THE JOURNAL OF BIOLOGICAL CHEMISTRY

Vol. 266, No. 8, Issue of March 15, pp. 4951-4958, 1991
Printed in U.S.A.

A Range of Catalytic Efficiencies with Avian Retroviral Protease Subunits Genetically Linked to Form Single Polypeptide Chains*

(Received for publication, August 16, 1990)

Diane Bizub‡, Irene T. Weber§, Craig E. Cameron¶‖, Jonathan P. Leis¶, and Anna Marie Skalka‡**

From the ‡Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111, the §National Cancer Institute-Frederick Cancer Research and Development Center, ABL-Basic Research Program, Frederick, Maryland 21701, and the ¶Case Western Reserve University, School of Medicine, Cleveland, Ohio 44106

Molecular modeling based on the crystal structure of the Rous sarcoma virus (RSV) protease dimer has been used to link the two identical subunits of this enzyme into a functional, single polypeptide chain resembling the nonviral aspartic proteases. Six different linkages were selected to test the importance of different interactions between the amino acids at the amino and carboxyl termini of the two subunits. These linkages were introduced into molecular clones of fused protease genes and the linked protease dimers were expressed in *Escherichia coli* and purified. Catalytically active proteins were obtained from the inclusion body fraction after renaturation. The linked protease dimers exhibited a 10–20-fold range in catalytic efficiencies (V_{max}/K_m) on peptide substrates. Both flexibility and ionic interactions in the linkage region affect catalytic efficiency. Some of the linked protease dimers were 2–3-fold more active than the nonlinked enzyme purified from bacteria, although substrate specificities were unchanged. Similar relative efficiencies were observed using a polyprotein precursor as substrate. Mutation of one catalytic Asp in the most active linked protease dimer inactivated the enzyme, demonstrating that these proteins function as single polypeptide chains rather than as multimers.

Retroviruses encode a protease in polyprotein precursors which also include domains that ultimately become the structural and enzyme components of the infectious virion. The protease is essential for replication and is responsible for cleaving these precursors to release individual viral proteins, including itself, in a process known as maturation. Mature viral proteases can be purified from virions (Alexander *et al.*, 1987), produced synthetically (Schneider and Kent, 1988), or in bacteria by recombinant DNA technology (Kotler *et al.*, 1988a). The amino acid sequence, biochemical, and crystallographic analyses have shown that retroviral proteases are related to the aspartic protease family which includes the human enzymes pepsin, renin, and cathepsin D (Miller *et al.*, 1989).

A major structural difference between the retroviral and the cell-derived, pepsin-like proteases is that the latter contain a single polypeptide chain with two structurally similar domains, while the former are composed of two separate, identical subunits. It has been proposed that the viral proteins may resemble an ancient progenitor protease from which the pepsin-like enzymes may have evolved by a process of gene duplication and fusion (Tang *et al.*, 1978; Pearl and Taylor, 1987). In the viral enzymes, two protease subunits interact to form a symmetric dimer and the active site is formed by residues from both subunits (Fig. 1*A*). The active sites of retroviral and nonviral aspartic proteases superimpose with a root mean square deviation of 0.5 Å or less for approximately 80 atoms (Jaskólski *et al.*, 1990). Despite this remarkable degree of structural similarity, the viral proteases have a very different specificity for substrate, and turnover numbers that are several orders of magnitude lower than those of the pepsin-like enzymes.

In this report we describe our use of molecular modeling based on the crystal structure of RSV[1] protease, to design a series of linkers which join two protease subunits in a single polypeptide chain resembling the nonviral aspartic proteases. The linkers were designed with different lengths, from 0 to 4 residues, to investigate the importance of flexibility at the protease-protease junction and, in some cases, amino acid substitutions were chosen to test the effect of the salt bridges formed by the free NH_2 and COOH termini.

DNA encoding the linkers, together with cloned protease sequences were assembled to produce the desired fusions of tandemly repeated protease sequences and the fusion proteins (called linked protease dimers) expressed in *Escherichia coli*. The linked protease dimers were purified from *E. coli* and, upon renaturation, were found to be active proteases with substrate specificity similar to the unlinked wild-type enzyme. However, the catalytic properties of the linked dimers varied in a way which suggests that both flexibility and ionic interactions are important in the region of linkage. Some of the linked protease dimers appeared to be 2–3 times more active than the nonlinked enzyme, but none of them approached the high specific activities exhibited by the single polypeptide cell-derived aspartic proteases. Thus, this difference must be ascribed to other structural or sequence differences in the cellular and viral enzymes. Finally, we have demonstrated that mutations can be introduced asymmetrically into the protease subunits. The ease of construction, purification, and renaturation of the linked protease dimers will facilitate more

* This work was supported in part by United States Public Health Grants CA-49042, CA-48703, RR-05539, CA-06927 (to A. M. S.), CA38046 (to J. P. L.), by the National Cancer Institute, Department of Health and Human Services, under Contract No. N01-C01-74101 with ABL (I. T. W.), and also by an appropriation from the Commonwealth of Pennsylvania (to A. M. S.) The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‖ Recipient of a Predoctoral Fellowship GM13628 from the National Institutes of Health.

** To whom correspondence should be addressed: Fox Chase Cancer Ctr. Institute for Cancer Research, 7701 Burkholme Ave., Philadelphia, PA 19111. Tel.: 215-728-2490; Fax: 215-728-2778.

[1] The abbreviations used are: RSV, Rous sarcoma virus; AMV, avian myeloblastosis virus; PR, retroviral protease.

detailed studies of the proteolysis reaction which is essential for virus replication.

MATERIALS AND METHODS

Modeling Procedure—The structure of RSV protease (Jaskólski *et al.*, 1990) was examined on an Evans and Sutherland computer graphics system using the program FRODO (Jones, 1978). The linkers were tested by modeling the changes and additions at the dimer interface. Surface loops were built between Leu-124 and Leu-1′ of the other subunit. The bond lengths and angles were idealized using FRODO refinement, and interatomic distances were checked to avoid close contacts. In all cases reported here it was possible to connect the subunits without extreme distortions or unfavorable contacts. Several of the linkers have amino acid changes that permit formation of salt bridges that were previously made by the free COOH and NH_2 termini. In these cases, the amino acid side chain of the substituted residue was adjusted to form the desired salt bridge. The modeling was done in a qualitative manner to test the different designs, rather than to predict accurate structures.

Bacterial Cells—Expression plasmids were initially introduced into *E. coli* MC1061 (Casadaban and Cohen, 1980). This bacterial strain contains the plasmid pRK248cIts which expresses a temperature-sensitive bacteriophage λ repressor protein used to control expression of the cloned gene from the bacteriophage λ P_L promoter on the expression plasmid (Crowl *et al.*, 1985). After confirmation of positive clones by diagnostic restriction enzyme digestion and protein production, DNA from the expression vectors was introduced into *E. coli* strain AR68 (Baker *et al.*, 1984; Stickler *et al.*, 1989). This protease-deficient strain contains a defective λ prophage with a temperature-sensitive bacteriophage λ repressor (λcI857 ΔB*am* ΔH1) and is also defective in the heat shock response.

