

Development of Broad-Spectrum Halomethyl Ketone Inhibitors Against Coronavirus Main Protease 3CL^{pro}

Usman Bacha¹, Jennifer Barrila¹, Sandra B. Gabelli², Yoshiaki Kiso³, L. Mario Amzel² and Ernesto Freire^{1,*}

¹Department of Biology, The Johns Hopkins University, Baltimore, MD 21218, USA

²Department of Biophysics and Biophysical Chemistry, The Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

³Department of Medicinal Chemistry, Center for Frontier Research in Medicinal Science, Kyoto Pharmaceutical University, Yamashina-ku, Kyoto 607-8412, Japan

*Corresponding author: Ernesto Freire, ef@jhu.edu

Coronaviruses comprise a large group of RNA viruses with diverse host specificity. The emergence of highly pathogenic strains like the SARS coronavirus (SARS-CoV), and the discovery of two new coronaviruses, NL-63 and HKU1, corroborates the high rate of mutation and recombination that have enabled them to cross species barriers and infect novel hosts. For that reason, the development of broad-spectrum antivirals that are effective against several members of this family is highly desirable. This goal can be accomplished by designing inhibitors against a target, such as the main protease 3CL^{pro} (M^{pro}), which is highly conserved among all coronaviruses. Here 3CL^{pro} derived from the SARS-CoV was used as the primary target to identify a new class of inhibitors containing a halomethyl ketone warhead. The compounds are highly potent against SARS 3CL^{pro} with K_i 's as low as 300 nM. The crystal structure of the complex of one of the compounds with 3CL^{pro} indicates that this inhibitor forms a thioether linkage between the halomethyl carbon of the warhead and the catalytic Cys 145. Furthermore, Structure Activity Relationship (SAR) studies of these compounds have led to the identification of a pharmacophore that accurately defines the essential molecular features required for the high affinity.

Abbreviations: CoV, Coronavirus; SARS, severe acute respiratory syndrome; HCoV, human coronavirus; MOE, molecular operating environment; RTI, respiratory tract infection; NL-63, human coronavirus Netherlands 63; HUK1, human coronavirus HKU1; TCEP, Tris[2-carboxyethyl] phosphine; EDTA, ethylenedinitrilotetraacetic acid.

Key words: biophysical chemistry, calorimetric techniques, drug design, drug discovery, protease and ligands (substrate/inhibitor)

Received 14 February 2008, revised and accepted for publication 27 May 2008

Coronaviruses (CoVs) are responsible for more than 30% of all respiratory tract infections (RTIs), affecting both the upper and lower respiratory tract (1)^a. Coronaviruses were previously thought to cause only benign respiratory infections with infection rates peaking during the winter months. The emergence of the most virulent member of this family, SARS-CoV, was a harsh deviation from this belief (2–4). The discovery of two new species of CoV that infect humans, NL-63 and HCoV-HKU1 in 2004 and 2005, respectively, confirm the high rate of mutagenesis and genetic recombination within Coronaviridae (5,6). Evolutionary insights gained from sequence comparisons of different strains of SARS-CoV have shown that the virus has a mutation rate of 8.26×10^6 /nt/day (about one third that of HIV) (7). This high rate of mutation often leads to host swapping or to the generation of novel CoVs, posing a significant challenge to the development of broad-spectrum inhibitors (8).

An appropriate target for the development of broad-spectrum anti-coronavirals should be both indispensable to the viral life cycle and conserved among CoVs. The main viral protease 3CL^{pro} plays a key role in viral transcription and propagation of progeny virions (3,9–12). Inhibition of this enzyme with a general cysteine protease inhibitor has been shown to inactivate viral replication in mouse hepatitis virus (MHV) and human coronavirus 229E (HCoV-229E) infected cells (13,14). 3CL^{pro} is a three-domain cysteine protease, which predominantly occurs as a dimer in solution. Previous studies have concluded that dimerization is required for the activity of the protease (15–18). The catalytic residue Cys 145 and His 41 are located in a cleft between the first two domains (Figure 1) that comprise a highly conserved active site cavity. Although the overall sequence identities of 3CL^{pro} among members of the coronaviral family is only 40–50%, the three-dimensional structure of the proteases are very similar (Figure 1) (19). Sequence conservation is more pronounced in certain regions of 3CL^{pro}. One such region identified is a cluster of serine residues consisting of Ser 139, Ser144 and Ser 147, adjacent to the active site of SARS 3CL^{pro} (20). Subsequent alanine mutagenesis experiments on these serine residues, performed in our laboratory, indicated that this cluster plays a major role in not only the activity of the protease but also in its ability to dimerize (16).



Figure 1: The aligned structures of SARS 3CL^{pro} (PDB ID 1Q2W), HCoV 229E 3CL^{pro} (PDB ID 1P9S), IBV 3CL^{pro} (PDB ID 2Q6D), and TGEV 3CL^{pro} (PDB ID 1LVO) are shown in ribbon representation and coloured blue, yellow, red and green respectively. An arrow points to the position of the active site cavity. The C α of each structure was used in the alignment and all structures were aligned to the SARS 3CL^{pro} structure. The RMSD values from the alignment were 0.889 Å for HCoV 229E 3CL^{pro} (over 247 C α), 1.302 Å for IBV 3CL^{pro} (over 247 C α), and 1.8 Å for TGEV 3CL^{pro} (over 272 C α).

Another region with a high degree of conservation consists of the residues determining substrate specificity in the S1, S2 and S4 subsites of the active site cavity of 3CL^{pro} [(19) and unpublished work from our laboratory]. Although the catalytic Cys and His residues are absolutely conserved in the main proteases of all CoVs, the high similarity of these sites in 3CL^{pro} among members of the three coronaviral families is also well established (3,19,21). Substrate specificity studies have shown high conservation in the residue preference at the corresponding site of the substrate as well (22). Moreover, substrate sequences derived from the N-terminal autocleavage site of SARS 3CL^{pro} have been shown to be processed with equal efficiency by proteases from other coronaviruses when compared to SARS-CoV (19). Therefore, the design of broad-spectrum inhibitors against coronaviral main proteases based on substrate mimetics appears to be a feasible strategy for drug development.

Recently, substrate mimetics with a trifluoromethyl ketone warhead that specifically targeted the active site of SARS 3CL^{pro} were identified (23). In an attempt to optimize those leads, the affinity of a library of halomethyl ketone compounds with various P1 substitutions towards 3CL^{pro} were evaluated. Here we show the results of the characterization of the five best compounds from this screen. The data presented include the determination of the binding mechanism, the thermodynamic dissection of the process and the selectivity against a panel of other proteases. Our results indicate that a 1000-fold improvement in affinity towards 3CL^{pro} can be achieved by modifying the halogenation of this warhead and the substitution at the P1 position, and by reducing the compound size. Of particular interest is Compound **4**, which inhibits the protease by forming an initial reversible complex followed by a much slower irreversible reaction between Cys 145 and the adjacent halomethyl resulting in a thioether linkage. The crystallographic structure of 3CL^{pro} in complex with Compound **4**, indicates novel P n -S n interactions. An accurate pharmacophore model has been derived from the affinity (K_i) profiles of the compounds studied in this work. Experimental validation of the predictability of this model performed using a commercially available compound library indicates that the pharmacophore has an effectiveness of 95% in selecting molecules with activity better than 1000 μ M against SARS 3CL^{pro}.

Materials and Methods

Protein purification

Recombinant SARS 3CL^{pro} was expressed as a soluble fraction in BL21 Star DE3 *E. coli* competent cells (Invitrogen, Carlsbad, CA, USA). The construct begins with residue Ser1, and therefore does not contain the full N-terminal auto-cleavage site of the protein. Cells were grown in LB supplemented with ampicillin (50 μ g/mL) at 37 °C, induced with 1 mM IPTG when the optical density (as determined by absorbance at 600 nm) was 0.8 or greater, and harvested after 4 h. Cells were re-suspended in lysis buffer (50 mM potassium phosphate (pH 7.8), 400 mM NaCl, 100 mM KCl, 10% glycerol, 0.5% Triton-X, and 10 mM imidazole). The cells were broken by sonicating on ice for short pulses of one second followed by 3 seconds off for a total of 16 min. Cell debris was collected by centrifugation (20 000g at 4 °C for 45 min). The supernatant was filtered using a 0.45 μ m pore size filter (Millipore, Billerica, MA, USA) and applied directly to a nickel affinity column (HiTrap Chelating HP, Amersham Biosciences, Piscataway, NJ, USA) that had been pre-equilibrated with binding buffer (50 mM sodium phosphate, 0.3 M NaCl, 10 mM imidazole, pH 8.0). The protease was eluted with a linear gradient of 50 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, pH 8.0. After elution, the protein was buffer exchanged into 10 mM Tris-HCl pH 7.5, and loaded onto a Q-sepharose anion exchange column (Amersham Biosciences). The protease was eluted with a gradient of 10 mM Tris-HCl, 1 M NaCl, pH 7.5. The pooled fractions containing 3CL^{pro} were exchanged into storage buffer (10 mM sodium phosphate, 10 mM NaCl, 1 mM Tris[2-carboxyethyl] phosphine (TCEP), 1 mM EDTA, pH 7.4) and digested for 48 h at 4 °C with enterokinase (Invitrogen, 0.1 units per 112 μ g of protease) to remove the N-terminal polyhistidine tag. The enterokinase was removed by incubation with EK-away resin (Invitrogen). The reaction mixture was passed through a nickel affinity column to remove undigested prote-

ase. The protease was exchanged into storage buffer, concentrated to 10 mg/mL and used immediately for experiments. The sample was more than 95% pure, as assessed by SDS-PAGE.

