Proteolytic processing of a human astrovirus nonstructural protein

David Kiang and Suzanne M. Matsui

Division of Gastroenterology, Center for Clinical Sciences Research, Room 3115, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, CA 94305-5187, USA, and VA Palo Alto Health Care System, Palo Alto, CA 94304, USA

To analyse the activity of the putative 3C-like serine protease encoded in open reading frame (ORF)-1a of human astrovirus serotype 1 (HAstV-1), ORF-1a was transcribed and translated in vitro. Translation products, identified by immunoprecipitation with specific antibodies against recombinant C-terminal ORF-1a fragments, include the full-length 101 kDa (p101) protein and a 38 kDa band (p38). In addition, a 64 kDa protein (p64) was immunoprecipitated by an anti-FLAG antibody when a FLAG epitope was inserted at the N terminus of the ORF-1a product. Mutation of the amino acids predicted to form the catalytic triad of the HAstV-1 3C-like serine protease (Ser-551, Asp-489, His-461) resulted in undetectable levels of p38, supporting the involvement of the HAstV-1 3C-like serine protease and the importance of these amino acids in the processing of p101 into p38 and p64. N-terminal deletions of up to 420 aa of p101 that did not involve the predicted 3C-like serine protease motif did not alter levels of p38 compared to wild-type. C-terminal deletion analysis localized p38 to the C terminus of ORF-1a. Mutation of the P1 residue of the predicted cleavage site, which is conserved among known human and sheep astrovirus sequences, resulted in no detectable p38, supporting cleavage at the Gln-567 $\rightarrow$ Thr-568 dipeptide. These results suggest that p101 is cleaved into an N-terminal p64 fragment and a C-terminal p38 product at Gln-567 $\rightarrow$ Thr-568 in a process that is dependent on the viral 3C-like serine protease.

Introduction

Human astroviruses, first described as a cause of infantile gastroenteritis in 1975 (Appleton & Higgins, 1975; Madeley & Cosgrove, 1975), belong to the Astroviridae family of nonenveloped single-stranded positive-sense RNA viruses (Monroe et al., 1995). Recent epidemiological studies indicate that HAstV is a medically important cause of acute gastroenteritis worldwide (Cruz et al., 1992; Gaggero et al., 1998; Herrmann et al., 1991; Lopez et al., 2000; McLver et al., 2000; Medina et al., 2000; Mitchell, 1999; Naficy et al., 2000; Steele et al., 1998). Eight serotypes of HAstV have been described to date with serotype 1 being the most prevalent (Lee & Kurtz, 1994; Taylor et al., 2001).

The complete sequences of HAstV-1 (L23513, Z25771), -2 (L13745), -3 (AF141381) and -8 (Mendez-Toss et al., 2000), sheep astrovirus (SAstV; Y15937, AF260508), turkey astrovirus (TArtV) (Koci et al., 2000) and avian nephritis virus (ANV; an avian astrovirus) (Imada et al., 2000) have been determined. The ~6.8 kb polyadenylated viral genome is composed of three open reading frames (ORFs): 1a, 1b and 2. The structural proteins, encoded by ORF-2, are expressed from a 2.4 kb subgenomic RNA that is coterminal with the 3' end of the genome and detectable in infected cells (Lewis et al., 1994; Matsui et al., 1993; Monroe et al., 1993; Willcocks & Carter, 1993). Putative nonstructural proteins (NSPs), including a 3C-like serine protease motif and an RNA-dependent RNA polymerase (RdRp) motif, are encoded by ORF-1a and ORF-1b, respectively (Jiang et al., 1993; Lewis et al., 1994). Expression of the RdRp is regulated by (−1) ribosomal frameshifting (Lewis & Matsui, 1996; Marczinke et al., 1994), a mechanism used by HIV-1, other retroviruses and coronaviruses (Bredenbeek et al., 1990; Jacks, 1990). ORF-1a also encodes a putative nuclear localization signal (NLS) (Jiang et al.,...
Fig. 1. HAstV-1 ORF-1a product and recombinant constructs. The ORF-1a product (aa 1–920) with known motifs and epitopes is depicted at the top. TM, predicted transmembrane helices; SP, serine protease motif (aa 454–569) with catalytic triad H-461, D-489, S-551 indicated; NLS, nuclear localization signal (aa 662–682); IRE, immunoreactive epitope (aa 757–899). The proposed cleavage site at Q567 → T568 is indicated by the arrow. Lines labelled 1aC2 and 1aC1 indicate regions of the ORF-1a product that were expressed in *E. coli* as GST fusion proteins (asterisks represent the GST fusion partner) and used for polyclonal antiserum production. (A) Constructs pAV1a, pAV1a/Δprot, pFLGAV1a; (B) five N-terminal deletion constructs; (C) two C-terminal deletion constructs. In (A), the dotted boxes indicate that these constructs extend into ORF-1b (to aa 1024), HA represents the substitution of HAstV-1 aa 546–554 (TQDGMSGAP) by the HA epitope (YPYDVPDYA) and the diamond represents the FLAG epitope. For all panels, calculated molecular size (in kDa) of the expressed ORF-1a products is indicated in the right-hand column.

