RSC Advances

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2023, **13**, 30052

Computational medicinal chemistry applications to target Asian-prevalent strain of hepatitis C virus†

Rashid Hussain, *[a] Zulkarnain Haider,[a] Hira Khalid,*[a] M. Qaiser Fatmi,[b] Simone Carradori, [c] Amelia Cataldi[c] and Susi Zara[c]

Received 11th July 2023
Accepted 4th October 2023

DOI: 10.1039/d3ra04622b

rsc.li/rsc-advances

Hepatitis C Virus (HCV), affecting millions of people worldwide, is the leading cause of liver disorder, cirrhosis, and hepatocellular carcinoma. HCV is genetically diverse having eight genotypes and several subtypes predominant in different regions of the globe. The HCV NS3/4A protease is a primary therapeutic target for HCV with various FDA-approved antivirals and several clinical developments. However, available protease inhibitors (PIs) have lower potency against HCV genotype 3 (GT3), prevalent in South Asia. In this study, the incumbent computational tools were utilized to understand and explore interactions of the HCV GT3 receptor with the potential inhibitors after the virtual screening of one million compounds retrieved from the ZINC database. The molecular dynamics, pharmacological studies, and experimental studies uncovered the potential PIs as ZINC000224449889, ZINC000224374291, and ZINC000224374456 and the derivative of ZINC000224374456 from the ZINC library. The study revealed that these top-hit compounds exhibited good binding and better pharmacokinetics properties that might be considered the most promising compound against HCV GT3 protease. Viability test, on primary healthy Human Gingival Fibroblasts (HGFs) and cancerous AGS cell line, was also carried out to assess their safety profile after administration. In addition, Surface Plasmon Resonance (SPR) was also performed for the determination of affinity and kinetics of synthesized compounds with target proteins.

Introduction

Hepatitis C Virus (HCV) is a member of the Flaviviridae family that was first discovered in 1989.[1] According to the World Health Organization (WHO), HCV is responsible for more than 185 million infections worldwide, making it a significant global public health issue.[2] The virus is most commonly transmitted through contact with infected blood, such as sharing of needles among injection drug users, and unsafe medical procedures, including blood transfusions and organ transplants prior to the implementation of screening procedures. It can also be transmitted through unprotected sexual contact, perinatally from mother to child during childbirth, and in rare cases through occupational exposure to infected blood. HCV infections can lead to chronic hepatitis, cirrhosis, liver failure, and liver cancer, highlighting the importance of effective prevention and treatment strategies.[3,4]

The HCV genome exhibits high genetic diversity, with eight major genotypes and 87 subtypes identified to date.[5,6] This genetic diversity is due to the high mutation rate in the HCV genome, *i.e.*, 10^{-3} substitutions per site per year.[7] It contributes to the difficulty in developing effective vaccines and antiviral therapies against HCV. However, understanding the structure and function of the HCV genome provides insights into viral replication and pathogenesis, and can inform the development of new therapeutic strategies to combat HCV infections. The genome of the Hepatitis C Virus (HCV) is a single-stranded, positive-sense RNA molecule that contains approximately 9.6 kilobases in length. The RNA genome consists of an open reading frame (ORF), 5′ untranslated region (UTR), and 3′ UTR.

HCV encodes a single polyprotein that is processed into at least 10 individual proteins, including six non-structural (NS) proteins and three structural proteins. The NS proteins consist of ion channel (p7), auto-protease (NS2), protease and helicase (NS3), co-factor (NS4A), membrane-associated protein (NS4B), phosphor-protein (NS5A), and RNA-dependent RNA polymerase (NS5B). These proteins are responsible for various functions in viral replication, assembly, and immune evasion. NS3 has both protease and helicase activity and is essential for the replication of the HCV genome. NS5B is an RNA-dependent RNA polymerase, which is essential for viral replication, and is also the target of several antiviral drugs. NS5A is a multi-functional protein that plays a critical role in viral replication, assembly,

[a]*Department of Chemistry, Forman Christian College University, Lahore-54000, Pakistan. E-mail: hirakhalid@fccollege.edu.pk; rashid.bioinfo@gmail.com*

[b]*Department of Biosciences, COMSATS University Islamabad, Park Road, Chak Shahzad, Islamabad 45600, Pakistan*

[c]*Department of Pharmacy, "G. d'Annunzio" University of Chieti-Pescara, via dei Vestini 31, 66100 Chieti, Italy*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3ra04622b

and modulation of host immune responses. The three structural proteins include capsid (C) and envelope proteins (E1 and E2). The capsid protein forms the nucleocapsid core of the viral particle, while the envelope proteins are responsible for viral entry into host cells and are the main targets of neutralizing antibodies. E2 is also involved in viral attachment to host cells, while E1 is required for viral fusion with host cell membrane.[8–10] To cure HCV, it's important to identify the most suitable drug target. There are various tools available for this purpose, including the prediction of chokepoints for drug-target identification. This process involves identifying metabolic reactions that either consume a unique substrate or produce a unique product, which can then be used as potential drug targets.[11–13] In general, both the non-structural and structural proteins of HCV are crucial to the viral life cycle and represent critical targets for antiviral therapies. Inhibition of NS3/4A protease and NS5B polymerase has been shown to be effective in treating HCV infection, and ongoing research is focused on developing new drugs that target other HCV proteins, including NS5A and the envelope proteins.[14–16]

Currently, available drugs for HCV treatment are not equally effective against all genotypes. Most of these drugs are designed to target genotype 1, while little attention has been given to developing drugs specific to genotype 3a. This is partly due to the lack of a crystal structure of NS3 GT3, which makes it difficult to design genotype-specific drugs. Recent studies have identified specific mutations at key residues that are responsible for the lower response of genotype 3a to existing drugs.[7] On the other hand, *in silico* calculations are among essential therapeutic strategies, particularly where the experimental structure of the target proteins has not been revealed yet. In addition, molecular modeling methods help us understand drug-target interactions and discover novel drug candidates.

Our recent studies have used several *in silico* tools to target HCV NS3 protease.[17–19] Similar studies were carried out for quest of finding suitable drugs against different diseases.[20–25] Finally, to determine the safety profile of the best-in-class compounds, we assayed them at two fixed concentrations (10 and 50 μM) and discrete time points (48 and 72 h) on healthy primary Human Gingival Fibroblasts (HGFs) and a cancerous cell line (AGS, gastric adenocarcinoma) by means of the MTT test. The former cell type has been selected as they represent the first cell population to be in contact with the compounds after oral administration.[26]

Material and methods

The general workflow of the research project is given in the flowchart (Fig. 1). Initially, the computational studies were conducted to find potential inhibitors against the drug target HCV NS3 GT3. The computational results were then validated using experimental methods. The methods and techniques used in computational and experimental studies are given below.

Computational studies

Homology modelling. The homology modeling approach was used to predict the 3D model of the protease domain of NS3 protease genotype 3a using the crystal structure of HCV NS3 protease genotype 1b as a template (PDB ID: 4I31), having 100% sequence coverage and 78% sequence identity. The primary sequence of HCV NS3 genotype 3a was retrieved from NCBI (GenBank accession: AEV46286). SWISS-MODEL, an automated protein structure homology modeling server, was used to model the protein. The target and the template sequence were aligned using the ClustalW alignment tool.[27] Finally, the modeled structure of NS3 GT3a was further evaluated for compatibility of various structural parameters using comparative assessment tools like Ramachandran Plot.[28]

Compounds database. The clinically validated compounds reported against HCV NS3 protease include paritaprevir, glecaprevir, grazoprevir, telaprevir, voxilaprevir simeprevir, and boceprevir, which were taken from Drugbank (**https://go.drugbank.com/**) and used as a control for the in-house compounds. The ZINC database (ZINC is not Commercial) was used for retrieving one million compounds from ZINC15 (**https://zinc15.docking.org/**).[29]

Virtual screening. The virtual screening of the retrieved compounds was performed by UCSF DOCK 6 [30] in the following steps:

(1) Receptor and ligand structure preparation: The model protein of HCV NS3 GT3 was opened using UCSF Chimera. The Dock prep module of Chimera was used for receptor preparation.

(2) Sphere generation and selection: The binding groove sphere of 3 Å was generated by using the knowledge of docked ligand of the template protein PDB ID: 4I31.

(3) Grid generation: The grid around the receptor's active site was generated by keeping the distance between grid points along each axis.

(4) Docking: Rigid ligand docking was performed in which the ligand was kept completely rigid during the orientation step.

Molecular dynamics simulations. The refinement of the modeled NS3 GT3 was obtained through MD simulations using GROMACS 5.1.1.[31] The modeled protein was checked for missing residues/atoms and then initialized by generating topologies using the OPLS-AA/L all-atom force field.[32,33] The system was solvated in an explicit water cubic box using a 3-site Simple Point Charge (SPC) model. Periodic boundary conditions (PBC) were implemented to circumvent boundary effects caused by the finite size of the system. The minimum distance between the protein system and the box edge was set to at least 1.0 nm to avoid any artifact incurred by the minimum image convention. The whole system was neutralized by adding 6 Cl^- ions to the environment. The plan was then energy minimized using 50 000 steps of the steepest descent minimization algorithm to avoid any bad contacts generated while solvating the system. To stabilize the environment equilibration of the system was conducted in two phases. The first phase was conducted under an NVT ensemble (constant number of particles, volume, and temperature) by keeping the temperature at 300 K and pressure coupling off. In the second phase, NPT ensemble (constant number of particles, pressure, and temperature) was used to keep pressure coupling at 1 bar. The Leapfrog integrator

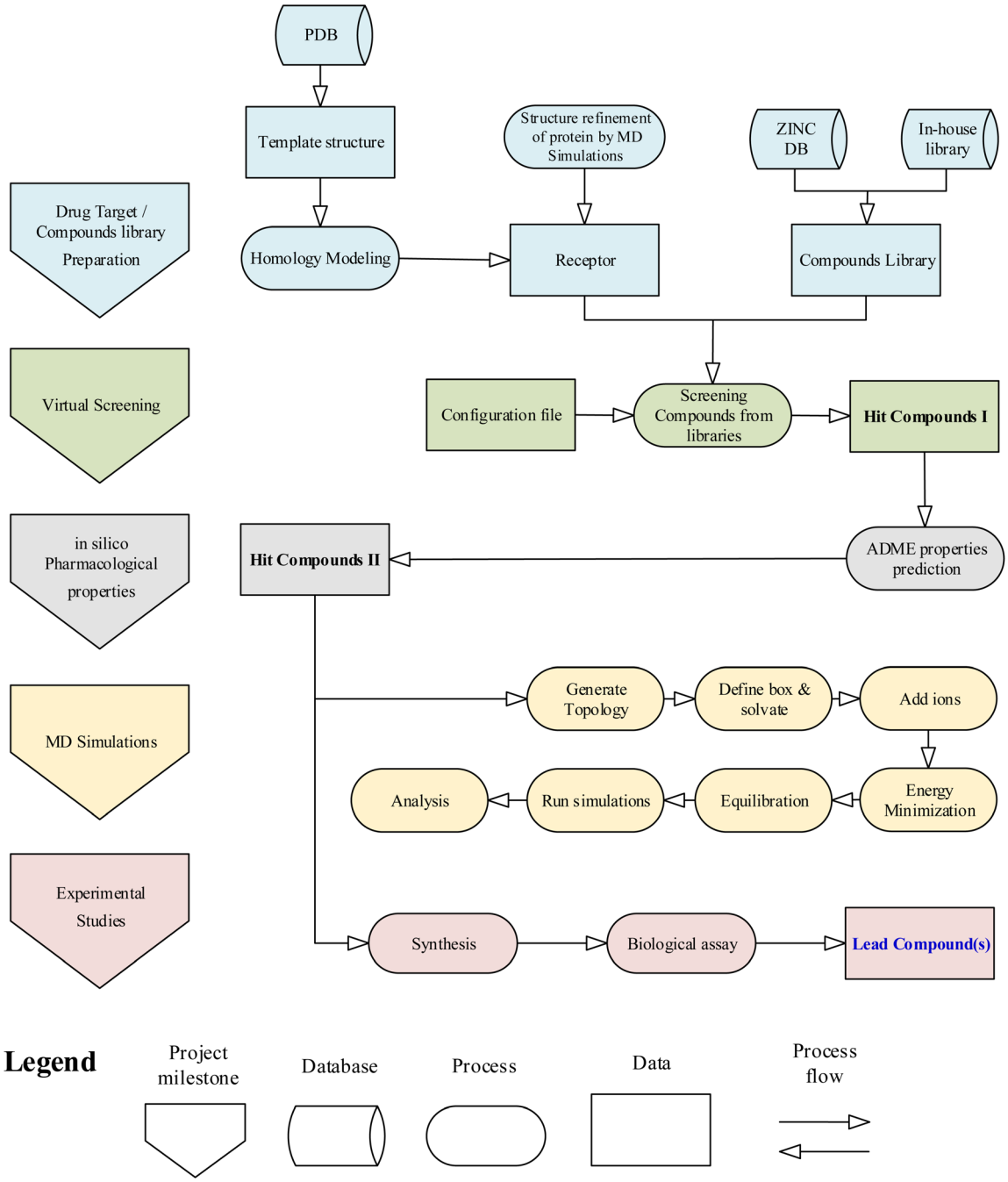


Fig. 1 Flowchart of research methodology.

