Expert Review of Proteomics

ISSN: 1478-9450 (Print) 1744-8387 (Online) Journal homepage: https://www.tandfonline.com/loi/ieru20

Sounds interesting: can sonification help us design new proteins?

Sebastian L. Franjou, Mario Milazzo, Chi-Hua Yu & Markus J. Buehler

To cite this article: Sebastian L. Franjou, Mario Milazzo, Chi-Hua Yu & Markus J. Buehler (2019): Sounds interesting: can sonification help us design new proteins?, Expert Review of Proteomics, DOI: 10.1080/14789450.2019.1697236

To link to this article: https://doi.org/10.1080/14789450.2019.1697236

Accepted author version posted online: 22 Nov 2019.

Publisher: Taylor & Francis & Informa UK Limited, trading as Taylor & Francis Group

Journal: *Expert Review of Proteomics*

DOI: 10.1080/14789450.2019.1697236

Special Report

Sounds interesting: can sonification help us design new proteins?

Sebastian L. Franjou [1,2,3,§], Mario Milazzo [1,4,§], Chi-Hua Yu [1,5,§], and Markus J. Buehler [1]
[§]These authors contributed equally.

[1] Laboratory for Atomistic and Molecular Mechanics (LAMM), Massachusetts Institute of Technology, Cambridge, MA, United Stated of America

[2] Music and Theater Arts, Massachusetts Institute of Technology, Cambridge, MA, United Stated of America

[3] Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, United Stated of America

[4] The BioRobotics Institute, Scuola Superiore Sant'Anna, Pisa, Italy

[5] Department of Engineering Science, National Cheng Kung University, Tainan, Taiwan

Correspondence:
Markus J. Buehler
Laboratory for Atomistic and Molecular Mechanics (LAMM), Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, Massachusetts 02139 United States of America
Email: mbuehler@mit.edu

Abstract

Introduction: The practice of turning scientific data into music, a practice known as sonification, is a growing field. Driven by analogies between the hierarchical structures of proteins and many forms of music, multiple attempts of mapping proteins to music have been made. Previous works have either worked at a low level, mapping amino acid to notes, or at a higher level, using the overall structure as a basis for composition.

Areas covered: We report a comprehensive mapping strategy that encompasses the encoding of the geometry of proteins, in addition to the amino acid sequence and secondary structure information. This leads to a piece of music that is both more complete and closely linked to the original protein. By using this mapping, we can invert the process and map music to proteins, retrieving not only the amino acid sequence but also the secondary structure and folding from musical data.

Expert opinion: We can train a machine learning model on "protein music" to generate new music that can be translated to new proteins. By selecting a proper dataset and conditioning parameters on the generative model, we could tune the *de novo* protein with high level parameters to achieve certain protein design features.

Keywords: Biomateriomics; design; music; proteins; amino acids; sonification; sound; materiomusic

Article highlights

1. Artist and scientists have been collaborating to make music out of scientific data in a process known as sonification. The numerous analogies between proteins and music have made protein sequences a popular source material for such a work.

2. Proteins and Western music both share a similar hierarchical structure, in which a small number of elementary blocks (e.g., notes for music, amino acids for proteins) assemble in various ways to make a complex functional system.

3. Mapping amino acids to notes leads to very simple music pieces. A more complex mapping, which considers the secondary structure as well as local folding information of proteins, enables the generation of more articulated music. A set of rules enforces a musical order in the generated music.

4. The amino acids - notes mapping is a unified/reversible mapping: this opens up to new approaches for creating new proteins in, so far, unexplored ways.

5. More holistic mappings are possible, such as using the transposed frequencies of the normal vibrational modes of a protein to synthesize new sounds or provide a scale as a basis for a microtonal composition.

6. By using a neural network trained on music mapped from proteins, we can generate music that will be mapped to new types of proteins.

