

Ligand Accessibility Insights to the Dengue Virus NS3-NS2B Protease Assessed by Long-Timescale Molecular Dynamics Simulations

Thales Kronenberger^{+,*}[a, b] Mateus Sá Magalhães Serafim^{+, [c]} Arun Kumar Tonduru,^[b] Vinícius Gonçalves Maltarollo,^[d] and Antti Poso^[a, b]

Dengue is a tropical disease caused by the dengue virus (DENV), with an estimate of 300 million new cases every year. Due to the limited vaccine efficiency and absence of effective antiviral treatment, new drug candidates are urgently needed. DENV NS3-NS2B protease complex is essential for viral post-translational processing and maturation, and this enzyme has been extensively studied as a relevant drug target. Crystal structures often underestimate NS3-NS2B flexibility, whereas

they can adopt different conformational states depending on the bound substrate. We conducted molecular dynamics simulations (~30 μ s) with a non- and covalently bound inhibitor to understand the conformational changes in the DENV-3 NS3-NS2B complex. Our results show that the open-closing movement of the protease exposes multiple druggable subpockets that can be investigated in later drug discovery efforts.

Introduction

Dengue virus (DENV), a member of the *Flavivirus* genus and *Flaviviridae* family, is an arthropod-borne pathogen responsible for a large disease burden worldwide. Estimates indicate that over 3 billion people are at risk of DENV infections worldwide with approximately 284–528 million new infections annually, 500,000 severe cases, and 20,000 deaths per year.^[1–3] These DENV infections primarily occur in tropical and subtropical climates, where the transmitting vector, the *Aedes aegypti* mosquito, is prevalent. DENV infections may develop into a haemorrhagic fever in severe cases of the disease. As there are

no efficient vaccines or antiviral treatments available for this disease, new treatment strategies and identification of potential drug candidates is urgently needed.^[2,4]

The DENV genome is composed of a positive-sense single-stranded RNA that is directly translated into a single polyprotein, in which its complex topology becomes embedded into endoplasmic reticulum (ER) membranes.^[5,6] The polyprotein cleavage occurs co- and post-translationally resulting in three structural proteins (capsid protein C, prM and protein E) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5).^[7]

Non-structural proteins are mostly multifunctional, and for example, NS3 contains both RNA helicase and RNA annealing activities. Structurally, NS3 is composed of a protease (NS3_{PRO}), a helicase (NS3_{HEL}) as well as an RNA triphosphatase domains, with the protease being responsible for cleavage of the polyprotein, together with cellular peptidases. The protease and helicase domains interact with the viral genome, potentially bridging the replication and capsid assembly steps for nucleocapsid formation.^[8–10] It is also well established that uncoupling the sequential order of these two proteolytic domains reduces the nucleocapsid incorporation into the virus particle. This indicates that polyprotein processing does regulate the viral assembly.^[11–13]

The NS3_{PRO} domain, herein referred to only as NS3 for short, belongs to the S7 family of serine proteases, with a conserved catalytic triad; His51, Asp75 and Ser135.^[14] In terms of mechanism, the reactive Ser135 triggers the proteolytic activity by generating a tetrahedral intermediate with His51, which is stabilized by the interaction with Asp75. This intermediate conformation then attacks the electrophilic atoms of the substrate, followed by the cleavage and release of substrate fragments.^[15]

It is worth mentioning that DENV's NS3, similarly to other flaviviruses proteases, requires a peptide cofactor (NS2B) to be


[a] Dr. T. Kronenberger,⁺ Prof. A. Poso
Department of Medical Oncology and Pneumology,
University Hospital of Tübingen
Otfried-Müller-Strasse 14, 72076 Tübingen
(Germany)
E-mail: kronenberger7@gmail.com


[b] Dr. T. Kronenberger,⁺ A. Kumar Tonduru, Prof. A. Poso
School of Pharmacy, University of Eastern Finland
Kuopio 70211 (Finland)

[c] M. Sá Magalhães Serafim⁺
Departamento de Microbiologia
Universidade Federal de Minas Gerais (UFMG)
Av. Antônio Carlos, 6627, Pampulha, CEP 31270-901
Belo Horizonte (Brazil)

[d] Dr. V. Gonçalves Maltarollo
Departamento de Produtos Farmacêuticos
Universidade Federal de Minas Gerais (UFMG)
Av. Antônio Carlos, 6627, Pampulha, CEP 31270-901
Belo Horizonte (Brazil)

[†] These authors contributed equally to this work.

 Supporting information for this article is available on the WWW under <https://doi.org/10.1002/cmdc.202100246>

 © 2021 The Authors. ChemMedChem published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

fully active. Structural studies have revealed that NS2B participates in the catalytically active conformation of the DENV protease, as a β -hairpin region wraps around the NS3_{PRO} core^[14] (Figure 1A illustrates this complex organization with the NS3 in grey and NS2B in cyan).

The NS3-NS2B complex prefers substrates with basic residues (Arg or Lys) at the P₁ and P₂ core sites, followed by amino acids with short side chains (Gly, Ser, or Ala) at the P₁' site^[16] (Figure 1). This preference can only be explained by the substrate interaction with residues from both NS3 and NS2B proteins, which underlies the relevance of their interaction.

Illustrating this fact, the inhibitor peptide NDL (Bz-Nle-Lys-Arg-Arg-H) has been employed to study the binding mode of covalently bound inhibitors in the NS3-NS2B protease from DENV-3 and WNV (Figure 1A,C). NDL covalently binds to Ser135 and it is further stabilized by a series of interactions between its positively charged moieties and residues in the NS3 (Figure 1C,D).

Interestingly, the NS3-NS2B complex exhibits either an open or closed configuration, which is believed to change upon ligand binding. For example, NDL in the crystal structure (3U1I) displays a closed conformation where the NS2B beta-hairpin

covers the active site (Figure 1A). However, the trypsin substrate-mimetic inhibitor aprotinin binds an open conformation of the NS3_{PRO} (3U1J),^[14] in which the NS2B hairpin does not cover the active site. This indicates the so-called "induced-fit" mechanism of catalysis, where substrates and substrate-mimicking inhibitors induce conformational changes in the protein leading to the active conformation.^[17] Differences between both conformations are not limited to changes in the NS2B conformation and previous comparative work discusses that associating "open" and "closed" to activity might be an oversimplification.^[18]

The crucial role of processing viral polyproteins into their final functional entities highlights this protease potential as a drug target.^[19] However, most of the DENV NS3-NS2B inhibitors described in the literature display high molecular weight peptide-mimetic structures with poor bioavailability,^[20,21] which prevents their clinical use.

To identify new NS3-NS2B inhibitors, it is essential to understand the motion of the protease domain as well as the drug-protein interactions. Here, by simulating the apostructure of NS3-NS2B like also both a covalently and non-covalently bound inhibitors (~10 μ s each), we aim to compare the conformational landscape of NS3 system. Analysis of the trajectories allowed us to clarify subtle differences in ligand binding profiles, highlighting the relevance of NS2B residues for its stabilization. Furthermore, we were able to identify putative druggable pockets for further drug discovery work.

Results

Structure selection and simulation setup

During the DENV maturation process, the N-terminal portion of NS3 is anchored in the vicinity of the endoplasmic reticulum and non-covalently interacting with the NS2B.^[22] In order to mimic this behaviour, several studies have utilized an artificial construct linking NS3 and NS2B via a glycine-rich loop to experimentally evaluate their activity.^[23] Although previously co-crystallized flavivirus' NS3_{PRO} structures are available, most of them are either incomplete or do present this artificial glycine-rich loop (indicated as GlyL, Table 1).

