

DOI: <https://doi.org/10.24297/jbt.v9i.9109>

A Mathematical Graph-Theoretic Model of Single Point Mutations Associated with Sickle Cell Anemia Disease

Edem K. Netsey¹, Samuel Kakraba^{2,*} Samuel M. Naandam¹, Aayire C. Yadem³

¹ Department of Mathematics, University of Cape Coast, Cape Coast, Ghana, West Africa

² Department of Mathematics and Statistics, College of Science, Technology, Engineering and Mathematics, Eastern Kentucky University, Richmond, KY 40475, U.S.A

³ Department of Applied Science (Physics), University of Arkansas at Little Rock, Little Rock, AR 72205, U.S.A

***Correspondence:** Samuel Kakraba, Ph.D., Department of Mathematics and Statistics, College of Science, Technology, Engineering and Mathematics, 521 Lancaster Ave., 310 Wallace Building, Eastern Kentucky University, Richmond, KY 40475-3133, Telephone: (+1)-859-622-2478, Email: Samuel.Kakraba@eku.edu

Abstract

Folding and maintenance of protein is critical to protein function. Mutations in the amino acid sequence of a protein might prevent it from assuming the appropriate conformation needed for proper functions. Many diseases like cystic fibrosis and sickle cell anemia disease (SCD), among others, arise from single point mutations in the respective proteins. How a single point mutation might lead to a global devastating consequence on a protein remains an intellectual mystery. SCD is a genetic blood-related disorder resulting from mutations in the beta chain of the human hemoglobin protein (simply, β -globin), subsequently affecting the entire human body. Higher mortality and morbidity rates have been reported for patients with SCD, especially in sub-Saharan Africa. Clinical management of SCD often requires specialized interdisciplinary clinicians. SCD presents a major global burden, hence an improved understanding of how single point mutations in β -globin results in different phenotypes of SCD might offer insight into protein engineering, with potential therapeutic intervention in view. By use of mathematical modeling, we built a hierarchical (nested) graph-theoretic model for the β -globin. Subsequently, we quantified the network of interacting amino acid residues, representing them as molecular system of three distinct stages (levels) of interactions. Using our nested graph model, we studied the effect of virtual single point mutations in β -globin that results in varying phenotypes of SCD, visualized by unsupervised machine learning algorithm, the dendrogram.

Keywords: Anemia–sickle cell disease (SCD), beta-globin (β -globin), hemoglobin protein (1A3N), graph theory, graph-theoretic model (nested graph), mutation GLU6VAL (E6V), homozygous HbSS, unsupervised machine learning, clustering

Introduction

Protein engineering requires deep insight into protein folding and misfolding, thereby underscoring the importance of experimental and computational models to study such complex network and systems. There is limited understanding of how a single point mutation in the β -globin prevents correct folding and maintenance of hemoglobin and subsequently affects the entire protein. Protein folding and maintenance are critical for ensuring proper protein function (Yagishita et al., 2008). Many factors (genetic or non-genetic) can prevent correct protein folding, leading to a protein's inability to perform the appropriate structure-related functions (Cooper & Hausman, 2007). Even though the human body has several biological corrective and clearance mechanisms, in extreme instances, these mechanisms are unable to undo the effect of such misfolded proteins that might arise from severe mutations. Misfolded uncorrected proteins aggregate in and around the cells. Many diseases including SCD, COVID-19 and neurodegenerative diseases (such as Parkinson's, Alzheimer's, Amyotrophic lateral sclerosis) show elevated levels of protein aggregates resulting from mostly severe mutations (Ayyadevara et al., 2017; Balasubramaniam et al., 2019; Kakraba et al., 2019). Conventionally, a mutation refers to changes in amino acid sequences of a protein, occurring by either addition, insertion, deletion or switching of positions in the amino acid residue(s) (Griffiths et al., 2002). Single point mutations (a mutation in only one amino acid residue) may result in varying degrees of severity and phenotypic clinical manifestations (Thom et

al., 2013). Illustratively, mutation in the Huntington gene, Parkinson's (e.g. LRRK2, PARK7), cystic fibrosis (e.g. N1303K, deltaF508), diabetes mellitus, sickle-cell anemia (e.g. E6V, V23I, K82N, K95E, E6K, E26K) among other diseases, all result from single point mutations in associated proteins (Abou-Sleiman et al., 2004; Chakravorty & Williams, 2015; Chernoff et al., 1954; Erer et al., 2016; Nuytemans et al., 2010).

SCD is a genetic hemoglobin disorder affecting many people in different populations, diagnosed at birth upon receiving either one or two sickle cell genes (one from each parent) resulting in continuous production of faulty hemoglobin molecules. The most severe form of the disease, homozygous HbSS, occurs when the faulty hemoglobin producing gene is passed on from both parents to the newborn (Da Guarda et al., 2020). The faulty hemoglobin occurs because of a single point mutation involving the substitution of glutamic acid with valine at the sixth position (denoted as Glu6Val or E6V) of the β -globin. In contrast to healthy red blood cells (RBCs) (that appear 'disc-like round'), sickled RBCs appears crescent, less flexible, less oxygenated (due to reduced cell surface and volume), stiff and have the general tendency to stick together, thus preventing smooth blood transport which may end up clogging the blood vessels (Hartl, D. L., & Ruvolo, 2012). Sickled RBCs have a shorter life span of 10 to 20 days instead of the typical 90 to 120 days for healthy RBCs (Kanter & Kruse-Jarres, 2013; Neumayr et al., 2019). A characteristic symptom of SCD is anemia (low RBCs cells count) leading to fatigue, occasional difficulty in breathing and retarded growth in children. Pulmonary hypertension associated with SCD is a serious complication estimated to affect 6-11% of patients with SCD. The world health organization (WHO) estimates the global prevalence of sickle cells trait to be about 5%, comprising mainly the sickle cell disease and thalassemia. Yearly, about 30,000 newborns are diagnosed with significant hemoglobin disorders. In Africa, which carries a major percentage of the global distribution of SCD (SCD prevalence 20% to 45%), several children born with SCD die by the age of five (5). In the United States (US) SCD prevalence is about 1 in every 16,300 Hispanic American and 1 in about every 365 African American births. SCD is among 8 out of 240 specific causes of disease that accounted for 100,000 deaths from 1990 to 2013 worldwide.

