TWO VARIANTS OF SLC6A1 ON AMINO ACID 451 (D451E AND D451G) ASSOCIATED WITH DEVELOPMENTAL DELAY AND EPILEPSY

Benjie Blair[2], Cynthia Stenger[1,*], Jared Painter[1], Jeremy W Prokop[3,4], Sara Woodley[1], Jenna Ridlen[2], and Richard Watkins[2]

[1]Department of Mathematics, University of North Alabama, UNA Box 5051, Florence, AL 35632, USA

[2]Department of Biology, Jacksonville State University, 700 Pelham Rd. N. Jacksonville, AL 36265, USA

[3,]Department of Pediatrics and Human Development, College of Human Medicine, Michigan State University, Grand Rapids, MI 49503, USA

[4,]Department of Pharmacology and Toxicology, Michigan State University, East Lansing MI 48824, USA

*Corresponding Author: Cynthia Stenger (clstenger@una.edu)

ABSTRACT

We present two case studies on the same allele and amino acid. The result of both variants is early onset of developmental delay and seizures. SLC6A1 is associated with an autosomal dominant early onset seizure and epileptic encephalopathy and intellectual disability. Genomic studies reported in ClinVar revealed two variants D451E and D451G. The amino acid substitution suggests that glycine would be predicted to be slightly more detrimental simply based on the physical differences between aspartic acid and glycine. Glutamic acid is similar to aspartic acid in structure and both are negatively charged amino acids. Structural and evolutionary assessments establish these variants represent a loss of function to the protein. Compiled metrics through custom tools on sequence, structure, and protein dynamics combined with PolyPhen2, Provean, SIFT, and Align-GVGD reveal these variants to rank in the top functional outcome changes relative to gnomAD, TOPMed, and ClinVar variants known to date. It is not known if these patients are resistant to multiple epileptic drugs; however, it has been noted that other variants in the same vicinity as these two respond better to valproic acid in controlling the seizures. This is consistent with additional groups studying SLC6A1 variants within patients.

INTRODUCTION

SLC6A1 encodes for a GABA transporter responsible for the reuptake of GABA from the interstitial space around the synapse after neural stimulation. A rare genetic disorder SLC6A1 epileptic encephalopathy is caused by a dominant, de novo mutation in SLC6A1, resulting in the loss of the transporter protein function. This results in development of seizures in early childhood with mild to moderate intellectual disability (ID) and behavioral disorders found in severe autism (Mattison, et al, 2018; Johannesen, et al, 2018; Yuan et al, 2017). Mutations in SLC6A1 have been found in association with myoclonic-atonic epilepsy, and other symptoms including language impairment, dystonia, and schizophrenia. However, the significance of many variants is currently unknown (Carvill, et al, 2015; Rees et al, 2020; Zech et al, 2017). For example, recent research has shown that although G443D is a far outlier for functional impact and structural fold

contribution, it has a critical loss-of-function role (Devries, et al, 2020). Their novel strategy is recommended for future SLC6A1 variant assessments and applied in this study.

Compiled metrics and statistics comparing these variants to all the known variants listed in ClinVar (Landrum et al., 2016), TOPMed, and GnomAD (Lek et al., 2016) as well as evolutionary comparison to 225 species has established a method for determining the impact of variants as it relates to other known samples. A 21-codon impact score statistical analysis was developed for detection and prediction of individual motif involvement in the development of disease. We utilize the YASARA set of tools (Krieger et al., 2009) to model changes in the protein.

Data was compiled for variants of known pathogenicity and variants of unknown significance (VUS) from databases of patients with a variant in this gene. The VUS D451E and D451G (Table 1) were compared to all known variants by analyzing and interpreting this data. Molecular dynamics simulations (mds) provided additional information about the variant impact on protein movement in a computationally derived cellular environment by analyzing movement trajectories.

CASE PRESENTATION

The D451G variant of SLC6A1 was submitted by HudsonAlpha Institute for Biotechnology (Huntsville, AL) and is listed as a variant of uncertain significance with myoclonic-atonic epilepsy as a de novo mutation with resulting autism spectrum disorder, moderate intellectual disability, seizures, and speech delay in a male in his early 20s. [NM_003042.4(SLC6A1):c.1352A>G (p.Asp451Gly)] (ClinVar Miner). Funding for this sequencing was obtained from National Human Genome Research Institute (NHGRI), UM1HG007301.

