Treffry, A., Harrison, P. M., Cleton, M. I., deBruijn, W. C., & Mann, S. (1987) *J. Inorg. Biochem.*, 1-6.
Watt, G. D., Frankel, R. B., & Papaefythymiou, G. C. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 3640-3643.
Watt, G. D., Frankel, R. B., Spartalian, K., & Steifel, E. I. (1986) *Biochemistry 25*, 4330-4336.
Weir, M. P., Gibson, J. F., & Peters, T. J. (1984) *Biochem. J. 223*, 31-38.
Williams, J. M., Danson, D. P., & Janot, C. H. R. (1978) *Phys. Med. Biol. 23*, 835-851.
Yang, C.-Y., Meagher, A., Huynh, B.-H., Sayers, D. E., & Theil, E. C. (1987) *Biochemistry 26*, 497-503.

# Substrate Analogue Inhibition and Active Site Titration of Purified Recombinant HIV-1 Protease

Alfredo G. Tomasselli,[‡] Mary K. Olsen,[§] John O. Hui,[‡] Douglas J. Staples,[‡] Tomi K. Sawyer,[‡] Robert L. Heinrikson,*[,‡] and Che-Shen C. Tomich[§]

*Biopolymer Chemistry and Molecular Biology Research Units, The Upjohn Company, Kalamazoo, Michigan 49001*
*Received June 21, 1989; Revised Manuscript Received August 9, 1989*

ABSTRACT: The aspartyl protease of human immunodeficiency virus 1 (HIV-1) has been expressed in *Escherichia coli* at high levels, resulting in the formation of inclusion bodies which contain denatured insoluble aggregates of the protease. After solubilization of these inclusion bodies in guanidinium chloride, the protease was purified to apparent homogeneity by a single-step reverse-phase HPLC procedure. The purified, but inactive, protein was denatured in 8 M urea and refolded to produce the active protease. Enzyme activity was demonstrated against the substrate H-Val-Ser-Gln-Asn-Tyr-Pro-Ile-Val-OH, modeled after the cleavage region between residues 128 and 135 in the HIV gag polyprotein. With this substrate, a $V_{max}$ of 1.3 ± 0.2 μmol/(min·mg) and $K_M$ of 2.0 ± 0.3 mM were determined at pH 5.5. Pepstatin (Iva-Val-Val-Sta-Ala-Sta-OH) and substrate analogues with the Tyr-Pro residues substituted by Sta, by Phe$\Psi$[$CH_2N$]Pro, and by Leu$\Psi$[CH(OH)$CH_2$]Val inhibited the protease with $K_I$ values of 360 nM, 3690 nM, 3520 nM, and <10 nM, respectively. All were competitive inhibitors, and the tightest binding compound provided an active site titrant for the quantitative determination of enzymatically active HIV-1 protease.

Human immunodeficiency virus (HIV), a member of the retrovirus group, has been recognized as the causative agent of acquired immunodeficiency syndrome (AIDS) (Weiss et al., 1985; Gallo & Montagnier, 1988). The molecular organization of the HIV genome comprises *gag*, *pol*, and *env* as genes necessary for viral replication (Dickson et al., 1984; Ratner et al., 1985). During viral replication, these genes are expressed as polyproteins which undergo enzymatic cleavage to generate the functional proteins of the mature virus. Genetic and biochemical studies have demonstrated that a virally encoded protease is responsible for the release of protease (autoproteolysis), reverse transcriptase, integrase, and other proteins from the *gag-pol* fusion proteins (Crawford & Goff, 1985; Katoh et al., 1985; Darke et al., 1988; Krausslich & Wimmer, 1988; Le Grice et al., 1988). Recently, the HIV-1 protease has been proven to be essential for viral maturation; site-directed mutagenesis of the protease active site leads to the loss of viral infectivity (Kohl et al., 1988).

The HIV-1 protease consists of only 99 amino acids, and its demonstrated activity and homology to aspartyl proteases such as pepsin and renin (Katoh et al., 1987; Seelmeier et al., 1988; Loeb et al., 1989) led to inferences regarding the three-dimensional structure and mechanism of the enzyme that have since been borne out experimentally (Pearl & Taylor, 1987; McKeever et al., 1989; Miller et al., 1989; Navia et al., 1989; Weber et al., 1989). To obtain quantities of HIV-1 protease needed for biophysical and structural characterization and for testing of inhibitors, a number of laboratories have produced the enzyme either by chemical synthesis (Copeland & Oroszlan, 1988; Nutt et al., 1988; Schneider & Kent, 1988) or by expression in *Escherichia coli* (Farmerie et al., 1987; Debouck et al., 1987; Giam & Boros, 1988; Graves et al., 1988; Hansen et al., 1988; Mous et al., 1988; Darke et al., 1989; Krausslich et al., 1989; Meek et al., 1989). Some of these preparations have been used to test synthetic peptides as substrates and inhibitors (Billich et al., 1988; Darke et al., 1988; Kotler et al., 1988; Moore et al., 1989; Meek et al., 1989; Richards et al., 1989).

