Articles

Proteolysis of an Active Site Peptide of Lactate Dehydrogenase by Human Immunodeficiency Virus Type 1 Protease†

Thaddeus A. Tomaszek, Jr.,‡ Michael L. Moore,§ James E. Strickler,∥ Robert L. Sanchez,§ J. Scott Dixon,⊥ Brian W. Metcalf,ℱ Anne Hassell,∥,+ Geoffrey B. Dreyer,‡ Isobel Brooks,# Christine Debouck,# and Thomas D. Meek*,‡

Departments of Medicinal Chemistry, Peptidomimetics Research, Macromolecular Sciences, Physical and Structural Chemistry, Chemical and Biological Sciences, and Molecular Genetics, SmithKline Beecham Pharmaceuticals, 709 Swedeland Road, King of Prussia, Pennsylvania 19406

Mitchell Lewis

The Johnson Research Foundation, Department of Biochemistry and Biophysics, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Received May 5, 1992; Revised Manuscript Received August 6, 1992

ABSTRACT: The muscle and heart lactate dehydrogenase (LDHs) of rabbit and pig are specifically cleaved at a single position by HIV-1 protease, resulting in the conversion of 36-kDa subunits of the oligomeric enzymes into 21- and 15-kDa protein bands as analyzed by SDS–PAGE. While the proteolysis was observed at neutral pH, it became more pronounced at pH 6.0 and 5.0. The time courses of the cleavage of the 36-kDa subunits were commensurate with the time-dependent loss of both quaternary structure and enzymatic activity. These results demonstrated that deoligomerization of rabbit muscle LDH at acidic pH rendered its subunits more susceptible to proteolysis, suggesting that a partially denatured form of the enzyme was the actual substrate. Proteolytic cleavage of the rabbit muscle enzyme occurred at a decapeptide sequence, His-Gly-Trp-Ile-Leu*Gly-Glu-His-Gly-Asp (scissile bond denoted throughout by an asterisk), which constitutes a "strand–loop" element in the muscle and heart LDH structures and contains the active site histidyl residue His-193. The kinetic parameters K_m, V_{max}/K_mE_t, and V_{max}/E_t for rabbit muscle LDH and the synthetic decapeptide Ac-His-Gly-Trp-Ile-Leu*Gly-Glu-His-Gly-Asp-NH_2 were nearly identical, suggesting that the decapeptide within the protein substrate is conformationally mobile, as would be expected for the peptide substrate in solution. Insertion of part of this decapeptide sequence into bacterial galactokinase likewise rendered this protein susceptible to proteolysis by HIV-1 protease, and site-directed mutagenesis of this peptide in galactokinase revealed that the Glu residue at the P2′ was important to binding to HIV-1 protease. Crystallographic analysis of HIV-1 protease complexed with a tight-binding peptide analogue inhibitor derived from this decapeptide sequence revealed that the "strand–loop" structure of the protein substrate must adopt a β-sheet structure upon binding to the protease. The Glu residue in the P2′ position of the inhibitor likely forms hydrogen-bonding interactions with both the α-amide and γ-carboxylic groups of Asp-30 in the substrate binding site.

The *gag* and *pol* genes of HIV-1[1] respectively encode the structural *gag* proteins of the virion core (p17, p24, p9, and p7) and the essential enzymes of retroviral replication: protease, reverse transcriptase, ribonuclease H, and integrase (endonuclease). The initial translation products of these genes are two precursor polyproteins, $Pr55^{gag}$ and $Pr160^{gag\text{-}pol}$. During maturation of the HIV-1 virion, $Pr55^{gag}$ and $Pr160^{gag\text{-}pol}$ are proteolytically processed by HIV-1 protease to form the protease itself, from $Pr160^{gag\text{-}pol}$, and the other retroviral proteins. HIV-1 protease is an aspartic protease which is composed of a unique homodimeric structure (Meek et al., 1989; Wlodawer et al., 1989). In HIV-1, the complete processing of the $Pr55^{gag}$ and $Pr160^{gag\text{-}pol}$ apparently requires only eight discrete cleavages (Darke et al., 1988; Meek et al., 1990), suggesting that the digestion of these large protein substrates is limited by the stringent specificity of the enzyme either for certain amino acids at the cleavage sites or for the specific conformations effected by these peptide sequences

† Supported in part by NIH Grants GM-39526 (I.B.) and AI 24845. A preliminary account of some of this work was presented at the Gordon Research Conference on Enzymes, Coenzymes and Metabolic Pathways, Meridan, NH, July 4, 1991.

* Author to whom correspondence should be addressed.
‡ Department of Medicinal Chemistry.
§ Department of Peptidomimetics Research.
∥ Department of Macromolecular Sciences.
⊥ Department of Physical and Structural Chemistry.
ℱ Department of Chemical and Biological Sciences.
+ Present address: Glaxo Inc., 5 Moore Dr., Research Triangle Park, NC 27709.
# Department of Molecular Genetics.

[1] Abbreviations: AcHGWILG(A)EHGD-NH_2, acetyl-His-Gly-Trp-Ile-Leu-Gly(Ala)-Glu-His-Gly-Asp-NH_2; Boc, *tert*-butyloxycarbonyl; Cbz, benzyloxycarbonyl; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; HIV-1, human immunodeficiency virus, type 1; HPLC, high-performance liquid chromatography; LDH, lactate dehydrogenase; RM, rabbit muscle; Mes, 2-(*N*-morpholino)ethanesulfonic acid; NADH, nicotinamide adenine dinucleotide (reduced form); SDS–PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; TBDMS, *tert*-butyldimethylsilyl.

0006-2960/92/0431-10153$03.00/0

within these polyprotein substrates. The active site of the protease may accommodate six to eight amino acids in its binding cleft (Fitzgerald et al., 1990; Miller et al., 1989; Erickson et al., 1990).

However, the amino acids found at the P4–P4′ positions [nomenclature of Schechter and Berger (1967)] of the proteolytic cleavage sites of $Pr55^{gag}$ and $Pr160^{gag\text{-}pol}$ are surprisingly diverse. While the scissile dipeptides of the HIV-1 proteolytic processing sites are generally hydrophobic (Try*Pro, Leu*Ala, Met*Met, Phe*Leu, Phe*Pro, Phe*Tyr, and Leu*Phe), the residues in the other positions vary widely. Henderson et al. (1988) have categorized the cleavage site residues recognized by HIV-1 protease and other retroviral proteases into three classes: Class 1 sites contain a Tyr(Phe)-*Pro scissile dipeptide, and a preference for Asn and a hydrophobic residue at P2 and P2′, respectively. Class 2 sites contain an Arg at P4 and Phe–Leu at the P1′–P2′ positions. Class 3 sites contain Glu or Gln at P2′ and have a preference for hydrophobic residues at the P2, P1, and P1′ positions. More recently Pettit et al. (1991) have narrowed into two classes the cleavage site sequences in retroviral polyprotein substrates. Poorman et al. (1991) have performed a statistical analysis of the frequency with which a given residue is found at the P4–P4′ positions of protein substrates and have concluded that the highest stringency for a given residue is found at the P2, P1, and P2′ positions. These findings with protein substrates are borne out with specificity studies of oligopeptide substrates of HIV-1 protease (Drake et al., 1988; Moore et al., 1989; Phylip et al., 1990; Richards et al., 1990; Konvalinka et al., 1990; Pettit et al., 1991; Poorman et al., 1991), which also confirm that heptapeptides (P4–P3′) comprise the minimal size for a good substrate. The diversity of residues at the cleavage sites of both protein and peptide substrates suggests that factors other than primary structure, such as the secondary and tertiary structures found at the processing sites within a protein substrate, may play a role in the recognition of a protein cleavage site by HIV-1 protease.

To date, very little is known about the role of protein structure in the recognition of a cleavage site by a protease. Apart from HIV-1 protease itself, no structural information is currently available for its natural substrates, $Pr55^{gag}$ and $Pr160^{gag\text{-}pol}$. Several nonviral proteins, such as *Pseudomonas* exotoxin (Tomasselli et al., 1990), the filament proteins vimentin and desmin (Shoeman et al., 1990), covalently modified ribonuclease A (Hui et al., 1990), calcium-depleted calmodulin (Tomasselli et al., 1991a), and others (Tomasselli et al., 1991b), are specifically cleaved by HIV-1 protease, and the sequences of their cleavage sites are overwhelmingly of the "class 3" type. In order to study the role of protein structure in proteolytic site recognition by HIV-1 protease, we sought to identify a protein substrate from a large array of proteins of known three-dimensional structure and to define the structural features other than the primary structure of the characterized cleavage sites which impart specificity. Here, we report the elucidation of a highly conserved, active site peptide sequence found within the mammalian lactate dehydrogenases that comprises a novel proteolytic processing site for HIV-1 protease and discuss elements of its three-dimensional structure that contribute to its recognition by HIV-1 protease.

EXPERIMENTAL PROCEDURES

Enzyme and Chemicals. Recombinant HIV-1 protease was obtained from the PRO4 expression vector in *Escherichia coli* strain AR58 as previously described (Debouck et al., 1987, Strickler et al., 1989). The protease was purified, either by the method of Strickler et al. (1989) or by modifications of this procedure as described by Grant et al. (1991), from the supernatants of bacterial crude extracts to >95% homogeneity, as judged by both amino acid analysis and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE). The purified protease was routinely stored at −20 °C in 20 mM Tris–HCl (pH 7.5), 1 mM DTT, 1 mM EDTA, 200 mM NaCl, and 40% glycerol (enzyme storage buffer) at protein concentrations of 10–30 μg/mL.

Protein concentrations were measured either by use of the BCA protein reagent of Pierce Chemical Co. or, for rabbit muscle lactate dehydrogenase, by spectrophotometric detection at 280 nm [A^{280nm} (1% solution, 1 cm) = 12.3; Pesce et al. (1967)]. For HIV-1 protease, protein concentrations were also determined by analytical reversed-phase HPLC. The concentrations of active protease in the purified enzyme preparations were routinely determined by active site titration as described (Grant et al., 1991), and typically, 80–100% of the active sites of HIV-1 protease were found to be active in the enzyme preparations used in these studies.

Lactate dehydrogenases from rabbit muscle, pig muscle, chicken muscle, and *Bacillus stearothermophilus* were purchased from either Boehringer Mannheim Biochemicals or Sigma Chemical Co. and used without further purification. NADH, NAD, sodium pyruvate, and oxamic acid were obtained from Sigma. EDTA (Gold Label grade) was purchased from Aldrich Chemical Co. Trifluoroacetic acid (Sequanal grade) was purchased from Pierce Chemical Co. Solvents used for HPLC were from Burdick-Jackson Laboratories. Buffers, Triton X-100, and DTT were obtained from Sigma Chemical Co. All other chemicals were of the highest available quality.

Buffers. The following buffers were used: 50 mM 2-(*N*-morpholino)ethanesulfonic acid (Mes) (pH 6.0), 1 mM EDTA, 0.2 M NaCl, 1 mM DTT, and 0.1% (v/v) Triton X-100 (MENDT buffer); 50 mM each of glycine, sodium acetate, Mes, Tris–HCl (variable pH), 1 mM DTT, 1 mM EDTA, 0.2 M NaCl, and 0.1% (v/v) Triton X-100 (GAMT-NEDT buffer); 50 mM each of sodium acetate, Mes, Tris, 1 mM EDTA, 1 mM DTT, and 200 mM NaCl (pH 5–7) (AMT-EDN buffer).

Peptide Substrates. Ac-His-Gly-Trp-Ile-Leu-Gly-Glu-His-Gly-Asp-NH_2 (AcHGWILGEHGD-NH_2), Ac-Trp-Ile-Leu-Gly-Glu-His-NH_2 (Ac-WILGEH-NH_2), and Ac-His-Gly-Trp-Ile-Leu-Ala-Glu-His-Gly-Asp-NH_2 (Ac-HGWIL-AEHGD-NH_2) were prepared by solid-phase synthesis on benzhydrylamine resin and cleaved from the resin with anhydrous HF at 0 °C. The oligopeptides were purified to homogeneity by either thin-layer chromatography or reversed-phase HPLC, and structures were confirmed by amino acid composition and fast-atom-bombardment mass spectrometry.

Synthesis of Ac-His-Gly-Trp-Ile-Leuψ[CHOHCH$_2$]Gly-Glu-His-Gly-Asp-NH$_2$. Proton magnetic resonance (NMR) spectra were recorded at 250 MHz on a Bruker AM 250 spectrometer. Desorptive chemical ionization mass spectra (MS) were obtained with a Finnigan-MAT quadrupole mass spectrometer. Flash chromatography was performed with E. Merck Kieselgel 60. HPLC was performed with Dynamax silica (5 μM) (Rainin).

(*a*) *(6S)-8-Methyl-6-[(tert-butyloxycarbonyl)amino]-5-oxonon-1-ene* (*1*). A solution of 10 g of (*tert*-butyloxycarbonyl)-L-leucine *N*,*O*-dimethylhydroxylamide (prepared by the procedure of Goel et al. (1989) in 100 mL of ether was

added over 40 min to a freshly prepared solution of 4-butenylmagnesium bromide (0.2 mol) in 100 mL of ether maintained at 0 °C. After 2 h, dilute HCl was added. Extractive workup followed by flash chromatography (10% ethyl acetate in hexanes) provided **1** (8.0 g, 75% yield): NMR ($CDCl_3$) δ 5.79 (1 H, m), 5.0 (3 H, m), 4.30 (1 H, m), 2.59 (2 H, m), 2.32 (2 H, m), 1.69 (1 H, m), 1.40 (9 H, m), 1.34 (2 H, m), 0.95 (6 H, dd); MS (DCl/NH_3) m/z 270 $(M + H)^+$, 231, 214, 186, 170.

