Hindawi Publishing Corporation
Computational Biology Journal
Volume 2013, Article ID 938169, 9 pages
http://dx.doi.org/10.1155/2013/938169

Research Article

Graph-Theoretic Models of Mutations in the Nucleotide Binding Domain 1 of the Cystic Fibrosis Transmembrane Conductance Regulator

Debra J. Knisley,[1,2] Jeff R. Knisley,[1,2] and Andrew Cade Herron[1]

[1] *Department of Mathematics and Statistics, East Tennessee State University, Johnson City, TN 37614, USA*
[2] *Institute for Quantitative Biology, East Tennessee State University, Johnson City, TN 37614, USA*

Correspondence should be addressed to Debra J. Knisley; knisleyd@etsu.edu

Received 30 November 2012; Revised 12 March 2013; Accepted 12 March 2013

Academic Editor: Alessandra Lumini

Copyright © 2013 Debra J. Knisley et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cystic fibrosis is one of the most common inherited diseases and is caused by a mutation in a membrane protein, the cystic fibrosis transmembrane conductance regulator (CFTR). This protein serves as a chloride channel and regulates the viscosity of mucus lining the ducts of a number of organs. Although much has been learned about the consequences of mutations on the energy landscape and the resulting disrupted folding pathway of CFTR, a level of understanding needed to correct the misfolding has not been achieved. The most common mutations of CFTR are located in one of two nucleotide binding domains, namely, the nucleotide binding domain 1 (NBD1). We model NBD1 using a nested graph model. The vertices in the lowest layer each represent an atom in the structure of an amino acid residue, while the vertices in the mid layer each represent the residue. The vertices in the top layer each represent a subdomain of the nucleotide binding domain. We use this model to quantify the effects of a single point mutation on the protein domain. We compare the wildtype structure with eight of the most common mutations. The graph-theoretic model provides insight into how a single point mutation can have such profound structural consequences.

1. Introduction

Cystic fibrosis is the most common genetic disorder in the Caucasian population. Cystic fibrosis (CF) is caused by a single point mutation in the cystic fibrosis membrane conductance regulator (CFTR) protein [1–4]. CFTR is a chloride channel located in the apical membrane of epithelial cells and plays a fundamental role in transepithelial salt and water movement [5]. A mutation of this protein affects a number of organs in the body such as lungs, pancreas, reproductive organs, and colon. The viscosity of the mucus that lines the ducts of these organs is altered by the increased salt levels resulting in sticky mucus plugs that disrupt the normal function of these organs. A mutation in the CFTR protein occurs in approximately one in every twenty individuals in the Caucasian population and there are more than one thousand nine hundred different reported mutations of CFTR resulting in different levels of severity of clinical consequences [6]. Although there are a large number of reported mutations of CFTR, the deletion of phenylalanine at position 508 (ΔF508) occurs in more than 90% of the CF population [7]. The ΔF508 mutation prevents the correct folding of the protein and consequential degradation [7, 8]. Thus, this mutation results in one of the more severe phenotypes. Once considered a fatal disease, knowledge about the mechanisms and clinical consequences of the mutations of this membrane protein has raised the expected life span to nearly forty years. Despite the extended life expectancy, the quality of life for people with cystic fibrosis is still very affected and the variation in life expectancy is very pronounced.

CFTR is a member of the ATP-Binding Cassette (ABC) transporters. There are forty-nine human ABC transporters, including the multidrug resistance protein (MDR) which thwarts many efforts to utilize chemotherapy to treat various cancers [9]. Similar to most ABC transporters, CFTR contains two membrane spanning domains and two nucleotide binding domains (NBD1 and NBD2). CFTR also contains a regulatory domain (R) which none of the other ABC

transporters have [9]. It is well known that the role of CFTR is multifaceted, serving as both a transporter and an ion channel and regulating the activity of other channels such as the epithelial sodium channel, ENaC [10]. Control of CFTR channel activity is modulated by the phosphorylation of the R domain by protein kinase A. Although a great deal has been learned about the control of CFTR channel gating by phosphorylation and ATP binding/hydrolysis, details about the specific interactions remain unknown and may require the knowledge of the complete 3D structure of the entire protein [11–14]. Due to the size of the protein, this has proven to be difficult.

It is interesting to note that when ΔF508 is successfully folded in vitro, there is very little change in the energy landscape and the folded protein is relatively stable [15]. Thus a great deal of the focus on the treatment of cystic fibrosis is directed towards finding a means by which the protein can escape the degradation tag [1, 16, 17]. This approach alone has proven to be difficult and most likely insufficient because the key question remains: Why is the deletion of this particular amino acid, F508, so catastrophic? In [6] it was demonstrated that self-chaperoning activity is diminished as a consequence of the missing phenylalanine residue. Furthermore, it has been shown that cross-linking between NBD1 which contains F508 and a cytoplasmic loop between two of the membrane spanning helices is critical to the gating mechanism and this is disrupted by the deletion of ΔF508 [3]. Consequently, novel methods that employ a combination of knowledge about the energy landscape and structural information coupled with functional attributes are needed.

A mutation that results in the absence of a single residue in a protein structure has a profound local effect, but how this local perturbation manifests to a global one as in the case of ΔF508 remains unclear. Protein structures in general are replete with mutations that result in a single amino acid substitution or deletion, many of which cause no disruption in the synthesis and functionality of the protein. It is still not clearly understood how and why each of these mutations of CFTR, ΔF508 in particular, causes such a profound effect. To address the key question, how the deletion of phenylalanine at position 508 results in the complete loss of function of CFTR, we build a graph-theoretic model of NBD1, the domain containing ΔF508. With the vertex-weighted hierarchical graph representation of the protein domain NBD1, we present a method to model the effect of a single point mutation of a protein. Using the hierarchical, vertex-weighted graph, we define novel combinatorial descriptors based on these vertex-weights. We employ these graph-theoretic measures to quantify the consequences of nine mutations of CFTR's NBD1, including ΔF508. Each mutation results in a distinct set of graph-theoretic measures that are both local and global and capture the underlying structural "network" consequences of the mutation. Our model reveals a process by which a local change can produce a significant global change. Once we identify the combinatorial measure at the global level that distinguishes a particular mutation, we can reverse our steps to the lower level to see which structural changes in the intermediate level were responsible for these global changes.