Current treatment options are case and symptoms dependent. In the US, Hydroxyurea is prescribed to alleviate episodic pain sensation in SCD patients (Naghavi et al., 2015; Wastnedge et al., 2018). Long time use of Hydroxyurea has been shown to reduce mortality rate by 40% (Neumayr et al., 2019). However, certain adverse events associated with long term use of Hydroxyurea, including alopecia (hair loss), headache, nausea, weight gain and carcinogenic potential continue to impact its global use in SCD clinical management (Salinas Cisneros & Thein, 2020; Segal et al., 2008; Steinberg et al., 2003; Strouse et al., 2008). Periodic RBCs transfusion is an alternative therapy to manage SCD. Adverse immune response to donor antigens has been associated with RBC transfusions (Neumayr et al., 2019). Recently, L-glutamine was approved for management of SCD by the US FDA (Gardner, 2018; Neumayr et al., 2019; Strouse et al., 2008; Wastnedge et al., 2018). Though L-glutamine has been shown to reduce the occurrence of pain crises related to SCD, the frequency of administration of the drug and monthly usage cost (~\$3000 USD) may be challenging to patients (Quinn, 2018). For complete treatment, the state-of-the art is bone marrow therapy or hematopoietic stem cell transplant (HSCT). HSCT procedure is an intensive and risky procedure, ultimately relying on availability of a matching donor. Although there are ongoing research efforts on gene editing therapy for SCD treatment, there are also medical ethics-related challenges and regulatory issues impacting the necessary technological advances required (Gardner, 2018; Neumayr et al., 2019). This underscores the importance of a continuous search for novel treatment for SCD. We are of the view that a great deal of understanding of how a single point mutation in the beta chain of the hemoglobin protein leads to different phenotypes of SCD is critical to protein engineering, drug discovery and design to address the increasing challenges presented by SCD.

Graph theory (a branch of discrete mathematics), concerned with the study of mathematical structures (graphs), is used for pairwise-object relationship modeling. It has been applied to model and study many dynamic systems in an attempt to gain insight into how such complex network systems function (Kyrtos & Baras, 2012). Quite recently, few works have used mathematical graph-theoretic modeling for examining the effect of single point mutations in a protein domain that results in disease with varying phenotypes and clinical manifestations (Kakraba, 2015; Kakraba & Knisley, 2016; Knisley et al., 2013). Presently, to the best of our knowledge, graph-theoretic modeling has not been employed to examine the effect of single point mutations on the β -globin that result in SCD with various clinical manifestations. In this work, we extend mathematical graph-theoretic

modeling approach in our previous works (Knisley et al., 2013; Kakraba, 2015; Kakraba & Knisley, 2016) to the β -globin and examine the effect of single point mutations associated with SCD in β -globin of the hemoglobin.

In modeling the β -globin, we constructed a 3-level hierarchical vertex-weighted model for the β -globin. We obtained new vertex and edge weighted combinatorial molecular indices for our graph-theoretic model of the β -globin. The molecular descriptors were computed based on standard graph theory definitions and authors-adopted definitions for vertex-weighted graphs. The resulting impact of six (6) single point mutations associated with varying degrees of severity and phenotype of sickle cell anemia in the β -globin are measured by use of the nested graph model. Each virtual single point mutation results in distinct combinatorial invariants (indices or quantitative molecular descriptors) that reflect the local and global effect of each specific mutation on the entire protein. In this way, we captured the underlying consequential structural local and global network changes that stem from each single point mutation in the β -globin. Arguably, typical of our mathematical graph-theoretic modeling is the ability to rationally model how a local perturbation (mutation) results in small or huge global effect on the β -globin.

Materials and Methods

Graph-theoretic model for β -globin

Although, graph-theoretic modeling adopted in this work is similar to previously used methods (Kakraba, 2015; Kakraba & Knisley, 2016), this work focuses on the beta chain of the hemoglobin protein and broaden previously published combinatorial descriptors to incorporate edge-weight assignments. We adopted some molecular indices from the molecular database from Table 1-3 (Kakraba, 2015; Kakraba & Knisley, 2016) and incorporated other standard and author-adopted graph invariants in this work. Therefore, the molecular descriptors of the subdomain graphs (resulting from the unique partitioning into subsequences), the resulting β -globin graph, and the mutations examined in this work are specific to the β -globin.

Subsequence partition of the β -globin

The fasta sequence of the protein corresponding to the β -globin (PDB ID: 1A3N) was retrieved from the protein databank (Tame & Vallone, 2000; *The Protein Data Bank*, 1998). Using UCFS Chimera (Pettersen et al., 2004) plug-in built for Cystoscape (Shannon et al., 2003), we visualized and partitioned the β -globin into ten (10) non-overlapping subsequences represented by $G_i, i = 1 \dots 10$. To preserve essential biological information contained in the secondary structures of the β -globin, we avoided cutting through binding sites, beta strands and alpha helices during partitioning of the β -globin into subsequences. We also ensured each subsequence contained only one type of secondary structure; either a beta strand, an alpha helix, turn, bend or a loop with no more than 25 amino acid residues. A loop region was tolerated to include turns, a 3/10-helix, and an alpha helix with less than 6 residues of amino acids. Table 1 is the resulting non-overlapping subsequences for each subdomain and reason for such partitioning of the β -globin.

Table 1. Subsequence Partition for the β -globin

Subdomain	Subsequence	Amino Acid Sequence	Reason/ Rationale
G₁	1.....17	VHLTPEEKSAVTALWGK	Coil, alpha helix, turn
G₂	18.....35	VNVDEVGGEALGRLLVY	Coil, alpha helix, bend
G₃	36.....48	PWTQRFFESFGDL	3/10 helix, coil, binding site
G₄	49.....57	STPDVVMGN	Bend, alpha helix, coil
G₅	58.....73	PKVKAHGKKVLGAFSD	Alpha helix
G₆	74.....79	GLAHLG	Binding site

G₇	80.....94	NLKGTFATLSELHCD	Alpha helix, binding site
G₈	95.....117	KLHVDPENFRLLGNVLCVLAHH	Turn, coil, alpha helix, binding site
G₉	118.....123	FGKEFT	3/10 helix, bend
G₁₀	124.....146	PPVQAAYQKVVAGVANALAHKYH	Alpha helix, turn, coil

Subdomain graphs corresponding to the subsequences of the β -globin

The subsequences corresponding to each subdomain of the β -globin were submitted to I-TASSER for *Ab-initio* modeling, and the results used in Cystoscape to generate the interaction or network graphs. Altogether, ten (10) subdomain graphs were generated corresponding to the ten subsequences in Table 1.

To observe more interactions between the amino acid residues in the subdomain graphs of the β -globin, we chose a proximity measure of six (6) angstroms. We determined the endpoints by the center of mass of each amino acid residue in the 3D structure of β -globin. The wildtype subdomain graph and mutated subdomain graph (E6V) of subsequence corresponding to G₁ are shown in **Fig.1** (refer to the **appendix 1** for other subdomain graphs generated for β -globin). Mutated subdomain graphs were obtained by submitting the mutated subsequences to I-TASSER (Zhang, 2007b, 2007a) for *ab initio* predictive modeling, and subsequently using Cystoscape for the corresponding network graphs (refer to the **appendix 2** for other mutated subdomain graphs generated for β -globin).

