Cell, Vol. 86, 835–843, September 6, 1996, Copyright ©1996 by Cell Press

Structure of the Human Cytomegalovirus Protease Catalytic Domain Reveals a Novel Serine Protease Fold and Catalytic Triad

Ping Chen, Hideaki Tsuge,* Robert J. Almassy, Cindy L. Gribskov, Susumu Katoh,* Darin L. Vanderpool, Stephen A. Margosiak, Christopher Pinko, David A. Matthews, and Chen-Chen Kan
Agouron Pharmaceuticals
3565 General Atomics Court
San Diego, California 92121

Summary

Proteolytic processing of capsid assembly protein precursors by herpesvirus proteases is essential for virion maturation. A 2.5 Å crystal structure of the human cytomegalovirus protease catalytic domain has been determined by X-ray diffraction. The structure defines a new class of serine protease with respect to global-fold topology and has a catalytic triad consisting of Ser-132, His-63, and His-157 in contrast with the Ser-His-Asp triads found in other serine proteases. However, catalytic machinery for activating the serine nucleophile and stabilizing a tetrahedral transition state is oriented similarly to that for members of the trypsin-like and subtilisin-like serine protease families. Formation of the active dimer is mediated primarily by burying a helix of one protomer into a deep cleft in the protein surface of the other.

Introduction

Human cytomegalovirus (HCMV) is a β herpesvirus belonging to the Herpesviridae family, which includes herpes simplex virus -1 and -2 (HSV-1, HSV-2) and Epstein Barr virus (EBV) from the α and γ herpes subgroups, respectively. Viral infection by HCMV is common, with 40%–80% of the human population infected prior to adulthood (Gold and Nankervis, 1982). For immunocompromised individuals, HCMV is a serious pathogen that can cause fatal organ damage.

All herpesviruses encode a protease that cleaves a major component of the intermediate capsid, the assembly protein precursor, during capsid maturation. This proteolytic processing is essential for virion maturation (Preston et al., 1983; Liu and Roizman, 1991). Herpesvirus proteases are, therefore, considered to be attractive molecular targets for antiviral drug development. HCMV protease is a 708 amino acid protein encoded by the UL80 open reading frame. The gene for the assembly protein precursor, although nested inside of and 3′-coterminal with the UL80 open reading frame (Welch et al., 1991), is transcribed separately. Thus, the full-length HCMV protease (UL80) and the assembly protein precursor share an identical sequence of 373 amino acids.

HCMV protease cleaves the assembly protein precursor and itself between Ala-643 and Ser-644 at the C-terminal maturation (M-) site. Additionally, HCMV protease–catalyzed autoproteolysis occurs between Ala-256 and Ser-257 at the release (R-) site. The released 28 kDa catalytic domain comprises amino acid residues 1–256 and retains full protease activity. Only β herpesviruses have additional autoproteolytic cleavage sites within the protease catalytic domain. For HCMV protease, these internal cleavage (I-) sites are between, in the first instance, Ala-143 and Ala-144; and in the second instance, Ala-209 and Ser-210 (Jones et al., 1994). In transfections (Welch et al., 1993) and in viral infections (Jones et al., 1994), proteolytic cleavage of UL80 occurs at the M-site first, followed in order by the R-site and then the I-sites. Mutant HSV-1 viruses that fail to cleave the protease at the R-site or the assembly protein precursor at the M-site do not produce infectious virus (Matusick-Kumar et al., 1995).

Herpesvirus proteases share considerable amino acid sequence homology. There are five cysteine residues in the HCMV protease catalytic domain, none of which form disulfide bonds. Only one cysteine, Cys-161 in HCMV protease, is conserved among all herpesvirus proteases. It has been reported that the enzyme loses catalytic activity if Cys-161 is oxidized to form a disulfide bond with Cys-138 (Baum et al., 1996) or modified by alkylating agents (Burck et al., 1994). However, based on modification of reactive serine residues by di-isopropylfluorophosphate (Dilanni et al., 1994), herpesvirus proteases have been proposed to be serine proteases, although they have no amino acid sequence homology with any other known serine proteases. Mutagenesis results for HCMV protease (Cox et al., 1995) suggest that His-63, Glu-122, and Ser-132 may form a catalytic triad required for proteolytic activity. Recently, we and others have independently demonstrated that the active HCMV protease catalytic domain forms an obligate dimer with a K_d of 5 μM (Margosiak et al., 1996; Darke et al., 1996).

We are interested in HCMV protease as an antiviral drug target and a member of an apparently novel class of serine protease. Using a recombinant enzyme purified from Escherichia coli as described previously (Pinko et al., 1995), we have determined the three-dimensional structure of HCMV protease catalytic domain.

Results and Discussion

Overall Structure and Novel Catalytic Triad

The catalytic domain of HCMV protease is an α/β protein consisting of a central core comprising two orthogonal four-stranded β sheets surrounded by eight α-helices (Figure 1). A sequence alignment for representative members of the herpesvirus protease family, along with secondary structure assignments based on the three-dimensional structure of HCMV protease, are given in Figure 2. Twisting of the two central β sheets brings them close together at one corner, where β3 in the top sheet bends approximately 90° at conserved Gly-70 and passes directly across into the other sheet, forming what

*Permanent address: Japan Tobacco Inc., Central Pharmaceutical Research Institute, 1-1, Murasaki-cho Takatsuki, Osaka 569, Japan.

