

Development of a Novel Pharmacophore Model Guided by the Ensemble of Waters and Small Molecule Fragments Bound to SARS-CoV-2 Main Protease

Pawan Kumar^[a] and Debasisa Mohanty^{*[a]}

Abstract: Recent fragment-based drug design efforts have generated huge amounts of information on water and small molecule fragment binding sites on SARS-CoV-2 M^{Pro} and preference of the sites for various types of chemical moieties. However, this information has not been effectively utilized to develop automated tools for *in silico* drug discovery which are routinely used for screening large compound libraries. Utilization of this information in the development of pharmacophore models can help in bridging this gap. In this study, information on water and small molecule fragments bound to M^{Pro} has been utilized to develop a novel Water Pharmacophore (Waterphore) model. The Waterphore model can also implicitly represent

the conformational flexibilities of binding pockets in terms of pharmacophore features. The Waterphore model derived from 173 apo- or small molecule fragment-bound structures of M^{Pro} has been validated by using a dataset of 68 known bioactive inhibitors and 78 crystal structure bound inhibitors of SARS-CoV-2 M^{Pro}. It is encouraging to note that, even though no inhibitor data has been used in developing the Waterphore model, it could successfully identify the known inhibitors from a library of decoys with a ROC-AUC of 0.81 and active hit rate (AHR) of 70%. The Waterphore model is also general enough for potential applications for other drug targets.

Keywords: Waterphore · Pharmacophore · SARS-CoV-2 · Main Protease · Binding hotspots


1 Introduction

The recent outbreak of SARS-CoV-2 has necessitated efforts by the scientific community to design potent drugs targeting proteins that play essential roles in the viral life cycle and its survival within the host.^[1] One of the best-characterized drug targets is the Main Protease (M^{Pro}), also called 3CL^{Pro} (3-chymotrypsin-like cysteine protease) that is involved in the multiple cleavage process of the polyproteins ORF1a and ORF1ab encoded by the viral genome.^[2] Main protease is a cysteine protease having a non-canonical Cys-His catalytic dyad. It has a high degree of sequence and structural similarity to M^{Pro} proteins from other Coronaviruses.^[2] The current anti-M^{Pro} inhibitor development project reports that the PF-00835321 compound, which was earlier identified for SARS M^{Pro}, is also active against SARS-CoV-2 M^{Pro} protein.^[3] The Pfizer team is now conducting a clinical trial for this compound to determine the safety profile (<https://www.pfizer.com/science/coronavirus>). The curative potential of anti-HIV drugs Lopinavir-Ritonavir (Kaletra[®]) which have been identified by drug repurposing studies, remains minimal.^[4] High-throughput screening (HTS) of FDA-approved anti-viral drugs and focused collection of proteases inhibitors have identified HCV drugs, Boceprevir and telaprevir, GC-376 and calpain inhibitors II and XII as promising M^{Pro} inhibitors having activity in the low micromolar range.^[5,6] To assist in finding new drug candidates for M^{Pro}, several *in silico* studies have

also been reported.^[7-9] However, most of them have used standard drug discovery or molecular modeling software-based generic criteria for inhibitor design and a considerable volume of structural information on SARS-CoV-2 M^{Pro} has not been fully utilized.

The binding affinity of a ligand or inhibitor to a given receptor depends on several factors like conformational flexibility of the ligand/receptor,^[10] shape/geometry of the binding pocket and available specificity determining residues in the binding site^[11] etc. Out of the available interaction space for the ligand in a specific receptor, binding hotspot regions significantly contribute to the overall binding energy.^[12] Several computational approaches have been successfully developed^[13] for different well-known target proteins to characterize major binding hotspots critical for ligand/inhibitor binding.^[14,15] Apart from binding pocket residues, water molecules also play an important role in ligand recognition and stability of the

[a] P. Kumar, D. Mohanty
National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi – 110067, India
Tel: +91-11-26703749
Fax: +91-11-26742125
E-mail: deb@nii.ac.in
deb@nii.res.in

 Supporting information for this article is available on the WWW under <https://doi.org/10.1002/minf.202100178>

protein-ligand complex.^[16] Bound water molecules in the ligand-binding sites of protein crystal structures emulate the key element of interactions, thereby provide important information and understanding that a ligand should possess for high-affinity binding to a targeted protein.^[17,18] Furthermore, a water-filled cavity compels potential small molecules to compete with occupied water molecules at the binding site and consequently, their displacement to the bulk solvent affects the thermodynamic signature of the binding free energy.^[19] Since pharmacophore-based methods are more versatile and straightforward to implement as per the requirement,^[20] Pharmacophore modeling can, in principle, be used to utilize structurally conserved water binding sites in inhibitor design studies. However, in most cases, 3D-pharmacophore models are developed using inhibitor-bound crystal structures, while water or small molecule fragment-bound crystal structures are not used. Only a limited number of studies have reported apo-receptor based pharmacophore models.^[21–24] Such models can also incorporate receptor flexibilities by utilizing multiple conformations of receptors, as shown in the case of HIV-integrase, HIV-Protease and DHFR.^[25,26] Similarly, novel methods like pharmacophore modeling using Site-Identification by Ligand Competitive Saturation (SILCS) approach^[23,24] can utilize apo-protein structures to include protein flexibility and desolvation effects.

In this study, we have utilized the available crystal structures of ligand-free as well as ligand-bound crystal structures of Main protease (M^{pro}) from SARS-CoV-2 to analyze the distribution of the water molecules in the substrate-binding site and other small fragment molecules to map potential hotspot regions in the substrate-binding site of M^{pro} . The identified water-binding clusters have been further utilized to define displaceable water binding sites and corresponding pharmacophore features. Based on these results, a novel rule-based protocol, Crystal Water/Fragment Ensemble-based Pharmacophore (CWFEP) or Waterphore, has been developed. Since the waterphore model does not use any information from inhibitor bound crystal structures of SARS-CoV-2 or other known inhibitors of M^{pro} identified by biochemical assays, the waterphore model has been benchmarked using these data as independent test sets. After establishing high prediction accuracy of the waterphore model, its utility in identifying new high-affinity inhibitors for M^{pro} has been investigated by prediction of novel M^{pro} inhibitors from DrugBank, Protease inhibitor library and anti-viral compound library.

2 Materials and Methods

2.1 Compilation and Preparation of Crystal Structures of M^{pro}

As of 20th January 2021, a total of 251 crystal structures of SARS-CoV-2 M^{pro} were available in the Protein Data Bank

(<https://www.rcsb.org/>). Downloaded structural data were manually scrutinized for missing residues information in the substrate-binding site and further divided into three categories, 1) Apo structures – M^{pro} structures having no ligand bound in the substrate-binding site 2) small fragment bound structures – all structures having bound small-fragment molecule in the substrate-binding site and 3) ligand-bound complexes – M^{pro} structures having inhibitor like bound ligand molecules in the substrate-binding site. To differentiate the fragment molecules from ligand/inhibitor molecules, criteria based on molecular weight (MW) was employed. From the bimodal MW distribution (Figure S1), small molecules having MW less or equal to 300 Da were selected as fragment molecules and the rest compounds were classified as ligand/inhibitor molecules. A significant collection of the fragment-bound complexes were deposited by Diamond Light Source employing XChem crystallographic fragment screening platform.^[27] Explored fragment molecules span the whole substrate binding site and can be used to understand the all-possible available interaction pattern to design inhibitors with improved binding affinity.

Out of the 251 M^{pro} structures, 79 were classified as apo- M^{pro} structures, 94 as fragment-bound M^{pro} structures and 78 were selected as ligand/inhibitor bound complexes. Among the 94 fragment-bound complexes, 54 fragments were covalently bonded with the catalytically essential Cys-145 residue of M^{pro} . Along with the small molecules, M^{pro} structures also had a large number of bound water and Dimethyl Sulfoxide (DMS) molecules around the substrate or inhibitor binding region (Table S1). All selected PDB structures were prepared using the 'Prepare Proteins' module of the DS2020.^[28] This module takes care of the alternative conformations, the addition of missing hydrogen atoms, the correctness of the bond order and correct protonation states of titratable residues. This module also optimizes the hydrogen bond network at selected pH and this optimization includes assigning hydrogen atoms according to the calculated pK values, optimize tautomeric forms of histidine residues, flipping amide groups of Asn/Gln and the carboxyl group of protonated Asp/Glu to get lowest energy conformer.