1. Introduction

Proteins are the elementary building blocks of life, and present one of the most challenging areas of biological sciences when it comes to predicting how the sequence and its chemical makeup relate to their functional properties [1]. There is a hierarchy innate in the makeup of protein structures: primary, composed of the amino acids sequence; secondary, arranged as α-helix and β-sheets; tertiary, named polypeptide chain, and quaternary, organized as an assembly of polypeptide chains. Amino acids, composing the primary structure of proteins, are classified in 20 types in eukaryotes with backbones composed of an organized sequence of nitrogen, oxygen and carbon atoms [2].

Just like proteins, many other human-made systems like language, art or music are hierarchical systems that rely on the use of universal building blocks like characters, or harmonic waves, to construct complex functional systems (**Figure 1**) [3]. Likewise, the 20 amino acids are combined to make a large number of proteins, the 12 notes of the chromatic scale are combined in different ways to make up most of Western music. In fact, earlier work has shown that we can translate, reversibly, proteins to other hierarchical manifestations such as music, and thereby conduct design in other domains, as illustrated in **Figure 1B**. Similar to the way sounds are sometimes designed in the frequency domain and then synthesized back to the time domain using the Fourier transform, a reversible mapping allows to design proteins in the "music domain", and map them back to the "protein domain" afterwards. For instance, by translating a protein sequence to music, it may be possible to design new music and then translate that new music back into proteins. This avenue of translational hierarchical design will be discussed in this brief paper.

The question of what is sound, and where does it originate from, also brings us to a more fundamental discussion of musical instruments and in that context, the construction of complex hierarchical systems. In fact, the design of musical instruments has traditionally relied on exploiting properties of properly shaped materials. From carved bone and wood, to the more recently employed electronical means of sound generation, composers, musicians and researchers have looked for innovative solutions to create audible acoustic vibrations [4,5]. After having deeply exploited the properties of traditional materials such as wood or metals at a macro scale, eventually mimicked by electronic synthesizers [6], instrument makers and researchers have been lately exploring the surrounding world to create a new generation of instruments based on the transposition of materials' hierarchical structures, up to the micro/nano scale, into music. This approach, known as sonification, consists of turning inaudible data into sounds, and it often requires striking a balance between faithfully representing the data and imposing musical constraints on the generated sounds. Often, there is a desire to ensure the final result is to some degree harmonious according to certain cultural concepts. Notable recent applications of sonification in the protein biomaterials area concern spider webs [7] and proteins [2,8]. Recent work has also explored the possibility of a reverse mapping between data and sound, using music to design physical objects such as proteins [8]. If a one-to-one mapping can be found to generate music from data, then reversing the operation allows for the exploration of new design avenues.

2. The hierarchical vibrational nature of proteins

Due to the weak interatomic force of the atoms, which are connected strongly by their backbone, proteins (like other molecules) naturally vibrate, and are excited by external forces much more easily than more rigid materials such as graphene or diamond [9]. In fact, proteins lend themselves particularly well to sonification, since both their physical properties and structure can be associated with sound features and music theory [8]. A recent study by Qin *et al.* investigated the vibrational spectrum of a large number of proteins in the Protein Data Bank through Normal Mode Analysis [2]. By shifting the vibrational modes into audible range using the concept of transpositional equivalence, it is possible to preserve the ratios between the frequencies and, thus, define the unique timbre of a protein-base instrument [2].

Additional higher-level, and more 'functional' analogies can be drawn between the structures of proteins and music. They are both made of a small number of different discrete units: amino acids for proteins, notes or chords for a piece of music, organized in space for proteins, and in time for music (**Figure 2A**). In this analogy, secondary structure, the local organization of amino acids in space, is mapped to rhythm, the local organization of notes in time (**Figure 2B**), allowing to "listen to a protein" the way one listens to a song [8]. Although this approach has been tried in different research works [10], reversible mapping and hierarchical transformations have rarely been explored.