Earlier efforts to model proteins as networks with graphs were introduced in [18, 19]. In an earlier work by Haynes et al., we introduced the use of the domination number of a graph to quantify a biomolecule [20]. We used the domination number and variations of the domination number to classify small tree graphs (4–6 vertices) as either RNA-like or not RNA-like. We found the domination number of a graph to be a better means to quantify the structural properties of secondary RNA structure than the second smallest eigenvalue utilized by the RNA database RAG [21]. In [22], we use the domination number, coupled with other graph invariants, to quantify the amino acid residue structures in order to build a predictive model for protein-ligand binding affinity. Although both of these were successful, the authors noted the shortcomings of graphical invariants as molecular descriptors. Namely, in all graphical invariants, the weights of the vertices are assumed to be one. In fact, the reason that these measures are called invariants is because they are invariant under isomorphism. However, two vertex-weighted graphs whose (nonweighted) structures are isomorphic may have *different* values when these measures incorporate the vertex-weights. The measures that we define, although derived from well-established graphical invariants [23, 24], are no longer invariant under isomorphism since the weights of the vertices are incorporated into the definition of the measure. Consequently we have termed these values *combinatorial descriptors*. Our method of building nested vertex-weighted graphs is described below.

2. Materials and Methods

2.1. Overview of Graph-Theoretic Model. We model NBD1 with a series of nested graphs. First, each of the twenty most common amino acids is the modeled as a graph. Given an amino acid, the backbone and central carbon atom are represented by a single vertex and each of the atoms in the corresponding amino acid residue structure is represented by a vertex. Vertices in the residue are weighted by the nearest integer value of the mass of the corresponding atom. Edges are determined by molecular bonds and hydrogen atoms are ignored. Using each of these hydrogen suppressed models of the twenty most common amino acids, we obtain twenty corresponding vectors of descriptors based on the graph-theoretic measures: weighted domination, weighted diameter, circumference and weighted periphery. We also use a measure of polarity found in [25] and a measure of hydrophobicity found in [26]. These measures were used and can be found in [22].

We next partition the sequence of CFTR that corresponds to the NBD1 domain into eight subsequences. In particular, we obtain the following subsequences denoted by $S_i : S_1$, S_2, S_3, S_4, S_5, S_6, S_7, and S_8. In determining these subsequences, we are guided by the secondary structures of the protein. Each subsequence contains one and only one type of secondary structure, either a beta strand, an alpha helix, or a loop. The loop regions may contain turns, a 3/10-helix or an alpha helix with no more than 6 residues. The corresponding subsequences of each subdomain are provided in Table 1.

Table 1: Subsequences of the sequence for NBD1 of CFTR.

Subdomain	Subsequence	Subdomain	Subsequence
S_1	ISFCSQFSWIMPGT 488–501	S_5	KDNIVL**G**EGGITL**S** 536–549
S_2	IKENIIFGVSYD 502–513	S_6	E**G**QQAKISLARAVY 550–563
S_3	EYRYRSVIKA 514–523	S_7	KDADLYLLDSPFG 564–576
S_4	CQLEEDISKFAE 524–535	S_8	YLDVLTTEKEIFESCVCKL 577–594

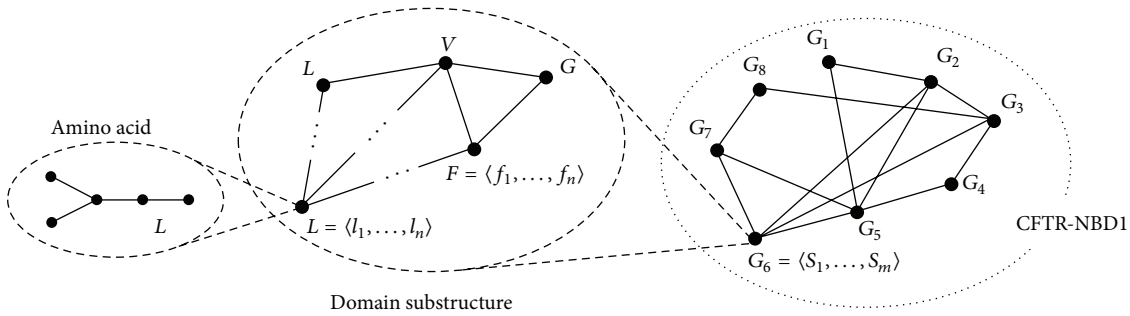


Figure 1: The nested graph model.