Variant D451E (SCV000742016 was submitted by Ambry genetics (Aliso Viejo, CA)) is listed as a variant of uncertain significance. It is also listed as a possible hereditary disease, a germline disease resulting in seizures, movement disorders, dysmorphic features, FTT/undergrowth, and hypotonia. Additionally, childhood onset of cardiovascular, dental, dermatologic, musculoskeletal, neurologic, and audiologic effects. [NM_003042.4(SLC6A1):c.1353C>G (p.Asp451Glu)] (ClinVar Miner).

There is limited availability to the case histories of these two variants. Generally, pathogenic effects from SLC6A1 are noticeable by two years of age and many of the variants are resistant to multiple types of the most current seizure medications (Devries, et al, 2020). It has been noted in close variants (G443D) that one of the oldest antiseizure drugs on the market, valproic acid, is more effective with variants of this gene (Devries, et al, 2020). Valproic acid does require significant monitoring to ensure side effects are not a confounding factor.

MATERIALS AND METHODS

Variants were assessed through our previously published sequence-to-structure-to-function workflow (Prokop et al. 2017), comparing the patient variant to all gnomAD (Lek et al., 2016), TOPMed, and ClinVar (Landrum et al., 2016) missense variants for SLC6A1. All variants were assessed with PolyPhen2 (Adzhubei et al. 2010), Provean (Choi and Chan 2015), SIFT (Ng and Henikof 2003), and Align-GVGD (Tavtigian et al. 2006). A total of 20 nanoseconds of molecular dynamics simulations were run on a lipid membrane embedded SLC6A1 protein model and both variants were modeled from this template using the AMBER03 force field (Duan et al. 2003).

RESULTS

The SLC6A1 protein model with 599 amino acids was generated and embedded within a lipid membrane (Fig. 1 A, B). The model is shown in grayscale with the 451-residue highlighted in red, all other molecular markers and membrane have been hidden to give a clear 3D image of the model (Fig. 2). For each amino acid position, the conservation score was examined for linear motif conservation. This was done using a 21-codon sliding window additive scoring system where the scores of 10 amino acids before and after each position were used to find the most conserved linear motifs within this protein. Position 451 was among the highest conservation scores (Fig. 3). Variant impact shown in the scatterplot for all gnomAD/TOPMed, and ClinVar variants reveals the D451G and D451E mutations both fall near several other ClinVar variants with high variant impact scores (Fig. 4). The impact score, shown in box plot format, demonstrates that the D451G has slightly higher impact than the D451E, yet both are near the top of the likely pathogenic to pathogenic groups (Fig. 5). The mds revealed a root mean squared fluctuation (RMSF) of variant groups where both the variants at amino acid 451 cluster with the pathogenic and likely pathogenic groups (Fig. 6). The conservation score for 225 species of animals for SLC6A1 was calculated and conservation scores for the 21-codon window (10 below and 10 above) for position 451 showed high conservation in the region surrounding 451 (position 11). This suggests that it is highly likely to induce changes in protein function if an amino acid is substituted into the site (Fig. 7). The mds data supports D451 to be critical in folding of the protein, with movement below average. The global deviation chart shows movement of D451G is above the wildtype (Fig. 8) This data supports D451G to be a loss-of-function site for SLC6A1. This is consistent with reported neurological conditions of the patients and the models. SLC6A1 has been associated with myoclonic atonic epilepsy with some variability in epilepsy types, while most all patients have intellectual disability similar to this patient (Johannesen et al. 2018).

DISCUSSION

In conclusion, we identified variants in SLC6A1, namely D451G and D451E, that were associated with seizures and developmental delay. Based on comparison to all known variants in SLC6A1, D451 was found to be an outlier for functional impact and structural fold contribution indicating a potential loss-of-function role in SLC6A1. The sequence-to-structure-to-function strategy was useful for VUS assessment and is recommended for future screening projects. Our statistical analysis of the data suggests that both 451 variants are most likely pathogenic in nature. The VUS have different physiochemical properties and the structure of D451G is less similar to the original amino acid than D451E. The case study indicates D451G was associated with patient reports of *late* development neurological disorders and seizures. In contrast, D451E is a negative charge to negative charge mutation, and is associated with *early* onset of developmental delay and seizures. Overall, analysis suggests mutations at position 451 play a role in the development of the SLC6A1 epileptic encephalopathy with myoclonic-atonic seizures disorder.

The fact there is such a difference in the amino acids and variations across the length of the protein gives reason to conjecture that mutation at this site most likely hinders the function in some way. Additionally, it is conjectured that the transmembrane domains are aligned very specifically and any change in alignment of these domains will result in lower transport of the normal molecules. However, more research must be conducted to confirm these hypotheses. Now that the simulation has been developed it may be possible to model potential drug-ligand interactions to determine which drugs may be more effective with certain groups of variants. More work needs to be

completed before stringent protocols for determining the VUS status of this gene can be developed, but this study represents a step toward establishing that protocol.

