Allosteric modulator identification from South African Natural Compounds Database (SANCDB) against SARS-CoV-2 M^{pro} protein in the presence of its evolutionary mutations

Özlem Taştan Bishop

Research Unit in Bioinformatics (RUBi),

Department of Biochemistry and Microbiology,

Rhodes University, Makhanda, South Africa

https://rubi.ru.ac.za/

RHODES UNIVERSITY

CAiSMD – 9-11 March 2022

RUBi

Research Unit in Bioinformatics

South African Natural Compounds Database

https://sancdb.rubi.ru.ac.za/

Hatherley *et al. J Cheminform (2015) 7:29*
DOI 10.1186/s13321-015-0080-8

Journal of Cheminformatics

DATABASE

Open Access

SANCDB: a South African natural compound database

Rowan Hatherley[1], David K Brown[1], Tho
Kevin A Lobb[1,2] and Özlem Tastan Bisho

Diallo *et al. J Cheminform (2021) 13:37*
https://doi.org/10.1186/s13321-021-00514-2

Journal of Cheminformatics

RESEARCH ARTICLE

Open Access

SANCDB: an update on South African natural compounds and their readily available analogs

Bakary N'tji Diallo[1], Michael Glenister[1], Thommas M. Musyoka[1], Kevin Lobb[1,2] and Özlem Tastan Bishop[1*]

AIM:

Identify allosteric inhibitors alternative to active site inhibitors.

How the behavior of allosteric inhibitors changes in the presence of the evolutionary mutations?

Allostery and mutations: Not commonly used biological phenomena in the early computational drug discovery stages.

Outline of study/talk

- **STEP 1:** Identification of M^{pro} mutations and their effects on the structure
- **STEP 2:** Identification of potential allosteric site(s) and effects of mutations on allosteric and active sites
- **STEP 3:** Identification of potential allosteric modulators (SANCDB) against reference structure
- **STEP 4:** Identification of functionally important residues and allosteric communication paths within protein-inhibitor complexes, and changes in the presence of mutations

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

Impact of Early Pandemic Stage Mutations on Molecular Dynamics of SARS-CoV-2 M^{pro}

Olivier Sheik Amamuddy, Gennady M. Verkhivker, and Özlem Tastan Bishop*

Cite This: https://dx.doi.org/10.1021/acs.jcim.0c00634 Read Online

Olivier Sheik Amamuddy

Computational and Structural Biotechnology Journal 19 (2021) 6431–6455

ELSEVIER

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

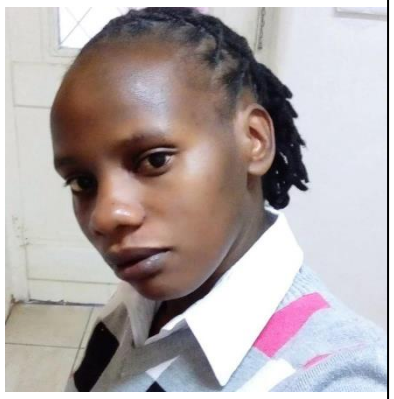
Novel dynamic residue network analysis approaches to study allosteric modulation: SARS-CoV-2 M^{pro} and its evolutionary mutations as a case study

Olivier Sheik Amamuddy [1], Rita Afriyie Boateng [1], Victor Barozi, Dorothy Wavinya Nyamai, Özlem Tastan Bishop *

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

Rita Boateng

Victor Barozi



Dorothy Nyamai

BASIC CONCEPTS

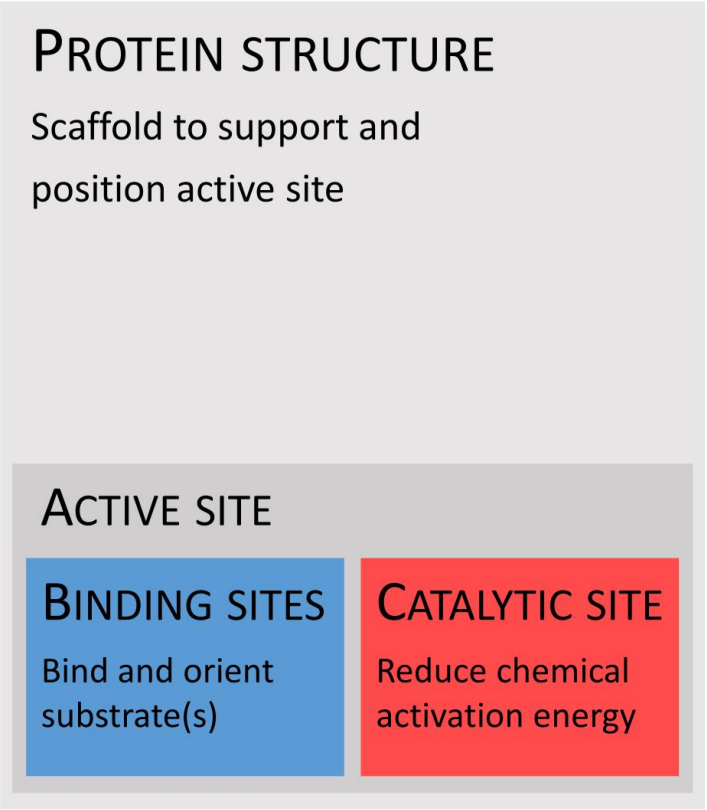
Two types of FDA approved drugs:

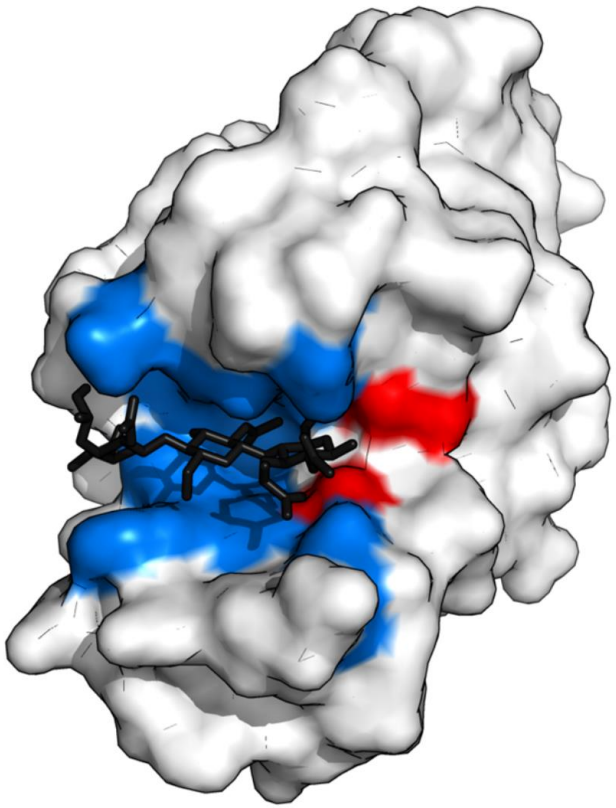
Orthosteric drugs
Allosteric drugs

Definitions – Orthosteric site & drug

Orthosteric site:
The functional site, e.g., active sites for enzymes, or protein-protein binding sites for receptors.

Orthosteric drugs bind to the functional site of a protein, compete with endogenous regulators and block the activity directly.



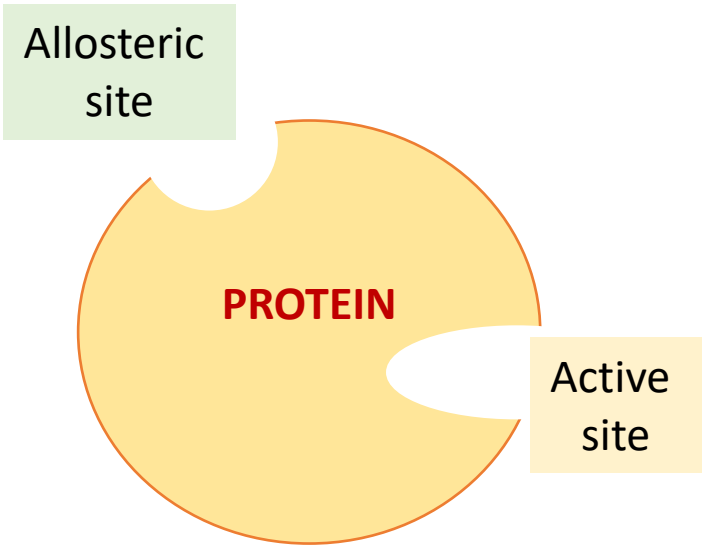


Definitions – Allostery, allosteric site & drug

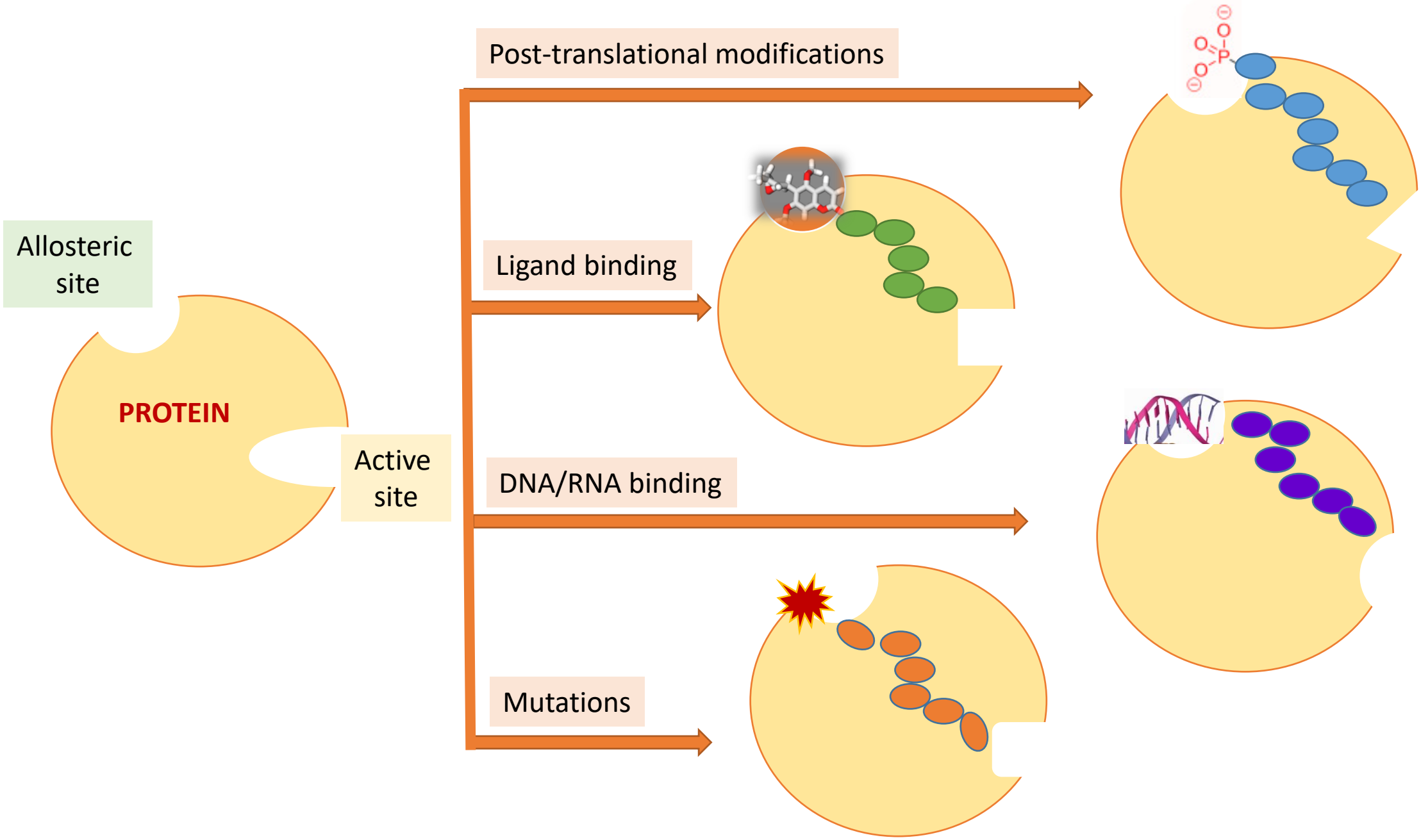
Two Greek words: "**allos**" - other/alternate & "**stereo**" - solid shapes.

Allosteric site: A site away from the orthosteric site but whose perturbation by binding an **effector** affects the conformation/function at the orthosteric site → **allosteric communication**.

Allosteric drugs modify the functional conformation and/or the active site of the target protein from a distance - allosteric site.

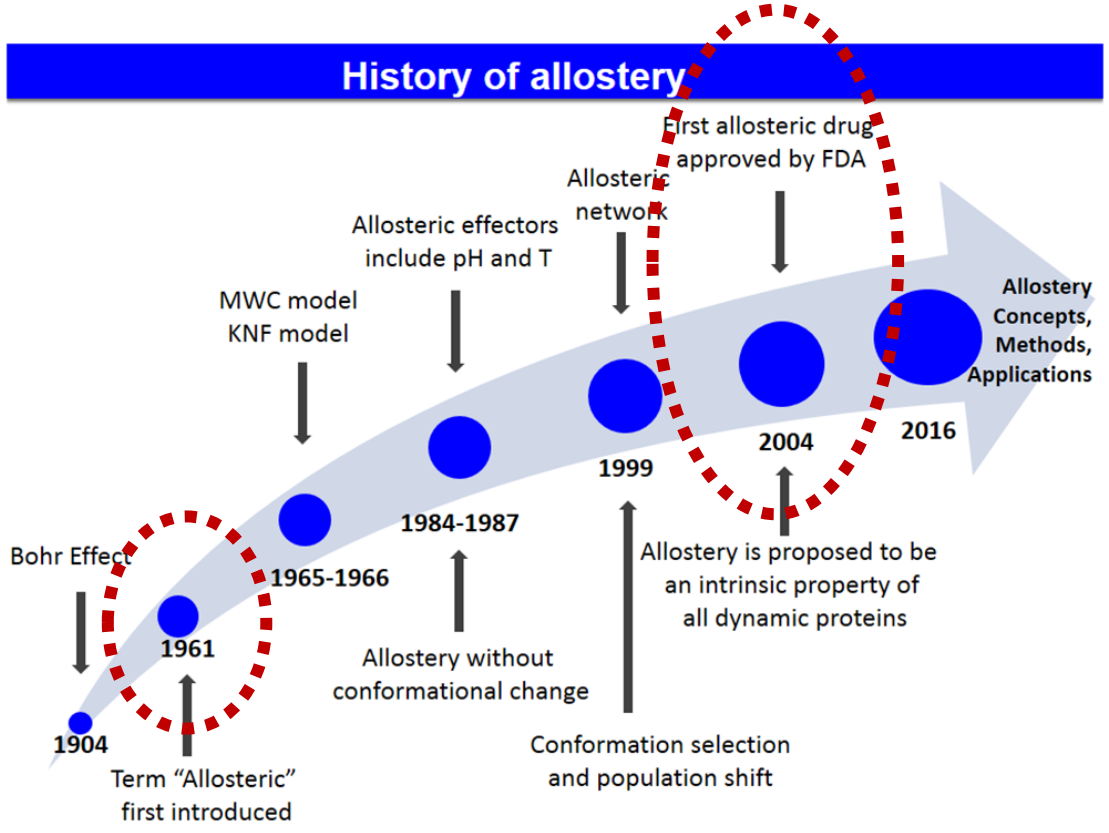


Perturbation is the key to allosteric behavior, and effectors that can cause perturbation are:



FDA approved drugs: Orthosteric versus allosteric

- The first discovery of allosteric systems > 60 years.
- The concept of using allosteric sites as drug targets is still not common.
- ~3700 FDA approved drugs.
- 19 of them allosteric drugs.
- 19 versus ~ 3681 drugs → hints at the difficulty of designing allosteric drugs.



