# Analysis of reported SCO2 gene mutations affecting cytochrome c oxidase activity in various diseases

**Radhika Chadha[1], Ritika Shah[2] & Shalini Mani[2]\***

[1]International Centre for Genetic Engineering and Biotechnology, New Delhi; [2]Department of Biotechnology, Jaypee Institute of Information Technology, Noida; Shalini Mani – Email: Mani.shalini@gmail.com; *Corresponding author

**Abstract:**
A large number of mutations have been reported in SCO2 (synthesis of cytochrome c oxidase) gene in association with COX deficiency reported in different diseases such as cardioencephalomyopathy, cardiomyopathy and Leigh syndrome. However, very few of these mutations have been functionally analyzed. SCO2 gene encodes for an essential assembly factor for the formation of cytochrome c oxidase (COX). It is a nuclear encoded protein that helps in transfer of copper ions to COX. This study is an attempt to understand the possible effect of these mutations on the structure and function of SCO2 protein, by using different *in silico* tools. As per Human Gene Mutation Database, total 11 non synonymous variations have been reported in SCO2 gene. Among these 11 variations, only E140K and R171W are functionally proven to cause COX deficiency. They have been used as controls in this study. The remaining variations were further analyzed using ClustalW, SIFT, PolyPhen-2, GOR4, MuPro and Panther softwares. As compared to the results of the controls, most of these variations were predicted to affect the structure of SCO2 protein and hence, may cause COX dysfunction. Thus, we hypothesize that these variations have the potential to result in a disease phenotype and should be investigated by subsequent functional analyses. This will help in an appropriate diagnosis and management of the wide spectrum of COX deficiency diseases.

**Key words**: SCO2, COX, *in silico*, OXPHOS, mitochondrial diseases, mutations

**Background:**
Oxidative phosphorylation (OXPHOS) is the main function of mitochondria and takes place in its inner membrane. There are five complexes that are involved in the respiratory chain. Cytochrome c oxidase (COX) is the Complex IV of this pathway. It catalyzes the transfer of reducing equivalents from cytochromec to molecular oxygen and utilizes the energy generated by this reaction to pump protons across the mitochondrial inner membrane. COX comprises of 13 subunits, two heme groups (a and a3), three copper ions (two in the CuA site and one in CuB site), a zinc ion, and a magnesium ion. The biogenesis of COX requires the interplay of two genomes mitochondrial DNA (mtDNA) andNuclear DNA (nDNA).MtDNAencodes the three larger subunits (COX I, COX II, and COX III) whereasnDNAencodes the remaining 10

smaller COX subunits and several accessory proteins (eg. SURF1, SCO1, SCO2), which are known to help in the proper assembly of all the 13 subunits of COX **[1]**. COX plays a vital role in the oxidative phosphorylation pathway. If Complex I is affected in the respiratory chain, then there is complex II for entry of electrons. But if COX is affected then the catalysis of the transfer of reducing equivalents from cytochrome c to molecular oxygen, is disturbed. There is no other complex which can contribute to this activity in the respiratory pathway. The energy generated by COX is necessary to perform a transmembrane proton-pumping activity. This proton pumping helps in maintaining the proton motive force which drives the synthesis of ATP with the help of ATP synthase. Thus, decreased production of ATP can lead to a wide spectrum of diseases in which organs with high energy requirements, such

# BIOINFORMATION

as brain and heart, are majorly affected. Hence, COX deficiency is a highly heterogeneous condition and leads to a wide clinical spectrum of diseases [2]. SCO2 gene (OMIM 604272.0008) is located on chromosome 22 (cytogenetic location: 22q13). Total length of this gene is2872 bp and it consists of four exons. This gene has four transcript variants. The fourth exon (size: 856 bp) is conserved in all the transcripts.SCO2 gene produces SCO2 protein, which is an assembly factor for COX and mainly takes part in copper (Cu) delivery. SCO2 protein is 266 amino acids long and its molecular weight is 25 KDa. SCO2 is a ubiquitously expressed protein and similar expression pattern is observed in different human tissues [3].
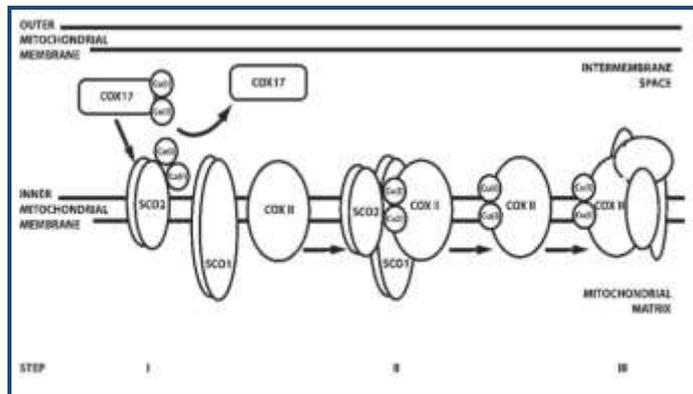


