m \times n$ matrix. A feasible solution to the system is a vector $x = (x_1, x_2, \ldots, x_n)$ with $x_i \geq 0$, for $i \in [n]$, and $Ax = b$.

Definition 3.2.7. Let $P = \{x \in \mathbb{R}^n ; Ax = b, \ x \geq 0\}$, where $A$ is an $m \times n$ matrix, $x = (x_1, x_2, \ldots, x_n)$. Let $B$ be a subset of $[n]$, with $|B| \leq m$ such that $A_B$, the matrix consisting of columns of $A$ that correspond to the indices of $B$, is invertible. A basic feasible solution is a feasible solution $x$ with

- $x_j = 0$, for $j \notin B$,

- $A_B x_B = b$, or $x_B = A_B^{-1} b$, where $x_B$ is the vector consisting of elements of $x$ restricted to the indices of $B$.

It is known that the convex hull of a set $S$ is the smallest convex set containing $S$ [7]. Also, every polytope is the convex hull of its extreme points and the extreme points are the vertices of the polytope [7].
Also, from Linear programming [7, 33], it is known that a vector \( r \) is an extreme point of the convex polyhedron \( S = \{ r : A r = b, \ r \geq 0 \} \) (\( A \) is \( n \times m \), \( \text{rank}(A) = n < m \), \( b \in \mathbb{R}^n \), and \( r \in \mathbb{R}^m \)) if and only if \( r \) is a basic feasible solution to \( A r = b \).

Denote by \( \mathcal{H}_m \) the intersection of the subspace of \( \mathbb{Q}_+^m = \{(r_1, r_2, r_3 \ldots r_m) : r_i \in \mathbb{Q}, r_i \geq 0 \ for \ i \in [m]\} \), and the hyperplane \( r_1 + r_2 + \ldots + r_m = 1 \). Definition 3.1.1 shows that the spectrum of a pot with \( m \) tile types is a subset of the set \( \mathcal{H}_m \) and the \( n \) hyperplanes (for each \( h, \sum_{t \in P} r_t z_t(h) = 0 \)).

The intersection of the hyperplanes and \( \mathbb{R}_+^m \) is a polytope, i.e., it is a convex hull of its vertices. Therefore the spectrum of any given pot is dense in the corresponding convex hull and contains all the vertices of the hull.

**Proposition 3.2.8.** The spectrum \( S(P) \) of a pot type \( P \) with \( |P| = m \) and corresponding set of sticky ends \( H \) with \( |H^+| = n \) is an intersection of \( n \) hyperplanes and the set \( \mathcal{H}_m \). Moreover the closure of a spectrum in Euclidian space is a convex hull whose vertices are rational points.

**Proposition 3.2.9.** a) A pot type is weakly satisfiable if and only if it admits a nonempty spectrum.

b) The closure in Euclidian space of the spectrum \( S(P) \) of a pot type \( P \) is a convex hull: if \( u, v \in S(P) \), and if \( z \in [0, 1] \), then \( z u + (1 - z) v \) is in the closure of \( S(P) \).

**Proof.**

a) Let \( P \) be a weakly satisfiable pot type and \( C = \langle T, S, J \rangle \) be a complete complex \( (T \neq \emptyset) \) assembled from tiles of types in \( P \). Denote by \( k_i \) the number of tiles in \( T \) of type \( t_i \) and let \( r_i = \frac{k_i}{|T|} \) (Note \( |T| > 0 \) and \( r_i \geq 0 \)). Obviously \( \sum_{i=1}^{m} r_i = 1 \). Since \( C \) is a complete complex, for every \( h \in H \) there are as many sticky ends of type \( h \) among the tiles in \( T \) as of type \( \hat{h} \). Consequently \( \sum_{i=1}^{m} k_i t_i(h) = \sum_{i=1}^{m} k_i t_i(\hat{h}) \), or by dividing both sides of the equality by \( |T| \), it follows that \( \sum_{i=1}^{m} r_i t_i(h) = \sum_{i=1}^{m} r_i t_i(\hat{h}) \).
From the definition of the spectrum it follows that the vector \( \mathbf{r} = (r_1, r_2, \ldots, r_m) \) is an element of the spectrum.

Conversely, if \( S(P) \neq \emptyset \), there exists a nonzero vector \( \mathbf{r} = (r_i : t_i \in P) \) of rational numbers in \( S(P) \). Each coordinate of \( \mathbf{r} \) can be written as \( r_i = \frac{q_i}{d_i} \) for \( q_i \geq 0 \) and \( d_i > 0 \) both integers. Denote by \( d = \text{lcm}(d_1, d_2, \ldots, d_m) \) and \( p_i d_i = d \) for an integer \( p_i \). Thus \( d \mathbf{r} = (dr_1, dr_2, \ldots, dr_m) = (p_1 q_1, p_2 q_2, \ldots, p_m q_m) \). A (not necessarily connected) complex that has \( p_i q_i \) tiles of type \( t_i \), for \( i \in [m] \) is a complete complex.

b) Follows immediately from the fact that the spectrum is a convex hull.

**Proposition 3.2.10.** The spectrum \( S(P) \) of a given pot \( P \) is either empty, a singleton or an infinite set.

**Proof.** The spectrum of an unsatisfiable pot is empty (Proposition 3.2.9). Since the spectrum is dense in a convex hull and includes the vertices of that convex hull, if it contains two points then it contains at least two vertices, and hence every rational point between those vertices, so the spectrum is infinite.

**Proposition 3.2.11.** Let \( P = \{t_1, t_2, \ldots, t_m\} \) be a pot type and \( S(P) \) its spectrum. For every extreme point, \( \mathbf{s} = \left( \frac{k_1}{d_1}, \frac{k_2}{d_2}, \ldots, \frac{k_m}{d_m} \right) \) of \( S(P) \), there exists a complete connected complex \( C = \langle T, S, J \rangle \in C(P) \) with \( d \frac{k_i}{d_i} \) tiles of type \( t_i \), for \( i \in [m] \) where \( d = \text{lcm}(d_1, d_2, \ldots, d_m) \) and \( \gcd(k_j, d_j) = 1 \), for \( j \in [m] \).

**Proof.** The ratio of the tiles of the complex \( C = \langle T, S, J \rangle \) consisting of \( d \frac{k_i}{d_i} \) tiles of type \( t_i \), for \( i \in [m] \) is \( \mathbf{s} = \left( \frac{k_1}{d_1}, \frac{k_2}{d_2}, \ldots, \frac{k_m}{d_m} \right) \) (Note that, \( \sum_{i=1}^m \frac{k_i}{d_i} = 1 \)). Consequently the number of tiles in the complex is

\[
|T| = \sum_{i=1}^m d \frac{k_i}{d_i} = d \sum_{i=1}^m \frac{d_i}{k_i} = d.
\]
Since \( s \in S \) (and from the definition of the spectrum), \( C \) is a complete complex. Next we have to show that \( C \) is a connected complete complex.

Assume that \( C \) is not a connected complete complex. Without loss of generality, we can assume that \( C \) consists of two nonempty complete complexes \( C_1 = \langle T_1, S_1, J_1 \rangle \) and \( C_2 = \langle T_2, S_2, J_2 \rangle \) with \( p_i \) and \( q_i \) the numbers of tiles of type \( t_i \), respectively. The spectrum points corresponding to \( C_1 \) and \( C_2 \) are \( s_1 = \left( \frac{p_1}{|T_1|}, \frac{p_2}{|T_1|}, \ldots, \frac{p_m}{|T_1|} \right) \) and \( s_2 = \left( \frac{q_1}{|T_2|}, \frac{q_2}{|T_2|}, \ldots, \frac{q_m}{|T_2|} \right) \). Since \( C \) consists of \( C_1 \) and \( C_2 \), then \( |T| = |T_1| + |T_2| \), and since both \( C_1 \) and \( C_2 \) are nonempty, \( |T_1| < d \) and \( |T_2| < d \).

First, let show that \( s_1 \neq s_2 \).

If \( \frac{p_i}{|T_1|} = \frac{q_i}{|T_2|} \) for each \( i \), then \( \frac{k_i}{d_i} = \frac{p_i}{|T_1|} + \frac{q_i}{|T_2|} = \frac{p_i + |T_2| p_i}{|T_1| + |T_2|} = \frac{p_i}{|T_1|} \), so \( k_i = \frac{p_i}{|T_1|} = \frac{q_i}{|T_2|} \), so \( s = \left( \frac{p_1}{|T_1|}, \ldots, \frac{p_m}{|T_1|} \right) \).

Since \( \gcd(k_i, d_i) = 1 \), then either \( \gcd(p_i, |T_1|) = 1 \) or \( \gcd(p_i, |T_1|) = r_i > 1 \). In the first case, when \( \gcd(p_i, |T_1|) = 1 \), \( p_i = k_i \) and \( d_i = |T_1| \), for every \( i \in [m] \). So, \( d_1 = d_2 = \ldots = d_m = d \), i.e., \( |T_1| = d \), which is contradicts with \( |T_1| < d \). In the second case, when \( \gcd(p_i, |T_1|) = r_i \), \( p_i = r_i k_i \) and \( |T_1| = r_i d_i \), for every \( i \in [m] \). Therefore, \( d_i \mid |T_1| \) for every \( i \), so \( \text{lcm}(d_1, d_2, \ldots, d_m) \mid |T_1| \), i.e., \( d \mid |T_1| \), so it contradicts \( |T_1| < d \). Hence, \( s_1 \neq s_2 \).

Next we will show that \( s \) can be written as a convex combination of \( s_1 \) and \( s_2 \), which contradicts the fact that \( s \) is an extreme point.

The convex combination on the points \( s_1 \) and \( s_2 \)

\[
\frac{|T_1|}{|T_1| + |T_2|} \left( \frac{p_1}{|T_1|}, \ldots, \frac{p_m}{|T_1|} \right) + \frac{|T_2|}{|T_1| + |T_2|} \left( \frac{q_1}{|T_2|}, \ldots, \frac{q_m}{|T_2|} \right) = \left( \frac{p_1 + q_1}{|T_1| + |T_2|}, \ldots, \frac{p_m + q_m}{|T_1| + |T_2|} \right)
\]

is equal to \( s \) (Note \( C \) has \( p_i + q_i \) tiles of type \( t_i \) and \( |T_1| + |T_2| \) number of tiles). This contradicts the fact that \( s \) is an extreme point. Therefore, \( C \) must be connected.

\[
\Box
\]

**Remark 3.2.12.** To every extremal point \( s = \left( \frac{k_1}{d_1}, \frac{k_2}{d_2}, \ldots, \frac{k_m}{d_m} \right) \), there might be more that one complex with \( \frac{k_i}{d_i} \) tiles of type \( t_i \), depending on the number of possible con-
Connections between the tiles. Similarly, there might be more than one complete complex associated with a given point from the spectrum of a given pot type.

**Definition 3.2.13.** Let \( P = \{t_1, t_2, \ldots, t_m\} \) be a pot type with its spectrum \( S \).

Let \( s_i = \left( \frac{k_{i1}}{d_{1i}}, \frac{k_{i2}}{d_{2i}}, \ldots, \frac{k_{im}}{d_{mi}} \right) \) for \( i \in [l] \) be the extreme points of \( S \) and let \( d_i = \text{lcm}(d_{1i}, d_{2i}, \ldots, d_{mi}) \).

The set \( S_i \) of complete complexes \( C \) consisting of \( d_{j}^{k_{ij}} \) tiles of types \( t_i \) is called the set of extreme complete complexes corresponding to \( s_i \) and the complexes are called extremal complexes.

**Definition 3.2.14.** A complex \( C = \langle T, J \rangle \) is called a minimal complete complex if there does not exist a complete complex \( C' = \langle T', J' \rangle \) with \( T' \subseteq T \) and \( T \neq T' \).

From Proposition 3.2.11 follows that every extreme complex is also a minimal complete complex. But the converse is not necessarily true, as we will see in Example 3.2.16.

**Proposition 3.2.15.** Let \( P = \{t_1, t_2, \ldots, t_m\} \) be a pot type, \( S(P) \) the spectrum of \( P \) with \( l \) extreme points, and \( C_1, C_2, \ldots, C_l \) be extreme complete complexes for the pot type corresponding to the extreme points. The vector of the number of tiles of every complete complex in \( C(P) \) is a linear combination of the vectors of the number of tiles of the extremal complete complexes.

**Proof.** Let \( C = \langle T, S, J \rangle \) be a complete complex built from tiles of types in \( P \), let \( r = (r_1, r_2, \ldots, r_m) \) be the vector of the ratios of the tile types in \( C \) and let the corresponding vector of the numbers of tiles be \( c = (c_1, c_2, \ldots, c_m) = |T|r \).

If \( r \) is an extreme point of \( S(P) \), then the number of tiles of \( C \) is a multiple of the number of tiles of the extremal complete complexes corresponding to that extreme point.

Let \( r \) be a non-extreme point of \( S(P) \), then \( r \) can be written as a convex combination of the extreme points in \( S(P) \), say \( r = \sum_{i=1}^{l} \lambda_i r_i \), for \( r_i \) ranging over the extreme points of \( S(P) \) and \( \sum_{i=1}^{l} \lambda_i = 1 \).
For $i \in [l]$, $\mathbf{r}_i = (r_{i1}, r_{i2}, \ldots, r_{im})$, where $r_{ij} \geq 0$ for $j \in [m]$ and $\sum_{j=1}^{m} r_{ij} = 1$. Let $r_{ij} = \frac{x_{ij}}{d_j}$ and $d_i = \text{lcm}(d_{i1}, d_{i2}, \ldots, d_{im})$, then $d_i \mathbf{r}_i = (d_{i1}r_{i1}, d_{i2}r_{i2}, \ldots, d_{im}r_{im}) = (c_{i1}, c_{i2}, \ldots, c_{im})$ and $\sum_{j=1}^{m} d_{ij}r_{ij} = d_i$, where $(c_{i1}, c_{i2}, \ldots, c_{im})$ is the vector corresponding to the extreme point $\mathbf{r}_i$. With other words, the minimal complete complex has $d_i$ tiles.

Denote $D = \text{lcm}(d_1, d_2, \ldots, d_l)$. For each $i \in [l]$, we can write $D = d_im_i$, for an $m_i \in \mathbb{N}$. The spectrum $\mathcal{S}(\mathbf{P})$ consists of points with rational coordinates, say $\lambda_i = \frac{p_i}{q_i}$, $i \in [l]$. Denote by $q$ the lcm$(q_1, q_2, \ldots, q_k)$, i.e., $q = s_iq_i$, for appropriate $s_i \in \mathbb{N}$. $i \in [l]$. Then,

\[
\mathbf{c} = |T|\mathbf{r} = \sum_{i=1}^{l} |T|\lambda_i(r_{i1}, r_{i2}, \ldots, r_{im}) = \sum_{i=1}^{l} |T| Dq \frac{p_i}{q_i} (r_{i1}, r_{i2}, \ldots, r_{im}) = \sum_{i=1}^{l} |T| d_im_is_iq_i \frac{p_i}{q_i} (r_{i1}, r_{i2}, \ldots, r_{im}) = \sum_{i=1}^{l} |T| p_is_im_i (r_{i1}d_i, r_{i2}d_i, \ldots, r_{im}d_i) = \sum_{i=1}^{l} |T| p_is_im_i d_i (c_{i1}, c_{i2}, \ldots, c_{im}).
\]

The vector of the number of tiles in $C$ can be written as a linear combination of the vectors of the number of tiles for the extremal complete complexes.

The previous proposition states that the vector of the number of tiles of every complete complex could be written as a linear combination of the vectors of the
number of tiles of the extreme complexes. This linear combination is not necessarily an integer combination. For a given pot type $P$, the set of complete complexes, $C(P)$ can be characterized through the extreme complete complexes. Not all minimal complexes are extreme complexes as we can see in Figure 3.1.

**Example 3.2.16.** Consider the Pot type from Figure 2.9 c. Its spectrum is the set of solutions to the following system of equations.

\[
\begin{align*}
r_1 + r_2 + r_3 &= 1 \\
3r_1 - 2r_2 - r_3 &= 0.
\end{align*}
\]

Hence, the spectrum is $S(P) = \{(u, 4u - 1, 2 - 5u) : \frac{1}{4} \leq u \leq \frac{2}{3}\}$. To find the extreme points, we need to find the basic feasible solution for the system given above, i.e., to find solutions to the following systems.

\[
\begin{bmatrix} 1 & 1 \\ 3 & -2 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 3 & -1 \end{bmatrix} \begin{bmatrix} r_1 \\ r_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ -2 & -1 \end{bmatrix} \begin{bmatrix} r_2 \\ r_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}
\]

for $r_1 \geq 0$, $r_2 \geq 0$, $r_3 \geq 0$.

The extreme points for the spectrum are $s_1 = \left(\frac{1}{4}, 0, \frac{3}{4}\right)$ and $s_2 = \left(\frac{2}{5}, \frac{3}{5}, 0\right)$. The minimal complete complexes for this pot type are given in the figure below.

![Figure 3.1](image)

**Figure 3.1:** a) The minimal complete complexes for the pot type given in Figure 2.9 c b) A complete complex that is minimal, but not extremal complete complex.
The spectrum point corresponding to $C_1$ is $\left(\frac{1}{4}, 0, \frac{3}{4}\right)$. The spectrum point corresponding to $C_2$ is $\left(\frac{2}{5}, \frac{3}{5}, 0\right)$. The spectrum point corresponding to $C_3$ is $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.

Although all three complexes are minimal one, $C_1$ and $C_2$ are extreme complexes, while $C_3$ is not.

**Corollary 3.2.17.** If the spectrum consists of only one point, $S(P) = \{(r_{t_1}, r_{t_2}, \ldots, r_{t_m})\}$, and if $r_{t_k} > 0$ for some $k \in [m]$, then every complete complex in $C(P)$ contains a tile of type $t_k$. Moreover, if $r_{t_k} > 0$ for all $k \in [m]$, $P$ is strongly satisfiable.

**Proof.** If the spectrum consists of only one point, then that point must be an extreme point. From Proposition 3.2.11 it follows that a minimal complete complex corresponding to the extreme point is connected and contains tiles of types $t_k$ for which $r_{t_k} > 0$. Consequently, every other complete complex is a linear combination of these complexes, i.e., the number of tiles of type $t_k$, for $k \in [m]$, of any other complete complex is a multiple of the number of tiles of type $t_k$ on the complexes corresponding to the extreme point.

If $r_{t_k} > 0$ for all $k \in [m]$, then the complete complex corresponding to the extreme point from the spectrum contains tiles of each type, so $P$ is strongly satisfiable.

From Corollary 3.2.17 we can conclude that if the spectrum $S(P)$, of a given pot type $P$, consists of only one point whose coordinates are positive, then every complete complex in $C(P)$ contains tiles of each type. Now let’s consider examples for the spectra of strongly satisfiable pot types.

**Example 3.2.18.** Consider the pot types depicted in Figure 2.9. Their spectra can be computed similarly as in Example 3.1.2. The spectrum of the pot type in Figure 2.9 (a) is $S = \{(1/2, 1/2)\}$, the spectrum of the pot type in Figure 2.9 (b) is $S = \{(1/2, 1/2)\}$, while the one for Figure 2.9 (c) is $S = \{(u, 4u - 1, 2 - 5u) : u \in \mathbb{R}, \frac{1}{4} \leq u \leq \frac{2}{5}\}$.

In two-dimensional space (corresponding to a pot type with exactly two tile types) the spectrum is a part of the line segment $(r_1 + r_2 = 1, 0 \leq r_1 \leq 1, 0 \leq r_2 \leq 1)$.
connecting the points \((0, 1)\) and \((1, 0)\). So the spectrum is either a point of that line segment, or it is the entire line segment, or it is the empty set. The spectrum is the entire line segment if and only if complementary sticky ends of same type on a single tile is allowed. Because, in this case the system of equations and inequalities becomes \(r_1 + r_2 = 1, 0 \leq r_1 \leq 1, 0 \leq r_2 \leq 1\), i.e., when a tile of the first type forms a complete complex, and a tile of the second type forms a complete complex. Since we do not allow that, the spectrum in two-dimensional case will always consist of only one point.

**Proposition 3.2.19.** The spectrum, \(S(P)\), of the pot type \(P = \{t_1, t_2, \ldots, t_m\}\) is a convex hull with vertices \(\{(1,0,0,\ldots,0), (0,1,0,\ldots,0), \ldots, (0,0,\ldots,0,1)\}\) if and only if each tile from a type in \(P\) is a complete complex.

*Proof.* Assume the spectrum \(S(P)\), of the pot type \(P = \{t_1, t_2, \ldots, t_m\}\) is a convex hull with vertices \(\{(1,0,0,\ldots,0), (0,1,0,\ldots,0), \ldots, (0,0,\ldots,0,1)\}\). From Proposition 3.2.15 follows that for each \(i \in [m]\) the complex \(C_i\) consisting of a tile of type \(t_i\) is connected complete complex. Because of that each tile from a type in \(P\) can form a complete complex.

Assume that each tile from a type in \(P\) is a complete complex. Then, for each \(h \in H\) \(z_{t_i}(h) = 0\), and hence we do not get any equation from the second condition in the definition of the spectrum. So, the spectrum of \(P\) is: \(S(P) = \{(r_1, r_2, \ldots, r_m) : r_i \geq 0, \text{ for } i \in [m] \text{ and } \sum_{i=1}^{m} r_i = 1\}\), which is a polytope with vertices \(\{(1,0,0,\ldots,0), (0,1,0,\ldots,0), \ldots, (0,0,\ldots,0,1)\}\).  

**Example 3.2.20.** All three examples have a three tile type pot type \(P = \{t_1, t_2, t_3\}\), and the set of sticky end types \(H = \{a, b, \hat{a}, \hat{b}\}\).

\[
z_{t_1} = (2, 1), \ z_{t_2} = (-1, 1), \ z_{t_3} = (0, -3),
\]
\[z_{t_1} = (3), \quad z_{t_2} = (-2), \quad z_{t_3} = (-1),\]

\[z_{t_1} = (0,0), \quad z_{t_2} = (0,0), \quad z_{t_3} = (0,0)\]

Figure 3.2: a) Strongly satisfiable pot type with spectrum \(\{(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})\}\), b) Strongly satisfiable pot type with spectrum \(S(\mathbf{P}) = \{(u, 4u - 1, 2 - 5u) : \frac{1}{4} \leq u \leq \frac{2}{5}\}\). c) Strongly satisfiable pot type with spectrum \(\{(1 - u - v, u, v) : 0 \leq u \leq 1, 0 \leq v \leq 1, u + v \leq 1\}\).

Note the spectrum contains vectors with rational entries, its closure in Euclidean space is bounded and therefore compact subset of \(\mathbb{R}^m\).

Figure 3.3: The closure of the spectrum of the pot type given in Example 3.2.20 b) is the line segment; the closure of the spectrum of the pot type given in Example 3.2.20 c) is the triangle bounded by the dotted lines along with its interior.
3.3 Algebraic Representation of the Spectrum

The spectrum is the intersection of $n$ hyperplanes (for each $h$, $\sum_{t \in P} r_t z_t(h) = 0$) and $H_m$. Hence it is the solution of $n$ homogeneous and $1$ non-homogeneous equations with $m$ variables over $\mathbb{Q}^+$. 

$$
\begin{align*}
    r_1 + r_2 + \cdots + r_m &= 1 \\
    z_{t_1}(h_1)r_1 + z_{t_2}(h_1)r_2 + \cdots + z_{t_m}(h_1)r_m &= 0 \\
    z_{t_1}(h_2)r_1 + z_{t_2}(h_2)r_2 + \cdots + z_{t_m}(h_2)r_m &= 0 \\
    \vdots & \quad \vdots \\
    z_{t_1}(h_n)r_1 + z_{t_2}(h_n)r_2 + \cdots + z_{t_m}(h_n)r_m &= 0.
\end{align*}
$$

(3.3.3)

An efficient way to solve this system is by the Gauss-Jordan elimination, which transforms the augmented matrix of system (3.3.3) into the row-echelon form. The computational complexity of solving this system with the aid of Gauss-Jordan elimination is $O(m^2n)$.