As a solution this issue decided to use homology modelling. We observed that structures from other viruses could be used as a template for the homology model. However, these structures either do not harbour an inhibitor (e.g. PDB: 2FOM, 2VBC and 2WHX) or harbour one not specifically bound to its catalytic site (e.g. 6MO0, 6MO1 and 6MO2), present mutations (e.g. 2WZQ, 3LKW, 4M9F, 4M9I and 4M9T) or were produced in different pH (e.g. 4M9K and 4M9M). In particular, DENV-3 protease bound to aprotinin (3U1J) does not form a closed conformation (β -hairpin region of NS2B wrapped around the NS3 core)^[14] and, therefore it was not selected.

The best structure for our purposes was 3U1I, as it was co-crystallized with a covalently bound inhibitor (NDL, Figure 1C,D) connected to Ser135. This structure has a closed conformation^[14] (Figure 1A), where the NS2B beta-sheet folds

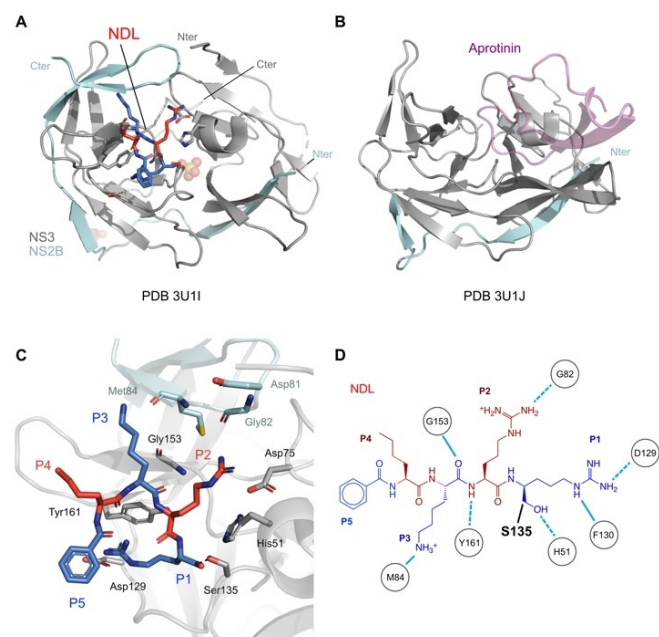


Figure 1. DENV-3 crystal structure illustrating the closed active conformation (PDB 3U1I) (A) and the open conformation (PDB 3U1J) (B), where the C-terminal of the NS2B cofactor becomes flexible and decouples from the active site. NS3 is represented as a grey cartoon, NS2B is represented as a cyan cartoon, the active site residues are highlighted as sticks and the covalently bound NDL inhibitor is depicted in red/blue. C) Inset of the active site bound to the covalent inhibitor NDL and its main interactions (PDB 3U1I). D) Example of peptide inhibitor NDL. From left to right, the following residues are represented Bz-Nle-Lys-Arg-Arg-H and numbered P₁–P₅, accordingly. In this representation residues displaying hydrogen bond interactions in the original NS3-NS2B-NDL co-crystallized structure (PDB 3U1I) are depicted as spheres and the hydrogen bonds with the main-chain amino acids are displayed as cyan lines, while interactions with the side chains are represented as dashed cyan lines. Ser135's covalent bond in the P₁ moiety is represented as a dark line.

Table 1. Crystallographic structures analysed for conformational changes, co-crystallized with peptides or peptide inhibitors. NS3_{PRO} refers to the protease domain of the NS3 protein, ignoring the helicase part of the protein, while NS2B_{CF} refers only to the hydrophilic loop of NS2B, which acts as a cofactor for the NS3_{PRO} activation.

Virus	PDB ID	Structural Information	Resolution (Å)
DENV-1	3L6P ^[a]	NS2B _{CF} -GlyL-NS3 _{PROΔ10}	2.20
DENV-2	2FOM	NS2B _{CF} -GlyL-NS3 _{PRO}	1.50
DENV-3	3U11 ^[b]	NS2B _{CF} -NS3 _{PRO} + NDL	2.30
DENV-3	3U1J	NS2B _{CF} -NS3 _{PRO} + Aprotinin	1.80
DENV-4	2WHX ^[c]	NS2B ₁₈ -GlyL-NS3 _{full}	2.20
DENV-4	2VBC ^[c]	NS2B ₁₈ -GlyL-NS3 _{full}	3.15
WNV	2FP7	NS2B _{CF} -GlyL-NS3 _{PRO} + NDL	1.68
WNV	2IJO	NS2B _{CF} -GlyL-NS3 _{PRO} + BPTI	2.30
WNV	3E90	NS2B _{CF} -GlyL-NS3 _{PRO} + NKK	2.45

[a] The construct has a 10 residues deletion in the N-terminus of NS3_{PRO}, to facilitate crystallization. [c] Missing NS3 Pro10-Ala16. [b] Instead of using the full hydrophilic loop from NS2B, a short sequence of 18 residues was used; this was done only to provide solubility to the expressed complex. NDL: Bz-Nle-Lys-Arg-Arg-H, BPTI: bovine pancreatic trypsin inhibitor, NKK: Naph-Lys-Lys-Arg-H, DENV: dengue virus, WNV: West Nile virus Bz-benzyl, Nle-norleucine, Naph-naphthalene

over the active site and is further stabilized by interactions at the P₄' to P₃ sites of the substrate. The NS3-NS2B structure 3U11^[14] was simulated in three different modes: i) with the NDL inhibitor covalently bound to Ser135, representing the inhibited state; ii) with the NDL in position for inhibition but non covalently bound; and iii) the apostructure with no inhibitor.

Inhibitor binding mode assessed by MD depends on the interaction with NS3 and NS2B

As an initial approach, we investigated the stability of the interactions between NS3-NS2B and the ligand along the entire MD trajectory. Slightly different interaction profiles (residues, bond types and frequencies) are observed when comparing the simulations with the non-covalently and covalently bound ligands (Figure 2A). In both systems, NDL is mainly stabilized by a pair of ionic interactions and hydrogen bonds between its basic side chains (P₁-P₂, Figure 2), and between the catalytic Asp75 (NS3) and Asp79 (NS2B) (Figure 2B). Simulations with the non-covalently bound ligand also features a prominent interaction of the P₂'s arginine moiety with Asp81 from the NS2B (Figure 2B).

Further, simulations with covalently bound ligand reveal that the ligand is stabilized by an interaction between the hydrogen of P₂ main-chain and the side chains of Asn152 and His51, as well as π - π interactions and hydrophobic contacts between the P₄-P₅ moieties and the Tyr161 residue (Figure 2A,C,D). These interactions are slightly different from the ones originally proposed in the crystal structure, which highlights the possibility of multiple binding modes for the ligand. Our data also suggests that Asn152 stabilizes the hydroxyl group of non-covalently bound peptide by forming a hydrogen bond (Figure 2A,B), and the hydrogen bond becomes more prevalent in the covalently bound systems (Figures 2C,D). It is worth mentioning that Asn152Ala NS3 mutants remain active in terms of proteolysis, but exhibit a significant decrease in their binding affinity for some ligands,^[24] which supports the relevance of this residue for stabilization even if not directly involved in the catalytic reaction.

This interaction profile is in agreement with the enzyme's preference for basic moieties in the P₁/P₂, as a result of its interactions with the Asp75 (NS3, interacting with the P₁) and Asp129 (NS2B, with P₂).^[24] These observations are consistent with the results of Kouretova *et al.*, (2017). It has been postulated that effective inhibitors in biochemical assays would therefore be positively charged, which in turn would lead to poor *in vivo* bioavailability.

To a further investigation, whether these differences in the ligand interaction were a result from the conformation changes in the protein, we decided to analyse the MD trajectories using principal component analyses.