ADDITIONAL INFORMATION

No IRB statement or ethics are of concern since all information was obtained from publicly available databases and the information was already FERPA certified.

No funding was provided for this analysis.

AUTHOR CONTRIBUTIONS

Benjie Blair and Cynthia Stenger were prime lead geneticists on the project. Jared Painter was responsible for molecular dynamics simulations. Jeremy W. Prokop was a mentor on all bioinformatic variant interpretation. Sara Woodley was a student investigator. Richard Watkins was involved in informatics and figure creation. Jenna Ridlen is a clinical DO who contributed her medical expertise. All authors contributed to the writing of the manuscript.

ACKNOWLEDGMENTS

We would like to acknowledge the help and support of Ms. Michelle Morris of HudsonAlpha for training and access to software.

REFERENCES

Adzhubei IA, Schmidt S, Peshkin L et al (2010) A method and server for predicting damaging missense mutations. Nat Methods 7:248–249. https://doi.org/10.1038/nmeth0410-248

Choi Y, Chan AP (2015) PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinform Oxf Engl 31:2745–2747. https://doi.org/10.1093/bioinformatics/btv195

Devries S, Mulder M, Charron JG, Prokop JW, Mark PR. SLC6A1 G443D associated with developmental delay and epilepsy [published online ahead of print, 2020 Jul 13]. *Cold Spring Harb Mol Case Stud*. 2020;mcs.a005371. doi:10.1101/mcs.a005371

Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, et al. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem* 24: 1999–2012.

Carvill GL, McMahon JM, Schneider A, et al. Mutations in the GABA Transporter SLC6A1 Cause Epilepsy with Myoclonic-Atonic Seizures. *Am J Hum Genet*. 2015;96(5):808-815. doi:10.1016/j.ajhg.2015.02.016

Johannesen KM, Gardella E, Linnankivi T, Courage C, de Saint Martin A, Lehesjoki A-E, Mignot C, Afenjar A, Lesca G, Abi-Warde M-T, et al. 2018. Defining the phenotypic spectrum of SLC6A1 mutations. *Epilepsia* 59: 389–402.

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

Landrum MJ, Lee JM, Benson M et al (2016) ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res 44: D862–868. https://doi.org/10.1093/nar/gkv1222

Lek M, Karczewski KJ, Minikel EV et al (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nature 536:285– 291. https://doi.org/10.1038/nature19057

Mattison KA, Butler KM, Inglis GAS, et al. SLC6A1 variants identified in epilepsy patients reduce γ-aminobutyric acid transport [published correction appears in Epilepsia. 2019 Aug;60(8):1751]. *Epilepsia*. 2018;59(9):e135-e141. doi:10.1111/epi.14531

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

Prokop JW, Lazar J, Crapitto G, Smith DC, Worthey EA, Jacob HJ. 2017. Molecular modeling in the age of clinical genomics, the enterprise of the next generation. *J Mol Model* 23: 75.

Rees E, Han J, Morgan J, et al. De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in schizophrenia. *Nat Neurosci*. 2020;23(2):179-184. doi:10.1038/s41593-019-0565-2

Tavtigian SV, Defenbaugh AM, Yin L et al (2006) Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. J Med Genet 43:295–305. https://doi.org/10.1136/jmg.2005.033878

Yuan FF, Gu X, Huang X, Zhong Y, Wu J. SLC6A1 gene involvement in susceptibility to attention-deficit/hyperactivity disorder: A case-control study and gene-environment interaction. *Prog Neuropsychopharmacol Biol Psychiatry*. 2017;77:202-208. doi:10.1016/j.pnpbp.2017.04.015

Zech M, Jech R, Wagner M, et al. Molecular diversity of combined and complex dystonia: insights from diagnostic exome sequencing. *Neurogenetics*. 2017;18(4):195-205. doi:10.1007/s10048-017-0521-9

AUTHOR INFORMATION

Benjie Blair: (256) 473-6804, bblair@jsu.edu, 700 Pelham Rd. N, Jacksonville, AL 36265
Jared Painter: (256) 765-4411, jpainter@una.edu, One Harrison Plaza, Florence, AL 35632
Jeremy W Prokop: (330) 573-3176, jprokop54@gmail.com, 400 Monroe Ave NW, Grand Rapids, MI 49503

Jenna Ridlen: (256) 782-5644, jridlen@jsu.edu, 700 Pelham Rd. N, Jacksonville, AL 36265