From Example 3.1.2 and Example 3.2.18 it can be seen that for satisfiable and weakly satisfiable pots (but not necessarily strongly satisfiable pots) vectors of the spectrum may have zero coordinates. If the spectrum of strongly satisfiable pot types is a singleton, then all of its coordinates are positive numbers (Proposition 3.3.2).

**Definition 3.3.1.** Let $A$ be a set of $n$ dimensional vectors. The support of $A$ is the set $\text{supp}(A) = \{i \in [n] : \text{there exists a vector } u = (u_1, u_2, \ldots, u_n) \in A \text{ such that } u_i \neq 0\}$. In other words, if $i \notin \text{supp}(A)$, then the $i^{th}$ coordinate of every point in $A$ is $0$.

**Proposition 3.3.2.** Suppose $S(P)$ is the spectrum of a given pot type $P$ with $|P| = m$

a) $\text{supp}(S(P)) = [m]$ if and only if $P$ is strongly satisfiable.

b) $\emptyset \neq \text{supp}(S(P)) \subsetneq [m]$ if and only if $P$ is weakly satisfiable but not strongly satisfiable.

**Proof.** a) $\text{supp}(S(P)) = [m]$ if and only if every tile has a positive probability of being on a complete complex, i.e., every tile type occurs on a complete complex,
which means that the pot type is strongly satisfiable.

b) $\text{supp}(\mathcal{S}(\mathbf{P})) \subsetneq [m]$ if and only if there is a coordinate that is zero in every vector of the spectrum, i.e., at least one tile type cannot be embedded into a complete complex, so the pot is not strongly satisfiable, but it is weakly satisfiable since it has a nonempty spectrum.

\[ l_t(h) = \begin{cases} 
1 & \text{if } t(h) \geq 1 \text{ or } t(\hat{h}) \geq 1 \\
0 & \text{otherwise}
\end{cases} \]

are called sticky ends vectors.

We denote by $l_h[i]$ the $i^{th}$ coordinate of the vector $l_h$.

**Definition 3.3.3.** Let $\mathbf{P}(H, \theta) = \{t_1, \ldots, t_m\}$ be a pot type. The $m$ dimensional vectors $l_h = (l_1(h), l_2(h), \ldots, l_m(h))$ such that such that

**Proposition 3.3.4.** Classification of pot types into weak satisfiability, satisfiability and strong satisfiability is in $\text{PTIME}$.

**Proof.** Let $\mathbf{P}$ be a pot type with $m$ tile types and $n$ sticky end types. In order to obtain the spectrum for the given pot we need to solve system (3.3.3) of $n + 1$ equations with $m$ variables. If there is a solution with all positive coordinates, then the spectrum is nonempty and from Proposition 3.3.2 it follows that the pot is strongly satisfiable. If there is a solution to the system (3.3.3) and $\text{supp}(\mathcal{S}(\mathbf{P})) \subsetneq [m]$, then the spectrum is nonempty and from Propositions 3.2.9 and 3.3.2 follows that the pot is weakly satisfiable but not strongly satisfiable. Therefore weak satisfiability and strong satisfiability are in $\text{PTIME}$.

Now suppose that the pot is weakly satisfiable but not strongly satisfiable, i.e., $\text{supp}(\mathcal{S}(\mathbf{P})) \subsetneq [m]$.

Consider the sticky end vectors $l_h = (l_t(h) : t \in \mathbf{P})$ for $\mathbf{P}$. If there exists a sticky end $h \in H$ such that $l_h[i] = 0$ for all $i \in \text{supp}(\mathcal{S}(\mathbf{P}))$, then $h$ could not be embedded
into any complete complex, hence the pot is weakly satisfiable, but not satisfiable (otherwise the pot is satisfiable). Consequently, if \( \text{supp}(\mathcal{S}(P)) \subseteq [m] \) and there exists an \( h \in H \) for which \( \sum_{i=1}^{m} l_i(h)s_i = 0 \) (where \( \mathcal{S}(P) = \{(s_1, s_2, \ldots, s_m)\} \)), then the pot type is weakly satisfiable. To check that one needs to form the dot products between the \( l \) vectors and the spectrum. If one of the dot product is 0, then the pot type is not satisfiable.

Hence, the computational complexity for classifying pot types according to their type if satisﬁability is \( O(m^2n) + O(mn) = O(m^2n) \), i.e it is in PTIME.

\[ \text{Corollary 3.3.5.} \ A \text{ pot type } P(H, \theta) \text{ is satisfiable if and only if for every } h \in H \ l_h[i] = 1 \text{ for every } i \in \text{supp}(\mathcal{S}(P)). \]

\[ \text{Example 3.3.6.} \ The \text{ pot types given in Figure 2.8 have same spectrum } \mathcal{S}(P) = \left(\frac{1}{2}, \frac{1}{2}, 0, 0\right) \text{ (therefore } \text{supp}(P) = \{1, 2\} \text{), although one of them (Figure 2.8 a)) is satisfiable, while the other (Figure 2.8 b) ) is only weakly satisfiable. Proposition 3.3.4 helps to classify them.} \]

For the pot type given in Figure 2.8 a) the \( l \) vectors are: \( l_a = (1, 1, 1, 0) \), \( l_b = (1, 1, 0, 1) \). The dot products between the spectrum and the \( l \) vectors are: \( \mathcal{S}(P) \cdot l_a = 1 \), and \( \mathcal{S}(P) \cdot l_b = 1 \). Therefore, this pot type is satisfiable.

For the pot type given in Figure 2.8 b) the \( l \) vectors are: \( l_a = (1, 1, 1, 1) \), \( l_b = (1, 1, 0, 0) \), and \( l_c = (0, 0, 1, 1) \). The dot products between the spectrum and the \( l \) vectors are: \( \mathcal{S}(P) \cdot l_a = 1 \), \( \mathcal{S}(P) \cdot l_b = 1 \), but \( \mathcal{S}(P) \cdot l_c = 0 \). Therefore, this pot type is not satisfiable.

\[ \text{3.4 Maple Program} \]

We wrote a Maple Program that computes the spectrum of a given pot type, the support, and that classify the pot type in one of the four classes. The program is given in the Appendix A. To describe the program, we will give an example the way program works for the Figure 3.2 a). For \( m \) we input the number of tile types, and
for $n$ we input the number of sticky end types. First we form a matrix $a = [a_{s,t}]$ is a matrix of size $(n + 1) \times (m + 1)$, where $a_{s,t} = z_{t,s} m b h_{t-1}$ for $s = 2, \ldots, n + 1, t \in [m]$, $a_{1,i} = 1$ for $i \in [m + 1]$, $a_{i,m+1} = 0$ for $i = 2, \ldots, n + 1$. We also need a matrix $L = [l_{s,t}]$ of size $n \times m$ for the sticky ends vectors, i.e., $l_{s,m} = 1$ if $t_m(h_s) > 0$ or $t_m(h_s') > 0$, otherwise $l_{s,m} = 0$. For the Figure 3.2, that matrices are:

$$a = \begin{bmatrix}
1 & 1 & 1 & 1 \\
2 & -1 & 0 & 0 \\
1 & 1 & -3 & 0
\end{bmatrix},$$

$$L = \begin{bmatrix}
1 & 1 & 0 \\
1 & 1 & 1
\end{bmatrix}.$$

The Maple program calculates the support of the pot, calculates the spectrum of the pot and classify the pot types.
4 Graph of a Pot with DNA complexes

4.1 Introduction

In this chapter we address several structural questions about assembled complexes. Some include: What kind of complexes can result from a given pot type? Could two different pot types have the same set of complete complexes? What kind of relations could be defined on the set of pot types? To help answer these as well as other questions, we approach self-assembly from graph theoretical point of view.

To every tile type from a pot of DNA molecules we assign a labeled multigraph, and we assign a labeled multigraph to the pot of DNA molecules and to every complex which can be produced from self-assembly. The main idea is to classify the type of complexes that can appear in a given pot.

We compare two pots types according to their tile types and according to the complete complexes that can be assembled. First we give definitions that we will use through the chapter.

Definition 4.1.1. A labeled multigraph $G$ is a quadruple $G = (V, E, l, L)$, where $V$ is the set of vertices, $E$ is the set of edges, $L$ is the set of labels, and $l$ is the labeling function $l : E \to L$ that assigns label to every edge. To every edge $e \in E$, we assign the vertex set of $e$, $\text{vs}(e)$, defined as the set of two vertices incident to the edge and the label $l(e)$.

Definition 4.1.2. For a given multigraph $G = (V, E, l, L)$, the degree of a vertex $w \in V$ with respect to the label $a$, is defined as

$$\deg(w, a) = |\{e \in E : l(e) = a \text{ and } w \in \text{vs}(e)\}|.$$
Definition 4.1.3. The labeled multigraph \( G_1 = (V_1, E_1, l_1, L_1) \) is homomorphic to the labeled multigraph \( G_2 = (V_2, E_2, l_2, L_2) \) if there exists a homomorphism \( h : V_1 \cup E_1 \cup L_1 \to V_2 \cup E_2 \cup L_2 \) such that \( h|_{V_1} = h_V : V_1 \to V_2 \), \( h|_{E_1} = h_E : E_1 \to E_2 \) and \( h|_{L_1} = h_L : L_1 \to L_2 \) is such that \( vs(h_E(e)) = h_V(vs(e)) \) for every \( e \in E_1 \), i.e., for \( u, w \in V_1 \) and \( e \in E_1 \), \( \{h_V(u), h_V(w)\} = vs(h_E(e)) \) whenever \( \{u, w\} = vs(e) \), and \( l(h_E(e)) = h_L(l(e)) \).

To ease the notation, for the homomorphisms between the graphs \( G_1 \) and \( G_2 \) we will use the notation \( h : G_1 \to G_2 \).

### 4.2 Definition of a Pot Graph

**Definition 4.2.1.** Let \( P(H, \theta) \) be a pot type. Define the pot graph \( G_P \) of \( P(H, \theta) \) as a labeled multigraph \( G_P = (V, E, l, H^+) \) as follows. Let \( V = \{v_t : t \in P\} \) be the set of vertices, \( E \) be the set of edges, and \( l : E \to H^+ \) be the set of labels with the following proviso: for each sticky end type \( h \in H^+ \) and each pair of tile types \( s, t \in P(H, \theta) \), \( s(h) > 0 \) and \( t(\hat{h}) > 0 \), there exists an edge \( e \in E \) with \( vs(e) = \{v_s, v_t\} \) and \( l(e) = h \).

**Remark 4.2.2.** For each \( s, t \in P(H, \theta) \) and each \( h \in H^+ \), in the pot graph \( G_P \) there is at most one edge connecting \( v_s, v_t \) with label \( h \).

Note that when the pot type is \( P = \{t_1, t_2, \ldots, t_m\} \), the corresponding pot graph, the set of vertices is defined as \( V = \{v_i : t_i \in P\} \). And since it known that the set of labels is \( H^+ \), we will omit it from the definition of the pot graph, i.e., from now on the pot graph will be denoted as \( G_P = (V, E, l) \).

**Example 4.2.3.** Figure 4.1 is an example of a pot type and its pot graph.
It follows from the above definitions that for a given tile $t \in \mathbf{P}$, if $t(h) > 0$ for $h \in H^+$ then its corresponding vertex $v_t$ will be adjacent to all vertices of different tile types $v_s$ for which $s(\hat{h}) > 0$, or if $t(\hat{h}) > 0$ its corresponding vertex $v_t$ will be adjacent to all vertices of different tile types $v_s$ for which $s(h) > 0$. Therefore for each $h \in H^+$

$$\deg(v_t, h) = \left( \sum_{s \in \mathbf{P}} I_s(\hat{h}) \right) I_t(h) + \left( \sum_{s \in \mathbf{P}} I_s(h) \right) I_t(\hat{h}),$$

where

$$I_t(h) = \begin{cases} 1 & t(h) > 0 \\ 0 & \text{otherwise}. \end{cases}$$

From the assumption that there is no tile with complementary sticky ends of types $h$ and $\hat{h}$, for every $h \in H$, in the above definition, either $I_t(h) = 0$ or $I_t(\hat{h}) = 0$.

For a given pot type $\mathbf{P}$ and its corresponding pot graph $\mathcal{G}_\mathbf{P}$, for each $t \in \mathbf{P}$, $\deg(v_t, h)$ does not depend on the number of sticky end types of type $h$ on $t$, i.e., does not depend on $|z_t(h)|$. The pot type in Figure 4.2 is strongly satisfiable, and $|z_{t_1}(h)| < \deg(v_{t_1}, h)$, $|z_{t_2}(h)| = \deg(v_{t_2}, h)$, and $|z_{t_3}(h)| > \deg(v_{t_3}, h)$. 
Remark 4.2.4. Not every labeled multigraph is a pot graph. The following example confirms that.

Without loss of generality let us assume that $t_1(h) > 0$, from the given pot type. Since there is an edge with vertex set $\{v_1, v_2\}$ labeled $h$, and there is an edge with vertex set $\{v_1, v_3\}$ labeled $h$, and since there is no edge with vertex set $\{v_1, v_4\}$ labeled $h$, we can conclude that $t_2(h) > 0$, $t_3(h) > 0$, and $t_4(h) > 0$. That means that there should exist an edge with vertex set $\{v_2, v_4\}$ labeled $h$ and an edge with vertex set $\{v_3, v_4\}$ labeled $h$. Since there is no edge with vertex set $\{v_2, v_4\}$ labeled $h$, this multigraph cannot be a pot graph.

With the following Proposition we classify the pot graphs, we give a necessary and sufficient condition for a graph to be a pot graph.

Proposition 4.2.5. A labeled multigraph $G = (V, E, l)$ ($l : E \to H^+$) is a pot graph
if and only if for every $h \in H^+$ the subgraph spanned by the edges with label $h$ is a complete bipartite graph.

Proof. Let $G_P = (V, E, l)$ be a pot graph. For every $h \in H^+$ construct two subsets of $V$, $V_h = \{v_t : t \in P, t(h) > 0\}$ and $\hat{V}_h = \{v_t : t \in P, t(\hat{h}) > 0\}$. By the assumption that no tile has sticky ends of type $h$ and $\hat{h}$ at the same time, $V_h$ and $\hat{V}_h$ are disjoint. In $G_P$ every vertex of $V_h$ is incident to every vertex of $\hat{V}_h$ by an edge labeled $h$ and there are no other edges labeled $h$ in $G_P$. No two vertices of $V_h$ (or $\hat{V}_h$) are adjacent in the subgraph spanned by the edges with label $h$. Therefore, if we denote with $E_h = \{e : e \in E$ and $l(e) = h\}$, then the graph $G = (V_h \cup \hat{V}_h, E_h)$ is a complete bipartite graph.

Conversely, suppose that $G = (V, E, l)$ is a labeled graph (with finite set of vertices and edges) and for every $h \in H^+$ the subgraph spanned by the edges with label $h$ is a complete bipartite graph $G_h = (V_h \cup \hat{V}_h, E_h)$, where $V_h \cap \hat{V}_h = \emptyset$. We define a pot type $P$ that has as many tile types as vertices in $V$. Since the set of vertices is finite, say $|V| = m$, we can number the vertices of $V = \{v_i : i \in [m]\}$. To every vertex $v_i \in V$ we assign a tile type $t_i$ defined in the following way: for every $h \in H^+$, if $v_i \in V_h$ then $t_i(h) = \deg(v_i, h)$, and if $v_i \in \hat{V}_h$ then $t_i(\hat{h}) = \deg(v_i, h)$, otherwise $t_i(h) = t_i(\hat{h}) = 0$. Consider the pot graph $G_P = (V_P, E_P, l_P)$ of $P$. The set of vertices for the pot graph is $V_P = \{v_i : i \in P, i \in [m]\}$, therefore $V_P = V$. For every $h \in H^+$, if $t_i(h) \geq 1$ and $t_j(\hat{h}) \geq 1$, there will be an edge $e \in E_P$ with $\text{vs}(e) = \{v_i, v_j\}$ and $l(e) = h$. From the definition of the sets $V_h$ and $\hat{V}_h$ it follows that where $v_i \in V_h$ and $v_j \in \hat{V}_h$, i.e., $e \in E_h$. Hence, $E_P \subseteq E_h$.

Suppose a sticky end $h \in H^+$ is given, and an edge $e \in E_h$ with $\text{vs}(e) = \{v_i, v_j\}$. Since $G_h$ is a complete bipartite graph it will be the case that $v_i \in V_h$ and $v_j \in \hat{V}_h$, or $v_i \in \hat{V}_h$ and $v_j \in V_h$. From the definition of the pot it follows that $t_i(h) \geq 1$ and $t_j(\hat{h}) \geq 1$, or $t_i(\hat{h}) \geq 1$ and $t_j(h) \geq 1$, i.e., there exists an edge in $G_P$ labeled $h$ incident to $v_i$ to $v_j$, i.e, $e \in E_P$. Hence $E_h \subseteq E_P$, from where it follows that $G = G_P$. \[\blacksquare\]
Suppose we are given a complex $C = \langle T, S, J \rangle$, where $T$ is the set of tiles, $S$ is the set of sticky ends, and $J$ is the set of connections. We define a function $\text{type} : T \cup S \rightarrow P \cup H$ such that $\text{type}(h) = h$ if the sticky end $h \in S$ is of type $h \in H$ and $\text{type}(t) = t$ is the tile $t$ is of type $t \in P$. Besides assigning pot graphs to a given pot type, we also assign graphs to complexes.

**Definition 4.2.6.** Let $C = \langle T, S, J \rangle$ be a complete complex of $C(P)$. Define the complete complex graph $G_C$ of $C$ as a multilabeled multigraph $G_C = (V_C, E_C, l_C)$, as follows. $V_C = \{v_t : t \in T\}$ is the set of vertices, $l_C : E_C \rightarrow H^+$ is the set of labels. The set of edges, $E_C$, is defined such that for every connection $c = \{(t, h), (t', \hat{h})\} \in J$ for $t, t' \in T$, $\text{type}(h) = h \in H^+$, $\text{type}(\hat{h}) = \hat{h} \in H^-$, there exists an edge $e \in E_C$ with $\text{vs}(e) = \{v_t, v_{t'}\}$ and $l(e) = h$.

From the definition of a complete complex graph it follows that for every $t \in T$, $\deg(v_t, h) = |z_{\text{type}(t)}(h)|$.

**Example 4.2.7.** Several elements from the set of complete complex graphs arisen from the pot type given in Example 4.1 are given in Figure 1.4.

![Figure 4.4: Elements from the set of complete complex graphs for the pot type given in Example 4.1](image)

**Definition 4.2.8.** Let $C = \langle T, S, J \rangle$ be a complex over $P(H, \theta)$. Define the complete complex graph $G_C$ of $C$ as a multilabeled multigraph $G_C = (V_C, E_C, l_C)$ with finite set of vertices and edges. The set of vertices $V_C$ is partitioned into a disjoint union $V_C = V_T \cup$
\(V_H,\) where \(V_T = \{v_t : t \in T\}\) and \(V_H = \{v_h : \exists t \in T \text{ for which } t(h) > 0, \ h \in S\) is of type \(h \in H\) and \((t, h) \notin c\) for any \(c \in J\). The set of edges \(E_C,\) which is also partitioned into a disjoint union \(E_C = E_T \cup E_{T,H},\) together with the labeling function \(l_C : E_C \to H^+\) are defined such that

- for every connection \(c = \{(t, h), (t', \hat{h})\} \in J\) for \(t, t' \in T,\) type\((h) = h \in H^+\) and type\((\hat{h}) = \hat{h} \in H^-\), there exists an edge \(e \in E_T\) with \(vs(e) = \{v_t, v_{t'}\}\), and \(l(e) = h\).

- for every \(t \in T\) such that \(t(h) > 0\) and \((t, h) \notin c\) for some sticky end \(h \in S\) of type \(h \in H\) and \(c \in J\), there exists an edge \(e \in E_{T,H}\) with \(vs(e) = \{v_t, v_h\}\), and \(l(e) = h\) if \(h \in H^+\) or \(l(e) = \hat{h}\) if \(h \in H^-\).

**Remark 4.2.9.** Every complete complex graph is a complex graph for which the set of edges \(V_H\) is empty.

**Remark 4.2.10.** The tile graph, \(G_t,\) corresponding to a tile \(t,\) is a complex graph with \(V_T = \{v_t\}\) and \(V_H = \{v_h : \text{type}(h) = h, t(h) > 0\}, E_T = \emptyset\) and for every \(h \in H^+\).

As we mentioned previously, the main motive for modeling pot types with pot graphs was to study the outcomes of the process of self-assembly with tools that we are familiar with and tools that can help us in understanding of the process. The next subsection, we show that the definitions used for pot graphs and complex graphs are very natural, and that can be very easily established homomorphism between a complex graph and pot graph of a same pot type.

### 4.3 Homomorphisms

In this section, we show that the graph of every complete complex from a given pot type \(P(H, \theta)\) is homomorphic to the pot graph of \(P(H, \theta)\). The definition of equivalent and similar pot types are given in this section, and we show that equivalent pot types have isomorphic pot graphs, and equivalent sets of complex graphs. At the end we show that for every pot type \(P\) there exists a similar pot type \(\hat{P}\) (\(P\) and \(\hat{P}\) have isomorphic complex graphs) whose tile types have sticky ends of distinct types.
Lemma 4.3.1. Let $G_P = (V_P, E_P, l_P)$ be the pot graph of the pot $P(H, \theta)$ and let $G_C = (V_C, E_C, l_C)$ be the complete complex graph of a complete complex $C = \langle T, S, J \rangle \in \mathcal{C}(P)$. There exists a homomorphism $\varphi : G_C \to G_P$ defined in the following way:

- $\varphi_V(v_t) = v_{\text{type}(t)}$, for $v_t \in V_C$
- $l(\varphi_E(e)) = l(e)$, for $e \in E_C$.

Proof. $\varphi$ is well defined:

If $v_t = v_{t'}$, for $v_t, v_{t'} \in V_C$ then $t$ and $t'$ must be of the same type, therefore $\varphi_V(v_t) = \varphi_V(v_{t'})$.

If $e = e'$, for $e, e' \in E_C$, then $v_s(e) = v_s(e')$ and $l(e) = l(e')$. Since $\varphi$ is a homomorphism and $\varphi_V$ is well defined it follows that $\varphi_V(v_s(e)) = \varphi_V(v_s(e'))$ and $l(\varphi_E(e)) = l(e) = l(e') = l(\varphi_E(e'))$ i.e., $v_s(\varphi_E(e)) = v_s(\varphi_E(e'))$ and $l(\varphi_E(e)) = l(\varphi_E(e'))$. By Remark 4.2.2, there are no two edges in $G_P$ with the same vertex set and same label. Therefore, $\varphi_E(e) = \varphi_E(e')$.