Conformational changes in the NS3-NS2B dynamics upon ligand binding

It is known from the available crystal structures (Supporting Information, Figure S2) that, in the absence of a substrate or active site inhibitor, the C-terminal portion of NS2B can adopt an "open" conformation and become disordered.^[25-27] This conformational change in NS2B after ligand binding has been described by short MD simulations and by several NMR studies using a linked construct.^[25-27] After substrate binding to the NS3 active site, the C-terminal portion of NS2B interacts with the NS3 core, closing the NS3 active site with the so-called active "closed" conformation. However, in our simulations, we observed that the protease can adopt four major states (based on the K-means clusterization, Supporting Information: Figure S3): State I – a tightly closed state which is exclusive of the system simulated with the covalently bound ligand (Figure 3A); State II – a semi-closed state where the NS3's N-terminal is unfolded (Figure 3B); State III – a second semi-closed state where NS3's N-terminal loses the beta-sheet fold but remains near the core protein (Figure 3C); and State IV – a fully open state, where the NS2B's C-terminal uncovers the active site (Figure 3D). With exception of state I, which seems almost exclusive for systems with the covalently bound ligand, the occurrence of the states in the simulations happens independently from the presence of ligand binding, though with different frequencies (Figure 3A-D

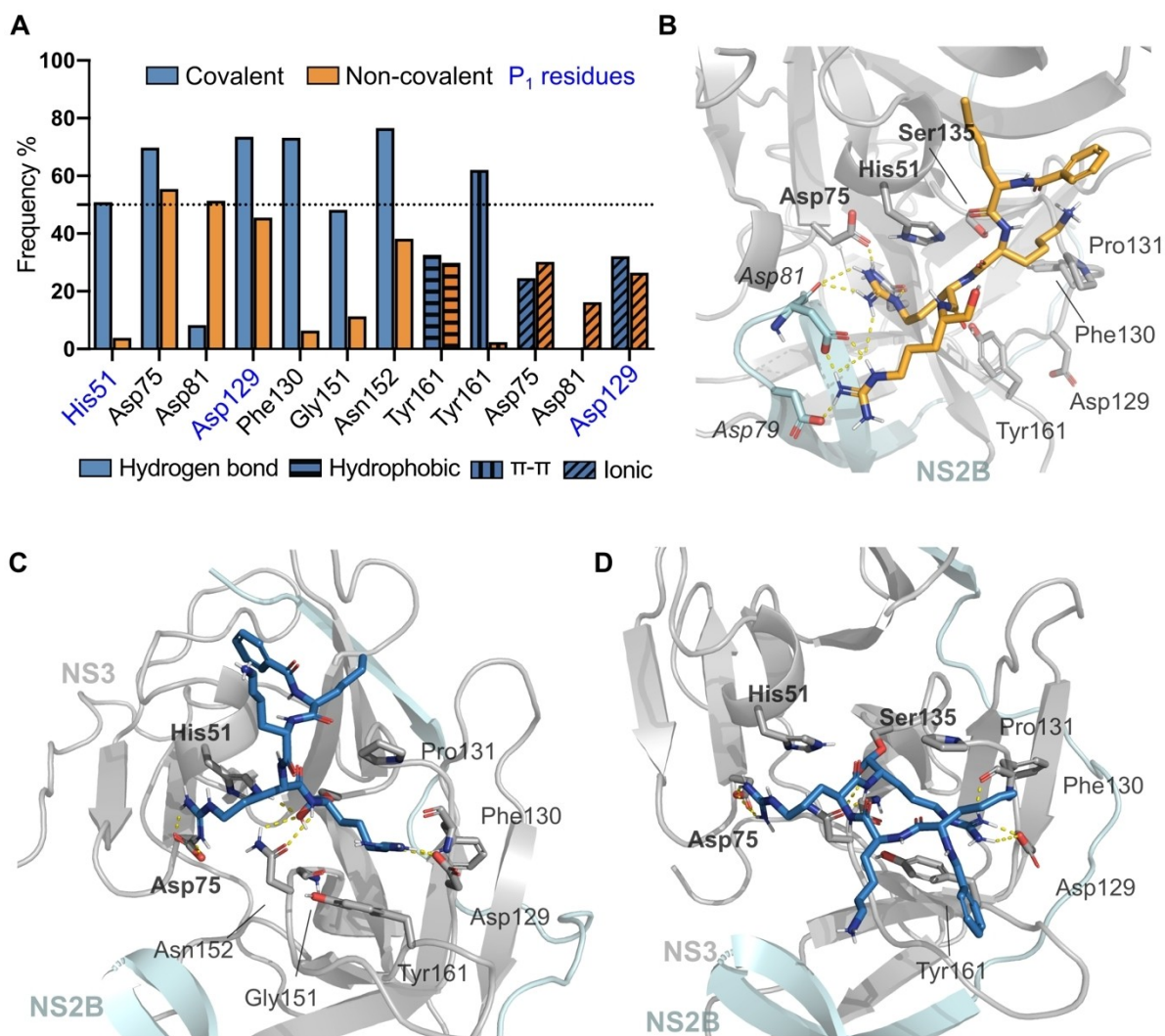


Figure 2. Interactions of DENV-3 protease with consensus peptide, both covalently and non-covalently bound. A) Interaction pattern comparing simulations with the covalently bound peptide against the non-covalently bound peptide, bars represent the frequency of interaction during the simulation. Residues belonging to the P₁ site are labelled in blue. Other P-sites were not significantly enriched in the simulations. The non-covalently bound ligand mainly manifests a single stable conformation (B), while the covalently bound system explores two different binding modes (C,D). Dashed yellow lines designate polar interactions. The protease's residues are coloured according to the atom types (protein carbons, light grey for the NS3 and cyan for the NS2B; nitrogen, blue; oxygen, red).

and Supporting Information Figure S3). These conformations are distinct, where the closed-ones are stabilized by the interaction between the negatively charged amino acids of NS2B's C-terminal loop, and the NS3's lysine residues (Lys73 and Lys74, Figure 3E).

In a non-covalently bound system, the active site covering the beta-sheet lid of the C-terminal region of NS2B (highlighted in purple in Figure 4E), has a lower root mean square fluctuation (RMSF) in comparison to the simulations with the apostructure and covalently bound ligands (Figure 4A–E). According to the PCA analysis, the high flexibility in this region is in agreement with the opening and exposure of the active site (Figure 4F).

Assessment of binding pockets and sites occupancy: druggability potential

Our microsecond-scale MD simulations make it possible to study transient binding site properties. We decided to utilize SiteMap analyses along the simulated trajectories to identify changes in the average site volume (Figure 5A) between the states. The SiteMap can also predict the potential druggability of the pockets, ranking them by the SiteScore values. These values range from 0–1.5, with values higher than 0.8 suggesting druggable pockets. Interestingly, the differences in volume seem to inversely affect the predicted site's druggability, where more compact sites had higher SiteScore values (Figure 5B). Accordingly, State II and IV display a more accessible and

