Structural Basis for the Development of SARS 3CL Protease Inhibitors From a Peptide Mimic to an aza-Decaline Scaffold

Kenta Teruya,1 Yasunao Hattori,2 Yasuhiro Shimamoto,2 Kazuya Kobayashi,2 Akira Sanjoh,3 Atsushi Nakagawa,4 Eiki Yamashita,4 Kenichi Akaji2
1Department of Neurochemistry, Tohoku University Graduate School of Medicine, Aoba-Ku Sendai 980-8575, Japan
2Department of Medicinal Chemistry, Kyoto Pharmaceutical University, Yamashina-Ku, Kyoto 607-8412, Japan
3R&D Center, Protein Wave Co, Nara 631-0006, Japan
4Institute for Protein Research, Osaka University, Suita, Osaka 565-0871, Japan

Received 31 August 2015; revised 22 October 2015; accepted 2 November 2015
Published online 17 November 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bip.22773

ABSTRACT: Design of inhibitors against severe acute respiratory syndrome (SARS) chymotrypsin-like protease (3CLpro) is a potentially important approach to fight against SARS. We have developed several synthetic inhibitors by structure-based drug design. In this report, we reveal two crystal structures of SARS 3CLpro complexed with two new inhibitors based on our previous work. These structures combined with six crystal structures complexed with a series of related ligands reported by us are collectively analyzed. To these eight complexes, the structural basis for inhibitor binding was analyzed by the COMBINE method, which is a chemometrical analysis optimized for the protein–ligand complex. The analysis revealed that the first two latent variables gave a cumulative contribution ratio of $r^2 = 0.971$. Interestingly, scores using the second latent variables for each complex were strongly correlated with root mean square deviations (RMSDs) of side-chain heavy atoms of Met49 from those of the intact crystal structure of SARS-3CLpro ($r = 0.77$) enlarging the $S_2$ pocket. The substantial contribution of this side chain (~10%) for the explanation of pIC50s was dependent on stereochemistry and the chemical structure of the ligand adapted to the $S_2$ pocket of the protease. Thus, starting from a substrate mimic inhibitor, a design for a central scaffold for a low molecular weight inhibitor was evaluated to develop a further potent inhibitor. © 2015 Wiley Periodicals, Inc. Biopolymers (Pept Sci) 106: 391–403, 2016.

Keywords: SARS 3CL protease; inhibitor; aza-decaline; SARS 3CL protease; crystal structure complexed with inhibitor; aza-decaline scaffold; COMBINE analysis for inhibitor design, hydrophobic interaction

This article was originally published online as an accepted preprint. The “Published Online” date corresponds to the preprint version. You can request a copy of any preprints from the past two calendar years by emailing the Biopolymers editorial office at biopolymers@wiley.com.

INTRODUCTION

Although the primary epidemic of severe acute respiratory syndrome (SARS)1–3 was eventually brought under control, the recent identification of a SARS CoV (coronavirus)-like virus in Chinese bats4,5 indicates that there are natural
reservoirs for this virus. Since no effective therapy exists for these viral infections, developing anti-SARS agents against future outbreaks remains a formidable challenge. In the early stage of the infection, SARS 3CL protease (3CLpro) is a key enzyme to cleave polyproteins to yield functional polypeptides for replication.\(^6\)\(^,\)\(^7\) The 3CLpro is a cysteine protease containing a Cys-His catalytic dyad. Because of its functional importance in the viral lifecycle, 3CLpro is considered an attractive target for the structure-based design of drugs against SARS. Thus, numerous inhibitors of 3CLpro have been reported including peptide-mimics\(^8\)\(^–\)\(^12\) and small molecules derived from natural products,\(^13\)\(^–\)\(^15\) anti-viral agents,\(^16\)\(^,\)\(^17\) anti-malaria agents,\(^18\) or high throughput screening.\(^19\)\(^–\)\(^22\)