Processing of polyproteins by virally encoded proteases is a necessary event in the replication of many positive-stranded RNA viruses (Dougherty & Semler, 1993). The 3C and 2A protease of picornaviruses and many others belong to the trypsin family of proteases (Bazan & Fletterick, 1988). For picornaviruses, successful infection of susceptible cells is dependent upon proteolytic processing of a single large polyprotein, and the 3C protease is involved in the majority of these cleavage events (Palmenberg, 1990). An ORF-1a-encoded 3C-like serine protease motif of astrovirus was identified by alignment with 3C and other 3C-like protease motifs (Jiang et al., 1993). By analogy, the HAstV 3C-like protease likely plays an important role in processing astroviral proteins produced during infection.

Substrate cleavage by serine proteases involves the catalytic triad consisting of His, Asp and Ser. For serine proteases, the Ser serves as the nucleophile by contributing an electron to the carbonyl carbon of the hydrolysed peptide bond. It is unusually reactive and, like the His and Asp residues, has a conserved location in the spatial configuration of the active site of the protease (Dougherty & Semler, 1993).

At present, limited information is available about astrovirus NSPs, and the mechanisms by which they are processed are not known. In this report, we present evidence for proteolytic activity of the HAstV-1 3C-like serine protease encoded by ORF-1a. Among potential cleavage products detected by *in vitro* expression of ORF-1a, two major products – the focus of this study – were sensitive to mutations of the 3C-like serine protease: an N-terminal cleavage product (p64) and a C-terminal cleavage product (p38). N- and C-terminal deletion analyses were used to further characterize this proteolytic
process. The results of this study provide a detailed view of the processing of the ORF-1a product in vitro.

Methods

**Plasmid constructs.** DNA cloning was performed using established procedures (Maniatis et al., 1982). The transcription vector pTM1 (kindly provided by B. Moss, NH, Bethesda, MD, USA) contains a promoter for T7 RNA polymerase, a T7 terminator and an encephalomyocarditis internal ribosomal entry site (IREs) to enhance translation (Moss et al., 1990). pAV1α/Δprot and pAV1α (Fig. 1A) were derived from pAVIC (Geigenmüller et al., 1997) and kindly provided by U. Geigenmüller-Günirke. These constructs contain the entire ORF-1α (nt 86–3156) of HAstV-1 placed into the Ncol and Sfil sites of pTM1. In the expressed product of pAV1α/Δprot, a 9 aa region of ORF-1α (aa 540–554, TQDGMSGAP), containing Ser551 of the 3C-like protease catalytic triad and two predicted substrate-binding amino acids (Thr-546 and Gln-547), is replaced by the influenza virus haemagglutinin nonapeptide epitope (YPYDVPDYA, Boehringer Mannheim), which is not antigenically active in this context.

N-terminal deletion constructs, pAV1α/NΔ1–/NΔ5 (Fig. 1B), were generated by PCR using one of five positive oligonucleotides (AV1α-F1, 5’ TATGGATCCATGGGACGATGCTTAC 3’, AV1α-F2, 5’ TATGGATCCATGGCCTGGTTTCTTACCCA G 3’, AV1α-F3, 5’ TATGGAATTCATGGCATATG 3’, AV1α-F4, 5’ TATGGAATTCATGGCATATG 3’, AV1α-F5, 5’ TATGGAATTCATGGCATATG 3’) in conjunction with a 3’ negative-sense primer, AV1α-R1 (5’ TTCCACGCTATCAAATGTTG 3’). The resulting PCR products were digested with Ncol, and ligated into the Ncol and Sfil sites of pTM1.

Constitutive truncations, pAV1α/CA1 and CA2 (Fig. 1C), were produced by digestion of pAV1α with SacI (nt 2720) and Sfil (nt 1657)/Stud, respectively. For each digest, the fragment containing pTM1 and the 5’ end of the astrovirus sequence was purified. T4 DNA polymerase was used to remove the 3’ overhang in the SacI-digested fragment and to fill in the 5’ overhang in the Sfil-generated fragment. The resulting plasmids were then cycled by ligating the blunt ends.

