Vary, Eds. (American Society for Microbiology, Washington, D.C. 1978), p. 131.
11. W. C. Johnson, C. P. Moran, Jr., R. Losick, *Nature (London)* **302**, 800 (1983).
12. C. D. B. Banner, C. P. Moran, Jr., R. Losick, *J. Mol. Biol.* **168**, 351 (1983).
13. W. G. Haldenwang and R. Losick, *Nature (London)* **282**, 256 (1979).
14. J. L. Wiggs, M. Z. Gilman, M. Chamberlin, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2762 (1981); M. Z. Gilman and M. Chamberlin, *Cell* **35**, 285 (1983).
15. W. G. Haldenwang, N. Lang, R. Losick, *Cell* **23**, 615 (1981).
16. R. Lardick, V. Vaughn, E. T. Lau, R. A. Van Bogeln, J. W. Erickson, F. C. Neidhardt, *ibid.* **38**, 175 (1984); A. Grossman, J. W. Erickson, C. Gross, *ibid.*, p. 383.
17. J. Pero, J. Nelson, T. D. Fox, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1589 (1975); J. Duffy and E. P. Geiduschek, *Nature (London)* **270**, 28 (1977); R. Tjian and J. Pero, *ibid.* **262**, 753 (1976); G. Spiegelman and H. R. Whiteley, *J. Biol. Chem.* **249**, 1476 (1974); C. Talkington and J. Pero, *Virology* **83**, 365 (1978).
18. M. S. Rosenkrantz, S. Rong, A. L. Sonenshein, in preparation; J. Trempy, D. Plummer, W. G. Haldenwang, in preparation.
19. J. H. Miller, W. S. Reznikoff, A. E. Silverstone, K. Ippen, E. R. Signer, J. R. Beckwith, *J. Bacteriol.* **104**, 1273 (1970); M. J. Casadaban and S. N. Cohen, *J. Mol. Biol.* **138**, 198 (1980); C. E. Donnelly and A. L. Sonenshein, in *Molecular Cloning and Gene Regulation in Bacilli*, A. T. Ganesan, S. Chang, J. A. Hoch, Eds. (Academic Press, New York, 1984), p. 63.
20. P. Zuber and R. Losick, *Cell* **35**, 275 (1983).
21. D. S. Goldfarb, S.-L. Wong, T. Kudo, R. Doi, *Mol. Gen. Genet.* **191**, 319 (1983); S.-L. Wong, C. Price, D. S. Goldfarb, R. Doi, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1184 (1984).
22. M. Stephens, N. Lang, K. Sandman, R. Losick, *J. Mol. Biol.* **176**, 333 (1984).
23. S. Mongkolsuk and P. Lovett, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3457 (1984).
24. S. Goldrick and P. Setlow, *J. Bacteriol.* **155**, 1459 (1983); J. Mason and P. Setlow, *ibid.* **157**, 931 (1984).
25. H. C. Wong, H. E. Schnepf, H. R. Whiteley, *J. Biol. Chem.* **258**, 1960 (1983); A. Klier, F. Fargotte, G. Rapoport, *EMBO J.* **1**, 791 (1982).
26. H. Hirochika, Y. Kobayashi, F. Kawamura, H. Saito, *J. Bacteriol.* **146**, 494 (1981); J. Weir, E. Dubnau, N. Ramakrishna, I. Smith, *ibid.* **157**, 405 (1984).
27. E. Ferrari, D. J. Henner, J. A. Hoch, *ibid.* **146**, 430 (1981).
28. P. J. Youngman, J. B. Perkins, R. Losick, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2305 (1983).
29. P. Tomich, F. Y. An, D. B. Clewell, *J. Bacteriol.* **141**, 1366 (1980).
30. J. B. Perkins and P. Youngman, *Plasmid* **12**, 119 (1984).
31. P. Youngman, J. B. Perkins, K. Sandman, in *Genetics and Biotechnology of Bacilli*, A. T. Ganesan and J. A. Hoch, Eds. (Academic Press, New York, 1984).
32. P. Youngman, J. B. Perkins, R. Losick, *Mol. Gen. Genet.* **195**, 424 (1984).
33. N. Kleckner, J. Roth, D. Botstein, *J. Mol. Biol.* **116**, 125 (1977).
34. J. Kuner and D. Kaiser, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 425 (1981).
35. B. Ely and R. Croft, *J. Bacteriol.* **149**, 620 (1982).
36. J. E. Beringer, J. L. Beyon, A. V. Buchanan-Wallaston, A. W. B. Johnston, *Nature (London)* **276**, 633 (1978); R. Simon, U. Priefer, A. Puhler, *Biotechnology* **1**, 784 (1983).
37. M. Casadaban and S. Cohen, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4530 (1979).
38. V. Bellofatto, L. Shapiro, D. Hodgson, *ibid.* **81**, 1035 (1984); L. Kroos and D. Kaiser, personal communication.
39. P. Youngman, J. B. Perkins, R. Losick, *Plasmid* **12**, 1 (1984).
40. C. Kenyon and G. C. Walker, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2816 (1980).
41. The use of MUG on agar plates as an indicator substrate for β-galactosidase was suggested to us by David Gelfand of the Cetus Corporation.

RESEARCH ARTICLE

Redesigning Trypsin: Alteration of Substrate Specificity

Charles S. Craik, Corey Largman, Thomas Fletcher
Steven Roczniak, Philip J. Barr
Robert Fletterick, William J. Rutter

Studies relating the structure of the serine proteases to their functions as catalysts of amido and ester bond cleavage have influenced current ideas concerning the mechanism of enzyme action. X-ray crystallographic analyses of eukaryotic and prokaryotic members of this family of enzymes (*1, 2*) have revealed a common three-dimensional structure. The results of structural analysis and ancillary data on selective chemical modification have revealed the location and function of essential residues for zymogen activation (*3*) and catalysis (*4*). The catalytic mechanism appears to be conserved among the members of this extensive family. The diverse activities of the serine proteases are the result of the different sets of amino acids that are utilized by each enzyme for substrate binding.