Liu and Nussinov, 2016;
https://doi.org/10.1371/journal.pcbi.1004966

Difficulty is due to:

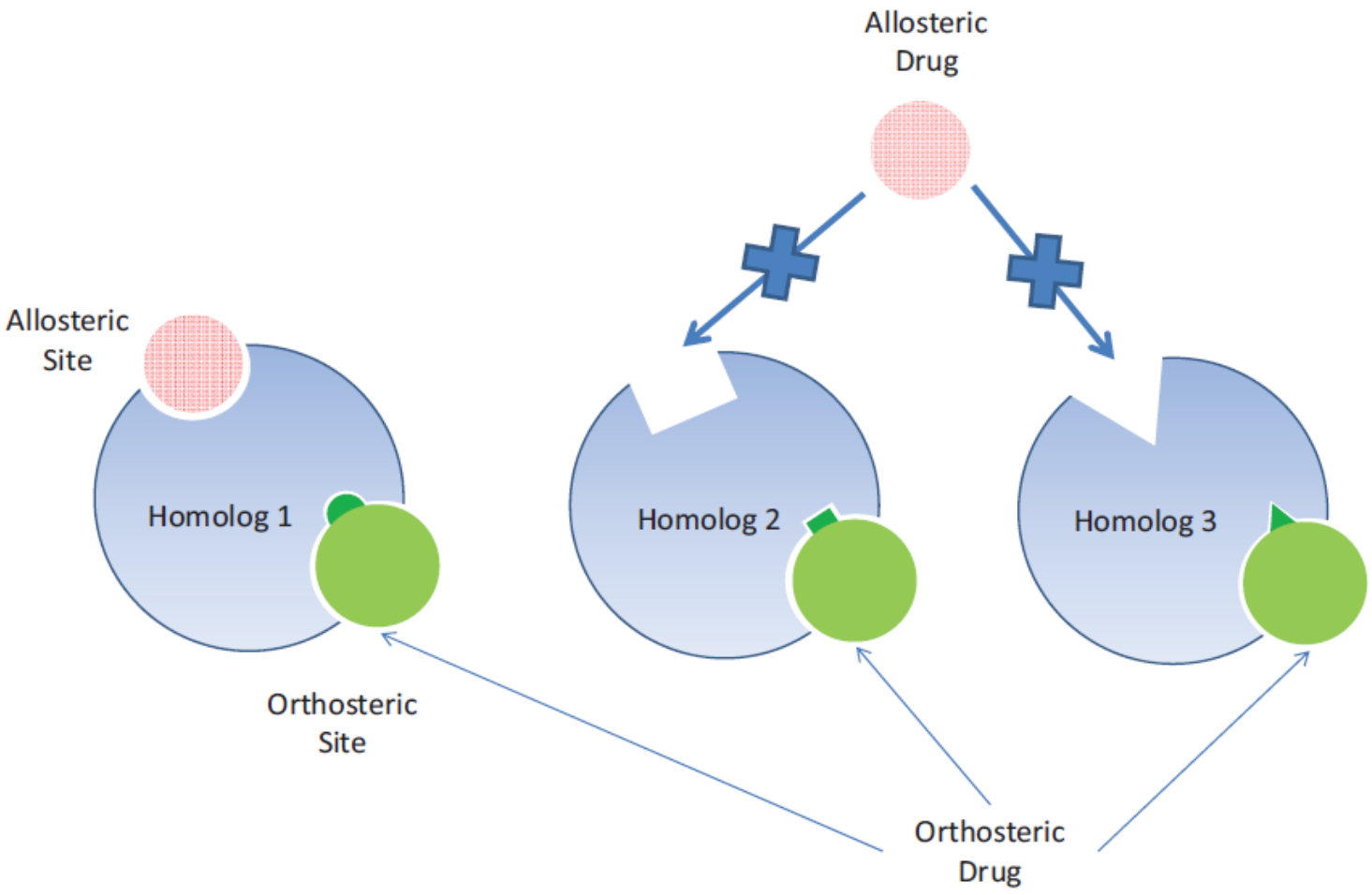
- Identification of allosteric sites (shallow, cryptic pockets …) – serendipity

- Demonstration of the effects of allosteric modulators
  - inhibitor
  - activator

Allosteric drugs have benefits over orthosteric drugs:

- ✓ Allosteric sites are less conserved compared to active sites, therefore allosteric modulators are highly specific, hence may be less toxic to host.

Allosteric drugs versus orthosteric drugs: *Specificity - toxicity*

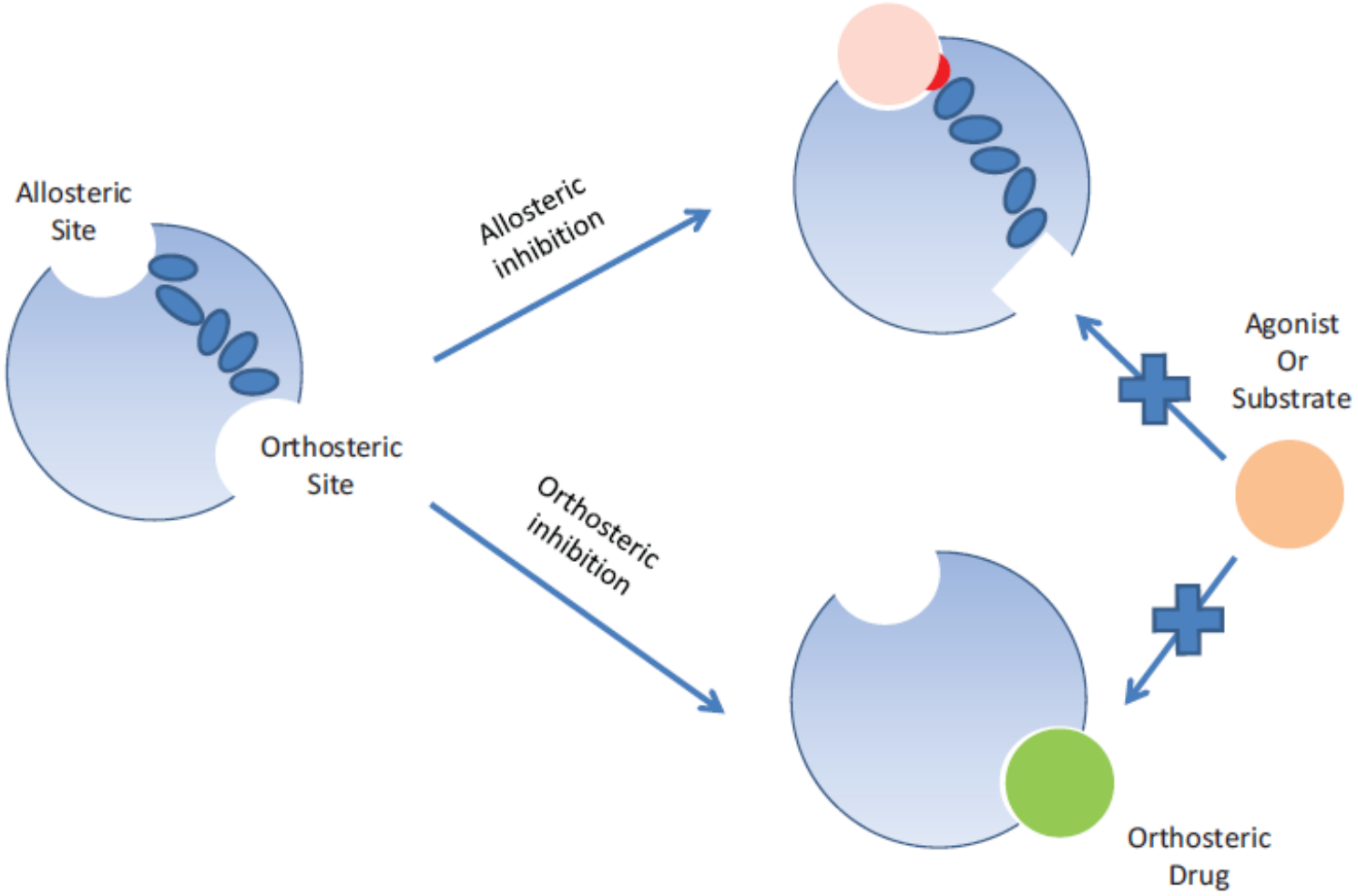


Nussinov and Tsai, Current Pharmaceutical Design, 2012

Allosteric drugs have benefits over orthosteric drugs:

✓Unlike orthosteric drugs that compete with the substrate and cofactors, allosteric drugs can be active even in the presence of the native substrates, and ***thus reduce the chances of pathogen developing resistance by increasing the substrate concentrations and/or mutations.***

Allosteric drugs versus orthosteric drugs: *Resistance*



Nussinov and Tsai, Current Pharmaceutical Design, 2012

Resistance → brings us to other biological phenomenon:

MUTATIONS

NONSYNONYMOUS SUBSTITUTION
a nucleotide mutation that alters the amino acid sequence of a protein

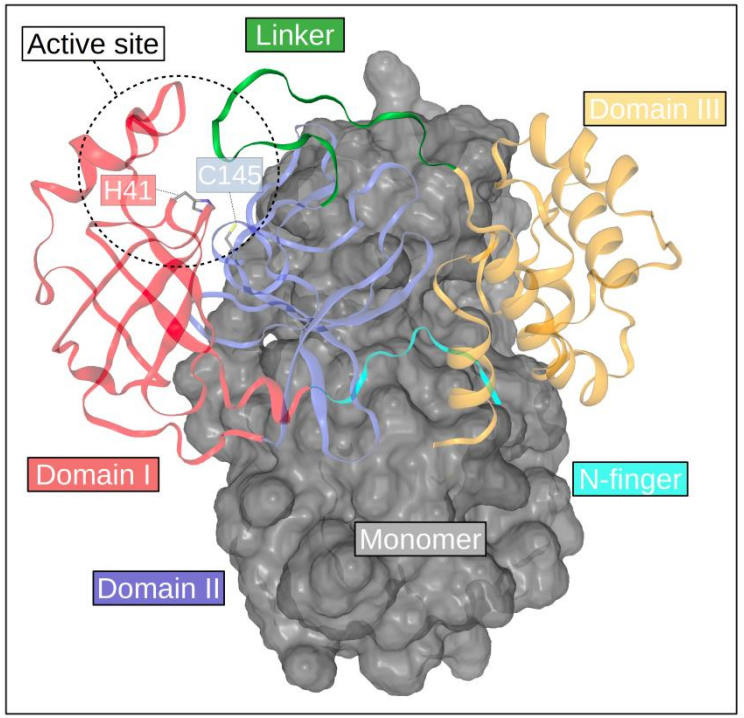
SYNONYMOUS SUBSTITUTIONS
a nucleotide mutation that does not alter amino acid sequences

MISSENSE MUTATIONS
mutations in a single nucleotide that result in the substitution of a different amino acid, resulting in a change to the protein encoded.

NONSENSE MUTATIONS
a mutation in the DNA sequence causes a protein to terminate prematurely by changing the original amino acid to a stop codon.

STEP 1:
Identification of mutations and effects on the reference M^{pro} protein structure

Main Protease - M^{pro}



- 3C-like protease

- a conserved protein present in all members of the *Coronavirinae* subfamily

- does not have a human homolog – good drug target

- **functions as a homodimer**

- Each monomer is call “protomer”

- SARS-CoV M^{pro} homolog
  - Homo-dimerization plays an important role in the catalytic activity of M^{pro}
  - Only one of the dimers was shown to be active at a time in SARS-CoV M^{pro}

Main Protease - M^{pro}

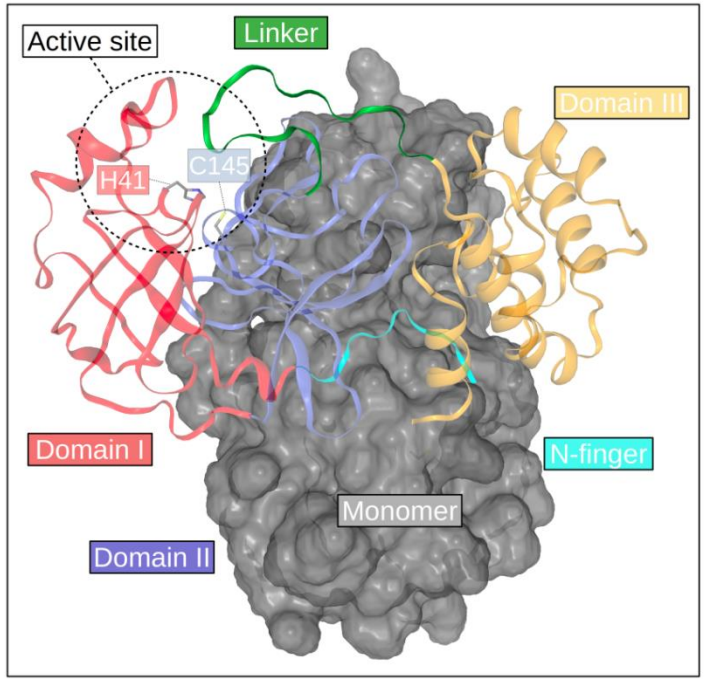
- comprises three domains (I-III)

- chymotrypsin-like domain I (residues 10-99)
- picornavirus 3C-protease like domain II (residues 100-182)

form a hydrophobic substrate binding site, with catalytic residues HIS41 and CYS145

- domain III (helical domain) (residues 198-303)

connected to domain II by a 15 residue linker loop, and was shown to regulate enzymatic activity in SARS-CoV.