was used to integrate the Newtonian equation of motion with 25,000,000 steps. SHAKE algorithm was used to fix all bond distances involving hydrogen atoms; therefore, the time step was increased to 2 femtoseconds (fs), making the total duration of the simulation 50 ns. The Lenard–Jones equation was used to calculate van der Waal's interactions. The short-range neighbour list cut-off, short-range electrostatic cut-off, and short-range van der Waal's cut-off were fixed at 1 nm. The conformations of the homology model generated during the 50 ns MD simulation were compared with those obtained for simulations of the template crystal structure, 4I31.pdb, performed using the same parameters and conditions.

To assess the results of virtual screening by USCF DOCK6, MD simulations of top-hit compounds were carried out at 100 ns. In addition, the hit compounds complexed with the modeled protein were undergone for protein–ligand complex simulations. The topologies of the receptor and each top-hit compound were prepared separately and then joined into a single GROMACS file. The system is finally prepared and run after solvation, ionization, energy minimization, and equilibration.

Hardware & software. The homology model of NS3 protease was used from our previous studies.[18,19] The docking studies were carried out through the Linux operating system (Ubuntu 18.04.5 LTS, x86_64) with remotely accessed virtual machines

with a range of processing power. The list of software and hardware used in the study is listed in Appendix-A, Tables S1 and S2.†

Experimental studies

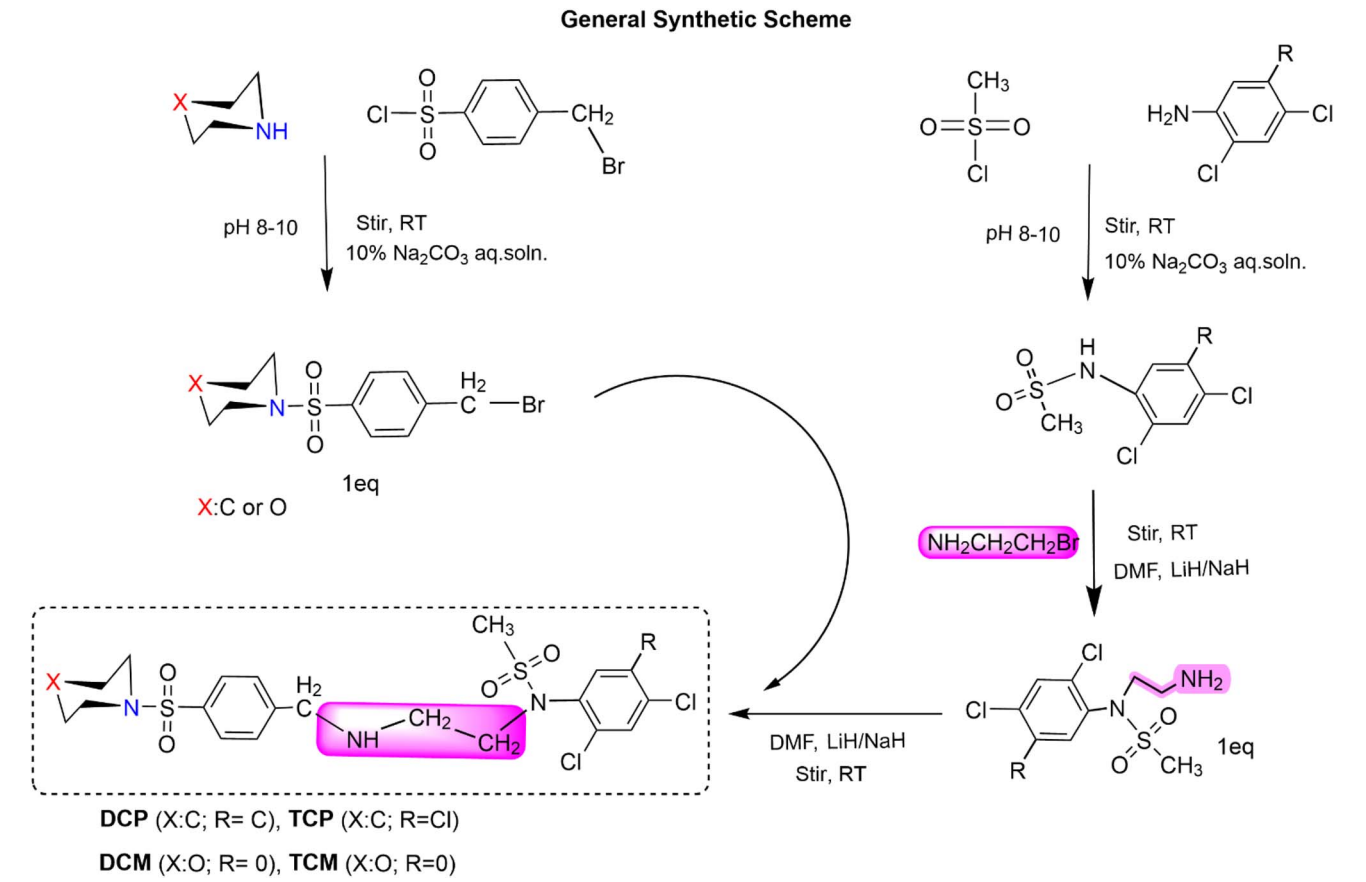
Organic synthesis of the hit compounds. The chemicals and solvents were purchased from Sigma Aldrich and Alfa Aesar and used for experimental work without further purification. Silica GEL G TLC plates were used to monitor all reactions, and the spots were detected under UV lamps of long and short wavelengths (model UVGL-minor light multiband UV-254/366). In addition, the purity of synthesized compounds was also checked by using silica gel G TLC plates.

Synthesis schemes of the hit compounds. The synthesis of the hit compounds was performed after the optimization of top compounds obtained after the results of virtual screening. The synthetic scheme of each hit compound is shown in the figures: the general synthetic scheme of the hit compounds is given in Scheme 1. The synthesis of fragments of hit compounds are given in the appendix (Appendix-B, Fig. S1 and Table S1†).

Synthesis of N-(2-(4-(piperidin-1-ylsulfonyl)benzylamino)ethyl)-N-(2,4,5-trichlorophenyl)methanesulfonamide (TCP). 1-(4-(Bromomethyl)phenylsulfonyl)piperidine (0.1749 g) was taken in a round-bottomed flask (150 mL) and dissolved into 5% DMF (15 mL). *N*-(2-Aminoethyl)-*N*-(2,4,5-tri-chlorophenyl) methanesulfonamide (0.15 g) was added to it and stirred at room temperature for 8 hours. Lithium hydride (0.002 g) was also added as a catalyst. TLC (hexanes, acetate; 80 : 20) showed a single spot. The reaction mixture was quenched with chilled water, and the product precipitated, filtered, and dried.

Synthesis of N-(2-(4-(morpholinosulfonyl)benzyl amino)ethyl)-N-(2,4,5-trichlorophenyl)methanesulfonamide (TCM). 4-(4-(Bromomethyl)phenylsulfonyl)morpholine (0.17375 g) was taken in the round-bottomed flask (150 mL) and dissolved into 5% DMF (15 mL). *N*-(2-Aminoethyl)-*N*-(2,4,5-trichlorophenyl)methanesulfonamide (0.15 g) was added to it and stirred at room temperature for 8 hours 15 minutes. Lithium hydride (0.002 g) was also added as a catalyst. TLC (hexanes, acetate; 80 : 20) showed a single spot. The reaction mixture was quenched with chilled water, and the product precipitated, filtered, and dried.

Synthesis of N-(2,4-dichlorophenyl)-N-(2-(4-(piperidin-1-ylsulfonyl)benzylamino)ethyl)methanesulfonamide (DCP). *N*-(2-Aminoethyl)-*N*-(2,4-dichlorophenyl)methanesulfonamide (0.34 g) was taken in a round-bottomed flask (150 mL) and dissolved into 5% DMF (15 mL). 1-(4-(Bromomethyl)phenylsulfonyl) piperidine (0.38 g) was added to it and stirred at room temperature for 10 hours. Lithium hydride (0.002 g) was also added as a catalyst. TLC (hexanes, acetate; 80 : 20) showed a single spot. The reaction mixture was quenched with chilled water, and the product precipitated, filtered, and dried.



Scheme 1 General synthetic scheme of the hit compounds.

Synthesis of N-(2,4-chlorophenyl)-N-(2-(4-(morpholinosulfonyl) benzyla-mine)ethyl)methanesulfonamide (DCM). *N*-(2-Aminoethyl)-*N*-(2,4-chlorophenyl)methanesulfonamide (0.35 g) was taken in the round-bottomed flask (150 mL) and dissolved into 5% DMF (15 mL). 4-(4-(Bromomethyl)phenylsulfonyl)morpholine (0.4 g) was added to it and stirred at room temperature for 6 hours. Lithium hydride (0.002 g) was also added as a catalyst. TLC (hexanes, acetate; 80 : 20) showed a single spot. The reaction mixture was quenched with chilled water, and the product precipitated, filtered, and dried.

Biological evaluations

Cell culture

HGF and AGS culture. A total of 10 healthy donors, who undergone to third molars extraction, signed the informed consent according to the Italian Law and to the Ethical Principles for Medical Research code including Human Subjects of the World Medical Association (Declaration of Helsinki). The project was approved by the Local Ethical Committee of the University of Chieti (Chieti, Italy, approval number. 1173, approved on 31/03/2016). Gingiva biopsies were rinsed in phosphate-buffered saline (PBS), placed in Dulbecco's modified Eagle's medium (DMEM), cut into smaller pieces, and cultured in DMEM, with 10% foetal bovine serum (FBS), 1% penicillin/streptomycin and 1% fungizone (all purchased from Merck Life Science, Milan, Italy). After 10 days of culture, fungizone was removed from the medium, and cells cultured until 5–8 passages. AGS human gastric adenocarcinoma cell line (ECACC 89090402, Merck Life Science, Milan, Italy) was cultured in Ham's F12 medium with 10% of FBS, 1% of penicillin/streptomycin, and 1% of L-glutamine (all purchased from Merck Life Science). Both cell cultures were maintained at 37 °C within an incubator in the presence of 5% (v/v) CO_2.

HGF and AGS treatment. For each compound, a stock solution of 0.1 M was prepared using DMSO as a vehicle. Then, the stock solution was diluted in DMEM or Ham's F12 medium (for HGFs and AGS, respectively) to obtain intermediate solutions of 100 μM and final solutions of 50 and 10 μM for HGFs and 50 μM for AGS. To exclude DMSO cytotoxicity, the final concentration of DMSO within the culture medium was kept at 0.05%. The HGFs and AGS cells were seeded at 6700 and 8000 cells per well, in a 96 multiwell plate, respectively. After 24 h from seeding, the medium (DMEM and Ham's for HGFs and AGS, respectively) was replaced by a fresh one containing compounds at 10 and 50 μM for HGFs. In AGS culture newly synthesized compounds were administered at 50 μM. Treatments were maintained from 48 to 72 h within an incubator in a humified atmosphere in the presence of 5% (v/v) CO_2 at 37 °C.

