Homologous Modeling of Nonfunctional Human Gulonolactone Oxidase Remnant

Neil Berkel

Abstract

This work involves the homologous modeling of the nonfunctional human gulonolactone oxidase remnant that is believed to exist in humans using computational modeling software. Genomic studies have established the amino acid sequence of a functional gulonolactone oxidase present in rodents. This information is used to model the nonfunctional enzyme remnant most probably still present in humans. 3-D structural analysis of the modeled enzyme in comparison to the functional enzyme is performed to discuss the reasons why this remnant has become nonfunctional. This study helps with the understanding of the ways enzymes might lose activity through evolution.

Introduction

Vitamin C is known to be an essential dietary nutrient. It is a cofactor required for the activity of many enzymes in the body. Ascorbic acid, the reduced form of vitamin C, is an effective antioxidant due to its electron donating nature and its ready conversion to the reduced state. Vitamin C is also believed to decrease the risk of chronic diseases, such as heart disease, cancer, eye diseases, and neurodegenerative conditions (Jacob and Sotoudeh, 2002).

Humans cannot synthesize vitamin C due to the lack of functional gulonolactone oxidase enzyme which catalyzes the last step of the synthesis (Nishikimi and Yagi, 1991). However, this enzyme is present in mice and they can synthesize vitamin C (Jenness *et al.*, 1984). It is reasonable to assume that there could be a nonfunctional gulonolactone oxidase present in humans which lost its activity after evolutionary split from rodents. The purpose of this project is to model the nonfunctional human enzyme based on the structural analysis of the functional gulonolactone oxidase.

Methods

The freeware used for this project was Swiss-PdbViewer (DeepView) version 4.1. The interface of this software allowed the user to analyze several proteins at once. The proteins were superimposed using the ‘magic fit’ tool. This tool compared the primary sequences of two proteins and selected the best matching fragments of amino acid sequences before superimposing the proteins. Similarly, there was an ‘iterative magic fit’ tool that went through several cycles of improving the fit between the two proteins before superimposing them. More cycles simply improved the superposition of the proteins.

DeepView also calculated the ‘threading energy’ of the superimposing process. ‘Threading energy’ showed how closely a created protein fold compared to a known homologous protein fold. The model was submitted to the PRODRG server maintained at the University of Dundee and underwent an ‘energy minimization’ calculation. The energy minimized model was then optimized by Swiss Model’s automated protein structure homology-modeling server.

Swiss-Model was also used to find templates for a known protein sequence. Templates are similar protein sequences that have a known 3-D structure. A similar function was performed using the National Center for Biotechnology Information (NCBI) database. On this database, a Basic Local Alignment Search Tool (BLAST) was used first to find the protein sequences of interest, then to search for the similar protein sequences with a known 3-D structure.

Obtaining the Human Gulonolactone Oxidase Sequence

The initial sequence for gulonolactone oxidase in *Mus musculus* (house mouse) was found from the NCBI database. The reference sequence found was NP_848862.1 (Kim *et al.*, 2013). The sequence is:

MVHGYKGVQFQNWAKTYGCSPEMYYQPTSVGEVREVLALARQQNKKVKVVGGGHSPSDIACTDGFMIHMG
KMNRVLQVDKEKKQVTVEAGILLTDLHPQLDKHGLALSNLGAVSDVTVGGVIGSGTHNTGIKHGILATQV
VALTLMKADGTVLECSESSNADVFQAARVHLGCLGVILTVTLQCVPQFHLLETSFPSTLKEVLDNLDSHL
KKSEYFRFLWFPHSENVSIIYQDHTNKEPSSASNWFWDYAIGFYLLEFLLWTSTYLPRLVGWINRFFFWL
LFNCKKESSNLSHKIFSYECRFKQHVQDWAIPREKTKEALLELKAMLEAHPKVVAHYPVEVRFTRGDDIL
LSPCFQRDSCYMNIIMYRPYGKDVPRLDYWLAYETIMKKFGGRPHWAKAHNCTRKDFEKMYPAFHKFCDI
REKLDPTGMFLNSYLEKVFY

