Structure-Based Analysis of Single Nucleotide Variants in the Renin-Angiotensinogen Complex

David K. Brown, Olivier Sheik Amamuddy, Özlem Tastan Bishop
Grahamstown, South Africa

ABSTRACT

Background: The renin-angiotensin system (RAS) plays an important role in regulating blood pressure and controlling sodium levels in the blood. Hyperactivity of this system has been linked to numerous conditions including hypertension, kidney disease, and congestive heart failure. Three classes of drugs have been developed to inhibit RAS. In this study, we provide a structure-based analysis of the effect of single nucleotide variants (SNVs) on the interaction between renin and angiotensinogen with the aim of revealing important residues and potentially damaging variants for further inhibitor design purposes.

Objectives: To identify SNVs that have functional and potentially damaging effects on the renin-angiotensinogen complex and to use computational approaches to investigate how SNVs might have damaging effects.

Methods: A comprehensive set of all known SNVs in the renin and angiotensinogen proteins was extracted from the HUMA database. This dataset was filtered by removing synonymous and missense variants and using the VAPOR pipeline to predict which variants were likely to be deleterious. Variants in the filtered dataset were modeled into the renin-angiotensinogen complex using MODELLER and subjected to molecular dynamics simulations using GROMACS. The residue interaction networks of the resultant trajectories were analyzed using graph theory.

Conclusions: This research identified important SNVs in the interface of RAS and showed how they might affect the function of the proteins. For instance, the mutant complex containing the variant P40L in angiotensinogen caused instability in the complex, indicating that this mutation plays an important role in disrupting the interaction between renin and angiotensinogen. The mutant complex containing the SNV A188V in renin was shown to have significantly increased fluctuation in the residue interaction networks. D104N in renin, associated with renal tubular dysgenesis, caused increased rigidity in the protein complex comparison to the wild type, which probably in turn negatively affects the function of RAS.

The renin-angiotensin system (RAS) is responsible for the regulation of blood pressure and sodium homeostasis [1]. This is achieved by producing angiotensin II, a potent, 8-residue vasoconstrictor, which causes the arterioles to constrict, resulting in increased blood pressure. Angiotensin II also stimulates the release of aldosterone, which increases the rate at which sodium ions are reabsorbed into the blood [2].

When RAS is activated, the juxtaglomerular cells in the kidneys secrete renin into the blood. Once in the blood, renin cleaves angiotensin I, a 10-residue peptide, from a plasma protein known as angiotensinogen, which originates in the liver. Angiotensin I is then converted to angiotensin II by angiotensin-converting enzyme, which cleaves a further 2 residues from the former [2].

Hyperactivity of RAS has been linked to high blood pressure (hypertension), congestive heart failure, and kidney disease. As such, various classes of drugs have been developed to inhibit this system including angiotensin-converting enzymeinhibitors [3], angiotensin receptor blockers [4], and renin inhibitors [5].

In this study, we have used structural bioinformatics and network analysis techniques to investigate the effect of nonsynonymous single nucleotide variants (SNVs) on renin-angiotensinogen interaction to identify important residues and potentially damaging SNVs.

MATERIAL AND METHODS

Data retrieval

Sequences and suitable Protein Data Bank (PDB) [6] structures for renin and angiotensinogen were identified via a search of the Human Mutation Analysis (HUMA) [7] database. The amino acid sequences were then downloaded from Uniprot. Structures were evaluated for suitability based on their coverage of the target sequences and their PDB validation metrics, before being downloaded from the PDB. All available SNVs for each protein were then downloaded from the HUMA database.

The authors report no relationships that could be construed as a conflict of interest.
This work is supported by the National Institutes of Health Common Fund under grant number U41HG006941 to H3ABioNet and National Research Foundation (NRF), South Africa, (grant number 93690).
The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the funders.
Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.gheart.2017.01.006.
From the Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University, Grahamstown, South Africa.
Correspondence: Ö. Tastan Bishop (o.tastanbishop@ru.ac.za).

GLOBAL HEART

VOL. 12, NO. 2, 2017
ISSN 2211-8160/$36.00.
http://dx.doi.org/10.1016/j.gheart.2017.01.006

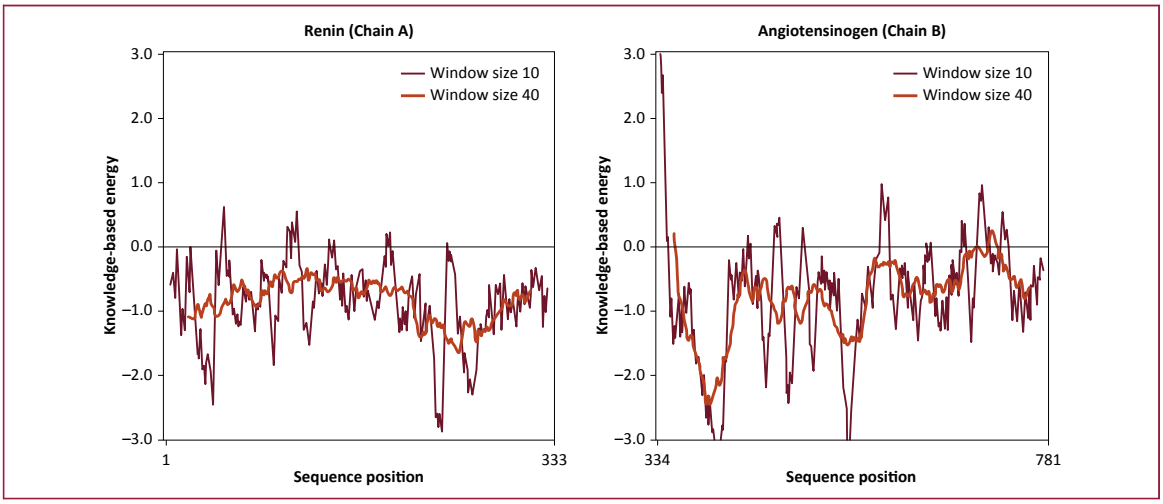

FIGURE 1. PROSA results for the renin-angiotensinogen complex.

Homology modeling: wild type

To account for missing residues in existing experimental structures, the renin-angiotensinogen complex was modeled using MODELLER [8]. The complex has been solved in 2X0B, which was used as the main template. Renin was covered by chain A of 2X0B. Three additional templates, 2WXY, 2WXW, and 2WXZ were used to cover gaps in chain B (i.e., angiotensinogen). Alignment of the templates to the target sequences was performed using PROMALS3D [9].

After aligning, the first 73 residues of the renin target sequence and the first 32 residues and last 3 residues of the angiotensinogen target sequence were not covered by the templates. These residues were trimmed from the alignment as a result. One hundred models were then generated using very slow refinement.

Model evaluation

One hundred models of the renin-angiotensinogen complex were generated. The top 3 models were then selected based on their DOPE *z*-score [10], and further evaluated using PROCHECK [11], VERIFY3D [12], and PROSA [13] to validate that they were indeed accurate models. The best model was then chosen based on the combination of these results.

SNV filtering

The SNV data set, obtained from the HUMA database, contains all SNVs from dbSNP [14] that could be mapped to the renin and angiotensinogen protein sequences based on their chromosome coordinates. As such, the resulting data set contained synonymous and nonsense SNVs.