A musical piece obtained by directly mapping amino acids to either chords or notes will likely lack either the conventional concepts of melody or harmony [8]. This problem can be addressed by taking the analogy further. Such a direct mapping represents what a traveler would hear while walking along the main chain of the protein, with each amino acid producing a different note. However, within this perspective, the traveler would also hear the sound of other parts of the chain, as the chain folds back on itself and ventures close to where the traveler is. Incorporating this element in our mapping produces several simultaneously sounding parts, completing the original accompaniment with a melody, and producing a more complete piece of music. The ways by which the protein geometry is expressed as sound provides one possible avenue of coding its geometry, especially the relationship between amino acid sequence and structure and folding. By using this encoding approach, we could design and fold proteins by composing music, opening up a new possible method to design new proteins. One possible avenue, using recurrent neural networks trained on protein music, was explored in [8].

3. Conclusions

As discussed in this article, multiple attempts at protein sonification have been made in earlier work, usually focusing on the amino acid sequence. By introducing a reversible mapping between proteins and music that takes into account the overall geometry of the protein, in addition to primary and secondary structure, we propose a new mapping that is potentially more musically rich closer to the source protein. This new strategy also allows for reverse mapping music to proteins, which opens new possibilities for protein design. This novel sonification approach that connects science and music has the potential to be reciprocal, since protein design can now draw from musical composition, as music has drawn from scientific data through traditional sonification.

4. Expert Opinion

Sonification is often used as a way to make new sounds or musical compositions based on scientific data. By introducing a reverse mapping, we can do the opposite, and use music as a design tool for making proteins. Any piece of music following the conventions of our "protein music" contains the sequence of a hypothetical protein, with its secondary structure and folding – the latter encoded in the form of a contact map of geometrically close amino acids by multiple musical voices. Empowered by the emergent discipline of machine learning (ML), one can adopt a ML model to study the embedded patterns in this sonified protein data and use them to generate new proteins. Moreover, we could design de novo proteins by using tools previously reserved to music, such as interpolating between two proteins. The representation of proteins as music encodes the amino acid sequence as well as secondary structure and folding, and allows us to conveniently embed all of this information into a type of data that can be well processed by a neural network. This makes this representation especially well-suited to machine-learning-powered applications: for example, by training a neural network on specific types of protein, we could design new proteins that share characteristics with the proteins of the dataset. Some of these concepts have been explored in [8], where a ML approach was used to compose new music from musical data derived from proteins that was translated back into proteins (for audio examples, please see [11]).

The results achieved in this field so far open to a number of question and avenues of research, tightening the connection between materials design and music. In [2], the authors interpreted the vibrational modes of a protein as peaks in the spectrum of a single note, in order to create new timbres, and offering new compositional avenues. However, in an alternative interpretation each frequency could instead be read as a different note, turning a protein into a *microtonal scale*. Most of Western music is based on the major and minor scales in equal temperament, meaning we use a small number of similarly organized subsets of the 88 piano keys [12]. However, *microtonal music* uses notes that are not playable on a piano, allowing the use of an essentially infinite number of possible scales, each one with a different sound and hence the potential to make unique sounding music. Another way to listen to a protein, as well as a way to write novel music, consists of using a scale based on the frequencies of the normal modes of such a protein as a basis for a microtonal composition. This method is quite different from the one previously mentioned, as it does not map a protein to a single, well-defined piece of music. Rather, it allows the protein to determine a certain set of parameters (viz., the microtonal scale) that a composer or improviser could employ to create a piece. Such a process, although different from most sonification attempts, enables greater human interaction, and perhaps a more musically meaningful end product. Composers have taken inspiration from natural processes such as birdsong [13] and stochastic processes [14], and the degrees of freedom ensured by such a mapping would allow composers to base their work on natural processes and materials that do not originally have any audible characteristics. Furthermore, such mappings could still be reversible, and possibly allow to reverse map a greater variety of music into proteins. For instance, it would be possible to outline a new design theory for smart materials based on music hierarchies able to optimize specific properties such as toughness or strength.

