Foot-and-mouth disease virus (FMDV) causes a widespread and economically devastating disease of domestic livestock. Although FMDV vaccines are available, political and technical problems associated with their use are driving a renewed search for alternative methods of disease control. The viral RNA genome is translated as a single polypeptide precursor that must be cleaved into functional proteins by virally encoded proteases. 10 of the 13 cleavages are performed by the highly conserved 3C protease (3Cpro), making the enzyme an attractive target for antiviral drugs. We have developed a soluble, recombinant form of FMDV 3Cpro, determined the crystal structure to 1.9 Å resolution, and analyzed the cleavage specificity of the enzyme. The structure indicates that FMDV 3Cpro adopts a chymotrypsin-like fold and possesses a Cys-His-Asp catalytic triad in a similar conformation to the Ser-His-Asp triad conserved in almost all serine proteases. This observation suggests that the dyad-based mechanisms proposed for this class of cysteine proteases need to be reassessed. Peptide cleavage assays revealed that the recognition sequence spans at least four residues either side of the scissile bond (P4–P4′) and that FMDV 3Cpro discriminates only weakly in favor of P1-Gln over P1-Glu, in contrast to other 3Cpro enzymes that strongly favor P1-Gln. The relaxed specificity may be due to the unexpected absence in FMDV 3Cpro of an extended Gln. The relaxed specificity may be due to the difficulty in distinguishing vac
cinated from infected animals, which currently acts as an impediment to the resumption of international livestock trade (1). The use of vaccines is also hampered by the existence of seven distinct serotypes and multiple subtypes of the virus; vaccine stocks must therefore be constantly updated to protect against the circulating strains most likely to give rise to new outbreaks (3). Moreover, at least 3–4 days are required following vaccination for a protective response to develop, during which time outbreaks may gain a wider foothold (1, 3). In view of these problems, which are exacerbated by the emergent threat that the virus may be used as a bioterrorism agent (1, 4), additional measures for disease control are required. One promising strategy is to develop drugs to inhibit viral enzymes, such as proteases, that are necessary for replication and propagation of the disease, an approach that has been adopted with some success for human immunodeficiency virus (5) and is now being applied to a number of other viral targets (6, 7). In the case of FMDV, such antiviral drugs could be deployed as soon as an outbreak is detected and thereby provide a potent adjunct to vaccines for faster control of incipient epidemics.

FMDV, a picornavirus related to poliovirus (PV), human rhinovirus (HRV), and hepatitis A virus (HAV), possesses a single-stranded, positive sense RNA genome containing one open reading frame that is translated as a polypeptide precursor. The cleavage of this polypeptide into the functional proteins required for viral replication is performed by virally encoded proteases. The most important of these is the conserved 3C protease (3Cpro), which in FMDV performs 10 of the 13 cleavages (4); this enzyme is therefore an appealing target for antiviral drugs.

Structural studies on HAV, HRV, and PV 3Cpro enzymes have established that these belong to an unusual family of cysteine proteases that have an overall fold similar to chymotrypsin (8–10). They invariably possess a conserved Cys-His-Asp/Glu catalytic triad at the active site, which is similar to the Ser-His-Asp triad found in the vast majority of serine proteases. In serine proteases, the His residue serves to prime the Ser side chain for nucleophilic attack on the scissile bond by abstracting a proton; the role of the Asp is to provide electrostatic stabilization of the resulting positive charge on the peptide bond by abstracting a proton; the role of the Asp is to provide electrostatic stabilization of the resulting positive charge on the peptide bond.
Structure and Specificity of FMDV 3Cpro

The structures of picornaviral 3C enzymes indicate that their active sites deviate in significant details from the canonical Ser-His-Asp configuration, leading to suggestions that they also differ mechanistically. In HAV 3Cpro, the catalytic Asp is found to be directed away from the active site and is proposed not to play a significant role in catalysis; its function is possibly replaced by a tyrosine residue that is suggested to be negatively charged (8, 12). In HRV and PV 3C proteases, which are very similar to one another, the position normally occupied by the catalytic Asp is taken by Glu (9, 10), a substitution that is extremely rare in serine proteases (11). These unexpected features have generated new hypotheses regarding the catalytic mechanism of picornaviral 3C proteases that have yet to be fully resolved (8–10, 12, 13). At the same time, structural studies have also provided a detailed structural basis for the search for effective inhibitors (14–16). In order to shed further light on the possible catalytic mechanism of chymotrypsin-like cysteine proteases and to establish a molecular framework for anti-FMDV drug design, we have developed a soluble, recombinant form of FMDV 3Cpro, determined the crystal structure, and analyzed its proteolytic specificity.

