

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/6233114>

Substrate Specificity Profiling and Identification of a New Class of Inhibitor for the Major Protease of the SARS Coronavirus †,‡

ARTICLE *in* BIOCHEMISTRY · AUGUST 2007

Impact Factor: 3.02 · DOI: 10.1021/bi0621415 · Source: PubMed

CITATIONS

35

READS

14

9 AUTHORS, INCLUDING:



[Michael McDowell](#)

New Mexico State University

7 PUBLICATIONS 124 CITATIONS

[SEE PROFILE](#)



[Colleen Jonsson](#)

University of Tennessee

126 PUBLICATIONS 2,392 CITATIONS

[SEE PROFILE](#)



[Charles S Craik](#)

University of California, San Francisco

335 PUBLICATIONS 15,205 CITATIONS

[SEE PROFILE](#)

Substrate Specificity Profiling and Identification of a New Class of Inhibitor for the Major Protease of the SARS Coronavirus^{†,‡}

D. H. Goetz,[#] Y. Choe,[#] E. Hansell,[§] Y. T. Chen,^{||} M. McDowell,[⊥] C. B. Jonsson,[⊥] W. R. Roush,^{||}
J. McKerrow,[§] and C. S. Craik,^{*,#}

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143, Pathology Department, University of California, San Francisco, San Francisco, California 94158-2550, Department of Chemistry and Biochemistry, The Scripps Research Institute Florida, Jupiter, Florida 33458, and Department of Biochemistry and Molecular Biology, Southern Research Institute, Birmingham, Alabama 35205

Received October 13, 2006; Revised Manuscript Received May 1, 2007

ABSTRACT: Severe acute respiratory syndrome (SARS) is an emerging infectious disease associated with a high rate of mortality. The SARS-associated coronavirus (SARS-CoV) has been identified as the etiological agent of the disease. Although public health procedures have been effective in combating the spread of SARS, concern remains about the possibility of a recurrence. Various approaches are being pursued for the development of efficacious therapeutics. One promising approach is to develop small molecule inhibitors of the essential major polyprotein processing protease 3Clpro. Here we report a complete description of the tetrapeptide substrate specificity of 3Clpro using fully degenerate peptide libraries consisting of all 160 000 possible naturally occurring tetrapeptides. The substrate specificity data show the expected P1-Gln P2-Leu specificity and elucidate a novel preference for P1-His containing substrates equal to the expected preference for P1-Gln. These data were then used to develop optimal substrates for a high-throughput screen of a 2000 compound small-molecule inhibitor library consisting of known cysteine protease inhibitor scaffolds. We also report the 1.8 Å X-ray crystal structure of 3Clpro bound to an irreversible inhibitor. This inhibitor, an α,β -epoxyketone, inhibits 3Clpro with a k_3/K_i of $0.002 \mu\text{M}^{-1} \text{s}^{-1}$ in a mode consistent with the substrate specificity data. Finally, we report the successful rational improvement of this scaffold with second generation inhibitors. These data provide the foundation for a rational small-molecule inhibitor design effort based upon the inhibitor scaffold identified, the crystal structure of the complex, and a more complete understanding of P1–P4 substrate specificity.

A newly discovered, highly infectious coronavirus has been shown to be the causative agent of severe acute respiratory syndrome (SARS).¹ The SARS-associated coronavirus (SARS-CoV) is a potentially lethal pathogen, and much effort has been spent developing effective therapeutics. DNA-based vaccination using the structural spike protein has

so far been promising with reports of protective immunity in mice (1). An intranasal vaccine has also been shown to protect monkeys (2). While vaccines represent a potentially effective first line of defense against a recurrence of SARS, the speed at which SARS can spread necessitates the availability of drugs to treat those who are already infected.

[†] This work was supported in part by the National Institutes of Health (NIH) grant GM56531 to C.S.C., a National Research Service Award GM069243 to D.H.G., an NIH Tropical Disease Research Unit Grant AI035707 to W.R.R., and the Sandler Family Supporting Foundation.

[‡] Coordinates for the structure discussed in this manuscript have been deposited in the protein databank under pdb code 2OP9.

* To whom correspondence should be addressed: University of California San Francisco, Mission Bay Genentech Hall, MS 2280, 600 16th St. San Francisco, California 94143. Phone 415-476-8146. Fax 415-502-8298. E-mail craik@cgl.ucsf.edu.

[#] Department of Pharmaceutical Chemistry, University of California.

[§] Pathology Department, University of California.

^{||} The Scripps Research Institute Florida.

[⊥] Southern Research Institute.

¹ Abbreviations: SARS, severe acute respiratory syndrome; SARS-CoV, SARS coronavirus; kb, kilobase; PLP, papain-like protease; 3Clpro, 3C like protease; IBV, avian infectious bronchitis virus; TGEV, transmissible gastroenteritis virus; MHV, murine hepatitis virus; TLS, torsion libration screw; PS-SCL, positional-scanning synthetic combinatorial library; ACC, 7-amino-4-carbamoyl-coumarin; CBZ, benzylloxycarbonyl; CMK, chloromethylketone; IC₅₀, concentration at which 50% of cells display cytopathic effect; EC₅₀, concentration at which viral replication is inhibited by 50%.

The SARS-CoV is a positive-sense single-stranded RNA virus (3). The 30 kilobase (kb) long genome is predicted to encode 10 open reading frames (4, 5). The first open reading frame at 21 kb is referred to as the replicase and encodes two overlapping polyproteins Orf1a and Orf1b (5). Functionally active proteins are released from the polyproteins through proteolytic processing by two virally encoded thiol proteases. A papain-like protease (PLP) is predicted to cleave the polyprotein in four places, and a chymotrypsin-like cysteine protease (3Clpro), often called the major protease, is predicted to cleave the polyprotein at 11 sites (4). Both proteases are thought to be essential for viral replication, and they share significant homology with proteases of other coronaviruses including avian infectious bronchitis virus (IBV), transmissible gastroenteritis virus (TGEV), and murine hepatitis virus (MHV). The 11 3Clpro cleavage sites are characterized by a P1-Gln, P2-Leu/Phe motif. Of these, the most preferred cleavage site contains the P6–P1 sequence T–S–A–V–L–Q (6). The P1-Gln specificity is conserved

in all known coronavirus 3Clpro cleavage sites and is considered an absolute requirement for cleavage by this enzyme (7).

3Clpro has previously been expressed and characterized (8). The enzyme is an obligate dimer with one independent active site per monomer. In addition, structural data have been reported including structures bound to a peptidyl chloromethyl ketone (CMK) (9). Finally, small molecule inhibitors have been identified and characterized (8, 10–19). Here, we report the non-prime side tetrapeptide specificity of the SARS-CoV including evidence for a previously undescribed P1 substrate specificity that is conserved among related proteases of the coronavirus family. We also report the identification and characterization of a promising small-molecule inhibitor scaffold that inhibits this class of enzymes as well as viral production. Finally, we report structural data describing the mode of inhibitor binding of this new class of 3Clpro inhibitors that provides a structural rationale for the unique substrate specificity of the enzyme.