Kinetic assay

The activity of the SARS protease 3CL^{pro} was determined by continuous measurement kinetic assays using the fluorogenic substrate Dabcyl-Lys-Thr-Ser-Ala-Val-Leu-Gln-Ser-Gly-Phe-Arg-Lys-Met-Gln-Edans (Genesis Biotech, Taipei, Taiwan). The sequence of the peptide which was derived from the N-terminal auto-cleavage site of the protease is flanked by fluorescent groups, Dabcyl and Edans (24). The increase in fluorescent intensity upon substrate cleavage was monitored in a Cary Eclipse fluorescence spectrophotometer (Varian) using wavelengths of 355 and 538 nm for the excitation and emission, respectively. The experiments were performed in a buffer containing 10 mM sodium phosphate, pH 7.4, 10 mM NaCl, 1 mM TCEP, and 1 mM EDTA. Enzyme activity parameters, K_m and k_{cat} , were determined by initial rate measurements of substrate cleavage at 25 °C in 2% dimethyl sulfoxide (DMSO). The reaction was initiated by adding protease (final concentration 250 nM) to a solution of substrate at final concentration of 0–80 μ M to a total volume of 120 μ L in a microcuvette.

Inhibition assay

Compounds **1**, **2**, **3**, and **5** were purchased from Bachem (Bachem Corporation, USA) and Compound **4** was purchased from Fluka (Sigma-Aldrich Corporation, St Louis, MO, USA). Inhibition assays were performed under the same conditions as described in the Kinetic Assay section with increasing concentrations of substrate (5–20 μ M) in the presence of inhibitor (0–1.5 mM). This data was fit to the first-order rate exponential equation:

$$[P] = \frac{v_0}{k_{app}} (1 - e^{-k_{app}t}) + D \quad (1)$$

where [P] is the product fluorescence, v_0 is the initial velocity of rate of substrate cleavage, D is the displacement term to account for the fact that emission is not zero at the start of the assay measurement, k_{app} is the apparent rate constant of the reaction and t is the time in seconds. The k_{app} obtained was plotted as a function of the [I] in a linear relationship:

$$k_{app} = k_3 \left(1 + \frac{[I]}{K_i \left(1 + \frac{[S]}{K_m} \right)} \right) \quad (2)$$

where k_3 is the inactivation rate constant, K_i is the equilibrium inhibition constant, K_m is the Michaelis constant and [S] is the substrate concentration (19,25). Data from these continuous assays were analysed using the non-linear regression analysis software Origin.

Isothermal titration calorimetry

Isothermal titration calorimetry experiments were carried out using a high precision VP-ITC titration calorimetric system (Microcal Inc.,

Northampton, MA, USA). The enzyme solution in the calorimetric cell was titrated with inhibitor solutions dissolved in the same buffer (10 mM sodium phosphate, 10 mM NaCl, 1 mM TCEP, 1 mM EDTA, pH 7.4) with a 2% final DMSO concentration at 25 °C. The heat evolved after each ligand injection was obtained from the integral of the calorimetric signal. In order to compensate for the delayed k_{off} of the inhibitor, injections were spaced 1000 seconds apart. The heat because of the binding reaction between the inhibitor and the enzyme was obtained as the difference between the heat of reaction and the corresponding heat of dilution.

Trypsin inhibition assay

Selectivity of the compounds was determined by measuring their ability to inhibit commercially available Bovine pancreatic Trypsin (Sigma). Trypsin (final concentration 359 nM) and compound (0–100 μ M) were pre-incubated for 10 min prior to the start of the assay. The reaction was initiated by the addition of the chromogenic substrate, Na-Benzoyl-L-arginine ethyl ester hydrochloride (Sigma) with a final concentration of 200 μ M. The change in absorbance at 253 nm was detected using a Cary spectrophotometer (Varian). The experiments were performed in 50 mM sodium phosphate, pH 7.0, 5% DMSO at 25 °C.

Thrombin inhibition assay

Inhibition of Thrombin by the compounds was evaluated in experimental conditions similar to that of Trypsin. Human Thrombin (Sigma) was pre-incubated with compound for 10 min prior to the start of assay measurements. The final Thrombin concentration was 25 nM and the compound concentration varied from 0–100 μ M. The reaction was initiated by the addition of the chromogenic substrate Sar-Pro-Arg p-nitroanilide dihydrochloride (Sigma) with a final concentration of 208 μ M. The change in absorbance of the substrate upon cleavage by Thrombin was monitored at 405 nm over time using a Cary spectrophotometer (Varian). Experiments were conducted in 50 mM Tris, pH 7.4, 100 mM NaCl, and 5% DMSO at 25 °C.

Calpain inhibition assay

Inhibition against purified Calpain (BioVision, San Francisco, CA, USA) was measured using the fluorescent substrate Ac-Leu-Leu-Tyr-AFC. The reaction was initiated by the addition of 0.5 μ L of enzyme to a mixture containing reaction and extraction buffer that were provided by Biovision in the presence of 2% DMSO at 37 °C and 5 μ L substrate. A Calpain inhibitor, Z-LLY-FMK (provided by manufacturer) was used at a final reaction concentration of 100 nM as a standard to gauge the inhibition by the compounds in this study. The increase in fluorescence upon substrate cleavage was monitored using Cary Eclipse fluorescence spectrophotometer (Varian) using wavelengths of 400 and 505 nm for the excitation and emission, respectively.

Crystallization

Co-crystals of SARS 3CL^{pro} with Compound **4** were obtained by adding 98 μ L SARS 3CL^{pro} (6.5 mg/mL) in 10 mM Tris-HCl, 0.1 M NaCl, 1 mM EDTA, 1 mM TCEP pH 7.4 to 2 μ L Compound **4**

(100 mM) dissolved in 100% DMSO. The final concentration of the inhibitor was 2 mM. The mixture was incubated at room temperature for 15 min to allow for the protein and inhibitor to interact, and then was centrifuged for 5 min at $14\,000 \times g$ to remove any aggregates that had formed. A condition used to crystallize wild-type SARS 3CL^{PRO} was used as a starting condition (18). The best crystals grew in hanging-drop experiments with a 500 μL reservoir solution containing 0.7 M sodium malonate (pH 7.0), and 3–5% isopropanol. The drop was made using 2 μL of reservoir solution and 2 μL of the protein/inhibitor solution. Crystals appeared after 3 months at room temperature.

Data collection

SARS 3CL^{PRO}-Compound **4** co-crystals belong to space group P2₁2₁2 with cell dimensions $a = 106.66 \text{ \AA}$, $b = 45.16 \text{ \AA}$, and $c = 53.96 \text{ \AA}$. Data were collected from a crystal flash-frozen using 10% glycerol as cryoprotectant at beam line X6a at the National Synchrotron Light Source, Brookhaven National Laboratory. Intensity data were integrated, scaled, and reduced to structure factor amplitudes with the HKL2000 suite (26) as summarized in Table 4.

Structure determination and refinement

The structure of the co-crystal of SARS-3CL^{PRO} and Compound **4** was determined by molecular replacement with the program AmoRe (27) using wild-type SARS 3CL^{PRO} (pdb ID 2BX4) as the search model (28). Restrained refinement of the models was performed using REFMAC (29). Manual building was carried out using O (30) and water molecules were placed with Arp-WARP (31). The stereochemistry of the model was checked and analysed using PROCHECK and MolProbity (32).

Coordinates

The coordinates for the structure of SARS 3CL^{PRO} bound to Compound **4** have been deposited in the Protein Data Bank (accession number 3D62).