MATERIALS AND METHODS

Protein Purification and Crystallization—FMDV 3Cpro strain A1056 was expressed in BL21(DE3) pLyS8 Escherichia coli from a pET-28 vector modified to add a thrombin-cleavable N-terminal polyhistidine tag. Selenomethionine-labeled protein was produced in B834(DE3) E. coli grown in minimal medium supplemented with essential nutrients and selenomethionine (Molecular Dimensions). The expressed protein was purified on TALON resin (BD Biosciences), digested overnight with thrombin, reapplied to the TALON resin to remove the cleaved tag, and further purified by gel filtration on a Hi Load 16/60 Superdex 75 column (Amersham Biosciences). Peak fractions were concentrated to 10 mg/ml in 50 mM HEPES, pH 7.0, containing 200 mM NaCl, 1 mM EDTA, 1 mM β-mercaptoethanol. The purified protein contains residues 1–207 of the 213 amino acids of wild-type 3Cpro; it has an additional Gly at the N terminus remaining from the thrombin cleavage and one additional His at the C terminus. In addition it contains mutations C95K and C142S, which were designed via homology modeling to enhance solubility. Incorporation of these mutations into a 3B-3C fusion construct (17) had no effect on 3Cpro-mediated removal of the 3B peptides upon expression in E. coli (data not shown). The mutation C163A was also made to eliminate the activity of the active site nucleophile. Crystals were obtained by sitting drop vapor diffusion over a 100-μl reservoir containing 2 μl of protein with 0.5 μl of 10% (v/v) Anapoe-X-405 and 2.5 μl of reservoir solution prepared by mixing 75 μl of 1.5 M sodium citrate, pH 6.5, with 25 μl 0.5 M ammonium phosphate.

Data Collection, Phasing, and Refinement—Crystals (typically 250 × 250 × 50 μm) were transferred to a 60/40 (v/v) mix of reservoir solution and saturated sucrose and frozen in a stream of N2 gas at 10 K immediately prior to data collection. Multil wavelength x-ray diffraction data to 1.9 Å collected from crystals of SeMet-labeled protein were processed and scaled using the CCP4 program suite (18). The crystals belong to the R3 space group and have two molecules in the asymmetric unit related by a noncrystallographic 2-fold axis (Table I). Since the crystals exhibited hemihedral twinning (19) with a twinning fraction of around 0.2, the diffraction intensities were detwinned using DETWIN. Initial selenium sites were located from data collected at peak and remote wavelengths using SHELDX and SHELXE (20) and refined with SHARP (21). Solvent flattening, histogram matching, and 2-fold non-crystallographic symmetric were applied in DM (22) to improve the phases. An initial model encompassing about 75% of the polypeptide was built automatically using ARP/wARP (23), completed manually in O, and refined using CNS (Table I). The two independent molecules of 3Cpro in the asymmetric unit are very similar (root mean square deviation in Cα positions = 0.4 Å); in each chain, electron density is visible starting from residue 7 and ending at residue 205, although loops A142–146, B142–146, and B76–81 are also missing due to disorder.

Peptide Cleavage Assays—Soluble single (C95K) and double (C95K/C142S) cysteine mutant proteases for peptide cleavage assays were generated in a construct that generates a 3B1–3B2–3B3–3C fusion protein, which contains 3Cpro fused to the three upstream 3B peptides found in the FMDV polyprotein (though 3B1 is truncated at its N-terminus) (17). Upon expression in E. coli, the 3Cpro cleaved itself from the 3B peptides and was purified in a single step to >95% homogeneity using TALON resin via its nontaggable C-terminal polyhistidine tag. Fresh preparations of the single mutant 3Cpro exhibited significantly higher specific activity, but the loss of enzyme activity under assay conditions and a propensity for aggregation made it difficult to obtain reproducible results. The additional C142S substitution in the B2-C2 loop of the double mutant is absolutely required to obtain a soluble and stable enzyme for cleavage assays. Although this mutation may lie close to the substrate binding site (Fig. 1A), there were no significant differences in peptide cleavage specificity compared with the single mutant (data not shown).