Moreover, this tight connection between music and proteins naturally leads us to wonder if it would be possible to translate our favorite musical compositions, or different genres, into proteins. Although the current strategy requires the music to follow strict rules that prevent almost any known composition to be translated, an even higher-level analogy between music and protein could allow such mappings. Both proteins can be represented as a formal grammar [15,16]. Formal grammars are an analytical tool from linguistics designed to represent hierarchical structures. A sentence, for example, may contain a noun phrase, which may itself contain a noun and a preposition phrase, and so on. Both music and proteins share these attributes with language. The hierarchical organization of proteins into secondary, tertiary and quaternary structure echoes the way the chords in tonal music can be classified as having a local harmonic function, while being part of a more global functional region [16]. By abstracting music to the level of harmonic function, *i.e.*, the perceived tension and direction of the chords, we can parse a large number of tonal pieces from the common practice period composers such as Bach, Mozart, or Beethoven. Moreover, by using a well-designed set of rules to generate the music on a functional level (such as the ones outlined in [16]), we can ensure that the music generated from a protein will follow certain musical conventions that should lead to a harmonious result. These concepts can offer innovative approaches to design proteins or to evolve proteins from existing ones, and connecting disparate scientific and artistic disciplines to realize a new route for bio-inspiration.

Funding

This work was supported by the Massachusetts Institute of Technology Undergraduate Research Opportunities Program and the Massachusetts Institute of Technology Center for Art, Science & Technology (CAST), supported by the Andrew W. Mellon Foundation. Additional support from the US Office for Naval Research N00014-19-1-2375,US Air Force Office of Scientific Research FA9550-15-1-0514, and NIH U01 EB014976 is acknowledged.

Declaration of interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Reviewer disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

References

[1] Mathews CK, Van Holde KE (Kensal E, Ahern KG. Biochemistry. Benjamin Cummings; 2000.

[2] Qin Z, Buehler MJ. Analysis of the vibrational and sound spectrum of over 100,000 protein structures and application in sonification. Extreme Mechanics Letters [Internet]. 2019;29:100460. Available from: https://doi.org/10.1016/j.eml.2019.100460.

[3] Buehler MJ. Materials by design - A perspective from atoms to structures. MRS Bulletin. 2013;38.

[4] Wegst UGK. Bamboo and Wood in Musical Instruments. Annual Review of Materials Research [Internet]. 2008;38:323–349. Available from: http://www.annualreviews.org/doi/10.1146/annurev.matsci.38.060407.132459.

[5] Osaki S. Spider silk violin strings with a unique packing structure generate a soft and profound timbre. Physical Review Letters. 2012;108.

[6] Russ M. Sound Synthesis and Sampling. Sound Synthesis and Sampling. 2018.

[7] Su I, Qin Z, Saraceno T, et al. Imaging and analysis of a threedimensional spider web architecture. Journal of the Royal Society Interface. 2018;15.

[8] Yu C-H, Qin Z, Martin-Martinez FJ, et al. A Self-Consistent Sonification Method to Translate Amino Acid Sequences into Musical Compositions and Application in Protein Design Using Artificial Intelligence. ACS Nano. 2019;in revisio:7–9.

[9] Cranford SW, Buehler MJ. Biomateriomics. Proceedings of the National Academy of Sciences. 2012.

[10] Takahashi R, Miller JH. Conversion of amino-acid sequence in proteins to classical music: Search for auditory patterns. Genome Biology. 2007;8.

[11] Buehler MJ. SoundCloud repository [Internet]. Available from: https://soundcloud.com/user-275864738.

[12] Hyer B. Tonality. Oxford University Press; 2001.

[13] Griffiths P. Messiaen, Olivier. Grove Music Online. 2001;

[14] Hoffmann P. Xenakis, Iannis. Komponisten Lexikon. Springer; 2003. p. 699–701.

[15] Gimona M. Protein linguistics - a grammar for modular protein assembly - Gimona - 2007. 2006;7:411–414.