A pTM1 vector that allows for incorporation of a FLAG octapeptide (DYKDDDDK) in the N terminus of the expressed product was constructed by annealing two synthetic oligonucleotides (5’ CATGATGCATCCCCGGGATGCAATACAAGAGGACGATGACAGAACGCCATGGAGCT 3’ and 5’ CAGTGGGCCTGTCATGCGTCGTCCTTGTAGT-TCCATCCGCGATGAT 3’) to form a duplex containing Ncol and Sfil overhangs, which was inserted into the Ncol and Sfil sites of pTM1. The insertion of the duplex allowed for the original Ncol site of the pTM1 multiple cloning site to be destroyed, thereby facilitating the insertion of astrovirus sequence into a newly created Ncol site downstream of the FLAG marker. To facilitate future cloning, additional restriction sites (SacI and Clal) were also included upstream of the sequence encoding the FLAG epitope. The resulting construct, pFLG, was used to generate the clones pFLGAV1α and pFLGAV1α/Δprot by digesting pAV1α and pAV1α/Δprot with Ncol and Sfil and introducing the ORF-1α-containing fragments into the Ncol and Sfil sites of pFLG.

DNA sequences of all plasmid constructs were confirmed by an ABI Prism 377 automated fluorescence-based Taq cycle sequencing system using dye-labelled terminators. Sequencing was performed by the VA/DDC Automated DNA Sequencing Facility.

**Site-directed mutagenesis.** Mutations of the putative catalytic domain, the putative cleavage site and control sites were performed using the Transformer Site-Directed Mutagenesis kit (Clontech) according to the manufacturer’s instructions. The mutations were either directly introduced into pAV1α, or when size became a limitation, into a smaller shuttle plasmid containing a region of astrovirus sequence which encoded the 3C-like serine protease motif. To create the shuttle plasmid, pAV1α was digested with Ncol (nt 1024) and the 5’ overhangs of the linearized plasmid were filled using T4 DNA polymerase. The DNA was then digested with Sfil (nt 2720) to release a 1696 bp fragment which was purified and inserted into the Sfil/Ncol sites of pUC18. Mutations were made in the shuttle vector and digested with SnaBII (nt 1218) or BglII (nt 1657)/SaciI (nt 2710) to excise the region of astrovirus sequence containing the desired mutation. This fragment was substituted for the same region of pAV1α containing the wild-type astrovirus sequence by restriction digestion with the corresponding restriction enzymes and ligation. All mutations were confirmed by DNA sequence analysis. Five constructs were produced by this method: pAV1α/H469L, pAV1α/D489V, pAV1α/S551A, pAV1α/Q567P and pAV1α/I578M/I579M.

**In vitro transcription–translation.** Non-linearized astrovirus plasmids were transcribed and translated in vitro using the TnT coupled in vitro transcription–translation kit (Promega) with T7 RNA polymerase, as recommended by the manufacturer. Translated proteins were labelled with Tran[35S]label (ICN) at 30 °C for 90 min. Input template and other conditions were standardized for all experiments.

**Expression of ORF-1α fusion proteins in E. coli.** Recombinant ORF-1a fusion proteins were expressed from plasmid constructs pGEX1aC1 and pGEX1aC2 (kindly provided by the late Terry L. Lewis, PhD). The 1aC1 and 1aC2 expressed products span aa 706–920 and aa 448–691, respectively (Fig. 1) and together encompass the C-terminal half of the ORF-1a product. These proteins, which contain the 26 kDa glutathione S-transferase (GST) of Schistosoma japonicum as the fusion partner, were purified using glutathione-Sepharose 4B (Pharmacia) as previously described (Smith & Johnson, 1988). Induction of 1aC1 expression in Escherichia coli by IPTG (0.1 mM) resulted in the production of milligram quantities of soluble protein. 1aC2 was insoluble and required solubilization from the bacterial pellet by the addition of 1:5 % N-laurylsarcosine, 25 mM triethanolamine, 1 mM EDTA (Grieco et al., 1992). 1aC2 remained bound to the column matrix, despite elution with 5 mM glutathione, and therefore was released by boiling in SDS sample buffer for gel analysis.

**Antisera.** Anti-FLAG M2 monoclonal antibody (mAb; Sigma) was used for immunoprecipitation of N-terminal FLAG-tagged proteins. The recombinant glutathione S-transferase fusion proteins, 1aC1 and 1aC2 (Fig. 1), were used to immunize animals for production of polyclonal antisera anti-1aC1 and anti-1aC2. In the case of 1aC2, the glutathione–Sepharose beads containing the bound proteins were used for immunization (see above).