(*b*) *(5S)-7-Methyl-5-[(tert-butyloxycarbonyl)amino]-4-oxooctanoic Acid (2).* Potassium permanganate (14 g) was added to a vigorously stirring mixture of **1** (6.82 g) and $(n\text{-}Bu)_4NBr$ in benzene (200 mL), water (200 mL), and acetic acid (27 mL). After 2 h, the black mixture was cooled in ice and 200 mL of saturated aqueous $NaHSO_3$ was added. Filtration, extractive workup, and flash chromatography (80:20:3 n-hexane/ethyl acetate/acetic acid) provided **2** (6.8 g, 86% yield): NMR ($CDCl_3$) δ 5.0 (1 H, d), 4.35 (1 H, m), 2.85 (2 H, m), 2.65 (2 H, m), 1.7 (1 H, m), 1.45 (11 H, m), 1.0 (6 H, d); MS (DCl/NH_3) m/z 288 $(M + H)^+$, 249, 232, 188.

(*c*) *Benzyl (4S,5S)-4-(tert-Butyldimethylsiloxy)-5-[(tert-butyloxycarbonyl)amino]-7-methyloctanoate (3).* To a solution of **2** (4.0 g) in acetonitrile (60 mL) was added 1.8-diazabicyclo[5.4.0]undec-7-ene (2.0 mL) followed by benzyl bromide (2.25 mL). After 3 h, acidification, extractive workup, and flash chromatography (10% ethyl acetate in hexanes) afforded the crude γ-keto ester **2a**. $NaBH_4$ (0.11 g) was added to a solution of keto ester **2a** (2.10 g) in methanol (70 mL) at 0 °C. After 30 min, acetic acid (1 mL) was added. The mixture was diluted with water and extracted with CH_2Cl_2. The organic layer was dried over $MgSO_4$ and concentrated to afford the crude γ-hydroxy esters (2.1 g), which were then combined in dimethylformamide (12 mL) with imidazole (1.5 g) and *tert*-butyldimethylsilyl chloride (1.5 g). After 3 days, workup and flash chromatography (10% ethyl acetate in hexanes) provided an epimeric mixture of silyl ethers (3.9 g). The epimers were separated by HPLC (5:95 ethyl acetate/hexanes) to afford **3** (0.27 g) and its 4*R* diastereomer (1.31 g): NMR ($CDCl_3$) δ 7.35 (5 H, s), 5.12 (2 H, s), 4.52–4.30 (1 H, 2 d's), 3.66 (2 H, m), 2.43 (2 H, m), 1.78 (2 H, m), 1.58 (1 H, m), 1.43 (9 H, s), 1.29 (2 H, m), 0.9 (15 H, m), 0.4 (6 H, s); MS (DCl/NH_3) m/z 494 $(M + H)^+$, 411, 308.

(*d*) *(4S,5S)-4-(tert-Butyldimethylsiloxy)-5-[(tert-butyloxycarbonyl)amino]-7-methyloctanoic Acid (4).* Ester **3** (0.25 g) was stirred with 10% Pd on carbon (28 mg) in ethyl acetate (5 mL) under an atmosphere of H_2 for 12 h. Filtration and removal of solvent afforded **4** (0.22 g) as a glass. The NMR spectrum of **4** indicates conformational isomerism, as previously observed for related compounds (Dreyer et al., 1992): NMR ($CDCl_3$) δ 10.65 (1 H, br), 5.55 ($^2/_5$ H, d), 4.53 ($^3/_5$ H, d), 3.7–3.5 (2 H, m), 2.42–2.28 (2 H, m), 1.72 (2 H, m), 1.59 (1 H, m), 1.45 (9 H, d), 1.29 (2 H, m), 0.9 (15 H, d), 0.4 (6 H, s); MS (DCl/NH_3) m/z 404 $(M + H)^+$, 348,304.

Ac-His-Gly-Trp-Ile-LeuΨ[CHOHCH$_2$]Gly-Glu-His-Gly-Asp-NH_2 was obtained from **4** by solid-phase synthesis of benzylhydrylamine resin using previously described methods (Moore et al., 1989).

Electrophoresis Methods. The proteolytic products of LDH following treatment with proteases were separated by SDS–PAGE (15% polyacrylamide) on "minigels" (13 × 4 × 0.1 cm; Bio-Rad Mini-Protein apparatus). Proteins were visualized by staining with Coomassie Brilliant Blue. Protein species within Coomassie-stained gels were quantified by densitometric scanning using an LKB Ultroscan 250 laser densitometer equipped with a Hewlett-Packard 3390A digital integrator. The prestained, low molecular weight standards of BRL Laboratories were used: ovalbumin, 43 000; carbonic anhydrase, 29 000; β-lactoglobulin, 18 400; lysozyme, 14 300; bovine trypsin inhibitor, 6200; insulin, 3000.

Proteolysis of the Lactate Dehydrogenases. Proteolysis of the LDHs by HIV-1 protease was conducted at 37 °C. Mixtures (50–200 μL) containing AMT-EDN buffer (pH 5–7) and 0.5 mg/mL LDH (14 μM monomer) were preincubated for 10 min at 37 °C prior to initiation of reaction by the addition of 5–50-μL aliquots of HIV-1 protease (20–600 ng; 20–540 nM, final concentrations). At various times, 20-μL aliquots were withdrawn, and reaction was quenched by boiling 2 min in 20 μL of SDS treatment buffer [0.125 M Tris–HCl (pH 6.8), 20% glycerol, 10% 2-mercaptoethanol, 4% SDS, and 0.01% bromophenol blue). The quenched samples were then subjected to SDS–PAGE as described.

Initial Velocity Measurements. The initial velocities of the HIV-1 protease-catalyzed proteolysis of rabbit muscle LDH were determined by measurement of the remaining enzymatic activity of LDH following its treatment with HIV-1 protease. Reaction mixtures (0.1 mL) containing AMT-EDN buffer (pH 5.0–7.0) and 0.5 mg/mL rabbit muscle LDH (RM-LDH, 3.5 μM tetramer) were preincubated for 30 min at 37 °C in buffer prior to the addition of HIV-1 protease (30–700 nM, final concentrations). After various reaction times (2–120 min), 5–10-μL aliquots were diluted into 1 mL of 0.1 M potassium phosphate (pH 7.0) and incubated at room temperature for 30 min. Remaining LDH activity was then assayed by addition of 100-μL aliquots to 0.9-mL assay mixtures containing 0.1 M potassium phosphate (pH 7.0), 7.6 mM sodium pyruvate, and 0.2 mM NADH (equilibrated at 25 °C). LDH activity was determined spectrophotometrically by monitoring the continuous decrease in NADH absorbance ($\epsilon^{340nm} = 6220\ M^{-1}\ cm^{-1}$) at 25 °C. The resulting initial rates of LDH proteolysis were determined from time courses (0–10 min) of the percentage of remaining LDH activity, based on its initial concentration (as determined from its UV absorption spectrum). Initial rates were determined from these data in units of nmoles of LDH proteolyzed per minute, in which the amount of proteolyzed LDH was calculated from the percentage of LDH activity lost. Each rate was corrected for the amount of LDH activity lost not due to proteolysis using identical control samples which contained either no HIV-1 protease or protease inhibited with 10 μM Ala-Ala-PheΨ[CHOHCH$_2$]Gly-Val-Val-OMe [a competitive inhibitor; Dreyer et al. (1989)]. Initial rates at variable concentrations of RM-LDH (5–40 μM, monomer concentration) were evaluated as above and were obtained from time courses (0–10 min) of the loss of LDH enzymatic activity following preincubation with HIV-1 protease at each concentration of LDH.

The peptidolytic activity of HIV-1 protease upon the oligopeptide substrates AcHGWILGEHGD-NH_2, Ac-HG-WILAEHGD-NH_2, and Ac-WILGEH-NH_2 were measured using an HPLC-based assay under conditions similar to those which were previously described (Meek et al., 1989; Moore et al., 1989). The fraction of reaction was calculated as the ratio of (product formed)/(product formed + substrate remaining) from digital integration of the following pairs of substrate and product peaks detected at 220 nm: Ac-HGWIL/AcHGWILGEHGD-NH_2, Ac-HGWIL/Ac-HGWILAEH-GD-NH_2, and Ac-WIL/Ac-WILGEH-NH_2. The products

GEHGD-NH_2, AEHGD-NH_2, and GEH-NH_2 did not appear in the chromatographic retention profile, probably due to their elution in the apparent void volume.

Aqueous solutions (≤10 mM) of each oligopeptide substrate were freshly prepared for each experiment from lyophilized material. The solution concentrations of the substrates in solution were determined from their UV spectra by assuming extinction coefficients of ϵ^{275nm} = 5500 M^{-1} cm^{-1} for tryptophanyl-containing peptides.

The time course of the protease-catalyzed peptidolytic reaction for each oligopeptide substrate was ascertained under all of the experimental conditions described below. Initial rate data were in most cases obtained from single time points ($t \leq 20$ min) at which substrate turnover was low (≤20%).

Inhibition constants for potent inhibitors of HIV-1 protease were determined as described (Grant et al., 1991). The pH dependence of V_{max}/K_mE_t and V_{max}/E_t of Ac-HGWILAE-HGD-NH_2 was determined by the method of Hyland et al. (1991).

Protein Sequencing. A sample containing RM-LDH which had been extensively cleaved by HIV-1 protease was subjected to sequence analysis by automated Edman degradation using an Applied Biosystems Model 470A gas-phase Sequenator. Sequencing of the N-terminus of the 15-kDa proteolytic fragment was performed on an unpurified sample, owing to the low mole fraction of HIV-1 protease in the sample and the fact that the N-terminus of the native LDH is acetylated. The resulting phenylthiohydantoin-amino acids (PTH-amino acids) were analyzed as described (Strickler et al., 1989).

Size-Exclusion Chromatography. The apparent molecular weights under nondenaturing conditions of RM-LDH and its proteolytic fragments were determined by size-exclusion chromatography. Proteins were separated by HPLC on a Bio-Sil TSK-400 column (300 mm × 7.5 mm; Bio-Rad), in which the isocratic mobile phase consisted of 50 mM KH_2PO_4 and 300 mM KCl (pH 7.0) at a flow rate of 0.5 mL/min. Analytical samples (100 μL) contained 50–100 μg of LDH. Peak detection was conducted simultaneously at 220 and 280 nm. The protein standards used, cited below with their corresponding molecular weights and elution times, were obtained from Bio-Rad: thyroglobulin (M_r = 670 000; 19.2 min), γ-globulin (M_r = 158 000; 22.0 min), ovalbumin (M_r = 44 000; 24.5 min), myoglobin (M_r = 17 000; 27.0 min), and vitamin B-12 (M_r = 1350; 32.1 min).

Data Analysis. Kinetic data were fitted to eq 1 using the FORTRAN programs of Cleland (1979). Values of $V_{max}/$

$$v = V_{max}A/(K_m + A) \quad (1)$$

K_m and V_{max} were obtained by fitting of initial velocity data at variable concentrations of the oligopeptide substrates to eq 1. In the equations below, v is the initial velocity, V_{max} is the maximum velocity, K_m is the Michaelis constant, and A is the concentration of variable substrate.

Tight-binding inhibition data from plots of residual enzyme activity (v_i/v_0) vs inhibitor concentration (I_t) at each fixed protease concentration (E_t) were fitted to eq 2, in which α is

$$v_i/v_0 = [\alpha E_t - I_t - K_i' + [(K_i' + \alpha E_t - I_t)^2 + 4K_i'I_t]^{1/2}]/2\alpha E_t \quad (2)$$

the fraction of active enzyme (0–1.0), E_t is the nanomolar enzyme concentration as determined by the chromatographic method described above, and K_i' is the apparent inhibition constant, which equals $K_i(1 + A/K_m)$. Use of eq 2 assumes linear competitive inhibition. Equation 2 was analyzed by using the SUPERFIT package, a nonlinear regression analysis utilizing the method of Marquardt (1983) and developed in-house.

Crystallization of the HIV-1 Protease–Ac-His-Gly-Trp-Ile-Leuψ[CHOHCH$_2$]Gly-Glu-His-Gly-Asp-NH$_2$ Complex and Data Collection. Cocrystals of purified HIV-1 protease complexed with Ac-His-Gly-Trp-Ile-Leuψ[CHOHCH$_2$]-Gly-Glu-His-Gly-Asp-NH_2 were prepared by mixing the inhibitor with the protease (7.2 mg/mL) at an enzyme: inhibition stoichiometric ratio of 1:2. Crystals were grown by adding equal volumes (3 μL) of the precipitating solution, containing NaOAc buffer (50 mM NaOAc (pH 5.2–5.6), 1 mM DTT, 1 mM EDTA, and 50 mM NaCl) and 0.2–0.6 M ammonium sulfate, with the solution of the protein–inhibitor complex (NaOAc buffer). By use of the hanging drop vapor diffusion method, the protease–inhibitor complex was mixed with the precipitating solution on a cover glass, inverted, and sealed with silicone vacuum grease (Dow-Corning) over a reservoir containing 1 mL of the precipitating solution. Samples were then incubated at 22 °C for 1–2 weeks, and rectangular-shaped crystals appeared within 4–10 days. These crystals were screened by precession photography, and the symmetry was consistent with the orthorhombic space group $P2_12_12$ with unit cell dimensions at a = 58.4 Å, b = 86.8 Å, and c = 46.3 Å. These crystals were isomorphous with those reported by Fitzgerald et al. (1990) for the acetyl-pepstatin–HIV-1 protease complex.

The cocrystals were prepared for data collection using standard techniques. Three-dimensional diffraction data were recorded to a nominal resolution of 2.2 Å using a Siemans multiwire X-ray detector equipped with a rotating anode X-ray source. X-rays were generated at 40 kV and 50 mA using 200-μm focal cup with a graphite monochromator. Data were recorded by collecting a series of 10′ oscillation frames, in which each frame was collected for a period of 400 s. A total of 1200 frames of data were collected from three crystals, and the data were corrected for radiation damage and absorption effects. The resolution of the diffraction varied from crystal to crystal and the effective resolution of the data was 2.5 Å.