Pharmacophore generation

Pharmacophore models were generated using the Pharmacophore Application module in MOE 2006.0804 (Quebec, Canada). Being a binary model, a threshold of 1000 μM was selected based on the distribution of K_i data of the compounds in our database. Compounds with activity (K_i in our case) lower than the threshold were chosen as active and those with potency higher than the threshold were inactive. A low energy multi-conformational database of all the compounds in the library were generated using the MMFF94x force field, with a cutoff on the strain energy to be $<4 \text{ kcal/mol}$. The pharmacophore annotation scheme PPCH_ALL, provided by MOE, was used to calculate the planar, polar, charged, and hydrophobic features including all hydrophobes for the conformation library^b. The model was optimized in a training set database of 22 compounds, with 15 active compounds ($K_i < 1000 \mu\text{M}$) and seven inactive compounds ($K_i > 1000 \mu\text{M}$). The structure and activity of all 22 compounds are shown in Table S1 of the Supplementary Material. Based on this breakdown, an ideal model would select only

the active compounds (positive controls) from the training database and not the inactive compounds (negative controls) thereby providing an ideal enrichment factor (R_{ideal}) of 1.5 based on eqn. 3. Flexible alignment of the lowest energy conformation of the most active compounds led to the identification of their critical ligand features. The basic level of this model was the arrangement of the annotation features in a three-dimensional array as shown in Figure 6A. The second level of complexity was further added to this model defining the exclusion regions; compounds with features protruding into these areas were excluded from the model. The final selection criterion was the placement of an external shell which defined the maximum conformational space that can be sampled by molecules. By gradual refinement, the ability of the model to discriminate between active versus inactive compounds was improved. This was gauged by the calculation of an observed enrichment factor (R_{observed}) using eqn. 3 for the training models and the pharmacophores were optimized until $R_{\text{observed}} = R_{\text{ideal}}$. The final model effectiveness was calculated using eqn. 4.

$$R = \frac{\left(\frac{\text{Active Hits}}{\text{Total Hits}} \right)}{\left(\frac{\text{Total Active}}{\text{Total Database}} \right)} \quad (3)$$

$$\text{Model Effectiveness} = \frac{R_{\text{observed}}}{R_{\text{ideal}}} \times 100 \quad (4)$$

Virtual screening

The pharmacophore model was used as a template for virtual screening of commercially available databases provided by MOE 2006.0804 containing a total of 1 000 000 compounds. The screening was carried out using MOE and resulted in 40 hits (18). These 40 compounds were purchased and measured for inhibition against SARS 3CL^{PRO} in a fluorogenic assay that was explained earlier. Of the 40 compounds tested, 38 had a $K_i < 1000 \mu\text{M}$, including two false positives with a $K_i > 1000 \mu\text{M}$. Based on the actives false positives that were selected by the model, R_{observed} , R_{ideal} and the effectiveness were calculated using eqns 3 and 4 respectively. Product 1 was purchased from ChemDiv (ChemDiv, Inc., San Diego, CA, USA), Product 2 was purchased from Sigma (Sigma-Aldrich Corp., USA), Product 3 was purchased from ChemBridge (ChemBridge Corp., USA), Product 4 was purchased from Florida Center for Heterocyclic Compounds (University of Florida, USA), and Product 5 was purchased from Interchim (Montluçon, France).

Results and Discussion

Catalytic mechanism and lead generation

Cysteine proteases, like serine proteases, employ a mechanism involving the formation of an acyl-enzyme intermediate that is hydrolysed via the formation of a tetrahedral adduct (33,34). According to this mechanism, the thiol group of the catalytic Cys 145 in SARS 3CL^{PRO} initiates a nucleophilic attack on the carbonyl flanking the sessile peptide bond of the substrate with the imidazole ring of the His 41 side chain acting as a general base. This proton is later donated to the leaving group of the tetrahedral intermediate (35,36).

Substrate specificity studies of SARS 3CL^{pro} have indicated that the S1 subsite shows preference for Gln at the P1 site of the substrate. The large S2 subsite, on the other hand, can accommodate Leu, Ile, Phe, Val and Met in the P2 position. The S3 subsite is not conserved among coronaviruses and the P3 residue at this site is generally solvent exposed and therefore not well defined. The S4 subsite favours a hydrophobic side chain to fit into this cavity; a position that is usually occupied by Ala in the native substrate (18,34). Based upon this information, we previously generated a library of substrate mimetics linked to the trifluoromethyl ketone warhead that showed moderate affinities towards 3CL^{pro} (23). The best compound (KNI-30001, shown in Figure 2) had Glu at P1, Leu at P2, and Val at P3 and was characterized by a K_i of 116 μM .

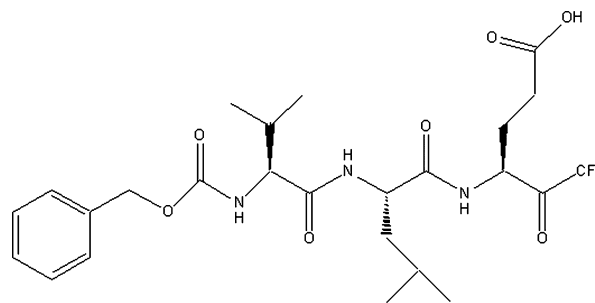


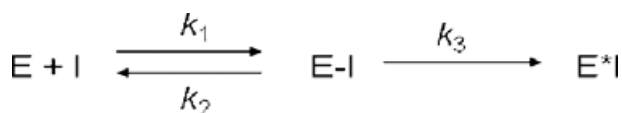
Figure 2: The chemical structure of KNI-30001. The structures were generated using ChemDraw Ultra 6.0 (Cambridge Software). The compound consists of a trifluoromethyl ketone warhead with a Glu in the P1 position, Leu at P2 and Val at P3.

Structure-based optimization of compounds

In order to optimize KNI-30001, the effect of each of the components of the scaffold on the overall affinity towards 3CL^{pro} was evaluated. Three specific aspects of the scaffold were emphasized: (i) the halogenation of the warhead, (ii) the compound size, and (iii) the substitution at the P1 position of the scaffold. Mono-peptide, dipeptide and tripeptide mimetics with modifications in both the halogen content of the warhead and the P1 position were selected and screened against 3CL^{pro} using an *in vitro* kinetic assay. The results for the best five compounds from this screen are shown in Table 1 along with the general scaffold used for optimization. Compounds **1–3** have the same monochloromethyl ketone warhead with alterations at the R₂ site on the scaffold (corresponding to the P1 position). Compounds **1, 2** and **3** have a Phe, naphthalene and a *p*-fluoro phenyl derivative at the P1 position respectively. Compound **4** has a monobromomethyl ketone warhead and an aliphatic substitution at the P1 position. Compound **5** was a dipeptide with a formic acid methyl ester at P1 position and Val at R₃ site (corresponding to the P2 position).

Inhibition results of the compounds

The general reaction scheme used to analyse the inhibition kinetics is shown below in Scheme 1:



Here the E-I indicates a reversible enzyme (E) and inhibitor (I) complex which subsequently undergoes an inactivation step to form the irreversible E*I complex. The K_i for the reversible step is measured as the ratio of k_2/k_1 , and k_3 is the rate-limiting inactivation step. A similar scheme was used earlier to measure the inhibition of halomethyl ketones to other cysteine proteases (37).

A possible mechanism for the inactivation step (Scheme 2) is thought to be initiated by the thiolate imidazolium ion pair at the active site towards the warhead carbonyl, leading to the formation of a thiohemiketal complex which subsequently undergoes an alkylation reaction to form the irreversible product (38). The halomethyl

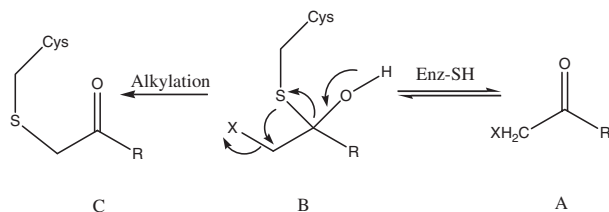
Table 1: Kinetic inhibition data for Compounds **1–5**

Compound	R ₁	R ₂	R ₃	K_i (nM)	k_3 ($\times 10^{-2}$ /second)
1		Cl		306 \pm 10	1.5 \pm 0.1
2		Cl		371 \pm 15	2.8 \pm 0.5
3		Cl		380 \pm 31	1.8 \pm 0.7
4		Br		400 \pm 71	$\ll 0.005$
5	F		Cbz-Val	512 \pm 25	1.6 \pm 0.1

Cbz = benzyloxycarbonyl.

The scaffold shown defines the core structure of the compounds.

group (A in Scheme 2) makes the adjacent ketone group more susceptible to nucleophilic attack. In the presence of a thiol group, the warhead attains the tetrahedral conformation that resembles the enzyme–substrate intermediate formed during substrate catalysis (B in Scheme 2). Over time, intramolecular rearrangement leads to the alkylation with the adjacent carbon of the halomethyl group with the final structure (C in Scheme 2). Another possibility is the attack of the thiolate ion on the halomethyl carbon adjacent to the warhead carbonyl which then leads to the alkylated product.