Peptide substrates were synthesized on an Advanced ChemTech Apex 396 multiple peptide synthesizer using a standard Fmoc (N-(9-fluorenyl)methoxycarbonyl) tert-butyl peptide synthesis strategy (24) and 1-hydroxybenzotriazole/2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluroniumhexafluorophosphate activation and purified to >96% homogeneity by reverse-phase HPLC using a Delta-Pak C18 Cartridge (15 μm, 25 × 100 mm). The identity of the peptides was confirmed by mass spectrometry.

Samples of peptide (50 μl) were incubated with the 3Cpro enzyme at 37 °C in 0.1 mM phosphate buffer, pH 7.4, containing 0.01 mM diithreitol, 1 mM EDTA, and 5% glycerol. The extent of reaction was determined using HPLC and analysis of the mixture by integration of the peaks representing the starting and product peptides. The identity of the product peptides was verified by MS analysis and/or comparison with HPLC traces of peptide standards. In all cases where cleavage occurred, this was found to be at the expected P1/P1’ junction. For the alamine-scanning experiments, reactions were quenched after a 4-h incubation by the addition of trifluoroacetic acid to 2.5% (v/v). Relative hydrolysis rates were calculated from the data assuming that the hydrolysis followed apparent first order kinetics.

RESULTS

Crystallography and Structure Determination—Recombinant FMDV 3Cpro strain A1056 may be highly expressed in E. coli (17, 25), but we found that the native protein precipitates relatively rapidly following purification. Storage at 5–10 mg/ml over a period of days led to the formation of a proteinaceous skin over the top of the solution, which makes this form incompatible with crystallization studies. Extensive mutagenesis trials of amino acids predicted to be surface-exposed (on the basis of modeling using HAV 3Cpro, the nearest homologue of the known structure (8)) were performed in an effort to enhance protein solubility. Dynamic light scattering analysis of the aggregation properties of the resulting mutant enzymes, coupled with activity assays, eventually yielded a soluble, active variant.2 In this mutated FMDV 3Cpro, two solvent-exposed cysteine residues (C95K and C142S) have been replaced, and the C terminus has been trimmed by six residues (for details, see “Experimental Procedures”). For crystallization, the active site nucleophile was also replaced to eliminate proteolytic activity (C163A). The refined structure, solved to 1.9-Å resolution by multil wavelength anomalous dispersion, has an Rfree of 24.6% and good stereochemistry (Table I).

Structure of FMDV 3Cpro and Comparison with Other Proteases—FMDV 3Cpro adopts a chymotrypsin-like fold, which typically consists of two β-barrel domains, each composed of a pair of four-stranded anti-parallel β-sheets that pack together to form a shallow peptide binding cleft (Fig. 1A). According to the naming convention adopted for these proteases, strands B and E within each domain usually contribute to both sheets. Although the other picornaviral 3C proteases (8–10) and the 3C-like NiA protease from tobacco etch virus (TEVpro; 16%) have similar structures and specificities, there are no significant differences in peptide cleavage specificity compared with the single mutant (data not shown).

2 Birtley, J. R., and Curry, S. (2005) Acta Crystallogr. Sect. D., in press.

Downloaded from http://www.jbc.org/ by guest on July 24, 2018
In particular, the B2 strand does not extend to participate in the E2-F1-C2-B2 β-sheet found in the C-terminal β-barrel of other 3Cpro structures (Fig. 1A); instead, this segment of polyepitope (residues 137–139) is separated from the C2 strand and pairs up with the residues preceding the A2 strand (residues 107–109) to form a small β-sheet (A2–B2) that is not observed in the HAV, HRV, or PV 3C proteases or in other picornaviruses. The A1-B1, C1-D1, and A2-B2 loops in FMDV 3Cpro also differ significantly from the conformations observed in other picornaviral 3C proteases (Fig. 1E), and although they project away from the active site, they may contribute to substrate specificity.