MTT metabolic activity test. After 48 and 72 h of culture an MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay was carried out. The MTT test measures the viable cells' capability to transform MTT into a violet formazan salt. At the established experimental time points, the culture medium was added of MTT 10% (Merck Life Science, Milan, Italy) and incubated at 37 °C for 5 h for HGFs and for 4 h for AGS. To dissolve formazan salts plate was probed in DMSO for 30 min at 37 °C, then read at 540 nm wavelength through a microplate reader (Multiskan GO, Thermo Scientific, Waltham, MA, USA). The obtained values were normalized with values derived from cells treated with DMSO (vehicle).

Statistics. Statistical analysis was performed using the GraphPad 7 software (GraphPad Software, San Diego, CA, USA) by means of Ordinary One-Way ANOVA followed by posthoc Tukey's multiple comparison tests.

SPR assay. The purified DNA sequence encoding the Hepatitis C Virus (HCV) (serotype 1a, isolate H77) NS3 (NP_803144.1) (Thr1356-Thr1459) was expressed with a GST tag at the N-terminus was purchased from Sino Biological enzyme was immobilized on flow channels 2 and 4 of a CM5 sensor chip using modified GST-coupling with running buffer HBS-EP (10 mM HEPES, 150 mM NaCl, 0.05% surfactant P-20, pH 7.4) using a Biacore S200 instrument. Flow channels 1 and 3 were used as control surfaces. The HCV-NS3 enzyme was diluted in 10 mM sodium acetate (pH 5.0) and immobilized after sensor surface activation with 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC)/*N*-hydroxy succinimide (NHS) mixture followed by ethanolamine (pH 8.5) blocking on unoccupied surface area.

The selected compound was initially prepared as 10 mM DMSO stock solutions, and compound solutions with a series of increasing concentrations (2.4–1500 5-fold dilution) were applied to all four channels at a 30 μL min^{-1} flow rate at 25 °C. Sensorgrams were analysed using BIA evaluation software 3.0, and response unit difference (ΔRU) values at each concentration were measured during the equilibrium phase. All data were double referenced with both blank surface and zero compound concentration responses and fitted with the steady-state affinity equation below where y is the response, Y_{max} is the maximum response and x is the compound concentration.[34] Refer to Appendix-D† for pH scouting.

$$y = \frac{y_{max} \times x}{K_D + x}$$

Results and discussion

Computational studies

Virtual screening of one million ZINC library compounds. One million compounds retrieved from the ZINC database were screened against modelled NS3 protease GT 3a Using UCSF DOCK6. Due to computational resource constraints, the whole compound library of one million compounds was subjected to

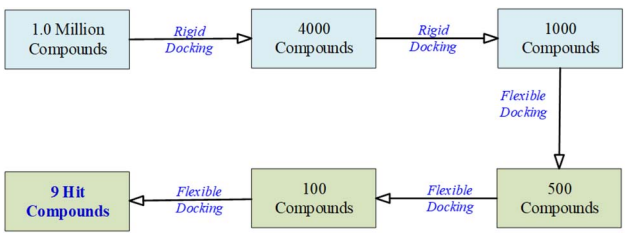


Fig. 2 Workflow of virtual screening.

Table 1 Virtual screening results and pharmacokinetics studies of top hit compounds. Pharmacokinetics studies of the top compounds. The pharmacokinetics parameters are defined beneath the table[a]

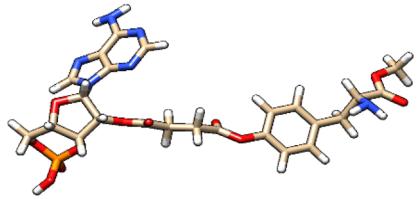
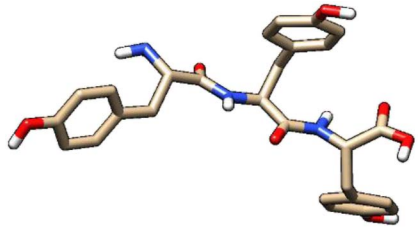
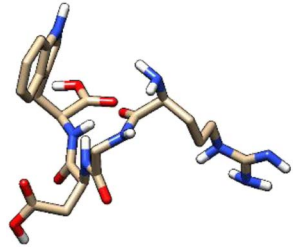
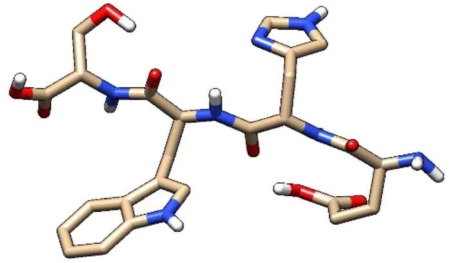
S. no.	Molecule ID	Structure	DOCK 6 results	Molecular properties
1	ZINC000100685029		Grid score: −62.41 DOCK rotatable bonds: 12 Molecular weight: 606.48 Formal charge: 0.010 Grid VDW energy: −48.31 Grid ES energy: −14.10 Internal energy repulsive: 17.97	milog *P*: −1.11 TPSA: 239.57 *n*atoms: 42 MW: 606.49 *n*ON: 17 *n*OHNH: 5 *n*rotb: 12 Volume: 488.37
2	ZINC000005273907		Grid score: −59.34 DOCK rotatable bonds: 16 Molecular weight: 507.54 Formal charge: 0.02 Grid VDW energy: −47.01 Grid ES energy: −12.32 Internal energy repulsive: 15.78	milog *P*: −0.87 TPSA: 182.20 *n*atoms: 37 MW: 507.54 *n*ON: 10 *n*OHNH: 8 *n*rotb: 11 Volume: 452.86
3	ZINC000003917816		Grid score: −58.56 DOCK rotatable bonds: 19 Molecular weight: 532.55 Formal charge: -6.70×10^{-7} Grid VDW energy: −46.64 Grid ES energy: −11.92 Internal energy repulsive: 17.74	milog *P*: −4.31 TPSA: 265.61 *n*atoms: 38 MW: 532.56 *n*ON: 15 *n*OHNH: 12 *n*rotb: 16 Volume: 469.75
4	ZINC000101149671		Grid score: −56.99 DOCK rotatable bonds: 19 Molecular weight: 557.56 Formal charge: 0.01 Grid VDW energy: −48.16 Grid ES energy: −8.83 Internal energy repulsive: 15.16	milog *P*: −4.48 TPSA: 252.62 *n*atoms: 40 MW: 557.56 *n*ON: 15 *n*OHNH: 10 *n*rotb: 15 Volume: 484.05

Table 1 (*Contd.*)

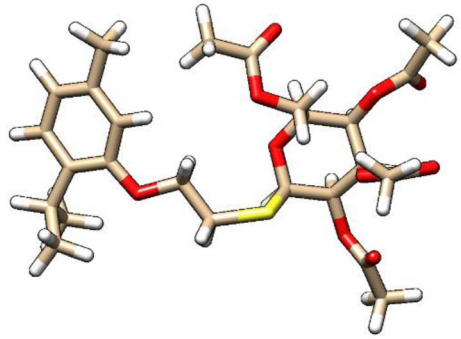
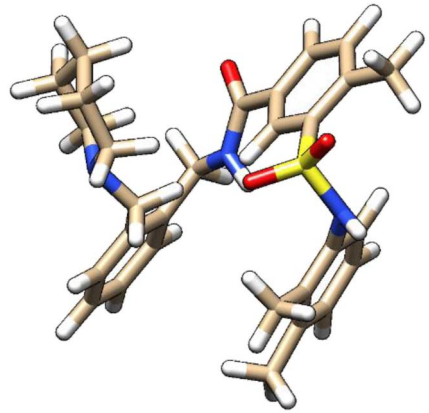
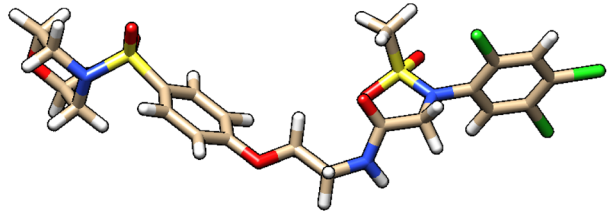
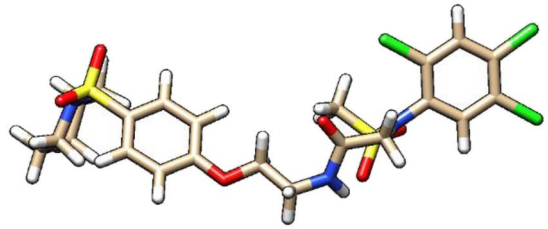
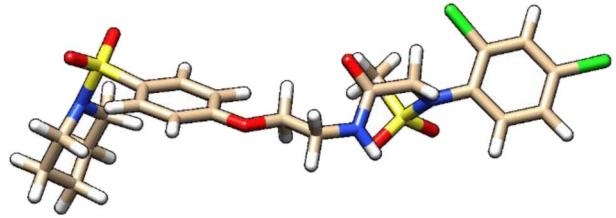
S. no.	Molecule ID	Structure	DOCK 6 results	Molecular properties
5	ZINC000101574832		Grid score: −56.18 DOCK rotatable bonds: 15 Molecular weight: 540.63 Formal charge: −0.02 Grid VDW energy: −51.30 Grid ES energy: −4.88 Internal energy repulsive: 33.20	milog *P*: 3.99 TPSA: 123.69 *n*atoms: 37 MW: 540.63 *n*ON: 10 *n*OHNH: 0 *n*rotb: 15 Volume: 488.47
6	ZINC000224822442		Grid score: −54.43 DOCK rotatable bonds: 9 Molecular weight: 505.68 Formal charge: −0.02 Grid VDW energy: −47.58 Grid ES energy: −6.85 Internal energy repulsive: 37.15	milog *P*: 5.14 TPSA: 78.50 *n*atoms: 36 MW: 505.68 *n*ON: 6 *n*OHNH: 2 *n*rotb: 8 Volume: 471.32
7	ZINC000224449889		Grid score: −54.12 DOCK rotatable bonds: 11 Molecular weight: 600.92 Formal charge: 0.03 Grid VDW energy: −53.63 Grid ES energy: −0.48 Internal energy repulsive: 17.71	milog *P*: 3.75 TPSA: 122.33 *n*atoms: 36 MW: 600.93 *n*ON: 10 *n*OHNH: 1 *n*rotb: 10

Table 1 (*Contd.*)

S. no.	Molecule ID	Structure	DOCK 6 results	Molecular properties
8	ZINC000224374291		Grid score: −53.86 DOCK rotatable bonds: 11 Molecular weight: 598.95 Formal charge: 0.009 Grid VDW energy: −53.50 Grid ES energy: −0.35 Internal energy repulsive: 16.58	Volume: 457.18 milog *P*: 4.82 TPSA: 113.09 *n*atoms: 36 MW: 598.96 *n*ON: 9 *n*OHNH: 1 *n*rotb: 10
9	ZINC000224374456		Grid score: −53.38 DOCK rotatable bonds: 11 Molecular weight: 564.51 Formal charge: -1.49×10^{-7} Grid VDW energy: −51.35 Grid ES energy: −2.03 Internal energy repulsive: 45.04	Volume: 464.99 milog *P*: 4.21 TPSA: 113.09 *n*atoms: 35 MW: 564.51 *n*ON: 9 *n*OHNH: 1 *n*rotb: 10
				Volume: 451.46

[a] milog *P*: log *P* (octanol/water partition coefficient), TPSA: molecular polar surface area, MW: molecular weight, nON: number of hydrogen bond acceptors, nOHNH: number of hydrogen bond donors, nrotb: number of rotatable bonds.

rigid docking on remote virtual servers. The scheme of the virtual screening is given below. Initially, the top 4000 compounds with better grid scores were selected and re-docked again to shortlist the top 1000 compounds. Onward, these compounds were kept on docking through flexible docking (anchor and grow algorithm) until the top nine compounds came out as hit compounds (Fig. 2). Finally, the top nine compounds were selected having grid scores greater than −53. The attributes of these top compounds, such as rotatable bonds, molecular weight, electrostatic interaction energy, and repulsive energy, were noted and compared (Table 1). The ZINC ID of the selected hit compounds are ZINC000100685029, ZINC000005273907, ZINC000003917816, ZINC000101149671, ZINC000101574832, ZINC000224822442, ZINC000224449889, ZINC000224374291, and ZINC000224374456.