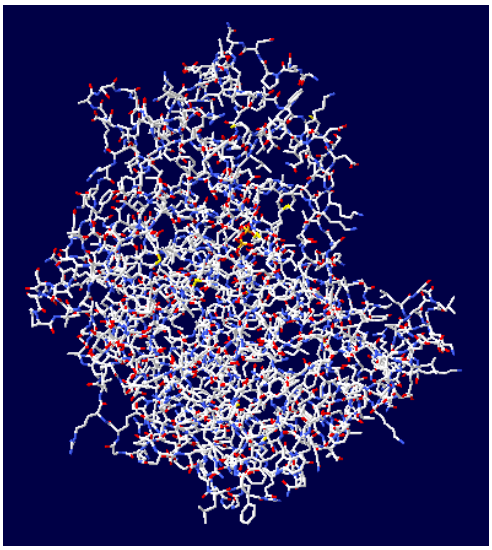
There are 440 residues in this reference sequence. After collecting this reference sequence for mouse gulonolactone oxidase, the sequence for a nonfunctional gulonolactone oxidase in humans could be searched. By performing a BLAST search on NP_848862.1 and searching specifically for *Homo sapiens* protein sequences, the highest scored sequence was obtained. The highest scored sequence was NP_055577.1. This sequence is for 24-dehydrocholesterol reductase (Ishida *et al.*, 2013).

MEPAVSLAVCALLFLLWVRLKGLEFVLIHQRWVFVCLFLLPLSLIFDIYYYVRAWVVFKLSSAPRLHEQR
VRDIQKQVREWKEQGSKTFMCTGRPGWLTVSLRVGKYKKTHKNIMINLMDILEVDTKKQIVRVEPLVTMG
QVTALLTSIGWTLPVLPELDDLTVGGLIMGTGIESSSHKYGLFQHICTAYELVLADGSFVRCTPSENSDL
FYAVPWSCGTLGFLVAAEIRIIPAKKYVKLRFEPVRGLEAICAKFTHESQRQENHFVEGLLYSLDEAVIM
TGVMTDEAEPSKLNSIGNYYKPWFFKHVENYLKTNREGLEYIPLRHYYHRHTRSIFWELQDIIPFGNNPI
FRYLFGWMVPPKISLLKLTQGETLRKLYEQHHVVQDMLVPMKCLQQALHTFQNDIHVYPIWLCPFILPSQ
PGLVHPKGNEAELYIDIGAYGEPRVKHFEARSCMRQLEKFVRSVHGFQMLYADCYMNREEFWEMFDGSLY
HKLREKLGCQDAFPEVYDKICKAARH

The sequence includes 516 residues with an expectation value of 7×10^{-11}. The residues were 31% identical and 53% similar. Also, there was only a single gap in the sequence.

Modeling Mouse Gulonolactone Oxidase

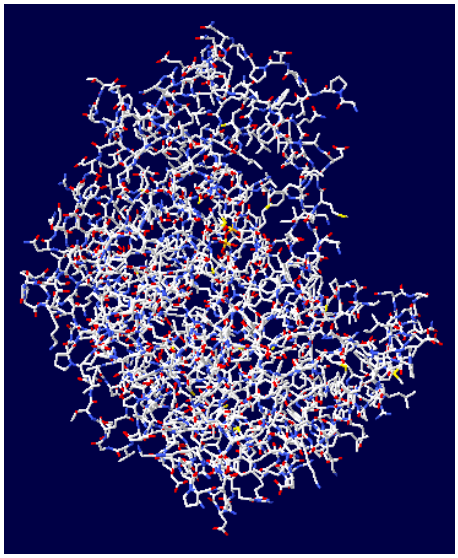
The 3-D structure of the gulonolactone oxidase from *Mus musculus* is not available, because it has not yet been determined. Therefore the first step involved the homologous modeling of the functional rodent enzyme. Templates with known 3-D structure were matched to the mouse sequence using Swiss-Model. The two templates that showed the optimum match were cholesterol oxidase from *Chromobacterium* sp. DS-1 (CHOLOX) (3JS8.1.A) and cholesterol oxidase from *Brevibacterium sterolicum* (2I0K.1.A). The 3-D models of these bacterial enzymes were displayed using Swiss-Model. These templates were 'magic fitted' using 3JS8 as the reference protein since this template was the best match for the mouse sequence. The template structure of 3JS8 is shown in Views 1 and 2.



View 1: *Solvent-Stable Cholesterol Oxidase Model.* This enzyme is commonly used for the detection of cholesterol in blood samples (Sagermann *et al.*, 2010). The image displays the protein in CPK (standard) colors. The entire protein is displayed.

View 2: *Solvent-Stable Cholesterol Oxidase Model in Ribbon Diagram.* The cholesterol oxidase is displayed in ribbon diagram. The ribbon is colored by secondary structure.

The template structure of 2I0K is shown in Views 3 and 4.



View 3: *Cholesterol Oxidase Model from Brevibacterium sterolicum with His121Ala Mutant Protein.* This protein catalyzes the oxidation of cholesterol-5-en-3-one followed by isomerization to cholesterol-4-en-3-one (Lim *et al.*, 2006). The image displays the protein in CPK colors. The entire protein is displayed.