Cynthia Stenger: (256) 765-4286, clstenger@una.edu, One Harrison Plaza, Florence, AL 35632

TABLES AND FIGURES REFERENCED IN RESULTS

TABLE 1. Genomic Findings

Gene	Variant	Zygosity	Variant Classification	Inheritance
SLC6A1	D451G	Heterozygous	VUS	De Novo
SLC6A1	D451E	Heterozygous	VUS	De Novo

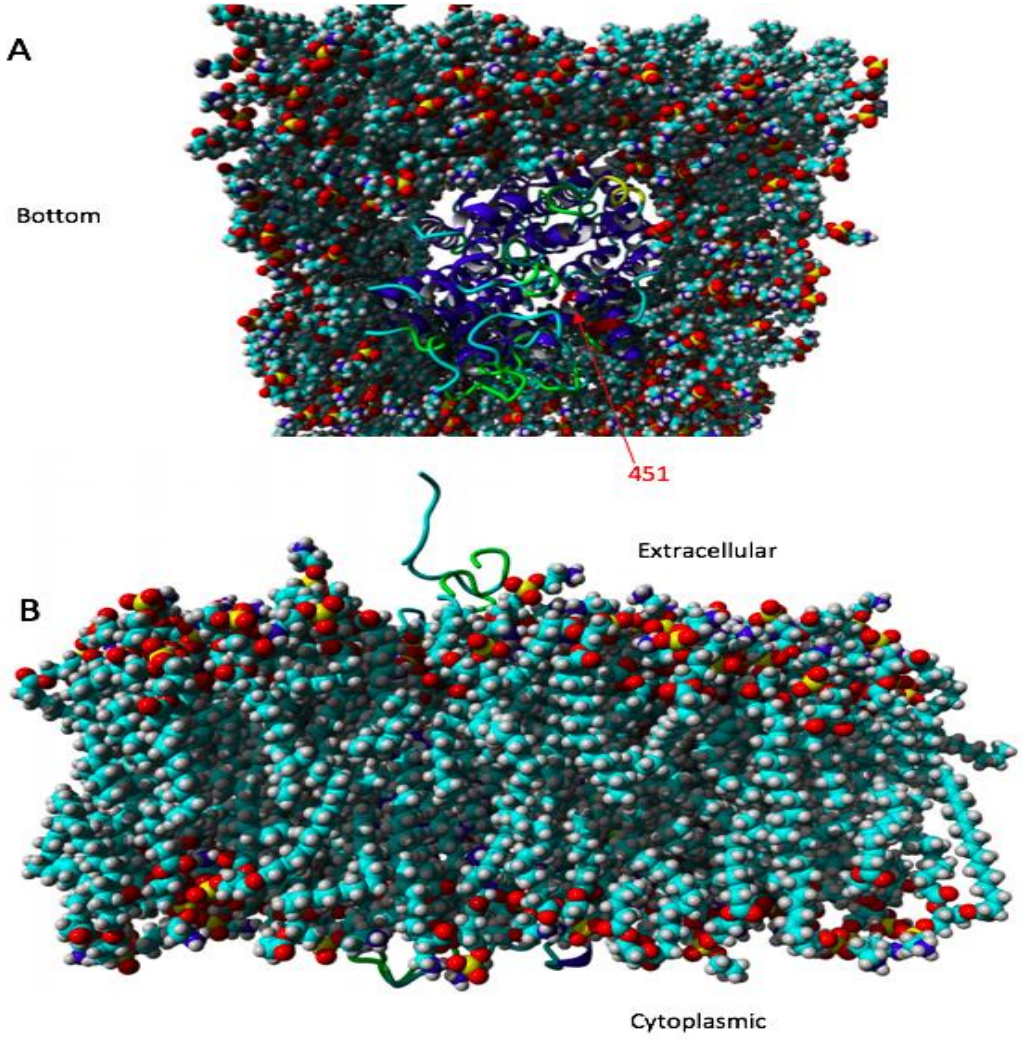

Figure 1. SLC6A1 variant analysis. A-B) Structural model of SLC6A1 (multicolor) in a lipid membrane **A.** looking up from the cytoplasmic side. **B.** Side view (cross section) of the protein modeled in a typical phospholipid membrane.