**Radioimmunoprecipitation and SDS-PAGE.** To immunoprecipitate ORF-1a-derived radiolabelled products, immunoprecipitation buffer (50 mM Tris–HCl, 150 mM NaCl, 5 mM EDTA, 1% Triton X-100, 0.2% deoxycholate, 0.1% SDS, pH 7.2), ORF-1α or FLAG-specific antibody, and protein A immobilized on Sepharose CL-4B beads (Sigma) were added to the TnT reaction and mixed at 4 °C for 18 h. The immunoprecipitates were washed extensively with immunoprecipitation buffer, high salt immunoprecipitation buffer containing 0.65 M NaCl and TNE buffer (10 mM Tris–HCl, 100 mM NaCl, 1 mM EDTA, pH 8.0). Immunoprecipitated proteins were solubilized in equal volumes of 2 × SDS sample buffer and separated by SDS–PAGE (12% Tris–glycine or 4–12% Bis–Tris gradient where indicated). Except where noted, gels were processed for fluorography with Intensify (NEN) and exposed for autoradiography.
Alignment of amino acid sequences and prediction of protease cleavage site. Regions of ORF-1a containing the 3C-like serine protease motifs of HAstV serotypes 1–3 and 8 (accession nos L23513, Z25771, L13745, AF141381 and AF260508), SAstV (Y15937), TAstV (AF206663) and ANV (AB033998) were aligned with those of equine arterivirus (P19811), tobacco etch virus (P04517), hepatitis C virus (P26604) and murine hepatitis virus (X73559) as well as the 3C protease of human rhinovirus type 14 (HRV14) (P03303) and hepatitis A virus (HAV, P26580) using the BLAST program from SeqWeb v1.1. Wisconsin Package v10.0 (Genetics Computer Group, University of Wisconsin). The alignment was performed with a gap weight of 6, a gap extension penalty of 1 and a maximum sequence input range of 5000, using the bloom 62 scoring matrix. Amino acids predicted to be members of the 3C-like serine protease catalytic triads were assigned higher priority for alignment. Putative substrate-binding residues preceding the catalytic serine or cysteine were manually aligned.

Analysis of the HAstV-1 ORF-1a polypeptide sequence was performed using the NetPicoRNA prediction server (version 1.0, Center for Biological Sequence Analysis) to predict a cleavage site similar to picornaviral 3C proteases. Cleavage score output from the neural network is reported in the interval 0–000 and scores above 0 are considered as potential cleavage sites. Database searches for homologous proteins were performed with BLASTP and FASTA from the Wisconsin Package mentioned above or BLAST 2.0 from the National Center for Biotechnology Information, National Institutes of Health.

Results

Conservation of the predicted astrovirus 3C-like serine protease motifs

The predicted serine protease motifs of HAstV serotypes 1–3, 8, SAstV, TAstV and ANV were aligned with two well-characterized picornavirus 3C proteases, HRV14 (Matthews et al., 1994) and HAV (Allaire et al., 1994), and with 3C-like protease motifs of other positive-strand RNA viruses (EAV, HCV, TEV and MHV) (Fig. 2). The amino acids of the catalytic triad (His, Asp, Ser/Cys) were conserved for all the 3C and 3C-like protease motifs analysed in this report. For astroviruses, the predicted catalytic triad consists of His-461, Asp-489 and Ser-551. Over the 123 aa region of HAstV compared, a high degree of conservation was observed. HAstV-V-1 differed from HAstV-2 at two sites (aa 493 and aa 515), from HAstV-3 at one site (aa 493) and from HAstV-8 at four sites (aa 470, aa 477, aa 493 and aa 532). Among the amino acids implicated in substrate binding, Thr-546 and His-566 were highly conserved among the viral sequences compared. The two glycines on each side of the catalytic serine/cysteine were identical for all proteases compared in Fig. 2.

Evidence of proteolytic processing by the HAstV-1 3C-like serine protease

To identify the precursor and potential cleavage products, the ORF-1a product was expressed from pAV1a (Fig. 1A) by a coupled in vitro transcription–translation system containing nuclease-treated reticulocyte lysate in the presence of Tran^35S-

label (ICN). Immunoprecipitation of the radiolabelled translation products with C-terminal antibody anti-1aC1 (Fig. 1), followed by SDS–PAGE analysis (Fig. 3, lane 3), revealed an intense 101 kDa band (p101) that corresponds to the predicted full-length ORF-1a product, as well as bands at 48 (p48) and 38 (p38) kDa, and faint bands at 46 (p46) and 41 (p41) kDa.