Since Compounds **1–3** have similar warheads, the differences in their activity can be attributed only to the changes in the P1

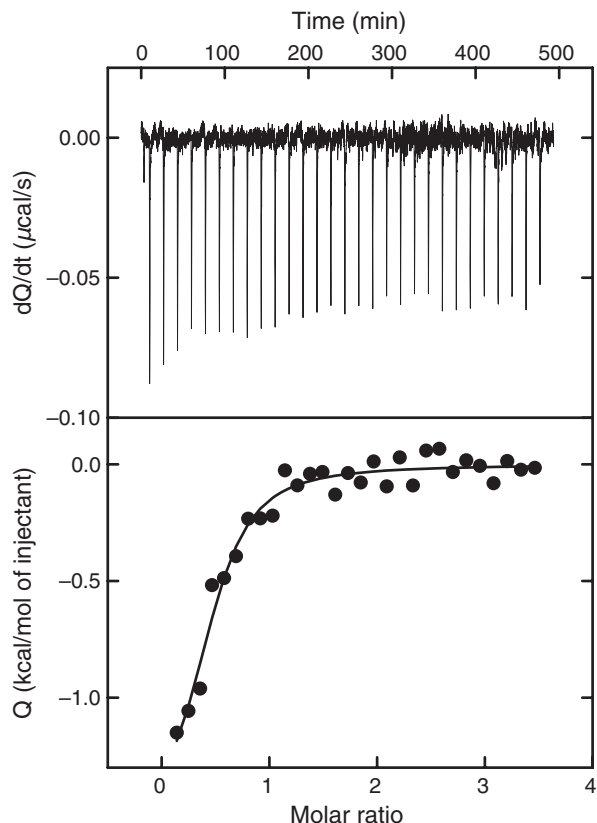


Figure 3: Calorimetric titration of SARS 3CL^{pro} with Compound **4**. In this experiment, each peak represents the injection of 10 μ L of 3CL^{pro} (125 μ M) into the calorimetric cell (1.4272 mL) containing Compound **4** at a concentration of 8 μ M. The experiment was performed at 25 °C in buffer containing 10 mM sodium phosphate, 10 mM NaCl, 1 mM TCEP, 1 mM EDTA, pH 7.4, with a 2% final DMSO concentration.

Table 2: The selectivity of Compounds 1–5 as measured against trypsin, thrombin, and calpain

Compound	K_i against 3CL ^{pro} (nM)	K_i against trypsin (μ M)	K_i against thrombin (μ M)	K_i against calpain (μ M)
(A)				
1	306	>100	>100	10
2	371	>100	75	9
3	380	>200	150	8
4	400	>200	>200	15
5	512	>200	>200	20
(B)				
Compound	Selectivity against trypsin	Selectivity against thrombin	Selectivity against calpain	
1	>300	>300	32	
2	>300	202	24	
3	>500	394	21	
4	>500	>500	38	
5	>300	>400	39	

(A) The measure affinities (K_i) of Compounds **1–5** for 3CL^{pro}, trypsin, thrombin and calpain. (B) Selectivity of Compounds 1–5 with respect to trypsin, thrombin and calpain.

position. The compound with the highest inhibitory potency, Compound **1** had a K_i of 306 ± 10 nM as shown in Table 1. The naphthalene substitution at the P1 position of Compound **2** reduced the potency to a K_i of 371 ± 15 nM. The addition of fluorine at the para position of the phenyl ring at P1 did not change the potency of Compound **3** ($K_i = 380 \pm 31$ nM). Although Gln is traditionally present at this position, a hydrophobic moiety is also highly tolerated. This tolerance is consistent with previous studies where the modifications to the P1 position included a lactam ring in the S stereochemistry (19), keto-glutamine analogs with a phenyl group at P1 (39) and α,β unsaturated ester (40); all of which showed a stark improvement in the inhibitory potency of the compounds to the protease. Compound **4**, with a bromomethyl ketone warhead and an aliphatic substitution at the P1 position had a K_i value of 400 ± 71 nM and may also bind in a similar conformation. Compound **5**, with a monofluoromethyl ketone warhead had a K_i value of 512 ± 25 nM. Altering the methyl ester at the P1 position of this compound into a carboxylic acid, completely rendered it inactive with a $K_i > 1000$ μ M (Table S1 in Supplementary Material). This indicated that the larger footprint occupied by the dipeptidic compounds leads to an altered conformation of the compound when compared to mono-peptidic compounds. The P1 substitution in this orientation fits into a hydrophobic pocket that is more sensitive to structural changes. This hydrophobic moiety provides additional van der Waals interactions at this site that improve the affinity of the ligands.

Previous reports have shown that halomethylketone compounds react with thiols to form thioethers (37). It is also well documented that these warheads form methyl phosphonium salts with reducing agents such as phosphines. Phosphinomethyl ketone compounds were previously shown to inhibit cysteine and serine proteases (41). This raised the possibility that the compounds presented in this

study may be interacting with the protease due to the reaction with tris(2-carboxyethyl)phosphine (TCEP) present in the buffer. However, no loss in inhibition was observed in kinetic measurements of Compounds **1–5** performed in the absence of TCEP. This result indicates that the active species in the inhibition of the protease is the halomethyl ketone and not the phosphinomethyl ketone.

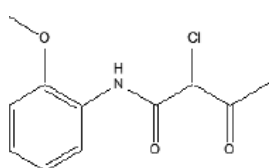
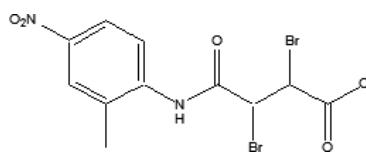
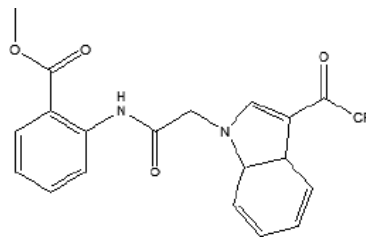
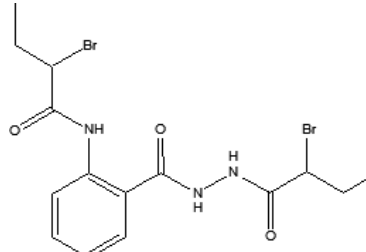
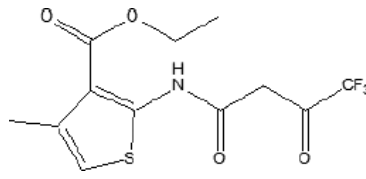
Altering the halogen substitution of the warhead also had a substantial effect on the reactivity of the inhibitors. Substrate analogues with chloromethyl ketones have been shown to inhibit Cathepsin B and Papain whereas fluoromethyl ketones have shown activity against Caspases, Calpains and Cathepsin B as well (37,42–44). Furthermore, NMR experiments have indicated that fluoromethyl and chloromethyl ketones are able to activate the carbonyl carbon of the ketone facilitating the formation of a thio-hemiketal at the active site of cysteine proteases (37). A high rate of inactivation (k_3) is related to the ability of the warhead to accept a nucleophilic attack by the thiol side chain leading to the eventual alkylation. The inactivation constant k_3 for Compound **2** ($2.8 \pm 0.5 \times 10^{-2}$ /second) was almost twice that of Compound **1** ($1.5 \pm 0.1 \times 10^{-2}$ /second) and **3** ($1.8 \pm 0.7 \times 10^{-2}$ /second) (Table 1). The larger P1 moiety of Compound **2** may orient the warhead in a conformation more favourable for reacting with the thiol side chain of Cys 145. The inactivation constant of the dipeptide compound **5** was $1.6 \pm 0.1 \times 10^{-2}$ /second, which was similar to Compounds **1** and **3**. The k_3 of Compound **4** was too small to be measured accurately in the kinetic assay ($\ll 0.005 \times 10^{-2}$ /second), indicating that the irreversible step is very slow and that for several hours the compound behaves as a reversible inhibitor. In fact, the activity of the

Table 3: Crystallographic data collection and refinement statistics

Data collection	
Space group	P2 ₁ 2 ₁ 2
Cell dimensions	106.66, 45.16, 53.96
	<i>a</i> , <i>b</i> , <i>c</i> (Å)
Resolution (Å)	50.00–2.70 (2.80–2.70) ^a
<i>R</i> _{sym} or <i>R</i> _{merge}	0.077(0.505)
<i>I</i> / σ <i>I</i>	29.3 (3.9)
Completeness (%)	98.8 (98.4)
Redundancy	5.0 (5.7)
X-ray Source	X6a
Wavelength	0.9537
Refinement	
Resolution (Å)	50.0–2.70
No. reflections	7.138
<i>R</i> _{work} / <i>R</i> _{free}	0.259/0.362
	(0.299/0.497)
Number of atoms	
Protein	2.350
Ligand/ion	15
Water	53
<i>B</i> -factors	
Protein	36.67
Ligand/ion	49.22
Water	48.03
RMS deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.112

^aParenthesis indicate values for the highest resolution shell.

Table 4: Pharmacophore validation^a

Compound	Structure	MW	<i>K_i</i> (μM)
Product 1		241.67	4.5 ± 1
Product 2		409.01	13 ± 3
Product 3		404.34	27 ± 2
Product 4		449.14	23 ± 4
Product 5		323.29	34.5 ± 9

^aThe pharmacophore was screened through 1 000 000 compounds to result in 40 hits. Only the *K_i* values of the best compounds from the hits are shown.

protease was completely recovered after incubation in 20 μM Compound **4** (50 times the *K_i*) for 10 min. In the case of complete irreversible inhibition, the protease would have been inactivated at such high inhibitor concentrations. In the case of reversible interaction however, activity can be recovered when the inhibitor concentration is diluted. Irreversibility in the enzyme activity was only noticed after incubation times exceeding 12 hours at a high concentration of Compound **4** (data not shown).

