Colin King

3D Protein Modeling of CYB5R3 Gene Mutations in Recessive Congenital Methemoglobinemia

Introduction:

Recessive Congenital Methemoglobinemia (RCM) is a rare disease that is associated with the CYB5R3 gene. Located on the 22nd chromosome, this protein coding gene calls for the formation of the enzyme NADH-cytochrome *b5* reductase 3, which is responsible for reducing methemoglobin to hemoglobin to bind and carry oxygen [1]. Methemoglobin is the oxidized form of hemoglobin and contains ferric iron atoms, which cannot bind and carry oxygen [1]. Iron atoms in hemoglobin can naturally oxidize to the ferric state when releasing oxygen to cells. The cytochrome *b5* reductase enzyme uses NADH to reduce this oxidized hemoglobin (methemoglobin) back to a ferrous state, where it can bind and carry oxygen once more [1]. A mutation of the CYB5R3 gene can cause a deficiency of NADH-cytochrome *b5* reductase 3 enzymes, causing an increase in methemoglobin in the bloodstream. This increase in methemoglobin subsequently causes an oxygen deficiency, resulting in methemoglobinemia.

Methemoglobinemia has two variations, Type I (erythrocyte specific) and Type II (generalized). In Type I methemoglobinemia, activity of the NADH-cytochrome *b5* reductase 3 enzyme is reduced only in erythrocytes (red blood cells) [2]. With reduced activity, the enzyme NADH-cytochrome *b5* reductase 3 cannot efficiently convert ferric iron to ferrous iron, which causes an increase in methemoglobin and a decrease in hemoglobin [3]. This decrease in hemoglobin causes a decrease in the amount of oxygen delivered to tissues in the body.

In Type II methemoglobinemia, mutations in the CYB5R3 gene cause total loss of enzyme activity. This causes a significant increase in methemoglobin and decrease in hemoglobin in red blood cells [3]. This results in problems throughout the body, such as neurodegeneration, encephalopathy, and mental retardation, among many other medical conditions [3].

In this tutorial, 3D protein modeling of the CYB5R3 gene which codes for NADH-cytochrome *b5* reductase 3 will be examined to assess how mutations possibly lead to conformational changes. The process to acquire the reference transcript sequence for the gene will be demonstrated, and the mutations will be manually inserted into this sequence to represent mutated forms. The reference transcript sequence and subsequent variant sequences will then be assembled into three dimensional models to explore the change in enzyme conformation.

Tools used: MutationTaster and Robetta

How to access data:

NCBI Gene ID: 1727, https://www.ncbi.nlm.nih.gov/gene/1727/
NCBI Reference Sequence ID: https://www.ncbi.nlm.nih.gov/nuccore/NM_000398.7/
Ensembl Sequence ID: ENST00000352397.10
Ensembl Gene ID: ENSG00000100243

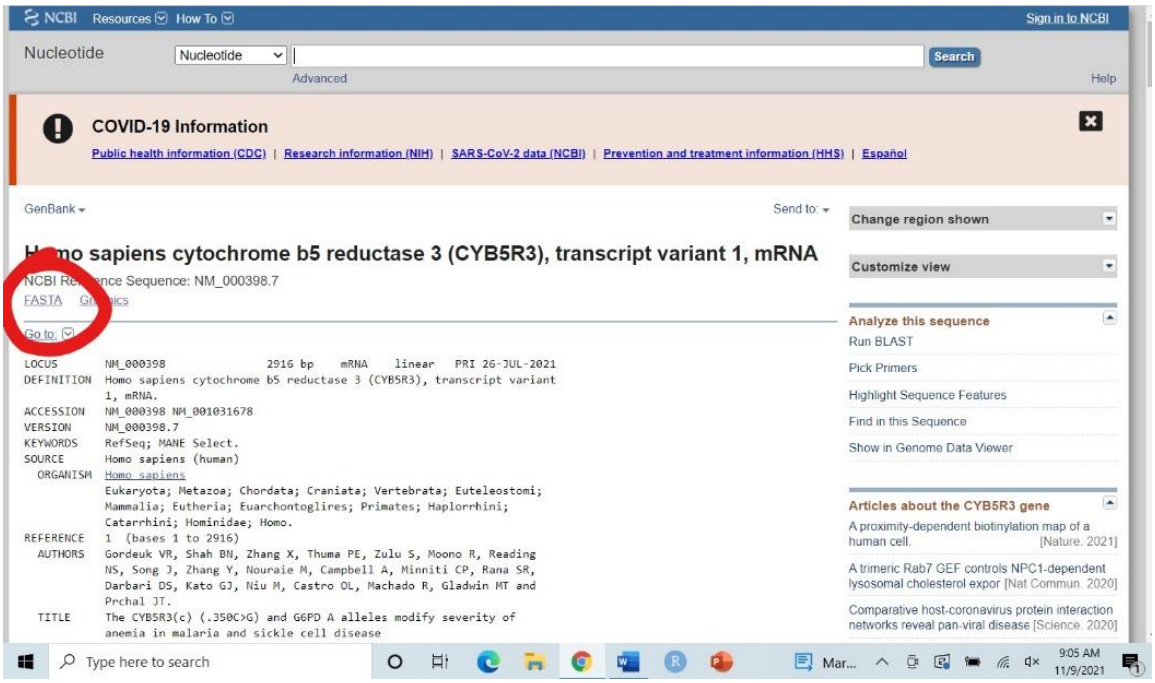
To examine mutations that cause RCM Types I and II, a table of selected mutations has been assembled in Table 1 below. Note that these are not all the mutations associated with this disease, but ones that have been selected to be used in a tutorial for analysis. These mutations were selected as they represent amino acid changes, indicating a possible change in protein structure.

