Elucidating the pathogenicity of missense variants in the nucleotide-binding and transmembrane protein domains of ABCG5 associated with sitosterolemia

Authors: Heiberger, Helena E., Stenger, Cynthia L., Morris, Michele, Truitt, Jared, and Terwilliger, Luke

Source: BIOS, 95(4) : 251-258

Published By: Beta Beta Beta Biological Society

URL: https://doi.org/10.1893/BIOS-D-24-00001

BIOS

Research Article
https://doi.org/10.1893/BIOS-D-24-00001

Elucidating the pathogenicity of missense variants in the nucleotide-binding and transmembrane protein domains of ABCG5 associated with sitosterolemia

Helena E. Heiberger[1], Cynthia L. Stenger, PhD[2], Michele Morris, PhD[3], Jared Truitt[4], and Luke Terwilliger[4]

[1]Department of Biology, University of North Alabama, Florence, AL, 35633
[2]Department of Mathematics, University of North Alabama, Florence, AL, 35633
[3]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806
[4]Department of Computer Science, University of North Alabama, Florence, A, 35633

Correspondence to: clstenger@una.edu

Abstract

Consuming vegetables poses a challenge for individuals with sitosterolemia, a rare autosomal recessive genetic disorder characterized by heightened intestinal absorption and decreased biliary excretion of plant sterols. Pathogenic mutations in the ATP Binding Cassette Subfamily G Member 5 (*ABCG5*) and ATP Binding Cassette Subfamily G Member 8 (*ABCG8*) genes result in sitosterolemia. *ABCG5* and *ABCG8* form the ATP-binding cassette transporter protein ABCG5/ABCG8 that functions to efflux plant sterols and cholesterol from the liver and small intestine. This study seeks to predict the pathogenicity of the missense swaps G91E, F399C, R419C, and R419G. These variants were selected based on specific placement within the functional domains of ABCG5. Pathogenicity scores were compared to two pathogenic variants using Mutation Assessor, MetaLR, REVEL, CADD, PolyPhen, and SIFT. ConSurf predicted the amino acid position of each missense swap to be conserved, buried, structural, or exposed. Molecular dynamics simulations revealed differences in movement between the variants and the WT. Our results predict that these variants are pathogenic regarding sitosterolemia. These findings contribute to the understanding of genetic factors influencing sitosterolemia and underscore the importance of further investigations to elucidate the clinical implications of these variants for improved diagnostic and therapeutic strategies in managing this rare genetic disorder.

Keywords: in silico pathogenicity prediction; nucleotide-binding domain; premature heart atherosclerosis; transmembrane domain; variants of uncertain significance; xanthoma

How to Cite: Heiberger HE, Stenger CL, Morris M, Truitt J, Terwilliger L. 2024. Elucidating the pathogenicity of missense variants in the nucleotide-binding and transmembrane protein domains of ABCG5 associated with sitosterolemia. Bios 95:251–258. DOI:10.1893/BIOS-D-24-00001.

Received 9 January 2024; revised 4 October 2024; accepted 7 November 2024.

Introduction

You might have heard the common phrase, "eat your vegetables." However, approximately 1 in 50,000 people struggle to digest agricultural products due to a rare autosomal recessive genetic disorder called sitosterolemia (Farzam and Morgan 2023). Sitosterolemia is characterized by increased intestinal absorption and decreased biliary excretion of plant sterols and cholesterol (Park et al. 2014). Plant sterols, also called phytosterols, are cholesterol-like molecules that are important components of the cell membrane in plants (Du et al. 2022). Patients with sitosterolemia illustrate extreme phenotypic heterogeneity ranging from asymptomatic to premature cardiac death (Yoo et al. 2016). Symptoms can include xanthoma, premature heart atherosclerosis, hematologic manifestations, and high plant sterol levels in the blood.

It is important to diagnose people with this disorder because symptoms can worsen over time (Yoo et al. 2016). However, there is a similar and more common disease called familial hypercholesterolemia that leads to clinical misdiagnosis and confusion for physicians. Genetic testing for pathogenic mutations in the ATP Binding Cassette Subfamily G Member 5 (*ABCG5*) and ATP Binding Cassette Subfamily G Member 8 (*ABCG8*) genes can diagnose people with sitosterolemia. *ABCG5* and *ABCG8* form the ATP-binding cassette transporter protein ABCG5/ABCG8 that functions to efflux plant sterols and cholesterol from the liver and small intestine. Evidence suggests that genetic mutations in *ABCG5* are more prevalent in those of Asian descent, but especially in Chinese or Japanese patients