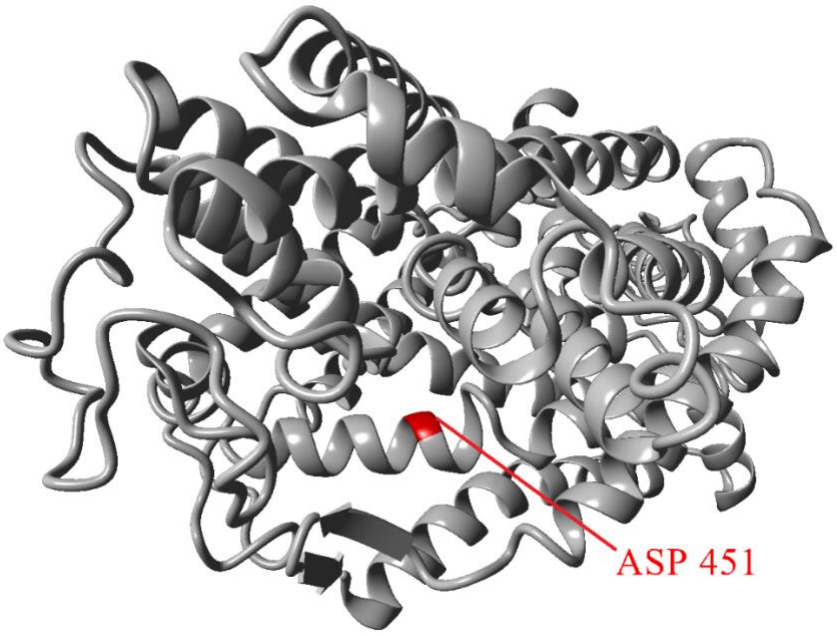

Figure 2. The protein is shown up close with the 451 position marked. The model was performed in a membrane matrix but that has been hidden for ease of viewing. This demonstrates that the variant location is close to the end of an alpha helix and is packed into the middle of the protein.

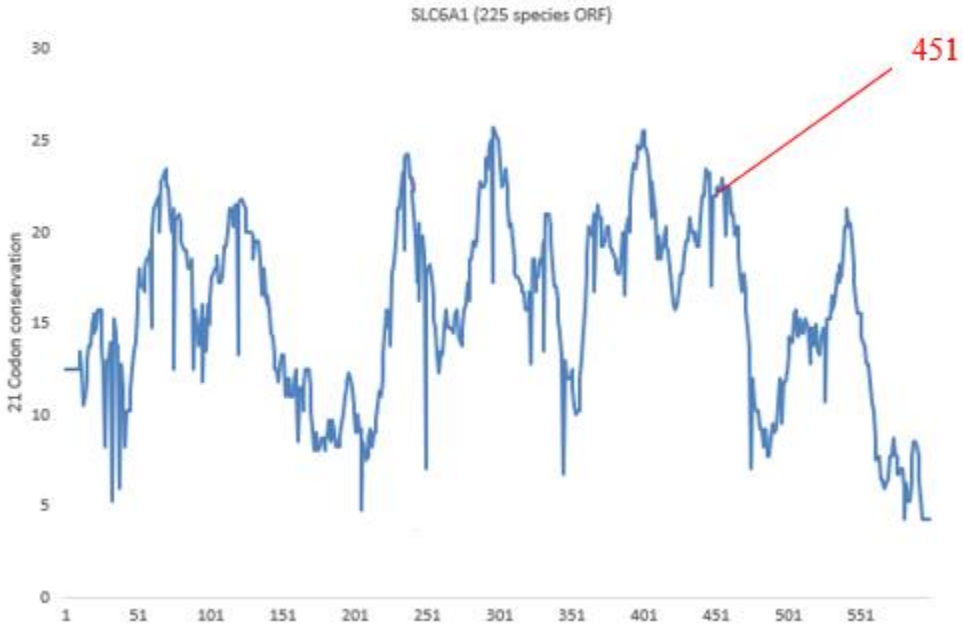

Figure 3. ORF (open reading frames). Deep evolutionary analysis using 225 species open reading frame sequences for SLC6A1. The plot shows a sliding window calculation for each site (plus ten up and downstream), finding the most selected and conserved linear motifs within the gene. Amino acid 451 is showed in red.

