Spring 4-2015

Coding DNA into Music: An Alternate Way of Analysis

Samuel Fesenmeier
University of Dayton

Coding DNA into Music:

An Alternate Way of Analysis

Honors Thesis

Samuel Fesenmeier

Departments: Biology and Music

Advisors: Dr. Mark Nielsen Ph.D., Dr. Tobias Rush D.M.A.

April 2015

Coding DNA into Music:

An Alternate Way of Analysis

Honors Thesis

Samuel Fesenmeier

Department: Biology, Music

Advisors: Mark Nielsen, Ph.D., Dr. Tobias Rush, D.M.A.

April 2015

Abstract

In the study, we are analyzing the human genome in order to determine patterns that may tell something about how DNA functions. Patterns require an explanation: it is highly improbable that they are random occurrence. These patterns may hint to something about how DNA functions. There are known patterns already discovered in DNA. For example, in the coding portion, three base pairs translate to a specific amino acid. In the non coding portion, however, specific patterns are not as simple.

We will search for patterns by applying a coding system that turns DNA into music. Music may serve as a powerful tool because we will be able to use to analyze long codings in a short period of time, and the entire phrases will connect in some way. When you listen to a song the order of the chords and phrases have an impact on the entire sound. DNA is the same way; the individual sequencing tie in together to replicate and transcript our genetic material. We will use known knowledge of DNA to base our system of musical coding off of, such as the individual nucleotide sequencing, amino acid coding and protein binding sites. Specific musical assignments will be given to each of the pairings. For example, an adenine base pair could be assigned a C major chord. The DNA will then be played using a computer programing software. The DNA sequences will then be heard and analyzed for specific patterns.

Acknowledgements

Thank you to all the people who helped me with this project, especially Dr. Nielsen whom had no doubts in believing this was all possible from the start. Thank you also to the Honors Department as well as the Biology and Music Departments. The many people involved in these programs have been of much guidance over the years

UNIVERSITY of DAYTON

Table of Contents

Introduction

We are in the midst of a genomic revolution. The Human Genome has been sequenced and new genomes are being sequenced every month. With all this information available, the need to interpret the sequences has been raised. Statistical analysis and experimentation have helped to unveil some of the core function of our DNA, but there is much more to be learned about how sequences function.

How exactly does a genome act to generate a person, chimp, spider, or any life form? It is understood that DNA is comprised of three basic types: coding, non-coding and repeat. The coding sequence is consistent throughout all life forms and determines the amino acid composition of proteins. The repeat portions are viral in origin and contain no function. Non-coding portions of a protein are the most difficult to read and interpret. Unlike coding where base pairs determine a specific amino acid, there lacks a strong correlation between non-coding sequence and function. Furthermore, computational analysis has lacked the power to distinguish signal from noise. For example, there can be up to 50,000 base pairs of DNA that are involved in the expression of one gene, some can change with no effect while others cannot, and it is not clear why. Testing these sequences experimentally is not feasible, considering the 50,000 base pairs as well as specific combinations of base pairs needed to be tested.

This study will take a completely different approach to studying DNA sequences. This will be done through coding DNA into music. By applying a system that codes known information of DNA into sound, this system could prove to be a powerful means of finding patterns in DNA with unknown function. It is possible that hearing the sequences could allow our brains to pinpoint patterns that are not found through computational or experimental analysis. Music may also open up the expression of complex patterns that are visually hidden. The application of DNA to music will also allow for very long segments to be analyzed in a short period of time. Ideally, discovering certain patterns will lead to candidate sequences that warrant further study.

Why use music as a medium for analyzing DNA? Most importantly, music evokes affect in the brain that entails a novel way of perception. For example, simply looking at a written composition may tell very little about how a piece sounds or how the pitches connect, similarly to how visualizing DNA may not tell us everything. There is an aspect of music that brings all elements together, almost like a story. In the case of the regulatory sequence, it all could read in a

manner to position the polymerase on the promoter. By writing the song of DNA, it possible that connections can be made between sequence and function simply by listening.

A second reason for why music will be an effective tool is that biology and music are very similar in the nested structure of their component interactions. Each contain various components that build together and overlap to "sing" either an intricate piece or replication of cells. These similarities allow for the creation of an isomorphism/relations between the two, which is used to analyze the DNA with music. Table 1 and 2 indicate the following comparisons made in the study.