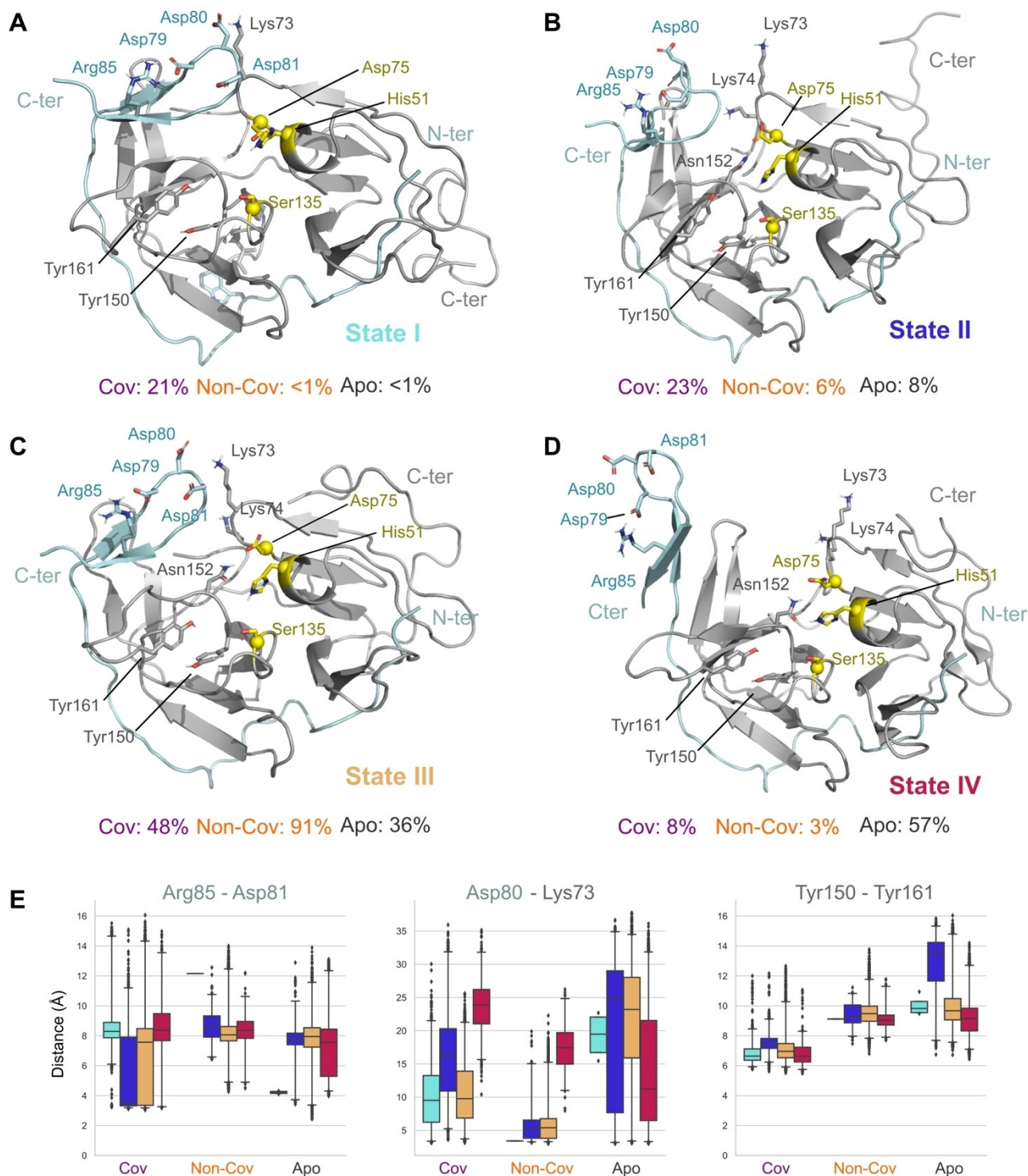


Figure 3. DENV-3 protease conformational states. A) State I, a tightly closed state, is only seen with the ligand covalently bound system. B) State II, a semi-closed state, where the NS3's N-terminal is unfolded. C) State III, the semi-closed state where NS3's N-terminal is losing the beta-sheet fold, but remains near the core protein D) State IV, fully open with the NS2B's C-terminal uncovers the active site. The frequency of each conformation for covalent bound ligand (Cov), non-covalent bound ligand (Non-Cov) and apostructure (Apo) is shown below the structural representation. The grey cartoon represents the NS3 protease domain and light blue represents the NS2B, yellow spheres represent the position of the active site and relevant residues are depicted as sticks. E) box plot representation of the distance variation along the analysed trajectory and plotted by each cluster for relevant amino-acid pairs. Lys73-Asp80 was chosen to represent the charge-clamp holding the NS2B in the closed conformation. Arg85-Asp81 distance variation represents the NS2B's hairpin bending and conformational changes and Tyr150-Tyr161 describes hydrophobic pair relevant for the NS3 folding.

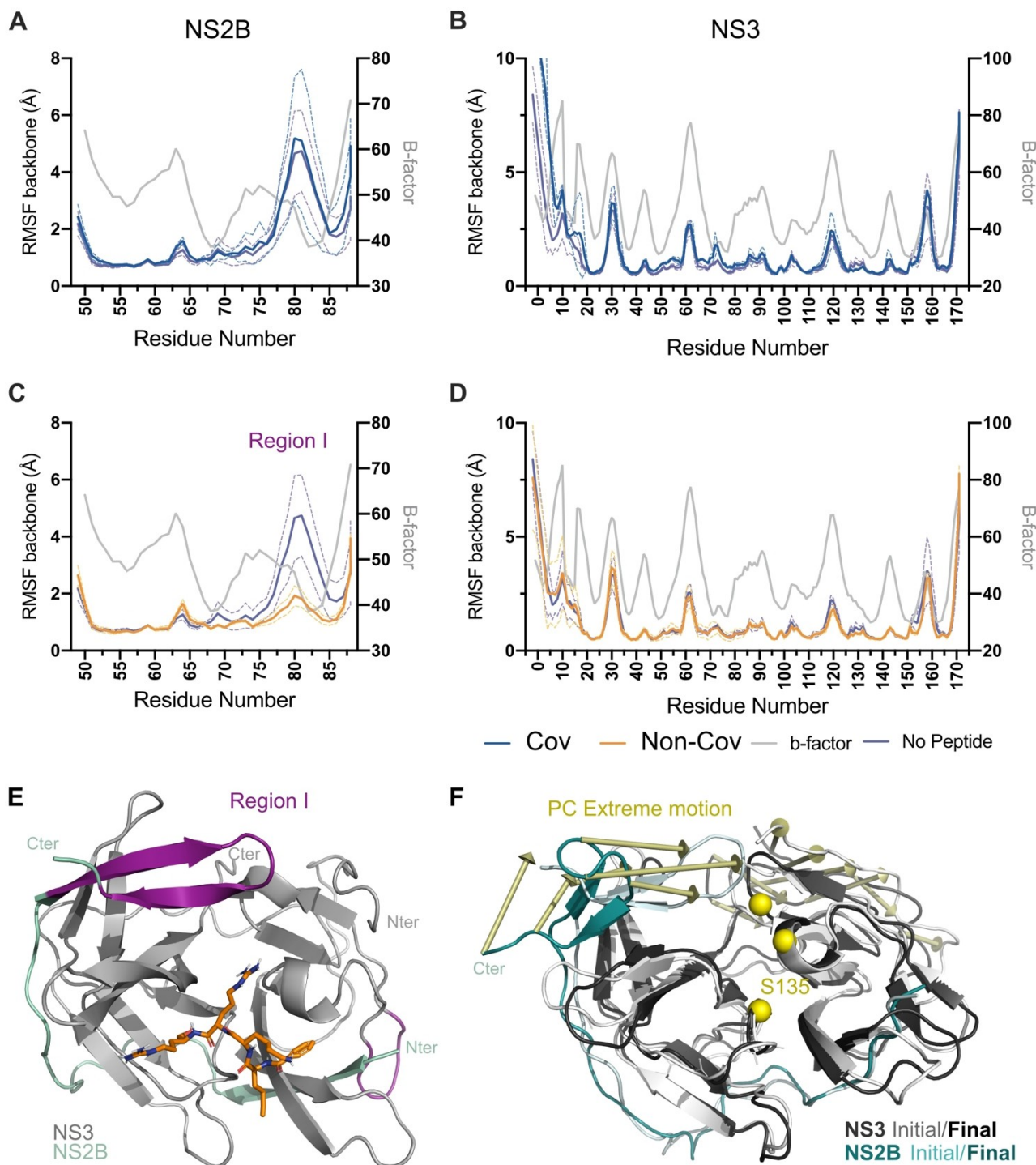


Figure 4. DENV protease core (NS3) mobility influences the interaction partner (NS2B) and vice-versa. Comparison between the protein b-factor from the crystal structure DENV-3 protease (3U11, in grey on the right axis) and the average residue fluctuations obtained from the RMSF of DENV's backbone atoms. RMSF values were compared between the average of the replicates of the system with (blue) and without peptide (purple, in the left axis). Dashed lines represent the standard deviating limits for the NS2B (A) and NS3 (B), while orange lines represent results from simulations with non-covalently bound ligand (C, NS2B) and (D, NS3). E) The region that deviates in the presence of the peptide are highlighted in purple in the NS2B co-activator beta-sheet between (residues 75 and 85). F) The extreme movements of PC1 common for systems in the presence of ligands. In E) and F) NS3 is coloured as grey/black and NS2B is coloured as cyan, while golden arrows indicate the extreme movements related to each principal component. Yellow spheres represent the position of the catalytic triad.

exposed surface, resulting in lower SiteScore values for these states.