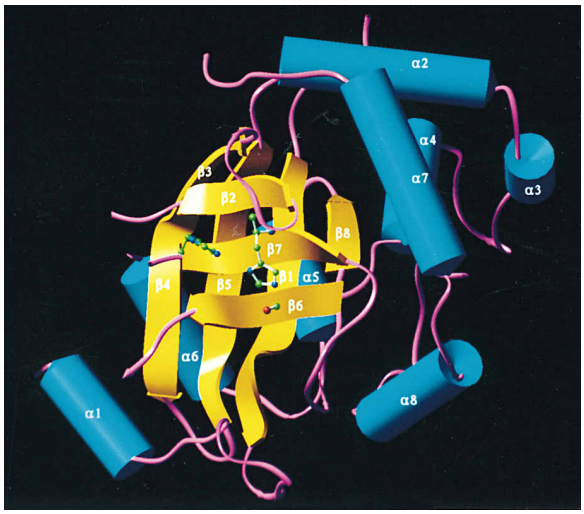

Figure 1. View of the HCMV Protease Catalytic Domain Oriented to Highlight the Core β Sheets

Strands are shown in yellow, α-helices in blue, and connecting loops in purple.

has been termed a right-handed triple-stranded corner (Efimov, 1992). Individual β strands in the lower sheet are longer and extend out from underneath those in the upper sheet, where they are covered at one end by helix α1 and the extended loop connecting β1 to α1 and at the other end by helices α2 and α4. A pair of other helices (α5 and α6) shield the back side of the large lower β sheet. The remaining helices (α3, α7, and α8) along with portions of α2 and α4, are positioned near one edge of the β sandwich where the two sheets splay apart, creating a pronounced depression in the protein surface. Dimerization of HCMV protease is mediated by burying four turns of helix α7 from one protomer into this complementary cleft on the other (Figure 3).

His-63 and Ser-132 are located on the solvent-exposed surface of the upper β sheet. The imidazole side chain of His-63 is positioned above the sheet and within hydrogen-bonding distance of the Ser-132 hydroxyl. Both residues are conserved among all known amino acid sequences of herpesvirus proteases. On the basis of amino acid sequence alignments and mutagenesis experiments, it has been proposed that His-63, Ser-132, and Glu-122 form a catalytic triad in HCMV protease (Cox et al., 1995) similar to the well-studied Ser-His-Asp catalytic triads in trypsin-like and subtilisin-like serine proteases. Crystallographic evidence supports the identification of His-63 and Ser-132 as the essential catalytic couple. Glu-122 on α4, however, is more than 20 Å from His-63 in both protomers and therefore cannot contribute directly to substrate binding or catalysis. In fact, there are no acidic residues in the vicinity of His-63 that could participate in forming a classical catalytic triad. A second conserved histidine (His-157) hydrogen bonds to the imidazole of the His-63 in each of the four crystallographically independent HCMV protease polypeptide chains, suggesting that a previously unrecognized variation of the classical Ser-His-Asp catalytic triad exists in HCMV protease comprising Ser-132, His-63, and His-157.

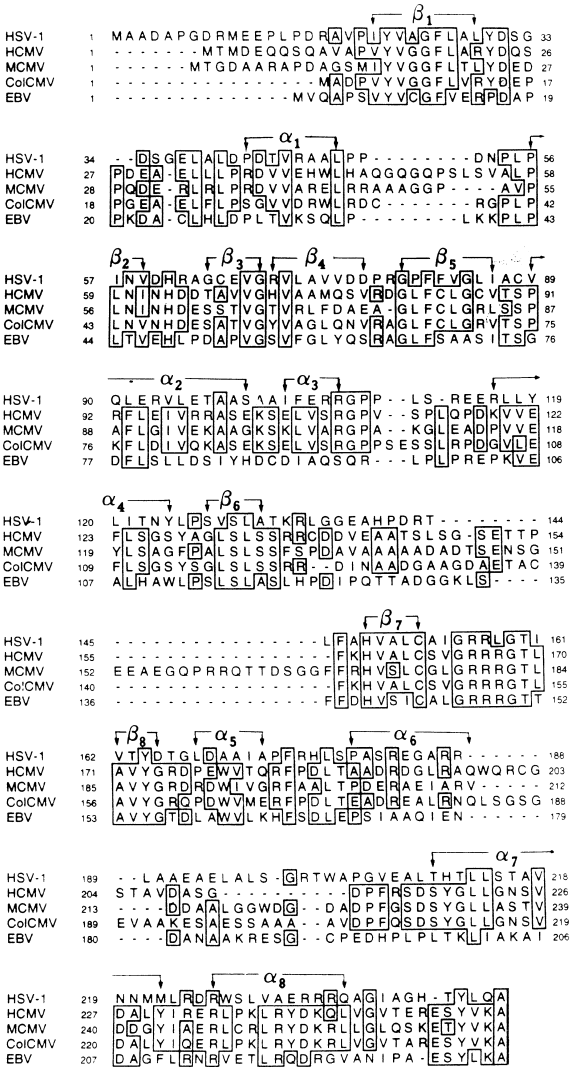

Figure 2. Secondary Structure Assignments for the HCMV Protease Catalytic Domain and Sequence Alignments for Catalytic Domains of Representative Herpesvirus Proteases

α-helices and β strands are marked as α1–α8 and β1–β8, respectively. HSV-1, herpes simplex-1; HCMV, human cytomegalovirus strain AD169; MCMV, mouse CMV; ColCMV, simian CMV strain Colburn; EBV, Epstein Barr virus.

Residues 1–7, 46–55, 139–154, and 200–211 are not modeled in the current structure because electron density for these residues is weak and poorly defined. Sequence alignments (see Figure 2) indicate that the unmodeled amino acids have little homology with residues in corresponding connecting loops of other herpesvirus proteases. A pair of the loops span known autoproteolytic cleavage sites at residues 143 and 209 (Jones et al., 1994) and are likely to be flexible and loosely structured in solution.

Transition-State Stabilization

Substrate hydrolysis by serine proteases occurs through a covalent tetrahedral intermediate resulting from attack of the active-site nucleophile on the carbonyl carbon of the scissile bond. In all known serine

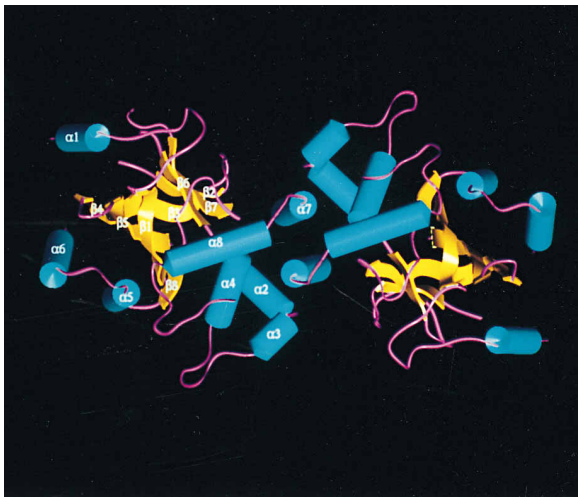

Figure 3. HCMV Protease Catalytic Domain Dimer, Viewed Down the 2-Fold Axis

α-helices are shown in blue, and the two central β sheets are shown in yellow.