In the virtual screening process, the transition from an initial pool of 1000 compounds to a refined set of 9 hit compounds during the flexible docking stages involved a series of steps designed to identify compounds with progressively stronger binding affinities.

To accomplish this, flexibility was introduced by adjusting the number of orientations or poses during flexible docking. This adaptability further refined the selection from 1000 to 9 hit compounds. As the process advanced from 1000 to 500 to 100 compounds, the number of orientations or poses was increased to explore a broader conformational space. This approach facilitated a comprehensive evaluation of each compound's interactions with the NS3 protease receptor, capturing a wide range of binding modes.

The initial set of 1000 compounds, selected from the secondary rigid docking, underwent flexible docking using UCSF DOCK6's anchor and grow algorithm. Consistent parameters were applied to all compounds, with a focus on grid scores and binding affinities as the selection criteria. From this stage, compounds with the most favorable grid scores and binding affinities were identified. Those with the highest scores were considered for further refinement, resulting in the retention of the top 500 compounds.

Subsequently, a more detailed analysis of binding energies and receptor interactions was conducted for the top 500 compounds, maintaining the same docking parameters. This process allowed a focus on compounds with even stronger binding potential, leading to the selection of the top 100 compounds.

The top 100 compounds underwent intensive evaluation, using the same consistent docking parameters. A comprehensive analysis of their binding modes, energies, and interactions with the NS3 protease receptor was performed, ultimately resulting in the identification of the top nine compounds with the highest binding affinities, considered as the hit compounds.

Pharmacokinetics studies of top hit compounds. The physicochemical parameters related to drug-likeness, adsorption, distribution, metabolism, and excretion (ADME) were calculated for the top hit compounds using Molinspiration to assess

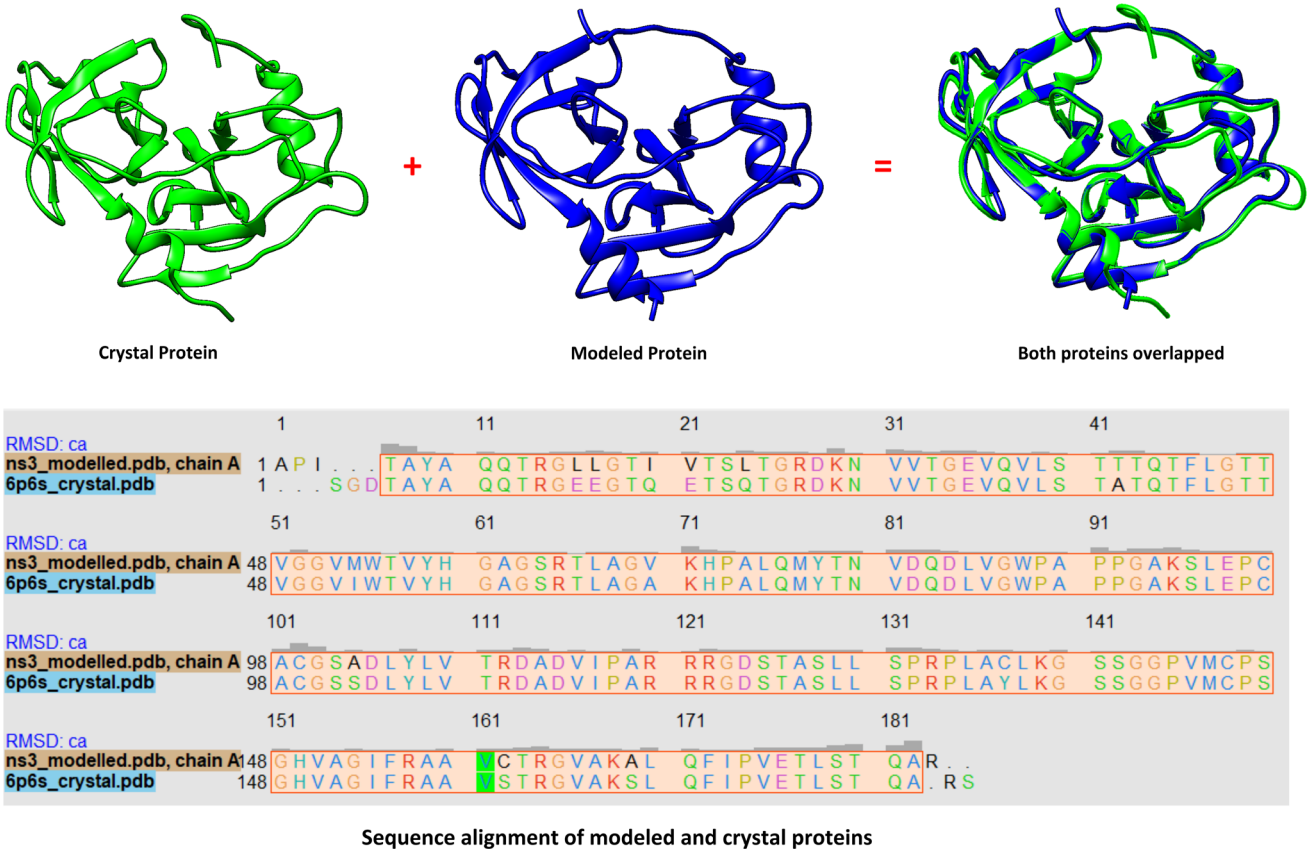


Fig. 3 Comparison between the modelled (blue) and the crystal structure (green) of NS3 protease.

Table 2 Docking score results of the modelled and crystal proteins

Molecule ID/name	Score of Modeled structure	Score of crystal structure
ZINC000100685029	−62.41	−59.62
ZINC000005273907	−59.34	−52.02
ZINC000003917816	−58.56	−63.77
ZINC000101149671	−56.99	−58.79
ZINC000101574832	−56.18	−50.33
ZINC000224822442	−54.43	−48.15
ZINC000224449889	−54.12	−48.30
ZINC000224374291	−53.86	−50.07
ZINC000224374456	−53.38	−50.13
Paritaprevir	−55.25	−55.62
Glecaprevir	−26.46	−44.46
Grazoprevir	−46.85	−64.67
Telaprevir	−50.90	−49.79
Voxilaprevir	−60.29	−42.60
Simeprevir	−59.09	−45.89
Boceprevir	−34.73	−46.79

their pharmacokinetics properties (Table 1). All the hit compounds conform to the Molinspiration parameter except for a few deviations. The compounds ZINC000100685029, ZINC000005273907, ZINC000003917816, ZINC000101149671, ZINC000101574832, and ZINC000224822442 have slightly more hydrophobicity which affects drug absorption, bioavailability, hydrophobic drug–receptor interactions, metabolism of molecules, as well as their toxicity (Table 1). The remaining three compounds, ZINC000224449889, ZINC000224374291, and ZINC000224374456, demonstrated favorable properties to conform with the Lipinski rule of 5. However, the molecular volume of the latter is a little bit high, which is also essential to occupy the wide binding site of NS3 protease. Thus, ADME studies showed that compounds ZINC000224449889, ZINC000224374291, and ZINC000224374456 exhibited good pharmacokinetic properties and therefore taken for further experimental studies to validate the theoretical study's findings.

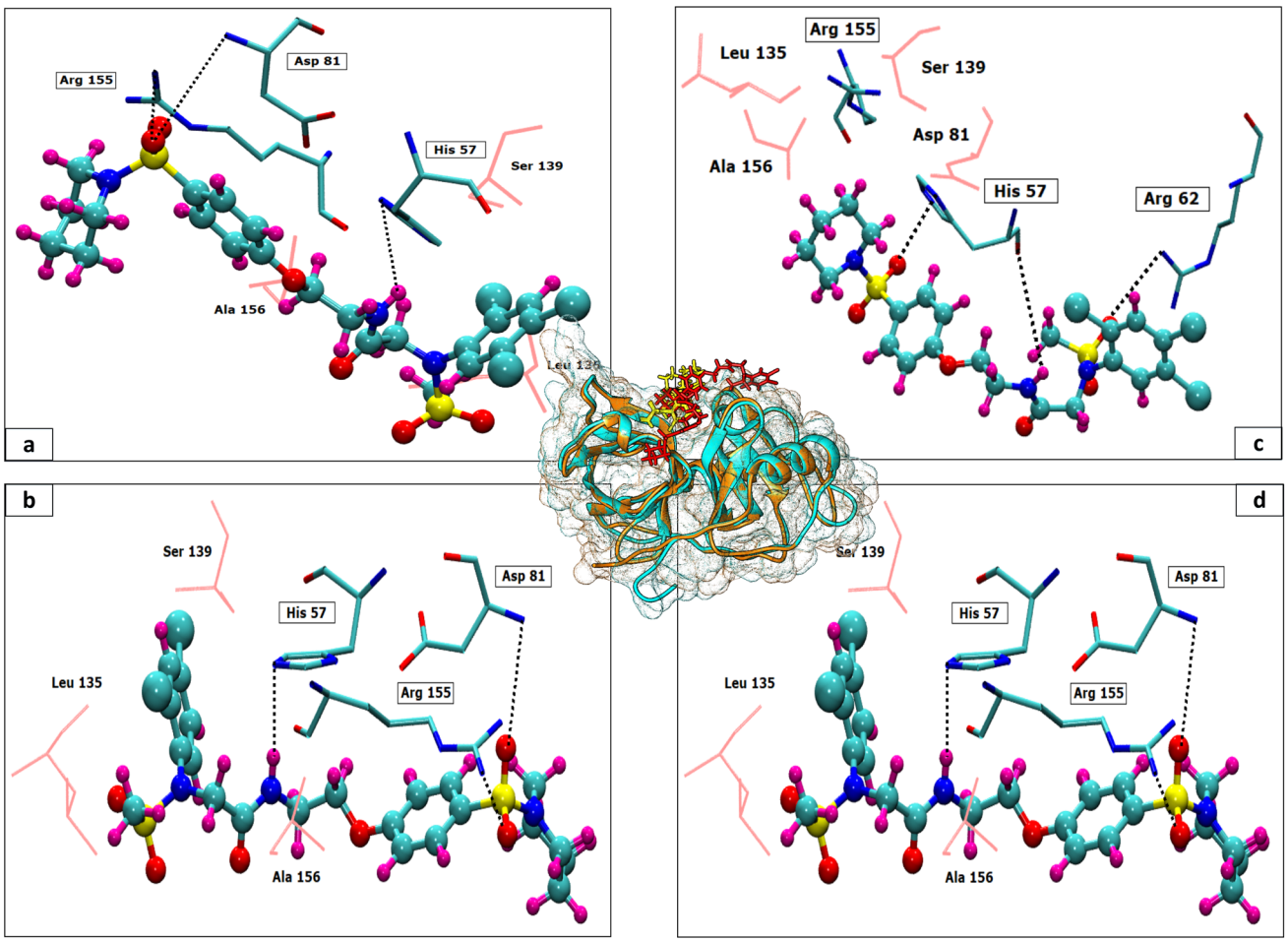


Fig. 4 The modelled protein (orange) and the crystal protein (cyan) are overlapped surrounded by meshed surface of both proteins in respective colors (center). The docked pose of the hit compound ZINC000224374291 and ZINC000224374456 for the modelled protein (yellow) and the crystal protein (red) are shown on the binding groove of the overlapped proteins. The interaction images of ZINC000224374291 (a) and ZINC000224374456 (b) with the modelled protein are given. Similarly, interaction images of both the compounds *i.e.*, ZINC000224374291 (c) and ZINC000224374456 (d), are shown in complex with the crystal protein.