$\varphi$ is homomorphism: Let $e \in E_C$ with $v_s(e) = \{v_t, v_{t'}\}$ and $l(e) = h$. There exists a connection $c = \{(t, h), (t', \hat{h})\} \in J$, where $h$ is a sticky end of type $h \in H^+$, $\hat{h}$ is a sticky end of type $\hat{h} \in H^-$, $t$ is a tile of type $t \in P$ and $t'$ is a tile of type $t' \in P$. Without loss of generality we may assume that $t(h) > 0$ and $t'(\hat{h}) > 0$, from where it follows that $t(h) > 0$ and $t'(\hat{h}) > 0$ i.e., by the definition of the pot graph it follows that there is an edge $e_P \in E_P$ with $v_s(e_P) = \{v_t, v_{t'}\} = \{\varphi_V(v_t), \varphi_V(v_{t'})\}$ and $l(\varphi_E(e)) = l(e_P) = h = l(e)$.

Example 4.3.2. Consider the pot graph given in Figure 4.1 and one complete complex graph from the Figure 4.4, say the one given in the Figure 4.5.
The homomorphism $\varphi : G_C \to G_P$, defined as: $\varphi(v_1) = v_1$, $\varphi(v_2) = v_2$, $\varphi(v_3) = v_4$, $\varphi(v_4) = v_3$, and $\varphi(e) = e$ for $e \in \{a, b, c\}$, maps the complete complex graph from Figure 4.5 in the pot graph in Figure 4.1.

The following proposition is a characterization of complete complexes, i.e., we give necessary and sufficient condition for a graph to be a complete complex graph.

**Proposition 4.3.3.** Suppose a pot type $P(H, \theta)$ (with its pot graph $G_P = (V_P, E_P, l_P)$) and a labeled multigraph $G = (V, E, l)$ $(l : E \to H^*)$ are given. If there exists a homomorphism $\varphi : G \to G_P$ satisfying: $\varphi_V(v) = v_t$ if and only if $\deg(v, h) = |z_t(\varphi_L(h))|$, then $G$ is a complete complex graph for that pot.

**Proof.** Suppose a pot type $P = \{t_1, t_2, \ldots, t_m\}$ and a labeled multigraph $G = (V, E, l)$ are given. Let $\varphi_V : V \to V_P$ be a homomorphism such that, $\varphi_V(v) = v_t$ if and only if $\deg(v, h) = |z_t(\varphi_L(h))|$ for every $v \in V$.

For every $v_t_j \in \varphi_V(V)$, $j \in [m]$, there exists a set $\{w_{t_j}^1, w_{t_j}^2, \ldots, w_{t_j}^k\} \in V$ such that $\varphi_V(w_{t_j}^i) = v_t_j$ for $i \in [k]$. Let $T_j = \{t_{w_1}^j, t_{w_2}^j, \ldots, t_{w_k}^j\}$ be sets of tiles, such that all tiles in $T_j$ are of type $t_j$, for $v_{t_j} \in \varphi_V(V)$ and $j \in [m]$.

Next, we construct a complex $C = (T, S, J)$, where the set of tiles is $T = \bigcup_{v_{t_j} \in \varphi_V(V)} T_j$ while the set of connections, $J$, is defined in the following manner. To $v_{t_j} \in \varphi_V(V)$, every edge $e \in E$ with $\mathfrak{h}(e) = \{w_{t_j}^i, w_{t_j}^j\}$ and $l(e) = \text{type}(h)$ we associate a sticky end $h_e \in S$ such that where $\text{type}(h_e) = \text{type}(h)$. To every edge $e \in E$ with $\mathfrak{h}(e) = \{w_{t_j}^i, w_{t_j}^k\}$ and $l(e) = \text{type}(h)$ we associate a connection $c = \{(t_{w_i}^j, h), (t_{w_k}^j, h')\} \in J$ where $\text{type}(h') = \theta(\text{type}(h))$. From the definition of the homomorphism it follows
that there are as many connections between $t_{w_r}^i$ and $t_{w_s}^j$ via sticky end type $h$ as the edges between $w_{r}^t$ and $w_{s}^t$ labeled $\varphi_L^{-1}(h)$ (note that $\varphi_L$ is a bijection that maps $H^+$ into $H^+$). Because of the homomorphism, if $e \in E$ with $\mathfrak{v}s(e) = \{w_{r}^t, w_{s}^t\}$ there is an edge $e_P \in E_P$ with $\mathfrak{v}s(e_P) = \{v_{r_k}, v_{t_j}\}$ labeled $l(\varphi_E(e))$ in the pot graph $G_P$. From the definition of the pot graph it follows that $t_i(l(\varphi_E(e))) > 0$ and $t_j(l(\varphi_E(e))) > 0$ (or $t_i(\theta(l(\varphi_E(e))) > 0$ and $t_j(l(\varphi_E(e))) > 0$), i.e., $t_{w_r}^i(l(e)) > 0$ and $t_{w_s}^j(l(e)) > 0$ (or $t_{w_r}^i(\theta(l(e))) > 0$ and $t_{w_s}^j(l(e)) > 0$).

If $C$ is not a complete complex, then there exists a tile $t_{w_q}^k \in T$ and a sticky end $h$ of type $h \in H$ such that $t_k(h) > 0$, but $(t_{w_q}^k, h) \notin c$ for any $c \in J$, i.e., for that $h$, $w_{t_q}^k$ does not have incident edge. Therefore $\deg(w_{t_q}^k, h) < |z_{t_k}(h)|$, which is not possible. So $C$ is a complete complex.

Next we will show that $G$ is isomorphic to the complete complex graph of $C$. Suppose $G_C = (V_C, E_C, l_C)$ is the complete complex graph of $C$. Then $V_C = \{v_{t_q}^i : t_{w_q}^i \in T\}$ and $E_C \in E_C$, with $\mathfrak{v}s(e_C) = \{v_{w_r}, v_{t_w}\}$ and $l(e_C) = \text{type}(h)$, if and only if $\{(t_{w_q}^i, h), (t_{w_q}^j, h')\} \in J$.

We define a function $\phi : V \rightarrow V_C$ by $\phi(w_{t_q}^i) = v_{t_q}^i$. By the construction of the complex $C$ it follows that $V_C = V$, so $\phi$ is a bijection. It is homomorphism because: $e \in E_C$ with $\mathfrak{v}s(e) = \{v_{w_r}, v_{t_w}\}$ and $l(e) = h$ if and only if $\{(t_{w_q}^i, h), (t_{w_q}^j, h')\} \in J$, for $\text{type}(h) = h$ if and only if there exists an edge $e' \in E$ with $\mathfrak{v}s(e') = \{w_{r}^t, w_{s}^t\}$ and $l(e') = \varphi_L^{-1}(h)$.

By the homomorphism defined in Lemma 4.3.1, $G$ is homomorphic to $G_P$.

Then we proceed to define when two pot types are isomorphic. In order to do that first we define isomorphism between two tile types (one from each pot type) and sticky end types, and then define when two pots are equivalent. Two pot types are equivalent if there is a bijection $\psi$ between the sets of sticky end types and a bijection between the pot types, so that if a tile type $t$ is mapped to a tile type $t'$, then there is a bijection from the sticky ends of $t$ to those of $t'$ such that a sticky end of type, say $h$, is assigned to one of type $\psi(h)$.
Definition 4.3.4. Let \( t_1 \) be a tile type of the pot type \( P_1(H_1, \theta_1) \) and \( t_2 \) be a tile type of the pot type \( P_2(H_2, \theta_2) \). The tiles \( t_1 \) and \( t_2 \) are isomorphic (notation: \( t_1 \cong t_2 \)) if there exists a graph isomorphism \( \sigma \) from the tile graph \( G_{t_1} \) to the tile graph \( G_{t_2} \).

When two tiles, \( t_1 \) and \( t_2 \) are isomorphic, we say that the isomorphism \( \sigma \) between \( G_{t_1} = (V_{H_1} \cup \{v_{t_1}\}, E_1, l_1) \) and \( G_{t_2} = (V_{H_2} \cup \{v_{t_2}\}, E_2, l_2) \) preserves sticky ends if for every \( h_1 \) and \( h_2 \) of the same sticky end type \( h \in H_1 \) and \( \sigma(v_{h_1}) = v_{h'_1}, \sigma(v_{h_2}) = v_{h'_2}, \) \( v_{h_1}, v_{h_2} \in V_{H_1}, v_{h'_1}, v_{h'_2} \in V_{H_2}, h'_1 \) and \( h'_2 \) are of the same sticky end type \( h' \in H_2, \) i.e., \( \deg(v_{t_1}, h) = \deg(v_{t_2}, h') \).

Example 4.3.5. Consider the Figure 4.6. The isomorphism between \( t \) and \( t' \) preserves sticky ends, while the isomorphism between \( t \) and \( t'' \) does not preserve sticky ends.

![Figure 4.6: t \cong t' via an isomorphism that preserves sticky ends. t \cong t'' via an isomorphism that does not preserves sticky ends.](image)

Definition 4.3.6. The pairs \((H_1, \theta_1)\) and \((H_2, \theta_2)\) are isomorphic (notation: \((H_1, \theta_1) \cong (H_2, \theta_2)\)) if there is bijection \( f : H_1 \to H_2 \) such that the following diagram commutes.

\[
\begin{array}{ccc}
H_1 & \xrightarrow{f} & H_2 \\
\downarrow{\theta_1} & & \downarrow{\theta_2} \\
H_1 & \xrightarrow{f} & H_2 \\
\end{array}
\]

\[f \circ \theta_1 = \theta_2 \circ f\]

We require the diagram to commute, because we want all Watson-Crick connections to be preserved. For the rest of the dissertation we fix \( f \) to be the isomorphism defined above.
Proposition 4.3.7. The isomorphism between the sticky ends is an equivalence relation.

Proof. ≃ is reflexive.

Given \((H_1, \theta_1)\). Consider the identity map \(f : H_1 \to H_1\), i.e., \(f(h) = h\) for every \(h \in H_1\). Then for any \(h \in H_1\), \(f(\theta_1(h)) = \theta_1(h) = \theta_1(f(h))\), i.e., \(f \circ \theta_1 = \theta_1 \circ f\).

≃ is symmetric.

Assume \((H_1, \theta_1) \cong (H_2, \theta_2)\). There exists a bijective map \(f : H_1 \to H_2\) such that \(f \circ \theta_1 = \theta_2 \circ f\). Since \(f\) is a bijection, it follows that \(f^{-1} : H_2 \to H_1\) is a bijection, and \(\theta_1 \circ f^{-1} = f^{-1} \theta_1\), i.e., \((H_2, \theta_2) \cong (H_1, \theta_1)\) via the map \(f^{-1}\).

≃ is transitive.

Let \((H_1, \theta_1) \cong (H_2, \theta_2)\) and \((H_2, \theta_2) \cong (H_3, \theta_3)\). There exists a bijection \(f : H_1 \to H_2\) such that \(f \circ \theta_1 = \theta_2 \circ f\), and a bijection \(g : H_2 \to H_3\) such that \(g \circ \theta_2 = \theta_3 \circ g\). The bijection \(h : H_1 \to H_3\) defined as \(h = g \circ f\) satisfies \(h \circ \theta_1 = g \circ f \circ \theta_1 = g \circ \theta_2 \circ f = \theta_3 \circ g \circ f = \theta_3 \circ h\). Therefore \((H_1, \theta_1) \cong (H_2, \theta_3)\).

Denote by \(\mathcal{G}(P) = \{G_C : C \in \mathcal{C}(P)\}\) the set of all complete complex graphs of \(P(H, \theta)\).

Denote by \(G(P) = \{G_C : C \text{ is a complex over } P\}\) the set of all complex graphs of \(P(H, \theta)\).

Finally we define when two pot types are equivalent. The idea behind pot type equivalence is to transfer all the properties of one pot type to the one equivalent to it, no matter how different tile types or sticky end types they might have.

Definition 4.3.8. The pot types \(P_1(H_1, \theta_1)\) and \(P_2(H_2, \theta_2)\) are equivalent (notation \(P_1 \cong P_2\)) if \((H_1, \theta_1) \cong (H_2, \theta_2)\) via the isomorphism \(f\), and there is a bijection \(\psi : P_1 \to P_2\) such that for every \(t \in P_1\), \(G_t \cong G_{\psi(t)}\) via a unique isomorphism \(\sigma\) that preserves sticky end types, i.e., for every \(h \in H_1\), \(\sigma_L(h) = f(h)\).
For the above definition, the isomorphism to preserve sticky end types means: if \( P_1(H_1, \theta_1) \cong P_2(H_2, \theta_2) \) via bijection \( \psi \), and for \( G_t = (v_t \cup V_{H_1}, E_{t,H_1}, l_t) \), \( G_t \cong G_{\psi(t)} \) via isomorphism \( \sigma \), then for every \( v_h \in V_{H_1} \), \( \sigma(v_h) = v_h' \in V_{H_2} \) such that \( \text{type}(h') = f(h) \in H_2 \) whenever \( \text{type}(h) = h \in H_1 \), and \( \sigma(v_t) = v_t' \), where \( \text{type}(t') = \psi(t) \) and \( G_{\psi(t)} = (v_t' \cup V_{H_2}, E_{\psi(t),H_2}, l_{\psi(t)}) \).

**Example 4.3.9.** Let us show that the pot types given in Figure 4.7 are equivalent. They are defined as follows: \( P_1 = \{t_1, t_2\} \), \( H_1 = \{a, b, \theta_1(a), \theta_1(b)\} \); \( P_2 = \{t'_1, t'_2\} \), \( H_2 = \{x, y, \theta_2(x), \theta_2(y)\} \). The corresponding tile graphs are defined as usual.

![Figure 4.7: The pot types \( P_1 \) and \( P_2 \) are equivalent](image)

First, \( (H_1, \theta_1) \cong (H_2, \theta) \) via the isomorphism \( f : H_1 \rightarrow H_2 \), defined as \( f(a) = x \), \( f(b) = \theta_2(y) \), \( f(\theta_1(a)) = \theta_2(x) \), and \( f(\theta_1(b)) = y \). Obviously, \( f \circ \theta_1 = \theta_2 \circ f \), since \( f(\theta_1(a)) = \theta_2(x) = \theta_2(f(a)) \) and \( f(\theta_1(b)) = y = \theta_2(\theta_2(y)) = \theta_2(f(b)) \).

Second, there exists a bijection \( \psi : P_1 \rightarrow P_2 \) defined as \( \psi(t_1) = t'_1 \) and \( \psi(t_2) = t'_2 \).

Third, \( G_{t_1} \cong G_{t'_1} \) and \( G_{t_2} \cong G_{t'_2} \) via the isomorphism \( \sigma \) defined as \( \sigma(v_1) = v'_1 \), \( \sigma(v_2) = v'_2 \), \( \sigma(a) = x \), \( \sigma(\theta_1(a)) = \theta_2(x) \), \( \sigma(b) = \theta_2(y) \), and \( \sigma(\theta_1(b)) = y \).

**Definition 4.3.10.** For the pot types \( P_1(H_1, \theta_1) \) and \( P_2(H_2, \theta_2) \), the sets of complete complex graphs are equivalent, \( \mathcal{G}(P_1) \cong \mathcal{G}(P_2) \), if there exist two functions \( \phi_1 : \mathcal{G}(P_1) \rightarrow \mathcal{G}(P_2) \) and \( \phi_2 : \mathcal{G}(P_2) \rightarrow \mathcal{G}(P_1) \) such that for every \( G_{C_1} \in \mathcal{G}(P_1) \), \( \phi_1(G_{C_1}) \cong G_{C_1} \), and for every \( G_{C_2} \in \mathcal{G}(P_2) \), \( \phi_2(G_{C_2}) \cong G_{C_2} \).

Here we need to point out that in the above definition the complete complex graphs are not necessarily isomorphic with a unique isomorphism. For example if \( G_{C_1} \) and
$G_{C_2}$ are two complete complete graphs from $\mathcal{G}(P_1)$, then we could have $\phi_1(G_{C_1}) \cong G_{C_1}$ via an isomorphism $\sigma_1$ and $\phi_1(G_{C_2}) \cong G_{C_2}$ via another isomorphism $\sigma_2$.

**Example 4.3.11.** Consider the pot types given in Figure 4.8.

![Two pot types](image)

Figure 4.8: Two pot types $P_1$ and $P_2$ with equivalent sets of complete complex graphs

Their corresponding sets of complete complex graphs $\mathcal{G}(P_1) = \{G(C_1), G(C_2)\}$ and $\mathcal{G}(P_2) = \{G(C'_1), G(C'_2)\}$ are given in Figure 4.9. The graphs $G(C_1)$ and $G(C'_1)$ are isomorphic via the isomorphism $\sigma_1 : G(C_1) \to G(C'_1)$ defined as $\sigma_1(v_1) = v'_1$, $\sigma_1(v_3) = v'_3$, and $\sigma_2(v_4) = v'_4$. The graphs $G(C_2)$ and $G(C'_2)$ are isomorphic via the isomorphism $\sigma_2 : G(C_2) \to G(C'_2)$ defined as $\sigma_1(v_2) = v'_2$, $\sigma_1(v_3) = v'_3$, and $\sigma_2(v_5) = v'_6$. Obviously, the isomorphisms $\sigma_1$ and $\sigma_2$ are different, but the sets of complete complex graphs are equivalent.
Figure 4.9: The sets of complete complex graphs for the pot types given in Figure 4.8

Although the pot types $P_1$ and $P_2$ have equivalent sets of complete complex graphs, they are not equivalent.

Similarly we can define when the sets of two complex graphs are isomorphic.

**Definition 4.3.12.** For the pot types $P_1(H_1, \theta_1)$ and $P_2(H_2, \theta_2)$, the sets of complex graphs are equivalent, $G(P_1) \cong G(P_2)$, if there exist two functions $\phi_1 : G(P_1) \to G(P_2)$ and $\phi_2 : G(P_2) \to G(P_1)$ such that for every $G_{C_1} \in G(P_1)$, $\phi_1(G_{C_1}) \cong G_{C_1}$, and for every $G_{C_2} \in G(P_2)$, $\phi_2(G_{C_2}) \cong G_{C_2}$.

Also, in this definition, it is not necessarily true that two complex graphs are isomorphic via a single isomorphism, but rather several different ones.

**Proposition 4.3.13.** If the pot types $P_1(H_1, \theta_1)$ and $P_2(H_2, \theta_2)$ are equivalent, then the corresponding sets of complex graphs are equivalent, i.e.

$$P_1 \cong P_2 \Rightarrow G(P_1) \cong G(P_2).$$

*Proof.* Let $P_1 \cong P_2$, so $(H_1, \theta_1) \cong (H_2, \theta_2)$ (i.e., there exists a bijection $f : H_1 \to H_2$ such that $f \circ \theta_1 = \theta_2 \circ f$) and there exists an bijection $\psi : P_1 \to P_2$ s.t. for every $t \in P_1$, $G_t \cong G_{\psi(t)}$ via an isomorphism $\sigma$.

Suppose the tile types $s$ and $t$ from $P_1$ have complementary sticky end of type $h \in H_1$ such that $t(h) > 0$ and $s(\theta_1(h)) > 0$. Consider the tile graphs $G_t$ and $G_s$ from
\(G(P_1)\) and their equivalent tile graphs \(G_{\psi(t)}\) and \(G_{\psi(s)}\). There exists an edge \(e_t \in G_t\) with \(\text{vs}(e_t) = \{v_t, v_h\}\), \(\text{type}(h) = h\), and an edge \(e_s \in G_s\) with \(\text{vs}(e_s) = \{v_s, v_h\}\), where \(\text{type}(h) = \theta_1(h)\). Because \(G_t \cong G_{\psi(t)}\) and \(G_s \cong G_{\psi(s)}\), there exists an edge \(e'_t \in G_{\psi(t)}\) with \(\text{vs}(e'_t) = \{v'_t, v'_{h'}\}\), \(\text{type}(t') = \psi(t)\), \(\text{type}(h') = f(h)\), and an edge \(e'_s \in G_{\psi(s)}\) with \(\text{vs}(e'_s) = \{v'_s, v'_{h'}\}\), \(\text{type}(s') = \psi(s)\), \(\text{type}(h') = \theta_2(f(h))\). From the construction of the tile graphs follows that \(\psi(t)(f(h)) > 0\) and \(\psi(s)(\theta_2(f(h))) > 0\), i.e., the tile types \(\psi(t)\) and \(\psi(s)\) from \(P_2\) can connect via the sticky end \(f(h)\). With other words if tiles of types \(t\) and \(s\) can connect via a sticky end type \(h\), then tiles of the types \(\psi(t)\) and \(\psi(s)\) can connect via the sticky end of type \(f(h)\).

Next we will show that to every complex graph from \(G(P_1)\), there exists a unique complex graph in \(G(P_2)\) isomorphic to it.

Let \(G_C = (V_T \cup V_H, E_T \cup E_{T,H}, l) \in G(P_1)\), then there exists a complex \(C = (T, S, J)\) of \(P_1\) s.t. \(G_C\) is a complex graph of \(C\). We can construct a complex \(C' = (T', S', J')\) of \(P_2\) that is isomorphic to \(C\). \(C'\) has tiles of types that are isomorphic to the tile types in \(P_1\) such that the number of tiles in \(T\) of type \(t\) is equal to the number of tiles in \(T'\) of type \(\psi(t)\) in \(P_2\). Hence, we can define a bijection \(g : T \rightarrow T'\), \(g(t) = t'\) if \(\psi(t) = t'\), where \(\text{type}(t) = t\) and \(\text{type}(t') = t'\). The complex \(C'\), for each \(h \in H\), the number of sticky ends of type \(h\) on the tile in \(T\) is equal to the number of sticky ends of type \(f(h)\) on the tiles in \(T'\). We can define a bijection \(k : S \rightarrow S'\) such that \(k(h) = h'\) if \(\text{type}(h) = h\), \(\text{type}(h') = f(h)\), and if \(t(h) = 1\), then \(g(t)(k(h)) = 1\).

The set of connections for \(C'\) is defined as \(J' = \{(t', h'), (s', \tilde{h}')\} : \text{type}(h') = h', \text{type}(\tilde{h}') = \theta_2(h'), \{(g^{-1}(t'), k^{-1}(h')), (g^{-1}(s'), k^{-1}(\tilde{h}'))\} \in J, \text{where type}(k^{-1}(h')) = f^{-1}(h') \text{and type}(k^{-1}(\tilde{h}')) = f^{-1}(\theta_2(h')).\) (The above discussion follows the definition of the set of connections for \(C'\).)

Therefore the complex graph for \(C'\), \(G_{C'} = (V_{T'}, v_{H'}, E_{T'} \cup E_{T',H'}, l_{C'})\), is defined as the following: \(V_{T'} = \{v'_{g(t)} : v_t \in V_T\}\), \(V_{H'} = \{v'_{k(h)} : v_h \in V_H\}\), \(v_t \in \text{vs}(e)\) for \(e \in E_{T,H}\). The set of edges and the set of labels are defined as \(e' \in E_{T'}\) with \(\text{vs}(e') = \{v'_s, v'_t\}\) and label \(l_{C'}(e)\) if and only if there exists and edge \(e \in E_T\) with
\( \text{vs}(e) = \{v_{g^{-1}(s)}, v_{g^{-1}(t)}\} \) and \( f(l_C(e)) = l_{C'}(e') \). An edge \( e'_t \) with \( \text{vs}(e'_t) = \{v'_t, v'_{h'}\} \) (type\( (h) = h \)) and label \( l_{C'}(e'_t) \) is in \( E_{T', H'} \) if and only if there exists an edge \( e_t \in E_{T, H} \) with \( \text{vs}(e_t) = \{v_{g^{-1}(t')}, v_{k^{-1}(h')}\} \) (type\( (h') = f^{-1}(h) \)) and \( f(l_C(e_t)) = l_{C'}(e'_t) \). In other words, there exists a bijection between \( G(P_1) \) and \( G(P_2) \), i.e., to every complex graph \( \mathcal{G}_C \in G(P_1) \) we can bijectively correspond a complex graph \( \mathcal{G}_{C'} \in G(P_2) \). Next we will show that these two graphs are isomorphic, \( \mathcal{G}_C \cong \mathcal{G}_{C'} \).

