## Missense Mutation Muddle

Have you received a genetics report summary that looks something like this?

"The T60P missense variation is likely a disease causing mutation but without further evidence of the pathogenic nature of this variant we are not able to rule out the possibility that T60P is a benign coding polymorphism."

So, the mutation in this case is responsible for the disease in your child……unless it isn't. Confusing, isn't it? Sadly, such a report is not unusual and points to certain limits of genetic analyses. The mutation described above was identified in *RPS19* when it was the only gene known to be involved in DBA. With somewhere between 10-15 genes now known, and with scientists sequencing whole exomes (expressed regions of the genome, millions of bases of DNA) and whole genomes (the entire genome, billions of bases of DNA) in DBA patients in patients where the genes affected are unknown, we are going to be faced with more and more sequence variations where judgment calls will need to be made regarding their pathogenic nature. With mutations that can block the production of a candidate protein (nonsense and frameshift mutations) it is sometimes relatively straightforward to infer that a sequence change will eliminate the protein and therefore, its function. Moreover, if this function is known to be relevant to DBA pathogenesis, like something involved in ribosome biogenesis, one could feel confident stating that this is a disease causing mutation. But such judgment calls become far more complicated when all that has happened with a mutation is a change of one amino acid for another in a protein's sequence, a so-called missense mutation.

Rps19 is a string of 145 amino acids. At each position in this sequence the genetic code determine which of 20 common amino acids will be incorporated into the growing protein as it is made. A mutation may change the genetic code for the protein and insert an amino acid at one of these positions that differs from the one found in a healthy individual. This would be classified a missense mutation (missed-sense, where sense is used in the context of the normal meaning of the genetic code for this protein). Going back to the genetic report above, the sense readout at position 60 in Rps19 is to insert the amino acid threonine (T). The missense mutation in this patient instead replaces the threonine (T) with a different amino acid proline (P).

The million dollar question for this patient is whether this substitution does something that disrupts the protein's structure and in doing so interferes with its function.

# Biochemistry 101

So where to begin to determine if this change might be the pathogenic lesion in this patient? Well, one place to start is to look to nature and ask do other organisms have an Rps19 protein, and if so, are their structures similar to the human protein, particularly at position 60 where the change of interest occurs. The idea here is if proline is found at position 60 in Rps19 proteins from other organisms, this change may be allowable for maintenance of proper structure and function for the human protein and so rather unlikely that it is a pathogenic mutation.

It turns out that Rps19 proteins are found in all animals, plants and even archaebacteria. Archaebacteria are found in thermal vents and other extreme locations on the planet, so Rps19 proteins and their genes have been with us for billions and billions (we all have a little Carl Sagan in us) of years. And these different genes for Rps19 can change with time, through the acquisition of mutations. This random acquisition of mutations is Darwin's variation, so in neo-Darwinian terms, mutations in genes give rise to the variation upon which selective forces are imposed. And with time, Rps19 genes begin to change, being adapted to the needs of the organism in which they are found.

For example, a bacterium living in a thermal vent at temperatures near where water boils would presumably need a protein able to withstand these high temperatures, whereas a chicken living out back in the hen house would not need a protein able to withstand such high temperatures. So what does high temperature do to a protein? Well, let's take an egg from our chicken adapted for life in our back yard. Inside the egg is a clear goo surrounding the yolk. This goo is enriched in protein. Now, if we were to boil the egg, the protein in the goo denatures with heat, loses its structure and crashes out of solution forming a solid mass, the egg white. This loss of structure when a protein is denatured with heat is associated with a loss of function. If instead our chicken had been adapted to live in a thermal vent with temperatures approaching 212°F, boiling the egg and then cracking would reveal that nothing is happened, the goo is as slimy as ever, because our chicken protein had been adapted to withstand high temperatures. So, mutational change allows organisms the flexibility to adapt to new environments.

But I digress. Let us go back to Rps19. Rps19 proteins from different organisms all have functions in ribosome synthesis that are similar, so those parts of the protein involved in these similar functions might very well be the same for all organisms. In contrast, other parts of the protein not so critical for these functions might change in such a way as to make the protein more adaptable to specific needs of the organism (like our thermotolerant bacteria, or chicken). So in comparing the proteins between humans and other organisms, some amino acids along the sequence will be the same and other different, depending on how necessary they are for the most fundamental shared properties of the Rps19 protein.

Regarding the potential pathogenicity of the T60P change, we can turn to position 60 and ask how much variation there is at this position when comparing Rps19 sequences from different organisms. The thought here is that if position 60 of Rps19 shows a lot of variation in sequence it may not be so critical for function, which again might argue against the T60P variation being pathogenic. So let's take a look at the sequence of amino acids around position 60 of Rps19 proteins from widely distant organisms: humans, archaebacteria, worms, and yeast. Figure 1 shows that in the short stretch of Rps19 around position 60 some positions have exactly the same amino acid in each organism (shown in blue boxes). This is pretty impressive as these organisms have existed independently of one another for billions of years acquiring dozens of mutations changes throughout the Rps19 protein, and yet these sequences haven't changed at all. Thus it is inferred that these amino acids play critical functions in the most basic properties and functions of the Rps19 protein. But there are other positions in this region where there is variation, where the amino acids at that these positions differ from one organism to another. Moreover, one of these positions is position 60 (shown in red). In humans, there is a T (threonine), archaebacteria I (isoleucine), worms L (leucine) and yeast V (valine). Hmmm, so the position where the

patient's change has occurred shows a certain amount of flexibility, with allowable substitutions identified that apparently do not interfere with function.