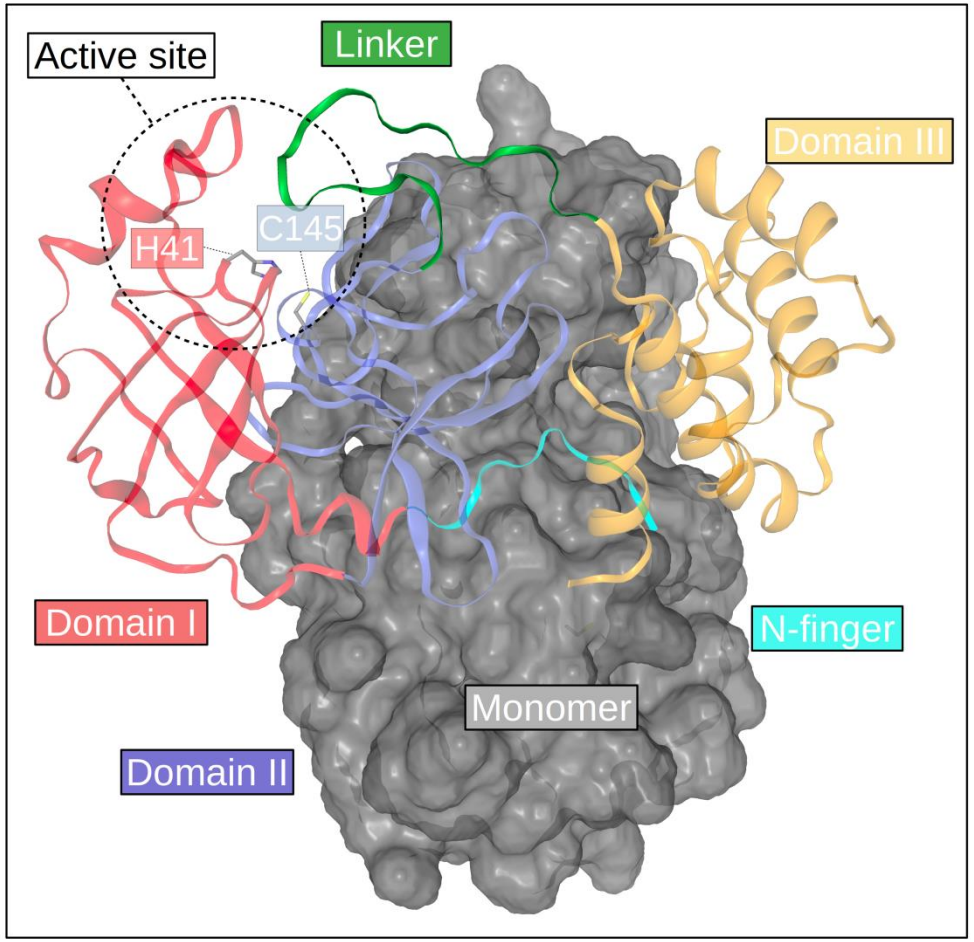


Main Protease - M^{pro}

- The majority of the dimer contact interface is a result of interactions present between domain II (protomer 1) and the N-finger (protomer 2).

- In the same manner, the N-finger from protomer 1 contacts domain II from protomer 2.

- Dimer stability!



Reference sequence/structure: PDB ID: 5RFV
50 distinct mutant sequences were filtered: GISAID

Table 1: List of missense mutations in SARS-CoV-2 M^{pro}. Samples with noticeably large differences in C_α RMSD from the reference protease are marked with an asterisk.

GISAID ID	Mutation	GISAID ID	Mutation	GISAID ID	Mutation
EPI_ISL_425319*	A7V	EPI_ISL_425284*	A116V	EPI_ISL_424470*	T196M
EPI_ISL_420422	G15D	EPI_ISL_421005*	P108S	EPI_ISL_423642	T201A
EPI_ISL_420181	G15S	EPI_ISL_422860*	A129V	EPI_ISL_419256*	L220F
EPI_ISL_425242	G15S,D48E	EPI_ISL_420579	P132L	EPI_ISL_421506	L232F
EPI_ISL_423772*	M17I	EPI_ISL_425655	T135I	EPI_ISL_425235	A234V
EPI_ISL_425342	V20L	EPI_ISL_420182	I136V	EPI_ISL_426097*	K236R
EPI_ISL_421312*	T45I	EPI_ISL_420510*	N151D	EPI_ISL_416720*	Y237H
EPI_ISL_425839*	M49I	EPI_ISL_415503	V157I	EPI_ISL_425886*	D248E
EPI_ISL_418269*	R60C	EPI_ISL_426028	V157L	EPI_ISL_418075	A255V
EPI_ISL_420306	K61R	EPI_ISL_417413	C160S	EPI_ISL_422919	I259T
EPI_ISL_421763*	A70T	EPI_ISL_418082	A173V	EPI_ISL_423725*	A260V
EPI_ISL_413021	G71S	EPI_ISL_423288	P184S	EPI_ISL_425498*	V261A
EPI_ISL_415643	L89F	EPI_ISL_420241*	P184L	EPI_ISL_421380*	A266V
EPI_ISL_420059	K90R	EPI_ISL_419710*	A191V,L220F	EPI_ISL_420610*	N274D
EPI_ISL_419756	P99L	EPI_ISL_415610*	A193V	EPI_ISL_425643	R279C
EPI_ISL_425132	Y101C	EPI_ISL_421515*	T198I	EPI_ISL_422184*	S301L
EPI_ISL_419984*	R105H	EPI_ISL_423007	T190I		

D48E variant (from sample EPI ISL 425242) lead to a novel “TSEEMLN"" motif at the substrate binding flap, which may have an impact on substrate binding affinity or even specificity.

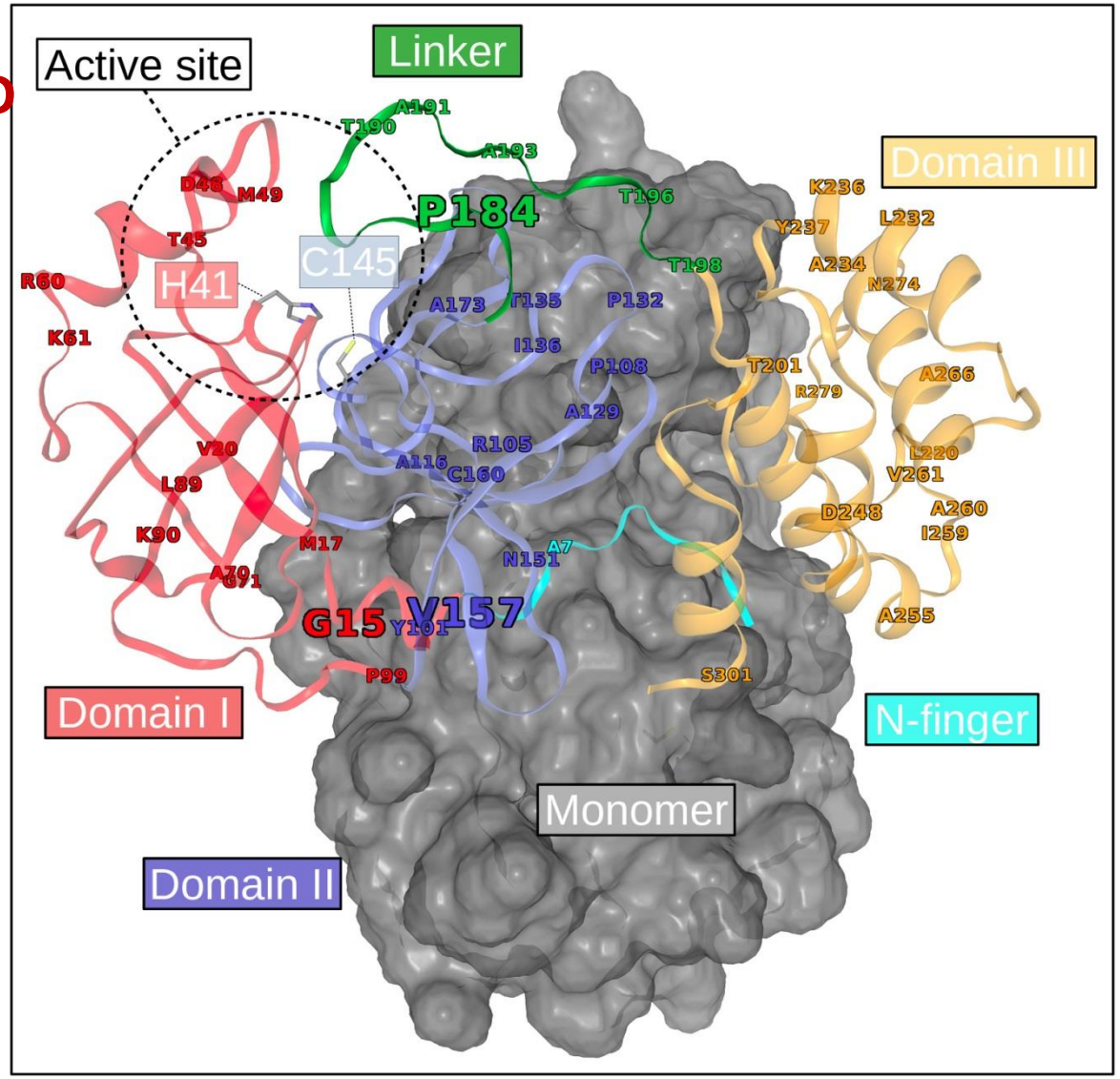


Figure 1: Mapping of the positions showing unique mutations from the reference M^{pro} sequence. For clarity, domains (I-III) are coloured (red, blue and orange respectively) only for one of the monomers, while the other is represented as a grey surface. The domain linker region is in green and the N-finger is in cyan. The size of the labels denotes the number of unique mutations recorded at that position.

Estimation of the protein backbone flexibility from MD using C_{alpha} RMSD

Some examples to highly divergent ones:

A7V A266V A70T M17I A116V

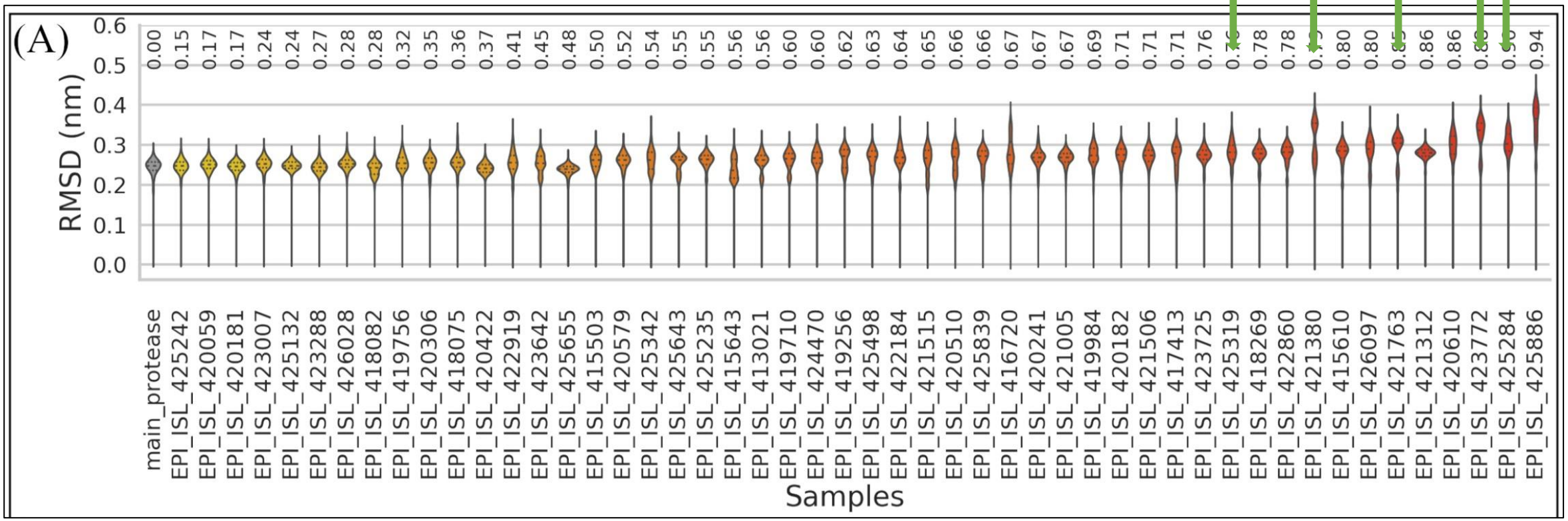


Figure 3: (A) Violin plots of C_α RMSD values for the reference (in grey) and the mutant (coloured in blue) M^{pro}, showing the 25th, 50th and 75th percentiles in dotted lines inside the kernel density plots. Distributions are scaled by area, and have been sorted by the KDE D_J distance (shown above each distribution) computed between the each sample and the reference protease.

- A7V occurs on the N-finger, which is a critical region for M^{pro} dimer stability.
- M17I occurs on an internal loop that connects a strand to a helix in domain I.
- A70T occurs on solvent-exposed loop in the same domain.
- A116V occurs in a buried strand within domain II.