The determination of the crystal structure of HAV 3Cpro unexpectedly revealed that residue Asp84, which was predicted to form the catalytic triad along with His44 and Cys172, is actually redirected away from the active site by a salt bridge interaction with the side chain of Lys302 (8) (Fig. 2). Tyr144, which points toward the catalytic His in this enzyme was proposed to compensate for this unusual arrangement by providing sufficient electrostatic stabilization to activate the charge relay system (8, 14), and, based on homology modeling, it was recently proposed that FMDV 3Cpro would operate a similar system (28). However, the sequence alignment based on the structure reported here shows that neither Tyr144 nor Lys302 from HAV 3Cpro is conserved in the FMDV enzyme (Fig. 3). Instead, Asp84 in the crystal structure of FMDV 3Cpro makes a hydrogen bond interaction with His46 of the catalytic triad, and the enzyme therefore has a Cys-His-Asp triad that is very similar in configuration to the equivalent triad in the 3C-like TEVpro (26) and to the Ser-His-Asp triad found in the vast majority of serine proteases (11) (Fig. 2). This similarity extends to encompass the side-chain hydroxyl group of Ser192 in FMDV 3Cpro (Ser214 in chymotrypsin), which is positioned to stabilize the conformation of the side chain of the triad residue Asp84, an interaction that is conserved in many chymotrypsin-like serine proteases (11). Intriguingly, this Ser residue is replaced by Val192 in HAV 3Cpro, loss of this interaction may therefore contribute to the reorientation of the catalytic Asp84 in that enzyme.

The face of the protease opposite to the active site contains a basic patch that is implicated in RNA binding during picornaviral RNA replication (29–31). As in other picornavirus 3C proteases, this is centered on the conserved motif (K/R)(F/V)RDI (95KVRDI99 in the present structure, due to the Cys95→W mutation introduced to aid solubility; exceptionally for atheroviruses, the motif has the sequence CVRDI in wild-type FMDV A1041, which could conceivably be due to a single base substitution during passage of virus stocks in tissue culture) (32, 33). The basic residues of the motif, Lys95 and Arg97, are exposed on the surface of FMDV 3Cpro and form the core of a basic patch that also contains Lys98, Arg100, Arg102, Arg126, and Lys203.

**Peptide Cleavage Specificity**—Comparison of the cleavage rates of peptides corresponding to the 10 polyproteins described above revealed a surprising degree of variation (Table II). The most susceptible peptides correspond to the VP1/2A and 2B/2C junctions, which were completely cleaved within 6 h, whereas peptides corresponding to the VP2/VP3, 2C/3A, 3A/3B1, 3B1/3B2, and 3B2/3B3 junctions were uncleaved after a 24-h incubation under our experimental conditions (see “Experimental Procedures”). The VP1/2A junction in HRV and PV is the primary cleavage site of a different viral protease, 2Apro (32). The 2B/2C cleavage site is also found to be one of the most efficiently processed protein junctions in peptide cleavage assays performed with PV and HAV 3Cpro (34, 35), although it is not efficiently cleaved as a peptide by HRV 3Cpro (35). The primary 3Cpro cleavage site in vivo is variously reported to occur at the 2A/2B, 2B/2C, 2C/3A, or 3A/3B junctions (32, 33, 36, 37); clearly, factors such as the folding of the viral polyprotein precur-