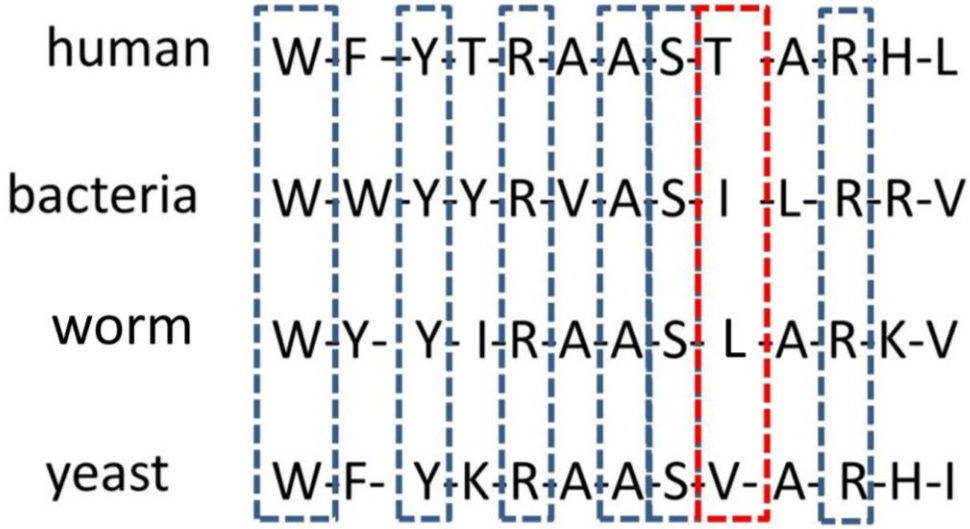


**Figure 1 - Sequence alignment of Rps19 proteins.** The alignment is shown for positions 50 to 64. Blue boxes show regions of absolute sequence conservation. The red box highlights position 60 in the Rps19 protein sequence. One letter codes are used for amino acids. *Editorial note:* the worm sequence is only relevant within the red box.

It's now worth taking a little time to look at the different amino acids found at position 60 (Figure 2).

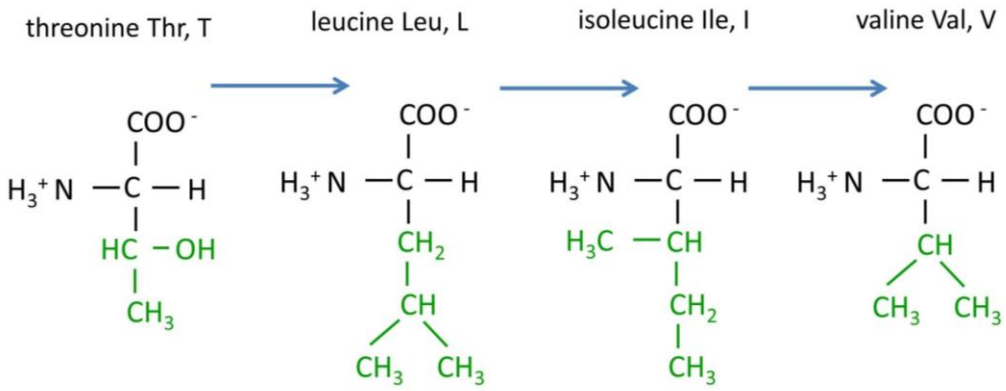


**Figure 2 – Amino acids found at position 60 in the Rps19 sequence**. Regions shown in green are unique, differentiating one amino acid from another. Regions found in black are used to join amino acids together, and are the same for each amino acid.

You can see with the exception of threonine, the parts of these amino acids shown in green, which are unique for each amino acid, are all very similar. They are composed of carbon and hydrogen, so called hydrocarbons. I don't know about you but when I think of hydrocarbons I think of oil. And we all know what happens when you mix oil with water, they separate forming

two distinct phases. Why, because hydrocarbons are hydrophobic, water hating! And where you find these types of amino acids in proteins is on the interior of the protein structure, away from water. What this suggests is that position 60 in the Rps19 structure is facing toward the inside of the protein shielded from water, and as long as you have a water hating amino acid at this position you may be OK. In looking at the 3D structure of Rps19, position 60 is found on the backside of the central helical structure shown with the arrow in Figure 3. This is exactly where you would expect to find a water hating amino acid since it is on the backside of the helix facing away from the aqueous environment into the interior of the protein. Even this statement has a little wiggle room, as the threonine in the human protein has an OH group which can interact favorably with water, but of the amino acids that like water, it likes water the least. So there you have it, position 60 is flexible as long as you put in an amino acid that isn't terribly thrilled about the prospect of interacting with water.