N-finger can adopt a range of equilibrium conformations, with some showing protomer symmetries and others not

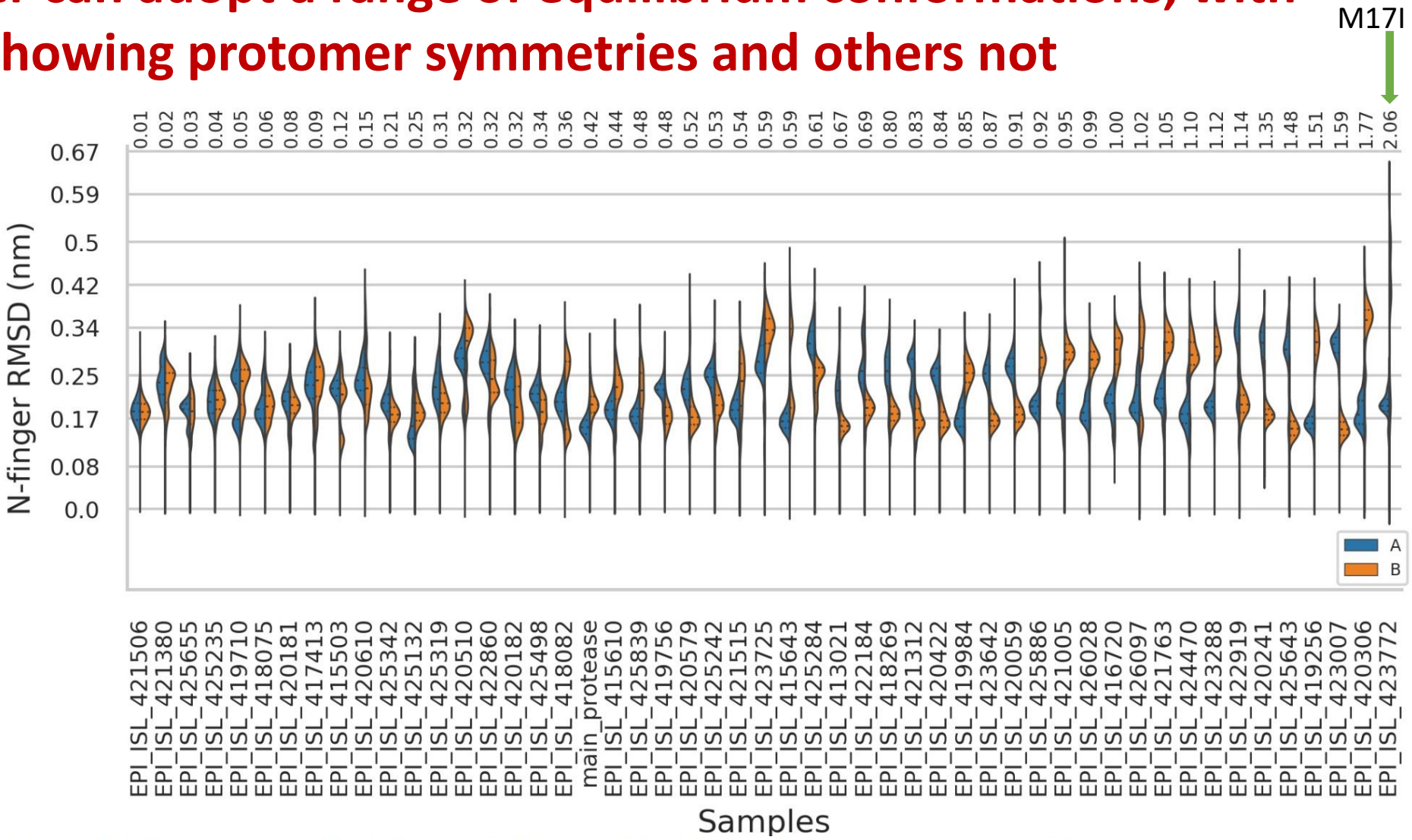


Figure 5: Kernel density distributions of RMSD values for the N-finger region across the mutant and reference protease complexes. The violin plots are split in two for each protein sample, showing the RMSD values for chains A (in blue) and B (in red). The tips of the distributions mark the minimum and maximum values for both chains combined in each complex. Samples have been sorted by increasing median distance between the chains, also shown (in Å) at the top of each sample distribution.

Interprotomer distances are estimated using COM distance

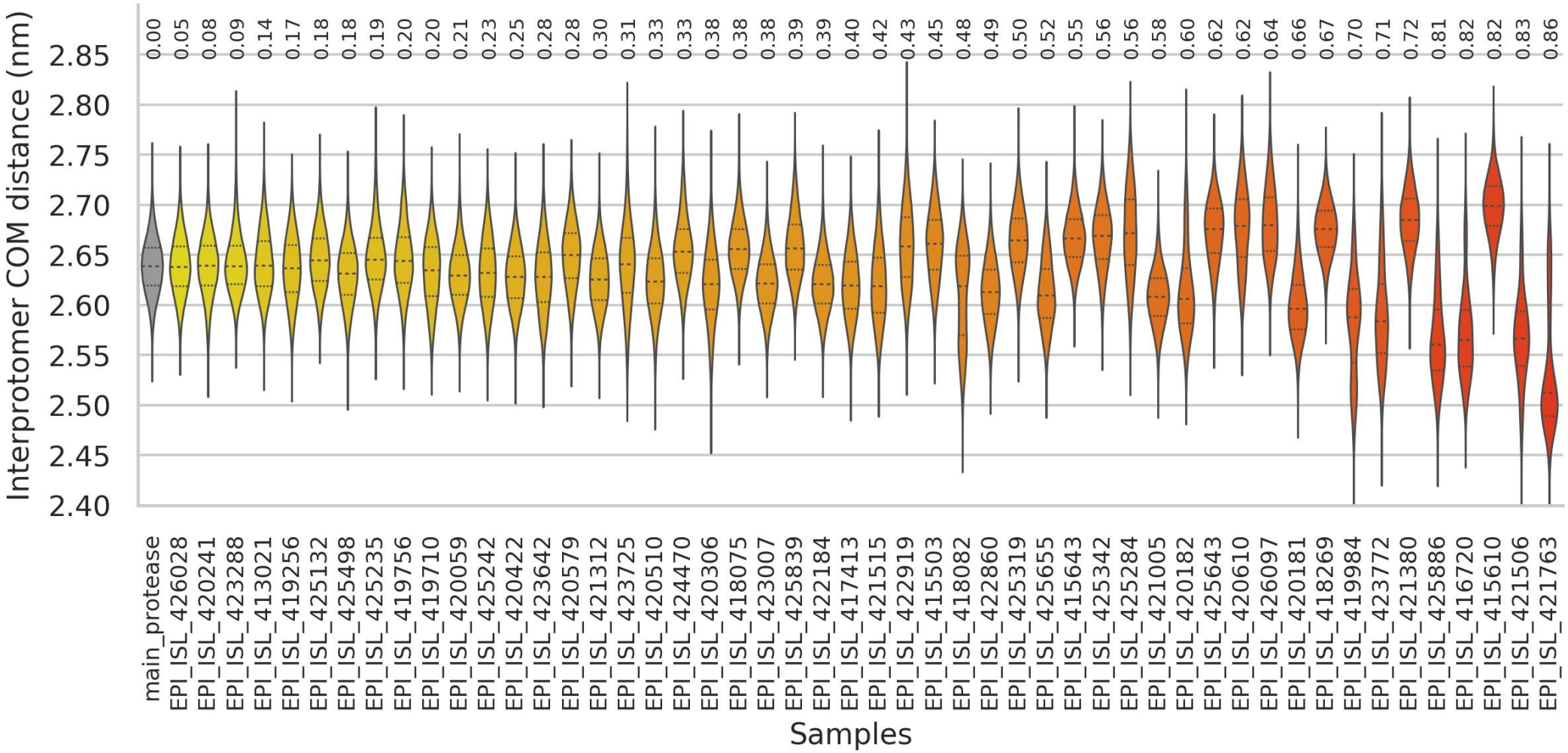


Figure 8: Distributions of interprotomer COM distances across samples, arranged in ascending order of the KDE d_J. The reference protease is in grey while the mutants are coloured from yellow to red, in increasing order of distance from the reference KDE.

Inter-domain angles in each protomer of Mpro are calculated

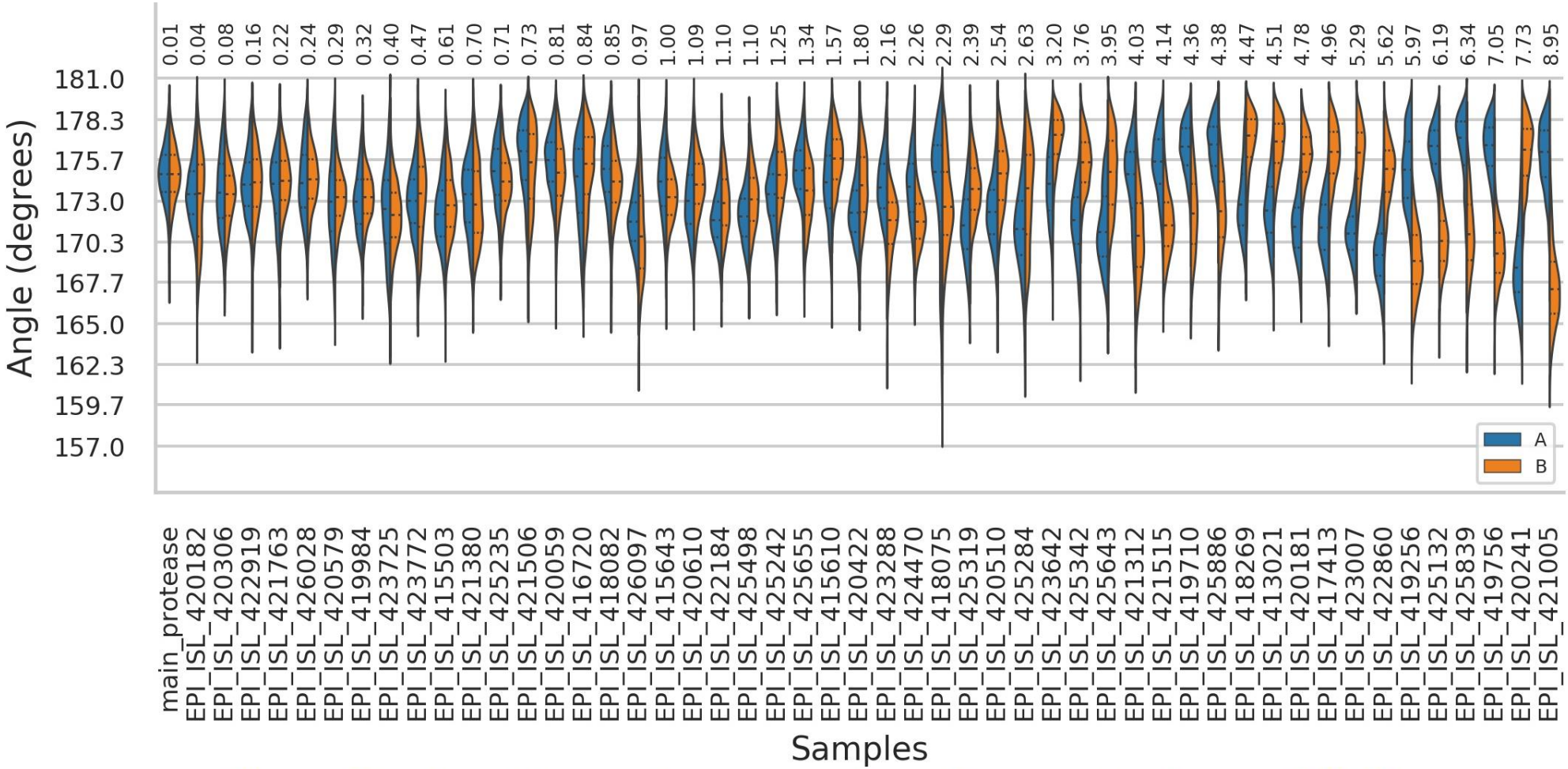


Figure 10: Kernel density distributions of inter-domain angles (domains I-II-III) across the mutant and reference protease complexes. The violin plots are split in two for each protein, showing the inter-domain angles for chains A (in blue) and B (in red). The tips of the distributions mark the minimum and maximum values for both chains combined in each protein complex.

STEP 2:
Identification of potential allosteric site and effects of mutations on allosteric and active site

Dual allosteric pockets coinciding with various stabilizing and functional components of the substrate binding pocket is identified

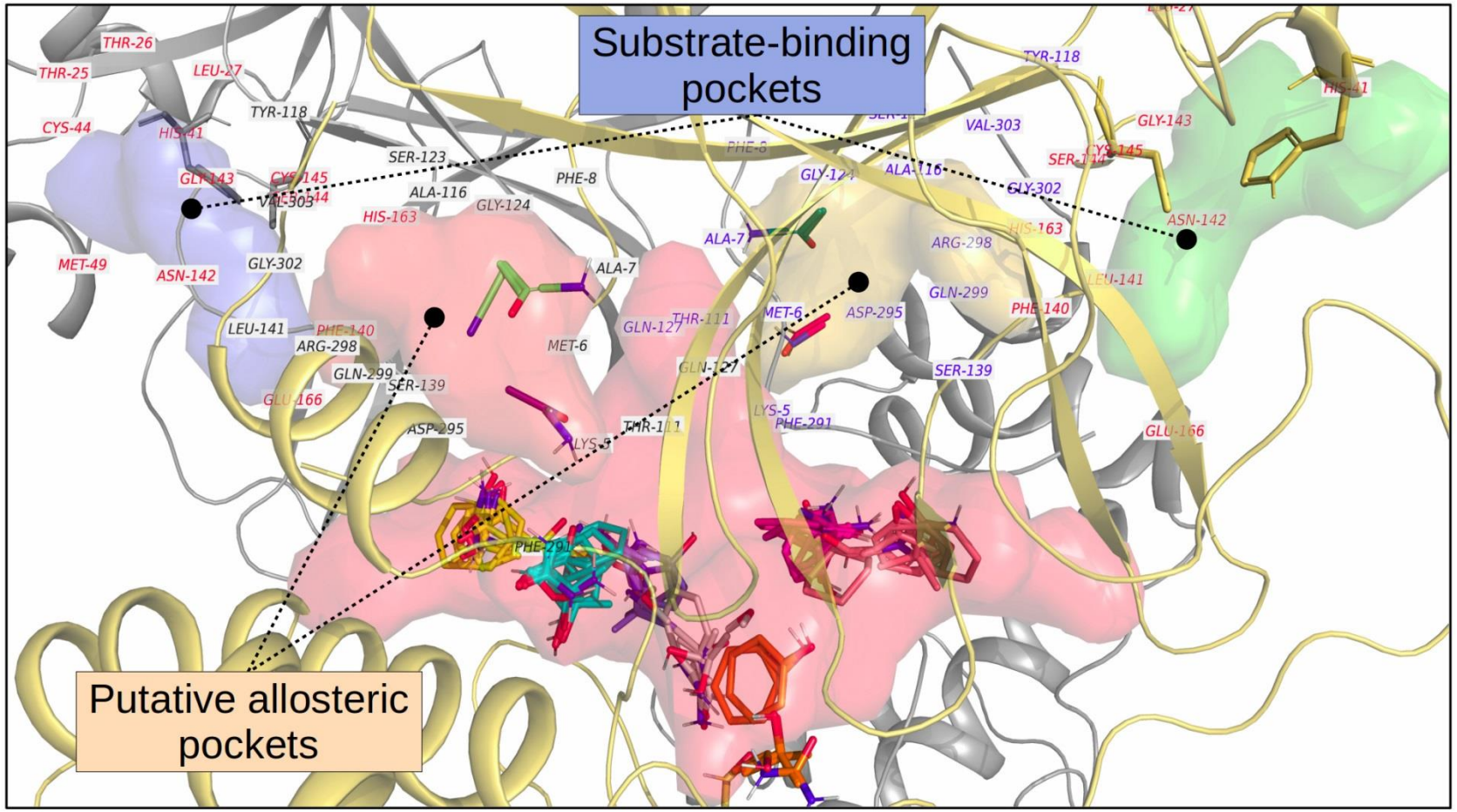


Figure 12: Pocket detection using combined predictions from FTMap and PyVOL. FTMap probes are shown as stick figure representations, while those from PyVOL are shown as surfaces. The protomers are depicted as cartoon representations, in grey and light orange.