Biology	Music
Nucleotides	Notes
Amino Acids	Pitches
Alpha Sheet Beta Sheet	Chords
Domains	Keys
Proteins	Song

Table 1

Amino Acids (nucleotide triplets)	Scale degrees
Nonpolar	I, IV, V
Polar	ii, iii, vi
Stop	vii*

Secondary Structure	Chords
Alpha Helix	I chord
Beta Sheet	V chord

Domain	Key
First Domain	Major Scale
Second Domain	Major Scale+ 8 half steps
Third Domain	Major Scale- 3 half steps

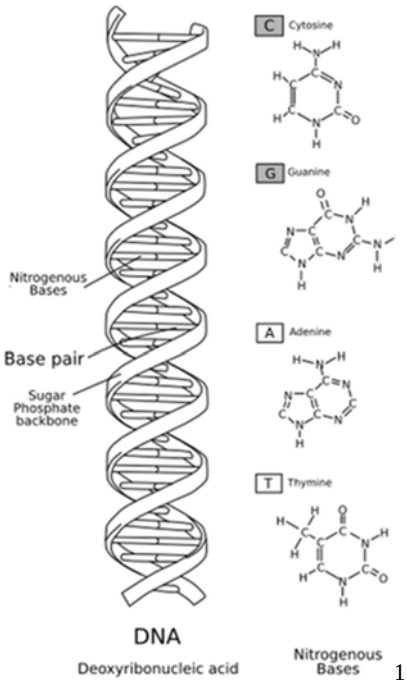
Protein	Song

Quatenary Structure	Proteins played together

Table 2

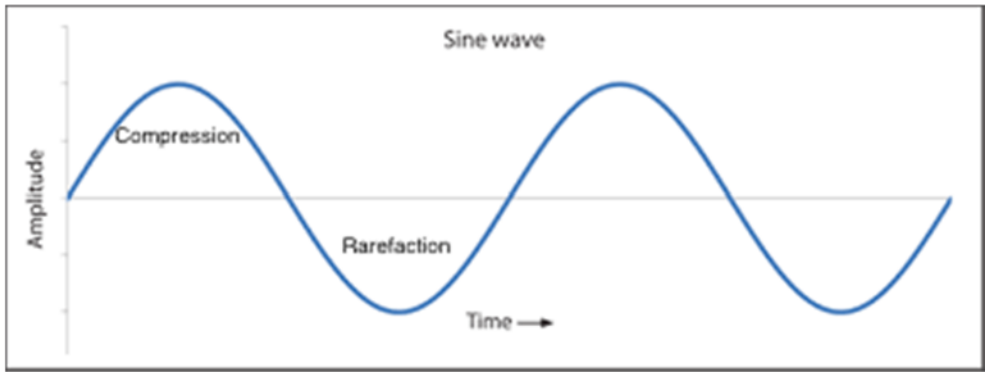
The Nesting of Biology and Music

The core of all life lies in Deoxyribonucleic acid, or DNA. Guanine, Cytosine, Thymine and Adenine are what comprise DNA and millions of these molecules intertwine to form a double Helix. Adenine only links with Thymine and cytosine with guanine. All of these molecules are held together with a phosphate back bone.



1

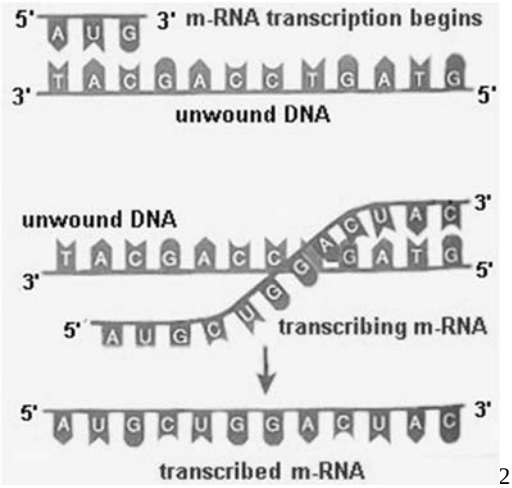
Similarly, the core of music lies in sound waves. These waves consist of a few core elements. All sound waves contain a compression phase and rarefaction phase, which occur consecutively over and over again to produce the wave. One rarefaction followed by a compression combine to form what is called a cycle. All waves share this quality of essentially "up and down" movement, which is similar to how all life have DNA as the core.



