Novel dynamic residue network analysis approaches to study homodimeric allosteric modulation in SARS-CoV-2 M^{pro} and in its evolutionary mutations

*Olivier Sheik Amamuddy[#], Rita Afriyie Boateng[#], Victor Barozi, Dorothy Wavinya Nyamai and Özlem Tastan Bishop**

Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University, Makhanda, South Africa

# Equally contributed first authorship

* Corresponding Author: Özlem Tastan (o.tastanbishop@ru.ac.za)

ABSTRACT

The rational search for allosteric modulators and the allosteric mechanisms of these modulators in the presence of evolutionary mutations, including resistant ones, is a relatively unexplored field. Here, we established novel *in silico* approaches and applied to SARS-CoV-2 main protease (M^{pro}) as a case study. First, we identified six potential allosteric modulators (SANC00302, SANC00303, SANC00467, SANC00468, SANC00469, SANC00630) from the South African Natural Compounds Database (SANCDB) bound to the allosteric pocket of M^{pro} that we determined in our previous work. We also checked the stability of these compounds against M^{pro} of laboratory strain HCoV-OC43 and identified differences due to residue changes between the two proteins. Next, we focused on understanding the allosteric effects of these modulators on each protomer of the reference M^{pro} protein, while incorporating the symmetry problem in the functional homodimer. In general, asymmetric behavior of multimeric proteins is not commonly considered in computational analysis. We introduced a novel combinatorial approach and dynamic residue network (DRN) analysis algorithms to examine patterns of change and conservation of critical nodes, according to five independent criteria of network centrality (*betweenness centrality (BC), closeness centrality (CC), degree centrality (DC), eigencentrality (EC)* and *katz centrality (KC)*). The relationships and effectiveness of each metric in characterizing allosteric behavior were also investigated. We observed highly conserved network hubs for each averaged DRN metric on the basis of their existence in both protomers in the absence and presence of all ligands, and we called them *persistent hubs* (residues 17, 111, 112 and 128 for averaged *BC;* 6, 7, 113, 114, 115, 124, 125, 126, 127 and 128 for averaged *CC*; 36, 91, 146, 150 and 206 for averaged *DC*; 7, 115 and 125 for *EC*; 36, 125 and 146 for *KC*). We also detected ligand specific signal changes some of which were in or around functional residues (i.e.

chameleon switch PHE140). Using *EC persistent hubs* and ligand introduced hubs we identified a residue communication path between allosteric binding site and catalytic site. Finally, we examined the effects of the mutations on the behavior of the protein in the presence of selected potential allosteric modulators and investigated the ligand stability. The hit compounds showed various levels of stability in the presence of SARS-CoV-2 M^{pro} mutations, being most stable in A173V, N274D and R279C, and least stable in R60C, N151D V157I, C160S and A255V. SANC00468 was the most stable compound in the 43 mutant protein systems. We further used DRN metric analysis to define cold spots as being those regions that are least impacted, or not impacted, by mutations. One crucial outcome of this study was to show that *EC* centrality hubs form an allosteric communication path between the allosteric ligand binding site to the active site going through the interface residues of Domain I and II; and this path was either weakened or lost in the presence of some of the mutations. Overall, the results of this study revealed crucial aspects that need to be considered in drug discovery in COVID-19 specifically and in general for rational computational drug design purposes.

KEYWORDS

Evolutionary mutations; allosteric modulators; homodimeric protein; natural compounds; dynamic residue network analysis; MD-TASK; MDM-TASK-web; *betweenness centrality, closeness centrality, degree centrality, eigencentrality* and *katz centrality*

1. INTRODUCTION

With the advent of COVID-19, researchers across the world reacted quickly with the proposal of multiple potential inhibitors designed to abrogate viral protein activity using rational drug design approaches and wet lab experiments. This concept primarily involves targeting critical viral life-cycle proteins [1–4]. The SARS-CoV-2 main protease (M^{pro}) protein plays a crucial role in the viral maturation cycle by lysing itself (autocatalysis) and other viral polyproteins [5]. This presents SARS-CoV-2 M^{pro} as a key drug target for designing wide-spectrum [6,7] anti-COVID-19 inhibitors or allosteric modulators that terminate the viral replication cycle [8]. Among the multitude of studied COVID-19 related proteins, the active site of SARS-CoV-2 M^{pro} has been extensively targeted by virtual screening of both natural and non-natural compounds [9–11]. In contrast, the rational search for allosteric modulators of the protein is still relatively unexplored. Additionally, allosteric mechanisms in the presence of mutations have not yet been a subject of focus in drug screening. In our previous study, a potential dual allosteric pocket of SARS-CoV-2 M^{pro} was identified through multiple *in silico* tools in the presence of 50 early pandemic mutations [12]. These two pockets are mirrored across the dimer interface and are individually composed of residues from each protomer. As a continuation of our previous SARS-CoV-2 M^{pro} work [12], here we aim to set up alternative innovative therapeutic concepts on identification of allosteric modulators in the presence of early evolutionary mutations of the virus. These concepts are explained under three subsequent sections:

PART I: Here, we identified potential allosteric modulators for the dimeric SARS-CoV-2 M^{pro} protein, at a protonation state corresponding to pH 7.0, by screening it against 625 South African natural compounds [13,14]. Parallel to this, we also docked the natural compounds against the M^{pro} protein of one of the seven human coronaviruses, HCoV-OC43. Previously, HCoV-OC43 was

suggested as a model to study SARS-HCoV without the need for Biosafety Level 3 facilities [15]. This strain is, indeed, commonly used as the laboratory strain. Thus, using *in silico* techniques we wanted to see if similar results would be obtained from the M^{pro} protein in each strain. This analysis sheds light on potential considerations to factor in when transferring findings from HCoV-OC43 to SARS-CoV-2, using the M^{pro} allosteric site as example.

PART II: Next, our focus was on understanding the allosteric effects of the selected hit compounds (PART I) on each protomer of the reference M^{pro} protein (wild type, WT). From our previous study, we came across the problem of protein symmetry, where we observed that protomer dynamics could be switched between identical copies of a protomer in a homodimer. Symmetry correction was performed then by aligning single equilibrium conformations. In the current study, we investigated the phenomenon in greater detail using a combinatorial approach to examine patterns of change and conservation of critical nodes, according to five independent criteria of network centrality (*betweenness centrality (BC), closeness centrality (CC), degree centrality (DC), eigencentrality (EC)* and *katz centrality (KC)*), used as averages. While doing so, we also investigated the relationships and effectiveness of each metric in characterizing allosteric behavior. We hypothesized that allosteric change might be expressed through complex routes involving intraprotomeric and interprotomeric combinations of critical residues. By monitoring the centrality patterns of these residues across the homodimer under the influence of intrinsic (e.g. protein mutations and ligand binding) and extrinsic (simulation parameters) factors during molecular dynamics (MD) simulations, we aimed to extract further details from the homodimer state of the protease. To our knowledge this phenomenon is not typically addressed in the case of homodimeric protein complexes. While we believe that the same phenomenon

exists at the homomultimeric level, a less complex case involving allosterically bound dimeric M^{pro} is investigated herein.

Further, we, for the first time, introduced the concept of analyzing globally central nodes (i.e. the 5% most central nodes measured across all samples) for each of the five metrics of dynamic residue networks (DRNs). The metrics comprised averaged versions of *BC, CC, DC, EC* and *KC.* Even though some of these metrics were, previously, used for static protein structure analysis [16–18] to our knowledge, this is the first study that gathers together information from these metrics over molecular dynamics (MD) simulations. Additionally, the hub data was itself reformulated as a set of network graphs, which were queried in order to decipher the complex patterns of hub conservation and transition (according to each DRN metric) from the apo state to one that is allosterically occupied.

PART III: In the third part of the study, we examined the effects of the mutations on allosteric behavior of the protein in the presence of selected potential allosteric modulators and investigated ligand stability. Structure-based drug discovery approaches have been successfully used for the design of many orthosteric drugs [19] and to some extent allosteric modulators [20] for the treatment of communicable and non-communicable diseases. A good example is that of HIV protease inhibitor [21]. However, consideration of the impact of evolutionary mutations of pathogens, including those linked to drug resistance, is mostly undetermined in rational drug design. Depending on their position and physicochemical properties, mutations can modulate protein behavior by altering their stability and/or affinity to other interacting biological molecules [22–25]. A more complex, yet subtle phenomenon may be observed at the level of entropic effects of mutations, whereby differences may be seen at the level of the rate of visiting certain states, and not by the mere presence or absence of a defined state (or set thereof) [26–28]. A

classic case is the distance effect of pathogenic mutations that keep the function of the protein while gaining resistance [23,29]; hence the purpose of this part of the study is to understand the effect of evolutionary mutations in the attempt of COVID-19 rational drug design. We believe the information gleaned here may help to develop drugs that could potentially minimize the risks of having premature drug inactivation; and may reduce potential drug resistance effects to provide a longer-lasting treatment option.

For that purpose, mutant protein-allosteric modulator complexes were subjected to 20 ns all-atom MD simulations at a fixed pH, and the results were, then, evaluated in the same manner as introduced in the second part of the article. The potential effectiveness of the allosteric modulators was identified in the presence of some of the early pandemic mutations of the protein. Even though no solid evidence of the effect of these mutations has been reported, involving them in drug development might help further our understanding of the enzyme's mechanics and pre-empt the most worrying feature of mutations: drug resistance.

Overall, the results of this study revealed crucial aspects that need to be considered in structure-based drug discovery such as the way in which the allosteric modulators should be identified; and how the stability of these modulators should be considered in the presence of mutations. We further argue that consideration of potential asymmetric behavior of homodimer proteins; of the novel DRN approaches and data analysis that are presented here would be applicable and useful in any computational drug discovery research.

2. MATERIALS AND METHODS

2.1. Preparation of the reference and mutant SARS-CoV-2 M^{pro} and HCoV-OC43 M^{pro} structures

The three-dimensional (3D) structure of the SARS-CoV-2 M^{pro} was retrieved from Protein Data Bank (PDB)[30] (PDB ID: 5RFV [31]), and its dimeric unit assembled as described in our previous study [12]. In this study, we also utilized a set of 50 SARS-CoV-2 M^{pro} mutant proteins that were prepared in our previous study [12]. The list of mutations that were acquired from the Global Initiative on Sharing All Influenza Data (GISAID) [32] as described in our previous work is presented in **Table S1**[53].

5RFV was further used as a template to model the 3D structure of the human coronavirus strain (HCoV-OC43) M^{pro} via MODELLER, using the automodel function parameterized with a slow refinement with loop deviation of 2.5 Å [33]. This protein is a close homolog of the SARS-CoV-2 M^{pro}, and the strain is generally used in inhibition assays in the laboratory. Prior to homology modelling, the HcoV-OC43 protein sequence was retrieved from the replicase polyprotein 1a record available from UniProt (Entry ID: P0C6U7; position 3247-3549), and was aligned against the sequence and structure of 5RFV using PROMALS3D [34]. The model with the lowest z-DOPE score was selected from a parallel run of 50 models. The PROPKA tool under the PDB2PQR algorithm [35] was, then, utilized to assign protonation states of all the proteins at a pH of 7. The calculations were done with the AMBER force field [36].

Based on the assembled and protonated SARS-CoV-2 M^{pro} dimeric structure, all 50 mutations were inserted using BIOVA Discovery Studio Visualizer [37]. This approach was utilized to minimize structural variations across the proteins. All mutated structures were subsequently protonated using the same procedure as for the reference structure.

2.2. High-throughput virtual screening of SANCDB compounds against M^{pro} proteins

A total of 623 compounds were first obtained from the South African natural compound database (SANCDB) [13,14]. Partial charges were assigned to compounds and the protonated proteins using

the Gasteiger-Hückel protocol in AutoDockTools (ADT) [38]. The AutoDock/Vina plugin from PyMOL was used to place the docking grid around the dimeric SARS-CoV-2 M^{pro} reference protein. A docking box size of 65 x 71 x 80 Å with a grid spacing of 1 Å was centered at coordinates (0.00, 0.65 and 0.00). The exhaustiveness of 1000 was used, and the maximum number of docking poses was increased to 20. Blind docking (BD) simulations were performed in parallel, with 12 cores per job at the Center for High-Performance Computing (CHPC) using the QuickVina-W program [39]. After having docked the SANCDB compounds, the ligand *PDBQT* files were split into their separate poses before being converted to PDB format. Preliminary filtering was then applied using an in-house C++ script to every file to retain ligand poses that had a centroid distance of less than 10 Å to any of the allosteric pockets irrespective of binding energy. The pre-filtered poses were then manually curated in PyMOL (version 2.4) [40] to remove those that did not localize to the allosteric pocket. For each of the filtered ligands, the number of poses was tallied and ranked in ascending order of binding energy [41,42]. The top six compounds from the SARS-CoV-2 M^{pro} were then short-listed based on residue interactions for their respective lowest energy poses. HCoV-OC43 M^{pro} underwent the same steps, to be used as a comparator.

2.3. Molecular dynamics simulations protocol of M^{pro} and mutant systems

100 ns all-atom molecular dynamics (MD) simulations were conducted using GROMACS (version 2019) [43] for SARS-CoV-2 M^{pro} reference protein and the HCoV-OC43 strain homolog protein both in the absence and presence of six hit compounds bound at the previously identified allosteric site. In order to investigate the effect of mutations, 50 ligand-bound SARS-CoV-2 M^{pro} mutants were similarly taken into 20 ns MD for each of the six compounds. GROMACS-compatible structure and ligand topology input files were derived using the AMBER03 force

field [36] and the ACPYPE tool [44] respectively. A total of 314 systems [(reference protein x 6) + (homolog protein x 6) + Apo-reference protein + Apo-homolog protein + (50 mutant x 6 compounds)] were solvated using the TIP3P water model [45] in a cubic box, with a minimum distance of 1 nm between the box edge and the protein. All systems were subsequently neutralized with 0.15 M NaCl. Solvated systems were first minimized for 5000 steps using the steepest descent algorithm until the relaxed systems converged to a maximum force of 1000 kJ/mol/nm. Following minimization, systems were equilibrated at constant number, volume and temperature (NVT) at 300 K temperature and constant volume using the modified Berendsen thermostat algorithm [46] followed by NPT (constant number of particles, pressure and temperature) at 1 bar pressure and constant volume and temperature ensemble using the Parrinello–Rahman barostat algorithm [47] In all ensembles, systems coupling groups and time constraints were set at 0.1 ps. All bonds were constrained under the LINCS holonomic constraints algorithm [58] whereas the Particle-mesh Ewald (PME) algorithm [48] was set to include the contribution of long-range electrostatic interactions. The overall MD protocol was carried out on the Center for High-Performance Computing (CHPC), Cape Town, South Africa using 384 cores with total CPU hours of ~2921472. Structure coordinates were written after every 10 ps and periodic boundary conditions (PBC) were removed.

2.4 Calculation of dynamic residue network metrics

To study the effect of ligand binding on the active site, as well as on inter-and intra-domain residue dynamics over the course of MD simulations, dynamic residue network analysis (DRN) was done using MDM-TASK-web scripts [49]. DRN [50] was applied on the last 10 ns trajectories of the apo and ligand-bound M^{pro} systems. Residue network analysis uses graph theory concepts and characterizes residues in a protein structure in which each amino acid is represented as a

node and inter-connected residues (C_β - C_β and in Gly C_α - C_α atoms) are depicted as edges based on a specified cut-off distance (6.7 Å) [50]. DRNs were analysed based on five metrics; averaged *betweenness centrality* (*BC*), averaged *closeness centrality* (*CC*), averaged *degree centrality (DC)*, averaged *eigencentrality* (*EC*) and averaged *katz centrality* (*KC*) *via the cal_network.py script* incorporated in the web server, MDM-TASK-web [49]. Each of the metrics is a time-averaged summary of the network metrics obtained during MD simulations.

The averaged *BC* metric is defined as how often a residue is traversed along the shortest paths connecting every other residue pairs [51]. This metric was calculated based on the equation:

$$\underline{BC}(v) = \frac{1}{m}\sum_{i=1}^{m}\sum_{s,t\in V}\frac{\delta(v_i\, t_i|v_i)}{\delta(s_i,t_i)} \quad \textbf{(1)}$$

where $\delta(s,t|v)$ symbolises the number of shortest paths bridged between a residue n and other nodes s and t. $\delta(s,t)$ denotes the averaged shortest paths existing between residues *s* and *t* where *s* and *t* are part of the set V, which comprises the set of all nodes, while *m* indicates the overall number of frames.

Averaged *closeness centrality* (*CC*) of a residue is calculated as the reciprocal of the average number of the shortest paths linking a residue n and all other residues in the network.

$$CC\ (x) = \frac{1}{\sum_y d(v,u)} \quad \textbf{(2)}$$

where $d\ (v, u)$ is the total distance between residue *v* and all other residues *u*.

Additionally, metric *degree centrality (DC)* defines the number of neighboring nodes (the local connectivity) around a given node. It is normalized by both the number of nodes in the network and the number of MD frames. The equation for computing the averaged *DC* is as follows:

$$\underline{DC}(k) = \frac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}A_{ijk} \quad \textbf{(3)}$$

where n indicates the number of residues, m denotes the number of frames; A_{ijk} indicates adjacency in time frame i, being 1 if residues with indices j and k are adjacent and 0 otherwise.

Eigencentrality (EC) measures the high centrality given to high degree residue, or to a residue that is connected to other high degree residues. The procedure for calculating *EC* is summarized here, and further details are in the literature [52]. The formula for the computation of *EC* for a single residue i for the k^{th} frame is as follows:

$$EC_{ik} = \epsilon_k^{-1} \sum_{t=1}^{n} A_{ijk} \cdot EC_{jk} \quad \textbf{(4)}$$

The weighted multiplication operation between the adjacency matrix A is repeated against the vector *EC* until convergence. A_{ij} is an adjacency, k is a frame, EC_{ik} is the j^{th} component of the *EC* vector for the k^{th} frame, and n is the number of nodes. The averaged *EC* for the i^{th} node is then computed from the matrix of *EC* likewise using MDM-TASK-web as follows:

$$\underline{EC_i} = \frac{1}{m} \sum_{k=1}^{m} EC_{ik} \quad \textbf{(5)}$$

Lastly, the *Katz* centrality (*KC*) measures the relative degree of influence of a residue i within connected residues in a network. The procedure for calculating *KC* is summarized here, and further details are in the literature [52]. The *KC* of node i is

$$KC = \alpha \sum_j A_{ij} x_j + \beta \quad \textbf{(6)}$$

where A represents the adjacency matrix and *KC* is the eigenvector computed by NetworkX in MDM-TASK-web. α and β denoted the attenuation factor and weight assigned to the immediate neighbors of node i. The same metric is computed for each frame before averaging the value across frames for each residue.

2.5. Identification of top 5% global high network centrality residues

DRN metrics were computed for the reference and the mutant SARS-CoV-2 M^{pro} samples using MDM-TASK-web for both the apo and the six ligand-bound complexes. In order to estimate

residue hubs, all related samples that were to be compared were combined in order to have a common scale. Therefore, for each individual DRN metric, the data points of samples (apo, mutant and ligand-bound) belonging to that metric were concatenated into a single vector, which was sorted in descending order to focus on nodes of highest overall centrality. The top 5% of these values were extracted [304 residues x 2 chains x (1 apo reference + 6 bound reference) systems x 0.05 = 212 elements]. The value at this index was used as a threshold for the selection of entries from the original data set. Then, each of the original matrices was searched for any component that is greater than or equal to that minimum number. To accomplish that, a binary matrix was built that contained the number "one" for any cell that satisfied the condition, and from which the row sums were then computed, in order to select any row with a non-zero row sum. This generated a set of row indices that were used to subset the original matrix of centrality values. In this manner, the globally high network centrality values were obtained in the presence of their counterpart values in other samples, thus showing how the hubs perform sample-wide. This approach was performed separately for each of the 5 metrics.

2.6. Application of binary logic to investigate protomer hub combinations from DRN analysis

For each DRN metric, a global network was built using as nodes the detected globally central hubs for all of the reference protein states (ApoA/B, SANC00302A/B, SANC00303A/B, SANC00467A/B, SANC00468A/B, SANC00469A/B), which have as labels components of the protein state and protomer to which each hub residue belongs. These labels were inserted as nodes, and undirected edges were created from them by linking their respective hub nodes to them. As this global network was too dense to analyze, a sub-network was extracted for each individual complex and was merged to the apo protomers. In this way, one could identify

whether a hub was shared, gained or lost from the apo state upon ligand binding. This representation was applied and analyzed in a systematic manner (according to Table 1) to investigate whether ligand binding had any effect, as we posited that the effects of a ligand's presence in the allosteric site may manifest itself not only in the bound protomer, but also in the unbound one. In this way it was possible to track patterns of hub conservation and divergence.