Here we report the high-level expression of the HIV-1 protease in *E. coli*; accumulation of the enzyme results in its deposition as insoluble aggregates. A rapid HPLC purification of the protease from these solubilized inclusion bodies, followed by a denaturation/refolding step, yields a homogeneous preparation of active enzyme. This enzyme has been characterized structurally and kinetically, and inhibition constants have been determined with pepstatin and several substrate-based derivatives as inhibitors of the protease. One tight-binding substrate analogue provided titration of the functional catalytic sites in our purified recombinant protease preparation.

## Materials and Methods

*Enzymes, Oligonucleotides, and Plasmid Constructions.* Restriction endonucleases, T4 DNA ligase, and T4 poly-

[‡] Biopolymer Chemistry Research Unit.
[§] Molecular Biology Research Unit.

0006-2960/90/0429-264$02.50/0 

nucleotide kinase were purchased from New England Biolabs, and Klenow fragment of *E. coli* DNA polymerase I was from Boehringer Mannheim. Enzyme reactions were performed according to recommendations of the respective suppliers.

Oligonucleotides were synthesized, purified, and assembled as described (Tomich et al., 1989).

Plasmids were constructed and screened in *E. coli* DH1. Transformation, isolation of DNA fragments, small- and large-scale plasmid preparations, and other basic techniques were adopted from methods described by Maniatis et al. (1982). Cells containing plasmids with the *trp* promoter were grown in the presence of 100 μg/mL tryptophan to repress the promoter. Plasmid constructs were analyzed with restriction endonucleases and by sequencing according to the dideoxy chain-termination method on double-stranded templates (Wallace et al., 1981).

*Induction and Analysis of HIV-1 Protease Expression.* The expression vector pURA-PR3 was transformed into *E. coli* BSt1-c (obtained from Richard Kirschner, The Upjohn Co.) for expression of HIV-1 protease. Except during induction, all cultures were grown at 30 °C with aeration in Luria broth with 0.2% glucose, 100 μg/mL tryptophan, and 100 μg/mL ampicillin. Induction of expression was carried out by dilution of an overnight culture 100- to 200-fold into Luria broth containing 100 μg/mL ampicillin, allowing it to grow at 30 °C for 1 h, and then shifting to 37 °C for 4–5 h. Cells were harvested by centrifugation and stored as a frozen cell pellet or sonicated cell extract. Sonicated extracts were prepared in a buffer composed of 10 mM Tris-HCl at pH 7.4 containing 1 mM EDTA (TE buffer). The expression of HIV-1 protease was monitored by observation of inclusion body formation with a phase contrast microscope and analyzed by SDS–polyacrylamide gel electrophoresis.

*HIV-1 Protease Assay.* The synthetic peptide H-Val-Ser-Gln-Asn-Tyr-Pro-Ile-Val-OH (GSP) served as the substrate for the measurement of HIV-1 protease activity. This peptide corresponds to the sequence from residue 128 to 135 in the HIV *gag* protein (Ratner et al., 1985). Cleavage of the synthetic peptide, as well as the *gag* protein, takes place at the Tyr–Pro bond. HIV-1 protease activity was measured at 30 °C in 50 mM sodium acetate, pH 5.5, containing 10% glycerol, 5% ethylene glycol, and 0.1% Nonidet P-40 in a total volume of 50 μL. After 10 and 20 min of incubation, 20-μL samples were withdrawn into a vial containing 30 μL of 1% trifluoroacetic acid (TFA) and subjected to HPLC analysis. HPLC was carried out with a Vydac $C_4$ column (0.46 × 15 cm), eluting with a linear gradient of 0–30% acetonitrile over a period of 25 min at a flow rate of 1.0 mL/min.

*Isolation of HIV-1 Protease from Inclusion Bodies.* Cells from *E. coli* BStl-c(pURA-PR3) cultures induced to express HIV-1 protease were harvested by centrifugation at 10000*g* for 10 min, resuspended in TE buffer to a cell density of about 20 $A_{550}$/mL, and sonicated with a Branson sonifier for a total of 5 min. Particulate matter, including inclusion bodies and unbroken cells, was collected from the sonicated cell extract by centrifugation at 10000*g* for 10 min; pellets were resuspended in TE buffer and resonicated. The inclusion bodies were then collected by low-speed centrifugation. All manipulations were at room temperature.

The inclusion bodies, isolated from 200–500 mL of *E. coli* culture, were suspended in 200 μL of 6 M guanidinium chloride, sonicated for 10 min, and placed on a shaker for 60 min at room temperature. The suspension was centrifuged at 10000*g* for 10 min, and the clear supernatant was diluted with 1 volume of solvent B [0.15% TFA in 2-propanol/acetonitrile (7:3 v/v)]. This solution was then submitted to RP-HPLC on a Vydac $C_4$ column (0.46 × 25 cm) eluted at a flow rate of 1.0 mL/min with a gradient of increasing concentrations of acetonitrile and 2-propanol. The column was developed with a binary system made up of solvent A (0.15% TFA in water) and solvent B over a period of 33 min with a linear gradient from 30% to 55% B.