proteases, esterases, and lipases and in papain-like cysteine proteases as well, distortion of the carbonyl of the substrate is stabilized by strong hydrogen bonds between the developing oxyanion and the amide groups of the enzyme. If the hydrolytic mechanism of HCMV protease proceeded through a tetrahedral intermediate, we would expect to find protein architecture for oxyanion recognition in the vicinity of Ser-132. Among all enzymes cited above, oxyanion stabilization involves either two or three hydrogen bond donors, one of which invariably is the backbone amide of the serine (or cysteine) nucleophile itself or the amino acid immediately following it. In HCMV protease, the backbone amide group of Ser-132 and those of adjacent residues in β6 are involved in antiparallel hydrogen bonding to main-chain carbonyl groups of residues in β7 (Figure 4). Therefore, none of these backbone amides is available for hydrogen bonding to the developing oxyanion. A rigid loop in HCMV protease, consisting of seven residues (163–169) connecting β7 to β8, is positioned above Ser-132 and forms the molecular scaffolding for oxyanion stabilization. The signature sequence in this loop of G-R-R-X-G-T is absolutely conserved in all 14 known amino acid sequences of herpesvirus proteases (Gibson, 1994). The backbone NH of Arg-165 points directly into a small pocket just above Ser-132 Oγ, which is occupied by an ordered water molecule in each of the four crystallographically independent polypeptide chains. This water hydrogen bonds to both the Arg-165 backbone amide and to Ser-132 Oγ. Its position probably indicates the approximate location of the oxyanion of the tetrahedral intermediate during catalysis. The guanidinium side chains of Arg-165 and Arg-166 extend outward approximately parallel to one another across the top of the oxyanion binding site. Their geometry relative to the "oxyanion hole" and to each other is maintained by charge-mediated hydrogen bonds to acidic residues in the loop connecting β1 to α1 for Arg-165 and by hydrogen bonding to the backbone carbonyl oxygen atoms of conserved Leu-32 and Leu-133 for Arg-166. The crystallographic evidence suggests that these conserved arginine residues provide a cationic environment for stabilization of the oxyanion in the transition state. It is possible that substrate binding could induce conformational changes that would allow either or both arginine side chains to hydrogen-bond directly with the oxyanion. But this would involve disrupting preexisting protein hydrogen bonds and would therefore provide less transition-state stabilization than equivalent preformed scaffolding.

Comparison with Other Serine Proteases

The HCMV protease structure defines a new class of serine protease with an overall polypeptide fold distinct from that of other hydrolytic enzymes. However, the active site is remarkably similar in certain respects to those for proteins in the trypsin-like and subtilisin-like families of serine proteases. Geometrical similarities for representative members of these families can be seen in Figure 5, where functionally equivalent atoms in HCMV protease, trypsin (Protein Data Base entry 1smf), and subtilisin (PDB entry 2sec) are superposed. A structural template for HCMV protease consisting of the His-63 imidazole, Ser-132 Oγ, His-157 Nε2, and the backbone amide of Arg-165 aligns geometrically to corresponding templates in trypsin and subtilisin with root-mean-square differences of 0.38 Å and 0.76 Å, respectively. These results indicate that, while the backbone folds are quite different, the catalytic machinery for activating the serine nucleophile, stabilizing the tetrahedral transition state, and protonating the leaving group nitrogen of the substrate is similarly disposed in all three enzymes.

An obvious difference between HCMV protease and the trypsin-like and subtilisin-like proteases is the substitution of histidine for aspartic acid as the third member of the catalytic triad. A negatively charged aspartic acid is thought to be important for stabilizing a charged imidazole in the transition state. Replacement of this aspartate with neutral amino acids in trypsin and subtilisin reduces kcat by around 10^3–10^4-fold for both enzymes (Corey and Craik, 1992; Carter and Wells, 1988). His-157 mutants of herpesvirus proteases have not been characterized kinetically, but in some cases the mutated enzymes retain activity in transfection assays with assembly protein precursor (Cox et al., 1995). This suggests that the role of His-157 may be more to position His-63 for proton extraction from Ser-132 Oγ and subsequent delivery to the leaving group nitrogen of the substrate, rather than for electronic stabilization of the transition state. The kcat for D102N mutant trypsin is 3.0/min, which is close to that for HCMV protease (18/min; Pinko et al., 1995). A potentially interesting mutation of HCMV protease would be to change His-157 to glutamic acid (an aspartic acid side chain would be too short for hydrogen bonding with Nδ1 of His-63) to test the hypothesis that an acidic amino acid at this position of the catalytic triad could elevate kcat.

The HCMV Protease Dimer and Enzyme Activity

The catalytic domain dimer of HCMV protease has a dissociation constant of 5–8 μM (Darke et al., 1996;