Hierarchical (Nested Graph) Model for the β -globin

Finally, we denote each of the subdomains with a vertex in a domain graph of the β -globin (**H**). From the protein sequence submitted to UCFS chimera software (Pettersen et al., 2004), we obtained the network interaction data and using Cystoscape (Shannon et al., 2003), network (interaction) graph was generated for the β -globin. Our hierarchical graph has three (3) layers, namely, the lowest level, middle level, and the top level. The lowest level comprises 20-small vertex-weighted amino acids, corresponding to the 20 most essential amino acids. The middle level graph denotes the subgraphs corresponding to the subsequences of the β -globin. We generated ten (10) distinct vertex-weighted subgraphs (each vertex in the subgraph denoting an amino acid) in which the weights of the vertices and edges are graph invariants (molecular descriptors) obtained from the amino acids of the lower-level graph. In the top-level graph, each subdomain graph (subgraph from the middle level) collapses into a single weighted vertex (node). The weights assigned to each subdomain graph as a vertex or node in the top-level graph are based on vertex and edge weighted molecular descriptors of each subdomain graph. The edges/links seen in the β -globin nested domain graph are determined by a nearness or proximity measure threshold distance of 6 angstroms (**6Å**) between any two adjacent residues of the graph. The molecular

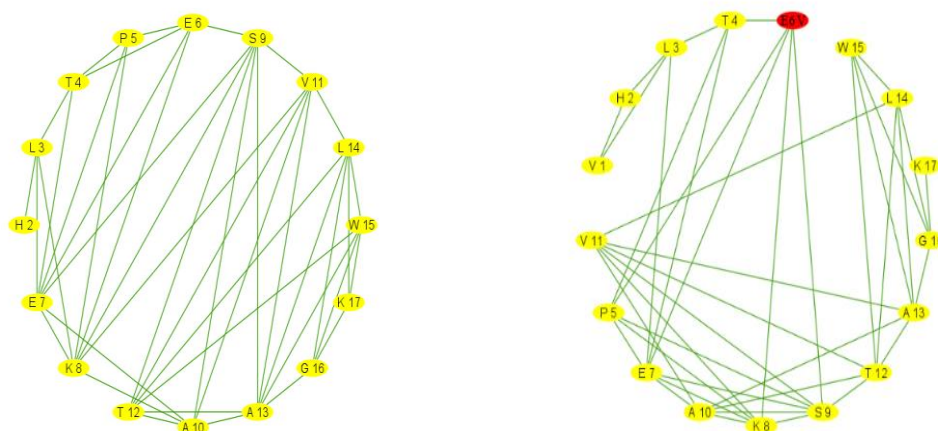


Figure 1. Subdomain (wildtype) G₁ (left) and mutated subdomain graph (E6V) G₁ (right)

descriptors particular to our β -globin model were then calculated for further network analysis. **Figure 2** shows a model for the β -globin based on the graph-theoretic approach applied in this work.

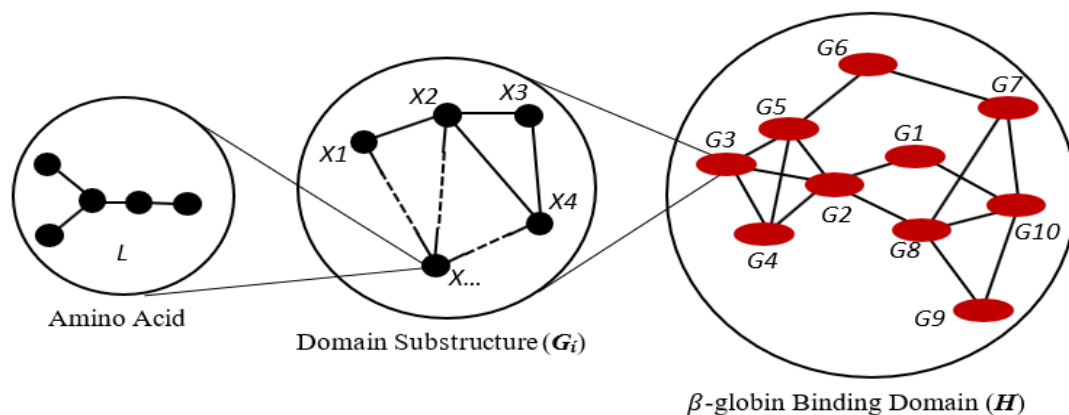


Figure 2. A Hierarchical (Nested) Graph Model for β -globin of Hemoglobin

Graph-theoretic invariants (molecular descriptors) for Subdomain and Top-level Graph

We computed weighted graph invariants for subdomain graphs for the corresponding subsequences aimed at describing the local properties of the subgraphs. These values become the wildtype values of the subdomain graphs, and subsequently incorporated as vertex-edge weights in the top-level graph (**H**). The molecular descriptors calculated for the wildtype values of the subdomain graphs are depicted in Table 2. Descriptor e5 was based on assignment of molecular descriptors d13 from our previous work (Kakraba & Knisley, 2016). Weighted domination number, weighted betweenness, and weighted eccentricity measures were used in assessing the impact of specific nodes in the graphs to determine their influences within the graph networks. For a given subdomain or top-level graph, we also considered the absolute value of the change in molar mass per average edge (ΔM_d) along the edge between connected vertices R_i and R_j , using equation (1) in which \bar{d} is the average weighted degree.

$$\Delta M_d = \left| \frac{M(R_i) - M(R_j)}{\bar{d}} \right| \times 10^{-3} \text{Kgmol}^{-1} \quad (1)$$

Table 2: Weighted Molecular Descriptors of Subdomain Graphs for β -globin

Subdomain	e1	e2	e3	e4	e5	e6	e7	e8	e9	e10	e11
G₁	17.000	43.000	84.000	15.000	24.000	5.000	4.941	0.255	18.000	0.032	0.229
G₂	18.000	52.000	104.000	15.000	36.000	6.000	5.778	0.235	22.333	0.026	0.221
G₃	13.000	24.000	48.000	10.000	42.000	6.000	3.692	0.222	17.846	0.035	0.241
G₄	9.000	21.000	42.000	7.000	7.000	3.000	4.667	0.407	4.222	0.085	0.313
G₅	16.000	45.000	90.000	11.000	7.000	5.000	5.625	0.255	17.375	0.032	0.237
G₆	6.000	10.000	20.000	5.000	15.000	2.000	3.333	0.583	1.667	0.153	0.399
G₇	15.000	40.000	80.000	13.000	24.000	5.000	5.333	0.264	16.000	0.035	0.244
G₈	23.000	59.000	118.000	19.000	49.000	9.000	5.130	0.153	50.609	0.015	0.178
G₉	6.000	11.000	22.000	5.000	12.000	2.000	3.667	0.667	1.333	0.163	0.400
G₁₀	23.000	62.000	124.000	19.000	48.000	8.000	5.391	0.170	43.826	0.016	0.188