In the course of our own studies on SARS 3CLpro and its inhibitors,\(^23\) we found that mature SARS 3CLpro is sensitive to degradation at the 188Arg/189Gln site, and the Arg188Ile mutant is resistant to this degradation. As a result, this stable mutant protease makes it feasible for evaluating designed and synthesized inhibitors. Exploring useful inhibitors for SARS 3CLpro based on the substrate revealed that peptide aldehyde is a facile for designing an inhibitor for SARS 3CLpro. Through combinations of the systematic substitution of peptide aldehyde amino acid and the crystallographic analysis of the complexes, a leading inhibitor (Table I, entry 1) emerged. This peptide-based inhibitor has all of its side-chain structures differing from the substrate non-prime site sequence, Thr-Ser-Ala-Val-Leu-Gln, except at the

| Compound entry No. | Chemical Structure of Ligand | Complex PDB Code | pIC\(_{50}\) | Reference |
|--------------------|------------------------------|------------------|-----------|-----------|
| 1                  | ![Image]                     | 3AW0             | 5.244     | 8         |
| 2                  | ![Image]                     | 3AVZ             | 7.187     | 8         |
| 3                  | ![Image]                     | 3ATW             | 7.009     | 8         |
| 4                  | ![Image]                     | 4WY3             | 3.620     | 24        |
| 5                  | ![Image]                     | 4TWY             | 3.967     | 24        |
| 6                  | ![Image]                     | 4TWW             | 4.201     | 24        |
| 7                  | ![Image]                     | 5C5O             | 3.668     | This work |
| 8                  | ![Image]                     | 5C5N             | 3.561     | This work |
| 9                  | ![Image]                     | 3AW1             |          | 8         |
P<sub>3</sub> site where the side-chain was directed outward. Especially, substitution of Leu to cyclohexyl alanine (Cha) P<sub>2</sub> of the peptide inhibitor gave a higher increase in inhibitory potency to SARS 3CL pro (Table I, entries 1 and 2). Moreover, crystallographic analysis of the inhibitor complex provided an extended design of both deletion at the P<sub>5</sub> residue and mutation providing a hetero atom at the P<sub>4</sub> position of the inhibitor. Evaluation of IC<sub>50</sub> in the protease and crystallographic analysis revealed that this down size design was clearly realized without any substantial decrease in potency (Table I, entry 3).

Based on this wide range surveillance of the surroundings of the active sites by the peptide aldehyde inhibitor, a novel non-peptide inhibitor based on interactions at the S<sub>1</sub> and S<sub>2</sub> sites of SARS 3CL<sup>pro</sup> was designed and synthesized. In the study, focusing on the fact that P<sub>1</sub> and P<sub>2</sub> sites interact with the catalytic dyad and hydrophobic pocket, respectively, compounds were designed by connecting the P<sub>2</sub> site cyclohexyl group of the substrate-based inhibitor to the main-chain at N-acyl-parts of an aza-decaline based inhibitor via a methylene linker. This strategy gave aza-decaline as a central scaffold of the inhibitor with expected restricted conformational diversity of the N-acyl group. The designed compounds showed substantial inhibitory activity. By X-ray crystallographic studies (Figure 1), it was confirmed that the aza-decaline inhibitors were at the active site cleft of 3CL<sup>pro</sup> and the aza-decaline scaffold was inserted into a large S<sub>2</sub> pocket, while most of the S<sub>2</sub> pocket was occupied. Interestingly, the stereochemistry on the aza-decaline scaffold, (4aS,8aR) and (4aR,8aS), showed clear differences in inhibitory activities for SARS 3CL<sup>pro</sup> (Table I entries 4-6). Thus, starting from the substrate mimic peptide aldehyde as a leading compound, we have been developing nonpeptidic inhibitors by extracting fine structure–activity relationships from these crystallographic analyses. However, this stereostructure–activity relationship was not fully evaluated because of a shortage of samples for structural comparison.

The COMparative BI Nd ing Energy (COMBINE) method developed by Ortiz is a powerful pipeline for the decomposition of the ligand–receptor interaction energies into a series of terms, extracting latent variables (LV) for the explanation of the interactions and then, assigning weights to selected ligand–residue interactions using multivariate statistics and partial least square (PLS) analysis to correlate with a set of ligand–receptor complexes data on the experimental activities or binding affinities. Gil-Redondo and colleagues developed gCOMBINE, which is a user-friendly graphical interface, to perform the analysis with defined input files and sophisticated result presentations. Since the ligand-induced structural alternation is taken into account, ligand–receptor interactions can be better understood by chemometrical analysis. This method has been applied to various analyses of quantitative structure–activity relationships.

To get further insight into the relations between the configuration of the side chain of the aza-decaline scaffold and the structure of the complexes, we analyzed two new complexes in which the ligands were (4aS,8aR)- and (4aR,8aS)- aza-decaline scaffold. Then, these structures and our previously reported six
crystal structures of a complex with a series of ligands and one crystal structure without ligands were collectively referred to derive a quantitative model.