We then proceeded to verify pockets using representative conformations for each state/ligand combination using the

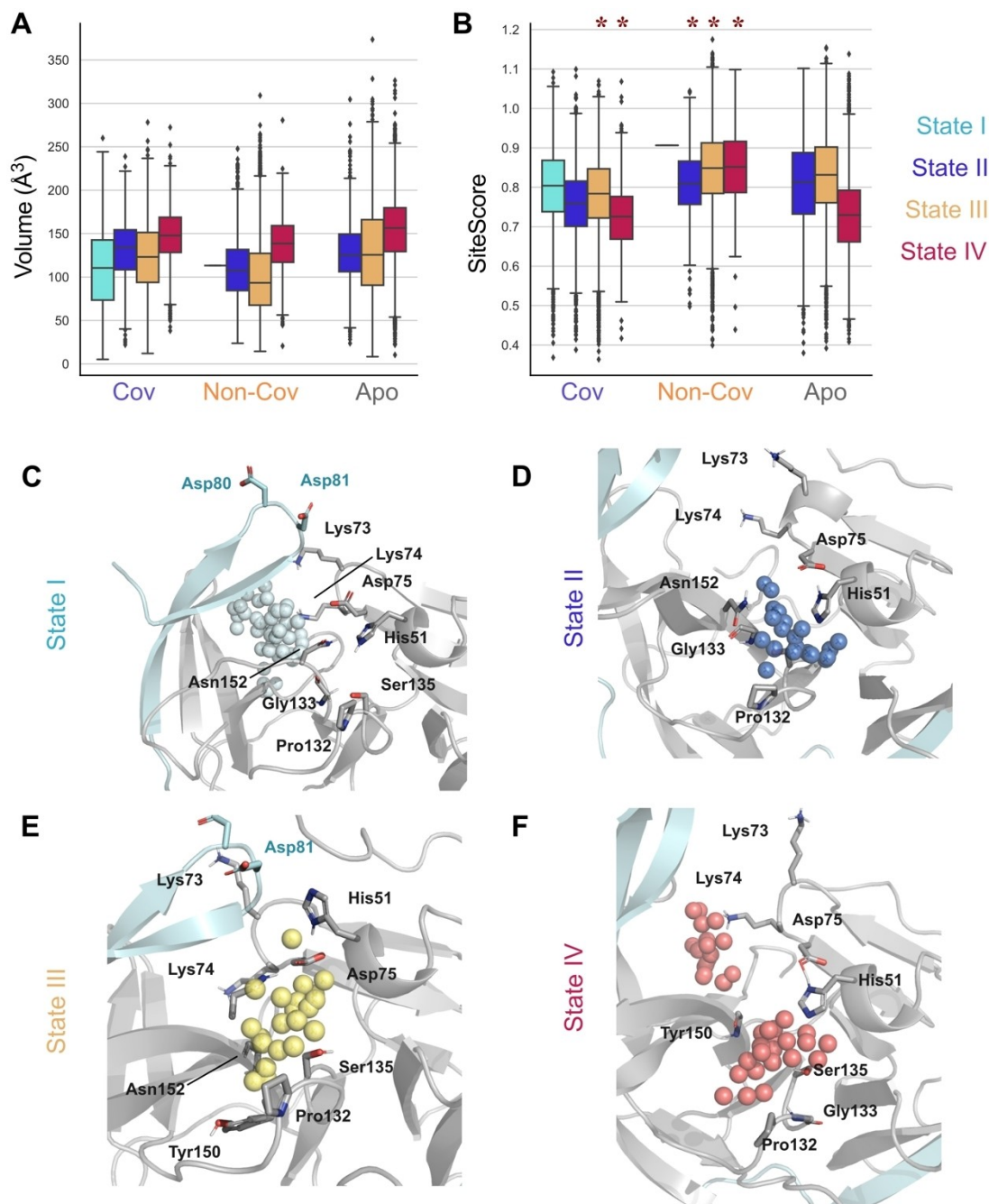


Figure 5. Large variation in the catalytic site volume influences its druggability and most prominent druggable region. DENV-3's NS2B anionic pocket. Box plot representation of the volume (A) and druggability (B) along the analysed trajectory (plotted by each state cluster). Druggability variation for the active site pocket was assessed by the SiteScore calculated with the SiteMap. Solid lines represent the average and the boxes amplitude the quartiles, and outliers are highlighted, red asterisks represent significance in comparison to the respective apostructure distribution, as calculated by Mann-Whitney test, p -value < 0.01 is represented by *. The identified pockets from the PrankWeb server and SiteMap analyses, represented as spheres (C–F) for the different protein states discussed on the PCA. Amino acid residues found in common between the different predictions are depicted as sticks.

website PrankWeb. This server searches for potential ligand-binding points on a protein surface, predicting the chemical propensity of these points. A feature vector is calculated by physicochemical and geometric properties of surrounding atoms and residues, which are then clustered, resulting in an individual ranking score.^[28] The PrankWeb server predictions varied among the clusters and conformations, but at least one

pocket was predicted to all states evaluated (Table 2). The results with the P2Rank method from PrankWeb also showed higher scores for covalent bound states. Herein they were visually depicted by one representative conformation of each state (I, II, III and IV, represented as spheres in Figure 5C–F). Pockets with higher scores were shown for the covalent (1.57 to

Table 2. Potential pocket's scores and binding sites predicted by the P2Rank method in PrankWeb to the three different modes assessed (apostructure, covalent bond to Ser135 and non-covalently bound). Ligands were manually removed, and each conformation state was individually evaluated.

Conformation state	Best predicted pocket score	Total of potential binding sites (residues)
Apostructure (State IV)	4.54	n = 37
Apostructure (State II)	2.37	n = 16
Apostructure (State I)	2.59	n = 12
Covalent (State IV)	1.57	n = 16
Covalent (State III)	3.71	n = 11
Covalent (State II)	2.49	n = 14
Covalent (State I)	7.98	n = 24
Non-covalently bound (State IV)	5.22	n = 24
Non-covalently bound (State III)	5.08	n = 30
Non-covalently bound (State II)	3.66	n = 20
Non-covalently bound (State I)	2.02	n = 11

7.98) and non-covalently bound (2.02 to 5.22) in comparison to the apostructure states (2.37 to 4.54).

In addition to the importance in the states' differences, pocket analyses identified similar concomitant residues predicted as potential binding sites, including Asn152 for all states evaluated; His51, Pro132, Gly133, Ser135, Gly151 and Asn152 for all apostructure states; Asn152 for all non-covalently bound states; and Ile123, Asn152, Ile165 and Ala166 for all covalent states (Supporting Information, Table 1).

Taken together, these results suggest that the druggable active site presents a unique residue (Asn152) shared among all conformational states. Also, this binding site is not found directly in the covalent or apostructure protein states. Residue Asn152 is a potential target in the protease active site, as it was conserved between the open and closed states, and corresponding P2Rank scores were similar in between all binding modes and protease conformations. Even so, it is important to mention that the structural accessibility of an inhibitor to the active site can also play a key role in the complex.^[17] To the best of our knowledge, this is the first time that this approach has been applied to comparatively evaluate the pocket sites that have importance to the different DENV-3 protease conformations.