Table 1. Hub combination possibilities for any given residue between two dimers. A tick symbol (✓) denotes the presence of a hub from a given protomer, while a cross (x) denotes absence of that same hub from a chain. Apo - A: Apo protein, protomer A; Apo - B: Apo protein, protomer B; Complex A: Protomer A of protein-ligand complex; Complex B: Protomer B of protein-ligand complex

	Apo - A	Apo - B	Complex A	Complex B	Score	Interpretation
1	✓	x	x	x	1	Potential ligand effect inferred by asymmetry
2	x	✓	x	x	1	Potential ligand effect inferred by asymmetry
3	x	x	✓	x	1	Potential ligand effect inferred by asymmetry
4	x	x	x	✓	1	Potential ligand effect inferred by asymmetry
5	✓	✓	x	x	2	Complete hub loss: ligand effect
6	x	✓	✓	x	2	Inconclusive effect
7	x	x	✓	✓	2	Hub gain on ligand presence
8	✓	x	✓	x	2	Inconclusive effect
9	✓	x	x	✓	2	Inconclusive effect
10	x	✓	x	✓	2	Inconclusive effect
11	✓	✓	✓	x	3	Potential ligand effect inferred by asymmetry
12	x	✓	✓	✓	3	Potential ligand effect inferred by asymmetry
13	✓	x	✓	✓	3	Potential ligand effect inferred by asymmetry
14	✓	✓	x	✓	3	Potential ligand effect inferred by asymmetry
15	✓	✓	✓	✓	4	No ligand effect from symmetry
16	x	x	x	x	0	Not applicable

3. RESULTS AND DISCUSSION

3.1. Revisiting the structure of M^{pro} and mutants

The SARS-CoV-2 M^{pro} protein comprises 306 residues [2] and is active in its dimeric state at a pH of 7.0 [6,53]. The dimeric functional state regulates catalytic turnover using the subunit flip-flop mechanism where the two monomers are used alternately for acylation and deacylation steps [54,55]. Each monomer (designated protomer A and B) harbors three distinct domains (I-III) [2,10] and contains a His-Cys catalytic dyad signature (His41 and Cys145) located within a well-defined hydrophobic substrate-binding site between domain I and II (**Figure 1**). The catalytic dyad residues are key for hydrolysis in which His41 functions as a general base [6,56]. SARS-CoV-2 M^{pro} domains I (residue 10-99) and II (100-183) consist of antiparallel β-barrel structure[2] that form the catalytic domains of the protein as the active site was located between domain I and II. Domain III (198-303) is predominantly antiparallel α-helices [53,57] and connected to the catalytic domains by a long loop region (184-197). This domain is involved in the regulation of enzymatic activity of the virus [58]. The interaction interface, which is crucial for dimerization and enzymatic activity, is formed between domain II of protomer A and N-finger region (1-9) of protomer B and vice versa [56,59]. These two N-finger signatures interact with Glu166 to maintain the correct orientation of the substrate-binding site subsite S1. The N-finger feature is similar to that of previously reported M^{pro} from other coronaviruses [8,53,60,61]. Each protomer has subsites (S1 – S5) located in the active site cavity and the active site cavity comprises the following residues: 25^{62}, 27, $41^{2,6,63}$, 44^{62}, 45, 46^{62}, $49^{2,6,62,63}$, $54^{2,62}$, $140^{2,6,62}$, $141^{2,62}$, $144^{6,63}$, $145^{2,6,63}$, $163^{2,6,62,63}$, $164^{2,62}$, $165^{2,62,63}$, $166^{2,6,62,63}$, $167^{2,62}$, $168^{2,6,62,63}$, $172^{2,62,63}$, $187^{2,62}$, 188^{62}, $189^{2,6,62,63}$, $190^{2,62}$, $191^{2,62}$, $192^{2,62}$.

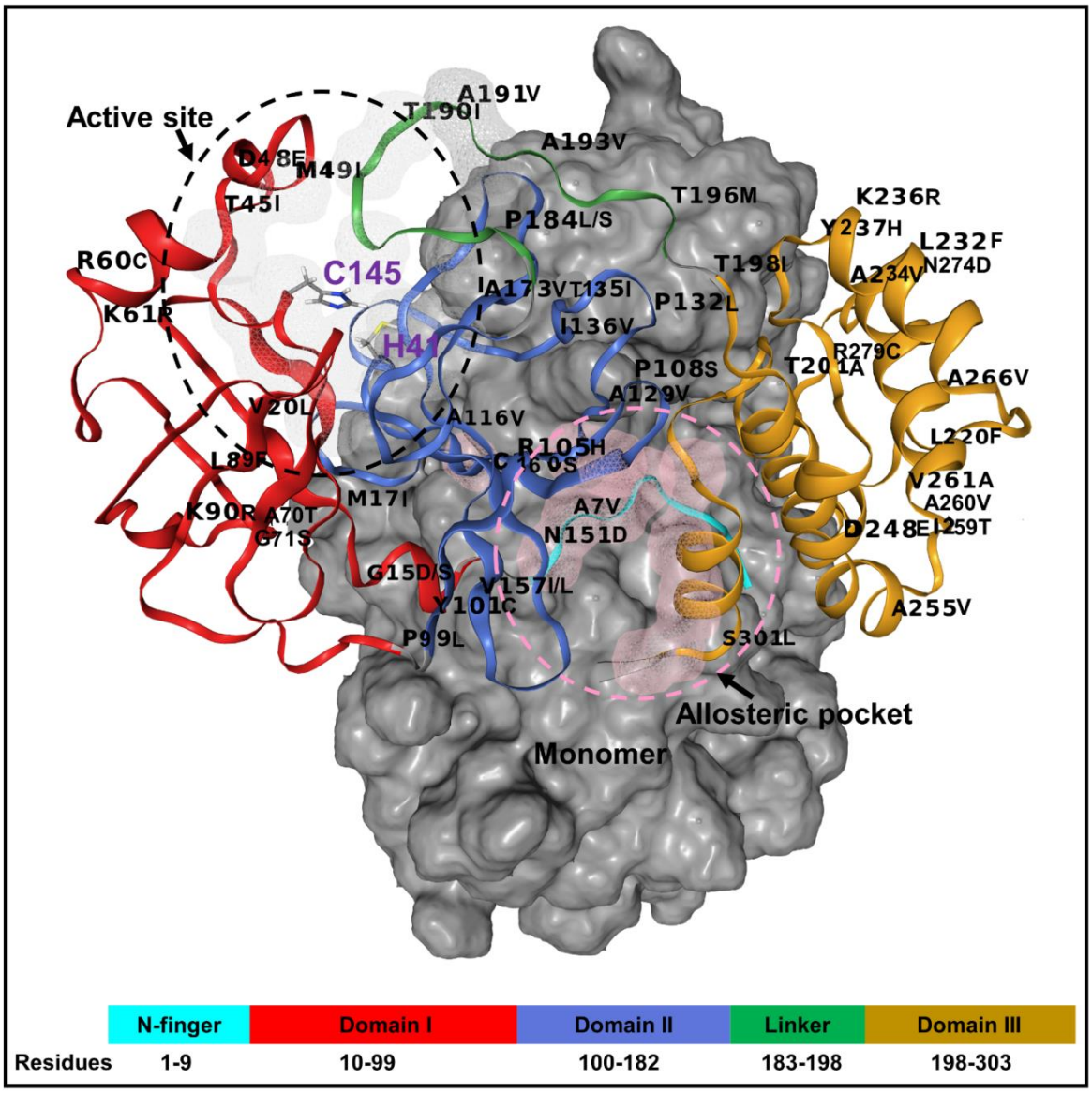

Figure 1. A structural representation of the homo-dimeric nature of SARS-CoV-2 M^{pro}. The structural domains (I-III) are shown in red, royal blue and orange cartoons, respectively. The N-finger region (residue 1-9) and the long loop connecting domain II to III (linker) are indicated in cyan and green colors, respectively. The substrate-binding pocket and allosteric pocket on a monomer are illustrated in grey and pink wireframe and dotted lines respectively. The distribution of SARS-CoV-2 M^{pro} mutations identified from the GISAID database [32] is labeled on the structure.

In our previous study, we identified dual allosteric pockets located at the interface of protomer A and B (**Figure 1**), that concur with key residues for functional dimerization and enzymatic activities [53]. The residues of this allosteric pocket of SARS-CoV-2 M^{pro} are 116, 118, 123, 124, 139, and 141 on protomer A and residues 5–8, 111, 127, 291, 295, 298, 299, 302, and 303 on protomer B; or vice versa. We also demonstrated that there is a correlation between substrate binding site and predicted allosteric sites and this correlation changes in the presence of some of the studied 50 mutations which spanned several secondary structures in M^{pro} domains as well as the N-finger and linker region (**Figure 1**).

PART I:

3.2. Identification of allosteric modulators against dimeric SARS-CoV-2 M^{pro} protein

We identified six compounds in SARS-CoV-2 and 15 compounds in HCoV-OC43 by blind docking and preliminary filtering of the 625 SANCDB compounds against the dimeric M^{pro} proteins (**Figure 2A**). The high degree of search exhaustiveness increased the likelihood of finding certain binding poses more than once, despite having less favorable binding energy scores. This approach draws from the idea of the use of pose clustering in AutoDock [64], as we have noticed that non-equilibrium binding energy scores tend to be affected by the length of the ligand. The poses corresponding to either copy of the allosteric site were tallied for each compound to be compared across all hit compounds in both coronavirus strains. As seen in **Figure 2A**, the lowest energy hits for the mirrored allosteric site occur in HCoV-OC43 but are not the most abundant hit compounds. Of notable interest are compounds SANC00209, SANC00210 and SANC00211, which are halogenated monoaromatic terpenoids produced from the marine alga *Plocamium corallorhiza*, with anti-proliferative properties. The four most

abundant hits for the SARS-CoV-2 allosteric site (SANC00467, SANC00468, SANC00469 and SANC00630) occurred in both coronavirus strains, despite showing less favorable energy scores. While SANC00467, SANC00468 and SANC00469 all come from *Drimia robusta* [65,66], SANC00630 is from *Senecio oxyodontus* [67]. All of them are monophenolic compounds. The fact that this allosteric site tends to bind to different aromatic compounds may suggest that this may be a scaffold for small molecules to target. Compounds SANC00302 and SANC00303 did not fare as well as the other compounds in terms of energy scoring of number of poses; however, we decided to carry them forward for MD analysis to cross-check their stability. The latter two compounds are halogenated indoles from *Distaplia skoogi* that have shown moderate cytotoxicity against cancerous cells [68].

The ligand RMSD graph of the last 10 ns of 100 ns MD simulations (**Figure 2B**) identified that these six compounds have relatively different behavior in the M^{pro} protein of SARS-CoV-2 and of HCoV-OC43; while SANC00467, SANC00468 and SANC00469 demonstrated the most stable single conformation (unimodal distribution) in SARS-CoV-2 protein. Ligand RMSDs of the 100 ns simulations are presented in Figure S1. The different behavior of the compounds can be attributed to the varying compound-protein residue interaction profiles as observed in docking (**Figure 2C**, Table S2) as well as the residue differences of the two homolog proteins at the allosteric sites (**Figure 2D**). ALA7, PHE8, GLN127, PHE291 and ARG298 of SARS-CoV-2 M^{pro} are replaced with VAL7, ASN8, HIS127, LEU291 and GLN298 in the M^{pro} of HCoV-OC43 lab strain (**Figure 2D**). In SARS-CoV-2, residues ALA7 and PHE8 comprise the N-finger region crucial for dimer stabilization [69]. GLN127, PHE291 and ARG298 have been reported to play important roles in dimerization and enzymatic activity in SARS-CoV M^{pro} [70].

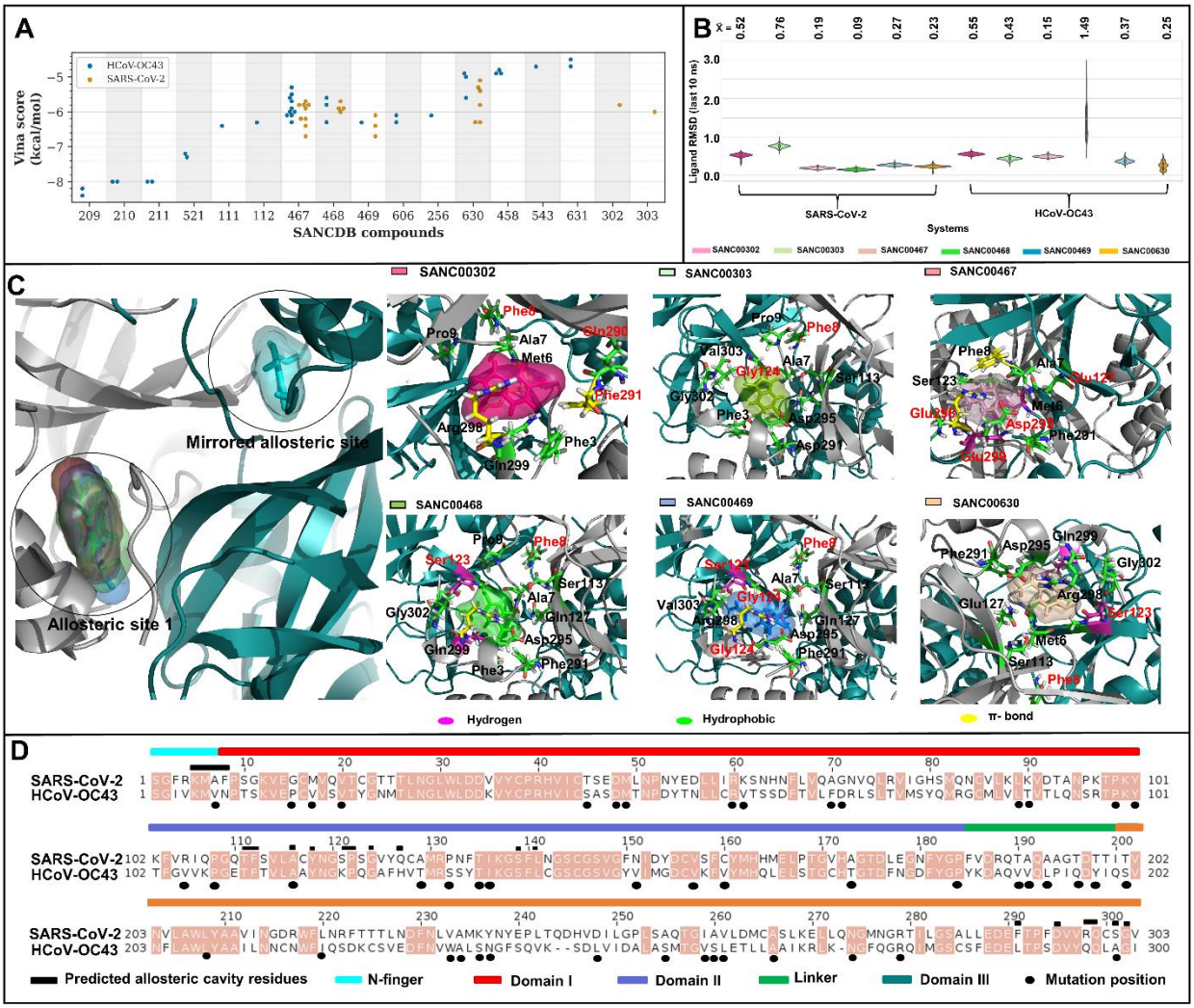

Figure 2. Overall binding profile of ligands in the allosteric site of SARS-CoV-2 and HCoV-OC43 strains. (A) Scatter plot of selected allosteric site ligands and their respective binding energies to SARS-CoV-2 (orange) and HCoV-OC43 (blue). (B) Kernel density plots of the ligand RMSD values over the last 10 ns of the 100 ns MD simulations. (C) Protein-ligand molecular interactions in SARS-CoV-2. Residue contribution from protomer A and B are labelled in black and red respectively. (D) Conservation of residues between the two strains.

In SARS-CoV-2 M^{pro}, several ligand interactions (such as hydrogen bond, hydrophobic and Pi interactions) with allosteric site residues were observed (Table S2). Compounds SANC00467,

SANC00468, SANC00469 and SANC00630 formed at least two hydrogen bonding interactions with some polar residue side chains (MET6, SER123, GLN299 and VAL303) that may play a role in ligand stabilization and retention within the pocket. The replacement of valine by a longer side chain in isoleucine at position 303 in HcoV-OC43, suggests that the site in HCoV-OC43 may not behave in the same way as that of SARS-CoV-2. On the other hand, at least seven hydrophobic interactions were observed across all modulators, indicating the enrichment of hydrophobic interactions in allosteric sites. The substitution of hydrophobic residue Phe8 and polar residues (SER121, SER301) in SARS-CoV-2 M^{pro} with polar ASN8, charged LYS121 and the hydrophobic residue LEU301 in HCoV-OC43 may explain the different ligand-binding patterns.

Our results indicated that the use of this strain for experimentation on allosteric modulation in SARS-CoV-2 M^{pro} may have some limitations.

PART II:

3.3. Identification of hub residues while considering symmetry in homodimers

Depending on the level of resolution desired for the analysis of homodimers, comparing MD-simulated pairs of a homodimeric protein can introduce conceptual challenges. For instance, one cannot easily know with certainty whether protomer A (or sections thereof) in one dimer behaves the same as its homologous position in protomer A in the second dimer. While a simpler protomer assignment approach based on permuted structural alignments was used in our earlier work [12] for single conformations, our attempt here investigates this issue in more depth, firstly by isolating potential hubs, and secondly by producing a representation of all the possible hub node combinations **(Table 1)** in order to obtain a scheme by which hub node importance can be

assessed. While a hub is generally accepted as a high connectivity (degree) node, it has also been used to mean high *BC* [71], but can also be understood as any node that may cause non-negligible topological alterations to a network when removed [72]. In this analysis the term is used in its more general sense to mean any node that forms part of the set of highest centrality nodes, here arbitrarily specified as the top 5% centrality nodes measured across all related samples, for any given averaged centrality metric. This procedure differs from the identification of 1 to 2 standard deviation from the mean or top 5% residues in individual samples that we generally used in our previous studies [73–75], in that it considers the strongest actors across samples and shows how other non-hub residues behave at the homologous position. We assume that investigating hub transitions in this manner is more likely to detect the most significant shifts in residue importance when exposed to a particular environment. We also used this approach to be able to handle a large amount of data analysis.

Figure 3 shows the heat maps of the five DRN metrics for the reference protease in the absence and presence of ligand binding; with the designated ligand-bound allosteric pocket of the dimer always being referenced as protomer A. Specifically, the ligand was assigned to protomer A based on its proximity to a terminal alpha helix in the same chain. In the case of SANC00467, where the allosteric compound had bound protomer B, the protomer label was swapped.

Preliminary examination of **Figure 3** showed that there are some residues that preserve their hub statuses. We, here, introduce the following terms: (1) *Constitutive hub*: If a hub is present in both protomers of the reference protein and remains as a hub irrespective of the apo or a ligand-bound state, it will be called a *constitutive hub* (see **Table 1**; *score 4*); (2) *Persistent hub*: If a hub remains across all systems compared, then the hub will be called *persistent*; in Part II, across all

systems would be apo protomers and all ligand-bound dimers of reference protein, and in Part III it would be both protomers of the reference and mutant proteins with a specific ligand. (3) *Super-persistent hub*: In Part III, we will use the concept of a *"Super-persistent hub"*, meaning that the hub is persistent across all the ligands considered in both reference and mutant proteins. Most of the *constitutive* and *persistent* hubs are metric-specific giving a different perspective to the network. As the five averaged centrality metrics refer to different measures of importance within a network, these terms will be used with respect to a given centrality metric and will not be shared between them.