(Shen et al. 2022), while Caucasians commonly exhibit mutations in *ABCG8* (Tada et al. 2018). *ABCG5* encodes a protein called ABCG5 and *ABCG8* encodes ABCG8. ABCG5 and ABCG8 combine to form the ABCG5/ABCG8 heterodimer sterol efflux transporter protein. ABCG5/ABCG8 is co-expressed in the apical membrane of enterocytes in the small intestine and the canalicular membrane of hepatocytes in the liver to prevent the accumulation of dietary sterols (Karpen and Karpen 2017). ABCG5/ABCG8 is an ATP-binding cassette transporter that hydrolyzes ATP to power the excretion of neutral sterol molecules (Zein et al. 2019). Members of this protein family share a common structure including two-nucleotide binding domains that mediate ATP binding and hydrolysis and two transmembrane domains that enable the substrate to be translocated across the membrane (Procko et al. 2009). ABCG5 contains six transmembrane domains (Lee et al. 2001) and a nucleotide-binding domain. The nucleotide-binding domain contains the Walker A motif, Walker B motif, and Signature motif that are important for nucleotide binding and hydrolysis (Higgins 1992; Zhang et al. 2006). Mutations in functional residues in the Walker A and Walker B motifs have prevented biliary sterol secretion (Zein et al. 2019; Zhang et al. 2006) while mutations in the transmembrane domain have hindered mature ABCG5 presence in the cell (Graf et al. 2004).

Numerous variants of uncertain significance (VUS) are situated within the nucleotide-binding and transmembrane domains, remaining elusive in their classification as either pathogenic or benign. Consequently, their exclusion from the realm of clinical diagnosis for sitosterolemia underscores the challenges posed by their ambiguous nature. The goal of this project was to assess the relative pathogenicity of VUS in these functional regions to further understand critical residues in these domains and to improve the diagnosis of sitosterolemia. We hypothesize that the VUS G91E, F399C, R419C, and R419G in ABCG5 are pathogenic due to their location in these protein domains.

Methods

Although mutations in both the *ABCG5* and *ABCG8* genes are implicated in sitosterolemia, *ABCG5* was selected because it has 136 clinical submissions, while *ABCG8* has a smaller frequency of 107 submissions on SimpleClinvar (Perez-Palma et al. 2019). Four VUS of ABCG5 were chosen based on their location in functional regions in the protein. To determine VUS that are in functional regions in ABCG5, we selected the Q9H222-1 isoform and observed various features on UniProt (The UniProt Consortium 2023). The Family & Domains section on UniProt lists the amino acid position location of the ABC transporter domain (nucleotide-binding domain) and the ABC transmembrane type-2 domain. The nucleotide-binding domain includes the Walker A motif, Walker B motif, and Signature motif (Higgins 1992; Zhang et al. 2006). The Walker A motif has the consensus sequence "GXXGXGKT" where X is a random amino acid (Lee et al. 2016; The UniProt Consortium, 2023; Zhang et al. 2006). We aligned the Walker A sequence with the Q9H222-1 FASTA file from UniProt to determine its location in the protein. UniProt provides the ATP-binding site sequence as GSSGSGKT, which corresponds to the Walker A motif location. Furthermore, UniProt predicted the ATP-binding site according to rules designated by PROSITE with the accession code PRU00434 (Sigrist et al. 2013). We selected G91E, F399C, R419C, and R419G for this study because they are in these functional domains according to UniProt. Specifically, G91E was determined to be in the ATP-binding site in the nucleotide-binding domain while F399C, R419C, and R419G are in the ABC transmembrane type-2 domain.

We used UniProt to determine the Reference SNP ID (rsID) for R419C (rs771475759), R419G (rs771475759), F399C (rs1369215333), and G91E (rs749587717). We then searched for the rsID for each VUS on Ensembl (Ensembl, 2023, Version 109) to retrieve pathogenicity scores from these in silico tools: Sorting Intolerant From Tolerant (SIFT; Ng and Henikoff 2003), PolyPhen-2 (Adzhubei et al. 2010), Combined Annotation Dependent Depletion (CADD; Kircher et al. 2014), Rare Exome Variant Ensemble Learner (REVEL; Ioannidis et al. 2016), and MetaLR (Dong et al. 2015). The pathogenicity scores of the VUS were compared to the pathogenicity scores of two pathogenic variants, R419P and R419H. ConSurf was used to estimate the evolutionary rate of the amino acid positions in the ABCG5 to determine critical residues in the protein structure (Ashkenazy et al. 2016). To do this, the PDB structure of ABCG5 called 5DO7 retrieved from UniProt was submitted to ConSurf using default parameters.

Homology protein models were generated to visualize the ABCG5 wild-type (WT) and VUS. In YASARA (Version 23.9.29), we conducted homology modeling of ABCG5 using a FASTA file called Q9H222.FASTA that was retrieved on UniProt from the Q9H222-1 isoform. UniProt selected this isoform to be the canonical sequence. The generated WT ABCG5 homology model contained 651 residues. We used YASARA to color the protein based on functional domains for visualization and generated the variant proteins by swapping amino acids corresponding to the VUS, which generated Protein Data Bank (PDB) files. These PDB files were cleaned using the em_runclean macro. Energy minimization was conducted in YASARA on the WT ABCG5 and VUS ABCG5 homology models. The energy minimized homology models were used to conduct molecular dynamic simulations to simulate the protein movement in the cellular environment to infer how the VUS could be impacting protein structure. MDS were conducted using YASARAstructure program (Version 23.9.29; Krieger et al. 2009; Krieger and Vriend 2014; Krieger and Vriend

2015) in the AMBER14 force field (Hornak et al. 2006) under default parameters and the md_analyze macro to obtain the data. The modeling simulation placed the energy minimized homology model in a non-membrane bound aqueous environment where movement was measured in global fluctuation (RMSF) and global deviation (RMSD) in angstroms (Å) over a period of 20 ns. One trial was conducted for the ABCG5 WT and each variant to assess differences in RMSF and RMSD between the WT and the VUS.

