Active Site Cavity of Herpesvirus Proteases Revealed by the Crystal Structure of Herpes Simplex Virus Protease/Inhibitor Complex‡

Susan S. Hoog,§ Ward W. Smith,§ Xiayang Qiu,§ Cheryl A. Janson,|| Brian Hellmig,|| Michael S. McQueney,|| Kevin O'Donnell,§ Daniel O'Shannessy,|| Anthony G. DiLella,⊥ Christine Debouck,⊥ and Sherin S. Abdel-Meguid*,§

Departments of Macromolecular Sciences, Protein Biochemistry, and Molecular Genetics, SmithKline Beecham Pharmaceuticals, King of Prussia, Pennsylvania 19406

Received May 28, 1997; Revised Manuscript Received August 25, 1997⊗

ABSTRACT: Human herpes simplex virus type 1 (HSV-1) and type 2 (HSV-2) are responsible for herpes labialis (cold sores) and genital herpes, respectively. They encode a serine protease that is required for viral replication, and represent a viable target for therapeutic intervention. Here, we report the crystal structures of HSV-1 and HSV-2 proteases, the latter in the presence and absence of the covalently bound transition state analog inhibitor diisopropyl phosphate (DIP). The HSV-1 and HSV-2 protease structures show a fold that is neither like chymotrypsin nor like subtilisin, and has been seen only in the recently determined cytomegalovirus (CMV) and varicella-zoster virus (VZV) protease structures. HSV-1 and HSV-2 proteases share high sequence homology and have almost identical three-dimensional structures. However, structural differences are observed with the less homologous CMV protease, offering a structural basis for herpes virus protease ligand specificity. The bound inhibitor identifies the oxyanion hole of these enzymes and defines the active site cavity.

Members of the human herpesvirus family are responsible for a variety of diseases from subclinical infections to fatal diseases in the immunocompromised and immunosuppressed. They are divided into three subfamilies: the α subfamily includes herpes simplex virus type 1 (HSV-1)[1] and type 2 (HSV-2) and varicella-zoster virus (VZV); the β includes cytomegalovirus (CMV), human herpesvirus-6 (HHV-6), and HHV-7; and the γ includes Epstein−Barr virus (EBV) and HHV-8. Herpesviruses vary greatly in their biological properties. Although all herpesviruses remain latent in a specific set of cells, the exact cell in which they remain latent varies from one virus to another, and they all differ with respect to the clinical manifestations of diseases they cause. For example, CMV infection can result in life-threatening acute disease states in congenitally infected infants, immunosuppressive transplant recipients, and AIDS patients (*1*). VZV is a neurotropic virus responsible for chickenpox, shingles, and postherpetic neuralgia. HHV-6 and HHV-7 have been associated with the childhood disease roseola, and recently HHV-8 has been linked to the development of Kaposi's sarcoma (*2*). The herpes simplex viruses were the first of the human herpesviruses to be discovered and are among the most intensively investigated of all viruses (*3*). Particularly, genital herpes simplex virus infection (HSV-2) is of increasing public health importance. The recurrent nature of the infection, its differing clinical manifestations, and complications such as aseptic meningitis and neonatal infection, are of great concern to patients and health care providers (*4*).

An essential step in herpesvirus capsid assembly is the proteolytic processing of the assemblin protein designated ICP35 in HSV-1 (*5*). This processing is catalyzed by a virally encoded serine protease that contains the assemblin protein at its C terminus. The protease then catalyzes its own cleavage to produce a 247 residue amino-terminal domain that has full catalytic activity (*6*, *7*). The amino acid sequence of the protease catalytic domain is relatively conserved within each subfamily, but less so between the different subfamilies. The amino acid sequence of the HSV-1 protease catalytic domain is 91% and 50% identical to that of HSV-2 and VZV, respectively, while it is only 26% identical to that of CMV protease. There is little sequence homology to other known proteins, including the absence of the conserved G-X-S/C-G-G sequence for chymotrypsin-like and G-T-S-M/A for subtilisin-like proteases (*8*, *9*). Recently the crystal structures of unliganded CMV and VZV proteases have been solved (*10*−*14*). These structures reveal a new fold for the herpes proteases that has not been reported for any other serine protease, and an active site consisting of a novel catalytic triad in which the third member of the triad is a histidine instead of an aspartic acid.

Here we report the crystal structures of the herpes simplex virus proteases in the presence and absence of the covalent inhibitor diisopropyl phosphate (DIP). The inhibitor helps to better delineate the enzyme active site cavity including the oxyanion hole. Comparison with other herpes protease structures allows a better understanding of ligand specificity and binding.

‡ The coordinates have been deposited in the Brookhaven Protein Databank, Accession Number 1AT3.

* Correspondence should be addressed to this author. E-mail: sherin_s_abdel-meguid@sbphrd.com.

§ Department of Macromolecular Sciences.

|| Department of Protein Biochemistry.

⊥ Department of Molecular Genetics.

⊗ Abstract published in *Advance ACS Abstracts,* October 15, 1997.

