Spring 2014

Structure vs. Properties Using Chemical Graph Theory

Tabitha N. Williford

STRUCTURE VS. PROPERTIES USING CHEMICAL GRAPH THEORY

by

TABITHA WILLIFORD

(Under the Direction of Hua Wang)

ABSTRACT

Chemical graph theory began as a way for mathematicians to bring together the areas of the Physical Sciences and Mathematics. Through its use, mathematicians are able to model chemical systems, predict their properties as well as structure-property relationships. In this dissertation, we consider two questions involving chemical graph theory and its applications. We first look at tree-like polyphenyl systems, which form an important family of compounds in Chemistry, particularly in Material Science. The importance can be seen in LEDs, transmitters, and electronics. In recent years, many extremal results regarding such systems under specific constraints have been reported. More specifically are the sub-categories of such systems with extremal Wiener indices. We provide a labeling of the vertices on each hexagon, which facilitates the illustration of a tree-like polyphenyl system with its corresponding tree structure. This approach helps to characterize the extremal tree-like polyphenyl systems with respect to the Wiener index (ones that minimize or maximize the Wiener index). This method can also be used to order these systems and the results will aid in predicting the physical properties of compounds. We further compare the study between tree-like polyphenyl systems that resulted from different tree structures. We then focus on the application of a general weighted distance-based indexing system, similar to but more complicated than the Wiener index, to devise a way to determine the binding of proteins in

biochemical systems. Proteins, composed of amino acids, are important biological molecules that dictate a wide range of functions on a cellular level. Although proteins have various basic sequences, proteins can be created to have similar functions. We compare peptide sequences that govern protein binding to methylated DNA. We do this by first devising an index for each amino acid and compare the overall index of a peptide sequence in order to characterize the binding capability between original proteins and mimics.

Key Words: Chemical Graph Theory, Wiener Index, tree structures, polyphenyl, topological indices, protein binding

2009 *Mathematics Subject Classification*: 05C05, 05C78, 92E10

STRUCTURE VS. PROPERTIES USING CHEMICAL GRAPH THEORY

by

TABITHA WILLIFORD

B.S. in Biology

B.S. in Mathematics

A Thesis Submitted to the Graduate Faculty of Georgia Southern University in Partial Fulfillment

of the Requirement for the Degree

MASTER OF SCIENCE

STATESBORO, GEORGIA

2014

STRUCTURE VS. PROPERTIES USING CHEMICAL GRAPH THEORY

by

TABITHA WILLIFORD

Major Professor: Hua Wang

Amanda Stewart

Committee: Emil Iacob

Electronic Version Approved:

May 9, 2014

ACKNOWLEDGMENTS

I wish to acknowledge and thank my advisors, Drs. Wang and Stewart, whose encouragement, guidance, and support have not only helped me professionally, but also in my personal development as I have conducted my research. I appreciate their flexibility and willingness to work in between departments that allowed me to combine two of my passions, mathematics and chemistry. I would so like to thank my other committee member, Dr. Emil Iacob, for guidance on this project.

I also wish to acknowledge all those who have had a major influence on my life in regards to my growth as a student in love with mathematics and the sciences. This list includes but is not limited to: Dr. Argene Claxton, Mrs. Mary Gentry, and Mrs. Emily Heath, who helped to cultivate my love for science and mathematics, push me to excellence, and copious amounts of encouragement early in my educational career. Drs. Alina Iacob, Lissa Leege, and William Irby from my undergraduate career at Georgia Southern, who provided support and gave me my first love of higher educational learning. I would also like to thank all those in the Mathematics, Chemistry, and Biology Departments, who supported and encouraged me through my time here at Georgia Southern obtaining two undergraduate degrees and a Masters degree, although countless others considered the task 'crazy'.

Next, I would especially like to thank my family for their continued emotional support and encouragement. I specifically like to acknowledge my parents and my grandmother, who early on instilled in me the love of learning and who I owe so much to for making me the person I am today. I would also like to thank my sisters, for their support, they were always there, even for late-night phone calls or rants, during my studies even if they had no idea what I was talking about at the time. I also thank my Georgia Southern friends and family for all their motivation, help with research, and for providing so much happiness over the past six years. I would not be where I

am today, especially with my sanity intact, without them.

Last and certainly not least, I would like to thank my Heavenly Father who has made everything possible.

Partial funding was provided through the Sims Foundation (# 245307).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND BACKGROUND

Chemical Graph Theory is an area of mathematics that spans both the mathematical and chemical worlds in their scope and application. In this dissertation, we address two questions concerning both structures of chemicals and their properties, in particular tree-like polyphenyl systems and peptide binding. In order to discuss this research, We first need to introduce various topics that are essential to its understanding and application. We begin with a brief history of chemical graph theory, followed by preliminary information on tree structures, polyphenyl compounds, and the Wiener index. We then introduce the concepts involving proteins and their formation, along with MBD Proteins and their binding to methylated DNA.

1.1 Chemical Graph Theory

Chemical Graph Theory began as a way to combine the natural sciences with mathematics. Chemical graphs were first used in the later eighteenth century as the basic understanding of matter and particles were being discovered. A Scottish chemist, William Cullen, first termed them as affinity diagrams in lecture notes to represent forces between pairs of molecules in chemical reactions [26]. In Cullens work, as well as later chemical graph theorists, numbers or symbols were used between interacting atoms to denote different magnitudes, whether it be gravitational or attraction forces [14]. However, the lines between pairs were not placed as bonds until later in the 19th century when the atom began to be studied more in depth. Some early structural formula for chemical graphs was constructed by Arthur Cayley and James Sylvester [4, 25]. Cayley used kenograms, or alkane tree graphs, to identify different isomers and is most likely the first to publish results on molecular graphs. Molecular graphs are also called chemical graphs and are a structural representation of a chemical com-

pound, where vertices are used to represent atoms and edges to represent chemical bonds. Sylvester then looked at chemicographs and the conditions for the existence of chemical graphs themselves. In the years after, the area of chemical graph theory has had periods of great interest followed by periods of abandonment. The idea of topological indices used for hydrocarbons, began with Hermann Kopp, who summed atoms of different types in molecules to determine volumes and densities of molecules [15]. Later, Harry Wiener used a special distance-based index to discover a relationship between physical properties of alkenes and their boiling points. He first termed this as path number and defined it as the sum of the distances between any two carbon atoms in the molecule, in terms of Carbon-Carbon bonds. This index was later known as the Wiener Index and is one of the most widely known indices used to characterize tree structures [30].

1.2 Tree Structures

Tree structures were first studied by Sir Arthur Cayley in 1857, who also termed kenograms. Tree structures are defined as a connected, acyclic graph or graphs with no cycles. Each point, or vertex, on the graph is called a node and segment is termed a branch (edge in mathematical graph theory). Due to the acyclic nature of the graphs each node is connected to at least one other node by only one edge. Therefore, a tree with n nodes has n-1 edges. In some texts, the points of connection can also be known as forks [13]. The final nodes on each of the branches are known as leaves. One of the most well known formulas concerning tree structures was Cayley's Formula. This formula identifies the number of different trees that can be constructed for n vertices. The simplest version of the formula is

$$|T_n| = n^{n-2}$$

For each of these systems, no two nodes share more than one branch and no

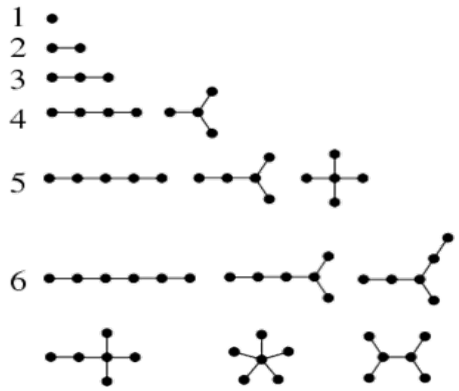

Figure 1.1: Examples of Tree Structures

one branch is attached to one node and no more than two nodes share a branch (i.e. simple graphs). Tree structures are very often used to represent hierarchical data. One example would be in computer science, such as an R-tree for spatial access methods, or in biology, such as evolutionary trees or cladograms.