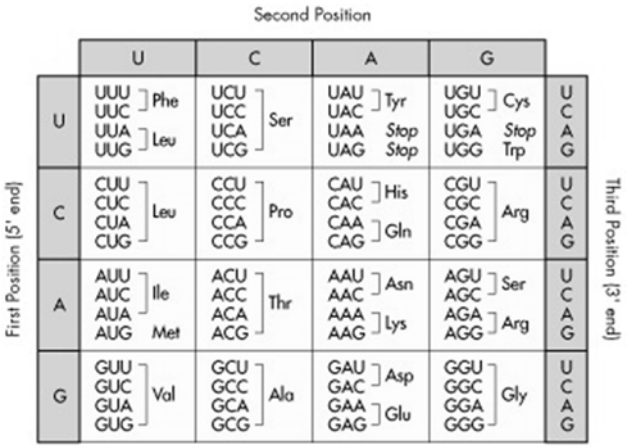
In biology, the complexity of life starts with the specific order of adenine, cytosine, guanine and thymine. The specific order is what allows for variability in all life forms. For example, a chimp and a human both contain DNA, but specific regions vary in the order of molecules and these order of difference is what makes us have larger brains and chimps have more hair.

From the DNA comes transcription, which occurs through a process of splitting the double latter and the formation of RNA through an enzyme. Next, RNA polymerase copies the DNA and sends to ribosomes for translation in a protein. Organisms are essentially walking sacs of protein, making protein the key foundation for life.

When the DNA splits, the broken spots are filled with the same four molecules with the exception that thymine is replaced with Uracil. RNA stays single stranded and is sent off in use of other function.



For the coding region of DNA, the RNA molecules go on to form amino acids. These are formed from the specific ordering and are created by units of three molecules. For example, three adenines in a row would produce the amino acid lysine. There are 64 combinations of triplets of RNA molecules, but there are only 20 amino acids. The reason for this is that some of the triplets account for the same amino acid. The following chart shows what the triplets translate to. The "start" and "stop" codons signify the regions where the chains begin and end.



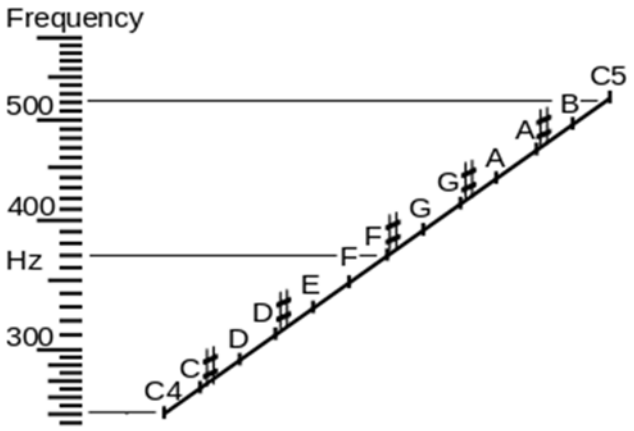
3

Complexity arises from sound waves when the frequencies, or length of cycles, are altered. Frequency is measured in Hertz, which is a unit of cycles per second. Changes in frequencies lead to changes in pitch, which is what makes a note sound high or low. The higher the frequency, the higher the pitch. For example, a bird chirping has a high pitch and high frequency, but a tuba plays notes of low frequency and pitch.

Just as combinations of nucleotides form amino acids in the coding region of DNA, certain combinations of frequencies form melodies and harmonies, which are essential elements for music. Melodies are the central themes in music that essentially the pieces are based upon. They are the progressions of notes that are most recognizable in a piece. Melodies are most brought about in popular music with the lead vocalist. The progression of melodies are relatable to chemical pathways, where each reaction progresses to form the final result or "song." Harmonies are the combination of multiple frequencies play simultaneously with each other. These are what give music depth and help to add strength to the melodies. Of course, there are many elements of biology that are "harmonious," such as multiple organ systems working together simultaneously to make the "orchestra" of an organism.

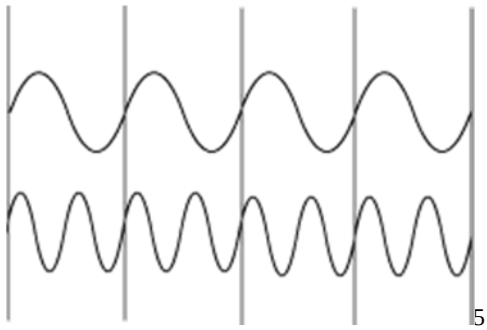
Throughout history, composers have created systems for naming specific frequencies of sound. The scale most commonly used in western music is called the seven-tone scale, which

consists of 7 pitches (A,B,C,E,F,G) as well as intermediate pitches between these notes which are flatted (b) or sharped (#). A note that is flat is a certain interval below the note by itself and a note that is sharped is a note that is higher pitched by the same certain interval than the note. They include A# (Bb) , C# (Db), D# (Eb), F# (Eb), and G# (Ab). The notes in parenthesis are "enharmonic" or "same-sounding" as the previous note. This is the same concept as in the different triplet codons that code for the same amino acid. The following chart represents frequencies that account for pitches in the 7-tone scale.



