13C-NMR sonification of human insulin: a method for conversion of amino-acid sequences to music notes

Majid Monajjemi 1,*

1Department of chemical engineering, Central Tehran Branch, Islamic Azad University, Tehran, Iran

*corresponding author e-mail address: Maj.monajjemi@iauctb.ac.ir

ABSTRACT

The concept of molecular sonification comprises total steps of methods that convert the physical data derived from chemical systems into acoustical music. NMR data of the 13C are especially well suited data sources for Insulin sonification. Even though their resonant frequencies are typically in the MHz region, the resonant frequencies span around kHz. The human insulin is consisting of 51 amino acids which can be divided into 7 series of amino acids for seven octaves of notes. During NMR calculation with ab-initio methods, these signals are routinely mixed down into the audible frequencies ranges, rendering the need for any additional frequencies transpositions unnecessary. By this work, insulin protein sequences into musical notes to reveal auditory algorithms have been converted. Calculation and optimization of 20 amino acids have been done and the total frequencies of each amino acid have been converted to 20 music notes and distinguishing those using variations of chemical shifts including pitch, time duration length of notes and even rhythm have been accomplished.

Keywords: Protein, Insulin, artificial intelligence, sonification, classical music notes.

1. INTRODUCTION

Converting molecular properties to sound is important for science information, therefore in the Sonification, sounds are produced from chemical properties in the bio molecular systems with the goal of facilitating data interpretation [1,2]. Sonification methods were used to DNA generating, to amino acids sequences, and to protein folding [3]. In all subjects, the mechanism of conversion has been done based on molecular properties through a suitable algorithm related to invisible phenomena with a visual inspection. Although the biological subjects have been started several billion years ago when life appeared on earth [4], it is possible to encode them into a sequence of chords and melodies. Obviously, the conversion from genomics to music would open a possible to encode them into a sequence of chords and melodies.

1.1. Convert NMR spectroscopy to music.

Molecular specification can be turned through sonification of atomic resonant processes directly into sound, including Infra-red spectroscopy "IR" or nuclear magnetic resonance "NMR". Via IR spectroscopy, it can be measured the vibrational behavior of molecules and it has been applied for its use as a sound source for spectroscopy’s sonification in theoretical musical backgrounds.
One general trait of all these sonification methods is that artistic selection has to be made during the sonification mechanism including musical’s notes, pitches, melodies, Rythms and Chords to be designed for different chemicals images. In this work, $^{13}$C-NMR spectroscopy is applied as a novel source for protein specification towards sonification. In contrast to IR, $^{13}$C-NMR spectroscopy experiments the frequencies of the nuclear signals which can be converted straightly into the audible described the sonification strategies which used NMR data in acoustic music composition [13].

1.2. Insulin specifics.

Insulin is a polypeptide hormone secreted by beta cells of the Pancreas which is one of the major anabolic hormones (Fig.1) Insulin hormones regulate the metabolism of fats, carbohydrate and proteins via increasing the suction of those compounds, especially glucose from the blood into liver and muscle’s cell. In the mentioned tissues the glucose is converted into glycogen, triglyceride or both in the liver, respectively. Circulating insulin also control the synthesis of proteins in several tissues. High insulin amount as an anabolic hormone causes to convert the small molecules in the blood into large molecules inside the cells and low amount level in the blood has an opposite effect towards catabolism. Simultaneous by increasing glucose in blood the beta cells secrete insulin into blood and as soon as glucose level decreases, secretion of insulin is inhibited. In contrast, alpha cell has activities based on beta cells, increasing secretion when blood glucose is low and decreasing secretion when blood glucose is high [14, 15]. Glucagon via stimulating the liver to produce glucose by gluconeogenesis has the opposite direction of insulin. The secretion of glucagon and insulin in response to the blood glucose condensation is the primitive mechanism of glucose homeostasis [14].