1.2.1 Polyphenyl Compounds

Polyphenyl compounds are synthetically or naturally derived compounds composed of multiple phenyl/benzene rings. These compounds can be hard to isolate and characterize due to the high variability and probability of impurities associated with their synthesis [27]. Therefore, the properties that have been reported are usually determined in large batches or are a mixture of the various formations. These compounds have been known to be useful in the area of Material Science, which include organic light emitting diodes, catalysts, and transmitters, along with some biological applications [17]. They have also been used in molecular models of graphene as well as discotic liquid crystals due to their higher solubilities, high thermal stability and lowered melting points [9]. The integration of polyphenyl compounds to polymer backbones have been shown in various studies to increase the high glass transition temperature (T_g), lower the degree of molecular association, and even create a trans-

parent film all the while conserving the properties of the original polymer [9]. It will be helpful to be able to predict properties of these compounds due to the associated applications, costly nature of synthesis, and purification techniques as well as the wide ranges of properties and conformations. Topological indices have been used as a convenient abstraction of chemical structures and have shown strong correlations with the chemicals physical properties. However, we are more concerned with tree-like polyphenyl systems which are very common in this group of compounds.

1.2.2 Tree-like Polyphenyl Systems

A polyphenyl system Z is "tree-like" if each vertex of Z lies on exactly one hexagon and the graph obtained by contracting each hexagon into a vertex is a tree (Figure 1.2).

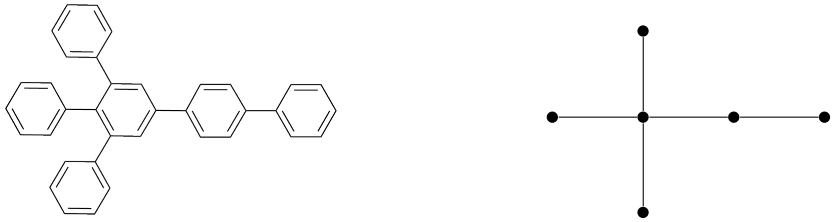

Figure 1.2: A tree-like polyphenyl system and the corresponding tree

Two adjacent vertices in the tree structure correspond to the two hexagons joined by an edge. Each vertex (on a hexagon) is incident to no more than one of the edges joining hexagons. For instance, Figure 1.3 shows an example of a structure that is not a tree-like polyphenyl systems. The structure has two vertices that are shared by more than one hexagon.

Different tree-like polyphenyl systems may be reduced to the same tree. For instance, Figure 1.4 shows a tree-like polyphenyl system reduced to the same tree

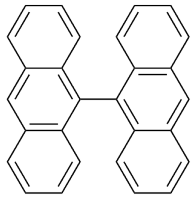

Figure 1.3: Example of structure that is not a tree-like polyphenyl system

structure (as that in Figure 1.2) after contraction of hexagons.

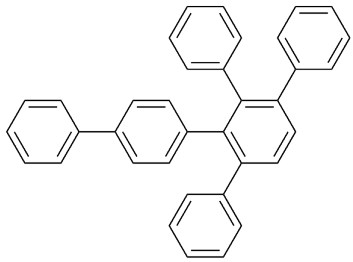

Figure 1.4: A different polyphenyl system

1.2.3 Wiener Index

Topological indices have been used as a convenient abstraction of chemical structures and have experimentally shown strong correlations with the chemical's physical properties. Throughout the years, numerous such indices are proposed, known as the chemical indices, for various categories of chemical structures. One of the most well-known such indices is due to and later named after Wiener [30]. This index is defined as the sum of the lengths of the shortest paths between all pairs of vertices in a chemical graph. This topological index can be used to determine how dense a chemical graph is as well as interactions between atoms in a chemical, for instance interactions that occur when boiling a substance. A boiling point occurs when the movement of the atoms in a substance becomes so rapid that the interactions of the atoms involved change, more specifically the interactions between electrons in the

electron clouds. As a molecule becomes larger, there are more pairs of interactions. Therefore the distance is also greater, which in turn increases the boiling point due to the increase in energy to break those bonds. The Wiener index is a good foundation for determining if there is a correlation between a molecule's structure and properties because it takes into account distances between atoms, more specifically their electron clouds. So when the molecule increases, the Wiener Index also increases.

In the past decade, many studies have been conducted on the Wiener Index. Both maximal and minimal Wiener indices have been characterized for trees with a given number of vertices, degree sequences, and those with a given order with only vertices of two different degrees [16, 29, 32, 33]. The extremal structures and edge-Wiener indices have also been studied among polyphenyl chains, along with the maximal and minimal degree distances [6, 12, 16]. Explicit formulas have also been derived for each of these [2, 6]. Characterization of spiro and polyphenyl hexagonal chains have also been computed as well as to establish a relationship between Wiener indices of these chains to their polycyclic aromatic hydrocarbons to determine the extremal graphs [16]. In addition to these few studies, there are certainly many more studies on the Wiener index and related concepts on trees. In recent years, similar studies have been conducted on some specific tree-like polyphenyl systems (see [2]) and related questions on such systems have been of interests (see [6, 31]).

1.3 Amino Acids, Proteins, and their Binding

1.3.1 Proteins and Amino Acids

Proteins are biological macro-molecules that dictate nearly every process that occurs in the cell for life to exist. These macro-molecules exist in various shapes and sizes and have a wide array of functions. These functions range from the oxygen carrying

hemoglobin located in blood cells, to the keratin in fingernails, to proteins such as polymerase that aids in the synthesis of DNA, the basis of our genetic information. Even though proteins have various functions and structures in nature they are all created from the same basic unit, amino acids. Amino acids are the basic building blocks of proteins. There are 20 common amino acids. The first, Asparagine, was discovered in 1806, while the last, Threonine, was discovered over twenty years later in 1938. All of the 20 amino acids have the same basic structure, a carboxyl group (COO^-) and an amino group (NH_3^+) bonded to a Carbon, called the α-Carbon. Attached to this α-Carbon are the various R groups or side chains that determine the size, charge, and structure of each amino acid [20]. These R groups are what scientists use to classify amino acids, whether the side groups are non-polar, aromatic, or polar (uncharged, positively charged, or negatively charged)(Figure 1.5)

The process for which these amino acids become proteins is called translation, where peptide bonds are created between two amino acids though a condensation reaction. This process forms peptides. When these peptides combine, proteins are formed and take the form of α-helices or β-sheets. A family of proteins containing β-sheets are Methyl-C_pG binding domains, or MBD proteins, that bind to methylated DNA.

1.3.2 Methylated DNA and MBD Proteins

Methylated DNA occurs when a methyl group is added to cytosine or adenine on a DNA strand. In humans, DNA is usually modified through cytosine methylation at C_pG sites by DNA methyltransferase (DNMT1) [1]; where Cytosine and Guanine nucleotides occur next to each other. This modification has been seen to occur in 60% to 90% of all C_pG sites [7]. DNA Cytosine methylation has been linked to the regulation of the X-chromosome, genomic imprinting, carcinogenesis, and gene

