Structural basis for specific recognition of substrates by sapovirus protease

Masaru Yokoyama[1‡], Tomoichiro Oka[2,3‡], Hirotatsu Kojima[4], Tetsuo Nagano[4], Takayoshi Okabe[4], Kazuhiko Katayama[2], Takaji Wakita[2], Tadahito Kanda[1†] and Hironori Sato[1]*

[1] Pathogen Genomics Center, National Institute of Infectious Diseases, Tokyo, Japan
[2] Department of Virology II, National Institute of Infectious Diseases, Tokyo, Japan
[3] Food Animal Health Research Program, Ohio Agricultural Research and Development Center, Department of Veterinary Preventive Medicine, The Ohio State University, Wooster, OH, USA
[4] Open Innovation Center for Drug Discovery, The University of Tokyo, Japan

Sapovirus (SaV) protease catalyzes cleavage of the peptide bonds at six sites of a viral polyprotein for the viral replication and maturation. However, the mechanisms by which the protease recognizes the distinct sequences of the six cleavage sites remain poorly understood. Here we examined this issue by computational and experimental approaches. A structural modeling and docking study disclosed two small clefts on the SaV protease cavity that allow the stable and functional binding of substrates to the catalytic cavity via aromatic stacking and electrostatic interactions. An information entropy study and a site-directed mutagenesis study consistently suggested variability of the two clefts under functional constraints. Using this information, we identified three chemical compounds that had structural and spatial features resembling those of the substrate amino acid residues bound to the two clefts and that exhibited an inhibitory effect on SaV protease in vitro. These results suggest that the two clefts provide structural base points to realize the functional binding of various substrates.

Keywords: sapovirus protease, substrate recognition, P1 and P4 amino acid residues, 3-D models, amino acid diversity, mutagenesis, 3-D pharmacophore, inhibitor screening

INTRODUCTION

Sapovirus (SaV) is a non-enveloped RNA virus that belongs to the family Caliciviridae and causes gastroenteritis in humans and swine (Chiba et al., 1979, 2000; Guo et al., 1999; Hansman et al., 2007). The SaV genome is a single-stranded RNA that encodes two or three open reading frames (ORFs; Liu et al., 1995; Noel et al., 1997; Numata et al., 1997; Guo et al., 1999; Robinson et al., 2002). The ORF1 encodes six non-structural proteins (NS1, NS2, NS3, NS4, NS5, and NS6-7) and a structural protein, the capsid protein (VP1; Oka et al., 2006, 2009). The NS6-7 protein contains the chymotrypsin-like protease domain (the 3C-like protease; Oka et al., 2005a,b, 2007; Robel et al., 2008) and the RNA-dependent RNA polymerase domain (the 3-D-like polymerase; Fullerton et al., 2007; Bull et al., 2011). The ORF1 precursor protein is post-translationally cleaved at six sites by the 3C-like protease (Oka et al., 2005b, 2006).

The SaV 3C-like protease domain comprises 146 amino acid residues (Oka et al., 2007). This enzyme cleaves the peptide bonds of specific dipeptides, such as the glutamic acid/glycine (E/G), glutamine/glycine (Q/G), and glutamic acid/alanine (E/A; Oka et al., 2006). However, these dipeptide motifs exist in the non-cleaved

sites of the ORF1 polyprotein, indicating that additional amino acid residues are required for the specific recognition of substrates (Oka et al., 2006). In this regard, calicivirus proteases have a large cavity that can accommodate substrate peptides with several amino acid in lengths (Nakamura et al., 2005; Zeitler et al., 2006; Oka et al., 2007). It is conceivable that these substrate amino acid residues around the cleavage sites, termed the P4, P3, P2, P1, P1′, P2′, P3′, and P4′ sites, are all involved to some extent, either directly or indirectly, in the recognition and cleavage by protease. However, there must be a division of roles: previous studies on the calicivirus proteases consistently suggest more extensive involvement of the substrate amino acid residues upstream of the peptide bond of the cleavage sites in the cleavage by proteases (Wirblich et al., 1995; Sosnovtsev et al., 1998; Hardy et al., 2002; Belliot et al., 2003; Scheffler et al., 2007; Robel et al., 2008). In the case of SaV, the substrate P1 and P4 amino acid residues in particular are physicochemically more conserved among different SaV strains (Oka et al., 2009) and more sensitive to the substitutions (Robel et al., 2008; Oka et al., 2009). Therefore, these amino acid residues may provide the specific contact sites with SaV protease. However, due to the lack of structural

information on SaV protease and its substrates, such interaction remains unclear.

Recent advances in the hardware and software for biomolecular simulation and bioinformatics have rapidly improved the precision and performance of these techniques. We have applied some of these techniques, in combination with experimental methods, to understand the structural and evolutionary basis of the virological phenomena (Oka et al., 2007, 2009; Motomura et al., 2008, 2010; Naganawa et al., 2008; Shirakawa et al., 2008; Yokoyama et al., 2010, 2012; Ode et al., 2011; Sakuragi et al., 2012). In this study, by combining methods of homology modeling, the automated ligand docking, Shannon entropy analysis, site-directed mutagenesis, and *in silico* screening of SaV inhibitors, we studied the structural basis for the substrate recognition by SaV protease.

MATERIALS AND METHODS
STRUCTURAL MODELING OF SaV PROTEASE DOCKED TO THE SUBSTRATE OCTAPEPTIDES
We first constructed a ligand-free protease domain model of the SaV Mc10 strain (Oka et al., 2005b; GenBank accession number: AY237420) by the homology modeling technique (Sanchez et al., 2000; Baker and Sali, 2001) as described previously (Oka et al., 2007). The modeling was performed using tools available in the Molecular Operating Environment (MOE; Chemical Computing Group, Inc., Montreal, QC, Canada). As the modeling template, we used the high-resolution crystal structure of norovirus 3C-like protease at a resolution of 1.50 Å [Protein Data Bank (PDB) code: 2FYQ; Zeitler et al., 2006] because, like SaV, the norovirus belongs to the family Caliciviridae, and thus the protease sequence shows a higher identity to the SaV protease sequences (about 25% identity) than to the other available 3C-like protease sequences of viruses. We applied the multiple sequence alignment approach (Baker and Sali, 2001) using the reported 3C-like proteases to minimize misalignments of the target and template sequences, as described previously (Oka et al., 2007; Shirakawa et al., 2008). The sequences used for the alignment included those of the rhinovirus 3C-like protease (PDB code: 1CQQ; Matthews et al., 1999), the poliovirus 3C-like protease (PDB code: 1L1N; Mosimann et al., 1997), and the hepatitis A virus 3C-like protease (PDB code: 1QA7; Bergmann et al., 1999). The alignment was done with the alignment tool MOE-Align, and homology modeling was done with the tool MOE-Homology in MOE. We optimized the 3-D model thermodynamically via energy minimization using the MOE and an AMBER99 force field (Ponder and Case, 2003). We further refined the physically unacceptable local structure of the models based on a Ramachandran plot evaluation using MOE. The 3-D models of the six octapeptides corresponding to the six cleavage sites of the ORF1 precursor protein of the SaV Mc10 strain (NS1/NS2, NS2/NS3, NS3/NS4, NS4/NS5, NS5/NS6-7, and NS6-7/VP1) were constructed using the Molecular Builder module in MOE. Subsequently, the thermodynamically and physically optimized protease models were used to construct protease-substrate complex models. Individual octapeptide models were docked with the optimized SaV protease domain model described above, using the automated ligand docking program ASEDock2005 (Goto et al., 2008) operated in MOE as described previously (Yokoyama et al., 2010). Default setting in ASEDock2005 was applied for the search of

the candidate docking structures, and the structures with the best docking score expressed by the arbitrary docking energy (U_{dock}) in ASEDock2005 (Kataoka and Goto, 2008) were selected for the analysis of the protease-substrate interaction sites.

ANALYSIS OF AMINO ACID DIVERSITY WITH INFORMATION ENTROPY
The amino acid diversity at individual sites of the SaV protease domain was analyzed with Shannon entropy scores as described previously (Sander and Schneider, 1991; Mirny and Shakhnovich, 1999; Oka et al., 2009). The amino acid sequences of the protease domain of various human SaV strains from different geographic regions in the world were obtained from GenBank (the number of sequences is 19; accession numbers: X86560, AY694184, AY237422, AY237423, AY646853, AY646854, AJ249939, AY237420, AY237419, AY646855, AY603425, AJ786349, DQ058829, DQ125333, AY646856, DQ125334, DQ366344, DQ366345, DQ366346). The amino acid diversity within the SaV protease population was calculated using Shannon's formula (Shannon, 1948):

$$H(i) = - \sum_{x_i} p(x_i) \log_2 p(x_i) \quad (x_i = G, A, I, V, \ldots),$$

where $H(i)$, $p(x_i)$, and i indicate the amino acid entropy (H) score of a given position, the probability of occurrence of a given amino acid at the position, and the number of the position, respectively. An H score of zero indicates absolute conservation, whereas 4.4 bits indicates complete randomness. The H scores were displayed on the 3-D structure of the SaV protease model constructed above.