Construction of RSV Protease and Linked Protease Dimer Expression Clones—A previously described protease expression vector, pPR (Kotler *et al.*, 1988a), was used to construct the first linked protease dimer, designated PR-PR, by an oligonucleotide mutagenesis protocol (Morinaga *et al.*, 1984; DeChiara *et al.*, 1986). An *Nhe*I site was introduced in the 5′ end of the coding sequence of one protease vector (PR 5′ *Nhe*I) and a similar site was introduced into the 3′ end of the coding sequence of a second vector (pPR 3′ *Nhe*I). To construct pPR 5′ *Nhe*I, the oligomer 5′-AGGAGGAATTAATATGCTAGCGATGACAATGGAA-3′ was added to a mixture of linearized (*Pvu*I-digested) and gapped (*Bgl*II- and *Bss*HII-digested) DNAs from pPR (see Fig. 2*A*). The mixture was denatured and renatured. The DNA was repaired by the Klenow fragment of DNA polymerase I, ligated, and introduced into *E. coli* MC1061:pRK248cIts. Bacterial transformants were first screened by hybridization to the above oligodeoxynucleotide, which was end-labeled with [γ-^{32}P]ATP catalyzed by polynucleotide kinase. Clones which exhibited positive hybridization to the labeled oligodeoxynucleotide were shown to contain sequences corresponding to the newly created *Nhe*I restriction site (*underlined*). pPR 3′ *Nhe*I was constructed by mixing linearized (*Pvu*I-digested) and gapped (*Bgl*II- and *Acc*I-digested) DNA from pPR with the oligodeoxynucleotide 5′-GCTCCGCTTGACAAACCTGCTAGCGTAACAGCTGCCTCGCGCGTTT-3′. Bacterial colonies were screened as above. To construct pPR-PR, a 0.73-kilobase pair *Nhe*I-*Acc*I fragment from pPR 5′- *Nhe*I was ligated to a 2.8-kilobase pair *Nhe*I-*Acc*I fragment from pPR 3′ *Nhe*I.

To construct pPR(T122D,L124G)-PR, the oligodeoxynucleotide 5′ - GGCCTAGGGCTCCGCCTCGACAATGGGCTCGCGATGACAATGGAA-3′ was added to a mixture of linearized (*Pvu*I-digested) and gapped (*Bgl*II- and *Acc*I-digested) DNAs from pPR-PR. The mixture was treated as described above and introduced into *E. coli*. After positive hybridization to radioactively labeled oligomer, plasmid DNAs from the selected bacterial colonies were tested for loss of the *Nhe*I site and acquisition of the newly created *Taq*I and *Ban*II sites (*underlined*).

A Lys codon linker was inserted into pPR(T122D,L124D)-PR to create pPR(T122D,L124D)-K-PR as follows. Oligodeoxynucleotide 5′-CGCCTCGACAATGGGAAGCTTGCGATGACAATGGAA-3′ was added to a mixture of linearized (*Pvu*I-digested) DNA from pPR-(T122D,L124G)-PR and gapped (*Ban*II-digested) DNA from pPR-PR and treated as above. Colonies that showed hybridization to the radiolabeled oligomer were screened for plasmids with sequences that contained a newly created *Hin*dIII site (see Fig. 2*B*).

To create the expression vectors for PR-GGD-PR, PR(L124D)-GGD-PR, and PR-GGGG-PR, linearized (*Pvu*I-digested) and gapped (*Ban*II-digested) DNAs from pPR-PR was mixed with oligodeoxynucleotides 5′-TCCGCTTGACAAACCTCGGGGGCGATCTAGCGATGACAATGG-3′, 5′-CCGCTTGACAAACGATGGCGGAGATCTAGCGATGACAATGG-3′, and 5′-CCGCTTGACAAACCTGGGCGGCGGAGGCCTAGCGATGACAATGG-3′, respectively. Positive PR-GGD-PR, PR(L124D)-GGD-PR, and PR-GGGG-PR clones were screened for acquisition of sequences which contained diagnostic *Ava*I, *Bgl*II, and *Stu*I restriction enzyme sites, respectively (*underlined*).

The construction of bacterial expression vector pPR(D37S) was previously described (Leis *et al.*, 1989). pPR(D37S) was used to construct a linked protease dimer in which the active site in only the amino-terminal domain was mutagenized. A 370-base pair *Eco*RI fragment was isolated from PR-GGGG-PR which contained the junction between the linked protease subunits, including the linker. This fragment was ligated to *Eco*RI-linearized and phosphatase-treated DNA from pPR(D37S). The DNA from this expression vector PR(D37S)-GGGG-PR was screened by digestion with restriction enzyme *Sst*I (diagnostic for the D37S mutation) and by digestion with *Stu*I (diagnostic for the GGGG linker). Orientation of the 370-base pair *Eco*RI fragment was determined by digestion with restriction enzyme *Pst*I.

Induction of Bacterial Cells for Protein Production—In our standard procedure, an overnight culture of AR68 cells was grown at 30 °C in LB broth supplemented with 20 μg/ml tryptophan containing 50 μg/ml ampicillin and 15 μg/ml tetracycline. The cells were diluted 1/50 or 1/100 in the same medium and grown to an optical density at 600 nm of 0.65. While agitating the culture flask slowly, an equal amount of medium at 58 °C was added, and the cells grown at 37 °C for an additional 4 h. The bacterial cells were collected by centrifugation at 8000 rpm for 10 min at 4 °C (JA-14 rotor; Beckman). The medium was decanted and the cell pellet resuspended in TE buffer (10 mM Tris, pH 7.4, 1 mM EDTA). The bacteria were again collected by centrifugation, the buffer was decanted, and the cell pellet stored at −20 °C until use.