**Figure 1:** Flow of electrons with the aid of SCO1 and SCO2 (Leary *et al.,* 2004)

SCO2 with COX17 and SCO1 protein helps in forming the active COX and these proteins are responsible for the insertion of Cu into the COX holoenzyme **(Figure 1).** Cu generates cytotoxic free radicals when free within the cell. Thus, specific chaperons or assembly factors are needed for storage or incorporation of Cu into the target molecules. As SCO2 plays a critical role in COX formation; it has been associated with a number of COX deficiency diseases. For an instance, defects in SCO2 are the cause of fatal infantile cardioencephalomyopathy with COX deficiency [3]. A number of studies have been carried out to find out the mutations present in the SCO2 gene in patients with COX deficiency (**Table 1).** However, there is a lack of studies which explore the functional relevance of these variations and propose their effects on COX activity. It is imperative to do systematic studies which could highlight the importance of SCO2 variations, before initiating *in vitro* and *in vivo* functional studies. The preliminary investigation of a polymorphism can be performed with the help of a precursory computational analysis using different *in silico* tools. This approach can help in increasing the efficiency of molecular studies by narrowing down the potential pathogenic mutations. It will also pave path for functional characterization of non-validated variations. In the current study, we have predicted the possible effects of all the non-validated reported SCO2 gene mutations, on COX activity. Both validated and non-validated variations have been analyzed and compared to relatively predict their possible role in COX deficiency.

## Methodology:
### Data Collection
The present study incorporates a number of databases and software tools to collect information. The Human Gene Mutation Database (HGMD) constitutes a comprehensive core

collection of data on germ-line mutations in nuclear genes underlying or associated with human inherited disease (http://www.hgmd.org). It was used to obtain the specific mutations in SCO2 which have been reported in cases of COX deficiency. It provided eleven missense mutations that have been reported till now in SCO2 gene **Table 1 (see supplementary material).** Protein sequence for SCO2 was obtained from NCBI.

### Data Analysis
As per literature, out of these eleven mutations, E140K and R171W have already been functionally validated **[3-6].** These variations were used as controls for our study. Amongst the remaining nine non-validated mutations, three mutations (W36X, Q53X and R90X) are predicted to result in truncated SCO2 protein and were not included in our study. Six non-validated mutations (C133Y, L151P, V160G, M177T, G193S and S225F) and two validated mutations (E140K and R171W) were analyzed and compared using various bio-informatics tools.

### Mutation analysis
To analyze the importance of all the mutations in SCO2 protein sequence, multiple sequence alignment was done using ClustalW. It is a general purpose multiple sequence alignment program for DNA or proteins available at the EMBL-EBI server (http://www.ebi.ac.uk/Tools/msa/clustalw2/). It aligns multiple sequences and highlights areas of similarity.

### Secondary structure prediction
To analyze the effect of mutations on the secondary structure of SCO2 proteins, Gor IV was used. Gor IV is the fourth version of GOR secondary structure prediction method (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html). It uses all possible pair frequencies within the window of 17 amino acid residues to predict alpha helix, beta sheet, turn, or random coil secondary structure at each position.