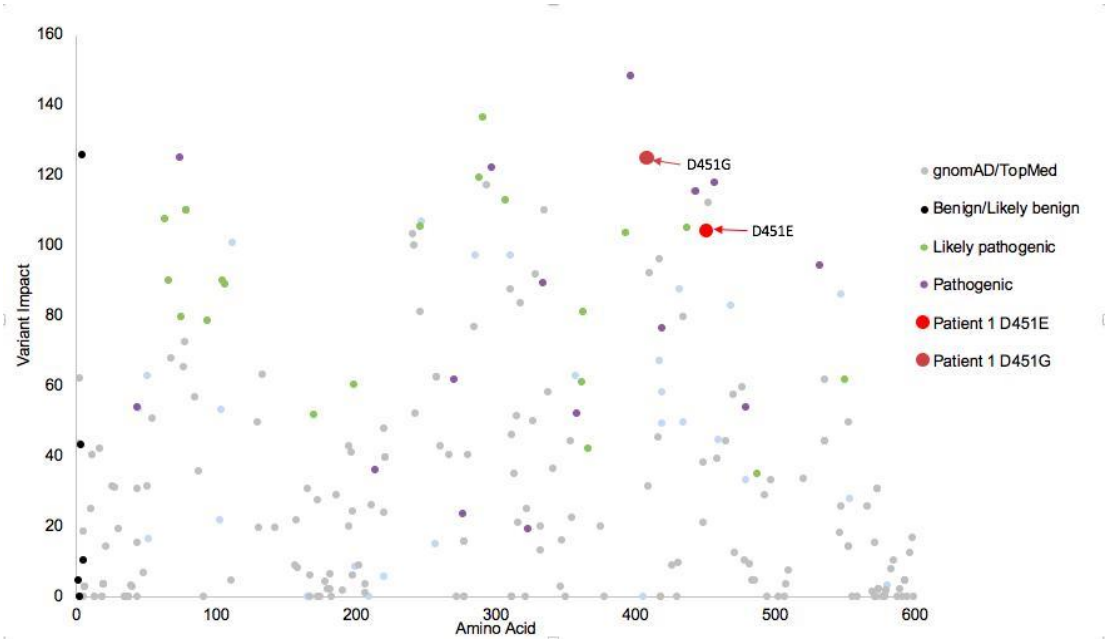

Figure 4. Scatterplot with variant impact scoring for all gnomAD/TOPMed (gray), ClinVar (benign-black), likely pathogenic-(light green), pathogenic-magenta), and patient 1 (cherry red), patient 2 (Maroon) (dots for patients were enlarged slightly) variants for SLC6A1.

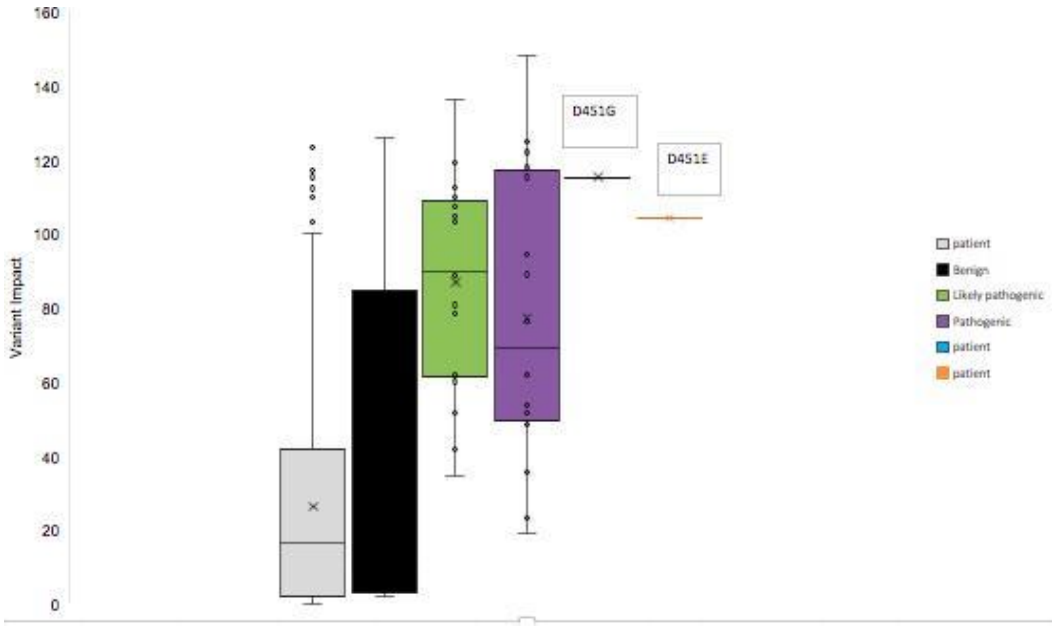

Figure 5. The impact difference between the two variants can be easily seen in a box and whisker plot where both variants are near the top of the pathogenic grouping. Box and whisker plot for each group plotted series 1 are unclassified (lite gray), Benign/likely benign in black, likely pathogenic green, pathogenic purple. Variant D451G and variant D451E.