[1] Abbreviations: HSV, herpes simplex virus; VZV, varicella-zoster virus; CMV, cytomegalovirus; HHV, human herpesvirus; EBV, Epstein−Barr virus; DIP, diisopropyl phosphate; DFP, diisopropyl fluorophosphate; Ni $^{2+}$NTA, nickel nitrilotriacetic acid−agarose; Tris, tris(hydroxymethyl)aminomethane; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; MALDI, matrix-assisted laser desorption and ionization; PEG, polyethylene glycol.

S0006-2960(97)01269-5 CCC: $14.00

Table 1: HSV-2/DIP Structure Determination Statistics[a]

data set	resolution (Å)	observed	unique	complete (%)	R_{merge}	R_{iso}	no. of sites	phasing power	R_{Cullis}
native	2.5	78502	15788	90.8	0.095	–	–	–	–
$KAu(CN)_2$	3.0	31511	10071	98.2	0.078	15.9	2	1.14	0.74
$LuCl_3$	3.0	31104	9964	97.3	0.078	12.4	1	1.51	0.79
$PrCl_3$	3.0	31645	10179	99.0	0.090	13.6	1	1.4	0.82
$Yb_2(SO_4)_2$	4.0	10495	3787	78.7	0.088	14.6	1	2.01	0.72
$GdCl_3$	4.0	18308	4268	96.6	0.166	15.7	1	1.56	0.79
$SmCl_3$	4.0	25172	4327	97.6	0.114	8.8	1	1.47	0.86

[a] $R_{merge} = \sum|I - \langle I\rangle|/\sum\langle I\rangle$, where I is the observed intensity and $\langle I\rangle$ is the average intensity of multiple observations. $R_{iso} = \sum|F_{PH} - F_P|/\sum F_P$. Phasing power = rms isomorphous difference/rms residual lack of closure. $R_{Cullis} = \sum|FH_o - FH_c|/\sum|FH_o|$, where FH_o and FH_c are the observed and calculated heavy atom structure factor amplitudes for centric reflections. R-factor $= \sum||F_o| - |F_c||/\sum F_o$. MIR overall mean figure of merit (15−3.0 Å): 0.62. Overall figure of merit after phase combination: 0.67. Mean figure of merit following density modification (100−3.0 Å): 0.838. X-PLOR refinement, 10−2.5 Å: no. of reflections used ($F > 2\sigma$), 14 605; R-factor, 20.5%; no. of protein atoms (non-H), 3364; free R-factor, 29.0%; mean coordinate error (SigmaA), 0.4 Å; average B-factor, 21.1 Å^2; rms deviation of Cα atoms between monomers, 0.164 Å; no. of solvent atoms, 43; rms bond length, 0.015 Å; rms dihedrals, 24.7°; rms bond angles, 1.9°; rms impropers, 1.5°.

EXPERIMENTAL PROCEDURES

General. Soluble HSV-1 and HSV-2 proteases were expressed in *E. coli* from a gene encoding the N-terminal 260 amino acids of the U_L26 gene, thereby generating a precursor form of the catalytic domain which contains an additional 13 amino acids beyond the release site (R-site). This was followed by a C-terminal hexa-histidine tag to facilitate purification by Ni^{2+}NTA affinity column chromatography. The protease was eluted from the Ni^{2+}NTA-agarose with buffer (50 mM Tris, 300 mM NaCl, pH 8.0) containing 250 mM imidazole. Autoprocessing of the hexa-histidine-tagged precursor was achieved by overnight incubation at 4 °C. The processed tag was removed using a Pharmacia Superdex 75 column. The autoprocessing was confirmed by protein SDS−PAGE and MALDI-mass spectrometry. The enzyme was stored in 50% glycerol at −20 °C.

HSV-2/DIP Complex. The irreversible inhibitor diisopropyl fluorophosphate (DFP) was added to the protease and incubated until > 98% was modified. Excess inhibitor was removed by Sephadex G-25 chromatography, and the material was concentrated for crystallization experiments. The highly reactive (P−F) linkage undergoes displacement by the enzyme's active site serine hydroxyl group. Once bound to the nucleophile, the ligand is referred to as DIP. Crystals of the HSV-2/DIP protease complex were grown using the hanging-drop method of vapor diffusion from a well containing 10% PEG 4000 (50% w/v) at pH 5.0, and a drop of 10 mg/mL enzyme mixed 1:1 with the well solution. The space group was determined to be $P2_12_12$ with cell dimensions a = 71.72 Å, b = 87.43 Å, c = 77.28 Å. The asymmetric unit of the crystal contains one protease dimer with a solvent content of ∼45%. The crystal used for data measurement was approximately 0.7 × 0.3 × 0.2 mm.

Data were collected on a Siemens area detector to 2.5 Å and processed using XDS (*15*). The structure was solved using a combination of multiple isomorphous replacement and molecular replacement methods. Derivatives were prepared by soaking the native crystals in a solution containing 0.1 mM $KAu(CN)_2$, 0.4 mM $LuCl_3$, 0.6 mM $PrCl_3$, 0.2 mM $Yb_2(SO_4)_2$, 0.5 mM $GdCl_3$, or 0.2 mM $SmCl_3$ for 1−2 days. Heavy atom positions were identified by difference Patterson and difference Fourier methods using the programs in the XtalView software package (*16*). Heavy atom refinement and determination of an initial set of phases were carried out using the programs in the CCP4 suite (*17*). Heavy atom statistics are summarized in Table 1.

