

Improvising Life

Questions about
Conceptual Design Constraints;
Research Paradigms

DRAFT – FOR DISCUSSION ONLY

Purpose of this Report

Identify some of the basic technical design issues, and see whether and how others have solved them, to:

- Establish the scope of the project and its phases, allowing prioritizing and efficient work.
- Gain a deeper understanding of the challenges and an optimum language for discussing them.
- Point to and prioritize potential collaborators and technologies.
- Not re-invent the wheel.
- Determine the feasibility of the project and components, so that we can give potential funding sources and collaborators confidence in the viability of the project.

Improvising Life



Inaugural & Primary Goal

To translate live, pre-set or improvised jazz performance into controlled video animation representations of DNA-related processes.

Overall Project Goals

- Translation. To develop a *system* (**input**→**black box**→**output**) through which live, and preset/improvised music can be translated into visually-represented molecular components and processes, from chemistry to DNA to protein function.
- **Aesthetics**. Both live performance—after practice and ear-training—and *desired* results, i.e. viable structures and functionality, should be aesthetically interesting and pleasing.
- Teaching. To excite people about and teach genetics, as well as music and improvisation, by giving them a sorcerer's apprentice-like ability to create molecules, and experience pleasurable feedback in success.
- Research. To create new, interesting proteins using aesthetics.

Two-Way Translation

- The initial goal focusses on a one-way path from performance to visual representation.
- Although this report uses the terms *input* and *output*, the bridge must be built from both sides. The black box can be filled in and correlations calibrated starting from either music or from models. The black box should function in either direction, e.g. correlations work both ways.
- The system could be *interactive* such that outputs include aural as well as visual representations
 - Recorded, in vitro, or computer-modeled components/process can be played aloud for the first time, or played back, to be able to analyze what worked.
 - Players could respond to the model's music in real-time
 - Playback could permit human-model improvisational collaborations.

Design Constraints

1. Translation
2. Aesthetics



Translation Constraints

- Input (aural)
- Black box
- Output (visual)

Input: Aural Performance

- One design challenge is to make a system that, after learning, ear-training, and practice, can be used easily, intuitively, and creatively.
- What are the performance rules?
- What kinds and velocity of bio-representational feedback to performers is desired? Will it be all visual?
- The project has already begun working with jazz performers using rules and correlations
 - Goals: Proof of concept, attracting people to the project through interesting and aesthetic results. (Yes?). The performer rules are...; the correlations are....
- Is useful sound capture an issue?

Black Box

- What are the correlations and grammars we want establish and which prioritize?
- What are the relevant algorithm paradigms? Will it use neural networks & AI?

Output

What elements and processes do we want to represent?

- Conceptual
 - Focus on big ideas vs. technical exactitude for teaching biology and creating exciting multi-media performances.
 - “[P]rotein music isn’t only for human ears. It can be used as AI input to analyse features and patterns of proteins, or even as [a way to create new proteins](#). But to be fair, the AI won’t care if music sounds good or not.” (Forbes article)
- One-to-one-correspondences, technical exactitude
- Aesthetically interesting
- Content
 - Well-understood biological processes:
 - Chemistry→DNA;
 - DNA transcription→RNA;
 - RNA translation→amino acid strings (AAS) and their side chains
 - Cutting-edge biological processes:
 - Amino acid (& sides) string→folding
 - Folded protein structure/chemistry→function

Aesthetic Constraints

- Golden Means. A challenge is to be able to musically represent components, structures, and processes such that successfully built components sound good to the performer. One goal would be to be able to use ear training to be able to create, recognize, and distinguish between functional elements and proteins.
- Alternate Constraint: Forgo conventional aesthetics.



Research Paradigms: Addressing Design Constraints

1. Protein Music, Tay, Chen et al., Singapore, 2021
2. Yu, Buehler et al., MIT, 2019
3. Additional practitioners with sound samples

Protein music

Tay, Chen, et al./Singapore

Protein music of enhanced musicality by music style guided exploration of diverse amino acid properties, Heliyon, 2021

Nicole WanNi Tay, Fanxi Liu, Chaoxin Wang, Hui Zhang, Peng Zhang, Yu Zong Chen

[Protein Music paper link](#)

In this project, the **input** is biological data including diverse amino acid properties and protein sequences contain recurring secondary structures and tertiary folds.

The **black box** processes the data *instantaneously* using correlations and features that add **musicality, following a specific musical style.**

The **output** is a musical representation of the data.