Since the serine residue within the active site of 3C-like serine proteases serves as the nucleophile during catalysis (Dougherty & Semler, 1993), substitution of this amino acid would be expected to inactivate or diminish the activity of the protease. To determine potential HAstV-1 3C-like serine protease-mediated cleavage products in the ORF-1a-derived p101, we examined the effect of a mutation (pAV1a/Δprot, Fig. 1A) in which a 9 aa region containing Ser-551 and two predicted substrate-binding amino acids (Thr-546 and Gln-547) were replaced. In addition, we determined the effects when Ser-551 of the putative catalytic triad was individually mutated by site-directed mutagenesis to Ala (pAV1a/S551A, Fig. 1A). The proteins produced by expression of pAV1a/Δprot and immunoprecipitation by anti-1aC1 included p101, p48, p46 and p41 at levels that remained comparable to wild-type ORF-1a (Fig. 3, lane 2). In addition, a faint band at 37 kDa was observed [confirmed in a 4–12% gradient gel to be a band distinct from p38, also present in the same intensity in the wild-type pAV1a (see Fig. 4, lane 1)]. Of note, p38 is no longer detectable, suggesting that mutation within the 3C-like protease motif interfered with p38 production. p38 was also no longer observed when pAV1a/S551A-expressed products were immunoprecipitated with anti-1aC1 (Fig. 3, lane 1). This supports the importance of Ser-551 in the cleavage of p101 to p38 by the HAstV-1 3C-like serine protease.

To further dissect the importance of the predicted catalytic triad on the processing of p101 to p38, the remaining 2 aa of the catalytic triad were also altered individually by site-directed mutagenesis. His-461 was substituted by Leu (pAV1a/H461L, Fig. 1A) and Asp-489 by Val (pAV1a/D489V, Fig. 1A). Small noncharged amino acids were chosen as substitutes to avoid disrupting the native folding of the protein. When C-terminal antibody anti-1aC1 was used to immunoprecipitate the proteins translated from pAV1a/H461L, pAV1a/H461L and pAV1a/S551A, all three mutants resulted in undetectable levels of p38 (Fig. 4, lanes 2–4, respectively) when compared to the wild-type (Fig. 4, lane 1). In contrast, p66, p48, p46, p41 and p37 intensities were not significantly affected by the S551A, H461L or D489V mutations, and are likely derived by mechanisms not involving the viral 3C-like serine protease.

To identify potential 3C serine protease cleavage sites which would result in p38 production, the HAstV-1 ORF-1a polypeptide sequence was analysed using the NetPicoRNA program (v1.0, Center for Biological Sequence Analysis). Based on the proteolytic activity of foot-and-mouth disease virus (FMDV) 3C protease, cleavage at Q567 of the HAstV-1 ORF-1a product was predicted (cleavage score: 0.586, range
for entire ORF-1a is 0·120–0·665). Cleavage at this site would result in a C-terminal protein fragment of ~38·8 kDa which is consistent with the size of the cleavage product p38. Site-directed mutagenesis of Gln-567 to Pro (at the P1 position of the predicted cleavage site) resulted in no detectable p38 production (Fig. 4, lane 6), suggesting that the protease is unable to process the pAV1a product containing this mutation. By contrast, mutations I578M and I579M (downstream of the cleavage site), which served as controls, did not suppress p38 (Fig. 4, lane 5) production.
Fig. 3. Identification of an ORF-1a-derived 38 kDa product. Detection of p38 by immunoprecipitation using anti-1aC1. Lanes 1–3 show pAV1a/S551A, pAV1a/Aprot and pAV1a, products, respectively, immunoprecipitated with anti-1aC1. The upper arrow points to p101 and the lower arrow points to p38. Molecular size standards are shown on the left.

Fig. 4. Effect of mutagenesis of the protease motif, catalytic triad residues and the predicted cleavage site. (A) pAV1a (lane 1), pAV1a/D489V (lane 2), pAV1a/H461L (lane 3), pAV1a/S551A (lane 4), pAV1a/Q567P (lane 5), and pAV1a/I578M/I579M (lane 6) products were expressed by in vitro transcription-translation, immunoprecipitated with anti-1aC1, and separated by a 4–12% Bis-Tris gradient gel. The upper arrow points to p101 and the lower arrow indicates p38. Molecular size markers in kDa are indicated to the left.

Fig. 5. Identification of the 64 kDa N-terminal cleavage product. ORF-1a constructs pAV1a (wild-type) and pAV1a/Aprot (mutated protease motif) were modified to include an 8 aa FLAG epitope at the N terminus (pFLGAV1a and pFLGAV1a/Aprot, respectively). Immunoprecipitated products from pFLGAV1a lysates (lanes 1 and 3) and pFLGAV1a/Aprot lysates (lane 2 and 4) are shown. Antibodies used for immunoprecipitation were anti-FLAG Ab (lanes 1 and 2) and anti-1aC2 (lanes 3 and 4). The upper arrow points to the full-length p101, the middle arrow p64, and the lower arrow p38. Molecular size markers in kDa are indicated to the left.