An asymmetry in compaction between substrate binding pockets coming from each protomer in each sample is observed

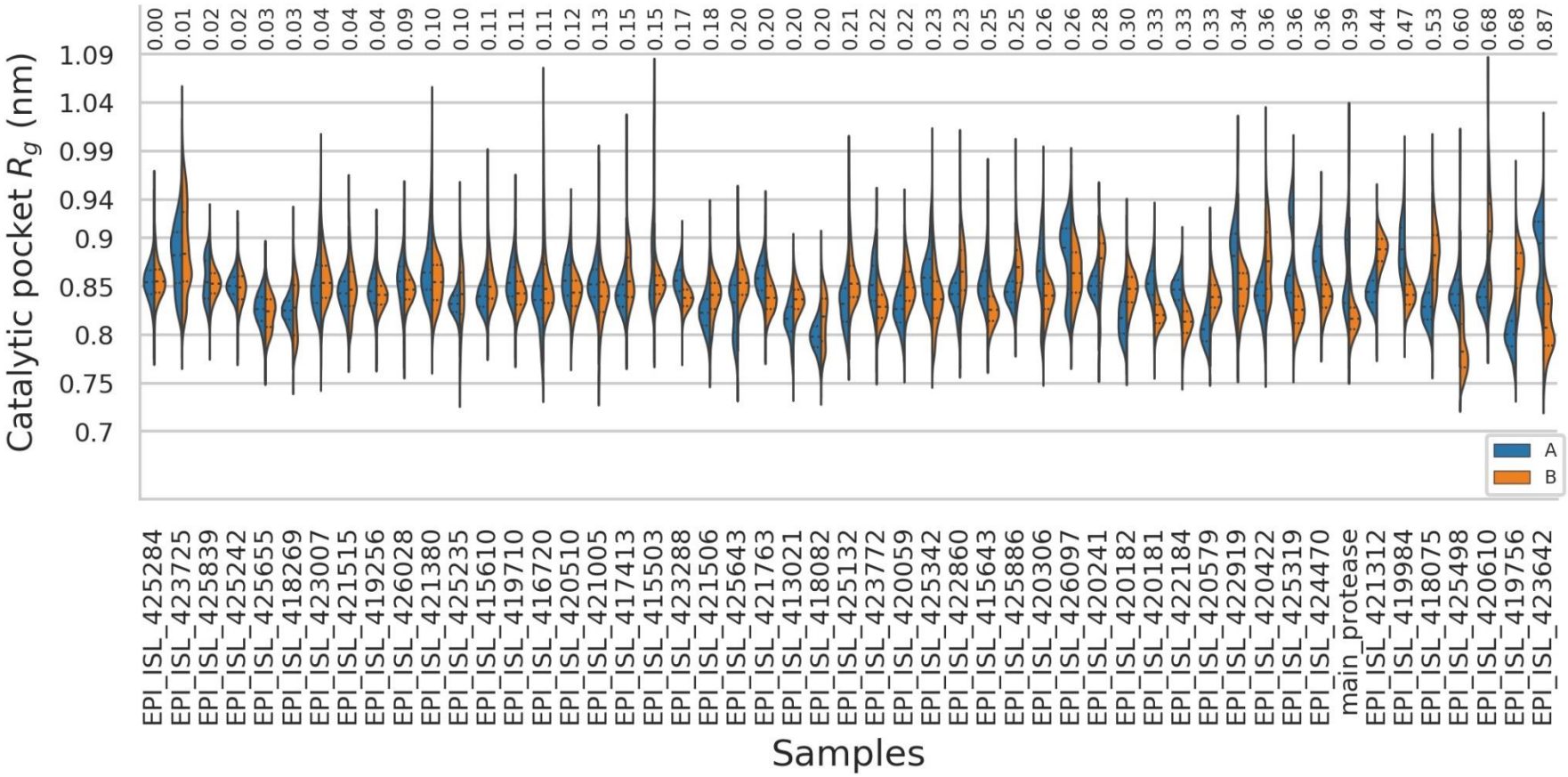


Figure 13: Kernel density distributions of R_g values for the substrate binding site from each protomer of M^{pro}, arranged in ascending order of difference in median from each chain. The differences, shown at the top are in nanometres. Chain A values are in red while chain B values are in blue. The maxima and minima are across protomers. Quartiles for each binding site are shown as dotted lines.

The interprotomer (allosteric) pockets may play an important role in affecting the degree of compaction of the binding cavity and vice-versa

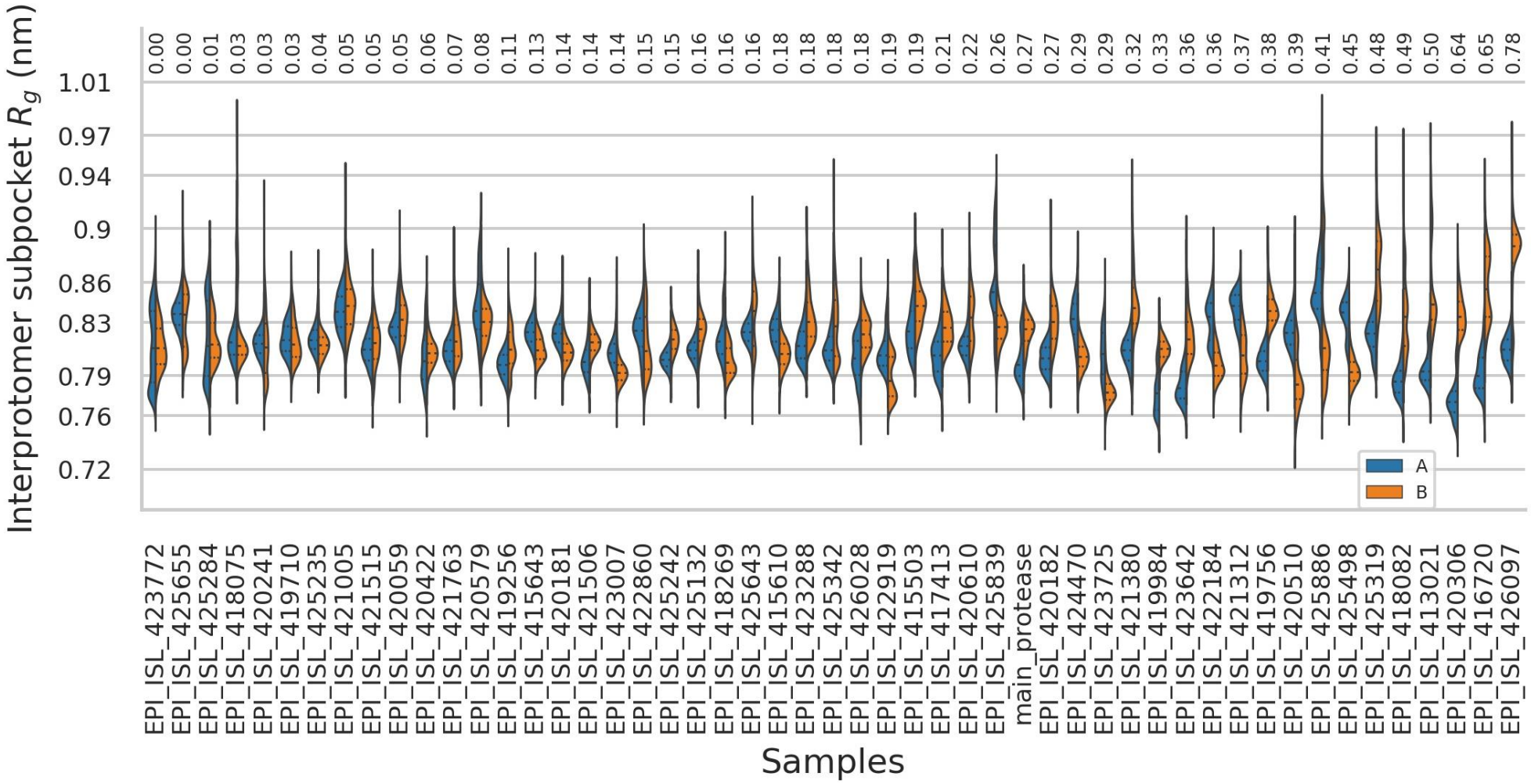
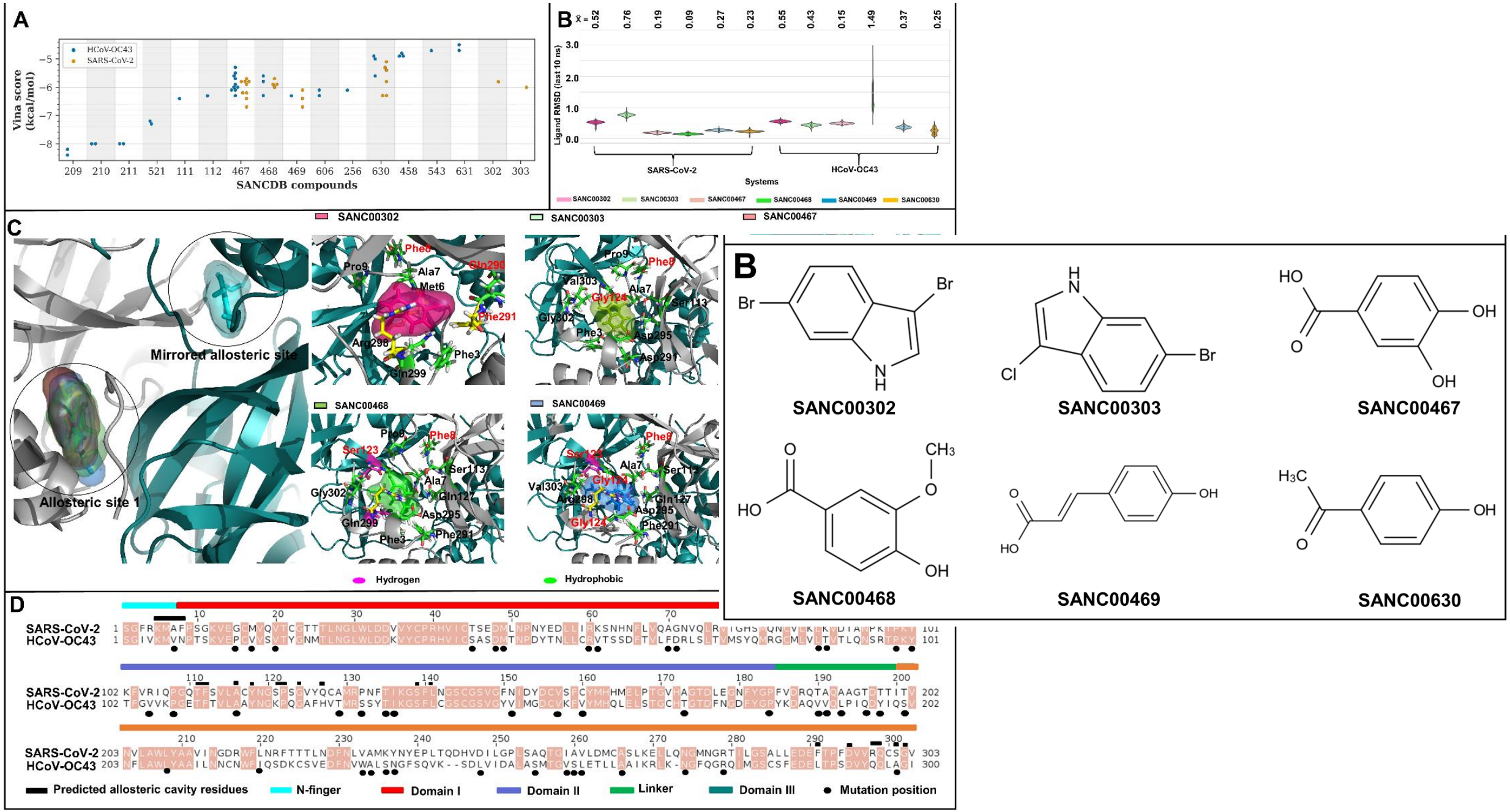


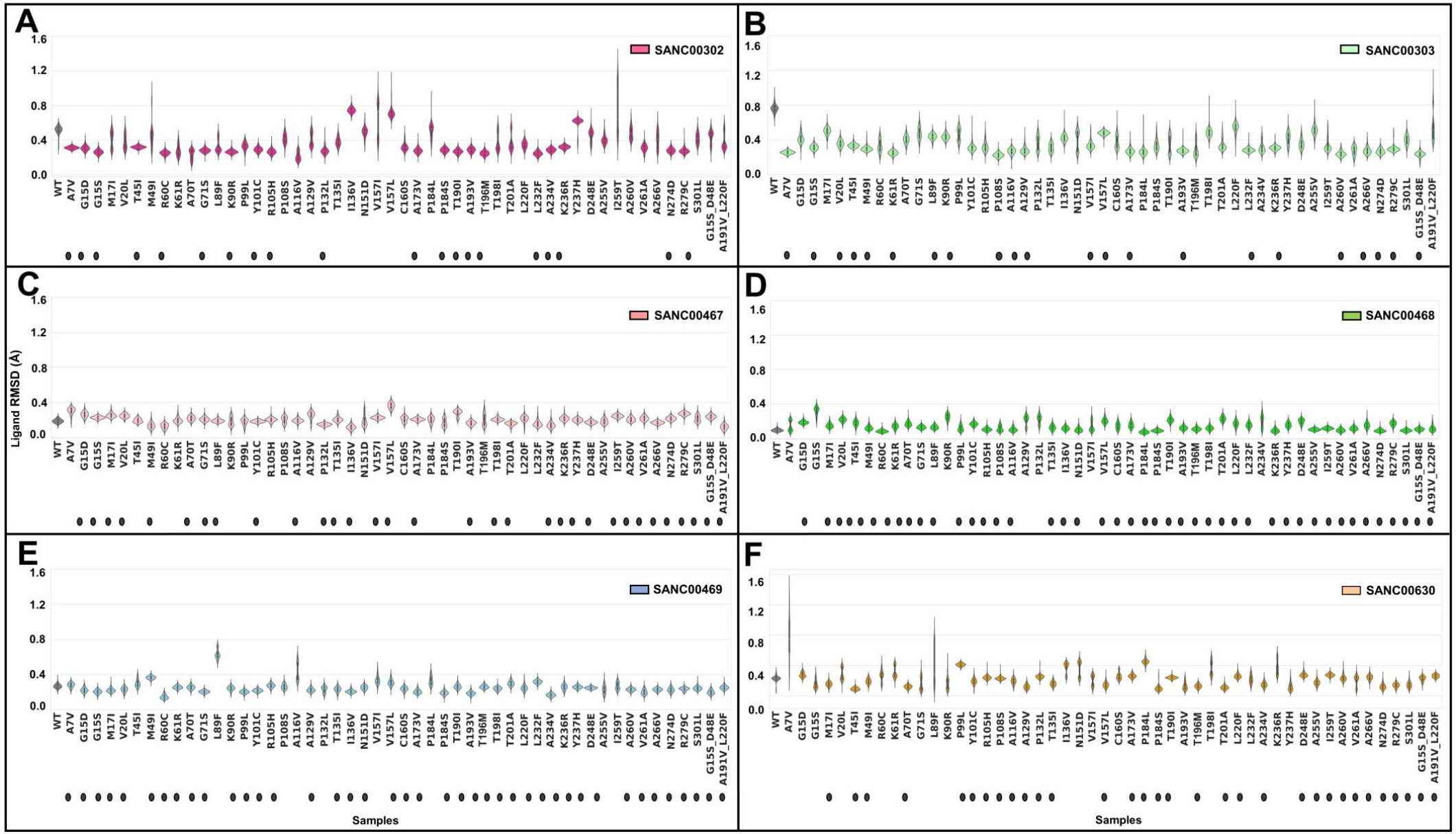
Figure 15: Kernel density distributions of R_g values across samples for the mirrored interfacial (and potentially allosteric) pockets. The differences, shown at the top are in nanometres.