Results

The objective of this study was to find variants of uncertain significance in the ABCG5 protein associated with sitosterolemia to improve diagnosis of this rare genetic disorder. All four VUS chosen in this study, G91E, F399C, R419C, and R419G, are in functional domains in ABCG5. In the Family & Domains section, UniProt lists the ABC transporter domain (nucleotide-binding domain) at amino acid positions 52-293 and ABC transmembrane type-2 domain at 388-645 (The UniProt Consortium 2023). R419C, R419G, and F399C are in the ABC transmembrane type-2 domain colored in green while G91E is located in the ATP-binding site in the Walker A motif within the ABC transporter domain colored in blue (Fig. 1I). The structures of the missense swaps were also illustrated (Fig. 1II).

Comparison of the in silico pathogenicity predictors demonstrate that the VUS had similar pathogenicity scores to the known pathogenic variants (Fig. 2). Scores close to 1 indicate high likelihood of pathogenicity or deleterious impact. For SIFT, the VUS and pathogenic variants all had a score close to 1. For PolyPhen-2, all variants had a score above 0.900 except for G91E with a score of 0.743. For CADD, all variants had a score above 0.90. For REVEL, R419C, R419G, R419H, and R419P had a score between 0.600 and around 0.700. F399C had a lower score of 0.423 while G91E had the highest score of 0.974. For MetaLR, all variants had scores around 0.600 except G91E had a much higher score of 0.996. For Mutation Assessor, R419C, R419G, R419H, and R419P had a score around 0.700. F399C had a lower score of 0.56 while G91E had a much higher score of 0.998.

ConSurf color coded the ABCG5 protein based on level of conservation (Fig. 3A). Additionally, ConSurf illustrated that the 4 VUS were located at conserved amino acid positions in ABCG5 (Fig. 3B). G91E, R419C, and R419G were located at amino acid positions containing a high conservation score of 9, and were characterized as buried and structural. In contrast, F399C was located at an amino acid position containing a conservation score of 7 and was characterized as exposed.

MDS simulated the movement of the WT and variant ABCG5 proteins in a non-membrane bound aqueous environment and quantified this movement in RMSD and RMSF in angstroms (Å) over 20 ns (Fig 4). There were slight differences in RMSD between R419 and R419C (Fig. 4A), F399 and F399C (Fig. 4B), G91 and G91E (Fig. 4C). R419 and R419C had alternating higher and lower RMSD throughout the simulation. G91 and G91E also had alternating higher and lower RMSD throughout the simulation. F399C had a general trend of higher RMSD compared to F399. The RMSF graphs illustrated minor differences of the WT ABCG5 compared to R419C (Fig. 4D), F399C (Fig. 4E), and G91E (Fig. 4F). The arrows pointing at the amino acid positions of the VUS indicated a slight difference from the WT

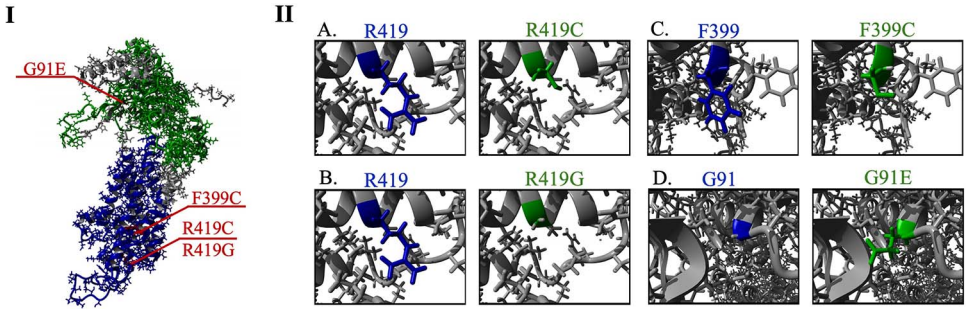
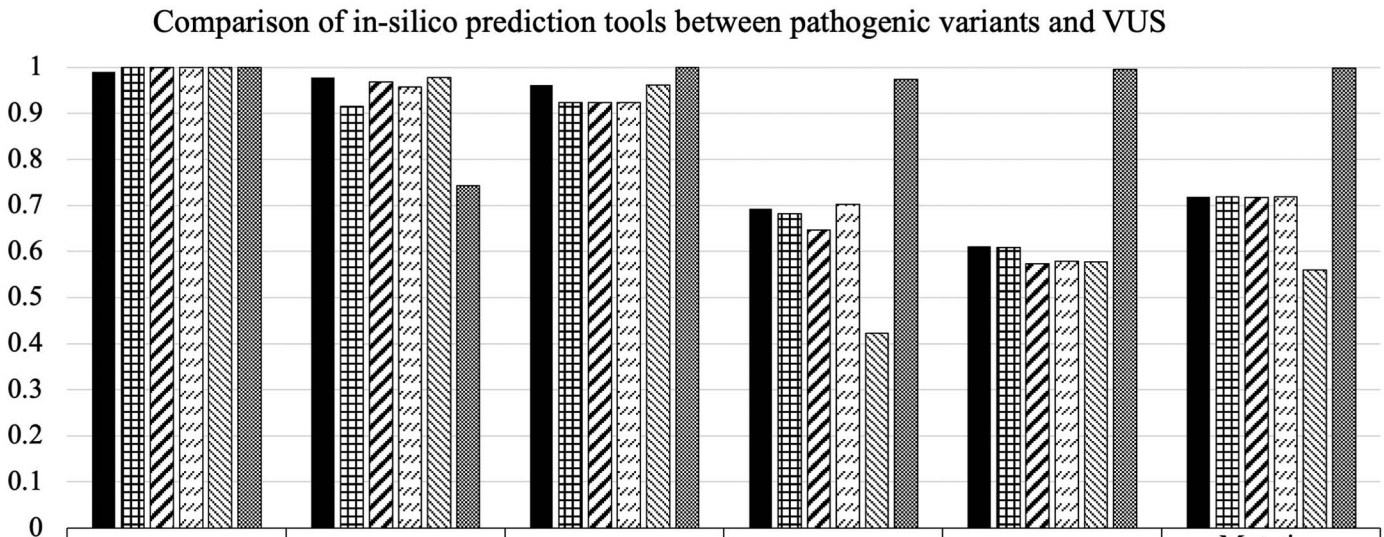