We model the structure corresponding to each S_i as a graph to obtain eight graphs which we denote by G_i. The vertices of G_i, each, represent a residue in the sequence S_i and edges are determined by a proximity measure of eight angstroms. The distance endpoints are determined by the center of mass of each residue in the 3D structure provided in 2BBO [27] in the Protein Data Bank [28]. For example, the subdomain graph G_2 which contains F508 is given in Figure 2. At the highest level, we represent the NBD1 as a single graph with eight vertices, one for each subdomain graph G_i. Each graph G_i is represented by a vertex and edges are determined by proximity in the 3D structure provided in [27]. The edges of the NBD1 CFTR Graph, or simply the domain graph G, are based on a proximity measure where the distance endpoints are determined by a threshold distance between any two residues of each subdomain. In conclusion, the nested graph has three layers. At the lowest level we have a collection of twenty small vertex-weighted graphs, one for each of the twenty most common amino acids. At the middle level, we have a collection of eight vertex-weighted graphs, G_i in which each vertex represents an amino acid and the weights of the vertices are the combinatorial descriptors of the amino acid graphs at the lower level. At the highest level, we have a vertex-weighted graph G that represents the nucleotide binding domain NBD1. The vertices, each, represent one of the subdomain graphs G_i and the weights assigned to these vertices are derived from the vertex-weighted graph descriptors of each G_i. Using these measures we obtain a final set of graph-theoretic-based measures for the domain graph G of NBD1. This modeling scheme is illustrated in Figure 1.

Having obtained a set of measures for the wildtype NBD1 domain, we select eight disease causing mutations found in the Cystic Fibrosis Mutation Databank [6] that occur in NBD1. Given that we have selected eight mutations to model, we now obtain a set of graph-theoretic measures for each mutation in the following way. To measure the global structural effect of a single point mutation, we first make the corresponding change at the residue level. This change affects one of the subdomains S_i. We obtain a new subdomain graph G_i of the effected subdomain by utilizing I-TASSER, an online protein folding server [29]. For example, in Figure 2, we show G_2 containing F508 and the graph with the predicted structural changes as a consequence of deleting F508.

For each mutation, we obtain a new graph for the subdomain where the mutation occurred. Note that an amino acid switch mutation produces a graph that contains a different residue together with different set of combinatorial descriptors. Since both the structure of G_i and possibly a vertex-weight for one vertex in G_i have changed, this changes the corresponding vector of vertex-weights for the vertex G_i in G. In this way we incorporate the graph-theoretic changes of a single point mutation with the predicted structural change by using the vertex-weights at each level. The edge set of the domain graph G remains unchanged, but the weights of the vertices are adjusted according to the structural (both vertex and edge) changes of the underlying subdomain graphs. Since our measures are based on vertex-weights, we obtain a new set of values associated with each mutation. These, together with the wildtype, provide a set of measures for nine distinct graphs. Using MATLAB [30], we calculate

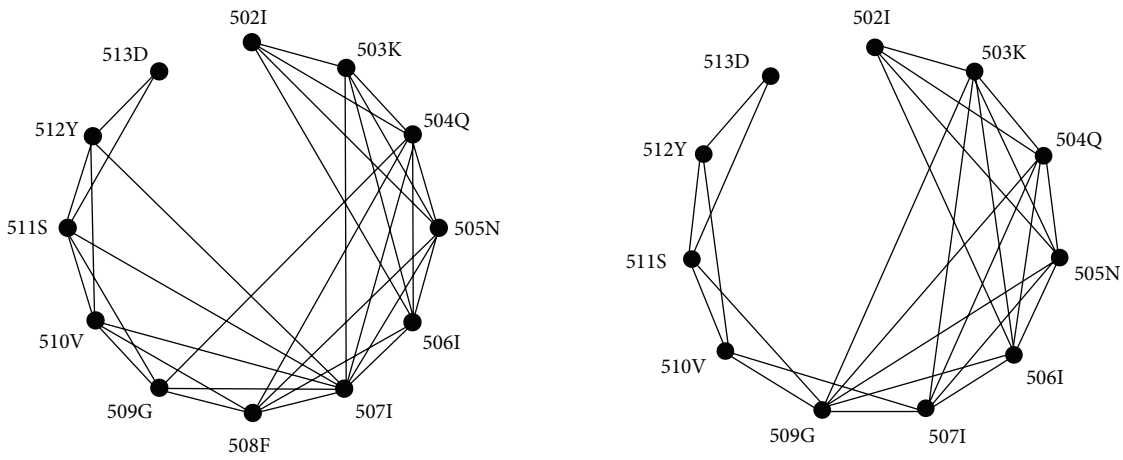


Figure 2: Subdomain graph G_2 with 508F and without 508F.

the differences of these nine distinct domain graph measures. We describe this process in more detail in the sections below.

2.2. Level One: The Amino Acid Graphs. We represent each of the hydrogen suppressed residue structures as a rooted, vertex-weighted graph. The central carbon atom of the amino acid serves as the root and the atomic masses measured to the nearest integer value serve as the vertex-weights. We quantify the weighted graphical structure using combinatorial measures from graph theory. In a combinatorial optimization setting, graph measures are typically defined by determining the maximum (minimum) values under a specified constraint. For example, a set of vertices is said to dominate a graph if all remaining vertices are adjacent to at least one vertex in the dominating set. Trivially, the entire vertex set is a dominating set; thus the optimal set is one of minimum cardinality. We adjust the definitions of three combinatorial measures to include the weights of the vertices and use these measures as biomolecular descriptors. We provide the definitions of these vertex-weighted combinatorial measures as follows.

(i) Weighted domination number: a vertex set S of a graph G is a set of vertices with the property that every vertex in the vertex set of G not in S adjacent to at least one vertex in S is a dominating set of vertices. A minimum dominating set in a vertex-weighted graph is dominating set whose vertex sum is a minimum. The *weighted domination number* is this minimum sum.

(ii) Weighted diameter: the distance between two vertices u and v is the length of the shortest path from u to v, which is equivalent to the number of vertices encountered along the shortest graph traversal from u to v, not including u as an encounter. The weighted distance is the minimum weight among all paths from u to v where the minimum weight is the smallest sum of the weights of the vertices encountered along the graph traversal from u to v, not adding the weight of u. The *weighted diameter* of a graph is the maximum weight among all minimum weighted distances in the graph.