[16] Rohrmeier M. Towards a generative syntax of tonal harmony. Journal of Mathematics and Music. 2011;

[17] Buehler MJ. Tu(r)ning weakness to strength. Nano Today. 2010;5:379–383.

[18] Bralley P. An Introduction to Molecular Linguistics. BioScience. 2006;46:146–153.

Figures and captions

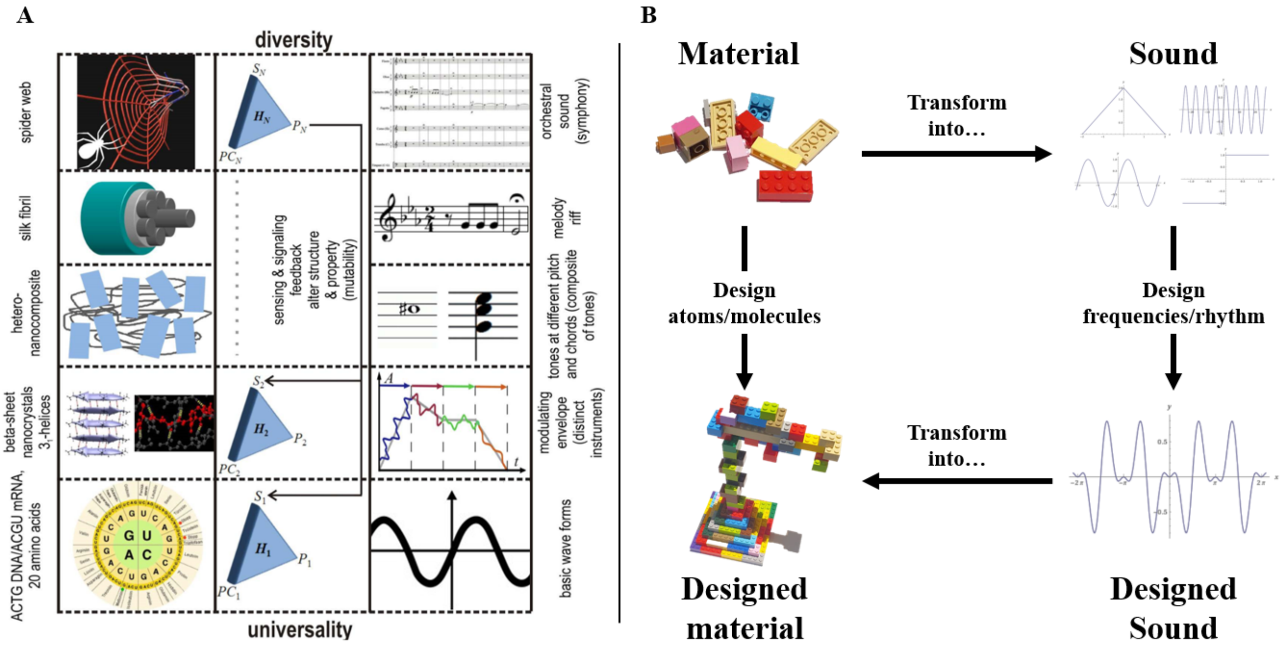

Figure 1. Panel A: Analogy of the hierarchical structure of proteins and music. Reproduced with permission from [17] © Elsevier. Panel B: Overview of the translational hierarchical design approach, where protein design is realized by translating the material representation of amino acid building blocks into sound, which serves as the basis for new design, and reverse translation.