The genes and complementary DNA's (cDNA's) for most of the pancreatic serine proteases, including trypsin (*5, 6*), chymotrypsin (*7*), elastase (*8*), and kallikrein (*9*), have been cloned and sequenced. We elected to substitute, delete, or insert the codons for specific amino acids in the coding regions of these genes. Subsequent expression of the modified genes and characterization of the amino acid–altered gene products provide means to evaluate critically the postulates about specific amino acid side chains as they relate to catalysis, substrate specificity, and zymogen activation of the serine proteases. Using this system, we have designed and partially characterized three trypsin mutants that influence substrate specificity. Two of the mutants may assume a zymogen-like conformation in the absence of substrate.

Modeling wild-type and mutant rat trypsins. Although the three-dimensional structure for rat pancreatic trypsin II is not known, the primary structure has 74 percent identity with bovine cationic trypsin (*5, 6*), whose three-dimensional crystal structure is known (*1*). Of the 57 differences in amino acids, 48 are at the solvent-accessible surface, 5 are conservative changes within the internal structure, and the other 4 (88 Ser→Ile; 112 Ala→Val; 181 Phe→Val; 183 Ala →Val) are nonconservative (*10*). Active site residues His^{57}, Asp^{102}, and Ser^{195} of rat and bovine trypsin are conserved and located in regions with sequence identity. Similarly, Asp^{189} at the base of the substrate-binding pocket, which is thought to confer the substrate specificity for arginyl and lysyl side chains upon trypsin, is present in the rat sequence. The glycine residues at positions 216 and 226, which seem to permit entry of large amino acid side chains into the hydrophobic pocket, are also conserved. Indeed, when the structure of rat trypsin is compared with that of bovine trypsin complexed with pancreatic trypsin inhibitor (*11*) (with Lys^{15} of the inhibitor in the substrate-binding pocket), there are no substitutions within 7.6 Å of the substrate side chain. We therefore used the bovine trypsin structures complexed with either the pancreatic trypsin inhibitor or benzamidine [an arginine analog that is an accurate structural model for arginine (*12, 13*)] as a model for the rat trypsin complexed with lysine and arginine substrates, respectively.

C. S. Craik, S. Roczniak, and W. J. Rutter are members of the Hormone Research Institute and the Department of Biochemistry and Biophysics, and R. Fletterick is a member of the Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143. P. J. Barr is a member of Chiron Corporation, 4560 Horton Street, Emeryville, California 94608. C. Largman and T. Fletcher are members of the Biochemistry Research Laboratories, Martinez Veterans Administration Medical Center, Martinez, California 94553. Correspondence should be sent to C. S. Craik.

Comparison of the related tertiary structures of trypsin and elastase (*14*) suggests that glycine residues 216 and 226 are appropriate initial targets for probing the structural basis of the substrate specificity of trypsin. The analogs of Gly^{216} and Gly^{226} in elastase are Val^{216} and Thr^{226} (*15*), which restrict the substrate specificity of elastase to small hydrophobic amino acids. Therefore, we expected that the substrate binding properties of trypsin would be altered if these residues were modified.

Abstract. *A general method for modifying eukaryotic genes by site-specific mutagenesis and subsequent expression in mammalian cells was developed to study the relation between structure and function of the proteolytic enzyme trypsin. Glycine residues at positions 216 and 226 in the binding cavity of trypsin were replaced by alanine residues, resulting in three trypsin mutants. Computer graphic analysis suggested that these substitutions would differentially affect arginine and lysine substrate binding of the enzyme. Although the mutant enzymes were reduced in catalytic rate, they showed enhanced substrate specificity relative to the native enzyme. This increased specificity was achieved by the unexpected differential effects on the catalytic activity toward arginine and lysine substrates. Mutants containing alanine at position 226 exhibited an altered conformation that may be converted to a trypsin-like structure upon binding of a substrate analog.*

If we assume that the structures of rat and bovine trypsin are identical, the three-dimensional coordinates of any amino acid at positions 216 and 226 can be modeled to determine the distances between the amino acid replacements at these sites and the rest of the protein-ligand complex. Table 1 presents a subset of these atomic distances for substitution of alanine at positions 216 and 226. Although the same binding pocket is shared by cationic substrates, the specific amino acid residues are employed differently for lysyl substrate binding than for arginyl substrate binding (Fig. 1). Both arginyl and lysyl substrates form hydrogen bonds with the water molecule 416, which is bound in the specificity pocket by hydrogen bonds to the backbone carbonyl oxygens of Trp^{215} and Val^{227}. The lysine side chain is bound in the specificity pocket by direct hydrogen bonds with Ser^{190} and by indirect hydrogen bonds (mediated through the water molecule 414) with Asp^{189}.

In contrast, the longer arginine side chain extends deeper into the binding pocket than the lysine side chain, displacing the water molecule 414 and forming a cyclic network of direct hydrogen bonds with Asp^{189}. The guanidinium group fills the base of the substrate-binding pocket. Small changes in the tertiary structure of the trypsin substrate complex will result from adding methyl groups at the interface between the ligand and the specificity pocket. Kinetic analysis of the mutant enzymes should show effects on the Michaelis constant (K_m), which relates to binding affinity, and on the catalysis constant (k_{cat}), which is a measure of catalytic activity. The catalytic activity is dependent on the substrate alignment, which is determined in part by the specificity pocket. The trypsin (216 Gly→Ala) mutant should show relatively better specificity for arginine because the water molecule 414 that is presumably displaced by Ala^{216} (Table 1 and Figs. 1 and 2) does not take part directly in binding of arginine. The (226 Gly→Ala) mutant should show relatively enhanced lysine specificity because there is more space to accommodate the steric conflicts of Asp^{189}, the methyl group at position 226, and the substrate at the base of the pocket (Figs. 1 and 2).