View 4: *Cholesterol Oxidase Model from Brevibacterium sterolicum with His121Ala Mutant Protein in Ribbon Diagram.* The ribbon is colored by secondary structure.

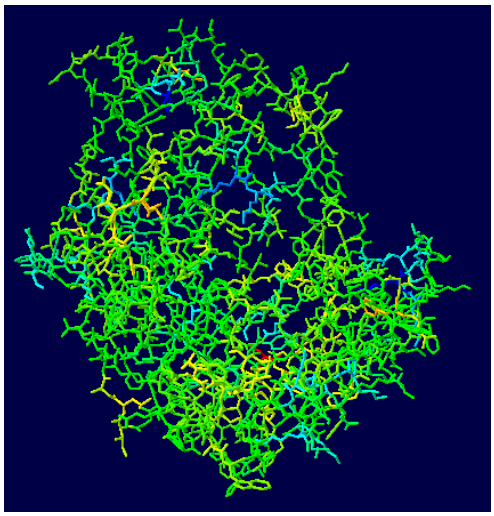
Analysis results and comparison of the two models are shown in Table 1.

Table 1: Comparison of 3JS8 and 2I0K templates by secondary structure, polarity, and charge.

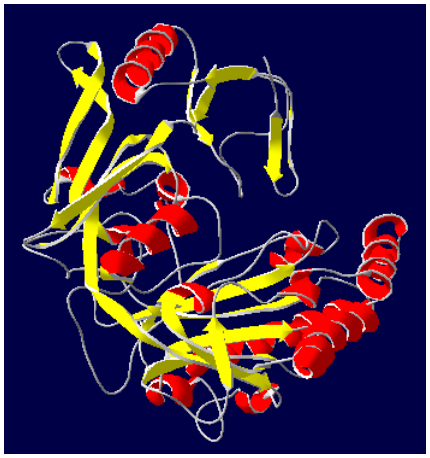
Characteristics of Templates		
Template	3JS8	2I0K
Total Residues	538	536
Alpha Helices	17	17
Beta Strands	22	22
Alpha Helix Residues	146 / 27%	144 / 27%
Beta Strand Residues	138 / 26%	141 / 26%
Polar Residues	132 / 25%	129 / 24%
Nonpolar Residues	310 / 58%	298 / 56%
Acidic Residues	44 / 8%	55 / 10%
Basic Residues	52 / 10%	54 / 10%

In the published 3-D structure by Swiss-Model, there were several gaps in the sequence. In the 3JS8 template, the first 2 residues were missing. In the 2I0K, 25 residues at two separate locations (53-57 and 330-349) were missing. The modeling process was performed to minimize the error due to sequence gaps. Within each template, there were sections of 9 and 5 antiparallel beta strands. Overall, the two template structures appear to be very similar after 'magic fitting' the 2I0K template to the 3JS8 template.

Superimposing these two templates enables the user to create a better model for the gulonolactone oxidase with the mouse sequence. The 3-D model for the gulonolactone oxidase was created by magic fitting it to the superimposed model of the two templates (see Views 5-10).

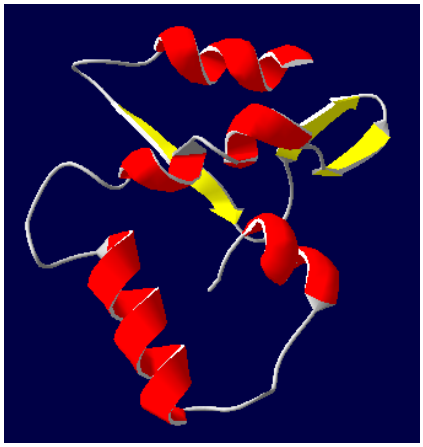


View 5: *Gulonolactone Oxidase Model in Mus musculus before Energy Minimization.* The oxidase protein is colored by threading energy. Blue and red colors represent the lowest and the highest threading energies, respectively. The majority of the residues are colored green indicating average energies.