### Predication of functional relevance of mutations
Several tools were used to determine a possible change in the protein function. SIFT (Sorting Intolerant from Tolerant) predicts whether an amino acid substitution affects the protein function (http://sift.jcvi.org/). Positions with normalized probabilities less than 0.05 are predicted to be deleterious; those greater than or equal to 0.05 are predicted to be tolerated.PolyPhen-2 (Polymorphism Phenotyping 2) is an automatic tool used for prediction of possible impact of an amino acid substitution on the structure and function of a human protein (http://genetics.bwh.harvard.edu/pph2/index.shtml). If the score is less than 0.5, prediction is benign else, the prediction is probably damaging. MuPro is used to predict how single site amino acid mutations affect protein stability (http://www.ics.uci.edu/~baldig/ mutation.htm). If the score is greater than 0, the prediction indicatesan increased stability otherwise there is decreased stability. Panther (Protein ANalysis THrough Evolutionary Relationships) classifies genes by their functions, using published scientific experimental evidence and evolutionary relationships to predict functions even in the absence of direct experimental evidence (http://www.pantherdb.org/). It calculates the subPSEC (substitution position-specific evolutionary conservation) score based on the alignment of evolutionarily related proteins. subPSEC scores are continuous values from 0

(neutral) to about −10 (most likely to be deleterious). $P_{deleterious}$ is also calculated which provides the probability of a mutation

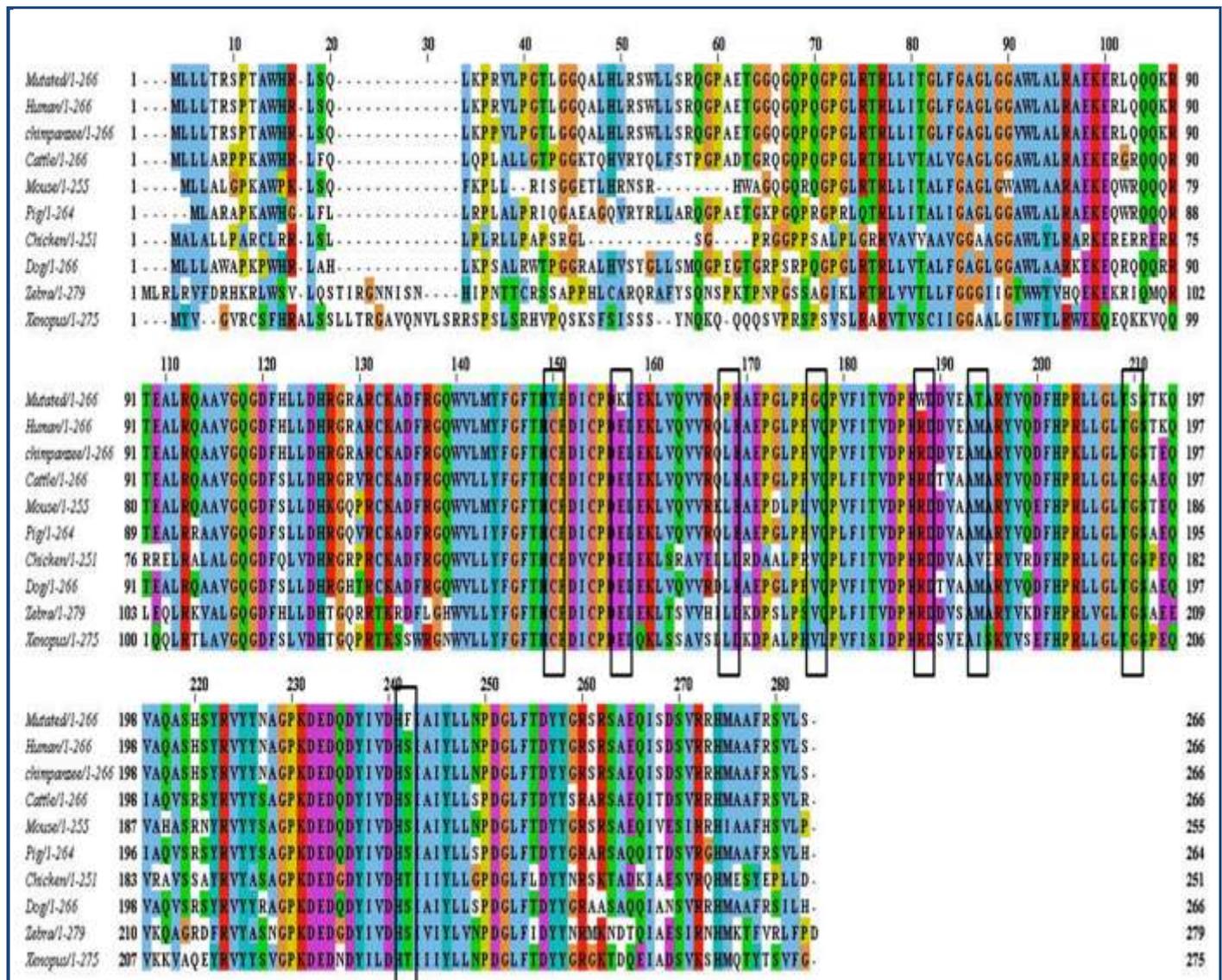being deleterious, where 1 indicates deleterious and 0 indicates not deleterious.