Table 1. Contacts (<3.8 Å) between alanine and the rest of the protein-ligand complex in the (216 Gly→Ala) and (226 Gly→Ala) trypsin mutants.

Substituted β carbon (position)	Atom contacted		Binding interactions	Distance (Å)
	Protein	Ligand		
216	$(H_2O)^{414}$		Lysyl	1.9
		Benzamidine C-1		3.3
		Benzamidine C-7		2.7
		Benzamidine N-1		2.3
		Lysine ε carbon		3.0
		Lysine ζ nitrogen		2.9
226	Asp^{189} δ-2 oxygen		Arginyl	2.8
	$(H_2O)^{416}$		Lysyl and arginyl	3.2
	Tyr^{228} hydroxyl			3.0
	Val^{227} nitrogen			3.2
	Benzamidine N-2			3.1

Expression and secretion. For the expression and mutagenesis experiments, a modified rat pancreatic trypsinogen gene was constructed by fusing the 5′ portion of the trypsin II gene (*5*) to the 3′ portion of a trypsin II cDNA clone (Fig. 3A). This resulted in a full-length copy of the trypsinogen coding sequence including the signal peptide, which was not present in the cDNA clone. The construction contained the first intervening sequence of the trypsin gene at amino acid position −2 within the signal peptide for evaluation of the possible role of the intron in mammalian cell expression.

We selected a mammalian system to maximize the probability that native folding of the protein with correct disulfide formation and secretion of the zymogen would occur. The early gene containing the T-antigen coding region of simian virus 40 (SV40) was replaced by the trypsin sequences (Fig. 3B), and the recombinant viruses (LSV-trypsin) were propagated to large numbers of copies in transformed monkey kidney cells (COS) as described for insulin (*16*) and the hepatitis B surface antigen (*17*). Northern blot analysis of polyadenylated RNA from COS cells infected with the recombinant virus revealed a major band of 910 nucleotides (Fig. 4A), which is the expected size for a transcript initiating at the SV40 early promoter start site and terminating at the trypsin polyadenylation site. The amount of trypsin-specific messenger RNA approximated that expressed in adult rat pancreas cells (Fig. 4A). A protein presumed to be trypsinogen could be precipitated from the extracellular medium of the infected COS cells by polyclonal antibodies to rat anionic trypsinogen (Fig. 4B). The immunoprecipitated material from the infected COS cells had a molecular weight of 30,000, as is expected for rat trypsinogen. A protein having these characteristics was also present in the extracellular medium of a transformed pancreatic tumor cell line, AR4-2J, that is known to express trypsinogen (*18*) (Fig. 4B). This indicates that the rat trypsin genes could be expressed heterologously and that the trypsin II gene encodes the anionic form of rat pancreatic trypsin. The identity and the amount of the secreted trypsinogen was confirmed. The results of radioimmunoassay, Western blot analysis (*19*), and active-site titration with 4-methylumbelliferyl-*p*-guanidinobenzoate (*20*) determined the concentration of the se-

creted trypsinogen to be approximately 1 μg per 10^6 cells, which is equivalent to more than 2×10^7 molecules per cell per 10 days. Similar expression was achieved for the mutant trypsinogens. This approximates the rate of accumulation of trypsinogen in the late embryonic pancreas (*21*) and was unaffected by removal of the intron from the trypsin minigene by site-specific mutagenesis (Fig. 3A).

Site-specific mutagenesis. Conversion of wild-type codons GGC for Gly^{216} and GGT for Gly^{226} to GCC for Ala^{216} and GCT for Ala^{226} (G, guanine; C, cytosine; T, thymine) was accomplished by oligonucleotide-directed, site-specific mutagenesis by methods adapted from Zoller and Smith (*22*). The full-length trypsin minigene was cloned into bacteriophage M13 (MP8) (*23*), and the single-stranded recombinant viral template was isolated. Hybridization of the template with an appropriate synthetic primer (see Fig. 5A) produced a partial duplex that could be extended by T4 DNA polymerase to form covalently closed DNA circles in the presence of the four nucleoside triphosphates and T4 DNA ligase. However, full-length circles were rarely present presumably because of secondary structure that blocked the replication of T4 DNA polymerase in vitro.

To increase the population of full-length circles and to increase the relative frequency of mutant DNA recovery, we exploited the single-stranded DNA binding properties of T4 gene 32 protein (*24*). Addition of this protein to the synthesis reaction permits DNA synthesis through regions of secondary structure that otherwise prevent full-length polymerization of the DNA duplex (Fig. 5B). The efficiency of this mutagenesis procedure varied but was usually between 20 and 40 percent. The mutant trypsinogen–M13 DNA was isolated and characterized as described (see legend to Fig. 5B). Mutant trypsin sequences in M13 bacteriophage were then recombined with the LSV-trypsin expression vector in an *Escherichia coli* strain JM101 bacterial host to permit homologous recombination between the common trypsin sequences. The LSV-trypsin recombinants were then screened with the isotopically labeled mutagenic oligonucleotide for the site-specific mutation and used for the expression of mutant trypsins (Fig. 5A) (*25*).