---

### Table I

| Parameters                              | Values               |
|-----------------------------------------|----------------------|
| Diffraction data and phasing            |                      |
| Space group                             | R3                   |
| a (Å)                                   | 141.59               |
| c (Å)                                   | 43.68                |
| Diffraction data and phasing            |                      |
| Resolution range (Å)                    | 35.3 to 1.9          |
| Independent reflections                 | 2572                 |
| Multiplicitya                           | 9.5 (9.1)            |
| Completeness (%)                       | 100 (100)            |
| R_{F}I                                  | 21.9 (3.1)           |
| R_{merge} (%)                          | 7.5 (64.4)           |
| R_{wo} (%)                              | 5.0 (37.1)           |
| Phasing power                           | 1.78                 |
| Phasing power                           | 1.66                 |
| Figure of merit (obs)                   | 0.49                 |
| Figure of merit (DM)                    | 0.80                 |
| Model refinement                        |                      |
| Nonhydrogen atoms/waters                | 2829/98              |
| R_{model}/R_{free} (%)                  | 21.7/24.6            |
| Root mean square deviation from ideal bond length (Å) | 0.006 |
| Root mean square deviation from ideal angles (degrees) | 1.31 |
| Ramachandran plot, % favored/additional (%) | 92/27.8 |
| Average B-factor (Å²)                   | 35.5                 |
| Protein Data Bank code                  | 2bhg                 |

*a Values for the outermost resolution shell (2.0 to 1.9 Å) are given in parentheses.

*b R_{merge} = 100\%\sum_{i,j}=I_{ij}/\sum_{j}I_{ij}, where I_{ij} represents the weighted mean intensity of the symmetry-related reflections.

*c R_{model} = 100\%\sum_{j}(|F_{o}|−|F_{c}|)/\sum_{j}|F_{o}|, where F_{o} and F_{c} represent the observed and calculated structure factors, respectively.

*d R_{free} is the R_{model} calculated using a randomly selected 5% sample of reflection data omitted from the refinement.
sor may affect site accessibility in infected cells. The two most susceptible peptides in the FMDV polyprotein contain the sequence KQL at positions corresponding to P2–P1/H11032. However, the 2C/3A peptide, which contains KQI at these positions, is resistant to cleavage, suggesting that amino acids outside the core sequence are also important for efficient cleavage.

To characterize the cleavage specificity of the enzyme in greater detail, alanine-scanning mutagenesis was performed on a 12-mer version of the VP1/2A peptide (IAPAKQ2LLNFDDL) that spans the P6–P6/H11032 positions. These experiments identified P4-Pro, P2-Lys, P1-Gln, and P4/H11032-Phe of this sequence as the most important residues for specific recognition by 3Cpro (Fig. 4), since substitution of these residues by Ala abrogated peptide cleavage under our conditions. The substrate recognition site therefore spans at least eight residues (P4–P4/H11032) and is longer than has previously been appreciated for 3C proteases. Substitution of P1/H11032-Leu and P2/H11032-Leu by Ala had smaller but still significant detrimental effects on cleavage. The P1 and P4 positions have previously been shown to be important in peptide substrates for HRV and HAV 3Cpro (38, 39) and are the most highly conserved positions in the FMDV polyprotein cleavage sites (40). Whereas the deleterious effect of Ala substitutions for P2-Lys and P4/H11032-Phe indicates that these residues contribute substantially to substrate recognition, it should be noted that they are not conserved in the polyprotein (Table II). The impact of alanine-scanning mutagenesis may well depend on the peptide sequence being probed, and further work is necessary to extract more detailed information on cleavage specificity. Nevertheless, the present results strongly suggest that the P4–P1 sequence (Pro-Ala-Lys-Gln) would serve as a promising template for the design of peptide-mimetic inhibitors.

Most picornavirus 3C proteases are generally specific for P1-Gln and discriminate strongly (48-fold in the case of HAV) against P1-Glu (38, 39). In marked contrast to these enzymes, FMDV 3Cpro is only weakly selective for Gln over Glu at this position (about 2-fold in rate) (Fig. 4). This is consistent with the occurrence of P1-Glu at many of the 3Cpro cleavage sites within the FMDV polyprotein (40) and the cleavage of eukaryotic initiation factor 4AI at a Glu-Val junction (17).