A molecular replacement solution was also identified by X-PLOR (*18*) using the structure of VZV protease (*14*) as a model. Each monomer in the search model was derived from a total of 177 amino acid residues comprising the core (residues 11−22, 46−91, 95−124, 137−183, and 189−230) of the VZV protease structure, with the side chains truncated to alanine. The rotation function calculation was carried out with data between 15 and 4.0 Å resolution with a maximum search vector length of 38 Å. The top peak in the rotation function from this dimer model was 4.9 σ. The translation function was calculated with data between 8 and 4 Å resolution. The top solution was at 9.3 σ. Rigid body refinement of the two monomers reduced the R-factor to 0.48 for all data to 3.5 Å. Using difference Fourier methods, phases derived from the molecular replacement solution were consistent with those generated from multiple isomorphous replacement. The combination of these two sets of phases resulted in an overall figure of merit of 0.67 calculated using SIGMAA (*17*). This was followed by 1 round of 150 cycles of noncrystallographic symmetry averaging, solvent flattening, and histogram matching, using *dm* (*17*), resulting in a new mean figure of merit of 0.83. The calculated electron density map following this procedure showed side chain density, derived solely from the MIR phases, that was very well-defined and allowed placement of almost all of the side chains in the original model using the program Xfit (*16*). Remaining effort was focused on building the missing 27% of the structure that was not part of the molecular replacement model. Two more rounds of density modification with the combined MIR phases allowed placement of an additional α-helix, several loops, and the DIP ligand in the active site. Phases were then extended to 2.5 Å followed by two more rounds of model building, where several more residues were added.

Three rounds of simulated annealing and conventional positional refinement in conjunction with manual adjustment of the model were performed. The X-PLOR refinement procedure included noncrystallographic symmetry restraints, where the restraint weight was gradually decreased as refinement progressed, allowing subtle differences in the monomers to be revealed. Annealed omit maps aided in the refinement of regions with ambiguous electron density. The present R-factor is 20.5% (free R-factor 29%), with good geometry (Table 1). In the Ramachandran plot generated

by PROCHECK (*19*), 85.9% are in most favored, 13.0% in additionally allowed, 0.5% in generously allowed, and 0.5% (Arg 113) in disallowed regions. The present model contains about 90% of the amino acids, with two surface loops at HSV-2 protease residues 104−110 and 134−140 not evident in the density. In addition, the first 16 N-terminal residues have no homology among the other herpes proteases and are not seen in this structure.

Unliganded HSV-2 Protease. The unliganded HSV-2 protease was crystallized from 20% PEG 8000, 0.1 M phosphate/citrate buffer at pH 4.2. Diffraction data were collected to 2.8 Å (94% complete; R_{merge} 9.5%) with space group and cell dimensions practically identical to the HSV-2/DIP crystal. The structure was determined by difference Fourier methods using the HSV-2/DIP protease structure, and refined to an *R*-factor of 22.4% with good geometry. The rms deviations are 0.017 Å for bond lengths, 2.1° for bond angles, 25.1° for dihedrals, and 1.9° for impropers. From the Ramachandran plot, 84.2% residues are in the most favored, 17.1% in additionally allowed, 1.1% in generously allowed, and 0.5% in disallowed regions.

HSV-1 Protease. Small crystals were grown from 9% PEG 8000 in Tris buffer at pH 8.5 at 4 °C over several months. The space group was *P1* with cell dimensions a = 79.6 Å, b = 81.2 Å, c = 93.4 Å, α = 115.5°, β = 98.4°, γ = 109.2°. The crystal used for data measurement was approximately 0.2 × 0.2 × 0.1 mm. The native data are 78.4% complete to 3.5 Å with an R_{sym} of 0.059. The asymmetric unit was calculated to have either three or four dimers (six or eight protease molecules).

Diffraction data were collected to 3.5 Å and processed using XDS. The HSV-1 protease structure was solved by molecular replacement using AMoRe (*17*). The model contained residues 17−103, 111−133, and 141−247 of the HSV-2/DIP structure. The rotation function was carried out with data between 8 and 4.0 Å resolution with a maximum search vector length of 31.3 Å. Only three pairs of peaks (reflecting the noncrystallographic symmetry of the dimer) were found with peak height greater than 0.5 times the maximum peak height. The translation function was calculated from 8−4 Å resolution by fixing the top solution in the P1 cell and searching for a second molecule, yielding a peak with a correlation coefficient of 41.0% and an *R*-factor of 38.5%. The top two solutions were then fixed to search for a third, yielding a peak of comparable height to the second solution. A search for a fourth solution showed smaller peaks of all about the same height. To see if there were three or four dimers in the cell, a rigid body fit was run to search for four dimers using the top three solutions with several of the similar peaks generated in the last translation function output. The fitting function yielded four peaks with a correlation coefficient of 45.6% and an *R*-factor of 37.1%. Alternatively, when the top three dimer solutions were fit, the correlation coefficient rose to 55.9% with an *R*-factor of 33.1%. Several other combinations of peaks were tried as controls, and none yielded satisfactory results as compared to the top three peaks. A packing diagram revealed a pseudotrimer of dimers. There were no overlaps between symmetry-related molecules.