Key: $e1$ = order of graph (number of nodes/vertices), $e2$ = size of graph (number of interactions/edges); $e3$ = total vertex-weighted degree; $e4$ = average edge-weighted domination number; $e5$ = vertex-weighted domination based on assignment of atomic number assignment to the vertices (based on $d13$), $e6$ = diameter; $e7$ = average degree value; $e8$ = average eccentricity value; $e9$ = average betweenness value; $e10$ = average closeness; $e11$ = average eigenvector

With these weights assigned to the edges of the top-level graph (**H**), we computed other graph centrality measures like weighted order of the graph, degree centrality, closeness centrality, eigenvector centrality and betweenness centrality (Ni et al., 2011) for subdomain graphs and top-level graph using Cystoscape (Shannon et al., 2003).

Virtual Single Point SCD Mutations in the β -globin

To assess the effect of single point mutations on the entire β -globin, we proceeded by selecting six (6) prevalent mutations known to be associated with mild or severe SCD phenotypes in the β -globin from literature (refer to Table 3). In computing the set of graph-theoretic indices corresponding to each mutation, we employed these procedures. First, starting from the residue in the subdomain graphs (**G_i**), assign amino acid descriptors from the molecular database (Kakraba, 2015; Kakraba & Knisley, 2016) as weights to the vertices of the subdomain graphs and calculate graph invariants specific to each subdomain graph. Use subdomain molecular descriptors from above as weights of the vertices or edges in the top-level graph and subsequently obtain graph invariants for the top-level graph (**H**). These invariants of the top-level graph represent the wildtype molecular descriptors. To perform the virtual single point mutations for the mild and severe phenotypes, return to each subdomain graph where a specific mutation occurs, and locate the corresponding amino acid at that position in the amino acid residue. Mutate the residue by substituting or deleting the mutant residue at the position (where appropriate) in the subdomain graph. Consequently, a 'mutant-specific vertex-weighted graph' results for that subdomain graph where the mutation occurred. Next, compute the corresponding graph-theoretic molecular descriptors or invariants for the new vertex-weighted subdomain graph ('mutated subdomain graph') arising from the mutation. Proceed to assign these new set of molecular descriptors to the top-level graph and recompute the molecular descriptors of the top-level graph resulting from the mutation. Each mutation results in graph-theoretic-specific molecular descriptors of the top-level graph. Through this approach, we determine the local and global effect of each single point mutation on the β -globin. The wildtype graph-theoretic molecular descriptors and those arising from the six (6) mutation provide indices useful for viewing the local and global effect of each single point mutations in the β -globin (*mutated subdomain graphs depicted in appendix 2*). The mutation-specific molecular descriptors for the top-level graph (β -globin) are calculated and reported in Table 4.

Table 3. Anemia-Sickle Cell-Causing Mutations in β -globin

Mutation	SCD Phenotype
Wildtype	No
GLU6VAL (E6V)	Severe
GLU6LYS (E6K)	Mild
VAL23ILE (V23I)	Mild
GLU26LYS (E26K)	Mild
LYS82ASP (K82N)	Mild
LYS95GLU (K95E)	Mild

Table 4. Graph Invariants for SCD-Specific Mutation in the β -globin

Measure	Wildtype	E6V	E6K	V23I	E26K	K82N	K95E
WMaxDeg	1.442	1.442	1.442	1.437	1.442	1.447	1.443
WMeanDeg	0.936	0.940	0.936	0.939	0.936	0.937	0.936
WMinDeg	0.243	0.263	0.244	0.248	0.243	0.243	0.243
WMaxEig	0.583	0.583	0.583	0.583	0.583	0.582	0.583
WMeanEig	0.248	0.248	0.248	0.247	0.248	0.248	0.248
WMinEig	0.037	0.036	0.037	0.036	0.037	0.036	0.037
WMaxClos	0.192	0.193	0.192	0.191	0.192	0.192	0.192
WMeanClos	0.152	0.153	0.152	0.152	0.152	0.153	0.152
WMinClos	0.100	0.101	0.100	0.100	0.100	0.100	0.100