To ensure the accuracy of the process, we sequenced the entire mutant DNA template (*26*) and found that it corre-

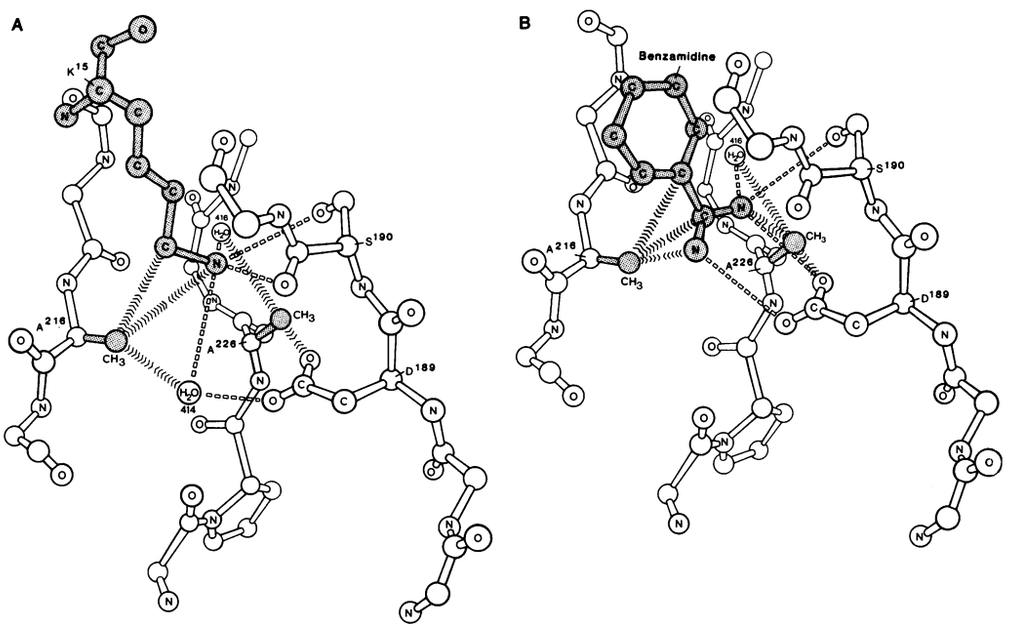


Fig. 1. Substrate-binding pocket of mutant trypsins. The backbone is shown for the chain segments, including amino acids 189 to 191, 214 to 217, and 224 to 227. Atomic coordinates for the drawing were obtained from the Brookhaven Protein Data Bank (entry set 3PTB). Both Ala^{216} and Ala^{226} as well as the ligands, Lys^{15}, and benzamidine are shaded. The rectangular boxes denote hydrogen bonds; the arrowheads denote selected short contacts (<3.8 Å). A, D, S, and K are the one-letter codes for alanine, aspartic acid, serine, and lysine, respectively. (A) The Lys^{15} of pancreatic trypsin inhibitor bound in the trypsin specificity pocket. (B) Benzamidine bound in the trypsin specificity pocket.

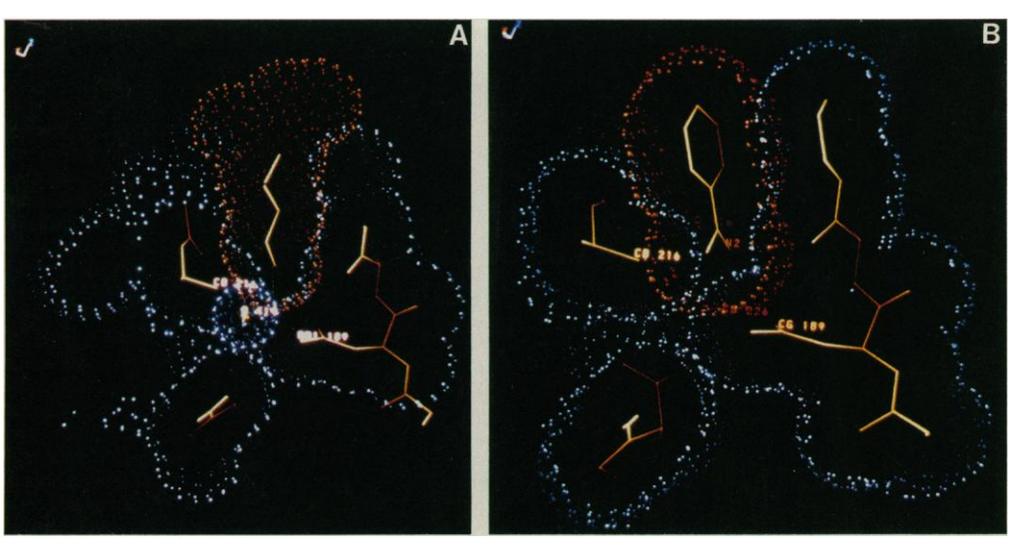


Fig. 2. Space filling representations of the substrate binding pocket of mutant trypsins. The atoms are shown with their van der Waals contact surfaces for the chain segments depicted in Fig. 1. The enzyme atomic surface is blue, and the ligand atomic surface is red. (A) View similar to that of Fig. 1A. The van der Waals surface of the β carbon of Ala^{216} overlaps that of the water molecule 414 and of Lys^{15}. (B) View similar to that of Fig. 1B. The van der Waals surface of the β carbon of Ala^{226} overlaps that of the guanidinium group of benzamidine. The van der Waals overlap of the γ carbon of Asp^{189} with the ligand is appropriate because a hydrogen bond is formed at this contact.

sponded to the sequence of the wild-type enzyme except at the point of mutation. As a further test of accuracy, the codon for Ala^{226} of the mutant trypsin was reverted back to the codon for Gly^{226}. The resultant protein had enzymatic properties similar to those of the wild-type protein.