4

Notice how there are two "C"'s indicated at different frequencies. In fact, all notes occur at multiple frequencies. The pitches can be played at different "octaves," which are the same note but they sound higher or lower than the other. What makes this possible is that the frequency of the note is either split for a higher pitch or doubled for a lower pitch. The "4" and the "5" indicate the specific octave of the note. The following picture represents a note that has been split to create an octave.



5

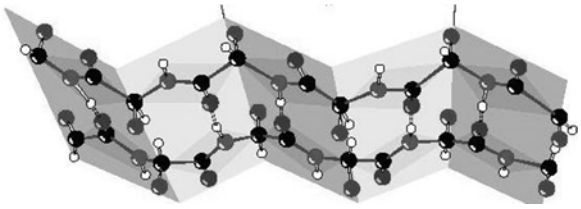
Like the twenty amino acids provide the building blocks for protein and structure, the twelve notes of the seven-tone scales create the core of many musical composition. Still, these building blocks do not tell the entire story. In music, there are still rests (the absence of sound), rhythm, intervals of notes, accidentals, chord progressions, structure of movements, scales, and many other elements to be accounted for. This is also true with amino acids; how these amino acids form and process into life requires millions of interactions, combinations, and even deletions that all derive from coding in the DNA. Many of these interactions are not completely understood, especially in the noncoding region. The same can be said for music; For example, why exactly Beethoven's 5^{th} symphony evokes emotions of anger is certainly up for questioning.

Beta Tubulin

Following the production of amino acid chains comes another essential layer of life: protein formation. The amino acids twist and combine in specific manor and function as a whole in a protein, which allow for even more complex processes in the body such as cell signaling or growth. A similarity in music to these are chords, which consist of three or more notes played together to produce a unifying sound. Chords, like proteins, create an essential part of virtually all music.

A specific case of proteins that was focused upon for the study is the beta tubulin protein, which is considered a globular protein and consists of various subunits. These proteins are what make up microtubules, which are essential for cell function. Microtubules are what provide the cell with system to hold other parts of the cell in place as well as give an interconnecting system of transport to the cell.

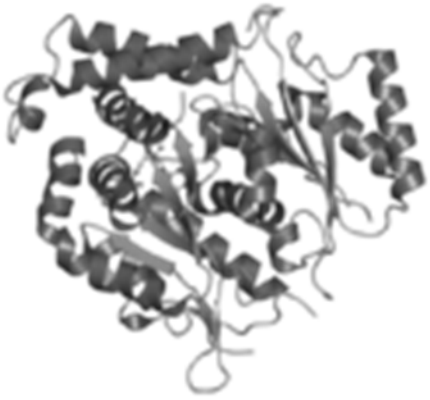
Beta tubulin contains specific patterns that allow for the specific folding of the protein which add to its function. The three types are alpha helixes, beta sheets and loops. An alpha helix is essentially a spiral of amino acids, while the beta sheet appears as a single plane that has been bent multiple times. Loops are the space that lies between the helices and sheets