*Chemistry and Analytical Methods.* The octapeptide substrate and three derivatives with the generic formula H-Val-Ser-Gln-Asn-Xaa-Yaa-Ile-Val-OH, where Xaa-Yaa (P1–P1′ at the cleavage site) refers to Sta, Phe$\Psi$[$CH_2$N]Pro, or Leu$\Psi$[CH(OH)$CH_2$]Val, were prepared by conventional methods of solid-phase peptide synthesis as previously reported (Sawyer et al., 1988). Chemical syntheses of Leu$\Psi$[CH(OH)$CH_2$]Val and Sta have also been reported by others (Boger et al., 1985; Evans et al., 1985; Woo, 1985; Fray et al., 1986; Kempf, 1986; Wuts et al., 1988). All synthetic peptides were characterized by amino acid analysis, RP-HPLC, and fast atom bombardment mass spectrometry (carried out with a Varian MAT-CH5-Df mass spectrophotometer).

SDS–polyacrylamide gel electrophoresis (SDS–PAGE) was carried out either as described by Laemmli (1970) or with a gel system having *N*,*N*′-diallyltartaramide as a cross-linker (Morse et al., 1978).

Amino acid analysis and protein sequencing were performed on a Beckman Model 6300 amino acid analyzer and an Applied Biosystems Model 470 sequencer as detailed by Tomasselli et al. (1989).

## Results and Discussion

*E. coli Expression of HIV-1 Protease.* The HIV-1 protease is encoded by the viral DNA sequence in a polycistronic arrangement. Production of the protease in *E. coli* requires that the DNA sequence be engineered to have a methionine codon at the 5′ end for initiation of protein synthesis and a translational stop codon at the 3′ end. To assure efficient synthesis of the protease in *E. coli*, the DNA coding sequence at the 5′ end was modified, without changing the amino acid residues, to minimize secondary structure in the mRNA and to optimize codon usage. This DNA sequence was then placed downstream of a strong *E. coli* promoter in a multicopy plasmid to direct expression of the protease. These manipulations involve the following steps. A 920 bp *Bgl*II–*Eco*RV fragment carrying the truncated protease precursor sequence [isolated from pBH10 (Ratner et al., 1985)] was cloned into the *Cla*I–*Bam*HI region of plasmid pBR322 along with two oligonucleotides which supply the beginning of the precursor protease sequence. From the resulting plasmid a 500 bp fragment containing the protease precursor sequence was isolated and inserted into the *Cla*I–*Hin*dIII region of the expression vector pSK101m (Tomich et al., 1989), together with two oligonucleotides to add a translational stop codon to the end of the protease sequence. Then the first half of the mature protease sequence was replaced with oligonucleotides to optimize codon preference for *E. coli* without generating secondary structures in the mRNA. In the resulting plasmid, pPR3, the coding sequence for HIV-1 protease, beginning with Pro-Gln-Ile-Thr... and ending with ...Val-Lys-Gln-Trp followed by a stop codon, is under the control of the strong promoter *trp* in a pBR322 background.

To obtain high-level expression of the HIV-1 protease, it was necessary to use a runaway vector (pURA, obtained from John Mott, The Upjohn Co.) which, upon induction by high temperature (above 37 °C), increases copy number by about 50-fold. Thus the *Eco*RI–*Hin*dIII fragment carrying the *trp*

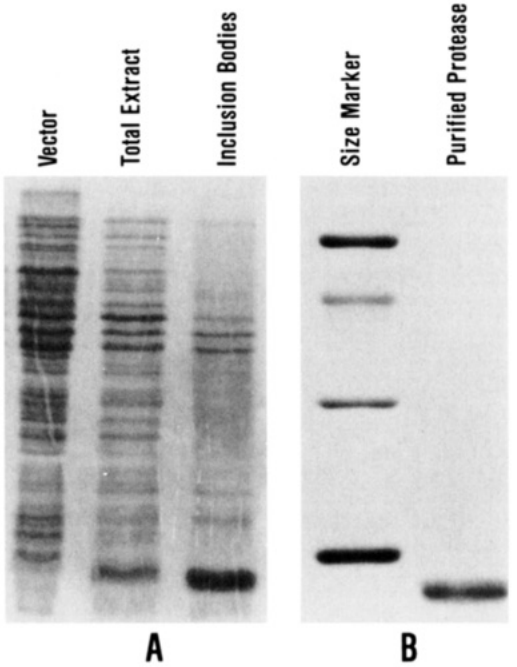


FIGURE 1: (A) SDS–PAGE of *E. coli* cell extracts. The vector lane = cells containing plasmid without the HIV protease sequence, total extract = cells containing the expression plasmid pURA-PR3, and inclusion bodies = inclusion bodies prepared from total extract. (B) SDS–PAGE of HIV protease purified in a single step by HPLC. Size markers ($M_r$) are bovine serum albumin (68 000), ovalbumin (45 000), chymotrypsinogen (25 000), and cytochrome *c* (12 500).

promoter and the coding sequence for HIV-1 protease was isolated from pPR3 and inserted into pURA. The resulting expression plasmid pURA-PR3 was transformed into *E. coli* BStl-c and induced for expression as described under Materials and Methods. High-level expression was obtained, as evident by a prominent protein band of about 10 000 daltons in SDS–PAGE which was not found in cell extracts from cultures harboring the vector without the protease sequence (Figure 1A). Phase contrast microscopy showed that cultures expressing the protease contained refractile inclusion bodies, suggesting that the protein had accumulated as insoluble aggregates within the *E. coli* cells.