Define a map \( \alpha : V_C \to V_{C'} \) such that \( \alpha(v_t) = v'_{g(t)} \) and \( \alpha(v_h) = v'_{k(h)} \). From the construction of the complex graph it is clear that \( \alpha \) is homomorphism.

\( \alpha \) is one-to-one: From the construction, if \( \alpha(v_s) = \alpha(v_t) \) follows that \( v'_{g(s)} = v'_{g(t)} \), so \( g(s) = g(t) \) and since \( g \) is bijection, \( s = t \), i.e. \( v_s = v_t \). If \( \alpha(v_{h_1}) = \alpha(v_{h_2}) \), then \( v'_{k(h_1)} = v'_{k(h_2)} \); it follows that \( k(h_1) = k(h_2) \) from where follows that \( h_1 = h_2 \) i.e. \( v_{h_2} = v_{h_2} \).

\( \alpha \) is onto: If \( v'_t \in V_{T'} \) then there exists \( v_{g^{-1}(t)} \in V_T \) s.t. \( \alpha(v_{g^{-1}(t)}) = v'_t \). If \( v'_{h'} \in V_{H'} \) then there exists \( v_{k^{-1}(h')} \in V_H \) s.t. \( \alpha(v_{k^{-1}(h')}) = v'(h') \).

\( \alpha \) is a homomorphism: Follows from the definition of \( \mathcal{G}_{C'} \).

We showed that for every \( \mathcal{G}_C \in G(P_1) \) there exists \( \mathcal{G}_{C'} \in G(P_2) \) satisfying \( \mathcal{G}_C \cong \mathcal{G}_{C'} \). Consequently there exists a map \( \phi_1 : \mathcal{G}(P_1) \to G(P_2) \), defined as \( \phi_1(\mathcal{G}_C) = \mathcal{G}_{C'} \) if \( \mathcal{G}_C \cong \mathcal{G}_{C'} \), i.e., \( \phi_1(\mathcal{G}_C) \cong \mathcal{G}_C \) for \( \mathcal{G}_C \in G(P_1) \).

Next we will show that there exists a map \( \phi_2 : G(P_2) \to G(P_1) \), such that \( \phi_2(\mathcal{G}_{C'}) \cong \mathcal{G}_{C'} \) for \( \mathcal{G}_{C'} \in G(P_2) \). In a very similar manner as previously in this proof, we show that for every complex graph \( \mathcal{G}_{C'} = (V_{T'}, V_{H'}, E_{T'}, E_{H'}, l_{C'}) \in G(P_2) \) with complex \( C' \) from \( P_2 \), we can build a complex \( C \) in \( P_1 \) whose graph, \( \mathcal{G}_C \), is isomorphic to \( \mathcal{G}_{C'} \). Hence \( \phi_2 \) is defined such that \( \phi_2(\mathcal{G}_{C'}) = \mathcal{G}_C \), i.e., \( \phi_2(\mathcal{G}_{C'}) \cong \mathcal{G}_C \).

For these reasons, we can conclude that \( G(P_1) \cong G(P_2) \).

\[ \text{Remark 4.3.14.} \] The converse is not true in general. The main reason why the converse fails is the multiple isomorphisms allowed between the sets of complete com-

56
plex graphs. Consider the pot types depicted on Figure 4.10. The pot types are not equivalent, while their sets of complex graphs are.

![Diagram of two pot types and their corresponding complex graphs](image)

Figure 4.10: The sets of complex graphs $G(P_1)$ and $G(P_2)$ are equivalent, but $P_1 \napprox P_2$

**Definition 4.3.15.** The pot types $P_1(H_1, \theta_1)$ and $P_2(H_2, \theta_2)$ are similar (notation $P_1 \sim P_2$) if their corresponding sets of complete complex graphs are equivalent $\mathcal{G}(P_1) \cong \mathcal{G}(P_2)$.

**Corollary 4.3.16.** If the strongly satisfiable pot types $P_1(H_1, \theta_1)$ and $P_2(H_2, \theta_2)$ are equivalent, then they are similar.

The proof immediately follows from the previous one.

**Proposition 4.3.17.** If the types $P_1(H_1, \theta_1)$ and $P_2(H_2, \theta_2)$ are equivalent, then the corresponding pot graphs, $G_{P_1}$ and $G_{P_2}$, are isomorphic.

**Proof.** Since $P_1 \cong P_2$, there exists a bijection $\psi : P_1 \to P_2$ such that for every $t \in P_1$, $G_t \cong G_{\psi(t)}$ and the isomorphism preserves sticky end types. Also, $(H_1, \theta_1) \cong (H_2, \theta_2)$, i.e., there exists a bijective function $f : H_1 \to H_2$ such that $f \circ \theta_1 = \theta_2 \circ f$.

Let use the following notation: $G_{P_1} = (V_{P_1}, E_{P_1}, l_{P_1})$ and $G_{P_2} = (V_{P_2}, E_{P_2}, l_{P_2})$, where $V_{P_1} = \{v_t : t \in P_1\}$ and $V_{P_2} = \{v'_t : t \in P_2\}$. Since $|P_1| = |P_2|$ it is clear that $|V_{P_1}| = |V_{P_2}|$.

Define a map $\gamma : G_{P_1} \to G_{P_2}$, s.t $\gamma(v_t) = v'_{\psi(t)}$.
\( \gamma \) is well defined: From the definition of the pot graph, if \( v_t = v_{t'} \), for \( v_t, v_{t'} \in V_{P_1} \), then \( t = t' \) i.e. \( \psi(t) = \psi(t') \). Hence \( v'_{\psi(t)} = v'_{\psi(t')} \) i.e. \( \gamma(v_t) = \gamma(v_{t'}) \).

\( \gamma \) is one-to-one: Assume \( \gamma(v_t) = \gamma(v_{t'}) \), by the definition of \( \gamma \) it follows that \( v'_{\psi(t)} = v'_{\psi(t')} \). Therefore \( \psi(t) = \psi(t') \). Since \( \psi \) is a bijection, \( t = t' \) i.e., \( v_t = v_{t'} \).

\( \gamma \) is onto: For every \( t' \in P_2 \), there exists \( t \in P_1 \) s.t. \( \psi(t) = t' \). Hence \( \gamma(v_t) = v'_{\psi(t)} = v'_{t'} \).

\( \gamma \) is a homomorphism: Let \( v_t \) and \( v_{t'} \) are adjacent via an edge labeled \( h \) (\( v_t \) and \( v_{t'} \) are vertices of \( G_{P_1} \)). By the definition of \( V_{P_1} \), it follows that either \( t(h) > 0 \) and \( t' (\theta_1(h)) > 0 \) or \( t(\theta_1(h)) > 0 \) and \( t'(h) > 0 \). Because \( \psi \) preserves sticky end types, these cases correspond to \( \psi(t)(f(h)) > 0 \) and \( \psi(t')(\theta_2(f(h))) > 0 \), or \( \psi(t)(\theta_2(f(h))) > 0 \) and \( \psi(t')(f(h)) > 0 \), respectively, i.e., \( v'_{\psi(t)} = \gamma(v_t) \) and \( v'_{\psi(t')} = \gamma(v_{t'}) \) are adjacent via an edge labeled \( f(h) \).

\[ \square \]

**Remark 4.3.18.** The converse in general is not true i.e., from \( G_{P_1} \cong G_{P_2} \) it doesn’t follow that \( P_1 \cong P_2 \). The pot types presented in Figure 4.11 a) have equivalent pot graphs, presented in Figure 4.11 b), but the pot types are not equivalent. The reason for that is that the set of complete complexes for the pot type \( P_1 \) is infinite, while the set of complete complexes for the pot type \( P_2 \) is singleton, i.e., \( P_1 \) and \( P_2 \) are not similar, therefore they are not equivalent.
In order for the converse of Proposition 4.3.17 to be true, we need to strengthen the hypothesis.

Proposition 4.3.19. Let \( t(h) \leq 1 \), \( \forall h \in H \) and \( \forall t \in P_1 \cup P_2 \). If \( G_{P_1} \cong G_{P_2} \) via isomorphism \( \sigma \), then \( P_1(H_1, \theta_1) \cong P_2(H_2, \theta_2) \) via isomorphism \( \psi \).

Proof. Let \( G_{P_1} \cong G_{P_2} \) via an isomorphism \( \sigma \). Define a map \( \psi : P_1 \to P_2 \) such that \( \psi(t) = s \) if and only if \( \sigma(v_t) = v_s' \). It is straightforward to show that \( \psi \) is a bijection. We should just show that \( \psi \) is a homomorphism and preserves sticky ends. Let \( v_t \) from \( G_{P_1} \) is adjacent to \( v_{t_1}, v_{t_2}, \ldots v_{t_k} \) via edges labeled \( h \). That means that either \( t(h) = 1 \) or \( t(\theta_1(h)) = 1 \). Without loss of generality lets take \( t(h) = 1 \), then \( t_1(\theta_1(h)) = 1, t_2(\theta_1(h)) = 1, \ldots, t_k(\theta_1(h)) = 1 \). Also because of the pot graph isomorphism follows that \( \sigma(v_t) \) is adjacent to \( \sigma(v_{t_1}), \sigma(v_{t_2}), \ldots, \sigma(v_{t_k}) \) via edges labeled \( \sigma(h) \). That means \( v_{\psi(t)}' \) is adjacent to \( v_{\psi(t_1)}', v_{\psi(t_2)}', \ldots, v_{\psi(t_k)}' \) via edges labeled \( \sigma(h) \), i.e. \( \psi(t)(\sigma(h)) = 1 \) (or \( \psi(t)(\theta_2(\sigma(h))) = 1 \)) and \( \psi(t_i)(\theta_2(f(h))) = 1 \) (or \( \psi(t_i)(\sigma(h)) = 1 \)) for \( i \in [m] \). Therefore the isomorphism preserves sticky ends i.e., \( \deg(v_t, h) = \deg(v'(\psi(t)), \sigma(h)) = 1 \) for all \( h \in H \) and \( t \in P_1 \). Hence \( P_1 \cong P_2 \).

Remark 4.3.20. If the pot types were only similar, but not equivalent then the corresponding pot graphs are not necessarily equivalent. Consider the following example.
depicted in the figures given below. The sets of complete complex graphs are equivalent since there exists a bijection $\phi : \mathcal{G}(P_1) \to \mathcal{G}(P_2)$ defined as: for $i \in [8]$, $\phi(G_{C_i}) = G_{C'_i}$, such that $G_{C_i} \cong G_{C'_i}$ for via different isomorphisms $\sigma_i : G_{C_i} \to G_{C'_i}$. However the pot types are not isomorphic since their pot graphs are not isomorphic.

\[ \begin{array}{c}
\text{P}_1 \\
\begin{array}{c}
\text{t}_1 \quad \text{a} \quad \text{t}_2 \quad \text{b} \quad \text{t}_3 \quad \text{c} \quad \text{t}_4 \quad \text{d} \\
\text{a} \quad \text{a} \quad \text{a} \quad \text{b} \quad \text{b} \quad \text{c} \quad \text{d} \\
\text{x} \quad \text{x} \quad \text{x} \quad \text{x} \quad \text{x} \quad \text{x} \quad \text{x}
\end{array}
\end{array} \\
\begin{array}{c}
\text{t}_5 \quad \text{e} \quad \text{t}_6 \quad \text{f} \quad \text{t}_7 \quad \text{g} \quad \text{t}_8 \quad \text{h} \\
\text{e} \quad \text{e} \quad \text{e} \quad \text{f} \quad \text{f} \quad \text{g} \quad \text{h} \\
\text{y} \quad \text{y} \quad \text{y} \quad \text{y} \quad \text{y} \quad \text{y} \quad \text{y}
\end{array}
\]

\[ \begin{array}{c}
\text{P}_2 \\
\begin{array}{c}
\text{t}_1 \quad \text{a} \quad \text{t}_2 \quad \text{b} \quad \text{t}_3 \quad \text{c} \quad \text{t}_4 \quad \text{d} \\
\text{a} \quad \text{a} \quad \text{a} \quad \text{b} \quad \text{b} \quad \text{c} \quad \text{d} \\
\text{x} \quad \text{x} \quad \text{x} \quad \text{x} \quad \text{x} \quad \text{x} \quad \text{x}
\end{array}
\end{array} \\
\begin{array}{c}
\text{t}_5 \quad \text{e} \quad \text{t}_6 \quad \text{f} \quad \text{t}_7 \quad \text{g} \quad \text{t}_8 \quad \text{h} \\
\text{e} \quad \text{e} \quad \text{e} \quad \text{f} \quad \text{f} \quad \text{g} \quad \text{h} \\
\text{y} \quad \text{y} \quad \text{y} \quad \text{y} \quad \text{y} \quad \text{y} \quad \text{y}
\end{array}
\]

The sets of complete complex graphs, $\mathcal{G}(P_1)$, $\mathcal{G}(P_2)$ for $P_1$ and $P_2$ are obviously equivalent, hence $P_1 \sim P_2$. 
But as we can see the pot graphs are not equivalent.

Also we would like to point out that this example can also serve as a counter example for the converse of Proposition 4.3.13, if $G(P_1) \cong G(P_2)$ it does not follow that $P_1 \cong P_2$. 
Next we will show that for every pot type \( P \), there exists a pot type \( \tilde{P} \) similar to \( P \) such that every tile type in \( \tilde{P} \) has sticky ends of distinct type. Hence, we could only consider pot types with tile types that have distinct sticky end types. Therefore, from Proposition 4.3.19 and Proposition 4.3.17 will follow that equivalent two pot types are equivalent if and only if their pot graphs are equivalent.

**Proposition 4.3.21.** For every pot type \( P(H, \theta) \) there exists a pot type \( \tilde{P}(\tilde{H}, \tilde{\theta}) \) such that every tile in \( \tilde{P}(\tilde{H}, \tilde{\theta}) \) has sticky ends of distinct type and \( P(H, \theta) \sim \tilde{P}(\tilde{H}, \tilde{\theta}) \).

*Proof.* Suppose \( |P| = m \), say \( P = \{ t_i : i \in [m] \} \), \( |H^+| = n \), say \( H^+ = \{ h_i : i \in [n] \} \), and \( \max_{t \in P} |z_t(h_i)| = k_i \) for \( i \in [n] \).

Construct a pot type \( \tilde{P}(\tilde{H}, \tilde{\theta}) \) defined in the following way:

- \( \tilde{H} = \tilde{H}^+ \cup \tilde{H}^- \), where \( \tilde{H}^+ \) is a disjoint union of the sets \( \tilde{H}^+_i = \{ h^i_j : j \in [k_i] \} \) and \( \tilde{H}^- \) is a disjoint union of the sets \( \tilde{H}^-_i = \{ \tilde{\theta}(h^i_j) : j \in [k_i] \} \).

- \( \tilde{\theta} : \tilde{H} \to \tilde{H} \) is an involution and for \( h^i_j \in \tilde{H}^+ \), \( \tilde{\theta}(h^i_j) \) and \( h^i_j \) can connect.

- For each \( t_j \in P \) we construct a set \( \tilde{P}_j \) of tile types in \( \tilde{P} \) defined in the following way. If \( t_j(h_i) = l_i \) for \( i \in [n] \) and \( h_i \in H^+ \), each tile type from \( \tilde{P}_j \) will have \( l_i \) sticky ends from the set \( \tilde{H}^+_i \), else if \( h_i \in H^- \) will have \( l_i \) sticky ends from \( \tilde{H}^-_i \). We assign one new tile type to each \( \prod_{i=1}^n \binom{k_i}{l_i} \) combination possible. With other words each tile type \( t' \in \tilde{P}_j \) will satisfy \( t'(h^i_1) + t'(h^i_2) + \cdots + t'(h^i_{k_i}) = t_j(h_i) \) for \( h^i_1, \ldots, h^i_{k_i} \in \tilde{H}^+_i \) if \( h_i \in H^+ \), or \( h^i_1, \ldots, h^i_{k_i} \in \tilde{H}^-_i \) if \( h_i \in H^- \). Every tile type in \( \tilde{P} \) has different sticky ends and since we define all the possible tile types in \( \tilde{P}_j \) such that each tile type has \( l_i \) sticky ends from \( \tilde{H}^+ \) or from \( \tilde{H}^- \), there are \( \prod_{i=1}^n \binom{k_i}{l_i} \) tile types in \( \tilde{P}_j \).

- \( \tilde{P} = \bigcup_{t_j \in P} \tilde{P}_j \)

CLAIM: \( P(H, \theta) \sim \tilde{P}(\tilde{H}, \tilde{\theta}) \)

62
Suppose a complete complex $C = \langle T, S, J \rangle$ in $C(P)$ is given, and suppose there are $r_i$ tiles of type $t_i$ in $T$, for $i \in [m]$, i.e., $T = \{t_1^1, \ldots, t_1^r_1, t_2^1, \ldots, t_2^r_2, \ldots, t_m^1, \ldots, t_m^r_m\}$ and $\text{type}(t_i^k) = t_i$, for $i \in [m]$ and $k \in [r_i]$. Let $G_C = (V_C, E_C, l)$ be the complete complex graph of $C$, where $V_C = \{v_t : t \in T\}$.

We construct a complex $\tilde{C} = \langle \tilde{T}, \tilde{S}, \tilde{J} \rangle \in C(\tilde{P})$, such that $|\tilde{T}| = |T|$ and there are $r_i$ tiles of types in $\tilde{P}_i$ in $\tilde{T}$, for $i \in [m]$. The complex graph is $\tilde{G} = (\tilde{V}, \tilde{E}, \tilde{l})$, for which $\tilde{V} = \{v'_t : t \in \tilde{T}\}$. Since $|\tilde{T}| = |T|$, from the definition of the complex graphs it follows that $|\tilde{V}| = |V|$.

Consider any two tiles from $T$, say $t_i$ (of tile type $t_i$), $t_j$ (of tile type $t_j$), such that they connect with $k_i$ connections via sticky ends $h_i$ and $\theta(h_i)$, for $h_i \in H_i^+$ and $l \in [n]$.

For the tiles in $T$ given above there exist two tiles in $\tilde{T}$, $\tilde{t}_i$ and $\tilde{t}_j$ such that $\text{type}(\tilde{t}_i) \in \tilde{P}_i$ and $\text{type}(\tilde{t}_j) \in \tilde{P}_j$ and they have $k_i$ complementary sticky ends from $\tilde{H}_i^+$ and $\tilde{H}_i^-$ respectively, say $\tilde{t}_i(h'_i) = \tilde{t}_j(\tilde{\theta}(h'_i)) = 1$ for $i \in [k_i]$. Based on the original tile connections, the other connections are established in a similar manner, i.e. there exists connections $\{(\tilde{t}_i, h'_i), (\tilde{t}_j, f'_i)\} \in \tilde{J}$, for $i \in [k_i]$, and $\text{type}(h'_i) = h'_i$, $\text{type}(f'_i) = \tilde{\theta}(h'_i)$. This construction always works, because $\tilde{P}_i$ has tiles with all the possible combinations of $l_i$ sticky ends from $\tilde{H}_i^+$ or $\tilde{H}_i^-$ depending whether $t_j(h_i) = l_i$ or $t_j(\theta(h_i)) = l_i$. / 

The number of connections between $t_i$ and $t_j$ is the same as the number of connections between $\tilde{t}_i$ and $\tilde{t}_j$. Note that the number of sticky ends on the tiles in $T$ is the same as the number of sticky ends on the tiles in $\tilde{T}$ (this comes from the definition of the tiles in $\tilde{P}$). Since for every connection in $C$, there exists exactly one connection in $\tilde{C}$, we can conclude that $\tilde{C}$ is a complete complex.

We define a map $\varphi : V \rightarrow \tilde{V}$ such that $\varphi(v_{t_i}) = v'_{t_i}$ and $\varphi(v_{t_j}) = v'_{t_j}$ if the above is satisfied. Therefore, the number of edges between $v_{t_i}$ and $v_{t_j}$ is the same as the number of edges between $v'_{t_i}$ and $v'_{t_j}$. Since $C$ and $\tilde{C}$ have the same number of connections, from the definition of complete complex graphs it follows that $|E| = |\tilde{E}|$. From here we can conclude that there are $x$ number of edges between $v_{t_i}$ and $v_{t_j}$ if
and only if there are $x$ number of edges between $v'_{t_i}$ and $v'_{t_j}$, i.e., $\varphi$ is an isomorphism.

**CONVERSE:** Suppose that $\tilde{C} = \langle \tilde{T}, \tilde{S}, \tilde{J} \rangle$ is a complete complex in $\mathcal{C}(\tilde{P})$ for which there are $r_i$ tiles (in $\tilde{T}$) from $\tilde{P}_i$ for $i \in [n]$ and some $r_i$. Suppose that the corresponding complete complex graph is $\tilde{G}_C = (\tilde{V}, \tilde{E}, \tilde{l})$.

We construct a complex $C = \langle T, S, J \rangle$ in $\mathcal{P}$ such that $T$ has $r_i$ tiles of type $t_i$. Consider any two tiles $\tilde{t}_i$ (of tile type in $\tilde{P}_i$) and $\tilde{t}_j$ (of tile type in $\tilde{P}_j$) from $\tilde{T}$, such that they connect with $w_l$ connections via sticky ends from $\tilde{H}_i^+$ and $\tilde{H}_i^-$ respectively, for $l \in [n]$.

For the tile types $t_i$ and $t_j$ in $\mathcal{P}$ given above we assign two tiles in $T$, $t_i$, $t_j$ such that $\text{type}(t_i) = t_i$ and $\text{type}(t_j) = t_j$.

From the construction of the pot type $\tilde{P}$, i.e., $\tilde{P}_i$ and $\tilde{P}_j$, it follows that $t_i(h_l) \geq w_l$ and $t_j(\theta(h_l)) \geq w_l$, for $l \in [n]$. To the given tiles $t_i$ and $t_j$ we assign $w_l$ connections via the sticky ends $h_l$ and $\theta(h_l)$ for $l \in [n]$.

Therefore, the number of connections between $\tilde{t}_i$ and $\tilde{t}_j$ is same as the number of connections between $t_i$ and $t_j$. Note that the number of sticky ends on the tiles in $T$ is the same as the number of sticky ends on the tiles in $\tilde{T}$ (this comes from the definition of the tiles in $\tilde{P}$). Since for every connection in $\tilde{C}$, there is one connection in $C$, we can conclude that $C$ is a complete complex.

We define a map $\phi : \tilde{V} \rightarrow V$ such that $\varphi(v'_{t_i}) = v_{t_i}$ and $\varphi(v'_{t_j}) = v_{t_j}$ if the above is satisfied. Therefore, the number of edges between $v'_{t_i}$ and $v'_{t_j}$ is the same as the number of edges between $v_{t_i}$ and $v_{t_j}$. Since $C$ and $\tilde{C}$ have same number of connections, from the definition of complete complex graphs it follows that $|\tilde{E}| = |E|$. From here we can conclude that there are $x$ number of edges between $v_{t_i}$ and $v_{t_j}$, i.e., $\phi$ is an isomorphism.