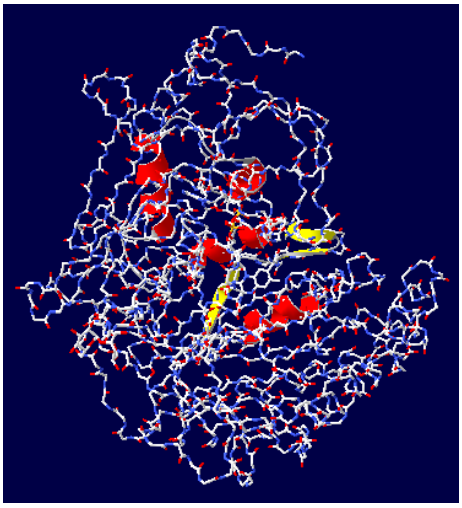


View 6: *Gulonolactone Oxidase Model in Mus musculus before Energy Minimization in Ribbon Diagram.* The ribbon is colored by secondary structure.

The dissimilarities in the secondary structure in gulonolactone oxidase protein and the superimposed model of the templates are shown in Views 7 (exclusively) and 8 (in the context of the protein).

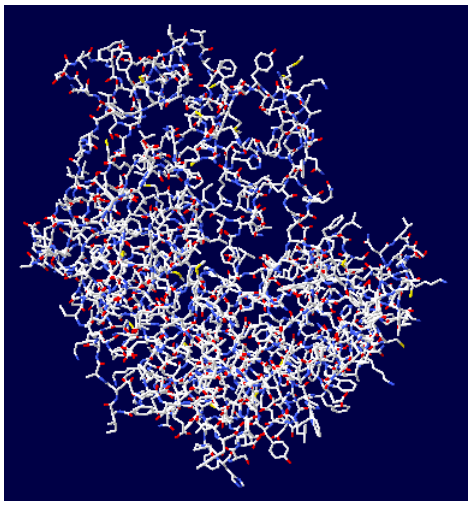


View 7: *Dissimilarities in the Gulonolactone Oxidase in Mus musculus before Energy Minimization.* The dissimilarities between the newly built gulonolactone oxidase in *Mus musculus* and the superimposed model of the two cholesterol oxidases are displayed in ribbon diagram colored by secondary structure. The dissimilar segments displayed are from the 3JS8 template since this chain extends further than the mouse model. The mouse sequence was fitted to the superimposed model of the templates leaving additional residues in the templates mismatched. The dissimilar segments contained more alpha helices than beta strands.



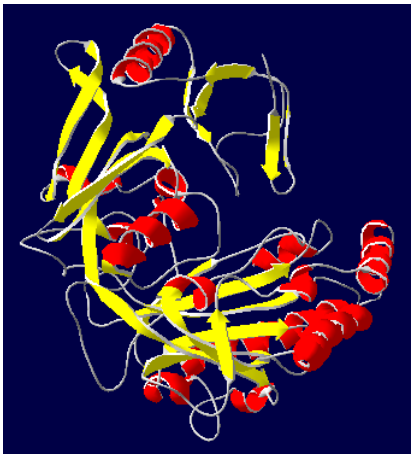
View 8: *Dissimilarities in the Context of the Entire Protein.* The dissimilarities from view 7 are displayed with respect to the entire 3JS8 template. The dissimilarities of the protein at the end of the chain are drawn in ribbon diagram (superimposed in wireframe), and the rest of the protein is in wireframe. The wireframe is modeled in CPK colors without showing the sidechains. The ribbon is displayed in secondary colors.

The initial structure of the newly developed protein appears to be very similar to templates. The major dissimilarities were found to be at the end of the chain. In view 5, it can be seen that the protein has relatively high threading energy. To reduce the threading energy, the model was submitted to Swiss-Model where the energy of the structure was minimized. The resulting model of the gulonolactone oxidase protein is shown in View 9.



View 9: *Final Gulonolactone Oxidase Model in Mus musculus after Energy Minimization.* Similar to view 5, the entire gulonolactone oxidase protein in *Mus musculus* is displayed in wireframe.

The wireframe is colored with CPK colors. This image is the final 3-D model of gulonolactone oxidase in *Mus musculus*.

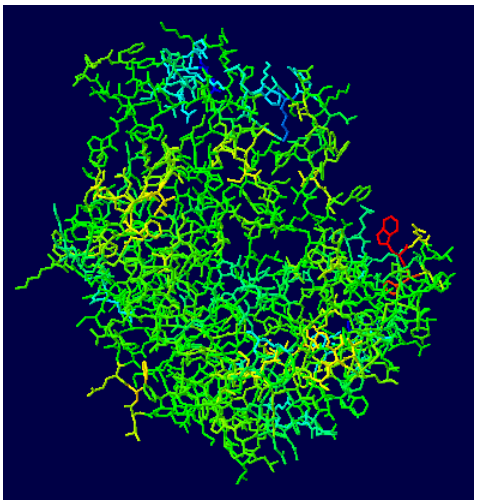


View 10: *Final Gulonolactone Oxidase Model in Mus musculus in Ribbon Diagram after Energy Minimization.* The gulonolactone oxidase protein is displayed in ribbon diagram colored by secondary structure.