**MATERIALS AND METHODS**

In the present study, two inhibitors were newly synthesized and examined for their inhibitory activity. These inhibitors were complexed with SARS 3CLpro and the complexes were subjected to X-ray structural analysis. After these structural analyses, another six complex structures containing previously developed inhibitors were subjected to the gCOMBINE, a chemometric analysis. Table I shows the chemical structures, PDB IDs, and the properties of the inhibitors.

**Synthesis and Determination of Inhibitory Activity of Compounds 7 and 8**

Inhibitors 7 and 8 were synthesized based on the Scheme 1 as previously described with modifications. Briefly, starting from an enantiomixture of alcohol a, previously reported product$^{24}$ by Diels-Alder reaction, the primary alcohol was converted to azide and then the amino group (b). Amine b was condensed with N-Boc-N-phenyl β-alanine (Ph-(Boc)bAla) to afford an enantio-mixture of compound c, a precursor of the cyclization reaction. Amide c, in which the nitrogen of the amide group located at 5-carbon apart from olefin, was cyclized with a Pd-based catalysis.$^{42}$ The obtained enantio-mixture of an aza-decaline derivative, olefin d was converted to diol and then to aldehyde by Lemieux-Johnson oxidation. The resulting aza-decaline-deriv aldehyde and H-His(Trt)-N(CH$_3$)OCH$_3$ were condensed via reductive amination to give a diastereo-mixture of compound e. This diastereo-mixture was separated and purified by flash column chromatography into 7e and 8e corresponding in stereochemistry to 7 and 8, respectively. Trt and Boc groups of 7e and 8e were removed and purified by a silica column. Finally, Weinreb amides were converted to aldehyde and then purified by RP-HPLC. Yields were 3% and 1% for 7 and 8, respectively, based on the enantio-mixture of olefin d.

For measurements of IC$_{50}$s of compounds 7 and 8, peptide substrate, H-Thr-Ser-Ala-Val-Leu-Gln-Ser-Gly-Phe-Arg-Lys-NH$_2$, (111 µM) in a reaction solution (25 µL of 20 mM Tris-HCl buffer pH 7.5 containing 7 mM DTT) was incubated with the R188I SARS 3CLpro (56 nM)$^{25}$ at 37°C for 60 min in the presence of various inhibitor concentrations at 37°C for 60 min. The cleavage reaction was monitored by analytical HPLC, a linear gradient of CH$_3$CN (10-20%) in
Development of SARS 3CL Protease Inhibitors

395

X-Ray Crystallography

The purified SARS 3CLpro in 20 mM Bis-Tris pH 5.5, 10 mM NaCl, and 1 mM DTT was concentrated to 8 mg/mL. Crystals of SARS 3CLpro were grown at 4°C using a sitting-drop vapor diffusion method by mixing it with an equal volume of reservoir solution containing 100 mM MES pH 6.2, 5-10% PEG20000, and 5 mM DTT. Cubic-shaped crystals with dimensions of 0.3 mm × 0.3 mm × 0.3 mm grew within 3 days. The crystals were then soaked for 24 hours with a reservoir-based solution of 100 mM MES pH 6.2, 5-8% PEG20000, and 5 mM DTT containing 3 mM of 7 or 8. Crystals were then transferred into a cryoobuffer of 100 mM MES pH 6.2, 10% PEG20000, 5 mM DTT, 15% ethylene glycol containing 3 mM of 7 or 8, and flash-frozen in a nitrogen stream at 100 K. X-ray diffraction data of SARS 3CLpro in complexes with inhibitor 7 or 8 was collected with a SPring-8, beamline BL44XU with a Rayonix MX300HE CCD detector at a wavelength of 0.900 Å. The structures of SARS 3CLpro in complex with inhibitors were determined by molecular replacement using the Molrep program with a R188I SARS 3CLpro structure (PDB code 3AW1) as the search model. Rigid body refinement and subsequent restrained refinement protocols were performed with the program Refmac 5.4 The Coot program was used for manual model rebuilding. Water molecules were added using Coot only after refinement of the protein structures had converged. Ligands generated on ILigand software were directly built into the corresponding electron density electron, and the model was then subjected to an additional round of refinement. The figures of structural representation in this paper were generated on Pymol or UCSF-Chimera software. Crystallographic data and results of refinement are summarized in Table II.

gCOMBINE Analysis

PDB files of the complex (corresponding to PDB ID: 3AW0, 3AWZ, 3ATW, 4TWY, 4TWZ, 4W3Y, 5C5O and 5C5N) were prepared as containing a single complex with a ligand. The complexes were imposed on the structure of SARS 3CLpro without a ligand (PDB ID: 3AW1) on the Match-Maker program based on the main chain Cz of the protease. Each complex was protonated and minimized under an amberFF4SB force field powered by MMTK and inhibitors were assigned AM1-BCC charges and fully optimized at the AM1 level using the MOPAC program on UCSF-Chimera. The proteases and inhibitors were saved separately.