**Example 4.3.22.** This is an example of a pot type $\mathcal{P}$ and its corresponding pot type $\tilde{P}$ that don’t contain tile types with the same sticky ends.
Figure 4.12: A pot type $\mathbf{P}$ and its corresponding pot type $\tilde{\mathbf{P}}$ that don’t contain tile types with the same sticky ends.

The corresponding sets of complete complex graphs.

$$G(\mathbf{P})$$

$$G(\tilde{\mathbf{P}})$$

Figure 4.13: Elements from the sets of complete complex graphs for the pot types $\mathbf{P}$ and $\tilde{\mathbf{P}}$ given in Figure 4.12.
Minimal and Maximal Complete Complex Graphs

Definition 4.3.23. Let $G_{C'}$ and $G_C$ be complex graphs of the same pot. The complex graph $G_{C'}$ is a covering complex graph of $G_C$ if there exists a surjective homomorphism $\pi : V_{C'} \to V_C$ which preserves the neighborhood of every vertex, i.e., for every vertex $v \in V_{C'}$ and every $h \in H$, $\deg(v, h) = \deg(\pi(v), f(h))$, where $f$ is the mapping defined in Definition 4.3.6.

Definition 4.3.24. The complex $C'$ is a covering complex of $C$ if the corresponding complex graph of $C'$, $G_{C'}$, is a covering complex graph of the corresponding complex graph of $C$, $G_C$.

Definition 4.3.25. Let $P$ be a pot. A complete complex $C = \langle T, J \rangle \in \mathcal{C}(P)$ is a minimal complete complex if every other complete complex $C' = \langle T', J' \rangle \in \mathcal{C}(P)$ such that $T$ and $T'$ have the same set of tile types, is a covering complex of $C$.

The set of minimal complete complexes of $\mathcal{C}(P)$ is denoted by $\mathcal{MC}(P)$.

Definition 4.3.26. Let $P$ be a pot. A complete complex $C = \langle T, J \rangle \in \mathcal{C}(P)$ is a maximal complete complex if it is not a covering complex of any other complex from $P$, and every other complete complex $C' = \langle T', J' \rangle \in \mathcal{C}(P)$ such that $T \subseteq T'$ and $J \subseteq J'$ is a covering complex of $C$. (viz, there are no bigger complexes with the same tiles, except the covering complexes.)

Proposition 4.3.27. The pot graph $\mathcal{G}_P$ of a strongly satisfiable pot $P$ with a single point spectrum is connected.

Proof. In Chapter 3 we discussed that if the spectrum of a strongly satisfiable pot $P$ is a single point, then every complete complex in $\mathcal{C}(P)$ contain a tile from every tile type. Consider two tiles $t_1$ and $t_2$ with corresponding vertices $v_1$ and $v_2$ in $\mathcal{G}_P$. If there is no path between $v_1$ and $v_2$, then $t_1$ and $t_2$ must be on two different complete complexes, which is not possible. Therefore, the pot graph of a strongly satisfiable pot with a single point spectrum is connected.

\[\blacksquare\]
Definition 4.3.28. A separating set of a complex $C = \langle T, J \rangle$ is a set of connections $S \subset J$ such that $\langle T, J - S \rangle$ is not a connected complex. The connectivity of a complex $C$, $k(C)$, is the minimum size of a separating set $S$ s.t. $J - S$ dissolves $C$ into two complexes. A complex is $k$-connected if its connectivity is at least $k$. A pot $P$ is $k$-connected if all of its complete complexes are $k$-connected.

Remark 4.3.29. If $P$ is a strongly satisfiable pot such that every incomplete complex has at least two free sticky ends, then every complete complex of $P$ is 2-connected.

Proposition 4.3.30. Let $P$ be a 2-connected strongly satisfiable pot type. The pot graph of $P$ is connected if and only if there is a connected complete complex $C \in \mathcal{C}(P)$ that contains all the tile types.

Proof. Suppose that the pot graph, $G_P$, of $P$ is connected and there is no complete complex in $\mathcal{C}(P)$ that contains all the tile types. We claim that there exist maximal connected complete complexes $C_i = \langle T_i, J_i \rangle$ for $i \in [n]$ such that for every tile type, there is a tile that is embedded in one of the maximal complete complexes and no two of the maximal complexes can have a tile of the same type, moreover no two of the maximal complexes can have a sticky end of the same type in common.

Let $C_{i_k} = \langle T_{i_k}, S_{i_k}, J_{i_k} \rangle$ and $C_{i_m} = \langle T_{i_m}, S_{i_m}, J_{i_m} \rangle$ be maximal complete complexes that have a sticky end of type $h \in H$ in common. There exist tiles $t_1, t'_1 \in T_{i_k}$ and $t_2, t'_2 \in T_{i_m}$ such that $t_1(h) > 0$, $t'_1(\hat{h}) > 0$, $t_2(h) > 0$, and $t'_2(\hat{h}) > 0$. Since $P$ is a 2-connected strongly satisfiable pot type, the tiles $t_1$ and $t'_2$ can connect, and the tiles $t_2$ and $t'_1$ can connect forming a complex $C = \langle T_{i_k} \cup T_{i_m}, S_{i_k} \cup S_{i_m}, J \rangle$. That is not possible, because the complexes $C_{i_k}$ and $C_{i_m}$ are maximal complete complexes. We can conclude that no two maximal complete complexes share a sticky end of the same type in common, therefore they cannot share a tile of the same type in common.
Let $G_{C_i}$ for $i \in [n]$ be the corresponding complete complex graphs for the maximal complete complexes, and $\varphi$ be the homomorphism from complete complex graphs and pot graphs. Then $\varphi(G_{C_i})$ for $i \in [n]$ are pairwise disjoint, since they don’t have tiles nor free sticky end of the same types. Also there is no vertex or an edge that is in $G_P$ that it is not in some of the $\varphi(G_{C_i})$. This is not possible since $G_P$ is connected. Hence there is a complete complex in $C(P)$ that contains tiles of every tile type.

Suppose that there exists a connected complete complex $C = \langle T, S, J \rangle \in C(P)$ that contains tiles of every tile type. The complete complex graph of $C$, $G_C$, is mapped onto with respect to the vertices to $G_P$ and since it is connected, for every $t, t' \in T$, there exists a path connecting them, i.e. for every $v_t, v_{t'} \in V$ there exists a path connecting them (because we correspond and edge to every connection). Therefore $G_P$ is connected.

The induced subgraph for a covering of a minimal complete complex will be the
same as the induced subgraph for the minimal complete complex. Therefore, from an induced subgraph we can obtain information only about the “minimal” complete complexes.

From the spectrum itself and from the support of a given pot $P$ we are not able to find all the information about the products of self-assembly. The study of pot graphs expands our view, and in conjunction with other studies can give more complete picture of the possible complexes assembling from a pot of tiles. The following is an example of two pot types with same spectrum and same support, but different complete complexes. We can conclude that they have different complete complexes only because the corresponding pot graphs are not equivalent.

Figure 4.15: pot types that are not similar, but with the same spectrum $S = \{(\frac{1}{2} - u, \frac{1}{2} - u, u, u) : 0 \leq u \leq \frac{1}{2}\}$ and same support $Supp = \{(1, 1, 1, 1)\}$

In this chapter we just scratched the surface of pot graphs. We plan to extend the this study and to connect it with the study of the spectrum and probability.
For better understanding of the self-assembly process as well as possibly to predict what we should expect to obtain within a pot with DNA molecules one needs to study the distribution of the possible outcomes of the process.

At the end of the experiment there may be some complete complexes of the desired type, some other complete complexes and some incomplete complexes. Since our major concern is the construction of complete complexes (of certain sizes), we want to explore the proportion of each complete complex, for the purpose of evaluating the results of the experiment. We approach this problem by considering a special case, graph assembly of uniformly distributed tiles, that can assemble into cyclic complexes. We believe that more complicated structures have additional geometric and other intrinsic constraints that would make our assumptions unrealistic and superfluous. Good conditions for studying such complicated structures remain to be discovered.

In this chapter we raise questions relating the probability of obtaining a cyclic complete complex and approach them using two different methods. Although this is a first theoretical study of the probability of self-assembled complexes, it is not the first study of the probability of obtaining cyclic molecules [13, 16, 17, 46, 47, 51, 52, 58, 59]. The distribution of linear and cyclic molecules provides a better understanding of the flexibility and intrinsic properties of the DNA molecules. Also the correlation between the concentration of the molecules and the probability of two sticky ends connecting can be better realized by the same distribution.

The DNA molecule in living organisms usually appears as a cyclic molecule of
different sizes. The configuration of those molecules and the probability of their formations is of interest to many biologists, chemists, mathematicians, and physicists.

The first paper concerning the probability of ring formation of DNA molecules appeared in 1950 by H. Jacobson and W. H. Stockmayer [16, 17]. They introduced the notion of ring closure probability of DNA molecules which is known as the $j$-factor. It is defined as the ratio of two equilibrium constants $K_c$ and $K_a$, where $K_c$ is the cyclization constant and $K_a$ is the bimolecular equilibrium constant for joining two molecules.

Few years later, J. C. Wang and N. Davidson [51, 52] used thermodynamic and kinetic properties of DNA molecules in the study of ring closure probability or cyclization. They measured the entropy and enthalpy change of DNA molecules during the interconversion from linear into circular molecules. Their experiments were similar to H. Jacobson and W. H. Stockmayer’s. Their work involved three different types of DNA molecules; one type had two sticky ends and those molecules were able to form cycles, while the other two types had one sticky end each (for example one of them had the sticky end on the left side, and the other one had it on the right side) and one blunt end, and those types of molecules were not able to configure into cycles, but rather to self-assemble into linear structures with two blunt ends.

D. Shore and R. L. Baldwin [46, 47] achieved a striking result showing that the $j$-factor mainly depends on the fractional twist of the DNA molecule: the difference between the total helical twist and the nearest integer. They were also able to distinguish different topoisomers obtained at the end of the experiment. In their research, they studied series of 12 linear DNA molecules that had complementary sticky ends on each side, and the DNA molecules differed only in their length.

A. Dugaiczyk, H. W. Boyer, and H. M. Goodman [13] studied the probability of a ring closure of DNA molecules. They considered only one type of DNA molecule in their experiments, and it was a linear molecule with complementary sticky ends on the sides. Their results showed that the $j$-factor depends on the contour length of the molecule, the random coli segment and the molar concentration of the solution.
The ring closure probability has been computed as a function of molecular length of a DNA molecule by Yamakawa and Stockmayer [58, 59]. They considered the molecule as a wormlike model and defined the $j$-factor as a Green’s function of the molecular length.

The cyclization of the DNA molecule has been the topic of scientific research for many years. Unfortunately we still don’t have a precise description of the process and of the conformation the DNA molecule takes. In the next few sections we outline the calculation of the probability of self-assembling cyclic molecules of different sizes.

For simplicity we avoid thermodynamic properties and consider only the self-assembly process for which all Watson-Crick connections are equally likely and no free sticky ends remain after the completion of the experiment. That means that after self-assembly has occurred only complete complexes are present. We propose a static model (similar to models studied in discrete probability theory) with uniformly distributed tiles, and Watson-Crick complementary pairs. Each sticky end has equal probability to connect.

The evolution of the self-assembly process is not examined, only distribution of obtained complete complexes at the end of the process is considered. We are mainly interested in the input and the output of the process. We provide some insight to the question: If we have a certain amount of tile molecules, how much of each complex types are there in the outcome?

In this chapter we only provide theoretical results, while experimental results are given in the next chapter. The existence of the probability space is demonstrated in Appendix B. In the next section we address the questions raised in the introduction by two different methods.

One Type Pot

To explore the static model, we begin with a basic set-up, pot type $P = \{t\}$ of only one tile type and $H = \{h, \hat{h}\}$ the set of sticky ends ($\hat{h}$ denotes the W-C complement of $h$), s.t. for the given tile type $t(h) = t(\hat{h}) = 1$. Assume that we have $m$ tiles in
Let \( t \) be a tile from the tile type \( t \), set

\[
P(t \text{ is on a cycle of length } k) = r_{m,k}.
\]

**Expectation**

Denote by:

- \( A_{m,t,k} \): the event that the tile \( t \) is on a \( k \)-cycle,
- \( I_{m,t,k} \): the associated indicator random variable,
- \( I_{m,k} = \sum_t I_{m,t,k} \): the number of tiles in \( k \)-cycles,
- \( X_{m,k} = \frac{I_{m,k}}{k} \): the number of \( k \)-cycles.

Then

\[
E(I_{m,t,k}) = P(I_{m,t,k} = 1) = r_{m,k}, \quad \text{and}
\]

\[
E(I_{m,k}) = \sum_t E(I_{m,t,k}) = m r_{m,k}, \quad \text{hence}
\]

\[
E(X_{m,k}) = \frac{1}{k} E(I_{m,k}) = \frac{m}{k} r_{m,k} \quad \text{is the expected number of } k - \text{cycles.}
\]

**Proposition 5.1.1.** Let \( P \) be a pot (of type \( P = \{t\} \)) which contains \( m \) 2-branched tiles of type \( t \) and \( H = \{h, \hat{h}\} \) be the set of sticky end types for the pot \( P \), such that \( t(h) = t(\hat{h}) = 1 \). Let \( X_{m,k} \) denote the number of cycles of length \( k \) in \( P \) and \( r_{m,k} \) the probability that a given tile from the pot \( P \) is on a cycle of length \( k \). Then

\[
E(X_{m,k}) = \frac{m}{k} r_{m,k}, \quad k \in [m].
\]

\[
\text{Var}(X_{m,k}) = \frac{m}{k} r_{m,k} (1 - r_{m,k}).
\]

**NOTE 1**

We assume that during the self-assembly process all Watson-Crick connections are established and that only complete complexes are obtained. That means after the
annealing process every tile is on a complete cycle, i.e., \( \sum_{k=1}^{m} r_{m,k} = 1 \). Moreover since we have a total of \( m \) tiles and \( \frac{mr_{m,k}}{k} \) \( k \)-cycles in the pot (in a \( k \)-cycle, \( k \) tiles are engaged)
\[
\sum_{k=1}^{m} k \frac{mr_{m,k}}{k} = m,
\]
should hold, which is anyway satisfied, since
\[
\sum_{k=1}^{m} k \frac{mr_{m,k}}{k} = m \sum_{k=1}^{m} r_{m,k} = m.
\]

**NOTE 2**
Based on the outcomes of some experiments, we can assume that probability of obtaining a smaller cycle is greater than the probability of obtaining a bigger cycle. Therefore probability of one tile being on a smaller cycle is greater than the probability of being on a larger cycle i.e., \( r_{m,1} \geq r_{m,2} \geq \cdots \geq r_{m,m} \).

**Proof. Variance**

From the definition of the events, since \( X_{m,k} = \frac{I_{m,k}}{k} \),
\[
\Var(X_{m,k}) = \frac{1}{k^2} \Var(I_{m,k}) = \frac{1}{k^2} \Var(\sum_{t} I_{m,t,k}).
\]

From the properties of the variance, it follows that
\[
\Var(X_{m,k}) = \frac{1}{k^2} \sum_{t_1,t_2} \Cov(I_{m,t_1,k}, I_{m,t_2,k}).
\]

When \( t_1 = t_2 \),
\[
\Cov(I_{m,t_1,k}, I_{m,t_1,k}) = \Var(I_{m,t_1,k}) = E(I_{m,t_1,k}^2) - (E(I_{m,t_1,k}))^2 = r_{m,k}(1 - r_{m,k}),
\]

74
while when \( t_1 \neq t_2 \),

\[
\text{Cov}(I_{m,t_1,k}, I_{m,t_2,k}) = E(I_{m,t_1,k}I_{m,t_2,k}) - E(I_{m,t_1,k})E(I_{m,t_2,k}).
\]

By the definition of the expectation,

\[
E(I_{m,t_1,k}I_{m,t_2,k}) = 1P(I_{m,t_1,k}I_{m,t_2,k} = 1) + 0P(I_{m,t_1,k}I_{m,t_2,k} = 0) = P(I_{m,t_1,k}I_{m,t_2,k} = 1) = P(I_{m,t_1,k} = 1)P(I_{m,t_2,k} = 1|I_{m,t_1,k} = 1).
\]

Given that the tile \( t_1 \) is on a \( k \)-cycle, the probability that \( t_2 \) is also on a \( k \)-cycle is equal to the probability that \( t_2 \) is on the same cycle as \( t_1 \) (which is one of the other \( k-1 \) tiles on the cycle) or it is one of the remaining \( m-k \) tiles that are joined in other \( k \)-cycle (each one with probability \( r_{m-k,k} \) being on a cycle of length \( k \)).

Therefore, \( P(I_{m,t_2,k} = 1|I_{m,t_1,k} = 1) = \frac{k-1}{m-1} + \frac{m-k}{m-1}r_{m-k,k} \).

Hence for \( k \leq \frac{m}{2} \) (in this case two tiles can be either on a same cycle or on two different ones) we obtain

\[
\text{Cov}(I_{m,t_1,k}, I_{m,t_2,k}) = r_{m,k} \left( \frac{k-1}{m-1} + \frac{m-k}{m-1}r_{m-k,k} \right) - r_{m,k}^2
= r_{m,k} \left( \frac{k-1}{m-1} + \frac{m-k}{m-1}r_{m-k,k} - r_{m,k} \right)
\]

while for \( k > \frac{m}{2} \) (in this case two tiles can be only on a same cycle),

\[
\text{Cov}(I_{m,t_1,k}, I_{m,t_2,k}) = r_{m,k} \frac{k-1}{m-1} - r_{m,k}^2 = r_{m,k} \left( \frac{k-1}{m-1} - r_{m,k} \right).
\]
Now,

\[ \text{Var}(X_{m,k}) = \frac{1}{k^2} \sum_{t_1, t_2} \text{Cov}(I_{m,t_1,k}, I_{m,t_2,k}) \]

\[ = \frac{1}{k^2} \left[ \sum_{t_1=t_2} \text{Cov}(I_{m,t_1,k}, I_{m,t_2,k}) + \sum_{t_1 \neq t_2} \text{Cov}(I_{m,t_1,k}, I_{m,t_2,k}) \right] \]

\[ = \frac{1}{k^2} \left[ mr_{m,k}(1 - r_{m,k}) + m(m-1)r_{m,k}\left(\frac{k-1}{m-1} + \frac{m-k}{m-1}r_{m-k,k} - r_{m,k}\right) \right] \]

**NOTE 3**

Usually in a pot consisting of a large number of tile molecules (say \(10^{15}\)) we expect that if a \(k\)-cycle is admitted, then almost surely we will observe many other \(k\)-cycles i.e., \(k \ll \frac{m}{2}\). Moreover if a \(k\)-cycle is obtained in the outcome of the pot, then with great probability cycles of smaller length than \(k\) are also obtained. Hence we can make our assumption even stronger and assume that only cycles of length much smaller than \(m\) are assembled in the pot i.e., \(k \ll m\). This will lead towards approximation of the probabilities \(r_{m,k}\) by

\[ \lim_{m \to \infty} \frac{r_{m-k,k}}{r_{m,k}} = 1, \]

\[ \text{Cov}(I_{m,t_1,k}, I_{m,t_2,k}) = r_{m,k}\left(\frac{k-1}{m-1} + \frac{m-k}{m-1}r_{m-k,k} - r_{m,k}\right) \]

for \(k \ll m\). Using the fact that for \(k \ll m\),

\[ \text{Var}(X_{m,k}) = \frac{m}{k^2} r_{m,k}(1 - r_{m,k} + k - 1 + (m-k)r_{m,k} - (m-1)r_{m,k}) \]

\[ = \frac{m}{k} r_{m,k}(1 - r_{m,k}). \]
Multi Type Pot

Let \( P = \{t_1, t_2, \ldots t_n\} \) be a pot type, and \( P \) a pot that contains \( m \) different 2-armed tiles from types in \( P \) from which an \( n \)-cycle can be constructed. Let \( H = \{h_1, h_2, \ldots h_n, \hat{h}_1, \hat{h}_2 \ldots \hat{h}_n\} \) be the set of sticky end types such that \( t_i(h_i) = t_i(\hat{h}_{i+1}) = 1 \) for \( i \in [n-1] \), and \( t_n(h_n) = t_n(\hat{h}_1) = 1 \). We assume to have uniformly distributed tile types i.e, \( n \) tiles from each type, and

\[
P(t \text{ is on a cycle of length } k) = r_k.
\]

From the selection of the tiles, it is clear that only cycles whose lengths are multiples of \( m \) are assembled, i.e, \( r_l = 0 \) if \( m \nmid l \).

Denote by

- \( A_{t,k} \): the event that the tile \( t \) is on a \( k \)-cycle,
- \( I_{t,k} \): the associated indicator random variable,
- \( I_k = \sum_t I_{t,k} \): the number of tiles in a \( k \)-cycle,
- \( X_k = \frac{I_k}{k} \): the number of \( k \)-cycles.

We can calculate

\[
E(I_{t,kn}) = P(I_{t,kn} = 1) = r_{kn}
\]

\[
E(I_{kn}) = \sum_t E(I_{t,kn}) = mn r_{kn}
\]

\[
E(X_{kn}) = E \left( \frac{I_{kn}}{kn} \right) = \left( \frac{mn}{kn} r_{kn} \right) = \frac{m}{k} r_{kn}.
\]

### 5.2 Another Method for Obtaining the Expected Number of Cycles

To explore the static model, we begin with a basic set-up, pot type \( P = \{t\} \) of only one tile type and \( H = \{h, \hat{h}\} \) (\( \hat{h} \) denotes the W-C complement of \( h \)), s.t. for the given tile type \( t(h) = t(\hat{h}) = 1 \). Assume that we have \( m \) tiles in the given solution.
Let $t$ be a tile from the tile type $t$, set

$$P(t \text{ is on a cycle of length } k) = r_k$$

$S_k$: a set of $k$ tiles from type in $P$

$A_{S_k}$: the event that the tiles from $S_k$ form a cycle of length $k$.

$X_{S_k}$: the associated random variable for $A_{S_k}$.

