Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Journal of Molecular Graphics and Modelling 112 (2022) 108122

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

Molecular dynamics simulations of the flexibility and inhibition of SARS-CoV-2 NSP 13 helicase

Bryan A. Raubenolt [a], Naeyma N. Islam [a], Christoper M. Summa [b], Steven W. Rick [a,*]

[a] Department of Chemistry, University of New Orleans, New Orleans, LA, 70148, USA
[b] Department of Computer Science, University of New Orleans, New Orleans, LA, 70148, USA

A B S T R A C T

The helicase protein of the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is both a good potential drug target and very flexible. The flexibility, and therefore its function, could be reduced through knowledge of these motions and identification of allosteric pockets. Using molecular dynamics simulations with enhanced sampling, we determined key modes of motion and sites on the protein that are at the interface between flexible domains of the proteins. We developed an approach to map the principal components of motion onto the surface of a potential binding pocket to help in the identification of allosteric sites.

1. Introduction

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has had a large health and socioeconomic impact worldwide [1–3]. During 2020, COVID-19, the disease caused by SARS-CoV-2, was among the three leading causes of death in the United States, along with heart disease and cancer [4]. The development of vaccines shows promise to end the pandemic, although therapeutics may also play a role [5]. Along with the Spike proteins, the non-structural proteins (NSP) main protease (3CLpro), papain-like protease (PLpro), RNA-dependent, RNA polymerase (RdRp), and helicase (NSP13) are the most promising targets for therapeutics against SARS-CoV-2 [6]. These proteins are highly conserved within the Coronavirus genus and important to the viral life cycle [7,8]. Helicase has a very high sequence conservation, making it a very promising target [9]. Helicase inhibitors, amenemevir [10] and pritelivir [11], have been developed for herpes simplex virus (HSV) and other studies have identified potential inhibitors of the closely related virus SARS-CoV [12–19].

SARS-CoV-2 NSP13 helicase has 596 residues and is 99% homologous to SARS-CoV helicase. The X-ray structure from SARS-CoV (PDB ID 6JYT) shows five domains [20]. The three domains, IA, 1B, and 2A make up a triangular base with the N-terminal zinc binding domain (ZBD) at the top, and a stalk region connecting the base to the ZBD (Fig. 1). Fig. 1 shows the SWISS-MODEL [21] predicted structure based on 6JYT. Recently, Newman, et al., determined the X-ray structure of SARS-CoV-2 NSP13 helicase, for the apo protein and in complex with phosphate, adenosine 5′-triphosphate, ATP, and other ligands [22]. The catalytic function of the helicase is to separate double stranded nucleic acids; the protein hydrolyzes nucleoside triphospate molecules (most often ATP), a chemical reaction that ultimately produces the mechanical energy necessary for the helicase to change conformations every step in the unwinding process [23]. The flexibility of the protein, key to its function, might make helicases a challenge for structural-based drug design [24], although a number of groups [25–29], have used computational docking and molecular dynamics (MD) simulations to identify potential SARS-CoV-2 inhibitors. The inhibitors are designed to bind to the ATP binding site of helicase [25–28], located between domains 1A, 1B, and 2A [25,30]. Another study designed inhibitors to bind to a potential allosteric site, involving resides of the ZBD, the stalk, and domain 1A [29]. Beck et al., used machine learning to find potential inhibitors without using structural information [31]. White, Lin, and Cheng demonstrated that two of their inhibitors inhibited helicase activity [26]. One simulation study, as part of a study of many SARS-Cov-2 proteins, using Markov state models and distributed computing, found a large domain motion [32].

For the SARS-CoV inhibitors [12–19] there is only a little known about the interactions between the molecules and the protein. The inhibitors bananin and its derivatives of Tanner and co-workers show a non-competitive mechanism [12] and mutagenesis experiments and molecular docking by Wang and co-workers suggest that the molecules interact with residues 257 and 259 on domain 1B [33]. The inhibitors of Adedeji and co-workers are also believed to be non-competitive, perhaps by inducing a structural change, although the binding site is not known [17]. Bismuth complexes may act as inhibitors by competing for the zinc binding sites [13]. Yu and co-workers reported inhibitors that modelling suggests bind to the ATP site [16]. For the other inhibitors, no binding

* Corresponding author.
E-mail address: srick@uno.edu (S.W. Rick).

https://doi.org/10.1016/j.jmgm.2022.108122
Received 30 September 2021; Received in revised form 31 December 2021; Accepted 3 January 2022
Available online 5 January 2022
1093-3263/© 2022 Elsevier Inc. All rights reserved.

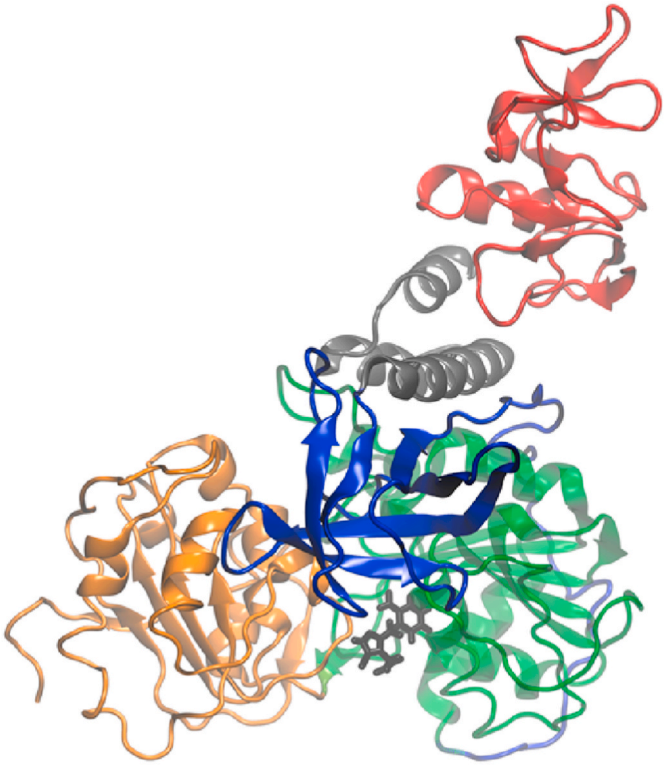

Fig. 1. SWISS-MODEL structure of SARS-CoV-2 helicase, showing domain 1A in green, domain 1B in blue, domain 2A in orange, the stalk domain in grey, the ZBD in red, and a potential inhibitor in grey near the bottom. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

sites have been proposed [14,15,18,19].

Inhibitor design could benefit from a better understanding of the flexibility of helicase, both with and without an inhibitor present. MD simulations, used to explore conformational space, can be aided using enhanced sampling methods. One general purpose, widely-used enhanced sampling approach is replica exchange [34,35]. In replica exchange (RE), simulations at a high temperature, which can more easily overcome high energy barriers, are run together with the simulations at the temperature of interest, with Monte Carlo swaps taken between the replicas. A limitation of RE is poor system size scaling. As the system size increased, the number of replicas required increases, making RE inefficient for large systems [36]. Both the utility and the limitations of RE are evident in a recent simulation of the SARS-CoV-2 spike protein [37]. This study required 46 replicas to span a temperature range of 310–350 K (and that was with many of the atoms restrained, effectively reducing the degrees-of-freedom of the system). Despite this, RE is a powerful method. Periole and Mark showed RE was 30–50 times more efficient than conventional MD in determining the melting curve for a small peptide in explicit solvent [38]. One method to increase the efficiency of RE is the Replica Exchange with Dynamical Scaling (REDS), which can reduce the number of replicas required by a factor of 8–10 [39,40].

In this work, we use computer simulations with the REDS method to understand the flexibility of SARS-CoV-2 NSP13 helicase and to predict alternative structures that the protein adopts. Simulations will be carried out for both the apo protein and the protein in complex with a potential inhibitor in the ATP site.

2. Methods

Examining the conformational flexibility of helicase presents a sampling challenge, as the time scales for the structural changes can be longer than typical simulation times, and a force field challenge, as the models must correctly weigh different, nearly energetically equal conformations [41]. Considerable effort has gone into the development of force fields for proteins [42,43]. The recently developed a99SB-*disp* model [44] has been shown to accurately reproduce the structure of a range of folded and disordered proteins [44], as well as a protein that contains both folded and disordered domains [45]. The a99SB-*disp* model is the result of many revisions of the Amber 99SB model [46], which was a revision of the Amber 94 force field [47]. The SB model revised the backbone φ/ψ dihedral terms, which corrected errors in the backbone secondary structure, particularly for glycine [46]. Best and Hummer revised the ψ dihedral terms to give the correct helix/coil populations, creating Amber 99SB* [48]. The side chain torsion potentials for four amino acids (isoleucine, leucine, aspartic acid, and asparagine) that showed disagreement with experimental data were reoptimized against quantum calculations, resulting in the a99SB*-ILDN model, also referred to as ff99SB*-ILDN [49]. To fix the helix propensities of charged residues, new charges from a restrained electrostatic potential (RESP) fit with constraints on the backbone charges to match neutral side chains, were assigned to charges for charged side chains, to give the a99SB*-ILDN-Q model [50]. The a99SB-*disp* model makes a number of changes to the a99SB*-ILDN-Q model, starting with enhanced dispersion interactions, through the Lennard-Jones potential, between the amide hydrogen and carbonyl oxygen backbone atoms. Side chain torsional terms were optimized against *ab initio* and experimental data. Side chain charges for the charged residues aspartate, glutamate, and arginine were re-parameterized, as was backbone torsional terms for glycine. Lastly, a new water model, a99SB-*disp* water with slightly stronger Lennard-Jones dispersion was developed to be used with the a99SB-*disp* model [44].