**Figure 2:** ClustalW results showing alignment of mutated human SCO2 gene sequence with other species. The missense mutations are highlighted and are found to be at completely conserved amino acid positions (except for M177T).

**Results & Discussion:**
Different bioinformatics tools have been used for the *in silico* analysis in this study. This has been done to ensure that the prediction for a particular mutation is correct and efficient. Probability of false positives is very common in computational work. Thus, it will help in better analysis of the results obtained. Also, along with the various non-validated reported mutations, functionally validated mutations (E140K and R171W) have also been used for comparison of the results with other mutations. E140K is a common change reported and very well validated in COX deficiencies **[3, 7, 8]**. When we compared the scores and predictions of E140K and R171W with the other non-validated mutations, we observed that the scores of E140K is less than that of the non-validated ones, according to almost all the tools. This helped us in elucidating that the non-validated mutations with almost similar or higher scores in comparison with validated mutations will also be deleterious and should show disease phenotype. The compiled results,

using all the tools are summarized in **Table 2 (see supplementary material).**

*Protein sequence alignment and analyses*
ClustalW results indicated that all the mutations are present at conserved amino acid positions, except M177T (**Figure 2**).

*Predication of structural and functional relevance*
E140K is a change from Glutamic acid to Lysine. It is a functionally proven mutation in COX deficiency reported firstly by Papadopoulou *et al.* in 1999 **[3].** As per this study, 3 patients with E140K mutation were suffering from a fatal infantile cardioencephalomyopathy and had decreased COX activity in their heart and skeletal muscles. In our analyses, SIFT and Polyphen-2 suggested that E140K is a damaging mutation, though the PANTHER score was less than the second functionally proven mutation, R171W, which is a change from arginine to tryptophan. R171W mutation was first reported in

# BIOINFORMATION

SCO2 gene, along with E140K and R90X, in 3 out of 10 patients clinically characterized with hypertrophic cardiomyopathy, muscle hypotonia, seizures and respiratory insufficiency **[6].** They concluded that these mutations result in a fatal infantile mitochondrial disorder characterized by hypertrophic cardiomyopathy and COX deficiency. After *in silico* analysis, this change has been predicted to be highly damaging by SIFT, Polyphen-2, MuPro and PANTHER. Hence, for both the validated mutations (used as controls), it is evident that they can be pathogenic in COX deficiency diseases. In a similar way, the rest of the mutations were analyzed. S225F was first reported along with E140K mutation and causes a change of serine to phenylalanine at 225 position **[3].** From the *in silico* data, it is predicted to be damaging by SIFT and Polyphen-2.Panther also indicated it to be deleterious with a score of 0.89, however, MuPro predicts increased instability.

