A proteomics-MM/PBSA dual approach for the analysis of SARS-CoV-2 main protease substrate peptide specificity

Gloria Gallo [a], Uilla Barcick [a], Camila Coelho [a], Murilo Salardani [a], Maurício F. Camacho [a], Daniela Cajado-Carvalho [b], Flávio V. Loures [a], Solange M.T. Serrano [b], Leon Hardy [c], André Zelanis [a], Martin Würtele [a,*]

[a] *Department of Science and Technology, Federal University of São Paulo, São José dos Campos, Brazil*
[b] *Laboratory of Applied Toxicology, Center of Toxins, Immune-Response and Cell Signaling (CeTICS), Butantan Institute, São Paulo, Brazil*
[c] *Department of Physics, University of South Florida, Tampa, United States*

ARTICLE INFO

ABSTRACT

The main protease M^{pro} of SARS-CoV-2 is a well-studied major drug target. Additionally, it has been linked to this virus' pathogenicity, possibly through off-target effects. It is also an interesting diagnostic target. To obtain more data on possible substrates as well as to assess the enzyme's primary specificity a two-step approach was introduced. First, Terminal Amine Isobaric Labeling of Substrates (TAILS) was employed to identify novel M^{pro} cleavage sites in a mouse lung proteome library. In a second step, using a structural homology model, the MM/PBSA variant MM/GBSA (Molecular Mechanics Poisson-Boltzmann/Generalized Born Surface Area) free binding energy calculations were carried out to determine relevant interacting amino acids. As a result, 58 unique cleavage sites were detected, including six that displayed glutamine at the P1 position. Furthermore, modeling results indicated that M^{pro} has a far higher potential promiscuity towards substrates than expected. The combination of proteomics and MM/PBSA modeling analysis can thus be useful for elucidating the specificity of M^{pro}, and thus open novel perspectives for the development of future peptidomimetic drugs against COVID-19, as well as diagnostic tools.

1. Introduction

SARS-CoV-2 spread worldwide rapidly, initiating the COVID-19 pandemic. Although several effective vaccines and small molecule inhibitors were rapidly developed, COVID-19 remains of great concern to global public health [1,2]. Due to several issues, including the emergence of new variants of SARS-CoV-2, global efforts for drug discovery to address the disease continue being of uttermost importance.

SARS-CoV-2 is an enveloped (+) single-stranded RNA coronavirus. Its genome size is about 29.9 kb in length. About two thirds of SARS-CoV-2 genome is occupied by orf1ab that encodes the non-structural proteins, while the remaining region next to the 3' end encodes the structural proteins [3]. Orf1ab is translated into two polyproteins. They

are processed by the virus's main protease M^{pro} (also called 3CL^{pro}) and a second papain-like protease (PL^{pro}) [4]. The structure of M^{pro} from SARS-CoV-2, a protein with 96% sequence identity to M^{pro} from SARS-CoV, was recently solved [5,6].

Because of its essential activity in the viral replication machinery, M^{pro}, a cysteine protease, is an excellent target for the rational development of SARS-CoV-2 medications [7–11]. Additionally, it is considered an interesting marker for diagnostic purposes [12,13]. Its off-target activity is especially intriguing, as it could be linked to SARS-CoV-2 pathogenicity [14]. For all these reasons, we here have opted for a dualistic approach. First, we employed Terminal Amine Isobaric Labeling of Substrates (TAILS; [15]) for the identification of M^{pro} cleavage sites in a complex proteome (mouse lung proteome, as described

herein). Then, as a second step, molecular modelling techniques via MM/PBSA free energy calculations were carried out to determine the amino acids that are most relevant for the interaction with protein substrates.

TAILS is an N-terminomic analytical approach that enables the profiling of the substrate repertoire of a given protease (degradome) in complex biological mixtures, therefore allowing subsite mapping based in the use of mass spectrometry and bioinformatics for the identification of (new) N-termini generated by active proteases. Since the protease is incubated with native substrates in complex biological mixtures such as cell lysates, long-range interactions (e.g. the interaction of protease exosites with their substrates) are potentially allowed under assay conditions. Such a feature is the main advantage of TAILS over approaches involving the incubation of protease with peptide libraries, in which mainly short-range interactions are allowed. Briefly, using dimethyl labeling, prime-side peptides (i.e those that were generated by active proteases in the sample) are labeled at their N-termini, therefore indicating the cleavage site. After the addition of a protease such as trypsin, peptides bearing free N-termini are produced. This latter set of peptides is then sulfo-modified and pulled out from the sample after strong cation exchange chromatography. The resulting peptide set (generated by active proteases in the sample) is submitted to mass spectrometry followed by bioinformatic analyses for protease specificity profiling.

The MM/PBSA and its variant MM/GBSA (Molecular Mechanics Poisson-Boltzmann Surface Area and Molecular Mechanics Generalized Born Surface Area) are popular methods used to determine the free binding energy between a protein and a ligand [16–19]. Essentially, these methods evaluate the free binding energy of the ligand from MD simulation snapshots using energy terms of the MD force field, but approximating the polar energy terms with the Poisson-Boltzmann (PB) equation, estimating the non-polar energy terms using the solvent-accessible surface area (SASA) method and roughly approximating the entropic terms using normal-mode analysis. They have proven to be reliable methods for the computation of binding energy leading to numerous applications in small-molecule ligand binding and drug design [20–22], protein-peptide interactions [23–26], protein-protein interactions [27,28] and protein design [29,30]. Although their accuracy in some systems is in the order of 10 kJ/mol, both computational improvements in the calculation of individual terms of the binding energy as well as carrying out many short MD simulations [31,32] has helped obtain more accurate estimations of binding energies up to the 1 kJ/mol range. As a whole, the accuracy of MM/PBSA methods ranks in between more simple semi-empirical molecular docking calculations and more computational demanding alchemical perturbation (AP) methods [33].

2. Material and methods

2.1. Protein expression and purification

SARS-CoV-2 M^{pro} (GenBank entry MT358641.1) was synthesized (GenScript, USA) and cloned into the pET21a plasmid to create an expression construct with an N-terminal cleavable His-tag for recombinant expression in *E.coli* BL21(DE3), purification by affinity chromatography and size exclusion chromatography after cleavage of the His-tag using TEV protease, as previously described [34].