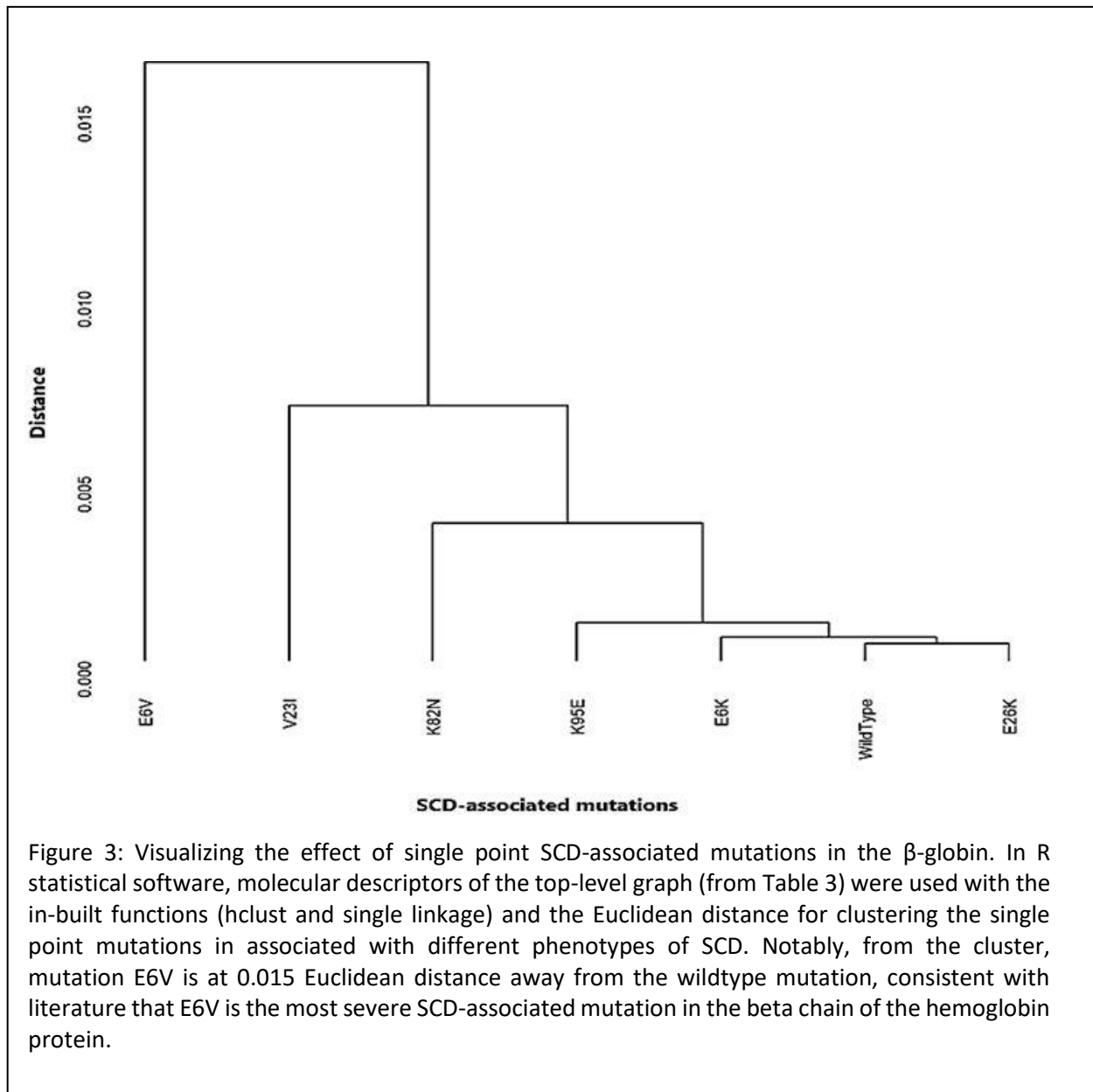
Key: *WMaxDeg* = maximum weighted degree; *WMeanDeg* = average weighted degree; *WMinDeg* = minimum weighted degree; *WMaxEig* = maximum weighted eigenvector; *WMeanEig* = average weighted eigenvector; *WMinEig* = minimum weighted eigenvector; *WMaxClos* = maximum weighted closeness; *WMeanClos* = average weighted closeness; *WMinClos* = mean weighted closeness

Results and Discussion

We used our graph-theoretic molecular weighted invariants or descriptors (refer to Table 4) computed for the wildtype, mild and severe mutations, and employed a statistical unsupervised machine learning algorithm (specifically, a hierarchical clustering) to visualize how each SCD single point mutation varies from the wildtype for the β -globin. In R statistical software (R Core Team, 2014), we used the single linkage function and the Euclidean distance for generating the dendrogram. Our choice of the single-linkage function stems from our effort to decrease any possible biases in the clustering of the single point mutations associated with SCD. In this way, we were able to visualize the effect of the virtual SCD single point mutations on the entire β -globin in relation to the wildtype. The dendrogram (depicted in **Fig. 3**) is the resulting visual representation of the effect of each single point mutation on the β -globin, and subsequently the entire hemoglobin protein.

Notably, our model (**Fig. 3**) classifies the severe mutation, E6V into entirely different and farthest cluster (distance of 0.015 Euclidean distance) away from the wildtype, as an expression of the profound impact of E6V on the β -globin. Consistent with literature and clinical manifestations, E6V (Glu6Val) is a typical sickle-cell anemia mutation of severe clinical outcomes. E6V or Glu6Val is the mutation occurring as a substitution of glutamic acid with valine at position 6 in the β -globin. Characteristic of E6V is the formation of abnormal hemoglobin S subunits that aggregates or sticks to each other. This results in formation of long, rigid rod-like molecules, thereby altering the characteristic functional shape of the red blood cells to form sickle or crescent shape. Consequentially, premature death of sickle-shaped cells arises, leading to shortage of red blood cells (anemia). Severe pain and damage to the organs happens when the rigid sickle-shaped cells block small blood vessels, preventing easy flow of blood (Cohen & Mahadevan, 2013; Hartl, D. L., & Ruvolo, 2012; Kim, 2014; Oksenberg et al., 2016; Parrow et al., 2017). In addition, our model suggests that V23I and K82N are significantly different from the wildtype, thereby warranting further research.

Literature review only shows a handful of such graph-theoretic models that have been used to study the effect of single point mutations that leads to different phenotypes of diseases in associated proteins. Like previous research (Knisley et al., 2013; Kakraba, 2015; Kakraba & Knisley, 2016) that applied graph-theoretic modeling to examine the effect of single point mutations that leads to the severe and the mild phenotypes of cystic fibrosis in the NBD1 and NBD2 of CFTR, the current research has brought to light how a single point mutation in the



beta chain of the hemoglobin protein might lead to the anemia sickle cell disease of varying clinical manifestations.

Conclusions

That single point mutations (local perturbations) in the beta chain of the hemoglobin consequentially translate into either mild or severe global effect on an entire protein has not been explicitly studied. To gain further insight into how single point mutations might impact the entire hemoglobin protein, we built a hierarchical (nested) graph-theoretic model for the β -globin. Subsequently, we quantified the network of interacting amino acid residues, representing them as molecular system of distinct stages (levels) of interactions. Using our nested graph model, we studied the effect of virtual single point mutations on the β -globin, visualized by unsupervised machine learning algorithm, the dendrogram. By employing such mathematical modeling, computational and systems biology methodology, we were able to visualize how a small perturbation such as a single point mutation affects the entire hemoglobin protein. Future research might consider addressing questions such as, why does mutation E6V cluster farthest from the wildtype in our model? Also, what inhibitors (corrector small molecules e.g., amino acid) or graph invariant(s) of the subdomain graph containing E6V would cause the severe

single point anemia-sickle-cell causing mutation (E6V) to cluster along or closer to the wildtype in our hierarchical graph-theoretic model? Such research questions, when addressed, will provide great insight from a computational therapeutic perspective on small molecule discovery and design to attenuate and or reverse the effects of the severe single point mutation (E6V) causing SCD.

Although, the present work focused on application of graph-theoretic model to examine the effect of single point mutations causing the mild and severe phenotype of SCD in the β -globin, the established mathematical modeling concept applied in this work is extendable to any disease arising from single point mutations in a protein like muscular dystrophy, COVID-19 (manuscript in preparation), among others. The existence of such a predictive graph-theoretic mathematical modeling that reflects systems biology might be informative in understanding unknown consequences of single point mutations in a protein domain, and aid in drug design for therapeutic intervention. This underscores the importance of computational models, especially mathematical graph-theoretic models in epidemiology. Our work adds to the existing knowledge on application of mathematical graph-theoretic modeling to study complex networks and systems biology, with potential therapeutic application in drug discovery and design.