Purification and analysis of wild-type and mutant trypsinogens. The trypsinogens were isolated from COS cell supernatants by affinity chromatography with the use of a column of pancreatic trypsin inhibitor–Sepharose or immobilized immunoglobulin G specific to rat anionic trypsinogen. Neither the mutant nor wild-type proteins initially exhibited activity when assayed with fluorogenic substrates, indicating that the enzymes had been purified as zymogens. After treatment with the natural activating enzyme enteropeptidase (*27*), the purified proteins catalyzed the hydrolysis of D-Val-Leu-Arg-aminofluorocoumarin. The eluted material before and after enteropeptidase activation migrated on sodium dodecyl sulfate (SDS) denaturing gels as trypsinogen and trypsin, respectively. This catalytic activity of the active enzymes can be blocked rapidly and completely by Phe-Ala-Arg-chloromethylketone (FARCK), an irreversible trypsin inhibitor (*28*). These results show that the zymogen form of the protease is expressed by the COS cells and can be purified readily from the tissue culture medium.

Nondenaturing gel electrophoresis (*29*) was used to analyze the wild-type and mutant trypsin products of enteropeptidase activation and to determine the effect of the substitutions of alanine for glycine on the conformation of the proteins. Trypsin migrated slower than trypsinogen (Fig. 6), perhaps partly be-

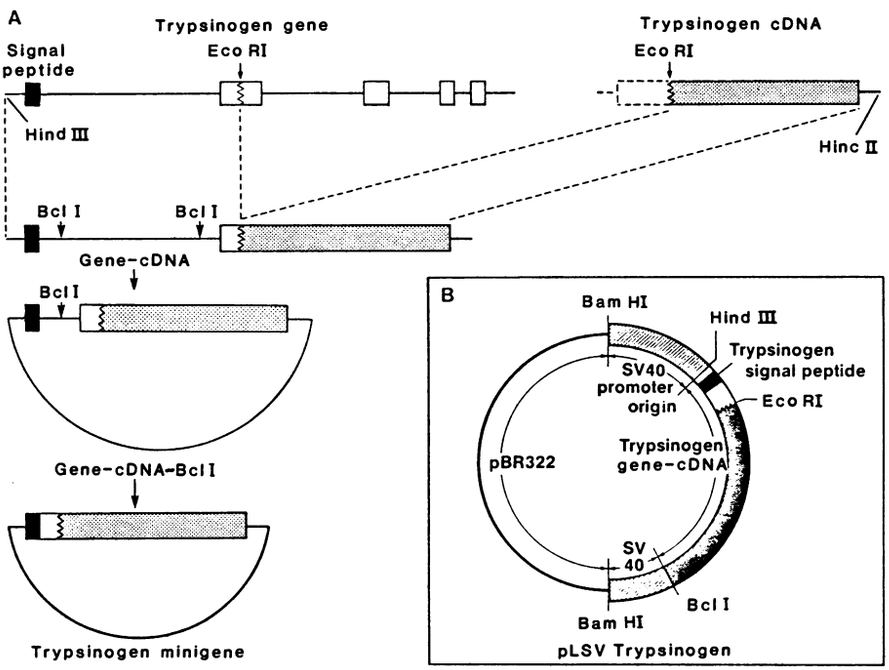


Fig. 3. Construction of an SV40-trypsinogen recombinant vector to express the trypsinogen gene. (A) Construction of a hybrid gene–cDNA sequence that encodes the entire rat pretrypsinogen sequence (*34*). The trypsinogen gene–cDNA–Bcl I fusion product was used for expression studies that required an intron, and the trypsinogen minigene was used for expression studies that did not require an intron. (B) Construction of pLSV trypsinogen (*35*) for transfection into permissive monkey COS cells (*36*).