Introduction

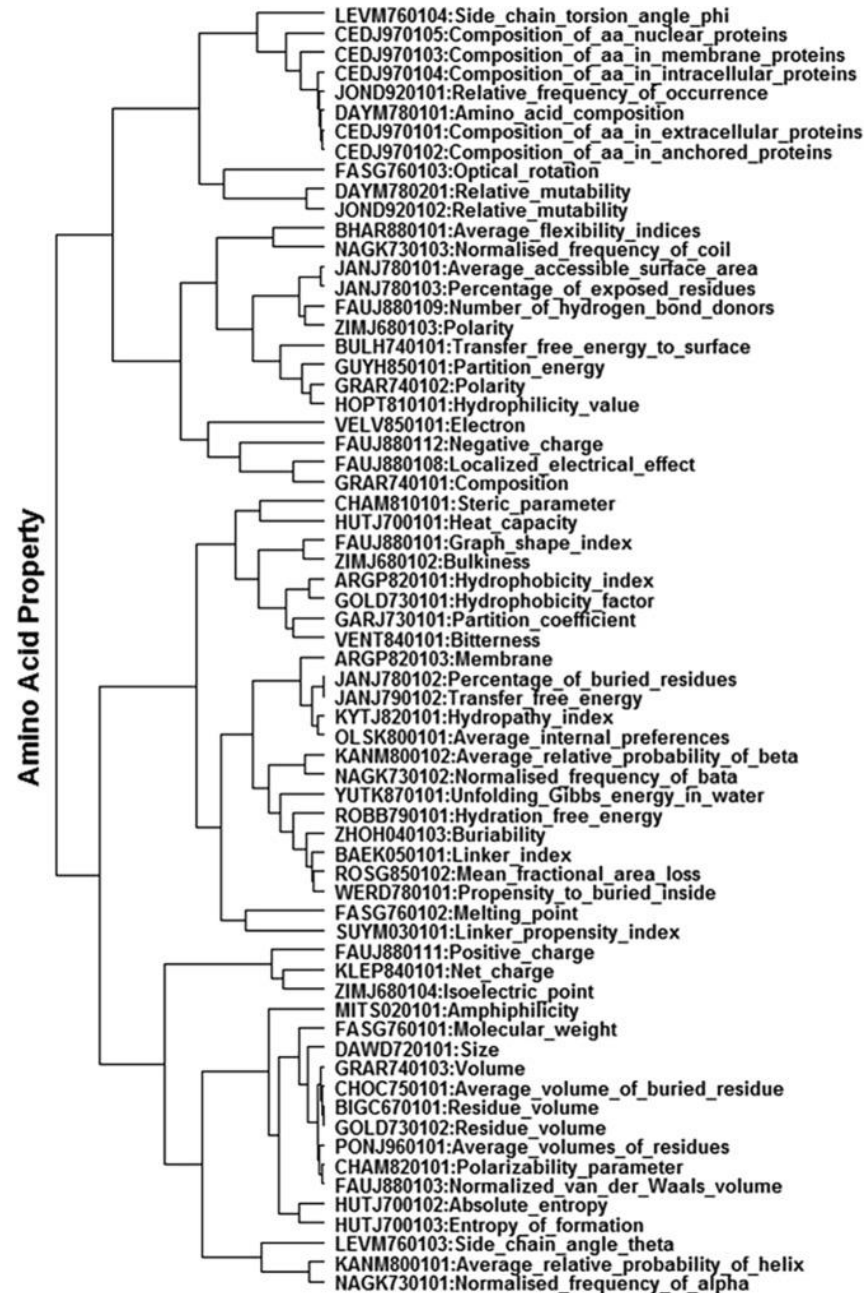
Despite the continuous development of protein-to-music algorithms, the musicality of protein music lags far behind human music.

Musicality may be enhanced by fine-tuned protein-to-music mapping guided by the features of specific music style. This mapping may be based on biologically-relevant amino acid properties, sequence patterns and variations, so as to robustly match but not overfit the musical features while maintaining biological implications.

We analyzed the features of a music style...and used the quantized musical features to guide broad exploration of diverse amino acid properties (104 properties, sequence patterns and variations) for developing a novel protein-to-music algorithm of enhanced musicality

Input

Example. Hierarchical clustering of amino acid properties.



Input, cont.

Based on the comparison of the quantitative profiles of the amino acid properties and the quantized *Fantasy-Impromptu* musical features, the following 11 amino acid properties were selected from the pool of 104 properties for generating a novel protein-to-music algorithm. These properties are the relative frequency of occurrence of the amino acid in proteins, amino acid composition, electron-ion interaction potential, net charge, bitterness, hydropathy, size, residue volume, rRNA binding propensity, mRNA binding propensity, and number of hydrogen bond donors. Based on these properties, a set of rules were devised for mapping the amino acids along a protein sequence to the musical features that lean towards the *Fantasy-Impromptu* style, which are described below.