Data Availability (excluding review articles)

Refer to the methods section for source of data from protein data bank

Author Contribution

Samuel Kakraba (SK) and Samuel M. Naandam (SMN) were responsible for concept, application, and interpretation of this work. Edem K. Netsey (EKN) performed works related to graph-theoretic model and generation of molecular descriptors under supervision of SK and SMN. Hierarchical clustering on dataset for the top-level graph invariants was implemented in R statistical software by SK. Manuscript was written by SK, Aayire C. Yadem (ACY), SMN and EKN.

Conflicts of Interest: Authors declare no conflict of interest for this present study.

Funding Statement: No funding was received for this research work.

Acknowledgments

We are thankful to Prof. Debra Knisley (East Tennessee State University, Johnson City, TN), Dr. Jeremy Chamberlain (Southern Arkansas University, Magnolia, AR), Dr. Russell Carpenter & Dr. Jamie Shaffer (Eastern Kentucky University, Richmond, KY) for their inputs on this work.

References

1. Abou-Sleiman, P. M., Healy, D. G., & Wood, N. W. (2004). Causes of Parkinson's disease: Genetics of DJ-1. In *Cell and Tissue Research*. <https://doi.org/10.1007/s00441-004-0922-6>
2. Ayyadevara, S., Balasubramaniam, M., Kakraba, S., Alla, R., Mehta, J. L., & Shmookler Reis, R. J. (2017). Aspirin-Mediated Acetylation Protects Against Multiple Neurodegenerative Pathologies by Impeding Protein Aggregation. *Antioxidants & Redox Signaling*. <https://doi.org/10.1089/ars.2016.6978>
3. Balasubramaniam, M., Ayyadevara, S., Ganne, A., Kakraba, S., Penthala, N. R., Du, X., Crooks, P. A., Griffin, S. T., & Shmookler Reis, R. J. (2019). Aggregate Interactome Based on Protein Cross-linking Interfaces Predicts Drug Targets to Limit Aggregation in Neurodegenerative Diseases. *IScience*, 20, 248–264. <https://doi.org/10.1016/j.isci.2019.09.026>
4. Chakravorty, S., & Williams, T. N. (2015). Sickle cell disease: A neglected chronic disease of increasing global health importance. *Archives of Disease in Childhood*. <https://doi.org/10.1136/archdischild-2013-303773>
5. Chernoff, A. I., Minnich, V., & Chongchareonsuk, S. (1954). *Hemoglobin E, a Hereditary Abnormality of Human Hemoglobin*. <https://doi.org/10.1126/science.120.3120.605>
6. Cohen, S. I. A., & Mahadevan, L. (2013). Hydrodynamics of hemostasis in sickle-cell disease. *Physical*

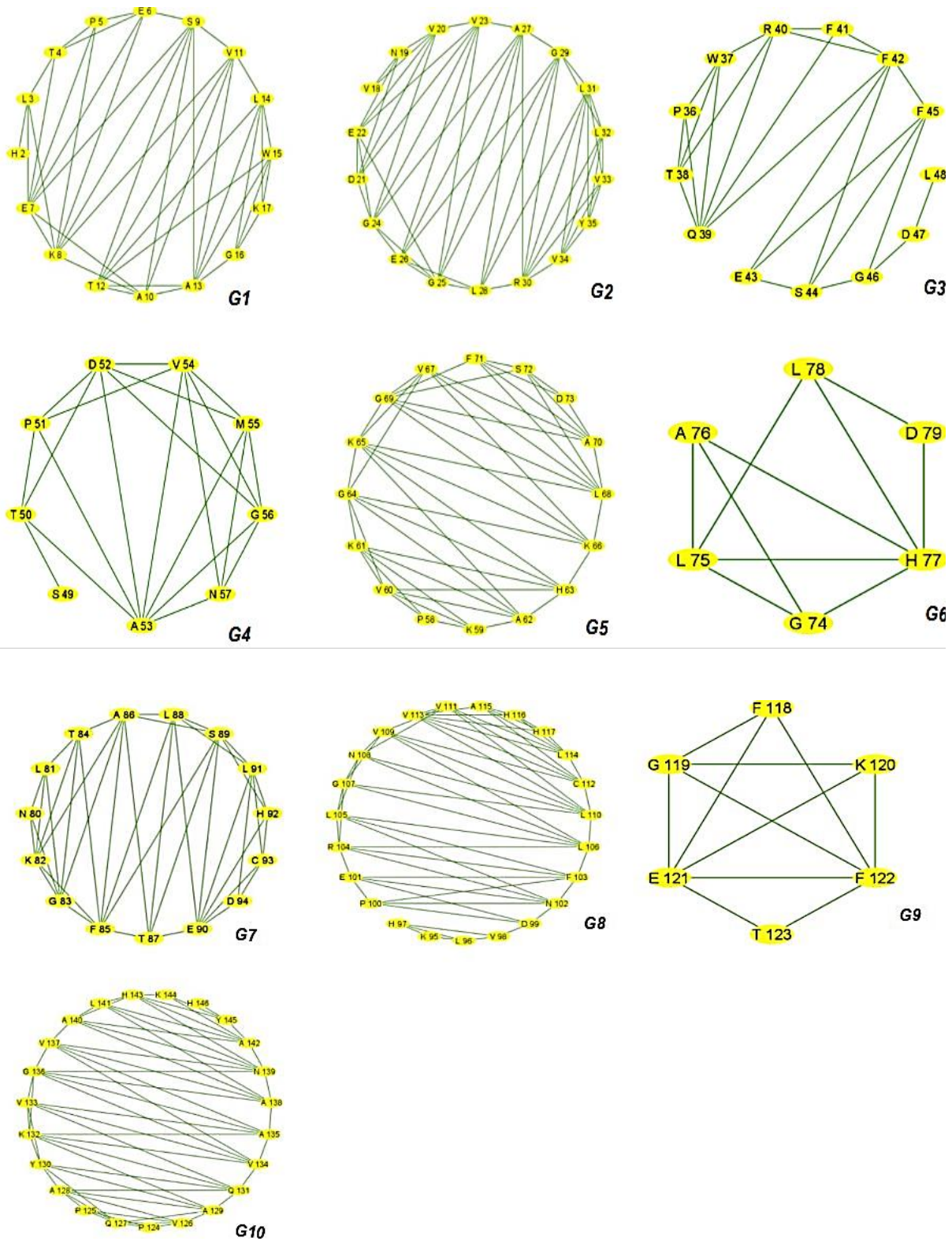
Review Letters. <https://doi.org/10.1103/PhysRevLett.110.138104>