2.2. N-terminome analysis - terminal amine isotopic labeling of substrates (TAILS)

N-terminome analysis was carried out using the Terminal Amine Isotopic Labeling of Substrates (TAILS) protocol [15] with modifications [35]. A lung lysate from C57BL/6 male mice (n = 2) was obtained after submitting the tissue samples to 3 cycles of 16,000 rpm of homogenization in a homogenizer (Bio-Gen PRO200 homogenizer PRO Scientific,

Oxford CT, USA), on ice bath using a lysis buffer (50 mM HEPES 150 mM NaCl 1 mM EDTA, pH 7.5). The tissue lysate was cleared by centrifugation (14 000 x g, 10 min, 4 °C) and the protein content of the supernatant was measured using the Bradford method. Five hundred micrograms of proteins were incubated for 12 h with 10 μg of recombinant M^{pro} or as the control with 50 mM HEPES buffer alone (pH 7.5) at 37 °C. Guanidine hydrochloride (GuHCl) was added to the mixture to a final concentration of 3 M, followed by the addition of dithiothreitol (DTT) to a final concentration of 5 mM. The mixture was incubated at 37 °C for 60 min. Iodoacetamide (IAA) was then added to a final concentration of 15 mM and the samples were incubated for 30 min at room temperature, in the dark. To quench the excess of IAA, DTT was added to a final concentration of 15 mM. N-termini were differentially labeled via stable-isotope reductive dimethylation with either light (for controls) or heavy (for M^{pro} experiments) formaldehyde solutions. Protein samples from each experimental condition were incubated overnight at 37 °C with either light or heavy sodium cyanoborohydride (NaBH$_3$CN, light, or NaBD$_3$CN, heavy) to a final concentration of 20 mM followed by the addition of formaldehyde $^{12}CH_2O$ (light) or $^{13}CD_2O$ (heavy) to a final concentration of 40 mM, resulting in mass differences of + 28.031300 Da and + 36.075670 Da for the light and heavy-labeled samples, respectively. The reaction was terminated by adding 1 M Tris (pH 6.8; to a final concentration of 100 mM) to each sample and the mixture was incubated for 2 h at 37 °C. Samples were then combined at a 1:1 ratio. Clean-up of samples was performed by the addition of ice-cold acetone (8 volumes) and methanol (1 vol), followed by the incubation of samples for 3 h at − 80 °C. After centrifugation at 14,000 g for 10 min, 4 °C, protein pellets were washed twice with one volume of ice-cold methanol and then resolubilized with 100 mM NaOH solution (final concentration of 2.5 mM), followed by the addition of HEPES buffer, pH 7.5, to a final concentration of 25 mM in 500 μL of reaction solution. Trypsin (Proteomics grade, Sigma, USA) was added at a 1:100 ratio (enzyme/substrate) and the mixture was incubated at 37 °C for 18 h. N-terminal peptides were enriched by negative selection using sodium 4-formylbenzene-1,3-disulfonate as described [35]. The peptide mixture was subjected to a three-step disulfomodification procedure [35]. Briefly, 4-formylbenzene − 1,3-disulfonate acid (4FDA) was added to the peptide mixture solution to a final concentration of 20 mM, followed by the addition of NaBH$_3$CN 20 mM (final concentration). The reaction was incubated at room temperature for 1 h. Two subsequent steps of incubation with 40 mM of 4FDA and 40 mM of NaBH$_3$CN, and 60 mM of 4FDA and 60 mM of NaBH$_3$CN were performed for 1 h and 18 h, respectively. The reaction was terminated by adding 1 M Tris (pH 6.8) to a final concentration of 100 mM, for 1 h at room temperature. After desalting using C-18 cartridges (3 M Empore™, SPE Extraction disks, USA) peptide samples were dried in a SpeedVac, resuspended in 300 μL of 1% acetic acid and 15% acetonitrile (ACN) and subjected to strong cation exchange (SCX) using SCX StageTips [36]. While disulfomodified (tryptic) peptides were mainly eluted in the flowthrough, N-terminal blocked peptides (either by chemical dimethylation or naturally acetylated), corresponding to natural or neo-N-terminal peptides, were eluted in two fractions collected after the addition of 200 mM and 500 mM of NaCl in 1% acetic acid 15% ACN, respectively. Finally, fractions were desalted using C18 StageTips, dried in a SpeedVac and redissolved in 50 μL of 0.1% formic acid prior to nanoflow liquid chromatography/-tandem mass spectrometry (LC−MS/MS) analysis.

2.3. Mass spectrometric analysis

An aliquot (5 μL) of the desalted peptide mixture was injected into a trap column packed with C18 (100 μm i.d. × 2 cm) for desalting with 100% solvent A (0.1% formic acid). Peptides were then eluted onto an analytical column (75 μm i.d. × 100 mm) packed *in house* with Aqua® C-18 5 μm beads (Phenomenex, USA). Nanoflow liquid chromatography was performed on an Easy nanoLC system (Thermo Fisher Scientific, USA) coupled to an LTQ-Orbitrap Velos mass spectrometer (Thermo

Table 1

Cleavage sites displaying the canonical specificity of M^pro identified by TAILS analysis. Six peptides spanning the canonical Gln at P1 site were identified by TAILS.

Cleavage site (P5-P5')[a]	Identified peptide (P1'-Pn')	Gene name	UniProt Accession	Protein name
GAEIQ↓DGRFN	DGRFNLFKVQQGR	FIP1	Q9D824–4	Pre-mRNA 3'-end-processing factor FIP1
GETEQ↓KRIRK	KRIRKKKAKKR	CA131	Q8CIL4	Uncharacterized protein C1orf131 homolog
SSPSQ↓SGDTQ	SGDTQTFAQKLQLR	GOGA4	Q91VW5	Golgin subfamily A member 4
QLNSQ↓LFVGG	LFVGGKSSRQKGFFGCIR	CTP5C	Q0V8T7	Contactin-associated protein like 5–3
KRSQQ↓EDDQE	EDDQEFFEDR	FBX43	Q8CDI2	F-box only protein 43
SSSVQ↓IDPPL	IDPPLSSWKDLRTFKQR	COSA1	Q2UY11	Collagen alpha-1(XXVIII) chain

[a] Since TAILS approach targets prime-side peptides, non-prime side positions (i.e. P5 to P1) were bioinformatically inferred from peptide identifications, after database searches using the MS data.

Fisher Scientific, USA). Peptides were loaded onto the column with solvent A (0.1% formic acid) and eluted with a 90 min linear gradient from 3% to 30% of solvent B (0.1% formic acid in acetonitrile) at a flow rate of 200 nL/min. Spray voltage was set at 2.1 kV, 200 °C and the mass spectrometer was operated in data dependent mode, in which one full MS scan was acquired in the m/z range of 400–1800 followed by MS/MS acquisition using Collisional Induced Dissociation (CID) of the ten most intense ions from the MS scan. MS spectra were acquired in the Orbitrap analyzer at 60,000 resolution (at 400 m/z). Dynamic exclusion was defined by a list size of 500 features and exclusion duration of 30 s. For the survey (MS) scan an AGC (Automatic Gain Control) target value of 1000,000 was set whereas the AGC target value for the fragment ion (MS/MS) spectra was set to 10,000 ions. The lower threshold for targeting precursor ions in the MS scans was 2000 counts.

2.4. Proteomics data processing

Mass spectrometric (RAW) data were analyzed within the Trans Proteomics Pipeline platform [37] (v.4.8;Build 201411201551–6764). Briefly, RAW files were converted to the mzXML file format and searched with Comet search engine [38] (version 2017.01, rev. 2) against the UniProt/SwissProt database restricted to the taxonomy 'Rodentia' (release 2021_04; 26,961 entries). Peptide identification was based on a search with mass deviation of the precursor ion of 20 ppm and the fragment mass tolerance was set to 0.4 Da. As primary amine dimethylation prevents the cleavage of trypsin at dimethylated-lysine residues, enzyme specificity was set to semi Arg-C and at least two missed cleavages were allowed. Separate searches were carried out to account for the two biological samples/labeling states that were used (control and protease- added, light and heavy dimethyl-labeled peptides, respectively). Free N-terminal peptides searching was carried out by selecting the light (+28.03 Da) and heavy (+36.07 Da) dimethylation as fixed modifications at peptide N-terminus and Lysine sidechains, depending on the biological sample analyzed (control or protease-added). Therefore, under the assay conditions, light and heavy singletons identified by tandem mass spectrometry represent background proteolysis or protease-generated peptides, respectively. For all searches cysteine carbamidomethylation was selected as fixed modification whereas methionine oxidation, glutamine/asparagine deamidation were selected as variable modifications. Peptide identification was accepted after estimating the False Discovery Rate calculated based on the score distributions in the output of the Comet search engine for each biological replicate. Search results were filtered with PeptideProphet to a > 99% confidence interval, corresponding to a False Discovery Rate (FDR) of less than 1%.