Structure Determination. The crystal structure of the HIV-1 protease–Ac-His-Gly-Trp-Ile-Leuψ[CHOHCH$_2$]Gly-Glu-His-Gly-Asp-NH_2 was solved by molecular replacement using model coordinates of the enzyme complexed with other hydroxyethylene isostere peptide analogues (Dreyer et al., 1992). The solution was consistent with the coordinates of the isomorphous crystal structure of the acetyl-pepstatin–HIV-1 protease complex provided by Fitzgerald (personal communication). The calculated structural amplitudes had a residual of 38% for data to 3-Å resolution. The preliminary electron density maps were used to improve the fit of our initial model. A summary of the crystallographic data is presented in Table I.

Structure Refinement. The improved model was initially refined with the restrained least squares program PROLSQ (Hendrickson, 1985). Since the two halves of the enzyme are chemically identical we imposed noncrystallographic symmetry constraints on the homodimer. After several cycles of least squares refinement, the R value dropped to 26% for data between 10- and 3-Å resolution. Electron density maps were recalculated with coefficients $2F_o - F_c$ and $3F_o - 2F_c$ using the refined phases, and the model was examined residue by residue. Several torsional angles were adjusted to optimize the fit to the electron density. The symmetry restraints were gradually removed and data with d spacings of >2.5-Å

Table I: Summary of Crystallographic Data for HIV-1 Protease–Ac-His-Gly-Trp-Ile-LeuΨ[CHOHCH$_2$]Gly-Glu-His-Gly-Asp-NH$_2$ Complex

Intensity Measurements and Refinement			
resolution range (Å)	no. of observations	no. of unique observations	residual
15.00–5.66	2992	843	23.0
5.66–4.00	3434	1427	18.2
4.00–3.27	3982	1810	16.4
3.27–2.83	4627	2189	19.3
2.83–2.50	3647	1797	22.1
15.00–2.50	18672	8066	19.2

Model Parameters	
rms deviation in bond lengths (Å)	0.021
rms deviation in angle distances (Å)	0.038
rms deviation in planar 1–4 distances (Å)	0.043

resolution were added. After further cycles of refinement, the residual was less than 24%. Difference electron density in the vicinity of the substrate(inhibitor) could be modeled in two orientations, analogous to that observed in the acetyl-pepstatin–HIV-1 protease crystals (Fitzgerald et al., 1990).

The initial model of the inhibitor was fit to the density with an extended conformation and positioned in the active site analogously to that of other hydroxyethylene isostere-containing peptide analogues (Dreyer et al., 1992). A modified, restrained least squares program was used to refine the enzyme structure and the two copies of the bound inhibitor simultaneously. The population of the inhibitor in each orientation was assigned an occupancy of 0.5 and not refined. Several additional cycles of refinement interspersed with graphical rebuilding reduced the residual to 22%. The final cycles of refinement were performed using the program XPLOR (Brunger et al., 1987), and the final residual was 19%. The numbering of the residues in this model are 1–99 and 101–199 for residues of the two HIV-1 protease monomers and 201–210 for the decapeptide analogue inhibitor.

In Vivo Proteolytic Processing of Recombinant Bacterial Galactokinase Engineered with Cleavage Sites of HIV-1 Protease. Synthetic oligonucleotides encoding various cleavage sites for HIV-1 protease and mutants thereof were inserted in-frame at the *Cla*I site within *E. coli* galactokinase (galK) (Debouck et al., 1985). Engineered galK was coexpressed with HIV-1 protease in *E. coli* using the double-plasmid system and induction with naldixic acid as previously described (Debouck et al., 1987; Mizrahi et al., 1989). The extent of galK processing by HIV-1 protease after 2 h of naldixic acid induction was monitored by subjecting the induced protein samples to SDS–PAGE followed by Western blotting using a rabbit polyclonal antibody to galK.

RESULTS

Cleavage of Lactate Dehydrogenases by HIV-1 Protease. As analyzed by SDS–PAGE, treatment of rabbit muscle LDH with HIV-1 protease at pH 5.0 resulted in the time-dependent appearance of two protein products of 21 and 15 kDa from the intact 36-kDa subunit of LDH (Figure 1). The 21-kDa protein band may contain two protein bands of nearly identical electrophoretic migration. At pH 5.0, the appearance of the 21- and 15-kDa protein fragments increased in a time-dependent manner (60–240 min) and increased at increasing levels of protease (20–140 nM; data not shown). Additional proteolytic products were apparent only at high levels (140 nM) of protease, at which point a third protein band of 9.4 kDa was weakly visible.

Under identical reaction conditions, HIV-1 protease cleaved heat-denatured rabbit muscle LDH into 21- and 15-kDa protein fragments to a considerably lesser extent, and additional protein bands were observed (Figure 1A, lane 5). This indicates that elements of the secondary and, possibly, quaternary structures of LDH are essential to its facile recognition by HIV-1 protease. The presence of a competitive, peptide analogue inhibitor, Ala-Ala-PheΨ[CHOHCH$_2$]Gly-Val-Val-OMe (K_i = 12 nM) at 10 μM completely inhibited proteolysis of rabbit muscle LDH at either pH (Figure 1A, lane 6). This demonstrates that the proteolysis of LDH results from HIV-1 protease and not a contaminating protease. A replicate set of samples as shown in Figure 1 but performed at pH 6.0 demonstrated similar results, but proteolysis occurred to a lesser extent in all samples (data not shown).

Likewise, pig muscle (Figure 1A, lane 8), rabbit heart (data not shown), and pig heart (data not shown) LDHs were cleaved into 21- and 15-kDa protein fragments at pH 5.0, although less readily than the rabbit muscle enzyme. Little cleavage of the pig muscle LDH was observable at pH 6.0. It is therefore likely that the cleavage sites observed in the rabbit muscle, pig muscle, and pig heart LDHs are similar, if not identical, but the cleavage site in the rabbit muscle enzyme is more accessible by the protease. Discrete proteolytic products of 21 and 15 kDa were not apparent upon treatment of the LDHs of chicken muscle (Figure 1A, lane 10) or chicken heart (data not shown) with HIV-1 protease as either pH 5.0 or 6.0.

Treatment of the LDH–NADH–Oxamate Complex by HIV-1 Protease. Oxamate is a competitive inhibitor (vs pyruvate) of LDH and binds tightly to E-NADH to form a ternary complex at pH <7.0 (Holbrook & Stinson, 1973). As shown in Figure 1B, at pH 6.0, the time-dependent cleavage of RM-LDH into 21- and 15-kDa protein fragments by HIV-1 protease is greatly decreased in the presence of inhibitory levels of NADH and oxamate under conditions at which the uncomplexed enzyme is extensively cleaved. These inhibitors also afford some protection from proteolysis at pH 7.0. However, the inhibition of the proteolysis by inhibitor complexation of RM-LDH is not as substantial as that observed by direct inhibition of the protease. This demonstrates that the scissile bond within RM-LDH is made less accessible to HIV-1 protease in the ternary inhibitory complex and suggests that it may be proximal to or contained within the the enzyme's active site.

Effect of pH on the Stability of RM-LDH and Its Susceptibility to Proteolytic Cleavage. RM-LDH was incubated at pH 5.0, 6.0, and 7.0 with HIV-1 protease, and the effects of limited proteolysis on enzymatic activity, primary structure, and quaternary structure were respectively assessed by spectrophotometric assay of residual enzymatic activity (pyruvate reduction) at pH 7.0, by analysis of the loss of the intact 36-kDa RM-LDH monomer by SDS–PAGE and by assessment of the loss of the RM-LDH oligomer by analytical gel filtration at pH 7.0 (Figure 2).

Untreated RM-LDH eluted from a calibrated TSK-400 column with a retention time of 25.3 min, corresponding to that of a 76-kDa protein. This apparent molecular weight may indicate adventitious retention of the native tetramer (m_r = 144 000) on the column, or that the tetramers have dissociated into active dimers upon exposure to 0.3 M KCl under the chromatographic conditions. Dissociation of tetramers of pig heart LDH at high NaCl concentrations has been reported (Yamamoto & Nakano, 1982). We hereafter refer to the 76-kDa protein species as "oligomers" of LDH.

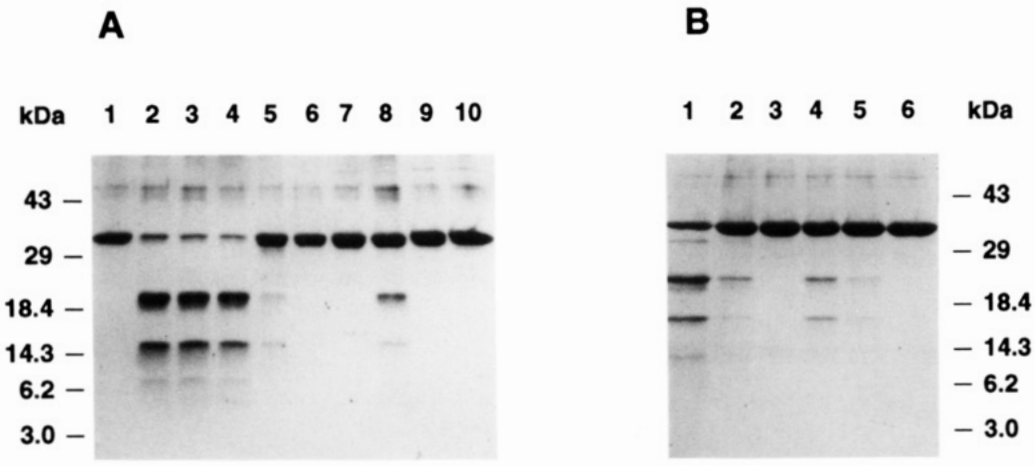

FIGURE 1: Proteolysis of LDHs by HIV-1 protease. Protein samples were prepared and analyzed by SDS–PAGE (15% polyacrylamide; staining by Coomassie Brilliant Blue) as described in Experimental Procedures. Relative molecular weights using protein standards are indicated. (A) Samples in AMT-EDN buffer (pH 5.0) contained 14 μM LDH and were incubated with 140 nM HIV-1 protease at variable times, unless noted otherwise: lane 1, RM-LDH, no protease, 240 min; lanes 2–4, RM-LDH at 60, 120, and 240 min; lane 5, heat-denatured RM-LDH at 240 min; lane 6, 240 min, containing 10 μM Ala-Ala-PheΨ[CHOHCH$_2$]Gly-Val-Val-OMe; lane 7, pig muscle LDH, no protease, 240 min; lane 8, PM-LDH, protease, 240 min; lane 9, chicken muscle LDH, no protease, 240 min; lane 10, CM-LDH, protease, 240 min. (B) Limited proteolysis of RM-LDH in the presence of NADH and oxamate. Samples in AMT-EDN buffer (variable pH) contained 14 μM RM-LDH; 700 nM HIV-1 protease samples in lanes 2 and 5 contained 10 mM oxamate and 1.0 mM NADH and those in lanes 3 and 6 contained 10 μM Ala-Ala-PheΨ[CHOHCH$_2$]Gly-Val-Val-OMe: lanes 1–3, pH 6.0, 120-min incubation; lanes 4–6, pH 7.0, 120-min incubation.

First, the pH stability of rabbit muscle LDH was assessed by preincubation of the enzyme at pH 5.0–7.0 in the presence of inhibited HIV-1 protease (containing 10 μM Ala-Ala-PheΨ[CHOHCH$_2$]Gly-Val-Val-OMe) followed by measurement of the residual enzymatic activity and remaining LDH oligomers at pH 7.0. As is shown in the odd-numbered lanes in Figure 2A, the presence of the inhibitor is sufficient to ablate all proteolysis of RM-LDH, indicating that these samples suffice as proteolysis-free control samples. In the absence of protease activity, the irreversible loss of enzymatic activity and quaternary structure following preincubation at slightly acidic pH is depicted in Figure 2B. At $t \leq 120$ min, RM-LDH is completely stable in terms of both oligomeric structure and enzymatic activity at pH 7.0, but both of these properties are lost progressively at decreasing solution pH. At pH 6.0 and 5.0, a commensurate loss in enzymatic activity and quaternary structure occurs to final values at 120 min of approximately 85% and 35%, respectively. In the gel filtration chromatograms in which the 76-kDa peak was diminished, the commensurate increase of a peak was observed at 30–32 min, corresponding to a molecular weight of $\leq$10 000. This probably represents a denatured, monomeric form of the enzyme, which is retained by the gel filtration column to yield a spuriously low apparent molecular weight. These results indicated that deoligomerization accompanies, or causes, inactivation of RM-LDH, such that the instability of RM-LDH at acidic pH results in part from a breakdown in quaternary structure.

As indicated in Figure 2C, proteolysis of RM-LDH is augmented at acidic pH. This finding is unsurprising given that HIV-1 protease is optimally active at pH 5–6 (Hyland et al., 1991). However, in addition to this, the inherent instability of RM-LDH at acidic pH undoubtedly contributes to its susceptibility to proteolysis. At neutral pH, at which the protein substrate is stable, minimal cleavage of the 36-kDa protein band was observed at 560 nM HIV-1 protease, and correspondingly, little loss of enzymatic activity or quaternary structure was evident. At pH 6.0 and 5.0, the rates of loss of both enzymatic activity and RM-LDH oligomer are essentially equal. At either pH, the rates of inactivation (or deoligomerization) of RM-LDH in the presence of protease (Figure 2C) exceeded both the rates of inactivation/deoligomerization in its absence (Figure 2B) and, apparently, the rates of proteolysis of the 36-kDa monomer (Figure 2C). At pH 5.0 and 6.0, 50% of enzymatic activity is lost at a point at which only 10% and 34%, respectively, of the subunits are proteolyzed. These results suggested that the "nicking" of one or more of the 36-kDa subunits contained within an oligomer enhances the inherent instability of the oligomer at acidic pH, resulting in oligomer dissociation and enzyme inactivation.

However, when one subtacts from the curves in Figure 2C the loss of enzymatic activity due to its instability at the acidic pH (curves from Figure 2B), the resulting corrected time courses of the loss of LDH activity due to proteolysis more closely matched the time courses of the diminution of the 36-kDa monomer. From this, we may conclude that a single proteolytic cleavage event is sufficient to inactivate one subunit of LDH, such that the protease-catalyzed loss of LDH enzymatic activity should suffice as a measure of enzyme cleavage.