7. Cooper, G. M., & Hausman, R. E. (2007). *The Cell: A Molecular Approach 2nd Edition*. In *Sinauer Associates*.
8. Da Guarda, C. C., Yahouédéhou, S. C. M. A., Santiago, R. P., Dos Santos Neres, J. S., De Lima Fernandes, C. F., Aleluia, M. M., Figueiredo, C. V. B., Fiuza, L. M., Carvalho, S. P., De Oliveira, R. M., Fonseca, C. A., Ndidi, U. S., Nascimento, V. M. L., Rocha, L. C., & Goncalves, M. S. (2020). Sick cell disease: A distinction of two most frequent genotypes (HbSS and HbSC). *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0228408>
9. Erer, S., Egeli, U., Zarifoglu, M., Tezcan, G., Cecener, G., Tunca, B., Ak, S., Demirdogen, E., Kenangil, G., Kaleagasi, H., Dogu, O., Saka, E., & Elibol, B. (2016). Mutation analysis of the PARKIN, PINK1, DJ1, and SNCA genes in Turkish early-onset Parkinson's patients and genotype-phenotype correlations. *Clinical Neurology and Neurosurgery*. <https://doi.org/10.1016/j.clineuro.2016.07.005>
10. Gardner, R. V. (2018). Sick cell disease: Advances in treatment. In *Ochsner Journal*. <https://doi.org/10.31486/toj.18.0076>
11. Griffiths, a. J. F., Gelbart, W. M., Lewontin, R. C., & Miller, J. H. (2002). *Modern Genetic Analysis*. <https://www.ncbi.nlm.nih.gov/books/NBK21417/>.
12. Hartl, D. L., & Ruvolo, M. (2012). *Genetics: Analysis Of Genes And Genomes* (8th ed.). Jones & Bartlett Learning.
13. Kakraba, S. (2015). *A Hierarchical Graph for Nucleotide Binding Domain 2*.
14. Kakraba, S., Ayyadevara, S., Penthala, N. R., Balasubramaniam, M., Ganne, A., Liu, L., Alla, R., Bommagani, S. B., Barger, S. W., Griffin, W. S. T., Crooks, P. A., & Shmookler Reis, R. J. (2019). A Novel Microtubule-Binding Drug Attenuates and Reverses Protein Aggregation in Animal Models of Alzheimer's Disease. *Frontiers in Molecular Neuroscience*. <https://doi.org/10.3389/fnmol.2019.00310>
15. Kakraba, S., & Knisley, D. (2016). A graph-theoretic model of single point mutations in the cystic fibrosis transmembrane conductance regulator. *JOURNAL OF ADVANCES IN BIOTECHNOLOGY*, 6(1), 780–786. <https://doi.org/10.24297/jbt.v6i1.4013>
16. Kanter, J., & Kruse-Jarres, R. (2013). Management of sickle cell disease from childhood through adulthood. *Blood Reviews*. <https://doi.org/10.1016/j.blre.2013.09.001>
17. Kim, H. C. (2014). Red cell exchange: Special focus on sickle cell disease. *Hematology*. <https://doi.org/10.1182/asheducation-2014.1.450>
18. Knisley, D. J., Knisley, J. R., & Herron, A. C. (2013). Graph-Theoretic Models of Mutations in the Nucleotide Binding Domain 1 of the Cystic Fibrosis Transmembrane Conductance Regulator. *Computational Biology Journal*, 2013, 938169. <https://doi.org/10.1155/2013/938169>
19. Kyrtos, C. R., & Baras, J. S. (2012). A graph-theoretic mathematical model for Alzheimer's disease: Using a systems biology approach. *IEEE 12th International Conference on Bioinformatics and BioEngineering, BIBE 2012*. <https://doi.org/10.1109/BIBE.2012.6399723>
20. Naghavi, M., Wang, H., Lozano, R., Davis, A., Liang, X., Zhou, M., Vollset, S. E., Abbasoglu Ozgoren, A., Abdalla, S., Abd-Allah, F., Abdel Aziz, M. I., Abera, S. F., Aboyans, V., Abraham, B., Abraham, J. P., Abuabara, K. E., Abubakar, I., Abu-Raddad, L. J., Abu-Rmeileh, N. M. E., ... Temesgen, A. M. (2015). Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*. [https://doi.org/10.1016/S0140-6736\(14\)61682-2](https://doi.org/10.1016/S0140-6736(14)61682-2)
21. Neumayr, L. D., Hoppe, C. C., & Brown, C. (2019). Sick cell disease: current treatment and emerging therapies. *The American Journal of Managed Care*.