MATERIALS AND METHODS

Protein Cloning, Expression, and Purification. cDNA from the Tor2 strain of the SARS-CoV was provided as a gift from Dr. Derisi, UCSF. 3Clpro from IBV and MHV were obtained from Dr. Baker at Loyola University Chicago and Dr. Britton of the Compton Laboratory in Newbury U.K. The gene encoding 3Clpro was cloned, expressed, and purified as described by Kuo et al. (8) with a yield of approximately 50 mg of purified protein per liter of culture. Various constructs containing amino and carboxy terminal purification tags were also attempted but found to be inferior to the native sequence primarily due to defects in dimerization. The purified and concentrated 3Clpro was stored in 25 mM Tris pH 7.5, 0.5 mM EDTA, 14 mM 2-mercaptoethanol, 100 mM NaCl at 10 mg/mL. For long-term storage, the enzyme was snap frozen in liquid N₂ and stored at –80 °C. Snap freezing was found to have no deleterious affect on enzyme activity.

Enzyme Assay, Substrate Library Profiling, and Inhibitor Screening. 3Clpro substrate specificity was determined using a complete diverse tetrapeptide substrate library (20) with 10 μM enzyme in each well in 100 mM Tris pH 7.5, 200 mM NaCl, 0.01% Brij-35, 1 mM 2-mercaptoethanol. This library contains a total of 160 000 unique sequences covering all possible tetrapeptides covalently linked via an amide bond to a fluorogenic amino-carbamoyl coumarin moiety. For single substrate kinetic assays, enzyme concentration was held constant at 35 nM, and substrate was varied from 0.12 to 250 μM. The initial reaction velocity was measured during the linear phase of the reaction by monitoring the time-dependent increase in fluorescence with an excitation wavelength of 360 nm and an emission wavelength of 480 nm in a 96-well microtiter plate fluorometer. The data were fit to the Michaelis–Menten equation using nonlinear regression in Kaleidograph. For our high-throughput screen, substrate concentration was kept constant at 25 μM. Candidate inhibitors were first assayed at 200 μM inhibitor concentration. Approximately 2000 compounds were tested. The inclusion of 0.01% Brij-35 detergent was found to be crucial for avoiding nonspecific inhibition by promiscuous inhibitors (21). Compounds exhibiting greater than 95%

inhibition of proteolytic activity were retested at a 10 μM inhibitor concentration.

Inhibitor Library. The inhibitor library is a collection of candidate cysteine protease inhibitors curated at U.C.S.F. by the Sandler Center for Basic Research in Parasitic Diseases (McKerrow, unpublished data). The individual inhibitors have been collected from several academic and industry collaborators. These inhibitors consist of a variety of electrophilic functional groups known to inhibit cysteine proteases including hydroxymethylketones, phenothiazines, thiosemicarbazides, acylamides, vinyl sulfones, and acyl hydrazines as well as reversible inhibitor scaffolds. The library primarily directed against proteases of *Plasmodium falciparum* and *Trypanosoma cruzi*. The *P. falciparum* and *T. cruzi* cysteine proteases falcipain and cruzain possess a P2-Leu preference suggesting its utility for screening against 3Clpro.

Inhibitor Kinetics. Substrate concentration was held constant at 100 μM, and enzyme concentration was tested at both 35 and 350 nM. Twelve different inhibitor concentrations were tested varying from 0.1 to 100 μM. Initial velocities were collected in triplicate as described above. Data were analyzed with the program dynafit (22) using a competitive inhibitor model. Because of the irreversible nature of the inhibitors studied, inhibition data are presented as a ratio of the rate of inhibition (k_3) versus the inhibition constant K_i .

Epoxyketone Synthesis. WRR 210 and WRR 211 were synthesized as described (23). WRR 182, 183, 485–488, 492, 493, 495 and 496 were prepared by established methods to synthesize other epoxy ketones with minor modifications (23–25). Treatment of Weinreb amide (Scheme 1, compound labeled **1**) with vinylmagnesium bromide, followed by ketone reduction provided 2:1 mixtures of *R* and *S* allylic alcohols, respectively. The Boc group was removed with trifluoroacetic acid, and the resulting amines were coupled to the appropriate Cbz-protected amino acids under standard peptide coupling conditions. Epoxidation of **4** with *m*-chloroperoxybenzoic acid, followed by oxidation of the epoxy alcohol using Dess–Martin periodinane provided both diastereomers of the targeted epoxy ketones which were separated by silica gel chromatography.

Crystallization Data Collection and Structure Determination. 3Clpro was cocrystallized with an equimolar concentration of inhibitor WRR 183 in 10% PEG 6000, 14 mM 2-mercaptoethanol, 5% DMSO, 10% glycerol, 150 mM sodium chloride, and 100 mM MES pH 6.0. Crystals were soaked in the synthetic mother liquor with 200 mM Tris pH 7.5 and 20% glycerol for 15 min prior to flash cooling in liquid nitrogen. Diffraction quality crystals grew in three weeks. Data were collected at the advanced light source beamline 8.3.1 at Lawrence Berkeley National Labs Berkeley, CA with 1.1 Å incident radiation in 180 1° oscillations at 100 K. A second pass with an increased detector to crystal distance was collected to better measure the low-resolution reflections. The space-group was determined to be *P2*₁. Data were processed, reduced, and scaled using MOSFLM and SCALA (26, 27). Phases were determined by molecular replacement using model 1Q2W in EPMR (28). The structure was refined against the observed data to 1.8 Å resolution in REFMAC5 (29). The structural model was built manually using the interactive program suite XtalView (30). Free atoms

Scheme 1: Synthesis of Epoxyketone Inhibitors

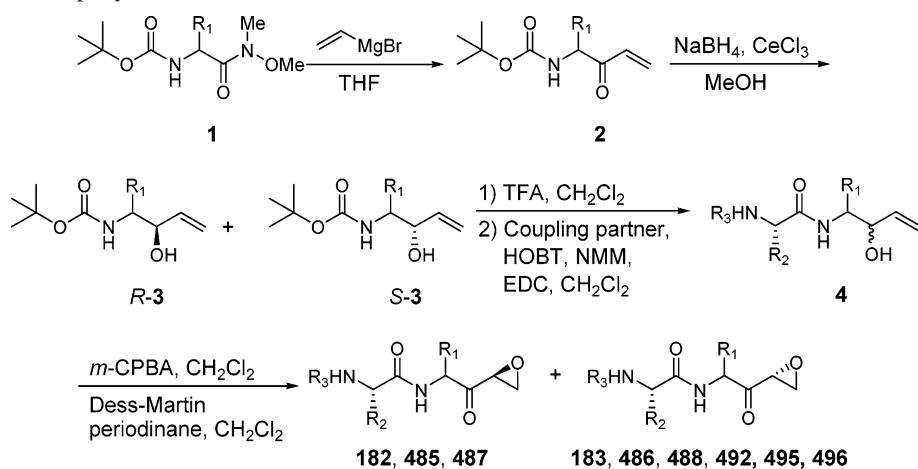


Table 1: Data Collection and Refinement Statistics

wavelength (Å)	1.1
resolution (Å)	60.0–1.8 (1.85–1.8)
cell dimensions (Å)	53.3 96.8 67.9 (P_{2_1})
unique reflections	56 987
R_{syn} (%)	0.054 (0.32)
completeness (%)	98.8 (97.0)
multiplicity	6.3 (4.8)
$I/\sigma I$	18.7 (3.8)
refinement	
protein atoms (2 molecules/au)	4665 (603 amino acids)
ligand atoms	58
solvent atoms	556
R_{cryst}	17.0 (20.3)
R_{free}	20.7 (27.6)
rmsd bonds	0.015 Å
rmsd angles	1.5°

density refinement as implemented in Arp/wARP (31) was used to reduce model bias in the electron density maps during the early rounds of model building. The bound inhibitor was modeled after the free R-factor was below 28%. Finally, torsion libration screw (TLS) refinement of rigid body motions (32) was used to model thermal motion of each protease monomer as well as each bound ligand (a total of four TLS groups). Waters were added prior to the last step of refinement. Statistics for the X-ray cocrystal structure are reported in Table 1. The coordinates of this structure have been deposited in the protein data bank (33).