Black Box 1A. – Quantization of Music

This is the secret sauce of the project. I give a flavor of it here.

Musical Style

Here we focused on the *Fantasy-Improptu* style, which is a style of solo piano compositions from the Romantic period [\[24\]](#). Music of this style is generally characterized by chromaticism, chords, a wide keyboard range and elicits strong emotions

An analysis of *Fantaisie-Improptu* [\[26\]](#) has revealed that it has a balanced structure of three sections. The first section presents a very strong characterized personality and forms a very important motif of the work on the left hand, where the whole left-hand part is constructed mainly on this motif structure. The second section displays a freer structure in D-flat major tonality, and the melody on the right hand has a slow and lyrical character. The third section begins exactly the same as the first section up to the hundred and eighteenth bar, followed by a C-sharp pedal ongoing for 9 bars.

Black Box 1B. Quantization of Music

Individual work quantization.

Four works of music "were quantized to extract the musical features of the *Fantasy-Impromptu* style by the following procedure: abc2midi program [\[31\]](#) was used for converting the MIDI files into the corresponding text files in a music notation known as ABC notation [\[32\]](#).

Six musical features, pitch, length, octaves, chords, dynamics and main theme, were quantitatively analyzed.

In counting the percentage frequency of the 12 pitches in an octave, all pieces were transposed to the base key of *C major* and octaves were ignored in counting the percentage occurrence of each pitch. In counting the percentage occurrence of each note length, the tempo and number of beats of music per bar was disregarded and all note lengths were expressed as a multiple or fraction of the length of a crotchet. In counting the percentage of the notes that occur per octave, the middle C was taken as the fourth C from the left of the keyboard, the notes before the first C key are in the A0 octave, while the notes before the second C key are in the C1 octave, until the last note of the piano at C8.

In counting the percentage occurrence of chords, the total number of chords in a piece was counted and the percentage frequency of occurrence of chords per piece was computed by dividing the number of chords against the total number of notes.

In quantizing the dynamics, i.e. the loudness of notes, the frequency of changes in loudness was counted in two groups, a gradual change (crescendo or diminuendo) and a sudden change.

Lastly, quantization was conducted on the first recurring melodic theme identified in each piece, to determine its characteristics for replicating such themes in the protein music."

Snippet from elsewhere: "*In mapping to the note length, there is a need to avoid a direct correlation with the changes of the pitches* [emphasis added]. Therefore, the mappings of the current amino acid to the note lengths were based on the corresponding properties of the frontward amino acid (i.e. the note length of the current amino acid was determined by the property of the frontward amino acid)

There are several more interesting paragraphs about this in the article.

Black Box 2. Correlations (example)

Table 2 The matching of the 20 amino acids to the pitch and note length in accordance with the *Fantasy-Impromptu* style.

Description	Matching relationship for the note pitch and length of the current amino acid						
Pitch of the right-hand	C	D	E	F	G	A	B
Matched amino acid based on relative frequency of occurrence of amino acid in proteins	T G D Y	L N	K Q	P R C	A I M E	V F H	S W
Pitch of the left-hand	C	D	E	F	G	A	B
Matched amino acid based on composition of amino acid in proteins	I F G P	A C H	Y K	V D	E Q N L	R M T	S W
Note length of the right-hand	2	3/2	1	1/2	1/4		
Matched amino acid before the current amino acid based on the size	G	D A	S C	N E T V	R K W Y F Q H M I L P		
Note length of the left-hand	2	3/2	1	1/2	1/4		
Matched amino acid before the current amino acid based on the residue volume	G	A S	C T	N D P V	Q E H I L K M F W Y		

Output: Audio Examples

- N.B. PowerPoint hyperlinks only work when you are in slide show mode (click “play from current slide”)
- Working Example: <https://www.eurekalert.org/multimedia/801148>
- Website 1. <http://bidd.group/html/music/index.html>
 - Includes music generator using protein info, example not currently working.
 - Add'l samples. Need application that can play MIDI files, <http://bidd.group/html/music/samples.html>
- Biology Inspired Art Server. Related work. <http://bidd.group/html/art/index.html>

Protein Music Takeaways

- This project's input is quantized DNA information. The interesting question is whether the input/output could be reversed, so that a piece of musical input could synthesize animations of amino acid sequences, etc. A second question is whether *live* music could be quantized in the black box.
- The project uses a variety of techniques and correlations we could use. They have published their code.
- The authors write that theoretically any musical style can be used.
- The translation is instantaneous.
- Published in September 2021, this may be considered state-of-the art.
- Yu Zong Chen, or someone else with the Bioinformatics & Drug Design group [BIDD] is a research group based in Department of Pharmacy, Faculty of Science, National University of Singapore. Or there may be someone associated with them working locally.