![Figure 1. Optimized structure of insulin.](Image)

If beta cell is annihilated by an auto-immune reaction, insulin can no longer be secreted into the blood which is called type 1 diabetes mellitus and specified by un-normal high blood glucose condensation. In type 2 the demolition of beta cell is less than in type 1 diabetes, and is due to an accumulation of amyloid in the pancreas, which probably interrupts its physiology. Although the pathogenesis of this kind of diabetes is not well known, it has been shown a decreased population of islet beta-cells due to the condensation of glucose in the blood [15]. As a result, the insulin concentrations, even when the blood sugar level is normal, are much higher than they are in healthy persons. The human insulin is consist of 51 amino acids, and has a molecular mass of 5808 Da. It is a dimer of two chains A and B (Scheme.1), which is attached together via disulfide bonds. Insulin's molecules differ somewhat among kinds of animals. Insulin from animal sources differs somewhat in mechanism from human insulin due to this difference. Porcine insulin is particularly close to the human type, and was vastly applied for treating type 1 diabetics before human insulin could be produced through recombinant DNA methods [14, 15].

1.3. Characteristics of sonified NMR.

Nuclear magnetic resonance is mostly applied in structure illumination and confirmation which are strongly sensitive to conformational changes in molecules. The human insulin is combined of 51 amino acids which can be divided into 7 series of amino acids for seven octaves of musical notes, including: Gly-Ile -Val-Glu-Gln –Cys –Cys–Thr as group (1). Thr-Ser-Ile-Cys-Ser-Leu-Tyr-Gln as group (2). Gln-Leu-Glu-Asn-Tyr-Cys-Asn-Phe as group (3), Phe- Val-Asn-Gln -His-Leu-Cys-Gly as group (4), Gly-Ser-His-Leu - Val-Glu-Ala-Leu as group (5), Leu-Tyr-Leu-Val-Cys-Gly-Glu-Arg as group (6) and Arg-Gly-Phe-Phe-Tyr -Thr-Pro-Lys, as group (7) which are shown in Figs 2-8. It is difficult to explain the details of NMR and accurate mechanisms to analyses data without introducing a wide concept of scientific subjects.

![Figure 2. Chemical shifts value of 8 amino acids in first octave.](Image)
which might be changed via Fourier transformation towards NMR spectrum, which is known as chemical shift with part per million or ppm unit. The conversion of NMR or FID spectrum to sound is NMR sonification which has been applied for insulin protein by this paper.

NMR spectroscopy is able for detecting any atoms and their isotopes with an odd number of protons or neutrons particles, $^{13}\text{C}$ is the most isotopes particularly suitable for sonification. This nucleus is the most commonly used in organic chemistry and database sites hosts spectra for more than 50,000 various chemicals found in the human body. Figures 2-8 exhibit $^{13}\text{C}$ NMR spectrum which converts the values to chemical shift units of ppm (deviation from a reference in parts per million) with Hertz. Here we will suppose which the reference frequencies are set to zero ppm.

Based on those atoms which are available in each mentioned of seven groups, frequencies clusters will occupy distinct frequencies. The number of signals in each spectrum relates to the complexities and structures of the molecules from one signal up even to 1000 or more peaks.

The peaks can be extended over a vast frequencies region or concentrated in narrow regions as seen in Figures 1-7.
$^{13}$C NMR exhibit a deviation of 1 ppm with frequency shift of 125 Hz and the frequencies ranges of $^{13}$C NMR peaks can appear between 0 – 30000 Hz which the most peaks generally locate above 12500 Hz. In contrast to proton NMR spectra, $^{13}$C NMR peak cannot be split and appears as a single frequency. For converting NMR spectrum to related sounds there are several ways based on mathematics of “Fourier transformation” and analyzing data through additive synthesis. The FID data in NMR calculation can be sonified straightly, through direct recording with related software in programs such as Matlab or Mathematica, from the output of the spectrum. Obviously the sonification of will leads to the most authentic molecular sounds which made from FID data, and occasionally are included of some useless of additional frequencies, arising from weak samples. FIDs Sounds consist of stochastic noises than sounds artificially created via additive synthesis. Therefore FIDs generated in a standard $^{13}$C NMR calculated sometimes decline within a short time.