For the three zinc ions in the ZBD and the complexing cysteine and histidine residues, we used the Zinc Amber Force Field (ZAFF) [51]. The ZAFF force field gave different charges for the backbone atoms for cysteine residues. To make the ZAFF force field compatible with a99SB-*disp* and have the same charges for the N,H, C, and O backbone atoms, we modified the charges of the Cα and Hα atoms for the cysteine residues complexing a zinc ion, by adding 0.0077*e* to both atoms, so that charge is conserved. The rest of the side chain charges were from the ZAFF force field. The backbone charges for the histidine residues from ZAFF were the same as a99SB-*disp*.

Simulations were carried out both for the apo enzyme and with a potential inhibitor in the ATP binding site. This inhibitor was identified through a combination of docking using the FRED and HYBRID programs [52], and molecular dynamics simulations, as described in the Supplementary Information. The Generalized Amber Force Field (GAFF) [53] parameters were used for the potential inhibitor, using AM1-BCC charges [54]. The apo simulation used 36002 and the complex simulation used 35993 water molecules, using the a99SB-*disp* model, in a rhombic dodecahedron box. Charge neutrality was maintained by adding 9 chloride ions for the apo and 10 chloride ions for the complex, using the a99SB-*disp* force field for ions.

The simulations were setup using the tleap and antechamber programs of AMBER 16 [55,56]. Those coordinate and topology files were converted to GROMACS input files using the ACPYPE program [57]. Parameter sets for the a99SB-*disp* model were downloaded from GitHub at https://github.com/paulrobustelli/Force-Fields.

Configurational space sampling is enhanced using the REDS method [39,40], which combines conventional replicas with replicas that have a modified potential energy. The modified, or scaled replicas, can span a large range of energy and are placed between two conventional replicas, so that fewer replicas are required. Replica exchange works by having a replica at a temperature high enough to overcome energetic barriers. Choosing the high temperature is important to the efficiency of replica exchange [58–61]. An optimal high temperature of 366 K was found by Rosta and Hummer for a small protein in water (the l-repressor fragment

$l_{6\text{-}85}$) [61] and Heo and Feig used a temperature of 360 K for the prediction of SARS-CoV-2 protein structures (but not including helicase) [62], suggesting that 360 K is a reasonable high temperature. Replica exchange simulations for eight SARS-CoV-2 proteins, but not helicase, used a temperature range of 310 K–350 K, with a number of replicas from 20 to 60 [63]. Our REDS simulations had conventional replicas, using the plain, unscaled replicas, at temperatures of 310, 319, 331, 345, and 360 K, with scaled replicas at 314, 325, 338, and 352 K, for a total of 9 replicas. These temperatures were determined using the method of Patriksson and van der Spoel [64]. RE would require 42 replicas for this system. The simulations used the REDS method as implemented in the GROMACS simulation package [65]. Details of this implementation are given elsewhere [40] and the additions to the GROMACS source code were downloaded from Gitlab at https://gitlab.com/csumma_uno/gromacs.

Initial coordinates of the protein were taken from the SWISS-MODEL [21] predicted structure (P0DTD1 PRO_0000449630, Model 1 [66]) based on the 6JYT [20] structure from SARS-CoV. The helicases from the two viruses have a close homology, differing only by one residue (570 which is an ILE in SARS-CoV and VAL in SARS-CoV-2). After our simulations were set-up and running, crystal structures of SARS-CoV-2 NSP13 were published [22]. The apo structure (PDB ID 7NIO) has two chains, with a Cα root-mean-square deviation (RMSD) of 0.15 nm. The SWISS-MODEL structure has a Cα RMSD of 0.20 nm with Chain A and 0.19 nm with Chain E of the 7NIO structure. Both Chain A and Chain E have unresolved residues, mostly in domain 1B. Also after our simulations were running, White, et al., published a paper pointing out problems with the 6JYT structure, including *cis* peptide bonds in the loop near the ATP binding site [26]. The 6JYT structure has two chains. The SWISS-MODEL structures used Chain B, which does not have *cis* peptide bonds in this loop. Comparisons between the SWISS-MODEL and the 7NIO structures are shown in Fig. S2. The loop does have a different structure from the 7NIO structures, but the residues that have important interactions with the ligand (LYS 288, ASP 374, and GLU 375) are in similar positions. The SWISS-MODEL structure was protonated using the H++ server [67]. The structure was solvated and chloride ions were added to give a neutral system. This structure was minimized using steepest decent for 1000 steps, followed by constant molecule number, volume, and temperature (NVT) simulations for 100 ps and then 100 ps at constant N, pressure, and T (NPT) for each of the temperatures.

The simulations used a 0.002 ps time step. All bonds involving hydrogens were constrained using the LINCS [68] and Settle [69] algorithms. Long-ranged electrostatics were treated using Particle mesh Ewald [70] with a Fourier mesh spacing of 0.16 nm, a real-space cut-off of 1.2 nm, and fourth-order interpolation. Lennard-Jones interactions used a 1.2 nm cut-off, with long-ranged dispersion corrections for energy and pressure. Temperature was maintained using stochastic velocity rescaling [71], with separate protein and solvent thermostating and a relaxation time of 0.1 ps. Pressure was maintained, at 1 bar, using a Parrinello-Rahman barostat [72] with a relaxation time of 2 ps and a compressibility of $4.5\text{x}10^{-5}$ bar^{-1}. Exchanges between replicas were attempted between adjacent replicas every 100 time steps. The simulations were run for a total of 150 ns. The last 100 ns were used as production runs for data analysis. Root mean square deviations and root mean square fluctuations were determined using gmx rms and rmsf utilities in Gromacs [73], respectively. Clustering of structures based on the root-mean-square-deviation (RMSD) with a 0.25 nm cut-off used the gmx clustering utility in Gromacs [73]. The solvent accessible surface area (SASA) was calculated with using the sasa utility in Gromacs [73], with a probe radius of 0.14 nm. Principal component analysis and correlation matrix calculations used the gmx covar utility in Gromacs [73]. The correlation matrix is defined as

$$C_{ij} = \langle \Delta r_i \cdot \Delta r_j \rangle \Big/ \left(\langle \Delta r_i^2 \rangle \langle r_j^2 \rangle \right)^{1/2} \quad (1)$$

where the angle brackets donate an ensemble average and Δr_i is the distance of atom *i* from its average position [74].

Using a structure that resulted from the lengthy MD simulations, additional binding pockets were explored using PrankWeb [75], as web interface for the P2Rank ligand binding site prediction method [76]. Docking into these sites was done using the program Autodock Vina [77], using 3868 compounds from the SWEETLEAD [78] library. The SWEETLEAD library consists of approved pharmaceuticals, illicit drugs, and traditional medicinal herbs. Spatial data file (SDF) input files were converted to PDBQT (protein data bank format plus partial charges and atom types) input for Autodock using OpenBabel [79].

3. Results

The top scoring compounds from the docking analysis were simulated for 100 ns, of which three stayed in the ATP site for the entire simulation length. Full details are given in the Supplementary Information. The top scoring inhibitor FCID1 was selected for further simulation using enhanced sampling. The initial placement of FCID1 is shown in Fig. 2.

Molecular simulation results. For the simulation of the complex, some of the replicas give trajectories where the FCID1 molecule leaves the binding pocket. For the analysis of the results of the complex that follows, only the bound structures are used. Structures are saved every 0.01 ns, so the entire 150 ns simulation has 15,000 saved structures. Of those 11,373 have FCID1 in the ATP site. In two of the nine replicas, the inhibitor will eventually leave the binding site. In the plots below showing the time evolution of selected properties, the absence of the data corresponding to the unbound data creates small gaps.

The structures for both the apo protein and the complex both change considerably from the initial structure (Fig. 3). The use of replica exchange results in sudden in changes in the root-mean-square deviation (RMSD) as configurations from different temperatures are swapped. Structures with an RMSD over 1 nm are observed. The RMSDs from the initial structure for the five domains, average over the last 100 ns of the simulations, are given in Table 1. The ZBD and the stalk domains are the most rigid, as indicated by the small RMSD and the small standard deviation of the RMSD. The 1A, 1B, and 2A domains show large structural changes, showing that the conformational changes involve not only motion of the domains relative to each other but structural changes within these three domains. Domain 2A has the largest RMSD. The apo and the complex show similar structural changes.