Inhibition of Viral Replication Assay. The effects of WRR 183 and WRR 495 on viral replication were assayed as previously described (34). Briefly, 50 µL of Vero E6 cells at 10 000 cells/well were dispensed into a 96-well plate and incubated for 24 h at 37 °C, 5% CO₂. Following 24 h, compound was added, and cells were infected by SARS-CoV (Tor2) at a concentration of 100* 50% tissue culture infection dose. Internal controls included media only, cells in media only, cells infected with virus, and virus infected cells treated with calpain inhibitor IV. After 72 h incubation, 100 µL of Promega CellTiter-Glo was added to each well, and the resulting luminescence was recorded. IC₅₀ and EC₅₀ values were then calculated as described by Severson et al. (34).

RESULTS

Tetrapeptide Specificity, Single Substrate Kinetics, and Inhibitor Discovery. Recombinant SARS-CoV 3Clpro was assayed for substrate specificity using a complete diverse

positional-scanning synthetic combinatorial library (PS-SCL) based on the 7-amino-4-carbamoyl-coumarin (ACC) fluorogenic leaving group. (Figure 1a) This technology provides an unbiased way of assigning the tetrapeptide substrate specificity of a given protease. As expected based on both biological polyprotein cleavage sites and substrate specificity against synthetic peptides, the enzyme exhibits a strong preference for P1 glutamine containing substrates and P2 leucine containing substrates. Surprisingly, the enzyme also shows a strong preference for P1 histidine containing substrates. We confirmed these results by synthesizing single substrates and comparing their kinetic constants (Figure 1b). Figure 1b also shows that 3Clpro has extended substrate specificity at P5 and P6 preferring hydrophobic amino acids such as leucine.

We also tested the P1 His containing substrate Ac-T-S-A-V-L-H-(ACC) against the related 3C-like proteases from IBV and MHV, to determine if this P1-His preference is conserved. These enzymes share functional as well as sequence and structural homology. Both enzymes exhibit similar strong specificity for polyprotein cleavage after Gln with a k_{cat}/K_m of $2.8 \pm 0.6 \text{ mM}^{-1} \text{ s}^{-1}$ and $2.2 \pm 0.4 \text{ mM}^{-1} \text{ s}^{-1}$ for IBV and MHV, respectively. Analysis of Ac-T-S-A-V-L-H-(ACC) shows that IBV and MHV 3C-like proteases cleave after P1-His with a k_{cat}/K_m of $19.9 \pm 3.1 \text{ mM}^{-1} \text{ s}^{-1}$ and $16.3 \pm 1.1 \text{ mM}^{-1} \text{ s}^{-1}$ for IBV and MHV, respectively, demonstrating that, unlike SARS-CoV 3Clpro, histidine is strongly preferred in the P1 position in comparison to P1-Gln.

The substrate specificity data were then used to perform a high-throughput in vitro screen of a focused cysteine protease specific inhibitor library using the optimized fluorogenic peptide substrate Ac-T-S-T-K-L-Q-ACC. After a first pass through the inhibitor library using an inhibitor concentration of 200 µM, 80 compounds out of a total of 2000 were chosen for further testing. These compounds were then subjected to a second round of screening at an inhibitor concentration of 10 µM. The most promising scaffold to result from our library screen was the dipeptidyl epoxyketone (23, 24) based on the reproducible time-dependent inhibition of proteolytic activity. Dipeptidyl derivatives of the parent compound from the library were synthesized and selected for further study (Figure 2).

The kinetic constants for the best inhibitor WRR 183 are $k_{\text{inact}} = 0.004 \pm 0.0003 \text{ s}^{-1}$ and $K_i = 2.2 \pm 0.2 \mu\text{M}$. WRR

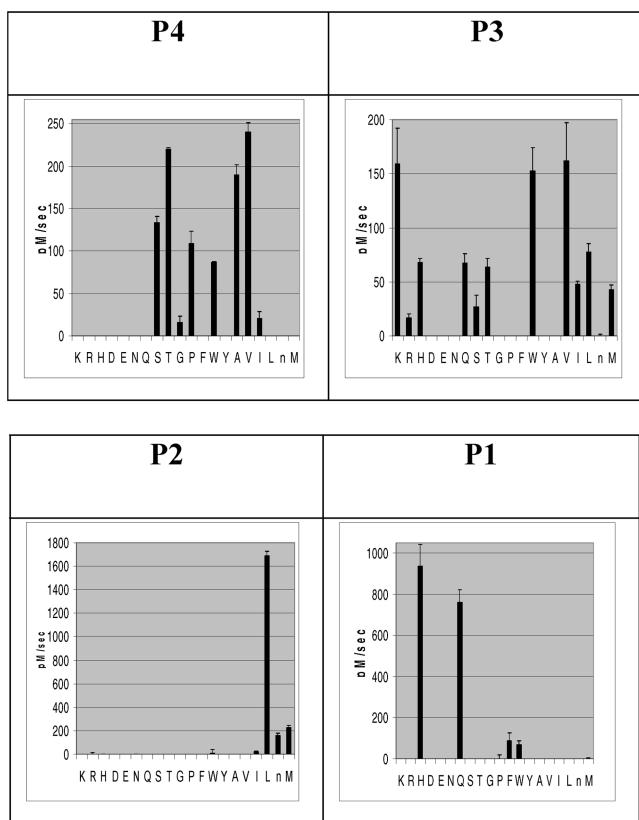
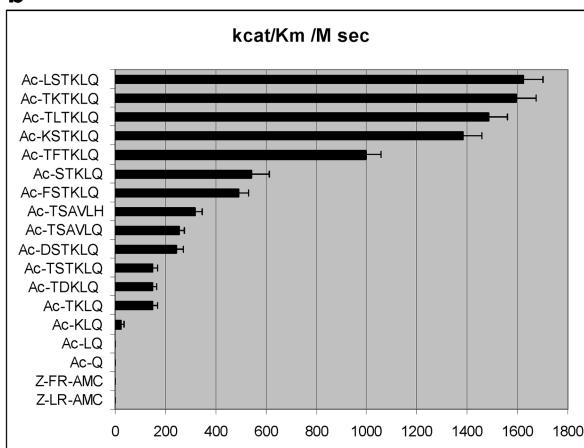
a**b**

FIGURE 1: Extended substrate specificity of SARS-CoV 3Clpro. (a) Results from P1, P2, P3, and P4 diverse positional scanning libraries. The y-axis represents the substrate cleavage rates in picomolar concentrations of fluorophore released per second. The x-axis indicates the amino acid held constant at each position, designated by the one-letter code (with n representing norleucine). (b) Single substrate kinetic data for SARS-CoV 3Clpro.