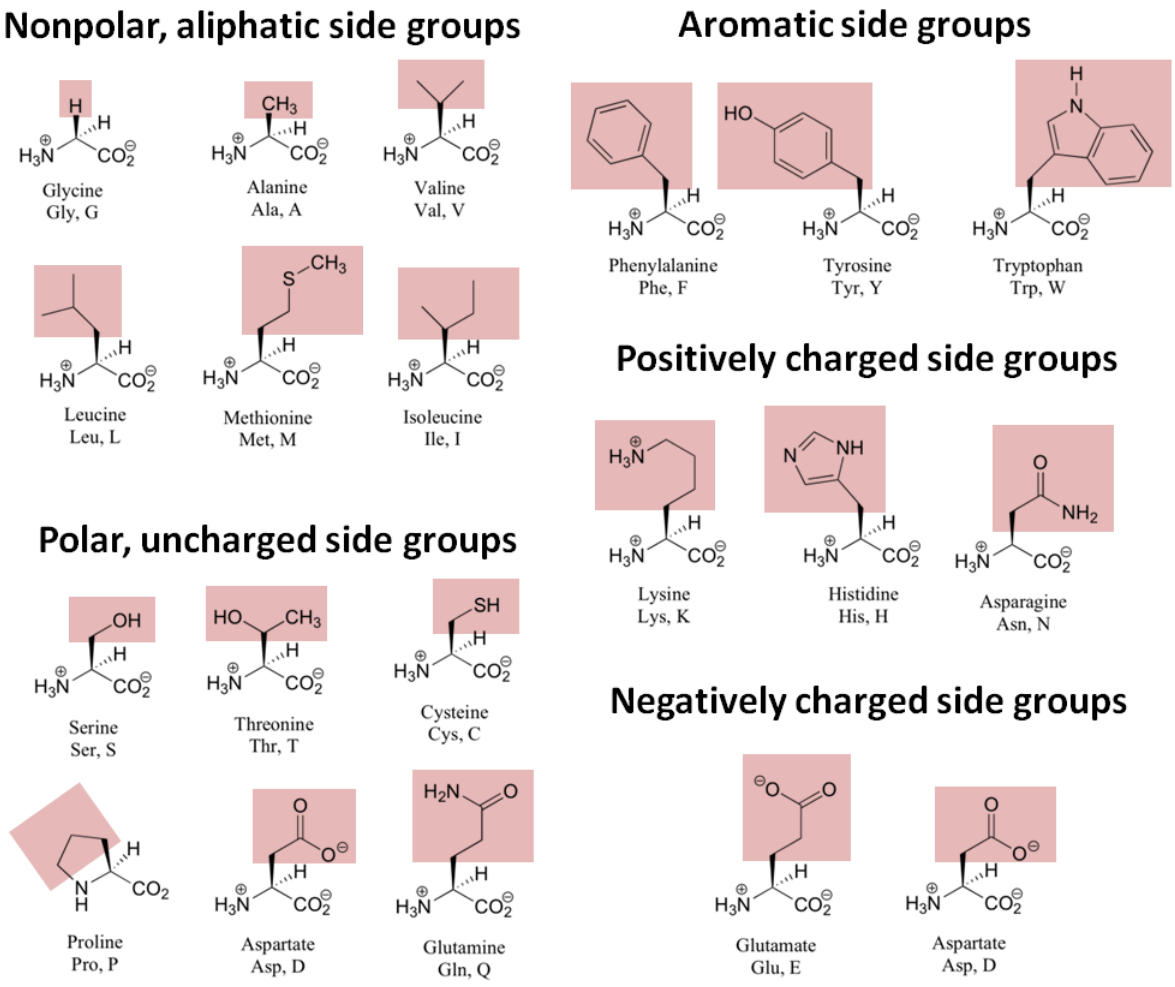

Figure 1.5: 20 Common Amino Acids

silencing. This methylation has also been shown to occur during embryogenesis and is highly important for neurological development [8]. Consequently, alterations at these methylated sites have been linked to various human diseases, such as Retts Syndrome and Fragile-X [24]. This methylation can affect gene expression though two mechanisms; direct interference of methyl-C_pG sites with the binding of transcription factors or as a group of proteins that bind methylated C_pG sites independently [3]. In mammals, there are five known MBD proteins that can recognize methylated DNA; MeCP2, MBD1, MBD2, MBD3, and MBD4. Each of these proteins functions in a similar way by regulating other complexes, and each contains a similar MBD region (Figure 1.6).

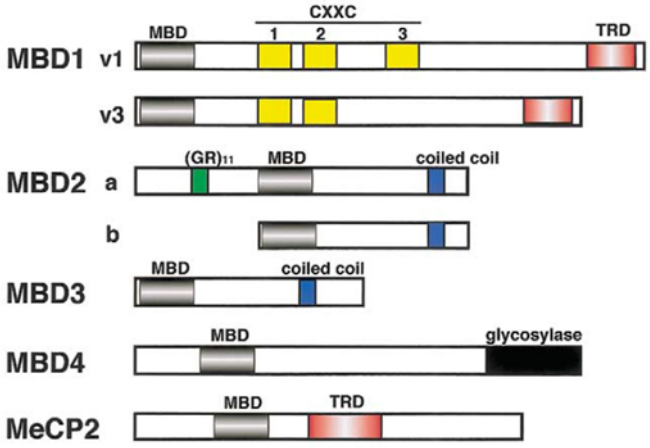

Figure 1.6: MBD region consistent throughout MBD proteins [19]

In the figure, the MBD regions are highlighted along with other regions included in these proteins, such as transcriptional repression domains (TRD), Cysteine rich domains (CxxC), and Glycine and Arginine repeats (GR). Even though these proteins share similar regions, they do have different regulatory functions including binding to methylated DNA to inhibit promoter activities of genes during transcription (MeCP2, MBD1 and MBD2), repressing transcription without specific binding (MBD3), and repairing DNA (MBD4). The study of these proteins is important because of their

importance for transcriptional regulation. Due to the link to various neurological disorders and cancer, these MBD proteins have been the focus of various studies to understand their mechanism of action and possible ways to prevent the alterations. For instance, a mutation in the MeCP2 gene is specifically linked to Rett Syndrome and MBD2 to the transcriptional silencing of hypermethylated genes in cancer [3, 18]. Learning more about the impacts of the alterations of the methylated sequences, the genes coding for MBD proteins, and their DNA binding will provide insight into correcting these imbalances that could cause various neurological diseases.

1.4 Purpose

A simple labeling system of hexagonal vertices will be presented that enables concise tree representations of tree-like polyphenyl systems. This system provides the explicit characteristics of the extremal structures that minimize or maximize the Wiener index among tree-like polyphenyl systems with the same underlying tree structure. The impact of different tree structures on the polyphenyl systems will be discussed through the consideration of pairs of adjacent hexagons as well as the results of simple systems to predicted physical properties.

With the knowledge of these MBD proteins and their binding to methylated DNA, a peptide mimic sequence can be developed to address the impact on the regulation of transcription. An index, similar to the Wiener index, will be devised for each of the 20 amino acids and used to determine a linear expression for a particular peptide sequence. This expression and the results will serve as a tool for comparing and quantifing the relationship of an original protein binding sequence and the created peptide mimic sequence. The information obtained from this study can be used to justify or improve a predicted sequence.

CHAPTER 2

TREE-LIKE POLYPHENYL SYSTEMS AND THEIR WIENER INDICES

2.1 Labeling of hexagonal vertices

For the purpose of distinguishing such systems with the same tree structure, we label the vertices on each hexagon with 1, 2, 4, 6, 5, 3 in the clockwise order as in Figure 2.1.

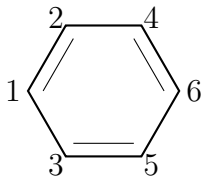

Figure 2.1: Labeling of an aromatic ring in a tree-like system

Remark 2.1. *This labeling of an aromatic ring, although seemingly unusual, emphasizes the importance of adjacent and opposite atoms of this aromatic ring in the tree-like system. The numbering is indeed coherent with the ordering of branching sizes when the Wiener index is minimized.*

For an edge connecting two vertices from different hexagons in a tree-like system, we label the two end of this edge to denote where is the hexagon connected to this edge. For instance, the system in Figure 1.2 can now be represented as Figure 2.2.

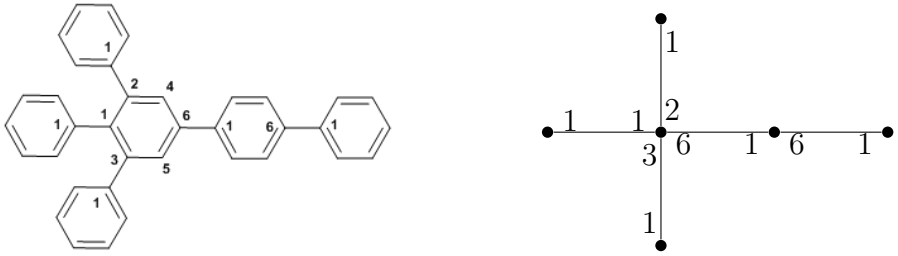

Figure 2.2: Labeling and tree representation with edge labels for Figure 1.2

We omit an edge label if it does not affect the tree-like system. In particular, we do not label the leaf-ends of pendant edges. Figure 2.3 shows another example with such labellings. Note that this example denotes a different system that shares exactly the same tree structure.

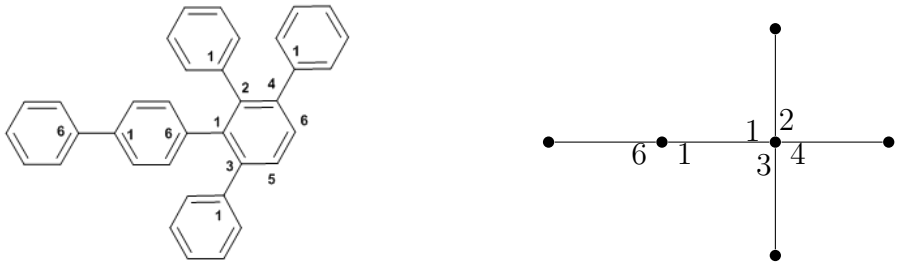

Figure 2.3: A different edge labeling pattern for Figure 1.4 with same tree-like structure

2.2 Tree-like polyphenyl system with a given tree structure

For a graph G with vertex set $V(G)$ and edge set $E(G)$, the Wiener index of G is defined as

$$W(G) = \sum_{u,v \in V(G)} d(u,v)$$

where $d(u,v)$ denotes the *distance* between u and v (the number of edges on the shortest path connecting u and v) and the sum goes over all unordered pairs of vertices in G.