TABLE 1. SNV data set for angiotensinogen

dbSNP ID	Residue Change	Location	Reason for Inclusion
rs539231427	H39R	Interface	• Highly damaging prediction by VAPOR • Interacts with position in renin where SNV occurs
rs746613821	P40L	Interface	• Highly damaging prediction by VAPOR • Interacts with position in renin where SNV occurs
rs41271499	L43F	Interface	• Highly damaging prediction by VAPOR • Interacts with position in renin where SNV occurs
rs760531325	E48K	Interface	• Interacts with position in renin where SNV occurs
rs751752211	S49G	Interface	• Interacts with position in renin where SNV occurs
rs377047370	S49N	Interface	• Highly damaging prediction by VAPOR • Interacts with position in renin where SNV occurs
rs201406560	A104T	Interface	• Interacts with position in renin where SNV occurs
rs767370325	M105V	Interface	• Interacts with position in renin where SNV occurs
rs756744141	D168Y	Interface	• Highly damaging prediction by VAPOR • Interacts with position in renin where SNV occurs

SNV, single nucleotide variation.

TABLE 2. SNV data set for renin

dbSNP ID	Residue Change	Location	Reason for Inclusion
rs868694193	D104N	Interface	• Highly damaging prediction by VAPOR • Interacts with position in angiotensinogen where SNV occurs
rs191049685	R148C	Interface	• Interacts with position in angiotensinogen where SNV occurs
rs371704012	R148H	Interface	• Interacts with position in angiotensinogen where SNV occurs
rs770190833	A188V	Interface	• Highly damaging prediction by VAPOR
rs752426689	L318R	Interface	• Interacts with position in angiotensinogen where SNV occurs
rs201922371	F319V	Interface	• Interacts with position in angiotensinogen where SNV occurs

SNV, single nucleotide variation.

Synonymous SNVs result in no change in the amino acid sequence. They are, therefore, unlikely to have any structural effects on the protein, and were removed from the data set. Nonsense SNVs, on the other hand, result in early stop codons. These SNVs are highly likely to be damaging as they usually result in nonfunctioning proteins. As such, there was little reason to further analyze these SNVs in this paper and they were also removed from the data set.

The Variant Analysis Portal (VAPOR) [15], a workflow that combines the results of PolyPhen-2 [16], Provean [17], PhD-SNP [18], PANTHER-PSEP [19], and FATHMM [20] to predict the functional effect of SNVs, was then used to predict which of the remaining SNVs were likely to cause conformational changes that would alter the functionality of the protein. If more than 1 program predicted that the SNV was neutral, we removed it from the data set. One exception to this is if there was a known disease-association in the HUMA database.

Lastly, we used the Protein Interactions Calculator [21] to determine which residues in the complex were interacting. The final data set contained all SNVs that were at the interface between the 2 proteins and were either predicted to be damaging or were interacting with a position at which a SNV occurred in the other protein.

Homology modeling: mutants

The SNVs identified in the previous step were introduced into the structures via homology modeling. Models were generated for each SNV individually. Additional models were then generated where SNVs occurred at the interacting positions in both renin and angiotensinogen. In these cases, the mutants were introduced to both protein sequences in the complex before being modeled. This resulted in a combination of models that accounted for each SNV occurring on its own as well as SNVs that might co-occur with SNVs in interacting positions—potentially compensating SNVs.

The same modeling methodology was used as for the wild type. 2X0B, 2WXY, 2WXW, and 2WXZ, were once again used as templates. Alignment was performed using PROMALS3D and 100 models were generated for each modeling run. The best model for each mutant and mutant combination was then picked based on DOPE *z*-scores. As these models were very similar to the wild type, it was not deemed necessary to repeat the other evaluation methods to confirm the quality.

Molecular dynamics

Homology models produced in the previous steps were free from any nonstandard residues or missing atoms and were thus ready for molecular dynamics (MD) simulations. The simulations were performed for each of the complexes using GROMACS 5.1 [22] on 480 CPU cores at the Centre for High Performance Computing (Cape Town, South Africa). The all-atom AMBER03 force field was employed for topology generation and energy calculations. Simple point

TABLE 3. Mutant models of the renin-angiotensinogen complex

Model	DOPE Z-Score	Renin (Chain A) Mutants	Angiotensinogen (Chain B) Mutants
ren_D104N	−1.20	rs868694193	
ren_R148C	−1.19	rs191049685	
ren_R148H	−1.20	rs371704012	
ren_A188V	−1.19	rs770190833	
ren_L318R	−1.18	rs752426689	
ren_F319V	−1.17	rs201922371	
ang_H39R	−1.18		rs539231427
ang_P40L	−1.18		rs746613821
ang_L43F	−1.18		rs41271499
ang_E48K	−1.18		rs760531325
ang_S49G	−1.21		rs751752211
ang_S49N	−1.21		rs377047370
ang_A104T	−1.20		rs201406560
ang_M105V	−1.18		rs767370325
ang_D168Y	−1.19		rs756744141
ren_D104N_ang_L43F	−1.20	rs868694193	rs41271499
ren_R148C_ang_E48K	−1.17	rs191049685	rs760531325
ren_R148C_ang_S49G	−1.19	rs191049685	rs751752211
ren_R148C_ang_S49N	−1.20	rs191049685	rs377047370
ren_R148H_ang_E48K	−1.18	rs371704012	rs760531325
ren_R148H_ang_S49G	−1.20	rs371704012	rs751752211
ren_R148H_ang_S49N	−1.20	rs371704012	rs377047370
ren_L318R_ang_A104T	−1.20	rs752426689	rs201406560
ren_F319V_ang_A104T	−1.19	rs201922371	rs201406560
ren_F319V_ang_M105V	−1.20	rs201922371	rs767370325

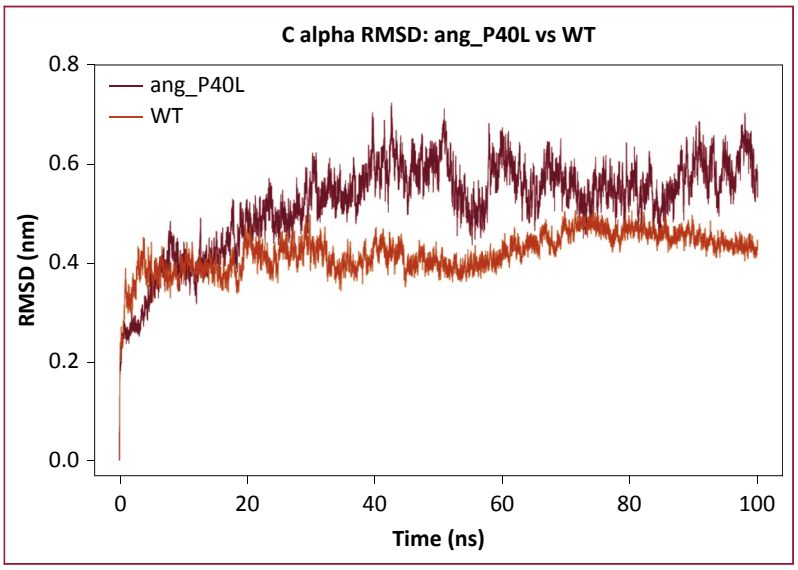

FIGURE 2. Root mean square deviation (RMSD) of the complex containing the single nucleotide variant, P40L, in angiotensinogen. This is the only mutation tested that destabilizes the complex.

charge water was added for solvation, after which the system was neutralized using 0.15 M NaCl in a triclinic periodic box with a clearance space of 1.5 Å from the protein. The system was then relaxed by energy minimization using the method of steepest descent with a force tolerance of 1000 kJ/mol/nm capped at an upper limit of 50,000 steps. Short-range cutoffs (van der Waals and Coulombic interactions) were set at 1 nm each, while long-distance electrostatics were handled by the particle-mesh Ewald algorithm [23]. Temperature was equilibrated at 310 K over a period of 100 ps, using the modified Berendsen thermostat [24]. Pressure equilibration ensued, using the Parrinello-Rahman barostat [25] to maintain the pressure at 1 bar. Production dynamics were finally run over 100-ns periods with 2 fs time steps, writing coordinates every 5,000 steps. The LINCS algorithm [26] was applied to correct for rotational bond lengthening during the MD runs.