*Purification and Denaturation/Refolding of Active HIV-1 Protease.* The purification of *E. coli* derived HIV-1 protease was carried out in a single step by RP-HPLC of solubilized inclusion bodies; active protease was then obtained by refolding from urea. Inclusion body preparations (see Materials and Methods) provided a source in which the HIV-1 protease is a major component (see Figure 1A). The inclusion bodies were extracted with 6 M guanidinium chloride and sonicated, and the clarified extracts were resolved by RP-HPLC (Figure 2). Identification of the HIV-1 protease in the RP-HPLC profile was based upon amino acid composition and sequence analysis, both of which were in agreement with values anticipated from the gene-derived protease primary structure (Ratner et al., 1985). More than 95% of the protease so purified began, as expected, with proline as the N-terminal residue. The remaining 5% of the preparation was accounted for as a component beginning with methionine at position 1. The methionine residue, which is necessary for the initiation of protein synthesis, was not removed completely by the processing enzyme, methionine aminopeptidase (Ben-Bassat et al., 1987), in *E. coli* cells overproducing the protease. SDS–PAGE analysis of the protease run under reducing conditions revealed a single band corresponding to $M_r$ = 10 000 (Figure 1B); similar staining was observed both with Coomassie blue and

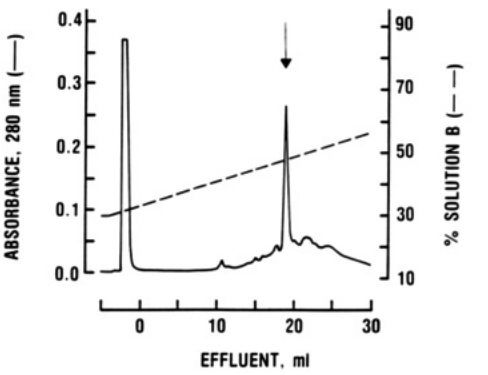


FIGURE 2: HPLC separation of solubilized inclusion bodies. The arrows designates the pure, but inactive HIV protease.

silver. Yields of the protease were in the range of 1–2 mg/L of *E. coli* cells.

The protease obtained in this single-step RP-HPLC procedure was pure as a protein but was enzymatically inactive. The purified material was lyophilized, and the residue was dissolved at a concentration of 5–15 mg/mL in 50 mM sodium acetate buffer, pH 5.5, containing 8 M urea, 1 mM EDTA, and 2.5 mM dithiothreitol (DTT). The completely soluble, unfolded, and reduced molecule was then refolded to an active protease by dilution with 9 volumes of 50 mM sodium acetate, pH 5.5, containing 1 mM EDTA, 2.5 mM DTT, 10% glycerol, 5% ethylene glycol, and 0.2% Nonidet P-40, at 4 °C. The resulting clear enzyme solution was allowed to stand at 4 °C for a few minutes and then stored at −80 °C. HIV protease refolded in this manner shows a specific activity of 1.0 ± 0.1 μmol of substrate hydrolyzed per minute per milligram of protein, consistent with the value of 1.2 published by Darke et al. (1989) for a soluble form of the enzyme expressed in *E. coli*. The enzyme can be stored frozen at −80 °C; freezing and thawing does not appear to lead to any significant loss of activity. It should be stressed that the enzyme can undergo autocatalytic breakdown when stored in solution, even at 4 °C. Storage at −80 °C is recommended to minimize autodegradation.

Overall, our expression–purification method has significant advantages to others described thus far in this rapidly expanding literature. Our high-level expression allows an easy purification procedure and high yield of the protease. Interestingly, the HIV-1 protease was active in 0.8 M urea, although protective agents such as glycols, thiols, and nonionic detergents did promote and preserve enzymatic activity. As first reported by Farmerie et al. (1987), we have observed that the protease produced in *E. coli*, together with HIV gag protein in the same operon, can process the gag polyprotein.

*Enzymatic Activity and Kinetic Characterization of HIV-1 Protease.* The automated RP-HPLC procedure for separation of the octapeptide substrate (GSP) and the tripeptide (H-Pro-Ile-Val-OH) and pentapeptide (H-Val-Ser-Gln-Asn-Tyr-OH) products has proven to be a highly reproducible assay for the enzyme. The course of hydrolysis of this octapeptide substrate is linear with time at least up to 10% conversion to products, and the reactions proceed smoothly to total hydrolysis. Dependency of the reaction on enzyme concentration is fairly linear in the range from 0.01 to 1.0 μM. The present assay is not sensitive enough to explore the enzyme concentration dependency much beyond 0.01 μM. Since all indications point to an enzyme dimer as the functional entity, this value serves at present as an upper limit of the dimer dissociation constant. Analysis of the substrate concentration dependency of GSP cleavage at pH 5.5 gave values of 1.3 ±