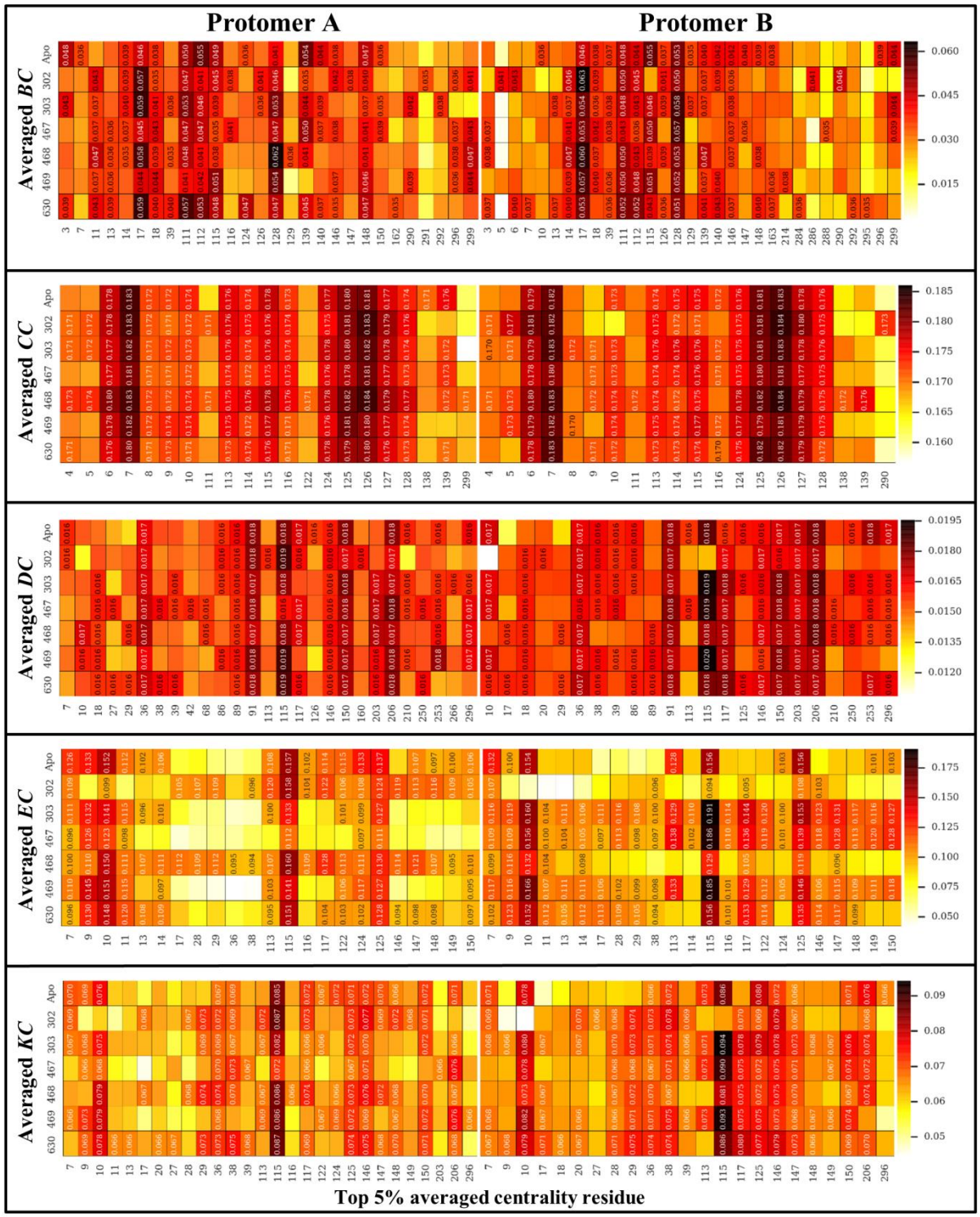

Figure 3. Heat maps for the potential hubs according to the global top 5% for each of the five DRN metrics, for the reference protein in the apo and the six allosterically bound states. Detected hubs are annotated with their centrality values, while their homologous residues in alternate

samples are not, but are only shown for the sake of comparison. For each metric, low to high centrality values are colored white, through yellow, orange and red to black. Measurements for the ligand-bound protomer (chain A) have been systematically presented on the left side, while those of the unbound protomer are on the right – this does not apply to the apo state.

3.4. Metric based investigation of *persistent hubs*

3.4.1. Betweenness Centrality

According to **Figure 3**, MET17, THR111, PHE112 and CYS128 hub residues were found to be unaltered from the reference protein apo state, or upon any selected ligand binding irrespective of protomer for the averaged *BC*. At individual ligand level, each of these hubs is *constitutive* and indicates that there is no ligand effect due to symmetry (Table 1, score 4).

For the entire system (apo + 6 ligand systems), these hubs are *persistent hubs* indicating that the allosteric modulators did not change the information path for these key residues; and any loss to these hubs may disrupt the communication.

Residues MET17 and CYS128 had been previously picked up from multiple simulations, but were not examined in depth in our previous work involving several M^{pro} mutants in the apo state [12]. The current analysis further showed that all conserved averaged *BC* hubs occurred as intrachain or interchain hinges within the dimer **(Figure 4A)**, both in the absence and presence of different allosterically bound compounds. Residue MET17 establishes intraprotomer contacts within the domain I/II interface by interacting with several residues of the beta hairpin. More specifically, it forms alkyl interactions with LEU115, PRO122 and CYS117. Of notable interest is the alpha helix that supports the N-finger. Being also part of the high *BC* hubs, it is possible that LEU115 and MET17 form an important bridge that relays interdomain information, potentially influencing N-finger stability, and by extension impacting the activity of the alternate

protomer. THR111 (from domain II) also plays a role in maintaining intraprotomeric interdomain stability by forming periodic H-bonds with ASP295 (from domain III), and at the same time mediating information flow. THR111 is also firmly bound to the other hub residue CYS128 via multiple hydrogen bonds and carbon H-bond. CYS128 is firmly seated on a beta-strand, forming non-bonded interactions with TYR126, VAL114 and PHE112.

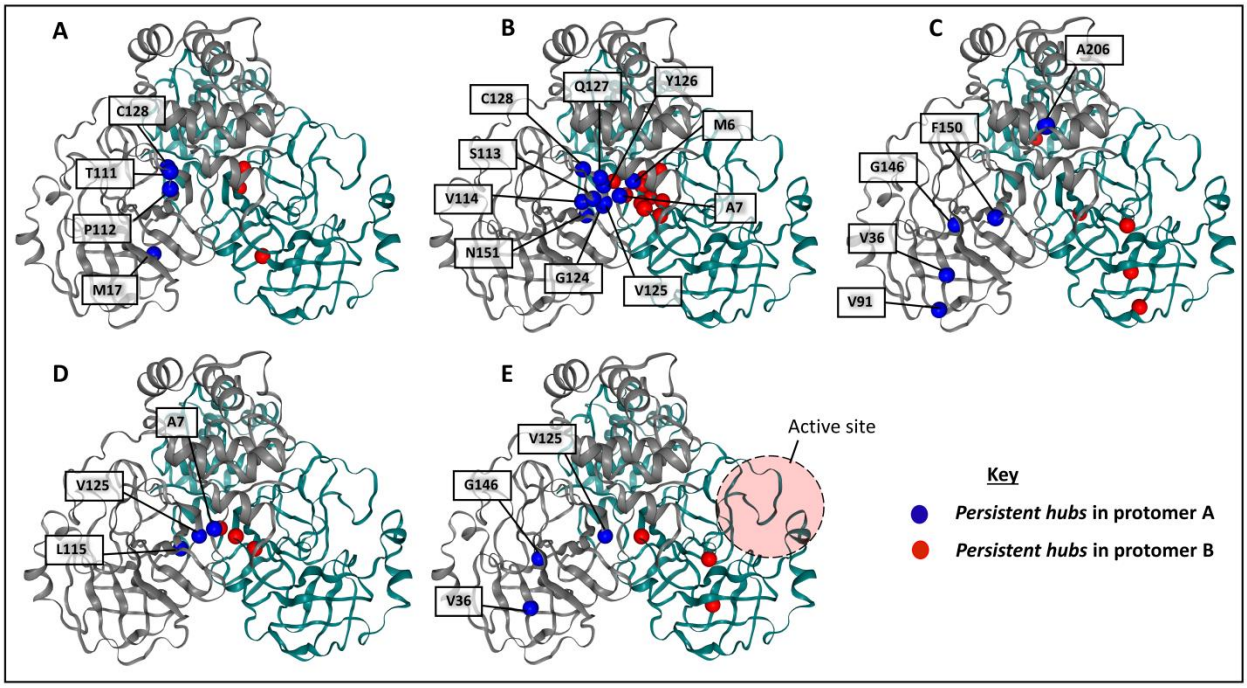

Figure 4. Cartoon representation of SARS-CoV-2 M^{pro} dimeric structure with the distribution of the *persistent hubs* as per five metrics of DRNs. (A) Averaged *BC*, (B) Averaged *CC*, (C) Averaged *DC*, (D) Averaged *EC*, (E) Averaged *KC*. Protomers A and B are shown as cartoon in teal and grey respectively. Protomer A *persistent hubs* are depicted in red spheres and protomer B ones are in blue. (F) Collective presentation of *persistent hubs* in orange spheres.

3.4.2. Closeness Centrality

CC is calculated as the inverse of the average of the shortest path length from the node to every other node; hence identifies the central nodes which are closer to most of the nodes. Previously we showed that residues with low average shortest path are correlated with the low mobility

(increased rigidity) of the protein [51]. Thus, high *CC* values are most likely to occur within the protein core. Previously, *CC* metric calculations over single static structures were used to identify active site residues with the support of other approaches, e.g. conservation, solvent accessibility [76,77] to distinguish them from residues located in the core. Here, our *persistent* averaged *CC hubs* are MET6, ALA7, SER113, VAL114, LEU115, GLY124, VAL125, TYR126, GLN127 and CYS128 **(Figure 3)**. Visual inspection of the residue mappings showed that they are all located at the vicinity of the dimeric center of mass found within the very stable domain II **(Figure 4B)**, as reported in our earlier work [12]. In the same work, ALA7 (part of the N-finger) was reported to be very rigid, and probably is the reason for the similar behavior in its immediate neighbor residue 6 within the same chain. SER113, VAL114, LEU115 and TYR126 are juxtaposed within the same beta sheet, supported by networks of H-bonding interactions, and are next to ALA7, which forms intraprotomeric alkyl interactions with VAL125 and interprotomeric H-bonds with VAL125 from the alternate protomer. More generally, these residues are mainly in direct contact with the center of the opposing protomer, and the reason for their high averaged closeness may be related to the way in which the protomers were reported to slide over each other, remaining at the same pivot point, centered at domain II (with the N-finger ALA7 also sandwiched in-between). Residue 7 appears crucial for maintaining the bulk of the averaged *CC* hubs.

3.4.3. Degree Centrality

DC defines the number of neighboring nodes around a given node, hence provides information on the local connectivity, but not how central it is in the entire network. *Persistent* averaged *DC* hubs in M^{pro} are comprised of residues VAL36, VAL91, GLY146, PHE150 and ALA206 **(Figure 3)**. 3D visualization of this metric showed that averaged *DC* tends to be concentrated at

the confluence of secondary structural elements, irrespective of inter- or intraprotomeric locations **(Figure 4C)**. VAL36 occurs within a beta sheet and establishes several types of non-bonded interactions with multiple residues within domain I, namely LYS88, LEU89, VAL68 and VAL18. Residue 91 occurs on another strand of the same beta sheet, next to VAL36, and establishes several types of non-bonded interactions with residues ASP34, LEU75, ARG76 and VAL77. Residue GLY146 is another *persistent DC* hub of potential significance, found next to the catalytic residue CYS145. It was found that GLY146 established durable intraprotomeric contacts with 10 residues, namely LEU27, ASN28, GLY29, CYS38, PRO39, CYS145, SER147, VAL148, MET162 and HIS163. This involved both intradomain (domain I) and interdomain interactions, and occurred in each protomer and both in the presence and absence of allosteric binding. The high averaged *DC* of GLY146 may be related to the fact that this area has to be kept relatively stable for the proper positioning of the catalytic residues CYS145 and HIS41, from domains II and I respectively. Given the presence of such a residue at the interface of domains I and II suggests that it may have a high *BC* as well, which is generally observed in both protomers A and B **(Figure 3)**. PHE150 similarly interacted with several residues in each protomer across all samples, and we observed high contact frequencies for 10 residues, namely VAL13, PHE112, SER113, VAL148, GLY149, ASN151, VAL157, SER158, PHE159 and CYS160, once more involving residues from domains I and II. ALA206 was similarly surrounded by 10 durable intraprotomeric contacts within domain III, being composed of residues VAL202, ASN203, LEU205, TRP207, TYR209, ALA210, PHE291, THR292, PRO293 and VAL296, in all cases. More generally, the shared high *DC* hubs seem to occur in each domain of the protein, probably due to their independent roles in maintaining the organization and integrity of the individual domains.

3.4.4. Eigencentrality

EC measures both the number of connections of a given node and its relevance in terms of information flow. It is based on a recursive allocation of centrality on the basis of nodes that draw importance from that of their successive connections, given that initial centrality is based on *DC*. Based on this calculation, one would expect high *EC* values to also have high *DC* values, or be in spatial proximity to high connectivity residues. However, we found that many of the high DC residues did not show up among the *EC* hubs, suggesting that *EC* is mostly gained via proximities to high *DC* residues, and do not necessarily have high connectivities themselves.

Persistent hubs of averaged *EC* for the M^{pro} reference protein comprised residues ALA7, LEU115 and VAL125 **(Figure 3)**. LEU115 is the only residue that maintained its importance according to averaged *DC* and *EC* measurements. Weighted residue contact analysis of this residue showed that LEU115 maintained high contact frequencies (>0.60) with residues CYS117, PRO122, VAL125 and SER147, irrespective of ligand binding. SER10 and VAL13 also showed high frequencies, except in the presence of SANC00302 where notable contact asymmetry was experienced; a similar pattern was observed for residues VAL148 and GLY149 in the presence of SANC00467. 3D visualization of the *EC* residues shows that it is concentrated around the interface of protomers A and B **(Figure 4D)**. The main message here is that high *DC* residues are sharing centrality to their immediate neighbors, and that the vicinity of the dimer interface seems to be the most residue-crowded area within the dimer. One should also bear in mind that centrality may also be coming from further degrees of separation. Other residues picked up as hubs in *DC* may be surrounded by fewer residues of high centrality, thus giving them less importance.

3.4.5. Katz Centrality

KC measures the relative degree of influence of a residue *i* within connected residues in a network. Irrespective of chain and ligand binding, nodes VAL36, VAL125 and GLY146 remained as hub nodes according to the averaged *KC* metric **(Figure 3)**.

Visualization of the averaged *KC* metric **(Figure 4E)** showed that this metric is an intermediate between averaged *EC* and averaged *DC*, with the former being more conservative than the latter when assigning relative node importance. *Persistent* averaged *KC hub* 125 was also central according to averaged *EC*; and VAL36, GLY146 were also *persistent hubs* according to averaged *DC*. The default attenuation coefficient (alpha = 0.1) appears to minimize the effect of more distant nodes in the network, such that it assigns centrality patterns intermediate to *DC* and *EC*.

In order to give an estimate of the hub similarities between those of *KC* and those from *DC* and *EC*, the Jaccard similarity coefficient (J) of hubs from protomer permutations was calculated, using as a rough estimate from the union of hubs across all states (ligand-bound and unbound) of the reference M^{pro}, for each of the protomers. The similarities were evaluated likewise: [J (KCA, KC'A), J (KCA, KC'B), J (KCB, KC'B) and J (KCB, KC'A)], where the subscript denotes protomer label and KC' denotes the complement of KC, in this case *DC* or *EC*. The hub similarities J (KC, DC) had a range of [0.375, 0.53] and those from J (KC, EC) had a range of [0.53, 0.76]. The observed ranges suggest that *KC* is more similar to *EC* than to *DC*. For comparison, J (KC, BC) and J (KC, CC) had ranges of [0.2, 0.3] and [0.16, 0.23], respectively denoting they tended not to share many hubs.

The reasons for the high centrality values for residues SER10, LEU115 and VAL125 are as explained in averaged *EC*, with the main difference being that the effect of distal nodes was

reduced due to the dampening coefficient. In this manner, averaged *KC* appears to improve the resolving power of averaged *EC*.

Overall, the heat map representation of the identified hubs according to the global top 5% foreach of the five DRN metrics **(Figure 3)** allowed us to identify *persistent hubs* according to each of the centrality measurements, suggesting that they are exposing different key aspects of mechanical signal transduction within the protease regardless of apo and allosteric ligand bound forms of both monomers **(Figure 4A-E)**. Collectively, these *persistent hubs* are spreading out from the allosteric site along the protein interface as well as to the antiparallel beta strands **(Figure 5)**. Even though previously it is not reported, we believe that the antiparallel beta strands, and especially the first two nearest to the dimer interface, are functionally highly important.

We also identified a number of changes to hub existence in the presence of potential allosteric modulators and these changes were investigated as explained in the next section.

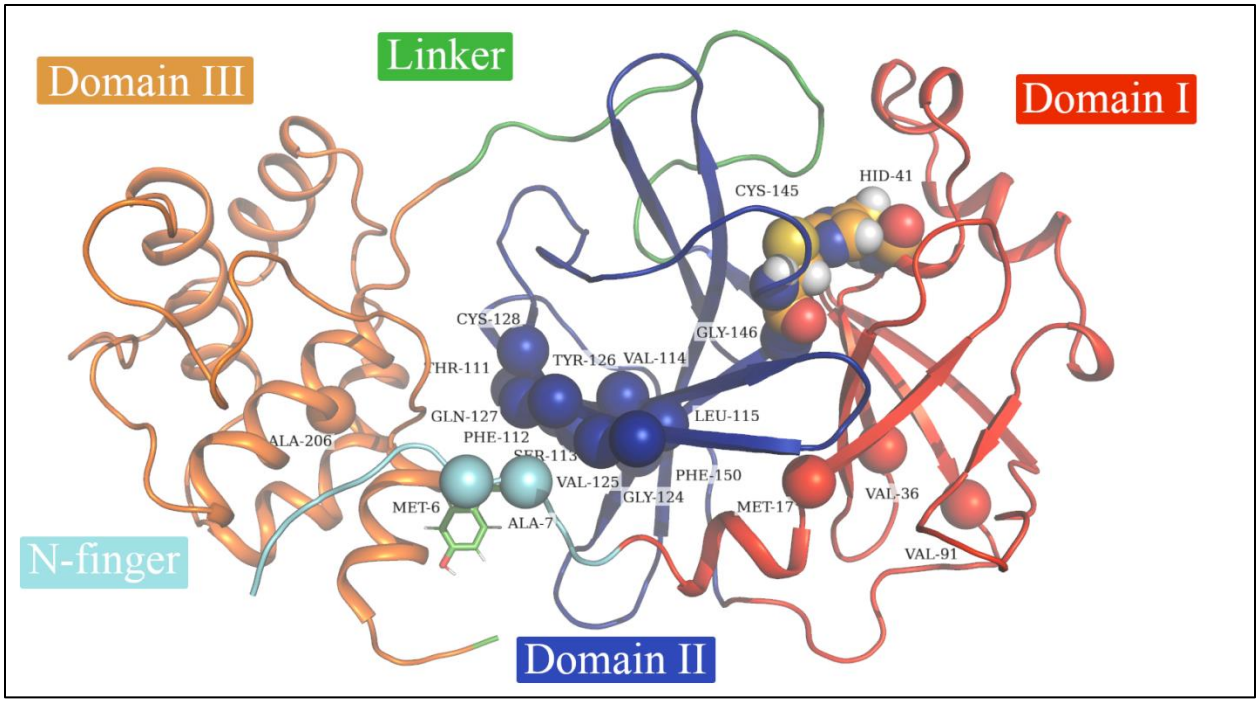

Figure 5: Cartoon representation of SARS-CoV-2 M^{pro} protomer A with the collective presentation of *persistent hubs* in spheres colored according to their domains. Catalytic residues are HIS41 and CYS145 are also depicted as spheres.

3.5. Establishing subnetworks for further investigation of hub changes upon allosteric binding within the reference homodimer

We also observed another layer of information within the homodimer, which exists due to the symmetry of the protomers, despite the adjustment made to present the ligand-bound protomer as the one left-hand side (protomer A) in **Figure 3**. We hypothesize that it is possible for a homodimer to switch states, because of their sequence identity. This is likely true for the apo state, but may also apply to the asymmetrically occupied allosteric sites, depending on how effective the allosteric pocket occupation is. This approach may also reveal if allosteric activity is manifested as a change of hub symmetry in the protein dynamics - for instance one or more hubs consistently appearing in one or even both of the protomers, when the allosteric site is

occupied by a ligand. For this reason, the same data set **(Figure 3)** is further analyzed using another concept that we demonstrate in this section.

For each allosterically bound ligand, a network was built using the detected centrality hubs as nodes, and the chains to which they belong. A subnetwork was then prepared by combining the edges from the apo protomers A and B, and those from the ligand-exposed dimer, while also tracking the protomer labels, given the ligands had settled at one chain of the blindly docked dimer (**Figures 6 – 10**). As indicated before the ligand was assigned to protomer A. In the case of SANC00467 where the compound bound to chain B, the chain label was swapped. The systematic hub representation was done to further investigate whether ligand binding had an effect, keeping a record of the chain labels, as we hypothesized that a ligand's presence in the allosteric site may manifest its effects not only in the bound protomer, but also in the unbound protomer, within the same dimer.