Analysis of MD trajectories

Preliminary root mean square deviation (RMSD) analysis revealed jumps that occurred in several complexes despite our first attempts to correct for periodic boundary conditions using molecular center of mass. Trajectories were thus sequentially corrected for periodic boundary conditions by first making the proteins whole, followed by the removal of jumps across boundaries and finally centering the protein inside the simulation box. RMSD was computed from Cα atoms after least square fitting along the respective protein backbones. Additionally, root mean square fluctuation (RMSF) calculations were evaluated to monitor residue motion (averaged over residue atoms) across the entire production run.

Dynamic network analysis (evaluation at SNV locations)

In order to calculate the residue interaction network (RIN) over each production MD frame, multi-PDB files were first generated using GROMACS tool for trajectory conversion. An in-house Python script was then designed to predict residue interaction by evaluating pairwise residue distances between all Cβ (or Cα in the case of glycine) atoms for each complex, looped across all frames. A cutoff distance of 6.7 Å [27] was used to define any residue-residue contact (contact = 1, no contact = 0) around designated SNV positions. A total of 10,001 frames (in PDB format) were traversed to generate edge lists that were used by an in-house R script—using the igraph library [28]—to produce a weighted adjacency matrix, which was converted to frequencies by dividing by the number of frames, similar to the approach used by Doshi et al. [29]. Edge weights were displayed as log2 rescaled values, while the contact frequencies were used as edge labels. The weighted RINs of the wild type proteins were compared to their mutant counterparts at analogous positions to assess gain or loss of contact.

Dynamic network analysis (ΔL)

The shortest path (L_{ij}) between 2 nodes, *i* and *j*, in a network is equal to the minimum number of edges that must be traversed to reach *j* from *i*. The L_i to a node, *i*, is the average of all the shortest paths to *i* from all other

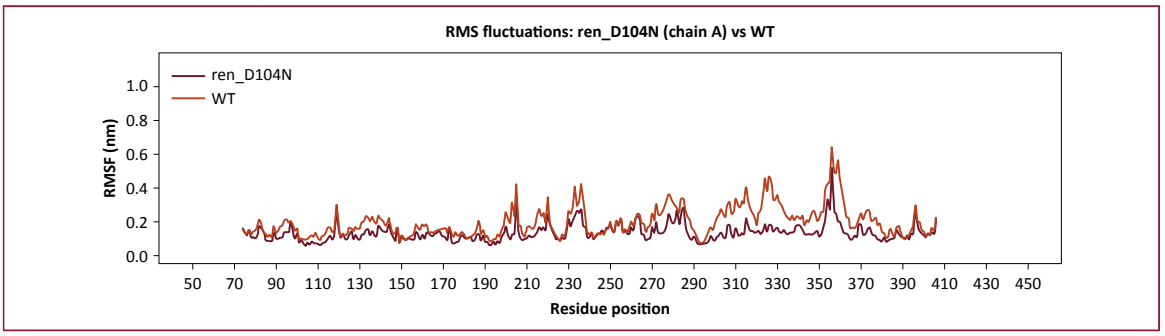

FIGURE 3. Root mean square fluctuation (RMSF) of ren_D104V. Increased rigidity is noticeable from residue 260 to 380 of renin (chain A) in disease-associated model, ren_D104V. WT, wild type.

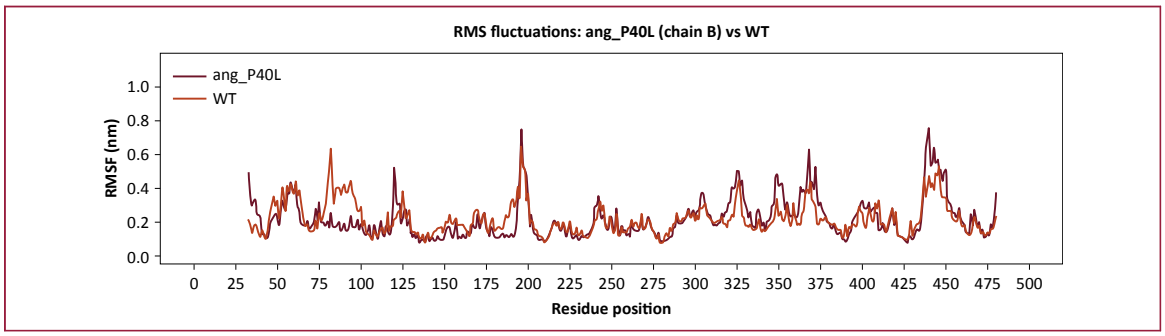

FIGURE 4. Root mean square fluctuation (RMSF) ang_P40L. Increased fluctuation is noticeable in ang_P40L between residue 300 and 380 of angiotensinogen (chain B).