Identification of an ORF-1a-derived N-terminal cleavage product

Kyte–Doolittle hydropathy plots of the ORF-1a product (Kyte & Doolittle, 1982) indicate that the N terminus is very hydrophobic. Attempts to produce adequate quantities of substrates for antibody production proved difficult. In the absence of specific antibodies to the N-terminal regions of ORF-1a, an 8 aa FLAG epitope was placed at the N terminus of the ORF-1a product. This epitope was chosen because its hydrophilic nature helps retain native conformation and function of the target protein. Immunoprecipitation of ORF-1a products containing the N-terminal FLAG epitope with anti-FLAG M2 mAb (Sigma) resulted in the detection of major bands at 101 and 64 kDa (p64) and minor bands at p48 and p41 (Fig. 5, lane 1). Bands corresponding to p101, p48 and p41, but not p64, were detectable when the pFLGAV1a/Aprot products were immunoprecipitated with the same antibody (Fig. 5, lane 2), indicating that generation of p64 is also dependent on the HAstV 3C-like serine protease. The faint band at ~38 kDa (Fig. 5, lane 1) may be due to some cross-reactivity between the anti-FLAG M2 mAb and an epitope in p38. (According to the manufacturer, cross-reactivity may be encountered with this mAb.) Anti-1aC2 antibodies specific for a more central
Processing of the HAstV-1 ORF-1a product

(A) (B)

Fig. 6. ORF-1a N- and C-terminal deletion analysis. (A) Immunoprecipitation of proteins expressed from N-terminal deletion constructs. ORF-1a deletion products expressed from pAV1a/N\textsuperscript{Δ}1–N\textsuperscript{Δ}5, respectively, containing deletions of 157, 282, 420, 473 or 517 aa from the N terminus were immunoprecipitated. Controls: pAV1a product immunoprecipitated with pre-immune serum (lane 1) and with anti-GST (lane 2), and pTM1 immunoprecipitated with anti-1aC1 (lane 3). Experimental: lane 5 shows pAV1a, lanes 6–10 pAV/N\textsuperscript{Δ}1–N\textsuperscript{Δ}5, respectively, and lane 11 pAV1a/Δprot; all immunoprecipitated with anti-1aC1 Abs. The upper arrow points to p101 and the lower arrow to p38. Molecular sizes of the full-length deletion products (pAV1a/N\textsuperscript{Δ}1–N\textsuperscript{Δ}5) are 84, 70, 55, 49 and 44 kDa, respectively, Lane 4, no sample. (B) ORF-1a C-terminal deletions. ORF-1a C-terminal deletions were made at the Sac\textsubscript{I} (pAV1a/C\textsuperscript{Δ}1) and Bgl\textsubscript{II} (pAV1a/C\textsuperscript{Δ}2) sites. Anti-1aC1 immunoprecipitated proteins from pAV1a, pAV1a/C\textsuperscript{Δ}1 and pAV1a/C\textsuperscript{Δ}2, pTM1 are shown in lanes 1–3 and 6 respectively. pAV1a/C\textsuperscript{Δ}1- and pAV1a/C\textsuperscript{Δ}2-expressed products were also immunoprecipitated with anti-1aC2 (lanes 4 and 5). Arrows show the 101 to 97 kDa shift for p101 and the 38 to 34 kDa shift for p38. Molecular size standards are indicated to the left.

region of the ORF-1a product than anti-1aC1 (Fig. 1) antibodies immunoprecipitated major bands at p101 and p64 and to a lesser degree p48, p41 and p38, from lysates derived from the expression of wild-type pFLGAV1a (Fig. 5, lane 3). p64 and p38 were undetectable when pFLGAV1a/Δprot products were detected with anti-1aC2 (Fig. 5, lane 4).

Amino-terminal deletion analysis

To determine the minimal sequence of the protease motif that preserves proteolytic activity (i.e. generates p38 at wild-type levels), a series of truncated polypeptides was synthesized (Fig. 1B) in the TnT system and immunoprecipitated with anti-1aC1. Full-length products of the deletion constructs, pAV1a/N\textsuperscript{Δ}1–N\textsuperscript{Δ}5 (Fig. 6A, lanes 6–10) were detected at 84, 70, 55, 49 and 44 kDa, respectively, as predicted (products of higher molecular mass than predicted full-length sizes, observed in some instances, may represent read-through products). The levels of p38 were not diminished for truncations that did not impinge upon the predicted protease motif, pAV1a/N\textsuperscript{Δ}1–N\textsuperscript{Δ}3 (Fig. 6A, lanes 6–8), when compared to the wild-type (Fig. 6A, lane 5). However, with further N-terminal truncations, pAV1a/N\textsuperscript{Δ}4–N\textsuperscript{Δ}5 (Fig. 6A, lanes 9 and 10), p38 was no longer detected. As a reference, expression of pAV1a/Δprot (where p37 is observed, but not p38) compared to wild-type are shown (Fig. 6A, lane 11).