Discussion

The NS3 proteases are highly flexible proteins, whose flexibility is often underestimated when only a single crystal structure is analysed. Especially the DENV's NS3 can adopt multiple substrate/inhibitor specific conformational states. Here we have used extensive MD simulations to study the DENV-3 NS3-NS2B complex, as a model for DENV proteases, to identify conformational changes occurring in the substrate/inhibitor binding. We sampled the conformational landscape by conducting simulations with a ligand (NDL) covalently and non-covalently bound to the catalytic Ser135 and compared these results with apostructure simulations. While there are clear differences

between each system, the most striking feature is the identification of distinct protein conformations.

From the start, the detected interplay between the NDL ligand interaction with residues from both NS3-NS2B and protein changes upon the ligand binding led us to systematically investigate conformational changes in the complex. We used a combination of principal component analyses and K-means clustering to classify the conformations explored during the simulation into four main States (I–IV). These states structurally vary in the NS2B conformation (open or closed), due to changes in the interaction between the NS2B's C-terminal beta-hairpin (Asp80 and Asp81), rich in negatively charged amino acids, and the NS3's lysine residues (Lys73 and Lys74). Open conformations where the NS2B hairpin is misplaced or unfolded have been previously observed in other crystal structures (Supporting Information, Figure S2).

Previously reported simulations conducted with the NS3-NS2B from WNV,^[29,30] starting from an initial open conformation, reported the movement of NS2B hairpin's closing as an irreversible process. Interestingly, the WNV open conformation displayed an improperly formed oxyanion hole (Thr132-Gly133), which flipped towards the Ser135, concomitantly with the closing of the NS2B hairpin. This resulted in the formation of the catalytically competent oxyanion hole in a closed conformation, which then remained stable until the end of the simulations (of around 80 ns).^[30] This observation is consistent with the mechanism of action of the catalytic triad in the flaviviruses' proteases, where the reactive serine oxyanion generates a tetrahedral intermediate necessary for the substrate cleavage.^[15] In agreement with that observation, the work of De Almeida *et al.*, (2013) also detected the involvement of a hydrogen bond with residue Gly133 between its main chain atoms and P₁ and P₂ in all conformations evaluated, and with the same NS3 structure assessed here (3U11).^[31]

We then proceeded on studying the potential druggability of these new conformations using both SiteMap and PrankWeb server. Interestingly the closed states (I and III) have, on average, smaller pocket volumes and higher druggability values. Studying the amino acid composition of the overlapping pockets suggests that the involvement of NS2B residues increases its druggability. The site prediction analyses also detected conserved ligand-binding sites on the protein's surface, extending it beyond the classical active site (*e.g.* Asn152). In PrankWeb, data showed only one residue (Asn152) concomitant to all conformation states, as well as specific residues concomitant to all covalent states (*e.g.* Ile123), and for the apostructure states (*e.g.* Gly133 and Gly151) (Supporting Information, Table 1). These results also corroborate previous findings which highlight Lys73, Lys74, Ile123 and Asn152,^[32] as well as Gly151 and Gly153,^[17] Gly133,^[31] and Asp75 as important binding sites on the protein's surface and are also consistent with the WNV protease interaction previously described.^[33,34] The work from Behnam *et al.* (2020), comparing different NS3-NS2B protease structures, describes the so-called subpocket B that extends beyond His51/Asp75 connecting it with the P₂ pocket.^[18]

Similarly, the work from De Almeida *et al.*, (2013) aimed to account for the possible influence of N- and C-terminus domains in the NS3 protease's active site in terms of conformational changes and/or substrate accessibility. They studied short simulations (30 ns) of a full-length homology model, whose conformations were then clustered into four families.^[31] However, due to the short simulation time, one can speculate whether those conformational differences were due to the initial system equilibration or whether the authors had underestimated the possibility of an even more flexible system.

Brecher *et al.*, (2016,^[24]) postulated the presence of an allosteric pocket in DENV-2 protease, composed of NS2B's 67–95 hairpin loop and the NS3_{PRO} 117–120 loop. This pocket would exist only in the closed-active conformation when the NS2B would be less flexible. In the open-inactive conformation, the active site would behave as a cleft, composed of NS3 residues such as Val146, Gly148, Leu149, Tyr150, Gly151 and Asn152, without support from the NS2B residues, which requires to be induced to a closed conformation to bind peptide inhibitors.

Based on our results, we suggest that DENV-3 protease would be a difficult target for classically designed drug-like small molecules, which would bind to a single protein conformation. One of the issues is the low druggability of open states, such as the State IV. Although the open conformation in State IV makes the active site more accessible, drug binding is hindered due to the strong desolvation effect. On the other hand, the more constrained States II and III are not very much better, as binding sites are too small for drug-like compounds. We believe that the best alternative is to have a covalent-binding, resembling the State I conformation. This is probably only possible if the initial binding occurs at State IV and, after complex formation, initiates conformational change to a more compact State.

The search of relevant conformations, among the heterogeneous set of structures available in the literature, have substantial implications on the understanding of this enzyme.^[18] We suggest that understanding this conformational shift is key for further drug discovery, since even induced-fit and flexible docking would be only relevant for fine tuning.

It would be also interesting to have a crystal structure of DENV protease bound to a lead compound with interactions in both P sites, since this would help in the further optimization in the design of inhibitors, especially considering P₂ and P₃ moieties, such as those available in this same structure (3U1I).^[35] For instance, Yao *et al.*, (2019)^[36] proposed pan-flavivirus inhibitors binding to an allosteric pocket of NS3. Their lead compound held the protease in a catalytically inactive open conformation, as confirmed by crystallization. This novel type of inhibitor could be expanded to allow more interactions with the NS2B cofactor, which may lead to conformational changes in the NS3-NS2B complex. Finally, we hope that our current study would enable the expansion of the DENV protease inhibitor discovery by providing novel interesting conformations to target as well as the dynamic description of the open-closing transition.

Conclusion

The DENV's NS3 protease crystal structures often underestimate its high flexibility, as this protein can adopt multiple conformational states depending on the ligand bound. Here we expand on the conformational changes occurring in the DENV-3 NS3-NS2B complex obtained from a long-timescale molecular dynamics simulation. We sampled the ligand-binding site conformational landscape by conducting MD simulations with the non- and covalently bound ligand states, as well as the apostructure.

We assessed which residues would compose potential druggable binding sites (Tyr161-NS3, Asp75 and Asp129, in the NS2B), in addition to evaluating which residues were involved in the establishment, stability and activity of the protease complex (e.g. Asn152). We observed that the interaction with Asp81 (NS2B) can be relevant for the recognition of the inhibitor and the transition towards the catalytically active closed conformation. This emphasizes the importance of the co-factor NS2B in the conformational switch that leads to activity.

Furthermore, we propose that inhibitors targeting the open NS3-NS2B conformations could benefit from the higher accessibility, despite the poorer druggability of these sites. Complementarily, inhibitors inducing closing conformational changes would be relevant for the protease inhibition.

Experimental Section

The NS3-NS2B protease bound to a covalent inhibitor (the ND1 inhibitor found in the PDB ID: 3U1I^[14]) was simulated in three different modes: i) with the ND1 inhibitor covalently bound to Ser135, representing the inhibited state; ii) with the ND1 in position for inhibition but non covalently bound; and iii) the apostructure with no inhibitor. This study aimed to identify different protease conformations in each system and evaluate their druggability potential for the development of new inhibitors.

Structure selection and system preparation

DENV-3 protease (wild-type PDB ID: 3U1I^[14]) was selected, among other available structures, for the molecular dynamics simulations based on the resolution (2.3 Å) and the presence of a peptide inhibitor (ND1), Bz-Nle-Lys-Arg-Arg-H, where Bz represents a benzyl group and Nle stands for norleucine.