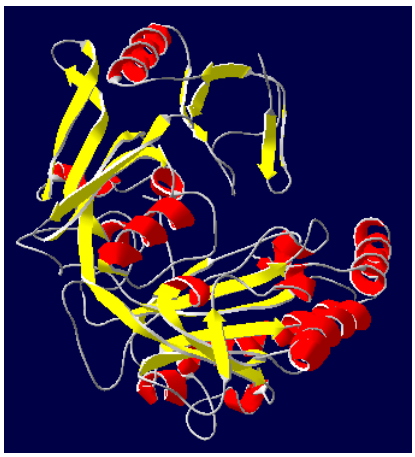
The final model for the gulonolactone oxidase protein in mice was further analyzed. It contained 440 total residues. There were 13 alpha helices (108 residues; 25%) and 18 beta strands (122 residues; 28%). Also, this protein contained 101 (23%) polar, 225 (51%) nonpolar, 45 (10%) acidic, and 69 (16%) basic residues within the protein. In comparison to the superimposed model of the two templates, there were a higher percentage of beta strands, but a lower percentage of alpha helices. The percentages of polar and acidic residues were very similar to the superimposed model of the templates. However, there was a slightly lower percentage of nonpolar residues and a higher percentage of basic residues.

Modeling Human Gulonolactone Oxidase

Designating this final model of mouse gulonolactone oxidase protein as a template, a potential nonfunctional gulonolactone oxidase model in humans could be created. By ‘magic fitting’ known human sequence to the reference mouse protein model, an initial human homolog was modeled (Views 11 and 12).

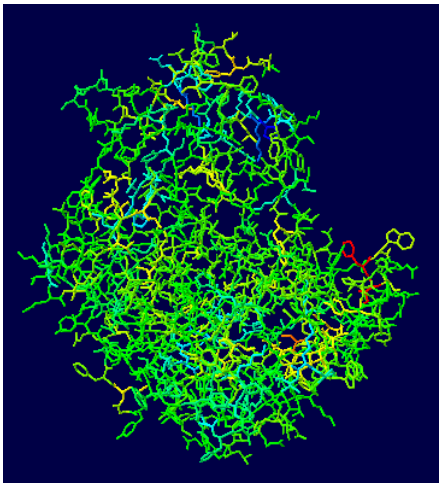


View 11: *Gulonolactone Oxidase Model in Homo sapiens before Energy Minimization.* The reference protein used was the gulonolactone oxidase in *Mus musculus*. The wireframe model is colored by threading energy. The majority of the residues are colored green indicating average energies.

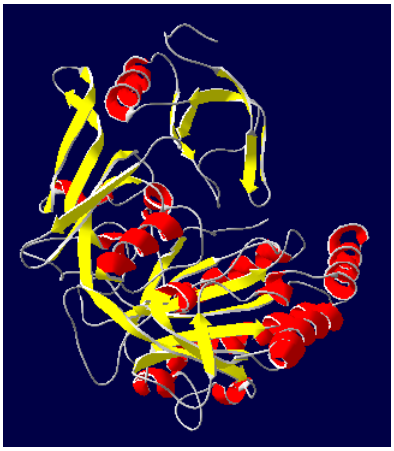


View 12: *Gulonolactone Oxidase Model in Homo sapiens before Energy Minimization in Ribbon Diagram.* The ribbon is colored by secondary structure.

The initial structure of the newly developed human model appears to be very similar to that of the mouse gulonolactone oxidase. The dissimilarity to the mouse protein was found to be the additional chain segments in the beginning (the first 46 residues) and at the end of the chain (the last 28 residues) as well as 1 additional residue in the middle. These residues are 1-46, 169, and 489-516. To reduce the amount of mismatches in modeling of the human nonfunctional gulonolactone oxidase, a new approach was taken. This time, the human sequence was ‘magic fitted’ to the superimposed model of the 3JS8 and 2I0K templates (Views 13 and 14).

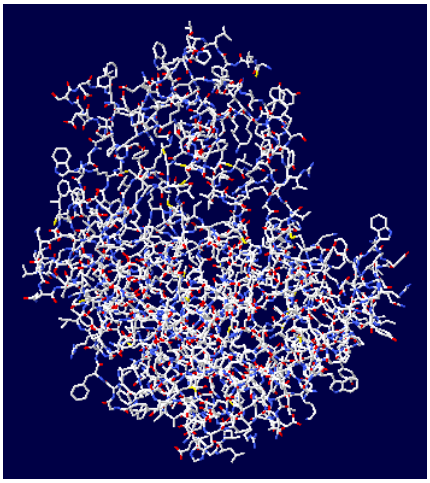


View 13: *Gulonolactone Oxidase Model in Homo sapiens with Superimposed Model of Two Templates as Reference.* The wireframe model is colored by threading energy. The majority of the residues are colored green indicating average energies.