(iii) Weighted periphery: a vertex whose distance from the center of a graph is maximum is known as a peripheral vertex. In this work, the graphs we use to model the residue chains were considered to be rooted at the central carbon atom of the backbone and we measured from the root as opposed to the center. Thus the *periphery* is the maximum weight among all vertices whose distance from the root is maximum.

We did not use the vertex-weights for the circumference, the fourth measure.

(i) Circumference: a cycle in a graph is a path whose begin (start) vertex equals the end vertex. Note that the standard graph-theoretic definition of a path requires that no vertex be repeated in the traversal of the path; hence in a cycle we make the exception for the begin/end vertex. The *circumference* of a graph is the length (or number of vertices) in the largest cycle of the graph. If a graph does not have a cycle, then we say that it has circumference number zero.

2.3. Level Two: The Subdomain Graphs G_i of NBD1. We represent each of these eight subsequences as a graph by reading in the 3D coordinates of the subsequence from the Protein Data Bank of the structure found in [27]. Each vertex represents a residue and the edges between residues are determined by a maximum threshold of 8 angstroms where the endpoints of the edges are the corresponding centers of mass of each residue. Table 1 contains the respective subsequences for each subdomain. In order to model a single point mutation, we replace a given amino acid with a different amino acid, or we delete a single amino acid. This new sequence of the subdomain determined by the mutation is then submitted to I-TASSER to obtain a predicted 3D structure of the mutated subdomain. Therefore, a single point mutation affects one and only one of the eight subdomain

graphs. The predicted structure is used to create the new subdomain graph S_i corresponding to the mutation. A large number of measures were calculated for each S_i based on the size and structure of these subdomain graphs. For this study we had selected a number of measures and used these as weights for the vertices of the domain graph G. We use these to define combinatorial measures for the G.

2.4. Level Three: The NBD1 Domain Graph G. The domain graph has eight vertices, one for each subdomain as defined in the above section. Edges are based on proximity of the 3D structure. Rather than using the centers of mass as we did for both levels one and two, here we use the backbone as reference points, and if any two alpha carbons are within threshold proximity, we apply an edge. Combinatorial descriptors for the domain graph are defined to measure the effects, both local and global, of the mutation. We rely on much of the work in Chemical Graph Theory for the selection of our measures. For example, many of the classical topological indices utilize the number of distinct paths of length 3 or overlapping paths of length 3 [31, 32]. We also select graphical invariants whose measures can reflect structural changes, even in a very small graph. The first combinatorial descriptor we define later measures the edge density. All of the definitions that we used for the combinatorial descriptors for this work are given as follows.

2.4.1. Total Circumference Degree Ratio Minimum Degree. Each of the vertices in the subdomain graphs are labeled with the circumference measure of the corresponding residue. The total circumference is the sum of these vertex-weights. The ratio of this subdomain total circumference to the subdomain degree in G is the vertex-weight of each (subdomain graph) vertex in G. The vertices of each G_i are weighted with this ratio. The circumference weighted degree of a vertex v in G is defined to be the sum of the weights of the neighbors of v and the minimum circumference weighted degree of G is denoted by $\delta_{\sum C/\deg}$.

Note. We do not consider any vertex to be in its own neighborhood. If a mutation occurs in a subdomain graph causing a change to that vertex-weight, it will not affect this measure of that subdomain. Rather the measure will be reflected in a change among its proximity neighbors. We define similar weighted degree measures below.

2.4.2. Total Hydrophobic Maximum Degree. Each vertex in the subdomain graphs G_i is labeled with the hydrophobic measure of the underlying residue graph. The total hydrophobic measure for G_i is the sum of these vertex-weights. The weighted degree of a vertex v in G is the sum of the weights of the neighbors of v in G. The maximum (total hydrophobic weighted) degree in G is denoted by $\Delta_{\sum H}$.

2.4.3. Total Hydrophobic Minimum Degree. Each vertex in the subdomain graphs G_i is labeled with the hydrophobic measure of the underlying residue graph. The total hydrophobic measure for G_i is the sum of these vertex-weights. The weighted degree of a vertex v in G is the sum of the weights of the neighbors of v in G. The minimum (total hydrophobic weighted) degree in G is denoted by $\delta_{\sum H}$.

2.4.4. Total Polarizability Maximum Degree. Each vertex in the subdomain graphs G_i is labeled with the *polarizability* measure of the underlying residue graph. The *total polarizability* measure for G_i is the sum of these vertex-weights. The weighted degree of a vertex v in G is the sum of the weights of the neighbors of v in G. The maximum (total polarizability-weighted) degree in G is denoted by $\Delta_{\sum P}$.

2.4.5. Total Polarizability Minimum Degree. Each vertex in the subdomain graphs G_i is labeled with the polarizability measure of the underlying residue graph. The total polarizabilty measure for G_i is the sum of these vertex-weights. The weighted degree of a vertex v in G is the sum of the weights of the neighbors of v in G. The minimum (total polarizability-weighted) degree in G is denoted by $\delta_{\sum P}$.

2.4.6. Total of the Edge/Vertex Ratio. Each vertex in G is weighted with the ratio of edges to vertices in the underlying subdomain graph. The sum of the vertices in G is the overall weight of the graph which we denote by $\Sigma e/v$.

2.4.7. Minimum Triangular Edge/Vertex-Weight. There are fourteen triangles in G. The weight of a triangle is the sum of the vertex-weights defining the triangle. We use the edge/vertex ratio for the vertex-weights as before mentioned and find the minimum weighted triangle in G which we denote by $\omega_{e/v}$.