Purification of Protein from Inclusion Bodies—Bacterial cells (pellets from 100 ml of induced culture) were resuspended in 5 ml of buffer (50 mM Tris, pH 8.0, 0.5 M NaCl, 1 mM EDTA, 5 mM 2-mercaptoethanol) and disrupted by sonication using four 40-s cycles at an output of 5.5 and 50% duty (Sonicata cell disruptor W-225R; Heat Systems Ultrasonics). The preparation was then subjected to centrifugation at 15,000 rpm (JA17; Beckman) for 30 min at 4 °C to collect bacterial inclusion bodies containing protease protein. The protease protein was extracted from the inclusion bodies by a modification of a procedure described by Manne *et al.* (1985) using 6 M urea containing 10 mM 2-mercaptoethanol at approximately 0.5 mg of protein/ml (~3 ml/100 ml of starting cell culture). The urea was removed in a decreasing stepwise fashion by dialysis (*versus* 4, 2, and 1 M urea with 2-mercaptoethanol, and then a buffer with no urea) in 22-mm DiaCell capsules containing a Spectropor-1 membrane with an M_r 6,000–8,000 cutoff (InstruMed Inc., Union Bridge, MD). Each dialysis was allowed to proceed for 1.5 h at 37 °C with gentle rocking. The final dialysis was in 25 mM sodium phosphate, pH 7.4, 150 mM NaCl, and 10 mM 2-mercaptoethanol. Samples were clarified by centrifugation in a microcentrifuge at full speed for 15 min and then stored at 4 °C. Protein concentration was determined using the Bio-Rad Protein Assay Kit according to the manufacturer's protocol.

Partial Purification of the Polyprotein Precursor Fragment Substrate RSV p36^{IN}—The cloning and expression of p36^{IN} (pFA3-RT36) was previously described (Alexander *et al.*, 1987). Bacterial cells containing the p36^{IN} plasmid were grown and processed as described above except that expression of p36^{IN} was induced by the addition of an equal volume of 65 °C medium and then grown for 4 h at 42 °C. The portion of p36^{IN} protein present in inclusion bodies was isolated as described above for the protease protein.

Immunoblotting of Purified Proteins—Purified proteins were mixed with 2 × sodium dodecyl sulfate-containing sample buffer, boiled for 5 min, and fractionated by electrophoresis in 8.5% or 12.5% denaturing polyacrylamide gels (Laemmli, 1970). Proteins were transferred to Immobilon-P transfer membrane (1PVH00010; Millipore) using the MilliBlot-SDE System (MBBDSDE00; Millipore) for 1 h according to the manufacturer's specifications. After transfer, the membrane was incubated in 25 mM sodium phosphate, pH 7.4, 0.5 M NaCl, 10% heat-inactivated horse serum (Gibco), and 0.1% Tween 20 for 45 min at 37 °C or overnight at 4 °C. The filter was incubated with either rabbit anti-purified AMV protease antiserum or rabbit anti-IN(RSV) antiserum (Alexander *et al.*, 1987) (diluted 1/3,000 and 1/5,000, respectively) at room temperature for 1 h. Antibodies to *E. coli* proteins were removed from anti-protease serum by adsorption to

acetone powder prepared from *E. coli* which contained the pEV-vrfII expression vector (Harlow and Lane, 1988). The filter was washed 3 times for 5 min each in 25 mM sodium phosphate, pH 7.4, 0.5 M NaCl, and 0.1% Tween 20. The Elite Vector ABC reagents (Vector Labs) were prepared according to the manufacturer's protocol using the same buffer as used with the primary antiserum and incubated with the filter for the appropriate time. The filter was washed once in 25 mM sodium phosphate, pH 7.4, 0.5 M NaCl, 0.1% Tween 20, once in 25 mM sodium phosphate, pH 7.4, 0.15 M NaCl, 0.1% Tween 20, and finally in 25 mM sodium phosphate, pH 7.4, 0.15 M NaCl. The antibody reaction was visualized using Indophane Blue as substrate according to the manufacturer's directions (VioMedics, Worcester, MA).

Digestion of p36^{IN} Protein—Partially purified bacterially produced p36^{IN} protein (60 ng) was mixed with purified bacterially produced protease and linked protease dimer proteins (450 ng) in a 30-μl reaction containing a final concentration of 0.1 M sodium citrate, pH 5.2, 2.0 M NaCl, and 1 mM EDTA. The reaction mixture was incubated at 37 °C for 15 min at which time samples (15 μl) were removed and added to an equal volume of water and 4 volumes of 2 × sample buffer. Samples were then loaded on a 8.5% denaturing polyacrylamide gel and subjected to electrophoresis.

Determination of Protease Activity—Protease (0.25 μg) protein was incubated in a 25-μl final volume of a reaction mixture containing 0.4 mM peptide substrate (see below), 2.4 M NaCl, 0.1 mM sodium phosphate, pH 5.9, at 37 °C for various lengths of time. The incubations were stopped with the addition of 170 μl of 0.5 M sodium borate, pH 8.5. The amount of cleavage of the peptide substrate was determined by adding 20 μl of 0.05% (w/v) fluorescamine (Hoffmann-La Roche) and measuring the fluorescence intensity at λ_{ex} of 386 nm and λ_{em} at 477 nm on a Perkin-Elmer LS-5B spectrofluorometer. The relative fluorescence intensity was converted to nanomoles of product formed using the following equation which defines the line of a standard curve obtained by reacting fluorescamine with standardized amino acid solutions: product formed (nmol) = (relative fluorescence − 2.0)/9.5.

Steady-State Kinetic Analysis of Protease—Protease protein was incubated with various concentrations of substrate (0.05–0.60 mM) as described above. No more than 20% of the substrate was allowed to be consumed during the course of any reaction. The initial velocity data were analyzed by double-reciprocal plots and the kinetic constants (K_m and V_{max}) calculated from a linear regression fit of the data. In each case the correlation coefficient, r, was greater than or equal to 0.95. The substrate concentration used in our standard peptide assays (0.4 mM) is only 2–3-fold greater than the calculated K_m values for the various proteases. Concentrations of substrates greater than 5 times the K_m showed decrease in activity probably due to problems in solubility of the peptides.

Measurement of Protein Concentration for Substrate Peptides—Peptides were synthesized and purified as previously described (Kotler *et al.*, 1988b). Stock solutions of peptides were prepared by dissolving lyophilized peptides in 1 mM 2-mercaptoethanol. Concentrations were determined by amino acid composition analysis. The peptides used in this study are as follows: PPAVS-LAMTMRR (NC-PR cleavage site); PATVL-TVALRR (PR-RT); and PPYVG-SGLYRR (p2-p10). Hyphens denote sites of cleavage.