The parametrization for the gCOMBINE was performed on the tLEAP module in AMBERTOOL14 programs. The PDB files of the receptor protease were cleaned by pdb2lambert. The antechamber suite has been developed to be used with the general AMBER force field for small molecules. The PDB files of the inhibitor were converted to mol2 files by antechamber, and fs commands and lib files for each inhibitor were generated by parmchk. Each file of the protease and inhibitor was combined and neutralized for each complex, parameter and topology files were generated.

The COMBINE program was used to decompose the interaction energy between the inhibitor and protein in each complex. That is, this program was used to calculate the Lennard-Jones and electrostatic interactions between the inhibitor and each protein residue on a per residue basis. gCOMBINE is a graphical interface that runs COMBINE. The gCOMBINE automatically arranges and constructs an X matrix in which the rows represent the different compounds studied, and the columns contain the residue-based energy information, which is separated into two blocks (van der Waals and electrostatic), plus an additional column (Y matrix) containing the experimental binding affinities. This X matrix was then projected onto a small number of orthogonal IVs using partial least-squares (PLS) analysis, and the original energy terms were given weights, according to their importance in the model, in the form of PLS pseudo coefficients. The higher these coefficients are, the more significant they are for explaining the variance in the experimental data. Thus, in this study, the van der Waals interactions, and the electrostatic interactions, between the inhibitor and each protein residue were selected to estimate the pIC50 value. The variables that were unimportant for activities were discarded, and the remaining variables were used to build...
Table II  Data Collection and Refinement Statistics for 3CL\textsuperscript{pro} and Its Inhibitor in Complexes With Compounds 7 and 8

| PDB ID | 5C5O | 5C5N |
|--------|------|------|
| Inhibitor name | 7    | 8    |
| Space group | C\textsubscript{121} | P\textsubscript{1} |
| Unit cell parameters | | |
| Length a | 107.80 | 54.66 |
| Length b | 82.09 | 58.66 |
| Length c | 53.27 | 68.10 |
| Angle α | 90 | 93.73 |
| Angle β | 104.35 | 103.40 |
| Angle γ | 90 | 106.51 |
| Resolution | 1.65 | 1.50 |
| Observations | | |
| Unique observations | 52746 | 119208 |
| Redundancy | 4.0 | 2.2 |
| Completeness | 84.8 | 95.0 |
| Mean I/σ(I) | 2.41 (at 1.65 Å) | 2.36 (at 1.50 Å) |
| R\text{merge} | 0.08 | 0.06 |
| Resolution range | 37.59–1.69 | 30.2–1.50 |
| R\text{cryst} | 0.29 | 0.25 |
| R\text{free} | 0.32 | 0.27 |
| RMSZ from ideal | | |
| Bond length | 0.87 | 0.99 |
| Bond angle | 0.92 | 1.01 |

the final PLS model. As a result of COMBINE analysis, gCOMBINE provided a summary of PLS weights, coefficients, rank, and files for PDB with coefficients. In the present study, the PLS ranks were calculated by 5.

RESULTS AND DISCUSSION

Design, Preparation, and Inhibitory Activities of Inhibitor 7 and 8

The compounds 7 and 8 were designed based on our previous structural analysis of 3CL\textsuperscript{pro} compounds, 1, 2, 3, 4, 5, and 6. Figure 1 shows a summary of our previous structural analysis of the complex with 1-6. The chemistry of the aza-decaline derived inhibitors including 7 and 8 was divided into three portions: histidine aldehyde (His-al), aza-decaline, and an N-acyl part (Figure 1B). The His-al part took an almost identical conformation and similarly interacted with the 3CL\textsuperscript{pro} over the complexes examined. Briefly, His-al is located at the active center of 3CL\textsuperscript{pro} as aldehyde and an imidazole ring interacted with the catalytic dyad, thiol of Cys, and imidazole of His, respectively. Thus, the His-al part functioned as a pharmacophore in these compounds. In contrast, the N-acyl part of the aza-decaline based inhibitors, 4, 5, and 6 were directed outward of 3CL\textsuperscript{pro} and then opposite to the P\textsubscript{3} to P\textsubscript{4} sites of peptide mimic inhibitors, 1, 2, and 3. Based on these observations, compounds 7 and 8 were designed as making an aromatic ring of the N-acyl part feasible to access the S\textsubscript{3} or S\textsubscript{4} pocket of 3CL\textsuperscript{pro} via a relatively flexible β-alanine linker and are expected to provide greater potency. The stereochemistry of compounds 7 and 8 corresponded to compounds 4 and 5, respectively.