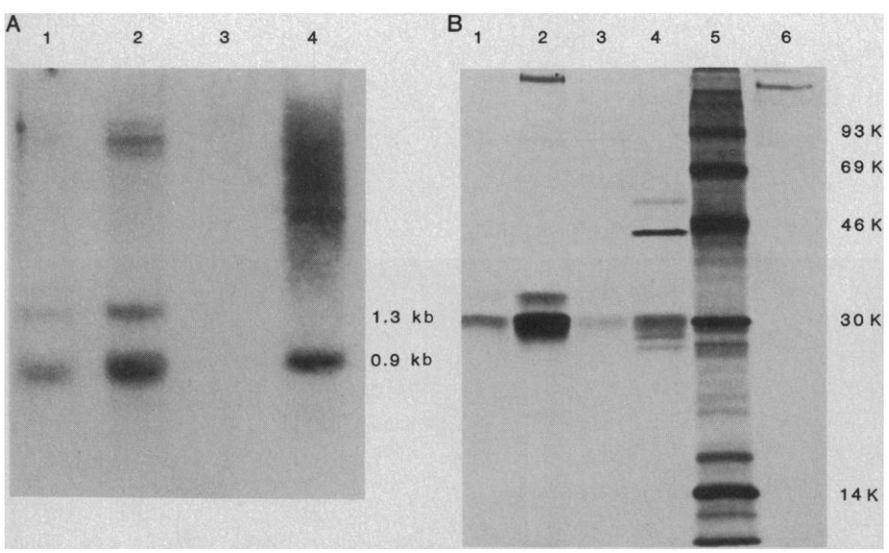


Fig. 4. Levels of expression of heterologously expressed rat pancreatic trypsinogen. (A) Transcription of the LSV trypsinogen recombinant. Confluent monolayers of COS cells were infected with 10 to 100 plaque-forming units of the LSV-trypsinogen virus stock per cell and then incubated for either 48 hours (lane 1) or 60 hours (lane 2) at 37°C. Control cells (lane 3) were infected with an LSV vector carrying the human insulin coding sequences instead of rat trypsinogen sequences. Cytoplasmic polyadenylated RNA (1 μg per lane) prepared as described (*37*) was denatured in 10 m*M* methyl mercury, subjected to electrophoresis through a 5 m*M* methyl mercury–1.5 percent agarose gel, and transferred to nitrocellulose paper. The resulting blots were hybridized to ^{32}P-labeled trypsinogen cDNA probes at 42°C in 50 percent formamide, 5× standard saline citrate [SSC; 1 × SSC is 0.15*M* NaCl and 0.015*M* sodium citrate (*p*H 7.0)], 0.1 percent SDS, 20 m*M* Na_2HPO_4-NaH_2PO_4 buffer (*p*H 6.5), 0.1 percent each of Ficoll, polyvinylpyrrolidone, and bovine serum albumin, and 10 μg of sonicated salmon sperm DNA per milliliter. Probes were labeled with ^{32}P by nick translation (*38*) to a specific activity of at least 10^8 count/min per microgram, and 10^6 to 5 × 10^6 count/min per filter was used for each hybridization. (Lane 4) Polyadenylated RNA (1 μg) purified from an adult rat pancreas as described (*39*). (B) Synthesis of trypsinogen in LSV-trypsinogen–infected COS cells. Monolayers of 10^6 COS cells were infected with LSV-trypsinogen virus stocks for 48 hours (lane 1) or 60 hours (lane 2). Cells were grown for 12 hours in a cysteine- and methionine-depleted medium and then were treated for 6 hours with [^{35}S]cysteine and [^{35}S]methionine (10 μCi/ml). Secreted proteins were analyzed by immunoprecipitation of the medium bathing the cells (200 μl) with rabbit antibodies to rat anionic trypsinogen and by electrophoresis on 12 percent SDS-polyacrylamide denaturing gels (*40*). (Lane 3) A portion (200 μl) of the labeled medium after purification on a pancreatic trypsin inhibitor–Sepharose column (see Fig. 5A). (Lane 4) The products of the immunoprecipitation with rat anionic trypsinogen from the labeled medium of AR4-2J cells, a rat exocrine pancreas tumor cell line (*18*). (Lane 5) [^{14}C]Methylated protein molecular weight markers (Amersham). (Lane 6) Control cells as in (A).

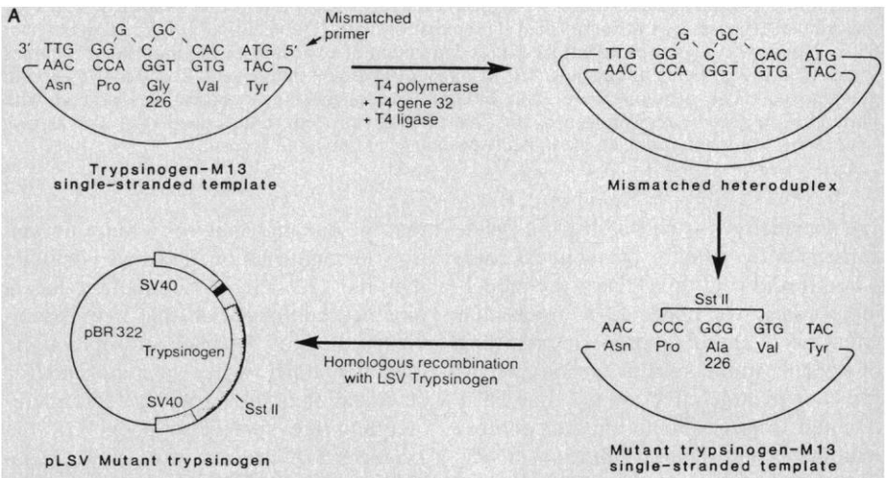
cause of the removal of the negatively charged, eight–amino acid activation peptide. The replacement of alanine for glycine at position 216 had little effect on the migration of either the zymogen or the activated form of the enzyme because these forms comigrated with wild-type trypsin. However, the migrations of trypsin (226 Gly→Ala) and (216, 226 Gly→Ala) were anomolous compared with the other activated proteins, even though the point mutations are uncharged substitutions. The activated trypsins (226 Gly→Ala) and (216,226 Gly→Ala) comigrated with trypsin upon incubation with the active site-directed inhibitor FARCK.

The amidase activities of the wild-type and mutant trypsins were assayed with the sensitive fluorometric peptide substrates D-Val-Leu-Lys-aminofluorocoumarin and D-Val-Leu-Arg-aminofluorocoumarin. Table 2 presents the kinetic measurements for the wild-type enzyme and the three mutants. The wild-type rat enzyme produced in COS cells had the same kinetic constants as the bovine enzyme (*30*). The revertant trypsin (226 Ala→Gly) had kinetic constants essentially identical to those for wild-type trypsin, which shows that the activity effects were due to the single–amino acid point mutation or mutations. Each modification produced an enzyme with a distinct set of kinetic measurements. The results showed a changed specificity (k_{cat}/K_m) of the mutant enzymes with respect to the wild type toward arginine and lysine peptide substrates: trypsin 216 Gly→Ala (arginine much greater than lysine); trypsin 226 Gly→Ala (lysine much greater than arginine); trypsin 216,226 Gly→Ala (arginine greater than lysine). However, these effects were not dominated by the K_m but by the k_{cat} values.

Discussion. Our results show that the wild-type trypsin was secreted into the extracellular medium of virally infected COS cells, that it was folded with proper disulfide formation, and that it was stable to cellular digestion and autodigestion. The problems sometimes encountered with the expression of eukaryotic genes in microbial systems were not observed; therefore, this approach seems applicable to a variety of eukaryotic and prokaryotic enzymes.