183 also inhibits MHV 3Clpro but shows no inhibition against IBV 3Clpro up to 20 μ M. This compound is an epoxide-based inhibitor with a P3 L-phenylalanine residue and an R configuration at C-2 of the epoxide group (Figure 2b). The stereochemistry of the epoxide was found to be crucial. WRR 182 and WRR 210, which are the C-2 (S) epoxide isomers of WRR 183 and WRR 211, are 10-fold less potent against the enzyme. These four compounds (Figure 2a–d) were evaluated for inhibition of viral replication in a tissue culture assay. WRR 182 and 183 inhibited viral replication with no detectable cytotoxicity, and WRR

183 inhibited viral replication >50% at 10 μ M. Finally, a second generation of inhibitors based on the WRR 183 scaffold was synthesized on the basis of computer modeling, substrate specificity data, and synthetic accessibility (Figure 2e–l) and tested for inhibition of 3Clpro. WRR 495 was found to inhibit 3Clpro with a k_3/K_i of 0.5 μ M^{-1 s⁻¹.}

WRR 183 and WRR 495 were tested in a tissue culture based assay for their affect on viral replication. Both compounds exhibited inhibition of viral replication at high concentration (>50 μ M). However, WRR 495 showed pronounced toxicity demonstrating toxic effects in 50% of cells at approximately 20 μ M. WRR 183 on the other hand was found to inhibit the virus with an EC₅₀ of 12 μ M, while exhibiting 50% toxicity at 60.56 μ M.

Mode of Inhibitor Binding. There are two molecules in the asymmetric unit of the WRR183:3Clpro crystal structure, and the inhibitor interacts with each monomer in an identical manner to within coordinate error except at the N-terminal serine residue. Exact angles and distances will only be reported for the interaction between monomer chain A and inhibitor chain C unless otherwise noted. WRR 183 is a peptidyl α,β -epoxyketone that inhibits 3Clpro by S-alkylation of the active site cysteine. Nucleophilic attack by the active site cysteine results in the opening of the epoxide ring and the formation of a 1.9 \AA thioether bond at C-2. (Figure 3a) This opening also results in extension of the atoms of the linearized oxirane ring past the oxyanion hole and into the P1 binding site. This has the effect of altering the expected register of binding. Thus, although WRR 183 has an Ala at P1, after catalysis a new P1 moiety is formed from the epoxide ring opening and Ala becomes P2. A hydrogen bond (3.2 \AA) is formed between the O1 terminal hydroxyl of the inhibitor and the backbone carbonyl of Leu 144. The terminal hydroxyl forms a hydrogen bond (3.0 \AA) to an ordered water molecule deep in the P1 pocket. This water is further stabilized by hydrogen bonds to N^ε of His 163 (2.9 \AA) and the N^δ of His 172 (3.4 \AA).

The P2 residue of the inhibitor WRR183 is an L-alanine. The side chain of this residue is oriented toward solvent and does not interact with any subsite, its closest approach being 4.9 \AA from the C' of Thr 25. However, the backbone amide hydrogen bonds with the catalytic His 41 epsilon nitrogen (3.1 \AA) and the carbonyl oxygen hydrogen bonds to the amide hydrogen atoms of Gly 143 (3.0 \AA) and Cys 145 (3.2 \AA). The side chain of the P3 Phe lies sandwiched in between the side chains of Met 49 and Met 165 forming favorable sulfur arene interactions (35). The terminal carbons of the phenylalanine side chain (C11–C13) make close contacts with the main chain atoms of residues 187 and 188. The carbonyl carbon of the benzyloxycarbonyl (CBZ) protecting group forms a hydrogen bond to the backbone amide of Glu 166 (3.1 \AA).

Two different conformations are seen at the N-terminus of the two monomers in the asymmetric unit. In monomer B, the formal positive charge of the Ser 1B NH₃⁺ points toward the side chain of Asn 214B (2.6 \AA) forming an intramolecular interaction. In monomer A, the formal positive charge of the N-terminus projects toward the S4 subsite of the neighboring monomer and is neutralized by the side chain of Glu 166 (3.5 \AA), forming an intramolecular salt bridge. This makes the Glu166 side chain effectively uncharged and the phenyl ring of the CBZ protecting group is positioned

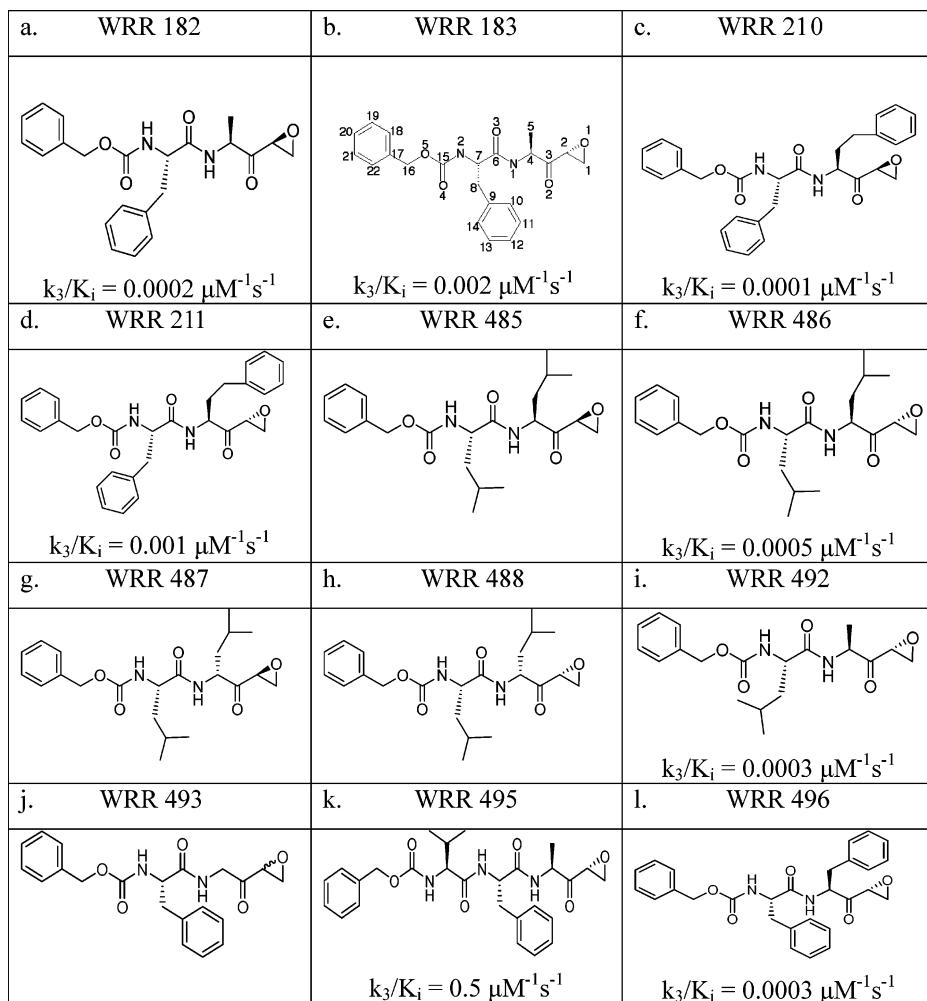


FIGURE 2: WRR 183 and derivatives. (a–l) Compound name is listed above each structure. Inhibition data are given for those compounds that showed inhibition in our initial screen. The atoms of WRR 183 are labeled as in the protein databank coordinate file.

to adopt extensive van der Waals contacts with the resulting hydrophobic pocket. However, the electron density for this ring is disordered in both monomers of our crystal structure making the precise orientation of this phenyl group ambiguous (Figure 3b).