However, here we consider polyphenyl systems Z with a given underlying tree structure T. First recall that the Wiener index of a tree T can also be represented by

$$W(T) = \sum_{(u,v) \in E(T)} n(u)n(v)$$

where $n(u)$ and $n(v)$ are the numbers of vertices in T that are closer to u or v respectively.

Following the same idea, we have

Proposition 2.2. *The number of times an edge $uv \in E(T)$ is used as part of a path in Z is*

$$(6n(u)) \cdot (6n(v)) = 36n(u)n(v).$$

The sum of these values for all edges in T is

$$36W(T).$$

Consequently,

$$W(Z) = 36W(T) + C(Z)$$

where $C(Z)$, the contribution to $W(Z)$ from hexagonal edges, is the only variable that we need to consider (since $W(T)$ is a constant when T is given).

Let the components resulted from removing the edges of a hexagon in Z be denoted by $Z_1, \ldots, Z_6$ (Figure 2.4) according to the labeling of the vertices on the aromatic ring, drawn here and throughout the rest of the article as a hexagon. Each component contains a polyphenyl system based around a central aromatic ring.

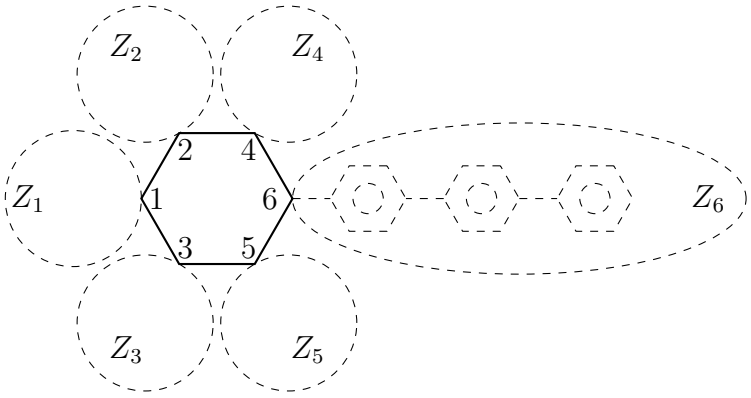

Figure 2.4: Z, represented by a hexagon and the resulted components

Take, for instance, a vertex $v_2 \in Z_2$ and a vertex $v_6 \in Z_6$, the contribution of edges on this hexagon to $d(v_2, v_6)$ is 2. Hence the total contribution of this hexagon to distances between vertices in Z_2 and Z_6 is $2z_2z_6$ where $z_i = |V(Z_i)|$ for $1 \leq i \leq 6$.

Taking all pairs of components into consideration, we have the contribution of this hexagon to $C(Z)$ as

$$\begin{aligned}
&(z_1z_2 + z_1z_3 + z_2z_4 + z_3z_5 + z_4z_6 + z_5z_6) \\
&+ 2(z_1z_4 + z_1z_5 + z_2z_6 + z_3z_6 + z_2z_3 + z_4z_5) \\
&+ 3(z_1z_6 + z_2z_5 + z_3z_4) \\
=&(z_1 + z_2 + z_3 + z_4 + z_5 + z_6)^2 - (z_1^2 + z_2^2 + z_3^2 + z_4^2 + z_5^2 + z_6^2) \\
&+ (z_1z_6 + z_2z_5 + z_3z_4) - (z_1z_2 + z_1z_3 + z_2z_4 + z_3z_5 + z_4z_6 + z_5z_6).
\end{aligned}$$

Note that, with given underlying tree structure and choice of the hexagon under consideration, both

$$(z_1 + z_2 + z_3 + z_4 + z_5 + z_6)^2 = |V(Z)|^2$$

and

$$(z_1^2 + z_2^2 + z_3^2 + z_4^2 + z_5^2 + z_6^2)$$

are constants. Hence we only need to focus our attention on

$$(z_1z_6 + z_2z_5 + z_3z_4) - (z_1z_2 + z_1z_3 + z_2z_4 + z_3z_5 + z_4z_6 + z_5z_6) \tag{2.1}$$

with given values of z_i.

We will show that, with given choices of Z_i's but flexibility to rearrange them, (2.1) is minimized when the components is arranged in a way such that

$$z_1 \geq z_2 \geq z_3 \geq z_4 \geq z_5 \geq z_6, \tag{2.2}$$

i.e., the "largest" component is attached to the hexagon at "1", the second largest at 2, etc..

Lemma 2.3. *The value of*

$$z_1z_6 + z_2z_5 + z_3z_4 \tag{2.3}$$

is minimized under condition (2.2)*.*

Proof. Without loss of generality, assume that $z_1 \geq z_i$ for any $2 \leq i \leq 6$. Supposing (for contradiction) that (2.2) does not hold, we have the following cases:

- If $z_6 > z_4$, consider the new system resulted from replacing Z_4 with Z_6 and Z_6 with Z_4. In the rest of this article we will simply refer to this operation as "switching" the corresponding components. Now the new value for (2.3) is

$$z_1 z_4 + z_2 z_5 + z_3 z_6. \tag{2.4}$$

  Comparing with the original value yields (2.4) − (2.3) as

$$z_1 z_4 - z_1 z_6 + z_3 z_6 - z_3 z_4 = (z_1 - z_3)(z_4 - z_6) \leq 0,$$

  showing that the new system bears a value for (2.3) that is at most as large.

- Similarly, if $z_6 > z_5$, switching Z_6 and Z_5 yields the same conclusion.

- If $z_6 > z_2$ (or $z_6 > z_3$), switching Z_2 and Z_6 (or Z_3 and Z_6) will not increase the value (2.3). The calculation is similar and we leave it to the reader.

Now we may assume that $z_1 \geq z_i \geq z_6$ for any $2 \leq i \leq 5$. Focusing on z_i for $2 \leq i \leq 5$ and the value of $z_2 z_5 + z_3 z_4$, through similar argument, it is easy to see that (z_2, z_5) and (z_3, z_4) must be paired such that the largest (i.e., z_2) and smallest (i.e., z_5) values are paired together. □

Remark 2.4. *Note that* (2.2) *is a stronger condition than what we needed here but nevertheless minimizes* (2.3).

For the second part of (2.1), we have the following through similar but slightly more complicated analysis.

Lemma 2.5. *The value of*

$$z_1 z_2 + z_1 z_3 + z_2 z_4 + z_3 z_5 + z_4 z_6 + z_5 z_6 \tag{2.5}$$

is maximized if and only if condition (2.2) *holds.*

Proof. Without loss of generality, assume that $z_1 \geq z_i$ for any $2 \leq i \leq 6$. Supposing that (2.2) does not hold:

- If $z_4 > z_2$, switching Z_2 and Z_4 yields a new value of (2.5) that is

$$z_1 z_4 - z_1 z_2 + z_2 z_6 - z_4 z_6 = (z_1 - z_6)(z_4 - z_2) \geq 0$$

  more than the original.

- If $z_4 > z_3$, switching Z_3 and Z_4, Z_5 and Z_6 (note that we are essentially "flipping" the portion $Z_3 Z_5 Z_6 Z_4$) yields a new value of (2.5) that is at least as large. The calculation is similar and we leave it to the reader.

- Similarly, the cases for $z_5 > z_3$ or $z_5 > z_2$ can be handled in completely analogous way as the previous two cases.

- If $z_6 > z_2$ or $z_6 > z_3$, switching Z_2 and Z_6 or Z_3 and Z_6 will yield new systems with non-decreasing (2.5).

Now we can assume that $z_1 \geq z_2 \geq z_3 \geq \max\{z_4, z_5, z_6\}$. Following the same arguments we have:

- If $z_6 > z_4$, switch Z_4 and Z_6.

- If $z_6 > z_5$, switch Z_5 and Z_6.