As described above, the heat-denatured RM-LDH and the stable ternary RM-LDH–NADH–oxamate complex are both more resistant to proteolysis than is the acid-treated form of the native enzyme. This suggests that neither the native, bound enzyme nor the fully denatured form of RM-LDH is the true substrate of HIV-1 protease. From this we propose that weakening or disruption of the quaternary structure of RM-LDH at acidic pH elaborates the exposure of the cleavage site to solvent, rendering it accessible to the protease.

Identity of Cleaved Peptide Sequence in Rabbit Muscle LDH. Amino-terminal microsequencing of the cleavage products of rabbit muscle LDH was greatly facilitated by the fact that the N-termini of the intact LDH monomers are apparently blocked. [The amino terminus of pig muscle LDH is acetylated (Kiltz et al., 1977)]. Edman sequencing of the protease-treated rabbit muscle LDH resulted in the following sequence: Gly-Glu-His-Gly-Asp-Ser-Ser-Val-Pro, which is identical to that found in pig muscle LDH at amino acids 191–199 (Kiltz et al., 1977) and in the translated nucleotide sequence of the rabbit muscle enzyme [occurring at amino acids 191–199 (Sass et al., 1989)]. From the known amino acid sequences of both rabbit and pig muscle LDHs, the apparent cleavage site occurs at a common amino acid

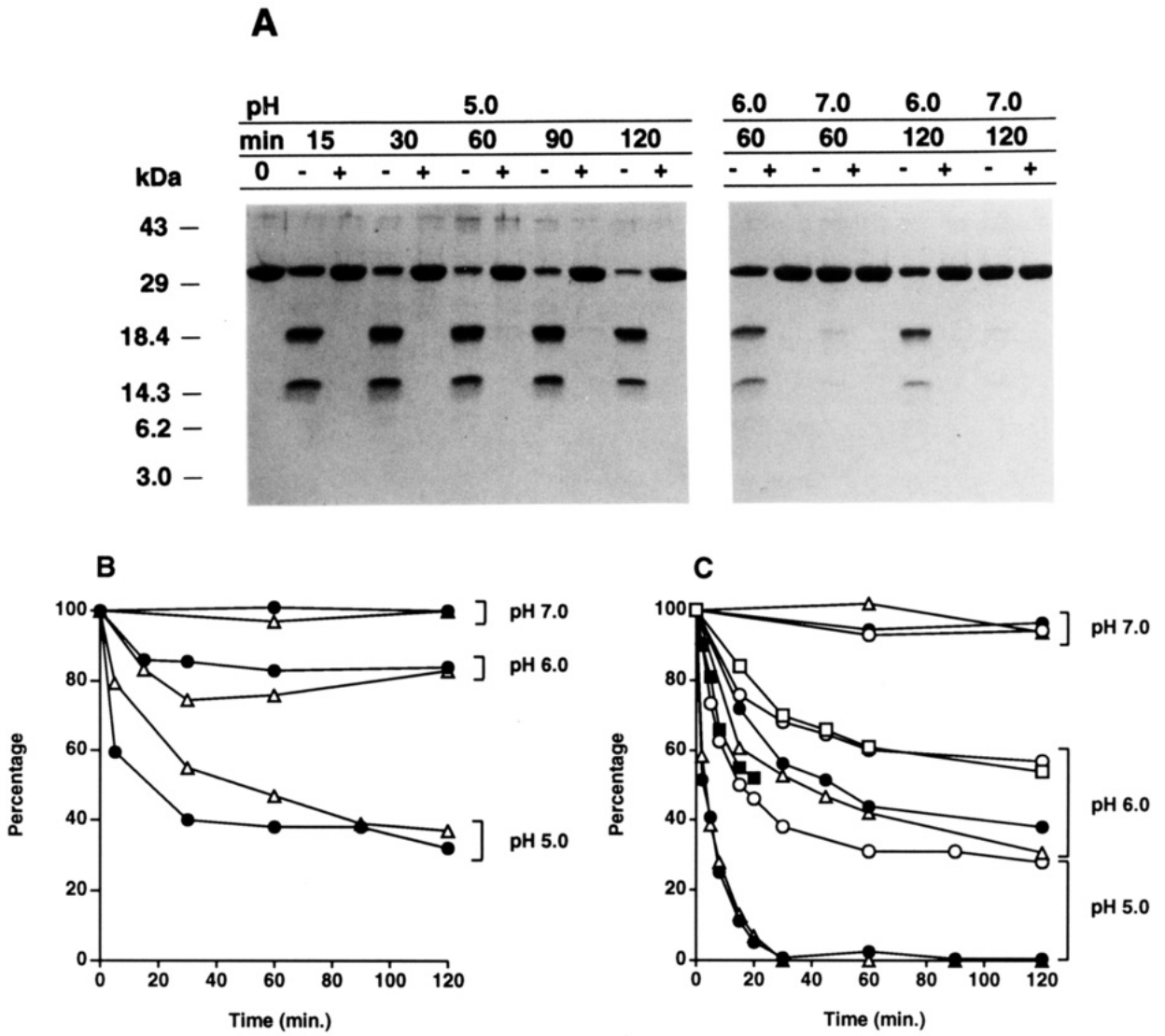

FIGURE 2: Effects of proteolysis of RM-LDH on enzymatic activity, primary structure, and quaternary structure. (A) SDS–PAGE of samples containing 14 μM RM-LDH (AMT-EDN buffer, pH 5.0–7.0), incubated with 140 nM HIV-1 protease for the indicated times and solution pH. Lanes labeled with 0, –, and + contained, respectively, no protease, no inhibitor, and 10 μM Ala-Ala-PheΨ[CHOHCH$_2$]Gly-Val-Val-OMe. (B) Stability of untreated RM-LDH at pH 5.0–7.0. Time course of percentage losses of RM-LDH enzymatic activity (closed circles) and oligomer concentration (open triangles; as determined from analytical gel filtration) at pH 5.0–7.0 for the samples shown in (A) (and additional samples) which contained 10 μM Ala-Ala-PheΨ[CHOHCH$_2$]Gly-Val-Val-OMe. (C) Time courses of protease-catalyzed percentage losses of RM-LDH enzymatic activity (closed circles), oligomer concentration (triangles), and 36-kDa subunits (open circles) at pH 5.0–7.0. Enzymatic activities and oligomer concentration were determined at pH 7.0 following preincubation at the given pH values. Samples were the same as those shown in (A) without 10 μM Ala-Ala-PheΨ[CHOHCH$_2$]Gly-Val-Val-OMe), and percentages of diminution of the 36-kDa subunits were determined by densitometric scanning of these (and other) SDS polyacrylamide gels as described in Experimental Procedures. Subtraction of the curves of residual RM-LDH activity in (B) from the corresponding curves in (C) (following proteolysis) resulted in the time courses in (C) at pH 5.0 (filled squares) and 6.0 (open squares), which are corrected for loss of enzyme activity not due to proteolysis.

sequence: His-Gly-Trp-Ile-Leu*Gly-Glu-His-Gly-Asp. Cleavage at the Leu-190*Gly-191 site of the pig and rabbit enzymes (both subunits consist of 332 residues) would result in protein fragments of 20.9 and 15.4 kDa. This is consistent with the relative molecular weights of the predominant proteolytic fragments observed upon treatment with HIV-1 protease: 21 and 15 kDa.

This unique proteolytic cleavage site in the LDHs of pig and rabbit muscle comprises a strongly conserved sequence in the mammalian LDHs and contains the active site residue His-193 (Table II). In the three-dimensional structure of both dogfish muscle (White et al., 1976) and pig heart LDHs (Grau et al., 1981), this cleavage site is contained in the sequence His-Gly-Trp-Ile-Leu-Gly-Glu-His-Gly-Asp found within the βG–βH structural element, which contains the catalytic residue His-195 (Figure 3). The sequence of this decapeptide is invariant in the heart and muscle LDHs of both pig and rabbit. In contrast, the chicken muscle LDH is not cleaved by HIV-1 protease. This may be due to the presence of Val at the P1 position of the decapeptide loop (amino acid 190), in that HIV-1 protease does not cleave peptides or proteins that contain a β-branched amino acid residue at its P1 position (Loeb et al., 1989; Phylip et al., 1990; Richards et al., 1990).

Kinetic Comparison of the LDH Proteolytic Cleavage Site in the Protein with the Synthetic Oligopeptide. The above results indicate that upon treatment of rabbit muscle LDH at pH 6.0 with HIV-1 protease the cleavage of the Leu-190*Gly-191 bond attends loss of enzymatic activity as measured at pH 7.0. The availability of a facile enzymatic assay therefore allows kinetic measurement of the rate of proteolysis of this dipeptide bond at variable concentrations of LDH, in that product formation is equivalent to a decrease in enzymatic activity. Under identical experimental conditions

Table II. Sequence Comparison of HIV-1 Protease Cleavage Site Found within βG Strand Element of Heart and Muscle LDHs[a]

LDH	scissile peptide at βG–βH region: βG sheet	βH sheet	residue nos.	proteolytic fragment sizes
rabbit muscle	**HALSC**	**HGWIL*GE*H*GD** SSVPV	aa 181–199	21-kDa, 15-kDa
rabbit heart	**HPSSC**	**HGWIL*GE*H*GD** SRLAV	aa 181–199	21-kDa, 15-kDa
pig muscle	**HPLSC**	**HGWIL*GE*H*GD** SSVPV	aa 180–199	21-kDa, 15-kDa
pig heart	**HPSSC**	**HGWIL*GE*H*GD** SSVAV	aa 183–202	21-kDa, 15-kDa
chicken muscle	**HPLSC**	**HGWIV*GQ*H*GD** SSVPV	aa 180–199	none
chicken heart	**HPTSC**	**HGWIL*GE*H*GD** SSVAV	aa 180–199	none
dogfish muscle	**HSCSH**	**HGWVI*GE*H*GD** SVPSV	aa 183–202	not determined

[a] Amino acid sequences taken from Sass et al. (1989), Kiltz et al. (1977), Torff et al. (1977), and Holbrook et al. (1975). The italic letter is the active site histidyl residue. Invariant residues are shown in bold lettering.

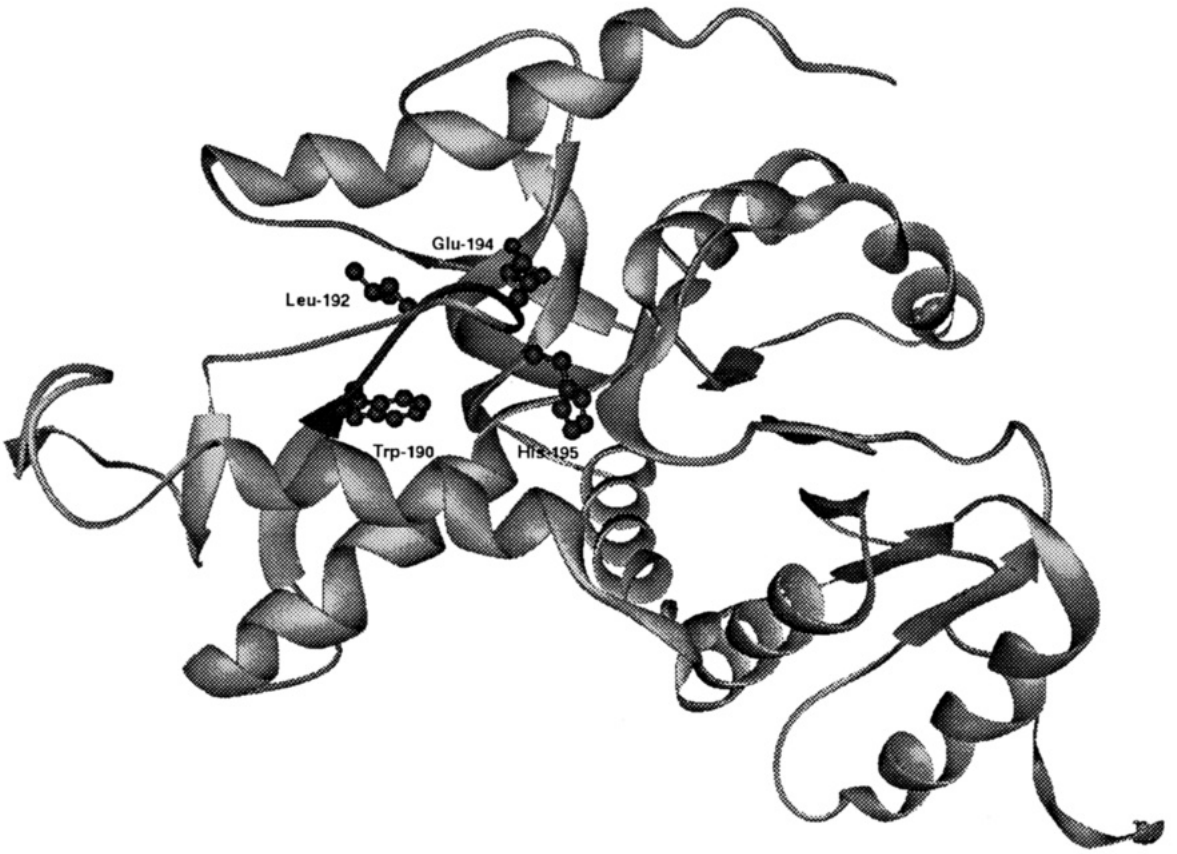

FIGURE 3: Three-dimensional structure of scissile peptide within the βG-βH element of pig heart LDH complexed with (*S*)-lactate–NAD (inhibitor not shown) (Grau et al., 1981).