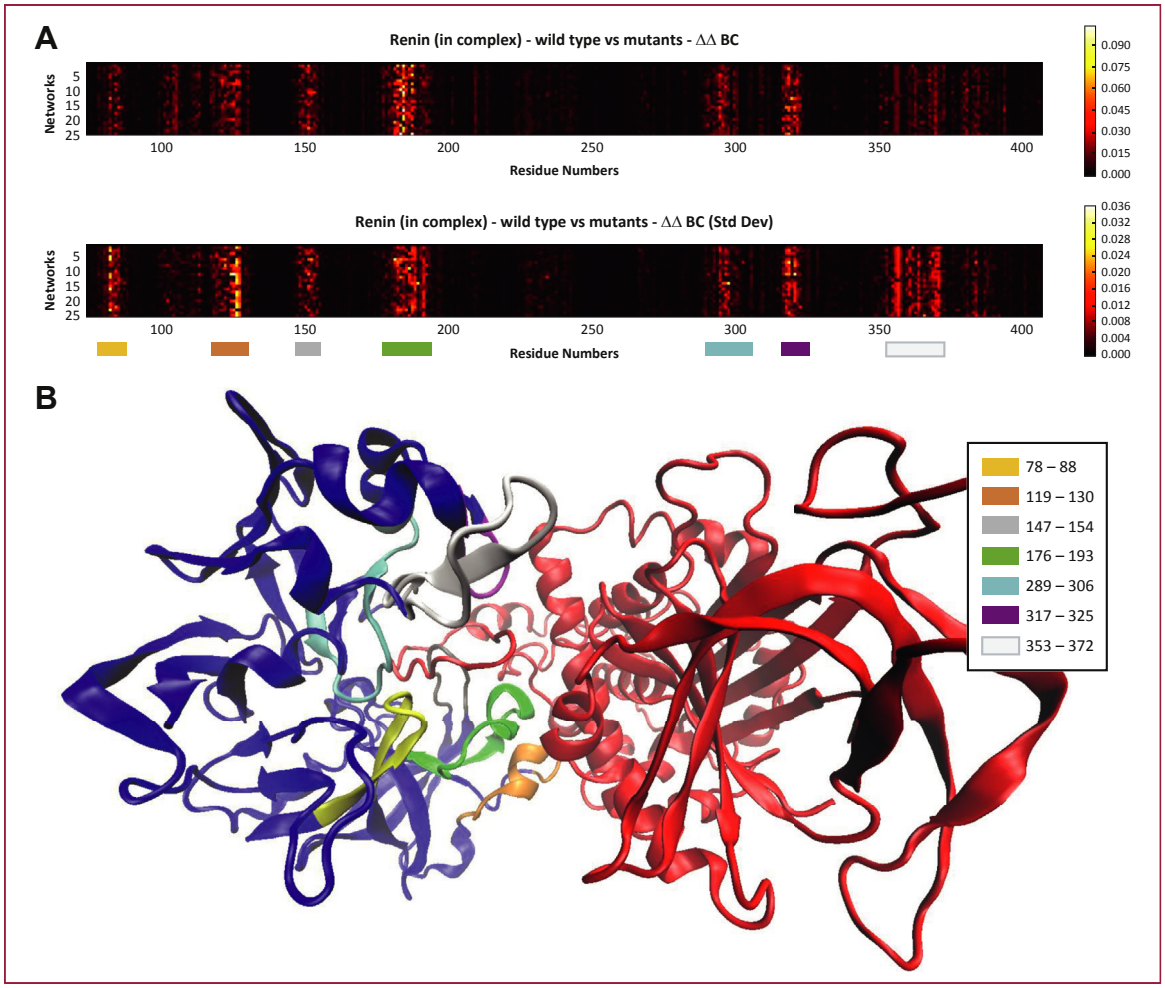

FIGURE 5. (A) A heat map depicting the absolute difference between change in betweenness centrality (ΔBC) in the wild type (first row) and ΔBC in the mutants (ΔΔBC) in chain A (renin) of the complex. Most differences are confined to the color-coded regions. (B) The renin-angiotensinogen complex. The areas consisting of large changes in BC in renin (red) have been mapped to the structure.

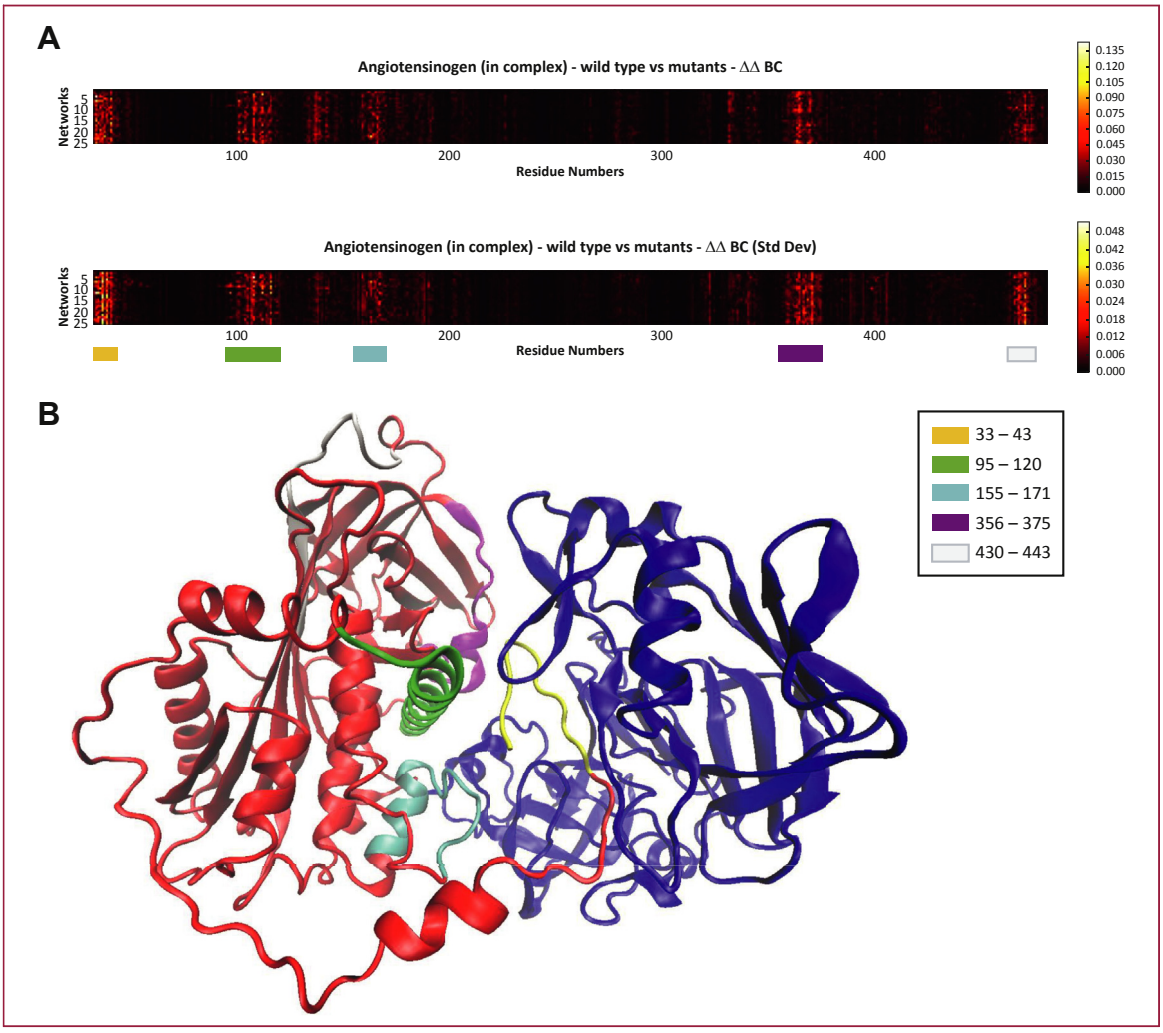

FIGURE 6. (A) A heat map depicting the absolute difference between change in betweenness centrality (ΔBC) in the wild type (first row) and ΔBC in the mutants (ΔΔBC) in chain B (angiotensinogen) of the complex. (B) The areas consisting of large changes in BC in angiotensinogen (blue) have been mapped to the complex structure.

nodes in the network. Given an RIN, where each node is a residue in a protein structure, Dijkstra's shortest path algorithm [30] can be used to produce an N × N matrix of all versus all shortest path lengths, where N is the number of residues in the protein. Getting the mean of each column in this matrix results in an N × 1 matrix, where each value corresponds to the average shortest path length to the respective residue.

For every MD run, an implementation of Dijkstra's shortest path algorithm in the NetworkX Python library [31], in conjunction with the NumPy python library [32], was used in a custom Python script to calculate the average path matrices for RINs at 10-ns intervals (including the RIN at time = 0). As such, for each 100-ns MD run, 11 average shortest path matrices were calculated.