Now we can assume that $z_1 \geq z_2 \geq z_3 \geq \max\{z_4, z_5\} \geq \min\{z_4, z_5\} \geq z_6$.

- If $z_5 > z_4$, switch Z_4 and Z_5.

Note that the value of (2.5) will strictly increase under the above assumptions and operations unless the corresponding z_i's are of the same value, we conclude that (2.2) is the necessary and sufficient condition to minimize (2.5). □

Lemmas 2.3 and 2.5 imply that the contribution $C(Z)$ from hexagonal edges are minimized when (2.2) holds for every hexagon. Together with Proposition 2.2, we have

Theorem 2.6. *With a given tree structure, the corresponding polyphenyl system has the minimum Wiener index if and only if condition* (2.2) *holds for every hexagon.*

Remark 2.7. *Theorem 2.6 asserts that, with a given underlying tree structure, to minimize the Wiener index of the corresponding polyphenyl system one simply need to arrange the outgoing edges of every hexagon according to the size of the attached components (i.e., the values of z_i's).*

With Theorem 2.6, one can easily check that Figure 2.3 provides a corresponding polyphenyl system that has the minimal Wiener index among all systems with the same underlying tree structure, i.e., Figure 2.5.

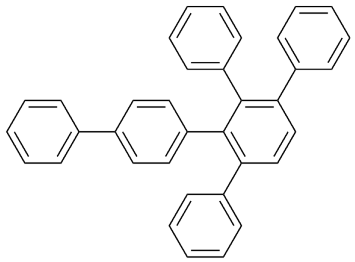

Figure 2.5: An extremal polyphenyl system that minimizes the Wiener index

Remark 2.8. *Although we focus our attention on the extremal structures in this section, our approach can be used to effectively compare the value of the Wiener indices of two isomeric tree-like polyphenyl systems even when they are not extremal. Examples of such application is shown in Section 2.4.*

2.3 Between Adjacent hexagons

In this section we consider the influence, from interchanging pendant branches of two adjacent hexagons, on the Wiener index of a tree-like polyphenyl system. First note that for any two adjacent hexagons as in Figure 2.6, permuting any of the branches (with the possibility of being empty) Z_{ij} ($i = 1, 2$ and $j = 1, 2, 3, 4, 5$) will not affect the contribution to $C(Z)$ from any other hexagons except the two under consideration in Figure 2.6.

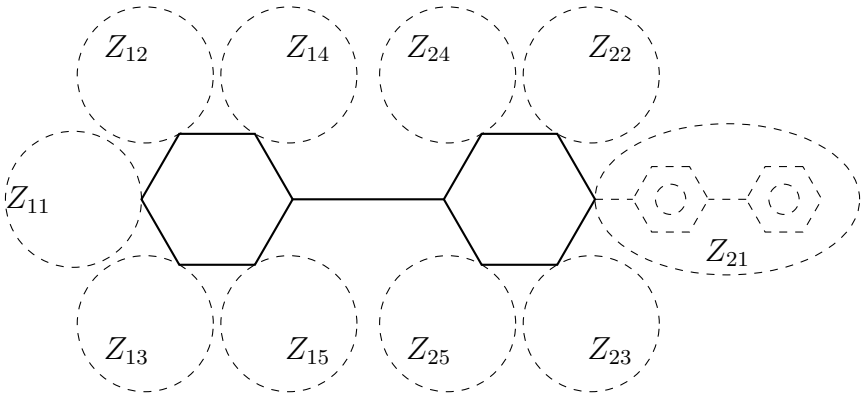

Figure 2.6: Adjacent hexagons and the resulting components

This contribution (from this pair of adjacent hexagons) can be calculated similarly as that from section 2.2 as:

$$\begin{aligned}
&\frac{5}{2}\left(\sum_{j=1}^{5} z_{1j} + \sum_{j=1}^{5} z_{2j}\right)^2 - \left(\sum_{j=1}^{5} z_{1j}^2 + \sum_{j=1}^{5} z_{2j}^2\right) \\
&+\sum_{i=1}^{2}\left(z_{i2}z_{i5} + z_{i3}z_{i4} - (z_{i1}z_{i2} + z_{i1}z_{i3} + z_{i2}z_{i4} + z_{i3}z_{i5})\right) \\
&-\frac{3}{2}\left(\sum_{j=1}^{5} z_{1j}\right)^2 - \frac{3}{2}\left(\sum_{j=1}^{5} z_{2j}\right)^2 \\
&+(z_{11} - z_{14} - z_{15})\left(\sum_{j=1}^{5} z_{2j}\right) + (z_{21} - z_{24} - z_{25})\left(\sum_{j=1}^{5} z_{1j}\right)
\end{aligned}$$

where $z_{ij} = |V(Z_{ij})|$.

Examining this expression, we have

1. The first line
$$\frac{5}{2}\left(\sum_{j=1}^{5} z_{1j} + \sum_{j=1}^{5} z_{2j}\right)^2 - \left(\sum_{j=1}^{5} z_{1j}^2 + \sum_{j=1}^{5} z_{2j}^2\right)$$
is a constant;

2. For any pair of adjacent hexagons in the system, one only needs to consider maximizing or minimizing the expression
$$\begin{aligned} f := & \sum_{i=1}^{2} (z_{i2}z_{i5} + z_{i3}z_{i4} - (z_{i1}z_{i2} + z_{i1}z_{i3} + z_{i2}z_{i4} + z_{i3}z_{i5})) \\ & - \frac{3}{2}\left(\sum_{j=1}^{5} z_{1j}\right)^2 - \frac{3}{2}\left(\sum_{j=1}^{5} z_{2j}\right)^2 \\ & + (z_{11} - z_{14} - z_{15})\left(\sum_{j=1}^{5} z_{2j}\right) + (z_{21} - z_{24} - z_{25})\left(\sum_{j=1}^{5} z_{1j}\right); \end{aligned}$$

3. Repeating (2), one can continue to increase or decrease the expression of f for pairs of adjacent hexagons. Note that in every step the value of $W(Z)$ will be strictly increased or decreased. Hence this process terminates in finite steps.

Remark 2.9. *In terms of the structural change of chemical compounds, the "switching" of Z_{ij}'s is merely breaking and forming bonds (ones that connect some Z_{ij} to one of the two hexagons). Among tree structures, it is known that a complete "chain decomposition" exists among the partially ordered set (ordered by the value of Wiener index) of trees of given order, where every pair of "adjacent" trees in a chain differ by only "breaking and forming" bonds at "adjacent locations". This offers an intuitive support for what is discussed above.*

2.4 Comparison with physical properties

We compare our theoretical studies with the predictions of physical properties of the following polyphenyls from hydrocarbons. Note that A, B, C, D are of the same size.

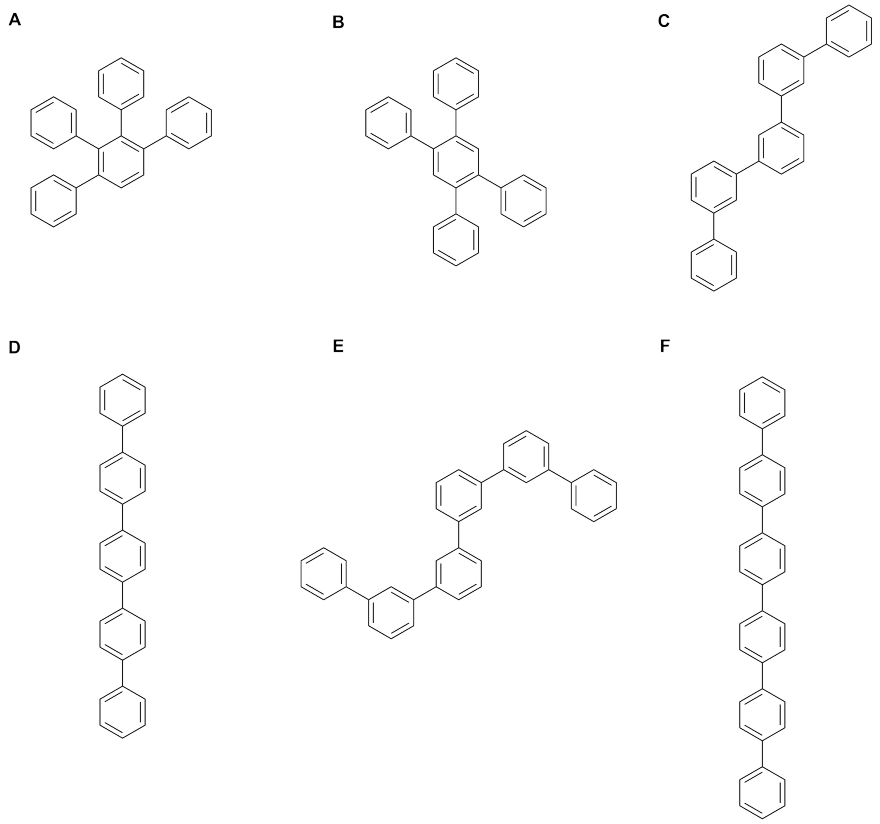

Figure 2.7: Examples of Polyphenyl Structures

Our discussions in sections 2.2 and 2.3 implies that

$$W(A) < W(B) < W(C) < W(D)$$

with A and D being the extremal cases that minimizes and maximizes the Wiener index respectively.