The hub data set was analyzed based on the logic described in **Table 1**. To simplify the terminology used to describe the presence of hubs within any combination of protomers, the term "*score*" is used to specify the number of protomers where the hub is present. In other words, if a hub is present in both protomers of the apo protein, the hub (irrespective of the DRN metric) will have a *score of 2*. Similarly, if a hub is present in each protomer of the apo dimer, and in each protomer of the ligand-bound dimer, this hub will have a *score of 4 (constitutive hub)*. Further, higher confidence was assumed on the basis of complete loss or complete gain of any hub in each *constitutive* protomer from either the apo protein or the ligand-exposed enzyme. Lower confidence was assumed when a single hub was gained (i.e. *score 0* to *1*, or *1* to *2*) or lost (i.e. *score 1* to *0*, or *2* to *1*) within a single protomer, out of all four protomers (i.e. the set of protomers: apo chains A/B and complex chains A/B), to account for the part stochasticity of MD

simulations. Higher confidence was given to these weaker signals when they were conserved across several ligand-bound states. Cases where asymmetric hub distributions occurred (i.e. a hub was found in only one protomer from each of the apo and the ligand-bound dimers) were ambiguous, given the fact that the apo dimer already expressed both hub states.

Here we will mainly focus on cases where we observed *score of 2* and *score of 1* gains or losses from each ligand-bound dimer with respect to the apo protein, in order to extract high likelihood ligand-induced changes.

3.5.1. Betweenness Centrality

While PHE140 was a *constitutive* hub in the presence of SANC00630, SANC00468 completely lost node 140 upon ligand binding (*score of 2* loss). SANC00630 gained hub 13 (*score of 2* gain) with respect to apo structure (**Figure 6**). The loss of *BC* from this "chameleon" switch residue (PHE140) suggests contact loss in its vicinity. VAL13, on the other hand, is found close to the N-finger – in a region where we previously reported lengthening and shortening of the alpha helix and suggested its possible involvement in dimer stability [12].

While there are many *score 1* hub gains and losses with each ligand-bound dimer (with respect to the dimeric apo protein), we report the ones which have the highest conservation among all these lower confidence cases, independent of the bound ligand. Hub residue GLY11 systematically changed from a *score of 0* in the apo dimer to a *score of 1* in the dimeric complexes, suggesting an increased use or stabilization around this residue upon ligand binding. Coincidentally, hub residue SER10 systematically transited from a *score of 1* in the apo state to a *score of 0* upon ligand binding. The fact that these two residues are next to each other, and in fact interact with their interprotomeric counterpart suggests a possible rerouting of information flow in their vicinity upon introduction of a ligand. *Score 2* to *1* (i.e. from the apo to the ligand-bound

dimer) changes appeared not as consistent, but showed some agreement on hub nodes being lost from one of the "chameleon switches" in subsite S1, similar to what was seen more strongly in the presence of SANC00468, where both nodes were lost. This was observed for residues SER139 and/or PHE140 when exposed to SANC00302, SANC00303, SANC00467 and SANC00469. *Score 1* to *0* changes were not observed when shifting the reference protein from an apo to a ligand-bound state, for any of the centrality metrics.

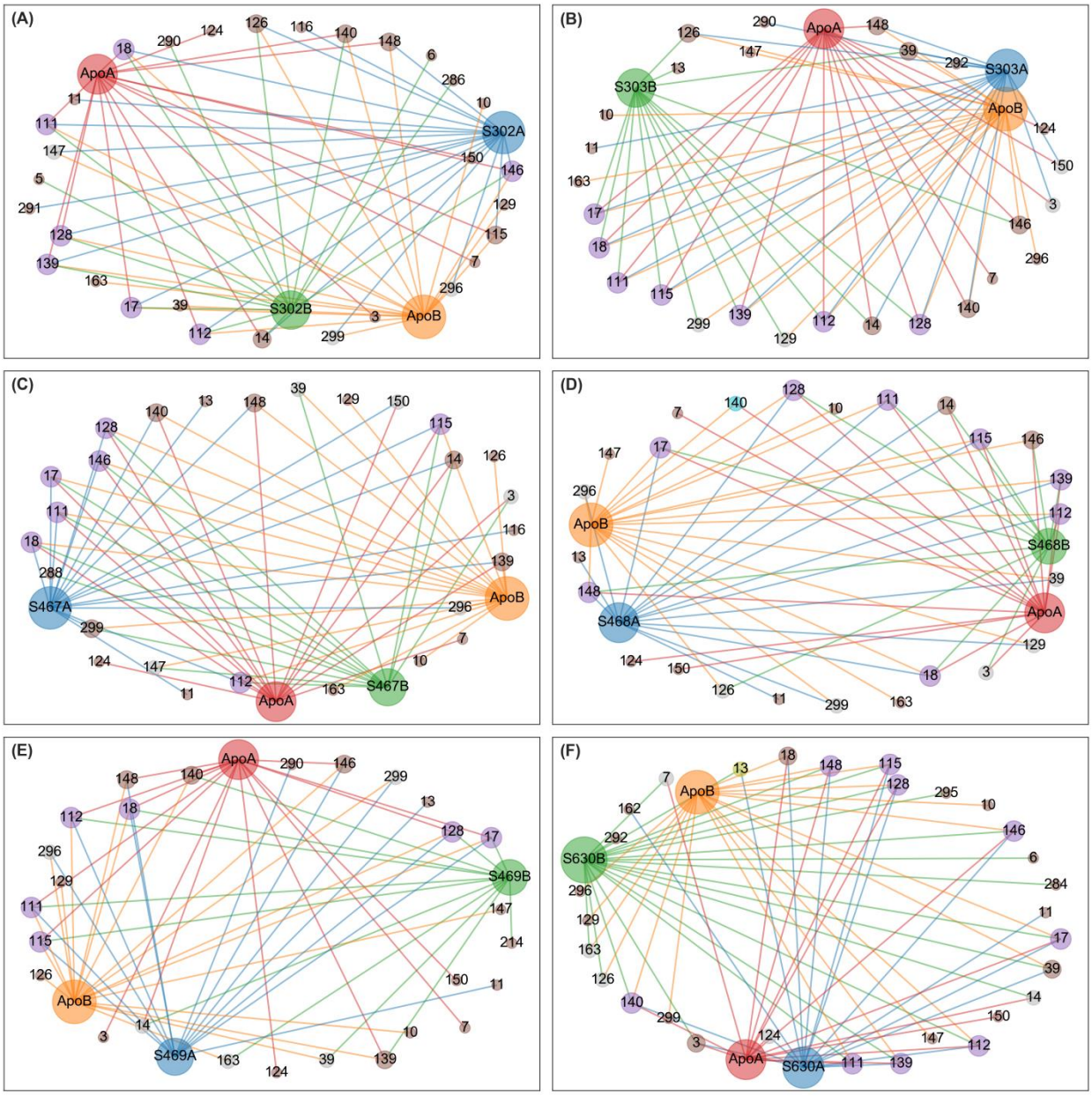

Figure 6. Averaged *betweenness centrality* hubs represented as a networks for both the apo state and the six allosterically-bound complexes. Each of the sub-figures represents the subnetwork obtained for each of the allosterically bound ligands, namely (A) SANC00302, (B) SANC00303, (C) SANC00467, (D) SANC00468, (E) SANC00469 and (F) SANC00630, when merged with the apo protein, in each case. Red, orange, blue and green nodes (and edges) depict the protomers (apo chains A and B, and complex chains A and B, respectively) to which a hub belongs. Each node is also scaled by its *score* – i.e. the number of edges it holds. Hubs that are present in all 4 protomers are in purple. *Score 2* loss and gains from the reference are colored yellow and cyan, respectively. *Score 1* losses and gains are colored brown. Inconclusive hubs are in grey.

3.5.2. Closeness Centrality

SER10 was a *constitutive hub* to five ligand-bound states, except SANC00302. *Score 2* gains of high *CC* hubs were observed for residue 4 and 5 in the presence of SANC00302, SANC00303 and SANC00468 – residue 4 also experienced a *score 1* gain in the presence of SANC00630, while residue 5 experienced a similar gain in the presence of SANC00469 (**Figure 7**). THR111 was also gained as a *score of 2* hub, only in the presence of SANC00468. GLY138, which is part of the S1 subsite, manifested itself as a hub in only one monomer of the apo protein, transitioning to a *score of 0* upon ligand binding in five out of the six bound states. Upon visual inspection, we find that this residue is next to residue ARG4 on the alternate protomer, even though they do not appear to interact via non-bonded interactions. By measuring the change in their C-alpha interprotomer distances [i.e. the distance between residue pair (4A, 138B) minus that between residues (4B, 138A)] in each of the apo and the ligand-bound M^{pro} showed us that one of the residue pairs from the apo form had a visibly larger variance in equilibrium distance compared to those all the ligand-bound proteins, as it had an interprotomer distance interquartile range (IQR) of 0.12 nm, compared to a maximum IQR of 0.07 nm across all six the ligand-bound states. The maximum upper quartile additionally informs us of the higher extent of dimeric

asymmetry for the residue pair for the apo protein (0.22 nm), compared to that observed in the ligand-bound proteins, which displayed a maximum value of 0.13 nm overall. While counterintuitive, it would seem that asymmetry favors a higher centrality at one of the GLY138-ARG4 (the N-finger from one protomer and domain II from the other) interfaces at the expense of the other in the apo state, while ligand occupation of only one of the detected allosteric sites, has a general stabilizing effect, which dilutes the centrality more evenly.

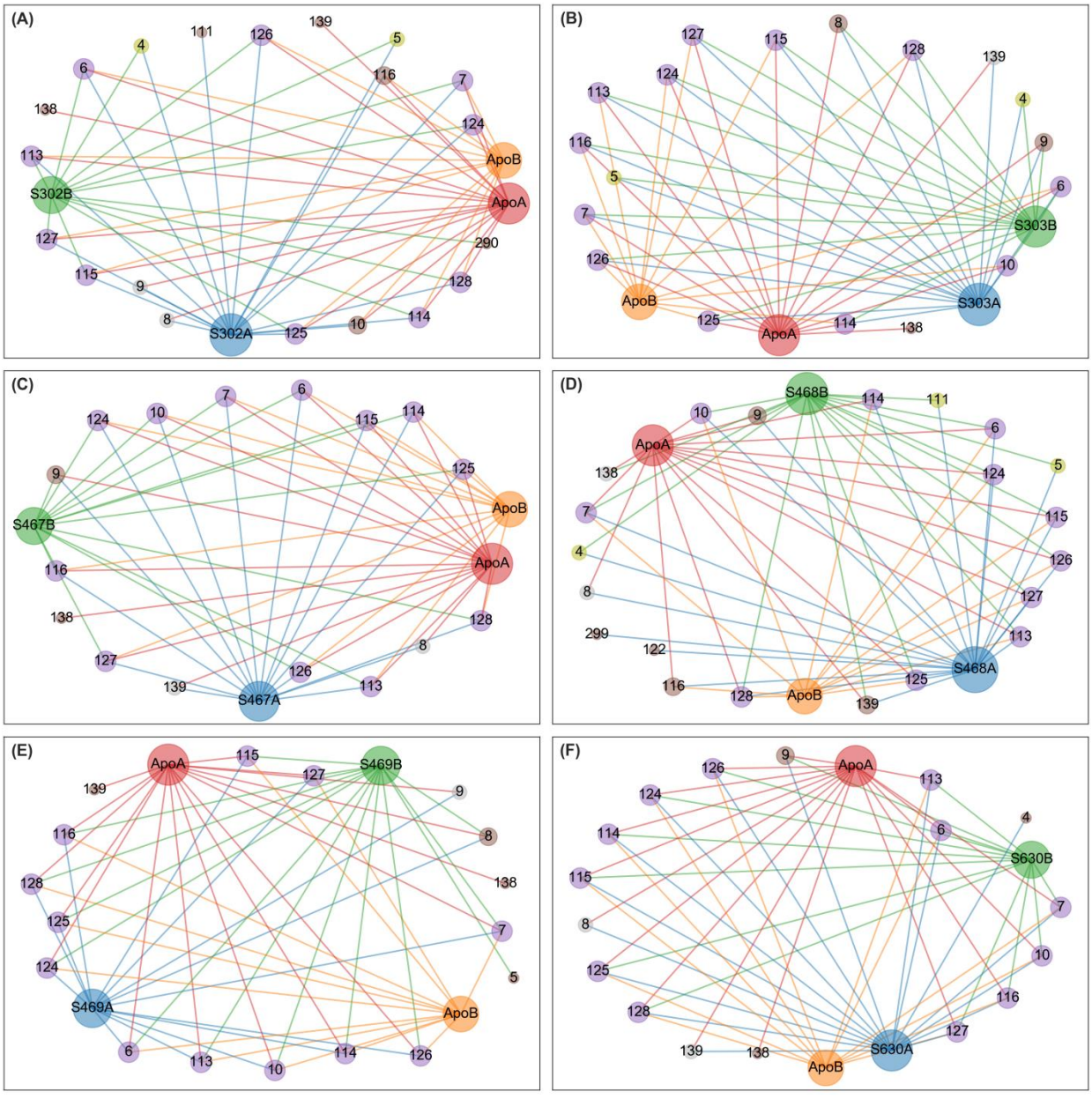

Figure 7. Averaged *closeness centrality* hubs represented as a networks for both the apo state and the six allosterically-bound complexes for the reference dimer. Each of the sub-figures represents the subnetwork obtained for each of the allosterically bound ligands, namely (A) SANC00302, (B) SANC00303, (C) SANC00467, (D) SANC00468, (E) SANC00469 and (F) SANC00630, when merged with the apo protein, in each case. Red, orange, blue and green nodes (and edges) depict the protomers (apo chains A and B, and complex chains A and B, respectively) to which a hub belongs. Each node is also sized by its *score* – i.e. the number of edges it holds. Hubs that are present in all 4 protomers are in purple. *Score 2* loss and gains from the reference are colored yellow and cyan, respectively. *Score 1* losses and gains are colored brown. Inconclusive hubs are in grey.

3.5.3. Degree Centrality

LEU115 was a *constitutive hub* to five ligand-bound states, except SANC00302 (**Figure 8**). Scaffold-related conservation patterns were not apparent using this metric, however some differences did occur. Residue LEU115, which occurs in proximity to the *persistent hub*, PHE150, was highly crowded and formed several durable contacts with its neighbors, namely VAL114, ALA116, CYS117, PRO122 and VAL125. LEU115 had a high frequency contact with PRO9 in only one chain in the presence of SANC00469 and a low frequency contact with VAL13 in only in one chain in the presence of SANC00302.

A *score 2* hub gain was experienced by VAL18 when exposed to SANC00303, SANC00467, SANC00468, SANC00469 and SANC00630. The same residue incurred a *score 1* gain in the presence of SANC00302. Upon contact visualization, we found the systematic significant increase in contact frequency between VAL18 and GLN69 in each protomer upon ligand binding. While their C-alpha distances were relatively similar throughout the apo and ligand-bound M^{pro} (averaging 0.59 nm), the C-beta distances were significantly larger in the apo (average of 0.69 and 0.70 nm in the apo protomers) compared to those of the ligand-bound states

(averages ranging from 0.63 to 0.65 nm), which suggests a rotational decrease of the C-beta distance upon ligand binding. A *score 2* gain was also experienced by residue GLY29 when exposed to SANC00468 and SANC00630. 3D visualization shows that GLY29 is H-bonded to VAL18, and together with GLN69 they form a geodesic path travelling directly across antiparallel beta strands. The proximity and arrangement of these three residues may suggest they may act in a concerted manner.

Score 2 losses were observed for VAL86 when exposed to SANC00467 and SANC00468; and for residues LEU253 and VAL296 in the presence of SANC00302, indicating that the connectivity around these areas was reduced. Conserved *score 1 to 0* changes were observed for residue TYR126 in the presence of ligand binding, suggesting a possible increase in local compaction in that area in the presence of any of the ligands.

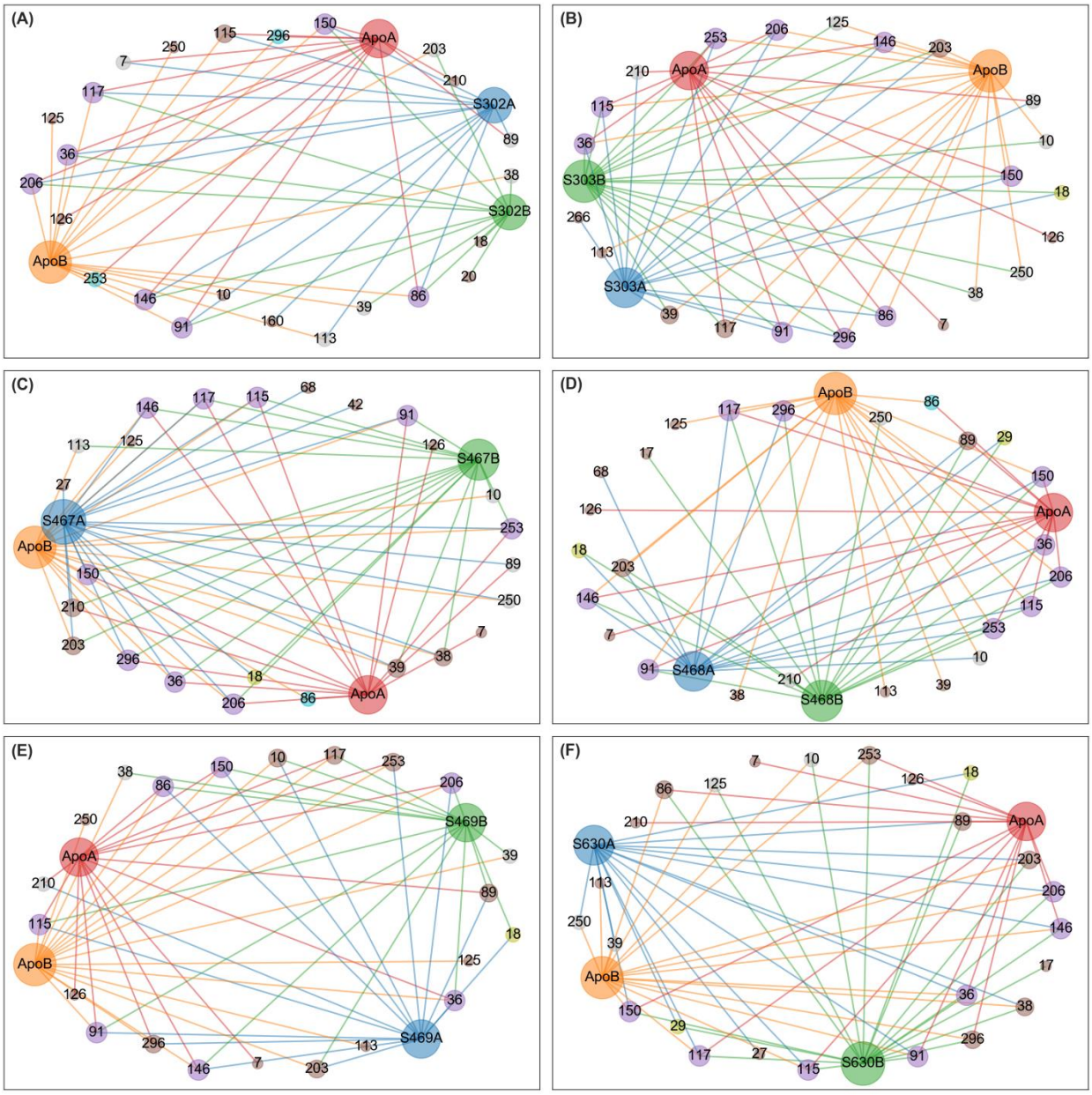

Figure 8. Averaged *degree centrality* hubs represented as a networks for both the apo state and the six allosterically-bound complexes. Each of the sub-figures represents the subnetwork obtained for each of the allosterically bound ligands, namely (A) SANC00302, (B) SANC00303, (C) SANC00467, (D) SANC00468, (E) SANC00469 and (F) SANC00630, when merged with the apo protein, in each case. Red, orange, blue and green nodes (and edges) depict the protomers (apo chains A and B, and complex chains A and B, respectively) to which a hub belongs. Each node is also sized by its *score* – i.e. the number of edges it holds. Hubs that are present in all 4

protomers are in purple. *Score 2* loss and gains from the reference are colored yellow and cyan, respectively. *Score 1* losses and gains are colored brown. Inconclusive hubs are in grey.