Table 1. Gene Mutations and RCM Type (Mutations originally chronicled by Gupta et al 2020)

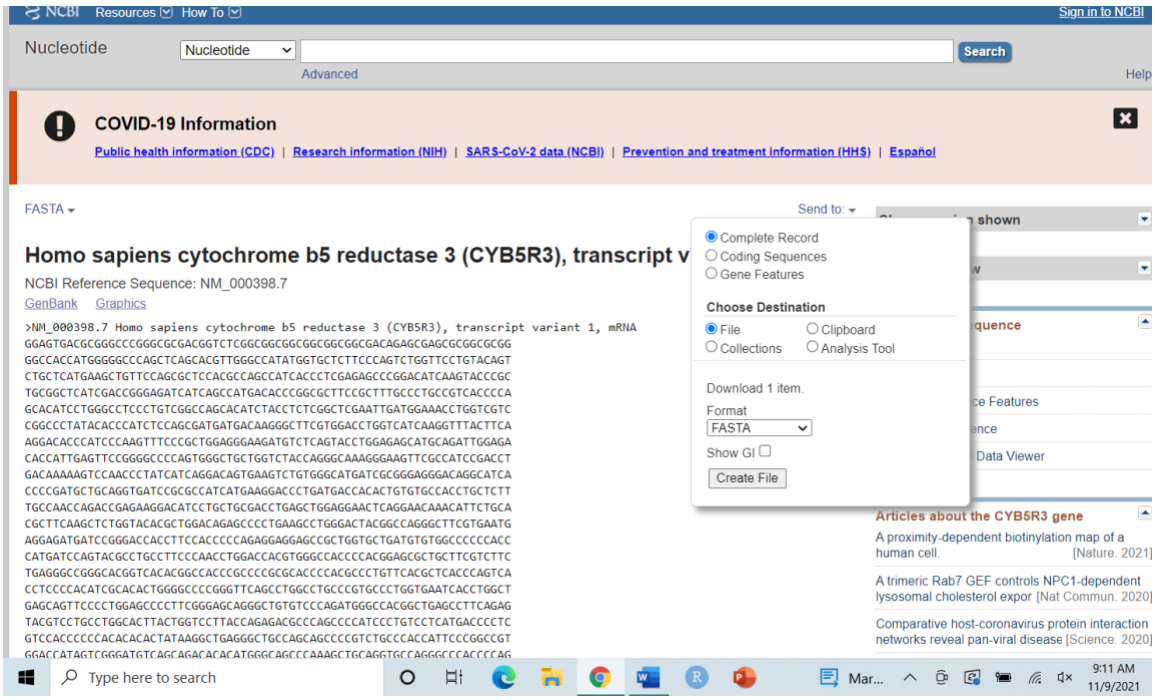
Nucleotide Variants	Amino Acid Variants	RCM Type	Original Source
c.136 C>T	Arg45Trp	I	Fermo et al (2008)
c.149 G>A	Arg50Gln	I	Dekker et al (2001)
c.190 C>G	Leu64Val	I	Gupta et al (2020)
c.287 C>A	Pro96His	II	Manabe et al (1996)
c.379 A>G	Met127Val	II	Kugler et al (2001)
c.610 T>C	Ile216Thr	II	Maran et al (2005)

Step 1: Downloading FASTA Reference Sequence.