The relative motion of the domains is determined by calculating the distance between the center of masses of the domains, using the Cα atoms (Fig. 4). The ZBD and stalk domains do not show any significant inter-domain flexibility, so that data is not shown. The distance between domains 1A and 1B is relatively constant over the simulation (Fig. 4A), For the apo-protein the 1A-1B distance tends to decrease with time, although it does sample some configurations near the end of time simulation where that distance is larger. The distance between domains 1A and 2A shows a larger change (Fig. 4C), fluctuating over distances of almost 1 nm. From the distribution of these distances (Fig. 4D), it can be seen that the complex fluctuates between two conformations, one with a 1A–2A distance around 3.1 nm and another around 3.7 nm. The apo protein shows a single broad distribution. The 1B-2A domains show the largest change in distance (Fig. 4E). Both the apo protein and the complex make transitions between a more open structure with a 1B-2A distance around 4.3 nm and the more closed conformation of the original structure, around 3.1 nm. The complex visit the open structure more frequently.

The structural fluctuations of an atom about its average position are given as the root mean square fluctuation (RMSF). The RMSF values for the Cα atoms are show in Fig. 5. The complex shows a slightly large RMSF, with an average of 0.42 nm for all atoms, compared to 0.37 nm for the apo protein. The residues at the C-terminus, which have a coil

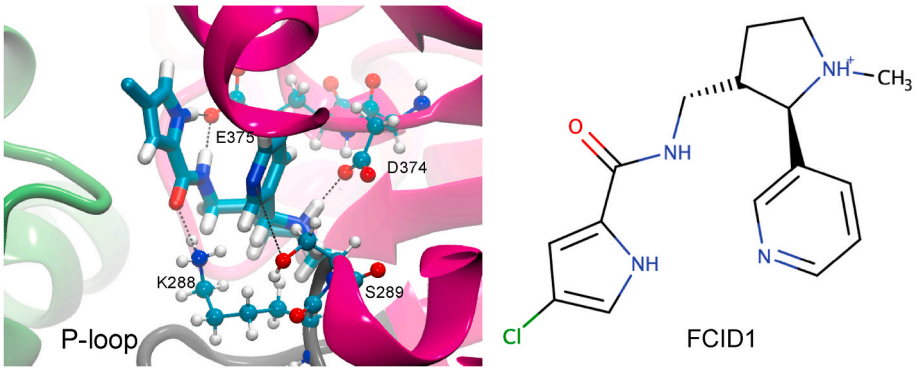

Fig. 2. Docked pose of FC1D1 in the ATP site, showing the major interactions with the protein side chains. The IUPAC name of FCID1 is (2R)-3-[(4-chloro-1H-pyrrol-2-yl)formamido]methyl-1-methyl-2-(pyridin-3-yl)pyrrolidin-1-ium.

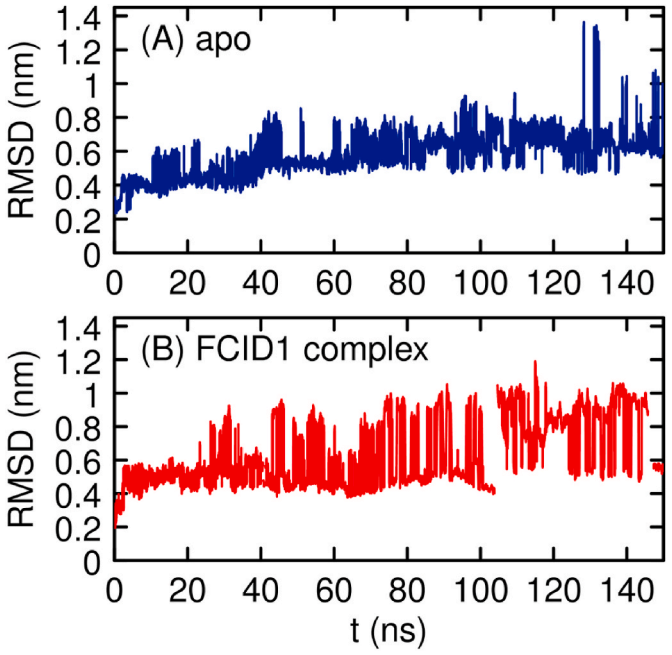

Fig. 3. Cα RMSD for the apo protein (A) and the protein/FCID1 complex as a function of time at 310 K.

Table 1
Average and standard deviation of the Cα RMSD for the total protein and each domain from the last 100 ns of the simulation.

	apo protein Cα RMSD (nm)		protein FCID1complex Cα RMSD (nm)	
	average	standard deviation	average	standard deviation
total	0.63 ± 0.04	0.10 ± 0.01	0.70 ± 0.14	0.22 ± 0.02
ZBD	0.168 ± 0.006	0.024 ± 0.003	0.17 ± 0.02	0.028 ± 0.007
stalk	0.104 ± 0.007	0.017 ± 0.001	0.091 ± 0.003	0.011 ± 0.001
1A	0.46 ± 0.05	0.11 ± 0.01	0.44 ± 0.08	0.18 ± 0.08
1B	0.37 ± 0.01	0.07 ± 0.02	0.33 ± 0.01	0.046 ± 0.009
2A	0.56 ± 0.06	0.12 ± 0.01	0.56 ± 0.03	0.047 ± 0.005

structure, have the largest RMSF. Domain 2A shows the largest fluctuations overall. There is a region with large RMSF in domain 1A, around residue CYS342, which is part of an extended coil region on the outside of the domain near the interface with domain 1B. Another region with a large RMSF is around residue VAL484 on domain 2A, also part of a coil.

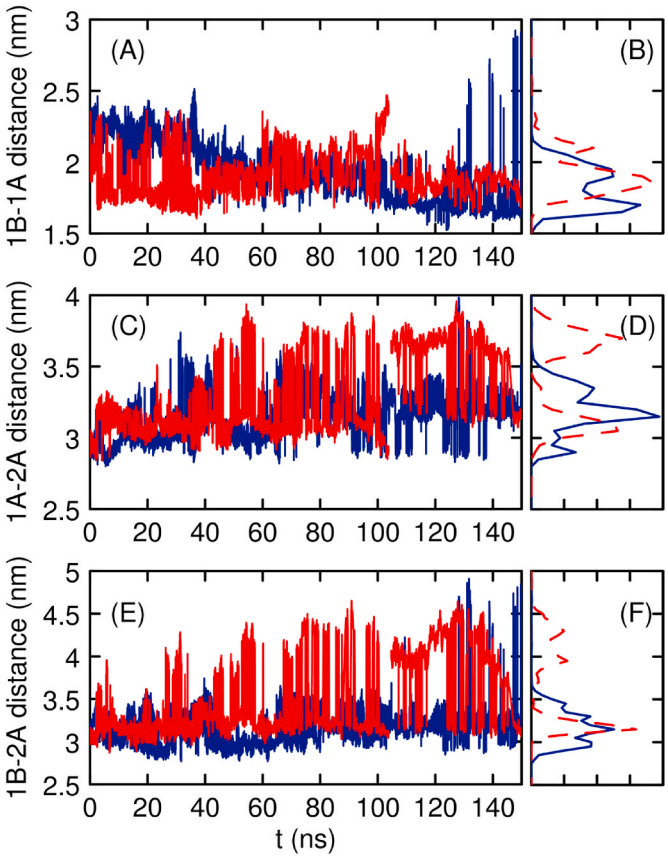

Fig. 4. Distance between the domains as a function of time (A,C,D) and the distributions of the distances from the last 100 ns (B,D,F). The blue lines are for the apo protein and the red lines are for the complex. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

The residues in the ATP binding site, particularly residues ASP374 and GLU375 have small RMSF values, around 0.2 nm. Other rigid regions are around residue 136, near the boundary between the Stalk and 1B domains, around residue PRO234, and around residue ASN388.