We also calculated the Shannon entropy by considering the physicochemical properties of amino acid residues, i.e., the chemical properties and size of side chains as described previously (Oka et al., 2009). For analysis of the diversity in the chemical properties, the amino acid residues were classified into seven groups: acidic (D,E), basic (R,K,H), neutral hydrophilic (S, T, N, Q), aliphatic (G, A, V, I, L, M), aromatic (F, Y, W), thio-containing (C), and imine (P). For analysis of the diversity in the size of side chains, the amino acid residues were classified into four groups: small (G, A, C, S), medium-small (T, V, N, D, I, L, P, M), medium-large (Q, E, R, K), and large (H, F, Y, W). The H scores were plotted on the 3-D structure of the SaV protease model.

SITE-DIRECTED MUTAGENESIS OF THE SaV PROTEASE DOMAIN
The detailed strategy of the mutagenesis for the SaV protease domain has been described previously (Oka et al., 2005b, 2006). Briefly, we used the full-length cDNA clone of the genome of the SaV strain Mc10 (pUC19/SaV Mc10 full-length; GenBank accession number: AY237420) as a starting material for the mutagenesis. We constructed nine SaV Mc10 full-length mutant cDNA clones. Site-directed mutagenesis was performed using a GeneTailor Site-Directed Mutagenesis System (Invitrogen). The oligonucleotides used for the site-directed mutagenesis were as follows (the codons corresponding to changed amino acid(s) are indicated in lowercase): for T1085A, 5′-GTGGTTGTCACAGTTgcaCACGTGGCCTCTGCG-3′; for Y1156A, 5′-ATCACGGTCCAGGGGgctCACCTGCGCATC

ATA-3′; for K1167A, 5′-ATGGATACCCAACAGcgCGTGGGGGACT
GTGGCAC-3′; for R1168A, 5′-GATACCCAACAAAGgctGGGGGAC
TGTGGCACAC-3′; for K1167E, 5′-ATGGATACCCAACAgagCGT
GGGGACTGTGGCAC-3′; for R1168E, 5′-ATGGATACCCAACAA
AGgagGGGGACTGTGGCACAC-3′; for K1167AR1168A, 5′-
ATGGATACCCAACAgcggcgGGGGACTGTGGCAC-3′; and for
K1167ER1168E, 5′-ATGGATACCCAACAgaggagGGGGACTGTG
GCAC-3′. The T1085AY1156A mutant was generated with the
above Y1156A primer using methylated DNA of the T1085A as
the template. All the mutant clones constructed were subjected to
the sequencing of the entire genomic cDNA region to verify the
absence of unnecessary mutations leading to amino acid changes.

IN VITRO TRANSCRIPTION-TRANSLATION ASSAY

In vitro transcription-translation with a rabbit reticulocyte
system was performed using the TNT T7 Quick for PCR
DNA kit (Promega, Madison, WI, USA) as described previ-
ously (Oka et al., 2005b). Briefly, a template for the *in vitro*
transcription-translation, containing the entire ORF1, was pre-
pared by PCR amplification using the full-length cDNA clone.
The primers used for the amplification were as follows. The for-
ward primer containing a T7 promoter sequence (underlined)
and a translation initiation codon (bold) was 5′-GGATCCTAA
TACGACTCACTATAGGGAACAGCCACC**ATG** gcttccaagccattcta
cccaatagag-3′; and the antisense primer containing a stop codon
(bold) was 5′-T$_{30}$**TTA**-ttctaagaacctaacggcccgg. The PCR product
(3 μl) was mixed with 20 μl of TNT T7 PCR Quick Master Mix
(Promega) and 2 μl of Redivue Pro-mix L- [^{35}S] *in vitro* cell-
labeling mix (GE Healthcare Biosciences, Piscataway, NJ, USA).
The mixture was incubated at 30°C for 3 or 16 h and subjected to
SDS-polyacrylamide gel electrophoresis (SDS-PAGE). The trans-
lation products separated by electrophoresis were blotted onto a
PVDF membrane (Immobilon-P; Millipore, Bedford, MA, USA)
using a semi-dry electroblotting apparatus (ATTO; Tokyo). The
radiolabeled proteins were detected by a BAS 2500 Bioimage
Analyzer (Fuji Film, Tokyo).

IMMUNOPRECIPITATION

For the detection of NS1 (p11) and NS5 (VPg), which were
undetectable with the above assay system, we performed immuno-
precipitation before the SDS-PAGE as described previously (Oka
et al., 2005b, 2006, 2009). Briefly, 10 μl of the *in vitro* transcription-
translation reaction mixture was diluted with 80 μl of RIPA lysis
buffer containing 50 mM Tris, pH 7.4, 150 mM NaCl, 0.25%
deoxycholic acid, 1% NP40, and 1 mM EDTA (Upstate, Lake
Placid, NY, USA) and incubated with 5 μg of anti-A (anti-NS1)
or ant-D (anti-NS5) antibodies raised against *E. coli*-expressed
recombinant proteins (aa 1–231 for NS1 and aa 941–1055 for
NS5; Oka et al., 2005b). After incubation for 1 h on ice, 25 μl
of a suspension of Protein A Magnetic Beads (New England
Biolabs) and 900 μl of RIPA buffer were added. The mixture
was gently rotated at 4°C for 1 h and then washed three times
with 1 ml of RIPA lysis buffer. The immunoprecipitated pro-
teins were resuspended in 20 μl of SDS-PAGE sample buffer
and heated at 95°C for 5 min prior to analysis with 5 to 20%
Tris-Gly polyacrylamide gel. The proteins were blotted onto an
Immobilon-P polyvinylidene difluoride membrane (Millipore).

Immunoprecipitated radioactive proteins were detected with a
Bioimage Analyzer BAS 2500 (Fuji Film).

THE CHEMICAL COMPOUND LIBRARY

Chemical compounds (139,369 compounds, molecular weights
42–2986) were obtained from the Open Innovation Center for
Drug Discovery (The University of Tokyo, Tokyo, Japan). The
compound library database of this center provides information
on the molecular formula, molecular weight, hydrogen-bond
donor-acceptor numbers, topological polar surface area (TPSA),
and other physicochemical parameters of the compounds for
pharmacophore-based *in silico* drug screening.

PHARMACOPHORE-BASED *IN SILICO* SCREENING

Pharmacophore-based *in silico* screening was done using tools
available in the MOE. We created a pharmacophore query with
a substrate feature using the Pharmacophore Query Editor tool
in MOE. Pharmacophore-based *in silico* screening was done by
the Pharmacophore Search module in the MOE using the created
query.

DRUG SUSCEPTIBILITY ASSAY

The susceptibility of SaV protease to the synthetic small chem-
ical compound was determined by means of an *in vitro* trans
cleavage assay as follows. A radiolabeled full-length Mc10 ORF1
polyprotein containing a defective protease (Pro^{mut}; Oka et al.,
2005b) or a non-radiolabeled partial Mc10 ORF1 polyprotein
(NS6-7-VP1) containing a functional protease (Pro^{wt}; Oka et al.,
2006) was separately expressed using the *in vitro* transcrip-
tion/translation system (Oka et al., 2011). The PCR primer pairs
used for the preparation of DNA template for the expression
of the NS6-7-VP1 were as follows. The forward primer was 5′-
GGATCCTAATACGACTCACTATAGGGAACAGCCACC**ATG**gctc
ccacaccaattgttac-3′, including the T7 promoter sequence (under-
lined) and a start codon (bold); and the antisense primer was
5′-T$_{30}$**TTA**-ttctaagaacctaacggcccgg, including a stop codon (bold).
Twenty microliter of the non-radiolabeled products containing
Pro^{wt} was mixed with 1 μl of 2 mM inhibitor candidate in DMSO
and incubated for 10 min at room temperature. Then 10 μl of the
radiolabeled full-length ORF1 polyprotein (Pro^{mut}) was added to
the Pro^{wt}-inhibitor mixture and incubated at 30°C for 20 h, and
subjected to the SDS-PAGE analysis as described above. To quan-
titate the proteolytic activity of the SaV protease, we measured the
intensity of the band corresponding to the NS4-NS5 intermediate
processing product with Typhoon 7500 (GE Healthcare), due to
the lack of overlapping non-specific products of the *in vitro* trans-
lation around the NS4-NS5. The chemical compound concentra-
tions resulting in a 50% reduction of the NS4-NS5 intermediate
protein production of the drug-free control were determined on
the basis of the dose-response curve and defined as the IC$_{50}$ values
of the SaV proteolysis activity.