Figure 1. YASARA ray-traced hires screenshots of a slow homology model of the ABCG5 protein. **I.** The nucleotide-binding domain is colored in green while the transmembrane domain is in blue. The G91E VUS location is indicated in the nucleotide-binding domain. The F399C, R419C, and R419G locations are indicated in the transmembrane region. **II.** YASARA was used to visualize the slow homology model of the ABCG5 WT and VUS. WT amino acids are colored blue and variant amino acids are colored green. **A**. On the left, WT arginine is located at the 419 amino acid position. On the right, arginine is swapped to a cysteine and is colored green. **B.** On the left, WT arginine is located at the 419 amino acid position. On the right, arginine is swapped to a glycine. **C.** On the left, WT phenylalanine is located at the 399 amino acid position. On the right, phenylalanine is swapped to a cysteine. **D.** On the left, WT glycine is located at the 91 amino acid position. On the right, glycine is swapped to a glutamic acid.



	SIFT	PolyPhen-2	CADD	REVEL	MetaLR	Mutation Assessor
R419C	0.99	0.978	0.96	0.693	0.611	0.718
R419G	1	0.915	0.92	0.682	0.609	0.718
R419H	1	0.968	0.92	0.646	0.573	0.717
R419P	1	0.957	0.92	0.702	0.579	0.718
F399C	1	0.978	0.96	0.423	0.578	0.56
G91E	1	0.743	1	0.974	0.996	0.998

Figure 2. Graph and chart containing in silico pathogenicity prediction results from SIFT, PolyPhen-2, CADD, REVEL, MetaLR, and Mutation Assessor. The scores of CADD and SIFT were normalized for comparison. The pathogenicity scores of the VUS R419C, R419G, F399C, and G91E were compared to the pathogenic variants R419P and R419H. A pathogenicity score close to 1 predicts a higher possibility of the variant being pathogenic.

protein. At these arrows, there was a difference in RMSF of 0.173 Å for R419 and R419C, a difference of 0.399 Å for F399 and F399C, and a difference of 0.292 Å for G91 and G91E. G91E had two regions with large differences in fluctuation from the WT. For G91 and G91E, there was a 2.316 Å difference at amino acid position 58, and there was a 5.818 Å difference at amino acid position 571. Comparatively, R419C and F399C did not have regions with this large of a difference.

Discussion

Sitosterolemia is a rare disorder and is often overshadowed by a more common disorder, familial hypercholesterolemia, which can make it difficult for patients to obtain clinical diagnosis. However, classifying VUS as pathogenic or benign can help with the diagnosis and treatment of rare disorders like sitosterolemia. In this study, we assessed the pathogenicity of the VUS G91E, F399C, R419C, and R419G of ABCG5. The VUS were selected based on location in important functional regions in the ABCG5 protein.

The pathogenic variants, R419P and R419H, are located at the apex of the transmembrane helix 2 (Zein et al. 2019). Two of the VUS in our study, R419C and R419G, are located at the same amino acid position as these pathogenic variants. R419P and R419H were found to hinder mature ABCG5 from being present in the cell (Graf et al. 2004). It is possible that R419C and R419G could have a similar outcome. According to the category PS1 of the American College of Medical Genetics and Genomics standards and guidelines, there is a "strong evidence of pathogenicity" when there is a "same amino acid change as a previously established pathogenic variant regardless of nucleotide change" (Richards et al. 2015). As such, this information paired with our results indicates that these two VUS are likely pathogenic.