2.4.8. Minimum Clique-Weighted P_3. We weight the vertices with the clique number of the underlying subdomain graph. The clique number of a graph is the order (number of vertices) of the largest complete subgraph. We then assign weights to the edges of the domain graph where the edge weight is the sum of the two end vertices. We now define the weight of a path of length two (or P_3) as the sum of the edge weights in the P_3. For example, the subdomain graphs S_1, S_9, and S_8 have clique numbers 4, 4, and 5, respectively, and so the two edges have weights 8 and 9 and the path has weight 17. Note that this approach gives more "weight" to the central vertex, which is intentional. We denote this by P_3^{ω}.

2.4.9. Minimum Double Domination Center. The eccentricity value of a vertex v in a graph is t if every vertex can be reached from v within a distance of t and not from any value less than t. The subgraph induced by the set of all vertices with minimum eccentricity in the graph is called the *center* of the graph. The vertices of the subdomain graph are labeled with the residue domination number and then the weighted domination number of the subdomain graph is calculated and becomes the weight of the vertex in the domain graph. We define the average weight of the center of the domain graph using this double-domination measure and we denote it by C_{γ}.

Table 2: Combinatorial descriptors.

	$\Delta_{\sum H}$	$\delta_{\sum H}$	$\Delta_{\sum P}$	$\delta_{\sum P}$	$\delta_{\sum C/\deg}$	$\omega_{e/v}$	$\sum e/v$	P_3^{ω}	C_γ	ΔI_p^d
WT	16.7	−22.5	14.754	7.57	6.857	6.447	21.602	17	36	14.855
ΔI507	12.2	−22.0	14.568	7.38	6.857	6.030	21.390	17	36	14.521
ΔF508	13.9	−22.0	14.464	7.28	6.307	6.030	21.390	17	36	14.455
G542A	16.7	−22.0	14.800	7.62	6.857	6.447	21.245	11	36	15.355
S549N	16.7	−22.0	14.826	7.64	6.857	6.447	21.387	17	36	14.730
S549I	16.7	−22.0	14.878	7.69	6.857	6.447	21.459	11	36	15.355
S549R	16.7	−22.0	14.983	7.80	6.857	6.447	21.459	17	36	14.730
G551D	14.0	−24.7	14.754	7.68	6.857	6.447	22.102	17	12	14.855
R560T	20.5	−18.2	14.754	7.39	6.857	6.447	22.102	17	36	14.855

2.4.10. Periphery Diameter Maximum Influence. Each vertex in the subdomain graph is weighted by the corresponding periphery value of the underlying residue. The weighted diameter of the subdomain graph is calculated and then this value is assigned to the corresponding vertex in the domain graph G. We determine the maximum weighted degree divided by the number of edges incident to each vertex to measure this periphery-diameter maximum influence, denoted by ΔI_d^p.

2.5. Modeling the Mutations. To model a single point mutation, such as the substitution mutation G542A, we find that 542 is in S_5. We change the residue in the specified position 542 from the amino acid G to A in S_5 and submit this short sequence to I-TASSER [29]. We use the returned predicted coordinates to construct the new G_5 subdomain graph. Thus, only one of the subdomain graphs will change, namely, G_5. Or, as in the case of Figure 2, we submit the sequence of S_2 without 508F to I-TASSER to obtain the 3D coordinates of the resulting predicted structure of ΔF508-S_2. We are therefore using the knowledge of the biophysical properties incorporated by the I-TASSER algorithm rather than simply relabeling the graph. Now that we have the predicted structural change, we incorporate the graph-theoretic change. Using the amino acid descriptors described earlier, we relabel the vertex with the new associated vector and recalculate the combinatorial measures of that subdomain using the combinatorial descriptors defined before to produce a new vector to be associated with that subdomain. Consequently, in the domain graph, the subdomain (vertex) receives a new set of values and we then recompute the measures of the NBD1 domain graph G. The values for each of the nine graphs are given in Table 2.

3. Results and Discussion

To analyze the results from the changes in the combinatorial measures due to a single point mutation, we employed MATLAB. We first calculate the p-distance for the data using the Euclidean distance measure. Using these results, we determined the default linkage and the corresponding dendrogram that is shown in Figure 3. The first resulting dendrogram reveals that G551D has graphical consequences more distinct than the other mutations due to its relative distance from all of the other mutations.

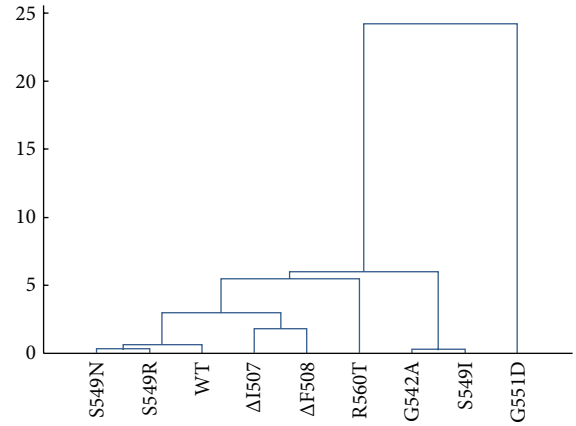


Figure 3: Dendrogram using all of the combinatorial measures.

It is well known that the consequences of G551D mutation on CFTR are distinct from ΔF508 [6]. Whereas CFTR is altogether absent from the membrane surface in patients with ΔF508, this is not the case for patients with the G551D mutation. The protein CFTR is present at the surface; however, the G551D-CFTR gating mechanism is faulty and thus the clinical consequences are similar. To correct the G551D defect, one needs to find a small molecule that can increase the efficiency of the gating mechanism which proved to be more easily addressed. At this time, G551D is the only mutation for which there exists a treatment that addresses the molecular consequences of the mutation rather than the clinical outcomes. Vertex Pharmaceuticals recently received FDA approval for a drug (ivacaftor) marketed as Kalydco, which uses a small molecule stabilizer of the mutant protein [33]. There are no other treatments available for the remaining mutations, although a combinatorial library of small molecules potential correctors exists for CFTR.