STRUCTURAL MODELING OF SaV PROTEASE DOCKED TO CHEMICAL
COMPOUNDS

Structural models of the chemical compounds were constructed
using the Molecular Builder tool in MOE. Individual compounds
were docked with the SaV protease domain model using the auto-
mated ligand docking program ASEDock2005 (Goto et al., 2008)
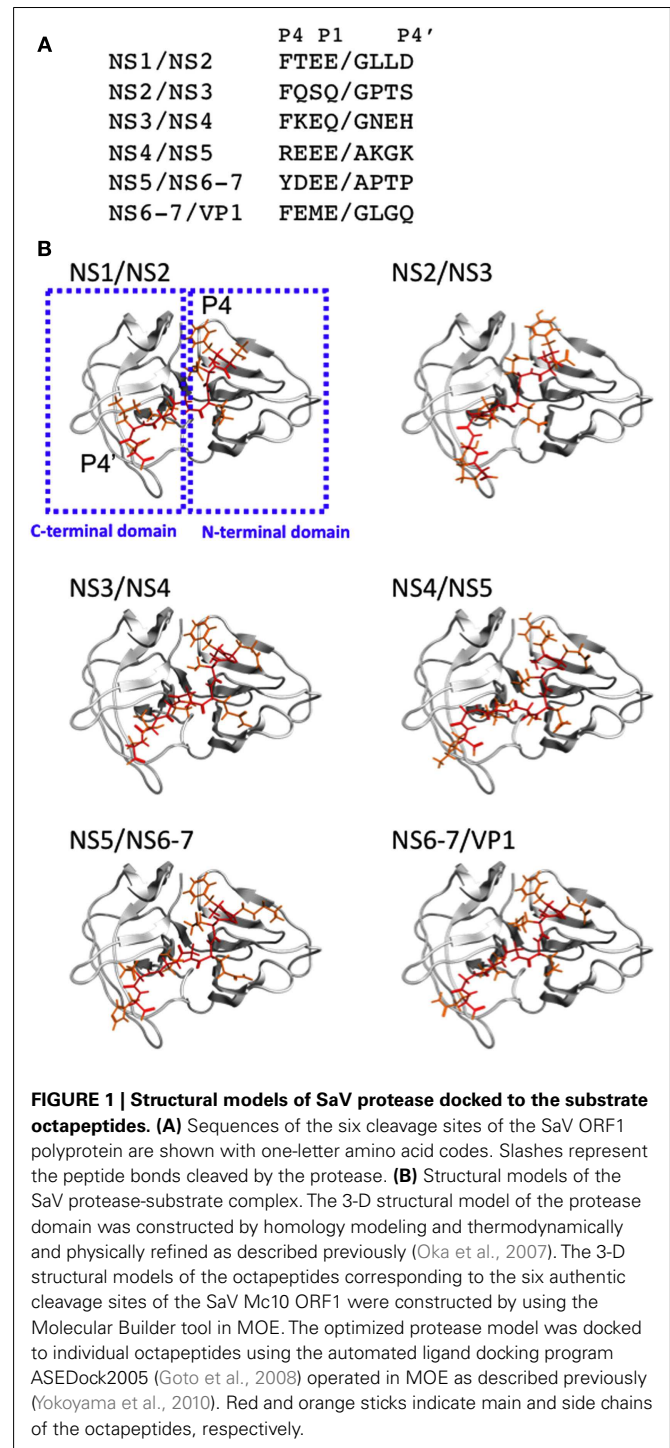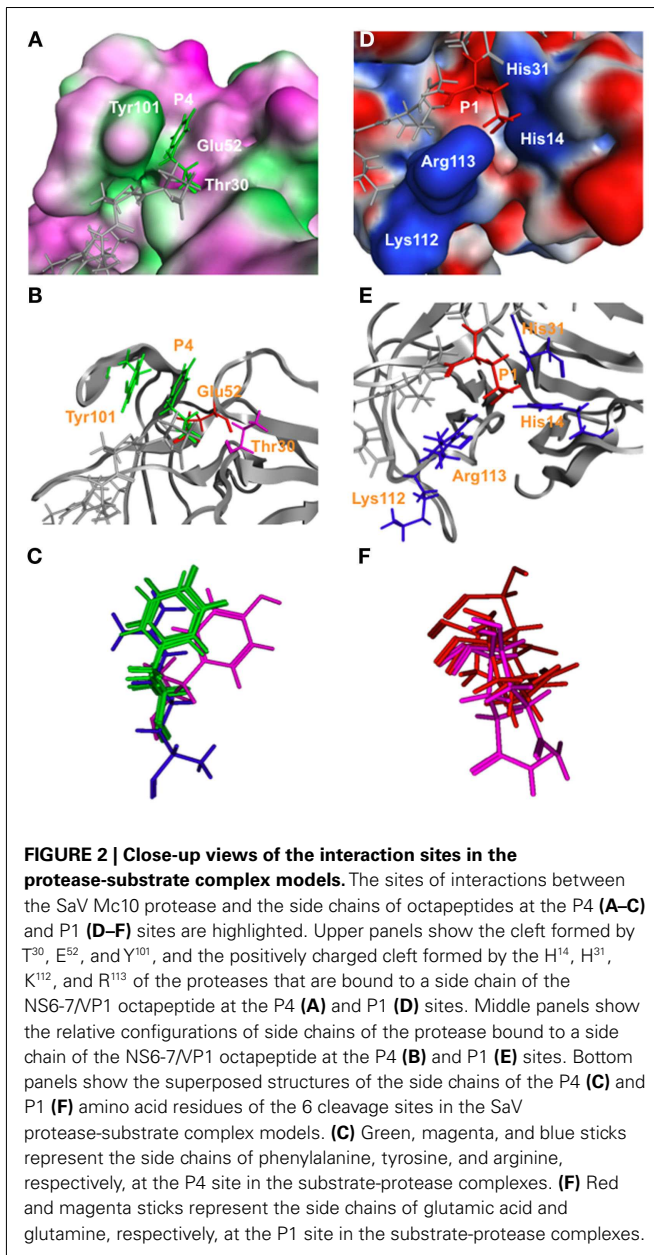operated in MOE as described above.

RESULTS

STRUCTURAL MODELING OF SaV PROTEASES DOCKED TO THE SUBSTRATE OCTAPEPTIDES

To obtain structural insights into the protease-substrate interactions at the atomic level, we constructed a 3-D model of the intact protease domain of the SaV Mc10 strain, which were docked to octapeptides corresponding to the six authentic cleavage sites (P4–P4′ sites) of the ORF1 polyprotein of the Mc10 strain (see Materials and Methods for details; **Figure 1**). The amino acid sequences of the six octapeptides are very different from each other (**Figure 1A**). Despite the variation, the octapeptides bound to the protease with the same orientation in the clefts of the protease (**Figure 1B**); the P1–P4 amino acid residues bound to the cleft between the N- and C-terminal domains, whereas the P1′–P4′ amino acid residues bound to the cleft on the C-terminal domain. The docking positions were functionally reasonable, because they allowed the cleavage sites of the octapeptides to be placed near the amino acids essential for the catalytic activity of the SaV protease, i.e., H^{31}, E^{52}, C^{116}, and H^{131} (Oka et al., 2005b). Other docking positions caused docking results with very poor docking scores and did not fulfill the functional requirement for the catalytic reaction.

The protease-peptide complex models disclosed two interaction sites that were common to the six bound peptides. First, the substrate amino acid residues at the P4 position were exclusively placed in a thin cleft, termed cleft 1, that was formed by threonine (T), glutamic acid (E), and tyrosine (Y) at positions 30, 52, and 101 of the protease domain (T^{30}, E^{52}, and Y^{101}, respectively; **Figures 2A–C**). An aromatic ring of the phenylalanine (F) or Y at P4 of the octapeptides of the NS1/NS2, NS2/NS3, NS3/NS4, NS5/NS6-7, and NS6-7/VP1 cleavage sites (**Figure 1A**) was positioned such that an aromatic stacking could be generated with the Y^{101} in the protease cleft 1 (**Figures 2A–C**). The steric configuration of the aromatic rings of the P4 amino acid residues in the bound state was very similar except for the Y of the NS5/NS6-7 cleavage site (**Figure 2C**). In the case of the NS4/NS5 peptide, the P4 amino acid is the arginine (R; **Figure 1A**) and was arranged near the side chain of the E^{52} (**Figure 2C**).

Second, the substrate amino acid residues at the P1 site were exclusively placed in a small positively charged cleft, termed cleft 2, that was formed by the histidine (H), H, lysine (K), and R at positions 14, 31, 112, and 113 of the protease domain (H^{14}, H^{31}, K^{112}, and R^{113}, respectively; **Figures 2D–F**). In four out of the six cleavage sequences the P1 amino acid is negatively charged (E; **Figure 1A**) that could interact electrostatically with the side chains of the positively charged cleft 2 of the protease (**Figure 2D**). In the case of the NS2/NS3 and NS3/NS4, the P1 amino acid was glutamine (Q; **Figure 1A**) which is hydrophilic and thus could cause electrostatic interactions via a polarized charge. The steric configuration of the side chains of the P1 amino acid residues at the bound state was very similar (**Figure 2F**). The simulated docking between the protease and the substrate having alanine substitutions at P1 and P4 positions resulted in a docking position similar to that for the wild-type substrate, whereas the docking score was reduced to about 1/2. Collectively, these results suggest that the interactions at the P1 and P4 sites of the substrates play a key role in the substrate recognition, as suggested in the previous experiments (Robel et al., 2008; Oka et al., 2009).