This reversible interaction of the bromomethyl ketone warhead of Compound **4** was somewhat unexpected as bromine derivatives are generally better leaving groups and therefore more reactive than

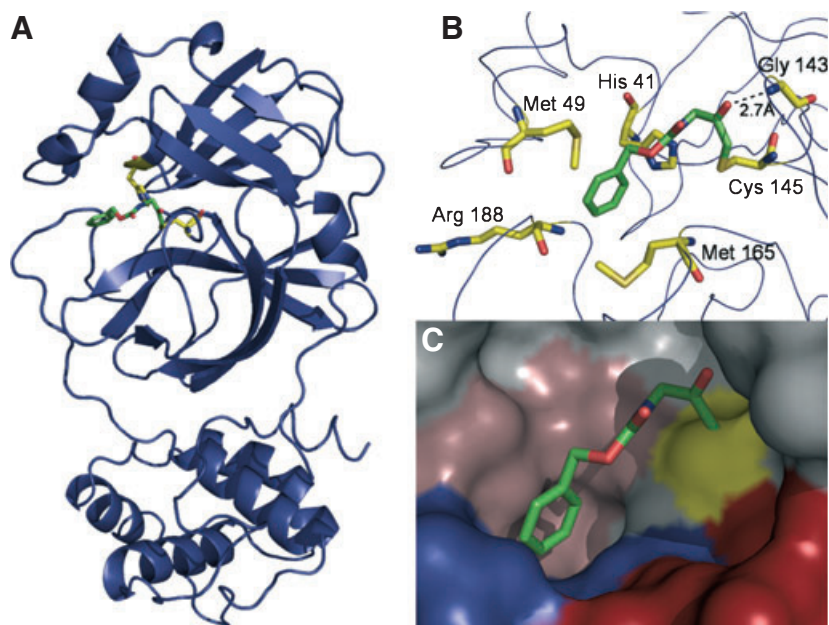


Figure 4: Location of the Compound **4** binding site. (panel A) Crystal structure of SARS 3CL^{pro} bound to Compound **4** showing the location of the ligand within the SARS protease monomer. The protease is depicted in ribbon representation with the catalytic residues His41 and Cys145 shown as yellow sticks. The ligand is shown in stick representation and is coloured by atom type with oxygen in red, nitrogen in blue, and carbon in green. (Panel B) Key residues involved in the protein–ligand interaction. The backbone of the protein is shown as blue ribbon. The side chains of the protein and ligand are coloured by atom type with oxygen in red, nitrogen in blue, sulfur in tan, and carbon in either green or yellow for the ligand and protein, respectively. (panel C) Orientation of the ligand within the active site cavity of SARS 3CL^{pro}. The ligand is depicted in stick representation and is coloured by atom type as in (panel A) and (panel B). The protein is shown in surface representation with each of the defined substrate subsites highlighted. S1 is shown in red, S2 in rose, and S4 in blue. Cys145 is depicted in yellow. Non-subsite residues are shown in grey.

chlorine or fluorine derivatives in SN2 reactions with nucleophiles such as thiol in cysteine proteases. The results from this study pointed to a much lower reactivity of this warhead with the formation of a reversible complex followed by an irreversible alkylation. Also, the inactivation constants (k_3) of the compounds in this study are lower than those reported for other dipeptidyl halomethyl ketones against human Cathepsin B (37). However, the rate of inactivation is dependent on the orientation of the compound warhead in the active site (37), which is different for 3CL^{pro}.

Isothermal titration calorimetry

The binding energetics of compounds **1–5** were also determined by isothermal titration calorimetry (ITC). The calorimetric titrations of compounds **1–3** and **5** were characterized by very large reaction heats (-18 kcal/mol), consistent with the formation of a covalent complex, as expected from the fast irreversible rates (k_3) for these compounds. Compound **4**, on the other hand, had a very different thermodynamic signature, consistent with the observation that the irreversible step of this compound is extremely slow and characterized by a time constant ($1/k_3$) larger than 300 min, i.e. within the calorimeter the binding reaction occurs under equilibrium conditions. Figure 3 shows the calorimetric titration of Compound **4** to 3CL^{pro}. The binding of Compound **4** is characterized by a small favourable binding enthalpy ($\Delta H = -1.6$ kcal/mol) and a favourable entropic

contribution ($-T\Delta S = -6.7$ kcal/mol) resulting in an overall Gibbs energy of -8.3 kcal/mol. The dissociation constant (K_d) determined calorimetrically amounts to 800 nM, which is close to the K_i value estimated from the kinetic inhibition data.

Inhibitor selectivity

Selectivity is a measure of the affinity of a compound against its intended target versus its affinity against other proteins, especially those belonging to the same class. Selectivity is defined as the following ratio:

$$\text{Selectivity} = \frac{K_i \text{ towards unwanted target}}{K_i \text{ towards intended target}} \quad (8)$$

Halomethyl ketone warheads have been shown to selectively target serine and cysteine proteases. Trypsin and thrombin are serine proteases that play a major role in the digestive system and the blood-clotting cascade respectively. The first two domains of 3CL^{pro} have an antiparallel β -barrel structure reminiscent of the chymotrypsin fold, which is also observed in picornavirus 3C proteases. Furthermore, the crystal structure of 3C protease from rhinovirus-14 showed the presence of two topologically equivalent six-stranded β -barrels that were similar to trypsin-like serine proteases such as thrombin (45). In order to investigate the selectivity of the

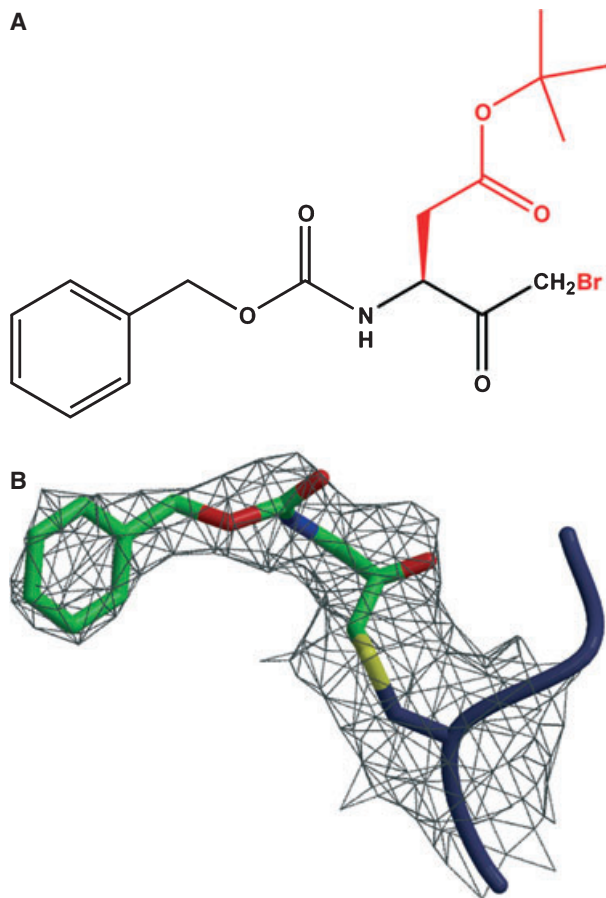


Figure 5: Structure of Compound **4**. (panel A) The chemical structure of Compound **4** is shown with the atoms of the ligand observed in the crystal structure coloured in black and the atoms not observed in red. (panel B) The $2F_o - F_c$ electron density map corresponding to Compound **4** bound to Cys145 is shown, contoured to 1 sigma. The protein and ligand are coloured by atom type with oxygen in red, nitrogen in blue, sulphur in yellow, and carbon in either green or blue for the ligand and protein, respectively.

inhibitors, their ability to inhibit serine proteases such as trypsin and thrombin, as well as a cysteine protease, calpain was tested. The results from these experiments are shown in Table 2a and b. The inhibitors were found to be highly selective towards SARS 3CL^{pro} when compared to the other three proteases. None of the inhibitors were active against trypsin even though the inhibition was measured until the solubility limit of each inhibitor was reached. Similar results were also obtained for inhibition measured against thrombin. In this case, only the K_i values for Compounds **2** and **3** were within detection limits and had values of 72 ± 20 and $150 \pm 30 \mu\text{M}$, respectively, indicating that these compounds were 200 and 400 times more selective towards 3CL^{pro} than thrombin (Table 2b).

The compounds showed a higher affinity for the cysteine protease Calpain compared to trypsin or thrombin. The chloromethyl ketones (Compounds **1–3**) had K_i values against Calpain of 10, 9, and $8 \mu\text{M}$

respectively (Table 2a). Compound **4** had a K_i of $15 \mu\text{M}$ against Calpain and Compound **5** had a K_i of $20 \mu\text{M}$. The calculated selectivity was 32, 24, 21, 38 and 39 for Compound **1**, **2**, **3**, **4** and **5** respectively towards 3CL^{pro} when compared to Calpain (Table 2b). The halomethyl ketone compounds had lower affinities towards Calpain despite the fact that the warhead is reactive towards the thiol side chain in the active site of cysteine proteases. As is the case for other type of warheads, the affinity of the compounds is not dominated by the reactive warhead but the interactions of the entire compound with the residues in the binding cavity. This observation indicates that these compounds can be further optimized to improve their potency and specificity towards 3CL^{pro}.