Carboxy-terminal deletion studies

To confirm that p38 is a C-terminal fragment of ORF-1a, two C-terminal deletion mutants (Fig. 1C) were constructed and lysates were immunoprecipitated with anti-1aC1 and anti-1aC2. pAV1a/C\textsuperscript{Δ}1 encoded an ORF-1a product that was estimated to be 5 kDa less than the wild-type product. Expression of this construct yielded a protein of 97 kDa, representing the full-length truncated product, and a protein of 34 kDa (Fig. 6B, lane 2), which is 4 kDa less than p38 from the wild-type ORF-1a (Fig. 6B, lane 1), suggesting that p38 is derived from the C terminus of ORF-1a. Reduction of p48, p46 and p41 by approximately 4–5 kDa also indicates a derivation from the C terminus of ORF-1a. The same truncated products were detected using anti-1aC2 (Fig. 6B, lane 4). pAV1a/C\textsuperscript{Δ}2 encoded an ORF-1a protein with a truncation of 396 aa from the C terminus (Fig. 1C), which deletes the 1aC1 epitope resulting in no detectable bands with anti-1aC1 (Fig. 6B, lane 4).
3) and detection of only the full-length product (approximately 48 kDa) with anti-1aC2 (Fig. 6B, lane 5).

**Discussion**

Proteolytic processing of viral polyproteins plays a crucial role in the replicative cycle of many viruses (Dougherty & Semler, 1993; Palmenberg, 1990). Many positive-stranded RNA viruses require several specific protease activities for polyprotein processing, some of which, including the 3C protease of picornavirus or the related 3C-like proteases, belong to the trypsin family of proteases (Bazan & Fletterick, 1988). For several of these viruses, the 3C or 3C-like protease is responsible for the cleavage of NSPs (Jore et al., 1988; Ypma-Wong et al., 1988). Analysis of the deduced amino acid sequence of ORF-1a has previously identified a 3C-like serine protease motif in HAstV (Jiang et al., 1993). The sequence of the protease motif is well conserved among the HAstV serotypes (98% identity between serotypes 1 and 2, 99% identity between serotypes 1 and 3, and 97% identity between serotypes 1 and 8). The HAstV-1 protease motif shows 31% identity with both TAstV and ANV motifs and 39% identity with SAsV. The sequence of HAstV-1 p38, based on the predicted cleavage site, is 95% identical with HAstV-2, 93% with HAstV-3 and 95% with HAstV-8 (Fig. 2).

Although astrovirus is known to contain a well-conserved 3C-like serine protease motif, little information is available regarding the processing of astrovirus NSPs. Gibson et al. (1998) observed a lack of processing of HAstV-2 (serotype 2) ORF-1a proteins containing the 3C-like serine protease motif in vitro, while Willcocks et al. (1999) detected proteins of 75, 34, 20, 6.5 and 5.5 kDa in HAstV-1-infected Caco-2 cells. The lack of consensus between these two studies may be attributable to the differences in the systems examined and antisera used.

In the current study, we demonstrated using a cell-free system that HAstV-1 ORF-1a encodes a protease that cleaves the full-length product p101 into an N-terminal p64 fragment and a C-terminal p38 fragment. The cleavage site between Gln-567 and Thr-568, predicted by an algorithm based on the FMDV 3C protease, was confirmed.

To characterize the viral protease encoded in ORF-1a of HAstV-1, we relied primarily on mutagenesis studies, as described for other viruses (Chambers et al., 1990; Snijder et al., 1996; Ziebuhr et al., 1997). We showed that when wild-type ORF-1a was expressed, products of 101 (full-length), 48, 46, 41, 37 and 38 kDa were identified using polyclonal antisera (anti-1aC1) specific for the C terminus of p101. Of the five fragments potentially derived by cleavage by the viral protease, the 38 kDa band was the only one that appeared to be sensitive to perturbation of the protease. p38 was undetectable after: (1) mutation of the protease motif by a Δprot mutation that replaced a region of 9 aa, containing Ser-551 and two substrate-binding amino acids (Thr-546 and Gln-547), with the 9 aa HA epitope (Fig. 3); (2) mutation of individual residues, Ser-551, His-461 and Asp-489, implicated as the catalytic triad (Fig. 4). Of note, each of the single amino acid mutations of the catalytic triad residues yielded no detectable p38 band, in contrast to the poliovirus 3C protease which displayed a certain tolerance for a His to Leu change (Baum et al., 1991).