	P4 P1 P4′
NS1/NS2	FTEE/GLLD
NS2/NS3	FQSQ/GPTS
NS3/NS4	FKEQ/GNEH
NS4/NS5	REEE/AKGK
NS5/NS6-7	YDEE/APTP
NS6-7/VP1	FEME/GLGQ

FIGURE 1 | Structural models of SaV protease docked to the substrate octapeptides. (A) Sequences of the six cleavage sites of the SaV ORF1 polyprotein are shown with one-letter amino acid codes. Slashes represent the peptide bonds cleaved by the protease. **(B)** Structural models of the SaV protease-substrate complex. The 3-D structural model of the protease domain was constructed by homology modeling and thermodynamically and physically refined as described previously (Oka et al., 2007). The 3-D structural models of the octapeptides corresponding to the six authentic cleavage sites of the SaV Mc10 ORF1 were constructed by using the Molecular Builder tool in MOE. The optimized protease model was docked to individual octapeptides using the automated ligand docking program ASEDock2005 (Goto et al., 2008) operated in MOE as described previously (Yokoyama et al., 2010). Red and orange sticks indicate main and side chains of the octapeptides, respectively.

AMINO ACID DIVERSITY OF HUMAN SaV PROTEASE

To obtain evolutionary insights into the protease-substrate interactions, we analyzed the amino acid diversity of the protease domain among various human SaV strains in the public database. Full-length human SaV protease domain sequences were collected from GenBank ($N = 19$) and used to calculate the Shannon entropy scores, H (Shannon, 1948), in order to analyze the diversity of individual amino acid residues in the SaV population

FIGURE 2 | Close-up views of the interaction sites in the protease-substrate complex models. The sites of interactions between the SaV Mc10 protease and the side chains of octapeptides at the P4 **(A–C)** and P1 **(D–F)** sites are highlighted. Upper panels show the cleft formed by T^{30}, E^{52}, and Y^{101}, and the positively charged cleft formed by the H^{14}, H^{31}, K^{112}, and R^{113} of the proteases that are bound to a side chain of the NS6-7/VP1 octapeptide at the P4 **(A)** and P1 **(D)** sites. Middle panels show the relative configurations of side chains of the protease bound to a side chain of the NS6-7/VP1 octapeptide at the P4 **(B)** and P1 **(E)** sites. Bottom panels show the superposed structures of the side chains of the P4 **(C)** and P1 **(F)** amino acid residues of the 6 cleavage sites in the SaV protease-substrate complex models. **(C)** Green, magenta, and blue sticks represent the side chains of phenylalanine, tyrosine, and arginine, respectively, at the P4 site in the substrate-protease complexes. **(F)** Red and magenta sticks represent the side chains of glutamic acid and glutamine, respectively, at the P1 site in the substrate-protease complexes.

FIGURE 3 | Diversity of SaV protease amino acid residues. The amino acid diversity at individual sites of the SaV protease domain was analyzed with information entropy as described previously (Oka et al., 2009). The Shannon entropy H was calculated with Shannon's formula (Shannon, 1948) based on amino acid residues **(A)**, chemical properties **(B)**, and the size of the side chain **(C)** using amino acid sequences of the SaV full-length protease domain from GenBank ($N = 19$). For analysis of the diversity in the chemical properties, the amino acid residues were classified into seven groups: acidic (D,E), basic (R,K,H), neutral hydrophilic (S, T, N, Q), aliphatic (G, A, V, I, L, M), aromatic (F, Y, W), thio-containing (C), and imine (P). For analysis of the diversity in the size of the side chain, the amino acid residues were classified into 4 groups: small (G, A, C, S), medium-small (T, V, N, D, I, L, P, M), medium-large (Q, E, R, K), and large (H, F, Y, W). The H scores are plotted on the 3-D structure of the SaV protease model, where an H score of zero indicates absolute conservation. Yellow and orange dotted circles indicate clefts 1 and 2, respectively.

as described previously (Oka et al., 2009). The H scores generally ranged from 0.0 to 0.6 bits (**Figure 3A**), indicating that the diversity of the SaV protease is relatively small, as seen in many viral enzymes. The variable sites were essentially located on the surface region, indicating that some exposed regions of the SaV protease allow amino acid changes (**Figure 3A**, greenish sites). Although less extensive, some variation was detected at Y^{101} and R^{113} in the clefts 1 and 2, respectively (**Figure 3A**, two dotted circles). However, when the H scores were calculated on the basis of chemical properties or the size of the amino acid residues, they were nearly zero throughout the substrate-binding cleft (**Figures 3B,C**). Similarly, the protease amino acid residues, which constitute a large cavity for the binding of entire octapeptides, were sometime variable but highly conserved in the context of the chemical properties
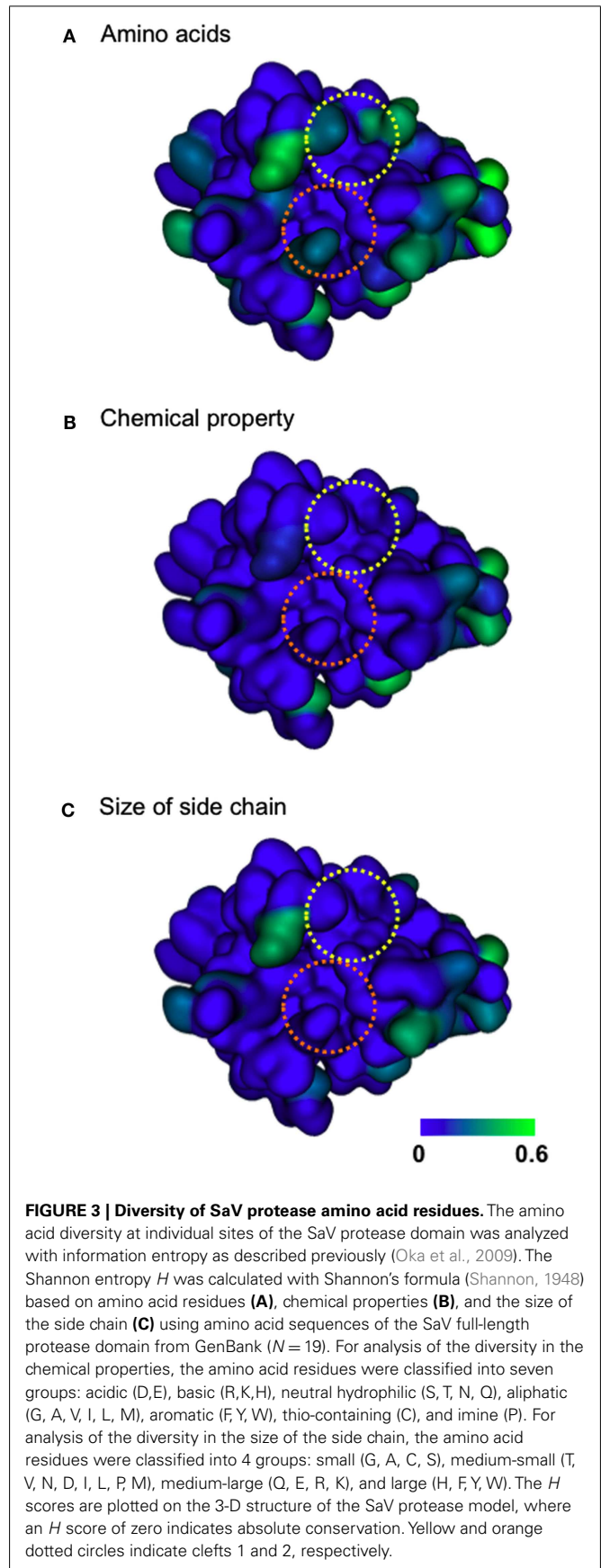
and sizes of the side chains (**Figures 1B** and **3**). Thus the SaV protease appears to restrict extensive changes in the shape and chemical properties of the substrate-binding surface for its survival.

SITE-DIRECTED MUTAGENESIS OF SaV PROTEASE

Consistent with the above structural and diversity data, we previously reported that the E^{52} in cleft 1, as well as H^{14} and H^{31} in cleft 2, are essential to maintain proper processing by SaV protease (Oka et al., 2007). To obtain further insights into the biological roles of clefts 1 and 2 in the proteolysis of the SaV precursor polyprotein, we performed additional site-directed mutagenesis using a full-length clone of SaV Mc10 strain (Oka et al., 2005b). The Mc10 ORF1 encodes a polypeptide of 2278 amino

acid residues, where the six cleavage sites have been experimentally determined (Oka et al., 2006; **Figure 4A**). A total of nine mutants of the SaV protease domain were constructed using the Mc10 ORF1. Full-length ORF1 precursor proteins having a single or double mutations in the protease domain were expressed using the *in vitro* transcription-translation system, and the processing products were analyzed by gel electrophoresis as described previously (Oka et al., 2005b, 2006, 2007, 2009). The Mc10 functional protease (Pro^wt) and a defective mutant completely lacking the proteolysis activity (Pro^mut; Oka et al., 2005b) were used as positive and negative controls of the proteolysis, respectively.