Table I: HIV-1 Protease Inhibitor Structure–Function Relationships

| | inhibitor | $K_I$ (nM) |
|---|---|---|
| pepstatin | Iva-Val-Val-Sta-Ala-Sta-OH | 362 ± 5 |
| U-85549E | H-Val-Ser-Gln-Asn-Sta-Ile-Val-OH | 3690 ± 90 |
| U-84645E | H-Val-Ser-Gln-Asn-Phe$\Psi$[$CH_2$N]Pro-Ile-Val-OH | 3520 ± 110 |
| U-85548E | H-Val-Ser-Gln-Asn-Leu$\Psi$[CH(OH)$CH_2$]Val-Ile-Val-OH | <10 |

0.2 μmol/(min·mg of protease) and 2.0 ± 0.3 mM for $V_{max}$ and $K_M$, respectively, values in close agreement with those published by Darke et al. (1989) for the same substrate.

Inspection of HIV-1 protease activity as a function of pH reveals an optimum at pH 6.1; therefore, in contrast to the significantly lower pH optimum of most aspartyl proteases, the value for HIV-1 protease is similar to that of renin (Poorman et al., 1986). A computer-fitted plot of the pH profile of HIV-1 substrate hydrolysis at constant ionic strength shows a dependence upon two dissociating groups of $pK_{a1}$ = 5.0 and $pK_{a2}$ = 7.2. Richards et al. (1989) have reported $pK_a$ values of 3.1 and 6.7; this difference from our results might be attributed to different substrate assay conditions (vide infra). On the basis of thermodynamic calculations (Tanford, 1961) we estimate that these two dissociating groups in the HIV-1 protease correspond to the two carboxylic acid moieties of each monomer-derived, catalytically important Asp residue and that these are separated by a distance of about 5 Å in the dimeric complex. Thus, it is perfectly reasonable to assign the $pK_{a2}$ dissociation which leads to enzyme inactivation to the carboxyl group of one protomer.

*HIV-1 Protease Inhibitors and Active Site Titration.* We synthesized several inhibitors based on the sequence of the octapeptide substrate and determined their $K_I$ values (Table I). Pepstatin was also investigated since it is a well-known inhibitor of aspartyl proteases. As has been shown with other preparations of the HIV-1 protease (Giam & Bores, 1988; Hansen et al., 1988; Seelmeier et al., 1988; Darke et al., 1989), our enzyme was competitively inhibited by pepstatin. A plot of pepstatin concentration versus the reciprocal of velocity with GSP was linear and afforded a $K_I$ value of 362 nM. This value is comparable to that of 1.0 μM reported by Darke et al. (1989). Substitution of statine (Sta) for the $P_1$–$P_1'$ Tyr-Pro of GSP yielded inhibitor U-85549E, which possessed a $K_I$ = 3690 nM. Therefore, the binding affinity of the Sta-modified GSP analogue was about 10-fold less than that of pepstatin. In contrast, $P_1$–$P_1'$ Phe$\Psi$[$CH_2$N]Pro substitution into the GSP template yielded an inhibitor (U-84645E, $K_I$ = 3520 nM) that was also about 10-fold less potent than pepstatin. Finally, a noteworthy improvement in binding affinity was achieved with the $P_1$–$P_1'$ Leu$\Psi$[CH(OH)$CH_2$]Val-modified GSP analogue U-85548E ($K_I$ < 10 nM) which provided an active site titrant of the HIV-1 protease (Figure 3). On the basis of the tight-binding affinity of U-85548E we were able to quantitate the amount of active protease present in our preparations and determined the average value to be in the range of 30–50%. Furthermore, these data allowed us to calculate the $V_{max}$ of the fully active enzyme; a Lineweaver–Burk plot analysis showed that a fully active protease would possess a $V_{max}$ = 4.9 ± 0.2 μmol/(mg·min) and a $K_M$ = 2.0 ± 0.3 mM under the assay conditions we have utilized (see insert, Figure 3).

Several papers have dealt with peptide substrates and inhibitors of the HIV-1 protease (Darke et al., 1988, 1989; Meek et al., 1989; Moore et al., 1989; Richards et al., 1989). All employ peptide substrates that are variants of the HIV *gag* sequence 128–135, as we have in the present study. Hydrolytic rate and binding constants may vary considerably depending upon the exact nature of the peptide substrate and differences in the assay conditions, particularly in ionic strength. Our $V_{max}$