Pronicki *et al.* (2010) investigated 18 patients with COX deficiency and one novel mutation M177T was found in SCO2 of one of the patient **[9].** M177T is a change from a non-polar methionine to a polar threonine. The *in silico* data we have for this novel mutation supports the decreased COX activity, as Polyphen-2, Panther and MuPro predict it to be damaging. All the mutations analyzed for secondary structure prediction by GORIV were predicted to be unable to change the structure, except for L151P which causes a change from helix to random coil. This variation was reported by Sacconi *et al.* (2003). They screened 30 patients with COX deficiency for mutations and found pathogenic changes L151P and E140K in a patient in SCO2 gene **[7].** The patients presented the classic phenotype of SCO2 defects with hypertrophic cardiomyopathy, encephalomyopathy and severely reduced COX activity. Thus, L151 is also predicted to be damaging by PolyPhen-2, MuPro and Panther. Moreover, Panther provides the highest deleterious score for this change: 0.98, comparable to the control R171W. Thus, clinical symptoms along with the *in silico* data suggest that it should be further investigated through functional validation. C133Y, a change of cysteine at the position 133 to tyrosine, is a novel mutation reported in a male neonate case of spinal muscular atrophy with COX deficiency **[10].** Post-mortem muscle, heart, and liver biopsies showed severe, moderate, and mild reductions in COX activity, respectively. This reduced activity is very well supported by the *in silico* analysis. SIFT and PolyPhen-2 predicted this mutation to be damaging and MuPro predicts decreased stability. It should be noted that Panther predicts this change to be deleterious with a score of 0.93, which is greater than Panther prediction score of the validated control mutation, E140K.

Knuf *et al.* (2007) reported a novel change V160G, in a girl with fatal infantile Cardioencephalomyopathy and the skeletal muscle biopsy of the patient revealed a significantly reduced COX activity **[11].** V160G causes the change of valine to glycine at position 160, which is a damaging change as predicted by

PolyPhen-2 and Panther. MuPro indicates a decreased stability of the protein due to this mutation. Thus, V160G seems to be one of the mutations that can cause severe damage in COX deficiency disease conditions. More recently, a novel mutation G193S has been reported in a patient with fatal infantile Cardioencephalomyopathy **[12].** It is a change from glycine to serine. PolyPhen-2 predicted this mutation to be damaging and MuPro predicts decreased stability. It should be noted that Panther predicts this change to be deleterious with a score of 0.95, which is greater than Panther prediction score of the validated control mutation, E140K and almost similar to that of R171W. It can be elucidated from the aforementioned studies that all the reported mutations analyzed in this study in SCO2 gene can probably be critical in causing COX deficiency.

**Conclusion:**
The functional analysis of the various reported mutations in SCO2 gene can be a potential approach for elucidating their pathogenic role in the disease phenotype. Thus, all the mutations predicted to be severely damaging in our study should be investigated further. The integrated analysis of mutations using different tools provides a clear picture of the possible impacts of genetic changes in SCO2. It provides a valid judgment and can serve as evidence to support an in-depth functional analysis. Moreover, before initiating the tedious laboratory experiments, selection of potential variants turns out to be highly cost-effective when working on complex disorders like COX deficiency diseases. These analyses can also be utilized in their clinical genetics, mainly for medical tests. Therefore, computational analyses like ours can serve as foundation blocks for breakthrough research in heterogeneous diseases and eventually help in their better management.

**References:**
**[1]** Zeviani M & Di Donato S, *Brain*. 2004 **127**: 2153 [PMID: 15358637]
**[2]** DiMauro S *et al. Adv Exp Med Biol.* 2012 **748**: 341 [PMID: 22729865]
**[3]** Papadopoulou LC *et al. Nat Genet*. 1999 **23**: 333 [PMID: 10545952]
**[4]** Leary SC *et al. Hum Mol Genet*. 2004 **13**: 1839 [PMID: 15229189]
**[5]** Verdijk RM *et al. Am J Med Genet*. 2008 **146**: 2822 [PMID: 18924171]
**[6]** Jaksch M *et al. Hum Mol Genet*. 2000 **9**: 795 [PMID: 10749987]
**[7]** Sacconi S *et al. Pediatr Res*. 2003 **53**: 224 [PMID: 12538779]
**[8]** Salviati L *et al. Arch Neurol*. 2002 **59**: 862 [PMID: 12020273]
**[9]** Pronicki M *et al. Eur J Paediatric Neurol*. 2010 **14**: 253 [PMID: 19879173]
**[10]** Tarnopolsky MA *et al. Am J Med Genet*. 2004 **125**: 310 [PMID: 14994243]
**[11]** Knuf M *et al. Acta Paediatr*. 2007 **96:** 130 [PMID: 17187620]
**[12]** Mobley BC *et al. Clin Neuropathol*. 2009 **28**: 143 [PMID: 19353847]