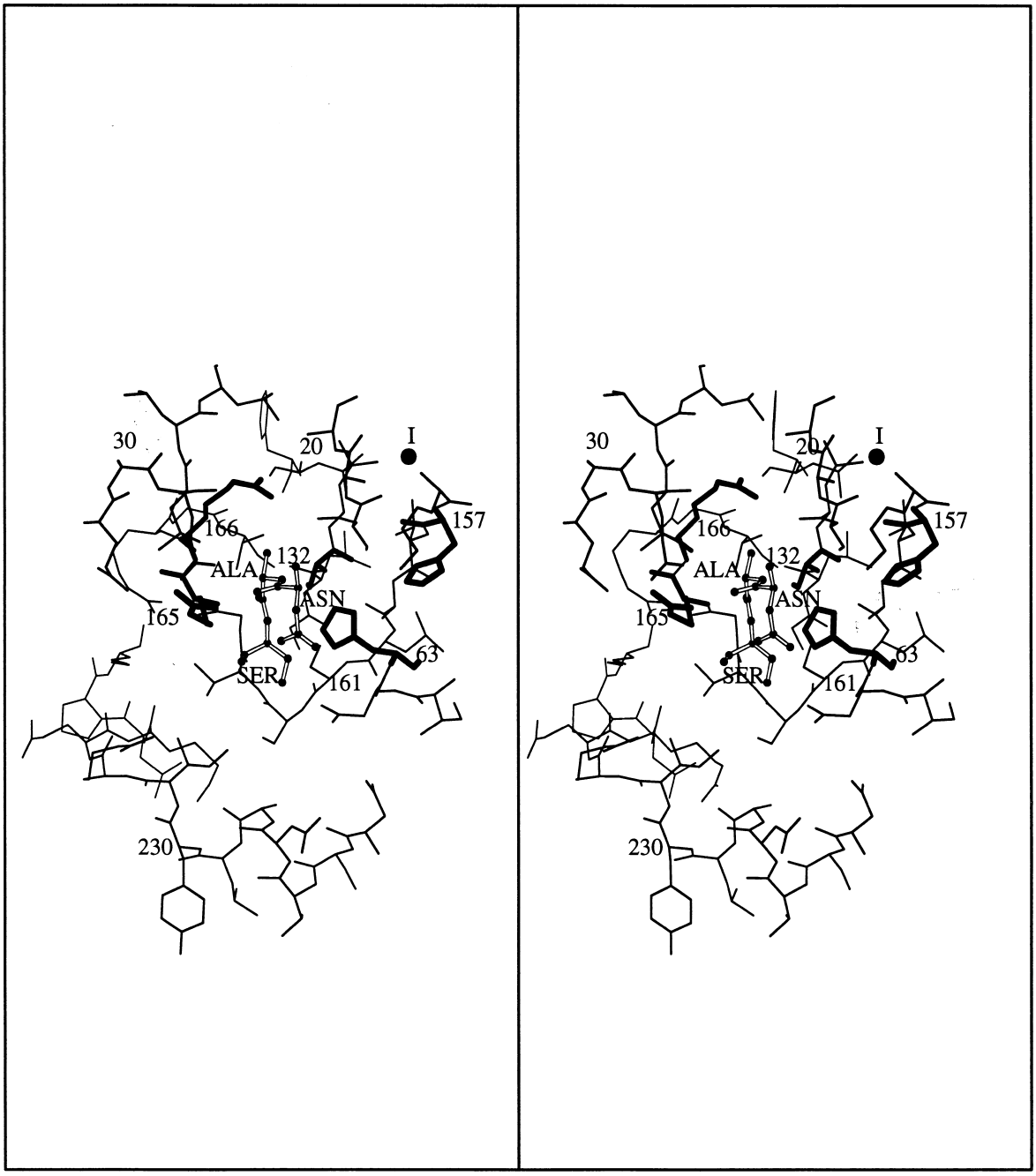

Figure 4. Stereo View of the HCMV Protease Active Site and Surrounding Residues

The catalytic triad, Arg-165, and Arg-166 are represented with thick bonds. Drawn as spheres with stick bonds is a model peptide showing how the P2 to P1' (Asn-Ala-Ser) portion of a substrate could be positioned within the active-site cleft. A larger sphere marks the iodine position in an iodotyrosine-containing tetrapeptide aldehyde inhibitor bound to the HCMV protease (see text).

Margosiak et al., 1996). The dimer has two well-separated active sites (35 Å between the side-chain hydroxyls of Ser-132 in each protomer) that may function independently. For small synthetic peptides, binding determinants probably reside within a single subunit, since the distal protomer approaches no closer than 12 Å to the scissile peptide bond of a substrate modeled into the other active site. Dimer formation is mediated by symmetrical docking of helix α7 of one subunit into a pronounced groove in the surface of the other subunit (Figure 6). Amino acids immediately N-terminal to α7 plus residues along one face of this helix directly interact with the loop that orients catalytically important His-63. Any repositioning of the "dimerization helix" α7 and its connecting loop upon dissociation into monomers could affect the placement of His-63 and inactivate the enzyme. The side-chain carboxylate of Glu-122 hydrogen-bonds to protein segments involved in dimer formation.

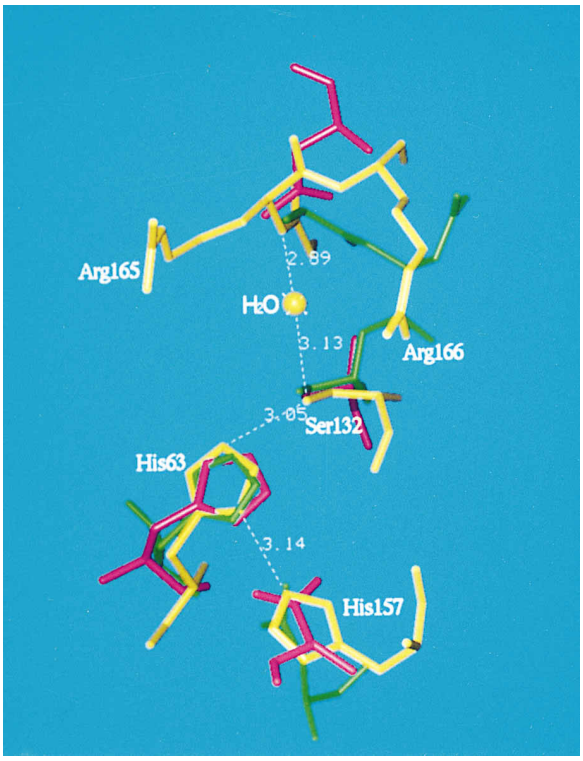

Figure 5. View of Superposed Catalytic Triads and Oxyanion Binding Sites for HCMV Protease (Yellow), Trypsin (Green), and Subtilisin (Pink)

Mutations of Glu-122 that result in loss of HCMV protease activity may produce structural changes at the dimer interface leading to the accumulation of inactive monomers.