F399C is also located in a transmembrane domain according to UniProt. Although no pathogenic variants are located at position 399, F399C is near the already classified pathogenic variant R389H. R389H is the most common mutation in ABCG5 and has been found only in Japanese patients (Lu et al. 2001). Furthermore, no mature ABCG5 was determined in cells expressing R389H (Graf et al. 2004). If F399C has a similar effect on ABCG5, this could hinder sterol transportation.

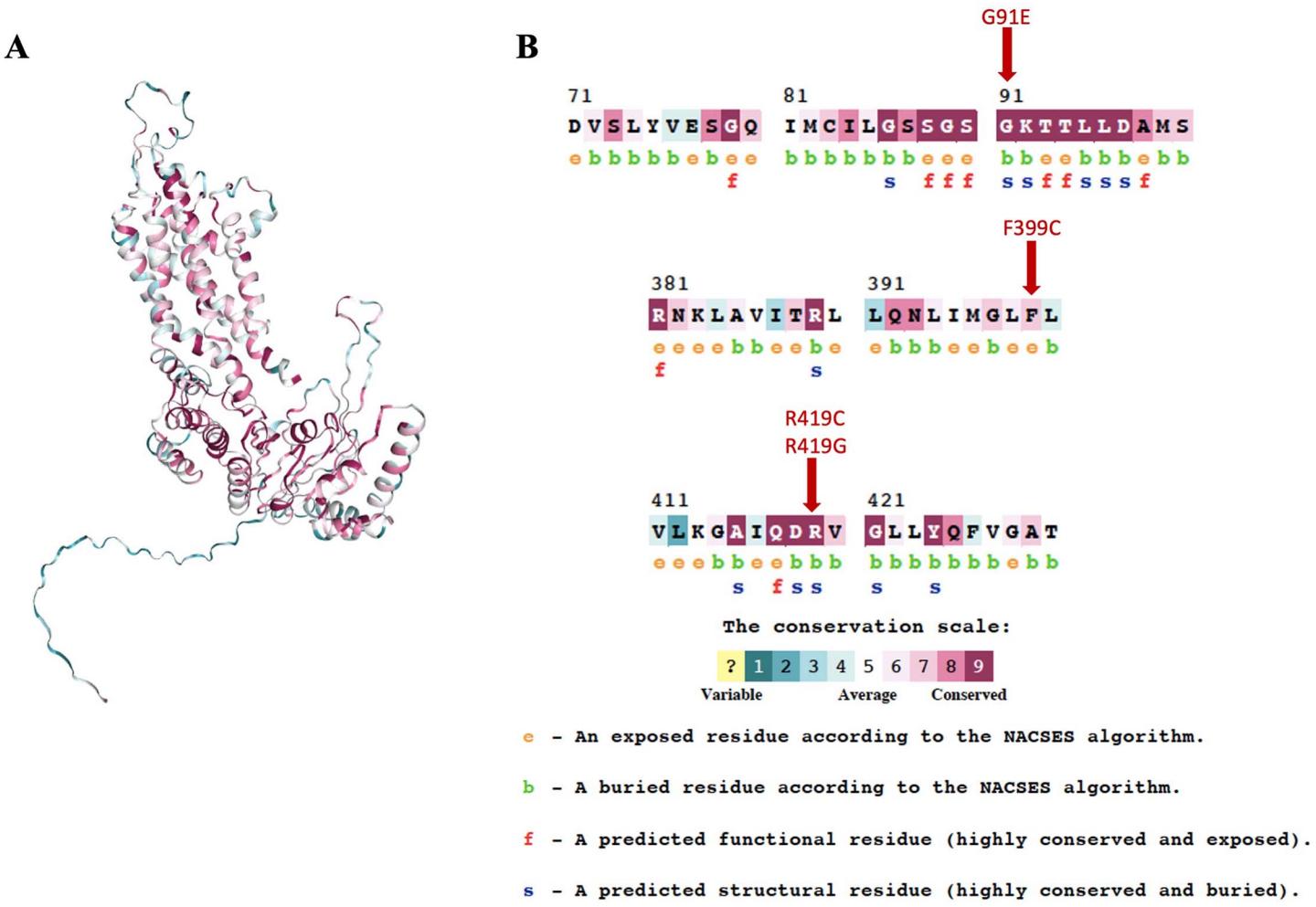


Figure 3. ConSurf analysis of ABCG5. **A.** 3D model of ABCG5 protein color-coded based on level of conservation. **B.** Amino acid positions in ABCG5 are colored based on level of conservation and labeled based on residue characteristics. The location of 4 VUS is indicated with red arrows.

G91E is located at the ATP-binding site in the Walker A motif within the nucleotide-binding domain of ABCG5 according to UniProt (The UniProt Consortium 2023). Mutations in Walker A have resulted in the prevention of biliary sterol secretion (Zein et al. 2019; Zhang et al. 2006). Given that the nucleotide-binding domain is responsible for ATP hydrolysis, which provides the energy required for sterol transport, it is reasonable to expect that mutations in this region could impair ABCG5 functionality. Specifically, the G91E mutation in the Walker A motif could disrupt ATP binding, thereby inhibiting hydrolysis and energy acquisition necessary for sterol transport.