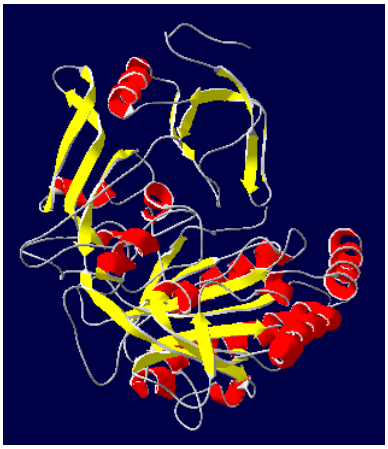


View 14: *New Gulonolactone Oxidase Model in Homo sapiens before Energy Minimization in Ribbon Diagram.* The ribbon is colored by secondary structure.

The new nonfunctional human gulonolactone oxidase model developed from the superimposed model of the 3JS8 and 2I0K templates looked very similar to both of the templates and the mouse gulonolactone oxidase model. In this model, only the first 35 additional residues were dissimilar. The amount of dissimilar residues has decreased significantly by using the superimposed model of the two templates as reference instead of the mouse gulonolactone oxidase model. View 13 shows that the protein has relatively high threading energy. To reduce this, energy minimization by Swiss-Model was completed. The potential final model of the nonfunctional gulonolactone oxidase protein in humans is shown in Views 15 and 16.



View 15: *Final Gulonolactone Oxidase Model in Homo sapiens after Energy Minimization.* Similar to View 13, the entire gulonolactone oxidase model in *Homo sapiens* is displayed in wireframe with CPK colors.



View 16: *Gulonolactone Oxidase Model in Homo sapiens after Energy Minimization in Ribbon Diagram.* The ribbon is colored by secondary structure.

The final model of nonfunctional gulonolactone oxidase in humans contained similar structural properties to the final model of the mouse gulonolactone oxidase as well as the superimposed model of the two templates.

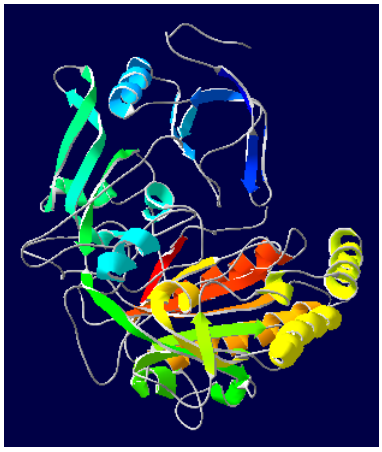
Table 2: Comparison of the nonfunctional gulonolactone oxidase in humans, the functional gulonolactone oxidase in mice, and the two templates by secondary structure, polarity and charge.

Characteristics of Four Enzymes				
Enzyme Source or Template	3JS8	2I0K	Mouse	Human
Total Residues	538	536	440	481
Alpha Helices	17	17	13	13
Beta Strands	22	22	18	19
Alpha Helix Residues	146 / 27%	144 / 27%	108 / 25%	112 / 23%
Beta Strand Residues	138 / 26%	141 / 26%	122 / 28%	116 / 24%
Polar Residues	132 / 25%	129 / 24%	101 / 23%	102 / 21%
Nonpolar Residues	310 / 58%	298 / 56%	225 / 51%	252 / 52%
Acidic Residues	44 / 8%	55 / 10%	45 / 10%	52 / 11%
Basic Residues	52 / 10%	54 / 10%	69 / 16%	75 / 16%

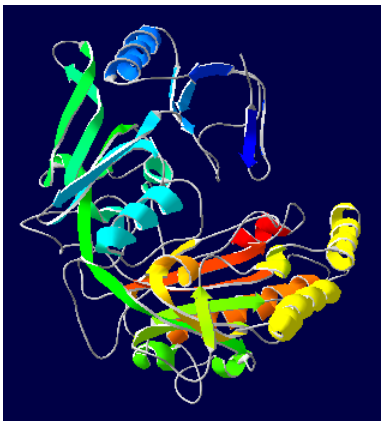
The major features of the secondary structure comparison included more beta strands but same number of alpha helices in the human model compared to the mouse model. The beta sheets present in the two templates and the mouse model were not quite the same as those in the human model. There was an additional beta strand in the beginning and end of the human protein chain. Also, the section of the mouse model that contained 5 antiparallel beta strands changed to 3 antiparallel beta strands in the human model.