Fourier coefficients were calculated using SIGMAA, followed by phase improvement by NCS averaging using *dm*. Electron density was visible for the residues unique to HSV-1 protease; however, there were similar disordered segments: 102−110, 134−143, and the first 14 residues of the N-terminus. The structure was rigid-body-refined from 10−3.5 Å using 12 346 reflections, giving an *R*-factor of 36.9%.

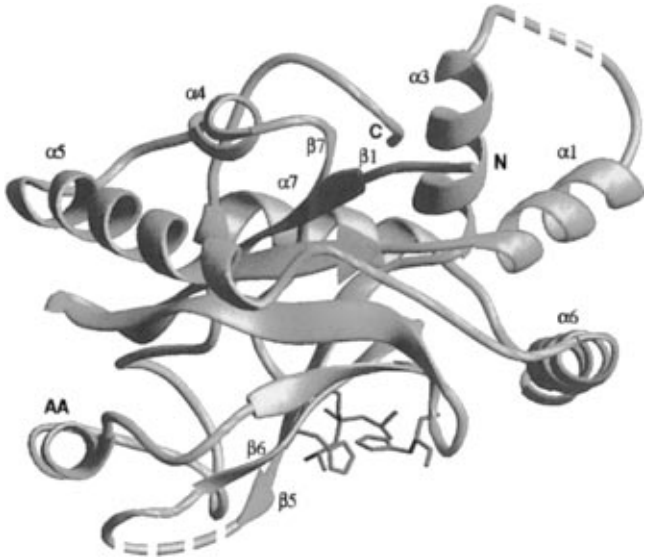

FIGURE 1: The HSV-2 protease monomer. Two disordered surface loops are shown as dashed lines (N-terminal residues 1−16 are disordered). The two histidines in the active site are shown in red, one on β6 and the other on the hairpin turn between β2 and β3, while the DIP covalently bound to the catalytic serine is in green on β5.

RESULTS AND DISCUSSION

The crystal structures of HSV-1, HSV-2, and HSV-2/DIP proteases were determined to 3.5 Å, 2.8 Å, and 2.5 Å resolution, respectively. The HSV-1 and HSV-2 proteases are homodimers with nearly identical subunits. The binding of DIP to HSV-2 protease does not alter the conformation of the enzyme. The rms deviation between the liganded and unliganded structures is 0.38 Å for Cα atoms. Thus, the following discussion will focus on description of only the crystal structure of the HSV-2/DIP protease complex.

The Overall Structure. The overall structure of the HSV-2 protease monomer (Figure 1) is similar to that of the previously reported structures of CMV and VZV proteases. It is comprised of a seven-stranded β-barrel with seven α-helices. The helices do not surround the barrel but rather cluster toward the edges. Helix AA seals one end while helices α1, α3, and α6 close the other. Helix α2 described in the CMV and VZV protease crystal structures (*10*, *14*) is disordered in the HSV-2 protease structure. The dimer interface of HSV-2 protease is made up of the interactions between helices α6 and α1 with the corresponding helices in the dimer mate. These same helices also interact in CMV protease but with a different orientation in relation to each other, resulting in a somewhat altered interface. Figure 2 shows that in CMV protease the α6 helices of each monomer are nearly coaxial. In HSV-2 protease, they are related by about 30°.

Catalytic Triad and DIP Binding. This is the first report of a crystal structure of a herpes protease−inhibitor complex. The presence of a ligand in the HSV-2 protease structure allows delineation of the enzyme active site cavity. Unlike trypsin, in which the active site cavity lies at the intersection of two domains, the active site cavity of HSV-2 protease

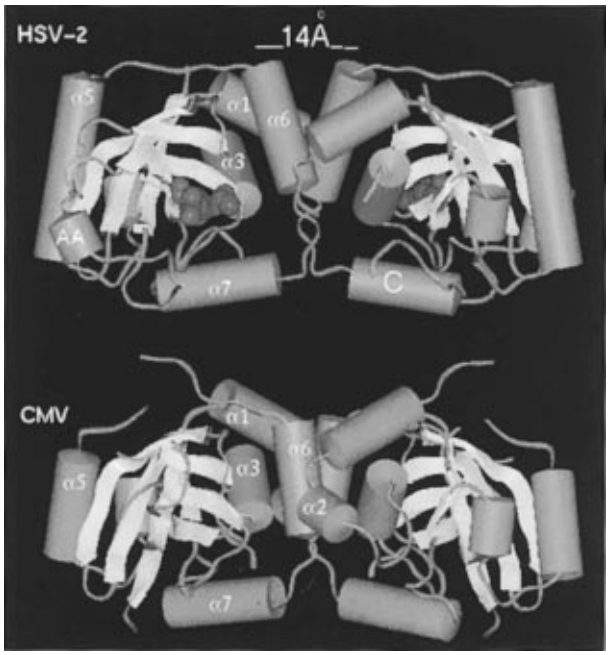

FIGURE 2: HSV-2 and CMV protease dimers with the monomers in about a 90° rotation from Figure 1. The two dimers were superimposed, and then separated to show how they differ. The DIP molecule (space-filling in blue) is shown to identify the position of the active site in HSV-2 protease. While the overall structures are similar, there are fewer disordered regions in HSV-2 as compared to CMV protease, with the presence of the loop containing helix AA and a longer α5 helix. The C-terminus of HSV-2 protease is labeled.