Modeling of a histidine in the S1 pocket of the protease bound to a P1 glutamine inhibitor (19) suggests that both histidine and glutamine can adopt the same binding mode in the S1 pocket (Figure 3c). In particular, both residues are capable of hydrogen bonding with the carbonyl O of Phe 140 and the N^ε of His 163 in the bottom of the binding pocket. Comparison of the S1 pockets of SARS-CoV 3Clpro and TGEV 3Clpro (36) suggests that the related IBV and MHV 3Clpro enzymes also adopt a similar binding mode but fails to explain the increased preference of these enzymes for P1-His preferences in comparison to SARS-CoV 3Clpro.

DISCUSSION

The substrate specificity of SARS-CoV 3Clpro (Figure 1a) shows a novel combination for P2 and P1 residues that may allow the design of extremely specific inhibitors. Although the P1 specificity profile clearly demonstrates an expected P1 glutamine specificity shown previously by inspection of known natural cleavage sites in vivo and against peptides in vitro, there is also an unexpected P1-His preference. This

P1-His preference is unlikely to be an artifact arising from a coumarin reporting group in the P1' position. First, of the over 30 serine and cysteine proteases we have profiled for tetrapeptide substrate specificity with this method (20, 37, 38), none exhibit a preference for P1-His and except for 3Clpro none exhibit an unexpected P1 preference. Second, extensive characterization of Granzyme B demonstrates that the absolute preference of the enzyme for P1-Asp is not affected by the use of a coumarin leaving group (39). Previously, it has been reported that this enzyme shows an absolute preference for P1-Gln (4, 6, 8, 16, 19, 40) demonstrating the advantage of using a fully degenerate tetrapeptide library to determine the P4 to P1 substrate specificity of a protease.

An exhaustive search of the MEROPS database (41) shows that no mammalian proteases cleave after P2-Leu P1-His. Figure 1b shows that the enzyme recognizes P1-His containing substrates with an equivalent k_{cat}/K_m as corresponding P1-Gln substrates. It is likely that a P1-His inhibitor may show increased specificity for 3Clpro over host proteases. The preference for P1-His substrates also raises the possibility that unidentified proteolytic sites encoded in the coronavirus or host genome exist. A search of possible protease sites using POPS (42) reveals a possible cleavage site in the RNA-dependent RNA polymerase nsp9 (GenBank ID NP_828869) of SARS-CoV.

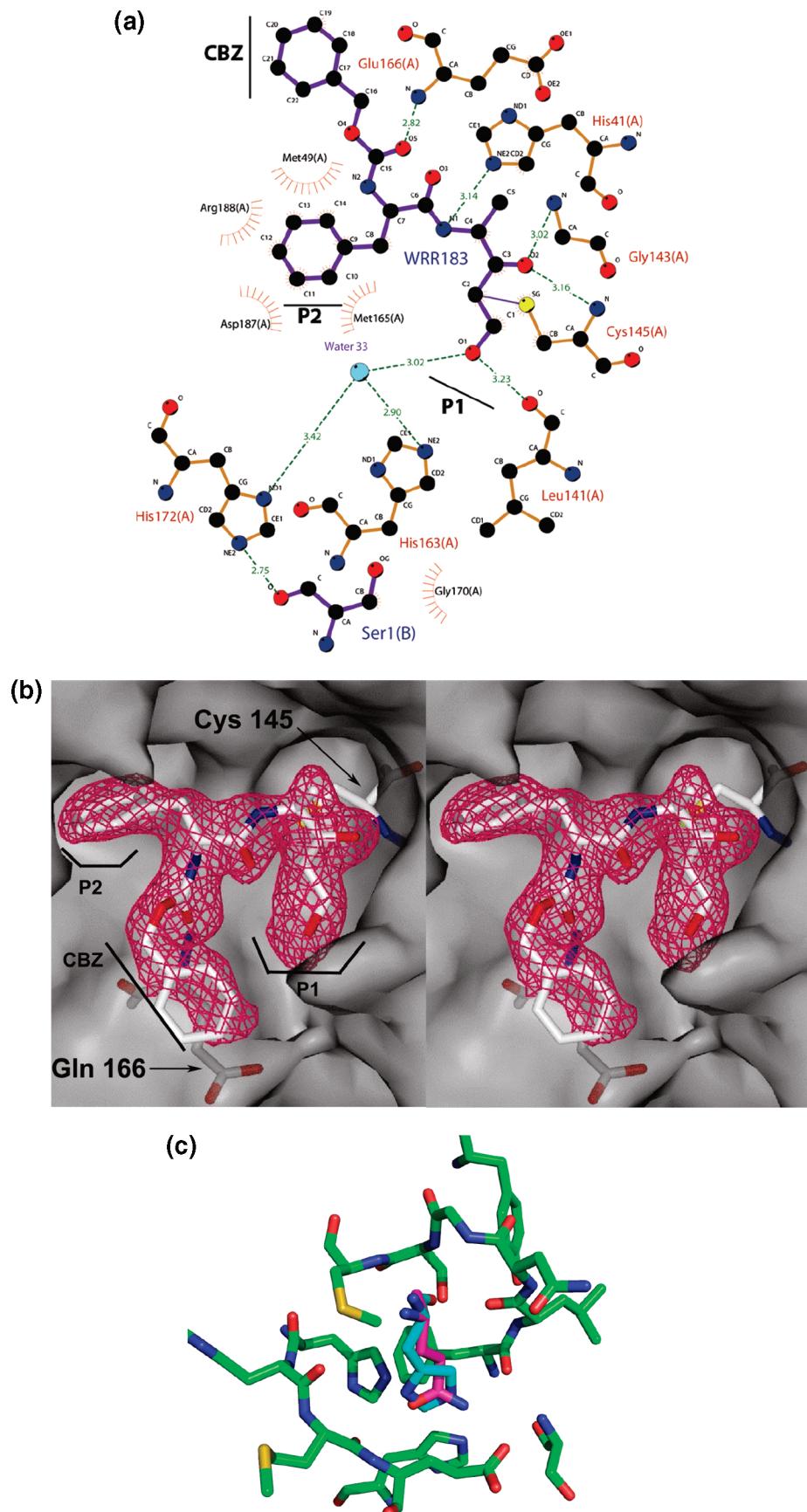


FIGURE 3: Structure of SARS-CoV 3Clpro bound to WRR 183. (a) Ligplot diagram of the interactions between WRR 183 (purple) and 3Clpro (orange). Dashed lines indicate hydrogen bonds, and numbers indicate distances in Å. Hydrophobic interactions are shown as arcs with radial spokes. (b) Cross-eyed stereoview of the WRR 183:3Clpro electron-density. The molecular surface of 3Clpro is rendered in gray. WRR 183 and protease residues Glu 166 and Cys 145 are represented as sticks and labeled. 2Fo-Fc electron density for WRR 183 contoured at 1.5 s is represented as magenta mesh. (c) Model of P1-His binding in the S1 pocket of 1UK4. Protease residues lining the S1 pocket are rendered as sticks with carbons colored green. P1-Gln of 1UK4 is rendered in as sticks with carbons colored magenta. P1-His is modeled and rendered as sticks with carbons colored cyan.

A subset of P5 and P6 preferences was analyzed through the use of single substrates based on the identity of P5 and P6 residues at known *in vivo* SARS-CoV cleavage sites. The increase in k_{cat}/K_m with substrates containing a Leu at either P5 or P6 suggests that the Leu may interact with the same residues of the SARS-CoV extended substrate binding site whether it is in P5 or P6. This phenomenon is similar to that found for the P3 or P4 (but not both) specificity for arginine found in membrane-type serine protease 1 (43). Figure 1b also clearly shows the strong requirement this enzyme has for extended substrate interactions before catalysis can occur. Unlike trypsin, for example, which can cleave ACC substrates with a single amino acid length, we do not detect SARS-CoV 3Clpro proteolytic activity with substrates shorter than two amino acids, and we detect only minimal activity against tripeptides.