The mean and standard deviation of these matrices were then calculated. The mean of the matrices represents the average L of each of the residues over the 100-ns run, while the standard deviation represents how much L of each residue fluctuates over the 100 ns. The values were plotted to an average L versus residue number graph using Matplotlib [33].

In addition to calculating L, ΔL/L was calculated for all residues after each 10ns interval by getting the difference between the L at time = 0 and L at the respective time point and normalizing by dividing by the original L. The mean and standard deviation of ΔL/L over the 100 ns were calculated in the same way as described previously and plotted using Matplotlib.

Dynamic network analysis (ΔBC)

Given a node, *i*, betweenness centrality (BC_i) is a measurement of how often shortest paths in a network run

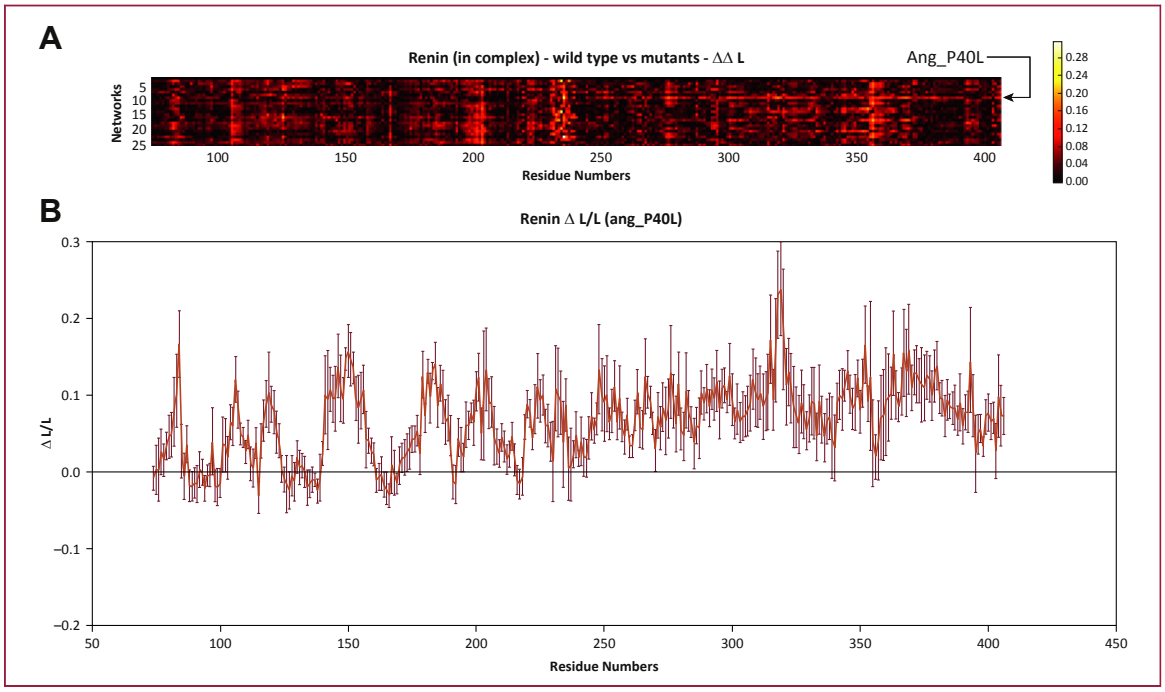

FIGURE 7. (A) Heat map representing the absolute difference in ΔL/L between the wild type renin and the mutants. The mutant complex, ang_P40L, stands out from residue 245 until the end of renin. (B) Closer inspection of ang_P40L shows an elevated *Δ*L/L for the later part of renin, while the standard deviation, depicted by the green error bars, remains relatively small.

through *i*. As such, it is a measure of how central a node is for efficient navigation of the network. BC can be calculated using Brandes algorithm [34].

For each MD run, an implementation of Brandes algorithm in the NetworkX library was used to compute BC for all RINs at 1-ns intervals (including the RIN at time = 0). This resulted in 101 *N x 1* matrices. As with *Δ*L/L, *Δ*BC was calculated by finding the difference between the BC at each time point and BC at time = 0 (BC values were already normalized using the NetworkX library

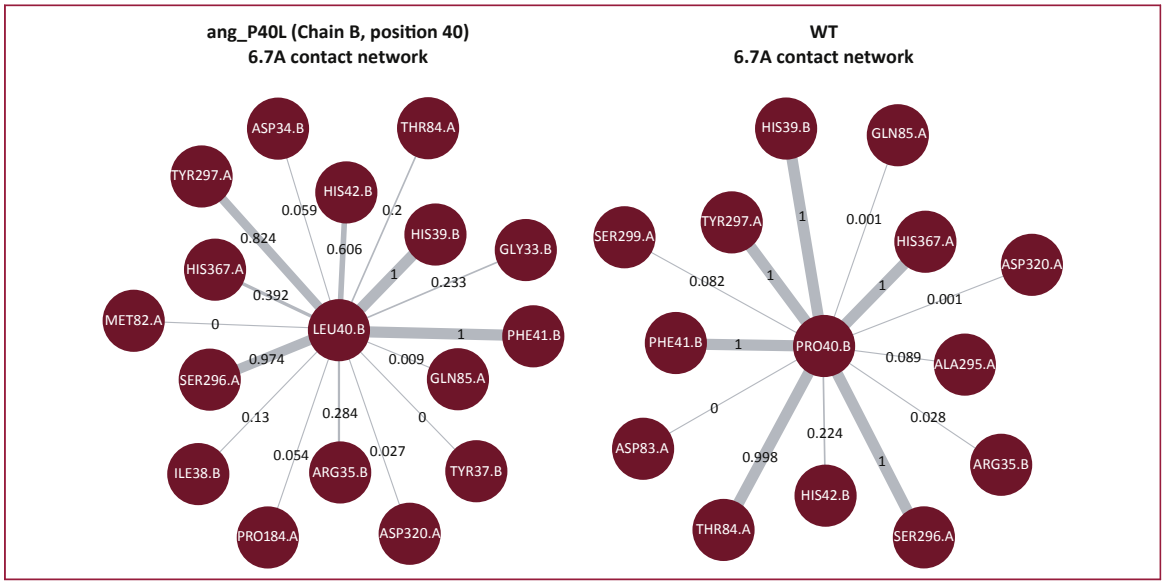

FIGURE 8. Residue contact map of ang_P40L (left) and the wild type (WT) (right) complexes. Interaction with HIS367.A and THR84.A is severely reduced in the mutant complex.