The structures from the last 100 ns of the simulations at 310K were clustered based on RMSD. The positions of the Cα atoms for residues 1–590, excluding the highly flexible C-terminus, were used. This resulted in 20 clusters for the apo protein and 11 for the complex. There is no overlap between the structures from the two separate simulations; no clusters contained structures form both the apo and complex simulations. Selected properties of the top 5 clusters are given in Table 2. The

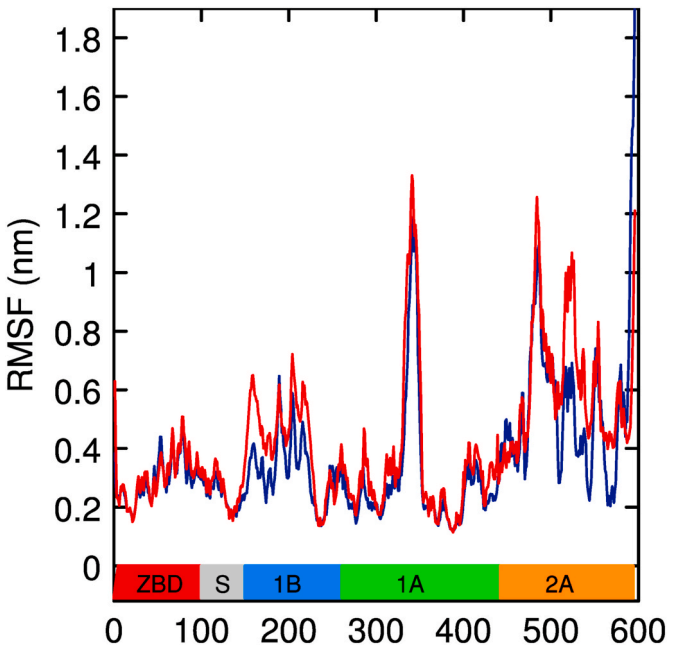

Fig. 5. Root mean square fluctuations of Ca atoms for the apo protein (blue line) and complex (red line). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Cluster rank, population fraction, RMSD from initial structure, distances between domains, and solvent accessible surface area (SASA).

rank	population fraction	RMSD (nm)	1B-1A distance (nm)	1A–2A distance (nm)	1B-2A distance (nm)	SASA (nm^2)
Original structure		0	2.092	3.000	2.991	296
Apo protein						
1	0.288	0.603	1.649	3.258	3.161	315
2	0.201	0.753	1.736	3.132	3.314	321
3	0.199	0.551	1.949	2.927	3.178	322
4	0.100	0.684	1.887	3.471	3.378	334
5	0.099	0.499	1.932	2.965	2.913	304
9	0.033	1.316	2.506	4.845	3.797	326
16	0.009	1.361	2.231	4.667	3.919	333
Complex						
1	0.374	0.806	1.909	4.228	3.650	344
2	0.346	0.422	2.102	3.174	3.073	331
3	0.084	0.976	1.884	3.938	3.642	335
4	0.069	0.938	1.825	3.929	3.630	333
5	0.068	0.536	1.810	3.135	3.206	319

top 5 clusters contained about 90% of the total structures. Also shown are properties for the two clusters for the apo protein that have the largest RMSD from the initial structure. For comparison, the properties of the SWISS-MODEL structure are given.

For the apo structures, the most populated cluster contains the structures that are predominant over the last 20 ns of the simulation, suggesting that this is the structure that is most stable. For the complex, the structures that are prevalent in the last 20 ns are in clusters 1, 3 and 5. The solvent accessible surface area for the structures from the simulations are larger than the initial structure, indicating, along with the domain distances, that the simulation structures are more open. Fig. 6 shows the structures of selected clusters, including those with the biggest populations and one with the largest RSMD from the original structure (cluster 9). A significant structural change involves motion of domain 1B (shown in blue). This domain moves between a structure closer to domain 2A (in orange) and a structure closer to the ZBD domain (in red). The motion involves hinge regions around residue TYR149, at the end of the stalk domain, and VAL232. Both of these regions have small fluctuations, bracketing a region of high fluctuations (Fig. 6). Domain 2A moves by bending around residue GLY439. The Markov state simulations using distributed computing found a large motion between domains, which they identified as domains 1A and 2A [32]. Upon inspection of Fig. S10 from that publication, the authors have mis-identified domain 1B as domain 1A, so that the large motion is actually between domains 1B and 2A.

The FCID1 inhibitor interacts with residues on both the 1A and 2A domains, as initially docked (Fig. 2). As the distances between the two domains increases, by over 1 nm, the molecule stays with domain 1A, where it has the strongest interactions, but loses interactions with residues on domain 2A. Fig. 7 shows the distances between the FCID1 atoms and the protein which are present in the original structure. There is a strong interaction between the ASP374 side chain and the cyclopentyl ammonium nitrogen (identified as N20), which is maintained during the length of the simulation (Fig. 8A). Another strong interaction is between the GLU375 side chain and two nitrogens on the inhibitor, the amide nitrogen (N18) and pyrrole ring nitrogen (N19). This interaction is also maintained during the simulation. Fig. 7B shows the N18- E375 distance; the N19-E375 distance is fairly constant as well. The Nζ atom of LYS288 interacts with the oxygen atom on the inhibitor. This interaction is not constantly maintained during the simulation (Fig. 8B). In cluster 1, the LYS288 Nζ and O atoms are in direct contact. In clusters 2–5, the atoms are over 0.5 nm apart, separated by solvent. The GLY538 Cα atom, part of domain 2A, has a close contact with a carbon atom (C3) on the cyclopentyl group. This interaction is lost for many of the simulation structures, including in clusters 1, 3 and 4, as domain 2A moves away from domain 1A (Fig. 9A).

The correlation of motion between different parts of the protein as given by the correlation matrix is shown in Fig. 10. Strongly positively correlated square regions corresponding to the five domains can be seen near the diagonal. The large 1A domain, from residues 261 to 442, is interrupted by the flexible region around residue CYS342, which does not have a strong correlation with the rest of domain 1A. Domain 2A, starting at residue GLN470, has a motion that is anti-correlated with all other domains, as indicated by the blue regions of Fig. 10. There is a region of positive correlation between residues 105 to 150 on the stalk domain and residues 370 to 430 on domain 1A. These regions are close to each other in the three dimensional structure. There is not a big difference between the apo and the protein complex, in comparing the upper left and the lower right sections of Fig. 8. The largest difference between the correlations of the apo and complex protein is the complex shows a larger negative correlation between domains 2A and the other domains.

To further quantify the motions of the protein, a principle component (PC) analysis was done, using all the trajectories, not only the data at 310 K. The principal component analysis for the apo protein shows that the first principal component involves a bending motion of domain 2A, towards domains 1A and 1B. The second principal component involves a twisting motion of domain 2A, combined with a hinge motion of domain 1B and motion of the flexible loop around residue 342. These motions are shown in Fig. 9. The PC eigenvalues are 35.2 nm^2 and 15.2 nm^2 for the first and second components, respectively. The eigenvalues sum to 117.0 nm^2, so the first two PCs represent 43% of all protein fluctuations. The projection of all the configuration (after the 50 ns of equilibration) for the apo and complex simulations onto the first two PC eigenvectors (from the apo simulations) are given in Fig. 10. The complex and the apo protein explore similar regions of PC2, but the complex does not explore very negative regions of PC1, corresponding to an open structure, with a large separation between domains 2A and 1B. Only one of the 9 replica trajectories of the apo protein visits the very open structures (with PC1 less than about 15), so it may be that the simulations of the complex have not had time to sample that region. It does appear that the path to large negative PC1 values goes through structures that have positive PC2 values. In order to reach a very open state, domain 2A has to twist.

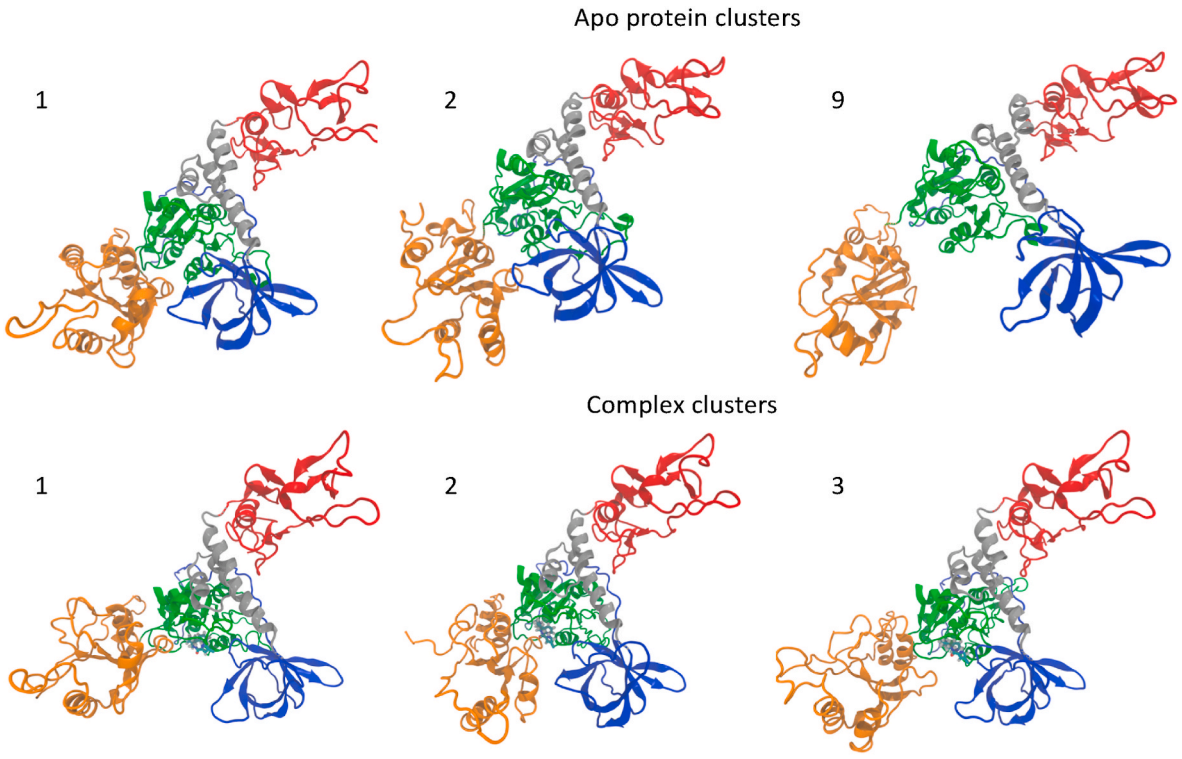

Fig. 6. Structures of selected clusters from the apo and cluster simulations, with domain 1A in green, domain 1B in blue, domain 2A in orange, the stalk domain in grey, and the ZBD in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