Beta Sheet

Alpha Helix

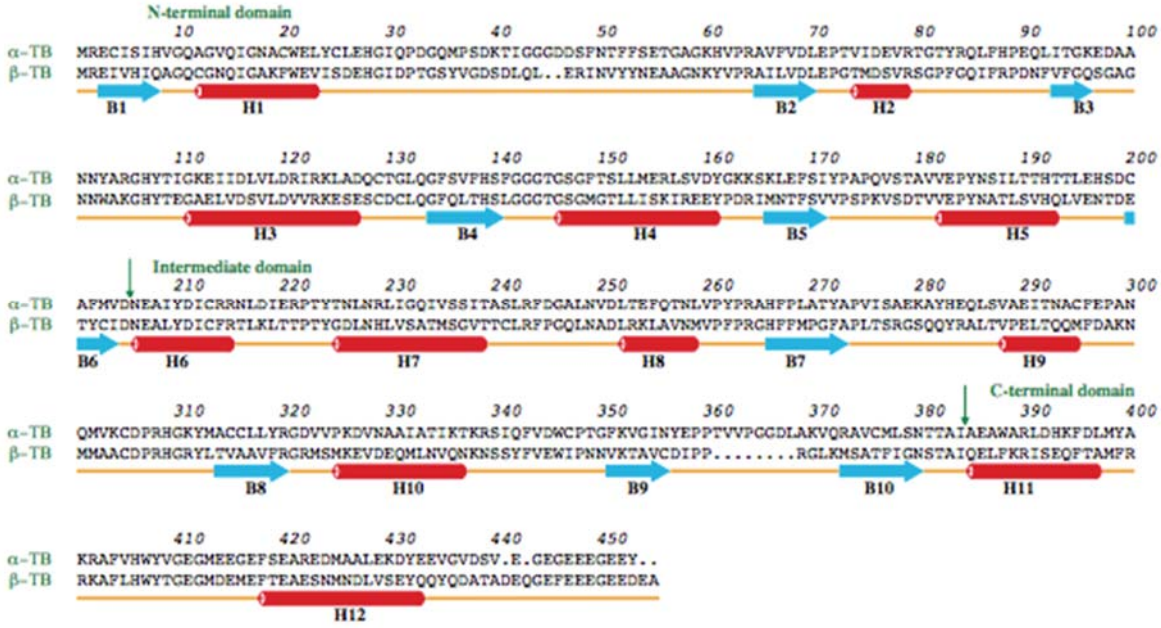
Similar distinct components can be found in music. Often times, songs consist of certain motives that are played throughout pieces. These motives add emphasis to the overall structure, just as alpha helixes and beta sheets do. Examples of such are specific chord progressions that are repeated. Other examples are melody lines that contain the same structure but are repeated throughout. These points of repeating are often crucial in a piece and make the song sound more holistic and complete. Similarly, the loops in beta tubulin allow for the holistic formation of the protein. The following figure shows the entire beta tubulin protein. Note the several beta sheets and alpha helixes in the structure.



6

Another division found in Beta Tubulin are what are called the domains. These are combinations of loops, helices and sheets that form a subunit of the entire protein. Beta tubulin consist of three of these: the N-terminal domain, intermediate domain and C-terminal domain. Each of them have a distinct function and are necessary for the overall function of beta tubulin. The commonality in music to these are key or structure changes in music. There are times in songs where the notes may shift to a different level of pitches. Though changes in harmony and structure causes contrast in the piece (such as in a modulation) , the different sections still are essential for the entire piece, just as the three domains are for the protein.

Below is a chart of beta tubulin containing all the elements just discussed. The red, circular portions are the alpha helixes, blue arrows are beta sheets, and the letters indicate the specific amino acids. The divisions of the domains are noted by the green arrows.



6

It is almost remarkable in the amount of ways DNA and music are related. By taking their similarities, a system of coding can be made in order to translate the amino acids, helices, sheets and other characteristics into music. In the study, this system will be used to code the different types of tubulin and test how their musical elements differ and relate.

From DNA to Music

The start of the project came from a very simplified way of coding. A computer programing system called XOJO was used to design a program that analyzed a DNA sequence by individual nucleotides. A sequence could be entered in and the program plays a note for every nucleotide. A timer was set up so that nucleotides would be read from left to right within a half second of one another. The following chart indicates the translation.

Amino Acid	Note Coding
Adenine	C
Thymine	E
Guanine	G
Cytosine	A

The reason for selecting the notes was that it is that all of the notes combine to produce a diatonic chord, which is called C major seventh. These notes can be played simultaneously or consecutively and produce a likeable sound. Similarly, the nucleotides work in a manor that reads them with harmony. The purpose was to determine whether the order of nucleotides translated to a musically enjoyable segment at a very low level of coding.

Next, a new program was created that would captivate more elements of tubulin into music. Using the polarity data, hydropathy index and the seven-tone scale pattern, a system was devised to code the amino acids into music. Nonpolar amino acids were given translations that consisted of the first (I), fourth (IV) and fifth (V) scale degree. Polar were given the second (ii), third (iii) and sixth (vi) scale degree indications while the stop codon was given the seventh degree (vii*). The chart below displaces the specifics when the DNA is played in C major.