Expectations

- **CASE 1: $S_1 = \{t_1\}$**

Since $X_{S_1}$ is the indicator random variable for the event $A_{S_1}$, $X_1 = \sum_{S_1} X_{S_1}$ will denote the number of cycles of length 1 in the pot i.e, monomers and

$$E(X_{S_1}) = P(X_{S_1} = 1) = r_1.$$ 

We have $m$ sets $S_1$ and by the linearity of the expectation the expected number of monomers in the pot is

$$E(X_1) = \sum_{S_1} E X_{S_1} = m r_1.$$ 

- **CASE 2: $S_2 = \{t_1, t_2\}$**

Since $X_{S_2}$ is the indicator random variable for the event $A_{S_2}$, $X_2 = \sum_{S_2} X_{S_2}$ will denote the number of cycles of length 2 in the pot i.e, dimers and

$$E(X_{S_2}) = P(X_{S_2} = 1)$$

Note: Given that $t_1$ is on a dimer, we know that it is linked to another tile, so as the probability that is linked to the tile $t_2$, for any particular tile $t_2 \neq t_1$, is

$$\frac{1}{m - 1},$$

so
\[
E(X_{S_2}) = P((t_1 \text{ and } t_2 \text{ are linked}) \land (t_1 \text{ is on a dimer})) \\
= P(t_1 \text{ is on a dimer})P(t_1 \text{ and } t_2 \text{ are linked} | t_1 \text{ is on a dimer}) \\
= r_2 \frac{1}{m-1} = \frac{r_2}{m-1}.
\]

We have \(\binom{m}{2}\) sets \(S_2\) and by the linearity of the expectation the expected number of dimers in the pot is

\[
E(X_2) = \sum_{S_2} EX_{S_2} = \binom{m}{2} \frac{r_2}{m-1} = \frac{m}{2}r_2.
\]

**CASE 3: \(S_3 = \{t_1, t_2, t_3\}\)**

\[
E(X_{S_3}) = P(X_{S_3} = 1)
\]

Note: Given that \(t_1\) is on a trimer, we know that it is linked to two other tile molecules, the probability that is linked via the sticky end of type \(\hat{h}\) to the tile \(t_2\), for some \(t_2 \neq t_1\) is \(\frac{1}{m-1}\). Given that \(t_1\) is on a trimer and that it is linked to \(t_2\), we know that it must be also linked to one other tile via the sticky end of type \(\hat{h}\), the probability that this other molecule is \(t_3\) is \(\frac{1}{m-2}\), for given \(t_3\). There are two possible trimers obtained from these tiles \((t_1t_2t_3\text{ and } t_1t_3t_2)\). Therefore

\[
E(X_{S_3}) = 2P((t_1 \text{ and } t_2 \text{ are linked}) \land (t_1 \text{ and } t_3 \text{ are linked}) \land (t_1 \text{ is on a trimer})) \\
= 2P(t_1 \text{ is on a trimer})P((t_1 \text{ and } t_2 \text{ are linked})|(t_1 \text{ is on a trimer})) \times \\
2P((t_1 \text{ and } t_3 \text{ are linked})|(t_1 \text{ is on a trimer}) \land (t_1 \text{ and } t_2 \text{ are linked})) \\
= 2r_3 \frac{1}{m-1} \frac{1}{m-2} = \frac{2r_3}{(m-1)(m-2)}.
\]

We have \(\binom{m}{3}\) sets \(S_3\) and by the linearity of the expectation the expected number
of trimers in the pot is

\[ E(X_3) = \sum_{S_3} E X_{S_3} = \binom{m}{3} \frac{2r_3}{(m - 1)(m - 2)} = \frac{m}{3} r_3. \]

**GENERAL CASE FOR** \( k \): \( S_k = \{t_1, t_2, \ldots, t_k\} \)

Calculating the expectation for this case will go the same way as for all previous cases up to permutation. This tile molecules can form \((k - 1)! \) “different” \( k \)-cycles. Therefore

\[ E(X_{S_k}) = (k - 1)! \frac{r_k}{(m - 1)(m - 2) \ldots (m - k + 1)} \]

\[ = (k - 1)! \frac{1}{(k - 1)!} \frac{r_k}{(m - 1) \binom{m - 1}{k - 1}} \]

\[ = \frac{m}{k} \frac{r_k}{\binom{m}{k}}. \]

We have \( \binom{m}{k} \) sets \( S_k \) and by the linearity of the expectation the expected number of \( k \)-cycles in the pot is

\[ E(X_k) = \sum_{S_k} E X_{S_k} \]

\[ = \binom{m}{k} \frac{m}{k} \frac{r_k}{\binom{m}{k}} \]

\[ = \frac{m}{k} r_k. \]

**Proposition 5.2.1.** Let \( P = \{t\} \) be a pot which contains \( m \) 2-branched tiles of type \( t \) and \( H = \{h, \hat{h}\} \) be the set of sticky end types for the pot \( P \), such that \( t(h) = t(\hat{h}) = 1 \). Let \( X_k \) denote the number of cycles of length \( k \) in the \( P \) and \( r_k \) the probability that a given tile is on a cycle of length \( k \). Then

\[ E(X_k) = \frac{m}{k} r_k, k \in [m]. \]
OBSERVATION

We assumed that during the self-assembly process all Watson-Crick connections are established and that only complete complexes are obtained. That means that after the annealing process every tile is on a complete cycle i.e., \( \sum_{k=1}^{m} r_k = 1 \). Moreover, since we have a total of \( m \) tiles and \( \frac{mr_k}{k} \) \( k \) cycles in the pot (in a \( k \)-cycle \( k \) molecules are engaged) we are not surprised that

\[
\sum_{k=1}^{m} k \frac{mr_k}{k} = m \sum_{k=1}^{m} r_k = m.
\]

Variance

- **GENERAL CASE FOR \( \text{VAR} \ (X_k) \ (2k \leq m) \):**

  \(- S_k \cap T_k = \emptyset\)

  We have \( \binom{m}{k} \binom{m-k}{k} \) choices for this kind of sets, and in this case

  \[
  E(X_{S_k}X_{T_k}) = P(X_{S_k} = 1, X_{T_k} = 1) = P(X_{S_k} = 1|X_{T_k} = 1)P(X_{T_k} = 1) = \frac{(m-k)r_k}{k^{\binom{m-k}{k}}} \frac{mr_k}{k^{\binom{m}{k}}}.
  \]

  \[
  EX_{S_k}EX_{T_k} = \frac{m^2}{k^2 \binom{m}{k}^2} r_k^2, \text{ hence}
  \]

  \[
  \text{Cov}(X_{S_k}, X_{T_k}) = \frac{(m-k)r_k}{k^{\binom{m-k}{k}}} \frac{mr_k}{k^{\binom{m}{k}}} - \frac{m^2}{k^2 \binom{m}{k}^2} r_k^2, \text{ and hence}
  \]

81
\[
\sum_{S_k \cap T_k = \emptyset} \text{Cov}(X_{S_k}, X_{T_k}) = \left( \binom{m}{k} \right) \left( \frac{m}{k} \right) \frac{(m-k)r_k}{k(m-k)} \frac{mr_k}{k(m-k)} - \left( \binom{m}{k} \right) \left( \frac{m}{k} \right) k^2 \binom{m}{k}^2 r_k^2
\]
\[
= \frac{m(m-k)}{k^2} r_k^2 - \left( \binom{m}{k} \right) \frac{m^2}{k^2} r_k^2.
\]

\[0 < |S_k \cap T_k| < k\]

We have \(\left( \binom{m}{k}^2 - \binom{m}{k} - \binom{m}{k} \left( \binom{m}{k} - \binom{m}{k} \right) \right) = \binom{m}{k} \left( \binom{m}{k} - \binom{m}{k} - 1 \right)\) choices for this kind of sets, and in this case

\[E(X_{S_k}X_{T_k}) = P(X_{S_k} = 1, X_{T_k} = 1)\]
\[= P(X_{S_k} = 1|X_{T_k} = 1)P(X_{T_k} = 1)\]
\[= 0, \text{ and}\]

\[EX_{S_k} EX_{T_k} = \frac{m^2}{k^2} \binom{m}{k}^2 r_k^2, \text{ hence}\]

\[\text{Cov}(X_{S_k}, X_{T_k}) = -\frac{m^2}{k^2} \binom{m}{k}^2 r_k^2, \text{ and hence}\]

\[\sum_{0 < |S_k \cap T_k| < k} \text{Cov}(X_{S_k}, X_{T_k}) = -\left( \binom{m}{k} \right) \left( \binom{m}{k} - \binom{m}{k} \right) - 1 \frac{m^2}{k^2} \binom{m}{k}^2 r_k^2\]
\[= -\frac{m^2}{k^2} r_k^2 + \left( \binom{m}{k} \right) \frac{m^2}{k^2} \binom{m}{k} r_k^2 + \frac{m^2}{k^2} \binom{m}{k}^2 r_k^2.\]

\[- S_k = T_k\]

We have \(\binom{m}{k}\) choices for this kind of sets, and in this case
\[
E(X_{S_k}X_{T_k}) = P(X_{S_k} = 1, X_{T_k} = 1) \\
= P(X_{S_k} = 1|X_{T_k} = 1)P(X_{T_k} = 1) \\
= \frac{m}{k^2(m)}r_k, \text{ and} \\
EX_{S_k}EX_{T_k} = \frac{m^2}{k^2(m)}r^2_k, \text{ hence} \\
Cov(X_{S_k}, X_{T_k}) = \frac{m}{k^2(m)}r_k \left(1 - \frac{m}{k^2(m)}r_k \right), \text{ and hence} \\
\sum_{S_k=T_k} Cov(X_{S_k}, X_{T_k}) = \left(\frac{m}{k}\right) \frac{m}{k^2(m)}r_k \left(1 - \frac{m}{k^2(m)}r_k \right) \\
= \frac{m}{k}r_k \left(1 - \frac{m}{k^2(m)}r_k \right) \\
= \frac{m}{k}r_k - \frac{m^2}{k^2(m)}r^2_k.
\]

CONCLUSION

\[
\text{Var}(X_k) = \sum_{S_k,T_k} Cov(X_{S_k}, X_{T_k}) \\
= \frac{m(m-k)}{k^2}r^2_k - \left(\frac{m-k}{k}\right) \frac{m^2}{k^2(m)}r^2_k - \\
\frac{m^2}{k^2}r^2_k + \left(\frac{m-k}{k}\right) \frac{m^2}{k^2(m)}r^2_k + \frac{m^2}{k^2}r^2_k - \frac{m^2}{k^2(m)}r^2_k \\
= \frac{m(m-k)}{(k)^2}r^2_k - m^2(k^2) + m^2(k^2) - \frac{m^2}{k^2}r^2_k + \frac{m^2}{k^2}r^2_k + m^2r^2_k - \frac{m^2}{k^2}r^2_k \\
= \frac{m}{k}r_k(1 - r_k).
\]
\[
\text{Var}(X_k) = \frac{m}{k} r_k (1 - r_k)
\]

**GENERAL CASE FOR VAR \((X_k)(2k > m):\)**

In this case the only possible outcome is one cycle of length \(m\). It is not possible to have two disjoint sets of \(k\) elements, just one with \(m\) elements \(S_m = T_m\). From the previous observation, we obtain

\[
\text{Var}(X_m) = \sum_{S_m = T_m} \text{Cov}(X_{S_m}, X_{T_m})
= \frac{m}{m} r_m - \frac{m^2}{m^2(m)} r_m^2
= r_m (1 - r_m).
\]

Our assumption is that \(r_m = 0\), so we can conclude that \(\text{Var}(X_m) = 0\), and 
\[
E(X_m) = \frac{m}{m} \frac{r_m}{m \binom{m}{m}} = 0.
\]

**5.3 Theoretical Base for the Experimental Results**

In this section we establish the theoretical base for the experimental results given in the next chapter. We begin with a pot containing uniformly distributed DNA molecules compatible of forming a cyclic graph structure. After the annealing process it is assumed that no free sticky ends remain. We prove that under certain probability conditions almost all structures represent the originally encoded graph i.e., the appearance of dimer (double cover) or trimer (triple cover) molecules is with small probability.

The main assumption of the model is that the probability \(r\) that the last of the possible connections within a complex appears after the other connections have been established is very high. Under these conditions we show that probability of appear-
ance of dimers and trimers in a pot designed to form monomer cycles approaches 0.

We start with a special case of obtaining cyclic molecules with three 2-armed tiles. This corresponds to building a triangle. For this purpose, we consider three different types of 2-armed tile types $P = \{t_1, t_2, t_3\}$ which contain 3 different types of complementary free sticky ends $H = \{h_1, h_2, h_3, \hat{h}_1, \hat{h}_2, \hat{h}_3\}$ These tiles are uniformly distributed in a pot and are capable of admitting a complete $K_3$ complex, meaning that we have equal amount of tiles from each tile type. We conveniently represent this amount with an integer $m$. The sticky end types are adequately arranged (see Fig. 5.1).

![Figure 5.1: (a) Three 2-armed tiles form a triangle which represents a $K_3$ complex. (b) Three tile graphs used in a pot to assemble $K_3$. (c) Complete complexes for this pot will be cycles of length divisible by 3. Cycles $K_3$ and $C_6$ are depicted.](image)

$$t_1(h_1) = t_1(\hat{h}_2) = 1,$$
$$t_2(h_2) = t_2(\hat{h}_3) = 1,$$
$$t_3(h_3) = t_3(\hat{h}_1) = 1.$$

With this kind of selection for the tile molecules, complete complexes that are obtained would be cyclic and would involve $3k$ tiles for some $k$ ($1 \leq k \leq m$). Tiles from types in $P$ are always assembling according to a specific pattern, $t_1t_2t_3$ repeatedly or $t_1t_3t_2$ repeatedly, depending on the orientation. We say that a cycle is of length $3k$, if it has $k$ tiles from each type adequately arranged, for example $t_1t_2t_3$ repeatedly.
times, such that the last $t_3$ tile is glued to the first $t_1$ tile. We will use the notation $C_{3k}$ for cycles of length $3k$, $k \geq 1$, (Note: For a cycle of length 3, we will use the notation $K_3$.)

We employ the probabilistic method, often used in random graph theory [18, 49] to obtain the results. We start our analysis by computing the probability of appearance of at least one $K_3$ complex in the pot described above.

**Proposition 5.3.1.** Let $P = \{t_1, t_2, t_3\}$ be a pot type and $P$ a pot which contains uniformly distributed 2-armed tiles of type in $P$ capable of admitting a $K_3$ complex. Let $X$ denote the number of complete $K_3$ complexes in $P$, and $r$ the probability that three connected tiles by two sticky ends will close in a complete $K_3$ complex. Under assumption that the conditional probability of the second connection is the same as the probability of the first, the expected number of $K_3$ complete complexes in the pot is

$$E(X) = mr;$$

moreover, when $r = 1$, $X = m$ almost surely, where $m$ denotes the amount of tiles in $P$ of each type.

**Proof.** For the proof we use the following notation:

- $S$: a set of 3 tiles from $t_3$, one of each type,
- $A_S$: the event that the tiles from $S$ form a complete $K_3$,
- $X_S$: the associated indicator random variable for $A_S$,
- $B_i$: the event that $h_i$ and $\tilde{h}_i$ will connect, for $i = 1, 2, 3$ and
- $\xi_i$: the associated indicator random variable for $B_i$.

For the set of sticky end types $H = \{h_1, h_2, h_3, \tilde{h}_1, \tilde{h}_2, \tilde{h}_3\}$ in $S$, the probability that the tiles from $S$ will form a complete $K_3$ is equal to the probability that all tiles
of $S$ would connect. That is,

$$P(\xi_1 = 1, \xi_2 = 1, \xi_3 = 1) = P(\xi_1 = 1)P(\xi_2 = 1|\xi_1 = 1)P(\xi_3 = 1|\xi_1 = 1, \xi_2 = 1)$$

$$= pp = p^2 r.$$  

(Notice we assume the conditional probability of the second connection is the same as the probability of the first).

Since $X_S$ is the indicator random variable for the event $A_S$, $X = \sum X_S$ will denote the number of complete $K_3$'s in the pot, and

$$E(X_S) = P(A_S) = p^2 r.$$  

We have $m^3$ sets $S$ and by the linearity of expectation the expected number of complete $K_3$'s in the pot is

$$E(X) = m^3 p^2 r.$$  

Ignoring the thermodynamic properties of the solution, the probability of one sticky end connecting with its complementary is $p = \frac{1}{m}$, from which it follows that $E(X) = mr$.

To calculate the variance for the number of complete $K_3$’s in the pot

$$\text{Var}(X) = \sum_{S,T} \text{Cov}(X_S, X_T)$$

we need to calculate the covariances first:

$$\text{Cov}(X_S, X_T) = E(X_S X_T) - E(X_S)E(X_T).$$

In order to do that we need to look at two sets $S$ and $T$, each one consisting of the three different tiles from the pot, one from each type. Again, for the analysis of the covariance we consider the case when $p = \frac{1}{m}$.

- Case 1: $S \cap T = \emptyset$
We have $m^3(m-1)^3$ choices for this kind of pair of sets, and in this case

$$E(X_S X_T) = P(X_S = 1, X_T = 1)$$

$$= P(X_S = 1|X_T = 1)P(X_T = 1)$$

$$= \frac{r}{(m-1)^2} \frac{r}{m^2} = \frac{r^2}{m^2(m-1)^2},$$

and

$$E(X_S)E(X_T) = \frac{r}{m^2} \frac{r}{m^2} = \frac{r^2}{m^4};$$

hence

$$Cov(X_S, X_T) = \frac{r^2}{m^2(m-1)^2} - \frac{r^2}{m^4} = \frac{r^2(2m-1)}{m^4(m-1)^2};$$

$$\sum_{S \cap T = \emptyset} Cov(X_S, X_T) = m^3(m-1)^3 \frac{r^2(2m-1)}{m^4(m-1)^2}$$

$$= \frac{r^2(2m-1)(m-1)}{m}.$$

- **Case 2:** $|S \cap T| = 1$

  We have $m^3\binom{m}{1}(m-1)^2 = 3m^3(m-1)^2$ choices for this kind of set, and we get:

  $$E(X_S X_T) = P(X_S = 1, X_T = 1)$$

  $$= P(X_S = 1|X_T = 1)P(X_T = 1)$$

  $$= 0,$$

  as $P(X_S = 1|X_T = 1) = 0$, and as $E(X_S)E(X_T) = \frac{r}{m^2} \frac{r}{m^2} = \frac{r^2}{m^4},$

  $$\sum_{|S \cap T| = 1} Cov(X_S, X_T) = -3m^3(m-1)^2 \frac{r^2}{m^4}$$

  $$= -3 \frac{(m-1)^2 r^2}{m}.$$

- **Case 3:** $|S \cap T| = 2$

88
We have $m^3\binom{3}{2}(m-1) = 3m^3(m-1)$ choices for this kind of pair of sets, so

$$E(X_SX_T) = P(X_S = 1, X_T = 1)$$

$$= P(X_S = 1|X_T = 1)P(X_T = 1) = 0,$$

and $E(X_S)E(X_T) = \frac{r}{m^2} \frac{r}{m^2} = \frac{r^2}{m^4}$, and hence

$$\sum_{|S \cap T| = 2} \text{Cov}(X_S, X_T) = -3m^3(m-1)\frac{r^2}{m^4}$$

$$= -\frac{3(m-1)r^2}{m}.$$

• Case 4: $|S \cap T| = 3$, i.e. $S = T$

We have $m^3$ choices for those kind of sets, so

$$E(X_S, X_T) = P(X_S = 1, X_T = 1)$$

$$= P(X_S = 1|X_T = 1)P(X_T = 1) = \frac{r}{m^2},$$

and $EX_SEX_T = \frac{r}{m^2} \frac{r}{m^2} = \frac{r^2}{m^4}$, and hence

$$\text{Cov}(X_S, X_T) = \frac{r}{m^2} - \frac{r^2}{m^4} = \frac{r}{m^2}(1 - \frac{r}{m^2})$$

$$\sum_{|S \cap T| = 3} \text{Cov}(X_S, X_T) = m^3 \frac{r(m^2 - r)}{m^4}$$

$$= \frac{r(m^2 - r)}{m}.$$

From the obtained information above, adding sums together, we obtain

$$\text{Var}(X) = \sum_{S, T} \text{Cov}(X_S, X_T) = mr(1 - r).$$

(Note that $\text{Var}(X) \geq 0$ if $m \geq 0$ and $r \leq 1$.)

When $r = 1$, $\text{Var}X = E(X - E(X))^2 = 0$, and since $X$ is a nonegative random
variable it follows that almost surely $X = E(X) = m$. That means almost surely only $K_3$ complexes are obtained in $P$.

When $r \to 1$, for every $\epsilon > 0$, $\lim_{r \to 1} P(|X - E(X)| \geq \epsilon) \leq \lim_{r \to 1} \frac{\text{Var}(X)}{\epsilon^2}$, i.e.,

$\lim_{r \to 1} P(|X - E(X)| \geq \epsilon) = 0$. Therefore, $\lim_{r \to 1} P(|X - E(X)| < \epsilon) = 1$.

The case when $p < \frac{1}{m}$ would result with incomplete complexes, and we do not consider this, but certainly we believe that such analysis may provide valuable information for understanding the self-assembly process.

To recapitulate, given $m$, depending on the amount of solution, and $r$, depending on the molecular dynamics, the expected number of tiles in $K_3$ cycles is $mr$, with standard deviation $\sqrt{mr(1-r)}$, the later being unobservable under contemporary laboratory conditions.

We can generalize the result (obtained for complete $K_3$) for circular complexes of any length. Consider a pot that contains $n$ 2-branched different tile types uniformly distributed, capable of forming a cycle of length $n$. 

90
In the previous chapter, we developed a theory for calculating the expected number of cyclic complexes. In order to check the validity of the theory, we did an experiment corresponding to the theory and in this chapter we present our findings.

The goal of the experiment is to calculate the probability of obtaining cyclic products when three 2-arm junction molecules are given. Each molecule is composed of three strands. One strand is 48 nucleotides long, another strand is 19 nucleotides long and the third one is 23 nucleotides long. (See the picture above.)

For flexibility of the molecules, bulges (6 T nucleotides) were added at the junc-
tions and also a nick (gap in the strand, i.e., a lack of a phosphodiester bond between two consecutive nucleotides) at the other side of the junction.

Each of the junction molecules is formed from three oligonucleotides, single stranded DNA molecules. Using the computer program SEQUIN we designed the strands for the molecules in such a way that there were no 5 bp mismatches and the number of 4 bp mismatches was minimal.

The designed sequences were the following:

**MOLECULE 1-M1**

STRAND # 1 (AS 11) CONSISTS OF:
C G T A G T C A C T G T G C G T C G C T G G T T T T T T T T
G T C G T T G A T G C T G A T A C A

STRAND # 2 (AS 12) CONSISTS OF:
A C C A G C G A C G C A C A G T G A C

STRAND # 3 (AS 13) CONSISTS OF:
C T G G T G T A T C A G C A T C A A C G A C A

**MOLECULE 2-M2**

STRAND # 1 (AS 21) CONSISTS OF:
C C A G C G A T G T C G T C A C T G T A G T A T T T T T T T T T
A T G G T A G C A C A C A C G C A T C A G

STRAND # 2 (AS 22) CONSISTS OF:
T A C T A C A G T G A C G A C A T C G

STRAND # 3 (AS 23) CONSISTS OF:
T C A A C T G A T G C G T G T G C T A C C A T

**MOLECULE 3-M3**

STRAND # 1 (AS 31) CONSISTS OF:
T T G A C T A C A A C A T C G C A G C A T C A T T T T T T T T T T T
G A C C A G C G T G T G C T A C T G T

STRAND # 2 (AS 32) CONSISTS OF:
T G A T G C T G C G A T G T T G T G T A G
STRAND # 3 (AS 33) CONSISTS OF:

T A C G A C A G T A G C A C A C G C T G G T C

Three oligonucleotides (strands # 3 from each molecule) were purchased from Integrated DNA Technology and then purified with the standard purification process in the laboratory. The remaining six oligonucleotides were designed to have an extra Phosphor attached to their 5' end, which is needed for the ligation process. Those strands were made in Dr. Seeman’s Laboratory by Rujie Sha.

The molecules were formed by annealing. Annealing is a process by which two complementary single stranded DNA molecules bond through the Watson-Crick complementarity. We formed the molecules by the fast annealing protocol (see Section 6.1). All three molecules formed well, as can be seen from the figures given below.