# BIOINFORMATION

## Supplementary material:

**Table 1**: Reported mutations in SCO2 gene associated with COX deficiency (data obtained from HGMD)

| No. | Amino acid position in SCO2 protein | NCBI rs ID | Disorder | Functional Studies |
|---|---|---|---|---|
| 1. | W36STOP CODON | rs121908508 | Cardioencephalomyopathy, fatal infantile [5] | Not validated |
| 2. | Q53STOP CODON | rs74315510 | Cytochrome c oxidase deficiency [3] | Not validated |
| 3. | R90STOP CODON | rs74315512 | Cytochrome c oxidase deficiency [6] | Not validated |
| 4. | C133Y | rs28937868 | Cytochrome c oxidase deficiency [7] | Not validated |
| 5. | E140K | rs74315511 | Cytochrome c oxidase deficiency [3] | Validated |
| 6. | L151P | rs137852511 | Cytochrome c oxidase deficiency [8] | Not validated |
| 7. | V160G | Not available | Cardioencephalomyopathy, fatal infantile [9] | Not validated |
| 8. | R171W | rs28937598 | Cytochrome c oxidase deficiency [6] | Validated |
| 9. | M177T | Not available | Cytochrome c oxidase deficiency [10] | Not validated |
| 10. | G193S | Not available | Cardioencephalomyopathy, fatal infantile [11] | Not validated |
| 11. | S225F | rs80358232 | Cytochrome c oxidase deficiency [3] | Not validated |

**Table 2:** *In silico* analysis of the reported validated and non-validated mutations in SCO2 gene.

| SCO2 mutations | CLUSTAL W Prediction | SIFT Prediction | Score | POLYPHEN-2 Prediction | Score | GOR4 | MUPRO Prediction | Score | PANTHER subPSEC | $P_{deleterious}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| E140K (as a control for the analysis) | Mutation at completely conserved amino acid position | Damaging | 0.00 | Probably damaging | 1.00 | c->c | Decrease Stability | -0.73 | -4.17 | 0.76 |
| R171W (as a control for the analysis) | Mutation at completely conserved amino acid position | Damaging | 0.00 | Probably Damaging | 1.00 | c->c | Decrease Stability | -1.0 | -8.31 | 0.99 |
| C133Y | Mutation at completely conserved amino acid position | Damaging | 0.00 | Probably damaging | .998 | c->c | Decreased stability | -0.11 | -5.6 | 0.93 |
| S225F | Mutation at completely conserved amino acid position | Damaging | 0.00 | Probably damaging | .999 | h->h | Increased stability | +0.82 | -5.1 | 0.89 |
| L151P | Mutation at completely conserved amino acid position | Not found | - | Probably damaging | 1.00 | h->c | Decreased stability | -1.0 | -7.0 | 0.98 |
| V160G | Mutation at completely conserved amino acid position | Not found | - | Probably damaging | 0.992 | c->c | Decreased stability | -1.0 | -5.6 | 0.93 |
| M177T | Mutation at an amino acid position conserved between groups of strongly similar properties. | Not found | - | Probably damaging | 0.970 | h->h | Decreased stability | -1.0 | -3.89 | 0.71 |
| G193S | Mutation at completely conserved amino acid position | Not found | - | Probably damaging | 1.00 | c->c | Decreased stability | -0.34 | -5.99 | 0.95 |

ClustalW: Comparing various sequences to find conserved amino acid locations. SIFT: Sorting Intolerant From Tolerant-If score < 0.05 mutations are predicted to be deleterious else are predicted to be tolerated. PolyPhen-2: Polymorphism Phenotyping v2- If score <0.5 prediction is benign, if score > 0.5 prediction is probably damaging. GOR4: Predicts change in secondary structure due to mutation-h referring a helix and c referring random coils. MuPro: Prediction of protein stability changes upon mutations-If score >0.0 prediction is increased stability else prediction is decreased stability. Panther: Protein analysis through evolutionary relationships, subPSEC: substitute position specific evolutionary conservation: based on alignment of evolutionary related proteins. $P_{deleterious}$ gives the probability of the mutation being deleterious where 1 is deleterious and 0 is not deleterious.