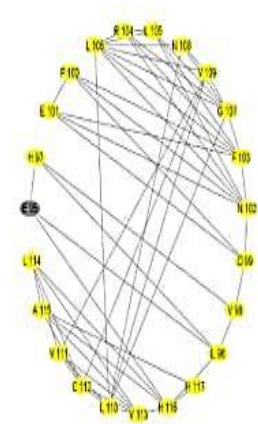
22. Ni, C., Sugimoto, C. R., & Jiang, J. (2011). Degree, Closeness, and Betweenness: Application of group centrality measurements to explore macro-disciplinary evolution diachronically. *Proceedings of ISSI 2011 - 13th Conference of the International Society for Scientometrics and Informetrics*.
23. Nuytemans, K., Theuns, J., Cruts, M., & Van Broeckhoven, C. (2010). Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes: A mutation update. In *Human Mutation*. <https://doi.org/10.1002/humu.21277>
24. Oksenberg, D., Dufu, K., Patel, M. P., Chuang, C., Li, Z., Xu, Q., Silva-Garcia, A., Zhou, C., Hutchaleelaha, A., Patskovska, L., Patskovsky, Y., Almo, S. C., Sinha, U., Metcalf, B. W., & Archer, D. R. (2016). GBT440 increases haemoglobin oxygen affinity, reduces sickling and prolongs RBC half-life in a murine model of sickle cell disease. *British Journal of Haematology*. <https://doi.org/10.1111/bjh.14214>
25. Parrow, N. L., Tu, H., Nichols, J., Violet, P. C., Pittman, C. A., Fitzhugh, C., Fleming, R. E., Mohandas, N., Tisdale, J. F., & Levine, M. (2017). Measurements of red cell deformability and hydration reflect HbF and HbA2 in blood from patients with sickle cell anemia. *Blood Cells, Molecules, and Diseases*. <https://doi.org/10.1016/j.bcmd.2017.04.005>
26. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *Computational Chemistry*.
27. Quinn, C. T. (2018). L-Glutamine for sickle cell anemia: More questions than answers. In *Blood*. <https://doi.org/10.1182/blood-2018-03-834440>
28. R Core Team. (2014). R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL [Http://Www.R-Project.Org/](http://www.R-project.org/).
29. Salinas Cisneros, G., & Thein, S. L. (2020). Recent Advances in the Treatment of Sickle Cell Disease. *Frontiers in Physiology*, 11(May), 1–15. <https://doi.org/10.3389/fphys.2020.00435>
30. Segal, J. B., Strouse, J. J., Beach, M. C., Haywood, C., Witkop, C., Park, H., Wilson, R. F., Bass, E. B., & Lanzkron, S. (2008). Hydroxyurea for the treatment of sickle cell disease. In *Evidence report/technology assessment*.
31. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11), 2498–2504. <https://doi.org/10.1101/gr.1239303>
32. Steinberg, M. H., Barton, F., Castro, O., Pegelow, C. H., Ballas, S. K., Kutlar, A., Orringer, E., Bellevue, R., Olivieri, N., Eckman, J., Varma, M., Ramirez, G., Adler, B., Smith, W., Carlos, T., Ataga, K., DeCastro, L., Bigelow, C., Sauntharajah, Y., ... Terrin, M. (2003). Effect of Hydroxyurea on Mortality and Morbidity in Adult Sickle Cell Anemia: Risks and Benefits Up to 9 Years of Treatment. *Journal of the American Medical Association*. <https://doi.org/10.1001/jama.289.13.1645>
33. Strouse, J. J., Lanzkron, S., Beach, M. C., Haywood, C., Park, H., Witkop, C., Wilson, R. F., Bass, E. B., & Segal, J. B. (2008). Hydroxyurea for sickle cell disease: A systematic review for efficacy and toxicity in children. *Pediatrics*. <https://doi.org/10.1542/peds.2008-0441>
34. Tame, J. R., & Vallone, B. (2000). The structures of deoxy human haemoglobin and the mutant Hb Tyr α 42His at 120 K. *Acta Crystallographica. Section D, Biological Crystallography*, 56(Pt 7), 805–811. <https://doi.org/10.1107/s0907444900006387>
35. *The Protein Data Bank*. (1998). <https://doi.org/10.2210/pdb1A3N/pdb>
36. Thom, C. S., Dickson, C. F., Gell, D. A., & Weiss, M. J. (2013). Hemoglobin variants: Biochemical properties and clinical correlates. *Cold Spring Harbor Perspectives in Medicine*. <https://doi.org/10.1101/cshperspect.a011858>

37. Wastnedge, E., Waters, D., Patel, S., Morrison, K., Goh, M. Y., Adeloje, D., & Rudan, I. (2018). The global burden of sickle cell disease in children under five years of age: A systematic review and meta-analysis. *Journal of Global Health*. <https://doi.org/10.7189/jogh.08.021103>
38. Yagishita, N., Yamasaki, S., Nishioka, K., & Nakajima, T. (2008). Synoviolin, protein folding and the maintenance of joint homeostasis. In *Nature Clinical Practice Rheumatology*. <https://doi.org/10.1038/ncprheum0699>
39. Zhang, Y. (2007a). Prediction Report Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins*. <https://doi.org/10.1002/prot>
40. Zhang, Y. (2007b). Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins: Structure, Function and Genetics*. <https://doi.org/10.1002/prot.21702>

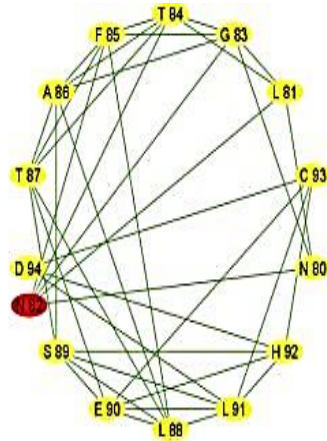
Appendix



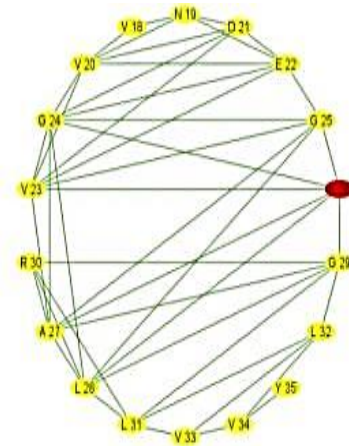
Appendix 1: Subdomain graphs corresponding to subsequences partition of the β -globin. Nested (interaction) graphs were obtained by submitting the subsequences (in Table 1) to I-TASSER for ab-initio modeling and used in Cytoscape for generation of the nested graphs (*refer to methods section for further details*)



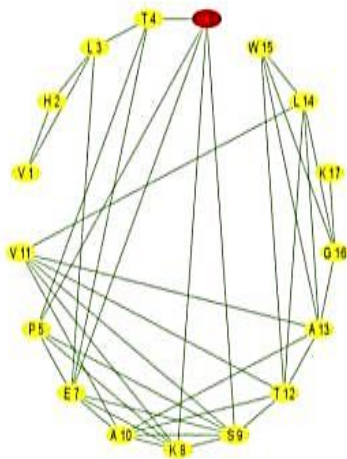
Mutated G8 (K95E)



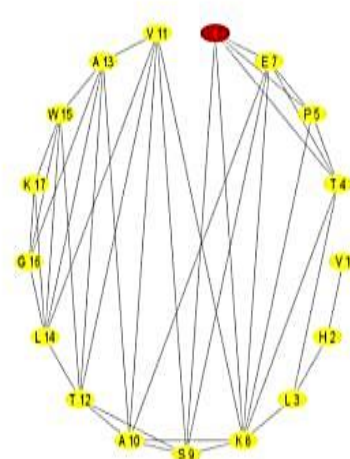
Mutated G7 (K82N)



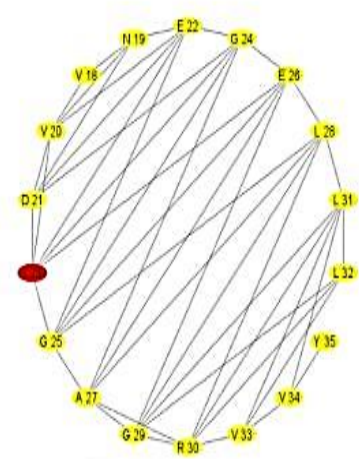
Mutated G2 (E26K)



Mutated G1 (E6V)



Mutated G1 (E6K)



Mutated G2 (V23I)

Appendix 2: Subdomain graphs corresponding to mild and severe SCD mutations. Nested (interaction) graphs were obtained by submitting the mutated subsequences to I-TASSER for ab-initio modeling and used in Cytoscape for generation of the nested graphs (*refer to methods section for further details*)