As for the template for the synthesis of the designed compounds, aza-decaline is an attractive scaffold which is restricted in configuration and can provide distinct and relative positions and directions with substituents. Considering this ability of extension, among the three parts of the basic inhibitor design, an aza-decaline was constructed by two cyclization steps utilizing different chemistries, Diels-Alder and diastereo-selective cyclization by PdCl\textsubscript{2}(AcCN)\textsubscript{2} for cyclohexyl and aza-cyclohexyl on the cyclohexyl, respectively. Since the second cyclization is feasible to react with a wide variety of substrates, the acylation is scheduled in the early steps. On the other hand, since the His-al part contains a labile group, aldehyde, the reductive amination is scheduled in the late steps. In the present study, the synthetic scheme worked well at least for synthesis of inhibitors 4-8. Production of a wide variety of inhibitors containing various stereochemistries is to be examined elsewhere based on these key reactions.\textsuperscript{42,59,60}

Although compounds 7 and 8 were designed based on the structural analysis mentioned earlier, the IC\textsubscript{50s} were not improved in comparison with those of the corresponding compounds 4 and 5, respectively. Compound 7 and 8 inhibited hydrolysis of SARS 3CL\textsuperscript{pro} in IC\textsubscript{50} 275 μM and 215 μM, respectively.

Structural Analysis of R188I SARS 3CL Protease Complexed with Compounds 7 and 8

To investigate the reasons for the deterioration in IC\textsubscript{50}, crystallographic analyses of the SARS 3CL\textsuperscript{pro} in complexes with compounds 7 and 8 were performed. Data collection and refinement statistics for the R188I SARS 3CL protease with compounds 7 and 8 are summarized in Table II.

The overall structure of 3CL\textsuperscript{pro} in complex with inhibitors 7 and 8 was similar to the other aza-decaline-based inhibitors (Figure 2A). Briefly, the aldehyde group and imidazole ring of His-al of these inhibitors took an almost identical conformation and similarly interacted with 3CL\textsuperscript{pro}. In contrast, the direction of the aza-substituents of the aza-decaline based inhibitors, 7 and 8 (Figure 2B), were outward from 3CL\textsuperscript{pro} and opposite to the P\textsubscript{3} to P\textsubscript{4} sites of peptide mimic inhibitors 1, 2, and 3 as in our previous inhibitors, 4, 5, and 6 (Figure 1A). Thus, crystallographic analyses of the SARS 3CL\textsuperscript{pro} in complexes with compounds 7 and 8 revealed that in our case the bAla was not functioning as a flexible linker to reach the Ph group of the
hydrophobic S3 pocket. These features were in good accordance with the deteriorated IC50s of 7 and 8, which may be due to a higher entropic discordance of 7 and 8 than of 4 and 5, respectively.

The trans aza-decaline scaffolds of 7 and 8 were adapted in the S2 pocket as deep as 4 and 5, respectively. Based on these structural analyses focusing on the chemical structure of the inhibitors, several points to increase potency were revealed. The aza-decaline structure is sufficient to be adapted in the S2 pocket, and stereochemistry on the aza-decaline is a determinant for orientation in the S2 pocket (Figure 3). Interestingly, the orientation of aza-decaline in the S2 pocket was not critical but was a determinant for relative potency; 4 and 7 were better than 5 and 8, respectively, in which these scaffolds almost crossed perpendicular to each other when complex structures were superimposed by the whole protein. Thus, the trans aza-decaline scaffold is suitable for embedding into the S2 pocket.

In comparison with peptide aldehyde inhibitors (Figure 4), to make the N-acyl-group of aza-decaline based inhibitors feasible to interact with the S3 and S4 sites, regioisomers such as 3,3-disubstituent or 3,4-substituted trans aza-decaline may be necessary.

