

# Listening to Biological Translation: Converting RNA to Music

**Ilana Shapiro**

Computer Science Department  
Pomona College  
Claremont, CA 91711 USA  
issa2018@mymail.pomona.edu

## Abstract

The sonification of genetic material is a little-explored mode of unconventional computation that bridges the divide between bioinformatics, computer science, and music, allowing bioinformaticians to perceptualize their data in a novel and illuminating manner. This paper presents BioMus, an original model for converting RNA to musical data in the form of MIDI piano chords. Genetic material from a variety of species is sourced from Nucleotide, a database of the NCBI's molecular biology database system Entrez. Then, codons inform the harmony in protein-coding regions, while individual nucleotides dictate chords in non-coding regions. Local keys are determined by amino acids in the protein-coding regions and carry over into the non-coding regions. Finally, an analysis of harmonic sequences in the resulting MIDI files is presented, as well as an attempt to classify species into the correct biological class based on this data. By mapping nucleotides and codons to chords and analyzing genetic material as music, BioMus thus gives scientists the means to uniquely conceptualize the process of biological translation.

ONE PARAGRAPH, max 250words

## Introduction

## Related Work

Ingalls et al. present

## Converting RNA to Music

BioMus's process of RNA sonification begins with the user specifying a desired species and gene, as well as the species' biological class. The species and gene are passed to the Biopython library to obtain the associated DNA sequence from Nucleotide, a database providing DNA and RNA sequences that is part of the NCBI's molecular biology database system Entrez. The resulting DNA is written to a text file in FASTA format, a standard textual format in bioinformatics in which nucleotides are represented with single characters corresponding to their bases. The file is placed inside a folder with the species' class name.

Next, a MIDI track is created using the MIDIUtil library with the tempo set to  $\text{♩} = 200$  BPM, and the DNA sequence is transcribed to RNA with Biopython. Recall the bases used in RNA are adenine (A), cytosine (C), guanine (G) and uracil (U), and have the pairings C-G and A-U. BioMus defines

the following mapping of individual bases to musical notes relative to the local key in Table 1

C	Tonic
G,A	Mediant
U	Dominant

Table 1: RNA Base Mappings

The musical key is minor in non-coding regions and major in protein-coding regions, with the initial key set to C minor. Before the start codon AUG is encountered, individual nucleotides outline the 3 notes of minor triad of the local key (base pairs form dyads, or 2-note chords). Then, during translation, the key is major and the volume doubles. The key is determined by the codon (see the AMINOACIDS dictionary above). Individual nucleotides no longer determine the notes; this is now dictated by the codons. Each time the key changes based on the codon, the major triad of that key is played, until a stop codon is reached. Then we remain in the same minor key as the previous stop codon, but the volume is halved again and individual nucleotides once again outline the notes of the new minor triad. Then this process can repeat if another start codon is then encountered'

The tonic, mediant, and dom

Throughout this paper, the chosen gene for MIDI generation and analysis is TP53, which codes for the tumor suppressor protein p53.

AUG	Methionine/ Start Codon	C
AUU, AUC, AUA	Isoleucine	C#
AAA, AAG	Lysine	D
ACU, ACC, ACA, ACG	Threonine	E♭
UUU, UUC	Phenylalanine	E
UGG	Tryptophan	F
UUA, UUG, CUU, CUC, CUA, CUG	Leucine	F#
CAU, CAC	Histidine	G
GUU, GUC, GUA, GUG	Valine	A♭

Table 2: Essential Amino Acids

AAU, AAC	Asparagine	A
GAU, GAC	Aspartate	B
GCU, GCC, GCA, GCG	Alanine	B

Table 3: Nonessential Amino Acids

UAU, UAC	Tyrosine	C
UGU, UGC	Cysteine	D
UCC, UCU, UCA, UCG, AGU, AGC	Serine	E
AGA, AGG, CGU, CGC, CGA, CGG	Arginine	F
CCU, CCC, CCA, CCG	Proline	G
CAA, CAG, GAA, GAG	Glutamine/ Glutamic acid	A
GGU, GGC, GGA, GGG	Glycine	B

Table 4: Conditionally Essential Amino Acids

UAA	C
UAG	E
UGA	G

Table 5: Stop Codons

## Obtaining Genetic Data

### Conversion to MIDI

### Analyzing MIDI for Harmonic Sequences

### Classification and Musicality of Species

### Conclusion

BioMus serves as a bridge between bioinformatics, computer science, and music by giving scientists the creative means to

### Acknowledgments

I am very grateful to Professor Zachary Dodds of Harvey Mudd College for his invaluable mentorship throughout this project.