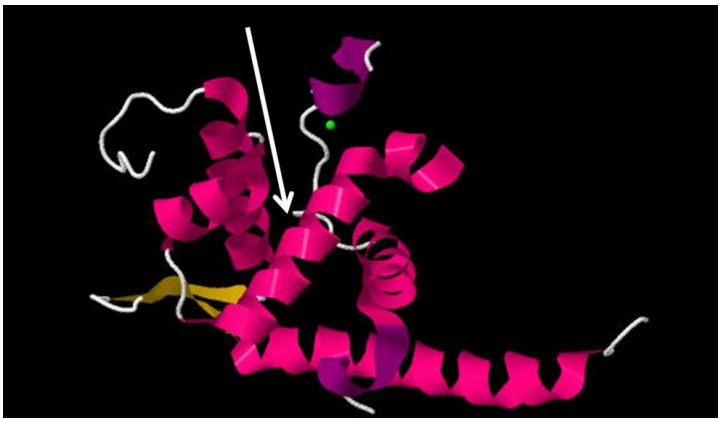

**Figure 3 – The three dimensional structure of Rps19.** Position 60 is found on the backside of the central pink helix facing the interior of the protein. If this region were illustrated with all atoms shown, the interior would be densely packed with water hating groups.

So now it's time to go back to the potentially pathogenic mutation in our DBA patient. T60P. P stands for proline. And if you look at the structure of proline in Figure 4 you'll note that it has a hydrocarbon side chain. Hmmm, again. Didn't I just say that position 60 is flexible and as long as this position contains an amino acid with a water-hating side chain the protein could very well be functional? So why did the genetic report indicate that this change is likely pathogenic?

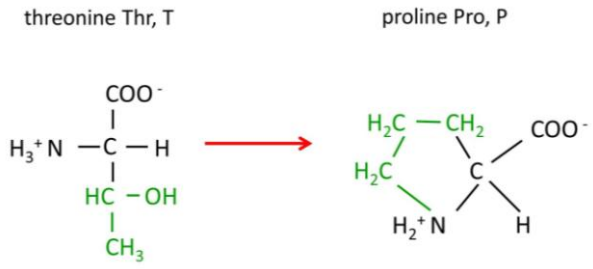


**Figure 4 – The structures of threonine and proline.**

Well, first of all, when we looked at the sequence of Rps19 from our diverse list of organisms we didn't see a proline there. So can we simplistically say that if we don't find it in our sequence comparisons it must not be an allowed change? I don't think so, there are thousands of organisms whose sequences for Rps19 are known and who's to say we've canvassed them all.

But there is something more insidious about proline. You'll note in Figure 4 that proline is kind of an odd duck among the amino acids because the atoms shown in green project from the central carbon like all the other amino acids, but then they wrap around and form a bond with nitrogen. This cyclic structure is unique to proline. To find out what makes proline stand out, we need to look at higher order elements of protein structure that go beyond just the order of amino acids along the chain. Protein chains fold into amazing structures with some parts of the chains forming spiraling helices, with others forming sheets and still others forming abrupt turns where the chains fold back on themselves (Fig. 3). Proline is the odd ball out in terms of forming the spiraling helical structures that are frequently found in proteins because the cyclic region in green constrains proline so it can't make the necessary rotations to form a helical structure. Oftentimes a proline will be used to disrupt a helix, so the chain at that point can assume a different type of structure. So prolines certainly have their place in protein structures. But if a protein was designed to have a helix at a particular point and a mutation inserts a proline into this region, the helix would most likely be disrupted and affect the overall three dimensional structure of the protein.

If you go back to the three dimensional structure of Rps19 shown in Figure 3, the white arrow points to where position 60 is found and low and behold, it is smack dab in the middle of a helix. So now we have a very different twist on the likely affects of this substitution on Rps19 structure. The helix that includes position 60 is almost certainly going to be affected by this substitution. Further, given that the other amino acids around position 60 are so highly conserved, this helical region is likely very critical for Rps19 function. Thus, when pressed, I suspect that 99 out of 100 biochemists polled as to whether this substitution could be pathogenic would answer, yes…….but. Are there exceptional cases were a proline substitution in a helical region does not disrupt the structure sufficiently to interfere with function? Yes, but don't they say the exception makes the rule?

So it's all a matter of degree of confidence with which a genetic report is making the call as to whether a mutation is pathogenic. In this case, T60P in Rps19, the genetics report states that this is very likely a pathogenic mutation (for the reasons outlined above). But they are unable to unequivocally rule out that it is not, so the qualifying clause in the genetics report. So the confidence meter in this case would be swung highly in favor of the mutation being pathogenic.

On the other hand, there are published reports of pathogenic mutations in Rps19 where I feel that the degree of confidence in the assertion that the mutations are pathogenic is far less certain than that discussed here.

The solution to the missense mutation muddle is ultimately to develop means of assessing the functions of proteins encoded by DBA genes so the effects of potentially pathogenic mutations can be tested directly. The good news is that these types of studies are ongoing in several laboratories around the world.