The hexapeptide substrate, designed on the basis of the tetrapeptide library profile, substrate solubility, synthetic accessibility, and known SARS protease cleavage sites, allowed us to perform a high-throughput fluorometric assay for inhibitors using a focused cysteine protease specific inhibitor library. The library contains many compounds designed to bind with a Leu or other hydrophobic moiety in the S2 pocket, suggesting its utility for 3Clpro inhibitor discovery.

This library contains a diverse array of chemotypes known to inhibit cysteine proteases. Interestingly, only the epoxy ketones were shown to potently inhibit 3Clpro. However, direct comparison of the relative efficacy of different electrophilic functional groups was not explored further. Our library also contains several derivatives of the chemically similar epoxysuccinate or E-64 inhibitors including E-64c and E-64d and those reported to inhibit cathepsins (44). We detected no inhibition of 3Clpro with these compounds, nor with any peptidyl vinyl sulfones (45). Molecular modeling suggests that this may be due to steric hindrance by the C-1 carboxylate of the epoxysuccinate class of inhibitors with residues lining the oxyanion hole.

Lee et al. (46) have recently published the structure of 3Clpro bound to an aza-peptide epoxide analogue. In their structure, the P1 residue (Gln) of the inhibitor lies in the S1 site. In contrast, in our structure the first residue of the inhibitor occupies the noncanonical S2 subsite. This may be because Gln is heavily preferred over the residues we have tested, namely, Ala, Phe, homo-Phe, and Gly at the P1 position. Unfortunately, we cannot test this hypothesis because both Gln and His at the P1 position in a WRR 183 analogue have proven to be synthetically inaccessible. However, because of this major difference in binding modes between the aza-peptide epoxides of Lee et al. and the epoxyketones described here, we consider these scaffolds to be fundamentally different.

Epoxyketones were previously designed and synthesized as inhibitors of the cysteine protease cruzain, but the C-2 (S) epimers 182 and 210 were substantially more active against that enzyme (23) demonstrating that the correct stereochemistry at the epoxide ring is essential for potent inhibition. The cocrystal structure with WRR 183 suggests that binding of the S epimer WRR 182 would position the C2 carbon too far from the active-site cysteine for efficient attack.

The structure also exhibits some differences when compared to the crystal structures of 3Clpro bound to the CMK pentapeptide Cbz-Val-Asn-Ser-Thr-Leu-Gln (9) and a peptidyl Michael acceptor (19). In the CMK complex, the authors describe two different pH switched active site conformations, one catalytically competent and one incompetent characterized by a collapsed P1-pocket and oxyanion hole. This incompetent form is present at low pH. In our structure, we do not see the incompetent conformation. Presumably, this is due to the soak of our crystals into synthetic mother liquor at pH 7.5 prior to flash cooling.

The ring-opening of the epoxide upon nucleophilic attack by the active-site cysteine results in the formation of an alcohol group in the P1 position. This hydroxyl group makes hydrogen bonds with residues at the bottom of the S1 pocket both directly and through a water molecule that mimic the mode of binding of substrate in the S1 pocket. These favorable interactions suggest that it may be possible to use this scaffold to synthesize potent noncovalent inhibitors of SARS-CoV 3Clpro. The P2-Ala of WRR183 binds at the noncanonical S2 binding site described by Yang et al. (9). This interaction is somewhat surprising because the side chain in both structures points toward bulk solvent belying the expected strong preference for leucine. Yang et al. hypothesize that this perhaps explains the relative lack of stringency for the P2 residue for Sars-CoV 3Clpro as compared to the homologous major cysteine proteases of TGEV and HCoV. However, our PS-SCL library results clearly demonstrate that in fact leucine is strongly preferred at P2. This suggests that while leucine is strongly preferred at P2, other amino acids, namely, Met and Phe, are tolerated but result in drastically lower cleavage rates. The work of Fan et al. (6) corroborates this hypothesis reporting a 50% decrease in cleavage for P2 Phe substrates and a >90% decrease in cleavage rates for Val and Met substitutions. However, a more thorough comparison between SARS-CoV, TGEV, and HCoV 3Clpro is still necessary for a full understanding. Regardless, the mode of binding of the P2-Ala with the side chain oriented toward solvent in our structure does explain how WRR 211 (P2-homo-Phe) can inhibit an enzyme with such a strong leucine preference at that site. The structural and biochemical data on these inhibitors suggest that considerable diversity will be tolerated at this site.

The P3-Phe interaction is likely the greatest determinant of specificity for this inhibitor. It interacts with a methionine clasp (Figure 3a) in the canonical S2 subsite identified by Yang et al. (9). This is also the same site at which the P2-Leu of the previously reported peptidyl Michael acceptor N2 (19) binds. A search of the relibase database of protein ligand interactions (47) shows that this favorable sulfur–arene interaction is similar to that found in calmodulin bound to trifluoperazine (48). Additionally, C-12, the terminal carbon of the P3-Phe side chain of the inhibitor approaches within 3.4 Å of the carbonyl oxygen of Arg 188 deep in the pocket (Figure 3a). The close contact suggests that a heterocyclic phenylalanine analogue capable of an electrostatic or hydrogen bonding interaction with this carbonyl oxygen might be favored.

It is unclear whether the CBZ group of WRR 183 plays an important role in the *in vitro* activity of WRR 183 against SARS-CoV 3Clpro. The contribution of the CBZ protecting

group is also unknown *in vivo*. Certainly, given the strong preference of 3Clpro for extended substrates, derivatization of this protecting group, either by replacement with a peptide bond and another amino acid or by a nonpeptide functional group, presents a facile way of exploring further diversity for this scaffold. To that end, WRR 495 was synthesized and found to be 275-fold more potent than the parent compound WRR 183. WRR 183 and 495 are identical except for the addition of the most preferred amino acid for the S3 pocket (Figure 1a). This suggests that the inhibitor improvement *in vitro* is due to the use of the optimal amino acid Val for interaction with the S3 pocket.

Validation of this scaffold as an effective inhibitor against replication of virus in a tissue culture model provided promising results for compound WRR 183. However, despite the 275-fold improvement of WRR 495 over WRR 183 against SARS-CoV 3Clpro *in vitro*, WRR 495 exhibited greatly increased toxicity and no increase in efficacy. This result demonstrates the increased complexity encountered when transitioning from an *in vitro* enzyme based assay to a cell-based assay and suggests there are other aspects of this scaffold that must be optimized.

Recently, there have been a number of successes in the search for efficacious therapeutics against the SARS-CoV. The work we report here reveals the complete tetrapeptide specificity of the 3Clpro drug target identifying an unexpected P1 specificity, provides a useful pharmacophore lead, identifies a low-molecular weight small molecule that inhibits the enzyme, and elucidates the molecular determinants of this inhibitor/enzyme interaction. Future work will focus on examining a possible *in vivo* role for the P1 histidine preference including whether or not it is a general phenomenon of coronaviral major proteases, identifying promising electrophilic functional groups other than epoxides, and examining the effects of various substituents at P2 and P3 of the WRR 183 scaffold on viral replication.

ACKNOWLEDGMENT

We would like to thank the staff at the Advanced Light Source beamline 8.3.1 for assistance with data collection.