A few experimental procedures exist that generate continuous signals or rapidly repeating FIDs. Via mathematical program it is possible to select particular frequencies from the sonified data and change the theme for each frequencies peak individually. Therefore, this action leads to more compatibility in the sounds creation processing and provides the construction of short bell-type sounds like murmur sounds. Using molecular sonification for providing a collection of applicable sounds, their utilization has to lie somewhere between two items first, it is possible to do the sounds unchanged and second it is possible to completely change the sonic characteristics of the starting material and convert to a new condition. Molecular sonification, as a sound creation method based on scientific approaches, might be useful tools for musical ideas that are describable via chemical mechanisms, such as Parkinson or Alzheimer’s disease. It can be argued that using calculating NMRs data as a source for sonification is more suitable than any indirect sonification method [16-18].

2. MATERIALS AND METHODS

Experiments methods of NMR &music sonification.
The start codon on mRNA inside ribosome is AUG, so natural all protein might contain methionine as starting amino acid, which means methionine acts as same as Sol’s key in musical notes. In addition the last amino acid in insulin is “Ala” which applied for ending of notes. Based on frequency calculations via DFT and ab-initio methods, the conversion of optimized energy of each amino acid to musical notes are listed in Table.1.

In abinitio calculation, the energies are related to specific methods and basis sets which are used in a model. In other words results during optimization of amino acids with any kind of methods or basis set are not equal. Consequently, the conversion of frequencies music notes for each amino acids might be different to data in table 1. Pitch, is one of the important components of musical notes which pivotal section of acoustics is related to it. In abinitio chemical calculation each methods including density functional theory, Moller -Plesset, Hartree- fock and semi empirical are related to a type of pitches. Treble clef symbol indicates that the second line from the bottom (the line that the symbol curls around) is "G". On any staff, the notes are always arranged so that the next letter is always on the next higher line or space. The last note letter, G, is always followed by another A.

In standard notation, a single musical sound is written as a note. The two most important things a written piece of music needs to tell you about a note are its pitch - how high or low it is - and its duration (time) - how long it lasts. To find out the pitch of a
written note, you look at the clef and the key signature, and then see what line or space the note is on. In NMR map the distances between chemical shift picks indicate the time scale between notes (Figs.10 & 11). A dot that is someplace other than next to the head of the note does not affect the rhythm. Other dots are articulation marks. They may affect the actual length of the note (the amount of time it sounds), but do not affect the amount of time it must be given.

Table 1. Frequency conversion of optimized amino acids via abinitio method (B3LYP/6-311G*) to acoustic music notes.

<table>
<thead>
<tr>
<th>Music Note</th>
<th>Amino Acids</th>
<th>Music Note</th>
<th>Amino Acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Trp</td>
<td>F</td>
<td>Gin</td>
</tr>
<tr>
<td>D</td>
<td>Met</td>
<td>G</td>
<td>Glu</td>
</tr>
<tr>
<td>E</td>
<td>Pro</td>
<td>A</td>
<td>Arg</td>
</tr>
<tr>
<td>F</td>
<td>His</td>
<td>B</td>
<td>Phe</td>
</tr>
<tr>
<td>G</td>
<td>Tyr</td>
<td>C</td>
<td>Ile</td>
</tr>
<tr>
<td>A</td>
<td>Leu</td>
<td>D</td>
<td>Ala</td>
</tr>
<tr>
<td>B</td>
<td>Val</td>
<td>E</td>
<td>Ser</td>
</tr>
<tr>
<td>C</td>
<td>Cys</td>
<td>F</td>
<td>Asn</td>
</tr>
<tr>
<td>D</td>
<td>Gly</td>
<td>G</td>
<td>Asp</td>
</tr>
<tr>
<td>E</td>
<td>Thr</td>
<td>A</td>
<td>Lys</td>
</tr>
</tbody>
</table>

Figure 9. Treble Clef Symbol of amino acids.

Figure 10. Time scales symbols for duration length of acoustic music notes.

Therefore for seven NMR groups the notes and time scales can be calculated (Fig.10). Time scale and note Lengths of group (1) based on distances between chemical shifts, each 10 ppm is one unite, as for instance among 200 to 170 there are 3 units so are equal to one half not plus to one quarter note. The simplest-looking note, with no stems or flags, is a whole note. All other note lengths are defined by how long they last compared to a whole note. A note that lasts half as long as a whole note is a half note. A note that lasts a quarter as long as a whole note is a quarter note. The pattern continues with eighth notes, sixteenth notes, and so on, each type of note is half the length of the previous type. Note lengths work just like fractions in arithmetic: two half notes or four quarter notes last the same amount of time as one whole note. Flags are often replaced by beams that connect the notes into easy-to-read groups.