When the ORF1 containing the Pro^wt was expressed, nine products corresponded in size to the mature proteins NS1, NS2, NS3, NS4, NS5, and VP1, and relatively stable intermediate proteins,
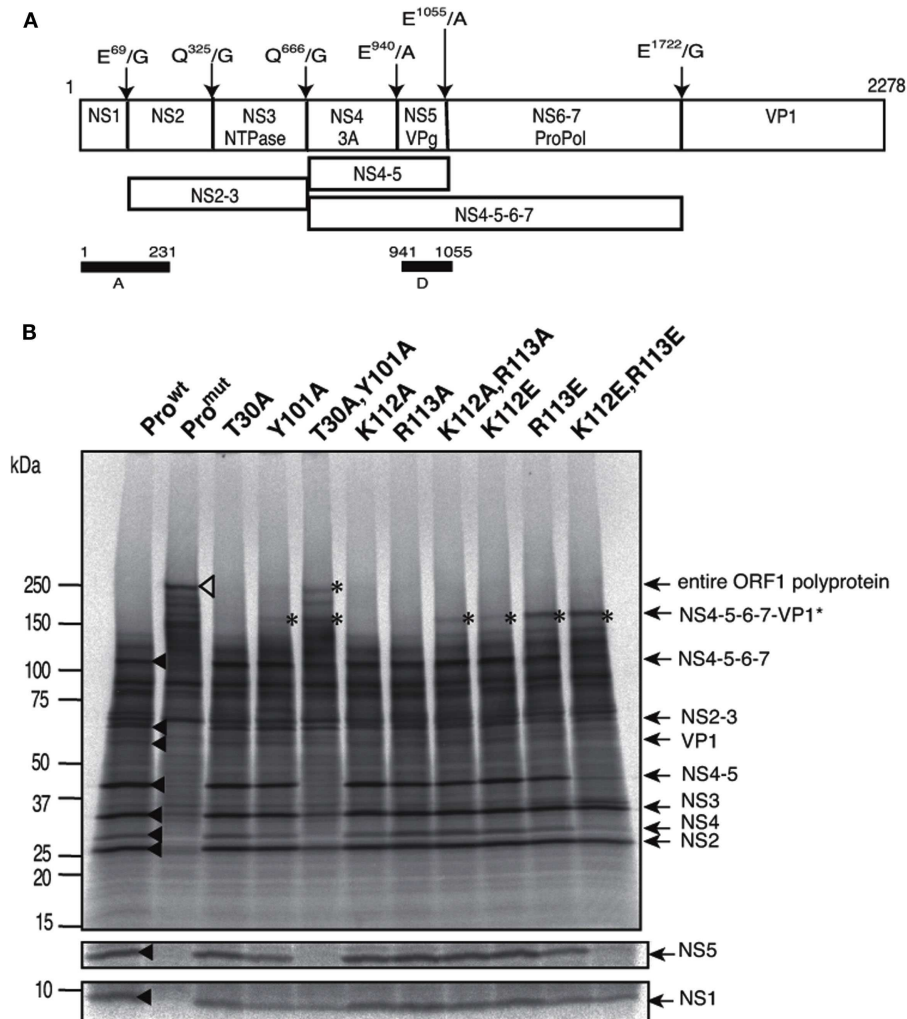
FIGURE 4 | Site-directed mutagenesis of the substrate interaction sites of SaV Mc10 protease. (A) Proteolytic cleavage map of the SaV Mc10 ORF1 polyprotein and the processing intermediates (Oka et al., 2006). Black bars indicate the protein segments, A and D, used to raise polyclonal antibodies for detection of the NS1 and NS5 proteins, respectively. **(B)** SDS-PAGE of ^35S-labeled *in vitro* translation products of SaV Mc10 ORF1 containing various protease mutants. NS1 and NS5 were detected by immunoprecipitation using

anti-A or anti-D polyclonal antibodies as described previously (Oka et al., 2005b, 2006, 2009). Mc10 ORF1 containing functional protease (Pro^wt) and a defective mutant lacking in the proteolysis activity (Pro^mut) were included as described previously (Oka et al., 2005b). Newly appearing products when compared to Pro^wt are indicated by asterisks. Size markers are shown on the left. Mc10 ORF1-specific proteins (Oka et al., 2005b, 2006) are shown on the right.

such as NS2-3, NS4-5, and NS4-5-6-7 were detected (**Figure 4B**, lane Pro^{wt}, black arrowheads; Oka et al., 2005b, 2006, 2009). These products were undetectable in the Pro^{mut} ORF1 sample, and instead a product corresponding to the ORF1 polyprotein was detected (**Figure 4B**, lane Pro^{mut}, open triangle; Oka et al., 2005b, 2006, 2009). A single alanine substitution at T^{30} in the cleft 1, K^{112} in the cleft 2, or R^{113} in the cleft 2 of viral protease resulted in a processing pattern similar to that of Pro^{wt} (**Figure 4B**, lanes T30A, K112A, and R113A). On the other hand, a single alanine substitution at Y^{101} in the cleft 1 (Y101A), a single acidic substitution at K^{112} or R^{113} in the cleft 2 (K112E and R113E), and double mutations in each cleft (T30AY101A and K112ER113E) resulted in abnormality of the precursor processing, i.e., an increase in accumulation of the full-length ORF1 polyprotein and/or the NS4-5-6-7-VP1 intermediate protein (**Figure 4B**, asterisks). In the samples expressing ORF1 with T30A/Y101A or K112E/R113E double mutations, processing products corresponding to the NS5 and NS4-5 disappeared almost completely (**Figure 4B**, lanes 5 and 11, respectively).
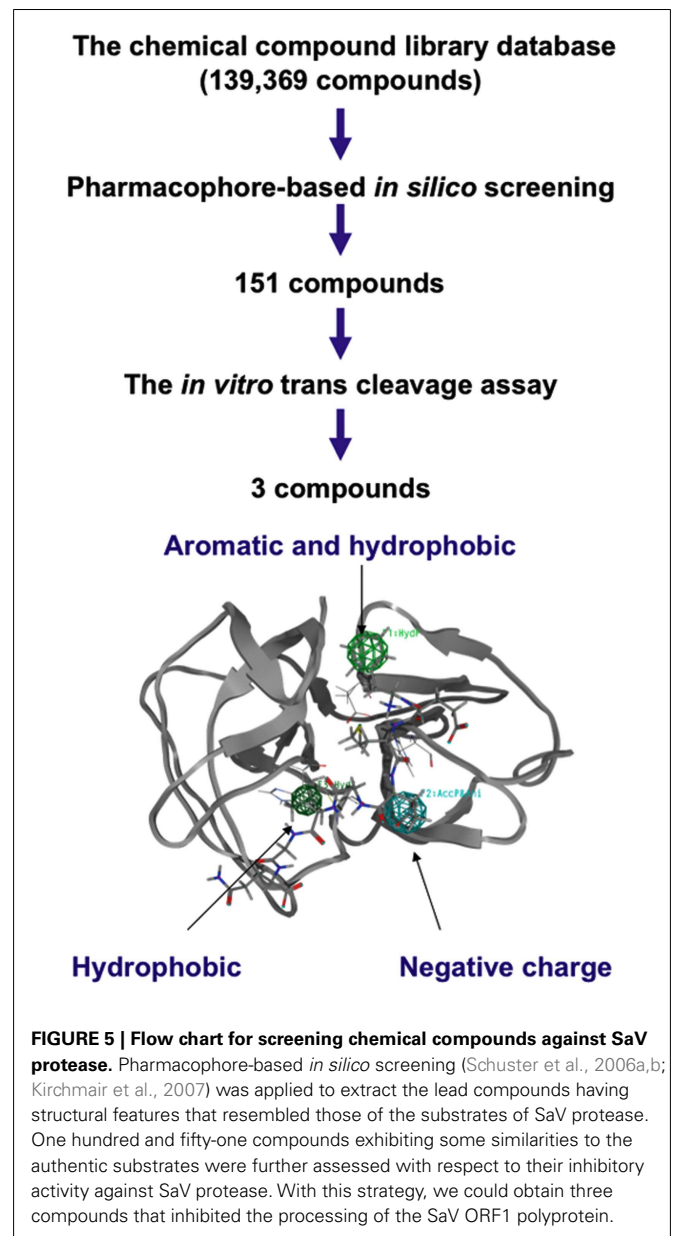
PHARMACOPHORE-BASED *IN SILICO* SCREENING FOR THE LEAD COMPOUNDS OF SaV PROTEASE INHIBITORS

To further assess the role of the clefts 1 and 2 in the ligand binding, we performed a pharmacophore-based *in silico* screening of protease inhibitors. A total of 139,369 compounds (molecular weights 42–2986) were screened for the lead molecules that contain an aromatic-ring-like portion resembling the P4 amino acid, a negatively charged portion resembling the P1 amino acid, and a hydrophobic portion resembling the P1′ amino acid, being arranged at similar 3-D positions with the authentic substrates (**Figure 5**). The hydrophobic portion resembling the P1′ amino acid was included to better mimic the authentic substrate structures. A total of 151 lead compounds matched to the category were then subjected to the *in vitro* trans cleavage assay of the SaV Mc10 ORF1 polyprotein. With this screening, we could obtain three compounds that inhibited processing of the SaV ORF1 at IC$_{50}$ values of 18.4–26.5 μM (**Figure 6**).