Since these mutations do not have their own respective FASTA sequences, the mutations must manually be inserted into the reference sequence. To do this, you must first download the FASTA sequence for the reference sequence from NCBI. You can do this by typing in the URL listed at the end of this tutorial for NCBI Reference Sequence ID or by clicking the link here. Once you get to this page, click the link to display the reference FASTA sequence.



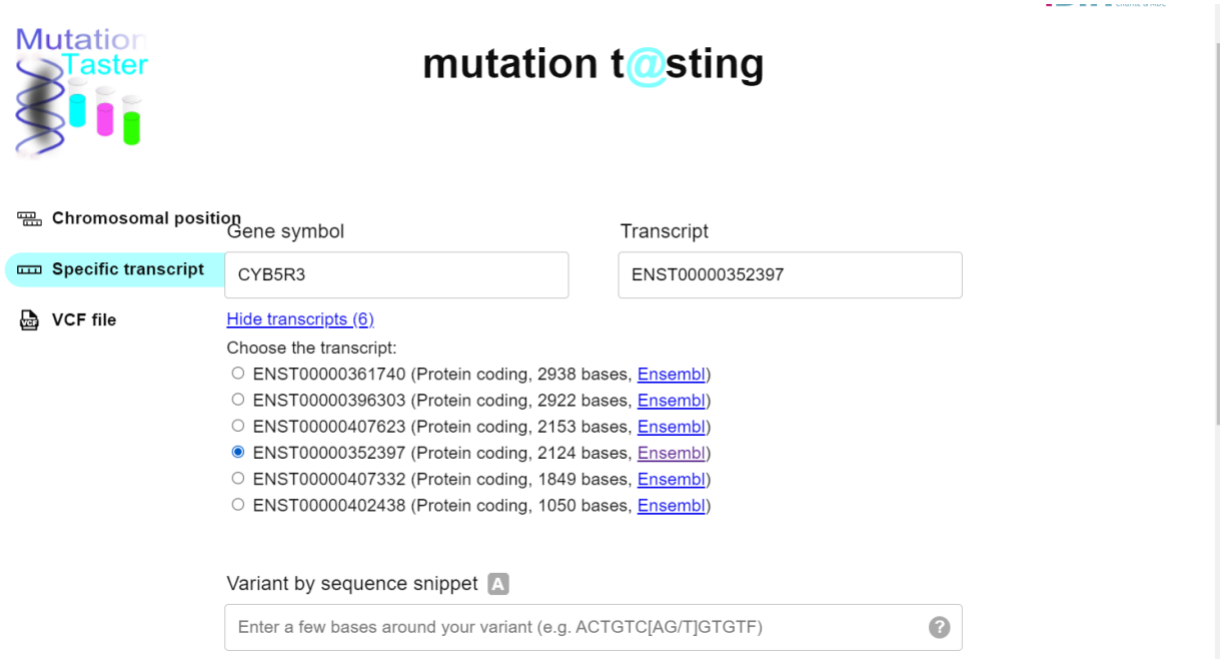
To download this sequence, click the prompt under send to. Select the options “Complete Record”, “File”, “FASTA”, and click create file. This will download the FASTA sequence to your device. Name this file CYBER_ref.txt. We will use this later.



Step 2: Mutation Insertion.

The next step in this project is to insert the specified mutations at the given locations in the reference sequence. You can do this by manually counting the number of nucleotides until you

get to the one that is changing, or you can use a tool to insert the mutation for you. In this tutorial we will be using a tool called MutationTaster. To use our sequence from above, fill in the search criteria in the main website interface as shown below.



The next step is to insert our mutations by nucleotide position. As an example, we will use the first mutation from Table 1. Scroll down on MutationTaster to the section labeled “variant by position”. In the position tab, input the position of the first mutation (136), and in the new base field, input the new nucleotide (T). Make sure that the coding sequence option is selected on the reference tab.

Clicking the Analyze button will take you to the results of your mutation. Among these results you will find the effect prediction, and comparisons between the reference sequence and your new mutated sequence. Copy and paste the “Mutated AA Sequence” into a new text file and name it after the mutation you just conducted. You will repeat this process for each mutation from Table 1.