Table III: Substrates and Inhibitors Based on LDH Cleavage Site[a]

substrate or inhibitor	K_m (μM)	V_{max}/E_t (s^{-1})	V_{max}/K_mE_t (mM^{-1} s^{-1})	K_i (nM)
rabbit muscle LDH	31 ± 5	0.42 ± 0.04	14 ± 2	
Ac-His-Gly-Trp-Ile-Leu-Gly-Glu-His-Gly-Asp-NH_2	50 ± 8	0.70 ± 0.05	14 ± 2	
Ac-His-Gly-Trp-Ile-Leu-Ala-Glu-His-Gly-Asp-NH_2	17.7 ± 0.9	5.1 ± 0.07	290 ± 15	
Ac-His-Gly-Trp-Ile-Leu-Gly-Ala-His-Gly-Asp-NH_2		not cleaved		
Ac-Trp-Ile-Leu-Gly-Glu-His-NH_2	5000 ± 2000	7 ± 2	1 ± 1	
Ac-His-Gly-Trp-Ile-LeuΨ[$CHOHCH_2$]Gly-Glu-His-Gly-Asp-NH_2[b]				5.2 ± 0.8

[a] All measurements made at 37 °C in AMT-EDN buffer (pH 5.0) as described in Experimental Procedures, except where noted. [b] Determined in MENDT buffer (pH 6.0).

the kinetic parameters of proteolysis of rabbit muscle LDH at variable concentrations of the enzyme (14–56 μM) were directly compared with those of the peptidolysis of the corresponding decapeptide substrate Ac-His-Gly-Trp-Ile-Gly-Glu-His-Gly-Asp-NH_2 from double-reciprocal plots of $1/v$ vs $1/S$ (Table III). The Michaelis constants (K_m), turnover numbers (V_{max}/E_t), and specificity constants (V_{max}/K_mE_t) of the protein and peptide substrates were remarkably similar. (Caveat: It is assumed here that the measured concentration of the protein substrate accurately reflects the true concentration of the form of the protein that is the actual HIV-1 protease substrate. Given that RM-LDH is nearly completely inactive under the assay conditions used, this assumption is reasonable.) These results indicated that the scissile portion of the βG–βH sheets of LDH and the corresponding decapeptide interact with HIV-1 protease by physically similar processes; that is, they have similar kinetic rates of binding, desorption, and catalysis. Such agreement suggests that either the decapeptide portion of LDH is conformationally mobile under conditions at which it is proteolytically labile or the

synthetic decapeptide possesses discrete secondary structure in solution which resembles that of its corresponding sequence in LDH. The finding that the substrate activity of RM-LDH is optimized under conditions at which it is structurally unstable favors the former explanation. This result is the first known demonstration of comparable kinetic parameters of a synthetic peptide substrate with that of an identical peptide sequence within a protein substrate of HIV-1 protease, such that in this instance, kinetic results from this and similar decapeptide substrates should accurately reflect the proteolytic processing of a protein substrate.

The V_{max}/K_mE_t value of AcWILGEH-NH_2 is 0.07 times that of both the protein and the decapeptide substrate. Since the active site of the protease should be able to accommodate all but the terminal residues of the decapeptide, the more efficient peptidolysis of the decapeptide compared to the hexapeptide substrate Ac-Trp-Ile-Leu-Gly-Glu-His-NH_2 suggests that the four additional residues of the decapeptide (P5, P4, P4′, and P5′), which are either outside of the protease substrate binding cleft or nearly so, may in fact be essential to secondary structure in solution or, more likely, may interact favorably with residues outside of the active site. Moreover, the equivalence of the kinetic parameters for the RM-LDH substrate and the decapeptide, but not the hexapeptide Ac-WILGEH-NH_2, demonstrates that the P5, P4, P4′, and P5′ residues of the decapeptide within the protein substrate make important contributions to recognition and binding by the protease. These putative additional interactions may account for the lower K_m value for the decapeptide substrate. On the other hand, the 10-fold higher value of k_{cat} for Ac-WILGEH-NH_2 is likely to result from a faster desorption rate of the peptidolysis products of this substrate compared to those of the decapeptide, which may be more slowly released due to interactions of their additional residues with the enzyme.

A decapeptide in which Gly at P1′ was replaced by Ala (Ac-HGWILAEHGD-NH_2) is among the best substrates of HIV-1 protease yet reported (K_m = 18 μM, V_{max}/E_t = 5 s^{-1}, V_{max}/K_mE_t = 290 000 M^{-1} s^{-1}). As the presence of P5, P4, P5′, and P4′ residues improves V_{max}/K_mE_t for a hexapeptide substrate by 14-fold, then the addition of even a small side chain at P1′ raises this value by an additional 21-fold. This improvement in V_{max}/K_mE_t is not reflected in the Michaelis constant and probably results from an increase in the rate of a catalytic step which would be reflected in both V_{max}/K_mE_t and V_{max}/E_t. The 10-fold increase in the latter value suggests that occupation of the S1′ site results in a binding conformation of the peptide which is more amenable to catalysis than when Gly is present and is therefore probably not a result of higher rates of desorption of its peptidolytic products. Occupation of the S1′ site probably allows more favorable positioning of the scissile amide bond, which is not reflected in the K_m value. This result is reminiscent of a variety of oligopeptide substrates of porcine pepsin, which possess similar Michaelis constants but widely disparate values of k_{cat} (Fruton, 1976). The substrate Ac-HGWILAEHGD-NH_2 is very similar to the proteolytic processing site KARVL*AEAM found in HIV-1 $Pr55^{gag}$, which has given rise to some of the best oligopeptide substrates yet reported for HIV-1 protease (Phylip et al., 1990). Replacement of the P2′ Glu of Ac-HGWILGEHGD-NH_2 with Ala resulted in a peptide which was not cleaved by HIV-1 protease.

The pH dependence of V_{max}/K_mE_t and V_{max}/E_t for Ac-HGWILAEHGD-NH_2 was also investigated (data not shown). The plot of log (V_{max}/K_mE_t) vs pH was characterized by a "bell-shaped" curve which displayed an apparent slope of 1 on the acidic side and a slope of −2 on the basic side. This plot differs from those of other oligopeptide substrates of HIV-1 protease at basic pH (Hyland et al., 1991) in that apparently two groups, rather than one, on either the enzyme or the substrate, must be protonated for this substrate to bind (Cleland, 1977). The plot of V_{max}/E_t vs pH was a wave-shaped curve in which the protonation of a single group of pK = 4.0 resulted in a lower value of V_{max}/E_t. This plot was similar to those of the other substrates. Fitting of the data for the log (V_{max}/K_mE_t) vs pH plot to log (V_{max}/K_mE_t) = log $(c/(1 + [H])/K_1 + K_2/[H] + K_3/[H]^2)$, in which c is the pH-independent value of V_{max}/K_mE_t and K_1, K_2, and K_3 are acid dissociation constants for the protonatable groups, resulted in values of pK_1 = 3.9, pK_2 = 5.4, and pK_3 = 6.3. While pK_1 and pK_3 are similar in value to the determined pK values for the active site aspartyl groups of the protease, using substrates such as Ac-Arg-Ala-Ser-Gln-Asn-Tyr-Pro-Val-Val-NH_2 (Hyland et al., 1991), pK_2 probably represents an acidic group on a AcHGWILAEHGD-NH_2 (either the Glu or Asp) residue or in the substrate binding site of the enzyme (such as Asp-29 or Asp-30) which must be protonated for the substrate to bind.

Structural Analysis of HIV-1 Protease Complexed to a Peptide Analogue of the LDH Proteolytic Cleavage Site. A synthetic peptide analogue of the LDH decapeptide was prepared which contained the inhibitory hydroxyethylene isosteric replacement for the scissile Leu-Gly dipeptide. As reported in Table III, the substrate analogue is a potent inhibitor of HIV-1 protease (apparent K_i = 5 nM). Interestingly, the inhibition constant of this peptide analogue is quite similar to that found for the protease by the renin inhibitor, H-261 [Boc-His-Pro-Phe-His-LeuΨ[$CHOHCH_2$]-Val-Ile-His; K_i = 15 nM at pH 5.0; Richards et al. (1989)]. H-261 is identical in sequence to the "LDH-derived" hydroxyethylene isostere at the P5, P1, and P3′ positions and similar at P3.

The space group of the crystal of the HIV-1 protease–Ac-His-Gly-Trp-Ile-LeuΨ[$CHOHCH_2$]Gly-Glu-His-Gly-Asp-NH_2 complex, $P2_12_12$, is identical to that reported for the acetyl-pepstatin complex of HIV-1 protease (Fitzgerald et al., 1990). Similar to the crystallographic data of other HIV-1 protease–inhibitor complexes (Miller et al., 1989; Erickson et al., 1990; Fitzgerald et al., 1990; Swain et al., 1990), and those structures which contain hydroxyethylene-bearing peptide analogues (Jaskolski et al., 1991; Dreyer et al., 1992), the decapeptide analogue inhibitor Ac-His-Gly-Trp-Ile-LeuΨ-[$CHOHCH_2$]Gly-Glu-His-Gly-Asp-NH_2 binds to the enzyme in an extended 31-Å chain which is normal to the pseudo-C-2 axis that relates the HIV-1 protease homodimer. The electron density which corresponds to the acetyl- and carboxamide-terminal groups are poorly resolved and are therefore omitted from the structural model. The inhibitory secondary hydroxyl group is centrally interposed within 3 Å of the four oxygen atoms of the active site aspartyl residues, Asp-25 and Asp-125. As with other inhibitors, the substrate binding "flaps", one contributed by each monomer, close down to bind around part of the inhibitor, and as expected, only the P3–P3′ residues (Trp-203–His-206) of the inhibitor appear to be fully bound within the enzyme active site (Figure 4). The structure of this inhibitor within the protease is remarkably similar to that of other inhibitors, having an extended β-conformation as described by Gustchina and Weber (1990). Fitzgerald et al. (1990) noted that the largest deviation from symmetry in the protein occurred in the "flap" region. We observed a similar distortion although our model does not position the carbonyl

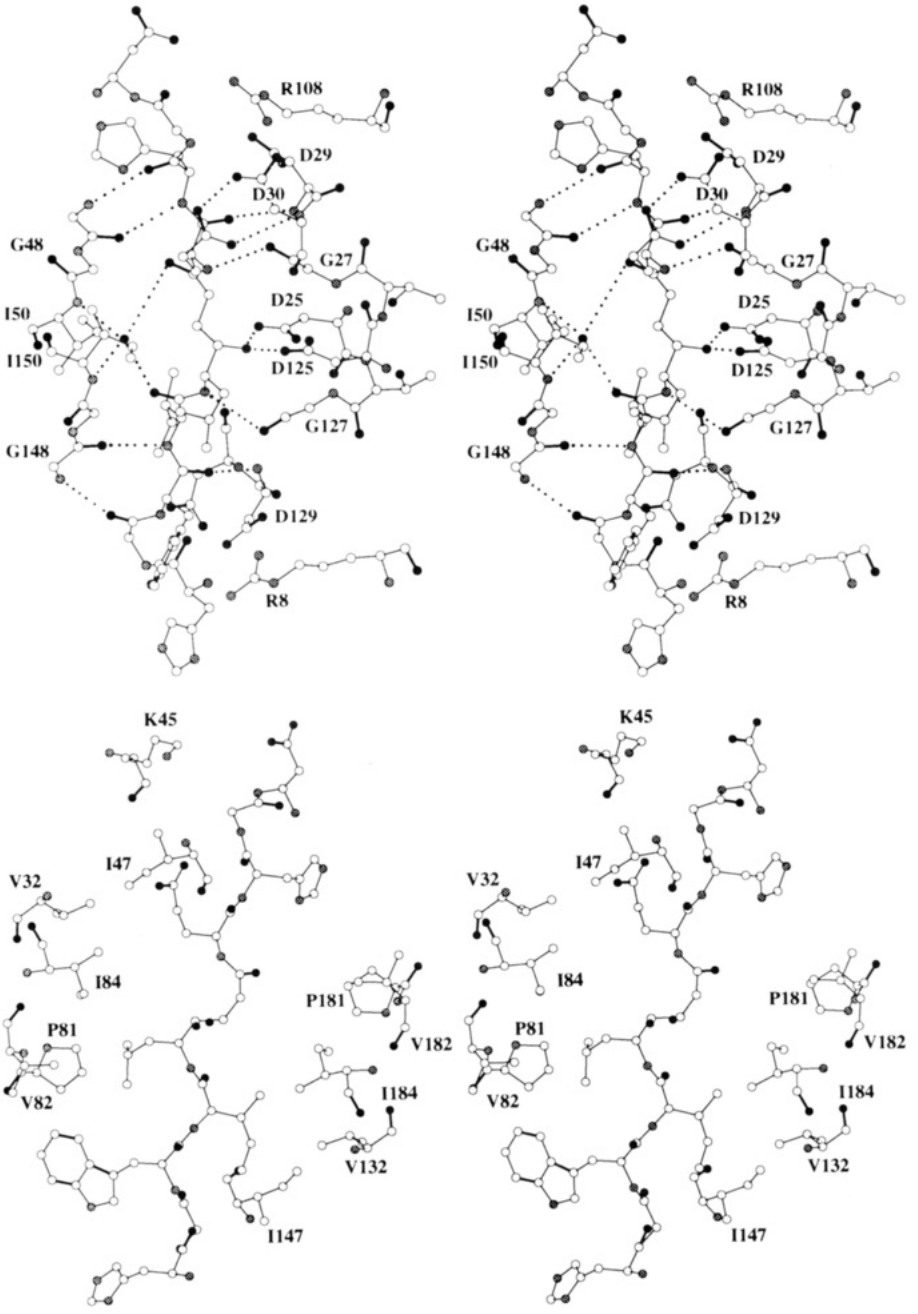

FIGURE 4: Stereo diagrams of hydrogen bonding and close contacts of the peptide analogue inhibitor Ac-His-Gly-Trp-Ile-LeuΨ-[CHOHCH$_2$]Gly-Glu-His-Gly-Asp-NH$_2$ bound to HIV-1 protease which differ by an approximate 90-deg rotation about the vertical axis of the inhibitor. (Top) Potential hydrogen-bonding interactions with backbone amides and side chains; (bottom) extended conformation of the inhibitor displaying protease residues which comprise the S2, S1, S1′, and S2′ binding pockets.

group of Ile-50 to accept a hydrogen bond from the amide of Gly-151. The residues that have the largest deviations in 2-fold symmetry in this structure are analogous to those reported by Fitzgerald et al. (1990). Interestingly, these are the same residues that are involved in lattice contacts within the crystal. Consequently, it is not possible to attribute the deviation in symmetry to the binding of the inhibitor.