lies in a shallow groove along the β-barrel surface (Figure 1). The catalytic mechanism of classical serine proteases involves an active site triad composed of a serine, a histidine, and an aspartic acid. As in the other known herpes protease structures, the crystal structure of HSV-2 protease reveals an active site composed of a serine (Ser 132{129}),[2] a histidine (His 157{148}), and another histidine (His 63{61}). Electron density around the active site shows the DIP covalently bound to Ser 132{129}(Figure 3a), consistent with chemical modification studies that identified this serine as the active site nucleophile in HSV-1 and HSV-2 proteases (*20*, *21*).

The active site of HSV-2 protease reveals a network of hydrogen bonds between the enzyme active site residues, the ligand, and two central water molecules (Wat1 and Wat2) (Figure 3b). The crucial elements of the active site are strikingly similar to chymotrypsin even though the two enzymes share no sequence homology and the overall tertiary structures are completely different. An overlay of the catalytic triad of γ-chymotrypsin bound to monoisopropyl phosphate (MIP) with that of the HSV-2/DIP protease structure shows this similarity (Figure 3c). This supports the hypothesis that there are many conserved features in their mechanism and general mode of substrate recognition despite large differences in their overall structure and catalytic residues. DIP binding causes a local conformational change in the active site in which His 63{61} does not hydrogen bond to Ser 132{129}, but instead to Ser 134{131} (2.5 Å) (Figure 3a,b). This is a slightly different hydrogen bonding network when compared to CMV protease (Figure 3d). Despite its location, Ser 134 has been found to be nonessential for catalysis in CMV protease (*22*). This position is also an Ala residue in the HSV-1 and EBV proteases (*22*).

The Oxyanion Hole. Functional groups comprising the oxyanion hole stabilize the transition state intermediate by forming hydrogen bonds to the negatively charged oxygen atom of the substrate. In HSV-2 protease, Wat1 and the amide nitrogen of Arg 165{156} act to stabilize the P═O oxygen of the DIP and define the oxyanion hole of the enzyme (Figure 3b). Wat1 is stabilized by hydrogen bonds to Wat2 (2.7 Å) and the carbonyl oxygen of Val 131{128} (2.8 Å). Wat 2 is held by hydrogen bonds to backbone atoms of Leu 133{130} (2.9 Å) and Leu 20{27} (2.9 Å) and the side chain of Arg 166{157}N_ϵ (3.2 Å). The alignment with CMV protease also shows a water molecule in the active site region of this enzyme, closely overlapping with Wat2 in the HSV-2 protease structure (Figure 3d). This water molecule maintains the same protein backbone hydrogen bonds as does Wat2 in HSV-2 protease and could help hold the side chain of Arg 166{157} in place. Arg 165{156} and Arg 166{157} are absolutely conserved in all herpes proteases and present an overall positive charge near the oxyanion hole. The side chain of Arg 166{157} further makes two hydrogen bonds to backbone atoms of Leu 133{130} (not shown) and Leu 32{38} both absolutely conserved in all herpes proteases. The P═O oxygen of MIP in the structure of γ-chymotrypsin is stabilized by a hydrogen bond to the amide nitrogen of Gly 193. In HSV-2 protease, the amide nitrogen of Arg 165{156} closely overlaps with that of Gly 193 even though the overall structure of the two enzymes is completely different (Figure 3c). Although we had proposed this region as the oxyanion hole in the CMV and VZV protease structures (*10*, *14*), this is the first direct experimental evidence showing that Wat1 and the Arg 165{156} amide nitrogen define the oxyanion hole.

Structural Implications for Ligand Specificity. Our interest in the structures of herpes proteases stems from our desire to design selective drugs to either one or all of these enzymes. Thus, a better understanding of enzyme ligand specificity should facilitate that goal. An alignment of the HSV-2, CMV, and VZV protease structures shows that the structures are most similar in the β-barrel region and differ mainly in surface loops. The structures of HSV-2 and VZV proteases, both members of the α family, are similar (*14*) whereas greater differences are seen between the HSV-2 and CMV proteases (Figure 4).

Herpes protease precursor molecules undergo autoproteolytic cleavage at two sites, and all between Ala and Ser residues (*23*). The first cleavage, near the carboxy terminus, is known as the maturation (M) site, while the second (R-site) is responsible for the release of the amino-terminal catalytic domain (*22*). Little is known about how the protease domain releases itself from the precursor. The active site and the C-terminus are on opposite sides within a protease monomer (Figure 1). Within a dimer (Figure 2), the C-terminus of one monomer is on the same side as the active site of the other monomer, with a well-defined groove connecting the two. However, not only are they 29 Å apart, but also considerable conformational change must occur in order for the C-terminus to position itself properly and

[2] Under Results and Discussion, we will use CMV protease numbering (as in ref *10*) to describe all HSV protease residues. In most cases, the HSV protease numbers will be shown in "{}" brackets. This should eliminate any future confusion, and will help standardize numbering of catalytic triad residues as has been done with the chymotrypsin family of serine proteases.