Tale of Two Created Enzymes

The structural properties of the two created enzymes, the mouse and human gulonolactone oxidase models, were compared. The secondary structure succession is shown in the human and mouse proteins in Views 17 and 18, respectively.



View 17: *Gulonolactone Oxidase Model in Homo sapiens in Secondary Structure Succession.* The protein starts with a loop that leads to a beta strand in blue. The chain ends near where it starts with another beta strand, but in red.



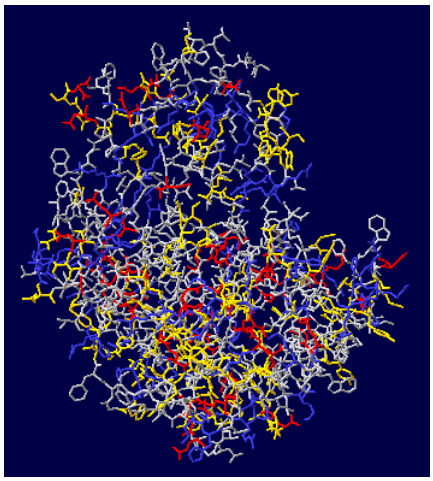
View 18: *Gulonolactone Oxidase Model in Mus musculus in Secondary Structure Succession.* The protein starts with a small loop that leads to a beta strand in blue. The chain ends near where it starts. However, it ends with an alpha helix in red.

It is clear that there is a major structural difference near the end of the chain as mentioned earlier (see Views 19 and 20).

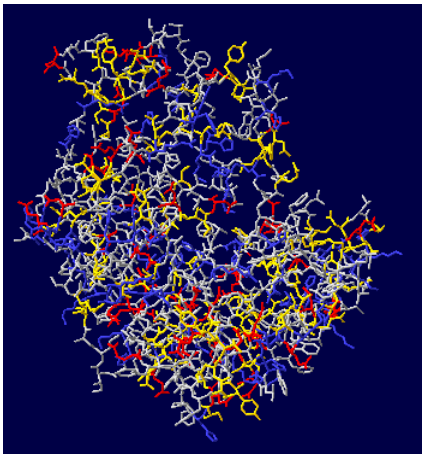
View 19: *End of Protein Chain in Nonfunctional Gulonolactone Oxidase Model in Homo sapiens.* This section of the protein contains residues 427-516. It is displayed in ribbon diagram colored by secondary structure. It has 3 beta strands and 1 alpha helix.

View 20: *End of Protein Chain in Gulonolactone Oxidase Model in Mus musculus.* This section of the protein contains residues 381-440. It is displayed in ribbon diagram colored by secondary structure. It has 2 beta strands and 1 alpha helix.

Comparison of Views 19 and 20 demonstrates some obvious differences. The human protein contains 1 additional beta strand and an extended alpha helix that the mouse protein does not. The polarities of the two compounds differ slightly as well. The comparison in polarity is shown in Views 21 and 22 for the human and mouse proteins, respectively.



View 21: *Gulonolactone Oxidase Model in Homo sapiens by Type.* The entire protein is displayed and colored by type. The nonpolar residues are grey, the polar residues are yellow. The acidic and the basic residues are red, and blue, respectively. The majority of the residues are nonpolar and are mostly located near the center of the protein.



View 22: *Gulonolactone Oxidase Model in Mus musculus by Type.* The entire protein is displayed and colored by type. The majority of the residues are nonpolar and are mostly located near the

center of the protein. There are slightly less nonpolar and acidic residues in *Mus musculus* than in *Homo sapiens*. However, *Mus musculus* contains slightly more polar residues than *Homo sapiens*. The amount of basic residues is about equal in both models.

Discussion

The overall polarities of the human and mouse protein models were very similar. Each protein was about 22% polar, 52% nonpolar, 11% acidic, and 16% basic. However, there appears to be a few differences at the potential active site. It is assumed that the active site is located at the crevice on the surface. In the mouse model's active site, there are some polar residues that are likely involved in enzyme-substrate interaction. Some of these polar residues in the active site of the mouse model are GLN 9, ASN 12, and SER 57. It appears that these polar residues mutated to LEU 44 (nonpolar), ASP 47 (acidic), and THR 92 (polar), respectively, in the human model. Mutations from polar residues to acidic or nonpolar residues could affect the functionality of the enzyme greatly that might lead to an inactive enzyme.