Since the project itself requires considerable precision, in order to calculate the percentage of molecules in a complex, one needs to read the results using a very sensitive method. One such method measures the percentages based on the radioactive counts (see Section 6.1). On a small portion of the strands #1 we attached a radioactive phosphate (P$^{32}$), in order to record (using Phosphor Imager) every complex formed. Hence, one strand of each junction molecule is radioactively labeled, and the percentage of molecules that form a complex can be easily determined for each complex. When we constructed each junction molecule we added 10% of the radioactively labeled strand #1 and 90% of the strand #1 that had regular Phosphate on its 5' end.

The complexes were formed by slow annealing of the junction molecules. We designed several test tubes with different concentrations of the molecules. The slow annealing process lasted for several hours and after that the complexes were ligated. (Ligation is a process by which a backbone of a DNA strand is recovered, i.e., if there is a nick in a DNA molecule such that there is a Phosphor on the 5’ end at the nick of the molecule, then a ligation enzyme seals the nick and recovers the backbone of the molecule.)
Figure 6.2: This is a 12\% native gel to check the formation of Molecule 1. A 10 nucleotide marker is in the first lane. In the second lane is Molecule 1, in the third lane is a complex consisting of the strands AS11 and AS12, in the fourth lane is a complex consisting of strands AS11 and AS13, in the fifth lane is a complex consisting of the strands AS12 and AS13, in the sixth lane is the strand AS11, in the seventh lane is the strand AS12, and in the eight lane is the strand AS13.
Figure 6.3: This is a 12% native gel to check the formation of Molecule 2. A 10 nucleotide marker is in the first lane. In the second lane is Molecule 2, in the third lane is a complex consisting of the strands AS21 and AS22, in the fourth lane is a complex consisting of strands AS21 and AS23, in the fifth lane is a complex consisting of the strands AS22 and AS23, in the sixth lane is the strand AS21, in the seventh lane is the strand AS22, and in the eight lane is the strand AS23.
Figure 6.4: This is a 12% native gel to check the formation of Molecule 3. A 10 nucleotide marker is in the first lane. In the second lane is Molecule 3, in the third lane is a complex consisting of the strands AS31 and AS32, in the fourth lane is a complex consisting of strands AS31 and AS33, in the fifth lane is a complex consisting of the strands AS32 and AS33, in the sixth lane is the strand AS31, in the seventh lane is the strand AS32, and in the eight lane is the strand AS33.
After the complexes were ligated, each test tube contained cyclic molecules as well as linear ones. In order to distinguish the linear ones from the cyclic, we split the solution of each test tube into two equal ones and added an Exo I and Exo III enzyme in one of the two halves. Exo enzymes work in the following way. If they recognize a nick in a molecule, they chop up the molecule, nucleotide by nucleotide, like the game packman. If an Exo enzyme is added into a test tube that contains both linear and circular molecules, the enzyme will destroy all linear molecules, leaving only the cyclic ones in the solution.

The results are obtained by measuring the sizes of the obtained complexes. A standard method for measuring the length of a DNA molecule is by gel electrophoresis. The electrophoresis technique is based on the fact that DNA molecules are negatively charged and if they are placed in an electric field, they will move (migrate) towards the positive electrode. The gel is prepared, and it is inserted between two 20 cm. long square glass plates. Before the gel thickens, a comb is inserted between the plates, to form wells (and later on lanes), and it is removed after the gel thickens. The DNA solution is put in the wells, and the gel is connected on an electric field.

In one well we add one half of a DNA solution that does not contain Exo enzymes, and in the lane next to it we add the other half of the solution, the one that contains Exo I and Exo III enzymes. This way it is easy to make the comparison which band of the first lane corresponds to a cyclic complex.

After the electric system is turned off, the results are obtained by exposing the gel to a PhosphorImager system. A PhosphorImager system is a quantitative imaging device that uses storage phosphor technology in life science imaging applications. It looks like a cassette with a white board in it, and the gel is exposed on the white board. The board counts the radioactivity emitted from the gel, and the results are obtained by scanning the board with a special scanner. Since the PhosphorImager system counts only radioactivity, the gel results will depend on the strand #1 (48 nucleotides long) from each molecule (since that strand was radioactively labeled). Hence, a linear monomer will be 48 nt., a linear dimer will have a length of 96 nt.,
a linear trimer and cyclic monomer will have a length of 144 nt., etc. Usually cyclic molecules travel slower than the linear ones so the small ones travel faster, etc., and we detected that cyclic 144 nt. runs similarly as linear 240 nt.

In the next section, a detailed description of the protocols used is given. In Section 6.2 we give the results.

6.1 The Protocol

Here is a more precise description of the protocols used in the experiment.

STEP 0: KINATION (Radioactive labeling) by the Phosphorylation

We mix together 1 µL DNA (1 pmole), 1 µL kination buffer, 1 µL labeled ATP, 6 µL dd H₂O, and 1 µL kinase. The reaction proceeds at 37 degree for about 1 hour. Then we inactivate the kinase, by leaving the solution on 90 degrees for 5 min. Afterwards, we filter the solution through the G-25 microspin column (Pharmacia) to remove unincorporated g-32P-ATP. Then we do phenol extraction and ethanol precipitation. At the end, the hot DNA is purified by a 10 to 15 % denaturing gel.

STEP 1: ANNEALING THE JUNCTION MOLECULES

For each molecule, we combine the three strands together that the molecule consists of (in 1 X TAE Mg ++ Buffer (12.5 µM, where µM stands for micro molar)) and then fast anneal them (5 min on 90°C, 15 min on 65°C, 20 min on 45°C, 20 min on 37°C and 20 min on room temperature).

STEP 2: ANNEALING THE COMPLEXES

After the junction molecules were formed, they were combined together to anneal in a solution containing 1X TAE Mg ++, 1X Ligase Buffer and dd H₂O depending on the desired concentration. They were annealing slowly in a period of more than 24 hours from 45°C to Room Temperature. (The temperature uniformly declined.)

In the first set of experiments, the volume was constant (30µL), therefore the counts in each concentration were different. In the second set of experiments, we used the same counts, therefore we used different volume for each concentration.

STEP 3: LIGATING THE COMPLEXES
After the complexes were annealed, we put the test tubes in the incubator (temperature 16°C) for 10 minutes, then we added 2µL of Ligase enzyme and left them in the incubator again for at least 16 hours.

STEP 4: EXO TREATMENT

After ligation, we divided the solution of each test tube into two equal parts. In one of them we added 1µL of Exo I and 1µL of Exo III and put that test tube for 1 hour on 37°C. The second one stayed the same.

STEP 5: GEL

We saw the results on an 8% denaturing gel, on 20 cm long square glass plates, which ran for at least an hour.

STEP 5: SCANNING THE RESULTS

We scanned the results using Phosphorimager and the accompanying software Molecular Dynamics.

6.2 Results

First set

First we did a set of 10 experiments with concentrations ranging from 10µM to 0.083µM. We set up the test tubes in the way given in Figure 6.2. For example, the junction molecule M1 is constructed by combining the strands AS11, AS12, AS13, and AS11*, which is the radioactively labeled strand, plus a buffer. In particular, in a test tube we combined 11 µL of 30 µM molecule AS11, 12 µL of 30 µM molecule AS12, 12 µL of 30 µM molecule AS13, and 1 µL of the molecule AS11* and 4 µL of TAE Mg++ (a buffer always constitutes 10 % of the total volume and it is necessary for the annealing process). In the test tube there is 12 × 30 = 360 pmols of each strand. Since, 1 pmol of AS11 + 1 pmol of AS12 + 1 pmol of AS13= 1 pmol M in the test tube there are 360 pmols of the molecule M1, in a 40 µL volume, i.e., the concentration of the molecule M1 in the test tube is \( \frac{360}{40} = 9 \mu M \).

To describe the way complexes were formed we use as an example the test tube
(T5*) whose concentration is 1 \( \mu M \). In the test tube T5*, 3\( \mu L \) of each of the three junction molecules was added, and also 1.8\( \mu L \) TAE Mg\(^{++}\) buffer was added (note that there was already 0.3\( \mu L \) of TAE Mg\(^{++}\) in the solution for each junction molecule), 3\( \mu L \) of Ligase buffer and 9 \( \mu L \) ddH\(_2\)O to dilute the solution for the desired concentration. The total volume in the test tube T5* is 31.8 \( \mu L \) and there are \( 3 \times 9 = 27 \) pmols of each complex, so the concentration on the test tube is \( \frac{27}{31.8} = 0.85 \mu M \).

In this set, we kept the volume of the solution in the test tubes the same (30\( \mu L \)), and therefore the number of radioactive counts was different.
Figure 6.5: (a) 10 nucleotide marker, (b) ligated molecules from the test tube T5 of 1 μM concentration, (c) exo treatment of the ligated product from test tube T5 of 1 μM concentration, (d) ligated molecules from test tube T4 of 1.5μM, (e) exo treatment of the ligated product from test tube T4 of 1.5μM, (f) ligated molecules from test tube T3 of 2μM, (g)exo treatment of the ligated product from test tube T3 of 2μM, (h) ligated molecules from test tube T2 of 2.5μM. (i) exo treatment of the ligated product from test tube T2 of 2.5μM.
Figure 6.6: (a) 10 nucleotide marker (b) ligated molecules from the test tube T10 of 0.1 µM concentration, (c) exo treatment of the ligated product from test tube T10 of 0.1 µM concentration, (d) ligated molecules from test tube T9 of 0.2 µM, (e) exo treatment of the ligated product from test tube T9 of 0.2 µM, (f) ligated molecules from test tube T8 of 0.3 µM, (g) exo treatment of the ligated product from test tube T8 of 0.3 µM, (h) ligated molecules from test tube T7 of 0.4 µM, (i) exo treatment of the ligated product from test tube T7 of 0.4 µM, (j) molecules from test tube T6 of 0.5 µM, (k) exo treatment of the ligated product from the test tube T6 of 0.5 µM.
Calculation on the percentages was done using the Software “Molecular Dynamics”. This software gives an estimate on the percentages based on the intensity of the bands. Here are the results.

<table>
<thead>
<tr>
<th></th>
<th>1 uM</th>
<th></th>
<th></th>
<th>1.5 uM</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>length</td>
<td>%</td>
<td>length</td>
<td>%</td>
<td></td>
</tr>
<tr>
<td>LINEAR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48 nt</td>
<td>11.63</td>
<td>9.88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>96 nt</td>
<td>12.01</td>
<td>9.76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>144 nt</td>
<td>14.23</td>
<td>10.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>192 nt</td>
<td>7.38</td>
<td>12.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>288 nt</td>
<td>4.93</td>
<td>4.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIRCULAR</td>
<td></td>
<td></td>
<td>CIRCULAR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>and LINEAR</td>
<td>144 nt</td>
<td>43.19</td>
<td>and LINEAR</td>
<td>144 nt</td>
<td>43.59</td>
</tr>
<tr>
<td>240 nt</td>
<td>40.81</td>
<td>40.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 uM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LINEAR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48 nt</td>
<td>18.07</td>
<td>13.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>96 nt</td>
<td>10.95</td>
<td>9.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>144 nt</td>
<td>13.94</td>
<td>11.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>192 nt</td>
<td>5.06</td>
<td>11.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>288 nt</td>
<td></td>
<td>4.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIRCULAR</td>
<td></td>
<td></td>
<td>CIRCULAR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>and LINEAR</td>
<td>144 nt</td>
<td>32.18</td>
<td>and LINEAR</td>
<td>144 nt</td>
<td>33.29</td>
</tr>
<tr>
<td>240 nt</td>
<td>34.82</td>
<td>34.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5 uM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LINEAR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48 nt</td>
<td>20.92</td>
<td>15.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>96 nt</td>
<td>12.08</td>
<td>11.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>144 nt</td>
<td>8.96</td>
<td>9.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>192 nt</td>
<td>10.77</td>
<td>8.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIRCULAR</td>
<td></td>
<td></td>
<td>CIRCULAR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>and LINEAR</td>
<td>144 nt</td>
<td>37.05</td>
<td>and LINEAR</td>
<td>144 nt</td>
<td>40.51</td>
</tr>
<tr>
<td>240 nt</td>
<td>36.95</td>
<td>36.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3 uM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LINEAR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>48 nt</td>
<td>5.51</td>
<td>9.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>96 nt</td>
<td>11.11</td>
<td>11.33</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>144 nt</td>
<td>10.86</td>
<td>8.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>192 nt</td>
<td></td>
<td>8.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIRCULAR</td>
<td></td>
<td></td>
<td>CIRCULAR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>and LINEAR</td>
<td>144 nt</td>
<td>48.93</td>
<td>and LINEAR</td>
<td>144 nt</td>
<td>43.22</td>
</tr>
<tr>
<td>240 nt</td>
<td>48.83</td>
<td>43.21</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ligation is a process with low yield. Unfortunately, that is our only way to check whether a complex is complete or not. If a given complex forms a cycle then it is complete, but sometimes a complex can form a cycle, but not all of the nicks at the sticky ends are ligated. In that case the complex appears as a linear one, although it is actually complete. When three junction molecules connect, they connect by their sticky ends. Each sticky end connection leaves one nick at the edge of the triangle. In order for those three molecules to form a cycle with no nicks, three nicks should be ligated and the probability of three ligations is not high. Hence in analyzing the results we assume that the circular molecules of length 144 nt. and linear molecules of length 144 nt. both formed a complete complex, just the first ones ligated three nicks, while in the second only two nicks were ligated. Hence if we interpretate the results
in the following way. For the concentration $2.5\mu M$ 44.32 % of cyclic monomers, for $2\mu M$ 46.12 % cyclic monomers, for $1.5\mu M$ 53.58 % cyclic monomers and for $1\mu M$ 57.42 %. From this it is clear that by decreasing the concentration, the number of cyclic monomers increases.

For $0.5\mu M$, 46.01 %, for $0.4\mu M$, 50.43 %, for $0.2\mu M$ 51.72 %. It is important to mention that the best way to compare the percentage of the cyclic monomers is by considering the solutions that run on a same gel. Since a software is used to obtain the percentages, the accuracy of the results depends on the accuracy of the software. In the test tube of $0.3\mu M$ for calculating the percentages of the cyclic molecules, the molecules of length 48 nt. were not taken into consideration, hence we are not considering the results for that test tube.

**Second set**

In the first set of experiments, we kept the volume of each solution the same, thus the amount of radioactive counts in each test tube was different. The results showed no correlation between the amount of cyclic triangles and concentration of the solution, so we changed the set-up of the test tubes for the next set of experiments. For the second set of experiments we decided to keep the amount of radioactive counts the same, and change the volume accordingly. This way from one look of the gel, one can see whether the number of cycles of length 3 increases as we decreased the concentration, which is what we assumed.

The lower concentration test tube had higher volume, consequently the higher concentration test tube had lower volume. Each buffer constitutes 10 % of the total volume, thus in a low concentration solution the amount of ions (from the buffers) was high and that prevents the DNA molecules from free movement in a gel. Although most of the gels were destroyed we were able to draw a conclusion from the results.

According to the theoretical analysis, we assumed that in a very low concentration, almost surely all of the junction molecules should become a part of a triangle. However, that was not the case in the previous set of experiments, and neither is it
We set up the test tubes in the following way:

**HOT EXPERIMENT 6**

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>M2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AS11 (SuM)</td>
<td>16 µL</td>
</tr>
<tr>
<td></td>
<td>AS12 (SuM)</td>
<td>20 µL</td>
</tr>
<tr>
<td></td>
<td>AS13 (SuM)</td>
<td>20 µL</td>
</tr>
<tr>
<td></td>
<td>AS11*(10 µM)</td>
<td>4 µL</td>
</tr>
<tr>
<td></td>
<td>TAE mg ++ Buff</td>
<td>6.67 µL</td>
</tr>
<tr>
<td><strong>Total Volume</strong></td>
<td>66.67 µL</td>
<td>Total Volume</td>
</tr>
<tr>
<td></td>
<td>Concentration</td>
<td>1.5 µM</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>M3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AS31 (SuM)</td>
</tr>
<tr>
<td></td>
<td>AS32 (SuM)</td>
</tr>
<tr>
<td></td>
<td>AS33 (SuM)</td>
</tr>
<tr>
<td></td>
<td>AS31*(10µM)</td>
</tr>
<tr>
<td></td>
<td>TAE mg ++ Buff</td>
</tr>
<tr>
<td><strong>Total Volume</strong></td>
<td>66.67 µL</td>
</tr>
<tr>
<td></td>
<td>Concentration</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>T1*-0.1µM</th>
<th>T2* -0.075 µM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M1</td>
<td>4 µL</td>
</tr>
<tr>
<td></td>
<td>M2</td>
<td>4 µL</td>
</tr>
<tr>
<td></td>
<td>M3</td>
<td>4 µL</td>
</tr>
<tr>
<td></td>
<td>Buffer (TAE Mg ++ 6 µL)</td>
<td>6 µL</td>
</tr>
<tr>
<td></td>
<td>Ligase buffer</td>
<td>6 µL</td>
</tr>
<tr>
<td></td>
<td>dd H2O</td>
<td>36 µL</td>
</tr>
<tr>
<td><strong>Total Volume</strong></td>
<td>60 µL</td>
<td><strong>Total Volume</strong></td>
</tr>
<tr>
<td></td>
<td>Concentration</td>
<td>0.10 µM</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>T3*-0.05µM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M1</td>
</tr>
<tr>
<td></td>
<td>M2</td>
</tr>
<tr>
<td></td>
<td>M3</td>
</tr>
<tr>
<td></td>
<td>Buffer (TAE Mg ++ 12 µL)</td>
</tr>
<tr>
<td></td>
<td>Ligase buffer</td>
</tr>
<tr>
<td></td>
<td>dd H2O</td>
</tr>
<tr>
<td><strong>Total Volume</strong></td>
<td>120 µL</td>
</tr>
<tr>
<td></td>
<td>Concentration</td>
</tr>
</tbody>
</table>

The results are the following.
Figure 6.7: (a) 10 nucleotide marker, (b) ligated molecules from the test tube T1 of $1\mu M$ concentration, (c) exo treatment of the ligated product from test tube T1 of $1\mu M$ concentration.
Figure 6.8: (a) 10 nucleotide marker, (b) ligated molecules from the test tube T2 of 0.75µM concentration, (c) exo treatment of the ligated product of 0.75µM concentration, (d) ligated molecules from test tube T3 of 0.5µM concentration, (e) exo treatment of the ligated product from test tube T3 of 0.5µM concentration.
Like we mentioned before, the best way to compare the percentages is by considering one gel at a time. In that case it is clear from this set of experiments that as the concentration decreases, the proportion of molecules that are in triangles increases. It is also obvious that besides the monomers (triangles), dimers also appear regardless of the concentration, i.e., we cannot conclude that in a very dilute solution, the probability of the last sticky end to close, \( r \), approaches 1.

Third set

We realized that the length on the gel plates was short, i.e. for better differentiation of the length of the DNA molecules, we repeated the process on longer gel plates (40 cm.).
Figure 6.9: (a) 10 nucleotide marker, (b) ligated molecules from the test tube of 3μM concentration, (c) exo treatment of the ligated product of 3μM concentration, (d) ligated molecules from the test tube of 1.5μM concentration, (e) exo treatment of the ligated product of 1.5μM concentration, (f) ligated molecules from the test tube of 1μM concentration, (g) exo treatment of the ligated product of 1μM concentration.
Figure 6.10: (a) 10 nucleotide marker, (b) ligated molecules from the test tube of 0.5μM concentration, (c) exo treatment of the ligated product of 0.5μM concentration, (d) ligated molecules from the test tube of 0.1μM concentration.
Figure 6.11: (a) 10 nucleotide marker, (b) ligated molecules from the test tube of 1µM concentration, (c) exo treatment of the ligated product of 1µM concentration, (d) ligated molecules from the test tube of 0.5µM concentration, (e) exo treatment of the ligated product of 0.5µM concentration, (f) ligated molecules from the test tube of 0.1µM concentration, (g) exo treatment of the ligated product of 0.1µM concentration.
The results were the following:

<table>
<thead>
<tr>
<th>Time (min)</th>
<th>Linear</th>
<th>Circular</th>
</tr>
</thead>
<tbody>
<tr>
<td>48</td>
<td>16.96%</td>
<td>31.24%</td>
</tr>
<tr>
<td>96</td>
<td>9.20%</td>
<td>1.69%</td>
</tr>
<tr>
<td>144</td>
<td>13.72%</td>
<td></td>
</tr>
<tr>
<td>192</td>
<td>11.29%</td>
<td></td>
</tr>
<tr>
<td>240</td>
<td>7.96%</td>
<td></td>
</tr>
<tr>
<td>288</td>
<td>4.71%</td>
<td></td>
</tr>
<tr>
<td>336</td>
<td>3.14%</td>
<td></td>
</tr>
<tr>
<td>384</td>
<td>2.43%</td>
<td></td>
</tr>
<tr>
<td>432</td>
<td>1.07%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time (min)</th>
<th>Linear</th>
<th>Circular</th>
</tr>
</thead>
<tbody>
<tr>
<td>48</td>
<td>8.22%</td>
<td>45.88%</td>
</tr>
<tr>
<td>96</td>
<td>5.86%</td>
<td>1.41%</td>
</tr>
<tr>
<td>144</td>
<td>15.78%</td>
<td></td>
</tr>
<tr>
<td>192</td>
<td>7.89%</td>
<td></td>
</tr>
<tr>
<td>240</td>
<td>3.44%</td>
<td></td>
</tr>
<tr>
<td>288</td>
<td>5.99%</td>
<td></td>
</tr>
<tr>
<td>336</td>
<td>2.91%</td>
<td></td>
</tr>
<tr>
<td>384</td>
<td>1.75%</td>
<td></td>
</tr>
<tr>
<td>432</td>
<td>1.42%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time (min)</th>
<th>Linear</th>
<th>Circular</th>
</tr>
</thead>
<tbody>
<tr>
<td>48</td>
<td>1.52%</td>
<td>53.51%</td>
</tr>
<tr>
<td>96</td>
<td>7.39%</td>
<td>0.74%</td>
</tr>
<tr>
<td>144</td>
<td>17.51%</td>
<td></td>
</tr>
<tr>
<td>192</td>
<td>7.64%</td>
<td></td>
</tr>
<tr>
<td>240</td>
<td>4.71%</td>
<td></td>
</tr>
<tr>
<td>288</td>
<td>4.31%</td>
<td></td>
</tr>
<tr>
<td>336</td>
<td>1.30%</td>
<td></td>
</tr>
<tr>
<td>384</td>
<td>0.78%</td>
<td></td>
</tr>
<tr>
<td>432</td>
<td>0.24%</td>
<td></td>
</tr>
</tbody>
</table>

6.3 Conclusion

From the several sets of experiments we did, we can conclude that with the decrease in the concentration the amount of cyclic monomers increases. Also, we can conclude that the number of cyclic dimers decreases with the increase in the concentration, but they do not disappear, as we assumed in the theoretical settings. Appearance of cyclic trimers was noticed in the last set of experiments (since the length of the gel plates was doubled), but it was not confirmed. We have to confirm the appearance of cyclic trimers as well as to check if there are cyclic molecules of larger size. Besides, examining the size of all the cyclic molecules, we should also examine the structure. In particular, whether they represent one big cyclic molecule, or they are two smaller cyclic molecules linked together, or one twisted molecule, etc.?