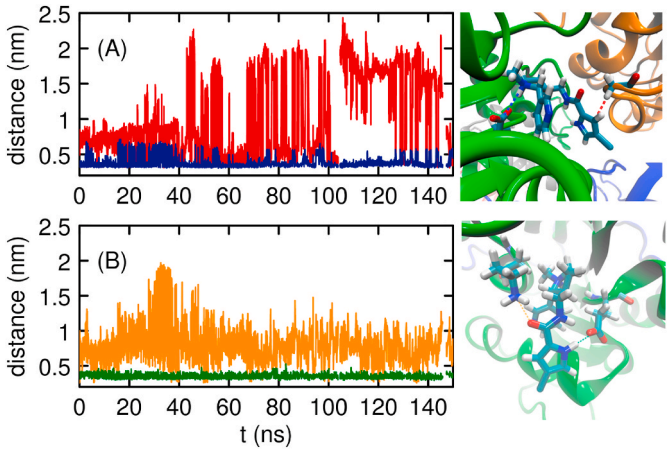

Fig. 7. (A) D374 Cγ - N20 (blue line) and G538 Cα - C3 (red line) distances. (B) E375Cδ - N18 (green line) and K288 Nζ - O (orange line) distances. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

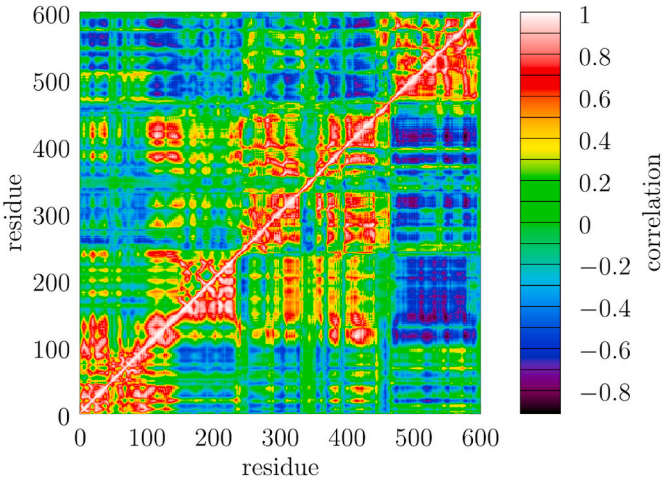

Fig. 8. Correlation matrix between positions of Cα atoms for the apo protein (upper left of the diagonal) and the complex (lower right of the diagonal), on left.

The large scale motion of the protein involves movement of domain 2A, relative to the other domains, including the neighboring domain 1A. A possible means to suppress the flexibility of the helicase might be designing an inhibitor to bind at the interface between domains 1A and 2A, in a region away from the both the ATP binding site and the nucleic acid binding channel. The nucleic acid binding channel is located between domain 1B on one side and domains 1A and 2A on the other [80]. Using the predominant structure of the apo protein from the clustering analysis, potential drug binding sites were identified using P2Rank [76]. Four, out of 17, binding sites were found at the interface between domains 1A and 2A, as shown in Fig. 10A. (Pockets 3 and 4 are close together but are separated by a narrow channel created by residues SER264, ASN268, and LYS462.) The pockets are numbered according to their rankings from PrRank, so that among the four sites in the region between domain 1A and 2A, pocket 1 is the highest ranked and pocket 4 is the lowest rank. Pocket 1 is the ATP binding site. In order to inhibit motion of domain 2A, the inhibitor would ideally be at a site with residues that are involved in domain motion but moving in opposite directions, as near a hinge or shear motion. The molecule would possibly inhibit the motion by interacting with the residues that need to move in opposite directions. The four binding sites can be assessed based on the motions of the residues in those binding sites using the results of the PCA. P2Rank returns the points of the solvent accessible surface (SAS) for each binding site. For a point *j* on the surface, the nearest residue is

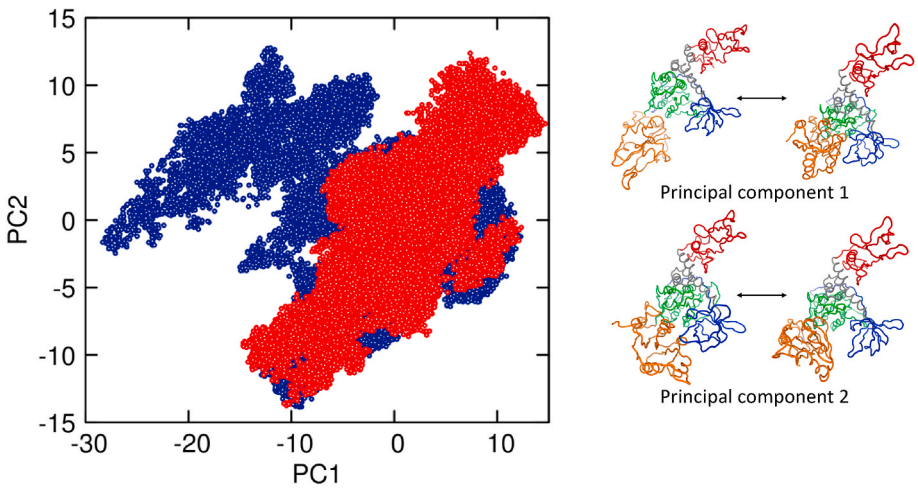

Fig. 9. Projection of the trajectory for the apo protien (blue dots) and the complex (red dots) onto the first and second principle components. The structures on the left correspond to negative values of the PC, so that negative values of PC1 correspond to a larger distance between domains 1B and 2A and positive values represents a rotation of the C-terminus part of domain 1B towards domain 2A. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

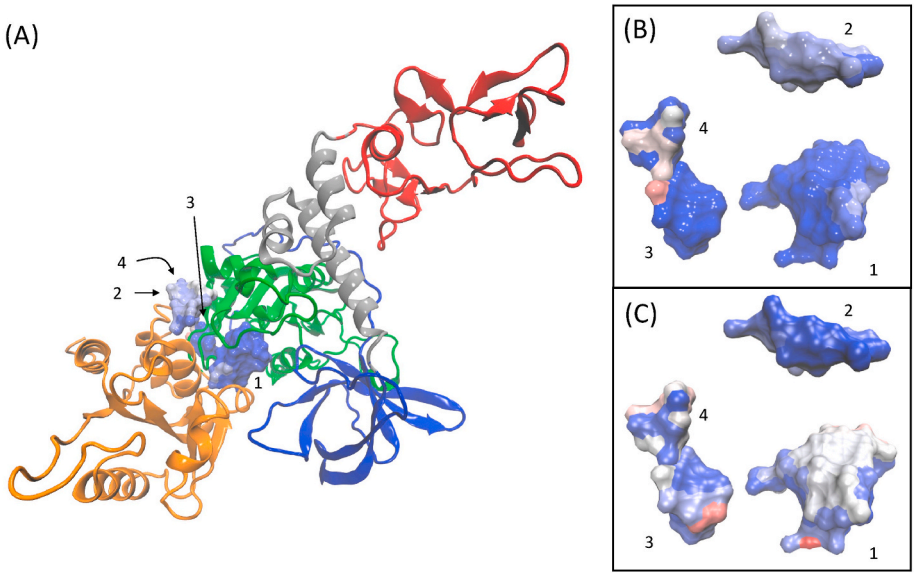

Fig. 10. (A) Four potential binding sites at the interface between domains 1A and 2A. Map for the movement of principal component 1 (B) and 2 (C) onto the surface of the binding sites, with blue denotes a positive values and red negative $d_\alpha(i)$ values. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

found, which is labeled i_j. The motion of that residue can be characterized from $q_\alpha(i_j)$ the eigenvector of principal component α for residue i_j. The total motion of the binding site can be found from

$$q_\alpha(total) = \sum_{j=1}^{N} q_\alpha(i_j) \Big/ N \tag{2}$$

where N is the number of SAS points for a specific pocket. The angle between the direction of motion of one point on the surface relative the average direction is a way to determine if a principal component of motion is moving the whole surface together, or if parts of the surface move in different directions. The cosine of this angle, given by

$$d_\alpha(i) = q_\alpha(i_j) \cdot q_\alpha(total) \,/\, \left(|q_\alpha(i_j)||q_\alpha(total)|\right) \tag{2.1}$$

can be mapped onto the SAS. Of the four pockets, site 4 shows the largest spread $d_\alpha(i)$ values, or most anisotropy, for the first PC, see Fig. 10 (B). Areas of the surface that move in the same direction as the average (positive $d_\alpha(i)$) of the pocket are shown in blue and regions that move in the opposite direction as shown in red. Pocket 4 has contact with SER264, ALA267, and ASN268 on domain 1A and LYS462 on domain 2A, which move in opposite directions in PC1. The pocket also makes contact with PHE437 and LYS460.