The VUS had similarly high pathogenicity scores with already classified pathogenic variants (Fig. 2). This result indicates that the VUS are predicted to be pathogenic. Intriguingly, G91E had higher scores for CADD, REVEL, MetaLR, and Mutation Assessor. This suggests that G91E is highly predicted to be disease-causing, which is consistent with its location in the critical ATP-binding site.

ConSurf has been shown to be an accurate tool for estimating the evolutionary rates of amino acids within proteins (Chorin et al. 2020; Mitchell et al. 2023). Missense mutations at conserved amino acid positions can hinder protein stability and function. We utilized ConSurf to estimate the evolutionary rates of the VUS, aiming to determine whether these missense variants could be pathogenic for sitosterolemia (Fig. 3). ConSurf revealed that the VUS are located at conserved residues in ABCG5, supporting the likelihood that these VUS are pathogenic.

MDS simulations of the ABCG5 WT and VUS proteins were performed an aqueous environment to predict how protein function might be impacted in a cellular context (Fig. 4). The MDS results illustrated minor fluctuations in movement, which may correspond to disruptions in normal biological function. Compared to the other VUS, G91E exhibited higher RMSF values at certain amino acid positions, which is consistent with its higher pathogenicity scores (Fig 3). However, while computationally expensive and time-consuming, additional MDS would be valuable for replicating the data and further assessing pathogenicity.