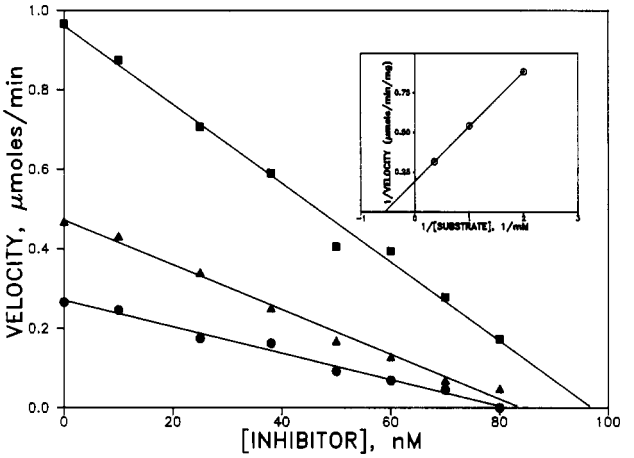


FIGURE 3: Main panel: Active site titration of HIV protease with the substrate-derived inhibitor Val-Ser-Gln-Asn-Leu$\Psi$[CH(OH)-$CH_2$]Val-Ile-Val. Velocity was determined at three different concentrations of substrate, 2.8, 1.0, and 0.5 mM, and expressed as μmol/(min·mg of active plus inactive protein). Insert: Velocity was recalculated as μmol/(min·mg of fully active enzyme) and plotted versus 1/[substrate] to calculate the theoretical of the fully active protease. The experiment was carried out at pH 5.5 with the conditions described under Materials and Methods.

and $K_M$ values agree closely with those published for GSP by Darke and co-workers. This is certainly due primarily to the fact that we use the same substrate, but also in part, perhaps, because we employ essentially identical assay conditions. Differences in $V_{max}$ values reported, for example, by Richards et al. (1989) and Meek et al. (1989) for similar peptides may result from subtle variations in substrate structure, as well as from the use of different detergents and high salt in their assay systems. The protease is known to be sensitive to ionic strength. It must be conceded that we do not know enough about protease specificity to explain variations in kinetic parameters as a function of alterations of peptide structure, even under the same assay conditions. Thus, in the paper by Moore et al. (1989), $V_{max}$ values vary by orders of magnitude in the protease hydrolysis of a set of systematically altered peptide substrates. Nevertheless, the $K_I$ reported by the latter group for a reduced-bond inhibitor differs only by a factor of 4 from that of a similar inhibitor determined in the present work. Our $K_I$ value of about 3500 nM for Val-Ser-Gln-Asn-Phe$\Psi$-[$CH_2$N]Pro-Ile-Val is 4-fold lower than the $K_I$ reported by Moore et al. (1989), for Ac-Ser-Gln-Asn-Phe$\Psi$[$CH_2$N]Pro-Val-Val-$NH_2$.

At present, it is difficult to factor out critical determinants in deriving these different sets of kinetic parameters, but we believe the variability seen from laboratory to laboratory would be minimized by adopting a more standardized protease assay. Part of this standardization process should include the protease itself, and active site titration of preparations with tightly bound inhibitors such as U-85548E described herein provides an important means to access the population of active enzyme molecules. It should be emphasized that this compound is the first HIV protease substrate-based inhibitor having a Leu$\Psi$-[CH(OH)$CH_2$]Val moiety at the $P_1$–$P_1'$ site. A structurally related $P_1$–$P_1'$ Leu$\Psi$[CH(OH)$CH_2$]Val-containing compound, Boc-His-Pro-Phe-His-Leu$\Psi$[CH(OH)$CH_2$]Val-Ile-His-OH

(H-261), has been reported to be a potent inhibitor ($K_I$ = 15 nM) against HIV protease (Richards et al., 1989). However, in contrast to H-261 which was first designed by Szelke et al. (1983) as a renin substrate-based inhibitor ($K_I$ = 0.1 nM against human renin), our GSP-based inhibitors are highly selective for the HIV protease (details to be reported elsewhere). It is noteworthy that the renin substrate peptide H-Pro-His-Pro-Phe-His-Leu-Val-Ile-His-D-Lys-OH (RSP; Poorman et al., 1986) is neither a substrate nor an inhibitor of HIV-1 protease. Thus, the choice of template as well as specific $P_1$–$P_1'$ modifications will undoubtedly be subjects of intensive research in the design of anti-AIDS drugs targeted to the HIV protease.

ACKNOWLEDGMENTS

We thank N. Hatzenbuhler for the synthesis of oligonucleotides, L. L. Maggiora for the synthesis of Phe$\Psi$[$CH_2$N]Pro, L. Baczynskyj for mass spectrometry, and R. A. Poorman and F. J. Kezdy for helpful suggestions regarding the kinetic analyses.

**Registry No.** GSP, 118488-98-5; U-85549E, 124020-60-6; U-84645E, 124020-61-7; U-85548E, 124020-62-8; pepstatin, 26305-03-3; aspartic proteinase, 78169-47-8.

REFERENCES

Ben-Basat, A., Bauer, K., Chang, S.-Y., Myambo, K. Boosman, A., & Chang, S. (1987) *J. Bacteriol. 169*, 751–757.

Billich, S., Knoop, M.-T., Hansen, J., Strop, P., Sedlacek, J., Mertz, R., & Moelling, K. (1988) *J. Biol. Chem. 263*, 17905–17908.

Boger, J., Payne, L. S., Perlow, D. S., Lohr, N. S., Poe, M., Blaine, E. H., Ulm, E. H., Schorn, T. W., LaMont, B. I., Lin, T.-Y., Kawai, M., Rich, D. H., & Veber, D. F. (1985) *J. Med. Chem. 28*, 1779–1790.