Compound **4** binds to SARS 3CL^{pro} through a thioether attachment

The structure of the wild-type SARS 3CL^{pro} bound to Compound **4** was determined using X-ray crystallography. Crystallization conditions were similar to those used before for the wild-type protease (18), although the length of time required for crystals to form in the presence of the ligand was longer than for the free protease (three months versus one week). The structure was determined by molecular replacement using SARS 3CL^{pro} (PDB ID 2BX4) (28) as a search model. Data collection and refinement statistics are summarized in Table 3. Crystals belong to the space group P2₁2₁2 and contained one monomer of SARS 3CL^{pro} per asymmetric unit. The final model was refined to a resolution of 2.7 Å with an R -value of 25.9% ($R_{\text{free}} = 36.2\%$). As observed in other structures of SARS 3CL^{pro}, the protein is organized in three domains with domains I and II (residues 3–184) structured into a chymotrypsin-like double β -barrel which forms the active site cavity of the protein. The catalytic domains are linked by a long loop (residues 185–200) to domain III (residues 201–306) of the protein which is comprised predominantly of α -helices. There are no significant differences between the liganded and unliganded structures of the protease (PDB ID 2BX4) (28) as evidenced by their superposition which yields an RMSD of 0.405 Å over 277 C ^{α} atoms.

The inhibitor binds within the substrate-binding cleft formed by the chymotrypsin-like fold of the enzyme (Figure 4A). Electron density for only a portion of the compound was observed in the final structure with the thiol of Cys 145 covalently bound to the carbon adjacent to the warhead carbonyl that was originally bonded to the bromine. Density for the portion of the compound corresponding to the R₂ (or P1) moiety of the ligand was missing. Figures 5A and 5B show the chemical structure of Compound **4** and the electron density in the $2F_o - F_c$ map. The absence of electron density for these atoms is consistent with mass spectrometry experiments which indicate that the compound begins to degrade after 6 h in solution, reflecting a mass consistent with a loss of the *tert*-butyl group in this position (data not shown). The remainder of the R₄ group may not be visible in the structure because of high flexibility in that region of the compound.

The key residues involved in the protein-ligand interaction are shown in Figure 4B. Compound **4** forms a 1.7 Å thioether attachment between the carbon that was originally bonded to bromine and the S ^{γ} of Cys145. In addition, the backbone amide of Gly

Table 5: Structure and activity of the compounds in the training set used to generate the pharmacophore are shown

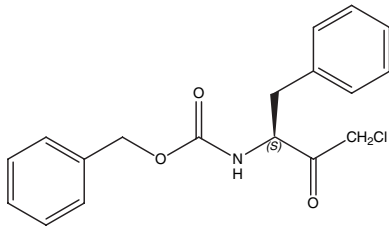
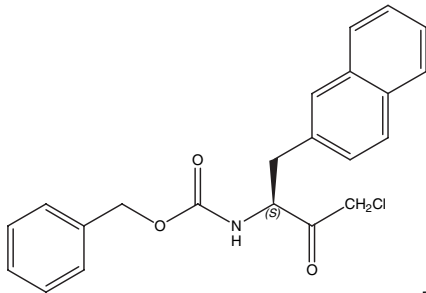
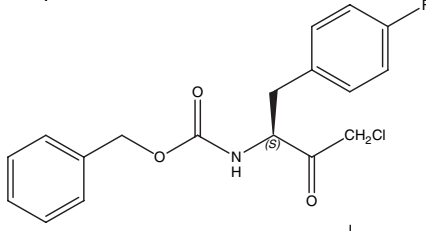
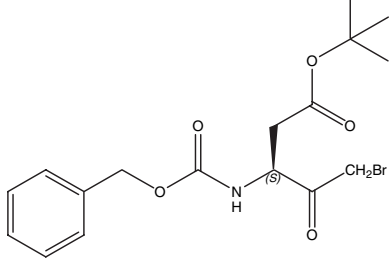
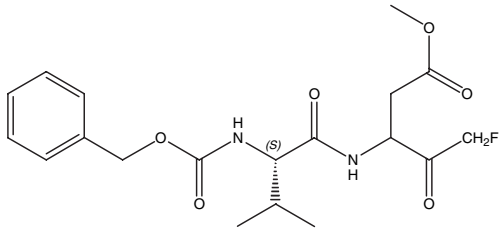
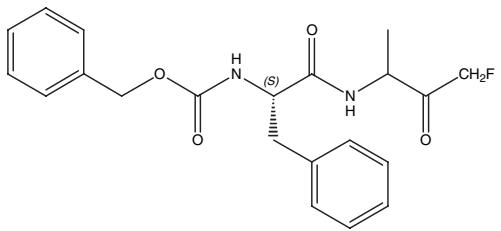
Compound Name	Compound Structure	K_i (μM)
1		0.31 ± 0.01
2		0.37 ± 0.01
3		0.38 ± 0.03
4		0.40 ± 0.07
5		0.51 ± 0.02
6		25.7 ± 4.6

Table 5: Continued

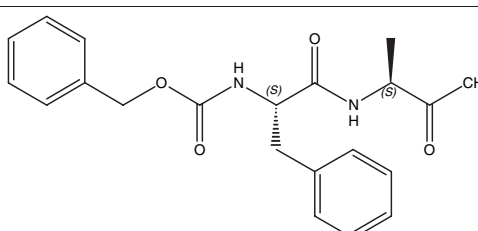
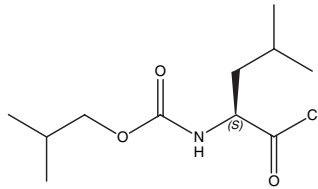
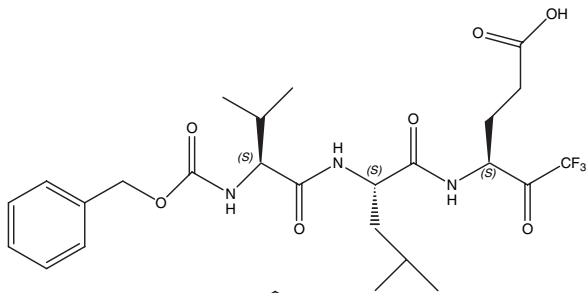
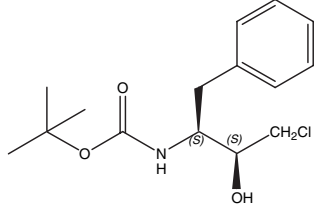
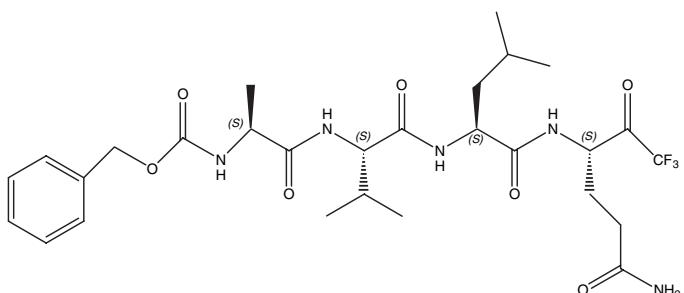
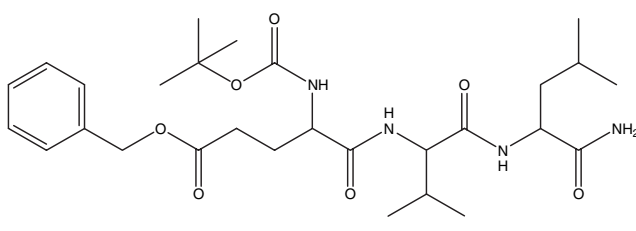
Compound Name	Compound Structure	K_i (μM)
7		64 ± 11.5
8		105 ± 16
9		116 ± 13
10		128 ± 44
11		134.5 ± 32
12		286 ± 62