2.5. Homology modelling

SARS-CoV-2 M^pro models were constructed using homology modeling with the MODELLER software [39] using a M^pro peptide complex structure (PDB ID 2Q6G) [40] as a template. This structure contains two chains of SARS-CoV M^pro proteins in complex with two TSAVLQSGFRK peptides. For homology modelling, both protomers of the M^pro dimer of SARS-CoV-2 were restrained on their C_α atoms to be

identical. Similarly, peptides were homology modeled using the peptides from 2Q6G as a template and their C_α atoms restrained to be identical.

2.6. 2.6 Molecular dynamics (MD)

MD simulations were carried out using the AMBER simulation software package [41]. The model was prepared with tleap using the ff99 AMBER force field, neutralized with Na^+ ions and solvated with the TIP3P water model in a cuboid integration box with 12 Å solvent margins. After structure minimization, a short heating run and a density equilibration run, several parallel 10 ns MD production runs were conducted with AMBER. Production runs were carried out with the GPU-version of pmemd at 300 K in time steps of 2 fs using a constant pressure with isotropic position scaling, Particle Mesh Ewald (PME) model of periodic boundary conditions with an 8 Å classical non-bonded cut-off, a Langevin thermostat and the SHAKE algorithm to constrain the bonds of all hydrogen atoms.

2.7. MM/GBSA

Essentially, a molecular dynamics (MD) simulations of protein-peptide complexes were carried out, and from single snapshots of the MD-trajectory, the binding free energy of the peptides to the protein ΔG_{bind} was calculated from protein-peptide complex MD simulations as an averaged difference between the free energies of the complex state $\langle G_{PL} \rangle$ and unbound states of the protein $\langle G_P \rangle$ and the peptide ligand $\langle G_L \rangle$:

$$\Delta G_{bind} = \langle G_{PL} \rangle - \langle G_P \rangle - \langle G_L \rangle$$

The states were evaluated using following approximation:

$$G = E_{bonded} + E_{el} + E_{VdW} + G_{polar} + G_{nonpolar} - TS$$

In the second equation, E_{bonded} is the bonded, E_{el} is the electrostatic and E_{VdW} is the Van der Waals energy component from the MM force field. G_{polar} is the polar solvation energy component. G_{polar} is approximated by using the generalized Born (GB) model approximation for solving the Poisson-Boltzmann (PB) equation. $G_{nonpolar}$ is the non-polar solvation energy contribution and is estimated using a linear approximation including the solvent accessible surface area (SASA). Finally, *TS* is the entropic term, which was approximated using normal mode analysis [41–43]. In detail, the MM/GBSA calculations were carried with the AMBER MMPBSA.py script [41] using the "OBC" model (igb=2) [44] together with mbondi2 atomic radii. At least 20 independent 10 ns production runs were carried out for each peptide. For each run, 25 snapshots for ΔH_{GBSA} and 5 snapshots for $T\Delta S$ calculations were evaluated. After averaging results, error values were calculated as SEM (Standard Error of the Mean) from the independent runs.

2.8. Ethics statement and animal model

Animal experiments were performed in accordance with the Brazilian Federal Law 11,794 establishing procedures for the scientific use of animals in accordance with the principles outlined by the Brazilian

Table 2

MM/GBSA analysis of M^{pro} binding peptides. The main energy components from the MM/GBSA calculations ΔH_{GBSA}, the conformational entropy term $T\Delta S$, as estimated with normal mode analysis and the total free binding energy ΔG are given in kcal/mol.

Peptide	ΔH_{GBSA} (kcal/mol)	TΔS (kcal/mol)	ΔG (kcal/mol)
GAEIQDGRFN	-64.5 + /- 0.9	-37.9 + /- 1.2	-26.6 + /- 1.3
GETEQKRIRK	-74.3 + /- 1.1	-47.5 + /- 1.0	-26.8 + /- 1.6
KRSQQEDDQE	-73.7 + /- 1.2	-45.3 + /- 1.1	-28.4 + /- 1.3
QLNSQLFVGG	-78.6 + /- 1.0	-42.7 + /- 1.6	-35.9 + /- 1.7
SSPSQSGDTQ	-56.2 + /- 1.1	-39.9 + /- 1.5	-16.3 + /- 1.6
SSSVQIDPPL	-62.2 + /- 0.7	-37.7 + /- 1.1	-24.5 + /- 1.4

College of Animal Experimentation (CONCEA), and the State Law establishing the Animal Protection Code of the State of São Paulo. All protocols adopted in this study were approved by the local ethics committee for animal experiments from the Federal University of São Paulo (Permit Number: 9552180122). All efforts were made to minimize suffering. Animals (C57BL/6 mouse) were sacrificed by intraperitoneal (i.p.) injection of ketamine and xylazine.

3. Results and discussion

3.1. M^{pro} specificity screening using TAILS

The N-terminomic analysis of the proteolytic activity of M^{pro} upon a mouse lung proteome library resulted in the identification of 58 unique cleavage sites. Thus, 58 peptides, which were absent in the control condition, were identified as heavy dimethylated-labeled (singleton) peptides (Table S1 and Table S2). Interestingly, out of the 58 unique cleavage sites, six contained Glutamine (Q) at the P1 position (Table 1). The cleaved proteins include contactin-associated protein like 5–3, a protein that is likely involved in cell adhesion and intracellular communication in the central nervous system; FIP1, a protein involved in pre-mRNA 3'-end-processing; F-box only protein 43, which is possibly involved in ubiquitination-mediated regulatory complexes [45]; golgin subfamily A member 4, a protein that is associated with vesicular trafficking by the Golgi apparatus, and collagen alpha-1(XXVIII) chain, which is possibly involved in damage repair processes [46]. These proteins might be *bona fide* substrates of M^{pro}. However, the physiological implications of their cleavage must be further validated. Thus, we chose to focus on structural aspects of M^{pro} subsite specificity.

The primary specificity of M^{pro} for residues containing a Glutamine at the P1 position was also observed in recent reports on M^{pro} cleavage specificity [14,47] and is in accordance with the primary specificity observed for SARS-CoV-2 and SARS-CoV main proteases [5,6,48]. Recent N-terminomics reports on M^{pro} specificity used a protease to proteome ratio ranging from 1:20–1:10 [14,47], whereas in this work we used a protease to proteome ratio of 1:50. This might explain why we identified fewer cleavage sites spanning the canonical M^{pro} preference of Q at P1. Moreover, for the cleavage sites containing Q at P1 position, we did not verify the recurrent presence of L at P2, as recently reported [14, 45]. On the other hand, most cleavage sites (~90%) varied in relation to the residues at the scissile bond (P1-P1' positions), and, notably, they deviated from the canonical presence of Q at P1 position in the set of identified peptides. In this context, we cannot rule out the activation of other proteases of mouse lung tissue upon incubation with M^{pro}, therefore resulting in peptides spanning distinct cleavage sites (i.e. other than Q at P1). Thus, we focused our structural analysis on peptides in which the preference of Q at P1 was observed (Table 1).