Pocket 4 shows the most anisotropy for the second PC as well, although pockets 1 and 3 also show a high anisotropy, as shown in Fig. 10 (C). In pocket 4, LYS460 and LYS462 move in opposite directions from ASN 268 for PC2. Pocket 3 has contact with ASN265 and ARG442, which move in opposite directions. Pocket 1, the ATP site, the residues GLY287, PHE291, GLY400, and GLN537 move in opposition to residues SER377, MET378, and ASP 274. Pocket 2, in contact with GLN281, ASP401, ALA403, and MET429, has a small anisotropy for both the first and second PC, meaning all those residues move in the same direction in those modes.

The PC map analysis suggests that pockets 3 and 4 might be good targets for a non-competitive inhibitor, which could limit the flexibility of domain 2A. If an inhibitor could simultaneously make contact with residues among SER264, ALA267, and ASN268 on domain 1A and LYS460 and LYS462 on domain 2A, it could reduce the movement of both PC1 and PC2. In pocket 3, an inhibitor in contact with ASN265 and ARG442 might reduce movement of PC2. From the correlation matrix residues 267 and 268 have a negative correlation (around −0.1) with

residues 460 and 462, while residues 265 and 442 have a positive correlation (around 0.5).

Pocket 4, while promising in terms of inhibiting domain motion, is not ranked highly by PrRank (13th out of 17 potential pockets). The druggability of the pockets was assessed through virtual screening. Docking into pocket 4 gave weakly binding compounds; in contrast, docking into pocket 3, which is ranked fourth out of the 17 pockets, gave more promising scores (details are in the Supplementary Information). A top scoring compound made hydrogen bonds with ASN265, THR440, and ARG442, as well as hydrophobic contacts between benzyl group and ALA446, VAL449, and the methyl group on THR440. It therefore makes significant contacts with residues on both domains 1A and 2A.

4. Conclusions

The simulations revealed that the NSP13 helicase, both with and without an inhibitor at the ATP site, is highly flexible, with an RMSD from the original structure of over 0.6 nm (see Fig. 3 and Table 1). Some structures have RMSD values over 1.3 nm (Table 1). The largest structural changes involve motion of domain 2A, relative to the other domains. Principal component (PC) analysis showed that about 43% of the fluctuations of the protein are captured in the first two principal components, both of which involve large amplitude motion of domain 2A. The flexibility of the helicase might be altered with an inhibitor to bind at the interface between domains 1A and 2A, in a region away from the ATP and the nucleic acid active sites. Four potential drug binding pockets between domains 1A and 2A were identified. We sought a binding site that could change the flexibility so we developed an approach that maps the motion of the atoms on the surface of each pocket onto the pocket's surface, as shown in Fig. 10. Maps were done using the motion as characterized either by the first or second PC. If a pocket is at the boundary between regions that tend to move in different directions, then the PC map would show negative and positive areas on the surface. The combination of the suitability of the pocket for drug binding and the indication from the PC map that it is involved domain motion, one pocket was identified as a promising allosteric site.

A number of studies have compared PCA from ligand bound and ligand free simulations, including those for helicase [81,82], to determine how does a ligand changes the principal components. We are looking at the reverse question, how does PCA point to a site where a ligand might influence the dynamics, and hence the PCAs. Other work has combined PCA and binding pocket characterization. The GRID/-CPCA method generates data for PCA for different ligand types in a pocket [83]. Other methods perform PCA on binding pocket grids from multiple MD structures and to analyze how a pocket changes shape [84–87]. The aim of these methods is to guide structure-based drug design and not necessarily to identify allosteric sites. Our approach differs in that it maps the relative motion of pocket surface atoms as determined from PCA onto the surface.

The proposed allosteric site is in a different location on the protein than some of the previously proposed allosteric sites for SARS-Cov and SARS-Cov-2 NSP13 helicase [29,33]. X-ray crystallographic fragment screening found three possible allosteric sites: between the stalk and ZBD domains, between the stalk and 1B domains, and between domains 1A and 2A [22]. The fragments in this third site (as determined from the 5RMB and 5RMF structures) makes contacts with residues GLN281, LYS430, PRO434, and PHY 437 on domain 1A and residues TYR457, ASP458, and LYS460 on domain 2A. These residues are in a similar region to one of the binding pockets identified from our simulations (pocket 3). Our results from the PCA analysis support the conclusion from Newman et al., [22] that this region could be a good target for allosteric inhibitor design.

Helicase NSP13 is part of a replication and transcription complex (RTC) together with other proteins, including NSP7, NSP8, and NSP12, as well as another copy of itself [80]. The helicase structure in that complex is likely to be different that it is in solution. However, *in silico* design has led to drugs that inhibit helicase ATPase activity *in vitro*. [26] Further studies would be needed to determine if molecules designed to bind to the allosteric site are effective inhibitors of helicase. While many structural details into the exact mechanism of replication remain yet to be discovers, the relevance of the range of motion of domain 2A observed in our studies provides some insight into how the helicase carries out its crucial role in the RTC.

Data and software accessibility

The REDS modifications to the GROMACS simulation package are available from Gitlab at https://gitlab.com/csumma_uno/gromacs. Input files for the simulations and structures from the cluster analysis are provided in associated content.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work used the Extreme Science and Engineering Discovery Environment (XSEDE) which is supported by National Science Foundation grant number ACI-1548562. The work was carried out at the Comet resource at the San Diego Supercomputer Center through allocation TG-CHE200066. Additional computational support from the Louisiana Optical Network Initiative is gratefully acknowledged.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmgm.2022.108122.

References

[1] Y.A. Helmy, M. Fawzy, A. Elaswad, A. Sobieh, S.P. Kenney, A.A. Shehata, The COVID-19 pandemic: a comprehensive review of taxonomy, genetics, epidemiology, diagnosis, treatment, and control, J. Clin. Med. 9 (2020) 1225.
[2] M. Nicola, Z. Alsafi, C. Sohrabi, A. Kerwan, A. Al-Jabir, C. Iosifidis, M. Agha, R. Agha, The socio-economic implications of the coronavirus pandemic (COVID-19): a review, Int. J. Surg. 78 (2020) 185–193.
[3] J.P.A. Ioannidis, Global perspective of COVID-19 epidemiology for a full-cycle pandemic, Eur. J. Clin. Invest. 50 (2020), e13423.
[4] S.H. Woolf, D.A. Chapman, J.H. Lee, COVID-19 as the leading cause of death in the United States, JAMA 325 (2021) 123–124.
[5] S.K. Mishra, T. Tripathi, One year update on the COVID-19 pandemic: where are we now? Acta Trop. 214 (2021) 105778.
[6] C. Wu, Y. Liu, Y. Yang, P. Zhang, W. Zhong, Y. Wang, Q. Wang, Y. Xu, M. Li, X. Li, M. Zheng, L. Chen, H. Li, Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods, Acta Pharm. Sin. B (2020).
[7] H. Yang, W. Xie, X. Xue, K. Yang, J. Ma, W. Liang, Q. Zhao, Z. Zhou, D. Pei, J. Ziebuhr, R. Hilgenfeld, K.Y. Yuen, L. Wong, G. Gao, S. Chen, Z. Chen, D. Ma, M. Bartlam, Z. Rao, Design of wide-spectrum inhibitors targeting coronavirus main proteases, PLoS Biol. 3 (2005) e324.
[8] E. De Clercq, Potential antivirals and antiviral strategies against SARS coronavirus infections, Expert Rev. Anti Infect. Ther. 4 (2006) 291–302.
[9] S. Habtemariam, S.F. Nabavi, M. Banach, I. Berindan-Neagoe, K. Sarkar, P.C. Sil, S. M. Nabavi, Should we try SARS-CoV-2 helicase inhibitors for COVID-19 therapy? Arch. Med. Res. 51 (2020) 733–735.
[10] K. Chono, K. Katsumata, T. Kontani, M. Kobayashi, K. Sudo, T. Yokota, K. Konno, Y. Shimizu, H. Suzuki, ASP2151, a novel helicase-primase inhibitor, possesses antiviral activity against varicella-zoster virus and herpes simplex virus types 1 and 2, J. Antimicrob. Chemother. 65 (2010) 1733–1741.
[11] A. Wald, L. Corey, B. Timmler, A. Magaret, T. Warren, S. Tyring, C. Johnston, J. Kriesel, K. Fife, L. Galitz, S. Stoelben, M.L. Huang, S. Selke, H.P. Stobernack, H. Ruebsamen-Schaeff, A. Birkmann, Helicase-primase inhibitor pritelivir for HSV-2 infection, N. Engl. J. Med. 370 (2014) 201–210.
[12] J.A. Tanner, B.J. Zheng, J. Zhou, R.M. Watt, J.Q. Jiang, K.L. Wong, Y.P. Lin, L. Y. Lu, M.L. He, H.F. Kung, A.J. Kesel, J.D. Huang, The adamantane-derived bananins are potent inhibitors of the helicase activities and replication of SARS coronavirus, Chem. Biol. 12 (2005) 303–311.