Table 5: Continued

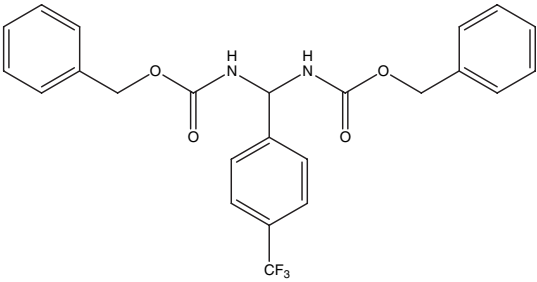
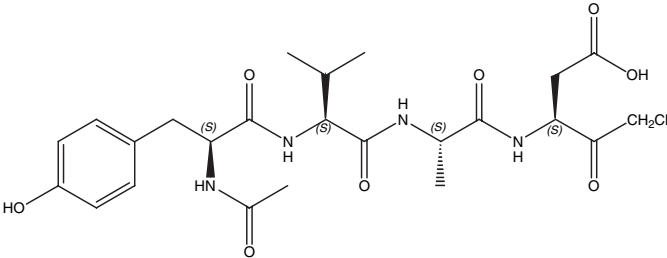
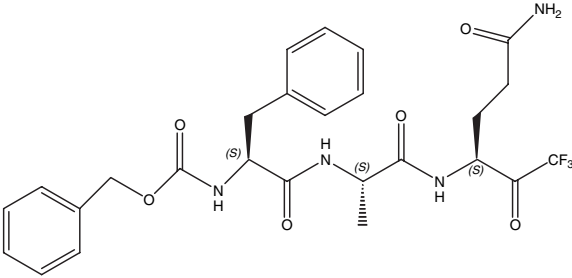
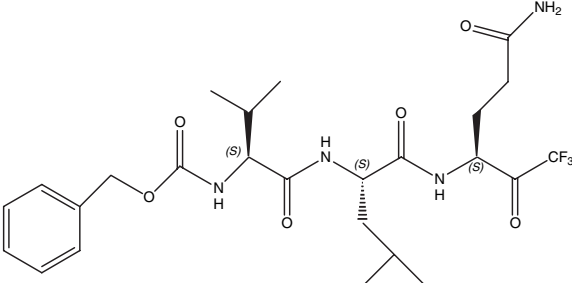
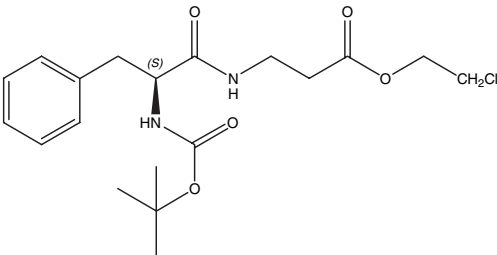
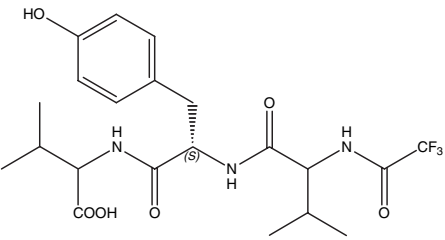
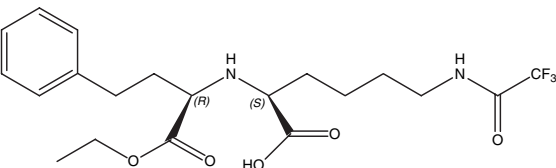
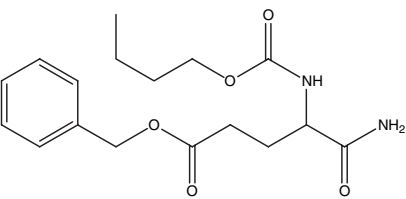
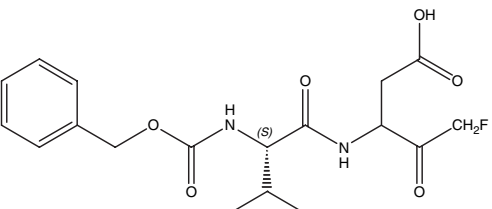
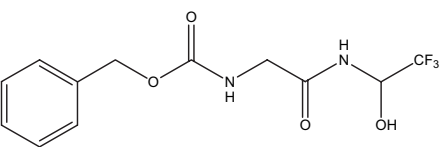
Compound Name	Compound Structure	K_i (μM)
13		366 ± 51
14		389 ± 97
15		844 ± 120
16		>1000
17		>1000

Table 5: Continued

Compound Name	Compound Structure	K_i (μM)
18		>1000
19		>1000
20		>1000
21		>1000
22		>1000

Compounds with $K_i < 1000 \mu\text{M}$ were assigned as active and compounds with $K_i > 1000 \mu\text{M}$ as inactive.

143 is positioned to donate a 2.7 Å hydrogen bond to the oxygen atom of the warhead ketone. The Cbz group of the ligand fits well into the S2 subsite as shown in Figure 4C, making non-bonding contacts to the side chains of His 41, Met 49, Met 165, and the main chain $C\alpha$ atom of Arg 188. The observation that the ligand covalently interacts with the enzyme is consistent with time-dependent inhibition experiments which indicate a bimodal mode of inhibition for this compound with the formation of a reversible complex followed by rearrangement to an irreversible complex after periods of time exceeding 6 h (Table 1). Together, these observations suggest that the initial binding

event is likely followed by the nucleophilic attack of the $S7'$ of Cys 145 on the carbonyl carbon of the warhead, resulting in the formation of a reversible tetrahedral complex which is subsequently followed by the slow, irreversible rearrangement to the thioether. The bromine is not observed in the structure because it is the leaving group during the rearrangement.

Structure-activity relationships

As an optimization tool, a pharmacophore model was generated based on the activity data of the training database containing the

halomethylketone compounds. The pharmacophore model defines the orientation of essential features that make up the scaffold of the library of lead compounds. The viability of each model during the optimization process was determined by accessing the ability to discriminate between true and false positives. In order to generate an accurate model, maximum structural diversity from the compounds in the training set was incorporated by setting a threshold value of 1000 μM . The pharmacophore model was generated based on this activity data combined with the structural information of the compounds (Table 5). The final, refined pharmacophore model as shown in Figure 6A contained the following features: A non-planar hydrophobic feature with a sphere radius of 1.1 Å (red), a planar donor feature with a sphere radius of 1.2 Å (purple), a planar hydrophobic or a non-planar hydrophobic feature with a sphere radius of 1.2 Å (blue), a planar acceptor or planar donor feature

with a sphere radius of 1.3 Å (yellow), and a planar hydrophobic feature with a sphere radius of 1.5 Å (green). This model selected only the active molecules in the training database providing a R_{observed} value of 1.5. For the selected threshold, the theoretical effectiveness of this model, calculated as the ratio of the R_{observed} and R_{ideal} , (eqn 4) was 100%. The applicability of this model can be better visualized by examining a molecule selected on the basis of the structural constraints imposed by the model (Figure 6B). A non-planar hydrophobic feature defined the region corresponding to the warhead of the molecule. The structural features defining the P1 subsite were a planar hydrophobic or a non-planar hydrophobic feature. This result agrees with the kinetic data discussed earlier and indicates that the compounds with the highest affinity to 3CL^{pro} have either a phenyl planar moiety or a flexible aliphatic chain (Table 1).



Library screening for model validation

The pharmacophore model defined above was validated by using a conformational database of approximately 1 000 000 compounds provided by MOE. The pharmacophore search of the test set resulted in the selection of 40 molecules as hits. The hypothesis for this binary model was that the selected molecules would inhibit 3CL^{pro} with a K_i of 1000 μM or lower; the unselected molecules will have a K_i greater than the threshold. Based on this hypothesis, the R_{ideal} was calculated to be 26315.8 using eqn 3. A common feature of the compounds that were hits was the presence of a halomethyl ketone group possessing an adjacent aromatic moiety. When the ability of the 40 selected compounds to inhibit 3CL^{pro} was measured, 38 out of 40 had K_i 's below 1000 μM and 2 were false positives with K_i 's greater than the threshold resulting in a R_{observed} of 25 000. Based on these results, an observed effectiveness of the model to 95% was calculated. The data from the best compounds from this validation step are shown in Table 4. These results provided evidence to the accuracy of the pharmacophore model. The compound with the highest affinity had a K_i of $4.5 \pm 1 \mu\text{M}$ which is encouraging, based on the fact that the compound was chosen from a database not specialized for halomethyl ketones, using a pharmacophore that had a broad threshold. This initial model performed well in both the training and a randomly chosen test set databases, establishing itself as an attractive scaffold that can be further optimized. Some unexpected warheads with moderate affinities such as Product 1 and Product 2 in Table 4 were also recovered. In the first optimization cycle, the activity data from this validation step was used to further refine the pharmacophore by reducing the threshold limit and consequently the sensitivity of the model. Eventually, a highly specialized pharmacophore will be developed that would select for high affinity compounds against 3CL^{pro} with features defined in the model.

Conclusions

The identification of inhibitors targeted towards the highly conserved main protease 3CL^{pro} (M^{pro}) of coronaviruses is an important step towards the development of new classes of antivirals. In this paper, we have shown that halomethyl ketones can be potent and selective inhibitors of the SARS protease 3CL^{pro}. While inhibitors like Compound **4** end up forming a covalent thioether complex, they do so in a very slow fashion (over 6 h) allowing the binding reaction to be controlled by reversible thermodynamic interactions. Compound **4** binds with favourable enthalpy and entropy changes. Since Compound **4** has a molecular weight of only 400.26 Da, it has the potential for much improved potency and selectivity.

Acknowledgments

This work was supported by grants from the National Institutes of Health GM 57144. Beamline X6a of the National Synchrotron Light Source, Brookhaven National Laboratory is also gratefully acknowledged.