Nonpolar	Hydropathy Index	Note
P	-1.6	G2
W	-0.9	C3
G	-0.4	F3
A	1.8	G3
M	1.9	C4
C	2.5	F4
F	2.8	G4
L	3.8	C5
V	4.2	F5
I	4.5	G5
Polar +		
R	-4.5	E2
K	-3.9	D3
Polar		
N	-3.5	E3
Q	-3.5	A3
H	-3.2	D4
Y	-1.3	E4
S	-0.8	A4
T	-0.7	E5
Polar -		
D	-3.5	D5
E	-3.5	A5
Stop		B4

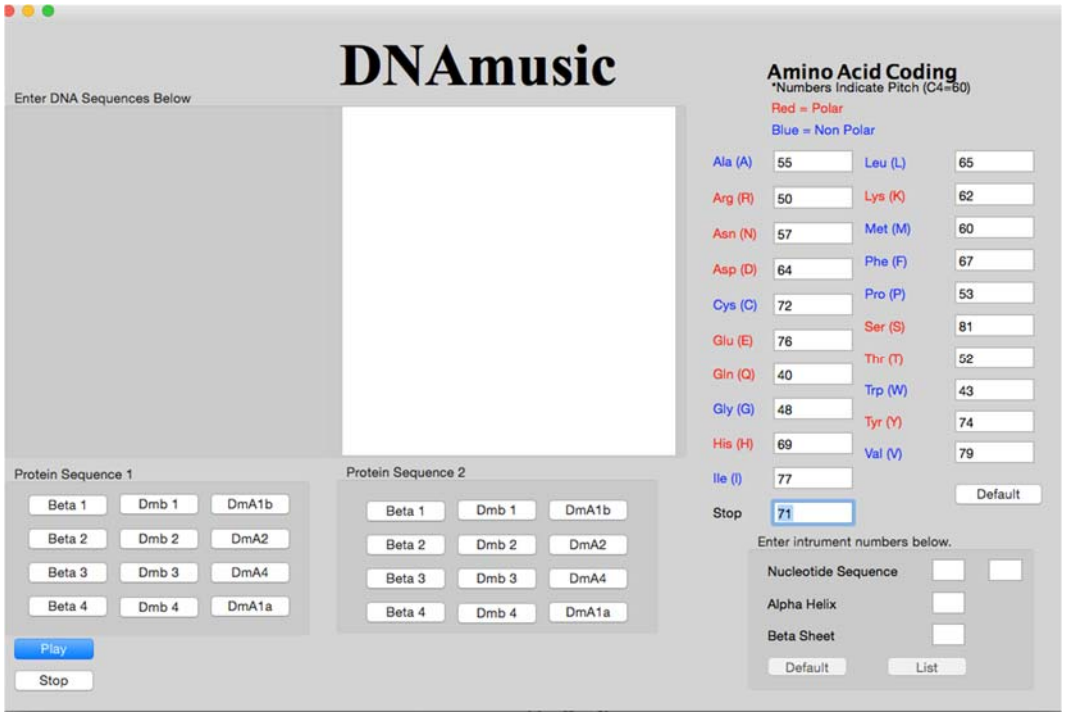
There is nothing relatively “minor” about nonpolar amino acids: The reason for choosing so was that they would be audibly distinguishable from polar. Another reason is that they are contrasting elements in music, just as polar and nonpolar have distinct qualities. Once the scale degrees were divided, the order of the pitches was determined in order of hydropathy index. In

general, the lower the hydropathy index, the lower the pitch. By trial and error, the octaves were chosen in a manor that was most audible.

Helices and sheets were accounted for by repeating chords. It was determined which amino acids the helices and sheets occurred over with the figure on pg. eleven. Over these areas, a V chord would be played for the beta sheet and a I chord would be played for an alpha helix. V and I chords are often what lay down the foundation for a musical piece. They are very distinct sounds, just as helices and sheets are distinctly visible. The timing in between the I and V chord also accounts for the loops, which gives the tubulin a form of rhythm.

