The Innovation

The sound of host-SARS-CoV-2 molecular interactions

José de la Fuente,[1,2,*] **Juan J. Pastor Comín,**[3] **and Christian Gortázar**[1]

[1]SaBio. Instituto de Investigación en Recursos Cinegéticos IREC-CSIC-UCLM-JCCM, Ronda de Toledo s/n, 13005 Ciudad Real, Spain
[2]Department of Veterinary Pathobiology, Center for Veterinary Health Sciences, Oklahoma State University, Stillwater, OK 74078, USA
[3]Centro de Investigación y Documentación Musical CIDoM-UCLM-CSIC, Facultad de Educación de Ciudad Real, Ronda Calatrava 3, 13071 Ciudad Real, Spain
*Correspondence: jose_delafuente@yahoo.com

Host- and virus-derived factors are key drivers of the COVID-19 pandemic (Gortázar et al.).[1] The interaction between the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike glycoprotein (S) and host cell membrane angiotensin I converting enzyme 2 (ACE2) receptor is highly dynamic, mediates virus infection, and is the main target for protective antibody response.[2]

Data sonification allows musical analyses of scientific questions

Scientists have advanced the identification of molecules across pathogens and host species and their molecular interactions. However, the big challenge now is to develop dynamic ways to analyze these data and formulate key questions on the function of these molecules/interactions in biological processes, such as pathogen infection and host immunity. To address this challenge, our resent research developed algorithms to approach music closer to scientific studies and challenges.[3] In this direction, scientists from the Massachusetts Institute of Technology applied sonification to the SARS-CoV-2 S-protein to contribute to the identification of drug and antibody targets (https://www.sciencemag.org/news/2020/04/scientists-have-turned-structure-coronavirus-music). To further advance the study of host-pathogen interactions, we considered the most recent processes of data sonification to use them not as esthetic instruments that can facilitate the construction of musical works, but to make a coherent transcription susceptible in all its dimensions to specific musical analyses of scientific questions.[3] The aim of our study was to provide an alternative and complementary approach for the study of SARS-CoV-2-host interactions and contribute to predictive models of animal host susceptibility to virus infection. In particular, we focused on the characterization of the interaction between SARS-CoV-2 S-protein receptor-binding domain (RBD) and host cell membrane ACE2 receptor (Figure 1).

Sonification to characterize RBD-ACE2 interactions

The RBD of SARS-CoV-2 initially identified variant WIV04/2019 S-protein (P0DTC2) that mediates viral entry into host cells was used together with human *Homo sapiens* (ENSG00000130234) and cat *Felis catus* (ENSFCAP00000052000) ACE2 proteins as a model. Human and cat were selected based on their susceptibility to SARS-CoV-2 infection,[4] but with high and low S-protein RBD-ACE2 affinity, respectively.[2] To determine the pitches for each amino acid, we used the algorithm previously developed to translate RNA code sequences into pitches.[3] This algorithm facilitates the expression of the nitrogenous bases as a short sequence of pitches coded by arbitrary attribution on each of the steps of the diatonic scale (Table S1). Therefore, each amino acid has a unique and distinct pitch translation, and the sequence of codons offers a line of tonal pitches whose reiterations, intervals, form, scope, tonal centers, inflection points, and reiteration of sounds allow a congruent musical analysis. To represent melodic motifs, a sequence of pitches must be characterized by a rhythmic structure that allows them to be understood as a unit. Each codon is metrically equivalent to a bar and has a differentiated rhythmic and melodic character.
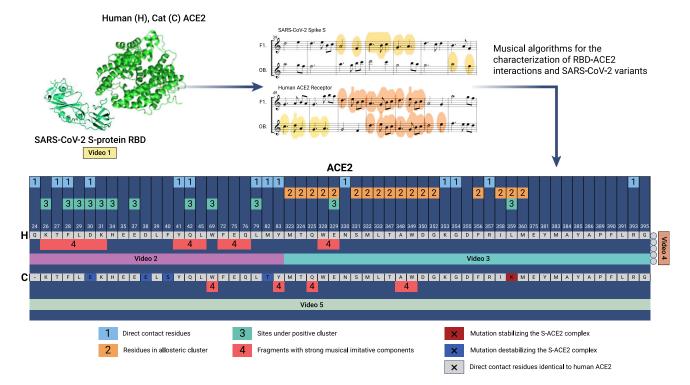

Figure 1. Melodic analysis of RBD-ACE2 interactions Methodological approach translating RBD-ACE2 interaction into musical algorithms for the characterization of RBD-ACE2 interactions and SARS-CoV-2 variants. Direct contact extended residues for human and cat ACE2 were obtained from Lam et al.[2] See also Videos S1, S2, S3, S4, and S5.

Three basic characteristics of the proposed musical analysis included linear analysis, replicability, and comparison from a polyphonic perspective. Linear analysis takes place once the pitches and rhythms have been defined and the genetic sequence can be read as a succession of rhythmic-melodic motifs that can configure larger units according to their own musical syntax of motif, semiphrase, phrase, period, etc. This allows the analysis of the sequence from a formal perspective, which can be characterized in coherent musical units. The replicability of this transcription independently of the genetic sequence, allows the comparison in strict musical terms between several genetic sequences. It is possible to implement a polyphonic perspective where genetic lines of different items can be seen as elements of a contrapuntal tissue that allows the observation, in a polyphonic context, of long sequences of unisons and melodic imitation effects even having long canonical structures that represent regions of amino acid homology between the compared sequences.