2.2 Pharmacophore Modeling Guided by the Ensemble of Bound Water and Fragment Molecules

In order to identify structurally conserved water binding sites and favorable hotspot regions in the substrate-binding site of M^{pro} , a pharmacophore modeling protocol was designed in which both fragment-bound and apo- M^{pro} structures were utilized. This ensemble-based pharmacophore modeling protocol was used to determine the Crystal Water/Fragment Ensemble-based Pharmacophore (CWFEP) features for the SARS-CoV-2 main protease (Figure 1). It consisted of three major steps.

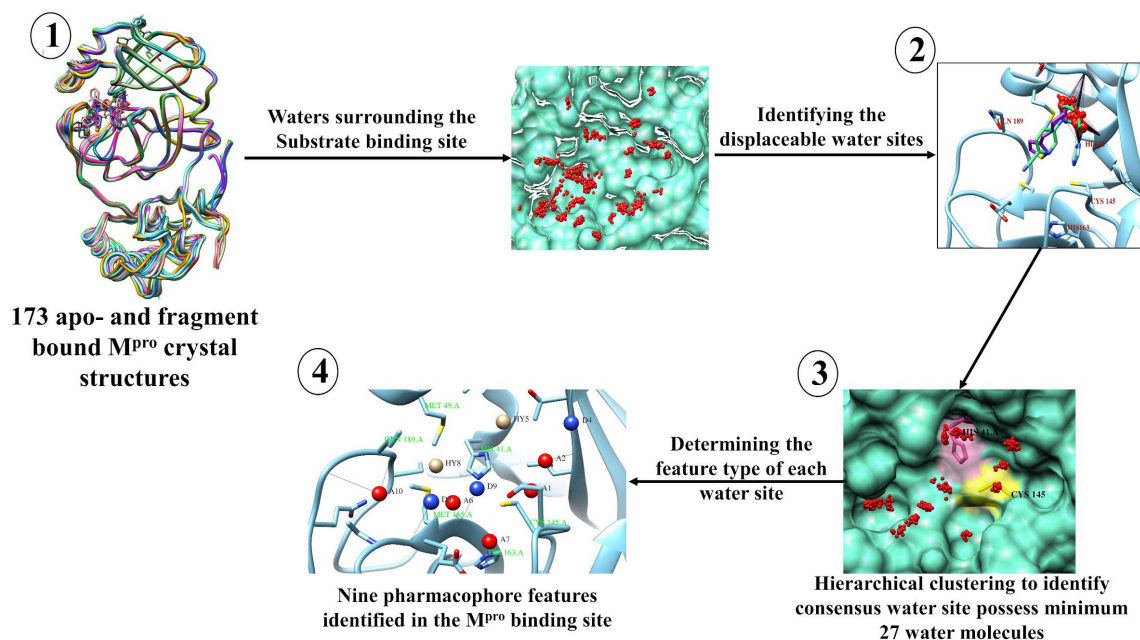


Figure 1. Waterphore identification protocol developed using crystal structure-bound waters and small molecules.

In the first step, a total of 173 structures were optimally superimposed onto the reference M^{PrO} structure 5R8T (resolution 1.27 Å) using coordinates of C^α atoms. The structural superimpositions were carried out using UCSF Chimera^[29] software. The number of bound water molecules on a given M^{PrO} structure typically ranged from 24 to 927. After determining the center mean position (CoM) of the substrate-binding site of M^{PrO} , all superimposed structures were analyzed to extract the water molecules lying within a distance of 8 Å from the CoM position (Figure S2). Binding site waters were further filtered to select only those which were at a minimum distance of 1.5 Å from the solvent accessible surface residues of the protein and formed hydrogen bonds with protein residues (Figure 1, panel 1). The minimum distance of 1.5 Å criterion was adopted because 1.4 Å water probe radius is generally used to mark the protein surface. So, any water molecule which is likely to be replaced by a ligand should not be part of the protein surface or buried in the interior of the protein.

The second step involved the identification of sites where bound water molecules can be displaced by the incoming ligand (Figure 1, panel 2). The ensemble of water and small molecule fragment binding sites were analyzed after they were transformed from different M^{PrO} structures to the reference structure. The visual analysis of the ensembles revealed that they formed three distinct ensembles in the substrate-binding site of M^{PrO} . One ensemble consisted of bound waters which were consistently found in the vast majority of the M^{PrO} crystal structures, while the second ensemble corresponded to sites which were occupied by small-molecule fragments in significant number of M^{PrO} structures in our dataset, hence this ensemble

will preserve less number of water molecules at those sites. This second category can be called “displaceable water ensemble” and the corresponding binding sites can be occupied by ligand molecules in ligand/fragment bound M^{PrO} structures. The third ensemble corresponded to sites where only small molecule fragments were found to be bound in the apo- M^{PrO} structures.

In the third step, all the water molecules in the “displaceable water ensemble” were clustered using an automated computational approach to identify displaceable water clusters which map to various sub-sites in the substrate-binding pocket of M^{PrO} (Figure 1, panel 3). Scipy package was used for euclidean distance metric-based hierarchical clustering using 3D coordinates of water molecules in the cluster. The complete linkage algorithm was used with an inconsistency cutoff of 2.2 Å. The clusters corresponding to water-binding sites occupied in a minimum of 27 out of the 173 apo- or fragment-bound overlaid M^{PrO} crystal structures were annotated as displaceable water clusters (Table S2). These cutoffs were selected to ensure that known subsites (S1, S2, S4 and S1') of M^{PrO} mapped to displaceable water clusters. The centers and the radii of these significantly populated displaceable water clusters were calculated from the coordinates of the water molecules which mapped onto a given cluster. The center of each such cluster was set to the average position of the oxygen atoms, and the radii were set based on the average RMSD of water molecules in the corresponding cluster. These clusters were used as features of water/fragment ensemble-based pharmacophore characterized by their positions and radii.

In the fourth step, the feature type (H-bond donor, acceptor etc.) of each water site was determined after analyzing the type of features projected by the fragment molecules at any particular water site (Figure 1, panel 4). All pharmacophore features encompassed by fragment molecule were calculated using the Rdkit package^[30] and water sites were annotated based on the maximally occupied feature type from the fragment molecules. This criterion was used to annotate the identified water cluster as a conserved sub-site in terms of pharmacophore feature types. For further analysis and representation purposes, waterphore features were represented in Catalyst query^[31] format and imported to the Discovery Studio 2020^[28] for all the analyses carried out in this study.

2.3 The Favorable Interaction Energy of the Water Feature

In order to quantify the interaction energy of each identified feature of the Waterphore, probe-based Molecular Interaction Field (MIF) was calculated using the Grid software^[13] from Molecular Discovery (<https://www.moldiscovery.com>). MIF calculation helps in finding favorable interaction energy spots accessible by different sizes and types of chemical probes. For MIFs analysis, the apo-structure of the M^{Pro} (PDB ID: 6YB7) was selected and as a preprocessing step, the structure was examined for missing residue information and all bound crystallographic water molecules, small molecules (such as DMS) and ions were removed. The prepared apo-structure was used for the MIF calculations by 22d version of GRID software. As per the standard protocol proposed by Goodford *et al.*^[32] for processing the receptor for molecular interaction field calculation, GRIN module of the GRID package was used to examine the correctness of the structure and to add hydrogens. An appropriate number of counterions were further added to neutralize the total charge of the receptor protein. The positions of the counterions were near the protein surface and three sodium ions were added to neutralize the total charge of M^{Pro} protein. The GRID box was centered around the M^{Pro} binding site and its dimensions were chosen to encompass all relevant binding site residues.

Two polar probes (Neutral Amide, Carbonyl) and a water probe were used to calculate MIF. The first two probes were selected because most protease inhibitors are peptide-like molecules and are very rich in these chemical groups. They are primarily involved in hydrogen bonding with residues in the binding pocket of the protease. Amide group can have a hydrogen bond donor (HBD) feature, while carbonyl can have a hydrogen bond acceptor (HBA) feature of the pharmacophore. The grid box of size 27 Å × 35 Å × 25 Å was used with a grid spacing of 1 Å. Other GRID input parameters were selected as default values. In this way, for each selected probe molecule, GRID had calculated the

23625 MIF points. One hundred grid points having the most favorable interaction energies from two polar probes were used to color the binding surface. On the other hand, a water probe was utilized to characterize the favorable interaction energies of the Waterphore features. Distance-based criteria were used to find the closest grid point for each feature.