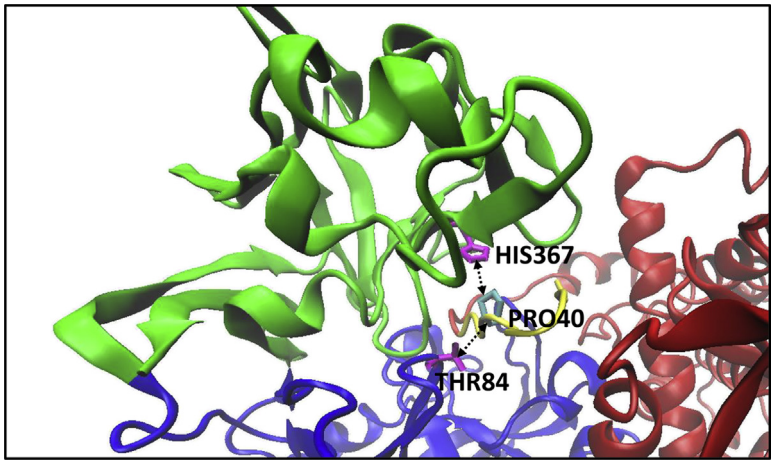

FIGURE 9. Interface of renin-angiotensinogen wild type. Residues HIS367 and THR84 in renin interact with PRO40 in angiotensinogen. In ang_P40L, Proline is replaced with Leucine and the interaction is severely reduced. In addition, ΔL/L for the region corresponding to residue 245 to 407 (green). As such, it appears that the loss of these 2 interactions affects the accessibility of this region.

function). The mean ΔBC and standard deviation of ΔBC for each residue were then calculated as described for ΔL/L.

RESULTS

Homology modeling: wild type

Homology models of the renin-angiotensinogen complex were generated using MODELLER. All models produced were evaluated using DOPE Z-Scores, PROCHECK, PROSA, and VERIFY3D. The top model of the complex had a DOPE z-score of -1.20, while PROCHECK calculated that 90.2% of residues were in most favored regions, and VERIFY3D calculated that 80.8% of residues had a 3-dimensional $-$ 1-dimensional score ≥ 0.2. PROSA results were also positive and are depicted in Figure 1.

SNV filtering

All 317 SNVs for renin and 212 SNVs for angiotensinogen were downloaded from HUMA. Synonymous and nonsense SNVs were removed from the data set. The remaining 130 renin SNVs and 198 angiotensinogen SNVs were submitted to VAPOR for functional predictions. For angiotensinogen, VAPOR produced results from PolyPhen-2, Provean, PhD-SNP, FATHMM, and PANTHER-PSEP (Appendix 1). If more than 1 of these tools predicted that an SNV would not have a deleterious effect, the SNV was removed from the data set. The only exception to this was if the SNV was associated with a disease in the HUMA database. In this case, it was retained. The same process was followed with renin (Appendix 2). In renin's case, PANTHER-PSEP was unable to produce results. The resulting data set held 85 damaging angiotensinogen SNVs and 39 damaging renin SNVs.

We further filtered this data set by only focusing on residues at the interface between renin and angiotensinogen. All damaging SNVs in the interface were retained in the data set. In addition, if a renin SNV was found to be interacting with an SNV in angiotensinogen, both were retained, regardless of the VAPOR predictions. The final data set contained 9 angiotensinogen SNVs (Table 1) and 6 renin SNVs (Table 2).

Homology modeling: mutants

In total, 25 mutant models of the renin-angiotensinogen complex were produced (Table 3). Models were named based on the SNV that they contained and the chain that the SNV occurred in. For example, in model ang_P40L, an SNV occurs at position 40 in angiotensinogen (chain B) and results in a substitution of proline for leucine. Similarly, in ren_A188V, an SNV occurs at position 188 in renin (chain A) and results in a substitution of alanine for valine.

In addition to introducing SNVs into a structure individually, mutant models were produced with combinations of SNVs. If an SNV occurred at a position in renin that interacts with a position in angiotensinogen where another SNV occurred, a model was produced containing both of these potentially compensating SNVs. For example, model ren_D104N_ang_L43F contains an SNV at position 104 in renin and position 43 in angiotensinogen and, in the wild type, these positions were calculated to be interacting.

Molecular dynamics

The wild type complex and the 25 mutant complexes were subjected to 100-ns MD simulations. Of the 26 simulations, 25 completed successfully. Only ren_L318R failed due to a bad contact with water. Due to already having a large amount of data and the fact that L318R in renin was not predicted to be damaging by VAPOR, we decided not to resolve this.

Of the 25 successful simulations (Appendix 3), only ang_P40L had still not stabilized after the full 100-ns (Fig. 2). The SNV, P40L, occurs in the loop region of angiotensinogen, which is cleaved by renin to become angiotensin 1. As such, the instability may indicate that this SNV plays an important role in the binding of renin to angiotensinogen.

RMSF was calculated for all successful MD runs (Appendix 4). The overall trend across mutant complexes revealed an increase in rigidity compared to the wild type, especially in the region from residue 260 to 380 in renin. This increased rigidity was noticeable in the complex containing the renin SNV, D104N (Fig. 3), which, according to HUMA, is associated with renal tubular dysgenesis.