3.2. MM/GBSA binding analysis

In order to estimate the free binding energy of the different detected M^{pro} binding peptides, MM/GBSA analysis was performed. To obtain reproducible data several 10 ns MD production runs were carried out.
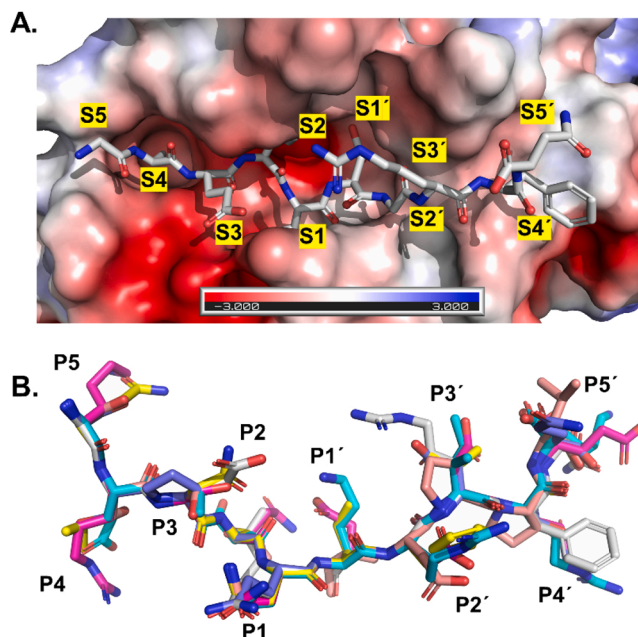

Fig. 1. Structural alignment of M^{pro} peptide complexes from homology modeling. A) The GAEIQDGRFN peptide identified in the TAILS analysis was homology modeled with MODELLER using a SARS-CoV peptide complex structure (PDB ID 2Q6G). The SARS-CoV-2 M^{pro} peptide binding site is shown as electrostatic potential surface. B) Structural superimposition of the six peptides displaying the canonical Gln at P1 position.

The obtained snapshots were then analyzed via AMBER's MMPBSA.py script. As a result, reproducible averaged free energy calculations with SEM values in the order of 1–2 kcal/mol could be obtained. The M^{pro} binding peptides together with their main MM/GBSA energy components are shown in Table 2. The obtained free binding energies are in a similar order of magnitude with those in a recent study of peptides interacting with the Zika virus protease. Interestingly, in that study, a significant correlation between free binding energies calculated with MM/PBSA and experimentally measured Michaelis-Menten constants K_m was obtained [23].

The homology modelled peptides are shown in Fig. 1. Basically, as expected the peptide binding site of M^{pro} shows high structural complementarity to the corresponding peptides, with particularly tight S1, S2 and S4 binding pockets. Contrasting with this, the S1', S2', S4' and particularly the S3 and S3' binding pockets are wider and more open with higher solvent exposure. The corresponding peptides bind with tight interactions to their respective S-sites through their P1, P2 and P4 side chains. The P3 and P5 side chains are mostly oriented towards the solvent. The C-terminal peptide half binds in a more solvent exposed mode, with P1', P2' and P4' interacting with the more open and solvent exposed S1', S2' and S4' sites. The P3' and P5' side chains are mostly solvent exposed.

Concerning the electrostatic potential surface, the N-terminal binding sites show a more negative electrostatic potential at the sites binding the N-terminal portion of the peptide (Fig. 1). Concerning the modeling of electrostatic binding enthalpies in the MM/GBSA studies of the peptides, more favorable electrostatics of peptide binding are mostly compensated by less favorable peptide polar solvation energies (Table 3). This indicates that peptide binding of M^{pro} is not determined exclusively by electrostatics, at least not in the MM/GBSA model applied.

3.3. Binding energy decomposition analysis

To further discern the importance of the amino acids to the free

Table 3

MM/GBSA energy components. Different energy components from the MM/GBSA analysis of the binding of the peptides displaying the canonical Gln at P1 position to SARS-CoV-2 M^{pro} are given in kcal/mol.