The results also show that the specificity (k_{cat}/K_m) of the mutant trypsin (216 Gly→Ala) shifted toward arginine substrates because of differential effects on k_{cat} but not on K_m. Modeling studies have shown that the methyl group of Ala^{216} interferes strongly with the binding of both arginine and lysine substrates. Whereas atoms required for binding (the binding set) of arginine are not disturbed by this additional methyl moiety, the displacement of the water molecule 414 disturbs the binding set of lysine. This may cause a slight distortion of the seating of lysine in the binding pocket and a less favorable position of the scissile bond, resulting in a lower value of k_{cat} for the lysine substrate.

For the mutant trypsin (226 Gly→Ala),



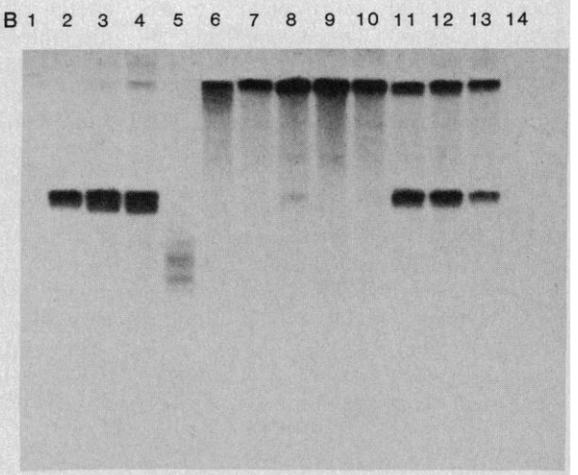


Fig. 5. Oligonucleotide-directed, site-specific mutagenesis of rat pancreatic trypsinogen. (A) Oligonucleotides were synthesized by means of solid-phase–phosphoramidite chemistry on a DNA synthesizer (Beckman System I or Applied Biosystems model 380A). The oligonucleotide 5′ CTTGG-TGTACACCGCGGGGT-TATCTGG 3′ primes the synthesis of a DNA strand encoding a trypsinogen molecule with alanine in place of glycine at position 226 (*41*). The third base change in codon 225 does not affect the amino acid encoded at this position (Pro) but provides a recognition sequence for the restriction enzyme Sst II. The covalently closed mismatched heteroduplex was used to transform competent JM101 cells (*42*). The recombinant bacteriophage from a plate containing 100 to 500 plaques was adsorbed to a nitrocellulose filter for 5 minutes, and the filter was baked at 80°C for 1 hour in a vacuum. The filter was hybridized to the isotopically labeled 5′ phosphorylated oligonucleotide used to initiate the mutagenesis reaction as described (*22*). After plaque purification, mutant trypsinogen–M13 single-stranded DNA was prepared from the recombinant bacteriophage, and the entire DNA insert was sequenced by the dideoxy chain terminating method (*26*). The mutant trypsinogen–M13 DNA was recombined with wild-type LSV trypsinogen by infecting JM101 bacteria that had been transformed with the pLSV trypsinogen plasmid (*25*). The pLSV mutant trypsinogen was isolated by hybridization to the synthetic oligonucleotide used to introduce the mutation and was verified by restriction enzyme digestion at the Sst II site and by sequence analysis. Similar methods were used to construct trypsin (216 Gly→Ala), trypsin (216,226 Gly→Ala), and trypsin (226 Ala→Gly) except that the mutagenic primer 5′ ACAGGCCA-TAGGCCCAGGAGAC 3′ was used to introduce the alanine codon change and a concomitant Xho I site at position 216 for the single and double mutants. The primer 5′ CTTGGTGTACA-CACCAGGGTTATCTGG 3′ was used to revert the alanine codon to a glycine codon at position 226. (B) Time course of primed M13 DNA synthesis with T4 DNA polymerase and T4 gene 32 protein. Samples were taken from a reaction mixture identical to that described (*42*) except that 64 μCi of [α-^{32}P]deoxyadenosine triphosphate and various amounts of T4 gene 32 protein were present. Samples were taken from the reaction mixture after various times and analyzed by alkaline 1 percent agarose gel electrophoresis (*43*). (Lanes 1 and 14) Linearized γ-^{32}P end-labeled M13 DNA. (Lanes 2, 3, and 4) Time course at 5, 20, and 60 minutes, respectively, in the absence of T4 gene 32 protein. (Lanes 5, 6, and 7) Time course at 5, 20, and 60 minutes, respectively, in the presence of 100 μg of T4 gene 32 protein per milliliter. (Lanes 8, 9, and 10) Time course at 5, 20, and 60 minutes, respectively, in the presence of 20 μg of T4 gene 32 protein per milliliter. (Lanes 11, 12, and 13) Time course at 5, 20, and 60 minutes, respectively, in the presence of 10 μg of T4 gene 32 protein per milliliter.