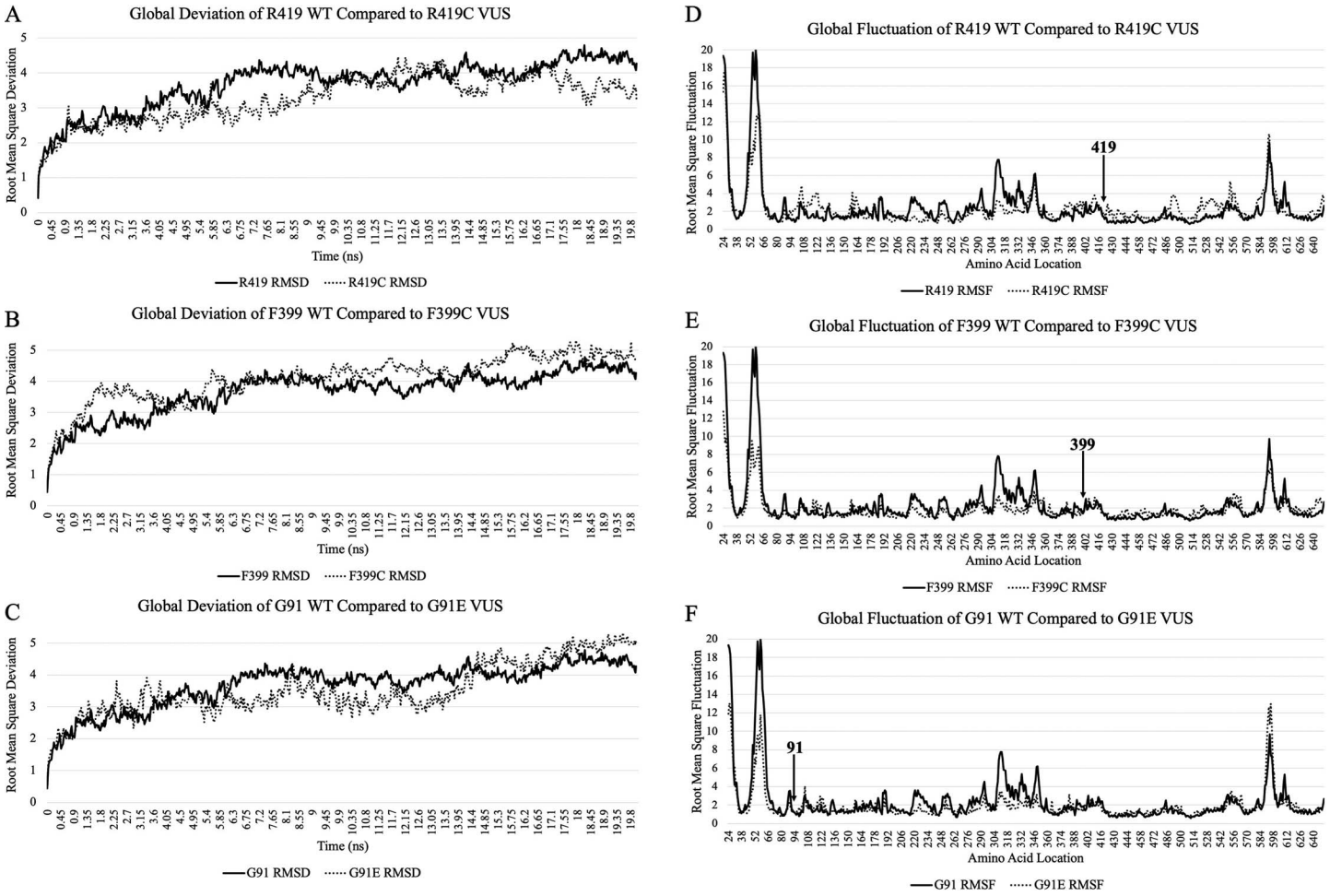


Figure 4. Graphs of molecular dynamics simulations. **A.** Root mean squared deviation (RMSD) measured in angstroms (Å) over 20 ns of R419 WT compared to R419C VUS. **B.** RMSD over 20 ns of F399 WT compared to F399C VUS. **C.** RMSD over 20 ns of G91 WT compared to G91E VUS. **D.** Root mean squared fluctuation (RMSF) measured in angstroms (Å) of each amino acid throughout the 20 ns simulation of the R419 WT compared to the R419C VUS. **E.** RMSF of each amino acid throughout the 20 ns simulation of the F399 WT compared to the F399C VUS. **F.** RMSF of each amino acid throughout the 20 ns simulation of the G91 WT compared to the G91E VUS.

Our findings and the location of these VUS in functional domains supports our hypothesis that these VUS are likely pathogenic regarding sitosterolemia. Moreover, these discoveries enhance our comprehension of the genetic elements that impact sitosterolemia end emphasizes the need for additional research to clarify the clinical significance of these variants. Such research could improve diagnostic and therapeutic approaches for effectively managing this rare genetic disorder.

Funding

This work is supported by the National Science Foundation under Grant Number 2120918. Any opinions, finding, conclusions, or recommendations expressed here are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. Nat Methods. 7(4):248–249. doi:10.1038/nmeth0410-248.

Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Research. 44(W1):W344–W350. doi:10.1093/nar/gkw408.

Ben Chorin A, Masrati G, Kessel A, Narunsky A, Sprinzak J, Lahav S, Ashkenazy H, Ben-Tal N. 2020. ConSurf-DB: An accessible repository for the evolutionary conservation patterns of the majority of PDB proteins. Protein Sci. 29(1):258–267. doi:10.1002/pro.3779.

Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Austine-Orimoloye O, Azov AG, Barnes I, Bennett R, et al. 2022. Ensembl 2022. Nucleic Acids Res. 50(D1):D988–D995. doi:10.1093/nar/gkab1049.

Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, Liu X. 2015. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet. 24(8):2125–2137. doi:10.1093/hmg/ddu733.

Du Y, Fu X, Chu Y, Wu P, Liu Y, Ma L, Tian H, Zhu B. 2022. Biosynthesis and the Roles of Plant Sterols in Development and Stress Responses. Int J Mol Sci. 23(4):2332. doi:10.3390/ijms23042332.

Farzam K, Morgan RT. 2024. Sitosterolemia (Phytosterolemia). In: StatPearls. Treasure Island (FL): StatPearls Publishing. [accessed 2024 Sep 14]. http://www.ncbi.nlm.nih.gov/books/NBK572142/.

Graf GA, Cohen JC, Hobbs HH. 2004. Missense mutations in ABCG5 and ABCG8 disrupt heterodimerization and trafficking. J Biol Chem. 279(23):24881–24888. doi:10.1074/jbc.M402634200.

Higgins CF. 1992. ABC Transporters: From Microorganisms to Man. Annual Review of Cell and Developmental Biology. 8 Volume (8, 1992): 67–113. doi:10.1146/annurev.cb.08.110192.000435.

Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins. 65(3):712–725. doi:10.1002/prot.21123.

Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D, et al. 2016. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am J Hum Genet. 99(4):877–885. doi:10.1016/j.ajhg.2016.08.016.

Karpen HE, Karpen SJ. 2017. 95 - Bile Acid Metabolism During Development. In: Polin RA, Abman SH, Rowitch DH, Benitz WE, Fox WW, editors. Fetal and Neonatal Physiology (Fifth Edition). Elsevier. p. 913-929.e4. [accessed 2024 Sep 14]. https://www.sciencedirect.com/science/article/pii/B9780323352147000950.

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 46(3):310–315. doi:10.1038/ng.2892.

Krieger E, Joo K, Lee Jinwoo, Lee Jooyoung, Raman S, Thompson J, Tyka M, Baker D, Karplus K. 2009. Improving physical realism, stereochemistry and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. Proteins. 77(Suppl 9):114–122. doi:10.1002/prot.22570.

Krieger E, Vriend G. 2014. YASARA View - molecular graphics for all devices - from smartphones to workstations. Bioinformatics. 30(20): 2981–2982. doi:10.1093/bioinformatics/btu426.

Krieger E, Vriend G. 2015. New ways to boost molecular dynamics simulations. J Comput Chem. 36(13):996–1007. doi:10.1002/jcc.23899.

Lee J-Y, Kinch LN, Borek DM, Wang Jin, Wang Junmei, Urbatsch IL, Xie X-S, Grishin NV, Cohen JC, Otwinowski Z, et al. 2016. Crystal structure of the human sterol transporter ABCG5/ABCG8. Nature. 533(7604):561–564. doi:10.1038/nature17666.

Lee MH, Lu K, Hazard S, Yu H, Shulenin S, Hidaka H, Kojima H, Allikmets R, Sakuma N, Pegoraro R, et al. 2001. Identification of a gene, ABCG5, important in the regulation of dietary cholesterol absorption. Nat Genet. 27(1):79–83. doi:10.1038/83799.