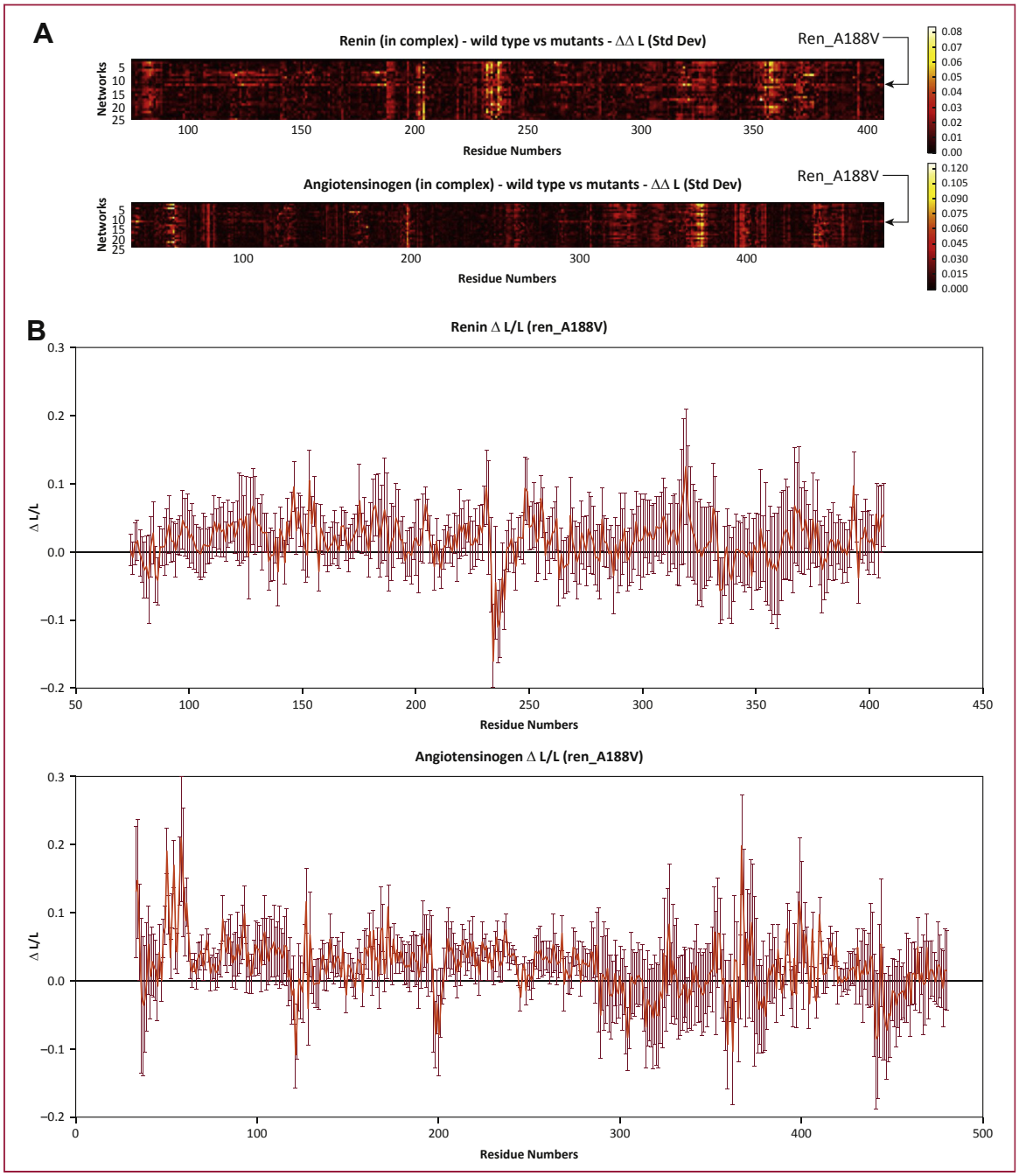

FIGURE 10. (A) A heat map depicting the difference in the standard deviation of the average ΔL/L between wild type and the mutants. The mutant complex, ren_A188V, appears to fluctuate more than the other complexes. (B) Large error bars in the ΔL/L plot indicate that ren_A188V fluctuated heavily over the 100 ns MD simulation.

The difference in rigidity is less pronounced in other complexes, however, such as ang_P40L (Fig. 4). Although ang_P40L was not able to stabilize over 100 ns, it seems that most of the fluctuation occurs in the angiotensinogen side of the complex. In angiotensinogen, ang_P40L has increased RMSF between residues 300 to 380. Only ren_A188V and ren_R148C_ang_E48K show signs of similar fluctuation in this region.

Almost all mutations resulted in increased rigidity from residue 75 to 100 in angiotensinogen. Additionally, increased rigidity was noticeable between residue 175 and 200 in angiotensinogen. The latter region is located on the

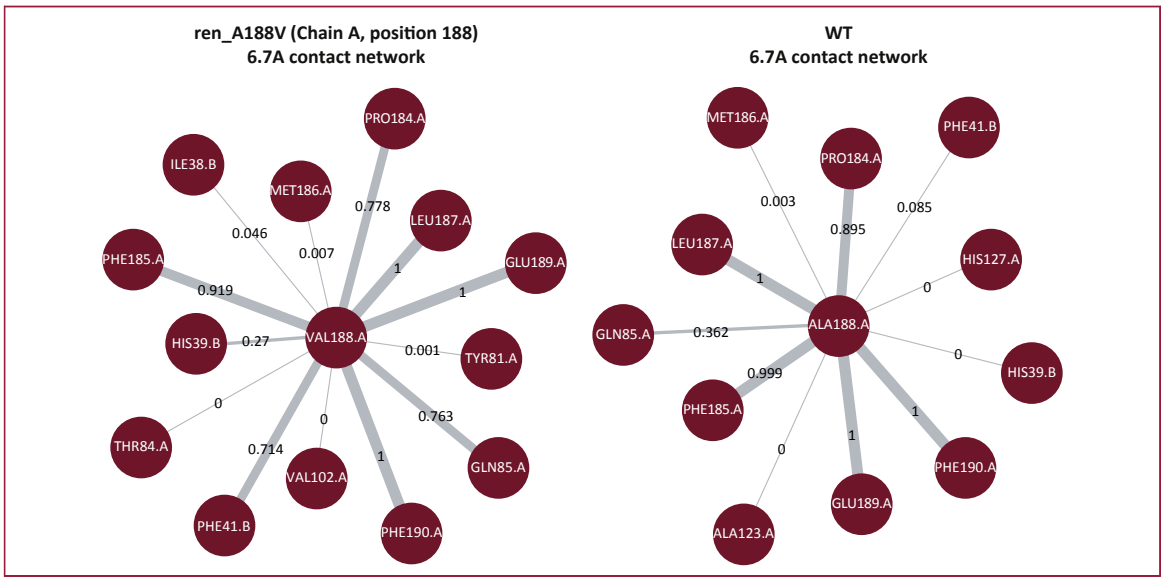

FIGURE 11. The residue contact map for ang_A188V (left) and the wild type (WT) (right). The interaction between PHE41.B occurs more often in the mutant complex.

inside of the former, while also being near to the interface of the 2 proteins. It is possible that mutations occurring in the interface help to hold this region steady, which in turn holds the region containing residues 75 to 100 steady.

A spike in the RMSF was seen between residue 435 and 450 in a number of the mutant models. This is a loop region far from the interface and, as such, high flexibility is not surprising here, nor is it likely to be damaging.

Network analysis

Changes in the RINs of the models were measured by calculating the average and standard deviation of ΔBC (Appendix 5) and ΔL/L (Appendix 6) for each MD simulation. The standard deviation provided a measurement of how much BC and L fluctuated over the course of the run, while the average values provided an idea of what regions experienced a more permanent increase or decrease in BC and L. Additionally, residue contact maps were calculated for the positions where the SNVs occurred (Appendix 7).

In terms of BC, mutations produced a fairly consistent pattern of changes to the network when compared to the wild type. In renin, significant changes could be seen in 7 regions, depicted in Figure 5. Interestingly, mutations resulted in significant changes to the network in regions far from where the SNV was located. Additionally, affected regions were centered around the interface between the proteins. This is not surprising, as the interface is a high traffic zone for network communication and, as such, mutations that shift these areas slightly could have a significant impact on interprotein communication.

Mutations have a similar effect on angiotensinogen. Changes in BC are most notable in 5 regions (Fig. 6). Although 4 of these regions were located at the complex interface, the region consisting of residues 430 to 443 (white) did not contain any interacting residues.

Regions in the complex revealed by the analysis of ΔBC may depict areas that reorganize to compensate for mutations. Such reorganization would be important for the protein to retain its function when mutations occur.

When analyzing ΔL/L, 2 mutants, in particular, caught the eye. The first was ang_P40L. The average ΔL/L for ang_P40L was high across most of the second half of the renin sequence (from approximately residue 245 until the end of the sequence), as can be seen in Figure 7. Looking at the contact network for residue 40 of angiotensinogen (Fig. 8), we can see that 2 prominent contacts in the wild type (PRO40.B-THR84.A and PRO40.B-HIS367.A) that are maintained consistently throughout the simulation of the wild type (99.8% and 100% of the time, respectively) are reduced to occasional contacts in the mutant (20% and 39.2% of the time, respectively). Upon further examination of the structure (Fig. 9), we see that THR84.A and HIS367.A are important residues for providing access to the part of the structure with increased ΔL/L. As such, it can be seen that the loss of these contacts results in decreased accessibility to the back half of renin—an increased ΔL/L means that the average path to these residues is increased. Further work is required to determine whether this loss in accessibility would result in negative effects, such as a reduced ability of renin to cleave angiotensin 1 from angiotensinogen.