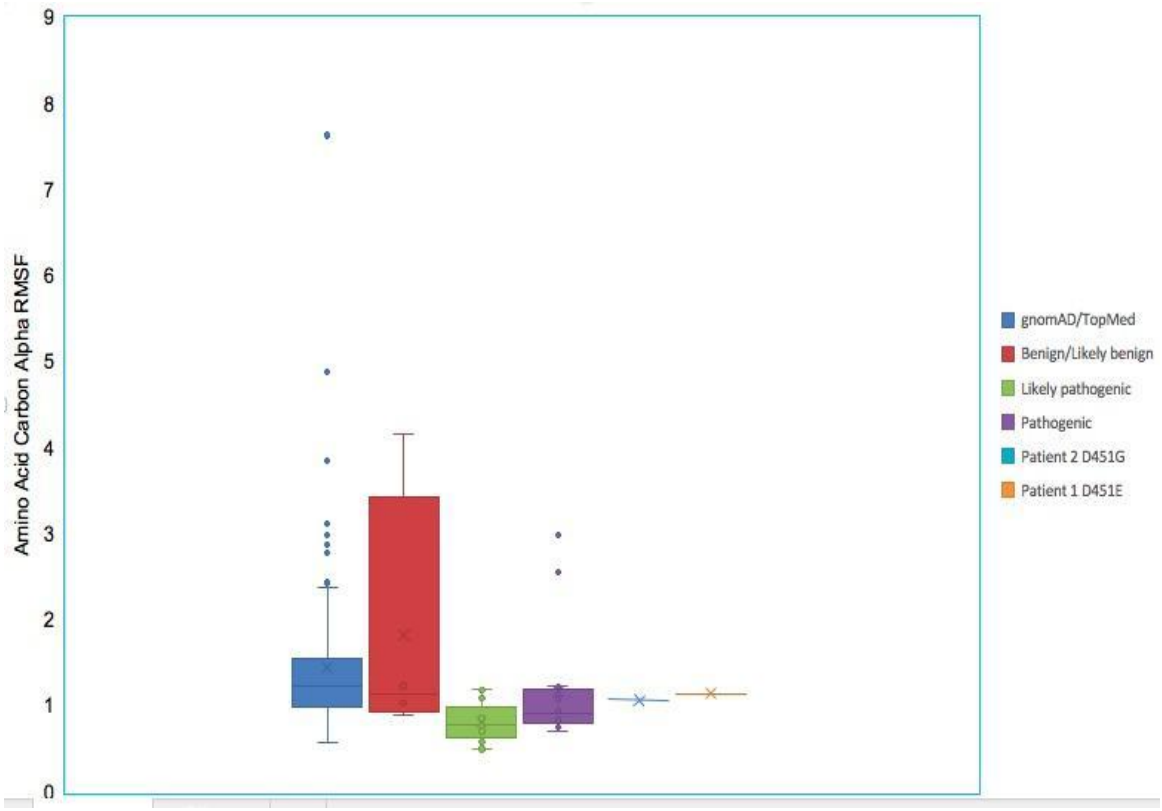

Figure 6. Root mean squared fluctuation (RMSF) of variant groups. With gnomAD/TOPMed (dark blue), Benign/likely benign (dark red), Likely pathogenic (green), pathogenic (purple), Patient 2 (light blue), patient 1 (orange).