Lu K, Lee M-H, Hazard S, Brooks-Wilson A, Hidaka H, Kojima H, Ose L, Stalenhoef AFH, Mietinnen T, Bjorkhem I, et al. 2001. Two Genes That Map to the *STSL* Locus Cause Sitosterolemia: Genomic Structure and Spectrum of Mutations Involving Sterolin-1 and Sterolin-2, Encoded by *ABCG5* and *ABCG8,* Respectively. The American Journal of Human Genetics. 69(2):278–290. doi:10.1086/321294.

Martin FJ, Amode MR, Aneja A, Austine-Orimoloye O, Azov AG, Barnes I, Becker A, Bennett R, Berry A, Bhai J, et al. 2023. Ensembl 2023. Nucleic Acids Research. 51(D1):D933–D941. doi:10.1093/nar/gkac958.

Mitchell JT, Johnson EB. 2023. Comparison of In Vitro and In Silico Assessments of Human Galactose-1-Phosphate Uridylyltransferase Coding Variants. Cureus. 15(1):e33592. doi:10.7759/cureus.33592.

Ng PC, Henikoff S. 2003. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 31(13):3812–3814.

Park JH, Chung IH, Kim DH, Choi MH, Garg A, Yoo E-G. 2014. Sitosterolemia presenting with severe hypercholesterolemia and intertriginous xanthomas in a breastfed infant: case report and brief review. J Clin Endocrinol Metab. 99(5):1512–1518. doi:10.1210/jc.2013-3274.

Pérez-Palma E, Gramm M, Nürnberg P, May P, Lal D. 2019. Simple ClinVar: an interactive web server to explore and retrieve gene and disease variants aggregated in ClinVar database. Nucleic Acids Res. 47(W1):W99–W105. doi:10.1093/nar/gkz411.

Procko E, O'Mara ML, Bennett WFD, Tieleman DP, Gaudet R. 2009. The mechanism of ABC transporters: general lessons from structural and functional studies of an antigenic peptide transporter. FASEB J. 23(5):1287–1302. doi:10.1096/fj.08-121855.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, et al. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 17(5):405–424. doi:10.1038/gim.2015.30.

Shen M-F, Hu Y-N, Chen W-X, Liao L-S, Wu M, Wu Q-Y, Zhang J-H, Zhang Y-P, Luo J-W, Lin X-F. 2022. Clinical and Genetic Analysis of a Family With Sitosterolemia Caused by a Novel ATP-Binding Cassette Subfamily G Member 5 Compound Heterozygous Mutation. Front Cardiovasc Med. 9:887618. doi:10.3389/fcvm.2022.887618.

Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I. 2013. New and continuing developments at PROSITE. Nucleic Acids Res. 41(Database issue):D344-347. doi:10.1093/nar/gks1067.

Tada H, Nohara A, Inazu A, Sakuma N, Mabuchi H, Kawashiri M-A. 2018. Sitosterolemia, Hypercholesterolemia, and Coronary Artery Disease. J Atheroscler Thromb. 25(9):783–789. doi:10.5551/jat.RV17024.

UniProt Consortium. 2023. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 51(D1):D523–D531. doi:10.1093/nar/gkac1052.

Williams K, Segard A, Graf GA. 2021. Sitosterolemia: Twenty Years of Discovery of the Function of ABCG5ABCG8. Int J Mol Sci. 22(5):2641. doi:10.3390/ijms22052641.

E-G. 2016. Sitosterolemia: a review and update of pathophysiology, clinical spectrum, diagnosis, and management. Ann Pediatr Endocrinol Metab. 21(1):7–14. doi:10.6065/apem.2016.21.1.7.

Zein AA, Kaur R, Hussein TOK, Graf GA, Lee J-Y. 2019. ABCG5/G8: a structural view to pathophysiology of the hepatobiliary cholesterol secretion. Biochem Soc Trans. 47(5):1259–1268. doi:10.1042/BST20190130.

Zhang D-W, Graf GA, Gerard RD, Cohen JC, Hobbs HH. 2006. Functional asymmetry of nucleotide-binding domains in ABCG5 and ABCG8. J Biol Chem. 281(7):4507–4516. doi:10.1074/jbc.M512277200.

Author Biographies

H.E.H.: My undergraduate research experiences in the Characterizing our DNA Exceptions (CODE) project has prepared me for rigorous and impactful scientific research. During CODE, I had the opportunity to participate in the Scholars Transforming through Research Program under the Council for Undergraduate Research. In this program, I had the wonderful opportunity to lobby for funding for undergraduate bioinformatic research with representatives from Alabama in Washington D.C. Furthermore, I have been able to present my research findings at the HudsonAlpha CODE Student Symposium in Huntsville, Alabama and at the National Conference on Undergraduate Research.

Conflict of Interest Statement

The authors declare no conflict of interest.

Author Contributions

H.E.H.: Conceptualizations, Data curation, Formal analysis, Investigation Methodology, Writing- original draft, review and editing. C.L.S.: Project administration, Supervision, Writing – review and editing. M.M.: Writing – review and editing. J.T.: Data curation. L.T.: Writing – review and editing, Data curation.