STEP 3:
Identification of potential allosteric modulators

SANCDB compounds have relatively different behavior in the M^{pro} protein of SARS-CoV-2 and of HCoV-OC43



Potential allosteric modulators in the presence of mutations



Six allosteric modulators were stable only on the 3 mutant systems

Isolate	Position	302	303	467	468	469	630	Consensus
EPI_ISL_413021	**G71S**	x		x	x	x		4
EPI_ISL_415503	**V157I**		x	x				2
EPI_ISL_415610	**A193V**	x	x	x	x	x		5
EPI_ISL_415643	**L89F**		x	x	x			3
EPI_ISL_416720	**Y237H**			x	x	x		3
EPI_ISL_417413	**C160S**				x	x		2
EPI_ISL_418075	**A255V**				x		x	2
EPI_ISL_418082	**A173V**	x	x	x	x	x	x	6
EPI_ISL_418269	**R60C**	x			x	x		2
EPI_ISL_419256	**L220F**				x	x	x	3
EPI_ISL_419710	**A191V, L220F**			x	x	x	x	4
EPI_ISL_419756	**P99L**				x	x	x	3
EPI_ISL_419984	**R105H**	x			x	x	x	4
EPI_ISL_420059	**K90R**	x	x			x		3
EPI_ISL_420181	**G15S**	x	x	x		x		4
EPI_ISL_420182	**I136V**			x	x	x		3
EPI_ISL_420241	**P184L**				x		x	2
EPI_ISL_420306	**K61R**		x		x	x		3
EPI_ISL_420422	**G15D**	x		x	x	x		4
EPI_ISL_420510	**N151D**				x	x		2
EPI_ISL_420579	**P132L**	x		x			x	3
EPI_ISL_420610	**N274D**	x	x	x	x	x	x	6
EPI_ISL_421005	**P108S**		x		x		x	3
EPI_ISL_421312	**T45I**	x	x		x		x	4
EPI_ISL_421380	**A266V**		x	x	x	x	x	5

Isolate	Position	302	303	467	468	469	630	Consensus
EPI_ISL_421506	**L232F**	x	x		x	x		4
EPI_ISL_421515	**T198I**			x	x	x		3
EPI_ISL_421763	**A70T**			x	x	x	x	4
EPI_ISL_422184	**S301L**				x	x	x	3
EPI_ISL_422860	**A129V**		x			x	x	3
EPI_ISL_422919	**I259T**			x	x		x	3
EPI_ISL_423007	**T190I**	x			x	x	x	4
EPI_ISL_423288	**P184S**	x			x	x	x	4
EPI_ISL_423642	**T201A**			x	x	x	x	4
EPI_ISL_423725	**A260V**		x	x	x	x	x	5
EPI_ISL_423772	**M17I**			x	x	x	x	4
EPI_ISL_424470	**T196M**	x			x	x	x	4
EPI_ISL_425132	**Y101C**	x		x	x	x		4
EPI_ISL_425235	**A234V**	x				x	x	3
EPI_ISL_425242	**G15S, D48E**		x	x	x	x	x	5
EPI_ISL_425284	**A116V**		x	x	x		x	4
EPI_ISL_425319	**A7V**	x	x			x		3
EPI_ISL_425342	**V20L**		x	x	x	x		4
EPI_ISL_425498	**V261A**			x	x	x		3
EPI_ISL_425643	**R279C**	x	x	x	x	x	x	6
EPI_ISL_425655	**T135I**			x	x	x	x	4
EPI_ISL_425839	**M49I**		x	x	x	x	x	5
EPI_ISL_425886	**D248E**			x	x	x	x	4
EPI_ISL_426028	**V157L**		x		x	x	x	4
EPI_ISL_426097	**K236R**	x	x	x	x	x		5
		20	**22**	**29**	**43**	**41**	**30**	

sancdb.rubi.ru.ac.za/compounds/468/

SANCDB

Search ▾ Tools ▾ Documentation ▾ Cite Us About Us ▾ Download Database

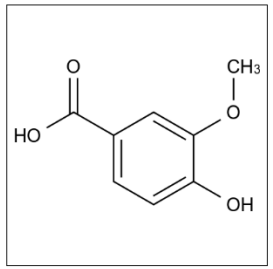
Search...

SEARCH OPTIONS

- Name
- SMILES
- Structure
- Properties
- Source Organism
- Classification
- Use
- Author
- Reference
- Advanced

SANC00468

Entry name:	4-Hydroxy-3-methoxybenzoic acid
Formula:	$C_8H_8O_4$
Molecular mass:	168.15
ChEMBL ID:	CHEMBL120568
ChemSpider ID:	8155
ZINC ID:	ZINC00338275
PubChem ID:	CID: 8468
CAS No.	121-34-6



2D Image 3D View

Downloads

View analogs

SMILES:

COc1cc(C(=O)O)ccc1O

Search PubChem SMILES

Scaffold SMILES:

c1ccccc1

References

- Koorbanally et al. (2004) Bufadienolides from Drimia robusta and Urginea epigea (Hyacinthaceae)
- Koorbanally et al. (2005) A novel homoisoflavonoid from Drimia delagoensis (Urgineoideae: Hyacinthaceae)

Classifications

- Aromatic acid
- Benzene and substituted derivatives (Classyfire)

Other Names

- 4-Hydroxy-3-Methoxybenzoic Acid
- Benzoic Acid, 4-Hydroxy-3-Methoxy-
- Vanillic Acid
- 2-Methoxy-4-Carboxyphenol
- 3-Methoxy-4-Hydroxybenzoic Acid
- Nsc 3987
- Nsc 674322
- Va
- M-Methoxy-P-Hydroxy-Benzoic Acid

Source Organisms

- Drimia robusta
- Drimia delagoensis

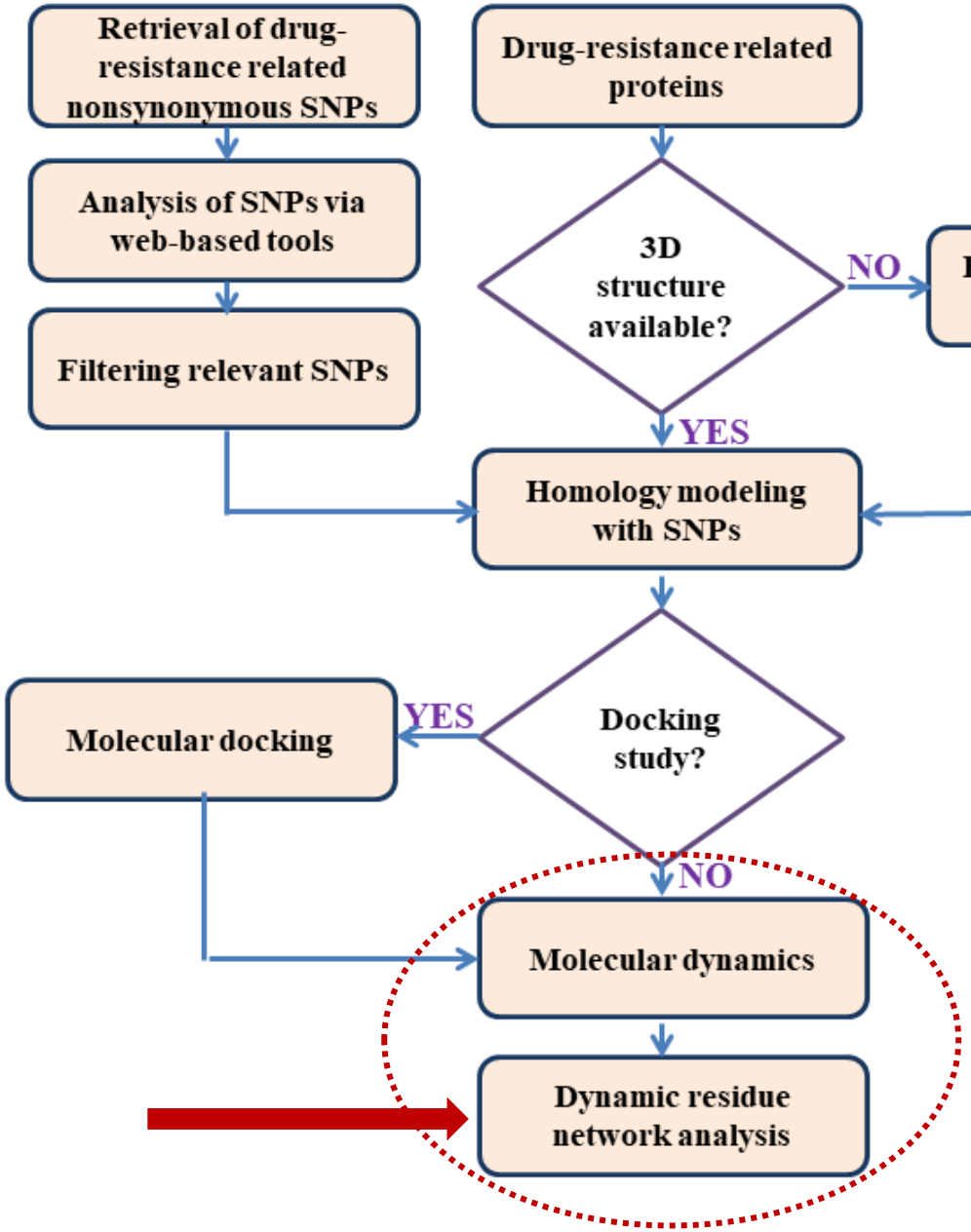
Compound Uses

None Recorded

STEP 4:
Hub (centrality) residues and allosteric communication paths (if any) and changes in the presence of mutations

{

Analysis of missense mutations: Proposed protocol



REVIEW

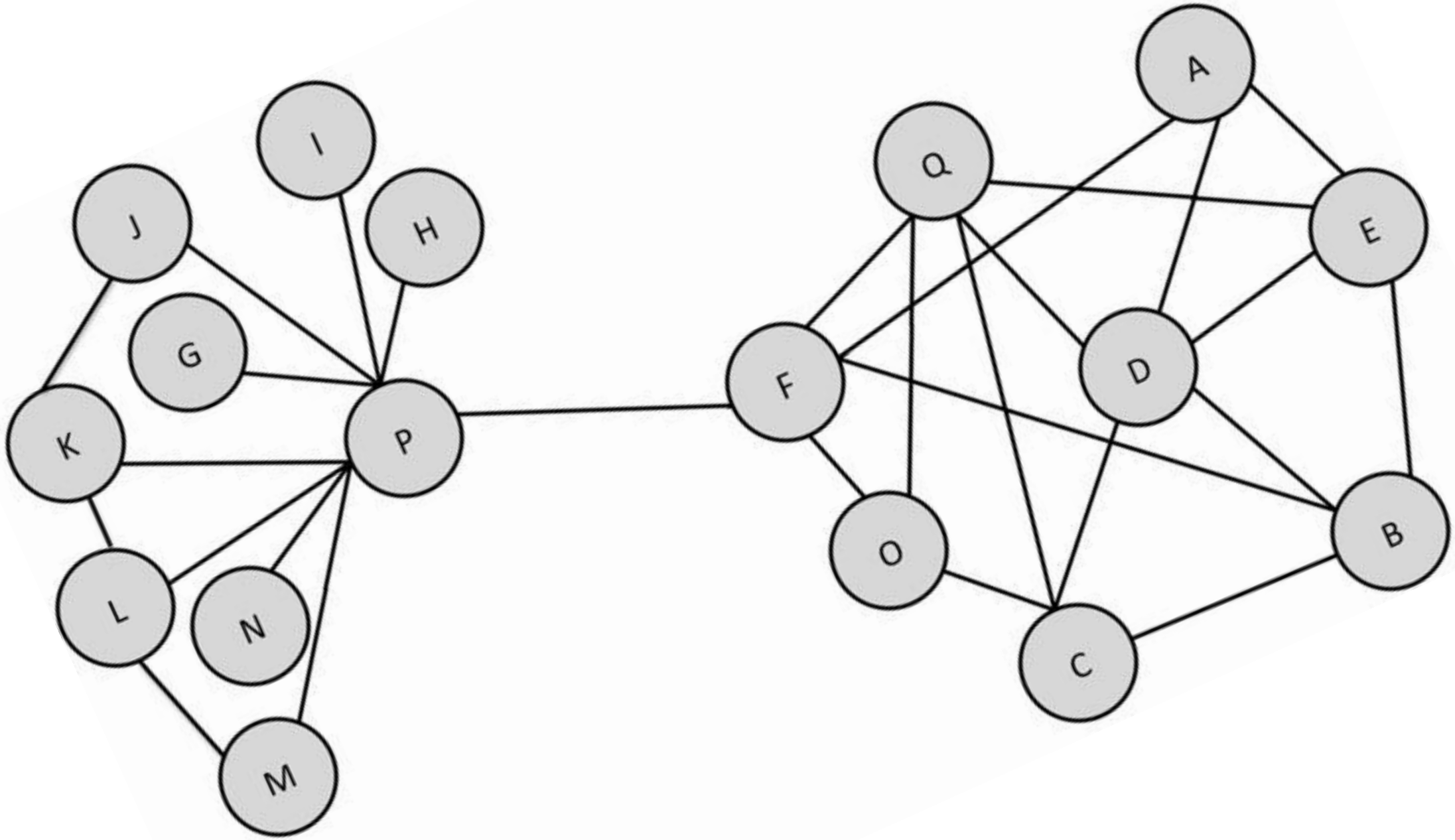
GLOBAL HEART, VOL. 12, NO. 2, 2017
June 2017: 151-161