2.4 Comparison of CWFEP with Interaction-based Pharmacophore Modeling

Interaction-based Pharmacophore (IBP) modeling is a well-known approach to identify bound ligand-dependent 3D-pharmacophore features.^[33] Though this approach determines the receptor-based features; feature exhaustiveness can be limited by the availability of only a few bound ligands/inhibitors from among diverse possible scaffolds of potential inhibitors. In order to analyze the comprehensiveness of CWFEP approach and compare its performance with interaction-based pharmacophore modeling, static protein-ligand interaction-based Pharmacophore (IBP) modeling^[20] was also performed for a set of 78 inhibitors bound protein complexes using the Discovery Studio (DS) 2020 software.^[28] This modeling aimed to identify which feature combinations are utilized more frequently in the inhibitor designing process while neglecting the other combinations. All complex structures were processed to add hydrogens and fix ionization states of amino acid residues using the Protein Prepare protocol of DS. The binding site sphere was defined as having a 10 Å radius to accommodate all surrounding residues. IBP module in DS was used to generate minimum of four and maximum ten pharmacophore feature sets with the default settings. The top pharmacophore model from each protein-ligand complex collectively was used for features comparison with CWFEP water pharmacophore model. All selected IBP models from each protein-ligand complex were overlaid with the Waterphore model to understand the distribution of the feature between two feature modeling approaches.

2.5 Generation of the Feature Availability Map of M^{Pro} by Analysis of the Known Inhibitor Dataset

Rigorous model validation is an essential requirement for enhancing the applicability of any pharmacophore model. To test how well Waterphore features can be mapped to the known inhibitors, known inhibitor-based profiling was done to understand feature types present in the bound inhibitor poses. This analysis also helped in identifying the pharmacophore features that were not explored yet among known inhibitors of M^{Pro}. For this feature profiling, 78 crystal bound and 68 assayed compounds extracted from the literature^[5,34–37] were used for features mapping using the Ligand Pharmacophore Mapping module of the DS2020.^[28]

For waterphore features mapping, the crystal-bound pose of ligands was used if available. For assayed compounds, 3D-chemical structures were prepared and up to 250 conformers for each compound were generated using the 'BEST' protocol of DS2020. All ten Waterphore features were used for feature mapping and a ligand with at least five features was analysed. In order to visualize the distribution of the features, a binary string named PharmString consisting of '1' for feature element mapped and '0' for absent/non-available feature, was created. PharmString will clearly depict which feature combinations are consistently utilized and which are the unexplored ones in the design of M^{Pro} inhibitors.

2.6 Assessment of Performance of Waterphore and Interaction Based Pharmacophores

Assessments of the performances of Waterphore as well as selected IBP models were carried out by calculating three different evaluation metrics such as enrichment factor (EF),^[38] which is a key parameter to understand the early enrichment during the virtual screening experiment. Active hit Rate (AHR) was used to report active compound retrieval rate and receiver operating characteristic (ROC) and corresponding area under curve (AUC)^[39] value was employed to depict as a global hit performance over the screen hit collection by calculating the true positive rate (TPR) as a function of false-positive rate (FPR) for different score cutoff. In general, EF reports the enrichment of hits against random screening, however, it is evident that EF depends on the number of Active compounds considered and the total size of the decoy dataset.^[40] Conversely, ROC-AUC is a well-known and commonly employed approach in machine learning to assess model performance.

Two different active compound datasets were collected. The first dataset consisted of 146 SARS-CoV-2 M^{Pro} inhibitors, out of which inhibitor-bound crystal structures were available for 78. The remaining 68 were known inhibitors of SARS-CoV-2 M^{Pro} as per experimental studies reported in the literature. The second dataset consisted of 70 known inhibitors of SARS-CoV-1 M^{Pro}, out of which SARS-CoV-1 M^{Pro} bound crystal structures were available for 30, while 40 were known inhibitors of M^{Pro} as per biochemical studies (Table S4&5). Decoy compounds were fetched from the zinc database using the Decoyfinder program.^[41] A maximum of 20 decoys were selected against each active compound having 20% or lower chemical structure similarity with respective active compounds in terms of Tanimoto score. No two decoys had more than 20% similarity in chemical structures. However, all physicochemical properties of the decoys were similar to the corresponding active compounds (Figure S3). The total number of compounds in the active and decoy dataset were 2778 targeting SARS-CoV-2 M^{Pro}, while SARS-CoV-1 set had 1378 active and decoy compounds. It may be noted that, due to the high degree

of sequence and structural similarities between M^{Pro} proteins from SARS-CoV-1 and SARS-CoV-2, several of the inhibitors of SARS-CoV-1 M^{Pro} have been shown to have inhibitory activity against M^{Pro} of SARS-CoV-2.^[6,35,42] Therefore, we designed the SARS-CoV-1 inhibitor dataset as an additional validation set. For all these compounds, 3D structures were generated using DS2020 software by considering their stereochemistry, ionization and tautomeric states, and a maximum of 250 conformations generation for each compound. The pharmacophore query for the waterphore was searched against these active/decoy compounds to investigate how well the waterphore model can distinguish between actives and decoys. Using the compounds identified by pharmacophore search in the decoy database, EF as well as AHR were calculated.

$$EF_{\text{subset size}} = \frac{\frac{\text{Hits}_{\text{sampled}}}{N_{\text{sampled}}}}{\frac{\text{Hits}_{\text{total}}}{N_{\text{total}}}}$$

$$\text{AHR} = \frac{\text{Total number of Active hits}}{\text{Number of Active in the database}}$$

Where Hits_{sampled} is the number of known active compounds in the hit list, N_{sampled} is the number of compounds in the hit list, Hits_{total} is the number of known actives in the full database and N_{total} is the total size of the database. Ranking of the screening hits was done based on the features fitting score. We have reported the EF at 1%, 5% and 10% of the full decoy library hits. ROC-AUC value has been calculated using the ranked compound hits dataset.

2.7 Chemical Library Selection and Virtual Screening for Novel M^{Pro} Inhibitors

Three different chemical libraries, namely DrugBank,^[43] Protease inhibitors from PubChem^[44] and antiviral library from the Enamine (<https://enamine.net/hit-finding/focused-libraries/antiviral-library>) were selected for identifying novel M^{Pro} hit compounds by screening with the Waterphore model. 'Prepare Ligands' module of the DS2020 software was used to prepare the compound libraries for virtual screening. All compounds were prepared at biological pH with all possible stereoisomers. In order to generate conformers for the prepared compounds, maximum of 250 conformers were generated with 20 kcal/mol energy threshold using the BEST module of the DS 2020. A two-step protocol was followed for efficient chemical library screening. In the first step, Discovery Studio specific Catalyst query was designed to have the core features (A1, D3, A6, A7, H8 and D9) along with binding shape information to identify the initial hit compounds. So, this query will ensure that all retrieved hits will explore the core catalytic region features and will be accommodated in the protease binding site. In

the second step, hits from the previous step were further screened using all Waterphore features to identify the hits having the maximum number of mapped features. In this way, compounds with the core catalytic features were extended to have the other features. All the Catalyst query-based screening has been performed using the 'Ligand Pharmacophore Mapping' module of Discovery Studio. Compound hits were collected from three different libraries and sorted based on the number of features mapped. Selected hits were further processed to narrow down the hit space based on the known inhibitors-based Physico-chemical properties driven threshold.