[13] N. Yang, J.A. Tanner, Z. Wang, J.D. Huang, B.J. Zheng, N. Zhu, H. Sun, Inhibition of SARS coronavirus helicase by bismuth complexes, Chem. Commun. (2007) 4413–4415.

[14] C. Lee, J.M. Lee, N.R. Lee, B.S. Jin, K.J. Jang, D.E. Kim, Y.J. Jeong, Y. Chong, Aryl diketoacids (ADK) selectively inhibit duplex DNA-unwinding activity of SARS coronavirus NTPase/helicase, Bioorg. Med. Chem. Lett 19 (2009) 1636–1638.

[15] M.K. Kim, M.S. Yu, H.R. Park, K.B. Kim, C. Lee, S.Y. Cho, J. Kang, H. Yoon, D. E. Kim, H. Choo, Y.J. Jeong, Y. Chong, 2,6-Bis-arylmethyloxy-5-hydroxychromones with antiviral activity against both hepatitis C virus (HCV) and SARS-associated coronavirus (SCV), Eur. J. Med. Chem. 46 (2011) 5698–5704.

[16] M.S. Yu, J. Lee, J.M. Lee, Y. Kim, Y.W. Chin, J.G. Jee, Y.S. Keum, Y.J. Jeong, Identification of myricetin and scutellarein as novel chemical inhibitors of the SARS coronavirus helicase, nsP13, Bioorg. Med. Chem. Lett 22 (2012) 4049–4054.

[17] A.O. Adedeji, K. Singh, N.E. Calcaterra, M.L. DeDiego, L. Enjuanes, S. Weiss, S. G. Sarafianos, Severe acute respiratory syndrome coronavirus replication inhibitor that interferes with the nucleic acid unwinding of the viral helicase, Antimicrob. Agents Chemother. 56 (2012) 4718–4728.

[18] J.B. Cho, J.M. Lee, H.C. Ahn, Y.J. Jeong, Identification of a novel small molecule inhibitor against SARS coronavirus helicase, J. Microbiol. Biotechnol. 25 (2015) 2007–2010.

[19] J.M. Lee, J.B. Cho, H.C. Ahn, W. Jung, Y.J. Jeong, A novel chemical compound for inhibition of SARS coronavirus helicase, J. Microbiol. Biotechnol. 27 (2017) 2070–2073.

[20] Z. Jia, L. Yan, Z. Ren, L. Wu, J. Wang, J. Guo, L. Zheng, Z. Ming, L. Zhang, Z. Lou, Z. Rao, Delicate structural coordination of the severe acute respiratory syndrome coronavirus Nsp13 upon ATP hydrolysis, Nucleic Acids Res. 47 (2019) 6538–6550.

[21] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T.A.P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: homology modelling of protein structures and complexes, Nucleic Acids Res. 46 (2018) W296–W303.

[22] J.A. Newman, A. Douangamath, S. Yadzani, Y. Yosaatmadja, A. Aimon, J. Brandao-Neto, L. Dunnett, T. Gorrie-Stone, R. Skyner, D. Fearon, M. Schapira, F. von Delft, O. Gileadi, Structure, mechanism and crystallographic fragment screening of the SARS-CoV-2 NSP13 helicase, Nat. Commun. 12 (2021) 4848.

[23] E. Delagoutte, P.H. von Hippel, Helicase mechanisms and the coupling of helicases within macromolecular machines. Part I: structures and properties of isolated helicases, Q. Rev. Biophys. 35 (2002) 431–478.

[24] A. Datta, R.M. Brosh Jr., New insights into DNA helicases as druggable targets for cancer therapy, Front. Mol. Biosci. 5 (2018) 59.

[25] M.U. Mirza, M. Froeyen, Structural elucidation of SARS-CoV-2 vital proteins: computational methods reveal potential drug candidates against main protease, Nsp12 polymerase and Nsp13 helicase, J. Pharm. Anal. (2020).

[26] M.A. White, W. Lin, X. Cheng, Discovery of COVID-19 inhibitors targeting the SARS-CoV-2 Nsp13 helicase, J. Chem. Phys. Lett. 11 (2020) 9144–9151.

[27] J.F. Borgio, H.S. Alsuwat, W.M. Al Otaibi, A.M. Ibrahim, N.B. Almandil, L.I. Al Asoom, M. Salahuddin, B. Kamaraj, S. AbdulAzeez, State-of-the-art tools unveil potent drug targets amongst clinically approved drugs to inhibit helicase in SARS-CoV-2, Arch. Med. Sci. 16 (2020) 508–518.

[28] S. Khater, G. Das, Repurposing Ivermectin to inhibit the activity of SARS CoV2 helicase: possible implications for COVID 19 therapeutics, OSF Preprints (2020).

[29] R. Adekunle, O. Olukemi, A. Titilayo, O. Amoge, B. Doofan, N. Ezinne, O. Uchenna, I. Zainab, H. Jamila, U. Adaku, Silico identification of potential allosteric inhibitors of the SARS-CoV-2 helicase, Trop. J. Nat. Prod. Res. 5 (2020) 165–177.

[30] G. Ligat, S. Da Re, S. Alain, S. Hantz, Identification of amino acids essential for viral replication in the HCMV helicase-primase complex, Front. Microbiol. 9 (2018) 2483.

[31] B.R. Beck, B. Shin, Y. Choi, S. Park, K. Kang, Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model, Comput. Struct. Biotechnol. J. 18 (2020) 784–790.

[32] M.I. Zimmerman, J.R. Porter, M.D. Ward, S. Singh, N. Vithani, A. Meller, U. L. Mallimadugula, C.E. Kuhn, J.H. Borowsky, R.P. Wiewiora, M.F.D. Hurley, A. M. Harbison, C.A. Fogarty, J.E. Coffland, E. Fadda, V.A. Voelz, J.D. Chodera, G. R. Bowman, SARS-CoV-2 simulations go exascale to capture spike opening and reveal cryptic pockets across the proteome, bioRxiv (2020), 2020.06.27.175430.

[33] Z. Wang, J.D. Huang, K.L. Wong, P.G. Wang, H.J. Zhang, J.A. Tanner, O. Spiga, A. Bernini, B.J. Zheng, N. Niccolai, On the mechanisms of bananin activity against severe acute respiratory syndrome coronavirus, FEBS J. 278 (2011) 383–389.

[34] Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, Chem. Phys. Lett. 314 (1999) 141–151.

[35] D.M. Zuckerman, Equilibrium sampling in biomolecular simulations, Annu. Rev. Biophys. 40 (2011) 41–62.

[36] H. Fukunishi, O. Watanabe, S. Takada, On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: application to protein structure prediction, J. Chem. Phys. 116 (2002) 9058–9067.

[37] M. Smith, J.C. Smith, Repurposing Therapeutics for COVID-19: Supercomputer-based Docking to the SARS-CoV-2 Viral Spike Protein and Viral Spike Protein-Human ACE2 Interface, 2020. ChemRxiv.

[38] X. Periole, A.E. Mark, Convergence and sampling efficiency in replica exchange simulations of peptide folding in explicit solvent, J. Chem. Phys. 126 (2007), 014903.

[39] S.W. Rick, Replica exchange with dynamical scaling, J. Chem. Phys. 126 (2007), 054102.

[40] S.W. Rick, G.J. Schwing, C.M. Summa, An implementation of replica exchange with dynamical scaling for efficient large-scale simulations, J. Chem. Inf. Model. 61 (2021) 810–818.

[41] S. Piana, K. Lindorff-Larsen, David E. Shaw, How robust are protein folding simulations with respect to force field parameterization? Biophys. J. 100 (2011) L47–L49.