We then analyzed how the lead compounds bound to the SaV Mc10 protease by docking simulation (**Figure 7**). As expected, these compounds were predicted to bind to the protease at the same interaction sites by which the authentic substrates bound to the protease. The aromatic-ring-like portion resembling the P4 amino acid bound to the thin cleft formed by T^{30}, E^{52}, and Y^{101} for the binding of the side chain of the P4 amino acid. The negatively charged portion resembling the P1 amino acid bound to the small positively charged pocket formed by the H^{14}, H^{31}, K^{112}, and R^{113} for the binding of the side chain of the P1 amino acid.
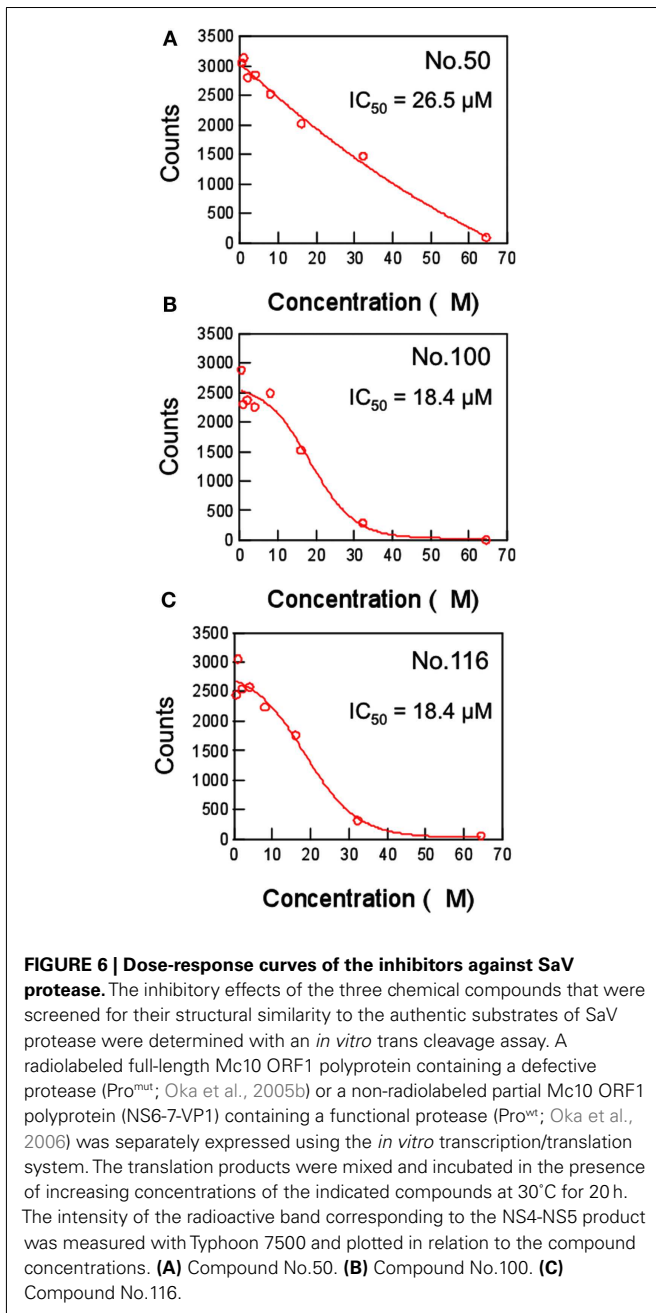
DISCUSSION

The viral proteins that support viral replication and make up the viral particle are often translated as part of polyprotein precursors. Viral protease catalyzes cleavage of the precursor protein and thus plays an essential role in the viral life cycle. In this study, by combining computational and experimental approaches, we studied the structural basis for the substrate recognition by SaV protease. The results obtained in this study were consistent with each other

FIGURE 5 | Flow chart for screening chemical compounds against SaV protease. Pharmacophore-based *in silico* screening (Schuster et al., 2006a,b; Kirchmair et al., 2007) was applied to extract the lead compounds having structural features that resembled those of the substrates of SaV protease. One hundred and fifty-one compounds exhibiting some similarities to the authentic substrates were further assessed with respect to their inhibitory activity against SaV protease. With this strategy, we could obtain three compounds that inhibited the processing of the SaV ORF1 polyprotein.
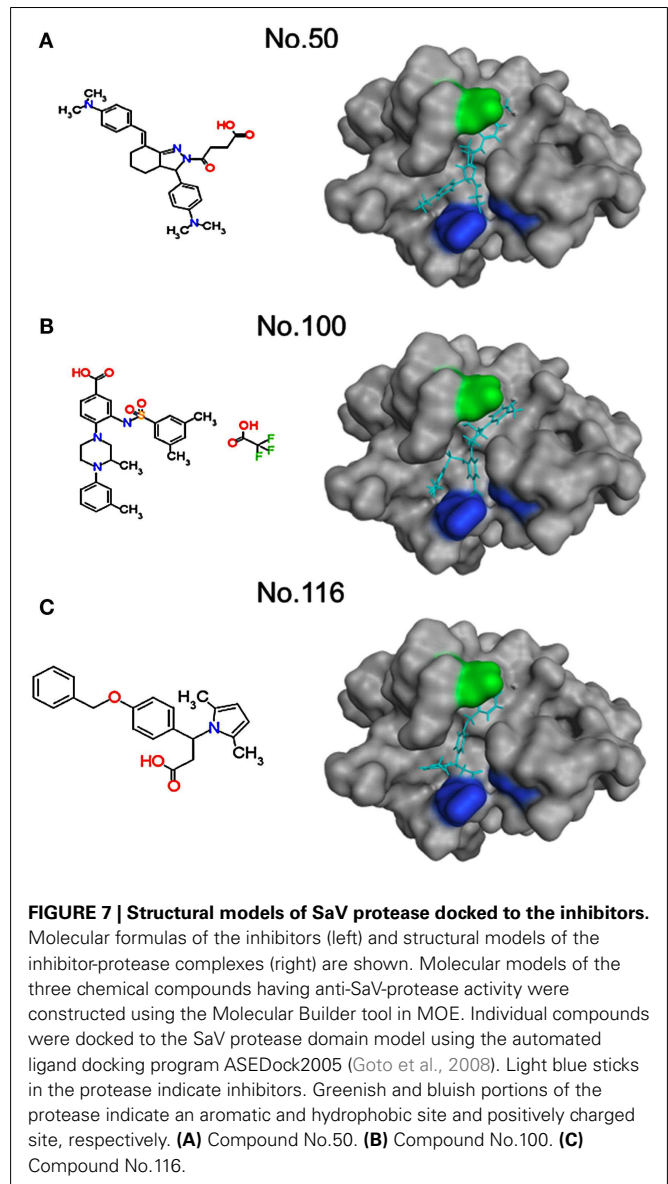
and disclosed novel structural base points of the protease for the attractive interactions with specific structures of ligands.

Using a homology modeling and a docking tool, we first examined the physical interactions of SaV protease and octapeptides corresponding to the six authentic cleavage sites of the SaV ORF1 polyprotein. Despite the marked sequence variation of the octapeptides, they were bound to the protease in the same orientation in the structural models (**Figure 1**). The results suggested that there might be common interaction sites that served as fulcrums to direct the orientation of the octapeptides. Consistently, the models disclosed two interaction sites that were shared with the six peptides and support the stable and functional binding of substrates to the catalytic cavity; the variable side chains at the P4 and P1 sites of the peptides were consistently bound to the two small clefts, termed clefts 1 and 2, respectively (**Figure 2**). The former

FIGURE 6 | Dose-response curves of the inhibitors against SaV protease. The inhibitory effects of the three chemical compounds that were screened for their structural similarity to the authentic substrates of SaV protease were determined with an *in vitro* trans cleavage assay. A radiolabeled full-length Mc10 ORF1 polyprotein containing a defective protease (Pro^mut; Oka et al., 2005b) or a non-radiolabeled partial Mc10 ORF1 polyprotein (NS6-7-VP1) containing a functional protease (Pro^wt; Oka et al., 2006) was separately expressed using the *in vitro* transcription/translation system. The translation products were mixed and incubated in the presence of increasing concentrations of the indicated compounds at 30°C for 20 h. The intensity of the radioactive band corresponding to the NS4-NS5 product was measured with Typhoon 7500 and plotted in relation to the compound concentrations. **(A)** Compound No.50. **(B)** Compound No.100. **(C)** Compound No.116.

FIGURE 7 | Structural models of SaV protease docked to the inhibitors. Molecular formulas of the inhibitors (left) and structural models of the inhibitor-protease complexes (right) are shown. Molecular models of the three chemical compounds having anti-SaV-protease activity were constructed using the Molecular Builder tool in MOE. Individual compounds were docked to the SaV protease domain model using the automated ligand docking program ASEDock2005 (Goto et al., 2008). Light blue sticks in the protease indicate inhibitors. Greenish and bluish portions of the protease indicate an aromatic and hydrophobic site and positively charged site, respectively. **(A)** Compound No.50. **(B)** Compound No.100. **(C)** Compound No.116.

participated in aromatic stacking interactions, whereas the latter participated in electrostatic interactions. These results are consistent with the previous findings that the P4 and P1 amino acid residues of the substrates play key roles in efficient proteolysis by SaV protease (Robel et al., 2008; Oka et al., 2009) and predicted that these two clefts could play a key role in substrate recognition via interactions with the P4 and P1 amino acid residues of substrates.