3 Result and Discussion

3.1 Crystal Water/Fragment Ensemble Pharmacophore (CWFE) Modeling

The major objective of this work was to investigate whether the pharmacophore model derived from the analysis of water and small molecule fragment binding sites in the dataset of 173 apo- and fragment-bound crystal structures of SARS-CoV-2 M^{pro} can identify the inhibitors from the decoys and also predict their bound conformations. The superimposition of 173 apo- and fragment-bound crystal structures on the reference structure (PDB ID: 5R8T) using backbone C^α atoms revealed low RMSD values in the range of 0.07 Å to 0.66 Å, thus indicating a high degree of structural conservation in general and also conservation of the overall geometry for the substrate-binding site (Figure S2). However, detailed analysis of the residue-wise RMSD values for each pair of superimposed structures revealed the presence of loop regions ASL1 (residues 22–53) and ASL2 (residues 184–194)^[45,46] in S4-subsite which exhibits much higher flexibilities compared to other regions in the substrate-binding cavity (Figure S4). The higher flexibility of the S4-subsite is also reflected in the corresponding water cluster, which has a much higher (3.6 Å) cluster radius than other water clusters. These results indicate the ability of the Waterphore model to include the inherent flexibilities present in the structures of receptor proteins.

Around the substrate-binding pocket, the subsites where water molecules bound in at least 27 crystal structures were annotated as conserved water binding sites. A total of nine such conserved water binding sites could be identified by the developed protocol (Table S6). All nine conserved displaceable binding sites were also occupied by the bound fragment molecules (Figure S5). These conserved water binding sites predominantly overlapped with the S1, S2, S4 and S1' subsites of the peptide-binding pocket of M^{pro}.

After implementing the rule-based Pharmacophore annotation protocol, water sites are categorized as hydrogen bond donor (HBD) and hydrogen bond acceptor (HBA).

One water cluster is classified as having a hydrophobic/Ring feature as this water cluster is mainly occupied by a ring containing moieties in the fragment-bound crystal structures of M^{pro} (Figure S5). As can be seen, five HBA, three HBD and one hydrophobic feature are identified. One more feature (hydrophobic) is placed manually at the S2 position as this feature position is occupied by the hydrophobic group of the most fragment molecules. So, a total of ten features were identified (Figure 2). Among these features, inter-features distance ranges from a minimum of 2.6 Å between A6 and D9 to maximum of 16.6 Å between D4 (near S1' subsite) to A10 (S4 subsite).

3.2 Molecular Interaction Fields (MIFs) Calculation

With the help of GRID software, MIF was calculated using the three different probe types, namely neutral amide, carbonyl and water. To decode the favorable interaction energy of each Waterphore feature, the MIF interaction energy of the water probe was examined within the 0.8 Å distance. The interaction energy of the nearest water probe is annotated as the interaction energy of each Waterphore feature (Table S6). This calculation shows that features located at the S1 and S4 positions have the most favorable interaction energies and while S2 position is more favourable for the hydrophobic interaction. This quantification further reflects that incoming ligands with corresponding features will contribute significantly to the binding energy upon displacement of the bound water clusters.

To further characterize the MIF-based binding hotspots in the protease binding site, one hundred MIF points with the most favorable interaction energies were selected for all three probes and shown in Figure S6. Water probe-based MIFs interaction energies range from –11.7 to –6.2 kcal/mol and mainly occupied the S1, S1' and S4 subsites. Catalytic dyad Cys-145 and His-41 are more populated with the water-MIF points. On the other side, polar probe such as amide (N) also follows the same pattern while carbonyl probe is distributed at the catalytic dyad site. Though MIF calculation is based on the single protease structure, theoretical descriptors (different probes-based MIF) agree with consensus water sites.

3.3 Waterphore vs. Ligand Interaction-based Pharmacophore Models

Ligand Interaction-based Pharmacophore (IBP) models depend on the single protein-ligand complex information for perceiving the pharmacophore features. IBP models from the multiple complexes of the particular target proteins will only highlight which features are more frequent and least frequent; however, this approach cannot discriminate which feature is more important than others. In order to compare the water-based hotspot guided

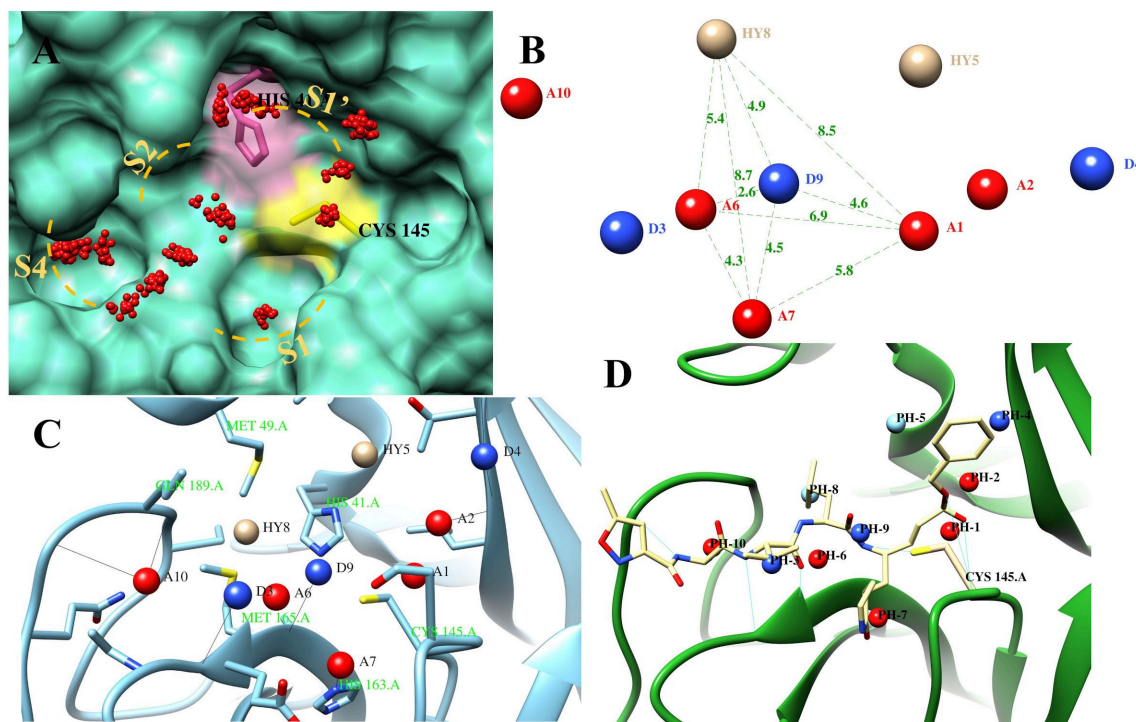


Figure 2. (A) Identified conserved water clusters in the substrate-binding cavity of the M^{pro} (B) Depiction of different pharmacophore features of waterphore and distances between them (C) Binding site residues interacting with Waterphore (D) Waterphore features mapped onto the M^{pro} bound crystal structure of N3 peptide inhibitor.

pharmacophore features (Waterphore features) with receptor-ligand interaction-based pharmacophore features, IBP models were developed for available 78 inhibitor bound crystal structures of SARS-CoV-2 M^{pro} . Of the 78 complexes, only 73 have features ranging from four to ten (Table S3), while the remaining five had less than four features. IBP models derived from peptide bound complexes N3, Telaprevir, TG-0203770, GRL-2420, and Boceprevir bound M^{pro} complexes have perceived maximum of ten features. Among all the available features for these complexes, only five features have made maximum consensus across the five IBP models (Figure S7). Figure 3 shows a comparison between waterphore and Interaction-based Pharmacophores (IBP) derived from bound ligand crystal structures. To decipher which feature is dominantly available among the IBP models and how well they overlaid with Waterphore features, features of IBP models were superimposed over the later one (Figure 3B). As can be seen, core features A1, D3, A6, H8, D9 are commonly occurring in both types of pharmacophore models, while A2, D4, R5, and A10 are among the features which are either absent or least frequent in the IBP models. This suggests the use of the Waterphore model has advantages over IBP in the case of M^{pro} pharmacophore searches. Hence, usage of Waterphore can potentially help in identifying novel inhibitors.

3.4 Search for Waterphore Features in Known M^{pro} Inhibitors

To highlight the availability of the CWFEF pharmacophore features among the known M^{pro} inhibitors, inhibitor-bound crystal structures of M^{pro} were used. Profiling of donor, acceptor and hydrophobic groups in known inhibitors was carried out to understand the feature types and size present in various inhibitor poses. The set of 78 crystal structures of the inhibitor bound with SARS-CoV-2 M^{pro} and 68 assayed compounds were used for feature mapping. As shown in Figure 3C, pharmacophore feature numbers A1 and D3, A6, A7, H8, and D9 are very consistent and form the core group occupying the S1, S1' and S2 subsites features while A2 and D4 are not covered at all. A10 feature at the subsite S4 derived from the most populated water cluster site and has the most favorable interaction energies were found to be least utilized among SARS-CoV-2 M^{pro} inhibitors. Thus, feature mapping onto the known inhibitors suggests designing of improved non-peptide inhibitors by employing pharmacophore features absent in known inhibitors.