Also of interest in the resulting dendrogram is the association of S549I with G542A, rather than with S549R

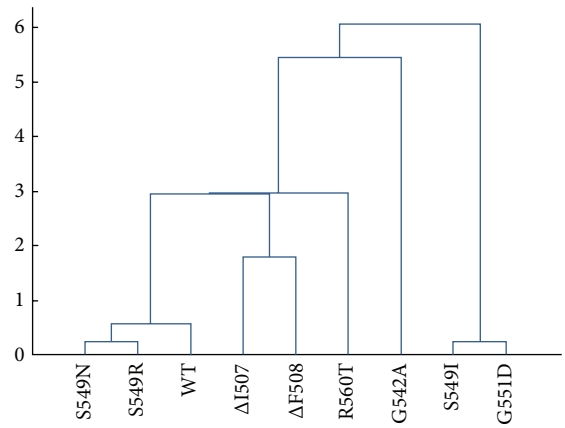


Figure 4: Dendrogram without the combinatorial measures based on double domination.

and S549N. To further investigate S549I, we use different combinations of descriptors and the consequences of these variations. In Figure 4, we show the result of removing the minimum double domination center measure from the model. In doing so, we now find that S549I is closely associated with G551D.

This points to an advantage of quantifying the mutations by combinatorial measures since we can now return to the measures to see which graph-theoretic properties are driving these distinctions. Removing one of the particular measures that brings S549I closer to G551D allows us to investigate exactly which amino acid residues are key in the changed measure. Consequently, since there is an effective treatment for G551D, we can reveal what must be addressed in order for this treatment to also be effective for S549I. In particular, by employing the principles of reverse engineering, we can identify which residues are critical for the calculation of the minimum double domination center. The effects of ivacaftor on CFTR channel gating and chloride transport were tested in cells expressing different CFTR gating mutations, among them S549N and S549R [33]. Ivacaftor was effective in addressing defects in CFTR channel gating but not nearly as effective as the correction achieved in G551D. Our model predicts that S541I is a more likely candidate to be corrected by ivacaftor since a slight adjustment in the combinatorial measures resulted in a closer association of S549I with G551D. Our model may also help determine which small molecules will be more effective for the S549N and S549R mutations.

Future work will include a complete list of vertices and the corresponding residues that participate in the calculation of the minimum double domination center C_γ as well as other associations that can be found. The results we report here are preliminary, but we feel that they are worth reporting due to the novelty of the approach. Perhaps the greatest distinction of the model is the realization that the structural properties of proteins can be quantified by graph-theoretic measures that are based solely on combinatorial optimization methods and these quantities do not rely on biophysical or biochemical properties. Thus, when these methods are coupled together, more information can be obtained than by either one alone. An increased understanding of the effects of a single-point mutation will help guide molecular targets for future drug design efforts.

Much of the emphasis in computational biology and computational chemistry involves the mining of large data sets in response to the deluge of data coming from the biological sciences arena. Necessarily, these modeling schemes and resulting algorithms do not transcend well to small data sets. Small data sets with very high similarity among the elements in the set cannot be mined using algorithms intended for large, more diverse sets. There has been a call, from the pharmaceutical industries especially, for more computational models that are designed for smaller, highly similar sets. High-throughput analysis may successfully reduce tens of thousands of potential small molecules down to hundreds of candidates. But given the high cost of clinical trials and the high rate of failure of many of these trials, additional models designed specifically for further refinement are a topic of research interest in structure-based drug design. With this work, we propose that combinatorial graph theory can be effectively utilized to define biomolecular descriptors when graphical invariants are modified to incorporate vertex-weights. These vertex-weights are not to be considered labels; rather they should contain structural information, especially with respect to the particular biomolecule under study.

4. Conclusion

Knowledge regarding the consequences of ΔF508 and other mutations is essential for the rational design of drugs for the treatment of cystic fibrosis. We have shown that meaningful information can be obtained by adding graph-theoretic modeling to the toolbox. This information, together with strategies to determine changes in the energy landscape, will help address the consequences of this mutation as well as provide a guide for applying this approach to other diseases caused by a single point mutation or possibly a small set of mutations.

Nucleotide binding domains of ATP-binding proteins are highly conserved and contain a well-described set of motifs. The most commonly occurring mutation that causes cystic fibrosis is located in the nucleotide binding domain 1 of the cystic fibrosis transmembrane conductance regulator. In order to guide the rational design of a corrector molecule for the mutant cystic fibrosis protein, more must be learned about the consequences of the mutation on the protein domain. It is now understood that the deletion of phenylalanine at 508 causes a number of disruptions, but the exact mechanisms are not fully understood. We expect that the information revealed in this work will provide a new direction in the work to find a cure for cystic fibrosis.

Given that the definitions are motivated by graphical invariants in graph theory such as in [23, 24] together with those defined in chemical graph theory [31, 32] and given the extensive amount of graphical invariants in mathematics, there exists a wealth of resources for novel combinatorial descriptors that can be utilized as quantifiers for biomolecular structures.

Abbreviations

CF: Cystic fibrosis
CFTR: Cystic fibrosis transmembrane conductance regulator
NBD1: Nucleotide binding domain 1
ΔF508: The deletion of phenylalanine at position 508.

Acknowledgment

The authors of this paper thank the authors of I-TASSER for their protein structure prediction tool.