Paasche et al.61 have pointed out that the low inhibition potencies of known covalently interacting inhibitors may, at least in part, be attributed to insufficient fostering of the proton-transfer reaction based on MM/MQ analysis of SARS 3CLpro. Unfortunately, our method of study was incapable of accessing this charge state. Cleary and colleagues62 showed that acetal or hemiacetal formation can be one of the components of the inhibitory potency for chymotrypsin. Among the eight complexes used in this study, structural refinements resulted in distances between the carbonyl carbon and sulfide ranging within 1.70–2.54 Å, which included distances judged as chemical bond formation by modeling software. There was, however, no correlation between this distance and potency ($r = -0.16$).

Quantitative Structure–Activity Relationship of R188I SARS 3CL Pro and Inhibitors

To investigate the structural basis of SARS 3CLpro based on our results, a chemometrical analysis was applied to the crystallographic structures of the complex with inhibitors 1-8. The ligands, including peptide aldehyde, 1, 2, and 3, which are large enough to cover nonprime sites from P1 to P5 and two pairs of diastereoisomers of trans aza-decaline 4 and 7 and 5 and 8, respectively, were expected to give fine structural diversity around the active site. The inputs for the COMBINE program are structural data including topology files, coordination files, and pharmacological activity correlated with ligand binding. In this study, the IC50 values were converted to negative logarithmic values, pIC50s, which correlated with the binding free energy of the inhibitors. In Table I, values ranging from 3.56 - 7.19 are listed with the inhibitor and the ligand chemical structure.

Based on the input static structures, the program computes residue-based van der Waals and electrostatic interactions.
between the residues in each complex and ligand for each complex (Figure 5). Ligand-binding pockets from S₁ to S₄ were clearly detected. Some of the pockets were located over the structural domains. The figure revealed that S₁ and S₂ contribute to inhibitor binding more than other pockets. Extensively, even in the region of S₁, in which all the examined inhibitors interacted with an identical His aldehyde structure, the calculated energies were not identical. This result illustrates that substitutions can affect fine interactions of the pharmacophore. These energy descriptors for each complex and the corresponding pIC₅₀ comprise the X and Y matrix for the gCOMBINE analysis, respectively. The COMBINE program does PLS analysis such that the composed X matrix is then projected onto a small number of orthogonal LVs in an optimized manner for protein–ligand interaction.

Table III shows a chemometric summary of the analysis for the indices of 5 LVs. The optimal dimensionality of the PLS models was determined by monitoring the cross-validation indexes as a function of the number of LVs extracted. The cross-validation procedure employed the leave-one-out method. The predictive ability of the resulting models was reported by both the cross-validated correlation coefficient ($q^2$) and the standard deviation of error in the predictions. The $q^2$ value served as the criterion to determine the optimal dimensionality of the PLS model. Thus, for further analysis, we decided to take ranks by 2 as our model. At this rank, the linear regression analysis reached $r^2 = 0.97$ (correlation coefficient). Similar to the primary component analysis, this indicates that the first and second LVs contribute 87% and 10% of the explanation of pIC₅₀ in this system, respectively. These results indicate that the first two LVs can explain the pIC₅₀ of complexes to a considerable extent.

The COMBINE program gave several files corresponding to each rank of LV. In fact, for each complex, predicted pIC₅₀ by
the first two LVs and experimental pIC$_{50}$ were in good correlation and clearly correlated with the corresponding experimental pIC$_{50}$ as shown in Figure 6. As a consequence, these 8 complexes composed a wide range of pIC$_{50}$s in the order of 3.5.

$P$ matrix (X-loading) in PLS to each LV reflects which and how much residue in a protein is contributing. Figure 7 shows a scatter plot of the part of the residues composing the first and second LVs in the horizontal and vertical axes, respectively. Among these elements which reflect the features of the first two LVs, the top 6 contributing residues for interaction were Pro$^{168}$, Glu$^{166}$, Met$^{165}$, Leu$^{167}$, Tyr$^{161}$, and Gln$^{192}$. On the other hand, Met$^{49}$ prominently contributed a large portion of the second LV.

For the first LV, residues mainly composing the S1 pocket and residues covering a wide range of binding sites were extracted. Thus, the first LV can be interpreted as it stands. Briefly, the first LV for inhibition reflects the fact that the pharmacophore interacts with active site fastening, and parts other than the pharmacophore interact with a wide range of binding sites as seen with peptide mimic inhibitors.

![FIGURE 5](image1)

**FIGURE 5** Interactions of 3CL$^{pro}$ and the inhibitor in each complex by residual resolution. Vertical and horizontal axes represent the computed energy of van der Waals (upper graph) and electrostatic (lower graph) interactions and the residue number of 3CL$^{pro}$, respectively. At the top, the schematic drawing of the secondary structure of 3CL$^{pro}$ along with the residue number is presented.