REFERENCES

- Subbarao, K., McAuliffe, J., Vogel, L., Fahlé, G., Fischer, S., Tatti, K., Packard, M., Shieh, W. J., Zaki, S., and Murphy, B. (2004) Prior infection and passive transfer of neutralizing antibody prevent replication of severe acute respiratory syndrome coronavirus in the respiratory tract of mice, *J. Virol.* **78**, 3572–3577.
- Bukreyev, A., Lamirande, E. W., Buchholz, U. J., Vogel, L. N., Elkins, W. R., St Claire, M., Murphy, B. R., Subbarao, K., and Collins, P. L. (2004) Mucosal immunisation of African green monkeys (*Cercopithecus aethiops*) with an attenuated parainfluenza virus expressing the SARS coronavirus spike protein for the prevention of SARS, *Lancet* **363**, 2122–2127.
- Marra, M. A., Jones, S. J., Astell, C. R., Holt, R. A., Brooks-Wilson, A., Butterfield, Y. S., Khattri, J., Asano, J. K., Barber, S. A., Chan, S. Y., Cloutier, A., Coughlin, S. M., Freeman, D., Girn, N., Griffith, O. L., Leach, S. R., Mayo, M., McDonald, H., Montgomery, S. B., Pandoh, P. K., Petrescu, A. S., Robertson, A. G., Schein, J. E., Siddiqui, A., Smailus, D. E., Stott, J. M., Yang, G. S., Plummer, F., Andonov, A., Artsob, H., Bastien, N., Bernard, K., Booth, T. F., Bowness, D., Czub, M., Drebot, M., Fernando, L., Flick, R., Garbutt, M., Gray, M., Grolla, A., Jones, S., Feldmann, H., Meyers, A., Kabani, A., Li, Y., Normand, S., Stroher, U., Tipples, G. A., Tyler, S., Vogrig, R., Ward, D., Watson, B., Brunham, R. C., Krajden, M., Petric, M., Skowronski, D. M., Upton, C., and Roper, R. L. (2003) The genome sequence of the SARS-associated coronavirus, *Science* **300**, 1399–1404.
- Gao, F., Ou, H. Y., Chen, L. L., Zheng, W. X., and Zhang, C. T. (2003) Prediction of proteinase cleavage sites in polyproteins of coronaviruses and its applications in analyzing SARS-CoV genomes, *FEBS Lett.* **553**, 451–456.
- Rota, P. A., Oberste, M. S., Monroe, S. S., Nix, W. A., Campagnoli, R., Icenogle, J. P., Penaranda, S., Bankamp, B., Maher, K., Chen, M. H., Tong, S., Tamin, A., Lowe, L., Frace, M., DeRisi, J. L., Chen, Q., Wang, D., Erdman, D. D., Peret, T. C., Burns, C., Ksiazek, T. G., Rollin, P. E., Sanchez, A., Liffick, S., Holloway, B., Limor, J., McCaustland, K., Olsen-Rasmussen, M., Fouchier, R., Gunther, S., Osterhaus, A. D., Drosten, C., Pallansch, M. A., Anderson, L. J., and Bellini, W. J. (2003) Characterization of a novel coronavirus associated with severe acute respiratory syndrome, *Science* **300**, 1394–1399.
- Fan, K., Wei, P., Feng, Q., Chen, S., Huang, C., Ma, L., Lai, B., Pei, J., Liu, Y., Chen, J., and Lai, L. (2004) Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase, *J. Biol. Chem.* **279**, 1637–1642.
- Lai, L., Han, X., Chen, H., Wei, P., Huang, C., Liu, S., Fan, K., Zhou, L., Liu, Z., Pei, J., and Liu, Y. (2006) Quaternary structure, substrate selectivity and inhibitor design for SARS 3C-like proteinase, *Curr. Pharm. Des.* **12**, 4555–4564.
- Kuo, C. J., Chi, Y. H., Hsu, J. T., and Liang, P. H. (2004) Characterization of SARS main protease and inhibitor assay using a fluorogenic substrate, *Biochem. Biophys. Res. Commun.* **318**, 862–867.
- Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G. F., Anand, K., Bartlam, M., Hilgenfeld, R., and Rao, Z. (2003) The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor, *Proc. Natl. Acad. Sci. U. S. A.* **100**, 13190–13195.
- Blanchard, J. E., Elowe, N. H., Huitema, C., Fortin, P. D., Cechetto, J. D., Eltis, L. D., and Brown, E. D. (2004) High-throughput screening identifies inhibitors of the SARS coronavirus main proteinase, *Chem. Biol.* **11**, 1445–1453.
- Chen, L., Gui, C., Luo, X., Yang, Q., Gunther, S., Scandella, E., Drosten, C., Bai, D., He, X., Ludewig, B., Chen, J., Luo, H., Yang, Y., Zou, J., Thiel, V., Chen, K., Shen, J., Shen, X., and Jiang, H. (2005) Cinanserin is an inhibitor of the 3C-like proteinase of severe acute respiratory syndrome coronavirus and strongly reduces virus replication *in vitro*, *J. Virol.* **79**, 7095–7103.
- Chen, L. R., Wang, Y. C., Lin, Y. W., Chou, S. Y., Chen, S. F., Liu, L. T., Wu, Y. T., Kuo, C. J., Chen, T. S., and Juang, S. H. (2005) Synthesis and evaluation of isatin derivatives as effective SARS coronavirus 3CL protease inhibitors, *Bioorg. Med. Chem. Lett.* **15**, 3058–3062.
- Jain, R. P., Pettersson, H. I., Zhang, J., Aull, K. D., Fortin, P. D., Huitema, C., Eltis, L. D., Parrish, J. C., James, M. N., Wishart, D. S., and Vederas, J. C. (2004) Synthesis and evaluation of keto-glutamine analogues as potent inhibitors of severe acute respiratory syndrome 3CLpro, *J. Med. Chem.* **47**, 6113–6116.
- Kao, R. Y., To, A. P., Ng, L. W., Tsui, W. H., Lee, T. S., Tsui, H. W., and Yuen, K. Y. (2004) Characterization of SARS-CoV main protease and identification of biologically active small molecule inhibitors using a continuous fluorescence-based assay, *FEBS Lett.* **576**, 325–330.
- Liu, Y. C., Huang, V., Chao, T. C., Hsiao, C. D., Lin, A., Chang, M. F., and Chow, L. P. (2005) Screening of drugs by FRET analysis identifies inhibitors of SARS-CoV 3CL protease, *Biochem. Biophys. Res. Commun.* **333**, 194–199.
- Shie, J. J., Fang, J. M., Kuo, T. H., Kuo, C. J., Liang, P. H., Huang, H. J., Wu, Y. T., Jan, J. T., Cheng, Y. S., and Wong, C. H. (2005) Inhibition of the severe acute respiratory syndrome 3CL protease by peptidomimetic alpha, beta-unsaturated esters, *Bioorg. Med. Chem.* **13**, 5240–5252.
- Shie, J. J., Fang, J. M., Kuo, C. J., Kuo, T. H., Liang, P. H., Huang, H. J., Yang, W. B., Lin, C. H., Chen, J. L., Wu, Y. T., and Wong, C. H. (2005) Discovery of potent amide inhibitors against the severe acute respiratory syndrome 3CL protease, *J. Med. Chem.* **48**, 4469–4473.
- Wu, C. Y., Jan, J. T., Ma, S. H., Kuo, C. J., Juan, H. F., Cheng, Y. S., Hsu, H. H., Huang, H. C., Wu, D., Brik, A., Liang, F. S., Liu, R. S., Fang, J. M., Chen, S. T., Liang, P. H., and Wong, C. H. (2004) Small molecules targeting severe acute respiratory