Similarly, our discussion in section 2.2 implies that

$$W(E) < W(F).$$

The following table shows some of the properties of these polyphenyls [34]. In particular, we see a clear correlation between the predicted boiling points, as well as enthalpy of vaporization and density, to the ordering according to the Wiener index.

Polyphenyls	Boiling Point (°C at 760 Torr or 1 atm)	Enthalpy of Vaporization (kJ/mol)	Density (g/cm^3)
A	466.7 ± 40	70.0 ± 0.8	1.091
B	508.5 ± 45	74.9 ± 0.8	1.091
C	567.7 ± 30	82.0 ± 0.8	1.091 ± 0.06
D	618.1 ± 35	88.3 ± 0.8	1.091 ± 0.06
E	646.7 ± 40	91.9 ± 0.8	1.102 ± 0.06
F	703.2 ± 45	99.2 ± 0.8	1.102 ± 0.06

Table 2.1: Predicted Properties of Polyphenyl Compounds (Data Provided from Advanced Chemistry Development Labs http://www.acdlabs.com)

2.5 Concluding remarks

When the underlying tree structure is given, the extremal systems of tree-like polyphenyl system can be specifically characterized using the Wiener index. When the systems

have the same chemical molecular formula, but different structural arrangements (isomers) that possibly provides different tree structures, the study is more complicated. However, a rough algorithm to study such questions is provided. The computational results from the study are compared with physical properties of some simple chemicals that test the validity of the method. The natural question of great importance is to consider the same question for systems that have all possible tree structures. Being able to characterize these structures provide a method for chemists to predict properties of tree-like polyphenyl compounds that could be dangerous or difficult to synthesize. This chemical index, along with various others, also has important applications as these indices are the foundations of chemical prediction and modeling software. It would be interesting to experimentally create these polyphenyl compounds to determine the effectiveness of the predictions.

Further research could include exploring other results using the Wiener index of stars and reduced trees. It is known that among general trees of given order, the star minimizes the Wiener index. We know the contributions from non-hexagonal edges to the Wiener index of a polyphenyl system are minimized when the reduced tree structure is a star. When the order of the reduced tree is at most 7, the star will indeed produce a feasible polyphenyl system.

As a first step of exploring the minimal Wiener index of such systems, trees with given order ≤ 7 and their corresponding polyphenyl systems can be explored through exactly the methods in this note. Note that variations of the tree structure will change the value of (2.1) for each hexagon. Hence more in-depth study is needed for more general structures.

A natural conjecture would be that, among tree-like polyphenyl systems of given order, the minimum (maximum) Wiener index is obtained when the underlying tree structure is extremal (with corresponding constraints such as maximum degree ≤ 6)

and condition (2.2) or (**??**) is satisfied.

CHAPTER 3
USING DISTANCE-BASED TOPOLOGICAL INDICES TO STUDY PROTEIN BINDING

3.1 MBD2 protein

The entire sequence for MBD2 can be found in the Protein Data Base or PBD [23]. The structure of human MBD2 is very similar to that of MBD1 determined by Ohki et. al through the isolation of *E. coli* that was then compared to other sequences in the MBD family [22]. The common sequence located in each of the MBD family members and MeCP2, in particular amino acids 147-215 in MBD2.

N-ESGKRMDCPALPPGWKKEEVIRKSGLSAGKSDVY

YFSPSGKKFRSKPQLARYLGNTVDLSSFDFRTGKM-C

Sequence common to all MBD proteins

However, the particular part of the sequence concerned with the actual binding to methylated DNA are amino acids 162-182,

KKEEVIRKSGLSAGKSDVYYF

MBD Sequence related to binding of Methylated DNA

This peptide sequence forms a loop region in between two beta sheets. These regions are indicated below, where the red amino acids are residues that interact with the DNA bases, the blue amino acids are those interacting with the DNA backbone, the highlighted yellow regions are those that form the beta sheets [22].

KKEEVIRKSGLSAGKSDVYYF

Regions of MBD2 Binding Sequence

Research currently being conducted in the Stewart Research Lab is studying the extension of the loop region in a particular sequence can impact its binding capability through the creation of mimic sequences. Using this idea of loop region extension and conservation of the binding regions, the following sequence was developed as a mimic for MBD2 and Figure 3.1 shows the predicted structure.

KKEEVIRKRQYSGLSAGWQKVRSDVYYF

MBD2 Mimic Sequence

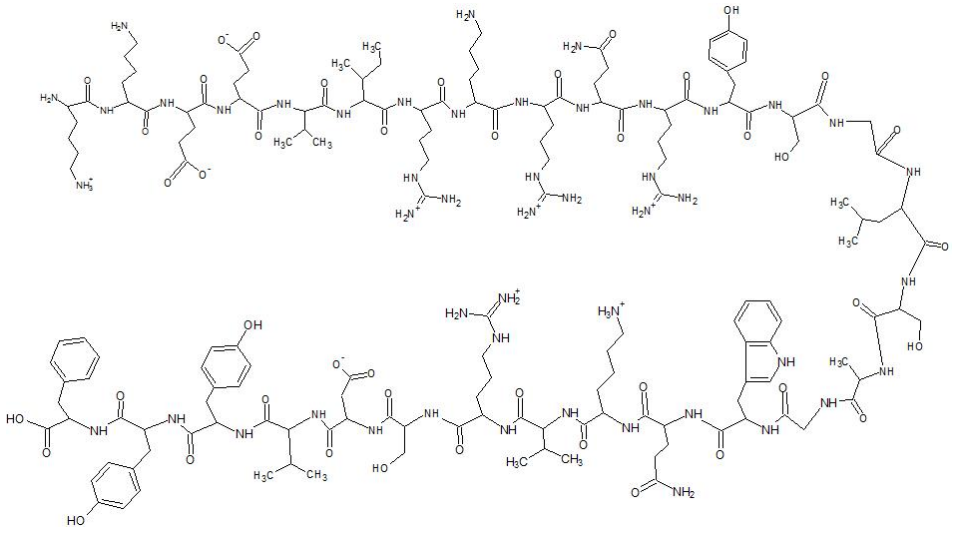

Figure 3.1: Structure of MBD2 Mimic Sequence

In comparing the two sequences (seen below), the beta sheet binding region for the original sequence and the two ends of the mimic are the same. These regions are conserved in an attempt to maintain similar or improved DNA binding affinities by

the peptide mimic.

KKEEVI	RK	SGLSAG	K	SDVYYF
KKEEVI	RKRQY	SGLSAG	WQKVR	SDVYYF

Comparison of MBD2 and MBD2 Mimic

3.2 Calculations for Mathematical Binding

We begin looking into the mathematical binding analysis by first determining the index for each particular amino acid.

3.2.1 Amino Acid Index

We started these calculations by taking the shortest path from each particular atom located in each side chain. The number of atoms, Carbon, Nitrogen, Oxygen, Sulfur, were counted and the distances for each of those atoms were calculated (Table 3.1 and Table 3.2).

The following tables (Tables 3.1 and 3.2) list the number of each atom type located in each side chain and the distances of each of those atoms. These distances are then summed up for each atom in Table 3.3.