3.5.4. Eigencentrality

PRO9 and SER10 were *constitutive hubs* to five ligand-bound states, except SANC00302 (**Figure 9**). GLY146 experienced a *score 2* gain in the presence of SANC00302 and SANC00630. The same was observed for CYS38 in the presence of SANC00302. While *score 1* gains from *1 to 2* were not completely conserved, hub score changes from *0 to 1* were conserved, comprising residues MET17, ASN28 and GLY29 in the presence of ligand binding, suggesting an increase in centrality in the vicinity of these residues. Visual inspection shows that MET17 is proximal to ASN28, which is next to GLY29 on a beta strand. The high averaged *EC* for MET17 is likely due to its high degree centrality combined to that of VAL18. It is possible that ligand binding further stabilizes its residue neighborhood, compared to the absence of occupation of the allosteric pocket. Hub residues ASN28 and GLY29 appear to draw centrality from the higher degree centrality residues VAL36 and VAL18. Together these domains I residues line the interface of domain II, indicating a possible stabilization around this area in the presence of an occupied allosteric pocket.

A very interesting communication path emerges when combining the *persistent* averaged *EC* hubs (ALA7, LEU115 and VAL125) and ones collectively gained by ligand binding (MET17, ASN28 and GLY29) (**Figure 10**). We thus describe the path MET7-VAL125-LEU115-MET17-GLY29-ASN28-HIS145, which originates from the N-finger to converge towards the catalytic HIS145, which is itself proximal to second catalytic residue CYS41. From these observations, it is possible that the intradomain MET17-LEU115 contact may be a crucial information path for the SARS-CoV-2 M^{pro}, as it plays a pivotal role in maintaining relaying information from the

allosteric pocket. This complements our previous observation of the bridging function of SER17 in the apo M^{pro} using averaged *BC*. Extending on the finding of a common path shared by ligand binding, we also describe an alternate path SER9-PRO10-LEU115-MET17-GLY29-ASN28-HIS145 being specifically used in the apo state and the ligand-bound states, with the exception of SANC00302. It is possible that this difference stems from the lack of stability of this compound in the pocket. The complete communication paths are further analyzed in Part III.

A *score 2* loss was observed by GLY149 in the presence of SANC00630. The same residue experienced a *score 2 to 1* change in the presence of the other compounds. This residue is found at the bifurcation of two beta hairpins that do not completely line up all the way into a beta sheet, close to a main contributor of degree centrality, residue LEU115. The generic decrease in averaged *EC* in at least one protomer points to a possible loss of contact in this hub's vicinity that occurs upon ligand exposure. Residue 9 was also lost as a *score 2* hub, only in the presence of SANC00302. *Score 1 to 0* changes were inconsistent.

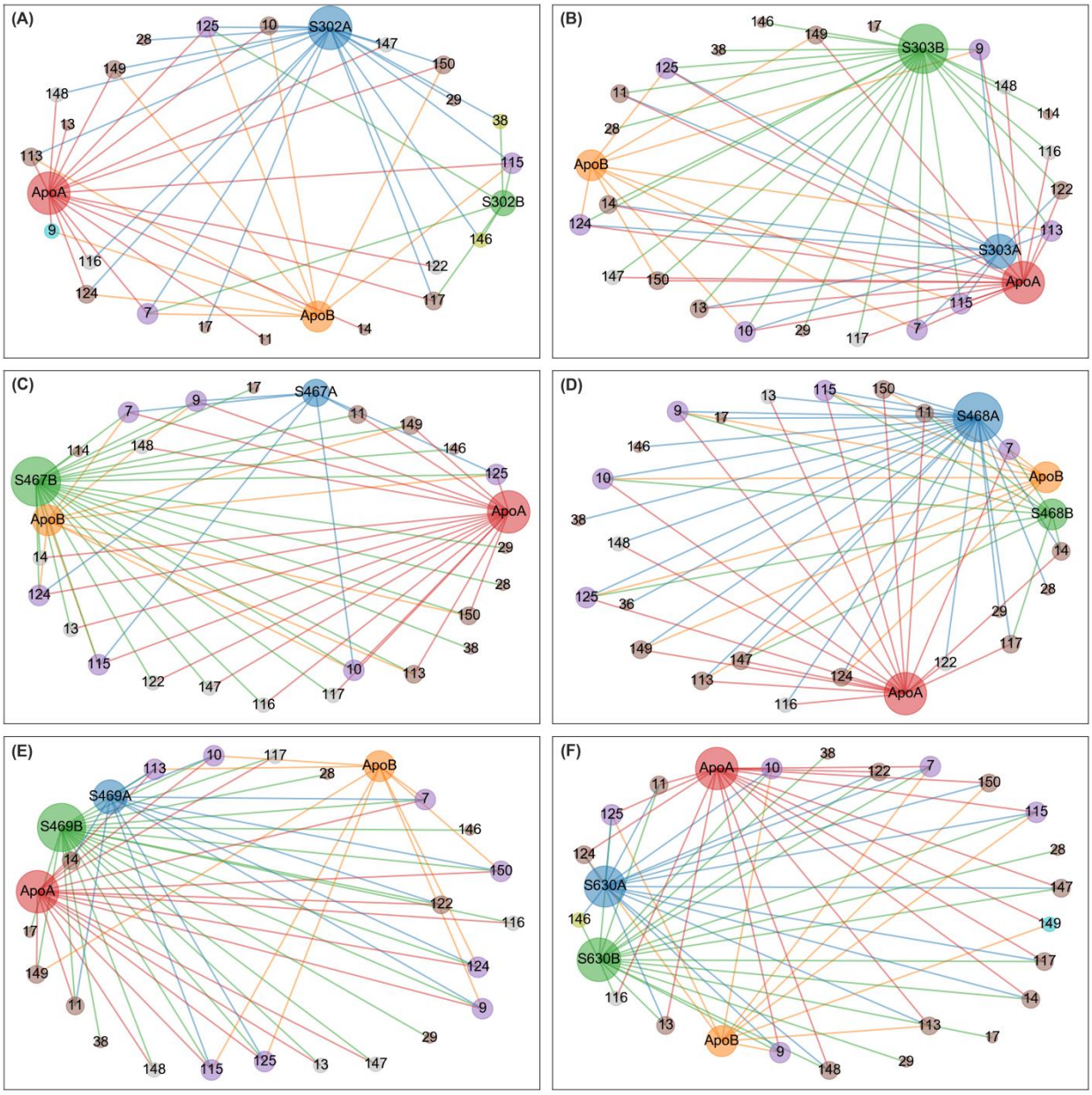

Figure 9. Averaged *eigencentrality* hubs represented as a networks for both the apo state and the six allosterically-bound complexes. Each of the sub-figures represents the subnetwork obtained for each of the allosterically bound ligands, namely (A) SANC00302, (B) SANC00303, (C) SANC00467, (D) SANC00468, (E) SANC00469 and (F) SANC00630, when merged with the apo protein, in each case. Red, orange, blue and green nodes (and edges) depict the protomers (apo chains A and B, and complex chains A and B, respectively) to which a hub belongs. Each node is also sized by its *score* – i.e. the number of edges it holds. Hubs that are present in all 4

protomers are in purple. *Score 2* loss and gains from the reference are colored yellow and cyan, respectively. *Score 1* losses and gains are colored brown. Inconclusive hubs are in grey.

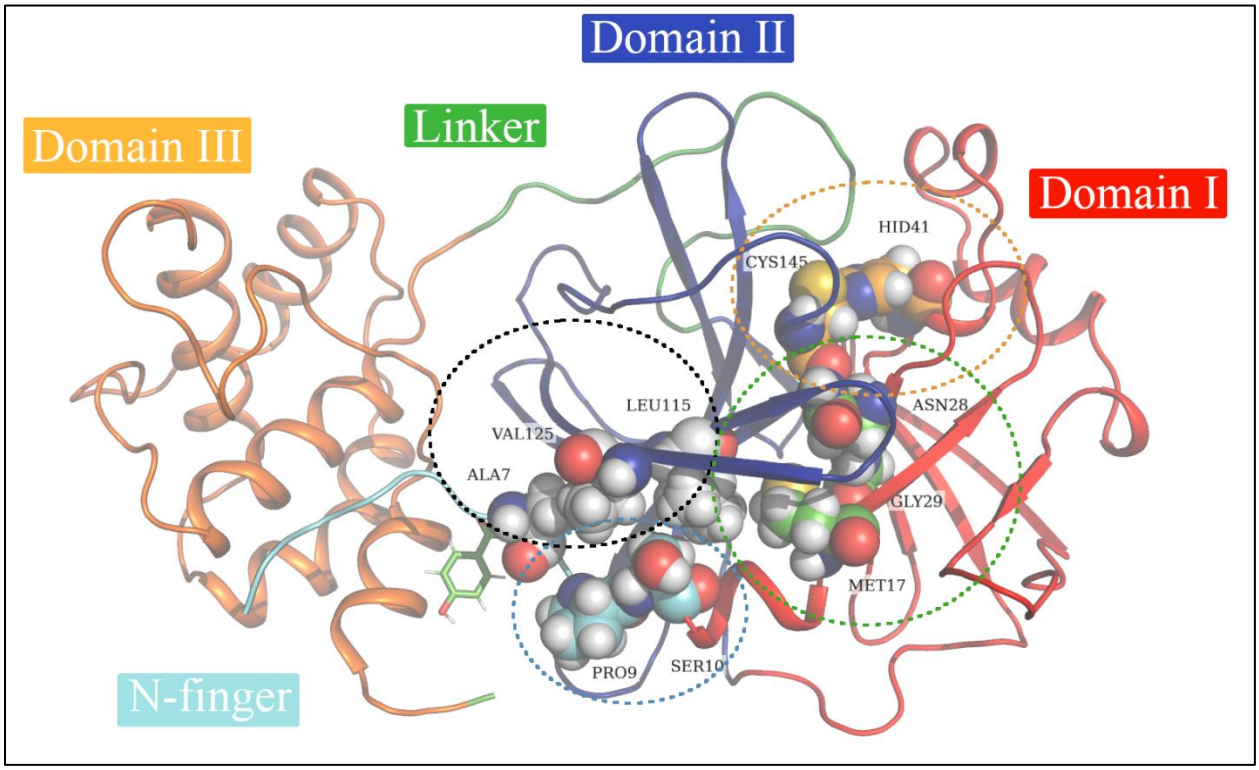

Figure 10. The path traced by averaged *EC* hubs, starting from the allosteric ligand towards the catalytic residue. The protease is depicted by a cartoon representation onto which the averaged *EC* hub residues are overlaid as sphere representations, together with the non-hub catalytic residues HIS41 and CYS145 (circled in orange). *EC persistent hub residues* are circled in black; the alternate path is circled in blue; and the one triggered by the binding of all ligands is circled in green. One of the compounds is also shown in stick figure representation, as an example.

3.5.5. Katz Centrality

SER10 and LEU115 were *constitutive hubs* to five ligand-bound states, except SANC00302 (**Figure 11**). SER10 was also a constitutive hub and LEU115 was a persistent hub in *EC*. *KC* hubs residues VAL36, GLY146, LEU115 were also central according to averaged *DC*.

Hub node 29 gained by a *score of 2* in the presence of SANC00302, SANC00303, SANC00468 and SANC00630, with respect to the apo protein, and the same residue incurred a *score 0 to 1* gain in the presence of SANC00467 and SANC00469, indicating that this residue gains in *KC* in at least one protomer upon ligand exposure. *Score 2* gains are also observed for residues 28 (in the presence of SANC00302 and SANC00468), 39 (in the presence of SANC00467), 17 (in the presence of SANC00468) and 20 (in the presence of SANC00630). *Score 2* losses comprised residue 7 in the presence of SANC00467 and SANC00468; and residue 10 in the presence of SANC00302. The similarities in hub response patterns in the presence of SANC00467 and SANC00468 may suggest that these changes may be related to their common scaffold, or to a similar conformational sampling.

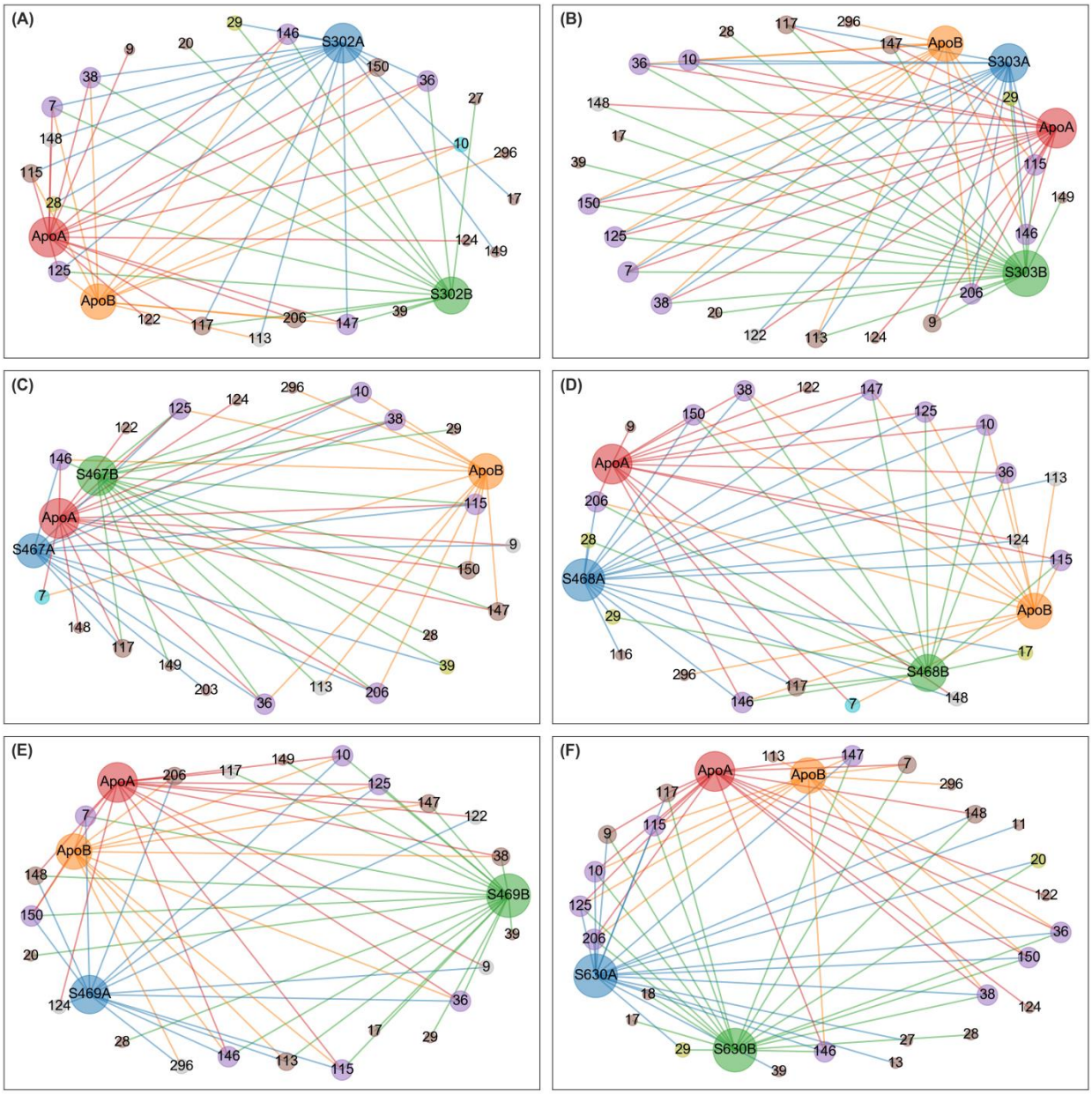

Figure 11. Averaged *Katz centrality* hubs represented as a networks for both the apo state and the six allosterically-bound complexes. Each of the sub-figures represents the subnetwork obtained for each of the allosterically bound ligands, namely (A) SANC00302, (B) SANC00303, (C) SANC00467, (D) SANC00468, (E) SANC00469 and (F) SANC00630, when merged with the apo protein, in each case. Red, orange, blue and green nodes (and edges) depict the protomers (apo chains A and B, and complex chains A and B, respectively) to which a hub belongs. Each node is also sized by its *score* – i.e. the number of edges it holds. Hubs that are present in all 4

protomers are in purple. *Score 2* loss and gains from the reference are colored yellow and cyan, respectively. *Score 1* losses and gains are colored brown. Inconclusive hubs are in grey.

Collectively, the analysis of hub transitions induced by allosteric binding in the homodimer via subnetworks showed that the DRN metrics shared similarities, but also emphasized importance in different ways. By focusing on sample-wide centrality at the expense of protein-specific importance, and possibly narrowing the information content, we extracted key features in the modulation of the mirrored allosteric pocket of M^{pro}. Changes in averaged *BC* hub patterns hinted at a possible rerouting of information flow from 10 to 11 induced by ligand stabilization. *CC* hub transitions lead to the observation of a ligand-induced stabilization, verified by the decreased interprotomer asymmetry of the distances between residues 4 and 138, compared to the apo protein. *DC* hubs detected a consistent decrease in compaction around TYR126, in the presence of any ligand, and further analysis showed an associated intraprotomeric side-chain rotation involving residues VAL18 and GLN89, upon ligand binding. The aggregation of *persistent* and gained averaged *EC* hubs revealed a common path connecting the allosteric occupation by ligands, to the active site, involving the interaction between MET17 and LEU115 (**Figure 10**). It is possible that due to its proximity to one of the *EC* hubs (VAL125) that TYR126 may have a role in the path traced via averaged *EC*. Averaged *KC* mainly showed similarities with averaged *EC* and *DC*.

PART III:

3.6. The stability of allosteric modulators in the presence of evolutionary mutations

Coronaviruses, including SARS-COV-2, depend on RNA-dependent RNA polymerase (RdRp) for RNA synthesis [8,78]. Due to the error-prone nature of RdRp, they can accumulate high rates of mutations, some of which may alter their virulence and antigenicity. As there is no drug for

COVID-19 yet, we do not know which of these mutations could affect drug efficacy, or cause drug resistance. Hence it is important to understand the potential effects of a range of mutations in hit identification studies. To date, only a few studies have considered the impact of SARS-COV-2 M^{pro} variations in apo protein [4,12,79]. However, to our knowledge, there is no systematic research incorporating hit identification with analysis of structural and functional effects of variations. Here, we further analyzed the behavior of six potential allosteric modulators that we identified in reference protein in the presence of early evolutionary mutations. In order to quickly examine the stability of ligands in a total of 300 mutant systems (6 ligands x 50 mutant proteins), the ligand poses via ligand RMSDs were calculated and ligand RMSD kernel plots produced for the last 10 ns of 20 ns simulations of mutant-ligand and of 100 ns MD simulations of reference protein-ligand complexes (**Figure 12**, **Figure S1**). Overall, all ligands were well anchored in the allosteric pocket of mutant proteins, as seen from the ligand RMSD median values below 2.0 Å. Regarding the variations in ligand motion, a more stable conformation (unimodal distribution) was observed across SANC00468, SANC00467 and SANC00469 bound to mutant proteins, followed by SANC00630 as compared to the conformational stability of SANC00302 and SANC00303 (**Figure 12**). This observation was in agreement with docking results where the first four compounds exhibited high stability through various hydrogen bond interactions with key allosteric site residues (**Figure 2C**, **Table S2**). A closer view of each ligand revealed the subtle movement of the bromide SANC00302 and SANC00303 and the hydroxyl groups of SANC00630 in some mutant proteins as seen from the bimodal distributions (**Figure 12**).

Figure 12 was further evaluated to calculate a consensus score across six ligands within each mutant system. For that, a table (**Table S3)** was prepared in which the y-axis contained individual mutant proteins and x-axis was for six ligands. For each ligand, kernel plots were

checked and the ligands with a unimodal distribution in each mutant protein system received a tick (✓) in the table; the selected ones are also indicated in **Figure 12** with black oval shape in the x-axis. Surprisingly, out of 50 mutant proteins, only three of them (A173V, N274D and R279C) received a consensus score of six **(Table S3)**, meaning all ligands in these mutant proteins stayed stable over the MD simulation. Over all the systems, the best performing ligands were SANC00468 and SANC00469, which gave highly stable motions for 43 and 41 mutant samples, respectively (**Table S3**).

The concept of allosteric effects of mutations and their role in the modulation of protein activity was previously discussed in literature [22,80]. Our results, here, demonstrated the importance of incorporating the mutation information in hit identification, as mutations might have distal-allosteric effects (allosteric mutations) to the ligand-binding site. As a next step, we further calculated the five DRN metrics (*BC*, *CC*, *DC*, *EC*, *KC*) for these 50 mutant systems complexed with each allosteric modulator and compared to the reference system as detailed in the next section.