RESULTS

Design Strategies—The crystal structure of RSV protease consists of two subunits in nearly identical conformations (Miller *et al.*, 1989; Jaskólski *et al.*, 1990). The amino (*N*) and carboxyl (*C*) termini are close together on the surface of the protein opposite to the active site region (Fig. 1*A*). The four termini (two from each subunit) form a four-stranded antiparallel β sheet in which the COOH terminus of one subunit is next to the NH_2 terminus of the other subunit in the dimer (Weber, 1990b). This arrangement made it seem feasible to design a single-chain molecule by connecting the subunits. The COOH-terminal residue in the first subunit, Leu-124, is separated by 7 Å from the NH_2-terminal residue in the second subunit, Leu-1′ (indicated by a prime to distinguish it from the first subunit). Asn-123 is closer to Leu-1′ since Leu-124 turns away from the plane of the β sheet, and the negatively charged COOH terminus forms an ionic interaction with the side chain of Arg-120′ (Fig. 1*B*). Molecular modeling suggested that it was possible to make a direct connection between Leu-124 and Leu-1′ by moving residue 124. This, the simplest and shortest possible linker (of zero length), is designated PR-PR (Fig. 2). Independent experiments in our laboratory[2] showed that deletion of Leu-124 does not eliminate RSV protease activity in an *in vitro* assay. Thus, an improvement on the design of PR-PR was to replace Leu-124 by Gly to allow a more flexible connection. In addition, the salt bridges formed by the charged termini were considered. In the linked dimer, one COOH terminus is joined to the other NH_2 terminus, so a negative and a positive charge are removed, eliminating salt bridges to Arg-120′ and Glu-92′. Therefore, for this second construct, Thr-122 was replaced by Asp so that the Asp-122 side chain could reform the ionic interaction with Arg-120′. This dimer, PR(T122D,L124G)-PR (Fig. 2), also has a zero length linker.

The next design includes the same changes described for the second construct and, in addition, a Lys residue was inserted between the two protease sequences to allow formation of a hairpin turn between Gly-124 and Lys. This addition was expected to increase the flexibility of the linker and to interact with the negative charge on Glu-92′ (PR(T122D,L124G)-K-PR; Fig. 1*C*).

The final series of linkers were designed to permit more flexible connections between the two original subunits. The longest linker was four amino acids in the construct PR-GGGG-PR (Fig. 1*D*, Fig. 2). This linked protease dimer does not include either of the salt bridges to the exposed charged side chains. In construct PR-GGD-PR (Fig. 2), an Asp was included to avoid reforming a possible cleavage site at the NH_2-terminal border of the second protease subunit. This protease dimer, with a 3-residue linker, also allows a flexible connection. A derivative, PR(L124D)-GGD-PR (Fig. 2), has the same linker, but also the possibility of reforming an ionic interaction between Asp-124 and Arg-120′. Our modeling suggests that the Asp at 124 can be easily repositioned to form this salt bridge with Arg-120′ instead of with the carboxylate at the terminus that has been removed with the L124D replacement.

Construction and Expression of Linked Protease Dimers in E. coli—Fig. 2*B* shows the nucleotide sequences in the region of fusion between the two DNA fragments encoding the protease subunits in our linked protease dimers. In each case, a diagnostic restriction site was incorporated to verify that the correct fusion had occurred and to readily distinguish one construct from another. In some cases this was accomplished without alteration in the amino acid sequence. In other cases alterations were intended and these are indicated in the figure in the single-letter code.

All of the vectors expressed protease proteins at high levels; most of it accumulated as a denatured form in inclusion bodies that could be separated from the bulk of cellular proteins by simple centrifugation procedures. After extensive washing of the pelleted inclusion bodies, the remaining protease proteins were solubilized by addition of 6 M urea containing 10 mM 2-mercaptoethanol. Renatured, enzymatically active proteins were obtained using a stepwise dialysis procedure. Fig. 3 shows results from denaturing polyacrylamide gel electrophoresis analyses which indicate that in each case the major protein is a single species. As shown previously (8), bacterially expressed protease (PR) and an enzymatically inactive protein, PR(D37S), that contains a substitution of Ser for Asp in the catalytic site, migrate to similar but not identical positions. It is noteworthy that a single amino acid substitution can lead

[2] H. Burstein, personal communication.

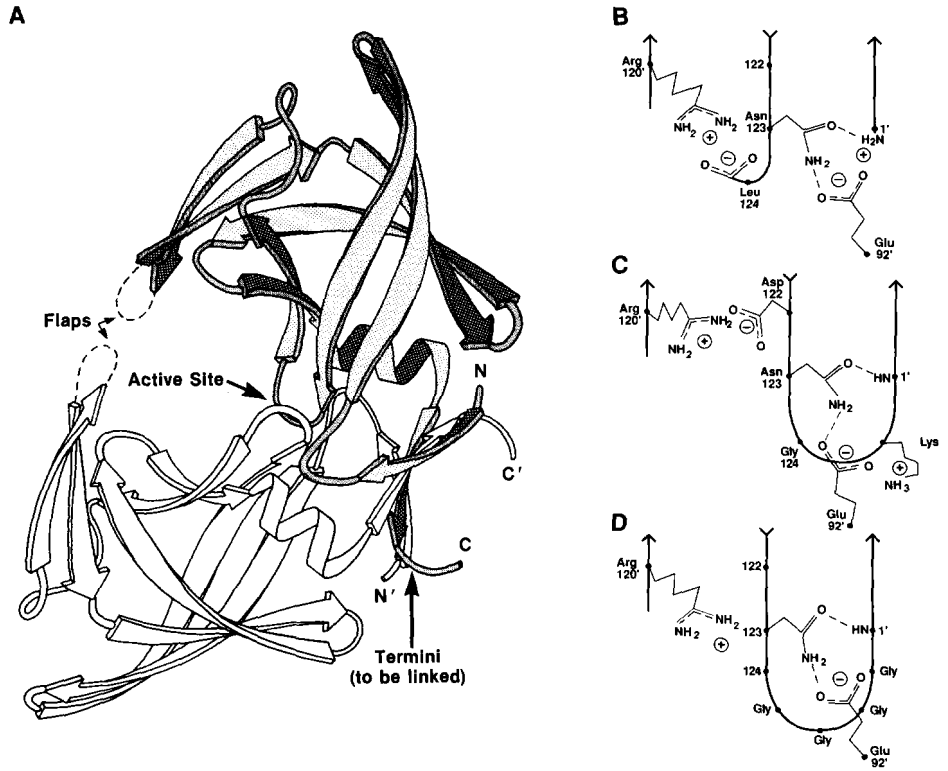

FIG. 1. **Structure of protease and design of linkers.** *A*, schematic representation of the dimer of RSV protease, a modification of the original drawing by Jane Richardson, in which one subunit is *shaded*. β strands are indicated by *curved arrows*, and a helix is also shown. About 10 residues at the tip of the flaps are not visible in the electron density map (Jaskólski *et al.*, 1990). The active site and the termini are indicated. The NH_2 and COOH termini in the dimer form a four-stranded β sheet (on the right side of the dimer). *B*, interactions formed by the COOH terminus of one subunit and the NH_2 terminus of the other subunit in the crystal structure of RSV protease dimer (Miller *et al.*, 1989; Weber, 1990). Three of the four β strands forming the four-stranded β sheet are shown as *thick lines* with *dots* at the position of each αC atom. Only the amino acid side chains that interact with the charged termini are shown. Residues from the second subunit are identified with a *prime*. The COOH terminus forms an ionic interaction with Arg-120′, and the NH_2 terminus forms an ionic interaction with Glu-92′. *C*, the junction region of PR(T122D,L124G)-K-PR is shown which contains one additional residue and alterations at 2 amino acid residues compared to the wild-type (*B*). Thr-122 is changed to Asp which is predicted to form an ionic interaction with Arg-120′. Leu-124 is changed to Gly to permit a more flexible turn. Lys is added between the original protease subunits to allow formation of a salt bridge with Glu-92′. The ionic interactions formed by the charged termini in the wild-type have been reformed. *D*, the junction region of PR-GGGG-PR is shown in a representation similar to that in *B*. Four glycines connect the two subunits. The ionic interactions formed by the charged termini in the wild-type have been eliminated.