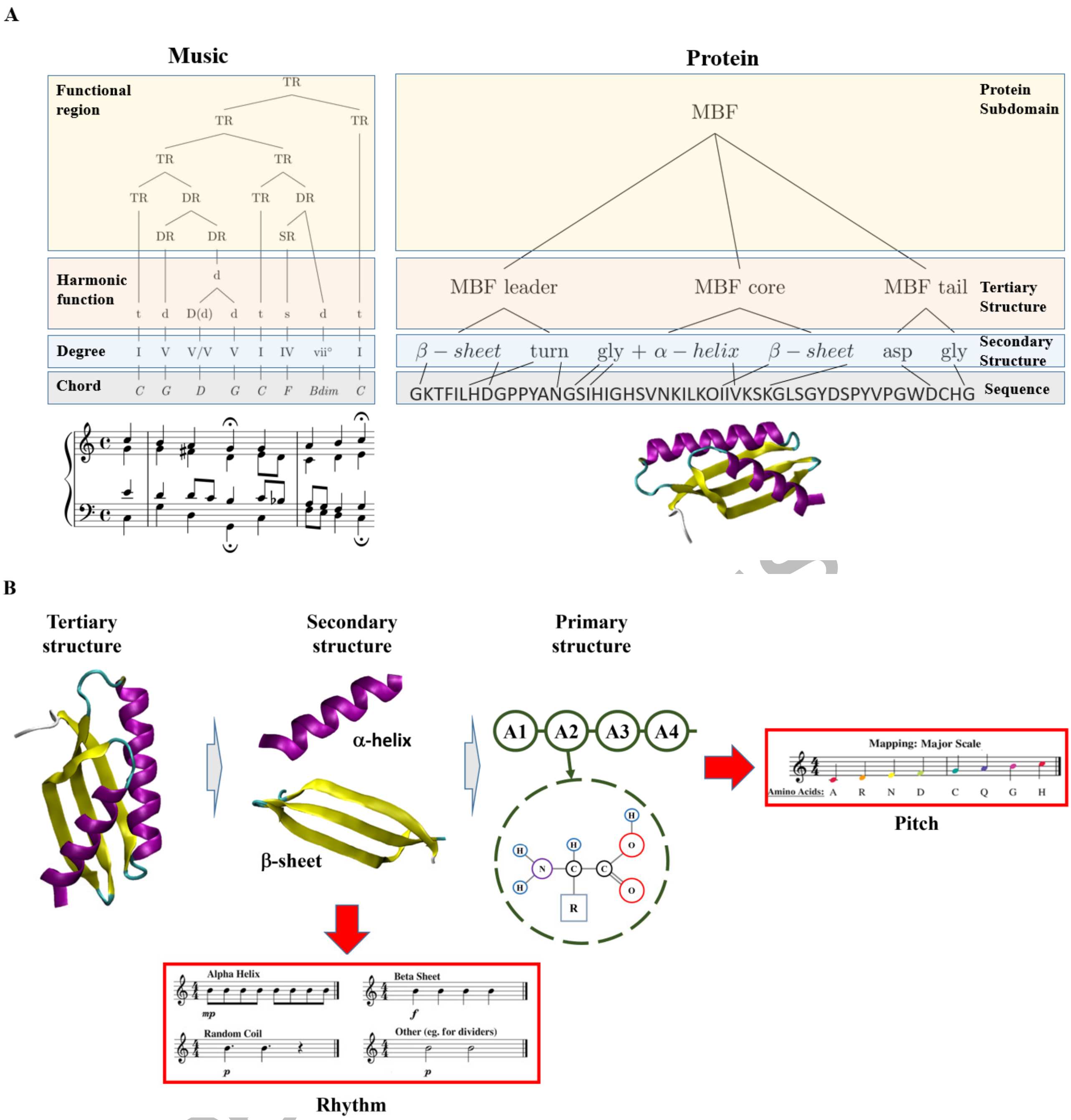

Figure 2. Proteins and music, focused on hierarchical organization and functional relationship between unitary building blocks. Panel **A**. As proposed by Rohrmeier [16], chords in tonal music can be analyzed by their relation to each other and to the key (Degree). These degrees have different levels of tension and musical direction (Harmonic Function), and can be grouped as sections exhibiting an overall function (Functional Region). Similarly, proteins can be analyzed in term of their structural properties (secondary and tertiary structure), which then give rise to functional properties (protein domains) [18]. Panel **B**. Possible sonification concept based on primary (Pitch) and secondary (Rhythm) structures, representing the folded geometry of a protein in music.