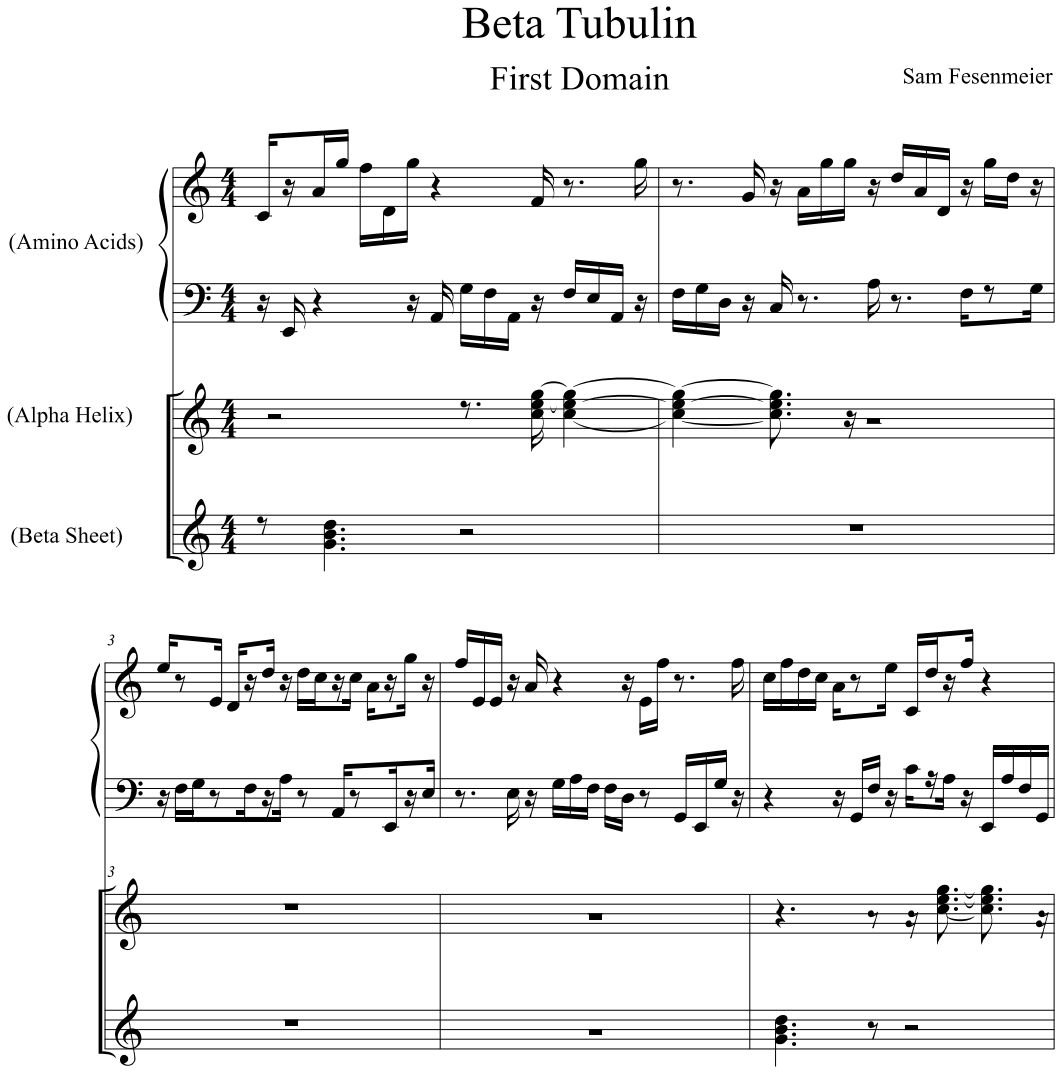
Domains were accounted for by specific key changes, or modulations. The first domain is played in C major. When the second domain is reached, the notes are played eight half steps up, or in A major. The third domain is played three steps below the original, which is A major an octave below the second the domain. Choosing the specific keys was done at random, but having these distinct changes allows the domain changes to clearly be heard.

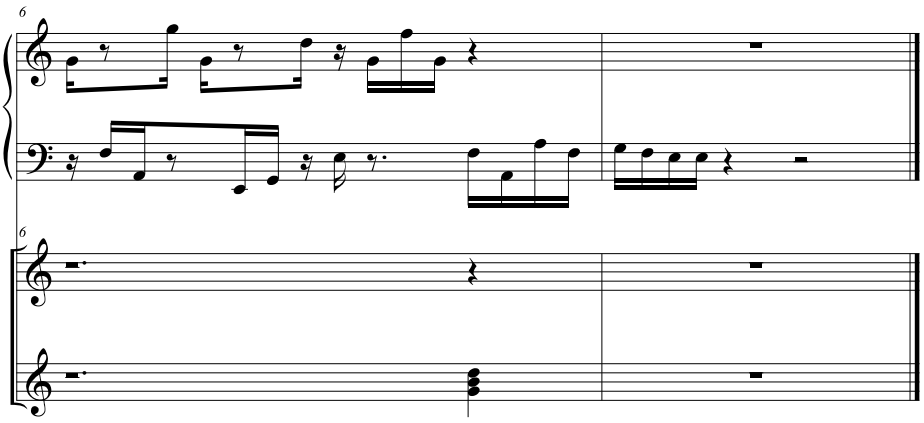
Combining all of these elements created the sound for the tubulin protein. More complexity was accounted for by creating a way to play multiple proteins at once. This created a sound for quaternary structure. Each of different tubulins were played over each other and the overall sound was analyzed. The picture below shows the interface of the final product.