to a significantly altered migration rate. This suggests that certain conformational differences are maintained even under these denaturing conditions. A small amount of slower migrating protein is seen above the major product in each of the lanes that contain bacterially expressed proteins. Immunoblot analysis (Fig. 4) shows that these apparently longer proteins are related to protease. The bacterial host is su^{ts}, and these upper bands could reflect a small amount of termination codon readthrough that occurs at the intermediate temperature (37 °C) used for induction. The linked protease dimer proteins migrate at rates similar to each other, but in some cases distinct differences are observed. We presume that this, and the fact that all of the migration rates are slower than would be predicted from the values for monomers, also reflects conformational restrictions. The results in Fig. 4 show that the major immunoreactive protein in each preparation is protease or a linked protease dimer. In some of the linked protease dimer preparations a very small amount of faster migrating immunoreactive protein is detected. This may result from breakdown of the full length proteins, internal initiation, or early termination of translation in the bacterial cells. No protease-related product was detected in the monomer region in any of the linked protease dimer preparations. From the results of Figs. 3 and 4 we estimate that the full length protease products are greater than 90% pure in each of our preparations.

Enzymatic Activity of the Linked Protease Dimers and Stimulation by High Salt Concentration—The experiment in Fig. 5, which compared the ability of the different proteins to cleave a peptide substrate as a function of salt concentration, shows that all of the linked protease dimers are active. They appear to fall into three general categories compared to the bacterially derived protease (PR): those which are less active (PR-PR and PR(T122D, L124G)-PR), those that have approximately equal activity (PR-GGD-PR and PR(L124D)-GGD-PR) and those which are more active (PR-GGGG-PR and PR(T122D, L124G)-K-PR). The virus-derived enzyme shows a similar dependence upon added salt for activity (Kotler *et al.*, 1989). Specific activities determined at optimal conditions are listed in Table I. The bacterially produced renatured protease (PR) is approximately one fourth as active as the virus-derived protease, presumably due to incomplete renaturation (Leis *et al.*, 1989). Comparison of the linked protease dimers show a 10–20-fold difference between the

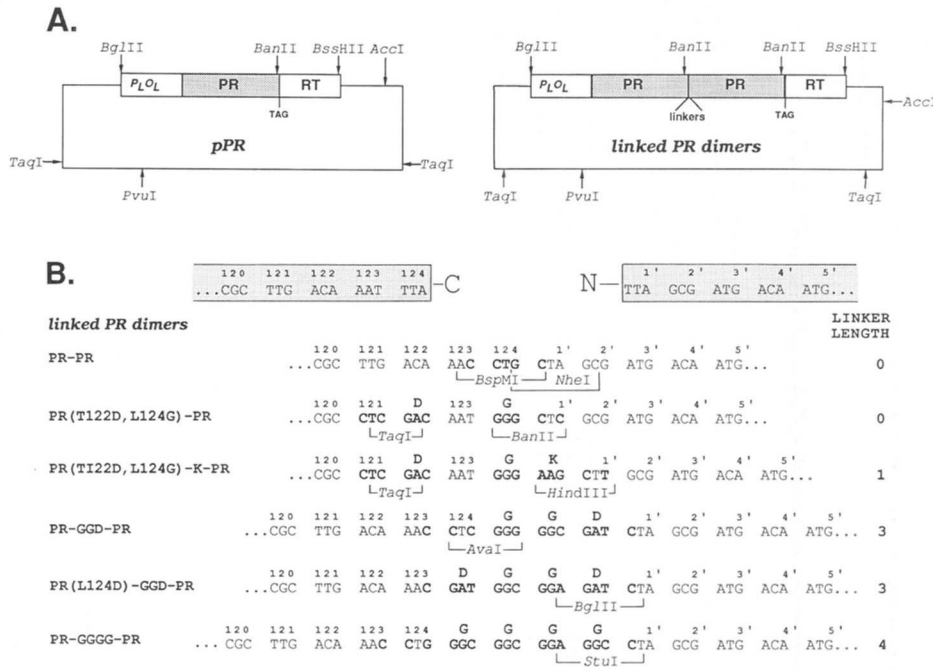

FIG. 2. **Genetic construction of the linked RSV protease dimers.** *A*, a diagramatic representation of the protease and linked protease dimer expression vectors, drawn to scale. Relevant restriction sites are indicated. P_LO_L is the bacteriophage λ promoter and operator region. The genetically engineered amber stop codon (TAG) is shown at the end of the protease domain. *B*, the nucleotide sequence encoding the COOH and the NH_2 termini of subunits 1 and 2 of RSV protease, respectively is shown at the *top*. Sequences at the fused regions of each of the six linked protease dimers are indicated *below*. The numbers above each codon identify the amino acids in each of the protease subunits. Letters in *bold* above the codons show amino acid substitutions or additions as linkers between the subunits. *Bold* letters in the nucleotide sequence indicate changes with respect to wild type or additions used to produce each linked protease dimer. The diagnostic restriction site which distinguishes each construct is also shown. Indicated to the *right* is the linker length for each linked protease dimer.

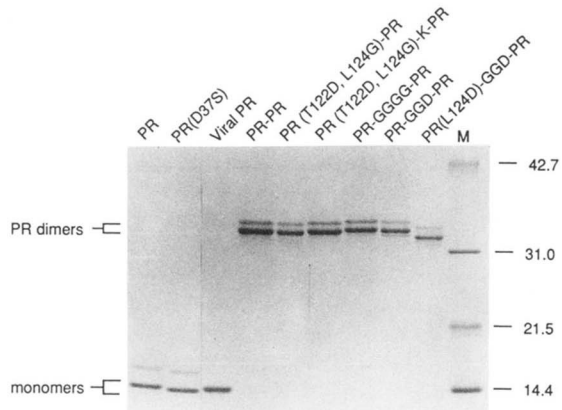

FIG. 3. **Gel analysis of purified protease and linked protease dimer proteins.** Bacterially produced protease and linked protease dimers (see Fig. 2) were solubilized from inclusion bodies and renatured as described under "Materials and Methods." Proteins (2 μg each) were mixed with 2 × sample buffer, boiled and fractionated on a 12.5% denaturing polyacrylamide gel. Proteins were visualized by Coomassie Blue staining. *Brackets* to the *left* show the position of the linked protease dimers and the protease monomers. *Lane M*, sodium dodecyl sulfate-polyacrylamide gel electrophoresis molecular weight standards, low range ($M_r \times 10^3$; Bio-Rad).