![FIGURE 6](image2)

**FIGURE 6** Scatter plot comparing experimental vs. predicted activities in COMBINE models for the eight complexes. The two LV model is represented. The straight line is (the diagonal) included in this plot as reference.

| Numbers of LVs | $r^2$  | $q^2$  |
|----------------|--------|--------|
| 1              | 0.863  | 0.602  |
| 2              | 0.971  | 0.632  |
| 3              | 0.993  | 0.577  |
| 4              | 1.000  | 0.570  |
| 5              | 1.000  | 0.571  |

Table III Extracted Statistical Summary of Chemometrical Analysis of 3CL$^{pro}$ Complexes

*Biopolymers (Peptide Science)*
Next, the S2 pocket was highlighted in the examining residues related with the second LV. Among these residues, it was revealed that configuration of the dominant Met\textsuperscript{49} is connected to the inhibitory effect of ligand binding. Figure 8A shows that the active site superimposed on the total 3CL\textsuperscript{pro} structures 1-9. Structure 9 contains no inhibitor. The catalytic Cys-His dyads were well overlapped. On the other hand, the configuration of Met\textsuperscript{49} side chains was separated into two groups. In one group, no ligand (9), Leu side chain (1), and (3R,4aS,8aR) aza-decaline (4 and 8) were involved. In the other group, the Cha side chain (2 and 3) and (3S,4aR,8aS) aza-decaline (5, 6 and 7) were involved. Inhibitors in the latter group had greater potency in comparison with the corresponding chemical structures. Thus, the S2 pocket was enlarged on the potent inhibitors by reorientation of the Met\textsuperscript{49} side chain. Figure 8B shows the correlation between the X score of each complex corresponding to the second LV model and each RMSD of the Met\textsuperscript{49} side chain heavy atom of complexes from that of structure 9. These values were strongly correlated with \( r = 0.77 \). Thus, the main contribution to the second LV was the orientation of the Met\textsuperscript{49} side chain followed by enlarging the S2 pocket. The isobutyl group of the Leu side chain was not sufficient for inducing this reorientation. Even if a large functional group such as aza-decaline occupied the S2 pocket, the angle between the plane consisting of the chair-chair configuration of aza-decaline and to Met\textsuperscript{49} was critical for the configurational change, which can be controlled by the stereochemistry of position 3 of the aza-decaline scaffold. These analyses revealed that the aza-decaline scaffold with stereochemistry of S at position 3 is a novel scaffold for designing an inhibitor to 3CL\textsuperscript{pro}. As mentioned in inhibitor-based analysis, optimized regioisomers need to be elucidated. The key feature of the interaction in the S2 pocket gives a novel strategy for developing a new inhibitor.

It is an important and unique strategy in COMBINE that multiple structures are involved in the analysis. This COMBINE strategy may, at least in part, represent the flexibility\textsuperscript{65-67} of the protein upon ligand binding. In this research, not protein dynamic structural changes, but configurational change of side chains by inhibitor binding and their substantial control of inhibitory potency were detected. This fact highlights the advantages of the COMBINE method. Together with ligand-based analysis and the COMIBINE analysis focusing on
protein structure, a peptide mimetic inhibitor and development of a small molecular weight molecule is feasible. These sets of training on COMBINE will be adapted to modeled complexes such as docking simulations.

In conclusion, peptidic protease inhibitors were able to increase potency when designed based on structural analysis at each designing step. For aza-decaline based inhibitors, the potencies decreased during design. Structural analysis revealed that orientation of the substituent on the aza-decaline scaffold was not suitable for interacting with the protein. The COMBINE analysis revealed that the volume of the S2 pocket perturbed by stereochemistry of the ligand is important for inhibitor potency. This is derived as the second LV (10% explanation of the X matrix). Further design and synthesis of inhibitors are expected.

PDB ID Codes of Entry for This Article
5C5O and 5C5N

The authors thank Ms. C. Teruya of Kyoto Pharmaceutical University for obtaining EI Mass spectra.