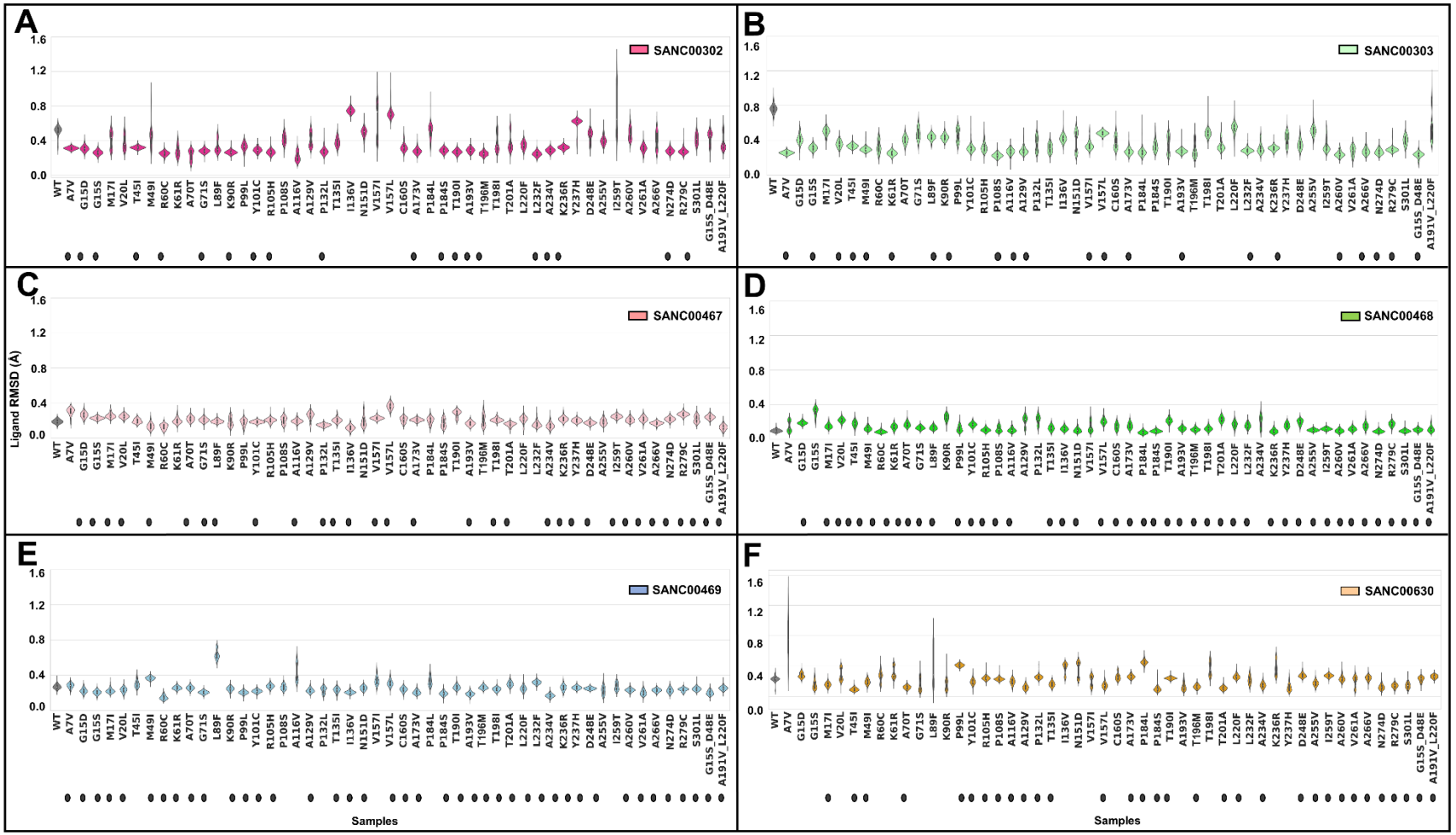

Figure 12. Kernel density distribution plot of ligand RMSD values in ligand-bound wildtype (WT) and mutant systems extracted from the last 10 ns simulation. Each panel A to F is for the ligand indicated.

3.7. *Persistent and super persistent hubs* of the averaged DRN metrics in the presence of mutations

In Part II, we identified the *persistent hubs* for each averaged DRN metric on the basis of their existence in both protomers in the absence and presence of all ligands. This gave us MET17, THR111, PHE112 and CYS128 for averaged *BC;* MET6, ALA7, SER113, VAL114, LEU115, GLY124, VAL125, TYR126, GLN127 and CYS128 for averaged *CC*; VAL36, VAL91, GLY146, PHE150 and ALA206 for averaged *DC*; ALA7, LEU115 and VAL125 for *EC*; VAL36, VAL125 and GLY146 for *KC* (**Table 2**; reference rows).

Here, to analyze the residue-residue communications, in the presence of potential allosteric modulators in mutant protein systems, we calculated the global top 5% averaged *BC, CC, DC, EC, KC* metrics for 51 protein systems (50 mutant protein systems and reference protein) (**Figure S2-S6**); and extracted the *persistent hubs* on the basis of their conservation in both reference protein and mutants bound to a specific ligand (**Table 2**). If a *persistent hub* is retained across all the ligand systems (in both protomers), then we called it a *super-persistent hub*.

In the case of averaged *BC* **(Table 2, Figure S2)**, we did not observe any *super-persistent hub*; however, MET17 retained as the main *persistent hub* in all protein-ligand systems except in the presence of SANC00630 in which the protomer B of the mutant M49I protein lost the hub node. Hub 111 was retained as *persistent hub* in the presence of SANC00468, SANC00469 and SANC00630; and hub CYS128 was *persistent* in the presence of SANC00302 and SANC00467. PHE112 remained as a *persistent hub* in all 51 protein systems complexed with SANC00630.

Super-persistent hubs of averaged *CC* for 51 protein systems of all allosteric modulators were observed for residues MET6, ALA7, SER113, LEU115, VAL125, TYR126 and GLN127 **(Figure S3)**. In the presence of SANC00302, mutant protein A7V and in the presence of

SANC00468, the mutant protein G15S lost hub 124 in their protomer B; hence GLY124 stayed as a *persistent hub* in only SANC00303, SANC00467, SANC00469 and SANC00630 ligand systems. Additionally, hub 128 stayed as *persistent* in all ligand systems except SANC00468 in which it was lost in protomer B of K61R, A193V, I259T and N274D mutant protein systems in the presence of SANC00468. We also observed a new *persistent hub* for residue 10 in the presence of SANC00303.

In the case of averaged *DC* **(Table 2, Figure S4)**, we did not observe any *super-persistent hub* over all the ligand systems. The key *persistent* averaged *DC hubs* in the presence of most allosteric ligands were comprised of residues 150 and 206. PHE150 was a *persistent hub* to all ligand system, but was missed as a hub in protomer B of P184L and A116V mutant proteins in the presence of SANC00467. ALA206 was again a *persistent hub* to all ligand systems, except SANC00468 in which the hub was missing in protomer A of L220F mutant protein. Interestingly *persistent hub* GLY146 was lost in the presence of all allosteric modulators.

Again, we did not observe any *super-persistent hub* in *EC* metric either **(Table 2, Figure S5)**. But a new *persistent hub* (residue 10) in the presence of SANC00303 and SANC00469 was obtained. LEU115 was retained as *persistent hub* in all ligand systems, except in the presence of SANC00468 (the hub node was lost in protomer B of double mutant protein (A191V, L220F)); and in the presence of SANC00630 (the hub node was lost in protomer B of two mutant proteins M49I and A193V). Interestingly, the *persistent hub,* ALA7, was lost in the presence of all allosteric modulators according to averaged *EC* metric.

LEU115 was also the key *persistent hub* according to *KC* metric, and it was only lost in the presence of SANC00302 due to absence of the hub node in protomer B of reference protein **(Table 2, Figure S6)**. Two new *persistent hubs* (residues 10 and 150) were introduced in the

presence of SANC00469. Interestingly, the *persistent hub,* VAL36 was lost across all allosteric modulators according to averaged *KC* metric.

Table 2: *Persistent hubs* (in grey) as observed in reference protein (apo and all ligand-bound states) and their comparison to 51 protein systems (reference protein + mutants) in the presence of each allosteric modulator. *Super-persistent hubs* are highlighted in orange. The hubs that are lost across all ligand systems are in pale blue.

Metric		HUB										
Betweenness Centrality (*BC*)	**Reference**	**17**			**111**		**112**			**128**		
	SANC00302	17								128		
	SANC00303	17										
	SANC00467	17								128		
	SANC00468	17			111							
	SANC00469	17			111							
	SANC00630				111		112					
Closeness Centrality (*CC*)	**Reference**	**6**	**7**		**113**	**114**	**115**	**124**	**125**	**126**	**127**	**128**
	SANC00302	6	7		113		115		125	126	127	128
	SANC00303	6	7	10	113	114	115	124	125	126	127	128
	SANC00467	6	7		113		115	124	125	126	127	128
	SANC00468	6	7		113		115		125	126	127	
	SANC00469	6	7		113		115	124	125	126	127	128
	SANC00630	6	7		113	114	115	124	125	126	127	128
Degree Centrality (*DC*)	**Reference**	**36**		**91**			**146**		**150**		**206**	
	SANC00302			91					150		206	
	SANC00303			91					150		206	
	SANC00467	36		91		115					206	
	SANC00468	36				115			150			
	SANC00469					115			150		206	
	SANC00630			91					150		206	
Eigencentrality (*EC*)	**Reference**	**7**					**115**		**125**			
	SANC00302						115		125			
	SANC00303				10		115		125			
	SANC00467						115					
	SANC00468											
	SANC00469				10		115		125			
	SANC00630											
	Reference		**36**				**125**		**146**			
	SANC00302								146			
	SANC00303					115						

Katz Centrality (*KC*)	SANC00467			115			
	SANC00468			115			
	SANC00469	10		115	125		150
	SANC00630			115			

In general, by tracking the conservation of the *persistent hubs* as defined PART II, we observed that the presence of the mutations affected the highly conserved communication hubs. This was evident by some of them being completely absent, i.e. hub 146 (*DC*), hub 7 (*EC*), hub 36 (*KC*). Some of the *persistent hubs* were lost in the presence of some ligand systems. We also observed newly introduced *persistent hubs* in some of the metrics, i.e. hub 10 (*KC*; SANC00499 and *EC*; SANC00303 and 469). The *super-persistent hubs* were only observed in the *CC metric* and this is probably because *CC* identifies short communication networks (the central nodes which are closer to most of the nodes).

3.8. Mutation cold spots via analysis of five DRN metrics

There are only limited studies about identification of mutation cold spots with varying definitions of what it means [62,81–83]. The techniques that have been used include *in silico* saturation mutagenesis, meaning mutating every residue to all the other 19 residues and predicting the change in stability [81]; or simply identifying regions where the mutations have not yet occurred in an organism [62]. Here we propose to use DRN metric analysis and define the cold spots as the regions that are the least affected or not affected at all, by mutations. In the previous sections, we introduced *persistent hubs* and *super-persistent hubs*, and we will consider the cold spots as being those hubs that are *super-persistent*, or almost so. The *super-persistent hubs* of *CC* metric are all located mainly in the interface of the dimer as well as in the first two antiparallel

beta strands. We believe, these regions should be strongly considered in structure based drug discovery.

3.9. Identification of ligand specific allosteric communication paths and changes in the presence of mutations

In this section we zoomed into the global top 5% averaged metric calculations for reference and 50 mutant protein systems in the presence of allosteric modulators (**Figure S2-S6**). We picked up two ligands as specific examples: SANC00302 being the least stable compound and SANC00468 being the most stable within all mutant systems. We specifically focused on the protomer A *EC* results **(Figure 13)** as a follow up on the allosteric communication path defined in Section 3.5.4, in which the path from allosteric site to the active site was defined via averaged *EC persistent hubs* (ALA7, LEU115 and VAL125) and hub *score* changes from 0 to 1 (MET17, ASN28 and GLY29) in the presence of ligand binding.

Protomer A of M^{pro}-SANC00302 reference protein – ligand complex has 18 centrality hubs for *EC* (residues 7, 10, 17, 28, 29, 38, 113, 115, 116, 117, 122, 124, 125, 146, 147, 148, 149, 150), including the path residues identified in Section 3.5.4 (**Figure 13**). When we collectively mapped these centrality hubs to the protein-ligand system, we had another very interesting observation: These centrality hubs form a communication path between the allosteric ligand binding site to the active site going through the interface residues of Domain I and II (**Figure 14A**). In the case of Protomer A of M^{pro}- SANC00468 reference protein – ligand complex, some new centrality hubs are gained (9, 11, 13, 14), and some lost (38, 149) compared to that of M^{pro}-SANC00302 system; totaling to 20 *EC* hub residues (7, 9, 10, 11, 13, 14, 17, 28, 29, 113, 115, 116, 117, 122, 124, 125, 146, 147, 148, 150) (**Figure 13**).

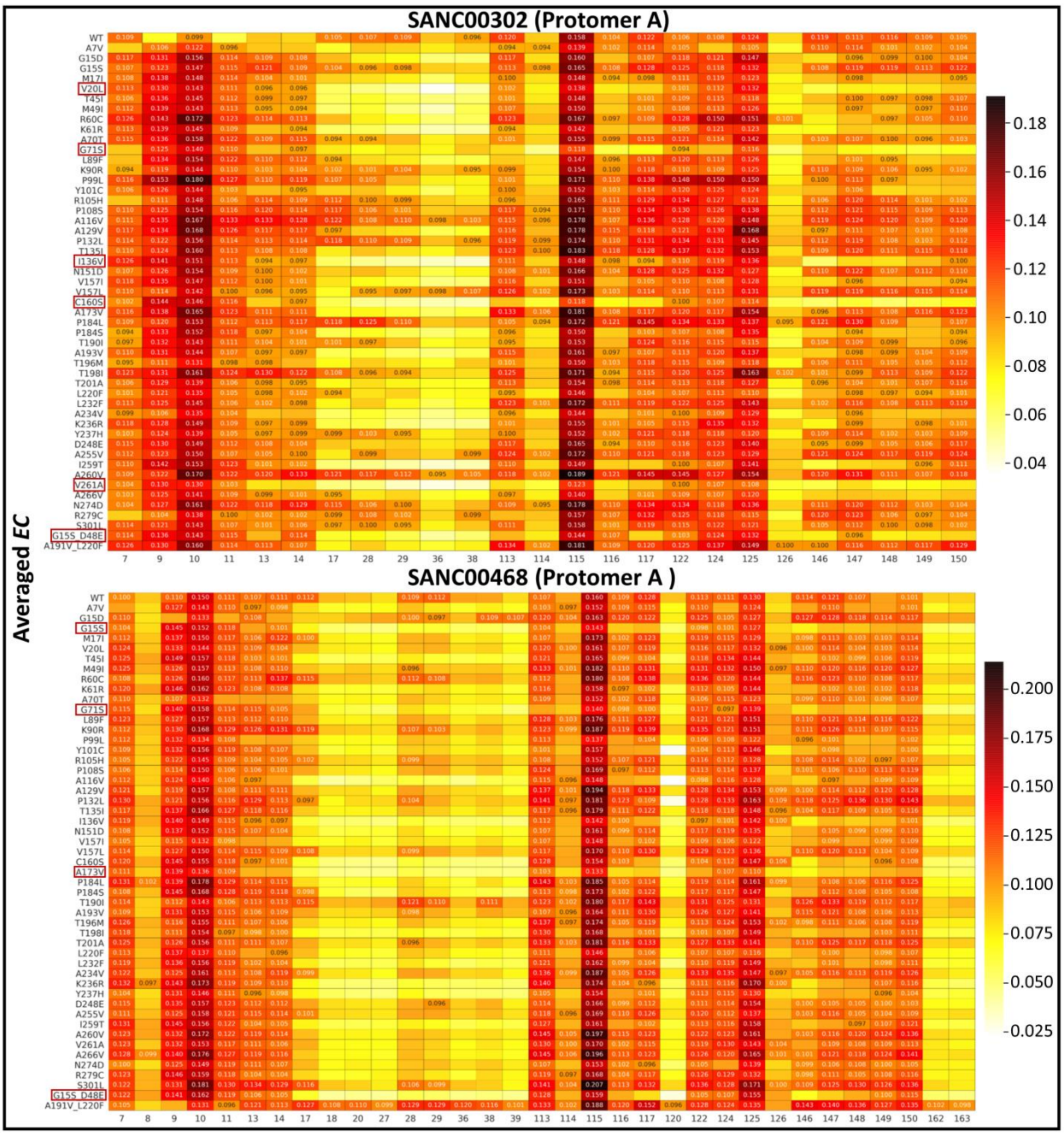

Figure 13. Heat map for the potential hubs according to the global top 5% for averaged *EC* metric for the reference and 50 mutant proteins in allosterically bound state to SANC00302 and SANC00468. Detected hubs are annotated with their centrality values, while their homologous residues in alternate samples are not, but are only shown for the sake of comparison. Low to high centrality values are colored white, through yellow, orange and red to black. Mutants demonstrating highly different centrality hub profiles are marked with red boxes.

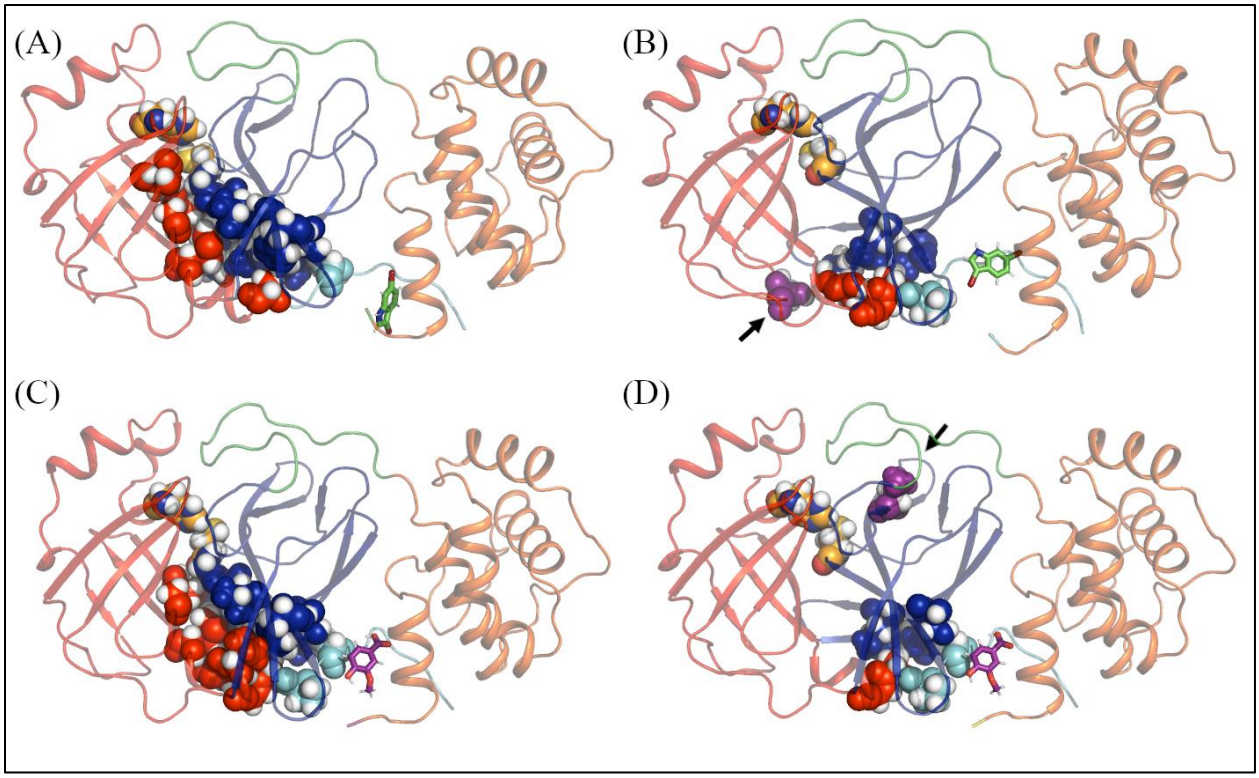

Figure 14. The communication path traced by averaged *EC* hubs, starting from the allosteric ligand towards the catalytic residue. The protease is depicted by a cartoon representation onto which the averaged *EC* hub residues are overlaid as sphere representations, together with the non-hub catalytic residues HIS41 and CYS145 (in orange). (A) M^{pro}-SANC00302 reference protein – ligand complex. Allosteric modulator is in green. (B) Mutant M^{pro} (G71S)-SANC00302 complex. Mutant (in purple) indicated by arrow. (C) M^{pro}- SANC00468 reference protein – ligand complex. Allosteric modulator is in purple. (D) Mutant M^{pro} (A173V)-SANC00468 complex. Mutant (in purple) indicated by arrow. Ligand positions are as observed after MD simulation.