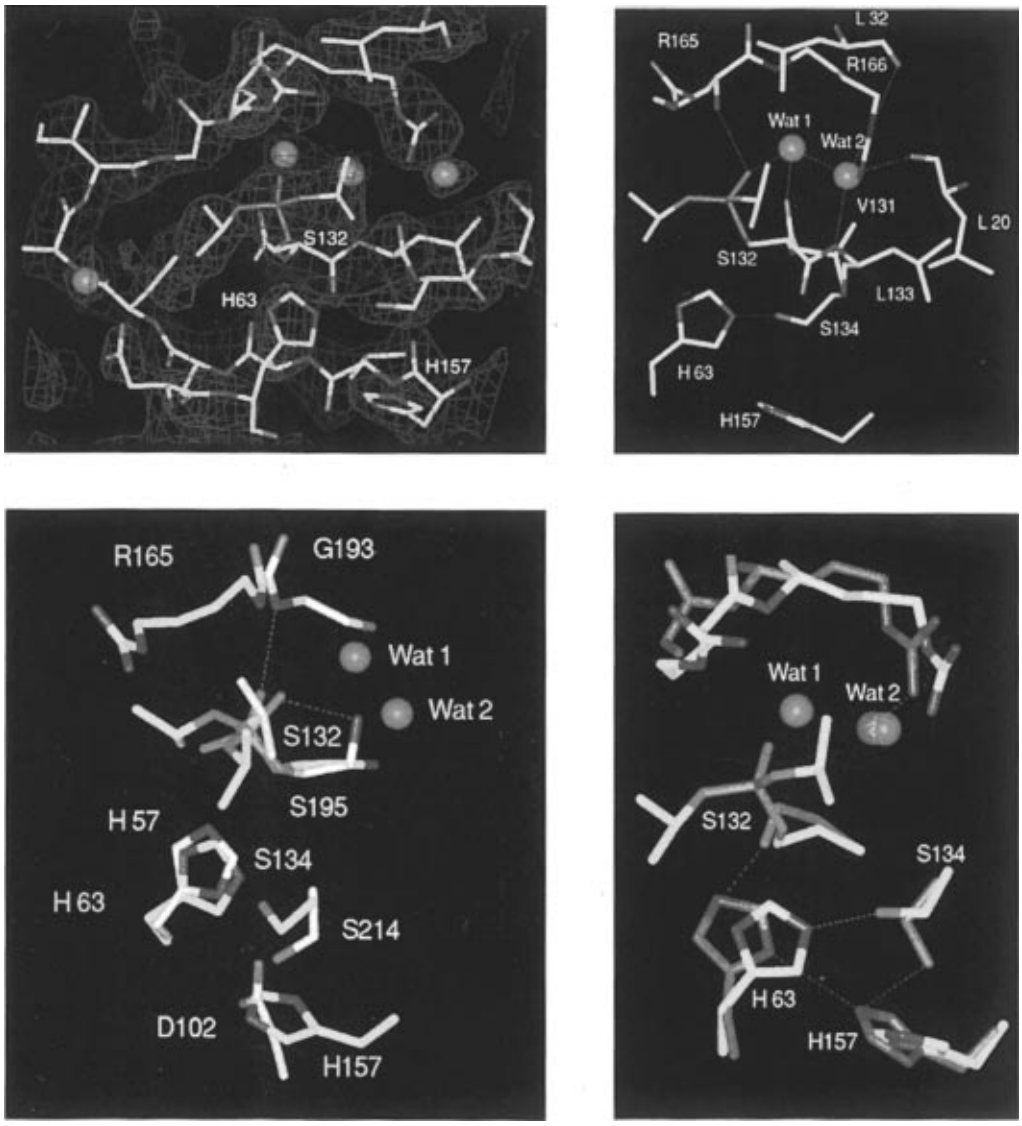

FIGURE 3: (a, top left) $2F_o - F_c$ electron density map (contoured at 1σ) showing a 9 Å sphere around the DIP molecule. Water molecules are shown as red spheres. (b, top right) Hydrogen bond network within the HSV-2/DIP protease active site. All active site residues are conserved between HSV-2 and CMV proteases except for Val 131 which is a leucine in CMV protease. (c, bottom left) Overlay of the active site of HSV-2/DIP protease and γ-chymotrypsin complexed with MIP (PDB Code 1GMH). The oxyanion hole is shown in chymotrypsin with hydrogen bonds between the P═O oxygen and the backbone nitrogens of Gly 193 and Ser 195 of γ-chymotrypsin. Carbon atoms and residue numbering are in gold for HSV-2 protease and in white for γ-chymotrypsin. (d, bottom right) An overlay of HSV-2 and CMV proteases. CMV protease carbon atoms and a water molecule are shown in green. The side chain of Arg 165 is missing in CMV protease.