The selected structure was prepared using Epik,^[37] by adjusting the ionization states of amino acid residues, especially taking into account the catalytic triad (Ser135, His51 and Asp75), in addition to fixing missing side chain atoms, and then further subjecting it to restrained energy minimization (PrepWiz, Maestro v2019.1). The missing segments Ser1-Lys15 from NS3PRO were modelled using Prime^[38] with standard options. Incomplete sequence N- and C-termini were capped and sulphate ions from the crystallization buffer were removed. Water molecules with more than two hydrogen bonds were retained. For the covalently bound system simulations, the bond with Ser135 was maintained. Non-covalently bound systems were generated by breaking the covalent bond with Ser135. Final structure was minimized using the PrepWizard using the OPLS3e force-field, with movements restrained to the maximum of 0.3 Å variation.

Molecular dynamics simulation

MD simulations were carried out using the Desmond engine^[39] with the OPLS3e force-field.^[40] The simulated system encompassed the protein-ligand complex, a predefined water model (TIP3P^[41]) as explicit solvent and counter-ions (Na⁺ or Cl⁻ adjusted to neutralize the overall system charge, around 4 or 5 Na⁺ atoms depending on the system). The entire system was treated in a cubic box with periodic boundary conditions, specifying the shape and the size of the box as 13 Å distance from the box edges to any atom of the protein (totalizing around 40,000 atoms between protein, ligand, solvent and ions). We used a time step of 1 fs and the short-range coulombic interactions were treated using a cut-off value of 9.0 Å, while the smooth Particle Mesh Ewald method (PME) handled the long-range coulombic interactions.^[42]

Initially, the system was minimized using Steepest Descent, followed by equilibration. All simulations were run in NPT ensemble (T = 310 K, Nosé-Hoover method; p = 1.01325 bar, Martyna-Tobias-Klein method) with default Desmond settings. Reversible reference system propagator algorithms (RESPA) integrator with 2 fs, 2 fs and 6 fs timesteps were used for bonded, near and far, respectively. A constant temperature of 310 K was maintained throughout the simulation using the Nose-Hoover thermostat algorithm, and the Martyna-Tobias-Klein Barostat algorithm was used to maintain 1 atm of pressure. After minimization and relaxation of the system, we proceeded with an independent production step of at least 1 μs for each system, with some particular systems being extended after evaluation. All MD simulations were performed at least in five independent runs with randomly generated seeds. In total, we studied 12 μs for DENV protease covalently bound to its ligand, 9 μs for the system with unbound ligand and 10 μs for the apostructure simulations. The representative structural frames were selected based on a visual inspection of the conserved interactions and principal component analyses (PCA) followed by cluster analysis.

Simulation trajectory analyses and principal component analyses (PCA)

In the PCA, the backbone of each frame was extracted and aligned using `trj_selection_dl.py` and `trj_align.py` scripts from Schrodinger (2019v4). Individual simulations from all the runs were merged using `trj_merge.py` into a final trajectory and CMS files, which were further used for the generation of the principal components. PCA was carried out employing protein C-alpha atoms as variables using `trj_essential_dynamics.py` script. The first principal component was found to explain 20.4%, the second 15.3% and the third 12.6%, of the total information (48.3%).

By using the first three principal components results from the previous analysis, we carried out K-means clustering analysis employing the KNIME analytics platform.^[43] Initially, aiming to define the optimal number of clusters, the Elbow method (also known as scree-plot) was carried out.^[44] In this step, ten clustering analysis were generated varying the number of clusters from 1 to 10. Then, the sum of squared distances between samples (frames from MD simulations) and the centroid of the defined cluster was calculated for each run. Then, the optimal number of clusters was defined by the point with a drastic decrease in the distance variance (therefore, "elbow-like" curve, as displayed in the scree-plot). Finally, a K-means analysis was carried out using the obtained optimal number of clusters.

Interactions and distances were determined using the Simulation Event Analysis pipeline implemented in Maestro (Maestro 2019v4) with standard bond definitions. Molecular dynamics trajectories

were visualized, and figures were produced using PyMOL (Schrödinger LCC, version 2.4). Trajectories and interaction data are available in the Zenodo repository (under the code: 10.5281/zenodo.3732941).

Identification of potential druggable Pockets

We used SiteMap^[45] to analyse the structures representing different conformational states of the DENV protease, aiming to identify and evaluate potential ligand-binding sites. SiteMap works by creating a grid of points separated by 1 Å. Each of those points is evaluated to check for one of the three cases: i) if they overlap with any protein atoms, ii) if they are insufficiently enclosed by the protein or iii) if they have a too small van der Waals interaction energy with the protein. If so, in any of those cases, this point is not considered for further analyses. The remaining points are then grouped into sites, and the enclosure, size, and hydrophilicity of the sites are calculated. These values are combined into a single scoring function called the SiteScore, that ranges from 0 to 1.5 (most likely to be druggable). SiteScore values higher than 0.8 were previously associated with druggable pockets.^[45]

Finally, a P2Rank analysis on the webserver PrankWeb was performed to compare different conformations states and map binding residues.^[28,46] Briefly, the method consists in a machine learning (ML) technique to predict binding sites. Employing a random forest (RF), the method predicts a chemical propensity bind of points (potential locations of binding ligand contact atoms) on a solvent-accessible protein surface. A feature vector describes these locations, which consists of physicochemical and geometric properties of surrounding atoms and residues (e.g. aromaticity and hydrophobicity). High propensity points are then clustered (i.e. pockets), and ranked based on the cumulative score of the cluster. Here, covalent and non-covalently bound conformations' ligands were manually removed before P2Rank analyses.

Acknowledgements

This work was supported by the UFMG intramural funds, CAPES (88887.595578/2020-00), CNPq (132732/2018-1) and FAPEMIG (APQ-03058-18) for the fellowship of M.S.M.S. The authors thank CSC – Finland for the very generous computational resources provided. The authors also thank Ewen MacDonald for critical reading and language corrections. Open access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: Dengue virus · Molecular dynamics simulation · NS3-NS2B protease

[1] "Dengue control. WHO | Epidemiology," can be found under <http://www.who.int/denguecontrol/epidemiology/en/> 2016, accessed on February 2021.

[2] S. Bhatt, P. W. Gething, O. J. Brady, J. P. Messina, A. W. Farlow, C. L. Moyes, J. M. Drake, J. S. Brownstein, A. G. Hoen, O. Sankoh, M. F. Myers, D. B. George, T. Jaenisch, G. R. W. Wint, C. P. Simmons, T. W. Scott, J. J. Farrar, S. I. Hay, *Nature* 2013, 496, 504–507.