Next, we looked at how these averaged *EC* hubs of protomer A change in the presence of mutations. In general, we observed that, there are more mutant cases where a large number of the centrality hubs is lost in the presence of SANC00302 than of SANC00468. Some examples of these cases from SANC00302 are V20L, G71S, I136V, C160S, V261A mutant proteins. We

further observed that the decreased number of *EC* hubs leads to weakened communication paths that are either weakened or totally lost. G71S-SANC00302 mutant system, for instance, could only maintain 7 *EC* hubs out of 18 (residues 9, 10, 11, 14, 115, 122, 125) **(Figure 14B)**. Extreme examples to the loss of the communication path in the presence of SANC00468 include G15S, G71S and A173V mutants. A173V-SANC00468 complex with 8 *EC* hubs (7, 9, 10, 11, 113, 115, 124, 125) is presented in **Figure 14D**.

4. CONCLUSION

In this study we have provided important novel insights towards computational drug discovery approaches and applied them to SARS-CoV-2 M^{pro} protein. Here, we will list these novel aspects and link to our findings for M^{pro} protein.

We previously proposed a post-hoc analysis approach of MD simulations using DRN analysis to consider the dynamic nature of functional proteins and protein-drug complexes and to probe the impact of mutations and their allosteric effects. We also established a tool for DRN [49]. We and others, in a number of publications, showed the effectiveness of our DRN approach [12,25,90–93,51,75,84–89]. In this study, for the first time, we investigated the relationships and effectiveness of five DRN metrics (*BC*, *CC*, *DC*, *EC* and *KC*) in characterizing key communication residues of reference M^{pro} protein and its allosteric behavior in the presence of potential allosteric modulators and evolutionary mutations. Further, we introduced the concept of analyzing globally central nodes (i.e. the 5% most central nodes measured across all samples) and developed an algorithm to pinpoint key hub residues, meaning any node that forms part of the set of highest centrality nodes for any given averaged centrality metric.

We investigated the hub transition when exposed to a particular environment (i.e. ligand binding) by considering these strongest actors (hubs) across samples and showed how other non-

hub residues behave at the homologous position. The key reason for using DRN analysis in M^{pro} protein was to tackle the problem of protein symmetry that we identified in our previous study [12], where we observed that protomer dynamics could be switched between identical copies of a protomer in a homodimer. In this study, we investigated the phenomenon in greater detail using a combinatorial approach to examine patterns of change and conservation of critical nodes, according to five independent criteria of network centrality. Asymmetric behavior of multimeric proteins, in general, is not considered in computational analysis. To our knowledge, this is the first study tacking this problem via the use of five DRN metrics, and emphasizing the importance of this aspect, especially while analyzing the allosteric behavior of a protein in the presence of ligands and mutations.

Applications of our approaches pinpointed a number of important aspects in SARS-CoV-2 M^{pro} protein: (1) we identified hubs that stayed the same in apo and upon a ligand binding (*constitutive hubs*) indicating that there is no ligand effect from symmetry; (2) we captured different *persistent hubs* from each metric, and collectively they gave us highly crucial functional residues which were spreading out from the allosteric site to the interface and antiparallel beta strands. We believe that the antiparallel beta strands, especially the first two near to the dimer interface, are highly crucial in the mechanical signal transduction; (3) we also looked at the symmetry problem and analyzed hub losses and gains in the presence of allosteric modulators. The identified residues informed us about specific communication changes due to the presence of ligand and allosteric communication residues. A few examples of hub gains and losses that we observed in functional residues are VAL13 (next to the N-finger), GLY 138 (part of S1 subsite) and PHE140 (chameleon switch). We also observed a number of hub transitions in antiparallel beta strands; (4) very interestingly, we showed that *EC* centrality hubs form ligand specific

communication paths between the allosteric ligand binding site to the active site going through the interface residues of Domain I and II.

In general, structure based drug discovery approaches have been used successfully for the design of many orthosteric drugs and to some extent of allosteric modulators. However, consideration of the impact of evolutionary mutations of pathogens is mostly undetermined in rational drug design; even though the information obtained may help to develop drugs that could circumvent or reduce potential drug resistance issues. Here, we applied this mostly neglected concept of computational drug discovery to identify potential allosteric modulators in the presence of 50 early evolutionary mutations of the SARS-CoV-2. We gained a number of observations: (1) stability of the ligands drastically changed in the presence of some of the mutations. R60C, N151D, V157I, C160S and A255V mutant proteins could only hold two compounds out of six stably. SANC00302 was the least stable compound (in 20 mutant systems) and SANC00468 was the most stable (in 43 mutant systems); (2) the *persistent hubs*, residue 7 (*EC*), 36 (*KC*) and 146 (*DC*), lost their importance in the network communications in the presence of mutations; (3) in the presence of mutants some new *persistent hubs* (residue 10 (*EC*), 115 (*DC* and *KC*), 150 (*KC*)) were gained; (4) Further, we defined *super-persistent hubs*, and we considered cold spots as being those hubs that are *super-persistent*, or almost so. These regions could be considered in structure based drug discovery; (5) in the presence of some of the mutations, the network communication within each protomer drastically differed from each other, emphasizing the asymmetric behavior of the dimer protein; (6) most importantly, the allosteric communication path, that was identified via *EC* centrality hubs, between the allosteric ligand binding site and the active site was lost in some of the mutant protein – ligand systems.

Collectively, our approaches offer routes for novel rational drug discovery methods and provide computationally feasible platforms (1) to determine globally central nodes that form part of the set of highest centrality nodes (hubs) for any given averaged centrality metric; (2) to identify key functional residues implicated in allosteric signaling in the presence of allosteric modulators; (3) to understand the potential asymmetric behavior of dimeric proteins under internal and external forces and to distinguish those introduced by ligand binding or by evolutionary mutations; (4) to utilize five DRN metrics to pinpoint cold spot residues that can potentially be chosen for structure guided drug discovery.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at

Table S1: List of mutations and sample IDs extracted from GISAID database.

Table S2: Molecular interactions established between hits and their respective allosteric site residues in SARS-CoV-2 M^{pro}. In black and red labels are residues from protomer A and B respectively.

Table S3: The distribution of ligand stability across mutant samples. The symbol '✓' indicates highly stable motion. The consensus score across each mutant system is the number of '✓' entries in that row.

Figure S1: (A) Kernel density plot of ligand RMSD of the whole 100 ns simulation in SARS-CoV-2 and HCoV-OC43 M^{pro}. (B) 2D structure of identified allosteric modulators.

Figure S2A-S2F: Heatmaps of the potential hubs as per to the global top 5% of the *BC* DRN metric for the reference (WT) and mutant systems in complex with allosteric

modulator: SANC00302, SANC00302, SANC00467, SANC00468 and SANC00630. *BC* measurements of ligand-bound protomer (chain A) are shown on the left while those of the unbound protomer B are on the right. Residue hubs are annotated with their centrality values, while their homologous residues in alternate samples are not. The colour scale from white, through yellow, orange and red to black indicates the degree of residue centrality.

Figure S3A-S3F: Heatmaps of the potential hubs as per to the global top 5% of the *CC* DRN metric for the reference (WT) and mutant systems in complex with allosteric modulator: SANC00302, SANC00302, SANC00467, SANC00468 and SANC00630. *CC* measurements of ligand-bound protomer (chain A) are shown on the left while those of the unbound protomer B are on the right. Residue hubs are annotated with their centrality values, while their homologous residues in alternate samples are not. The colour scale from white, through yellow, orange and red to black indicates the degree of residue centrality.

Figure S4A-S4F: Heatmaps of the potential hubs as per to the global top 5% of the *DC* DRN metric for the reference (WT) and mutant systems in complex with allosteric modulator: SANC00302, SANC00302, SANC00467, SANC00468 and SANC00630. *DC* measurements of ligand-bound protomer (chain A) are shown on the left while those of the unbound protomer B are on the right. Residue hubs are annotated with their centrality values, while their homologous residues in alternate samples are not. The colour scale from white, through yellow, orange and red to black indicates the degree of residue centrality.

Figure S5A-S5F: Heatmaps of the potential hubs as per to the global top 5% of the *EC* DRN metric for the reference (WT) and mutant systems in complex with allosteric modulator: SANC00302, SANC00302, SANC00467, SANC00468 and SANC00630. *EC* measurements of ligand-bound protomer (chain A) are shown on the left while those of the unbound protomer B are on the right. Residue hubs are annotated with their centrality values, while their homologous residues in alternate samples are not. The colour scale from white, through yellow, orange and red to black indicates the degree of residue centrality.

Figure S6A-S6F: Heatmaps of the potential hubs as per to the global top 5% of the *KC* DRN metric for the reference (WT) and mutant systems in complex with allosteric modulator: SANC00302, SANC00302, SANC00467, SANC00468 and SANC00630. *KC* measurements of ligand-bound protomer (chain A) are shown on the left while those of the unbound protomer B are on the right. Residue hubs are annotated with their centrality values, while their homologous residues in alternate samples are not. The colour scale from white, through yellow, orange and red to black indicates the degree of residue centrality.

Author Contributions

The project was conceived by ÖTB. The algorithms were developed by OSA. The calculations were performed by OSA, RAB, VB and DN. The data analysis was performed by OSA, RAB and ÖTB. The initial draft of the manuscript was written by OSA, RAB and ÖTB. All authors have given approval to the final version of the manuscript.

Funding Sources

O.S.A. is funded as a postdoctoral fellow by H3ABioNet, which is supported by the National Human Genome Research Institute of the National Institutes of Health under Award Number U24HG006941. R.A.B. is funded by DELTAS Africa Initiative under Wellcome Trust (DELGEME grant number 107740/Z/15/Z) for a Ph.D. fellowship. V.B is funded by the Grand Challenges Africa programme [GCA/DD/rnd3/03] African Academy of Sciences (AAS). This work is further supported by Funding for COVID-19 Research and Development Goals for Africa Programme (Grant number: SARSCov2-2-20-002) African Academy of Sciences (AAS). Both programmes of the AAS are implemented through the Alliance for Accelerating Excellence in Science in Africa (AESA) platform, an initiative of the AAS and the African Union Development Agency (AUDA-NEPAD). Grand Challenges Africa is supported by the Bill & Melinda Gates Foundation (BMGF), Swedish International Development Cooperation Agency (SIDA), German Federal Ministry of Education and Research (BMBF), Medicines for Malaria Venture (MMV), and Drug Discovery and Development Centre of University of Cape Town (H3D). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

Notes

The authors declare no competing financial interest.

DATA AND SOFTWARE AVAILABILITY

All data reported in this article are presented in the article and the Supporting Information section. Dynamic residue network analysis metric scripts are implemented in the MDM-TASK-

web platform (https://mdmtaskweb.rubi.ru.ac.za/) and are available at https://github.com/RUBi-ZA/MD-TASK/tree/mdm-task-web. MD simulations will be made available upon request.

ABBREVIATIONS

ACPYPE, AnteChamber PYthon Parser interface; *BC*, *betweenness centrality*; *CC*, *closeness centrality*; *DC*, *degree centrality*; DRN, dynamic residue network; *EC*, *eigencentrality*; GISAID, Global initiative on sharing all influenza data; HCoV-OC43, human coronavirus OC43; *KC, katz centrality*; M^{pro}, main protease; MD, molecular dynamics; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SANCDB, South African Natural Compound Database; WT, wildtype.

ACKNOWLEDGMENT

Authors acknowledge the use of the Centre for High-Performance Computing (CHPC), Cape Town, South Africa for the simulations.

REFERENCES

(1) Xiu, S.; Dick, A.; Ju, H.; Mirzaie, S.; Abdi, F.; Cocklin, S.; Zhan, P.; Liu, X. Inhibitors of SARS-CoV-2 Entry: Current and Future Opportunities. *J. Med. Chem.* **2020**, *63*, 2256–12274, DOI: 10.1021/acs.jmedchem.0c00502.

(2) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of Mpro from SARS-CoV-2 and Discovery of Its Inhibitors. *Nature* **2020**, *582*, 289–293, DOI: 10.1038/s41586-020-2223-y.

(3) Holshue, M. L.; DeBolt, C.; Lindquist, S.; Lofy, K. H.; Wiesman, J.; Bruce, H.; Spitters, C.; Ericson, K.; Wilkerson, S.; Tural, A.; Diaz, G.; Cohn, A.; Fox, L.; Patel, A.; Gerber, S. I.; Kim, L.; Tong, S.; Lu, X.; Lindstrom, S.; Pallansch, M. A.; Weldon, W. C.; Biggs, H. M.; Uyeki, T. M.; Pillai, S. K. First Case of 2019 Novel Coronavirus in the United States. *N. Engl. J. Med.* **2020**, *382*, 929–936, DOI: 10.1056/NEJMoa2001191.

(4) Zeng, L.; Li, D.; Tong, W.; Shi, T.; Ning, B. Biochemical Features and Mutations of Key Proteins in SARS-CoV-2 and Their Impacts on RNA Therapeutics. *Biochem. Pharmacol*. **2021**, *189*, 114424, DOI: 10.1016/j.bcp.2021.114424.

(5) Liu, C.; Zhou, Q.; Li, Y.; Garner, L. V.; Watkins, S. P.; Carter, L. J.; Smoot, J.; Gregg, A. C.; Daniels, A. D.; Jervey, S.; Albaiu, D. Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. *ACS Cent. Sci.* **2020**, *6*, 315–331, DOI: 10.1021/acscentsci.0c00272.

(6) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved α-Ketoamide Inhibitors. *Science* **2020**, *368*, eabb3405, DOI: 10.1126/science.abb3405.

(7) Sisay, M. 3CLpro Inhibitors as a Potential Therapeutic Option for COVID-19: Available Evidence and Ongoing Clinical Trials. *Pharmacol. Res*. **2020**, *156*, 104779, DOI: 10.1016/j.phrs.2020.104779.

(8) Wang, F.; Chen, C.; Tan, W.; Yang, K.; Yang, H. Structure of Main Protease from Human Coronavirus NL63: Insights for Wide Spectrum Anti-Coronavirus Drug Design. *Sci. Rep.*

2016, *6*, 22677, DOI: 10.1038/srep22677.

(9) Majumder, R.; Mandal, M. Screening of Plant-Based Natural Compounds as a Potential COVID-19 Main Protease Inhibitor: An in Silico Docking and Molecular Dynamics Simulation Approach. *J. Biomol. Struct. Dyn.* **2020**, 1–16, DOI: 10.1080/07391102.2020.1817787.

(10) Dai, W.; Zhang, B.; Jiang, X.-M.; Su, H.; Li, J.; Zhao, Y.; Xie, X.; Jin, Z.; Peng, J.; Liu, F.; Li, C.; Li, Y.; Bai, F.; Wang, H.; Cheng, X.; Cen, X.; Hu, S.; Yang, X.; Wang, J.; Liu, X.; Xiao, G.; Jiang, H.; Rao, Z.; Zhang, L.-K.; Xu, Y.; Yang, H.; Liu, H. Structure-Based Design of Antiviral Drug Candidates Targeting the SARS-CoV-2 Main Protease. *Science* **2020**, *368*, 1331–1335, DOI: 10.1126/science.abb4489.

(11) Wang, J. Fast Identification of Possible Drug Treatment of Coronavirus Disease-19 (COVID-19) through Computational Drug Repurposing Study. *J. Chem. Inf. Model.* **2020**, *60*, 3277–3286, DOI: 10.1021/acs.jcim.0c00179.

(12) Sheik Amamuddy, O.; Verkhivker, G. M.; Tastan Bishop, Ö. Impact of Early Pandemic Stage Mutations on Molecular Dynamics of SARS-CoV-2 M Pro. *J. Chem. Inf. Model.* **2020**, *60*, 5080–5102, DOI: 10.1021/acs.jcim.0c00634.

(13) Hatherley, R.; Brown, D. K.; Musyoka, T. M.; Penkler, D. L.; Faya, N.; Lobb, K. A.; Tastan Bishop, Ö. SANCDB: A South African Natural Compound Database. *J. Cheminform.* **2015**, *7*, 29, DOI: 10.1186/s13321-015-0080-8.

(14) Diallo, B. N.; Glenister, M.; Musyoka, T. M.; Lobb, K.; Tastan Bishop, Ö. SANCDB: An Update on South African Natural Compounds and Their Readily Available Analogs. *J.*

Cheminform. **2021**, *13*, 37, DOI: 10.1186/s13321-021-00514-2.

(15) St-Jean, J. R.; Jacomy, H.; Desforges, M.; Vabret, A.; Freymuth, F.; Talbot, P. J. Human Respiratory Coronavirus OC43: Genetic Stability and Neuroinvasion. *J. Virol.* **2004**, *78,* 8824-34, DOI: 10.1128/jvi.78.16.8824-8834.2004.

(16) Chea, E.; Livesay, D. R. How Accurate and Statistically Robust Are Catalytic Site Predictions Based on Closeness Centrality? *BMC Bioinform.* **2007**, *8*, 153, DOI: 10.1186/1471-2105-8-153.

(17) del Sol, A.; Fujihashi, H.; Amoros, D.; Nussinov, R. Residue Centrality, Functionally Important Residues, and Active Site Shape: Analysis of Enzyme and Non-Enzyme Families. *Protein Sci.* **2006**, *15*, 2120–2128, DOI: 10.1110/ps.062249106.

(18) Negre, C. F. A.; Morzan, U. N.; Hendrickson, H. P.; Pal, R.; Lisi, G. P.; Loria, J. P.; Rivalta, I.; Ho, J.; Batista, V. S. Eigenvector Centrality for Characterization of Protein Allosteric Pathways. *Proc. Natl. Acad. Sci.* **2018**, *115*, E12201–E12208, DOI: 10.1073/pnas.1810452115.

(19) Westbrook, J. D.; Burley, S. K. How Structural Biologists and the Protein Data Bank Contributed to Recent FDA New Drug Approvals. *Structure* **2019**, *27*, 211-217, DOI: 10.1016/j.str.2018.11.007.

(20) Sheik Amamuddy, O.; Veldman, W.; Manyumwa, C.; Khairallah, A.; Agajanian, S.; Oluyemi, O.; Verkhivker, G. M.; Tastan Bishop, Ö. Integrated Computational Approaches and Tools for Allosteric Drug Discovery. *Int. J. Mol. Sci.* **2020**, *21*, 847, DOI: 10.3390/ijms21030847.

(21) Drag, M.; Salvesen, G. S. Emerging Principles in Protease-Based Drug Discovery. *Nat. Rev. Drug Discov*. **2010**, *9*, 690–701, DOI: 10.1038/nrd3053.

(22) Guarnera, E.; Berezovsky, I. N. Allosteric Drugs and Mutations: Chances, Challenges, and Necessity. *Curr. Opin. Struct. Biol*. **2020**, *62,* 149-157, DOI: 10.1016/j.sbi.2020.01.010.

(23) Sheik Amamuddy, O.; Musyoka, T. M.; Boateng, R. A.; Zabo, S.; Tastan Bishop, Ö. Determining the Unbinding Events and Conserved Motions Associated with the Pyrazinamide Release Due to Resistance Mutations of Mycobacterium Tuberculosis Pyrazinamidase. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 1103–1120, DOI: 10.1016/j.csbj.2020.05.009.

(24) Ricatti, J.; Acquasaliente, L.; Ribaudo, G.; De Filippis, V.; Bellini, M.; Llovera, R. E.; Barollo, S.; Pezzani, R.; Zagotto, G.; Persaud, K. C.; Mucignat-Caretta, C. Effects of Point Mutations in the Binding Pocket of the Mouse Major Urinary Protein MUP20 on Ligand Affinity and Specificity. *Sci. Rep. Sci. Rep.* **2019**, *9*, 300, DOI: 10.1038/s41598-018-36391-3.

(25) Brown, D. K.; Sheik Amamuddy, O.; Tastan Bishop, Ö. Structure-Based Analysis of Single Nucleotide Variants in the Renin-Angiotensinogen Complex. *Glob. Heart* **2017**, *12*, 121–132, DOI: 10.1016/j.gheart.2017.01.006.