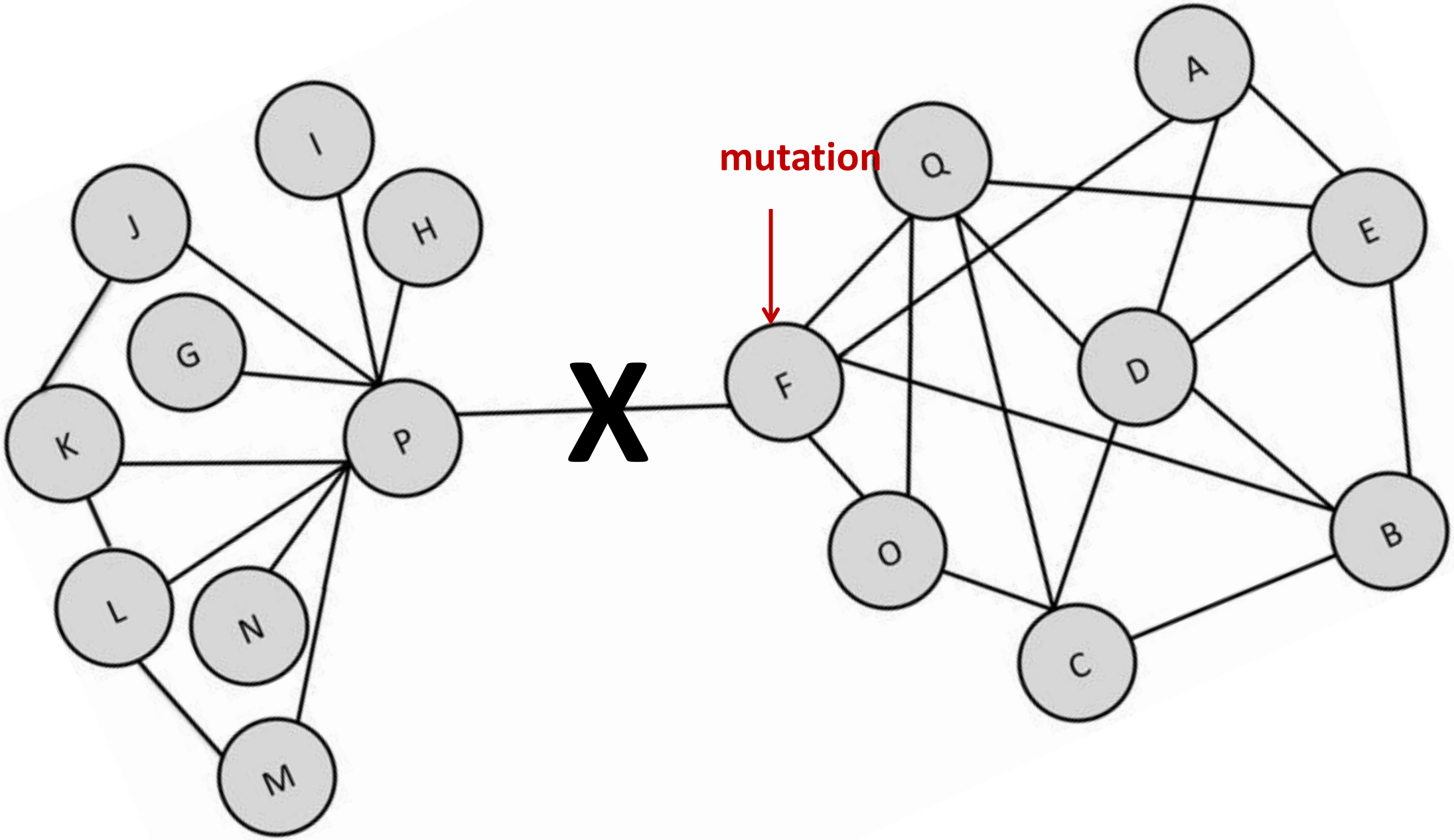
gREVIEW

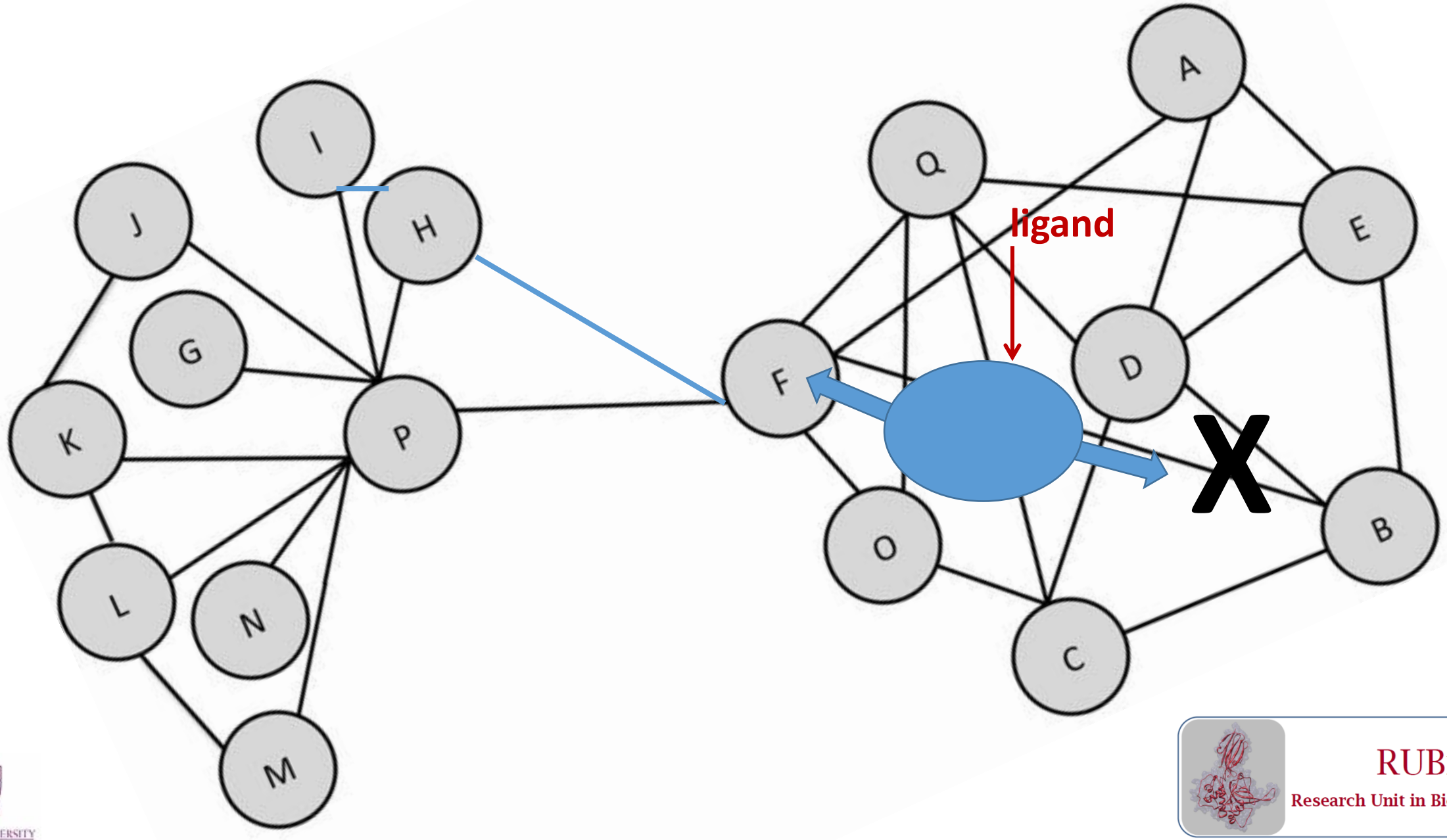
Role of Structural Bioinformatics in Drug Discovery by Computational SNP Analysis

Analyzing Variation at the Protein Level

David K. Brown, Özlem Tastan Bishop
Grahamstown, South Africa


I
J
H
G
K
P
L
N
M
A
Q
E
F
D
O
B
C


I
J
H
G
K
P
L
N
M
X
mutation
A
Q
E
F
D
O
B
C



RHODES UNIVERSITY

RUBi
Research Unit in Bioinformatics

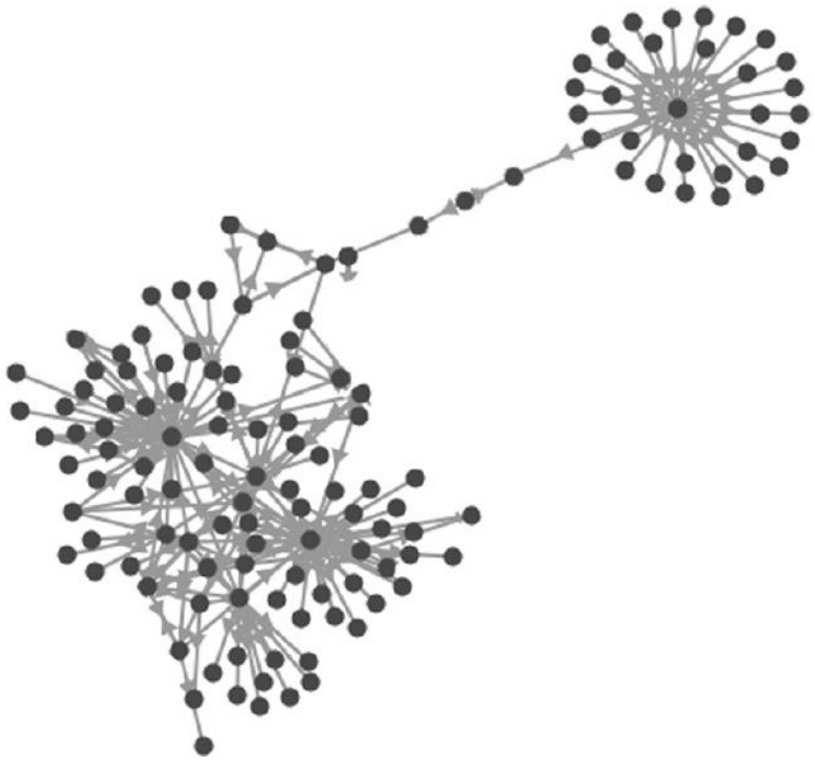


FIGURE 3.3

The node at the center of the cluster in the upper right would have a high degree centrality, even though it is far from the dense center of the network.

- ***betweenness centrality (BC),***
- ***closeness centrality (CC),***
- ***degree centrality (DC),***
- ***eigencentrality (EC),***
- ***katz centrality (KC)***

The averaged *BC* metric is defined as how often a residue is traversed along the shortest paths connecting every other residue pairs.

Averaged *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

DC defines the number of neighboring nodes (the local connectivity) around a given node.

***EC* measures the high centrality given to high degree residue, or to a residue that is connected to other high degree residues.**

KC measures the relative degree of influence of a residue i within connected residues in a network.

The term ‘‘**centrality**” is used as a measure of how **central a residue is in the protein network**, and several centrality metrics derived from the social sciences

Bioinformatics, 33(17), 2017, 2768–2771
doi: 10.1093/bioinformatics/btx349
Advance Access Publication Date: 31 May 2017
Applications Note

OXFORD

Structural bioinformatics

MD-TASK: a software suite for analyzing molecular dynamics trajectories

David K. Brown[1], David L. Penkler[1], Olivier Sheik Amamuddy[1], Caroline Ross[1], Ali Rana Atilgan[2], Canan Atilgan[2] and Özlem Tastan Bishop[1,*]

[1]Research Unit in Bioinformatics (RUBi), Department of Biochemistry ar
Grahamstown 6140, South Africa and [2]Faculty of Engineering and Natural
34956, Istanbul, Turkey

Computational and Structural Biotechnology Journal 19 (2021) 5059–5071

ELSEVIER

COMPUTATIONAL ANDSTRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

MDM-TASK-web: MD-TASK and MODE-TASK web server for analyzing protein dynamics

Check for updates

Olivier Sheik Amamuddy, Michael Glenister, Thulani Tshabalala, Özlem Tastan Bishop *

Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University, Makhanda 6140, South Africa

ARTICLE INFO

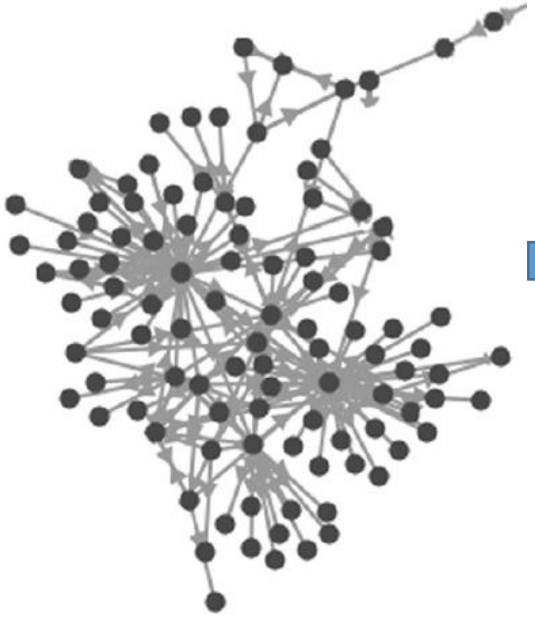
Article history:
Received 27 April 2021
Received in revised form 28 August 2021
Accepted 28 August 2021
Available online 02 September 2021