References

[1] K. Roberts, P. Cushing, P. Boisguerin, D. Madden, and B. Donald, "Computational Design of a PDZ domain peptide inhibitor that rescues CFTR activity," *PLOS Computational Biology*, vol. 8, no. 4, Article ID e1002477, 2012.

[2] A. Aleksandrov, P. Kota, L. Cui et al., "Allosteric modulation balances thermodynamic stability and restores function of ΔF508 CFTR," *Journal of Molecular Biology*, vol. 419, pp. 41–60, 2012.

[3] A. W. R. Serohijos, T. Hegedus, A. A. Aleksandrov et al., "Phenylalanine-508 mediates a cytoplasmic-membrane domain contact in the CFTR 3D structure crucial to assembly and channel function," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 9, pp. 3256–3261, 2008.

[4] L. He, A. A. Aleksandrov, A. W. R. Serohijos et al., "Multiple membrane-cytoplasmic domain contacts in the cystic fibrosis transmembrane conductance regulator (CFTR) mediate regulation of channel gating," *Journal of Biological Chemistry*, vol. 283, no. 39, pp. 26383–26390, 2008.

[5] T. C. Hwang and D. N. Sheppard, "Gating of the CFTR Cl- channel by ATP-driven nucleotide-binding domain dimerisation," *Journal of Physiology*, vol. 587, no. 10, pp. 2151–2161, 2009.

[6] The Cystic Fibrosis Mutation Database, http://www.genet.sickkids.on.ca/.

[7] A. W. R. Serohijos, T. Hegedus, J. R. Riordan, and N. V. Dokholyan, "Diminished self-chaperoning activity of the ΔF508 mutant of CFTR results in protein misfolding," *PLoS Computational Biology*, vol. 4, no. 2, Article ID e1000008, 2008.

[8] A. A. Aleksandrov, P. Kota, L. A. Aleksandrov et al., "Regulatory insertion removal restores maturation, stability and function of ΔF508 CFTR," *Journal of Molecular Biology*, vol. 401, no. 2, pp. 194–210, 2010.

[9] D. B. Luckie, J. H. Wilterding, M. Krha, and M. E. Krouse, "CFTR and MDR: ABC transporters with homologous structure but divergent function," *Current Genomics*, vol. 4, no. 3, pp. 109–121, 2003.

[10] B. K. Berdiev, Y. J. Qadri, and D. J. Benos, "Assessment of the CFTR and ENaC association," *Molecular BioSystems*, vol. 5, no. 2, pp. 123–227, 2009.

[11] G. Seavilleklein, N. Amer, A. Evagelidis et al., "PKC phosphorylation modulates PKA-dependent binding of the R domain to other domains of CFTR," *American Journal of Physiology—Cell Physiology*, vol. 295, no. 5, pp. C1366–C1375, 2008.

[12] H. A. Lewis, X. Zhao, C. Wang et al., "Impact of the ΔF508 mutation in first nucleotide-binding domain of human cystic fibrosis transmembrane conductance regulator on domain folding and structure," *Journal of Biological Chemistry*, vol. 280, no. 2, pp. 1346–1353, 2005.

[13] S. Y. Huang, D. Bolser, H. Y. Liu, T. C. Hwang, and X. Zou, "Molecular modeling of the heterodimer of human CFTR's nucleotide-binding domains using a protein-protein docking approach," *Journal of Molecular Graphics and Modelling*, vol. 27, no. 7, pp. 822–828, 2009.

[14] T. Hegedus, A. W. R. Serohijos, N. V. Dokholyan, L. He, and J. R. Riordan, "Computational studies reveal phosphorylation-dependent changes in the unstructured R domain of CFTR," *Journal of Molecular Biology*, vol. 378, no. 5, pp. 1052–1063, 2008.

[15] X. Wang, J. Matteson, Y. An et al., "COPII-dependent export of cystic fibrosis transmembrane conductance regulator from the ER uses di-acidic exit code," *Journal of Cell Biology*, vol. 167, no. 1, pp. 65–74, 2004.

[16] M. F. N. Rosser, D. E. Grove, and D. M. Cyr, "The use of small molecules to correct defects in CFTR folding, maturation, and channel activity," *Current Chemical Biology*, vol. 3, no. 1, pp. 100–111, 2009.

[17] N. Pedemonte, G. L. Lukacs, K. Du et al., "Small-molecule correctors of defective ΔF508-CFTR cellular processing identified by high-throughput screening," *Journal of Clinical Investigation*, vol. 115, no. 9, pp. 2564–2571, 2005.

[18] A. del Sol, H. Fujihashi, D. Amoros, and R. Nussinov, "Residues crucial for maintaining short paths in network communication mediate signaling in proteins," *Molecular Systems Biology*, vol. 2, Article ID 2006.0019, 2006.

[19] M. Habibi, C. Eslahchi, M. Sadeghi, and H. Pezashk, "The interpretation of protein structures based on graph theory and contact map," *Open Access Bioinformatics*, vol. 2, pp. 127–137, 2010.

[20] T. Haynes, D. Knisley, E. Seier, and Y. Zou, "A quantitative analysis of secondary RNA structure using domination based parameters on trees," *BMC Bioinformatics*, vol. 7, article 108, 2006.

[21] RAG: RNA-As-Graphs, http://www.biomath.nyu.edu/.

[22] D. Knisley and J. Knisley, "Predicting protein-protein interactions using graph invariants and a neural network," *Computational Biology and Chemistry*, vol. 35, no. 2, pp. 108–113, 2011.

[23] D. West, *Introduction to Graph Theory*, Prentice Hall, New York, NY, USA, 1996.

[24] J. Bondy and U. S. R. Murty, *Graph Theory*, Springer, New York, NY, USA, 2010.