(26) Nussinov, R.; Tsai, C. J. Allostery without a Conformational Change? Revisiting the Paradigm. *Curr. Opin. Struct. Biol*. **2015**, *30*, 17-24, DOI: 10.1016/j.sbi.2014.11.005.

(27) Sheik Amamuddy, O. Application of Machine Learning, Molecular Modelling and Structural Data Mining against Antiretroviral Drug Resistance in HIV-1, Ph.D Thesis,

Rhodes University, Makhanda, South Africa, **2019**.

(28) Guo, J.; Zhou, H.-X. Protein Allostery and Conformational Dynamics. *Chem. Rev.* **2016**, *116*, 6503–6515, DOI: 10.1021/acs.chemrev.5b00590.

(29) Sheik Amamuddy, O.; Bishop, N. T.; Tastan Bishop, Ö. Characterizing Early Drug Resistance-Related Events Using Geometric Ensembles from HIV Protease Dynamics. *Sci. Rep.* **2018**, *8*, 1–11, DOI: 10.1038/s41598-018-36041-8.

(30) Burley, S. K.; Berman, H. M.; Kleywegt, G. J.; Markley, J. L.; Nakamura, H.; Velankar, S. Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive; Wlodawer, A., Dauter, Z., Jaskolski, M., Eds.; Springer New York: New York, NY, **2017**; pp 627–641, DOI: 10.1007/978-1-4939-7000-1_26.

(31) Fearon, D., Owen, C.D., Douangamath, A., Lukacik, P., Powell, A.J., Strain-Damerell, C.M., Resnick, E., Krojer, T., Gehrtz, P., Wild, C., Aimon, A., Brandao-Neto, J., Carbery, A., Dunnett, L., Skyner, R., Snee, M., London, N., Walsh, M.A., von Delft, F. PanDDA Analysis Group Deposition SARS-CoV-2 Main Protease Fragment Screen. *Nat. Commun.* **2020**, *11,* 5047, https://www.ebi.ac.uk/pdbe/entry/pdb/5rfv.

(32) Shu, Y.; McCauley, J. GISAID: Global Initiative on Sharing All Influenza Data – from Vision to Reality. *Euro surveill.* **2017**, *22*, 30494, DOI: 10.2807/1560-7917.ES.2017.22.13.30494.

(33) Fiser, A.; Šali, A. Modeller: Generation and Refinement of Homology-Based Protein Structure Models. In *Methods in Enzymology*; Department of Biochemistry and Seaver Foundation Center for Bioinformatics, Albert Einstein College of Medicine, Bronz, New

York 10461, USA., 2003; Vol. 374, pp 461–491, DOI: 10.1016/S0076-6879(03)74020-8.

(34) Pei, J.; Kim, B. H.; Grishin, N. V. PROMALS3D: A Tool for Multiple Protein Sequence and Structure Alignments. *Nucleic Acids Res.* **2008**, *36*, 2295–2300, DOI: 10.1093/nar/gkn072.

(35) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* **2004**, *32*, W665-7, DOI: 10.1093/nar/gkh381.

(36) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Commun.* **1995**, *1*, 1-41, DOI: 10.1016/0010-4655(95)00041-D.

(37) Accelrys Software Inc. Discovery Studio Modeling Environment, **2012**, Release 3.5, San Diego, CA.

(38) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity-a Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219-3228, DOI: 10.1016/0040-4020(80)80168-2.

(39) Trott, O.; Olson, A. AutoDock Vina: Inproving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization and Multithreading. *J. Comput. Chem.* **2010**, *31* (2), *J. Comput. Chem.* **2010**, *31*, 455–461, DOI: 10.1002/jcc.21334.AutoDock.

(40) Schrödinger, LLC: New York 2015. *The PyMOL Molecular Graphics System, Version 2.4*, https://pymol.org/2/support.html?.

(41) McKinney, W. Pandas: Powerful Python Data Analysis Toolkit — Pandas 0.19.0+128.G43c24e6.Dirty Documentation. 2016.

(42) McKinney, W. Pandas: Powerful Python Data Analysis Toolkit, **2016**. https://pandas.pydata.org/pandas-docs/version/0.22/pandas.pdf. (accessed *2021-05-21)*.

(43) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25, DOI: 10.1016/J.SOFTX.2015.06.001.

(44) Sousa da Silva, A. W.; Vranken, W. F. ACPYPE - AnteChamber PYthon Parser InterfacE. *BMC Res. Notes* **2012**, *5*, 367, DOI: 10.1186/1756-0500-5-367.

(45) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954–9960, DOI: 10.1021/jp003020w.

(46) Lemak, A. S.; Balabaev, N. K. On The Berendsen Thermostat. *Mol. Simul.* **1994**, *13*, 177–187, DOI: 10.1080/08927029408021981.

(47) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190, DOI: 10.1063/1.328693.

(48) Petersen, H. G. Accuracy and Efficiency of the Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 3668, DOI: 10.1063/1.470043.

(49) Amamuddy, O. S.; Glenister, M.; Tshabalala, T.; Bishop, Ö. T. MDM-TASK-Web: MD-TASK and MODE-TASK Web Server for Analyzing Protein Dynamics. *Comput. Struct. Biotechnol. J.* **2021**, DOI: 10.1016/j.csbj.2021.08.043.

(50) Brown, D. K.; Penkler, D. L.; Sheik Amamuddy, O.; Ross, C.; Atilgan, A. R.; Atilgan, C.; Bishop, Ö. T.; Sheik Amamuddy, O.; Ross, C.; Atilgan, A. R.; Atilgan, C.; Tastan Bishop, Ö. MD-TASK: A Software Suite for Analyzing Molecular Dynamics Trajectories. *Bioinformatics* **2017**, *33*, 2768–2771, DOI: 0.1093/bioinformatics/btx349.

(51) Penkler, D. L.; Tastan Bishop, Ö. Modulation of Human Hsp90α Conformational Dynamics by Allosteric Ligand Interaction at the C-Terminal Domain. *Sci. Rep.* **2019** , *9*, 1600, DOI: 10.1038/s41598-018-35835-0.

(52) Hagberg, A.; Swart, P.; S Chult, D. Exploring Network Structure, Dynamics, and Function Using NetworkX. In *7th Python in Science Conference (SciPy 2008)*; Varoquaux, E., Vaught, T., Millman, J., Eds.; Los Alamos National Lab. (LANL), Los Alamos, NM (United States): Pasadena, CA USA, 2008; pp 1–15.

(53) Anand, K.; Palm, G. J.; Mesters, J. R.; Siddell, S. G.; Ziebuhr, J.; Hilgenfeld, R. Structure of Coronavirus Main Proteinase Reveals Combination of a Chymotrypsin Fold with an Extra α-Helical Domain. *EMBO J.* **2002**, *21*, 3213–3224, DOI: 10.1093/emboj/cdf327.

(54) Chang, G. G. Quaternary Structure of the SARS Coronavirus Main Protease. In *Molecular Biology of the SARS-Coronavirus* **2010,** 115–128, DOI: 10.1007/978-3-642-03683-5_8.

(55) Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods. *ACS Catal.* **2020**, *10*, 12544–12554,

DOI: 10.1021/acscatal.0c03420.

(56) Chen, S.; Hu, T.; Zhang, J.; Chen, J.; Chen, K.; Ding, J.; Jiang, H.; Shen, X. Mutation of Gly-11 on the Dimer Interface Results in the Complete Crystallographic Dimer Dissociation of Severe Acute Respiratory Syndrome Coronavirus 3C-like Protease: Crystal Structure with Molecular Dynamics Simulations. *J. Biol. Chem.* **2008**, *283*, 554-564, DOI: 10.1074/jbc.M705240200.

(57) Shi, J.; Wei, Z.; Song, J. Dissection Study on the Severe Acute Respiratory Syndrome 3C-like Protease Reveals the Critical Role of the Extra Domain in Dimerization of the Enzyme. *J. Biol. Chem.* **2004**, *279*, 24765–24773, DOI: 10.1074/jbc.M311744200.

(58) Shi, J.; Song, J. The Catalysis of the SARS 3C-like Protease Is under Extensive Regulation by Its Extra Domain. *FEBS J.* **2006**, *273*, 1035-45, DOI: 10.1111/j.1742-4658.2006.05130.x.

(59) Zhong, N.; Zhang, S.; Zou, P.; Chen, J.; Kang, X.; Li, Z.; Liang, C.; Jin, C.; Xia, B. Without Its N-Finger, the Main Protease of Severe Acute Respiratory Syndrome Coronavirus Can Form a Novel Dimer through Its C-Terminal Domain. *J. Virol.* **2008**, *82*, 4227–4234, DOI: 10.1128/JVI.02612-07.

(60) Yang, H.; Yang, M.; Ding, Y.; Liu, Y.; Lou, Z.; Zhou, Z.; Sun, L.; Mo, L.; Ye, S.; Pang, H.; Gao, G. F.; Anand, K.; Bartlam, M.; Hilgenfeld, R.; Rao, Z. The Crystal Structures of Severe Acute Respiratory Syndrome Virus Main Protease and Its Complex with an Inhibitor. *Proc. Natl. Acad. Sci.* **2003**, *100*, 13190–13195, DOI: 10.1073/pnas.1835675100.

(61) Yang, H.; Xie, W.; Xue, X.; Yang, K.; Ma, J.; Liang, W.; Zhao, Q.; Zhou, Z.; Pei, D.; Ziebuhr, J.; Hilgenfeld, R.; Yuen, K. Y.; Wong, L.; Gao, G.; Chen, S.; Chen, Z.; Ma, D.; Bartlam, M.; Rao, Z. Design of Wide-Spectrum Inhibitors Targeting Coronavirus Main Proteases. *PLoS Biol.* **2005**, *3*, e324, DOI: 10.1371/journal.pbio.0030324.

(62) Krishnamoorthy, N.; Fakhro, K. Identification of Mutation Resistance Coldspots for Targeting the SARS-CoV2 Main Protease. *IUBMB Life* **2021**, *73*, 670–675, DOI: 10.1002/iub.2465.

(63) Douangamath, A.; Fearon, D.; Gehrtz, P.; Krojer, T.; Lukacik, P.; Owen, C. D.; Resnick, E.; Strain-Damerell, C.; Aimon, A.; Ábrányi-Balogh, P.; Brandão-Neto, J.; Carbery, A.; Davison, G.; Dias, A.; Downes, T. D.; Dunnett, L.; Fairhead, M.; Firth, J. D.; Jones, S. P.; Keeley, A.; Keserü, G. M.; Klein, H. F.; Martin, M. P.; Noble, M. E. M.; O'Brien, P.; Powell, A.; Reddi, R. N.; Skyner, R.; Snee, M.; Waring, M. J.; Wild, C.; London, N.; von Delft, F.; Walsh, M. A. Crystallographic and Electrophilic Fragment Screening of the SARS-CoV-2 Main Protease. *Nat. Commun.* **2020**, *11*, 5047, DOI: 10.1038/s41467-020-18709-w.

(64) Forli, W.; Halliday, S.; Belew, R.; Olson, A. AutoDock Version 4.2. *Citeseer* **2012**, 1–66.

(65) Koorbanally, N. A.; Koorbanally, C.; Harilal, A.; Mulholland, D. A.; Crouch, N. R. Bufadienolides from Drimia Robusta and Urginea Epigea (Hyacinthaceae). *Chem. Inform.* **2005**, *36*, 1, DOI: 10.1002/chin.200516167.

(66) Koorbanally, C.; Mulholland, D. A.; Crouch, N. R. A Novel Homoisoflavonoid from Drimia Delagoensis (Urgineoideae: Hyacinthaceae). *Biochem. Syst. Ecol.* **2005**, *33*, 743-

748, DOI: 10.1016/j.bse.2004.11.009.

(67) Bohlmann, F.; Zdero, C. New Sesquiterpenes from Senecio Oxyodontus. *Phytochemistry* **1978**, *17*, 1591–1593, DOI: 10.1016/S0031-9422(00)94649-1.

(68) Bromley, C. L.; Parker-Nance, S.; De La Mare, J. A.; Edkins, A. L.; Beukes, D. R.; Davies-Colemanf, M. T. Halogenated Oxindole and Indoles from the South African Marine Ascidian Distaplia Skoogi. *South African J. Chem.* **2013**, *66,* 64–68, http://www.scielo.org.za/pdf/sajc/v66/15.pdf.

(69) Kneller, D. W.; Phillips, G.; Weiss, K. L.; Pant, S.; Zhang, Q.; O'Neill, H. M.; Coates, L.; Kovalevsky, A. Unusual Zwitterionic Catalytic Site of SARS-CoV-2 Main Protease Revealed by Neutron Crystallography. *J. Biol. Chem.* **2020**, *295, P17365-17373*, DOI: 10.1074/jbc.AC120.016154.

(70) Shi, J.; Sivaraman, J.; Song, J. Mechanism for Controlling the Dimer-Monomer Switch and Coupling Dimerization to Catalysis of the Severe Acute Respiratory Syndrome Coronavirus 3C-Like Protease. *J. Virol.* **2008,** *82,* 9, DOI: 10.1128/jvi.02680-07.

(71) Krukow, P.; Jonak, K.; Karpiński, R.; Karakuła-Juchnowicz, H. Abnormalities in Hubs Location and Nodes Centrality Predict Cognitive Slowing and Increased Performance Variability in First-Episode Schizophrenia Patients. *Sci. Rep.* **2019**, *9*, 9594, DOI: 10.1038/s41598-019-46111-0.

(72) Chapter 5 - Centrality and Hubs; Fornito, A., Zalesky, A., Bullmore, E. T. B. T.-F. of B. N. A., Eds.; Academic Press: San Diego, 2016; pp 137–161, DOI: 10.1016/B978-0-12-407908-3.00005-4.

(73) Manyumwa, C. V.; Bishop, Ö. T. In Silico Investigation of Potential Applications of Gamma Carbonic Anhydrases as Catalysts of Co2 Biomineralization Processes: A Visit to the Thermophilic Bacteria Persephonella Hydrogeniphila, Persephonella Marina, Thermosulfidibacter Takaii, and Thermus Thermophilus. *Int. J. Mol. Sci.* **2021**, *22*, DOI: 10.3390/ijms22062861.

(74) Amusengeri, A.; Tastan Bishop, Ö. Discorhabdin N, a South African Natural Compound, for Hsp72 and Hsc70 Allosteric Modulation: Combined Study of Molecular Modeling and Dynamic Residue Network Analysis. *Molecules* **2019**, *24*, 188, DOI: 10.3390/molecules24010188.

(75) Allan Sanyanga, T.; Nizami, B.; Bishop, Ö. T. Mechanism of Action of Non-Synonymous Single Nucleotide Variations Associated with α-Carbonic Anhydrase II Deficiency. *Molecules* **2019**, *24*, 3987, DOI: 10.3390/molecules24213987.

(76) Amitai, G.; Shemesh, A.; Sitbon, E.; Shklar, M.; Netanely, D.; Venger, I.; Pietrokovski, S. Network Analysis of Protein Structures Identifies Functional Residues. *J. Mol. Biol.* **2004**, *344*, 1135–1146, DOI: 10.1016/j.jmb.2004.10.055.

(77) Thibert, B.; Bredesen, D. E.; del Rio, G. Improved Prediction of Critical Residues for Protein Function Based on Network and Phylogenetic Analyses. *BMC Bioinform.* **2005**, *6*, 213, DOI: 10.1186/1471-2105-6-213.

(78) Snijder, E. J.; Decroly, E.; Ziebuhr, J. The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. *Adv. Virus Res*. **2016**, *96*, 59-126, DOI: 10.1016/bs.aivir.2016.08.008.

(79) Martin, R. W.; Butts, C. T.; Cross, T. J.; Takahashi, G. R.; Diessner, E. M.; Crosby, M. G.; Farahmand, V.; Zhuang, S. Sequence Characterization and Molecular Modeling of Clinically Relevant Variants of the SARS-CoV-2 Main Protease. *Biochemistry* **2020**, *59*, 3741–3756, DOI: 10.1021/acs.biochem.0c00462.

(80) Tee, W.-V.; Guarnera, E.; Berezovsky, I. N. On the Allosteric Effect of NsSNPs and the Emerging Importance of Allosteric Polymorphism. *J. Mol. Biol.* **2019**, *431*, 3933–3942, DOI: 10.1016/j.jmb.2019.07.012.

(81) Vedithi, S. C.; Rodrigues, C. H. M.; Portelli, S.; Skwark, M. J.; Das, M.; Ascher, D. B.; Blundell, T. L.; Malhotra, S. Computational Saturation Mutagenesis to Predict Structural Consequences of Systematic Mutations in the Beta Subunit of RNA Polymerase in Mycobacterium Leprae. *Comput. Struct. Biotechnol. J.* **2020**, *18,* 271-286, DOI: 10.1016/j.csbj.2020.01.002.

(82) Shirian, J.; Sharabi, O.; Shifman, J. M. Cold Spots in Protein Binding. *Trends Biochem. Sci*. **2016**, *41*, 739-745, DOI: 10.1016/j.tibs.2016.07.002.

(83) Naftaly, S.; Cohen, I.; Shahar, A.; Hockla, A.; Radisky, E. S.; Papo, N. Mapping Protein Selectivity Landscapes Using Multi-Target Selective Screening and next-Generation Sequencing of Combinatorial Libraries. *Nat. Commun.* **2018**, *9*, 3935, DOI: 10.1038/s41467-018-06403-x.

(84) Penkler, D. L.; Atilgan, C.; Tastan Bishop, Ö. Allosteric Modulation of Human Hsp90α Conformational Dynamics. *J. Chem. Inf. Model.* **2018**, *58*, 383–404, DOI: 10.1021/acs.jcim.7b00630.

(85) Amusengeri, A.; Tata, R. B.; Tastan Bishop, Ö. Understanding the Pyrimethamine Drug Resistance Mechanism via Combined Molecular Dynamics and Dynamic Residue Network Analysis. *Molecules* **2020**, *25*, 904, DOI: 10.3390/molecules25040904.

(86) Arifuzzaman, M.; Mitra, S.; Das, R.; Hamza, A.; Absar, N.; Dash, R. In Silico Analysis of Nonsynonymous Single-Nucleotide Polymorphisms (NsSNPs) of the SMPX Gene. *Ann. Hum. Genet.* **2020**, *84*, 54-71, DOI: 10.1111/ahg.12350.

(87) Xiao, F.; Song, X.; Tian, P.; Gan, M.; Verkhivker, G. M.; Hu, G. Comparative Dynamics and Functional Mechanisms of the CYP17A1 Tunnels Regulated by Ligand Binding. *J. Chem. Inf. Model.* **2020**, *60*, 3632–3647, DOI: 10.1021/acs.jcim.0c00447.

(88) Dehury, B.; Tang, N.; Mehra, R.; Blundell, T. L.; Kepp, K. P. Side-by-Side Comparison of Notch- And C83 Binding to γ-Secretase in a Complete Membrane Model at Physiological Temperature. *RSC Adv.* **2020**, *10*, 31215-31232, DOI: 10.1039/d0ra04683c.

(89) Keretsu, S.; Ghosh, S.; Cho, S. J. Molecular Modeling Study of C-Kit/Pdgfrα Dual Inhibitors for the Treatment of Gastrointestinal Stromal Tumors. *Int. J. Mol. Sci.* **2020**, *21*, 8232, DOI: 10.3390/ijms21218232.

(90) Fischer, A.; Häuptli, F.; Lill, M. A.; Smieško, M. Computational Assessment of Combination Therapy of Androgen Receptor-Targeting Compounds. *J. Chem. Inf. Model.* **2021**, *61*, 1001–1009, DOI: 10.1021/acs.jcim.0c01194.

(91) Wang, S.; Xu, Y.; Yu, X. W. A Phenylalanine Dynamic Switch Controls the Interfacial Activation of Rhizopus Chinensis Lipase. *Int. J. Biol. Macromol.* **2021**, *173*, 1-12, DOI: 10.1016/j.ijbiomac.2021.01.086.

(92) Ma, S.; Li, H.; Yang, J.; Yu, K. Molecular Simulation Studies of the Interactions between the Human/Pangolin/Cat/Bat ACE2 and the Receptor Binding Domain of the SARS-CoV-2 Spike Protein. *Biochimie* **2021**, *187*, 1-13, DOI: 10.1016/j.biochi.2021.05.001.

(93) Chebon-Bore, L.; Sanyanga, T. A.; Manyumwa, C. V.; Khairallah, A.; Bishop, Ö. T. Decoding the Molecular Effects of Atovaquone Linked Resistant Mutations on Plasmodium Falciparum Cytb-Isp Complex in the Phospholipid Bilayer Membrane. *Int. J. Mol. Sci.* **2021**, *22*, 2138, DOI: 10.3390/ijms22042138.