Amino Acid	No. Carbon atoms	distances	No. Oxygen atom	distances
Glycine (G)	0	0	0	0
Alanine (A)	1	1	0	0
Valine (V)	3	1,2,2	0	0
Leucine (L)	4	1,2,3,3	0	0
Methionine (M)	3	1,2,4	0	0
Isoleucine (I)	4	1,2,2,3	0	0
Phenylalanine (F)	7	1,2,3,3,4,4,5	0	0
Tyrosine (Y)	7	1,2,3,3,4,4,5	1	6
Tryptophan (W)	9	1,2,3,3,4,4,5,5,6	0	0
Serine (S)	1	1	1	2
Threonine (T)	2	1,2	1	2
Cysteine (C)	1	1	0	0
Proline (P)	3	1,2,2	0	0
Asparagine (N)	2	1,2	1	3
Glutamine (Q)	3	1,2,3	1	4
Lysine (K)	4	1,2,3,4	0	0
Arginine (R)	4	1,2,3,5	0	0
Histidine (H)	4	1,2,3,4	0	0
Aspartate (D)	2	1,2	2	3,3
Glutamate (E)	3	1,2,3	2	4,4

Table 3.1: Shortest Distance Calculations for Carbon and Oxygen Atoms in 20 Common Amino Acids

Amino Acid	No. Nitrogen atoms	distances	No. Sulfur atoms	distances
Glycine (G)	0	0	0	0
Alanine (A)	0	0	0	0
Valine (V)	0	0	0	0
Leucine (L)	0	0	0	0
Methionine (M)	0	0	1	3
Isoleucine (I)	0	0	0	0
Phenylalanine (F)	0	0	0	0
Tyrosine (Y)	0	0	0	0
Tryptophan (W)	1	4	0	0
Serine (S)	0	0	0	0
Threonine (T)	0	0	0	0
Cysteine (C)	0	0	1	2
Proline (P)	0	0	0	0
Asparagine (N)	1	3	0	0
Glutamine (Q)	1	4	0	0
positive				
Lysine (K)	1	5	0	0
Arginine (R)	3	4,6,6	0	0
Histidine (H)	2	3,4	0	0
Aspartate (D)	0	0	0	0
Glutamate (E)	0	0	0	0

Table 3.2: Shortest Distance Calculations for Nitrogen and Sulfur Atoms in 20 Common Amino Acids

Amino Acid	1-Letter Code	C	O	N	S	Index Sum
Non-polar						
Glycine	G	0	0	0	0	0C
Alanine	A	1	0	0	0	1C
Valine	V	5	0	0	0	5C
Leucine	L	9	0	0	0	9C
Methionine	M	7	0	0	3	7C + 3S
Isoleucine	I	8	0	0	0	8C
Aromatic						
Phenylalanine	F	22	0	0	0	22C
Tyrosine	Y	22	6	0	0	22C + 6O
Tryptophan	W	33	0	4	0	33C + 4N
Polar						
uncharged						
Serine	S	1	2	0	0	1C + 2O
Threonine	T	3	2	0	0	3C + 2O
Cysteine	C	1	0	0	2	1C + 2S
Proline	P	5	0	0	0	5C
Asparagine	N	3	3	3	0	3C + 3O + 3N
Glutamine	Q	6	4	4	0	6C + 4O + 4N
positive						
Lysine	K	10	0	5	0	10C + 5N
Arginine	R	11	0	16	0	11C + 16N
Histidine	H	10	0	7	0	10C + 7N
Negative						
Aspartate	D	3	6	0	0	3C + 6O
Glutamate	E	6	8	0	0	6C + 8O

Table 3.3: Index for 20 Common Amino Acids

The sum of these distances for each atom was then taken and to calculate the index for each amino acid. This sum is then multiplied by a set value for each type of atom (shown in Tables 3.3). For convenience, we leave the Carbon, Nitrogen, Oxygen, Sulfur values are left arbitrary. Using the indices calculated for each amino acid we can then look at the MBD2 structure to determine the index of entire peptide.

3.2.2 Calculation of Topological Index for MBD2 and mimic

Since the ends of the peptide are conserved and when we compare the original and the mimic we see a similar sequence of amino acids in the middle loop as well. Therefore we only need to be concerned with the parts that are different from the original. So in order to determine the index, we calculate the distances of the two dissimilar parts to the middle part from each side. (The highlighted regions)

KKEEVI	RK	SGLSAG	K	SDVYYF
KKEEVI	RKRQY	SGLSAG	WQKVR	SDVYYF

Comparison of structures

We do this in a similar way of calculating the indices for each of the amino acids. For each amino acid in the sequence on the two sides, we calculate the distances from each of these amino acids to the middle part.

For example, ook at the first Arginine (R), the distance of the alpha carbon from the Serine (S) located in the middle part is 18. We determine the distance by calculating the number of atoms along the backbone between the Nitrate group of the Serine to the α-Carbon of the Arginine.

For instance, Arginine has 4 Carbon atoms and 3 Nitrogen atoms. Looking at each particular atom, the distances from the α-Carbon of each Carbon is 1, 2, 3 and

5 and those for Nitrogen are 4, 6 and 6 (From Table 3.1 and 3.2). So when we add 18 to each of these distances, we have 19, 20, 21 and 23 for the Carbon atoms and 22, 24 and 24 for the Nitrogen atoms. Therefore the sum of these would be

$$83C + 70N. \tag{3.1}$$

However, we also add the distance of α-Carbon to the number of Carbons because it was not added previously for each amino acid. Therefore, we arrive at the total distance between the Arginine to the Serine to be

$$101C + 70N. \tag{3.2}$$

We continue the process in this fashion for each of the amino acids in both parts of the sequence to the middle part and arrive at the results in Tables 3.4 and 3.5. We then do the same calculations with the original MBD sequence Table 3.6 and Table 3.7.

Amino Acid	α-Carbon	Carbon	Oxygen	Nitrogen
R	18	19,20,21,23	0	22,24,24
K	15	16,17,18,19	0	20
R	12	13,14,15,17	0	16,18,18
Q	9	10,11,12	13	13
R	6	7,8,9,11	0	10,12,12
Y	3	4,5,6,6,7,7,8	9	0
W	3	4,5,6,6,7,7,8,8,9	0	7
Q	6	7,8,9	0	10
K	9	10,11,12,13	10	10
V	12	13,14,14	0	0
R	15	16,17,18,20	0	19,21,21

Table 3.4: Atom Distances for MBD2 Mimic Regions

Amino Acid	Sum Carbon	Sum Oxygen	Sum Nitrogen	Index
R	101	0	70	101C+70N
K	85	0	20	85C+20N
R	71	0	52	71C+52N
Q	42	13	13	42C+13O+13N
R	41	0	34	41C+34N
Y	46	9	0	46C+9O
W	63	0	7	63C+7N
Q	30	0	10	30C+10N
K	55	10	14	55C+10O+14N
V	53	0	0	53C
R	86	0	61	86C+61N

Table 3.5: Indices for MBD2 Mimic

Amino Acid	α-Carbon	Carbon	Oxygen	Nitrogen
R	3	4,5,6,8	0	7,9,9
K	6	7,8,9,10	0	11
K	3	4,5,6,7	0	8

Table 3.6: Atom Distances for MBD2

Amino Acid	Sum Carbon	Sum Oxygen	Sum Nitrogen	Index
R	26	0	25	26C+25N
K	40	0	11	40C+11N
K	25	0	8	25C+8N

Table 3.7: Indices for MBD2 Mimic

	leftside		right side
MBD2	66C + 36N	>	25C + 8N
MBD2 Mimic	386C+22O+189N	>	287C+10O+92N

Table 3.8: Index Comparison for MBD2 and MBD2 Mimic

3.2.3 Comparison of Indices

We begin our comparison of the two sequences by comparing the indices of parts. When we sum up the distances of each amino acid on the two sides of the MBD protein we get that the left side equals

$$66C + 36N. \tag{3.3}$$

and the right side equals

$$25C + 8N. \tag{3.4}$$

For the MBD Mimic, we see that the left side equals

$$386C + 22O + 189N \tag{3.5}$$

and the right side equals

$$287C + 10O + 92N. \tag{3.6}$$

In both cases we see that the left side is greater than the right side.

Therefore, the indices for the MBD2 and MBD2 Mimic sequences follow a similar mathematical pattern. This relationship can not only be seen in the comparison of the binding site, but also the comparison of the other regions. The similarity of the two sequences mathematically suggests the sequences will have also have a similar binding affinity. Synthesis of both MBD2 and MBD2 Mimic peptides with analysis using fluorescence binding studies will determine their binding affinity for methylated DNA. A comparison between the theoretical and experimental results can then be

conducted to provide invaluable information to the understanding and application in the prevention of neurological disorders.