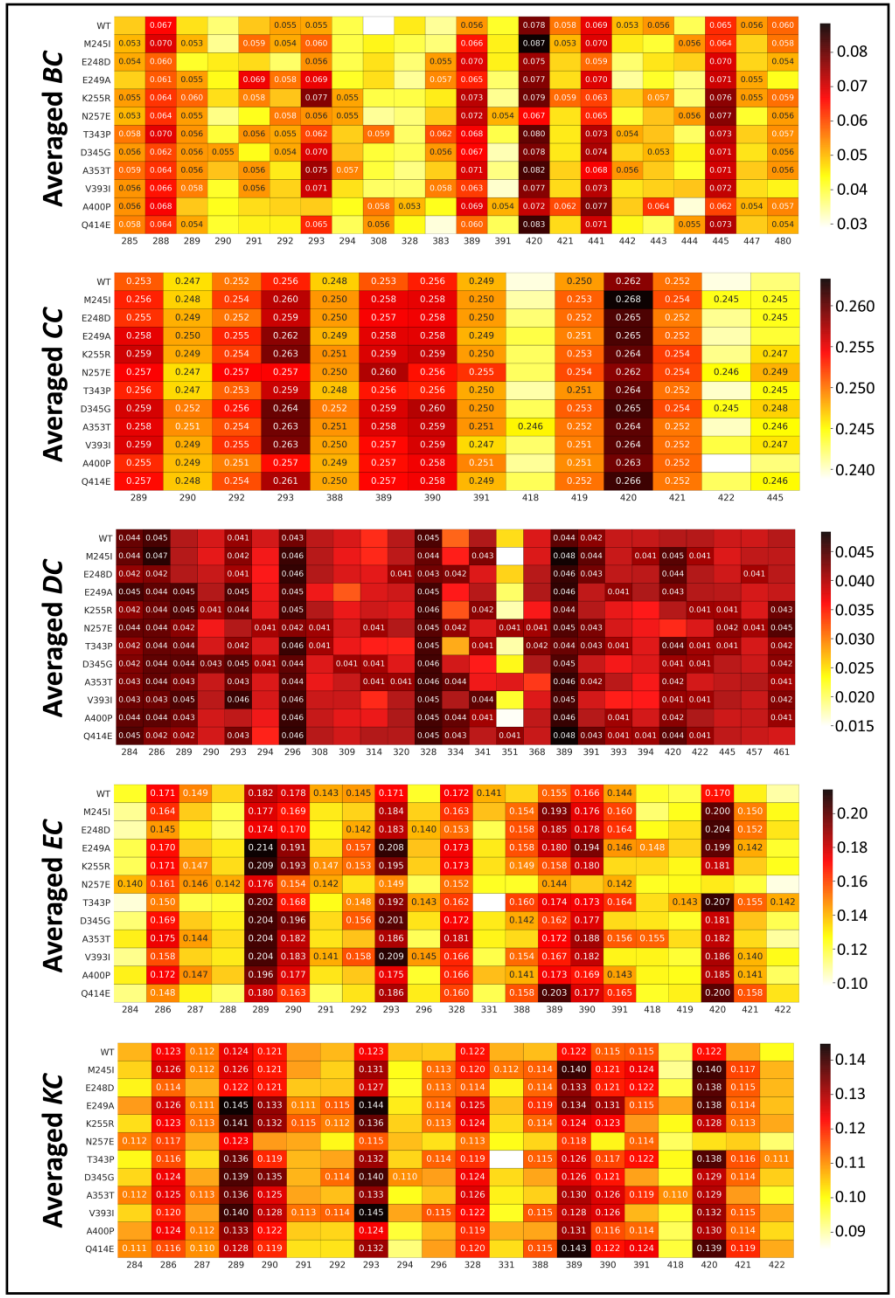
ABSTRACT

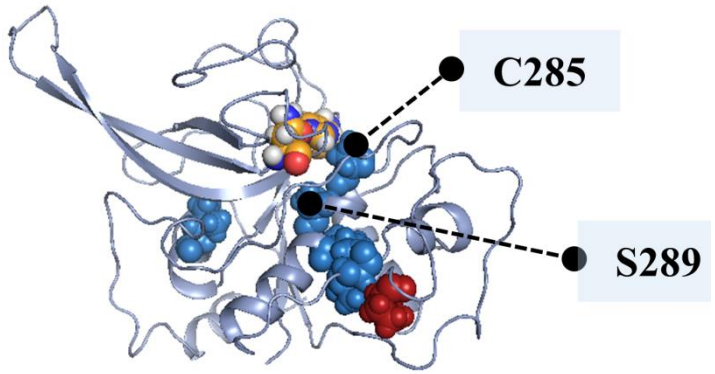
The web server, MDM-TASK-web, combines the MD-TASK and MODE-TASK software suites, which are aimed at the coarse-grained analysis of static and all-atom MD-simulated proteins, using a variety of non-conventional approaches, such as dynamic residue network analysis, perturbation-response scanning, dynamic cross-correlation, essential dynamics and normal mode analysis. Altogether, these tools allow for the exploration of protein dynamics at various levels of detail, spanning single residue pertur-

}

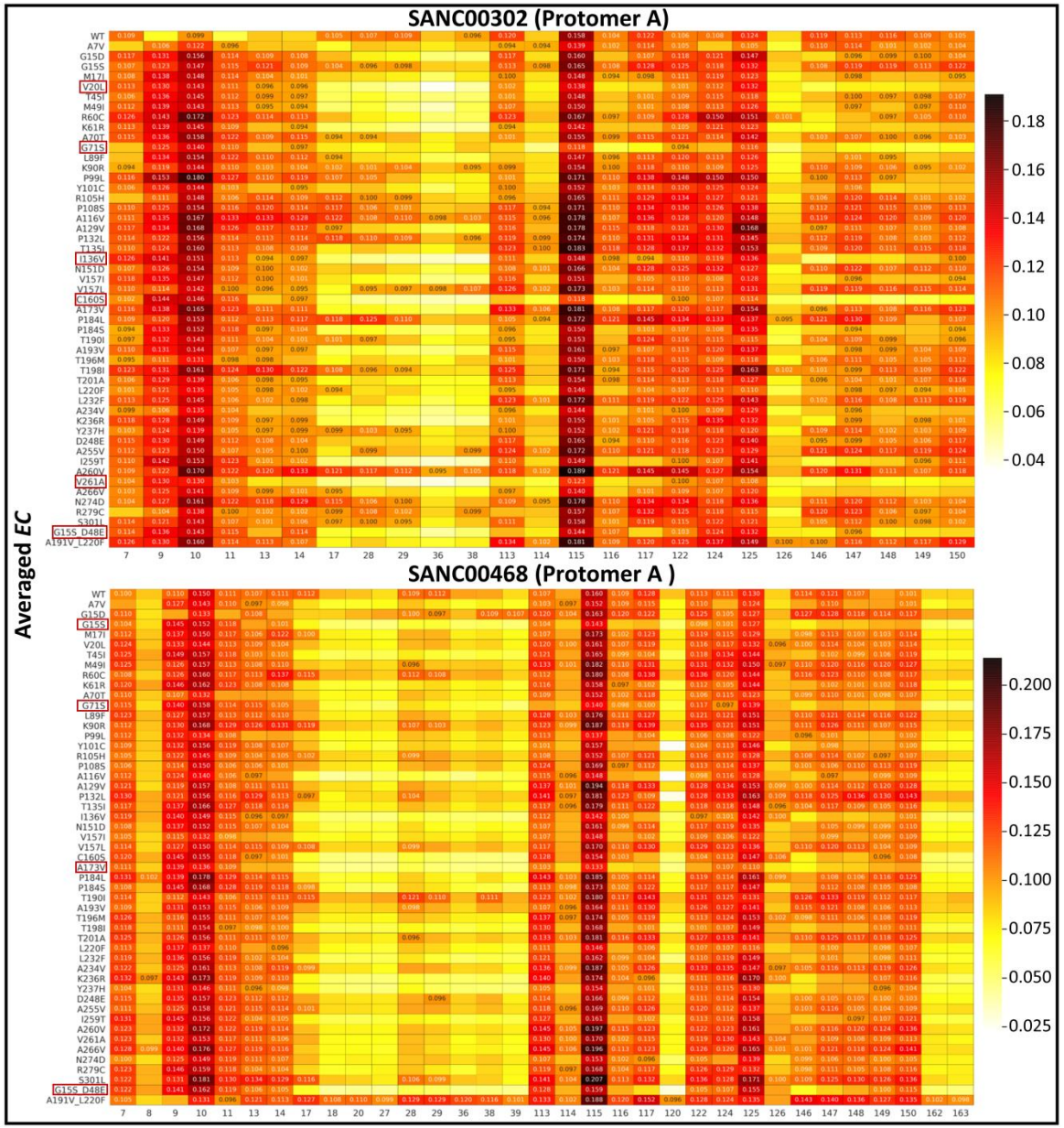
METHODOLOGY



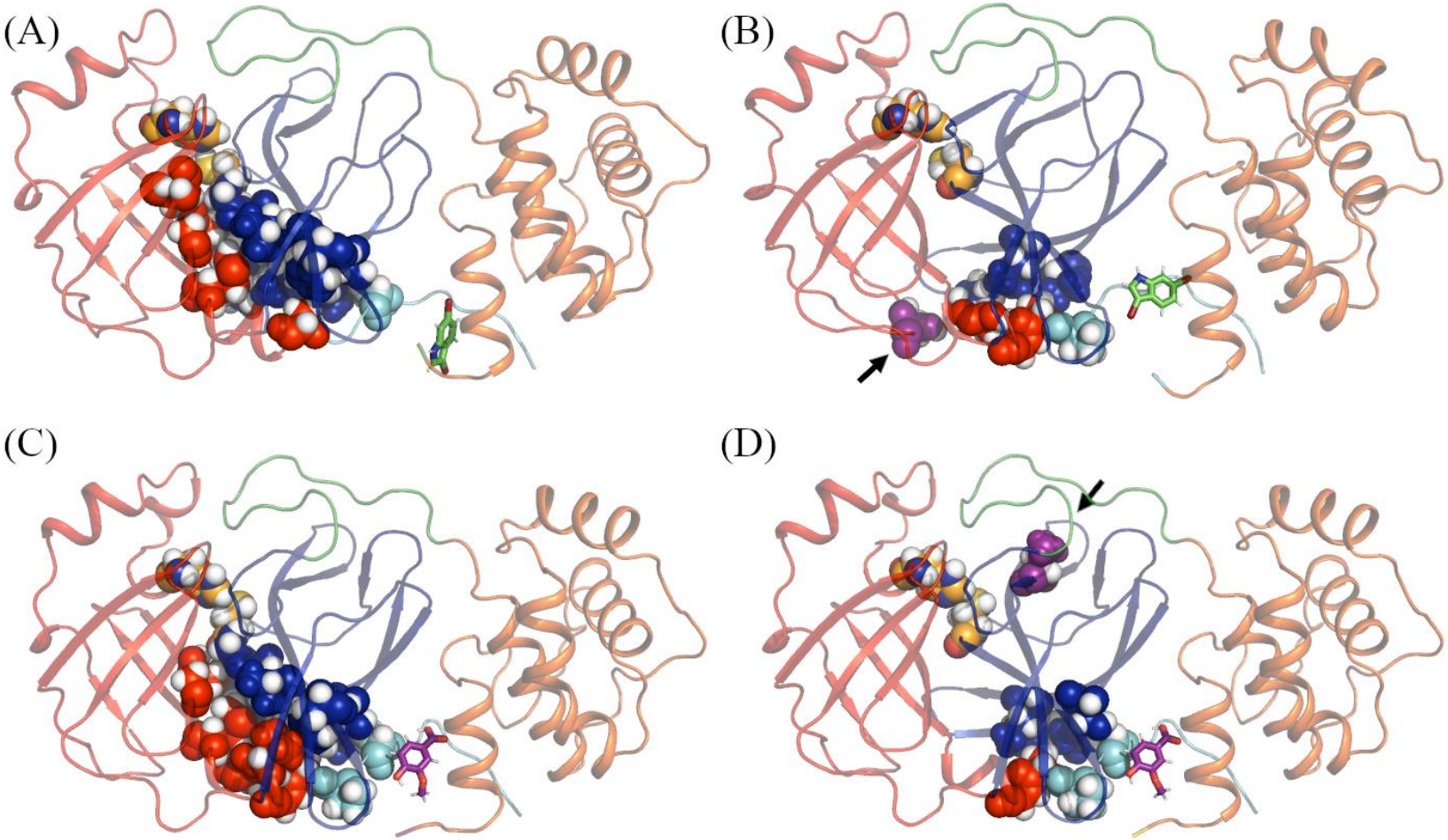




Heat map for the potential hubs according to averaged *EC* metric for the reference and 50 mutant proteins in allosterically bound state to SANC00302 and SANC00468.



The communication path traced by averaged *EC* hubs, starting from the allosteric ligand towards the catalytic residue



(A) M^{pro}-SANC00302 reference protein – ligand complex. Allosteric modulator is in green.
(B) (B) Mutant M^{pro} (G71S)-SANC00302 complex. Mutant (in purple) indicated by arrow.
(C) C) M^{pro}- SANC00468 reference protein – ligand complex. Allosteric modulator is in purple.
(D) (D) Mutant M^{pro} (A173V)-SANC00468 complex.

Conclusion

Collectively, our approaches offer routes for novel rational drug discovery methods and provide computationally feasible platforms

Acknowledgement

RHODES UNIVERSITY
Where leaders learn

NRF

NIH
National Institutes of Health

H3ABioNet

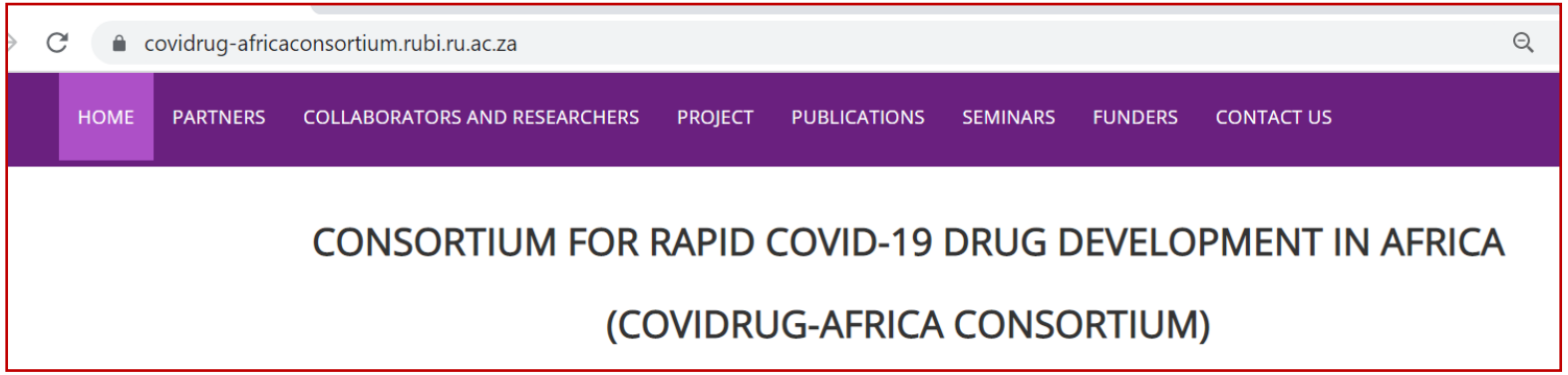
BILL & MELINDA GATES foundation

DELGEME
Training Leaders for a Malaria-free Africa

The African Academy of Sciences

wellcome

CHPC
CENTRE FOR HIGH PERFORMANCE COMPUTING



https://covidrug-africaconsortium.rubi.ru.ac.za/