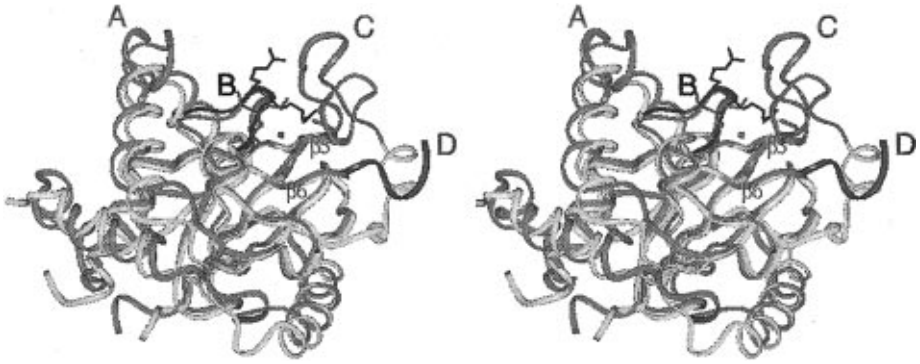

FIGURE 4: Stereoview of an overlay of HSV-2 (gold) and CMV (green) proteases. The DIP, Wat 1, and Wat2 are in red, and Arg 165 and 166 are in blue. Loops proposed to be involved in substrate recognition are identified and colored in HSV-2 protease: A (cyan), B (blue), C (magenta), and D (purple).

assume the correct orientation in the active site. Thus, the structure suggests that the protease is unlikely to act in *cis*.

Studies with both CMV (*24*) and HSV-1 (*23*, *25*) proteases have identified the P4−P1′ residues of the M-site as highly conserved between the two proteases, and having sequence-specific interactions with the S and S′ subsites of the protease. Figure 4 shows the active site cavity of HSV-2 and CMV protease as a shallow groove on the enzyme surface. The P

substrate residues could form an antiparallel β-sheet with β5 and β6, thus delineating the S subsite. A similar antiparallel hydrogen bonding pattern is seen in the trypsin family of serine proteases. In the herpes proteases, the P1 alanine could fit in a pocket created by the aliphatic portion of Arg 166 and Leu 32, while the P1′ serine residue could lie near the side chain of His 63 (Figures 4 and 3b). Upon substrate binding, Wat2 (present in both CMV and HSV-2 active sites) could be displaced and leave the guanidinium group of Arg 166 and the backbone amide of Leu 133 available for hydrogen bonding to the substrate residue at P2 (Figure 3b). A better understanding of peptide binding must await co-crystal structures with nonhydrolyzable peptides.

Evidence exists that substrate recognition/cleavage by the herpes proteases depends on more than a few residues around the cleavage site. Despite sharing the same core M-site sequence (VNA↓S), HSV-1 protease will not cleave at the CMV M-site; however, CMV protease will cleave at the corresponding HSV-1 site (*26*). The smallest peptide mimic of the CMV M-site that is cleaved by that protease is P4−P4′ (*24*), whereas 13 residues from P5−P8′ are required for cleavage by HSV-1 protease (*23*). This is surprising given the high sequence homology of residues lining the active site cavity of the two enzymes. It suggests that HSV-1 protease has a more extended substrate binding pocket, and that differences in substrate specificity between the two enzymes result from differences in loop conformations around the active site cavity. These loops show low sequence homology and are of differing lengths.

As with the chymotrypsin family of serine proteases, surface loops are proposed to influence substrate recognition of the herpes proteases (Figure 4). For example, loop D, following the active site serine between β5 and β6, could interact with noncore P substrate residues. This loop contains the I-site in CMV protease (*27*), an apparent cleavage site near the catalytic center that is unique to this enzyme. The loop, showing a high structural deviation between these enzymes, is mainly disordered in the CMV protease structure and partially missing in the HSV-2 and VZV protease structures, but would likely become ordered upon substrate binding. Loop C, disordered in the CMV protease structure (*10*), could also influence binding of P substrate residues. The role of these two loops in substrate recognition is also supported by substitution and deletion mutagenesis studies showing that changes in these loops alter R-site specificity (*22*). Loop B, following β6, includes the conserved arginines near the active site. The amino acid conservation in this region of the molecule suggests involvement in P1−P1′ sequence recognition, while loop A could influence interaction with the P′ residues beyond the core substrate sequence.

It has been postulated that the P2′−P8′ residues may play a conformational or structural role that is more length dependent than strictly sequence dependent in HSV-1 protease (*23*). This is in agreement with the shape of the active site cavity beyond S2′ that is large enough to accommodate a folded peptide.

The present structure, as a member of the herpes protease family, represents a new class of serine proteases. It sets an excellent and rather unique example of convergent evolution, reminiscent of examples from other protein families such as the nucleoside deaminases (*28*), the thioredoxin−glutathione reductases (*29*), and the lectins (*30*). The presence of the covalently bound transition state analog inhibitor DIP allows delineation of the active site cavity and the oxyanion hole, and permits greater understanding of ligand binding. Differences in substrate specificity among the herpes proteases appear to be influenced by surface loops in the vicinity of the active site.