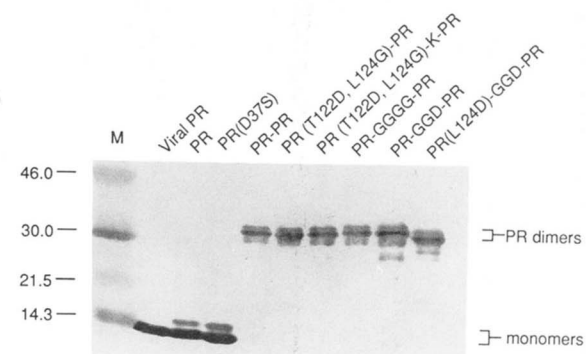

FIG. 4. **Immunoblotting of purified protease and linked protease dimers.** Bacterially produced protease and linked protease dimer proteins (see Fig. 2) were solubilized from inclusion bodies and renatured as described under "Materials and Methods." Proteins (50 ng each) were mixed with 2 × sample buffer, boiled and fractionated on a 12.5% denaturing polyacrylamide gel, transferred to Immobilon-P membrane, and reacted with rabbit anti-protease antisera as described under "Materials and Methods." *Brackets* to the *right* show the position of the linked protease dimers and the protease monomers. *Lane M*, Rainbow protein molecular weight markers ($M_r \times 10^3$; Amersham Corp.).

least active and the most active proteins.

We also asked whether the linked protease dimers could recognize protease target sites in a larger precursor fragment. Results of one such experiment, in which we measured the ability of the enzymes to cleave a COOH-terminal target (IN-p4) in bacterially produced $p36^{IN}$ to release mature IN protein, are shown in Fig. 6. We observed specific cleavage and the same general pattern of relative activities obtained with the peptide substrates shown in Fig. 5 and Table I. Longer incubations under these conditions led to eventual degradation of the IN protein product, and in other experiments (not shown), we observed an apparent autodegradation of the protease proteins themselves. Autodegradation of nonlinked HIV protease has been reported by Tomasselli *et al.* (1990).

Kinetic Properties of the Covalently Linked Enzymes—A steady-state kinetic analysis was carried out on the linked protease dimers to determine if the differences in relative

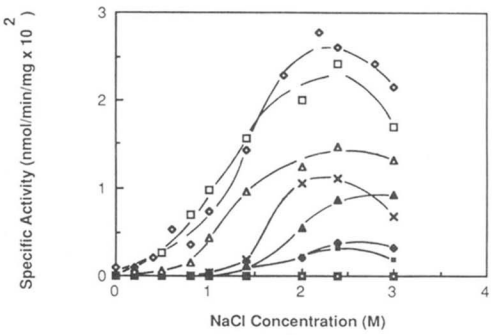

FIG. 5. **The effect of NaCl on the activity of purified protease and linked protease dimers.** Various forms of protease, as indicated below, were incubated with the NC-PR peptide substrate for 30 min at 37 °C as described under "Materials and Methods" except that the NaCl concentration was varied from 0 to 3 M. The amount of peptide cleavage was measured with fluorescamine as described under "Materials and Methods." Reactions contained the following amounts of bacterially derived enzyme: PR, 0.54 μg (▲); PR(D37S), 0.63 μg (□); PR-PR, 0.46 μg (■); PR(T122D,L124G)-PR, 1.26 μg (◆); PR(T122D,L124G)-K-PR. 0.46 μg (□); PR-GGD-PR, 0.5 μg (×); PR(L124D)-GGD-PR, 0.7 μg (△); PR-GGGG-PR, 0.54 μg (△). Data are expressed as specific activities.

TABLE I

Specific activities of the viral and bacterially expressed protease and linked protease dimers

The specific activity for each protease was determined using the NC-PR peptide substrate in the presence of 2.4 M NaCl as described under "Materials and Methods." Time of incubation was 30 min.

Protease tested	Specific activity
	nmol/min/mg protein
AMV protease	389
PR	93
PR(D37S)	0
PR-PR	15
PR(T122D,L124G)-PR	37
PR(T122D,L124G)-K-PR	242
PR-GGD-PR	93
PR(L124D)-GGD-PR	153
PR-GGGG-PR	277
PR(D37S)-GGGG-PR	0

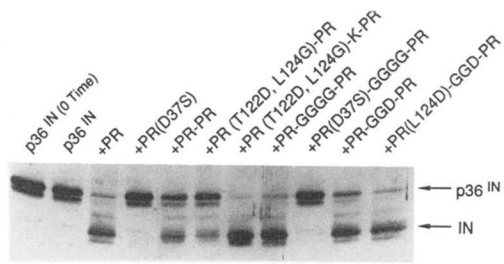

FIG. 6. **Cleavage of partially purified RSV p36^{IN} protein by purified protease and linked protease dimers.** RSV p36^{IN} protein was mixed with protease and linked protease dimer proteins as described under "Materials and Methods." The reactions were stopped by addition of 2 × sample buffer, boiled, and proteins fractionated on a 8.5% denaturing polyacrylamide gel, transferred to Immobilon-P membrane, and reacted with rabbit anti-IN(RSV) antisera (see "Materials and Methods" for details). To control for digestion by bacterial proteases that may be present in the partially purified p36^{IN} protein preparation, p36^{IN} protein was incubated in reaction buffer with and without protease and linked protease dimers or mixed (without prior incubation) with 2 × sample buffer, boiled, and applied to the gel (P36^{IN}, *0 Time*). *Arrows* to the *right* indicate the position of p36^{IN} and IN proteins.

specific activities summarized above reflected a greater affinity for substrates or an increased rate of catalysis. Table II lists the kinetic constants determined for the most active linked protease dimers and the NC-PR peptide substrate in the presence of 2.4 M NaCl, the optimal salt concentration for activity. With this substrate, the linked protease dimers, PR-GGGG-PR, PR(L124D)-GGD-PR and PR(T122D, L124G)-K-PR, show a decrease in K_m and an increase in V_{max} resulting in an approximately 2–3-fold increase in the apparent catalytic efficiency (V_{max}/K_m) compared to the bacterially derived protease (PR) (Table III). When assayed in lower salt concentration (1 M NaCl), a 2–3-fold increase in K_m and a decrease in V_{max} was observed for these enzymes. Thus, there was a 5–8-fold decrease in the overall efficiency of the reaction at suboptimal salt concentrations. There was approximately a 20-fold difference in the apparent catalytic activities of PR-PR (the least active), compared to PR-GGGG-PR (the most active) linked protease dimer.