References

1. Wat D. (2004) The common cold: a review of the literature. *Eur J Intern Med*;15:79–88.
2. Peiris J.S., Lai S.T., Poon L.L., Guan Y., Yam L.Y., Lim W., Nicholls J. *et al.* (2003) Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet*;361:1319–1325.
3. Rota P.A., Oberste M.S., Monroe S.S., Nix W.A., Campagnoli R., Icenogle J.P., Penaranda S. *et al.* (2003) Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science*;300:1394–1399.
4. Marra M.A., Jones S.J., Astell C.R., Holt R.A., Brooks-Wilson A., Butterfield Y.S., Khattra J. *et al.* (2003) The Genome sequence of the SARS-associated coronavirus. *Science*;300:1399–1404.
5. van der Hoek L., Pyrc K., Jebbink M.F., Vermeulen-Oost W., Berkhout R.J., Wolthers K.C., Wertheim-van Dillen P.M. *et al.* (2004) Identification of a new human coronavirus. *Nat Med*;10:368–373.
6. Woo P.C., Lau S.K., Chu C.M., Chan K.H., Tsoi H.W., Huang Y., Wong B.H. *et al.* (2005) Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol*;79:884–895.
7. Zhao G.P. (2007) SARS Epidemiology from Descriptive to Mechanistic Analyses. *Virus Res*; In press; doi: 10.1016/j.virusres.2007.01.010.
8. Jackwood M.W. (2006) The relationship of severe acute respiratory syndrome coronavirus with avian and other coronaviruses. *Avian Dis*;50:315–320.
9. Sawicki S.G., Sawicki D.L., Siddell S.G. (2007) A contemporary view of coronavirus transcription. *J Virol*;81:20–29.
10. Masters P.S. (2006) The molecular biology of coronaviruses. *Adv Virus Res*;66:193–292.
11. Snijder E.J., Bredenbeek P.J., Dobbe J.C., Thiel V., Ziebuhr J., Poon L.L., Guan Y. *et al.* (2003) Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J Mol Biol*;331:991–1004.
12. Ziebuhr J., Snijder E.J., Gorbalenya A.E. (2000) Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J Gen Virol*;81:853–879.
13. Ziebuhr J., Heusipp G., Siddell S.G. (1997) Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. *J Virol*;71:3992–3997.
14. Kim J.C., Spence R.A., Currier P.F., Lu X., Denison M.R. (1995) Coronavirus protein processing and RNA synthesis is inhibited by the cysteine proteinase inhibitor E64d. *Virology*;208:1–8.
15. Anand K., Palm G.J., Mesters J.R., Siddell S.G., Ziebuhr J., Hilgenfeld R. (2002) Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *EMBO J*;21:3213–3224.
16. Barrila J., Bacha U., Freire E. (2006) Long-range cooperative interactions modulate dimerization in SARS 3CL^{pro}. *Biochemistry*;45:14908–14916.
17. Chen S., Chen L., Tan J., Chen J., Du L., Sun T., Shen J. *et al.* (2005) Severe acute respiratory syndrome coronavirus 3C-like proteinase N terminus is indispensable for proteolytic activity but not for enzyme dimerization. Biochemical and thermodynamic investigation in conjunction with molecular dynamics simulations. *J Biol Chem*;280:164–173.

18. Hsu M.F., Kuo C.J., Chang K.T., Chang H.C., Chou C.C., Ko T.P., Shr H.L. *et al.* (2005) Mechanism of the maturation process of SARS-CoV 3CL protease. *J Biol Chem*;280:31257–31266.
19. Yang H., Xie W., Xue X., Yang K., Ma J., Liang W., Zhao Q. *et al.* (2005) Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol*;3:e324.
20. Bacha U., Barrila J., Velazquez-Campoy A., Leavitt S.A., Freire E. (2004) Identification of novel inhibitors of the SARS coronavirus main protease 3CL^{pro}. *Biochemistry*;43:4906–4912.
21. Anand K., Ziebuhr J., Wadhwani P., Mesters J.R., Hilgenfeld R. (2003) Coronavirus main proteinase (3CL^{pro}) structure: basis for design of anti-SARS drugs. *Science*;300:1763–1767.
22. Hegyi A., Ziebuhr J. (2002) Conservation of substrate specificities among coronavirus main proteases. *J Gen Virol*;83:595–599.
23. Sydnes M.O.H.Y., Sharma V.K., Takashi H., Bacha U., Barrila J., Freire E., Kiso Y. (2006) Synthesis of glutamic acid and glutamine peptide possessing a trifluoromethyl ketone group as SARS-CoV 3CL protease inhibitors. *Tetrahedron*;62:8601–8609.
24. Kuo C.J., Chi Y.H., Hsu J.T., Liang P.H. (2004) Characterization of SARS main protease and inhibitor assay using a fluorogenic substrate. *Biochem Biophys Res Commun*;318:862–867.
25. Tian W.X., Tsou C.L. (1982) Determination of the rate constant of enzyme modification by measuring the substrate reaction in the presence of the modifier. *Biochemistry*;21:1028–1032.
26. Otwinowski Z., Minor W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*;277:307–326.
27. Navaza J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr*;50:157–163.
28. Tan J., Verschuere K.H., Anand K., Shen J., Yang M., Xu Y., Rao Z. *et al.* (2005) pH-dependent conformational flexibility of the SARS-CoV main proteinase (M^{pro}) dimer: molecular dynamics simulations and multiple X-ray structure analyses. *J Mol Biol*;354:25–40.
29. CCP4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*;50:760–763.
30. Jones T., Zou J., Cowan S., Kjeldgaard M. (1991) Improved methods for binding protein models to electron density maps and the location of errors in these models. *Acta Crystallogr*;42:110–119.
31. Perrakis A., Harkiolaki M., Wilson K.S., Lamzin V.S. (2001) ARP/wARP and molecular replacement. *Acta Crystallogr D Biol Crystallogr*;57:1445–1450.
32. Laskowski R., MacArthur M., Moss D., Thornton J. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*;26:283–291.
33. Polgar L., Halasz P. (1982) Current problems in mechanistic studies of serine and cysteine proteinases. *Biochem J*;207:1–10.
34. Fan K., Wei P., Feng Q., Chen S., Huang C., Ma L., Lai B. *et al.* (2004) Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *J Biol Chem*;279:1637–1642.
35. Huang C., Wei P., Fan K., Liu Y., Lai L. (2004) 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. *Biochemistry*;43:4568–4574.
36. Lai L., Han X., Chen H., Wei P., Huang C., Liu S., Fan K. *et al.* (2006) Quaternary structure, substrate selectivity and inhibitor design for SARS 3C-like proteinase. *Curr Pharm Des*;12:4555–4564.
37. Rasnack D. (1985) Synthesis of peptide fluoromethyl ketones and the inhibition of human cathepsin B. *Anal Biochem*;149:461–465.
38. Brady K.D., Giegel D.A., Grinnell C., Lunney E., Talanian R.V., Wong W., Walker N. (1999) A catalytic mechanism for caspase-1 and for bimodal inhibition of caspase-1 by activated aspartic ketones. *Bioorg Med Chem*;7:621–631.
39. Zhang J., Pettersson H.I., Huitema C., Niu C., Yin J., James M.N., Eltis L.D. *et al.* (2007) Design, synthesis, and evaluation of inhibitors for severe acute respiratory syndrome 3C-like protease based on phthalhydrazide ketones or heteroaromatic esters. *J Med Chem*;50:1850–1864.
40. Ghosh A.K., Xi K., Grum-Tokars V., Xu X., Ratia K., Fu W., Houser K.V. *et al.* (2007) Structure-based design, synthesis, and biological evaluation of peptidomimetic SARS-CoV 3CL^{pro} inhibitors. *Bioorg Med Chem Lett*;17:5876–5880.
41. Mallamo J.P.B., Ron T., Ming W., Gregory J. (1997) Phosphorous-containing dipeptide inhibitors of cysteine and serine protease. US Patent Office. Patent Number CODEN: PIXXD2 WO 9703679 A1. February 6.
42. Drenth J., Kalk K.H., Swen H.M. (1976) Binding of chloromethyl ketone substrate analogues to crystalline papain. *Biochemistry*;15:3731–3738.
43. Waterhouse N.J., Finucane D.M., Green D.R., Elce J.S., Kumar S., Alnemri E.S., Litwack G. *et al.* (1998) Calpain activation is upstream of caspases in radiation-induced apoptosis. *Cell Death Differ*;5:1051–1061.
44. Yang W., Guastella J., Huang J.C., Wang Y., Zhang L., Xue D., Tran M. *et al.* (2003) MX1013, a dipeptide caspase inhibitor with potent in vivo antiapoptotic activity. *Br J Pharmacol*;140:402–412.
45. Matthews D.A., Smith W.W., Ferre R.A., Condon B., Budahazi G., Sisson W., Villafranca J.E. *et al.* (1994) Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell*;77:761–771.

Supplementary Material

The following supplementary material is available for this article:

Table S1. The structure and activity of the compounds in the training set used to generate the pharmacophore are shown. The compounds are arranged in descending order of affinity. 15 compounds with $K_i < 1000 \mu\text{M}$ were assigned as active and the compounds with $K_i > 1000 \mu\text{M}$ were inactive.

This material is available as part of the online article from: <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1747-0285.2007.00679.x>

(This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Notes

^aGwaltney Jr JM, Hayden FG. (2007) www.commoncold.org.In.

^bLin A. (2004) Overview of Pharmacophore Applications in MOE. Chemical Computing Group Inc. <http://www.chemcomp.com/journal/ph4.htm>. Access date 11/13/07.