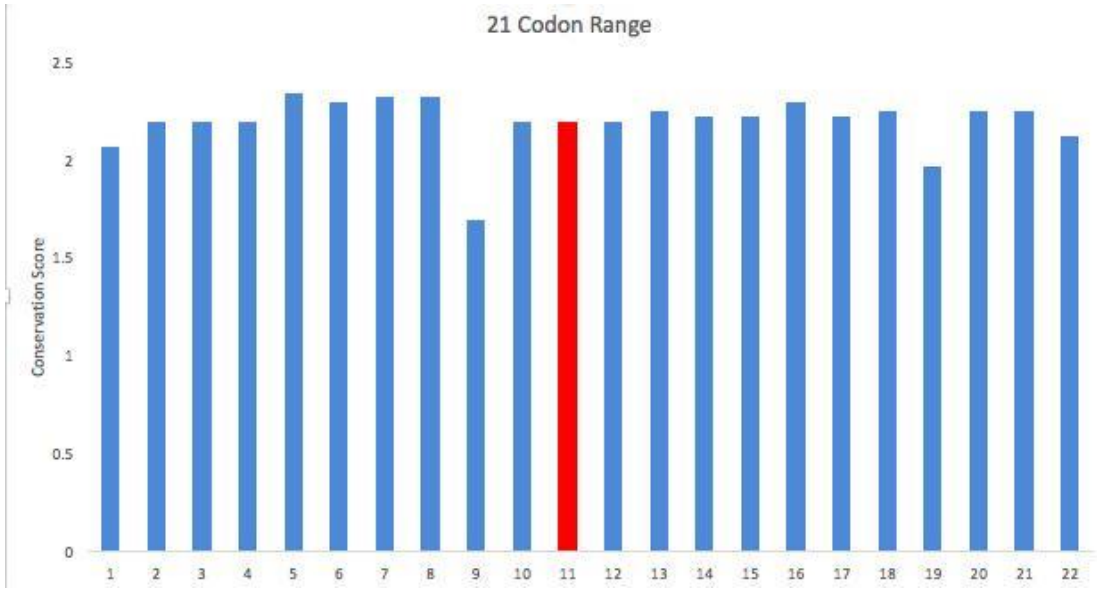

Figure 7. Zoomed in view of conservation for amino acid 451 (red) linear motif. The numbers above represent the percent of sequences with synonymous / nonsynonymous variants throughout evolution.

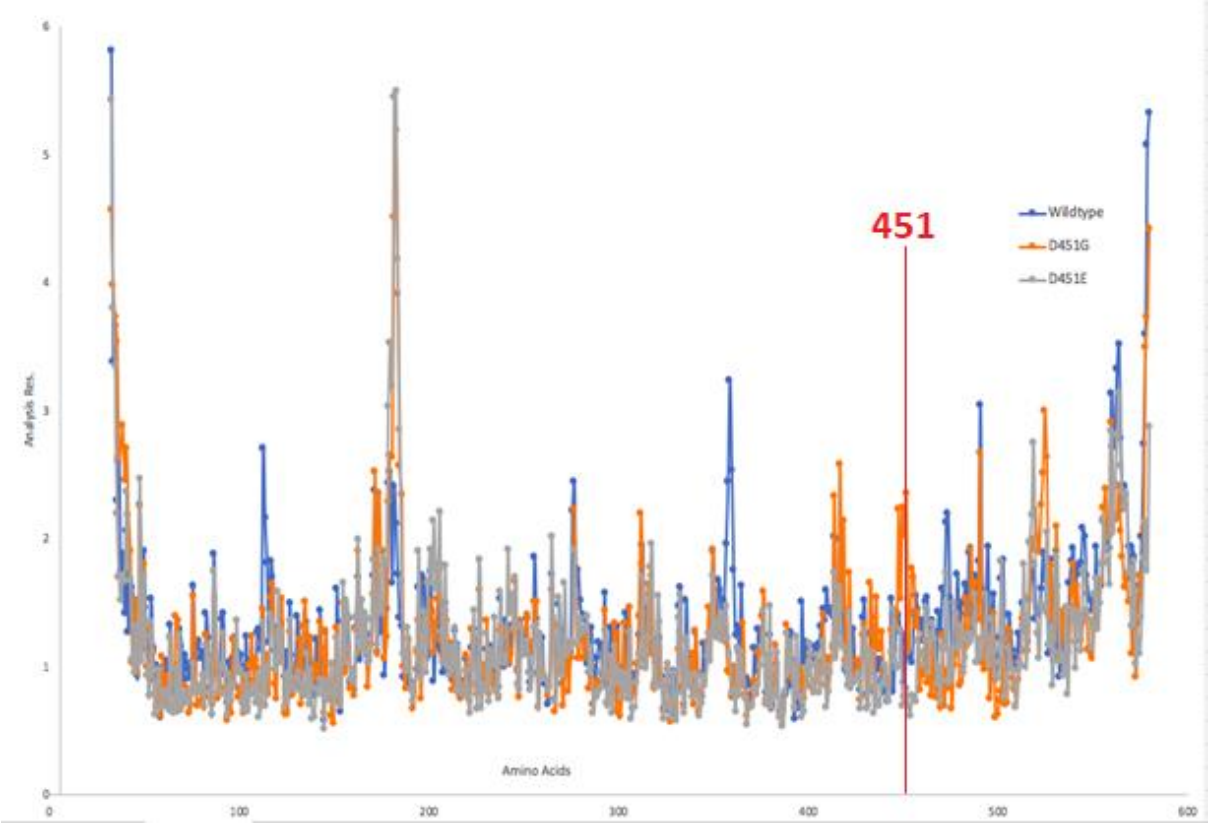

Figure 8. Global deviation of the wild type plus the variants shows D-G more affected than D-E as predicted by the biophysical properties of the amino acids.