**DISCUSSION**

We report here the first structural insights into FMDV 3Cpro. Our results confirm that the enzyme adopts the expected chymotrypsin-like fold but reveals several unexpected aspects of the structure that undoubtedly impact on its proteolytic activity. Strikingly, FMDV 3Cpro lacks the extended β-ribbon that folds...
Structure and Specificity of FMDV 3Cpro

Synthetic peptides spanning eight residues either side of the ten 3Cpro cleavage positions were incubated with enzyme (10 μM) as described under “Materials and Methods” and then analyzed for extent of cleavage by HPLC. Marked differences in cleavage rates were found, with rapid, complete cleavage occurring for peptides 3 and 4 within 6 h; partial cleavage was found for peptides 2, 9, and 10 after 24 h incubation; little or no measurable cleavage was found for the remaining peptides even after 24 h. At reduced enzyme concentration (1 μM), peptide 3 was found to hydrolyze slightly faster than peptide 4.

Table II

| Peptide Polyprotein junction | Sequence (P–P) | Cleavage |
|-----------------------------|---------------|----------|
| 1 VP2/VP3                   | AGERPGKKE    | GIPVPACA | – |
| 2 VP3/VP1                   | LPIDPRTQ     | TTTTGESA | + |
| 3 VP1/2A                    | KIIAPAQK     | LNPFDLLK | +++ |
| 4 2B/2C                     | DLERAEQK     | LKARDIND | +++ |
| 5 2C7A                      | SSHPIFKQ     | ISSPQKS  | – |
| 6 3A/3B1                    | AEDPQOAE     | GPVSPQLE | – |
| 7 3B1/3B2                   | RALPQQQ      | GPYAGQME | – |
| 8 3B2/3B3                   | VAPKPVVKE    | GPYEGQPK | – |
| 9 3B3/3C                    | ARNLITYE     | SGPAPTDL | + |
| 10 3C3D                     | VDPEEPHE     | GLVIVTRD | + |

Comparative cleavage of peptides corresponding to polyprotein junctions in FMDV A10p.

Fig. 2. Comparison of the active sites of 3C proteases shows that the catalytic triad found in FMDV 3Cpro most closely resembles that found in the serine protease chymotrypsin. To aid the comparison, the Cys315 side chain has been restored in the FMDV 3Cpro structure by modeling (since this is Ala in the crystal structure). Although the hydrogen-bonding geometry between Cys315 and His66 is nonideal, this may be due to small perturbations in the triad configuration resulting from the mutation to the smaller Ala side chain in the FMDV 3Cpro structure. The other crystal structures shown are chymotrypsin (Protein Data Bank code 4cha(55)), HAV 3Cpro (Protein Data Bank code 1hav(12)), and HRV 14 3Cpro (coordinates provided by D. Matthews).

TABLE II

| Peptide Polyprotein junction | Sequence (P–P) | Cleavage |
|-----------------------------|---------------|----------|
| 1 VP2/VP3                   | AGERPGKKE    | GIPVPACA | – |
| 2 VP3/VP1                   | LPIDPRTQ     | TTTTGESA | + |
| 3 VP1/2A                    | KIIAPAQK     | LNPFDLLK | +++ |
| 4 2B/2C                     | DLERAEQK     | LKARDIND | +++ |
| 5 2C7A                      | SSHPIFKQ     | ISSPQKS  | – |
| 6 3A/3B1                    | AEDPQOAE     | GPVSPQLE | – |
| 7 3B1/3B2                   | RALPQQQ      | GPYAGQME | – |
| 8 3B2/3B3                   | VAPKPVVKE    | GPYEGQPK | – |
| 9 3B3/3C                    | ARNLITYE     | SGPAPTDL | + |
| 10 3C3D                     | VDPEEPHE     | GLVIVTRD | + |
chymotrypsin-like proteases (including HRV 3C\textsuperscript{pro} (16) and TEV\textsuperscript{pro} (26)) complexed with peptides or peptide-based inhibitors indicates that the interaction between Tyr\textsubscript{143} and the catalytic His could not be maintained upon substrate binding (data not shown). It remains possible that the crystal structure of HAV 3C\textsuperscript{pro} represents a nonproductive form of the enzyme. Structural analysis of 3C\textsuperscript{pro}-peptide complexes should help to further our understanding of the enzyme mechanisms.