A summary of the close contacts made between HIV-1 protease and Ac-His-Gly-Trp-Ile-LeuΨ[CHOHCH$_2$]Gly-Glu-His-Gly-Asp-NH$_2$ can be seen in part in Figure 4. As with all HIV-1 protease–inhibitor complexes, the contacts made between residues in one monomer to the inhibitor on one side of the scissile bond are complemented by contacts between the identical residues of the other monomer and the other side of the scissile bond. The putative hydrogen bonds formed between the backbone amide groups of the protease and the inhibitor (for example, that between the carbonyl of Gly-27 and the amide of Glu-207 of the inhibitor) and between the "structured" water molecule (Miller et al., 1989) and the carbonyl oxygens of the P2 and P1′ groups of the inhibitor are conserved in all of the known HIV-1 protease–inhibitor complexes. Similarly, the P3–P3′ side chains of Ac-His-Gly-Trp-Ile-LeuΨ[CHOHCH$_2$]Gly-Glu-His-Gly-Asp-NH$_2$ occupy conserved subsites in the protease. As seen in other crystal structures, Pro-81, Val-82, and Ile-84 form the S1 subsite and are within 4 Å of atoms of the Leu-205 side chain. Interestingly, Trp-203 is also proximal to the S1 site and may also interact with these hydrophobic residues (in particular, Pro-81).

The P2 residue, Ile-204, as seen in other protease–inhibitor complexes, forms hydrogen bonds at its respective termini with the carbonyl oxygen of Gly-148 and the structured water

molecule and makes side-chain contacts with the residues which comprise the S2 site: Val-132, Ile-147, and Ile-184. As expected, the complementary residues on the other monomer provide the binding site for the P2′ residue, Glu-207, namely, Val-32, Ile-47, and Ile-84. It is interesting to compare in a single protease–inhibitor complex how the enzyme accommodates the Ile and Glu residues which occupy similar S2 and S2′ subsites. The CD1 carbon of Ile-204 and the OE1 and OE2 oxygens of Glu-207 appear to be within similar distances to the constellation of hydrophobic residues which comprise the S2 binding site. CD1 of Ile-204 and OE1 of Glu-207 are positioned nearly equidistant from CD1 of Ile-(1)47 and CG2 of Val-(1)32. Remarkably, the OE1 oxygen of Glu-207 is nearly as close to the OD1 oxygen of Asp-30 as CD1 of Ile-204 is to the corresponding oxygen of Asp-130. The comparison of the S2 and S2′ binding interactions with the inhibitor suggests that the potentially anionic side chain of Glu-207 is comfortably accommodated in a cluster of (presumably) anionic residues, namely, Asp-30 and Asp-29, at least as favorably as that of an uncharged, hydrophobic side chain.

The observation that the decapeptide substrate Ac-His-Gly-Trp-Ile-Leu-Gly-Glu-His-Gly-Asp-NH_2 has a substantially lower Michaelis constant and a higher value of V_{max}/K_mE_t than the hexapeptide substrate Ac-Trp-Ile-Leu-Gly-Glu-His-NH_2 suggests that there are important protease–substrate interactions in the P5, P4, P4′, and P5′ residues which reside partly on the outside of the substrate binding site. Upon examination of the HIV-1 protease–Ac-His-Gly-Trp-Ile-LeuΨ[$CHOHCH_2$]Gly-Glu-His-Gly-Asp-NH_2 structure, we note that there are no clear close contacts between the protein and these residues apart from the potential ionic interactions between the β-carboxylate of Asp-210 and the cationic side chains of Arg-108 (OD2 of Asp-210 is within 6.2 Å of NH1 of Arg-108) and Lys-45. A salt bridge formed between this carboxyl group and a cationic group on the enzyme could provide some of the binding energy necessary to influence these kinetic parameters and may also account in part for the lower value of V_{max}/E_t in the decapeptide, possibly due to a rate-limiting release of the carboxyl-terminal product resulting from this additional binding interaction. Similarly, N6 of His-201 is within 4 Å on NH1 of Arg-8. If one of these residues is uncharged, hydrogen-bonding interactions may exist.

Interestingly, the contacts between Arg-108 and Asp-29 in the present structure differ slightly from those observed in other crystal structures of HIV-1 protease–peptide inhibitor complexes. In structures reported by both Fitzgerald et al. (1990) and Miller et al. (1989), OD1 and OD2 of Asp-29 are in close contact (<3.5 Å) with the NE and NH1 nitrogens of Arg-108. Such "strong" interactions are not observed in our structure (Figures 4 and 7), possibly due in part to additional electrostatic influence on Arg-108 from the β-carboxylic group of the proximal Asp-210.

Proteolytic Processing of Recombinant Galactokinase Containing Genetically Inserted HIV-1 Protease Peptide Cleavage Sites. Previously, Debouck et al. (1987) demonstrated authentic proteolytic processing of a recombinant form of $Pr55^{gag}$ upon its coexpression with HIV-1 protease in *E. coli*. This protein substrate/HIV-1 protease coexpression system constitutes a useful bacterial model for viral polyprotein processing. In order to further investigate the recognition by HIV-1 protease of cleavage sequences within protein substrates, we chose to engineer individual cleavage sites within a heterologous protein, the *E. coli* galactokinase gene product (galK) (Debouck et al., 1985).

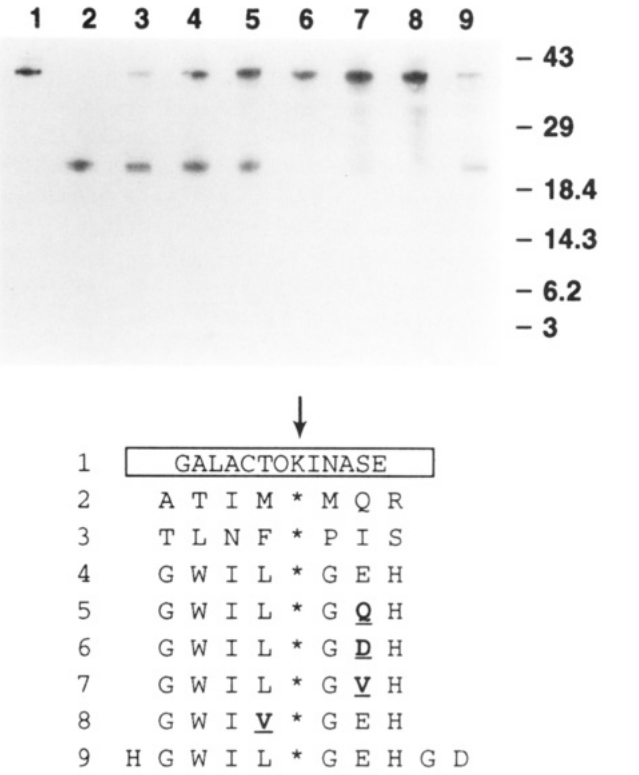

FIGURE 5: Processing of galactokinase mutants containing cleavage site insertions upon coexpression with HIV-1 protease in *E. coli*. Following induction of the double-plasmid vector for 2 h, bacterial crude extracts were prepared, subjected to SDS–PAGE, and immunoblotted with a rabbit polyclonal antibody. Wild-type galK (lane 1), and galK mutants containing the following oligopeptide insertions (cleavage site denoted by asterisk): ATIM*MQR (lane 2), TLNF*PIS (lane 3), GWIL*GEH (lane 4), GWIL*GQH (lane 5), GWIL*GDH (lane 6), GWIL*GVH (lane 7), GWIV*GEH (lane 8), and HGWIL*GEHGD (lane 9).

Galactokinase engineered with all HIV-1 $Pr^{gag-pol}$ cleavage sites was coexpressed with active HIV-1 protease using our double-plasmid expression system (Debouck, 1992). The processing of galactokinase was monitored by immunoblot analysis using a rabbit polyclonal antibody raised against galactokinase. Wild-type bacterial galactokinase was not processed by HIV-1 protease upon coexpression (Figure 5, lane 1). However, in all cases, galK mutants containing HIV-1 $Pr^{gag-pol}$ cleavage sites were efficiently cleaved upon induction of gene expression from the plasmids, and the resulting 20-kDa protein fragments were consistent with cleavage occurring at the inserted peptide (Debouck, 1992). By examining the extent of processing after short periods of induction and calculating the ratio of processed vs unprocessed galactokinase, we were able to rank the efficiency of cleavage by the HIV-1 protease at the various sites found in $Pr^{gag-pol}$ (Debouck, 1992). For example, the ATIM*MQR site found in HIV-1 $Pr55^{gag}$ was cleaved more efficiently in this system than the TLNF*PIS site found in $Pr160^{gag-pol}$ (Figure 5, lanes 2 and 3).

In an effort to better understand the activity of HIV-1 protease toward the LDH substrate, we inserted the cleavage site in the form of GWIL*GEH and HGWIL*GEHGD, and mutated versions of the former into galK. The mutations included Gln, Asp, and Val at position P2′ and Val at position P1. Upon coexpression of these constructs with HIV-1 protease, we observed that the wild-type LDH site GWIL*GEH was cleaved by the protease, as evidenced by the appearance of the smaller 20-kDa galK fragment (Figure 5, lane 4). The extent of processing was similar to that resulting from insertion of the TLNF*PIS HIV-1 polyprotein site

(compare lanes 3 and 4 in Figure 5). In kind, insertion of the decapeptide HGWIL*GEHGD into galK resulted in a substrate (Figure 5, lane 9). With respect to the mutated forms of the LDH site, only Gln in P2′ resulted in an equivalent amount of processing by the HIV-1 protease, whereas Asp and Val at this position or Val at the P1 position was not processed by the protease (Figure 5, lanes 5–8), even at longer times of induction (data not shown).

These results demonstrate that recognition of the heptapeptide GWILGEH by HIV-1 protease is not unique to the structure it assumes within the lactate dehydrogenases, such that this protein structural element is in itself likely to be sufficient to effect a conformation which may bind to the protease. It is important to bear in mind that while the three-dimensional structure of galactokinase has not yet been elucidated, the structure of the heptapeptide within this protein becomes irrelevant if the processing within the bacterial coexpression system occurs in an unfolded or partially folded form of the protein during translation. Nevertheless, if the specific proteolysis of the GWIL*GEH-galK, but not the wild-type galactokinase, occurred with a partially folded form of the enzyme, one may argue that this observation is analogous to that of the partially denatured form of RM-LDH.

Variation of some of the residues within the scissile heptapeptide of GWILGEH-galK also supported the specificity of HIV-1 protease for the LDHs as well as for peptide substrates. The inactivity of GWIV*GEH-galK is consistent with the known inability of the protease to process substrates which contain β-branched amino acids at the P1 position (Loeb et al., 1989; Phylip et al., 1990; Pettit et al., 1991). In kind, coexpressed HIV-1 protease failed to process the double mutant which contained respective substitutions of P1 Leu and P2′ Glu with Val and Gln (GWIV*GQH-galK) (data not shown). Interestingly, this heptapeptide sequence is identical to that of chicken muscle LDH, which also was not a substrate for the protease. These findings further demonstrate the fidelity of the protease for this heptapeptide sequence in two unrelated proteins.

The HIV-1 protease/GWILGEH-galK coexpression vector proved to be a useful probe of the amino acid requirements at the P2′ position of the heptapeptide fragment. Thus, while Glu and Gln are tolerated in the P2′ position, Val and Asp are not. Glu and Gln are the most commonly found P2′ residues in viral and nonviral protein substrates, while Asp never occurs at P2′ (Pettit et al., 1991; Poorman et al., 1991). Val occurs frequently at the P2′ position of peptide and protein substrates of HIV-1 protease such that its effect here may seem uncharacteristic (Moore et al., 1989; Phylip et al., 1990; Poorman et al., 1991).

DISCUSSION

Structural Aspects of the LDH Substrate. The mammalian lactate dehydrogenases from heart and muscle tissues are tetrameric enzymes of highly conserved three-dimensional structure despite significant differences in amino acid sequences between these isozymic forms (Holbrook et al., 1975; Eventoff et al., 1977). The active sites of the LDHs contain a histidyl residue (His-195 in dogfish muscle and pig heart LDHs; His-193 in pig muscle and rabbit muscle LDHs) found at a dipeptide loop which connects the βG and βH sheets (Holbrook et al., 1975; White et al., 1976; Grau et al., 1981). The primary structure of the antiparallel βG–βH loop is highly conserved among the various LDHs (see Table II), and its three-dimensional structure appears to be similar in both the pig heart and dogfish muscle enzymes in both the apo and

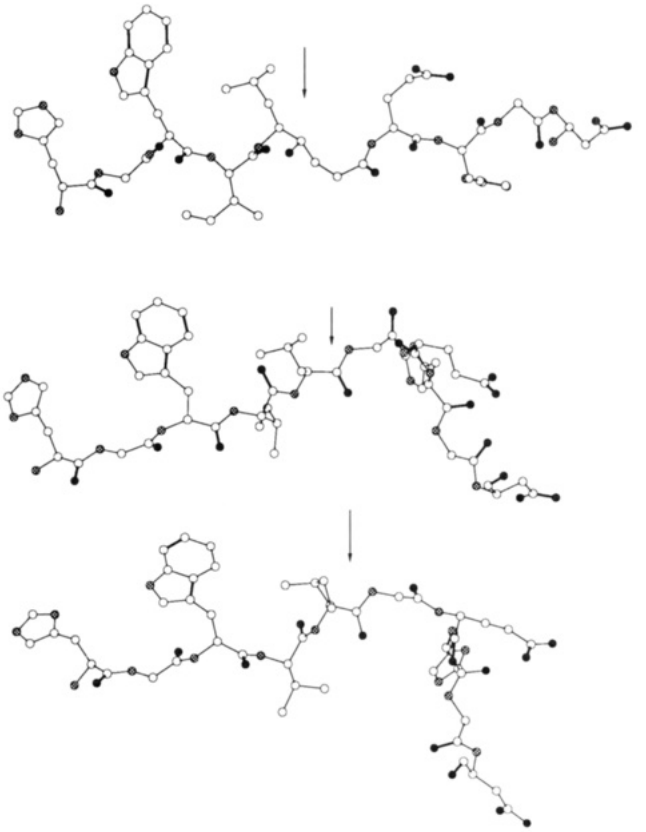

FIGURE 6: Structural comparison of decapeptide cleavage site within the HIV-1 protease-bound Ac-His-Gly-Trp-Ile-LeuΨ-[CHOHCH$_2$]Gly-Glu-His-Gly-Asp-NH$_2$ (top), inhibitor-complexed pig heart LDH [middle; Grau et al. (1981)], and dogfish muscle apo-LDH [bottom; White et al. (1976)]. The arrow indicates the position of the scissile bond.

complexed forms of the enzymes (Figure 6). Crystallographic studies of the pig muscle enzyme indicate an overall three-dimensional structure which is similar to these two well-characterized LDHs (Hackert et al., 1973), and by virtue of its high regional homology to the dogfish muscle and pig heart LDHs (Eventoff et al., 1977), it is assumed that the βG–βH strand–loop of the rabbit muscle enzyme is of similar structure and function.