The shape of the active site is also different in the human and mouse proteins. Both ends of the protein chain in the mouse gulonolactone oxidase appear to have extended outward in the human model by additional residues. At the end of the chain, these additional residues extend toward the active site. The additional 35 residues at the beginning of the human chain were not displayed due to the mismatch with the superimposed model of the two templates. It is possible that these additional residues at the beginning of the chain may also extend further toward the active site. These extensions to the active site can interfere with the entrance of the substrate to the crevice and/or leave less room for substrate binding which could lead to a nonfunctional protein.

There was some difficulty in the procedure, however. When finding the human protein sequence closest to the mouse protein sequence, it was difficult to choose which sequence to use. The best result from the BLAST search was used, but it may not be the actual sequence of nonfunctional gulonolactone oxidase in humans. Also, the choice of templates involved a similar situation. The two best results from Swiss-Model may not have been the most appropriate templates for the mouse protein. When modeling the human protein, there was some difficulty as well. It was assumed that the human protein should be modeled from the mouse protein as they are closely related. However, this led to a significant amount of mismatch as discussed earlier. There were many residues that were left out of the model. When the human protein was modeled from the same templates that the mouse protein was modeled from, the error due to the missing residues decreased. Therefore, the human version modeled from the superimposed model of the two templates was chosen for further use. However, the real nonfunctional gulonolactone oxidase could be more similar to the model based on the mouse gulonolactone oxidase.

Overall, the human nonfunctional gulonolactone oxidase and the mouse gulonolactone oxidase were successfully modeled using two template models. These models were cholesterol oxidase from *Chromobacterium* sp. DS-1 (CHOLOX) (3JS8.1.A) and cholesterol oxidase from *Brevibacterium sterolicum* (2I0K.1.A). The 3JS8 protein was the main reference protein during the modeling process. The created model enzymes for the mouse and human were compared to each other to determine why the human enzyme has evolved into a nonfunctional form. This work suggests that the two major reasons for the nonfunctionality of the human enzyme are the mutations and the shape of the active site. Polar residues that have mutated to nonpolar and acidic residues could change the enzyme-substrate interactions tremendously. The active site of the human enzyme model also appears to have less room for the substrate entrance due to a longer chain in the human enzyme. In conclusion, the active site structure of the gulonolactone oxidase enzyme seems to be altered enough to disrupt the enzymatic function completely, causing the enzyme to become inactive in humans.

References

Ishida, E.; Hashimoto, K.; Okada, S.; Satoh, T.; Yamada, M.; Mori, M. "Thyroid hormone receptor and liver X receptor competitively up-regulate human selective Alzheimer's disease indicator-1 gene expression at the transcriptional levels." *Biochem. Biophys. Res. Commun.*, **2013**, *432*, 513-518.

Jacob, R. A.; Sotoudeh, G. "Vitamin C function and status in chronic disease." *Nutr. Clin. Care*, **2002**, *5*, 66-74.

Jenness, R.; Birney, E. C.; Ayaz, K. L.; Buzzell, D. M. "Ontogenetic development of L-gulonolactone oxidase activity in several vertebrates." *Comp. Biochem. Physiol. B: Comp. Biochem.*, **1984**, *78*, 167-173.

Kim, H.; Bae, S.; Kim, Y.; Cho, C. H.; Kim, S. J.; Kim, Y. J.; Lee, S. P.; Kim, H. R.; Hwang, Y. I.; Kang, J. S.; Lee, W. J. "Vitamin C prevents stress-induced damage on the heart caused by the death of cardiomyocytes, through down-regulation of the excessive production of catecholamine, TNF-α, and ROS production in *Gulo (–/–)*$^{\text{Vit C-Insufficient}}$ mice." *Free Radical Biol. Med.*, **2013**, *65*, 573-583.

Lim, L.; Molla, G.; Guinn, N.; Ghisla, S.; Pollegioni, L.; Vrielink, A. "Structural and kinetic analyses of the H121A mutant of cholesterol oxidase." *Biochem. J.*, **2006**, *400*, 13-22.

Nishikimi, M.; Yagi, K. "Molecular basis for the deficiency in humans of gulonolactone oxidase, a key enzyme for ascorbic acid biosynthesis." *Am. J. Clin. Nutr.*, **1991**, *54*, 1203S-1208S.

Sagermann, M.; Ohtaki, A.; Newton, K.; Doukyu, N. "Structural characterization of the organic solvent-stable cholesterol oxidase from Chromobacterium sp. DS-1." *J. Struct. Biol.*, **2010**, *170*, 32-40.