Linked Protease Dimers Retain the Specificity of the Viral Protease—Table III shows a comparison of the catalytic efficiencies of a subset of the linked protease dimers with three of the ASLV peptide substrates, representing the NC-PR, PR-RT, or the p2-p10 cleavage sites. We also tested peptide substrates which represent the RT-IN and NC-p6a,b cleavage sites of the HIV polyprotein precursor. The HIV substrates were either not cleaved or cleaved at a low rate by the AMV protease or the bacterially derived proteases, although these substrates were very susceptible to cleavage by purified HIV 1 and HIV 2 proteases (data not shown). As summarized in Table III, the nonfused protease is more active on the NC-PR and PR-RT peptide substrates than on the p2-p10 substrate. A similar behavior was observed with the viral protease (data not shown) and by the linked protease dimers (Table III). Also, the relative efficiencies of the various linked protease dimers observed with the NC-PR peptide substrate was the same as with the PR-RT and p2-p10 peptide substrates.

TABLE II

Comparison of the kinetic constants of nonlinked and the most active linked protease dimers

Steady-state kinetic analysis was conducted as described under "Materials and Methods," using the NC-PR peptide substrate with 0.25 μg of each protease and in a reaction containing 2.4 M NaCl. Values represent the mean of three experiments. Numbers in parentheses are derived from assays in 1 M NaCl.

Enzyme	K_m	V_{max}
	mM	*μmol/min* × 10^5
AMV protease	0.23 (0.32)	8.8 (1.7)
PR	0.18 (0.37)	3.3 (0.8)
PR-GGGG-PR	0.13 (0.37)	6.9 (2.2)
PR(T122D,L124G)-K-PR	0.15 (0.36)	6.0 (1.7)
PR(L124D)-GGD-PR	0.13 (0.34)	3.8 (1.4)

TABLE III

Comparison of catalytic efficiency of nonlinked and linked protease dimers

Catalytic efficiencies were determined using the substrate peptides indicated as described under "Materials and Methods." Values represent the mean of three experiments.

Enzyme	V_{max}/K_m		
	NC-PR	PR-RT	p2-p10
	μmol/min/mM × 10^5		
PR	18.3	22.7	4.3
PR-GGGG-PR	53.1	65.2	9.2
PR(T122D,L124G)-K-PR	40.0	48.2	8.0
PR(L124D)-GGD-PR	29.2	35.2	5.2
PR(T122D,L124G)-PR	4.8	5.5	1.7

Thus we conclude that linking the two subunits of the protease, in the manner described for the three enzymes tested, does not alter their relative specificities for the normal substrates.

Mutations in Individual Domains of the Protease—The availability of a bacterial expression system for linked protease dimers now allows us to introduce mutations asymmetrically into one of the two domains. The D37S mutation, which substitutes a Ser for the active site Asp, was introduced into the 5′ subunit of the PR-GGGG-PR construct to produce PR(D37S)-GGGG-PR. Cleavage of precursor proteins (Fig. 6) or peptide substrates (Table I) was not detected with this mutant protein, even after incubation periods in which the substrate was completely digested by the unmodified PR-GGGG-PR (not shown). Thus, substitution of only one of the two catalytic Asp residues inactivates the linked protease dimer. These results indicate that the linked protease dimers form primarily intramolecular active sites, rather than the intermolecular active sites which could conceivably be formed by two linked protease dimers.

DISCUSSION

We have fused two copies of the viral protease sequence to produce a single polypeptide chain resembling cell-derived, pepsin-like proteases. A series of amino acid substitutions and/or insertions were introduced to provide spacer sequences connecting the COOH terminus of one subunit with the amino terminus of the second subunit, and to test the importance of flexibility and ionic interactions. All of the linked protease dimers show a single major band on SDS-gel electrophoresis. This, and the presence of catalytic activity suggests that upon renaturation the constructs fold into single-chain molecules resembling the wild-type protease dimer. The catalytic activity of the linked protease dimer is abolished if one of the active-site Asp-37 residues is changed to Ser, verifying that the active site is formed by residues from both protease domains in a single polypeptide chain. We note, however, that the efficiency of folding is less than optimal; the bacterially derived protease has a lower specific activity than the native viral enzyme. Denaturation and renaturation of the viral enzyme results in a similar reduction in specific activity (Leis *et al.*, 1989).

Although all of the purified linked protease dimers exhibited catalytic activity, the different constructs varied in activity from less than to approximately 2–3 times the specific activity of the bacterially derived nonfused protease dimer. The subunits that were covalently linked without a spacer sequence (PR-PR and PR(T122D,L124G)-PR) exhibited relatively low catalytic activity. This suggests that structural distortions in these fusions were unfavorable for catalysis. The level of catalytic activity of the linked protease dimer compared to the nonfused enzyme increased with the length and flexibility of the spacer. The most active enzyme is PR-GGGG-PR with a spacer of 4 glycines. Flexibility is predicted to be important in forming what we presume will be a hairpin turn between the two β strands 119 to 123, and 1′ to 5′. While one spacer Gly should be sufficient to allow a hairpin turn, catalysis is evidently favored by a longer spacer. If the various crystal structures of RSV protease and HIV protease with and without bound inhibitor are compared, the two subunits in the dimer are related by 178–180° in different cases, a conformational variation of about 2°. Our results showing increased activity for the PR-GGGG-PR construct are consistent with the idea that conformational adaptability is important for catalysis.

Flexibility is not the only important factor. The construct, PR(T122D,L124G)-K-PR, which is predicted to restore the charged interactions associated with the free termini in the wild-type protease dimer and to form a hairpin turn between Gly-124 and the Lys spacer, is nearly as active as PR-GGGG-PR. This suggests that the salt bridges formed to Arg-120′ and Glu-92′ (Fig. 1*B*) are also important for catalysis. This linked dimer is the only one in which both charge interactions may be restored. The relative specific activity of the linked protease dimer is also increased when only the negative charge is restored (*e.g.* PR(L124D)-GGD-PR compared to PR-GGD-PR). The charge interactions between the NH_2- and COOH-terminal β strands may be important for maintaining the conformation of this β sheet which in the unlinked wild-type enzyme contributes about 50% of the intersubunit hydrogen bond and ionic interactions in dimers of both HIV and RSV protease. The COOH terminus alone provides about 45% of the total surface area buried during dimer formation (Weber, 1990a). This suggests that the correct conformation of this β sheet is important for dimer formation and possibly also for catalysis. This β sheet is adjacent to some residues of the active site but is not directly part of it. Mutations of residues in this β sheet in HIV protease reduce enzymatic activity (Loeb *et al.*, 1989).