In general, proteases recognize as substrates the unstructured, external polypeptide regions of proteins, such that partial or complete denaturation of proteins tends to enhance their susceptibility to proteolysis. An initial proteolytic event may be thought to partially unravel the structure of a protein, thereby making it labile to subsequent cleavages. Superficially, the cleavage of rabbit muscle and pig muscle LDHs by HIV-1 protease at the Leu-190*Gly-191 dipeptide in the βG sheet does not conform to typical proteolysis because its cleavage site resides within a well-defined structural element of these enzymes which is not located on the protein surfaces in either their apoenzyme or enzyme–inhibitor(s)-complexed forms (Figure 3). One would expect that the βG-βH loop would be required to move outwardly from the unexposed part of the protein to allow the Leu-Gly dipeptide to be accessible to the deep active site cleft. Another problem arises from the fact that the residues of the substrate region of the βG–βH element, namely, Gly-187–Gly-194, are not in a continuous β-strand themselves, as would be required for the binding of a substrate at residues P4–P3′ to the active site of HIV-1 protease (Miller et al., 1989).

As shown in Figure 6, a comparison of the conformations of the scissile decapeptide within dogfish muscle apo-LDH

and inhibitor-complexed pig heart LDH and that of the HIV-1 protease-bound inhibitor Ac-His-Gly-Trp-Ile-LeuΨ-[CHOHCH$_2$]Gly-Glu-His-Gly-Asp-NH$_2$ clearly demonstrates important structural differences. The presence of a "turn" beginning at the P3′ position [Gly-196(194)] of both protein polypeptides is at variance with the extended β-sheet form of the bound inhibitor. While the residues from P5 to P1′ in the protein substrates overlay nicely with those of the decapeptide hydroxyethylene inhibitor, those on the carboxyl-terminal side of the scissile bond (P2′–P5′) depart from this extended conformation of the bound inhibitor, such that binding of this protein decapeptide to the protease would require linearization of this turn. Compared to the extended conformation of the protease-bound decapeptide inhibitor (Figure 6), the inhibitor-bound (Figures 3 and 6) and apoenzyme (Figure 6) forms of the two LDH decapeptides are similar in their nonlinear conformations. Accordingly, the conformational rearrangement of this βG–βH strand–loop structure into a single β-strand which could bind to the active site of HIV-1 protease could accompany the labilization of the quaternary structure at acidic pH.

We have applied the statistical analysis of Poorman et al. (1991) to the sequence of pig heart LDH. Surprisingly, five sites in this sequence are predicted to be more likely cleavage sites than the site actually observed (which has an h value of 0.05). Two of these predicted sites have h values of ≥0.6: RLNL*VQRN (Leu-110*Val-111; h = 0.64) and QQLN*-PEMG (Asn-211*Pro-212; h = 0.6), yet these sequences within rabbit muscle LDH are apparently not cleaved by HIV-1 protease under conditions that result in cleavage at the GWIL*GEHG site. Both of these peptides are situated on surface loops which are considerably more accessible than the GWIL*GEHG sequence.

How, then, do we account for the remarkable ability of HIV-1 protease to recognize this buried and "crumpled" cleavage site within LDH? It is clear from the present studies that RM-LDH is labile to denaturation at acidic pH, such that preincubation of this enzyme at pH <7.0 resulted in its loss of enzymatic activity (assayed at pH 7.0) and disruption of its tetrameric quaternary structure. The degree of susceptibility to specific proteolytic cleavage at Leu-190*Gly-191 accompanies this labilization. Moreover, while RM-LDH is only weakly cleaved by HIV-1 protease at its more stable pH (7.0), the proteolytic "nicking" does not result in a precipitous drop in enzymatic activity as is observed at the acidic pH values, suggesting that this nicking assists the unraveling of an already weakened protein structure. However, as described above, the heat-denatured RM-LDH is a much poorer substrate of the protease than in this acid-labile form. As a result, we propose that weakening or disruption of the quaternary structure of RM-LDH at mildly acidic pH elaborates the exposure to solvent of its cleavage site. In support of this, the enzyme is partially protected from specific proteolysis when its active site is occupied with oxamate and NADH, possibly by sequestering the target βG sheet within an inaccessible internal region of the enzyme, or simply by stabilizing the native structure.

It is also possible that the proteolytic cleavage site within the βG-sheet region of RM-LDH is somewhat conformationally mobile in the apoenzyme, perhaps more so than in the structures of the pig heart and dogfish muscle LDHs for which structure information is available. Indeed, the pig heart LDH is relatively poor substrate of HIV-1 protease despite an identical primary structure in the decapeptide. The comparable kinetic parameters between the presumably unstructured decapeptide substrate and RM-LDH also suggested that this decapeptide sequence within the protein is more loosely structured and solvent-exposed than the crystal structure data suggest.

Table IV: Homology of LDH Cleavage Site to "Class 3" Protein Cleavage Sites

consensus LDH site	GWIL*GEHD
	GWIL*AEHD
lentiviral polyprotein substrates[a]	
HIV-1 *gag*	ATIM*MQRG
HIV-1 *gag*	ARVL*AEAM
HIV-2 *gag*	PFAA*AQQR
HIV-1 *gag*	PRSG*VETT
SIV *gag*	ARLM*AEAL
HIV-1 protease[b]	QITL*WQRP
HIV-1 protease[b]	DQIL*IEIC
nonviral protein substrates	
calcium-depleted calmodulin[c]	GQVN*YEEF
	YEEF*VQMM
bovine pancreatic RNAse[d]	AAK^{succ}*FER
	IVA*^{cm}EG
Pseudomonas exotoxin[e]	ANL*AEEA
	GDAL*LERN
human vimentin[f]	SLNL*RETN
rabbit muscle actin[g]	TQIM*FETF
	SFIG*MESA
troponin C[g]	AEEL*AEIF
Alzheimer amyloid protein[g]	VEVA*EEEE
	LPVN*GEFS
	DSAD*AEED
prointerleukin 1β[g]	DDLF*FEAD
	PFIF*EEEP

[a] Henderson et al. (1988). [b] Strickler et al. (1989). [c] Hui et al. (1990); succ, succinylated; cm, carboxymethylated. [d] Tomasselli et al. (1990). [e] Tomasselli et al. (1991a). [f] Shoeman et al. (1990). [g] Tomasselli et al. (1991b).

Implications for Substrate Specificity of HIV-1 Protease. (a) P2, P1′, and P2′ Residues of the LDH Substrate are Optimal. From the present crystallographic analysis of the HIV-1 protease–Ac-HGWILΨ[CHOHCH$_2$]GEHGD-NH$_2$ complex, it is clear that the important protease–inhibitor binding interactions are found within the P4–P3′ heptapeptide sequence. The substrate Ac-HGWIL*GEHGD-NH$_2$ is probably also arranged as a β-sheet in the active site, as is the inhibitor, such that its binding requires "linearization" of the P2′–P3′ residues within the βG–βH sheet–loop element of LDH, which must be coaxed from the interior of the protein. The GWILGEH heptapeptide sequence is recognized and cleaved by HIV-1 protease when contained within two unrelated proteins, lactate dehydrogenase and engineered galactokinase, and is highly homologous with cleavage sites found within both HIV-1 polyproteins, HIV-1 protease itself, and other nonviral protein substrates (Table IV).

For both proteins, substitution of the P1 Leu by the β-branched Val ablates substrate activity, in concert with the results of identical substitutions within oligopeptide substrates. As described above, the GWILGEH heptapeptide within both the LDHs and galactokinase is likely to be structurally disordered and conformationally mobile, such that its primary structure becomes the major determinant in its recognition by the protease. Detailed analysis of protein substrates of HIV-1 protease of viral and nonviral origins has shown that the likelihood of Ile, Leu, and Glu at, respectively, the P2, P1, and P2′ positions is very high (Pettit et al., 1991; Poorman et al., 1991). Indeed, in the nonviral protein substrates, there is a preponderance of β-branched amino acids at P2, hydrophobic residues at P1, and either Glu or Gln residues at P2′ (Table IV). Accordingly, we conclude that the binding of the P2, P1, and P2′ residues constitutes the most important

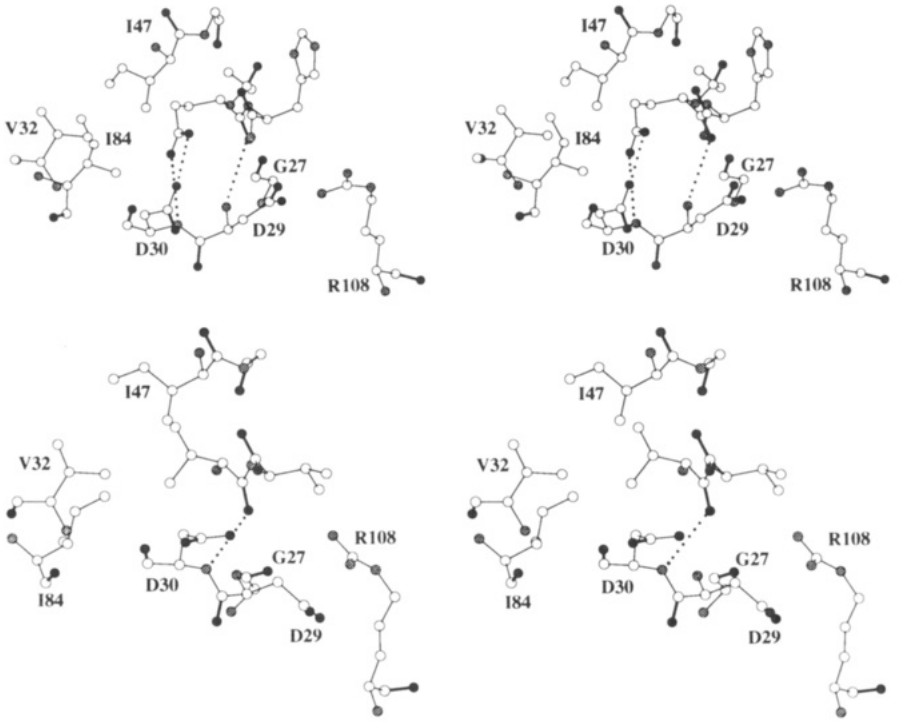

FIGURE 7: Stereo diagrams of the S2′ site of HIV-1 protease when complexed to HGWILΨ[CHOHCH$_2$]GEHGD-NH$_2$ (top) and AAFΨ[CHOHCH$_2$]GVV-OMe (bottom).

interactions between the decapeptide analogue inhibitor and the protease. This conclusion is strengthened upon use of the algorithm of Poorman et al. (1991) as discussed previously.

(*b*) *Glu Binding in the S2′ Site.* From the coexpression system, the S2′ site of the protease tolerates Glu and Gln in the heptapeptide substrates, but not Val and Asp. Likewise, replacement of Glu with Ala in Ac-HGWILGEHGD-NH$_2$ results in a peptide without substrate activity. Thus, the binding of Glu and Gln at the S2′ site is of paramount importance to the productive binding of substrates of this type. The inactivity of Val at S2′ is somewhat surprising considering that it is found in peptide (Moore et al., 1989) and protein substrates (Poorman et al., 1991) and peptide analogue inhibitors (Dreyer et al., 1989, 1992). The inactivity of Asp at P2′ demonstrates that the side-chain carboxylate must be properly positioned and probably is not involved in a favorable ionic interaction. That both the isosteric Glu and Gln side chains are well-tolerated suggest that either the protonation state of Glu is unimportant or it is protonated.

In Figure 7 are shown the S2′ binding sites of the AcHGWILΨ[CHOHCH$_2$]GEHGD-NH$_2$ and AAFΨ-[CHOHCH$_2$]GVV-OMe (Dreyer et al., 1992) complexes of HIV-1 protease. The P2′ Val residue is in close contact with Ile-84, Val-32, and Ile-47. In contrast, the side chain of P2′ Glu is positioned slightly away from these hydrophobic residues. Instead, OE1 of the γ-carboxylate is within hydrogen-bonding distance of and equidistant to OD1 and NH of Asp-30. From molecular modeling of peptide substrates which contain a Glu residue at the P2′ position, Griffiths et al. (1992) have recently proposed hydrogen-bonding interactions of the OE1 oxygen of the Glu residue with the amide hydrogens of Asp-29 and Asp-30. From our results it appears that, if unprotonated, OE1 of Glu-207 could accept a hydrogen bond from the NH of Asp-30. However, if protonated, this carboxylate could also donate a second hydrogen bond to OD1 of Asp-30 to form a bidentate nine-membered ring. In kind, P2′ Gln could also simultaneously accept and donate these hydrogen bonds. However, the β-carboxylate of P2′ Asp would be too far removed from Asp-30 to make these interactions. Asn is frequently found at the P2 position of peptide substrates and presumably is tolerated by the protease here, while Asp is not, due to a lack of anion–anion repulsive interactions with Asp-30. It is also apparent from the structures in Figure 7 that while the β-carboxylate of Asp-30 points away from the P2′ Val residue, it points toward the side chain of the P2′ Glu residue. The reason for the inability of HIV-1 protease to cleave the "LDH peptide" sequence when Glu P2′ is substituted with Val is unclear.