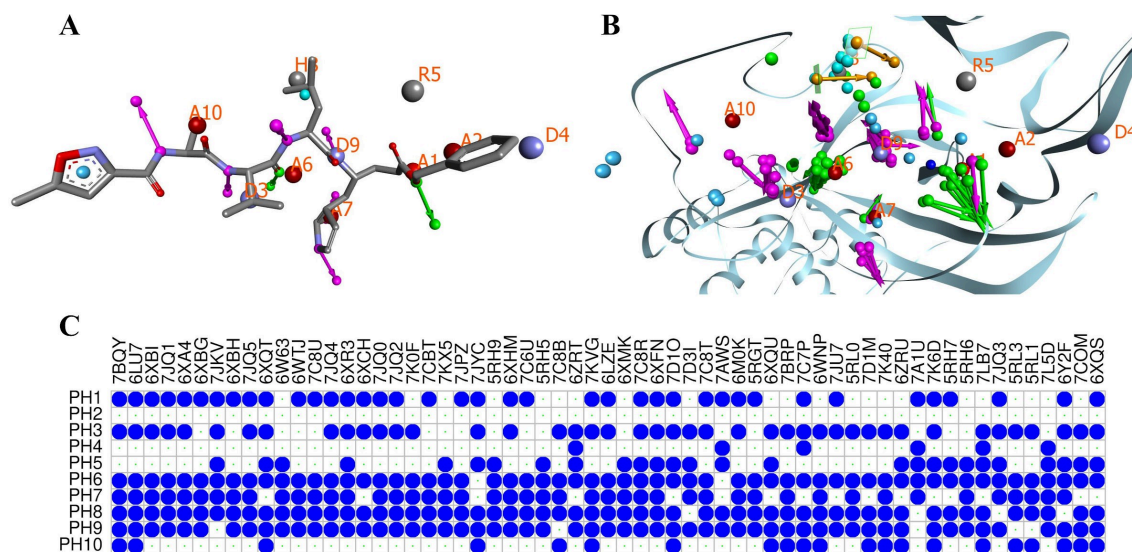


Figure 3. A) Both Waterphore and N3-M^{Pro} complex-based pharmacophore models have generated ten feature elements. Figure A) shows the 3D distribution of the Waterphore and N3 features overlaid with N3 peptide inhibitor in the protease binding cavity. Feature A1 and A2 are located near the Cys-145 which is catalytically essential and known to form the covalent bond with the inhibitor having the aldehyde or keto warhead. B) Depiction of the feature comparison between the CWFEP vs. static IBP models. 62 out of 78 complexes have generated the IBS models. All IBS features are overlaid to the CWFEP features to empathize with the comparative difference. CWFEP feature number A1, D3, A6, D7, D8 and D9 are mostly explored in the inhibitor designing however, A2, D4, R5, A10 features are either not utilized or least utilized. Further CWFEP features are used to profile the known inhibitors from the CoV-2 and figure C) shows that 1, 6, 7, 8 and 9 features as determined in figure B), are the most explored features.

3.5 Enrichment and ROC-AUC Calculation for Assessment of the Utility of Waterphore in Virtual Screening

Before carrying out the actual virtual screening of compound libraries using Waterphore, an assessment was done to evaluate the performance of the Waterphore model developed in the current study. Because of the sequence homology and structural similarity between M^{Pro} of SARS-CoV-2 and SARS-CoV-1, several inhibitors of SARS-CoV-1 M^{Pro} have also been shown to have activity against M^{Pro} of SARS-CoV-2. Therefore, we used known inhibitors of both SARS-CoV-2 and SARS-CoV-1 M^{Pro} to benchmark the predictive power of the Waterphore model. As mentioned in the methods section, two active compound datasets, SARS-CoV-2 (146 compounds) and SARS-CoV-1 (70 compounds) were combined with decoy set consisting of maximum of 20 decoy compounds per inhibitor to construct a decoy library of 2778 and 1387 respectively. The developed decoy library was further used for enrichment factor (EF), Active hit Rate (AHR) and ROC-AUC analysis. Both libraries were screened using the Waterphore model and five IBP models that encompassed all ten pharmacophore features. These IBP models corresponded to pharmacophores derived from crystal structures of M^{Pro} bound to N3, Telaprevir, TG-0203770, GRL-2420 and Boceprevir. All hits were collected having a minimum any five features mapped out of ten features. FitValue score, which depends on the number of features mapped and feature fitting

tolerance,^[47] calculated by Discovery Studio, was used as a criterion to sort the hits for all metrics calculation.

As can be seen from Figure 4, for the SARS-CoV-2 decoy library, the Waterphore model has achieved more than 19-, 11-, 11-times enrichment at EF1, EF5 and EF10, respectively with an Active hit rate (AHR) of around 70%. These results unambiguously demonstrate that the waterphore model predictions are highly enriched by actives over decoys. Compared to EF at different subsets, AUC represents the overall performance at different score cutoffs for predicting actives and ROC curve shows sensitivity or true positive rate (TPR) as a function of false-positive rate (FPR). As can be seen from Figure 4C waterphore model offers an AUC value of 0.81 and at optimum score cutoff, this model can predict active compounds with a sensitivity or TPR of above 77% at FPR of close to 20%. In contrast to the Waterphore model, the IBP models achieved lesser enrichment, AHR and ROC-AUC values (Figure 4), thus indicating the superior performance of Waterphore over IBP.

For SARS-CoV-1 dataset, the waterphore model also has EF1, EF5 and EF10 values 19, 10 and 8, respectively with AHR of 61% and ROC-AUC of 0.75 (Figure S8). This suggests that the waterphore model can efficiently identify potential inhibitors of SARS-CoV-2 M^{Pro} from among known inhibitors of SARS-CoV-1. These results are incredibly encouraging because the waterphore model has neither used any information from inhibitor bound crystal structures of M^{Pro} from SARS-CoV-1/2 nor information about known inhibitors

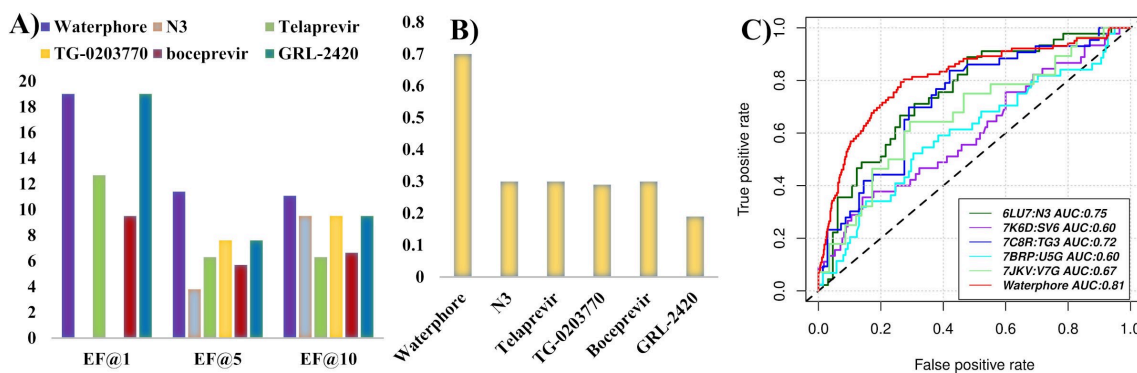


Figure 4. Comparison of the Enrichment Factors (EF) (A), Active Hit Rates (AHR) (B) and ROC curves (C) for the Waterphore (in dark red color) and IBP models on the dataset consisting of 146 known SARS-CoV-2 inhibitors and 2920 decoys. Among the top five IBP models, N3-peptide containing M^{pro} complex (in dark green color) has the maximum AUC value of 0.75, while SV6 (Telaprevir) (purple color) and U5G (Boceprevir) (cyan color) received the minimum AUC value of 0.60. The ROC curves for V7G (GRL-2420) and TG3 (TG-0203770) are shown in light green and blue colors respectively.

of M^{pro}. The Waterphore model derives its predictive power by utilizing water and fragment binding sites in substrate binding region obtained from XChem crystallographic fragment screen. The data on inhibitor-bound crystal structures of M^{pro} and other known inhibitors of M^{pro} serve as completely independent test sets. Therefore, it is encouraging to note that EF, AHR and ROC-AUC clearly demonstrate high prediction accuracy of Waterphore model on independent test set of compounds. Thus, the waterphore model can be used to screen large compound libraries quickly to identify novel inhibitors or repurposed molecules from among known drugs.