- syndrome human coronavirus, *Proc. Natl. Acad. Sci. U. S. A.* **101**, 10012–10017.
19. Yang, H., Xie, W., Xue, X., Yang, K., Ma, J., Liang, W., Zhao, Q., Zhou, Z., Pei, D., Ziebuhr, J., Hilgenfeld, R., Yuen, K. Y., Wong, L., Gao, G., Chen, S., Chen, Z., Ma, D., Bartlam, M., and Rao, Z. (2005) Design of wide-spectrum inhibitors targeting coronavirus main proteases, *PLoS. Biol.* **3**, e324.
20. Choe, Y., Leonetti, F., Greenbaum, D. C., Lecaille, F., Bogyo, M., Bromme, D., Ellman, J. A., and Craik, C. S. (2006) Substrate profiling of cysteine proteases using a combinatorial peptide library identifies functionally unique specificities, *J. Biol. Chem.* **281**, 12824–12832.
21. McGovern, S. L., Helfand, B. T., Feng, B., and Shochet, B. K. (2003) A specific mechanism of nonspecific inhibition, *J. Med. Chem.* **46**, 4265–4272.
22. Kuzmic, P. (1996) Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase, *Anal. Biochem.* **237**, 260–273.
23. Roush, W. R., Gonzalez, F. V., McKerrow, J. H., and Hansell, E. (1998) Design and synthesis of dipeptidyl alpha',beta'-epoxy ketones, potent irreversible inhibitors of the cysteine protease cruzain, *Bioorg. Med. Chem. Lett.* **8**, 2809–2812.
24. Der-Feng, G., Spaltenstein, A., Leban, J. J., Huang, J. J., Reinhardt, K. R., Viveros, O. H., Sigafoos, J., and Crouch, R. (1996) Design and synthesis of novel protease inhibitors. Tripeptide α',β'-epoxyketones as nanomolar inactivators of the proteasome, *Tetrahedron Lett.* **37**, 1343.
25. Wuts, P. G. M., Putt, S. R., and Ritter, A. R. (1988) Synthesis of the dipeptide hydroxyethylene isostere of Leu-Val, a transition state mimic for the control of enzyme function, *J. Org. Chem.* **53**, 4503–4508.
26. Leslie, A. G. W. (1992) Recent changes to the MOSFLM package for processing film and image plate data, *Joint CCP4 ESF-EAMCB Newslett. Protein Crystallogr.* **26**.
27. Collaborative Computational Project, N. (1994) The CCP4 Suite: Programs for Protein Crystallography, *Acta. Crystallogr. D. Biol. Crystallogr.* **50**, 760–763.
28. Kissinger, C. R., Gehlhaar, D. K., and Fogel, D. B. (1999) Rapid automated molecular replacement by evolutionary search, *Acta. Crystallogr. D. Biol. Crystallogr.* **55** (Pt 2), 484–491.
29. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta. Crystallogr. D. Biol. Crystallogr.* **53**, 240–55.
30. McRee, D. E. (1999) XtalView/Xfit-A versatile program for manipulating atomic coordinates and electron density, *J. Struct. Biol.* **125**, 156–165.
31. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement, *Nat. Struct. Biol.* **6**, 458–463.
32. Winn, M. D., Murshudov, G. N., and Papiz, M. Z. (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions, *Methods Enzymol.* **374**, 300–321.
33. Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J. D., and Zardecki, C. (2002) The Protein Data Bank, *Acta. Crystallogr. D. Biol. Crystallogr.* **58**, 899–907.
34. Severson, W. E., Shindo, N., Sosa, M., Fletcher, T., 3rd, White, E. L., Ananthan, S., and Jonsson, C. B. (2007) Development and validation of a high-throughput screen for inhibitors of SARS CoV and its application in screening of a 100,000-compound library, *J. Biomol. Screen* **12**, 33–40.
35. Meyer, E. A., Castellano, R. K., and Diederich, F. (2003) Interactions with aromatic rings in chemical and biological recognition, *Angew. Chem. Int. Ed.* **42**, 1210–1250.
36. Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J., and 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.
37. Debela, M., Magdolen, V., Schechter, N., Valachova, M., Lottspeich, F., Craik, C. S., Choe, Y., Bode, W., and Goettig, P. (2006) Specificity profiling of seven human tissue kallikreins reveals individual subsite preferences, *J. Biol. Chem.* **281**, 25678–25688.
38. Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A., and Craik, C. S. (2000) Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7754–7759.
39. Harris, J. L., Peterson, E. P., Hudig, D., Thornberry, N. A., and Craik, C. S. (1998) Definition and redesign of the extended substrate specificity of granzyme B, *J. Biol. Chem.* **273**, 27364–27373.
40. Thiel, V., Ivanov, K. A., Putics, A., Hertzog, T., Schelle, B., Bayer, S., Weissbrich, B., Snijder, E. J., Rabenau, H., Doerr, H. W., Gorbalenya, A. E., and Ziebuhr, J. (2003) Mechanisms and enzymes involved in SARS coronavirus genome expression, *J. Gen. Virol.* **84**, 2305–2315.
41. Rawlings, N. D., Tolle, D. P., and Barrett, A. J. (2004) MEROPS: the peptidase database, *Nucleic Acids Res.* **32**, D160–164.
42. Boyd, S. E., Pike, R. N., Rudy, G. B., Whisstock, J. C., and Garcia de la Banda, M. (2005) PoPS: a computational tool for modeling and predicting protease specificity, *J. Bioinform. Comput. Biol.* **3**, 551–585.
43. Takeuchi, T., Shuman, M. A., and Craik, C. S. (1999) Reverse biochemistry: use of macromolecular protease inhibitors to dissect complex biological processes and identify a membrane-type serine protease in epithelial cancer and normal tissue, *Proc. Natl. Acad. Sci. U. S. A.* **96**, 11054–11061.
44. Menard, R., Therrien, C., Lachance, P., Sulea, T., Qo, H., Alvarez-Hernandez, A. D., and Roush, W. R. (2001) Cathepsins X and B display distinct activity profiles that can be exploited for inhibitor design, *Biol. Chem.* **382**, 839–845.
45. Palmer, J. T., Rasnick, D., Klaus, J. L., and Bromme, D. (1995) Vinyl sulfones as mechanism-based cysteine protease inhibitors, *J. Med. Chem.* **38**, 3193–3196.
46. Lee, T. W., Cherney, M. M., Huitema, C., Liu, J., James, K. E., Powers, J. C., Eltis, L. D., and James, M. N. (2005) Crystal structures of the main peptidase from the SARS coronavirus inhibited by a substrate-like aza-peptide epoxide, *J. Mol. Biol.* **353**, 1137–1151.
47. Hendlich, M., Bergner, A., Gunther, J., and Klebe, G. (2003) Relibase: design and development of a database for comprehensive analysis of protein-ligand interactions, *J. Mol. Biol.* **326**, 607–620.
48. Vandonselaar, M., Hickie, R. A., Quail, J. W., and Delbaere, L. T. (1994) Trifluoperazine-induced conformational change in Ca(2+)-calmodulin, *Nat. Struct. Biol.* **1**, 795–801.

BI0621415