2.1. Computational details.

Each part of the Insulin has been optimized using abinitio with DFT calculations individually (Figs.1-8). The whole of insulin has been calculated with QM/MM methods and the systems have been simulated through semi empirical methods. The final molecular structures were computed using SCF calculations in order to find the optimal starting geometries, as well as the different energies for those vibrational spectrums. The DFT with the van der Waals density functional theory was investigated to a model of exchange-correlation calculation. All optimization of seven groups were performed by GAMESS-US package. The accurate calculations performed using m062x, m06-L, and m06 for normal mode analyzing. The m062x, m06-L and m06-HF methods are suitable for non-bonded calculations for non-covalent interactions between each part of seven groups. The ONIOM methods including 3 levels of high (H), medium (M) and low (L) calculations have been done. DFT methods were used for the high (H) layer and the semi empirical method of pm6 and Pm3MM was used for the medium and low layers, respectively. In this work, differences of force fields are debated through comparing density and energies with OPLS and AMBER force fields. In addition, a Hyper-Chem professional release 7.01 program has been applied for some additional keywords such as PM3MM, PM6 (pseudo=lanl2). All calculation and estimation both modeling and simulation have been done based on my previous works [19-46].
3. RESULTS

Here is reported a self-consistent field calculation through abinitio and DFT method and also sonification approaches [47-64] for translating amino acids sequences into audible music sounds and applying it to generate protein of insulin designing using artificial intelligence (Figs 11, 12).

![Figure 12: Time scale and note Lengths of group (2) based on distances. Between chemical shifts, each 10 ppm is one unit, as for instance among 200 to 170 there are 3 units so is equal to one half not plus to one quarter note. The sonification methods supposed here uses the optimized 20 vibrations of the amino acids to compute an audible representation of each natural amino acid. The NMR chemical shifts are transposed to the audible spectrum following the musical concept (Figs 11 &12). 13C NMR chemical shifts have been calculated for seven groups of amino acid’s sequences based on insulin structure [figs 2-8]. Generally, 13C NMR peaks might be occupied higher frequencies than 1H NMR. Via 13C NMR a combination of sounds created using chemical shifts data that might be occupied the whole audible spectrum. The acoustamic piece of spins has been created, exploring the aesthetic possibilities of NMR derived sounds. The use of sounds made by sonification of NMR data in musical compositions and sound art is almost unexplored. This trans-position approach enables us to correspond the relative values of the vibrational frequencies within each amino acid towards musical notes. The specific frequencies spectrum and sounds associated by each of the amino acids exhibit a type of musical scale that includes 20 tones. For making playable music, each tone associated with the amino acids is assigned to a special key on a musical instrument, which enables us to draw the sequences of amino acids in insulin or any other proteins into a musical format.

4. CONCLUSIONS

The structure of the human insulin being analyzed related to the features present in its NMR spectra. In addition, it is possible to choose proteins based on their structural features in order to create acoustic music notes. Based on this work, using the sonification methodology presented in this paper, it was possible to create an acoustic music composition based exclusively on publicly accessible NMR data. It is certain that NMR sonification as an acoustic sound creation methodology based on physical and chemical data has high potential to be a strong tool for any further controlling the human diseases such as Cancer, Alzheimer, Parkinson and diabetes in the global environment.

5. REFERENCES

5. Ohno, S.; Ohno, M. The all-pervasive principle of repetitious recurrence governs not only coding sequence construction but also human endeavor in musical composition. Immunogenetics 1986, 24, 71-78. [https://doi.org/10.1007/BF00373112]
7. Gena, P.; Strom, C. A physiological approach to DNA music. In: CADE 2001; Glasgow, UK: Glasgow School of UK
15. Koeslag, J.H.; Saunders, P.T.; Terblanche, E. A reappraisal of the blood glucose homeostat which comprehensively explains the type 2 diabetes mellitus-syndrome X complex. The Journal
6. ACKNOWLEDGEMENTS

Thanks of the Central Tehran Branch, Islamic Azad University for supporting computational software and all necessary equipment.

© 2019 by the authors. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).