3.6 Virtual Screening and Hit Mining: Finding Compounds Having Unexplored Waterphore Features

In silico high throughput screening against a particular target protein is a powerful approach to identify the novel inhibitors encompassing the required pharmacophore features. In the present study, Waterphore features were employed in two steps to identify novel inhibitors of M^{pro} protein from DrugBank, Protease inhibitor library and antiviral compound library. As discussed in the methods section, the first round of compound hits was collected from three different chemical libraries using the Catalyst query containing the core pharmacophore features coupled with the cavity shape information. Every ligand conformation was fitted to the pharmacophore features and sorted based on the calculated highest fit values. The pharmacophore-based screening reduces the chemical database to a large extent by implementing the required number of features and inter-feature distances criteria (Figure S10) and ensuring that compounds carry the right features at the appropriate position. In the second step, hits from the previous step were further screened to get the final list of potential M^{pro} inhibitors (Table S7). It is encouraging to note

that 68 out of 109 hits from DrugBank corresponded to approved drugs. Mapped features range from 5 to 7 for DrugBank, 6 to 8 for proteases, and 6 for the antiviral compound library. A list of 230 selected compounds from the waterphore based search of the PubChem Protease inhibitor library are provided in Table S8. These results demonstrate that the Waterphore approach developed in this work can be a powerful tool for drug repurposing studies.

Waterphore based virtual screening hits identified from the DrugBank consisted of ten protease inhibitors which had been developed as anti-HIV and anti-HCV agents. A Survey of the literature revealed that a set of 10 compounds of the identified protease inhibitors had been experimentally tested recently for their inhibitory activity against the SARS-CoV-2 (Table 1). As shown in Table 1, many of these protease inhibitors have EC₅₀ values in the micromolar range. These experimental studies provide indirect support that the Waterphore model can identify high-affinity inhibitors. We also performed the virtual screening of the DrugBank chemical library using the N3-peptide-protein complex (PDB-ID: 6LU7) driven IBP model and 22 hits were identified from the approved drug category out of which 12 hits belong to the antibiotics class of compounds and only one protease inhibitor Remikiren was identified. Hits comparison from the two models suggests that Waterphore features represent the key interaction elements that enhance its ability to successfully identify repurposed drugs or new inhibitors for SARS-CoV-2 M^{pro} protein. Along with hit analysis, we have also compared the screened hit poses of Boceprevir and Telaprevir with the corresponding poses in the known crystal structures and find good agreement in terms of orientation and conformation (Figure S9).

In order to analyze in detail the types of chemical scaffolds identified by waterphore model based on search in DrugBank and Protease libraries. A total of 298 hit compounds obtained from waterphore search in DrugBank

Table 1. Protease inhibitor drugs from the DrugBank which are predicted by Waterphore model to be potential inhibitors of SARS-CoV-2. The experimental activity of the identified compounds against SARS-CoV-2 in cell line-based assays or activity against M^{Pro} based on a literature survey is indicated. Results of cell line-based assay indicated as EC₅₀ and biochemical inhibition study against M^{Pro} indicated as IC₅₀ values.

DrugBank ID	Drug Name	MW	Number of features mapped	Experimental Activity*
DB00220	Nelfinavir	567.31	6	EC ₅₀ : 1.13 μM
DB01601	Lopinavir	628.36	5	EC ₅₀ : 5.73 μM IC ₅₀ : > 20 μM
DB01232	Saquinavir	670.38	6	EC ₅₀ : 8.83 μM IC ₅₀ : 31.4 μM
DB00932	Tipranavir	602.20	5	EC ₅₀ : 13.34 μM IC ₅₀ : 27.66 μM
DB00701	Amprenavir	505.22	5	EC ₅₀ : 31.32 μM IC ₅₀ : > 20 μM
DB01264	Darunavir	547.23	6	EC ₅₀ : 46.41 μM IC ₅₀ : 36.1 μM EC ₅₀ : 59.14 μM
DB00224	Indinavir	613.36	6	IC ₅₀ : > 200 μM
DB08873	Boceprevir	519.34	5	IC ₅₀ : 4.13 μM IC ₅₀ : 2.5 μM
	Remikiren	630.34	6	NA
DB00212	Telaprevir	679.40	6	IC ₅₀ : 10.7 μM
DB05521				

*Table S9 provides information on publications reporting the experimental EC₅₀/IC₅₀ values.

and PubChem protease library were clustered. A topological fingerprint was calculated and used for Butina algorithm-based clustering using the RDKit package.^[30] At 80% similarity cutoff, chemical clusters were identified in which the largest cluster had the ten compounds while 170 singleton clusters occurred (Figure S11). All the S-heterocyclic compounds were occurred together in the first cluster, while second and third clusters consisted of ethyl sulfamoyl derivatives and peptide derivatives respectively. Representative compounds from top twelve highly populated clusters have been shown in Figure S11 and they provide a snapshot of the diversity in chemical scaffolds identified by waterphore model. In order to evaluate the *in silico* binding affinity of the identified compounds for M^{Pro} representative compounds having the maximum number of mapped features from each cluster were further minimized in complex with the receptor using the *in situ* Ligand Minimization module of Discovery Studio followed by binding energy calculation of each minimized pose using CHARMM-like force field with Generalized Born implicit solvent model.^[48] So, in this way *in silico* binding affinities of the pharmacophore-screening based poses were evaluated and eleven novel M^{Pro} inhibitors compounds were selected from different compound clusters having known tested bioactivity against protease class of proteins (Figure 5). The known biological activity of these selected compounds are shown in Table S10. It is encouraging to note that they consist of inhibitors of HIV protease, thermolysin, renin, matrix metalloproteinase MMP9 and bone morphogenetic protein BMP1 and have low nM activity. These compounds

are potential novel M^{Pro} inhibitors predicted by our Waterphore model and are good candidates for experimental validation.

4 Conclusions

Even though large number of *in silico* prediction studies have reported inhibitors for SARS-CoV-2 M^{Pro} using standard drug discovery tools or simulation methods, information is available on large number of M^{Pro} crystal structures about water and drug-like fragment binding sites have not been effectively used in most of the *in-silico* screening studies. Since water molecules play a crucial role in the catalysis and substrate recognition of several enzymes, earlier studies have used information about bound water molecules in lead optimization of ligand-receptor systems.^[49] In this work, we have utilized the recently available crystal structures of large number of electrophilic fragment bound of M^{Pro} from PanDDA and XChem platform^[50] to systematically analyze the water and fragment binding sites in the active site pocket region. Using this information about conserved water binding sites and fragment binding sites, we have identified binding hotspots. The binding hotspots have been further utilized to develop a novel Crystal Water/Fragment Ensemble-based Pharmacophore (CWFE) or Water pharmacophore (Waterphore) modeling protocol, which can quickly screen large compound libraries to identify novel inhibitors for M^{Pro}. Waterphore approach only utilizes the apo-protein structure and bound water and