Crystal structure of HCV NS3 protease

At the time of synopsis approval from the Board of Advance Studies and Research (BASR) dated 23rd April 2019, the crystal structure of HCV NS3 protease GT3a was not yet revealed. So, the homology model of HCV NS3 protease GT3a was constructed using SWISS-MODEL[35] for computational studies. However, its structure was resolved by Timm, J. *et al.* and released by Protein DataBank on 10th Jun 2020 with PDB ID: 6P6S.[36] However, its paper is not published to date. So, it became incumbent to compare the results with the crystal structure obtained against the modeled structure of NS3 protease. Hence, the structure and the docking results against the modeled protein were compared with the crystal structure.

Comparison of crystal structure with the modelled protein

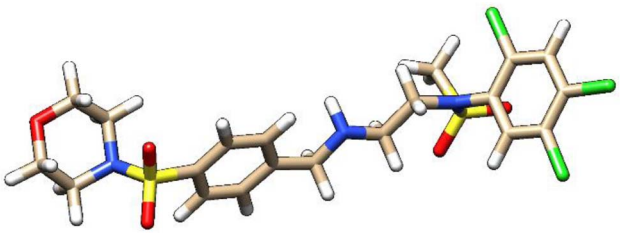
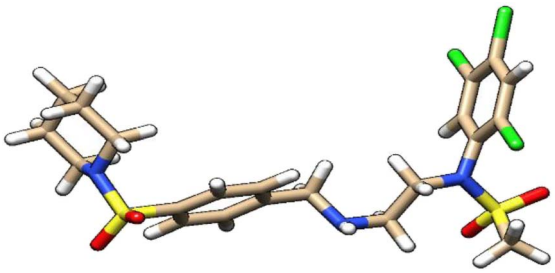
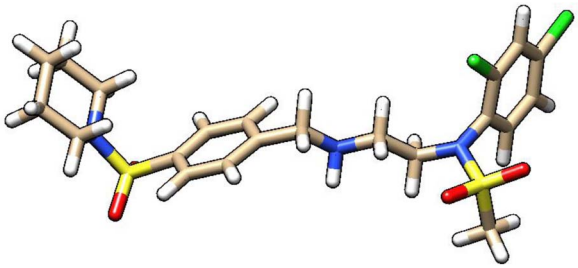
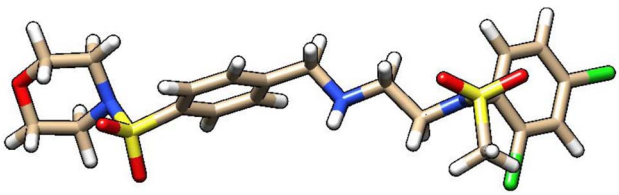
The 3D conformation of modeled NS3 protease and the recently reported crystal structure (PDB entry: 6P6S) exhibited similar coordinates, with an RMSD difference of 0.610 Å (Fig. 3). Furthermore, the sequence alignment of both proteins resulted in 100% sequence identity with an *E*-value of 2×10^{-145}. The top hit compounds were also docked against the crystal structure of NS3 protease. The seven clinically validated reported compounds were also taken into consideration to assess the difference between the results of both proteins and the variation in results between the hit and the reported compounds (Table 2). Interestingly, under the same parameters, most of the hit compounds and the reported compounds exhibited better grid scores in the case of the modeled protein than that of the

Table 3 Proteins ligand interactions of NS3 modelled structure and crystal structure

	Molecule ID	DOCK6 score		Hydrogen bonds	Electrostatic interactions	Vdw interactions
Modelled protein	ZINC000224374291	Grid score	−53.86	Ser139 with trichloro-benzene ring		
		Grid VDW energy	−53.50	His57 with amino group (3.87 Å)	Ala156 with carbonyl group	Leu135 with trichloro-benzene ring
		Grid ES energy	−0.35	Arg155 with sulfonamide (2.78 Å)		
		Internal energy repulsive	16.58			
	ZINC000224374456	Grid score	−53.38	His57 with amino group (2.82 Å)	Ser139 with trichloro-benzene ring	Leu135 with trichloro-benzene ring
		Grid VDW Energy	−51.35	Asp81 with sulfonamide (3.76 Å)	Ala156 with carbonyl group	
		Grid ES energy	−2.03	Arg155 with sulfonamide (2.59 Å)		
		Internal energy repulsive	45.04			
Crystal protein	ZINC000224374291	Grid score	−50.07			
		Grid VDW energy	−45.42		Ser139 with sulfonamide adjacent to piperidine ring	
		Grid ES energy	−4.65	Arg62 with sulfonamide (2.95 Å)	Arg155 with sulfonamide adjacent to piperidine ring	Leu135 with piperidine ring Ala156 with piperidine ring
		Internal energy repulsive	20.09			
	ZINC000224374456	Grid score	−50.13		Ser139 with sulfonamide adjacent to piperidine ring	
		Grid VDW energy	−44.66	His57 with sulfonamide (3.02 Å)		Leu135 with piperidine ring Ala156 with piperidine ring
		Grid ES energy	−5.46	Ala133 with sulfonamide (2.95 Å)	Arg155 with sulfonamide adjacent to piperidine ring	
		Internal energy repulsive	23.90			

crystal structure. Except for a single compound, ZINC000101149671, the rest of the hit compounds exhibited a better grid score against the modeled protein than the crystal structure. Similarly, the clinically reported compounds, telaprevir, voxilaprevir, and simeprevir, demonstrated a better grid score, *i.e.*, −50.90, −60.29, and −59.09 respectively, against the modeled protein compared to −49.79, −42.60, and −45.89 in the same order against the crystal structure. The compound Paritaprevir has almost the same grid score against both proteins (Table 2). Thus, the compound with a better grid score will have better binding with the target protein and better inhibitory potential. Furthermore, when the hit compounds and the clinically validated compounds (control) are compared, most of the hit compounds revealed better grid scores than the control compounds. In fact, the reported compounds, glecaprevir (−26.46), grazoprevir (−46.85), and boceprevir (−34.73), showed even lower grid scores than the hit compound with the lowest grid score *i.e.*, −53.38 (Table 2). For a more detailed comparison, the top two hit compounds ZINC000224374291 and ZINC000224374456, are shown as docked poses into the active site of the modeled and template protein overlapping each other (Fig. 4). The docking score and intermolecular interaction of both compounds are given in detail in Table 3 and shown in Fig. 4. The catalytic triad residues, His57, Asp81, and Ser139, are predominant in the active site of both proteins along with binding groove residues, especially Arg155 and Ala156. The compound ZINC000224374291 makes three hydrogen bonds and three electrostatic interactions with the modeled protein as well as the crystal structure protein. The residues that are involved in hydrogen bonding are His57 (3.87 Å), Asp81 (5.00 Å), and Arg155 (2.87 Å) with the modeled protein. Similarly, the same residues form hydrogen bonds with the crystal protein with the bond length 5.00 Å, 5.21 Å, and 2.95 respectively. In the case of the compound, ZINC000224374456, there are three hydrogen bonds with the modeled protein and two bonds with the crystal structure protein. His57 is the

Table 4 Organic synthesis of the hit compounds

ZINC ID	Modified compound name	Compound code	Compound structure
ZINC000224449889	*N*-(2-(4-(Morpholinosulfonyl) benzylamino)ethyl)-*N*-(2,4,5-trichlorophenyl) methanesulfonamide	TCM	
ZINC000224374291	*N*-(2-(4-(Piperidin-1-ylsulfonyl)benzylamino) ethyl)-*N*-(2,4,5-trichlorophenyl) methanesulfonamide	TCP	
ZINC000224374456	*N*-(2,4-Dichlorophenyl)-*N*-(2-(4-(piperidin-1-ylsulfonyl)benzylamino) ethyl)methanesulfonamide	DCP	
Derivative of DCP	*N*-(2,4-Dichlorophenyl)-*N*-(2-(4-(morpholinosulfonyl) benzylamino)ethyl) methanesulfonamide	DCM	

dominant residue involved in hydrogen bonds. Ser139 and Ala156 of modeled protein are predominant in electrostatic interactions whereas Ser139 and Arg155 are major residues involved in electrostatic interactions in the case of crystal protein. The residue Leu135 is mainly responsible for van der Waals interactions for both the proteins (Table 3 and Fig. 4).

Experimental studies

Synthesis and optimization of the hit compounds. Given the favorable results from virtual screening and *in silico* pharmacokinetic studies, the compounds ZINC000224449889, ZINC000224374291, and ZINC000224374456 have grid scores grid score −54.12, −53.86, and −53.38 were selected as the hit

Table 5 Docking score of the synthesized organic compounds

Compound code	DOCK6 score of modelled NS3 GT3	DOCK6 score of NS3 crystal structure
TCM	Grid score: −42.36 Grid VDW energy: −41.97 Grid ES energy: −0.39 Internal energy repulsive: 14.18	Grid score: −45.45 Grid VDW energy: −45.06 Grid ES energy: −0.39 Internal energy repulsive: 10.30
TCP	Grid score: −43.11 Grid VDW energy: −42.14 Grid ES energy: −0.97 Internal energy repulsive: 10.46	Grid score: −43.89 Grid VDW energy: −42.94 Grid ES energy: −0.94 Internal energy repulsive: 13.30
DCP	Grid score: −46.88 Grid VDW energy: −46.43 Grid ES energy: −0.45 Internal energy repulsive: 13.60	Grid score: −43.11 Grid VDW energy: −42.43 Grid ES energy: −0.68 Internal energy repulsive: 18.54
DCM	Grid score: −45.77 Grid VDW energy: −45.62 Grid ES energy: −0.14 Internal energy repulsive: 14.86	Grid score: −42.99 Grid VDW energy: −42.35 Grid ES energy: −0.63 Internal energy repulsive: 13.14

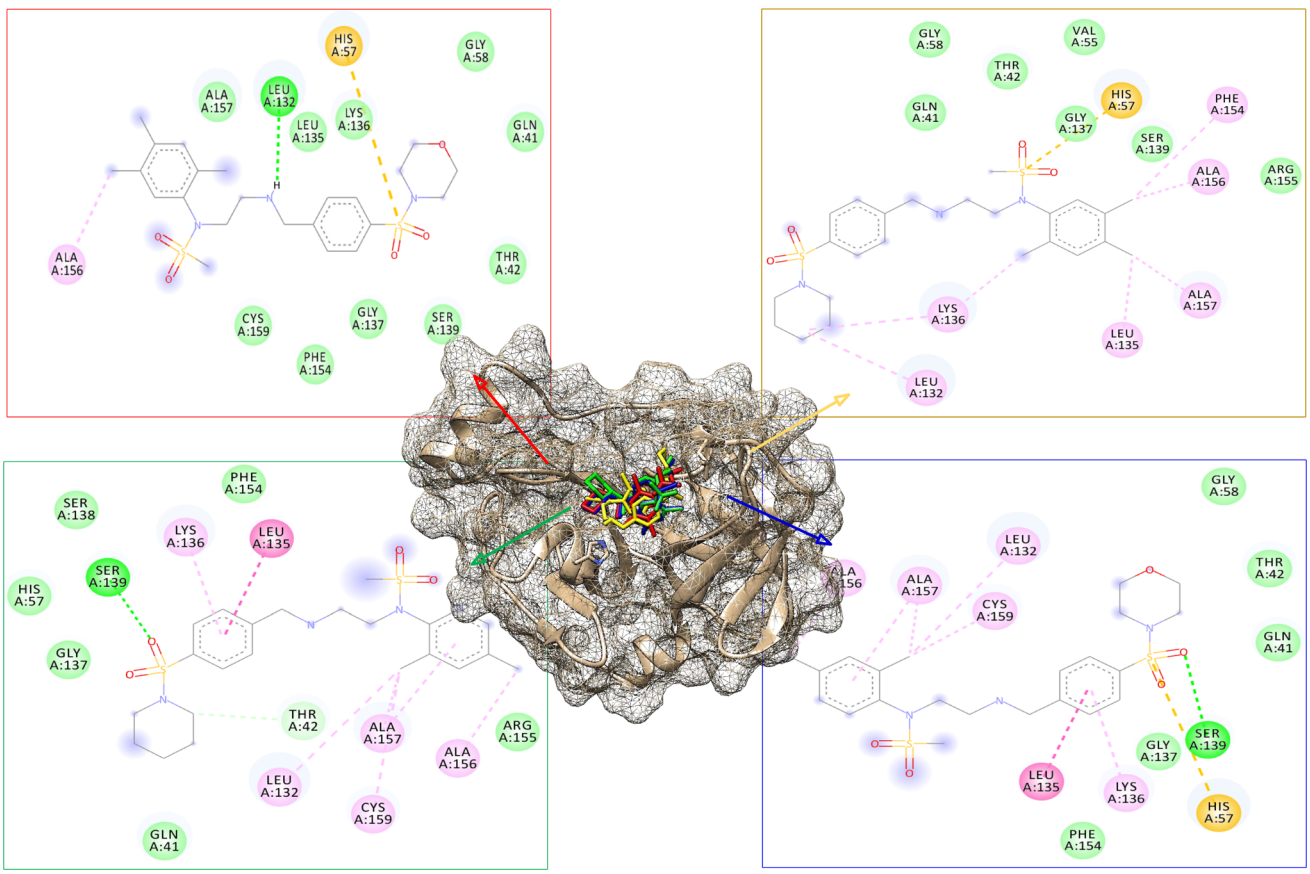


Fig. 5 The modified hit compounds are depicted as bound to the binding site of the modeled protein. The compound TCM is represented in red, TCP in yellow, DCP in green, and DCM in blue. Additionally, their 2D interaction images with the protein are displayed.