[42] P.E.M. Lopes, O. Guvench, A.D. MacKerell Jr., Current status of protein force fields for molecular dynamics simulations, Methods Mol. Biol. 1215 (2015) 47–71.

[43] J.W. Ponder, D.A. Case, Force fields for protein simulations, Adv. Protein Chem. 66 (2003) 27–85.

[44] P. Robustelli, S. Piana, D.E. Shaw, Developing a molecular dynamics force field for both folded and disordered protein states, Proc. Natl. Acad. Sci. U. S. A. 115 (2018) E4758–E4766.

[45] A. Kuzmanic, R.B. Pritchard, D.F. Hansen, F.L. Gervasio, Importance of the force field choice in capturing functionally relevant dynamics in the von Willebrand factor, J. Phys. Chem. Lett. 10 (2019) 1928–1934.

[46] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, Proteins 65 (2006) 712–725.

[47] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D. C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, J. Am. Chem. Soc. 117 (1995) 5179–5197.

[48] R.B. Best, G. Hummer, Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides, J. Phys. Chem. B 113 (2009) 9004–9015.

[49] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J.L. Klepeis, R.O. Dror, D. E. Shaw, Improved side-chain torsion potentials for the Amber ff99SB protein force field, Proteins 78 (2010) 1950–1958.

[50] R.B. Best, D. de Sancho, J. Mittal, Residue-specific alpha-helix propensities from molecular simulation, Biophys. J. 102 (2012) 1462–1467.

[51] M.B. Peters, Y. Yang, B. Wang, L. Fusti-Molnar, M.N. Weaver, K.M. Merz Jr., Structural survey of zinc containing proteins and the development of the zinc AMBER force field (ZAFF), J. Chem. Theor. Comput. 6 (2010) 2935–2947.

[52] M. McGann, FRED pose prediction and virtual screening accuracy, J. Chem. Inf. Model. 51 (2011) 578–596.

[53] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J. Comput. Chem. 25 (2004) 1157–1174.

[54] A. Jakalian, D.B. Jack, C.I. Bayly, Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation, J. Comput. Chem. 23 (2002) 1623–1641.

[55] D.A. Case, R.M. Betz, D.S. Cerutti, T.E. Cheatham, T.A. Darden, R.E. Duke, T. J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, P.A. Kollman, AMBER 2016, University of California, San Fransisco, 2016.

[56] J. Wang, W. Wang, P.A. Kollman, D.A. Case, Automatic atom type and bond type perception in molecular mechanical calculations, J. Mol. Graph. Model. 25 (2006) 247–260.

[57] A.W. Sousa da Silva, W.F. Vranken, ACPYPE - AnteChamber PYthon parser interfacE, BMC Res. Notes 5 (2012) 367.

[58] W. Zheng, M. Andrec, E. Gallicchio, R.M. Levy, SImulatung replica exchange simulations of protein folding with a kinetic network model, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 15340–15345.

[59] H. Nymeyer, How efficient is replica exhange molecular dynamics? An analytic approach, J. Chem. Theor. Comput. 4 (2008) 626–636.

[60] J.Z. Ruscio, N.L. Fawzi, T. Head-Gordon, How hot? Systematic convergence of the replica exchange method using multiple reservoirs, J. Comput. Chem. 31 (2009) 620–627.

[61] E. Rosta, G. Hummer, Error and efficiency of replica exchange molecular dynamics simulations, J. Chem. Phys. 131 (2009) 165102.

[62] L. Heo, M. Feig, Modeling of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) proteins by machine learning and physics-based refinement, bioRxiv (2020), 2020.03.25.008904.

[63] A. Acharya, R. Agarwal, M.B. Baker, J. Baudry, D. Bhowmik, S. Boehm, K.G. Byler, S.Y. Chen, L. Coates, C.J. Cooper, O. Demerdash, I. Daidone, J.D. Eblen, S. Ellingson, S. Forli, J. Glaser, J.C. Gumbart, J. Gunnels, O. Hernandez, S. Irle, D. W. Kneller, A. Kovalevsky, J. Larkin, T.J. Lawrence, S. LeGrand, S.H. Liu, J. C. Mitchell, G. Park, J.M. Parks, A. Pavlova, L. Petridis, D. Poole, L. Pouchard, A. Ramanathan, D.M. Rogers, D. Santos-Martins, A. Scheinberg, A. Sedova, Y. Shen, J.C. Smith, M.D. Smith, C. Soto, A. Tsaris, M. Thavappiragasam, A. F. Tillack, J.V. Vermaas, V.Q. Vuong, J. Yin, S. Yoo, M. Zahran, L. Zanetti-Polzi, Supercomputer-based ensemble docking drug discovery pipeline with application to covid-19, J. Chem. Inf. Model. 60 (2020) 5832–5852.

[64] A. Patriksson, D. van der Spoel, A temperature predictor for parallel tempering simulations, Phys. Chem. Chem. Phys. 10 (2008) 2073–2077.

[65] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX 1–2 (2015) 19–25.

[66] SWISS-MODEL. https://swissmodel.expasy.org/interactive/N2NgU3/models/02.

[67] R. Anandakrishnan, B. Aguilar, A.V. Onufriev, H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations, Nucleic Acids Res. 40 (2012) W537–W541.

[68] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: a linear constraint solver for molecular simulations, J. Comout. Chem. 18 (1997) 1463–1472.

[69] S. Miyamoto, P.A. Kollman, Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models, J. Comput. Chem. 13 (1992) 952–962.

[70] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys. 103 (1995) 8577–8593.
[71] G. Bussi, D. Donadio, M. Parrinello, Canonical sampoing through velocity rescaling, J. Chem. Phys. 126 (2007), 014101.
[72] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: a new molecular dynamics method, J. Appl. Phys. 52 (1988) 7182–7190.
[73] X. Daura, W.F. van Gunsteren, A.E. Mark, Folding-unfolding thermodynamics of a beta-heptapeptide from equilibrium simulations, Proteins 34 (1999) 269–280.
[74] T. Ichiye, M. Karplus, Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations, Proteins 11 (1991) 205–217.
[75] L. Jendele, R. Krivak, P. Skoda, M. Novotny, D. Hoksza, PrankWeb: a web server for ligand binding site prediction and visualization, Nucleic Acids Res. 47 (2019) W345–W349.
[76] R. Krivak, D. Hoksza, P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure, J. Cheminf. 10 (2018) 39.
[77] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2010) 455–461.
[78] P.A. Novick, O.F. Ortiz, J. Poelman, A.Y. Abdulhay, V.S. Pande, SWEETLEAD: an in silico database of approved drugs, regulated chemicals, and herbal isolates for computer-aided drug discovery, PLoS One 8 (2013), e79568.
[79] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, Open Babel: an open chemical toolbox, J. Cheminf. 3 (2011) 33.
[80] L. Yan, Y. Zhang, J. Ge, L. Zheng, Y. Gao, T. Wang, Z. Jia, H. Wang, Y. Huang, M. Li, Q. Wang, Z. Rao, Z. Lou, Architecture of a SARS-CoV-2 mini replication and transcription complex, Nat. Commun. 11 (2020) 5874.
[81] S.L. Badshah, N. Ahmad, A. Ur Rehman, K. Khan, A. Ullah, A. Alsayari, A. B. Muhsinah, N.M. Y, Molecular docking and simulation of Zika virus NS3 helicase, BMC Chem. 13 (2019) 67.
[82] W. Ma, K.D. Whitley, Y.R. Chemla, Z. Luthey-Schulten, K. Schulten, Free-energy simulations reveal molecular mechanism for functional switch of a DNA helicase, Elife 7 (2018).
[83] M.A. Kastenholz, M. Pastor, G. Cruciani, E.E. Haaksma, T. Fox, GRID/CPCA: a new computational tool to design selective ligands, J. Med. Chem. 43 (2000) 3033–3044.
[84] I.R. Craig, C. Pfleger, H. Gohlke, J.W. Essex, K. Spiegel, Pocket-space maps to identify novel binding-site conformations in proteins, J. Chem. Inf. Model. 51 (2011) 2666–2679.
[85] N. Desdouits, M. Nilges, A. Blondel, Principal Component Analysis reveals correlation of cavities evolution and functional motions in proteins, J. Mol. Graph. Model. 55 (2015) 13–24.
[86] D.B. Kokh, S. Richter, S. Henrich, P. Czodrowski, F. Rippmann, R.C. Wade, TRAPP: a tool for analysis of transient binding pockets in proteins, J. Chem. Inf. Model. 53 (2013) 1235–1252.
[87] J.R. Wagner, J. Sorensen, N. Hensley, C. Wong, C. Zhu, T. Perison, R.E. Amaro, POVME 3.0: software for mapping binding pocket flexibility, J. Chem. Theor. Comput. 13 (2017) 4584–4592.