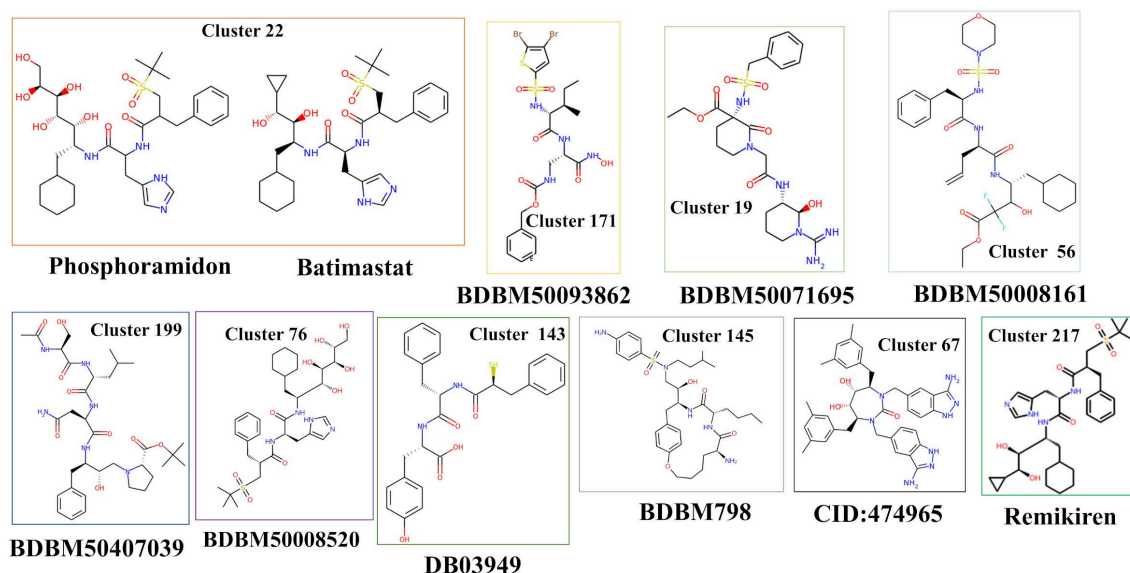


Figure 5. Eleven selected compounds having the maximum binding free energy and mapped features.

fragment molecule information to derive displaceable water molecules driven by pharmacophore features. Many of these pharmacophore features map onto the subsites (S1, S1' and S4) in the active site pocket of M^{pro} , which have the most favorable interaction energies could potentially be explored for the design of novel inhibitors. Waterphore model also implicitly represents conformational flexibilities of certain regions of binding site in terms of the higher radius of some of the pharmacophore features. Benchmarking analysis of the predictive power of Waterphore on a dataset of 146 known inhibitors in terms of enrichment factor (EF), hit rate (HR) and ROC-AUC reveal that it can successfully identify known inhibitors from among decoy compounds. Our benchmarking study also demonstrates the superior performance of Waterphore compared to interaction-based pharmacophores derived from ligand-bound crystal structures. This result is especially encouraging because, even though no known inhibitor information has been utilized in deriving the waterphore features, the ROC-AUC values are higher than some of the machine learning models^[51] which have been trained using known inhibitors data. Comparison of Waterphore features with crystal structure bound known inhibitors of SARS-CoV-1 and SARS-CoV-2 indicate that not all high scoring features of Waterphores have been explored in the design of known inhibitors of M^{pro} . Virtual screening of DrugBank, protease inhibitor and antiviral compound libraries using Waterphore reveal that its performance is superior to Interaction-based pharmacophore (IBP) models derived from inhibitor bound crystal structures of M^{pro} . Results from virtual screening by Waterphore demonstrate that, it can not only identify known drugs for repurposing against M^{pro} , but also aid in experimental studies by predicting diverse lead compounds which could be tested for their biological activity. Water-

phore approach developed in this work is general enough for its potential application to other drug targets. However, it is necessary that sufficient number of water and fragment-bound crystal structures should be available for the corresponding target. Therefore, there will be limitations in its application for less well-studied drug targets.

5 Funding

This work was supported by the Department of Biotechnology, Government of India grant to National Institute of Immunology, New Delhi. DM also acknowledges financial support from Department of Biotechnology (DBT), India under BTIS (BT/BI/03/009/2002 & BT/PR40325/BTIS/137/1/2020) and COE (BT/COE/34/SP15138/2015) projects and from National Supercomputing Mission, MeITY, India under NSM-NPGDD project (MeitY/R&D/HPC/2(1)/2014/CORP: DG:3191).

Acknowledgment

The authors would like to thank Dassault Systemes BIOVIA for providing complete BIOVIA Discovery Studio Academic Research Suite to carry out the SARS-CoV-2 related research. Authors would like to acknowledge Dr. Manisha Goyal, Delhi University for providing computational resources for part of this work. Authors would also like to thank Prof. Indira Ghosh for valuable inputs and critical comments on this work.

Conflict of Interest

None declared.

Data Availability Statement

No new data generated. The manuscript reports computational analysis of data available in public domain and relevant identifies and other details have been provided in supplementary information section.