This prediction was assessed by several analyses. If the clefts played essential roles in recognition of substrates, spontaneous mutations that alter profoundly the physicochemical properties of the clefts should be suppressed for viral survival. Consistently our Shannon entropy study using protease sequences of various SaV strains from the world shows that the amino acid residues forming the clefts 1 and 2 are variable but highly conserved in terms of the chemical properties or the sizes of side chains (**Figure 3**). The results indicate that these clefts tolerate mutations in nature but resist a range of mutations that markedly alter the chemical properties or the shapes of the cleft surface. The findings are consistent with the above structure-based prediction on the function of the clefts 1 and 2. These clefts are located on the surface of the large cavity of the protease. Therefore, the restrictions in the variation in two clefts are likely to be caused by functional constraints for the essential interactions.

Moreover, we examined whether a range of mutations that markedly alter the physicochemical properties of the clefts indeed could result in aberrant processing of the SaV precursor polyprotein. Our site-directed mutagenesis study showed that a single mutation in cleft 1 (T30A) or in cleft 2 (K112A or R113A) caused little detectable damage in the processing of the viral precursor

polyprotein, showing a tolerance to mutations as indicated by our information entropy study. Notably, however, (i) a single mutation that causes a loss of aromatic stacking interaction (Y101A) in the cleft 1, (ii) a single mutation that causes a loss of the electrostatic interaction in the cleft 2 (K112E or R113E), and (iii) double mutations within the clefts unexceptionally resulted in incomplete processing (**Figure 4**). The results indicate that the abnormal processing was caused only by single mutations that could extensively alter the chemical properties of the clefts. The data agree with the entropy data and again suggest the acceptability of variation in the two clefts under functional constraints.

Finally, we performed *in silico* screening of SaV protease inhibitors on the basis of the above structural and biological information. The screening of the 139,369 compounds *in silico* led to the identification of the 151 compounds that resembled the structural and spatial features of the P4 and P1 amino acid residues of authentic substrates (**Figure 5**). From them, we could experimentally identify the three compounds that inhibited proteolysis of the SaV precursor polyprotein *in vitro* (**Figure 6**). As expected, these compounds were predicted to bind to the SaV protease at the two clefts via similar attractive interactions with the authentic ligands (**Figure 7**). These results provide additional evidence that two clefts on the SaV protease cavity play a key role in the ligand recognition by providing the structural base points for the specific attractive interactions.

Notably, six cleavage sites of SaV precursor polyprotein also differ with respect to their susceptibility to the SaV protease, with the NS2/NS3, NS4/NS5, and NS5/NS6-7 sites being consistently more resistant to the cleavage than the NS1/NS2, NS3/NS4, and NS6-7/VP1 sites (Oka et al., 2005b, 2006, 2009). In this regard, it is of note that the P4 position of the NS4/NS5 site of human SaV is exclusively arginine instead of an aromatic amino acid (**Figure 1A**) and that this arginine is conserved in all human SaV strains reported thus far (Oka et al., 2005b, 2006, 2009). This substitution at P4 position will abolish the aromatic stacking interaction in the cleft 1 and thus will attenuate attractive interactions

between protease and the NS4/NS5 cleavage site. This possibility is well consistent with the experimental findings; the cleavage of the NS4/NS5 site is less efficient than that of the other sites (Oka et al., 2005b, 2006, 2009) and is more sensitive to the cleft 1 mutations than the other cleavage sites are (**Figure 4**, lane 5, NS5). Moreover, the attenuation of cleavage of the NS4/NS5 site was reversed simply by replacing the arginine with phenylalanine at the P4 site (Oka et al., 2009). These findings strongly suggest that the well-preserved arginine at the P4 position of the SaV NS4/NS5 cleavage site plays a key role in maintaining the distinct cleavability of precursor polyprotein by SaV protease.

In this study, we disclosed a novel 3-D pharmacophore containing two clefts on the cavity of the SaV protease, which can be used to identify the lead compounds of SaV protease inhibitors. SaV is one of the commonly detected pathogens in the acute gastroenteritis of both children and adults (Johansson et al., 2005; Harada et al., 2009; Iturriza-Gomara et al., 2009; Pang et al., 2009). Diarrhea is one of the greatest causes of mortality in children under age 5 in many countries (Boschi-Pinto et al., 2008), and the outbreaks of the acute gastroenteritis often seriously affects the clinical, economic, and social activities. Therefore, anti-viral compounds against SaV may be beneficial to some at-risk populations or communities. Thus far no anti-SaV inhibitors for the clinical use have been developed. Our findings will provide important clues to the unique specificity of the SaV protease, the regulation of SaV maturation, and the rationale design of anti-SaV inhibitors.

ACKNOWLEDGMENTS

REFERENCES

Baker, D., and Sali, A. (2001). Protein structure prediction and structural genomics. *Science* 294, 93–96.

Belliot, G., Sosnovtsev, S. V., Mitra, T., Hammer, C., Garfield, M., and Green, K. Y. (2003). In vitro proteolytic processing of the MD145 norovirus ORF1 nonstructural polyprotein yields stable precursors and products similar to those detected in calicivirus-infected cells. *J. Virol.* 77, 10957–10974.

Bergmann, E. M., Cherney, M. M., Mckendrick, J., Frormann, S., Luo, C., Malcolm, B. A., Vederas, J. C., and James, M. N. (1999). Crystal structure of an inhibitor complex of the 3C proteinase from hepatitis A virus (HAV) and implications for the polyprotein processing in HAV. *Virology* 265, 153–163.

Boschi-Pinto, C., Velebit, L., and Shibuya, K. (2008). Estimating child mortality due to diarrhoea in

developing countries. *Bull. World Health Organ.* 86, 710–717.

Bull, R. A., Hyde, J., Mackenzie, J. M., Hansman, G. S., Oka, T., Takeda, N., and White, P. A. (2011). Comparison of the replication properties of murine and human calicivirus RNA-dependent RNA polymerases. *Virus Genes* 42, 16–27.

Chiba, S., Nakata, S., Numata-Kinoshita, K., and Honma, S. (2000). Sapporo virus: history and recent findings. *J. Infect. Dis.* 181(Suppl. 2), S303–S308.

Chiba, S., Sakuma, Y., Kogasaka, R., Akihara, M., Horino, K., Nakao, T., and Fukui, S. (1979). An outbreak of gastroenteritis associated with calicivirus in an infant home. *J. Med. Virol.* 4, 249–254.

Fullerton, S. W., Blaschke, M., Coutard, B., Gebhardt, J., Gorbalenya, A., Canard, B., Tucker, P. A., and Rohayem, J. (2007). Structural and functional characterization of

sapovirus RNA-dependent RNA polymerase. *J. Virol.* 81, 1858–1871.

Goto, J., Kataoka, R., Muta, H., and Hirayama, N. (2008). ASEDock-docking based on alpha spheres and excluded volumes. *J. Chem. Inf. Model.* 48, 583–590.

Guo, M., Chang, K. O., Hardy, M. E., Zhang, Q., Parwani, A. V., and Saif, L. J. (1999). Molecular characterization of a porcine enteric calicivirus genetically related to Sapporo-like human caliciviruses. *J. Virol.* 73, 9625–9631.

Hansman, G. S., Oka, T., Katayama, K., and Takeda, N. (2007). Human sapoviruses: genetic diversity, recombination, and classification. *Rev. Med. Virol.* 17, 133–141.

Harada, S., Okada, M., Yahiro, S., Nishimura, K., Matsuo, S., Miyasaka, J., Nakashima, R., Shimada, Y., Ueno, T., Ikezawa, S., Shinozaki, K., Katayama, K., Wakita, T., Takeda, N., and Oka, T. (2009). Surveillance of pathogens in outpatients with

gastroenteritis and characterization of sapovirus strains between 2002 and 2007 in Kumamoto Prefecture, Japan. *J. Med. Virol.* 81, 1117–1127.

Hardy, M. E., Crone, T. J., Brower, J. E., and Ettayebi, K. (2002). Substrate specificity of the Norwalk virus 3C-like proteinase. *Virus Res.* 89, 29–39.

Iturriza-Gomara, M., Elliot, A. J., Dockery, C., Fleming, D. M., and Gray, J. J. (2009). Structured surveillance of infectious intestinal disease in pre-school children in the community: "The Nappy Study." *Epidemiol. Infect.* 137, 922–931.

Johansson, P. J., Bergentoft, K., Larsson, P. A., Magnusson, G., Widell, A., Thorhagen, M., and Hedlund, K. O. (2005). A nosocomial sapovirus-associated outbreak of gastroenteritis in adults. *Scand. J. Infect. Dis.* 37, 200–204.

Kataoka, R., and Goto, J. (2008). ASE-Dock – docking based on the shape of binding site. *Mol. Sci.* 2, NP008.

Kirchmair, J., Ristic, S., Eder, K., Markt, P., Wolber, G., Laggner, C., and Langer, T. (2007). Fast and efficient in silico 3D screening: toward maximum computational efficiency of pharmacophore-based and shape-based approaches. *J. Chem. Inf. Model.* 47, 2182–2196.

Liu, B. L., Clarke, I. N., Caul, E. O., and Lambden, P. R. (1995). Human enteric caliciviruses have a unique genome structure and are distinct from the Norwalk-like viruses. *Arch. Virol.* 140, 1345–1356.