compounds for organic synthesis. The position and number of chlorines on the benzene ring and attachment of morpholine or piperidine rings through the linker embodies the difference between the hit compounds. However, the bulky linker (highlighted in the table below) was reoptimized and shortened for better organic synthesis viability. As a result, the hit ZINC compounds were modified for optimized organic synthesis (Table 4). To keep the hit compounds' names simple and meaningful, the compound ZINC000224449889 was re-named trichloromorpholine abbreviated as TCM having three chlorines attached to the benzene ring and morpholine attached to it through the linker. Similarly, the other compounds ZINC000224374291 and ZINC000224374456 were renamed trichloropiperidine (TCP) and dichloropiperidine (DCP), respectively. To check the effect of morpholine moiety with the DCM, an additional compound was synthesized as dichloromorpholine (DCM), replacing piperidine with morpholine.

The modified compounds were re-evaluated against the modeled receptor and the crystal structure by performing their molecular docking, keeping the same parameters and conditions set for the hit compounds (Table 5). The intermolecular interactions between the modified hit compounds and the modeled protein are also depicted in Fig. 5. In the figure, Ser139 and His57 are identified as actively participating residues in the interactions between the protein and the hit compounds. Additionally, the Root Mean Square Fluctuation (RMSF) of the optimized compounds was calculated through MD simulations in the case of the modeled NS3 protease to further assess its flexibility inbound (complex with the hit compounds) and non-bound (single protein) (Fig. 6). The simulation result exhibit overall compactness in bound and non-bound form. More fluctuations were observed between the protein and compound DMC from atoms 300 to 600 and terminal residues in the case of compound DCP.

Characterization data of the hit compounds and their corresponding fragments. All four hit compounds (TCP, DCP, TCM, DCM) were purified by flash chromatography. Compounds were purified with a RediSep Rf Gold Silica Gel Disposable Flash column from Teledyne Isco (4 g, 18 mL min^{-1} from rate) with a gradient of EtOAc in hexanes of 0–100% in 7 min, eluting at 35% EtOAc with a 24–39% yield over 3 steps.

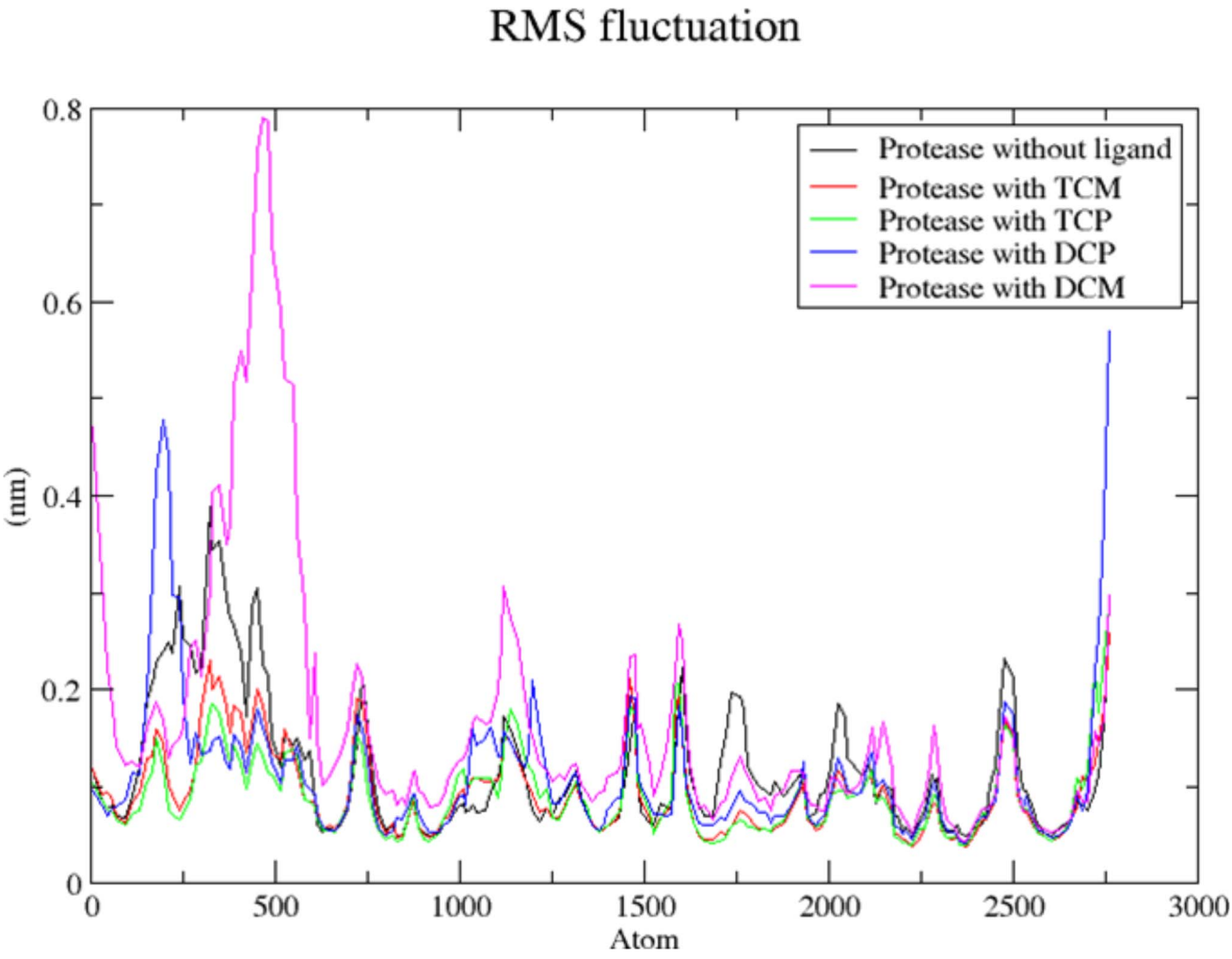
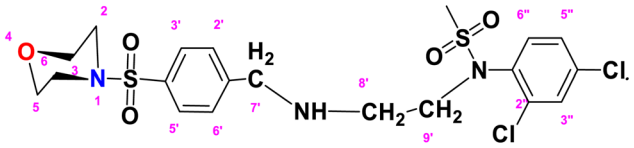


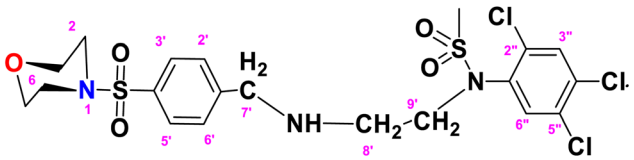
Fig. 6 RMSF analysis of NS3 protease in bound and non-bound form. The peaks of NS3 protease (black), The hit compounds, TCM (red), TCP (green), DCP (blue) and DCM (magenta) are shown.

N-(2,4-Dichlorophenyl)-N-(2-((4-(morpholinosulfonyl)benzyl) amino)ethyl)methane sulfonamide (DCM)



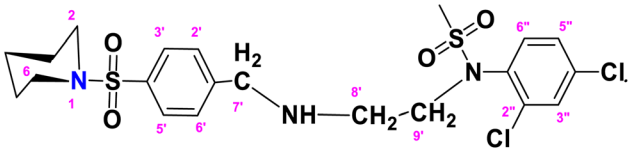
The structure of DCM with the formula $C_{20}H_{25}Cl_2N_3O_5S_2$ is given above. It is solid, white in color, having a molecular weight of 520.49 g mol^{-1}, a melting point of 125 °C, and is soluble in DMSO and chloroform. IR (cm^{-1}) V_{max}: 3256 (N–H), 2810 (Ar-H), 1336 (SO_2 stretching), 1163 (C–N stretching), 1091 (C–O stretching); HRMS (*m*/*z*): [M + 1] 522.46 (13%), 522.06 (24%), 281.98 (68%), 240.07 (73%), 147.97 (64%), 91.05 (100.0%). Anal. calcd: C, 45.98; H, 4.82; N, 8.04; S, 12.27. Found: C, 46.02; H, 4.85; N, 8.06; S, 12.30. 1H NMR (400 MHz, acetone) δ 7.73–7.49 (m, 2H, H-3′ & H-5′), 7.43 (dd, 1H, J = 8.3, 0.5 Hz, H-6′′), 7.35 (d, J = 2.4 Hz, 1H, H-3′′), 7.17 (dd, J = 8.6, 2.4 Hz, 2H, H-2′ and H-6′), 6.98 (d, J = 8.6 Hz, 1H, H-5′′), 3.92 (q, J = 7.1 Hz, 2H, CH_2-9′), 3.60–3.43 (m, 4H, H-3, H-5), 2.77–2.70 (m, 4H, H-2 and H-6), 1.92 (dq, J = 4.5, 2.3 Hz, 2H, CH_2-7′), 1.83 (s, 3H, CH3), 1.06 (t, J = 7.1 Hz, 2H, CH_2-8′). The spectra of DCM are given in Appendix-C.†

N-(2-((4-(Morpholinosulfonyl)benzyl)amino)ethyl)-N-(2,4,5-tri-chlorophenyl)methane sulfonamide (TCM)



The fragment TCM with the formula $C_{20}H_{24}Cl_3N_3O_5S_2$ is given above. It is solid, brown in color, having a molecular weight of 522.47 g mol^{-1}, a melting point of 95 °C, and is soluble in DMSO and chloroform. IR (cm^{-1}) V_{max}: 3251 (N–H), 2809 (Ar-H), 1337 (SO_2 stretching), 1168 (C–N stretching), 1085 (C–O stretching); HRMS (*m*/*z*): [M+] 556.90 (13.0%), 315.60 (98%), 319.94 (3.5%), 240.07 (73%), 147.97 (64%), 91.05 (100.0%). Anal. calcd: C, 43.13; H, 4.34; N, 7.55; S, 11.52. Found: C, 43.16; H, 4.37; N, 7.58; S, 11.55. 1H NMR (400 MHz, acetone) δ 7.60 (dd, J = 8.5, 1.5 Hz, 2H, H-3′ & H-5′), 7.27 (dd, J = 7.9, 1.6 Hz, 2H, H-2′ & H-6′), 6.93 (dddd, J = 18.6, 8.2, 6.7, 1.4 Hz, 2H), 6.75 (t, J = 7.8 Hz, 1H), 6.62 (d, J = 8.1 Hz, 1H), 6.22 (dd, J = 7.4, 1.1 Hz, 1H), 5.24 (s, 2H), 2.89 (s, 14H), 2.06 (p, J = 1.8 Hz, 4H). The spectra of TCM are given in Appendix-C.†

N-(2,4-Dichlorophenyl)-N-(2-(4-(piperidin-1-ylsulfonyl) benzylamino)ethyl)methane sulfonamide (DCP)