References

- [1] P. V'Kovski, A. Kratzel, S. Steiner, H. Stalder, V. Thiel, *Nat. Rev. Microbiol.* **2021**, *19*, 155–170.
- [2] A. K. Ghosh, M. Brindisi, D. Shahabi, M. E. Chapman, A. D. Mesecar, *ChemMedChem* **2020**, *15*, 907–932.
- [3] B. Boras, R. M. Jones, B. J. Anson, D. Arenson, L. Aschenbrenner, M. A. Bakowski, N. Beutler, J. Binder, E. Chen, H. Eng, J. Hammond, R. Hoffman, E. P. Kadar, R. Kania, E. Kimoto, M. G. Kirkpatrick, L. Lanyon, E. K. Lendy, J. R. Lillis, S. A. Luthra, C. Ma, S. Noell, R. S. Obach, M. N. O'Brien, R. O'Connor, K. Ogilvie, D. Owen, M. Pettersson, M. R. Reese, T. F. Rogers, M. I. Rossulek, J. G. Sathish, C. Steppan, M. Ticehurst, L. W. Updyke, Y. Zhu, J. Wang, A. K. Chatterjee, A. D. Mesecar, A. S. Anderson, C. Allerton, *bioRxiv* **2020**, *12.03.409441*, DOI: <https://doi.org/10.1101/2020.12.03.409441>.
- [4] B. Cao, Y. Wang, D. Wen, W. Liu, J. Wang, G. Fan, L. Ruan, B. Song, Y. Cai, M. Wei, X. Li, J. Xia, N. Chen, J. Xiang, T. Yu, T. Bai, X. Xie, L. Zhang, C. Li, Y. Yuan, H. Chen, H. Li, H. Huang, S. Tu, F. Gong, Y. Liu, Y. Wei, C. Dong, F. Zhou, X. Gu, J. Xu, Z. Liu, Y. Zhang, H. Li, L. Shang, K. Wang, K. Li, X. Zhou, X. Dong, Z. Qu, S. Lu, X. Hu, S. Ruan, S. Luo, J. Wu, L. Peng, F. Cheng, L. Pan, J. Zou, C. Jia, J. Wang, X. Liu, S. Wang, X. Wu, Q. Ge, J. He, H. Zhan, F. Qiu, L. Guo, C. Huang, T. Jaki, F. G. Hayden, P. W. Horby, D. Zhang, C. Wang, *N. Engl. J. Med.* **2020**, *382*, 1787–1799.
- [5] C. Ma, M. D. Sacco, B. Hurst, J. A. Townsend, Y. Hu, T. Szeto, X. Zhang, B. Tarbet, M. T. Marty, Y. Chen, J. Wang, *Cell Res.* **2020**, *30*, 678–692.
- [6] B. J. Anson, M. E. Chapman, E. K. Lendy, S. Pshenychnyi, R. T. D'Aquila, K. J. F. Satchell, A. D. Mesecar, *PREPRINT (Version 1) 01 May 2020*, available at *Research Square*.
- [7] A. Sharma, S. Goyal, A. K. Yadav, P. Kumar, L. Gupta, *J. Biomol. Struct. Dyn.* **2020**, 1–15.
- [8] S. T. Ngo, N. Quynh Anh Pham, L. Thi Le, D. H. Pham, V. V. Vu, *J. Chem Inf Model* **2020**, *60*(12), 5771–5780.
- [9] S. Kumar, P. P. Sharma, U. Shankar, D. Kumar, S. K. Joshi, L. Pena, R. Durvasula, A. Kumar, P. Kempaiah, Poonam, B. Rath, *J. Chem Inf Model* **2020**, *60*(12), 5754–5770.
- [10] S. J. Teague, *Nat. Rev. Drug Discovery* **2003**, *2*, 527–541.
- [11] N. Brooijmans, I. D. Kuntz, *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335–373.
- [12] P. J. Hajduk, J. R. Huth, C. Tse, *Drug Discovery Today* **2005**, *10*, 1675–1682.
- [13] P. J. Goodford, *J. Med. Chem.* **1985**, *28*, 849–857.
- [14] D. H. Hall, L. E. Grove, C. Yueh, C. H. Ngan, D. Kozakov, S. Vajda, *J. Am. Chem. Soc.* **2011**, *133*, 20668–20671.
- [15] M. L. Verdonk, J. C. Cole, R. Taylor, *J. Mol. Biol.* **1999**, *289*, 1093–1108.
- [16] R. Baron, P. Setny, J. A. McCammon, *J. Am. Chem. Soc.* **2010**, *132*, 12091–12097.
- [17] F. Spyrikis, M. H. Ahmed, A. S. Bayden, P. Cozzini, A. Mozzarelli, G. E. Kellogg, *J. Med. Chem.* **2017**, *60*, 6781–6827.
- [18] M. Maurer, C. Oostenbrink, *J. Mol. Recognit.* **2019**, *32*, e2810.
- [19] J. E. Ladbury, *Chem. Biol.* **1996**, *3*, 973–980.
- [20] M. P. A. Sanders, R. McGuire, L. Roumen, I. J. P. de Esch, J. de Vlieg, J. P. G. Klomp, C. de Graaf, *MedChemComm* **2012**, *3*, 28–38.
- [21] D. Schaller, D. Šribar, T. Noonan, L. Deng, T. N. Nguyen, S. Pach, D. Machalz, M. Bermudez, G. Wolber, *WIREs Comput. Mol. Sci.* **2020**, *10*, e1468.
- [22] P. Kumar, R. Kaalia, A. Srinivasan, I. Ghosh, *SAR QSAR Environ. Res.* **2018**, *29*, 1–19.
- [23] W. Yu, S. K. Lakkaraju, E. P. Raman, L. Fang, A. D. MacKerell Jr., *J. Chem. Inf. Model.* **2015**, *55*, 407–420.
- [24] A. D. MacKerell Jr., S. Jo, S. K. Lakkaraju, C. Lind, W. Yu, *Biochim. Biophys. Acta Gen. Subj.* **2020**, *1864*, 129519.
- [25] H. A. Carlson, *Curr. Pharm. Des.* **2002**, *8*, 1571–1578.
- [26] P. Ghanakota, H. A. Carlson, *J. Med. Chem.* **2016**, *59*, 10383–10399.
- [27] A. Douangamath, D. Fearon, P. Gehrtz, T. Krojer, P. Lukacik, C. D. Owen, E. Resnick, C. Strain-Damerell, A. Aimon, P. Abranyi-Balogh, J. Brandao-Neto, A. Carbery, G. Davison, A. Dias, T. D. Downes, L. Dunnett, M. Fairhead, J. D. Firth, S. P. Jones, A. Keeley, G. M. Keseru, H. F. Klein, M. P. Martin, M. E. M. Noble, P. O'Brien, A. Powell, R. N. Reddi, R. Skyner, M. Snee, M. J. Waring, C. Wild, N. London, F. von Delft, M. A. Walsh, *Nat. Commun.* **2020**, *11*, 5047.
- [28] D. S. Biovia, *BIOVIA Discovery Studio*, San Diego: Dassault Systèmes, **2020**.
- [29] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- [30] G. Landrum, *RDKit: Open-Source Cheminformatics*, <https://www.rdkit.org>.
- [31] P. W. Sprague, *Perspect. Drug Discovery Des.* **1995**, *3*, 1–20.
- [32] P. Goodford, in *Molecular Interaction Fields*, **2005**, pp. 1–25.
- [33] S. Y. Yang, *Drug Discovery Today* **2010**, *15*, 444–450.
- [34] R. Banerjee, L. Perera, L. M. V. Tillekeratne, *Drug Discov Today* **2020**, *26*(3), 804–816.
- [35] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, *Science* **2020**, *368*, 409–412.
- [36] W. Dai, B. Zhang, X. M. Jiang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jin, J. Peng, F. Liu, C. Li, Y. Li, F. Bai, H. Wang, X. Cheng, X. Cen, S. Hu, X. Yang, J. Wang, X. Liu, G. Xiao, H. Jiang, Z. Rao, L. K. Zhang, Y. Xu, H. Yang, H. Liu, *Science* **2020**, *368*, 1331–1335.
- [37] M. M. Ghahremanpour, J. Tirado-Rives, M. Deshmukh, J. A. Ippolito, C. H. Zhang, I. Cabeza de Vaca, M. E. Liosi, K. S. Anderson, W. L. Jorgensen, *ACS Med. Chem. Lett.* **2020**, *11*, 2526–2533.
- [38] H. Chen, P. D. Lyne, F. Giordanetto, T. Lovell, J. Li, *J. Chem. Inf. Model.* **2006**, *46*, 401–415.
- [39] A. N. Jain, A. Nicholls, *J. Comput.-Aided Mol. Des.* **2008**, *22*, 133–139.
- [40] M. Hernandez, G. Liang Gan, K. Linvill, C. Dukatz, J. Feng, G. Bhisetti, *J. Chem. Inf. Model.* **2019**, *59*, 4475–4485.
- [41] A. Cereto-Massague, L. Guasch, C. Valls, M. Mulero, G. Pujadas, S. Garcia-Vallve, *Bioinformatics* **2012**, *28*, 1661–1662.
- [42] B. Boras, R. M. Jones, B. J. Anson, D. Arenson, L. Aschenbrenner, M. A. Bakowski, N. Beutler, J. Binder, E. Chen, H. Eng, H. Hammond, J. Hammond, R. E. Haupt, R. Hoffman, E. P. Kadar, R.

- Kania, E. Kimoto, M. G. Kirkpatrick, L. Lanyon, E. K. Lendy, J. R. Lillis, J. Logue, S. A. Luthra, C. Ma, S. W. Mason, M. E. McGrath, S. Noell, R. S. Obach, M. N. O'Brien, R. O'Connor, K. Ogilvie, D. Owen, M. Pettersson, M. R. Reese, T. F. Rogers, M. I. Rossulek, J. G. Sathish, N. Shirai, C. Steppan, M. Ticehurst, L. W. Updyke, S. Weston, Y. Zhu, J. Wang, A. K. Chatterjee, A. D. Mesezar, M. B. Frieman, A. S. Anderson, C. Allerton, *bioRxiv* **2021**, 2020.2009.2012.293498.
- [43] D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, M. Wilson, *Nucleic Acids Res.* **2018**, *46*, D1074–D1082.
- [44] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, E. E. Bolton, *Nucleic Acids Res.* **2019**, *47*, D1102–D1109.
- [45] M. Bzowka, K. Mitusinska, A. Raczynska, A. Samol, J. A. Tuszynski, A. Gora, *Int. J. Mol. Sci.* **2020**, *21*, 3099.
- [46] J. Gossen, S. Albani, A. Hanke, B. P. Joseph, C. Bergh, M. Kuzikov, E. Costanzi, C. Manelfi, P. Storici, P. Gribbon, A. R. Beccari, C. Talarico, F. Spyraakis, E. Lindahl, A. Zaliani, P. Carloni, R. C. Wade, F. Musiani, D. B. Kokh, G. Rossetti, *ACS Pharmacol. Transl. Sci.* **2021**, *4*, 1079–1095.
- [47] O. Guner, O. Clement, Y. Kurogi, *Curr. Med. Chem.* **2004**, *11*, 2991–3005.
- [48] M. S. Lee, F. R. Salsbury, C. L. Brooks, *J. Chem. Phys.* **2002**, *116*, 10606–10614.
- [49] D. Chen, Y. Li, M. Zhao, W. Tan, X. Li, T. Savidge, W. Guo, X. Fan, *Phys. Chem. Chem. Phys.* **2018**, *20*, 24399–24407.
- [50] D. A. Erlanson, *Nat. Commun.* **2020**, *11*, 5048.
- [51] S. Mohapatra, P. Nath, M. Chatterjee, N. Das, D. Kalita, P. Roy, S. Satapathi, *PLoS One* **2020**, *15*, e0241543.

Received: June 28, 2021

Accepted: September 20, 2021

Published online on October 11, 2021