| Peptide	MM/GBSA Term	Ligand Energy	Ligand SEM	Receptor Energy	Receptor SEM	Total Energy	Total SEM
GAEIQDGRFN	ΔH_{GBSA}	-25.0	1.3	-39.8	0.6	-64.8	1.9
	$\Delta G_{nonpolar}$	-6.80	0.12	-4.83	0.04	-11.63	0.16
	ΔG_{polar}	38.2	4.3	21.4	1.2	59.7	5.5
	ΔE_{el}	-11.7	4.7	-11.7	1.4	-23.4	6.1
	ΔE_{VdW}	-44.7	0.9	-44.7	0.3	-89.4	1.1
GETEQKRIRK	ΔH_{GBSA}	-23.3	1.3	-51.8	0.6	-75.1	1.9
	$\Delta G_{nonpolar}$	-7.36	0.12	-5.20	0.03	-12.57	0.15
	ΔG_{polar}	160.8	5.4	130.2	1.4	290.9	6.8
	ΔE_{el}	-129.1	5.6	-129.1	1.7	-258.2	7.3
	ΔE_{VdW}	-47.6	0.9	-47.6	0.3	-95.3	1.2
KRSKQEDDQE	ΔH_{GBSA}	-22.5	1.7	-50.8	0.8	-73.4	2.5
	$\Delta G_{nonpolar}$	-7.29	0.11	-5.25	0.02	-12.55	0.13
	ΔG_{polar}	61.0	6.3	30.7	3.5	91.7	9.9
	ΔE_{el}	-26.6	7.1	-26.6	4.2	-53.2	11.2
	ΔE_{VdW}	-49.6	1.1	-49.6	0.3	-99.3	1.5
QLNSQLFVGG	ΔH_{GBSA}	-35.4	1.4	-43.0	0.5	-78.4	2.0
	$\Delta G_{nonpolar}$	-7.16	0.13	-4.95	0.03	-12.12	0.16
	ΔG_{polar}	90.2	2.5	80.5	1.3	170.7	3.8
	ΔE_{el}	-70.2	3.0	-70.2	1.6	-140.3	4.6
	ΔE_{VdW}	-48.4	1.0	-48.4	0.3	-96.7	1.4
SSPSQSGDTQ	ΔH_{GBSA}	-19.1	1.4	-37.0	0.6	-56.1	2.0
	$\Delta G_{nonpolar}$	-6.05	0.13	-3.99	0.03	-10.03	0.16
	ΔG_{polar}	53.1	3.0	33.3	1.3	86.4	4.3
	ΔE_{el}	-29.3	3.5	-29.3	1.6	-58.6	5.1
	ΔE_{VdW}	-36.9	1.1	-36.9	0.4	-73.8	1.5
SSSVQIDPPL	ΔH_{GBSA}	-26.2	1.2	-36.5	0.4	-62.7	1.6
	$\Delta G_{nonpolar}$	-6.28	0.10	-4.51	0.02	-10.79	0.12
	ΔG_{polar}	44.9	2.7	32.7	1.1	77.6	3.8
	ΔE_{el}	-21.7	3.1	-21.7	1.2	-43.4	4.3
	ΔE_{VdW}	-43.0	0.9	-43.0	0.4	-86.1	1.2
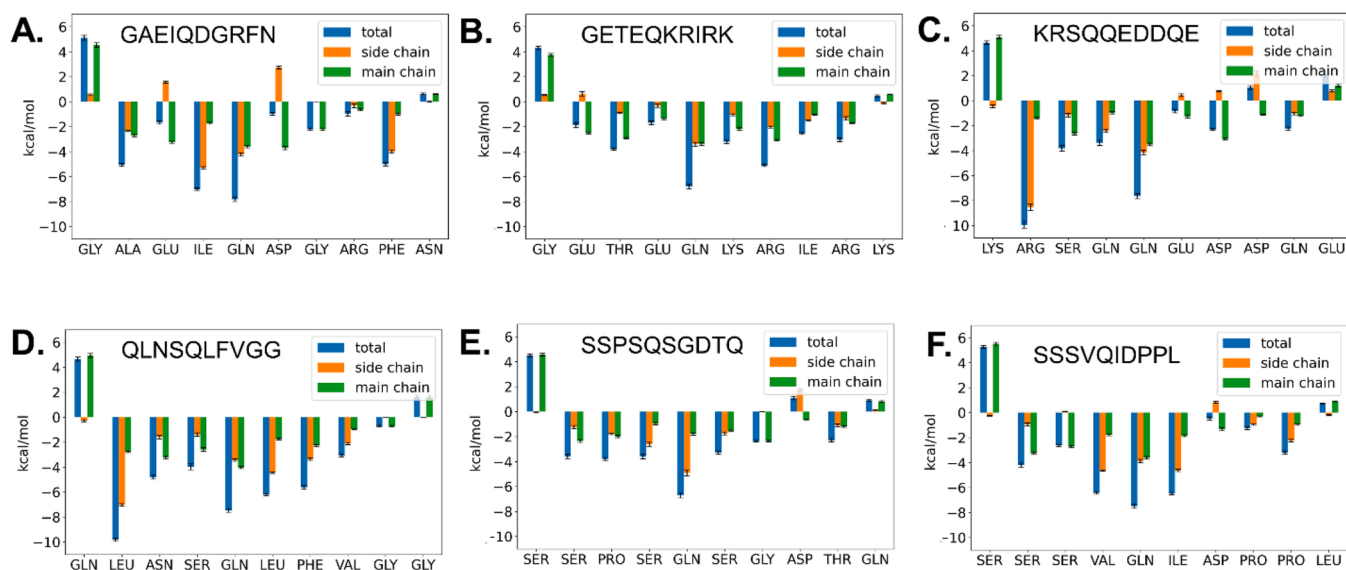
Fig. 2. Per-residue binding energy decomposition of peptides. A) GAEIQDGRFN. B) GETEQKRIRK. C) KRSQQEDDQE. D) QLNSQLFVGG. E) SSPSQSGDTQ. F) SSSVQIDPPL. Using MM/GBSA analysis, the interaction energy of each amino acid with the rest of the system was estimated. The importance of the residues at P1 and P2 positions become apparent. Variations at e.g. P4 and P1' positions could also explain improved binding of the corresponding peptides. Values are given in kcal/mol.

energy of binding, binding energy decomposition analysis was carried out with the AMBER MMPBSA.py script. The obtained results can be considered a measure of the interaction energy of each amino acid with the rest of the system. They are shown in Fig. 2. In this figure, the importance of the P1 and P2 residues, for peptide recognition by M^{pro} become apparent. To complement this analysis, in Fig. 3, the different MM/GBSA energy components, for both the peptides and the protein are highlighted. As shown in Table 2, due to the compensation effect of

electrostatic interaction energy by polar solvent interaction energies, in the employed MM/GBSA model, the binding is determined more by Van der Waals interactions, not electrostatic interactions.

Of the six peptides found in the mass spectrometry analysis containing the canonical Q at the P1 position, the modeled GAEIQDGRFN peptide shows average interacting affinity, as shown by the calculated ΔG value (−26.6 kcal/mol, Table 2). Per-residue binding energy decomposition analysis shows strong interaction of the protein with the
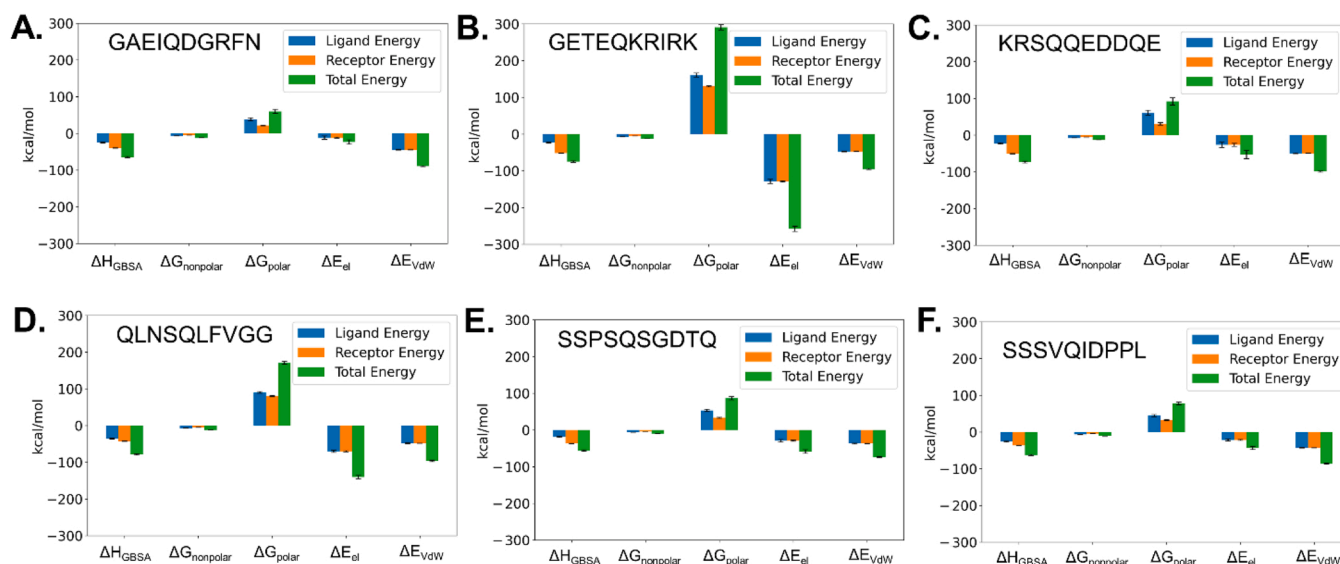
Fig. 3. MM/GBSA energy components of protein and peptide. A) GAEIQDGRFN. B) GETEQKRIRK. C) KRSQQEDDQE. D) QLNSQLFVGG. E) SSPSQSGDTQ. F) SSSVQIDPPL. Polar solvation energy (ΔG_{polar}) represents the electrostatic interaction between the solute and the continuum solvent. It mostly compensates favorable electrostatic interactions between peptide and protein (ΔE_{el}). Values are given in kcal/mol.

P2 Ile and P1 Gln side chains and to a lesser degree with the P4' Phe and P4 Ala side chains (Fig. 2A).