The S1 pocket that accommodates the P1 amino acid side chain of the substrate is composed mainly of residues from the C\textsubscript{2}-D\textsubscript{2} loop and the C-terminal end of strand E\textsubscript{2} (Figs. 1, 1E, and 5). This pocket contains His\textsubscript{181}, which is coordinated to Tyr\textsubscript{154} by a buried hydrogen bond, a feature that is observed in other 3C\textsuperscript{pro} enzymes (Figs. 3 and 5) (although in HAV 3C\textsuperscript{pro}, the equivalent Tyr residue is missing, and the His is instead hydrogen-bonded to a water molecule that is itself coordinated by a buried Glu side chain (8, 12)). It has been argued that this coordination allows the uncharged His to donate a hydrogen bond to the side-chain carbonyl oxygen of the P1-Gln of the peptide substrate (8, 10, 12). In HRV-2 3C\textsuperscript{pro}, structural studies with peptidomimetic inhibitors suggested that the selectivity for Gln over Glu at P1 may also require Thr\textsubscript{142} (equivalent to Thr\textsubscript{158} in FMDV 3C\textsuperscript{pro}), which donates a hydrogen bond from its side-chain hydroxyl to the P1-Gln side-chain carbonyl, whereas the main-chain carbonyl of this Thr accepts one from the P1-Gln side-chain.

**FIG. 3.** Structure-based sequence alignment of picornavirus 3C\textsuperscript{pro} enzymes. The sequence alignment was performed by the EBI SSM server (www.ebi.ac.uk/msd-srv/ssm/) using the known structures of 3C proteases from FMDV (strain A10\textsubscript{a}), HAV (strain HM-175), HRV (type 2), and PV (type 1 Mahoney). Residues that are identical in two or more sequences are shaded black; homologous residues are shaded gray. The secondary structure for FMDV 3C\textsuperscript{pro} is shown along the top. The asterisks indicate the positions of the C95K and C142S substitutions introduced to improve solubility (see “Materials and Methods”). The hash marks denote residues in the catalytic triad. Sequences flanking the S1 specificity pocket, which exhibit significant variation between different picornaviruses, are boxed.

**FIG. 4.** Effect of alanine-scanning substitution on cleavage activity by the double cysteine-substituted 3C\textsuperscript{pro} enzyme of a peptide substrate based on the sequence of the VP1/2A junction (IAP\textsubscript{AKQ} , ILNFDL). The chart compares rates of hydrolysis relative to that of the starting peptide for modified sequences that incorporate an alanine substitution. Where no bar appears, the observed rate was effectively zero. (Note that P6 and P3 are both alanine in the original sequence, so these positions are not represented.) The graph also shows the relative hydrolysis rate when the P1-Gln is replaced by Glu (denoted P1E).
amide group (16, 46) (Fig. 5B). An almost identical configuration, observed in the complex of TEVpro with a peptide substrate, was proposed to explain the P1-Gln selectivity of that enzyme (26) (although it should be noted that whereas P1-Gln is conserved at all TEVpro cleavage sites in the viral polyprotein (47) and the enzyme discriminates in favor or P1-Gln over most other amino acids, the particular effect of incorporating P1-Glu in substrates has not yet been tested (48)). However, the results presented here, coupled with more recent findings, indicate that the presence of this His-Tyr-Thr cluster in the S1 pocket does not fully account for P1-Gln selectivity. FMDV 3Cpro, which is not strongly selective for P1-Gln over P1-Glu, contains this cluster (Fig. 5A), whereas HAV 3Cpro, which is selective, lacks the Thr residue that is proposed to differentiate between Gln and Glu at P1. More tellingly, perhaps, the His-Tyr-Thr arrangement found in the S1 specificity pocket is conserved in a number of Glu-specific serine proteases (49–51), which tend to discriminate against P1-Gln and thus display the reverse selectivity to that seen in picornaviral 3C proteases. These serine proteases select for P1-Glu by placing a counter charge to the negative carboxyl of the P1-Glu side chain in the S1 pocket, either from a basic side chain or from the amino terminus of the polypeptide (49–52) or by using specific hydrogen bond contacts (Fig. 5C) (53). Thus, in picornavirus 3C proteases, the selection of P1-Gln over P1-Glu may simply be due to the absence of an appropriate counter charge. A definitive explanation for the more relaxed selectivity of FMDV 3Cpro at P1 is not yet possible, largely because the impact of remote interactions between the protease and its substrate are difficult to assess (11), but the more open active site in this enzyme that results from the absence of a B2-C2-helix-H9252-strand in the S1 pocket (Figs. 1E and 3), is perhaps surprising that picornavirus 3C proteases maintain a common preference for P1-Gln.