The results that we obtained from the experiments we plan to use them for studying the thermodynamics and kinetic properties of the molecule. In the next section, we represent some of the ideas that we want to pursue for future research. One idea is to use rate equations, and for that we need to know all the possible complexes, both linear and cyclic. We hope to extend this study from cyclic molecules, to any type of complexes, regardless of their structure.
7 Conclusion

Given the possibilities of establishing different connections, DNA molecules can connect in many different ways building different new structures. Although most of the research methods in this dissertation target DNA self-assembly, they are very general in a sense that they can be extended and applied to any other type of self-assembly. This could be done so that the elements that self-assemble will correspond to tiles, and the parts of them (or connection rules) by which the elements connect will correspond to sticky ends. Then the process of gluing and forming complexes should be defined similarly as we did in Chapter 2, and the other definitions and propositions should be changed accordingly.

In this dissertation we investigate fundamental questions related to DNA self-assembly and have developed the appropriate tools accordingly. The questions we have addressed are: What are the possible complexes obtained at the end of the self-assembly process? Which of those can be actually expected and with what probability? What are the necessary conditions to eliminate the byproducts? How to pre-design a test tube to minimize those by-products? What relations can be defined on test tubes?

Knowing that many times in scientific laboratories, besides a solution to a given problem, a lot of extraneous material (non-complete complexes) also appears, we addressed the problem of eliminating (or at least minimizing) the sludge in Chapter 3. We proved that a necessary condition for obtaining only complete complexes at the end of an experiment is to use the proper proportion of each type of molecules present. The set of vectors for the proper proportions is called “spectrum”, and
with the MAPLE program given in Appendix A it can be very quickly and easily
determined.

The pot types are classified in three classes according to possible components
that assemble in complete complexes, and based on the spectrum the classification is
PTIME computable. The question that arose recently is whether through the spec-
trum we can classify the possible complete complexes. Using computational geometry,
we showed that based on the extremal points of the spectrum, we can describe a por-
tion of the complete complexes, but not all of them. Next step to pursue is to find
ways to describe those random objects through algebraic representation, or through
algebraic and graphical representation combined. Another question that we should
address is if it is decidable whether or not a pot type contains a compete complex
that has tiles of each type. If we are able to classify all the complete complexes with
algebraic representation, then this question should be answered.

The graphical representation, given in Chapter 4, was developed to study the
products of self-assembly with familiar tools that help in understanding the process.
The graph model is used to determine what complete complexes can be assembled
from a given pot type as well as compare and classify the pot types themselves. It
represents an application of graph homomorphism theory. We only scratched the
surface of the topic, there is much more to be done in that direction. Immediate
questions that arise from this section are: Is there a connection between the pot
classification and graph representation, i.e., is there a way to classify the pot types
using their pot graphs? (We know, that we should not look at graph connectivity,
because it is possible for a strongly satisfiable pot type to have disconnected graph,
while an unsatisfiable pot type to have a connected graph.) Are minimal complex
graphs, prime graphs? Can we deduce some kind of conclusion for the maximal
complete complex graphs, based on the theory of prime graphs?

So as to determine the possible complexes at the end of the self-assembly process
and with what probability they occur, in Chapters 5 and 6 we concentrated on the
formation of cyclic complexes. In a quest to calculate the expected number of cyclic
molecules that show at the end of an experiment and the probability of their formation, we built a new random graph model that considers self-assembly. Naturally, the next step to pursue is to adjust the random graph model so as to include other molecules besides cyclic ones. In order to do that, first one should understand the number of different probabilities of the system and build the probability space accordingly. Although, the intermediate steps in the study of the cyclic molecules were known, we should perform a study for the model with arbitrary (not only cyclic) complexes. That will give better insight into the possible probabilities and the final distribution of the random structures. The study of the intermediate steps should be done with branching processes, to understand how one complex may evolve over time, and give a complete picture of the process and the final outcomes.

Thermodynamics hasn’t been incorporated in the random models. We are currently considering ways how to include it. The first approach that we took is with a system of rate equations.

Here is an idea how to include the system of rate equations. Assume we have a pot type with only one tile type, such that tiles can close in a complete complex, and there are exactly \( m \) tiles in the pot. Denote with \( C_n \) the number of cyclic complexes consisting of \( n \) tiles in step \( t \), \( L_n(t) \) the number of linear complexes consisting of \( n \) tiles in step \( t \), \( k_{i,j} \) the rate at which a linear complex of length \( i \) connects to a linear complex of length \( j \), and \( r_i \) the rate at which a cycle of length \( i \) closes. The following system of equations model the evolution of the self assembly process.

\[
\begin{align*}
\frac{dL_n(t)}{dt} &= \frac{1}{2} \sum_{i=1}^{n-1} k_{(n-i,i)} L_{n-i} I_i - r_n L_n \\
\frac{dC_n(t)}{dt} &= r_n L_n \\
\sum_{i=1}^{m} i(L_i + C_i) &= m, \text{ and} \\
L_1(0) &= m, C_1(0) = 0, L(\text{last step}) = 0, \sum_{i=1}^{m} iC_i(\text{last step}) = m
\end{align*}
\]

The system itself is difficult to solve, and we plan to study this system further.
In this dissertation we approached the self-assembly process through discrete mathematical theory. We addressed some of the fundamental questions related to the process, but there are more to be considered. We hope you enjoyed the journey, as we did, through this dissertation.
REFERENCES


Appendices

Appendix A - Maple Program

In this section we give the maple program for calculating the spectrum and a pot class of a given pot. What follows is a solution for the spectrum of the pot type explained in Fig 3.2.

This is a program that can calculate the spectrum of a given pot. Assume that a pot \( \mathbf{P} \) with \( m \) junction types and \( n \) sticky end types is given.

\[
\text{restart:}
\]

\[
\text{with(LinearAlgebra):}
\]

Input the number of junction types

\[
\text{m:=3:}
\]

Input the number of sticky end types

\[
\text{n:=2:}
\]

\[
\text{c:=0:}
\]

Enter the corresponding \( z(h) \) vectors. The coordinate for \( a[s, t]=z_{\{j, \{s\}\}}(h_{\{t-1\}}) \)

\[
\text{a:=Matrix([[1, 1, 1, 1], [2, -1, 0, 0], [1, 1, -3, 0]]);} \\
\text{L:=Matrix([[1, 1, 0], [1, 1, 1]]);} \\
\text{a :=} \\
\begin{bmatrix}
1 & 1 & 1 & 1 \\
2 & -1 & 0 & 0 \\
1 & 1 & -3 & 0
\end{bmatrix} \\
\text{L :=} \\
\begin{bmatrix}
1 & 1 & 0 \\
1 & 1 & 1
\end{bmatrix}
\]

122
A:=ReducedRowEchelonForm(a): evalm(A):
B:=Matrix(n+1,m):
for i from 1 to n+1 do
  for j from 1 to m do
    B[i,j]:=A[i,j]:end do: end do;B:
b:=Vector(n+1):
for i from 1 to n+1 do
  b[i]:=A[i,m+1]: end do: b;
if(evalb(Rank(A)=Rank(B)))=false
then c:=-1: end if:
r:=LinearSolve(B,b,free='t');
for i from 1 to m do
  eq[i]:=r[i]>=0;
end do:
Sys:=seq(eq[i], i=1..m);
d:=seq(t[i], i=1..m);
s:=solve({Sys},{d});
Sys := 0 ≤ \frac{1}{4}, 0 ≤ \frac{1}{2}, 0 ≤ \frac{1}{4}
d := t_1, t_2, t_3
s := \{t_1 = t_1, t_2 = t_2, t_3 = t_3\}
for i from 1 to m do
  for j from 1 to m do
    if evalb(s[i]=(t[j]=0))=true then r:=subs({t[j]=0}, eval(r))
    end if:
  end do: end do:

Supp:=[[]:for i from 1 to m do
  if evalb(r[i]>0)=false then Supp:=[op(Supp), 0];
  else Supp:=[op(Supp),1]; end if: end do: 'Supp'=Supp;

Supp = [1, 1, 1]

if evalb(c=0)=true then
  for i from 1 to n do
    for j from 1 to m do
      if evalb(Row(L, i)[j]=1 and Supp[j]=1)=true then c:=0:
        break;
      else c:=1: end if; end do;
    end do; end if; c;

  if evalb(c=0)=true then
    for i from 1 to m do
      if evalb(r[i]>0)=false then c:=2: break;
    end if; end do; end if;

  if evalb(c=0)=true then print("The given pot is strongly satisfiable and its spectrum is "); 'r'=r; s; end if;
  if evalb(c=1)=true then print("The given pot is weakly satisfiable and its spectrum is"); 'r'=r; s; end if;
  if evalb(c=2)=true then print("The given pot is satisfiable and its spectrum is"); 'r'=r; s; end if;
  if evalb(c=-1)=true then print("The given pot is not weakly satisfiable"); end if;

"The given pot is strongly satisfiable and its spectrum is 
\[ r = \begin{bmatrix} 1 \\ 4 \\ 1 \\ 2 \\ 1 \\ 4 \end{bmatrix} \]
\{t_1 = t_1, t_2 = t_2, t_3 = t_3\}

for i from 1 to m do
> e[i] := solve(r[i] = 0, \{d\});
> subs(e[i], eval(r));
> end do:
> sys1 := seq(e[i], i = 1 .. m):
Appendix B - Existence of the Probability Space

In this section of the dissertation we present a mathematically support for Chapter 5, i.e., a formal proof for the existence of the probability measure and the sample space, and to give an idea how the sample space should look like. As above we start with observation of the sample space ($\Omega_m$) consisting of the outcomes in the self-assembly process for a one type pot with $m$ tiles, $\mathbf{P}_m = \{t\}$. We assume that during the process all Watson-Crick connections are established and that only complete complexes are obtained. That means after the annealing process every tile is on a complete cycle. Also we assume that all Watson-Crick connections are equally likely. Specifically, an outcome is a complex consisting of a complex graph of $m$ vertices, as a 2-regular graph. Our assumptions assure that two isomorphic outcomes are equally likely.

Denote with $r_{m,k}$ the probability in $\Omega_m$, that a given tile is on a cycle of length $k$. The probability measure for this sample space we will define it recursively.

Consider two different tiles, say $t_1$ and $t_2$, from type $t \in \mathbf{P}_m$. Assuming all W-C connections are equally likely, for a fixed number $k \leq m$,

$$P(t_1 \text{ is on a cycle of length } k) = P(t_2 \text{ is on a cycle of length } k) = r_{m,k}.$$ 

Then we can define $r_{m,k}$ recursively. Fix $t_1$, $t_2$.

• For $1 \leq k < m$
\[ P(t_1 \text{ is on a cycle of length } k) \]

\[ = P(t_1 \text{ is on a cycle of length } k \mid t_1 \text{ and } t_2 \text{ are not on the same cycle}) \cdot P(t_1 \text{ and } t_2 \text{ are not on the same cycle}) \]

\[ + P(t_1 \text{ is on a cycle of length } k \mid t_1 \text{ and } t_2 \text{ are on the same cycle}) \cdot P(t_1 \text{ and } t_2 \text{ are on the same cycle}) \]

When \( t_1 \) is on a cycle of length \( k \), and \( t_1 \) and \( t_2 \) are not on the same cycle, then \( t_2 \) will be on any other cycle of length \( n \), for some \( 1 \leq n \leq m - k \). In that case

\[ P(t_1 \text{ is on a cycle of length } k \mid t_1 \text{ and } t_2 \text{ are not on the same cycle}) \cdot P(t_1 \text{ and } t_2 \text{ are not on the same cycle}) \]

\[ = \sum_{n=1}^{m-k} P(t_1 \text{ is on a cycle of length } k \mid t_1 \text{ and } t_2 \text{ are not on the same cycle, } t_2 \text{ is on a cycle of length } n) \]

\[ P(t_1 \text{ and } t_2 \text{ are not on the same cycle, } t_2 \text{ is on a cycle of length } n) \]

\[ = \sum_{n=1}^{m-k} P(t_1 \text{ is on a cycle of length } k \mid t_1 \text{ and } t_2 \text{ are not on the same cycle, } t_2 \text{ is on a cycle of length } n) \]

\[ P(t_1 \text{ and } t_2 \text{ are not on the same cycle, } t_2 \text{ is on a cycle of length } n) \]

\[ = \sum_{n=1}^{m-k} r_{m-n,k} \frac{m-n}{m-1} r_{m,n}. \]

When \( t_1 \) and \( t_2 \) are on the same cycle and is given that \( t_2 \) is on a cycle of length \( n = k \), then \( t_1 \) needs to be one of the remaining \( k - 1 \) tiles on that cycle. Therefore the probability \( t_1 \) is on a cycle of length \( k \) under the given conditions will be \( \frac{k-1}{m-1} \), i.e.,

127
\( P(t_1 \text{ is on a cycle of length } k | t_1 \text{ and } t_2 \text{ are on the same cycle}) \)

\[ P(t_1 \text{ and } t_2 \text{ are on the same cycle of length } k) = \frac{k - 1}{m - 1} r_{m,k}. \]

Merging both equations together we obtain

\[ P(t_1 \text{ is on a cycle of length } k) = r_{m,k} = \sum_{n=1}^{m-k} r_{m-n,k} \frac{m-n}{m-1} r_{m,n} + \frac{k - 1}{m - 1} r_{m,k} \]

Hence, for \( 1 \leq k < m \),

\[ r_{m,k} = \frac{m - 1}{m - k} \sum_{n=1}^{m-k} r_{m-n,k} \frac{m-n}{m-1} r_{m,n} \]

\[ = \frac{1}{m - k} \sum_{n=1}^{m-k} (m - n) r_{m-n,k} r_{m,n}. \]

• For \( k = m \),

\[ P(t_1 \text{ is on a cycle of length } m) = P(t_1 \text{ is on a cycle of length } m | t_1 \text{ and } t_2 \text{ are on the same cycle of length } m). \]

\[ P(t_1 \text{ and } t_2 \text{ are on the same cycle}) = P(t_2 \text{ is on a cycle of length } m). \]

Hence for \( k = m \), we do not obtain any new equation, instead we get that

\[ r_{m,m} = r_{m,m}. \]

As we said previously, we consider sample spaces that only have complete complexes, therefore every tile from the pot \( P_m \) must be on some cycle, i.e.,

\[ \sum_{k=1}^{m} r_{m,k} = 1. \]
The probabilities for the sample space $\Omega_m$ can be calculated from the following system of equations:

$$\begin{cases}
\sum_{k=1}^{m} r_{m,k} = 1 \\
r_{m,k} = \frac{1}{m-k} \sum_{n=1}^{m-k} (m-n) r_{m-n,k} r_{m,n}, \quad \text{for } 1 \leq k < m.
\end{cases}$$

Since we are taking the sample space to consider only of complete complexes, it is true that

$$\sum_{i=1}^{k} r_{k,i} = 1, \quad \text{for any } k \in \mathbb{N}.$$ 

Also, since in the space $\Omega_k$ it is not possible to have a cycle of size greater then $k$, it is true that $r_{k,n} = 0$ for all $k > n$. 

129
Let us expand the above equations.

\[ r_{m,1} = \frac{1}{m-1} \left[ (m-1)r_{(m-1),1}r_{m,1} + (m-2)r_{(m-2),1}r_{m,2} + (m-3)r_{(m-3),1}r_{m,3} + \ldots + \frac{3}{m-1}r_{3,1}r_{m,(m-3)} + \frac{2}{m-1}r_{2,1}r_{m,(m-2)} + \frac{1}{m-1}r_{1,1}r_{m,(m-1)} \right] \]

\[ r_{m,2} = \frac{1}{m-2} \left[ (m-1)r_{(m-1),2}r_{m,1} + (m-2)r_{(m-2),2}r_{m,2} + (m-3)r_{(m-3),2}r_{m,3} + \ldots + \frac{3}{m-2}r_{3,2}r_{m,(m-3)} + \frac{2}{m-2}r_{2,2}r_{m,(m-2)} \right] \]

\[ r_{m,3} = \frac{1}{m-3} \left[ (m-1)r_{(m-1),3}r_{m,1} + (m-2)r_{(m-2),3}r_{m,2} + (m-3)r_{(m-3),3}r_{m,3} + \ldots + \frac{3}{m-3}r_{3,3}r_{m,(m-3)} \right] \]

\[ \vdots \]

\[ r_{m,(m-3)} = \frac{1}{3} \left[ (m-1)r_{(m-1),(m-3)}r_{m,1} + (m-2)r_{(m-2),(m-3)}r_{m,2} + (m-3)r_{(m-3),(m-3)}r_{m,(m-3)} \right] \]

\[ r_{m,(m-2)} = \frac{1}{2} \left[ (m-1)r_{(m-1),(m-2)}r_{m,1} + (m-2)r_{(m-2),(m-2)}r_{m,2} \right] \]

\[ r_{m,(m-1)} = (m-1)r_{(m-1),(m-1)}r_{m,1} \]

(7.0.1)

We will transform this system into a homogeneous one.

\[
(m-1)r_{m,1} - r_{m,m-1} + \frac{m-2}{m-1}r_{m,1}r_{m,2} + \frac{m-3}{m-1}r_{m,1}r_{m,3} + \ldots + \frac{3}{m-1}r_{3,1}r_{m,(m-3)} + \frac{2}{m-1}r_{2,1}r_{m,(m-2)} + \frac{1}{m-1}r_{1,1}r_{m,m-1} = 0
\]

\[
\frac{m-2}{m-1}r_{m,2} + \frac{m-3}{m-2}r_{m,2}r_{m,3} + \ldots + \frac{3}{m-2}r_{3,2}r_{m,(m-3)} + \frac{2}{m-2}r_{2,2}r_{m,(m-2)} = 0
\]

\[
\frac{m-3}{m-3}r_{m,3} + \frac{m-2}{m-3}r_{m,3}r_{m,2} + \frac{m-3}{m-3}r_{m,3}r_{m,3} + \ldots + \frac{3}{m-3}r_{3,3}r_{m,(m-3)} = 0
\]

\[ \vdots \]

\[
\frac{m-1}{m-4}r_{m,(m-1),1} + \frac{m-2}{m-4}r_{m,(m-1),2}r_{m,2} + \frac{m-3}{m-4}r_{m,(m-1),3}r_{m,3} + \ldots + \frac{3}{m-4}r_{3,3}r_{m,(m-3)} = 0
\]

\[
\frac{m-2}{m-3}r_{m,(m-2),1} + \frac{m-3}{m-3}r_{m,(m-2),2}r_{m,2} + \frac{m-3}{m-3}r_{m,(m-2),3}r_{m,3} + \ldots + \frac{3}{m-3}r_{3,3}r_{m,(m-3)} = 0
\]

\[
(m-1)r_{m,(m-1),1} + 0 + 0 + \ldots + 0 + r_{m,(m-2)} + 0 = 0
\]

(7.0.2)

Multiplying by the reciprocal of the coefficient in front of the first term, we obtain the following
\begin{align*}
(r_{m-1}, 1 - 1)r_{m, 1} + \frac{m - 2}{m - 1}r_{m-1, 1}r_{m, 2} + \frac{m - 3}{m - 1}r_{m-1, 1}r_{m, 3} + \ldots + \frac{3}{m - 1}r_{m-1, 1}r_{m, m-3} + \frac{2}{m - 1}r_{m-1, 1}r_{m, m-2} + \frac{1}{m - 1}r_{m-1, 1}r_{m, m-1} &= 0 \\
r_{m-1, 2}r_{m, 1} + \frac{m - 2}{m - 1}r_{m-1, 2}r_{m, 2} + \frac{m - 3}{m - 1}r_{m-1, 2}r_{m, 3} + \ldots + \frac{3}{m - 1}r_{m-1, 2}r_{m, m-3} + \frac{2}{m - 1}r_{m-1, 2}r_{m, m-2} &= 0 \\
r_{m-1, 3}r_{m, 1} + \frac{m - 2}{m - 1}r_{m-1, 3}r_{m, 2} + \frac{m - 3}{m - 1}r_{m-1, 3}r_{m, 3} + \ldots + \frac{3}{m - 1}r_{m-1, 3}r_{m, m-3} &= 0 \\
\vdots \\
r_{m-1, (m-3)}r_{m, 1} + \frac{m - 2}{m - 1}r_{m-1, (m-3)}r_{m, 2} + \frac{m - 3}{m - 1}r_{m-1, (m-3)}r_{m, (m-3)} &= 0 \\
r_{m-1, (m-2)}r_{m, 1} + \frac{m - 2}{m - 1}r_{m-1, (m-2)}r_{m, 2} &= 0 \\
r_{m-1, (m-1)}r_{m, 1} &= 0
\end{align*}

\begin{align}
(r_{m-1}, 1) + r_{(m-1), 2} + \ldots + r_{(m-1), (m-2)} + r_{(m-1), (m-1)} - 1)r_{m, 1} + \\
\frac{m - 2}{m - 1}(r_{m-2}, 1) + r_{(m-2), 2} + r_{(m-2), 3} + \ldots + r_{(m-2), (m-3)} + r_{(m-2), (m-2)} - 1)r_{m, 2} + \\
\frac{m - 2}{m - 1}(r_{m-3}, 1) + r_{(m-3), 2} + r_{(m-3), 3} + \ldots + r_{(m-3), (m-4)} + r_{(m-3), (m-3)} - 1)r_{m, 3} + \\
\ldots + \\
\frac{2}{m - 1}(r_{2, 1} + r_{2, 2} - 1)r_{m, (m-2)} + \frac{1}{m - 1}(r_{1, 1} - 1)r_{m, (m-1)} &= 0
\end{align}

i.e., adding the equations we obtain

\begin{align*}
0r_{m, 1} + \frac{m - 2}{m - 1}0r_{m, 2} + \ldots + \frac{2}{m - 1}0r_{m, (m-2)} + \frac{1}{m - 1}0r_{m, (m-1)} &= 0
\end{align*}

i.e., we can choose \( r_{m, 1}, r_{m, 2}, \ldots r_{m, m} \) independently.
We have \(^{(m+1)}\binom{2}{2}\) equations and \(^{(m+1)}\binom{2}{2}\) unknowns, so we should be able to define all the probabilities. Since some of these equations are dependent, we have degrees of freedom and we have flexibility for choosing some of those probabilities based on the component in the solution space.

The sample space for a pot with tiles, is the set of all possible outcomes from the self-assembly process. So if \(S_m\) is a sample space for a pot with \(m\) tiles, \(E_{r,\epsilon}^k\) is the set of all outcomes whose proportion of tiles in \(k\)-cycles is between \(r - \epsilon\) and \(r + \epsilon\), we get

\[
\text{measure}(E_{r,\epsilon}^k) = P(E_{r,\epsilon}^k).
\]

Consider the pot type \(P = \{t\}\), with \(t(h) = t(\hat{h}) = 1\). Denote by \(P_i = \{t_1, t_2, t_3, \ldots, t_i\}\) the pot with \(i\) tiles from the type in \(P\), and by \(S_i\) the sample space for \(P_i\).

Given a sample space \(S_k\), we can define certain probabilities to each of the elements in the space, and this way for different probabilities we get different family of probabilities. One such family is uniform distribution of each elements in the probability space.

Since we can define one probability space, the existence of the probability space is proved.
Ana Staninska was born in Skopje, Macedonia. She received her B.S. degree in Mathematics at the University “Sv. Kiril i Metodij”, Skopje, Macedonia in 2000. She entered the Master program in Mathematics at the University of South Florida in 2001 and received her M.A Degree in 2003. She entered the P.h.D. program in Mathematics at the University of South Florida. She works on a mathematical model of DNA self-assembly. Her Ph.D. advisors are Dr. Nataša Jonoska and Dr. Gregory L. McColm. At USF, Ana taught several undergraduate courses as a Teaching Assistant. She spent three months working on a project in Dr. Ned Seeman’s Laboratory at NYU.

Ana likes all kinds of sports, especially biking, running and aerobics.