The fragment DCP with the formula $C_{21}H_{27}Cl_2N_3O_4S_2$ is given above. It is solid, white in color, has a molecular weight of 520.49 g mol^{-1}, a melting point of 125 °C, and is soluble in DMSO and chloroform. IR (cm^{-1}) V_{max}: 3277 (N–H), 2825 (Ar-H), 1334 (SO_2 stretching), 1166 (C–N stretching), 1091 (C–O stretching); HRMS (*m*/*z*): [M + 1] 520.08 (13%), 519.08 (24%), 281.98 (68%), 238.09 (75%), 147.97 (64%), 91.05 (100.0%). Anal. calcd: C, 48.46; H, 5.23; N, 8.07; S, 12.32. Found: C, 48.50; H, 5.28; N, 8.12; S, 12.37. 1H NMR (400 MHz, acetone) δ 7.75–7.33

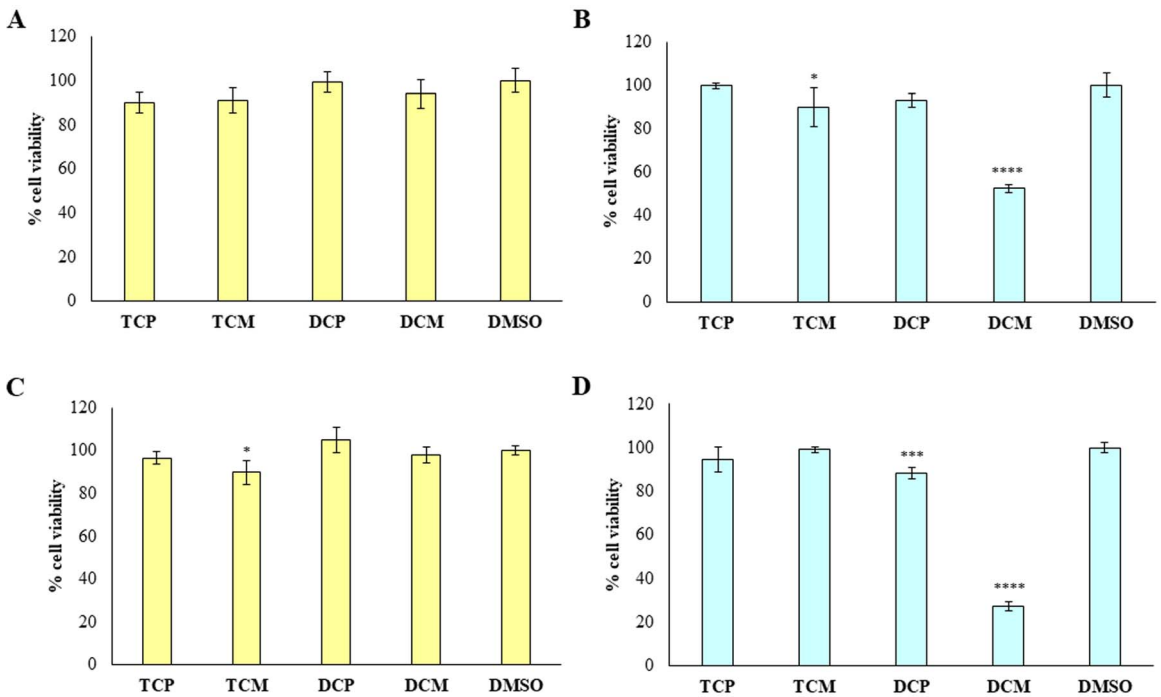


Fig. 7 MTT test on HGFs treated with the hit compounds. MTT test on HGFs treated with TCP, TCM, DCP and DCM compounds at 10 μM (left histograms) and 50 μM (right histograms) for 48 h (A and B) and 72 h (C and D); DMSO: control vehicle. Data are presented as mean% ± SD. The most representative of three different experiment is shown. (B) * *vs.* DMSO p = 0.0357. **** *vs.* DMSO p < 0.0001; (C) * *vs.* DMSO p = 0.0149; (D) *** *vs.* DMSO. p = 0.0002, **** *vs.* DMSO p < 0.0001.

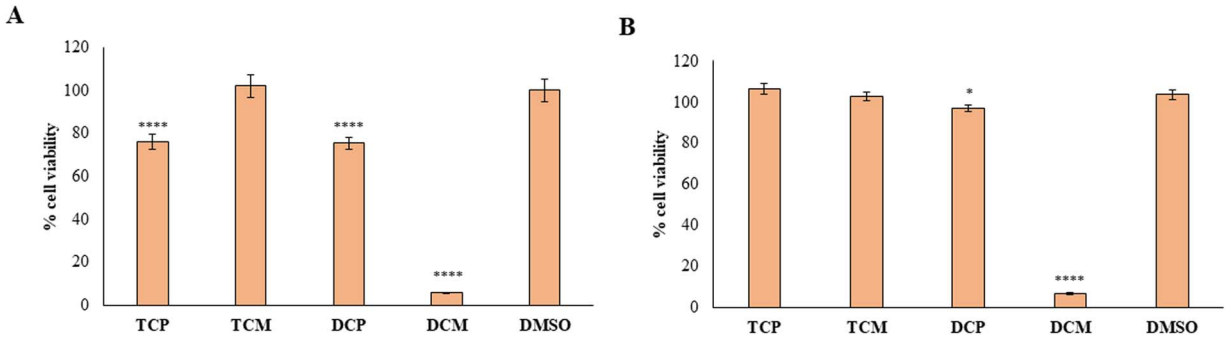


Fig. 8 MTT test on AGS treated with the hit compounds. MTT test on AGS treated with TCP, TCM, DCP and DCM at 50 μM for 48 h (A) and 72 h (B); DMSO: control vehicle. Data are presented as mean% ± SD. The most representative of three different experiment is shown. (A) **** *vs.* DMSO $p < 0.0001$; (B) * *vs.* DMSO $p = 0.0172$, **** *vs.* DMSO $p < 0.0001$.

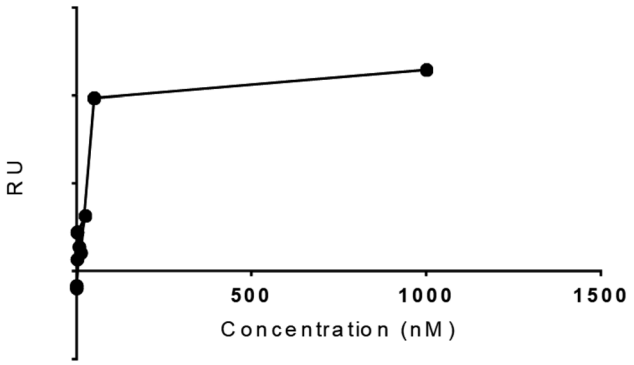
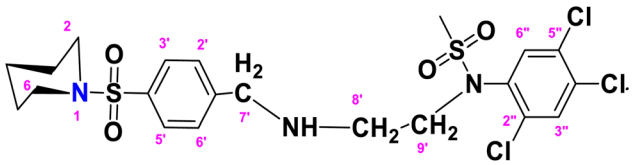


Fig. 9 Response *versus* concentration graph.

(m, 4H, H-3′, H-5′, H-3″ and H-6″), 7.17 (dd, $J = 8.6$, 2.4 Hz, 2H, H-2′, H-6′), 6.98 (d, $J = 8.6$ Hz, 1H, H-5″), 3.46 (s, 1H, H-7″), 2.72 (dt, $J = 18.6$, 6.8, 2.8 Hz, 4H, H-2 and H-6), 2.10–1.65 (m, 2H, H-4), 1.43 (p, $J = 5.7$ Hz, 2H, H-3 and H-5), 1.29 (d, $J = 5.5$ Hz, 2H, CH_2-9′), 1.16 (s, CH3, 3H), 1.06 (d, $J = 6.6$ Hz, 2H, CH_2-8′). The spectra of DCP are given in Appendix-C.†

N-(2-(4-(Piperidin-1-ylsulfonyl)benzylamino)ethyl)-N-(2,4,5-trichlorophenyl)methane sulfonamide (TCP)



The fragment TCP with the formula $C_{21}H_{26}Cl_3N_3O_4S_2$ is given above. It is solid, light brown in color, has a molecular weight of 554.94 g mol^{-1}, a melting point of 153 °C, and is soluble in DMSO and chloroform. IR (cm^{-1}) V_{max}: 3189 (N–H), 2814 (Ar-H), 1325 (SO_2 stretching), 1158 (C–N stretching), 1099 (C–O stretching); HRMS (*m*/*z*): [M+] 553.04 (15.0%), 315.59 (98%), 319.94 (3.5%), 238.32 (73%), 147.97 (64%), 91.05 (100.0%). Anal. calcd: C, 45.45; H, 4.72; N, 7.57; S, 11.56. Found: C, 45.50; H, 4.77; N, 7.62; S, 11.61. 1H NMR (400 MHz, DMSO) δ 7.71 (dd, $J = 8.3$, 6.6 Hz, 2H, H-3′ and H-5′), 7.62–7.52 (m, 2H, H-3″ and H-6″), 6.74–6.66 (m, 2H, H-2′ and H-6′), 4.55 (d, $J = 6.0$ Hz, 2H, CH_2-9′), 4.03 (brs, 2H, CH_2-7′), 3.62 (t, $J = 4.7$ Hz, 4H, H-2 and H-6), 3.16–3.10 (m, 2H, H-8′), 2.86 (q, $J = 6.0$ Hz, 4H, H-3 and H-5), 2.09 (s, 3H, CH_3), 1.51–1.24 (m, 2H, H-4). The spectra of TCP are given in Appendix-C.†

Biological assay results

To provide the safety profile of these four promising compounds, firstly they have been tested on HGFs, at 10 and 50 μM for 48 h, comparing the results with the vehicle DMSO, assumed as control. The results reported in Fig. 7 demonstrated that at 10 μM, the four compounds didn't affect the cell viability of healthy cells, disregarding the substitution pattern, with respect to DMSO (Fig. 7A). At 50 μM (Fig. 7B), TCP and DCP show the same result, whereas a statistically significant reduction in cell viability is evidenced when TCM and DCM are administered with respect to DMSO, with a major extent for DCM (52% of cell viability), even if the cell viability rate never goes under 50%. Thus, for the four compounds, we can assume an $IC_{50} > 50$ μM at 48 h. After 72 h of treatment, when the newly synthesized compounds are administered at 10 μM, a lower viability level is recorded in the presence of TCM (89.9% of cell viability), while, TCP, DCP, and DCM do not reveal any significant modification of HGF viability with respect to control sample (DMSO) (Fig. 7C). At 50 μM the previous trend recorded after 48 h was confirmed ($IC_{50} > 50$ μM), evidencing a slight but statistically significant alteration of the non-cancerous cells' viability (IC_{50} ranging between 10 and 50 μM) after DCP and DCM administration (Fig. 7D).

Then, the capability of the novel compounds to affect the viability of gastric adenocarcinoma AGS cell line at 50 μM (maximum concentration in the previous experiment after 48 (Fig. 8A) and 72 h (Fig. 8B) of treatment), has been determined. After 48 h of treatment, the cell viability percentage appears significantly reduced in the presence of TCP, DCP, and DCM compared to DMSO, with a major extent for DCM which leads to record an extremely low cell viability rate (5.7%). After 72 h of treatment, DCP still discloses a significant reduction of cell viability even if it is estimated to be of 96%, approximately. Conversely, DCM does not allow an AGS recovery considering that, after 72 h of treatment, the effect appears to be still strong keeping the cell viability level at very low percentages (6.8%). Thus, it can be argued that TCP, TCM, and DCP are well tolerated by AGS even if a slightly lower cell viability percentage with respect to HGFs

(IC_{50} > 50 μM) can be highlighted. On the contrary, DCM exerts a marked and pronounced effect on tumoral cell viability with respect to healthy HGFs (IC_{50} < 50 μM). These results pinpoint how slight differences in the substitution pattern (the chlorine atom at position 5 in the morpholino series) could influence the selection of the administration dose to avoid unpleasant side effects.

Surface Plasmon Resonance (SPR)

Determination of dissociation equilibrium constant (K_D) by SPR. SPR technique is an efficient biophysical method for the determination of affinity and kinetics of synthesized compounds with target proteins. The terms in SPR are a little different from conventional terminology. The ligand refers to the interactant attached to the sensor surface while the analyte is referred to as the interactant present in the sample solution injected over the surface. There are a number of coupling methods available for SPR studies depending on the target analyte and ligand.