Temporal and spatial regulation of herpesvirus protease activity in an infected cell could be regulated by dimerization. Nonoptimal matching of the HCMV protease oligomer interface is suggested by the presence of six water-filled cavities at the dimer interface and by low sequence homology for amino acids in α7 among all 14 known herpesvirus proteases. This could explain the weak association of the protease dimer, which may have a functional role in the viral life cycle. Proteolysis of the assembly protein precursor leading to its structural dissolution and entry of the viral genome occurs in the nucleus (Gibson et al., 1990) under conditions in which the local concentration of the viral protease must be considerably higher than in cytoplasm. Since HCMV protease and, most probably, other herpesvirus proteases are inactive as monomers, a weak association may keep the protease primarily monomeric in the cytoplasm, thus assuring that assembly protein precursors are not prematurely processed before being transported to the nucleus for construction of the viral capsid scaffold.

Substrate Binding

The substrate binding site of HCMV protease (see Figure 4) is an elongated shallow cleft in the surface of the protein extending in opposite directions from Ser-132, suggesting that the enzyme has recognition elements for amino acids on both sides of the scissile peptide bond of the substrate. This is consistent with results from studies of substrate specificity showing that efficient cleavage minimally requires amino acid residues spanning the P4 to P4' positions (Vinoid et al., 1994). Herpesvirus proteases cleave at several different but related sites (e.g., the M-, R-, and I-sites) having the general consensus sequence (V/I/L)XA↓(A/S) where P3 is usually valine, P1' is usually serine, and X at P2 can be asparagine, aspartate, glutamine, glutamate, or lysine (Gibson et al., 1994).

In attempting to model how peptide substrates bind, we were guided by the similar arrangement of catalytic residues in trypsin and HCMV protease and by the extensive database of X-ray structures that exist for trypsin-like serine proteases complexed with peptide inhibitors. Since the relative dispositions of the catalytic histidine imidazole, the nucleophilic serine, and the oxyanion hole

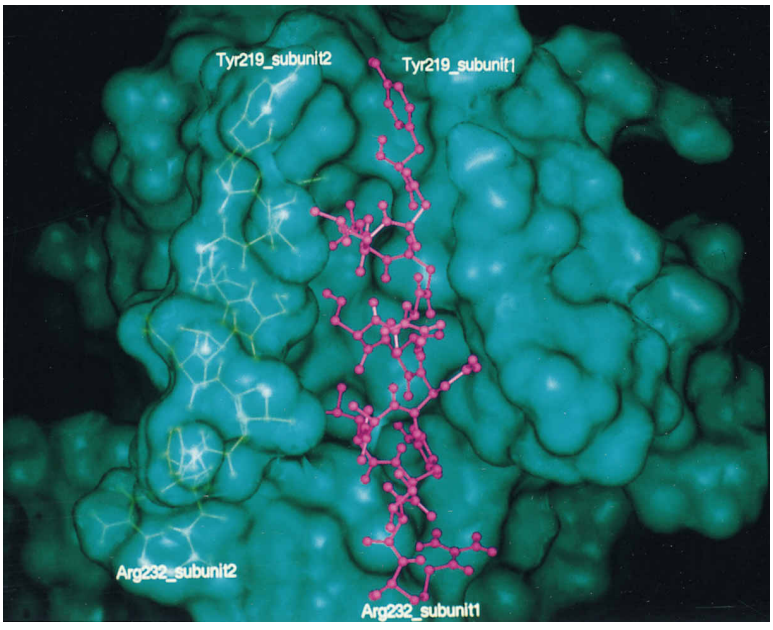

Figure 6. View of the HCMV Protease Dimer Interface Showing Helix α7 from One Protomer and the Molecular Surface of the Other Protomer

Helix α7 sits in a groove formed by five helices from the second subunit, one of which is α7′ (primed secondary structure elements refer to those in the second subunit). The surface of α7′ is directly to the left of α7.

Table 1. Statistics for the Crystallographic Analysis

Crystal	NativeI[1]	NativeII[2]	HgI[1]	HgII[2]
Soak concentration[3]	—	—	1/10K	1/1K
Soak time (hr)	—	—	90	74
Internal merging and scaling				
Resolution (Å)	3.5	2.5	3.5	2.9
Reflections measured	52932	117752	52551	115741
Unique reflection	12384	32525	12578	23431
Completeness (%)	86	91	91	98
Average I/σ	17.3	21.0	14.0	27.4
R_{sym}[4]	6.7	3.5	7.7	5.4
SIRSAS analysis				
Resolution (Å)			15−3.5	
Rcullis[5]			0.54	
Phasing power[6] (SIR/SAS)			3.1/1.94	
Figure of merit			0.676	
SAS analysis				
Resolution (Å)				15−3.5
Phasing power				1.90
Figure of merit				0.388

[1] Data collected on dual Xuong-Hamlin multiwire area detectors at 4°C.

[2] Data collected on MAR imaging plate at −170°C.

[3] Hygroscopic crystalline salt of $Na_2Hg(SCN)_4$ had absorbed an indeterminate amount of water. For heavy atom soaking, 1 μl of the dissolved mercury containing salt (unknown concentration) was added to 10 ml (HgI) or 1 ml (HgII) of the protein mother liquor.

[4] $R_{sym} = \sum_{h}\sum_{i=1}^{N} | I(h) - I(h)_i | / \sum_{h}\sum_{i=1}^{N} I(h)_i * 100$

where $I(h)_i$ is the ith measurement of reflection h and I(h) is the mean value of the N equivalent reflections.

[5] Rcullis = $\Sigma||F_{PH} +/- F_P| - F_{H(calc)}|/\Sigma|F_{PH} +/- F_P|$ for all centric reflections.

[6] Phasing power = root-mean-square ($|F_H|/E$), where $|F_H|$ is the heavy-atom structure factor amplitude and E is the residual lack of closure.

are the same in HCMV protease and trypsin, the handedness of the tetrahedral intermediates (Garavito et al., 1977) in the two cases must also be the same. This requires that peptide substrate residues on the N-terminal side of the scissile bond be bound in the upper portion of the active-site cleft, as viewed in Figure 4. We have recently determined a low resolution structure for HCMV protease in a second crystal form both as the apoenzyme and as a complex with the iodinated tetrapeptide aldehyde inhibitor (iodo-Y-V-N-A-aldehyde, Ki = 3 μM). Difference maps show a large peak at approximately the position expected for an iodine-containing tetrapeptide bound to the protease in an extended conformation.