Copeland, T. D., & Oroszlan, S. (1988) *Gene Anal. Tech. 5*, 109–115.

Crawford, S., & Goff, S. P. (1985) *J. Virol. 53*, 899–907.

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.

Darke, P. L., Leu, C.-T., Davis, L. J., Heimbach, J. C., Diehl, R. E., Hill, W. S., Dixon, R. A. F., & Sigal, I. S. (1989) *J. Biol. Chem. 264*, 2307–2312.

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.

Dickson, C., Eisenman, R., Fan, H., Hunter, E., & Teich, N. (1984) *RNA Tumor Viruses 2*, 513–648.

Evans, B. E., Rittle, K. E., Homnick, C. F., Springer, J. P., Hirshfield, J., & Veber, D. F. (1985) *J. Org. Chem. 50*, 4615–4625.

Farmerie, W. G., Loeb, D. D., Casavant, N. C., Hutchison, C. A., III, Edgell, M. H., & Swanstrom, R. (1987) *Science 236*, 305–308.

Fray, A. H., Kaye, R. L., & Kleinman, E. F. (1986) *J. Org. Chem. 51*, 4828–4833.

Gallo, R. C., & Montagnier, L. (1988) *Sci. Am. 259*, 40–48.

Giam, C.-Z., & Boros, I. (1988) *J. Biol. Chem. 263*, 14617–14620.

Graves, M. C., Lim, J. J., Heimer, E. P., & Kramer, R. A. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2449–2453.

Hansen, J., Billich, S., Schulze, T., Sukrow, S., & Moelling, K. (1988) *EMBO J. 7*, 1785–1791.

Katoh, I., Yoshinaka, Y., Rein, A., Shibuya, M., Odaka, T., & Oroszlan, S. (1985) *Virology 145*, 280–292.

Katoh, I., Yasunaga, T., Ikawa, Y., & Yoshinaka, Y. (1987) *Nature 329*, 654–656.

Kempf, D. J. (1986) *J. Org. Chem. 51*, 3921–3926.

Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A. F., Scolnick, E. M., & Sigal, I. S. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4686–4690.

Kotler, M., Katz, R. A., Danho, W., Leis, J., & Skalka, A. M. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 4185–4189.

Krausslich, H. G., & Wimmer, E. (1988) *Annu. Rev. Biochem. 57*, 701–754.

Krausslich, H.-G., Ingraham, R. H., Skoog, M. T., Wimmer, E., Pallai, P. V., & Carter, C. A. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 807–811.

Laemmli, U. K. (1970) *Nature 227*, 680–684.

Le Grice, S. F. J., Mills, J., & Mous, J. (1988) *EMBO J. 7*, 2547–2553.

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

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning; a laboratory manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

McKeever, B. M., Navia, M. A., Fitzgerald, P. M. D., Springer, J. P., Leu, C.-T., Heimbach, J. C., Herber, W. K., Sigal, I. S., & Darke, P. L. (1989) *J. Biol. Chem. 264*, 1919–1921.

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., Jaskolski, M., Rao, J. K. M., Leis, J., & Wlodawer, A. (1989) *Nature 337*, 576–579.

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

Morse, L. S., Pereira, L., Roizman, B., & Schaffer, P. A. (1978) *J. Virol. 26*, 389–410.

Mous, J., Heimer, E. P., & LeGrice, S. F. J. (1988) *J. Virol. 62*, 1433–1436.

Navia, M. A., Fitzgerald, P. M. D., McKeever, B. M., Leu, C.-T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L., & Springer, J. P. (1989) *Nature 337*, 615–620.

Nutt, R. F., Brady, S. F., Darke, P. L., Ciccarone, T. M., Colton, C. D., Nutt, E. M., Rodkey, J. A., Bennett, C. D., Waxman, L. H., Sigal, I. S., Anderson, P. S., & Veber, D. F. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7129–7133.

Pearl, L. H., & Taylor, W. R. (1987) *Nature 329*, 351–354.

Poorman, R. A., Palermo, D. P., Post, L. E., Murakami, K., Kinner, J. H., Smith, C. W., Reardon, I., & Heinrikson, R. L. (1986) *Proteins 1*, 139–145.

Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Jr., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C., & Wong-Staal, F. (1985) *Nature 313*, 277–284.

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

Sawyer, T. K., Pals, D. T., Mao, B., Staples, D. J., deVaux, A. E., Maggiora, L. L., Affholter, J. A., Kati, W., Duchamp, D., Hester, J. B., Smith, C. W., Saneii, H. H., Kinner, J., Handschumacher, M., & Carlson, W. (1988) *J. Med. Chem. 31*, 18–30.