REFERENCES
1. Lee, N.; Hui, D.; Wu, A.; Chan, P.; Cameron, P.; Joynt, G. M.; Ahuja, A.; Yung, M. Y.; Leung, C. B.; To, K. F.; Lui, S. E.; Szeto, C. C.; Chung, S.; Sung, J. J. Y. New Engl J Med 2003, 348, 1986–1994.
2. Drosten, C.; Günther, S.; Preiser, W.; van der Werf, S.; Brodt, H. R.; Becker, S.; Rabenau, H.; Panning, M.; Kolesnikova, L.; Fouchier, R. A. M.; Berger, A.; Burguière, A. M.; Cinatl, J.; Eickmann, M.; Escriou, N.; Grywna, K.; Kramme, S.; Manuguerra, J. C.; Müller, S.; Rickerts, V.; Stürmer, M.; Vieth, S.; Klenk, H. D.; Osterhaus, A. D. M. E.; Schmitz, H.; Doerr, H. W. New Engl J Med 2003, 348, 1967–1976.
3. Ksiazek, T. G.; Erdman, D.; Goldsmith, C. S.; Zaki, S. R.; Peret, T.; Emery, S.; Tong, S.; Urbani, C.; Comer, J. A.; Lim, W.; Rollin, P. E.; Dowell, S. E.; Ling, A. E.; Humphrey, C. D.; Shiieh, W. J.; Guarner, J.; Paddock, C. D.; Rota, P.; Fields, B.; DeRisi, J.; Yang, J. Y.; Cox, N.; Hughes, J. M.; LeDuc, J. W.; Bellini, W. J.; Anderson, L. J New Engl J Med 2003, 348, 1953–1966.
4. Li, W.; Shi, Z.; Yu, M.; Ren, W.; Smith, C.; Epstein, J. H.; Wang, H.; Cramer, G.; Hu, Z.; Zhang, H.; Zhang, J.; McEachern, J.; Field, H.; Daszak, P.; Eaton, B. T.; Zhang, S.; Wang, L. F. Science 2005, 310, 676–679.
5. Lau, S. K. P.; Woo, P. C. Y.; Li, K. S. M.; Huang, Y.; Tsoi, H. W.; Wong, B. H. L.; Wong, S. S. Y.; Leung, S. Y.; Chan, K. H.; Yuen, K. Y. P. Proc Natl Acad Sci USA 2005, 102, 14040–14045.
6. Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J. R.; Hilgenfeld, R. Science 2003, 303, 1763–1767.
7. Fan, K.; Wei, P.; Feng, Q.; Chen, S.; Huang, C.; Ma, L.; Lai, B.; Pei, J.; Liu, Y.; Chen, J.; Lai, L. J Biol Chem 2004, 279, 1637–1642.
8. Akaji, K.; Konno, H.; Mitsuji, H.; Teruya, K.; Shimamoto, Y.; Hattori, Y.; Ozaki, T.; Kusunoki, M.; Sanjoh, A. J Med Chem 2011, 54, 7962–7973.
9. Thanigaimalai, P.; Konno, S.; Yamamoto, T.; Koivai, Y.; Taguchi, A.; Takayama, K.; Yakushiji, F.; Akaji, K.; Kiso, Y.; Kasahara, Y.; Chen, S. E.; Naser-Tavakolian, A.; Schön, A.; Freire, E.; Hayashi, Y. Eur J Med Chem 2013, 65, 436–447.
10. Chuck, C. P.; Chen, C.; Ke, Z.; Chi-Cheng Wan, D.; Chow, H. F.; Wong, K. B. Eur J Med Chem 2013, 59, 1–6.
60. Crane, S. N.; Black, W. C.; Palmer, J. T.; Davis, D. E.; Setti, E.; Robichaud, J.; Paquet, J.; Oballa, R. M.; Bayly, C. I.; McKay, D. J.; Somoza, J. R.; Chauret, N.; Seto, C.; Scheigetz, J.; Wesolowski, G.; Massé, F.; Desmarais, S.; Ouellet, M. J Med Chem 2006, 49, 1066–1079.
61. Paasche, A.; Zipper, A.; Schäfer, S.; Ziebuhr, J.; Schirmeister, T.; Engels, B. Biochemistry 2014, 53, 5930–5946.
62. Cleary, J. A.; Doherty, W.; Evans, P.; Malthouse, J. P. G. BBA-Proteins Proteom 2014, 1844, 1119–1127.
63. Cheng, H. C. J Pharmacol Toxicol 2001, 46, 61–71.
64. Cheng, H. C. Pharmacol Res 2004, 50, 21–40.
65. Pastor, M.; Pérez, C.; Gago, F. J Mol Graph Model 1997, 15, 364–371.
66. Mou, T. C.; Gille, A.; Suryanarayana, S. R. Mol Pharmacol 2006, 70, 878–886.
67. Tomic, S.; Nilsson, L.; Wade, R. C. J Med Chem 2000, 43, 1780–1792.