In order to study possible substrate binding modes in more detail, we applied the same transformation that maps the catalytic template of trypsin onto that of HCMV protease to a modified version of the trypsin/Bowman Birk inhibitor complex X-ray structure (PDB entry 1smf), in which inhibitor side chains were changed to match a consensus sequence for the HCMV protease cleavage site. This comparison suggests that the P1 alanine of the substrate fits into a shallow buried pocket formed by conserved Leu-32 and aliphatic portions of the Arg-166 guanidinium side chain. On the leaving group side of the scissile bond, the P1' residue must be oriented with its side chain in a small pocket formed by Asn-62 and His-63. A large amino acid side chain can not be accommodated here, while a possible hydrogen bond between Ser at P1' and the side chain of Asn-62 is in agreement with the observed preference at P1' of Ser> Ala>Gly>>Thr (Vinoid et al., 1994). Residues N-terminal to the P1 alanine must align themselves along the extended strand β7 that forms the bottom of the active-site groove. In certain respects, this geometry is similar to that for substrate binding to trypsin-like serine proteases, where a similarly exposed β strand makes antiparallel β sheet–type hydrogen bonds with substrate in an extended conformation. For herpesvirus proteases, the P2 residue is variable, which is consistent with our positioning of the P2 side chain adjacent to His-63 but pointing into bulk solvent. The backbone carbonyl oxygen at P2 is positioned to hydrogen-bond to the guanidinium side chain of Arg-165. Beyond this point, the modeling becomes more problematic because without some adjustment of the protein, particularly in the loop connecting β1 to α1, which forms one edge of the binding cleft, it is difficult to satisfy all the substrate main-chain hydrogen bonding requirements. A detailed description of peptide binding to HCMV protease must await a high resolution cocrystal structure determination.

Implications

An X-ray structure of the HCMV protease catalytic domain reveals the protein to be a member of a new class of serine protease with respect to global fold, topology, and quaternary structure. The structure confirms that His-63 and Ser-132 are essential residues that together with a third amino acid, His-157, form a catalytic triad with similarities to the Ser-His-Asp triads in trypsin-like and subtilisin-like serine proteases. Recognition elements for oxyanion transition-state stabilization reside in a short highly conserved hexapeptide loop that provides both hydrogen bonding and charge-neutralization functionality. Although HCMV protease, trypsin, and subtilisin have unrelated amino acid sequences and structural scaffolds, geometrical arrangement of their catalytic triads and architecture for stabilizing the oxyanion transition state for peptide cleavage is nearly identical in the three enzymes, providing one of the few clear-

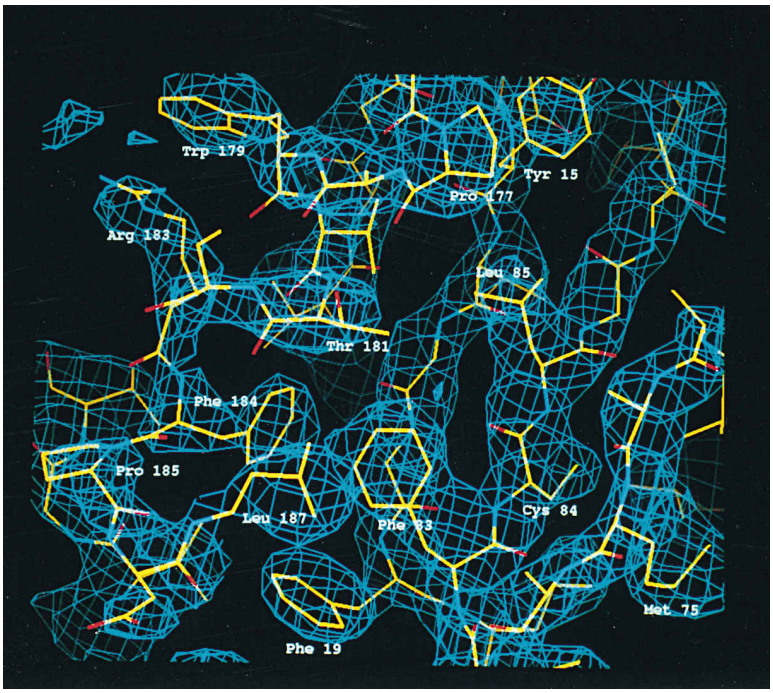

Figure 7. A Portion of the Averaged Solvent-Flattened ISAS Electron Density Map Calculated at 2.9 Å Resolution and Contoured at 2.5 σ

Superposed on the density is the final refined model.

cut examples of convergent evolution at the level of enzyme catalysis. An intriguing aspect of the structure is a four-turn α-helix that docks into a crevice from a second subunit to form active homodimeric HCMV protease.

The importance of HCMV protease for viral replication in infected cells is well documented, thus establishing the protein as a validated target for antiviral drug design. Our structure suggests several possible ways in which this goal could be accomplished. Because of the shallow specificity pockets at S1 and S1', it is likely that inhibitors will require covalent association with the catalytic nucleophile Ser-132 or nearby conserved Cys-161 (Figure 4). Others have shown that alkylation of Cys-161 or disulfide bond formation between Cys-161 and Cys-138 inactivates HCMV protease (Burck et al., 1994; Baum et al., 1996). The structure suggests that this is accomplished by blocking substrate access to the active site. Inhibitors with specificity for HCMV protease might take advantage of binding determinants in a pronounced pocket beyond S1' on the leaving group side of the active site. Alternatively, small molecules that interact with the α7 binding crevice could inhibit HCMV protease by shifting the equilibrium to favor inactive monomers. A strategy for disrupting dimer formation is attractive in this case because the oligomer interface is characterized by a deep helix binding pocket, but subunit association, at least for the catalytic domain of HCMV protease, is weak.