Modeling studies based on the structures of a range of 3C proteases and serine proteases complexed with peptides or peptide-based inhibitors reveal consistent backbone and side-chain configurations for the P1–P4 region (data not shown) and allow us to suggest plausible locations on the surface of FMDV 3Cpro for peptide interactions involving P2–P4 residues (Fig. 5A). The P2-Lys of the VP1/2A peptide extends to one side of the catalytic His46 and is likely to interact with the side-chain hydroxyl of Tyr190, a residue conserved in FMDV but not in other picornavirus 3C proteases; conceivably, this Tyr may also interact with Glu and His residues that occur at P2 in other FMDV 3Cpro cleavage sites (Table II). The P3 side chain is usually solvent-exposed and contributes little to specificity (11). The preference for a hydrophobic Pro side chain at P4 (40) is likely to be due to this side chain being pinned between the apolar moieties of Val188 and Tyr190. On the other side of the cleavage site, Ile30, Leu47, Ala122, and the aliphatic portion of Glu20 may provide hydrophobic contacts for the P1, P2, and P4 positions in the optimal substrate, consistent with the general preference for residues with a significant apolar component in these positions (Table II).

FMDV 3Cpro is an attractive target for drug design, since there
are no known cellular homologues in susceptible hosts. In addition, since FMDV 3C\textsuperscript{pro} sequences are highly conserved, it may be possible to develop protease inhibitors that have a broad activity against a range of different serotypes. For example, although FMDV capsid protein sequences from the seven serotypes are only 50–70% conserved, complicating vaccine formulation, their 3C\textsuperscript{pro} sequences are 82–85% identical. These figures rise to 91–97% if one excludes the three SAT serotypes that are usually confined to sub-Saharan Africa (2) and in any case form a distinct conserved group in terms of 3C\textsuperscript{pro} sequences (88–98% identical) (28). Mapping the variation found within 41 distinct FMDV 3C\textsuperscript{pro} sequences spanning all seven serotypes onto the 3C\textsuperscript{pro} structure (Fig. 6) reveals that these are almost entirely peripheral to the substrate-binding site. For the two residue positions closest to the substrate-binding site that exhibit some variation (Ile\textsuperscript{30} and Ala\textsuperscript{166}), the vast majority of this variation is due to differences between the SAT and non-SAT groups. Ile is substituted conservatively by Val in just one of 33 known non-SAT sequences and is strictly conserved as Leu in the SAT group; Ala is strictly conserved in the non-SAT group of viruses and substituted by Val in the SAT group. This emphasizes the highly conserved nature of the binding pocket on FMDV 3C\textsuperscript{pro} and highlights the potential for developing a broad spectrum inhibitor.

Acknowledgments—We thank Chris Proud, Wei Li, and Graham Belshaw for 3C\textsuperscript{pro} cDNA; David Barford and James Parker for help with dynamic light scattering measurements; David Matthews for providing coordinates of HRV14 3C\textsuperscript{pro}; and John Barton, Anne Dell, Maria Panico, and Howard Morris for mass spectral analysis. We are grateful to the staff at the BM14 (ESRF, Grenoble) and X13 (EMBL/DESY, Hamburg) synchrotron beamlines.
Crystal Structure of Foot-and-Mouth Disease Virus 3C Protease: NEW INSIGHTS INTO CATALYTIC MECHANISM AND CLEAVAGE SPECIFICITY
James R. Birtley, Stephen R. Knox, Agnès M. Jaulent, Peter Brick, Robin J. Leatherbarrow and Stephen Curry

J. Biol. Chem. 2005, 280:11520-11527.
doi: 10.1074/jbc.M413254200 originally published online January 14, 2005

Access the most updated version of this article at doi: 10.1074/jbc.M413254200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 53 references, 11 of which can be accessed free at
http://www.jbc.org/content/280/12/11520.full.html#ref-list-1