Schneider, J., & Kent, S. B. H. (1988) *Cell 54*, 363-368.
Seelmeier, S., Schmidt, H., Turk, V., & Von Der Helm, K. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6612-6616.
Tanford, C. (1961) in *Physical Chemistry of Macromolecules*, pp 548-561, Wiley, New York.
Tomasselli, A. G., Hui, J. O., Fisher, J., Zurcher-Neely, H., Reardon, I. M., Oriaku, E., Kezdy, F. J., & Heinrikson, R. L. (1989) *J. Biol. Chem. 264*, 10041-10047.
Tomich, C.-S. C., Olson, E. R., Olsen, M. K., Kaytes, P. S., Rockenbach, S. K., & Hatzenbuhler, N. T. (1989) *Nucleic Acids Res. 17*, 3179-3197.
Wallace, R. B., Johnson, M. J., Suggs, S. V., Miyoshi, K.-I., Bhatt, R., & Itakura, K. (1981) *Gene 16*, 21-26.
Weber, I. T., Miller, M., Jaskolski, M., Leis, J., Skalka, A. M., & Wlodawer, A. (1989) *Science 243*, 928-931.
Weiss, A., Hollander, H., & Stobo, J. (1985) *Annu. Rev. Med. 36*, 545-562.
Woo, P. W. K. (1985) *Tetrahedron Lett. 26*, 2973-2976.
Wuts, P. G. M., Putt, S. R., & Ritter, A. R. (1988) *J. Org. Chem. 53*, 4503-4508.

# Analysis of the Signals for Transcription Termination by Purified RNA Polymerase II†

Tom K. Kerppola and Caroline M. Kane*
*Department of Biochemistry, University of California, Berkeley, California 94720*
*Received June 19, 1989; Revised Manuscript Received August 16, 1989*

ABSTRACT: Eukaryotic RNA polymerase II recognizes certain DNA sequences as effective signals for transcription termination in vitro. Previously, we have shown that such termination occurs within T-rich sequences; however, not all T runs stop the enzyme nor is the efficiency of termination correlated with the length of the T run. Here we have investigated the sequence elements that signal transcription termination by purified RNA polymerase II. We have examined terminators located within introns of the human histone H3.3 gene and the human c-*myc* gene. Deletion analysis of the H3.3 termination region indicates that the sequences between −6 and +24 relative to the strongest termination site are sufficient to cause transcription termination. The minimal termination signal at this site has been localized to the sequence TTTTTTTC-CCTTTTTT in the nontranscribed strand. A similar but nonidentical sequence has been defined for the c-*myc* termination site. Since RNA polymerase II terminates transcription only within the first run of T residues in these sequences, at least part of the termination signal lies in downstream nontranscribed DNA sequences. Restriction fragment mobility analysis indicates that the H3.3 termination region contains a bend in the DNA helix. Oligonucleotides containing the minimal termination signals also cause restriction fragments to migrate with anomalous mobility. A region of the SV40 genome containing a previously characterized bend also causes RNA polymerase II to terminate transcription. We suggest that a structural element causing a bend in the DNA helix may be part of the signal for transcription termination by purified RNA polymerase II.

Transcription termination is an important mechanism for the regulation of gene expression in both prokaryotic and eukaryotic organisms. Termination in eukaryotes occurs downstream from the processing sites which generate the mature 3′ ends of transcripts for all three nuclear RNA polymerases (Platt, 1986; Sollner-Webb & Tower, 1986; Geiduschek & Tocchini-Valentini, 1988). In addition, intragenic blocks to transcript elongation have been detected in several genes transcribed by RNA polymease II (Bentley & Groudine, 1986; 1988; Eick & Bornkamm, 1986; Cesarman et al., 1987; Kerppola & Kane, 1988; Reines et al., 1987, 1989; Nepveu & Marcu, 1986; Wright & Bishop, 1989; Evans et al., 1979; Maderious & Chen-Kiang, 1984; Mok et al., 1984; Nevins & Wilson, 1981; Hay et al., 1982; Pfeifer et al., 1983; Skarnes et al., 1988; Kao et al., 1987; Resnekov & Aloni, 1989; Bender et al., 1987; Fort et al., 1987; Lattier et al., 1989; Bhat & Padmanaban, 1988). In some cases, these transcriptional blocks may regulate gene expression through conditional transcription termination (Bentley & Groudine, 1986; Eick & Bornkamm, 1986; Cesarman et al., 1987; Mok et al., 1984; Kao et al., 1987).

The study of transcription termination in eukaryotes is complicated by the rapid posttranscriptional processing of primary transcripts. Nuclear run-on analysis has been used to identify regions that block transcript elongation in several genes. However, with this method, it is difficult to identify the specific sites where RNA polymerase stops, and this method cannot distinguish between transcription termination and transcriptional pausing. It is also difficult to exclude the possibility that posttranscriptional processing destabilizes the nascent transcript. In order to circumvent these problems, we have used a defined in vitro transcription system that allows efficient transcription initiation by purified RNA polymerase II in the absence of accessory factors (Kadesch & Chamberlin, 1982). Using this transcription system, we have identified sites, which we designate intrinsic termination sites, where purified RNA polymerase II ceases elongation and releases the nascent

† This work was supported in part by a grant from the National Institutes of Health and in part by a predoctoral fellowship to T.K.K.

0006-2960/90/0429-269$02.50/0 