Experimental Procedures

Protein Expression and Purification

The catalytic domain of the HCMV protease was expressed in E. coli. Recombinant HCMV proteases were purified and proteolytic activity was determined as described previously (Pinko et al., 1995). Mutation of the A143Q eliminated autoproteolysis, allowing stable 28 kDa protease to be overproduced. A set of four additional mutations (C138L, D139G, R196S, and C202A) were introduced into the HCMV protease studied here for the purpose of optimizing crystallization properties. All five mutated residues are on the surface loops. Compared to A143Q mutant HCMV protease, the five-site mutant protease possesses 30% higher proteolytic activity.

Protein Crystallization and Data Collection

Crystals were grown at 13°C using a hanging drop vapor diffusion method. Aliquots (5 μl) of protein solution at a concentration of 10 mg/ml were mixed with equal volumes of a reservoir solution and sealed over individual wells filled with 1 ml of 0.2M MES (pH 6.0), 400 mM NaCl, 12% PEG 3.4K, 5 mM EDTA, and 20% DMSO. The crystals were shown to belong to space group $C222_1$, with cell parameters a = 105.74 Å, b = 113.38 Å, and c = 185.26 Å. Cell parameters for crystals frozen at −170°C were a = 104.48 Å, b = 111.89 Å, and c = 181.23 Å. There are four monomers in the asymmetric unit corresponding to a solvent content of 48%. Crystals diffracted well to 2.3 Å resolution. X-ray diffraction data were collected either at 4°C, using dual Xuong-Hamlin multiwire area detectors, or at −170°C, using a MAR imaging plate and processed by DENZO (Otwinowski, 1993). Data collection with the Xuong-Hamlin detectors was limited to 3.5 Å resolution because of the long c axis.

Heavy Atom Phasing

Initial heavy atom screening was carried out at 3.5 Å resolution using crystals cooled to 4°C. Diffraction data for a crystal soaked in $Na_2Hg(SCN)_4$ were used to calculate isomorphous and anomalous difference Patterson maps from which three consistent sites of heavy atom substitution were derived. Difference Fourier methods were used to locate five additional heavy atom sites. The positional and thermal parameters and relative occupancies for the eight heavy atom sites were refined using programs PHASES (Furey and Swaminathan, 1990) and HEAVY (Terwilliger and Eisenberg, 1983) at 3.5 Å. The handedness of the Hg positions was determined using the approach of Wang (1985). SIR and anomalous data at 3.5 Å were combined. Cycles (4) of solvent flattening were then carried out using phases calculated from the set of eight heavy atom positions in both the original and inverted configurations. The inverted Hg positions gave the correct hand and were used in the subsequent refinement. ISIRSAS maps revealed electron density for four monomers in the asymmetric unit. Helices (5) in each monomer were identified and modeled into ISIRSAS maps. Initial noncrystallographic (NC) symmetry operators relating the four molecules in

asymmetric unit were derived using the five helices in each molecule. NC symmetry operators were then refined by least squares using ISIRSAS maps at 3.5 Å. An iterative process of solvent-flattening, symmetry averaging, back-transformation, and phase combination was carried out. The resultant electron density maps were improved greatly, and an initial backbone model was built. However, attempts to fit individual amino acid side chains into these maps were unsuccessful. But this initial model was placed into a map generated from −170°C data, as described below.

Anomalous Patterson maps calculated using data from a $Na_2Hg(SCN)_4$-soaked crystal at −170° C showed good peaks; however, the corresponding isomorphous Patterson maps were of much poorer quality, perhaps because of nonisomorphism caused by freezing. Hg positions (8) were refined using only single anomalous scattering data at 4 Å. Several cycles of solvent flattening were carried out, and the phases were extended for the measured diffraction data from 4.0 Å to 3.5 Å. ISAS phases were then used to refine the Hg positions against 3.5 Å single anomalous scattering data. The partial model generated from 4°C data was then placed into the 3.5 Å ISAS map. NC operators as well as masks for each of the four monomers were generated from these partial models. NC symmetry operators were refined by least squares using ISAS maps at 4.0 Å. Again, an iterative process of solvent-flattening, symmetry averaging, back-transformation, and phase combination was carried out. Phases were extended for the measured diffraction data from 3.5 Å to 2.9 Å. Except for several loop regions, the resultant electron density maps showed good backbone density and very well-defined side chains (Figure 7). All phase calculations were performed using PHASES.

Model Building and Refinement

An initial model was built using program FRODO (Jones, 1978) and refined against data set NativeII using X-PLOR (Brünger, 1992). NC symmetry restraints were applied during the course of the refinement. The current crystallographic R factor is 22.45%, with R_{free} = 31.1% at 2.5 Å (30728 reflections with $F>2\sigma(F)$). The root-mean-square deviations from ideal bond lengths and angles are 0.02 Å and 3.5°, respectively. The average temperature factors are 24.85 $Å^2$ and 27.29 $Å^2$ for the protein-backbone and side-chain atoms, respectively. PROCHECK (Laskowski et al., 1993) revealed only one unfavorable (ϕ, φ) combination, with good main-chain and side-chain geometry compared with other well-refined protein structures at 2.5 Å resolution. All four crystallographically independent polypeptide chains include all atoms for residues 8–45, 56–135 (molecule 1 shows additional 136–138 residues), 155–199, and 212–256, as well as 135 ordered solvent molecules. Residues in three flexible surface loops, including two associated with internal cleavage sites at amino acids 143 and 209, have not been modeled into the current structure.

Acknowledgments

We thank the following individuals: S. Rahmati for bacterial cultures; W. Sisson and J. Barker for assisting in protein crystallization; Drs. M. McTigue, D. Knighton, M. Miyano, H. Ago, and E. Inagaki for help with data collection; Dr. S. Worland for valuable discussions; and Dr. Jay Davies II for critical reading of the manuscript. We also thank the staff of the synchrotron facility at the Photon Factory, Tsukuba, Japan.

Received June 28, 1996; revised July 30, 1996.

References

Baum, E.Z., Siegel, M.M., Bebernitz, G.A., Hulmes, J.D., Sridharan, L., Sun, L., Tabei, K., Johnston, S.H., Wildey, M.J., Nygaard, J., Jones, T.R., and Gluzman, Y. (1996). Inhibition of human cytomegalovirus UL80 protease by specific intramolecular disulfide bond formation. Biochemistry *35*, 5838–5846.