An 10% alanine mutant was created and was compared to the sound of the normal beta 1. A completely random mutant was also created and tested in the same manor. The alanine mutant was created by replacing every tenth amino acid with alanine. Alpha 1 and Beta 2 were also played simultaneously and the sound was analyzed. This reason for this is that these two work together in function. Further comparison was made by playing alpha 4 and beta 2. In the protein, these do not function together.

Results


Beta Tubulin
First Domain
Sam Fesenmeier
(Amino Acids)
(Alpha Helix)
(Beta Sheet)



Discussion

When comparing the alanine mutant and wild type of beta 1, no highly significant difference in sound was shown. Though they were slightly different, it would be very difficult to distinguish the two with a blind test. More specific coding would be needed to be able to distinctly hear the difference between the two. However, the completely random mutant did have distinct differences. The sound appeared to be more repetitive than the actual beta 1.

For alpha and beta tests, it was also difficult to hear for significant differences in sound between the two tests. There appears to be additive musical qualities in each when played over each other. It would be necessary to add more tones to the coding in order to distinguish dissonance between those that are not related in function in biology. By doing so, mutants and other functional qualities could easily be discovered through the musical tests.

Though the results of the two tests are essentially inconclusive, applying more specific coding to DNA could still show some promises. From the system used, musicality was brought about from the DNA regardless. Something about the order in which genetic molecules are assembled translates to musical qualities through coding. If the coding can become more specific towards normal, wild type connections in DNA, then the power of music can be used more thoroughly to make analysis in biology.

The study took the approach of analyzing the coding region of DNA. These are sections where there is clear structure to functional relationships. The next step will be to apply a system that will have the ability to play repeat and regulation sequences where the knowledge of relationships between structure and function are limited. Music could play an essential role in their analysis and bring about patterns not clearly visible through statistical analysis or specific laboratory studies.

Moving Forward

It is without doubt that music enhances a certain power in the brain. Listening to music uses several neural centers, including both the left and right sides of the brain.[7] Our study really only touched on the idea of harnessing the energy of music for use of other knowledge, and I cannot but help feel there is so much music can teach us so much more. I have always been a firm believer in the integration of music in all aspects of life, and I am very fortunate that I have been able to incorporate music into a means of studying biology.

Through my past two years of conducting this research, my mind has been open to many possibilities of how we as humans can learn. Yes, there will always be a need for standardized formats of learning and the current systems we use today. Still, I am very open to the idea using audible systems as a teaching tool. It is clear that we have thousands and thousands of written symbols to analyze complex issues. Why not apply audible cues symbolic of properties to learn as well?

If you have gained anything from this study, I hope it is that you see how learning can occur in a new light. The quest for new knowledge does not have to stop with music either; There could very well be other ways in which we can use creativity to take completely different approaches than the ones most often used now. As Einstein put it, “The important thing is not to stop questioning. Curiosity has its own reason for existing.” [8]

References

1.Retrieved from http://rosalind.info/media/problems/rna/RNA-DNA.png

2.Retrieved from http://upload.wikimedia.org/wikipedia/commons/thumb/6/65/Frequency_vs_name.svg/350px-Frequency_vs_name.svg.png

3.Retrieved from http://www.tritechresearch.com/shop//images/aachart.gif

4.Retrieved from http://www.prism.gatech.edu/~gh19/b1510/transla.jpg

5.Retrieved from http://cnx.org/resources/46f455d8daa0bd084c5c8db4e4061c0e/octavewaves.png

6.Nogales, E., Wolf, S., & Downing, K. (1998). Structure of the Alpha Beta Tubulin Dimer by electron crystallography. Letters to Nature, 391(1), 199-203.

7. Levitin, D. (2012). What Does It Mean to Be Musical? Neuron, 633-637.

8. Retrieved from http://www.brainyquote.com/quotes/authors/a/albert_einstein_2.html