In the presented study, we first tried the amine coupling method adopted from the Hyun Lee work.[37] The amine coupling method did not work for our Hepatitis C Virus (HCV-1a) NS3 protease/helicase immunodominant region protein (aa 1356–1459, GST tag). As the used ligand was not stable enough to support amine coupling. After amine coupling, we tried the GST coupling method for the SPR study of one of the hits TCM. It is important to find suitable immobilization pH prior to immobilization. For the two different PH buffers (sodium acetate buffer pH −5.0 and pH 4.5) were used. The response (RU) was dropped at a lower pH so the immobilization buffer of pH 5.0 was proceeded for immobilization. The sensorgram begins flattening out after the covalent coupling, which may contribute to the robustness of the assay. The graphs of pH scouting are given in Appendix-D.†

Following parameters were set for immobilization wizard:

Immobilization Setup

Chip type CM5

Flow cells per cycle 2

Flow cell 1, 2

Specify contact time and flow rate

Method S200 Amine

Ligand antiGST 30 μg/mL

Contact Time 420 (s)

Flow rate 10 (μL min^{-1})

The theoretical R_{max} value calculated is 54.78 from 1864.2 RU, MWA 556.91 Da, MWL 37900 Da and SM 2. The R_{max} value was calculated by using following equation:

$$R_{max}\frac{MWA}{MWL}\times SM\times RU$$

where, R_{max} = theoretical maximum binding capacity, MWA = molecular weight analyte (Da), MWL = molecular weight ligand (Da), SM = stoichiometric ratio (theoretical value of binding of analyte molecule to ligand, here is 1 : 2), RU = immobilized amount.

Single-cycle runs were used for small molecule binding analysis. The buffer used to prepare the protein samples was HBS-EP (20 mM HEPES, 150 mM NaCl, 10 mM $MgCl_2$, 0.01% Tween 20, pH = 7.4). For runs with small molecules, an additional 1% DMSO was added for solubility. GST capture kit conditions (Cytiva catalog number BR100223) were used to capture anti-GST antibody on both the sample and reference cells (7 min immobilization, 10 μL min^{-1} flow rate). Both surfaces had high-affinity sites capped with an additional 3 min of GST flowed over (5 μg mL^{-1} concentration, 5 μL min^{-1} flow rate) followed by regeneration (10 mM glycine, pH = 2.2). On the subtractive reference surface, GST was immobilized (10 μg mL^{-1}, 5 μL min^{-1}, 5 min). On the sample surface, GST-tagged HCV was immobilized in a similar fashion (10 μg mL^{-1}, 5 μL min^{-1}, 5 min). The wizard parameters for single-cell kinetics are given in Appendix-E.†

The experimental Rmax of TCM as function time was found 8.1 in FC 2–1 and K_D value 1.01×10^{-11}. The five concentrations 2.4, 12, 60, 300, and 1500 μM were formed for TCM to study the dose–response curve as shown (Fig. 9), the following curve shows the increase in response with sample concentration till 50 nM with RU 39.4 and keep increasing on higher concentration. These results gave encouragement to explore the other hits for the SPR binding and kinetics study. The dose response should be optimized further for more data points.

Conclusion

The computational chemistry, medicinal chemistry, and bioinformatics approaches have been employed to find protease inhibitors. This interdisciplinary methodology has enabled to analyze the problem under the wider scope. The study has revealed that theoretical results are corroborated by experimental findings. The molecular dynamics and pharmacokinetics studies revealed that the hit compounds ZINC000224449889, ZINC000224374291, and ZINC000224374456 and the derivative of ZINC000224374456 are potential drug contenders against HCV NS3 protease genotype 3a. The optimized compounds namely TCP, TCM, and DCP displayed a safe profile of cell viability (HGFs *versus* AGS) up to 50 μM, whereas DCM should be administered at lower concentrations.

We got the limited time facility for SPR and TCM was selected for SPR studies to evaluate the binding and kinetics of the compounds against GST-HCV NS/34A protein. The group was able to purchase NS3/4A 1b GST tag HCV protein while NS3/4A 3b genotype is not available as an isolated recombinant protein for assay. SPR assay was developed for HCV by hit and trial method. It was found that the GST capture approach is effective for the HCV SPR assay. There is a need for the

availability of our target genotype recombinant polyprotein for further exploration of hits for target-specific studies. Preliminary SPR results demonstrated that TCM was bound in 1 : 1 binding mode with the target protein and was found effective at 50 nM concentration. We aim to run SPR assays of other hits on the availability of SPR facility in future.

Institutional review board statement

The study was conducted in accordance with the Declaration of Helsinki and was approved by the Local Ethical Committee of the University of Chieti-Pe-scara (Chieti, Italy, approval number. 1173, approved on 31/03/2016).

Author contributions

Rashid Hussain: conceptualization; methodology; formal analysis; investigation; data curation; writing – original draft preparation; computational studies, preliminary synthesis. Zulkarnain Haider: writing, synthesis, characterization, investigation. Hira Khalid and Simone Carradori: writing – review and editing; project design and management, characterization; supervision, investigation; SPR studies. M. Qaiser Fatmi: review and editing; supervision. Susi Zara and Amelia Cataldi: cell-based assays.

Conflicts of interest

None.

Acknowledgements

IT department of Forman Christian College (A Chartered University), Lahore Pakistan, for cloud computing facility. Department of Medicinal Chemistry, University of Minnesota, Twin Cities, USA for high-resolution NMR and SPR facilities.

References

1 Q. L. Choo, G. Kuo, a J. Weiner, L. R. Overby, D. W. Bradley and M. Houghton, *Science*, 1989, **244**, 359–362.
2 K. Mohd Hanafiah, J. Groeger, A. D. Flaxman and S. T. Wiersma, *Hepatology*, 2013, **57**, 1333–1342.
3 C. Ferri, M. Sebastiani, D. Giuggioli, M. Colaci, P. Fallahi, A. Piluso, A. Antonelli and A. L. Zignego, *World J. Hepatol.*, 2015, **7**, 327–343.
4 I. Rusyn and S. M. Lemon, *Cancer Lett.*, 2014, **345**, 210–215.
5 S. M. Borgia, C. Hedskog, B. Parhy, R. H. Hyland, L. M. Stamm, D. M. Brainard, M. G. Subramanian, J. G. McHutchison, H. Mo, E. Svarovskaia and S. D. Shafran, *J. Infect. Dis.*, 2018, **218**, 1722–1729.
6 D. B. Smith, J. Bukh, C. Kuiken, A. S. Muerhoff, C. M. Rice, J. T. Stapleton and P. Simmonds, **https://talk.ictvonline.org/ictv_wikis/flaviviridae/w/sg_flavi/56/hcv-classification**.
7 K. Hara, M. M. Rivera, C. Koh, S. Sakiani, J. H. Hoofnagle and T. Heller, *J. Clin. Microbiol.*, 2013, 03344.
8 J. Gentzsch, C. Brohm, E. Steinmann, M. Friesland, N. Menzel, G. Vieyres, P. M. Perin, A. Frentzen, L. Kaderali and T. Pietschmann, *PLoS Pathog.*, 2013, **9**, e1003355.
9 R. Bartenschlager, F.-L. Cosset and V. Lohmann, *J. Hepatol.*, 2010, **53**, 583–585.
10 M. E. Major and S. M. Feinstone, *Hepatology*, 1997, **25**, 1527–1536.
11 C. M. Taylor, Q. Wang, B. A. Rosa, S. C.-C. Huang, K. Powell, T. Schedl, E. J. Pearce, S. Abubucker and M. Mitreva, *PLoS Pathog.*, 2013, **9**, e1003505.
12 T. U. Chae, S. Y. Choi, J. W. Kim, Y.-S. Ko and S. Y. Lee, *Curr. Opin. Biotechnol.*, 2017, **47**, 67–82.
13 M. R. Riaz, G. M. Preston and A. Mithani, *ACS Synth. Biol.*, 2020, **9**, 1069–1082.
14 A. Grakoui, D. W. McCourt, C. Wychowski, S. M. Feinstone and C. M. Rice, *J. Virol.*, 1993, **67**, 2832–2843.
15 A. E. Gorbalenya, A. P. Donchenko, E. V Koonin and V. M. Blinov, *Nucleic Acids Res.*, 1989, **17**, 3889–3897.
16 C. Failla, L. Tomei and R. De Francesco, *J. Virol.*, 1994, **68**, 3753–3760.
17 M. U. Ashraf, K. Iman, M. F. Khalid, H. M. Salman, T. Shafi, M. Rafi, N. Javaid, R. Hussain, F. Ahmad, S. Shahzad-Ul-Hussan, S. Mirza, M. Shafiq, S. Afzal, S. Hamera, S. Anwar, R. Qazi, M. Idrees, S. A. Qureshi and S. U. Chaudhary, *Med. Res. Rev.*, 2019, **39**, 1091–1136.
18 R. Hussain, H. Khalid and M. Q. Fatmi, *J. Comput. Biophys. Chem.*, 2021, **20**, 631–639.
19 R. Hussain, H. Khalid and M. Q. Fatmi, *Pure Appl. Chem.*, 2022, 1–10.
20 A. Śledź Pawełand Caflisch, *Curr. Opin. Struct. Biol.*, 2018, **48**, 93–102.
21 M. Umer and M. Iqbal, *World J. Gastroenterol.*, 2016, **22**, 1684–1700.
22 A. Ece, *J. Biomol. Struct. Dyn.*, 2020, **38**, 565–572.
23 L. Llorach-Pares, A. Nonell-Canals, M. Sanchez-Martinez and C. Avila, *Mar. Drugs*, 2017, **15**, 366.
24 K. Chan, N. Frankish, T. Zhang, A. Ece, A. Cannon, J. O'Sullivan and H. Sheridan, *J. Pharm. Pharmacol.*, 2020, **72**, 927–937.
25 A. Maryam, R. R. Khalid, S. C. Vedithi, E. C. E. Abdulilah, S. S. Ç\inaro\uglu, A. R. Siddiqi and T. L. Blundell, *Comput. Struct. Biotechnol. J.*, 2020, **18**, 1625–1638.
26 M. A. Alagöz, Z. Özdemir, M. Uysal, S. Carradori, M. Gallorini, A. Ricci, S. Zara and B. Mathew, *Pharmaceuticals*, 2021, **14**, 183.
27 K.-B. Li, *Bioinformatics*, 2003, **19**, 1585–1586.
28 G. N. Ramachandran, C. Ramakrishnan and V. Sasisekharan, *J. Mol. Biol.*, 1963, **7**, 95–99.
29 T. Sterling and J. J. Irwin, *J. Chem. Inf. Model.*, 2015, **55**, 2324–2337.
30 W. J. Allen, T. E. Balius, S. Mukherjee, S. R. Brozell, D. T. Moustakas, P. T. Lang, D. A. Case, I. D. Kuntz and R. C. Rizzo, *J. Comput. Chem.*, 2015, **36**, 1132–1156.
31 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–1718.

32 W. Kulig, M. Pasenkiewicz-Gierula and T. Róg, *Data Brief*, 2015, **5**, 333–336.
33 F. Jiang, C.-Y. Zhou and Y.-D. Wu, *J. Phys. Chem. B*, 2014, **118**, 6983–6998.
34 D. J. Wood, S. Korolchuk, N. J. Tatum, L.-Z. Wang, J. A. Endicott, M. E. M. Noble and M. P. Martin, *Cell Chem. Biol.*, 2019, **26**, 121–130.
35 T. Schwede, J. Kopp, N. Guex and M. C. Peitsch, *Nucleic Acids Res.*, 2003, **31**, 3381–3385.
36 *RCSB PDB – 6P6S: HCV NS3/4A protease domain of genotype 3a in complex with glecaprevir*, **https://www.rcsb.org/structure/6P6S**, (accessed 13 September 2023).
37 J. Ren, I. Ojeda, M. Patel, M. E. Johnson and H. Lee, *Bioorg. Med. Chem. Lett.*, 2019, **29**, 2349–2353.