Localization of the protease-sensitive p38 product to the C terminus was demonstrated by its detection with C-terminal-specific antibodies (anti-1aC1 and -1aC2, Fig. 1) and N- and C-terminal deletion analysis (Fig. 6). The N-terminal fragment, a protease-sensitive p64 product, was detected with anti-1aC2 and the anti-FLAG mAb which detected the FLAG epitope introduced at the N terminus of p101. Marked hydrophobicity of the N terminus of p101 precluded production of antibodies to this region and necessitated introduction of a tag to enable detection of N-terminal cleavage products.

The specific site at which cleavage occurs to yield p64 and p38 was estimated to be between aa 560 and 590, based on the sizes of the cleavage products. The canonical catalytic triad and the conserved substrate-binding Thr-546 and His-566 residues found in the astrovirus protease motif are also found in other viral 3C-like proteases, suggesting that astrovirus may share similar cleavage specificity. The typical cleavage site for 3C-like proteases follows Gln or Glu residues at the P1 position. Within the roughly estimated cleavage region suggested above, cleavage after Gln-567 (P1) would result in proteins of approximately 62.2 kDa (N-terminal fragment) and 38.8 kDa (C-terminal fragment). The NetPicoRNA program that identifies FMDV 3C protease cleavage sites predicted cleavage of p101 at the dipetide Gln-567 Thr-568. Mutation of Gln-567 to Pro resulted in no detectable p38 production, indicating that integrity at this site (i.e. preservation of the conserved Gln-567 at the P1 position) is required for proteolytic processing. In contrast, production of p38 was not affected by mutations introduced at nearby sites (e.g. I568M/I579M). Of note, Gln-567 Thr-568 is conserved among all the mammalian astrovirus sequences available for comparison: HAstV serotypes 1–3 and 8 and sheep astrovirus. Avian strains, TAstV and ANV, have a possible Gln Gly cleavage site in a similar region of their ORF-1a product. Taken collectively, these data strongly support cleavage at the HAstV-1 dipetide Gln-567 Thr-568.

The N-terminal boundary of the protease was also studied. We found that proteolytic activity was preserved despite deletion of up to 420 N-terminal amino acids. Further deletions that extend into the protease motif resulted in loss of proteolytic activity as manifested by undetectable levels of p38.

In summary, these studies provide evidence that the 3C-like serine protease motif encoded in ORF-1a is functional and mediates cleavage of p101 (full-length ORF-1a product) to p64 and p38. The predicted cleavage site, between Gln-567 and Thr-568, is at the C-terminal end of the 3C-like serine protease.
motif. The N-terminal cleavage product, p64, is predicted to encompass the four transmembrane helices previously described (Jiang et al., 1993) and the 3C-like serine protease motif. The transmembrane domains have been suggested to anchor the HAstV RNA replication complex to the membrane (Jiang et al., 1993). Membrane association may also lead to post-translational modifications. The C-terminal cleavage product, p38, contains an NLS (Dingwall & Laskey, 1991) and an immunoreactive epitope identified by antibodies to purified HAstV (Matsui et al., 1993). The NLS appears to be functional (Willcocks et al., 1999), but its role in the astrovirus life-cycle is not known. More work is needed to elucidate the precise function of p38 and p64 in infected cells. Given that the regions recognized by anti-1aC1 and anti-NAR (Willcocks et al., 1999) may overlap, it is very likely that p38 and the 34 kDa product identified by Willcocks et al. in astrovirus-infected Caco-2 cells may be related. Work is currently ongoing to address the origins of the p48, p46, p41 and p37 products, which appear to be derived from the C terminus of ORF-1a, but are insensitive to 3C-like serine protease mutations. Given the increasing recognition that human astroviruses are important enteric pathogens, further studies of the nonstructural genes are essential to an understanding of virus replication, polyprotein processing and ultimately pathogenesis.

We thank Ute Geigenmüller-Gnirke for many helpful discussions and for providing the pAV1a and pAV1a/Aprot constructs. We are also grateful to Terry L. Lewis for providing the pGEX constructs and to Teri Chew for assistance with production of antibody reagents.

This work was supported by the Office of Research and Development, Department of Veterans Affairs (Merit Review to S.M.M.) and United States PHS grant DK56339 (awarded to the Stanford Digestive Center). D.K. was supported by Stanford University Medical Center Dean’s Postdoctoral Fellowship Awards (1HAG303, 1HAG376) and PHS grants DK07056 (institutional training grant) and DK09629 (awarded to D.K.).

References


Received 18 April 2001; Accepted 18 September 2001