Compared to this peptide, the GETEQKRIRK peptide shows a similar binding affinity ($\Delta G = -26.8$ kcal/mol) in the modeling studies. In this case, the per-residue MM/GBSA energy decomposition analysis indicates that the less favorable substitutions Ile/Glu (at P2), Ala/Glu (P4) and Phe/Arg (P4') are mostly countered by several more favorable substitutions (e.g. Glu/Thr at P3, Asp/Lys at P1' and Gly/Arg at P2', Fig. 2A and B). Analyzing the ligand MM/GBSA energy components, improvement of the electrostatics for binding (ΔE_{el}) is reversed by less favorable polar solvent free energies (ΔG_{polar}, Table 3). On the side of the protein, electrostatics are more favorable in comparison with the GAEIQDGRFN peptide. However, the resulting more favorable MM/GBSA binding energy term (ΔH_{GBSA}) is weakened by a more unfavorable -$T\Delta S$ value (Table 3). Entropy evaluations with MM/PBSA or MM/GBSA methods must be interpreted carefully, due to obvious methodological limitations [16]. In the current analysis, most peptides showed however a similar conformational entropy (-$T\Delta S$) value of about 41.8 kcal/mol. Because this -$T\Delta S$ value is similar, the task of ranking the binding can be thus in most cases simplified in first approximation to comparison of the other binding energy values. The somewhat atypical -$T\Delta S$ value of 47.5 kcal/mol for the GETEQKRIRK peptide is possibly the result of its increased number of large, positively charged residues, which will induce more vibrational and rotational movements, including side-chain rotamer states [49,50]. These states are stabilized upon binding, leading to a reduction of conformational entropy. This effect is expected to be very significant when these residue side-chains bind to the interface of a protein. This is certainly the case of the all the three interacting residues of the KRIRK sequence, except the last lysine, which presumably remains mostly solvent exposed after peptide binding.

Compared with GAEIQDGRFN, the KRSQQEDDQE peptide binding energy decomposition shows binding energy improvements due mostly to the Ala/Arg (at P4) that are however compensated by Ile/Gln(P2) and Phe/Gln (P4') substitutions (Fig. 2C). Similar to GETEQKRIRK, the MM/GBSA energy components (Table 3) indicate that electrostatic improvements on the receptor side are neutralized by entropic disadvantages. As a whole the free energy of binding remains similar to the other described peptides ($\Delta G = -28.4$ kcal/mol).

The QLNSQLFVGG is the peptide with the most favorable ΔG value (-35.9 kcal/mol). Compared with GAEIQDGRFN, the most important favorable substitution is Ala to Leu at P4, which apparently fills the S4

subsite's steric restrains better. The other improvements are at the P1' (Asp to Leu), P2' (Gly to Phe) and P3' (Arg to Val) sites (Fig. 2D). On the other hand, QLNSQLFVGG has some less favorable substitutions at the P2 (Ile to Ser) and P4' (Phe to Gly) positions.

The last two peptides considered in this study are SSPSQSGDTQ ($\Delta G = -16.3$ kcal/mol) and SSSVQIDPPL ($\Delta G = -24.5$ kcal/mol). In the case of the below average ΔH_{GBSA} of SSPSQSGDTQ, MM/GBSA binding energy decomposition indicates most disadvantages in comparison with GAEIQDGRFN are due to an Ile/Ser substitution at P2 position and a Phe/Thr substitution at P4' position (Fig. 2E). In the case of SSSVQIDPPL, which has a slightly less favorable binding enthalpy when compared with GAEIQDGRFN, MM/GBSA binding enthalpy indicates only slight improvements mainly through the P1' Asp to Ile substitution in energy decomposition analysis (Fig. 2F). However, these small improvements are reversed by decrements in negative binding enthalpy in the protein part of the decomposition. These decrements are mostly related to both more unfavorable electrostatics (ΔE_{el} and ΔG_{polar}) and Van der Waals (ΔE_{VdW}) contacts being involved in the binding of the SSSVQIDPPL peptide (Table 3).

In conclusion, peptide binding to the M^{pro} active site does not seem to be driven mainly by electrostatics in any of the studied peptides. Furthermore, although rather specific, the site does show certain potential promiscuity towards binding variations of the peptides. From a structural point of view, the tight binding groove which binds to the N-terminal peptide shows correspondingly higher binding affinities in the MM/GBSA analysis than the more open C-terminal binding sites. As far as entropic calculations are reliable in the MM/GBSA method, the peptides GETEQKRIRK and KRSQQEDDQE, both containing several charged residues, show an above average unfavorable conformational entropy term.

3.4. Peptide design

The obtained results indicate that the ΔH_{GBSA} energy term can be significantly improved by substitutions. These subtle substitutions allowed us, based on this limited dataset, to extrapolate a peptide with the sequence GRNIQIFVFK. Using the MM/GBSA methodology, the following energetic values for this peptide were obtained for this peptide: $\Delta H_{GBSA} = -97.7 +/- 0.2$ kcal/mol, -$T\Delta S = 47.3 +/- 0.2$ kcal/mol, $\Delta G = -50.4 +/- 0.2$ kcal/mol. Peptide binding energy decomposition analysis confirms high energy of binding for the P4, P3, P2, P1,
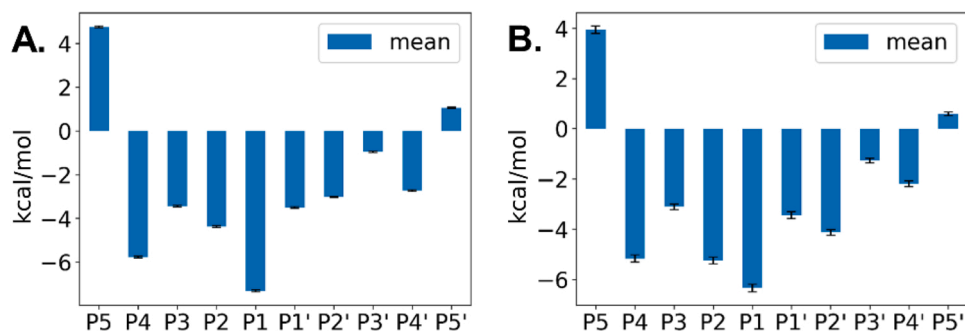
Fig. 4. Binding energy decomposition analysis averaged over all peptide positions. A) The data from the MM/GBSA analysis of the six peptides displaying the canonical Gln at P1 position obtained through TAILS analysis indicate the relative importance of the amino acid residues at the P1, P2 and P4 positions. B) Results from the 51 non-canonical peptides obtained, showing similar per-residue energy decompositions. Results are given in kcal/mol.

P1', P2' and P4' positions (Fig. S1A). Energy components from MM/GBSA indicates that strong electrostatic attraction is compensated by unfavorable polar solvation energy (Fig. S1C). The resulting high binding affinity indicates this approach can indeed be applied to obtain potent M^{pro} binding peptides.