ACKNOWLEDGMENT

We thank Lyn Gorniak and Arun Patel for activity assays, Jim Kane for fermentation studies, Catherine Peishoff, Christine Dabrowski, and Rick Keenan for useful discussions, and George Glover, Richard Jarvest, Hiro Nishikawa, and Martin Rosenberg for encouragement and support.

REFERENCES

1. Mocarski, E. S. (1993) in *The Human Herpesviruses* (Roizman, B., Whitely, R. J., and Lopez, C., Ed.) p 173, Raven Press, New York.
2. Levy, J. A. (1997) *Lancet 349*, 558−562.
3. Roizman, B., and Sears, A. E. (1993) in *The Human Herpesviruses* (Roizman, B., Whitely, R. J., and Lopez, C., Eds.) p 11, Raven Press, New York.
4. Corey, L., Adams, H. G., Brown, Z. A., and Holmes, K. K. (1983) *Ann. Int. Med. 98*, 958−972.
5. Gao, M., Matusick-Kumar, L., Hurlburt, W., DiTusa, S. F., Newcomb, W. W., Brown, J. C., McCann, P. J., Deckman, I., and Colonno, R. J. (1994) *J. Virol. 68*, 3702−3712.
6. Liu, F., and Roizman, B. (1991) *J. Virol. 65*, 5149−5156.
7. Liu, F., and Roizman, B. (1993) *J. Virol. 67*, 1300−1309.
8. Bazan, J. F., and Fletterick, R. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7872−7876.
9. Barr, P. J. (1991) *Cell 66*, 1−3.
10. Qiu, X., Culp, J. S., DiLella, A. G., Hellmig, B., Hoog, S. S., Janson, C. A., Smith, W. W., & Abdel-Meguid, S. S. (1996) *Nature 383*, 275−279.
11. Chen, P., Tsuge, H., Almassy, R. J., Gribskov, C. L., Katoh, S., Vanderpool, D. L., Margosiak, S. A., Pinko, C., Matthews, D. A., and Kan, C.-C. (1996) *Cell 86*, 835−843.
12. Shieh, H.-S., Kurumbail, R. G., Stevens, A. M., Stegeman, R. A., Sturman, E. J., Pak, J. Y., Wittwer, A. J., Palmier, M. O., Wiegand, R. C., Holwerda, B. C., & Stallings, W. C. (1996) *Nature 383*, 279−282.
13. Tong, L., Qian, C., Massariol, M., Bonneau, P. R., Cordingley, M. G., and Lagace, L. (1996) *Nature 383*, 272−275.
14. Qiu, X., Janson, C. A., Culp, J. S., Richardson, S. B., Debouck, C. M., Smith, W. W., and Abdel-Meguid, S. S. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 2874−2879.
15. Kabsch, W. J. (1993) *J. Appl. Crystallogr. 26*, 795−800.
16. McRee, D. E. (1993) *Practical Protein Crystallography*, Academic Press, San Diego.
17. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr. D50*, 760−763.
18. Brünger, A. T. (1992) *X-PLOR Version 3.1, A System for X-ray Crystallography and NMR*, Yale University Press, New Haven.
19. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
20. DiIanni, C. L., Stevens, J. T., Bolgar, M., O'Boyle, D. R., II, Weinheimer, S. P., and Colonno, R. J. (1994) *J. Biol. Chem. 269*, 12672−12676.
21. Pinto, I. L., West, A., Debouck, C. M., DiLella, A. G., Gorniak, J. G., O'Donnell, K. C., O'Shannessy, D. J., Patel, A., and Jarvest, R. L. (1996) *Bioorg. Med. Chem. Lett. 6*, 2467−2472.

22. Welch, A. R., McNally, L. M., Hall, M. R. T., and Gibson, W. (1993) *J. Virol. 67*, 7360−7372.
23. DiIanni, C. L., Mapelli, C., Drier, D. A., Tsao, J., Natarajan, S., Riexinger, D., Festin, S. M., Bolgar, M., Yamanaka, G., Weinheimer, S. P., Meyers, C. A., Colonno, R. J., and Cordingley, M. G. (1993) *J. Biol. Chem. 268*, 25449−25454.
24. Sardana, V. V., Wolfgang, J. A., Veloski, C. A., Long, W. J., LeGrow, K., Wolanski, B., Emini, E. A., and LaFemina, R. L. (1994) *J. Biol. Chem. 269*, 14337−14340.
25. McCann, P. J., O'Boyle, D. R., and Deckman, I. C. (1994) *J. Virol. 68*, 526−529.
26. Welch, A. R., Villarreal, E. C., and Gibson, W. (1995) *J. Virol. 69*, 341−347.
27. O'Boyle, D. R., II, Wager-Smith, K., Stevens, J. T., III, and Weinheimer, S. P. (1995) *J. Biol. Chem. 270*, 4753−4758.
28. Carter, C. W., Jr. (1995) *Biochimie 77*, 92−98.
29. Kuriyan, J., Krishna, T. S., Wong, L., Guenther, B., Pahler, A., Williams, C.H., Jr., & Model, P. (1991) *Nature 352*, 172−174.
30. Drickamer, K. (1995) *Structure 5*, 465−468.

BI9712697