Sonification

Yu, Buehler, et al./MIT

A Self-Consistent Sonification Method to Translate Amino Acid Sequences into Musical Compositions and Application in Protein Design Using Artificial Intelligence - ACS Nano 2019

Chi-Hua Yu; Zhao Qin; Francisco J. Martin-Martinez; and Markus J. Buehler (Buehler: Laboratory for Atomistic and Molecular Mechanics LAMM, MIT)

[Yu, Buehler 2019 paper in ACS Nano](#)
[MIT News article](#)

Yu, Buehler, et al./MIT

From the abstract (article not available to me)

We...translate **amino acid sequences** into audible sound, use the representation in the musical space to train a neural network, and then apply it to generate protein designs using artificial intelligence (AI).

Sonification Method

The sonification method proposed here uses the normal mode **vibrations of the amino acid building blocks of proteins** to compute an audible representation of each of the 20 natural amino acids.... The vibrational frequencies are transposed to the audible spectrum following the musical concept of **transpositional equivalence**, playing or writing music in a way that makes it sound higher or lower in pitch while retaining the relationships between tones or chords played. This transposition method ensures that the relative values of the vibrational frequencies within each amino acid and among different amino acids are retained.

The characteristic frequency spectrum and sound associated with each of the amino acids represents a type of musical scale that consists of **20 tones, the “amino acid scale”**. To create a playable instrument, each tone associated with the amino acids is assigned to a specific key on a piano roll, which allows us to map the sequence of amino acids in proteins into a musical score.

To reflect higher-order structural details of proteins, the volume and duration of the notes associated with each amino acid are defined by the **secondary structure of proteins**, computed using DSSP and thereby introducing musical rhythm. [DSSP refers to method of assigning & database of secondary structures.] Cont...

Yu, Buehler, et al./MIT, abstract continued

Representation & Data Structuring

We then train a recurrent neural network based on a large set of musical scores generated by this sonification method and use AI to generate musical compositions, capturing the innate relationships between amino acid sequence and protein structure.

We then translate the *de novo* musical data generated by AI into protein sequences, thereby obtaining *de novo* **protein designs that feature specific design characteristics**.

Demonstration & Teaching

We illustrate the approach in several examples that reflect the sonification of protein sequences, including multihour audible representations of natural proteins and protein-based musical compositions solely generated by AI.

Research Potential

The approach proposed here may provide an avenue for understanding sequence patterns, variations, and mutations and offers an outreach mechanism to explain the significance of protein sequences. The method may also offer insight into protein folding and understanding the context of the amino acid sequence in defining the secondary and higher-order folded structure of proteins and could hence be used to detect the effects of mutations through sound.

Downloadable demonstration app:

<https://play.google.com/store/apps/details?id=com.synth.aminoacidplayer>

Sonification example

- [Sonified 3D Protein of Snail](#) (sound clip w/explanation)

Markus Buehler

- Buehler says that after listening to the resulting melodies, he is now able to distinguish certain amino acid sequences that correspond to proteins with specific structural functions. “That’s a beta sheet,” he might say, or “that’s an alpha helix.”
- Markus J. Buehler is the McAfee Professor of Engineering at MIT, and a composer of experimental, classical and electronic music, with an interest in sonification. Using an approach termed "materiomusic", his artistic work explores the creation of new forms of musical expression - such as those derived from biological materials and living systems - as a means to better understand the underlying science and mathematics. One of his goals is to use musical and sound design as a novel and abstract way to model, optimize and create new forms of matter from the bottom up, and to assess cross-system design relationships.

Sonification Takeaways

- Again, this project moves from amino acid sequences to music.
- This project uses neural networks and AI.
- Is it obvious why they don't just assign one tone to each amino acid?
- Structure levels seems like a useful concept.
- As a local, Buehler could be an adviser or collaborator (or he may be too advanced for us).

Other interesting links

All with audio samples

- A.M. Clark, Tx Wesleyan, 2001 - <https://whozoo.org/mac/Music/>
- Linda Long, English?, 2001 - <http://www.molecularmusic.com/>
- Peter Gena, et al., '95-'99 protein music & current as artist, <https://www.petergena.com/>

One takeaway from these three practitioners is the value of using musical through-lines to accompany their abstract music, either emergent from the protein, or simply mixed in from a musical source.