Step 3: Making Protein Models.

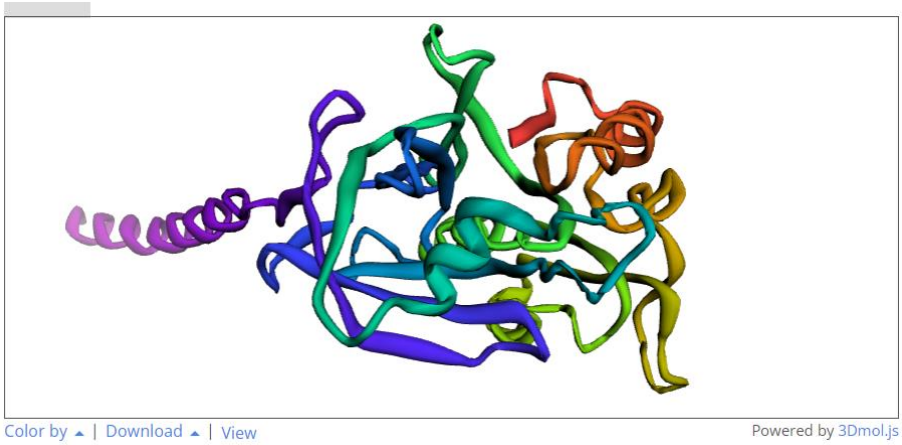
With our text files for each mutation now created, it is time to make three dimensional models to see how each mutation has changed the conformation of the enzyme coded by our CYB5R3 gene. We will be using Robetta. This website does require you to make an account but is completely free and can save protein builds; this is my preferred website for this type of protein analysis. After creating your account, use the dropdown menu under “structure prediction” and select submit. This will take you to the input page where you can upload the amino acid sequence for each mutated gene. The website should look like the one pictured below.

Next, copy and paste your mutated amino acid sequence into the input field for protein sequences. In the target name input, name your 3D model submission after the mutation that you are assembling; in this case it will be named c.136 C>T. Make sure that the model is the RoseTTAfold model. Before hitting submit, complete the math problem to verify that you are not a robot. With all fields filled in, click submit. The assembly will take a few hours to complete, and you will receive an email upon completion. You will repeat this process for each mutation sequence acquired from the previous step.

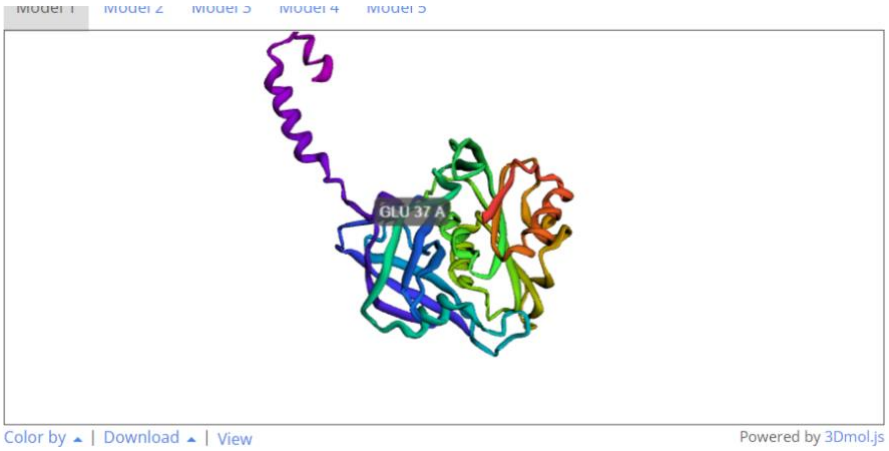
Step 4: Comparison of Protein Models.

After all protein models have been completed, their 3D structures can be compared to assess potential conformational changes as a result of their respective mutations. The models of each mutated sequence and the reference sequence model are shown below.