3.5. Non-canonical peptides

Additionally to the canonical peptides, MM/GBSA analysis was carried out with all the non-canonical peptides (i.e. peptides without a Q at P1) identified in TAILS analysis (Table S2 and Table S3). Interestingly, these peptides showed an average ΔG of − 27.3 + /− 1.2 kcal/mol similar to the peptides bearing the canonical cleavage sites in the MM/GBSA (−26.4 +/−0.4 kcal/mol) analysis. The average per-residue energy decomposition for the canonical peptide amino acids is given in Fig. 4A and for the non-canonical peptides in Fig. 4B. This image confirms the importance of the amino acid residues at the P4, P2 and P1 positions. Similar results could be obtained for the canonical as non-canonical peptides. Finally, a decomposition matrix with average decomposition enthalpies (ΔH$_{GBSA}$) was calculated for each of the 20 proteinogenic amino acids at each position (Table S4). Using this kind of analysis, an idealized peptide with the sequence PLIRRKLQYP could be extrapolated. MM/GBSA analysis led to following energetic values for this peptide with above average binding free energy: ΔH$_{GBSA}$ = −94.3 + /- 0.2 kcal/mol -TΔS = 51.8 + /− 0.2 kcal/mol ΔG = −42.5 + /− 0.2 kcal/mol. Energy decomposition analysis indicates high energy of binding for the P4, P3, P2, P1, P2' and P4' positions (Fig. S1B) and the MM/GBSA energy components indicate again that binding is not determined by electrostatic interactions (Fig. S1D). Due to a slightly less favorable binding enthalpy (ΔH$_{GBSA}$, −94.3 vs −97.7 kcal/mol) and entropy term (-TΔS, 51.8 vs 47.3 kcal/mol), possibly due to the presence of several large charged amino acids, this peptide has a lower binding energy than the peptide suggested containing the canonical Gln at position P1. Coincidentally, the results in Table S4 were partially corroborated by experimental peptide library cleavage data [51] and the native viral sequences cleaved by M^{pro} [10] in that e.g. our MM/GBSA studies indicate Leu in position P2 as one of the 3 top binders and Arg is the top binding amino acid in P3.

3.6. Final conclusions

Obtaining kinetical data from binding studies of substrates to proteases must be viewed with caution, as the binding of substrate is not usually the rate limiting step of the reaction. This can be exemplified in an extreme case by the well-studied peptidic protease inhibitors [52]. Another limitation of this study is related to the possibility that some of the detected peptides might have been generated by lung tissue proteases after incubation with M^{pro}. Notwithstanding, the modelling approach of peptides obtained by TAILS analysis described here can be

of significant interest for both the design of novel diagnostic peptides as well as inhibitors of this prominent SARS-CoV-2 drug target. We exemplify this briefly by suggesting two model peptides based on MM/GBSA energy decomposition analysis: The high affinity GRNIQIFVFK peptide and the non-canonical PLIRRKLQYP peptide, which binds somewhat weaker to M^{pro}. This approach can therefore be of use for the development of both potent inhibitors and selective substrates of M^{pro}. The modelling results obtained additionally indicate that the M^{pro} has a far higher potential promiscuity towards substrates than expected. Due to the methodological limitations stated above, this effect needs to be corroborated in further studies.

Funding

CRediT authorship contribution statement

Conception of the work, AZ, MW; experimental procedures, GG, UB, CC, MS, MC, DC, FL; analysis of data, SS, LH, AZ, MW; writing of manuscript, GG, SS, LH, AZ, MW.

Conflict of interest

The authors declare no conflict of interest.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.peptides.2022.170814.

References

[1] A. Hijikata, C. Shionyu, S. Nakae, M. Shionyu, M. Ota, S. Kanaya, T. Shirai, Current status of structure-based drug repurposing against COVID-19 by targeting SARS-CoV-2 proteins, BIOPHYSICS 18 (2021) 226–240.
[2] A. Saxena, Drug targets for COVID-19 therapeutics: ongoing global efforts, J. Biosci. 45 (2020) 87.
[3] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, et al., Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, Lancet 395 (2020) 565–574.
[4] R. Hilgenfeld, From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design, FEBS J. 281 (2014) 4085–4096.
[5] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, et al., Structure of M pro from SARS-CoV-2 and discovery of its inhibitors, Nature (2020) 1–5.
[6] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors, Science 368 (2020) 409–412.

[7] H.M. Mengist, T. Dilnessa, T. Jin, Structural basis of potential inhibitors targeting SARS-CoV-2 main protease, Front. Chem. 9 (2021), 622898.

[8] S. Gao, T. Huang, L. Song, S. Xu, Y. Cheng, S. Cherukupalli, D. Kang, T. Zhao, L. Sun, J. Zhang, P. Zhan, X. Liu, Medicinal chemistry strategies towards the development of effective SARS-CoV-2 inhibitors, Acta Pharm. Sin. B (2021). S2211383521003373.

[9] Sk.A. Amin, S. Banerjee, K. Ghosh, S. Gayen, T. Jha, Protease targeted COVID-19 drug discovery and its challenges: insight into viral main protease (Mpro) and papain-like protease (PLpro) inhibitors, Bioorg. Med. Chem. 29 (2021), 115860.

[10] S. Ullrich, C. Nitsche, The SARS-CoV-2 main protease as drug target, Bioorg. Med. Chem. Lett. 30 (2020), 127377.

[11] E.D. Clercq, G. Li, Approved antiviral drugs over the past 50 years, Clin. Microbiol. Rev. 29 (2016) 695–747.

[12] C. Moore, R.M. Borum, Y. Mantri, M. Xu, P. Fajtová, A.J. O'Donoghue, J.V. Jokerst, Activatable carbocyanine dimers for photoacoustic and fluorescent detection of protease activity, ACS Sens. 6 (2021) 2356–2365.

[13] E.E. Etienne, B.B. Nunna, N. Talukder, Y. Wang, E.S. Lee, COVID-19 biomarkers and advanced sensing technologies for point-of-care (POC) diagnosis, Bioengineering 8 (2021) 98.

[14] I. Pablos, Y. Machado, H.C.R. de Jesus, Y. Mohamud, R. Kappelhoff, C. Lindskog, M. Vlok, P.A. Bell, G.S. Butler, P.M. Grin, Q.T. Cao, J.P. Nguyen, N. Solis, S. Abbina, W. Rut, J.C. Vederas, L. Szekely, A. Szakos, M. Drag, J. N. Kizhakkedathu, et al., Mechanistic insights into COVID-19 by global analysis of the SARS-CoV-2 3CLpro substrate degradome, Cell Rep. 37 (2021), 109892.

[15] O. Kleifeld, A. Doucet, A. Prudova, U. auf dem Keller, M. Gioia, J. N. Kizhakkedathu, C.M. Overall, Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates, Nat. Protoc. 6 (2011) 1578–1611.

[16] S. Genheden, U. Ryde, The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities, Expert Opin. Drug Discov. 10 (2015) 449–461.

[17] W. Wang, P.A. Kollman, Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model, J. Mol. Biol. 303 (2000) 567–582.

[18] J. Srinivasan, T.E. Cheatham, P. Cieplak, P.A. Kollman, D.A. Case, Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate−DNA helices, J. Am. Chem. Soc. 120 (1998) 9401–9409.

[19] T. Tuccinardi, What is the current value of MM/PBSA and MM/GBSA methods in drug discovery? Expert Opin. Drug Discov. 16 (2021) 1233–1237.

[20] H. Sahakyan, Improving virtual screening results with MM/GBSA and MM/PBSA rescoring, J. Comput. Aided Mol. Des. 35 (2021) 731–736.

[21] G. Weng, E. Wang, F. Chen, H. Sun, Z. Wang, T. Hou, Assessing the performance of MM/PBSA and MM/GBSA methods 9 Prediction reliability of binding affinities and binding poses for protein–peptide complexes, Phys. Chem. Chem. Phys. 21 (2019) 10135–10145.