Linked dimers have been constructed in other systems, for single-chain antigen-binding proteins consisting of an antibody variable light chain sequence tethered to a variable heavy chain sequence (Bird *et al.*, 1988), and from connected genes of human superoxide dismutase (Hallewell *et al.*, 1989). These linked dimers showed nearly normal specific activity and specificity. The reason for the higher specific activity for two of our linked protease dimers compared to nonfused protease is not clear. The monomer-dimer equilibrium may influence activity. However, the specific activity of the viral protease is independent of protein concentration down to 100 nM, the limit of sensitivity for our assay (data not shown). At 375 nM, (the protein concentration used in our standard peptide assays), we would not expect protease activity to be greatly affected by an equilibrium whose K_d value is below 100 nM. An alternative explanation is that the structural constraints provided by tethering the two subunits favor catalysis. Conversely, distortions of the β sheet induced by some of the shorter linkers used in our constructs may be unfavorable for catalysis. Finally, it is possible that renaturation of the linked enzymes is simply more efficient than the nonlinked protease.

While the catalytic efficiencies (V_{max}/K_m) differed, the linked protease dimers had the same substrate selectivity as the nonfused protease. The relative activities on RSV substrates were the same for all of the linked protease dimers and the HIV peptides tested were poor substrates. All of the protease proteins showed a 5–8-fold increase in the catalytic efficiency in high salt concentration. This reflects both an increased affinity for substrate and increased catalytic rate. Since we do not detect changes in the monomer-dimer equilibrium of viral protease under our assay conditions, it is possible that the salt enhances favorable hydrophobic interactions and/or helps to stabilize intermediates along the reaction pathway. Even under optimal conditions, however, the activity of the viral protease is substantially less than the pepsin-like proteases. The data presented here indicate that simply connecting the protease subunits cannot explain this large disparity. The converse experiment also fails to shed light on these differences. Bianchi *et al.* (1990) have made a catalytically active, 135-residue NH_2-terminal fragment by autodigestion of pepsin, that corresponds approximately to one domain of pepsin and may be equivalent to one subunit of retroviral protease. The pepsin fragment showed a lowered

k_{cat} for hemoglobin, but was more active than pepsin against a synthetic peptide substrate (k_{cat}/K_m was increased 50-fold).

These new linked protease dimer constructs will make it possible to introduce additional mutations into individual domains of the enzyme. Their study should greatly facilitate further investigation of the structure/function relationships of this protein which is essential for viral replication and, in the human retroviruses, an important target for antiviral drug design.

Acknowledgments—We would like to thank Dr. Joyce Jentoft and H. Burstein for valuable discussions and P. Roat for excellent technical assistance.

Addendum—After our manuscript was submitted, DiIanni *et al.* (DiIanni, C. L., Davis, L. J., Holloway, M. K., Herber, W. K., Darke, P. L., Kohl, N. E., and Dixon, R. A. F. (1990) *J. Biol. Chem.* **265,** 17348–17354) reported construction of linked dimers of HIV PR in which subunits were joined by 1 or 2 Gly residues. The one dimer characterized had a GG linker and displayed similar or reduced activity on different substrates when compared with the nonlinked HIV protease.

REFERENCES

Alexander, F., Leis, J., Soltis, D. A., Crowl, R. M., Danko, W., Poonian, M. S., Pan, Y.-C. E., and Skalka, A. M. (1987) *J. Virol.* **61,** 534–542

Baker, T. A., Grossman, A. D., and Gross, C. A. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 6779–6783

Bianchi, M., Boigegrain, R. A., Castro, B., and Coletti-Previero, M.-A. (1990) *Biochem. Biophys. Res. Commun.* **167,** 339–344

Bird, R. E., Hardman, K. D., Jacobson, J. W., Johnson, S., Kaufman, B. M., Lee, S-M., Lee, T., Pope, S. H., Riordan, G. S., and Whitlow, M. (1988) *Science* **423,** 423–442

Casadaban, M. J., and Cohen, S. N. (1980) *J. Mol. Biol.* **138,** 179–207

Crowl, R., Seamans, C., Lomedico, P., and McAndrew, S. (1985) *Gene (Amst.)* **38,** 31–38

DeChiara, T. M., Erlitz, F., and Tarnowski, S. J. (1986) *Methods Enzymol.* **119,** 403–415

Hallewell, R. A., Laria, I., Tabrizi, A., Carlin, G., Getzoff, E. D., Tainer, J. A., Cousens, L. S., and Mullenbach, G. T. (1989) *J. Biol. Chem.* **264,** 5260–5268

Harlow, E., and Lane, D. (1988) *Antibodies: A Laboratory Manual,* pp. 632–633, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Jaskólski, M., Miller, M., Rao, J. K., Leis, J., and Wlodawer, A. (1990) *Biochemistry* **29,** 5889–5898

Jones, A. T. (1978) *J. Appl. Crystallogr.* **11,** 268–272

Kotler, M. Katz, R. A., and Skalka, A. M. (1988a) *J. Virol.* **62,** 2696–2700

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

Kotler, M., Danho, W., Katz, R. A., Leis, J., and Skalka, A. M. (1989) *J. Biol. Chem.* **264,** 3428–3435

Laemmli, U. K. (1970) *Nature* **227,** 680–685

Leis, J., Bizub, D., Weber, I. T., Cameron, C., Katz, R., Wlodawer, A., and Skalka, A. M. (1989) in *Current Communications in Molecular Biology: Viral Proteinases as Targets for Chemotherapy* (Kräusslich, H., Oroszlan, S., and Wimmer, E., eds) pp. 235–243, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Loeb, D. D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S. E., and Hutchinson, C. A., III (1989) *Nature* **340,** 397–400

Manne, V., Bekesi, E., and Kung, H. F. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 376–380

Miller, M., Jaskólski, M., Mohana Rao, J. K., Leis, J., and Wlodawer, A. (1989) *Nature* **337,** 576–579

Morinaga, Y., Franceschini, T., Inouye, S., and Inouye, M. (1984) *Biotechnology* **2,** 636–639

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

Schneider, J., and Kent, S. B. H. (1988) *Cell* **54,** 363–368

Strickler, J. E., Gorniak, J., Dayton, B., Meek, T., Moore, M., Magaard, V., Malinowski, J., and Debouch, C. (1989) *Proteins* **65,** 139–154

Tang, J., James, M. N. G., Hsu, I. N. Jenkins, J. A., and Blundell, T. L. *Nature* **271,** 618–621

Tomasselli, A. G., Olsen, M. K., Hui, J. O., Staples, D. J., Sawyer, T. K., Heinrikson, R. L., and Tomich, C.-S. (1990) *Biochemistry* **29,** 264–269

Weber, I. T. (1990a) *Proteins* **7,** 172–184

Weber, I. T. (1990b) *J. Biol. Chem.* **265,** 10492–10496