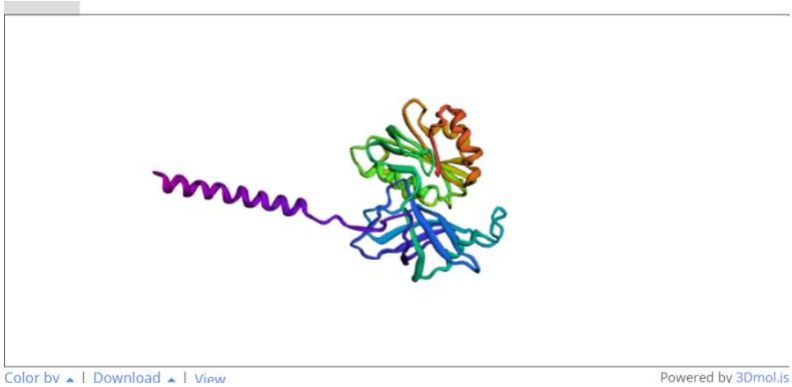
Reference Model:



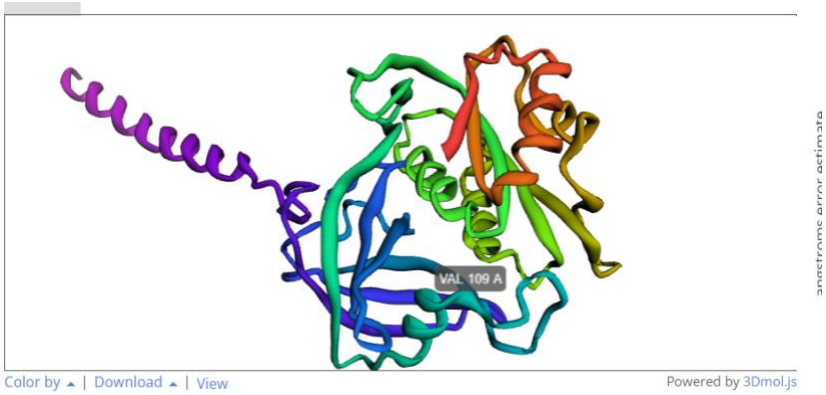
Mutation 136 C>T:



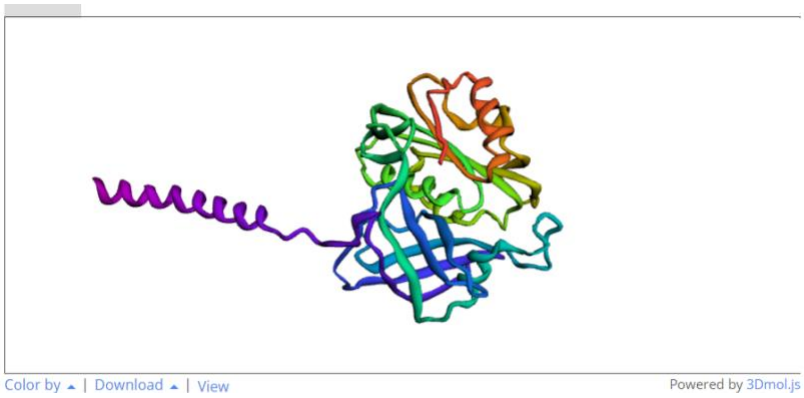
Mutation 149 G>A:



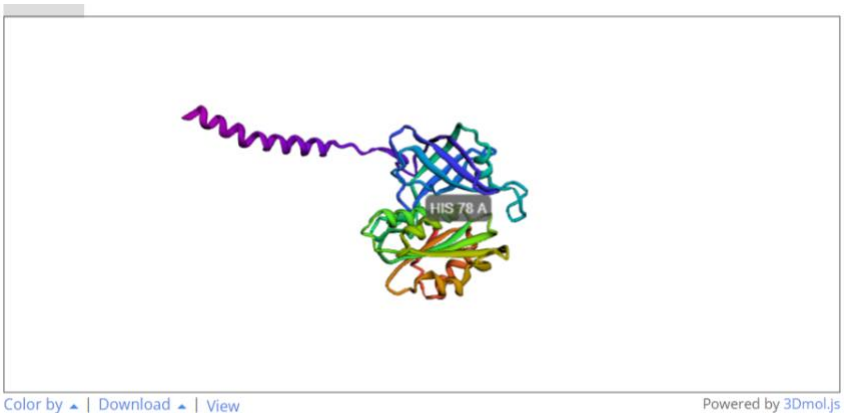
Mutation 190 C>G



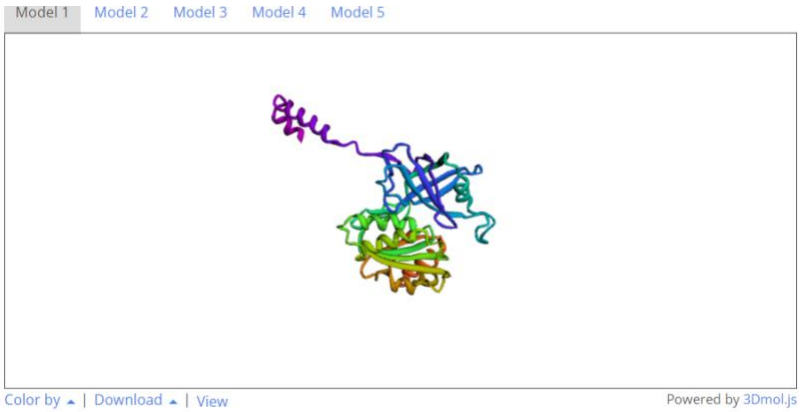
Mutation 287 C>A:



Mutation 379 A>G:



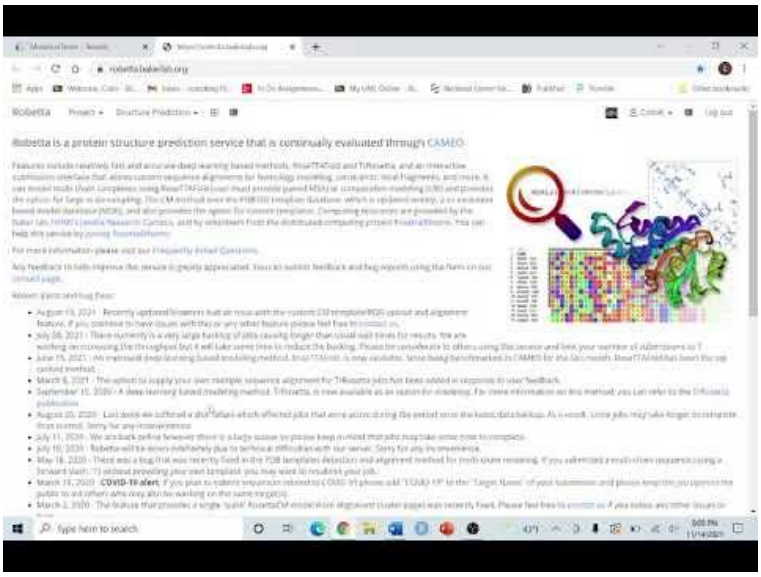
Mutation 610 T>C:



Discussion:

There do not appear to be major structural changes of the NADH-cytochrome-b5 enzyme among these mutations in comparison to the reference model. In some models, there appears to be a change in structure at the beginning of the amino acid chain, but this could be due to a high error estimate at the beginning of chains for each model assembly, per Robetta Lab's Angstrom Error Estimates. These mutations are all single nucleotide polymorphisms, so they may not affect enzyme conformation as much as insertions and deletions (indels) would. Indels are generally more noticeable because they change the length of the amino acid and can cause frame-shift mutations, meaning that the specific trio of nucleotides that code for an amino acid is altered in each read area. It is likely that the mutations used in this tutorial affect this enzyme in ways that we cannot see from examining the conformation of this enzyme.

Video Guide for the use of MutationTaster and 3D protein modeling: https://youtu.be/oAnLBAckthk



Link to NCBI Reference Sequence:

https://www.ncbi.nlm.nih.gov/nuccore/NM_000398.7/

Link to Mutation Taster:

https://www.genecascade.org/MutationTaster2021/#transcript

Link to Robetta Labs (protein modeling):

https://robetta.bakerlab.org/

References:

1. Ludlow JT, Wilkerson RG, Nappe TM. Methemoglobinemia. [Updated 2021 Sep 2]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK537317/
2. Lorenzo, Felipe R 5th et al. "Molecular basis of two novel mutations found in type I methemoglobinemia." Blood cells, molecules & diseases vol. 46,4 (2011): 277-81. doi:10.1016/j.bcmd.2011.01.005
3. "Autosomal Recessive Congenital Methemoglobinemia: Medlineplus Genetics." MedlinePlus, U.S. National Library of Medicine, 18 Aug. 2020, https://medlineplus.gov/genetics/condition/autosomal-recessive-congenital-methemoglobinemia/#references.
4. Gupta, V, Kulkarni, A, Warang, P, Devendra, R, Chiddarwar, A, Kedar, P. Mutation update: Variants of the CYB5R3 gene in recessive congenital methemoglobinemia. Human Mutation. 2020; 41: 737– 748. https://doi.org/10.1002/humu.23973