[22] G. Poli, C. Granchi, F. Rizzolio, T. Tuccinardi, Application of MM-PBSA methods in virtual screening, Molecules 25 (2020) 1971.

[23] B. Nutho, T. Rungrotmongkol, Binding recognition of substrates in NS2B/NS3 serine protease of Zika virus revealed by molecular dynamics simulations, J. Mol. Graph. Model. 92 (2019) 227–235.

[24] D.M. Isa, S.P. Chin, W.L. Chong, S.M. Zain, N.A. Rahman, V.S. Lee, Dynamics and binding interactions of peptide inhibitors of dengue virus entry, J. Biol. Phys. 45 (2019) 63–76.

[25] N.G. Azoia, M.M. Fernandes, N.M. Micaêlo, C.M. Soares, A. Cavaco-Paulo, Molecular modeling of hair keratin/peptide complex: using MM-PBSA calculations to describe experimental binding results, Proteins 80 (2012) 1409–1417.

[26] N. Ali, R. Khalil, M. Nur-e-Alam, S. Ahmed, Z. Ul-Haq, Probing the mechanism of peptide binding to REV response element RNA of HIV-1; MD simulations and free energy calculations, J. Biomol. Struct. Dyn. (2020) 1–10.

[27] Y. Sheng, Y. Yin, Y. Ma, H. Ding, Improving the performance of MM/PBSA in protein–protein interactions via the screening electrostatic energy, J. Chem. Inf. Model. 61 (2021) 2454–2462.

[28] R. Tang, P. Chen, Z. Wang, L. Wang, H. Hao, T. Hou, H. Sun, Characterizing the stabilization effects of stabilizers in protein–protein systems with end-point binding free energy calculations, Briefings in Bioinformatics, (2022) bbac127.

[29] E. Wang, H. Sun, J. Wang, Z. Wang, H. Liu, J.Z.H. Zhang, T. Hou, End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design, Chem. Rev. 119 (2019) 9478–9508.

[30] S. Reyaz, A. Tasneem, G.P. Rai, H.R. Bairagya, Investigation of structural analogs of hydroxychloroquine for SARS-CoV-2 main protease (Mpro): A computational drug discovery study, J. Mol. Graph. Model. 109 (2021), 108021.

[31] P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, T.E. Cheatham, Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models, Acc. Chem. Res. 33 (2000) 889–897.

[32] Y. Wang, S. Wu, L. Wang, Z. Yang, J. Zhao, L. Zhang, Binding selectivity of inhibitors toward the first over the second bromodomain of BRD4: theoretical insights from free energy calculations and multiple short molecular dynamics simulations, RSC Adv. 11 (2021) 745–759.

[33] E. Gallicchio, R.M. Levy, Recent theoretical and computational advances for modeling protein–ligand binding affinities, in: Advances in Protein Chemistry and Structural Biology, Elsevier, 2011, pp. 27–80.

[34] C. Coelho, G. Gallo, C. Campos, L. Hardy, M. Wurtele, Biochemical screening for SARS-CoV-2 main protease inhibitors, PLOS ONE (2020).

[35] Z.W. Lai, A. Gomez-Auli, E.J. Keller, B. Mayer, M.L. Biniossek, O. Schilling, Enrichment of protein N-termini by charge reversal of internal peptides, Proteomics 15 (2015) 2470–2478.

[36] J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips, Nat. Protoc. 2 (2007) 1896–1906.

[37] E.W. Deutsch, L. Mendoza, D. Shteynberg, J. Slagel, Z. Sun, R.L. Moritz, Trans-proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics, Prot. Clin. Appl. 9 (2015) 745–754.

[38] J.K. Eng, T.A. Jahan, M.R. Hoopmann, Comet: An open-source MS/MS sequence database search tool, Proteomics 13 (2013) 22–24.

[39] A. Fiser, A. Šali, Modeller: Generation and Refinement of Homology-Based Protein Structure Models, in: Methods in Enzymology, Elsevier, 2003, pp. 461–491.

[40] X. Xue, H. Yu, H. Yang, F. Xue, Z. Wu, W. Shen, J. Li, Z. Zhou, Y. Ding, Q. Zhao, X. C. Zhang, M. Liao, M. Bartlam, Z. Rao, Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design, J. Virol. 82 (2008) 2515–2527.

[41] D.A. Case, H.M. Aktulga, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T. E. Cheatham III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, C. Jin, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K. A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, Y. Xue, D.M. York, S. Zhao, P.A. Kollman, Amber 2021, University of California, San Francisco, 2021.

[42] D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The Amber biomolecular simulation programs, J. Comput. Chem. 26 (2005) 1668–1688.

[43] R. Salomon-Ferrer, D.A. Case, R.C. Walker, An overview of the Amber biomolecular simulation package: amber biomolecular simulation package, WIREs Comput. Mol. Sci. 3 (2013) 198–210.

[44] A. Onufriev, D. Bashford, D.A. Case, Exploring protein native states and large-scale conformational changes with a modified generalized born model, Proteins 55 (2004) 383–394.

[45] S. Shoji, N. Yoshida, M. Amanai, M. Ohgishi, T. Fukui, S. Fujimoto, Y. Nakano, E. Kajikawa, A.C.F. Perry, Mammalian Emi2 mediates cytostatic arrest and transduces the signal for meiotic exit via Cdc20, EMBO J. 25 (2006) 834–845.

[46] A. Arvanitidis, M.A. Karsdal, Type XXVIII Collagen, in: Biochemistry of Collagens, Laminins and Elastin, Elsevier, 2016, pp. 159–161.

[47] T. Koudelka, J. Boger, A. Henkel, R. Schönherr, S. Krantz, S. Fuchs, E. Rodríguez, L. Redecke, A. Tholey, N-terminomics for the identification of in vitro substrates and cleavage site specificity of the SARS-CoV-2 main protease, Proteomics 21 (2021), 2000246.

[48] C.-J. Kuo, P.-H. Liang, Characterization and inhibition of the main protease of severe acute respiratory syndrome coronavirus, ChemBioEng Rev. 2 (2015) 118–132.

[49] F. Maiello, G. Gallo, C. Coelho, F. Sucharski, L. Hardy, M. Würtele, Crystal structure of Thermus thermophilus methylenetetrahydrofolate dehydrogenase and determinants of thermostability, PLoS ONE 15 (2020), e0232959.

[50] B. Borges, G. Gallo, C. Coelho, N. Negri, F. Maiello, L. Hardy, M. Würtele, Dynamic cross correlation analysis of Thermus thermophilus alkaline phosphatase and determinants of thermostability, Biochim. Et. Biophys. Acta (BBA) - Gen. Subj. 1865 (2021), 129895.

[51] W. Rut, K. Groborz, L. Zhang, X. Sun, M. Zmudzinski, B. Pawlik, X. Wang, D. Jochmans, J. Neyts, W. Młynarski, R. Hilgenfeld, M. Drag, SARS-CoV-2 Mpro inhibitors and activity-based probes for patient-sample imaging, Nat. Chem. Biol. 17 (2021) 222–228.

[52] C.J. Farady, C.S. Craik, Mechanisms of macromolecular protease inhibitors, Chem. Eur. J. Chem. Bio. 11 (2010) 2341–2346.