[25] M. Charton and B. I. Charton, "The structural dependence of amino acid hydrophobicity parameters," *Journal of Theoretical Biology*, vol. 99, no. 4, pp. 629–644, 1982.

[26] J. Kyte and R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein," *Journal of Molecular Biology*, vol. 157, no. 1, pp. 105–132, 1982.

[27] H. A. Lewis, C. Wang, X. Zhao et al., "Structure and dynamics of NBD1 from CFTR characterized using crystallography and hydrogen/deuterium exchange mass spectrometry," *Journal of Molecular Biology*, vol. 396, no. 2, pp. 406–430, 2010.

[28] The Protein Databank, http://www.pdb.org/.

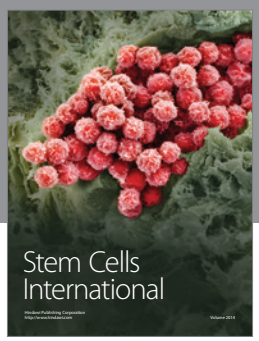
[29] I-TASSER, http://zhanglab.ccmb.med.umich.edu/I-TASSER/.

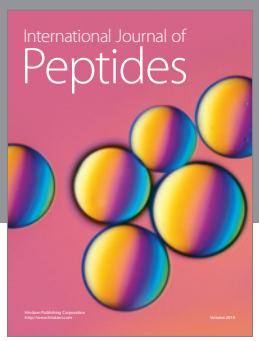
[30] MATLAB, http://www.mathworks.com/products/matlab/index.html.

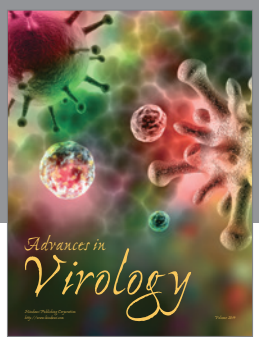
[31] D. Bonchev and D. H. Rouvray, *Chemical Graph Theory: Theory and Fundamentals*, Gordon and Breach, New York, NY, USA, 1991.

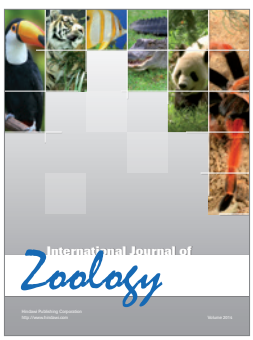
[32] N. Trinajstic, *Chemical Graph Theory*, CRC Press, Boca Raton, Fla, USA, 2nd edition, 1992.

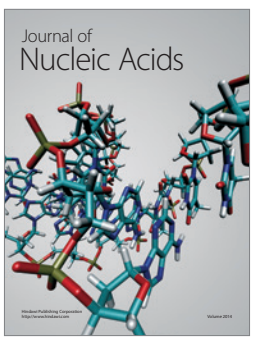
[33] H. Yu, B. Burton, C. Huang et al., "Ivacaftor potentiation of multiple CFTR channels with gating mutations," *Journal of Cystic Fibrosis*, vol. 11, no. 3, pp. 237–245, 2012.

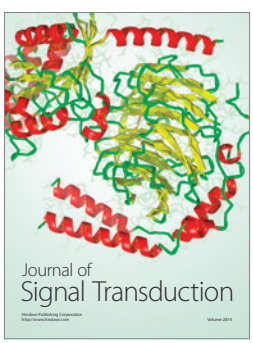


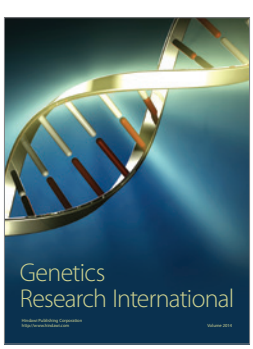


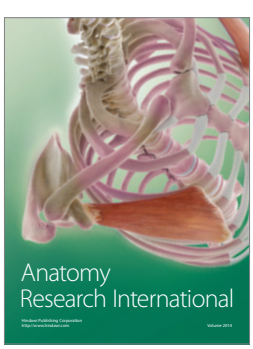


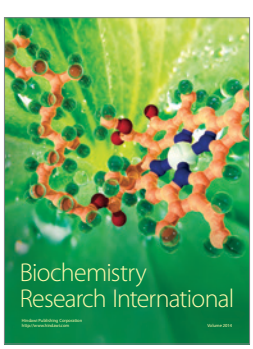


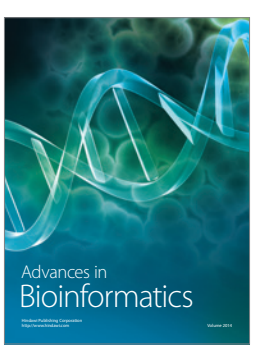


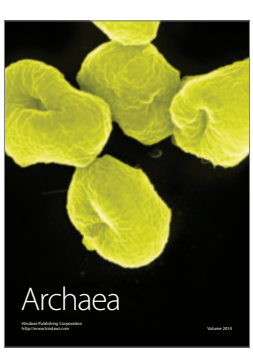


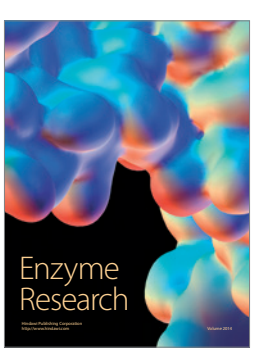


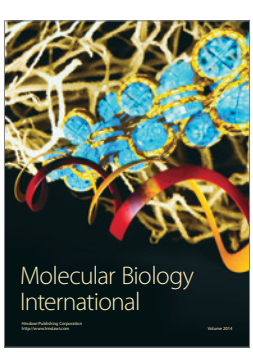












Hindawi

Submit your manuscripts at
http://www.hindawi.com