- [3] J. D. Stanaway, D. S. Shepard, E. A. Undurraga, Y. A. Halasa, L. E. Coffeng, O. J. Brady, S. I. Hay, N. Bedi, I. M. Bensenor, C. A. Castañeda-Orjuela, T.-W. Chuang, K. B. Gibney, Z. A. Memish, A. Rafay, K. N. Ukwaja, N. Yonemoto, C. J. L. Murray, *Lancet Infect. Dis.* **2016**, *16*, 712–723.
- [4] A. Wilder-Smith, E.-E. Ooi, O. Horstick, B. Wills, *The Lancet.* **2019**, *393*, 350–363.
- [5] X. Xie, S. Gayen, C. Kang, Z. Yuan, P.-Y. Shi, *J. Virol.* **2013**, *87*, 4609–4622.
- [6] S. Miller, S. Kastner, J. Krijnse-Locker, S. Bühler, R. Bartenschlager, *J. Biol. Chem.* **2007**, *282*, 8873–8882.
- [7] B. D. Lindenbach, H.-J. Thiel, C. M. Rice, *Fields Virology* **2006**, *52*, 1101–1152.
- [8] G. P. Pijlman, N. Kondratieva, A. A. Khromykh, *J. Virol.* **2006**, *80*, 11255–11264.
- [9] C. G. Patkar, R. J. Kuhn, *J. Virol.* **2008**, *82*, 3342–3352.
- [10] L. G. Gebhard, N. G. Iglesias, L. A. Byk, C. V. Filomatori, F. A. De Maio, A. V. Gamarnik, *J. Virol.* **2016**, *90*, 5451–5461.
- [11] S. M. Amberg, C. M. Rice, *J. Virol.* **1999**, *73*, 8083–8094.
- [12] E. Lee, C. E. Stocks, S. M. Amberg, C. M. Rice, M. Lobigs, *J. Virol.* **2000**, *74*, 24–32.
- [13] M. Lobigs, E. Lee, *J. Virol.* **2004**, *78*, 178–186.
- [14] C. G. Noble, C. C. Seh, A. T. Chao, P. Y. Shi, *J. Virol.* **2012**, *86*, 438–446.
- [15] A. K. Timiri, B. N. Sinha, V. Jayaprakash, *Eur. J. Med. Chem.* **2016**, *117*, 125–143.
- [16] I. E. Gouvea, M. A. Izidoro, W. a. S. Judice, M. H. S. Cezari, G. Caliendo, V. Santagada, C. N. D. dos Santos, M. H. Queiroz, M. A. Juliano, P. R. Young, D. P. Fairlie, L. Juliano, *Arch. Biochem. Biophys.* **2007**, *457*, 187–196.
- [17] A. E. Aleshin, S. A. Shiryayev, A. Y. Strongin, R. C. Liddington, *Protein Sci.* **2007**, *16*, 795–806.
- [18] M. A. M. Behnam, C. D. P. Klein, *Biochimie* **2020**, *174*, 117–125.
- [19] H. Steuber, R. Hilgenfeld, *Curr. Top. Med. Chem.* **2010**, *10*, 323–345.
- [20] Y.-S. Tian, Y. Zhou, T. Takagi, M. Kameoka, N. Kawashita, *Chem. Pharm. Bull.* **2018**, *66*, 191–206.
- [21] Q.-Y. Wang, S. J. Patel, E. Vangrevelinghe, H. Y. Xu, R. Rao, D. Jaber, W. Schul, F. Gu, O. Heudi, N. L. Ma, M. K. Poh, W. Y. Phong, T. H. Keller, E. Jacoby, S. G. Vasudevan, *Antimicrob. Agents Chemother.* **2009**, *53*, 1823–1831.
- [22] D. Luo, S. G. Vasudevan, J. Lescar, *Antiviral Res.* **2015**, *118*, 148–158.
- [23] W. W. Phoo, A. El Sahili, Z. Zhang, M. W. Chen, C. W. Liew, J. Lescar, S. G. Vasudevan, D. Luo, *Antiviral Res.* **2020**, *182*, 104900.
- [24] M. Brecher, Z. Li, B. Liu, J. Zhang, C. A. Koetzner, A. Alifarag, S. A. Jones, Q. Lin, L. D. Kramer, H. Li, *PLoS Pathog.* **2017**, *13*, e1006411.
- [25] X.-C. Su, K. Ozawa, R. Qi, S. G. Vasudevan, S. P. Lim, G. Otting, *PLoS Neglected Trop. Dis.* **2009**, *3*, e561.
- [26] X.-C. Su, K. Ozawa, H. Yagi, S. P. Lim, D. Wen, D. Ekonomiuk, D. Huang, T. H. Keller, S. Sonntag, A. Caffisch, S. G. Vasudevan, G. Otting, *FEBS J.* **2009**, *276*, 4244–4255.
- [27] L. Zhu, J. Yang, H. Li, H. Sun, J. Liu, J. Wang, *Biochem. Biophys. Res. Commun.* **2015**, *461*, 677–680.
- [28] L. Jendele, R. Krivak, P. Skoda, M. Novotny, D. Hoksza, *Nucleic Acids Res.* **2019**, *47*, W345–W349.
- [29] C. Kang, S. Gayen, W. Wang, R. Severin, A. S. Chen, H. A. Lim, C. S. B. Chia, A. Schüller, D. N. P. Doan, A. Poulsen, J. Hill, S. G. Vasudevan, T. H. Keller, *Antiviral Res.* **2013**, *97*, 137–144.
- [30] D. Ekonomiuk, A. Caffisch, *Protein Sci.* **2009**, *18*, 1003–1011.
- [31] H. de Almeida, I. M. D. Bastos, B. M. Ribeiro, B. Maigret, J. M. Santana, *PLoS One* **2013**, *8*, e72402.
- [32] H. Wu, S. Bock, M. Snitko, T. Berger, T. Weidner, S. Holloway, M. Kanitz, W. E. Diederich, H. Steuber, C. Walter, D. Hofmann, B. Weißbrich, R. Spannaus, E. G. Acosta, R. Bartenschlager, B. Engels, T. Schirmeister, J. Bodem, *Antimicrob. Agents Chemother.* **2015**, *59*, 1100–1109.
- [33] P. Erbel, N. Schiering, A. D'Arcy, M. Rénatus, M. Kroemer, S. P. Lim, Z. Yin, T. H. Keller, S. G. Vasudevan, U. Hommel, *Nat. Struct. Mol. Biol.* **2006**, *13*, 372–373.
- [34] C. A. Leonel, W. G. Lima, M. dos Santos, A. C. Ferraz, A. G. Taranto, J. C. de Magalhães, L. L. dos Santos, J. M. S. Ferreira, *Arch. Virol.* **2018**, *163*, 575–586.
- [35] K.-H. Lin, A. Ali, L. Rusere, D. I. Soumana, N. K. Yilmaz, C. A. Schiffer, *J. Virol.* **2017**, *91*, DOI 10.1128/JVI.00045–17.
- [36] Y. Yao, T. Huo, Y.-L. Lin, S. Nie, F. Wu, Y. Hua, J. Wu, A. R. Kneubehl, M. B. Vogt, R. Rico-Hesse, Y. Song, *J. Am. Chem. Soc.* **2019**, DOI 10.1021/jacs.9b02505.
- [37] J. C. Shelley, A. Cholleti, L. L. Frye, J. R. Greenwood, M. R. Timlin, M. Uchimaya, *J. Comput.-Aided Mol. Des.* **2007**, *21*, 681–691.
- [38] M. P. Jacobson, D. L. Pincus, C. S. Rapp, T. J. F. Day, B. Honig, D. E. Shaw, R. A. Friesner, *Proteins.* **2004**, *55*, 351–367.
- [39] K. J. Bowers, E. Chow, H. Xu, R. O. Dror, M. P. Eastwood, B. A. Gregersen, J. L. Klepeis, I. Kolossvary, M. A. Moraes, F. D. Sacerdoti, J. K. Salmon, Y. Shan, D. E. Shaw, in *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, ACM, New York, NY, USA, **2006**.
- [40] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyran, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel, R. A. Friesner, *J. Chem. Theory Comput.* **2016**, *12*, 281–296.
- [41] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.
- [42] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- [43] M. R. Berthold, N. Cebbron, F. Dill, T. R. Gabriel, T. Kötter, T. Meinel, P. Ohl, K. Thiel, B. Wiswedel, *ACM SIGKDD Explorations Newsletter* **2009**, *11*, 26.
- [44] "Topic Extraction – kilian.thiel," can be found under [https://hub.knime.com/knime/spaces/Examples/latest/08 Other Analytics Types/01 Text Processing/17 TopicExtraction with the ElbowMethod, n.d.](https://hub.knime.com/knime/spaces/Examples/latest/08%20Other%20Analytics%20Types/01%20Text%20Processing/17%20TopicExtraction%20with%20the%20ElbowMethod.n.d)
- [45] T. A. Halgren, *J. Chem. Inf. Model.* **2009**, *49*, 377–389.
- [46] R. Krivák, D. Hoksza, *J. Cheminf.* **2018**, *10*, 39.

Manuscript received: April 6, 2021

Accepted manuscript online: April 25, 2021

Version of record online: May 18, 2021