These findings are consistent with favorable interactions between these two carboxylate groups, suggesting that one or both are protonated. In support of this, we showed from the pH studies of the substrate AcHGWILAEHGD-NH$_2$ that either a substrate substituent or an enzymatic residue (other than the active site aspartyl groups) of pK = 5.4 must be protonated in order for this substrate to bind. This pK is high for a carboxylic side chain in a small peptide but is not unusual for such a group in an enzyme (Cleland, 1977). These results support the required protonation of either the P2′ Glu residue of the substrate or Asp-30 on the enzyme for binding of this substrate.

Analysis and graphical examination of a subset (an updated version of the set suggested by Wilmot & Thorton, 1988) of the Brookhaven crystallographic date base (Bernstein et al., 1977) indicates that there exist a number of Glu and Asp residue pairs with carboxylate oxygen distances of less than 3.5 Å. However, most of these occur at protein surfaces and have high thermal factors, indicating that their positions are relatively uncertain. Of the remainder, most occur at metal binding sites or in conjunction with a Lys residue. It is possible that Lys-45 plays this role in the HIV-1 protease–Ac-HGWILΨ[CHOHCH$_2$]GEHGD-NH$_2$ complex but the electron density in that region is not well-resolved. Of course, in the active sites of the aspartic proteases, there is such Asp-Asp pairing in which one of the carboxylates is protonated (Kostka, 1985).

It is therefore possible that it is the protonation of the γ-carboxylic group of Glu-194 within rabbit muscle LDH at pH ≤6.0 that elaborates a conformational change in the βG-βH strand–loop structure that allows binding of this region in an extended conformation within the active site of HIV-1 protease. Within the structure of dogfish muscle LDH, this Glu residue (Glu-192) is within 3.5 Å of the conserved Arg-106 and may in fact form a salt bridge to this residue at neutral pH. If present in the rabbit muscle LDH, such in interaction could be important to the βG-βH strand–loop structure and might be weakened by protonation of Glu-192. The resulting extended β-sheet conformation of this decapeptide sequence could be achieved at acidic pH values at which this enzyme is a good substrate for HIV-1 protease.

Conclusions. The discovery of a structurally defined proteolytic site in a protein substrate of HIV-1 protease which was subject to structural and kinetic investigation has allowed us to conclude the following: The necessity of conformational changes of the scissile peptide to assume binding to the protease, the kinetic equivalence of the protein and decapeptide substrates, and the enhancement of proteolytic lability of the substrate under conditions at which the protein is unstable all give support to the proposal that this protein substrate is recognized by the protease in a partially, but not fully, denatured form. Primary structure is therefore paramount to the binding of this protein substrate to the protease, and the side chains found at the P2, P1′, and P2′ positions contribute the most important protein–protease binding interactions. Thematic to this and other nonviral protein substrates of HIV-1 protease is a Glu or Gln residue at the P2′ position, which clearly constitutes an important element of both substrate recognition and binding to the protease.

ACKNOWLEDGMENT

We thank Michael D. Minnich and Dr. Jeffrey S. Culp for providing purified HIV-1 protease and protease–inhibitor complexes for crystallography, Dr. Steven A. Carr and the Mass Spectrometry group of SmithKline Beecham Pharmaceuticals for mass spectral analyses, and Dr. Renee DesJarlais for assistance with structural data. We thank Dr. Ingrid C. Deckman for providing some of the engineered galK substrates. We thank Wendy Crowell for assistance with preparation of the figures. Coordinates for the structural data disclosed in this paper have been deposited in the Brookhaven Protein Data Base.

REFERENCES

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535–542.

Brunger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science 235*, 458–460.

Cleland, W. W. (1977) *Adv. Enzymol. 45*, 273–385.

Cleland, W. W. (1979) *Methods Enzymol. 63*, 103–137.

Darke, P. L., Nutt, R. F., Brady, S. F., Garsky, V. M., Ciccarone, T. M., Leu, C.-T., Lumma, P. K., Freidinger, R. M., Veber, D. F., & Sigal, I. S. (1988) *Biochem. Biophys. Res. Commun. 156*, 297–303.

Debouck, C. (1992) in *Structure and Function of the Aspartic Proteinases: Genetics, Structures, and Mechanisms* (Dunn, B. M., Ed.) Advances in Experimental Medicine and Biology 306, pp 407–415, Plenum Press, New York.

Debouck, C., Riccio, A., Schumperli, D., McKenney, K., Jeffers, J., Hughes, C., Rosenberg, M., Heusterspreute, M., Brunel, F., & Davison, J. (1985) *Nucleic Acids Res. 13*, 1841.

Debouck, C., Gorniak, J. G., Strickler, J. E., Meek, T. D., Metcalf, B. W., & Rosenberg, M. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8903–8906.

Dreyer, G. B., Metcalf, B. W., Tomaszek, T. A., Jr., Carr, T. J., Chandler, A. C., III, Hyland, L. J., Fakhoury, S. A., Magaard, V. W., Moore, M. L., Strickler, J. E., Debouck, C., & Meek, T. D. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9752–9756.

Dreyer, G. B., Lambert, D. M., Meek, T. D., Carr, T. J., Tomaszek, T. A., Jr., Fernandez, A. V., Bartus, H., Cacciavillani, E., Hassell, A. M., Minnich, M., Petteway, S. R., Jr., & Metcalf, B. W. (1992) *Biochemistry 31*, 6646–6659.

Erickson, J., Neidhart, D. J., VanDrie, J., Kempf, D. J., Wang, X. C., Norbeck, D. W., Plattner, J. J., Rittenhouse, J. W., Turon, M., Wideburg, N., Kohlbrenner, W. E., Simmer, R., Helfrich, R., Paul, D. A., & Knigge, M. (1990) *Science 249*, 527–533.

Eventoff, W., Rossman, M. G., Taylor, S. S., Torff, H.-J., Meyer, H., Keil, W., & Kiltz, H.-H. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 2677–2681.

Fitzgerald, P. M. D., McKeever, B. M., VanMiddlesworth, J. F., Springer, J. P., Heimbach, J. C., Leu, C.-T., Herber, W. K., Dixon, R. A. F., & Darke, P. L. (1990) *J. Biol. Chem. 265*, 14209–14219.

Fruton, J. S. (1976) *Adv. Enzymol. Relat. Areas Mol. Biol. 44*, 1–36.

Goel, O. P., Krolls, U., Stier, M., & Kesten, S. (1989) *Org. Synth. 67*, 69–75.

Grant, S. K., Deckman, I. C., Minnich, M. D., Culp, J., Franklin, S., Dreyer, G. B., Tomaszek, T. A., Jr., Debouck, C., & Meek, T. D. (1991) *Biochemistry 30*, 8424–8434.

Grau, U. M., Trommer, W. E., & Rossman, M. G. (1981) *J. Mol. Biol. 151*, 289–307.

Griffiths, J. T., Phylip, L. H., Konvalinka, J., Strop, P., Gustchina, A., Wlodawer, A., Davenport, R. J., Briggs, R., Dunn, B. M., & Kay, J. (1992) *Biochemistry 31*, 5193–5200.

Gustchina, A., & Weber, I. T. (1990) *FEBS Lett. 269*, 269–272.

Hackert, M. L., Ford, G. C., & Rossman, M. G. (1973) *J. Mol. Biol. 78*, 665–673.

Henderson, L. E., Benveniste, R. E., Sowder, R., Copeland, T. D., Schultz, A. M., & Oroszlan, S. (1988) *J. Virol. 62*, 2587–2595.

Hendrickson, W. A. (1985) *Methods Enzymol. 115*, 252–270.

Holbrook, J. J., & Stinson, R. A. (1973) *Biochem. J. 131*, 739–748.

Holbrook, J. J., Liljas, A., Steindel, S. J., & Rossman, M. G. (1975) *Enzymes 11*, 191–292.

Hui, J. O., Tomasselli, A. G., Zurcher-Neely, H. A., & Heinrickson, R. L. (1990) *J. Biol. Chem. 265*, 21386–21389.

Hyland, L. J., Tomaszek, T. A., Jr., & Meek, T. D. (1991) *Biochemistry 30*, 8454–8553.

Jaskolski, M., Tomasselli, A. G., Sawyer, T. K., Staples, D. G., Heinrikson, R. L., Scheneider, J., Kent, S. B. H., & Wlodawer, A. (1991) *Biochemistry 30*, 1600–1609.

Kiltz, H. H., Keil, W., Griesbach, M., Petry, K., & Meyer, H. (1977) *Hoppe-Seyler's Z. Phys. Chem. 358*, 123–127.

Konvalinka, J., Strop, P., Velek, J., Cerna, V., Kostka, V., Phylip, L. H., Richards, A. D., Dunn, B. M., & Kay, J. (1990) *FEBS Lett. 268*, 35–38.

Kostka, V. (1985) *Aspartic Proteinases and Their Inhibitors* (Kostka, V., Ed.), Walter de Gruyter, Berlin.

Loeb, D. D., Hutchinson, C. A., III, Eagle, M. H., Farmerie, W. G., & Swanstrom, R. (1989) *J. Virol. 63*, 111–121.

Marquardt, D. W. (1983) *J. Soc. Ind. Appl. Math. 11*, 431–441.

Meek, T. D., Dayton, B. D., Metcalf, B. W., Dreyer, G. B., Strickler, J. E., Gorniak, J. G., Rosenberg, M., Moore, M. L., Magaard, V. W., & Debouck, C. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 1841–1845.

Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. W., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B. H., & Wlodawer, A. (1989) *Science 246*, 1149–1152.

Mizrahi, V., Lazarus, G. M., Miles, L. M., Meyers, C. A., & Debouck, C. (1989) *Arch. Biochem. Biophys. 273*, 347–358.

Moore, M. L., Bryan, W. M., Fakhoury, S. A., Magaard, V. W., Huffman, W. F., Dayton, B. D., Meek, T. D., Hyland, L., Dreyer, G. B., Metcalf, B. W., Strickler, J. E., Gorniak, J., & Debouck, C. (1989) *Biochem. Biophys. Res. Commun. 159*, 420–425.

Pesce, A., Fondy, T. P., Stolzenbach, F., Castillo, F., & Kaplan, N. A. (1967) *J. Biol. Chem. 242*, 2151.

Pettit, S. C., Simsic, J., Loeb, D. D., Everitt, L., Hutchinson, C. A. Q., III, & Swanstrom, R. (1991) *J. Biol. Chem. 266*, 14539–14557.

Phylip, L. H., Richards, A. D., Kay, J., Konvalinka, J., Strop, P., Blaha, I., Velek, J., Kostka, V., Ritchie, A., Broadhurst, A. V., Farmerie, W. G., Scarborough, P. E., & Dunn, B. M. (1990) *Biochem. Biophys. Res. Commun. 171*, 439–444.

Poorman, R. A., Tomasselli, A. G., Heinrikson, R. L., & Kezdy, F. J. (1991) *J. Biol. Chem. 266*, 14554–14561.

Richards, A. D., Roberts, R., Dunn, B. M., Graves, M. C., & Kay, J. (1989) *FEBS Lett. 247*, 113–117.

Richards, A. D., Phylip, L. H., Farmerie, W. G., Scarborough, P. E., Alvarez, A., Dunn, B. M., Hirel, Ph.-H., Konvalinka, J., Strop, P., Pavlickova, L., Kostka, V., & Kay, J. (1990) *J. Biol. Chem. 265*, 7733–7736.

Sass, C., Briand, M., Benslimane, S., Renaud, M., & Briand, Y. (1989) *J. Biol. Chem. 264*, 4076–4081.

Schechter, I., & Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157–162.

Shoeman, R. L., Honer, B., Stoller, T. J., Kesselmeier, C., Miedel, M. C., Traub, P., & Graves, M. C. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 6336–6340.

Strickler, J. E., Gorniak, J., Dayton, B., Meek, T., Moore, M., Magaard, V., Malinowski, J., & Debouck, C. (1989) *Proteins 6*, 139–154.

Swain, A. L., Miller, M. M., Green, J., Rich, D. H., Kent, S. B. H., & Wlodawer, A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8805–8809.

Tomasselli, A. G., Hui, J. O., Sawyer, T. K., Staples, D. J., FitzGerald, D. J., Chaudhary, V. K., Pastan, I., & Heinrikson, R. L. (1990) *J. Biol. Chem. 265*, 408–413.

Tomasselli, A. G., Howe, W. J., Hui, J. O., Sawyer, T. K., Reardon, I. M., DeCamp, D. L., Craik, C. S., & Heinrikson, R. L. (1991a) *Proteins 10*, 1–9.

Tomasselli, A. G., Hui, J. O., Adams, L., Chosay, J., Lowery, D., Greenberg, B., Yem, A., Deibel, M. R., Zurcher-Neely, H., & Heinrikson, R. L. (1991b) *J. Biol. Chem. 266*, 14548–14553.

Torff, H. J., Becker, D., & Schwarzwalder, J. (1977) in *Pyridine Nucleotide Dependent Dehydrogenases* (Sund, H., Ed.) pp 31–42, Walter de Gruyter, Berlin.

White, J. L., Hackert, M. L., Buehner, M., Adams, M. J., Ford, G. C., Lentz, P. J., Jr., Smiley, I. E., Steindel, S. J., & Rossman, M. G. (1976) *J. Mol. Biol. 102*, 759–779.

Wilmot, C. M., & Thorton, J. M. (1988) *J. Mol. Biol. 203*, 221–232.

Wlodawer, A., Miller, M., Jakolski, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., & Kent, S. B. H. (1989) *Science 245*, 616–621.

Yamamoto, S., & Nakano, E. (1982) *Int. J. Biochem. 14*, 159–163.