We will also do a similar study concerning other original and mimic sequences:

NF-κB p50-1

N-QRGFRWRYVCEGPSHGGLPG-C

NF-κB p50-2

N-QRGFRFRWVCEGPSHGGLPG-C

Each of these original/mimic pairs bind to a similar substrate, like MBD2 and the MBD2 Mimic. This allows for the further testing of these peptides and the method devised in the application to protein binding.

One of the above, Nuclear Factor-kappa B (NF-κB), is a transcription factor involved in many physiological processes and protein expression, including those linked to a variety of neurological diseases. Peptides will not be designed and synthesized to mimic the binding site of the NF-κB protein, which will act as an inhibitors for this protein and aid in the regulation of protein expression that have implications in the prevention of epilepsy as well as many other neurological disorders.

3.3 Concluding remarks and Future Research

The development of this method, similar to the Wiener index, allows for one to take a peptide sequence or other chemical structure and transform it into a linear expression. Calculating this index turns the atoms composing the amino acids in a peptide sequence into variables, which can then be used to determine the relationship between original sequences and mimics. This method can be used to determine this mathematical relationship prior to synthesis to guide chemist in the process of designing peptide mimics. Ultimately, the perfection of this approach adds a mathematical foundation

to the creation of peptides, not only in the prevention of neurological disorders, but the scientific community as a whole.

Further work, specifically to MBD2, would be to adjust the topological index after more research into the structure after synthesis. As some bonds are not only formed between adjacent amino acids, but also in other locations in a protein due to secondary structure. To do this we would need more information into the construction, formation, and interaction of the amino acids on either side of the loop region as well as the binding site.

Further study regarding the mathematical model could be to look at the atoms, not only as single variables, but as variables with powers. The inclusion of powers would turn our linear equations into higher order polynomials that could be compared to the idea of Graphic Polynomials, such as Matching and Characteristic Polynomials in Graph Theory. These polynomials have applications in molecular orbital theory, resonance theory, and statistical physics. The coefficients and zeros of characteristic and matching polynomials have been shown to be related to extent of branching in a molecule. A polynomial expression would create a model that could take into consideration the actual 3D structure of a protein and a better, more accurate prediction of the binding.

REFERENCES

[1] T. Bestor, A. Laudano, R. Mattaliano, V. Ingram, Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases, *J. Mol. Biol.* 203 (1988) 971–983.

[2] H. Bian, F. Zhang, Tree-like polyphenyl systems with extremal Wiener indices, *MATCH Commun.Math.Comput.Chem.* 61(2009) 631–642.

[3] S. Kriaucionis, A. Bird, DNA methylation and Rett syndrome, *Human Mol. Genetics*, 12(2)(2003) 221–227.

[4] A. Cayley, On the mathematical theory of isomers, *Philos. Mag.* 10(47)(1874) 444–446.

[5] A. Dobrynin, R. Entringer, I. Gutman, Wiener index of trees: Theory and applications, *Acta Appl. Math.* 66(2001) 211–249.

[6] Y. Dou, H. Bian, H. Gao, H. Yu, The Polyphenyl Chains with Extremal Edge-Wiener Indices, *MATCH Commun.Math.Comput.Chem.* 64(2010) 757–766.

[7] M. Ehrlich et. al, Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells, *Nucleic Acids Research*, 10(8)(1982) 2709–2721.

[8] G. Fan, L. Hutnick, Methyl-CpG binding proteins in the nervous system, *Cell Research*, 15(4) (2005) 255–261.

[9] E. Gagnon, T. Maris, P. Arsenault, K. Maly, J. Wuest, Structural Features in Crystals of Derivatives of Benzene with Multiple Contiguous Phenyl Substituents, *Crystal Growth and Design*, 10(2010) 648–657.

[10] I. Gutman, A property of the Wiener number and its modifications, *Indian J. Chem.* Sec. A 36(1997) 128–132.

[11] I. Gutman, B. Furtula, M. Petrović, Terminal Wiener index, *J. Math. Chem.* 46(2)(2009) 522–531.

[12] Y. Hao, M. Wang, Tree-like polyphenyl chains with extremal degree distance, *Digest Journ. of Nanomaterials and Biostructures*, 6(2)(2011) 739–745.

[13] F. Harary, Graph Theory, Reading, MA: Addison-Wesley,(1994).

[14] W. Higgins, A Comparative View of the Phlogistic and Antiphlogistic Theories, London, printed for J Murray,(1789).

[15] H. Kopp, Über den Zusammenhang zwischen der chemischen Constitution und einigen physikalischen Eigenschaften bei flüssigen Verbindungen, *Ann. Chem. Pharm.* 50(1844) 71–144.

[16] M. Fischermann, A. Hoffmann, D. Rautenbach, L. Székely, L. Volkmann, Wiener index versus maximum degree in trees, *Discrete Appl. Math.* 122(2002) 127–137.

[17] C. Lima, M. Rocha, A. Melo, L. Gomes, J. Low, L. Santos, Structural and Thermodynamic Characterization of Polyphenylbenzenes, *J. of Phy. Chem.* 115(2011) 11876–11888.

[18] X. Lin, W. Nelson, Methyl-CpG-binding Domain Protein-2 Mediates Transcriptional Repression Associated with Hypermethylated *GSTP1* CpG Islands in MCF-7 Breast Cancer Cells, *Cancer Res.* 63(2003) 498–504.

[19] M. Nakaoa, S. Sugiko, S. Yamamotoa, K. Okumurab, M. Shirakawac, N. Fujitaa, Regulation of transcription and chromatin by methyl-CpG binding protein MBD1, *Brain and Development*, 23(2001) 174–176.

[20] D. Nelson, D. Nelson, A. Lehninger, M. Cox, "Amino Acids, Peptides, and Proteins." *Lehninger Principles of Biochemistry*, 5th ed. New York: W.H. Freeman, 2008. 71–98.

[21] N. Schmuck, S. Wagner, H. Wang, Greedy trees, caterpillars, and Wiener-type graph invariants, *MATCH Commun.Math.Comput.Chem.* 68(2012) 273–292.

[22] I. Ohki et al., Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA, *Cell.* 105(4) (2001) 487–497.

[23] M. Gnanapragasam, et. al. "RCSB PDB - 2L2L Structure Summary." *Proc. Natl. Acad. Sci. U S A*. 108.18 (2011): 7487-492. *RCSB Protein Data Bank*. Web. 19 Mar. 2014. http://www.rcsb.org/pdb/explore/explore.do?structureId=2L2L.

[24] K. Robertson, DNA methylation in health and disease. *Nat. Rev. Genet.* 1 (2000) 11–19.

[25] J. Sylvester, On an application of the new atomic theory to the graphical representation of the invariants and covariants of binary quantics, with three appendices, *American J. of Math., Pure and Applied*, 1(1)(1878) 64–90.

[26] A. Thackray, Atoms and Powers, *Harvard University Press*, Cambridge, Massachusetts (1970) 266.

[27] M. Unroe, B. Reinhardt, One Pot Sythesis of *p*-Polyphenyls *via* the Intermolecular Cyclization of 3-Dimethyl-aminohex-5-en-1-ynes, *Synthesis.* 11(1987) 981–986.

[28] L.A. Szekely, H. Wang, T. Wu, The sum of distances between leaves of a tree and the 'semi-regular' property, *Disc. Math.* 311(2011) 1197–1203.

[29] H. Wang, The extremal values of the Wiener index of a tree with given degree sequence, *Discrete Applied Mathematics*, 156(2008) 2647–2654.

[30] H. Wiener, Structural determination of paraffin boiling point, *J. Amer. Chem. Soc.* 69(1947) 17–20.

[31] W. Yang, F. Zhang, Wiener Index in Random Polyphenyl Chains, *MATCH Commun.Math.Comput.Chem.* 68(2012) 3711–376.

[32] X.-D. Zhang, Q.-Y. Xiang, L.-Q. Xu, R.-Y. Pan, The Wiener index of trees with given degree sequences, *MATCH Commun.Math.Comput.Chem.* 60(2008) 623–644.

[33] X.-D. Zhang, Y. Liu, M.-X. Han, The Maximum Wiener Index of Trees with Given Degree Sequences, *MATCH Commun.Math.Comput.Chem.* 64(2010) 661–682.

[34] http://www.acdlabs.com/home/, ACD/Percepta Platform, *Advanced Chemistry Development.*