Matthews, D. A., Dragovich, P. S., Webber, S. E., Fuhrman, S. A., Patick, A. K., Zalman, L. S., Hendrickson, T. F., Love, R. A., Prins, T. J., Marakovits, J. T., Zhou, R., Tikhe, J., Ford, C. E., Meador, J. W., Ferre, R. A., Brown, E. L., Binford, S. L., Brothers, M. A., Delisle, D. M., and Worland, S. T. (1999). Structure-assisted design of mechanism-based irreversible inhibitors of human rhinovirus 3C protease with potent antiviral activity against multiple rhinovirus serotypes. *Proc. Natl. Acad. Sci. U.S.A.* 96, 11000–11007.

Mirny, L. A., and Shakhnovich, E. I. (1999). Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. *J. Mol. Biol.* 291, 177–196.

Mosimann, S. C., Cherney, M. M., Sia, S., Plotch, S., and James, M. N. (1997). Refined X-ray crystallographic structure of the poliovirus 3C gene product. *J. Mol. Biol.* 273, 1032–1047.

Motomura, K., Oka, T., Yokoyama, M., Nakamura, H., Mori, H., Ode, H., Hansman, G. S., Katayama, K., Kanda, T., Tanaka, T., Takeda, N., and Sato, H. (2008). Identification of monomorphic and divergent haplotypes in the 2006–2007 norovirus GII/4 epidemic population by genomewide tracing of evolutionary history. *J. Virol.* 82, 11247–11262.

Motomura, K., Yokoyama, M., Ode, H., Nakamura, H., Mori, H., Kanda, T., Oka, T., Katayama, K., Noda, M., Tanaka, T., Takeda, N., and Sato, H. (2010). Divergent evolution of norovirus GII/4 by genome recombination from May 2006 to February 2009 in Japan. *J. Virol.* 84, 8085–8097.

Naganawa, S., Yokoyama, M., Shiino, T., Suzuki, T., Ishigatsubo, Y., Ueda, A., Shirai, A., Takeno, M., Hayakawa, S., Sato, S., Tochikubo, O., Kiyoura, S., Sawada, K., Ikegami, T., Kanda, T., Kitamura, K., and Sato, H. (2008). Net positive charge of HIV-1 CRF01_AE V3 sequence regulates viral sensitivity to humoral immunity. *PLoS ONE* 3, e3206. doi:10.1371/journal.pone.0003206

Nakamura, K., Someya, Y., Kumasaka, T., Ueno, G., Yamamoto, M., Sato, T., Takeda, N., Miyamura, T., and Tanaka, N. (2005). A norovirus protease structure provides insights into active and substrate binding site integrity. *J. Virol.* 79, 13685–13693.

Noel, J. S., Liu, B. L., Humphrey, C. D., Rodriguez, E. M., Lambden, P. R., Clarke, I. N., Dwyer, D. M., Ando, T., Glass, R. I., and Monroe, S. S. (1997). Parkville virus: a novel genetic variant of human calicivirus in the Sapporo virus clade, associated with an outbreak of gastroenteritis in adults. *J. Med. Virol.* 52, 173–178.

Numata, K., Hardy, M. E., Nakata, S., Chiba, S., and Estes, M. K. (1997). Molecular characterization of morphologically typical human calicivirus Sapporo. *Arch. Virol.* 142, 1537–1552.

Ode, H., Yokoyama, M., Kanda, T., and Sato, H. (2011). Identification of folding preferences of cleavage junctions of HIV-1 precursor proteins for regulation of cleavability. *J. Mol. Model.* 17, 391–399.

Oka, T., Katayama, K., Ogawa, S., Hansman, G. S., Kageyama, T., Miyamura, T., and Takeda, N. (2005a). Cleavage activity of the sapovirus 3C-like protease in *Escherichia coli*. *Arch. Virol.* 150, 2539–2548.

Oka, T., Katayama, K., Ogawa, S., Hansman, G. S., Kageyama, T., Ushijima, H., Miyamura, T., and Takeda, N. (2005b). Proteolytic processing of sapovirus ORF1 polyprotein. *J. Virol.* 79, 7283–7290.

Oka, T., Murakami, K., Wakita, T., and Katayama, K. (2011). Comparative site-directed mutagenesis in the catalytic amino acid triad in calicivirus proteases. *Microbiol. Immunol.* 55, 108–114.

Oka, T., Yamamoto, M., Katayama, K., Hansman, G. S., Ogawa, S., Miyamura, T., and Takeda, N. (2006). Identification of the cleavage sites of sapovirus open reading frame 1 polyprotein. *J. Gen. Virol.* 87, 3329–3338.

Oka, T., Yamamoto, M., Yokoyama, M., Ogawa, S., Hansman, G. S., Katayama, K., Miyashita, K., Takagi, H., Tohya, Y., Sato, H., and Takeda, N. (2007). Highly conserved configuration of catalytic amino acid residues among calicivirus-encoded proteases. *J. Virol.* 81, 6798–6806.

Oka, T., Yokoyama, M., Katayama, K., Tsunemitsu, H., Yamamoto, M., Miyashita, K., Ogawa, S., Motomura, K., Mori, H., Nakamura, H., Wakita, T., Takeda, N., and Sato, H. (2009). Structural and biological constraints on diversity of regions immediately upstream of cleavage sites in calicivirus precursor proteins. *Virology* 394, 119–129.

Pang, X. L., Lee, B. E., Tyrrell, G. J., and Preiksaitis, J. K. (2009). Epidemiology and genotype analysis of sapovirus associated with gastroenteritis outbreaks in Alberta, Canada: 2004–2007. *J. Infect. Dis.* 199, 547–551.

Ponder, J. W., and Case, D. A. (2003). Force fields for protein simulations. *Adv. Protein Chem.* 66, 27–85.

Robel, I., Gebhardt, J., Mesters, J. R., Gorbalenya, A., Coutard, B., Canard, B., Hilgenfeld, R., and Rohayem, J. (2008). Functional characterization of the cleavage specificity of the sapovirus chymotrypsin-like protease. *J. Virol.* 82, 8085–8093.

Robinson, S., Clarke, I. N., Vipond, I. B., Caul, E. O., and Lambden, P. R. (2002). Epidemiology of human Sapporo-like caliciviruses in the South West of England: molecular characterisation of a genetically distinct isolate. *J. Med. Virol.* 67, 282–288.

Sakuragi, J. I., Ode, H., Sakuragi, S., Shioda, T., and Sato, H. (2012). A proposal for a new HIV-1 DLS structural model. *Nucleic Acids Res.* 40, 5012–5022.

Sanchez, R., Pieper, U., Melo, F., Eswar, N., Marti-Renom, M. A., Madhusudhan, M. S., Mirkovic, N., and Sali, A. (2000). Protein structure modeling for structural genomics. *Nat. Struct. Biol.* 7(Suppl.), 986–990.

Sander, C., and Schneider, R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins* 9, 56–68.

Scheffler, U., Rudolph, W., Gebhardt, J., and Rohayem, J. (2007). Differential cleavage of the norovirus polyprotein precursor by two active forms of the viral protease. *J. Gen. Virol.* 88, 2013–2018.

Schuster, D., Laggner, C., Steindl, T. M., Palusczak, A., Hartmann, R. W., and Langer, T. (2006a). Pharmacophore modeling and in silico screening for new P450 19 (aromatase) inhibitors. *J. Chem. Inf. Model.* 46, 1301–1311.

Schuster, D., Maurer, E. M., Laggner, C., Nashev, L. G., Wilckens, T., Langer, T., and Odermatt, A. (2006b). The discovery of new 11beta-hydroxysteroid dehydrogenase type 1 inhibitors by common feature pharmacophore modeling and virtual screening. *J. Med. Chem.* 49, 3454–3466.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423, 623–656.

Shirakawa, K., Takaori-Kondo, A., Yokoyama, M., Izumi, T., Matsui, M., Io, K., Sato, T., Sato, H., and Uchiyama, T. (2008). Phosphorylation of APOBEC3G by protein kinase A regulates its interaction with HIV-1 Vif. *Nat. Struct. Mol. Biol.* 15, 1184–1191.

Sosnovtsev, S. V., Sosnovtseva, S. A., and Green, K. Y. (1998). Cleavage of the feline calicivirus capsid precursor is mediated by a virus-encoded proteinase. *J. Virol.* 72, 3051–3059.

Wirblich, C., Sibilia, M., Boniotti, M. B., Rossi, C., Thiel, H. J., and Meyers, G. (1995). 3C-like protease of rabbit hemorrhagic disease virus: identification of cleavage sites in the ORF1 polyprotein and analysis of cleavage specificity. *J. Virol.* 69, 7159–7168.

Yokoyama, M., Mori, H., and Sato, H. (2010). Allosteric regulation of HIV-1 reverse transcriptase by ATP for nucleotide selection. *PLoS ONE* 5, e8867. doi:10.1371/journal.pone.0008867

Yokoyama, M., Naganawa, S., Yoshimura, K., Matsushita, S., and Sato, H. (2012). Structural dynamics of HIV-1 envelope Gp120 outer domain with V3 loop. *PLoS ONE* 7, e37530. doi:10.1371/journal.pone.0037530

Zeitler, C. E., Estes, M. K., and Venkataram Prasad, B. V. (2006). X-ray crystallographic structure of the Norwalk virus protease at 1.5-A resolution. *J. Virol.* 80, 5050–5058.