The melodic structure of the S-protein RBD that interacts with the cell ACE2 receptor was very coherent (i.e., leading thread for the listener to follow as the piece progresses; http://alanbelkinmusic.com/site/en/index.php/harmony-coherence-continuity/) (Figures S1A and S1B; Video S1). We distinguished first to fourth relevant musical motifs Q (caa), G (ggt), Y (tat), and N (aat) distributed into 3 sections, first section or introduction (bars 1−17), second or central section (bars 18−92), and third or final coda section with the reoccurrence of the first motif (bars 93−109). The greatest concentration of these motifs (up to 11 times in 23 measures) appeared at the end of the melodic line (bars 81−104). Coincidence in amino acids involved in RBD-ACE2 interactions and melody key residues and in second and third motifs were observed (Video S1).

The interaction between RBD and human ACE2 receptor was characterized using a polyphonic perspective by replicating the RBD in the extension of the ACE2 sequence (see methodology in the supplemental information). Within the polyphonic context, we found sequences of unisons, melodic imitations, and canonical structures that represent regions of amino acid homology between the compared sequences (Figure S2A). Bars involved in these thematic relations (n = 127) are distributed in 3 sections throughout the entire ACE2 sequence (amino acids 1−109, 322−433, and 698−end). These melodic imitation relationships were not restricted to the amino acids of the four motifs described above (Figure S1A). However, they coincide with ACE2 measures in the first (Q [caa], amino acids 336, 415, 803), second (G [ggt], amino acids 102, 105, 367, 368, 404, 405, 407, 422, 436), third (Y [tat], amino acids 23, 49, 50, 375, 718, 710, 746), and fourth (N [aat], amino acids 47, 48, 103, 370, 372, 718, 719, 745, 747) motifs. The greatest number of melodic imitations was identified in the first section (bars 1−109), with 42 measures involved. Fragments with a strong imitative component (bars 16−19, 26−32, 41−47, 48−51, 74−78, 101−104, 328−329, 336−338, 367−375, 404−412, 421−426, 714−729, and 795−805 were identified) (Figure S2A). The relevant musical motifs in RBS included first motif Q (caa), second motif G (ggt), third motif Y (tat), and fourth motif N (aat) (Figure S1B). These interactions can be viewed and listened in Videos S2, S3, and S4.

The interaction between RBD and cat ACE2 receptor was characterized applying the same polyphonic perspective used for human ACE2. When compared with human ACE2, fewer musical motivic relations (barely 44 bars involved) were observed with the involvement of only 13 bars in the first 109 bars (Figure S2B). Furthermore, the imitative relationship was sporadic or occasional, and unlike the RBD-human ACE2 interaction where the musical mimics occurred among a range of motifs other than the four delimited motifs in the S chain (Figure S2B), in the RBD-cat ACE2 interaction only the four musical motifs on the amino acids Q (caa), G (ggt), Y (tat), and N (aat) mediate most of the musical imitations. The musical imitations were located for the first motif on bars 325−326, second motif on bars 756−757, third motif on bars 427−428, 637, fourth motif on bars 58−70, 83−85, and 793, and third and fourth motifs in conjunction on bars 49−50 (Figure S2B). A different and significant melodic-imitative device was the one performed on A (gct) at bars 11−13, 129−131, and 348−349. These interactions can be viewed and listened in Video S5.

Melodically there are four motifs that give coherence to the SARS-CoV-2 S-protein RBD identified with the amino acids Q (caa), G (ggt), Y (tat), and N (aat). The RBD-human ACE2 interaction showed a high density of melodic-motoric imitations not restricted to these four amino acids, although they were highly participating and concentrated between bars 1−109, 323−423, and 698−808. However, in the RBD-cat ACE2 interaction there are hardly any melodic-motivic interactions, which were limited to the four motifs identified in RBD, with the only exception of certain imitations on the musical motif derived from A (gct). These musical imitative components when translated into biological information showed a higher correlation with ACE2 direct contact residues with RBD and sites under positive selection in human when compared with cat protein (Figure 1). Musical imitative components in ACE2 residues in the allosteric cluster were identified in both human and cat sequences (Figure 1).

Musical algorithms provide alternative and complementary methodological approaches

These results using musical algorithms support a stronger RBD-ACE2 interaction in humans than in cats, which identifies cats as susceptible to reverse zoonotic (human-to animal) virus transmission with low risk for human infection.[2,4] These results will be correlated with those from other approaches, such as changes in binding energy of the SARS-CoV-2 S-ACE2 complex in different host species.[2] However, new virus variants with mutations in S-protein may be more effective in evading immunity, thus increasing the risk for zoonotic virus transmission.[5] The musical algorithms described here provide alternative and complementary approaches to the study of protein-protein molecular interactions with implications in areas, such as host-pathogen interactions and quantum vaccinomics.[5] The application of these methods can contribute to the analysis and monitoring of SARS-CoV-2 virus variants generated in both humans and animal hosts with implications for the control of COVID-19 pandemic.

REFERENCES

1. Gortázar, C., Del-Río, F.J.R., Domínguez, L., and de la Fuente, J. (2020). Host or pathogen-related factors in COVID-19 severity? Lancet *396*, 1396−1397.
2. Lam, S.D., Bordin, N., Waman, V.P., et al. (2020). SARS-CoV-2 spike protein predicted to form complexes with host receptor protein orthologues from a broad range of mammals. Sci. Rep. *10*, 16471.
3. Artigas-Jerónimo, S., Comín, J., Villar, M., et al. (2020). A novel combined scientific and artistic approach for advanced characterization of interactomes: the Akiri/Subolesin model. Vaccines *8*, 77.
4. Shi, J., Wen, Z., Zhong, G., et al. (2020). Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. Science *368*, 1016−1020.
5. Sharun, K., Dhama, K., Pawde, A.M., et al. (2021). SARS-CoV-2 in animals: potential for unknown reservoir hosts and public health implications. Vet. Q. *41*, 181−201.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.