Brünger, A. (1992). X-PLOR version 3.1: a system for X-ray crystallography and NMR. (New Haven, Connecticut: Yale University Press.)

Burck, P.J., Berg, D.H., Luk, T.P., Sassmannshausen, L.M., Wakulchik, M., Smith, D.P., Hsiung, H.M., Becker, G.W., Gibson, W., and Villarreal, E.C. (1994). Human cytomegalovirus maturational proteinase: expression in Escherichia coli, purification, and enzymatic characterization by using peptide substrate mimics of natural cleavage sites. J. Virol. *68*, 2937–2946.

Carter, P., and Wells, J.A. (1988). Dissecting the catalytic triad of a serine protease. Nature *332*, 564–568.

Corey, D.R., and Craik, C.S. (1992). An investigation into the minimum requirements for peptide hydrolysis by mutation of the catalytic triad of trypsin. J. Am. Chem. Soc. *114*, 1784–1790.

Cox, G.A., Wakulchik, M., Sassmannshausen, L.M., Gibson, W., and Villarreal, E.C. (1995). Human cytomegalovirus proteinase: candidate glutamic acid identified as third member of putative active-site triad. J. Virol. *69*, 4524–4528.

Darke, P.L., Cole, J.L., Waxman, L., Hall, D.L., Sardana, M.K., and Kuo, L.C. (1996). Active human cytomegalovirus protease is a dimer. J. Biol. Chem. *271*, 7445–7449.

DiIanni, C.L., Stevens, J.T., Bolgar, M., O'Boyle, D.R., II, Weinheimer, S.P., and Colonno, R.J. (1994). Identification of the serine residue at the active site of the herpes simplex virus type-1 protease. J. Biol. Chem. *269*, 12672–12676.

Efimov, A.V. (1992). A novel super-secondary structure of β proteins: a triple-strand corner. FEBS Lett. *298*, 261–265.

Furey, W., and Swaminathan, S. (1990). "PHASES": a program package for the processing and analysis of diffraction data from macromolecules. PA33, Amer. Cryst. Assoc. Mtg. Summ. (series 2), *18*, 73.

Garavito, M.R., and Rossmann, M.G. (1977). Convergence of active center geometries. Biochemistry *16*, 5065–5071.

Gibson, W., Marcy, A.I., Comolli, J.C., and Lee, J. (1990). Identification of precursor to cytomegalovirus capsid assembly protein and evidence that processing results in loss of its carboxy-terminal end. J. Virol. *64*, 1241–1249.

Gibson, W., Welch, A.R., and Hall, M.R.T. (1994). Assemblin, a herpes virus serine maturational proteinase and new molecular target for antiviral. Perspect. Drug Discovery Design *2*, 413–426.

Gold, E., and Nankervis, G.A. (1982). Cytomegalovirus. In Viral Infections of Humans: Epidemiology and Control, Second Edition, A.S. Evans, ed. (New York: Plenum Publishing Corp.), pp. 167–186.

Jones, T.A. (1978). A graphics model building and refinement system for macromolecules. J. Appl. Cryst. *11*, 268–272.

Jones, T.R., Sun, L., Bebernitz, G.A., Muzithras, V.P., Kim, H.-J., Johnston, S.H., and Baum, E.Z. (1994). Proteolytic activity of human cytomegalovirus UL80 protease cleavage-site mutants. J. Virol. *68*, 3742–3752.

Laskowski, R.J., Macarthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. *26*, 283–290.

Liu, F., and Roizman, B. (1991). The herpes simplex virus-1 gene encoding a protease also contains within its coding domain the gene encoding the more abundant substrate. J. Virol. *65*, 5149–5156.

Margosiak, S.A., Vanderpool, D.L., Sisson, W., Pinko, C., and Kan, C.-C. (1996). Dimerization of the human cytomegalovirus protease: kinetic and biochemical characterization of the catalytic homodimer. Biochemistry *35*, 5300–5307.

Matusick-Kumar, L., McCann, P.J., III, Robertson, B.J., Newcomb, W.W., Brown, J.C., and Gao, M. (1995). Release of the catalytic domain No from the herpes simplex virus type-1 protease is required for viral growth. J. Virol. *69*, 7113–7121.

Otwinowski, Z. (1993). DENZO. In Data Collection and Processing of Proceedings of the CCP4 Study Weekend, L. Sawyer, N. Isaacs, and S. Bailey, eds. (Warrington, UK: SERC Daresbury Laboratory), pp. 56–62.

Pinko, C., Margosiak, S.A., Vanderpool, D., Gutowski, J.C., Condon, B., and Kan, C.-C. (1995). Single-chain recombinant human cytomegalovirus protease. J. Biol. Chem. *270*, 23634–23640.

Preston, V.G., Jonathan, A.V.C., and Rixon, F.J. (1983). Identification

and characterization of a herpes simplex virus gene product required for encapsidation of virus DNA. J. Virol. *45*, 1056–1064.

Terwilliger, T.C., and Eisenberg, D. (1983). Unbiased three-dimensional refinement of heavy atom parameters by correlation of origin-removed Patterson functions. Acta Cryst. *A39*, 813–817.

Vinod, V.S., Wolfgang, C.A., Veloski, C.A., Long, W.J., LeGrow, K., Wolanski, B., Emini, E.A., and LaFemina, R.L. (1994). Peptide substrate cleavage specificity of the human cytomegalovirus protease. J. Biol. Chem. *269*, 14337–14340.

Wang, B.-C. (1985). Resolution of phase ambiguity in macromolecular crystallography. Meth. Enzymol. *115*, 90–112.

Welch, A.R., McNally, L.M., and Gibson, W. (1991). Cytomegalovirus assembly protein nested gene family: four 3′-coterminal transcripts encode four in-frame, overlapping proteins. J. Virol. *65*, 4091–4100.

Welch, A.R., Hall, M.R.T., McNally, L.M., and Gibson, W. (1993). Herpesvirus proteinase: site-directed mutagenesis used to study maturational, release, and inactivation cleavage sites of precursor and to identify a possible catalytic site serine and histidine. J. Virol. *67*, 7360–7372.