The second interesting mutant was ren_A188V. In contrast to ang_P40L, the average ΔL/L remained relatively constant across this complex. However, when looking at

the standard deviation (Fig. 10), it can be seen that the ΔL/L of ren_A188V fluctuated heavily across most of renin and parts of angiotensinogen. Looking at the residue contact map of ren_A188V (Fig. 11), the VAL188.A interacts with PHE41.B more often than ALA188.A in the wild type (71.4% vs. 8.5% respectively). As PHE41.B is in a loop region, which is likely flexible, interacting with this residue may cause the renin structure to shift more. That the contact happens 71.4% of the time in the mutant also means that the bond is being continuously broken and reformed, which could also explain the fluctuation.

DISCUSSION

The RAS plays an important role in regulating arterial blood pressure and plasma sodium levels. Hyperactivity of this system has been linked to numerous conditions including hypertension, kidney disease, and congestive heart failure. The literature and human genome databases report a large number of variations linked to both proteins; renin and angiotensinogen. For drug development purposes, it is important to understand how these variations are affecting the system. Thus, in this study, we aimed to identify SNVs that have functional and potentially damaging effects on the renin-angiotensinogen complex. For this purpose we utilized a protocol recently proposed by our group [35], and used structural bioinformatics and network analysis techniques to analyze the effects of SNVs at the interface between renin and angiotensinogen.

A comprehensive set of all known SNVs (317 in renin and 212 in angiotensinogen proteins) was extracted from the HUMA database. This dataset was filtered by removing synonymous and missense variants and then using the VAPOR pipeline to predict which variants were likely to be deleterious. Further filtering was done to identify SNVs in the interface, which left us with a final data set of 9 angiotensinogen, 6 renin SNVs, and their combinations.

A total of 25 successful MD simulations were run to perform this analysis. All the MD trajectories were analyzed by RMSD and RMSF calculations and combined with network analysis. This research identified important SNVs and let us understand why certain variations have deleterious effects on the proteins by linking the structural variation(s) to the function. For instance, of the 25 MD runs, only the complex containing the SNV, P40L, in angiotensinogen did not stabilize after 100 ns. RIN analysis further showed that this mutation plays an important role in the interaction between renin and angiotensinogen, and this mutation caused a loss of interaction between position 40 in angiotensinogen and position 84 and 367 in renin. This resulted in a significant decrease in the accessibility of a large section of the renin structure.

Additionally, it was found that the complex containing the SNV, A188V, in renin was subject to significant fluctuation in ΔL/L. This was potentially due to increased interaction between position 188 in renin and position 41 in angiotensinogen.

Another important finding was the identification of the increased rigidity in some if the mutant complexes relative to the wild type. It was especially noticeable in the complex containing the D104N variation in renin. According to HUMA, this SNV is associated with renal tubular dysgenesis.

CONCLUSIONS

On the basis of these results, we found that using network analysis techniques in combination with MD provided useful insights into the function of the protein complex. It was especially useful for finding important interacting residues, such as PRO40 in angiotensinogen and HIS367 and THR84 in renin. However, areas of high and low ΔBC and ΔL/L do not correlate directly with RMSF. This may be due to the fact that RMSF measures the movement of a residue, whereas network analysis techniques such as BC and L also take the movement of surrounding residues into account. In other words, a change in the BC and L of a residue does not necessarily mean that that residue is, itself, moving. It might be that the residues around it are moving, affecting shortest paths that usually traverse it. As such, network analysis should be used in conjunction with RMSF to fully understand the movement of a protein or protein complex.

ACKNOWLEDGMENTS

The authors thank Canan Atilgan and Tandac Furkan Guclu for valuable discussions on network analysis.

REFERENCES

1. Kobori H, Nangaku M, Navar LG, Nishiyama A. The intrarenal renin-angiotensin system: from physiology to the pathobiology of hypertension and kidney disease. Pharmacol Rev 2007;59:251–87.
2. Paul M, Poyan MA, Kreutz R. Physiology of local renin-angiotensin systems. Physiol Rev 2006;86:747–803.
3. Sweitzer NK. What is an angiotensin converting enzyme inhibitor? Circulation 2003;108:16–8.
4. Barreras A, Gurk-Turner C. Angiotensin II receptor blockers. Proc (Bayl Univ Med Cent) 2003;16:123–6.
5. Israili ZH. Renin inhibitors as antihypertensive agents. Rev Latinoam Hipertens 2008;3:98–112.
6. Berman HM, Westbrook J, Feng Z, et al. The Protein Data Bank. Nucleic Acids Res 2000;28:235–42.
7. HUMA. Available at: https://huma.rubi.ru.ac.za/. Accessed October 1, 2016.
8. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
9. Pei J, Kim BH, Grishin NV. PROMALS3D: A tool for multiple protein sequence and structure alignments. Nucleic Acids Res 2008;36:2295–300.
10. Shen M-Y, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci 2006;15:2507–24.
11. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–91.
12. Eisenberg D, Lüthy R, Bowie JU. VERIFY3D: Assessment of protein models with three-dimensional profiles. Methods Enzymol 1997;277:396–406.
13. Wiederstein M, Sippl MJ. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 2007;35:W407–10.

14. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29:308–11.
15. VAPOR. Available at: https://huma.rubi.ru.ac.za/#vapor. Accessed October 1, 2016.
16. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248–9.
17. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS One 2012;7:e46688.
18. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics 2006;22:2729–34.
19. Tang H, Thomas PD. PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. Bioinformatics 2016;32:2230–2.
20. Shihab HA, Gough J, Cooper DN, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum Mutat 2013;34:57–65.
21. Tina KG, Bhadra R, Srinivasan N. PIC: protein interactions calculator. Nucleic Acids Res 2007;35:W473–6.
22. Abraham MJ, Murtola T, Schulz R, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 2015;1:19–25.
23. Darden T, York D, Pedersen L. Particle mesh Ewald: An N•log(N) method for Ewald sums in large systems. J Chem Phys 1993;98:10089.
24. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–90.
25. Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys 1981;52:7182–90.
26. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. J Comput Chem 1997;18:1463–72.
27. Ozbaykal G, Rana Atilgan A, Atilgan C. In silico mutational studies of Hsp70 disclose sites with distinct functional attributes. Proteins 2015;83:2077–90.
28. Csárdi G, Nepusz T. The igraph software package for complex network research. InterJournal Complex Syst:1695. Available at: http://www.necsi.edu/events/iccs6/papers/c1602a3c126ba822d0bc4293371c.pdf. Accessed October 1, 2016.
29. Doshi U, Holliday MJ, Eisenmesser EZ, Hamelberg D. Dynamical network of residue—residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. Proc Natl Acad Sci U S A 2016;113:4735–40.
30. Dijkstra EW. A note on two problems in connexion with graphs. Numer Math 1959;1:269–71.
31. Overview - NetworkX. Available at: https://networkx.github.io. Accessed October 1, 2016.
32. NumPy. n.d. Available at: http://www.numpy.org/. Accessed October 1, 2016.
33. Matplotlib: python plotting. Available at: http://matplotlib.org/. Accessed October 1, 2016.
34. Brandes U. A faster algorithm for betweenness centrality. J Math Sociol 2001;25:163–77.
35. Brown DK, Tastan Bishop Ö. Role of structural bioinformatics in drug discovery by computational SNP analysis: a proposed protocol for analyzing variation at the protein level. Glob Heart 2017 Mar 13 [E-pub ahead of print].