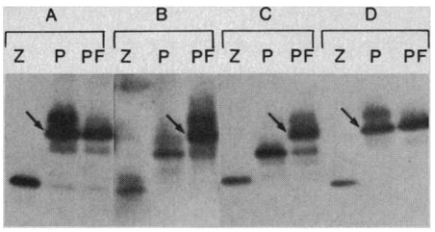


Fig. 6. Analysis of mutant trypsins by native gel electrophoresis. (A) Wild-type trypsinogen; (B) mutant trypsinogen (226 Gly→Ala); (C) mutant trypsinogen (216,226 Gly→Ala); and (D) mutant trypsinogen (216 Gly→Ala). For each protein analyzed, three samples were prepared: zymogen (unactivated enzyme, Z); porcine enteropeptidase–activated enzyme (P); and P treated with FARCK (PF). Portions of the wild-type and mutant trypsinogens were activated and assayed as described in the legend to Table 2. Inactivation of wild-type and mutant trypsins was accomplished by incubation with 2.36 m*M* FARCK for 20 minutes at 37°C. No activity could be monitored after 5 minutes of exposure to the inhibitors. Samples were subjected to electrophoresis at 4°C through a 10 percent native gel (*29*) and a 4.75 percent stacking gel. The proteins were then blotted onto nitrocellulose paper and probed with antibodies to anionic trypsinogen by the Western blot procedure described (*19*). The arrows denote the bands migrating at the expected position of activated trypsin.

the specificity was shifted toward lysine substrates by a factor of about 20 compared to the wild type. From the modeling studies, we expected a differential effect on K_m because of steric crowding of arginine in the specificity pocket (Table 1). The large effect on k_{cat} may have resulted from a shift to another conformational state. The substitution of alanine for glycine at position 226 also resulted in critically short contacts in the trypsin structure between Asp^{189} and the β carbon of Ala^{226} (Table 1). Because these contacts are not observed in the trypsinogen structure (*31*), we reasoned that the short contacts between the β carbon of Ala^{226} and the other residues of the mature enzyme may have resulted in trypsin (226 Gly→Ala) acquiring a trypsinogen-like conformation. The behavior of the enzyme on native gels supports this speculation. The altered electrophoretic behavior in the presence of the substrate analog FARCK suggests that the enzyme acquires a more normal trypsin conformation. Thus we postulate that the (226 Gly→Ala) mutant has a modified conformation that may resemble trypsinogen. In this conformation the oxyanion hole, which is important for stabilization of the tetrahedral intermediate, and the specificity pocket of the zymogen (*32*) are not in catalytic register, and the enzyme is inactive. The energy of binding the substrate drives the molecule to an active trypsin state, but the catalytic activity is still low. This can be most readily explained by an altered positioning of the substrate with respect to the catalytic apparatus so that the scissile bond is not aligned optimally. The selective effect against arginine may result from the steric crowding of Ala^{226}, which may hinder the guanidinium–aspartic acid hydrogen bonding network.

In contrast to the effects on K_m observed with the single-mutant enzymes, K_m for the double-mutant trypsin (216,226 Gly→Ala) enzyme was increased by a factor of 2 for lysyl substrates and by a factor of 15 for arginyl substrates compared to values for the wild type. Despite the apparent affinity of substrates, catalytic activity was virtually extinguished. This relatively weak effect on the apparent substrate binding seems inconsistent with the characteristics of the normal substrate binding pocket. We propose that this region of the molecule is altered in a major way to accommodate the substrates more effectively, although in a catalytically nonproductive mode (*33*). The mobility of the double mutant on denaturing gels suggests that it has a conformation similar to that of trypsin (226 Gly→Ala) and that substrate binding presumably induces native structure conformation because, in the presence of the analog FARCK, trypsin (216,226 Gly→Ala) comigrated with trypsin. However, the k_{cat} activities both decreased by a factor of 1000, which suggests that the orientation of the bound substrates to the catalytic site is even more distorted in this double-mutant enzyme than in trypsin (226 Gly→Ala).

These studies of the effects of replacing a hydrogen atom with a methyl group in the substrate-binding pocket of trypsin show that designing proteins by site-specific mutagenesis can help in elucidating the relation between protein structure and function. Further kinetic and structure analyses of mutant proteins are needed to explain fully the modified functional state.

Table 2. Kinetic measurements for wild-type and mutant trypsins. Portions of wild-type trypsinogen and the mutants (216 Gly→Ala), (226 Gly→Ala), revertant (226 Ala→Gly), and (216,226 Gly→Ala) were activated by incubation with purified porcine enteropeptidase (*27, 44*) at 37°C for 30 minutes in 28 m*M* sodium succinate and 2 m*M* $CaCl_2$ (*p*H 5.6). The concentrations of the activated trypsin ranged from 0.23 to 24 n*M* and were determined by radioimmunoassay, densitometric comparison of individual protein bands with standards of rat anionic trypsin after SDS gel electrophoresis and Western blotting (*19*), and active-site titration with 4-methylumbelliferyl-*p*-guanidinobenzoate (*20*). Assays were carried out on undefined mixtures of β trypsin and its autolysis products (*45*). Enzyme assays were performed at 37°C in 50 m*M* tris, 10 m*M* $CaCl_2$, and 1 percent dimethylformamide (*p*H 8.0) on a Perkin-Elmer LS5 spectrofluorometer. Fluorescence values were converted to micromoles of product by using a standard solution of 7-amino-4-trifluoromethylcoumarin. The substrates used were D-Val-Leu-Arg-aminofluorocoumarin (Arg) and D-Val-Leu-Lys-aminofluorocoumarin (Lys). The substrate concentrations were chosen so that the observed K_m values could be monitored over a 5- to 20-fold range. Kinetic data were analyzed by a program (*46*) that carries out a weighted linear and nonlinear least-squares regression analysis of data by using the Lineweaver-Burk and Michaelis-Menten equations, respectively. Data are reported ± experimentally determined estimated standard deviations.

Enzyme	Substrate	K_m (μ*M*)	k_{cat} (minutes^{-1})	k_{cat}/K_m (min^{-1} μ*M*$^{-1}$)	Arg/Lys*
Wild type	Arg	13.9 ± 0.2	1444 ± 11	104 ± 2	11.4 ± 0.9
	Lys	144 ± 5.0	1308 ± 20	9.1 ± 0.5	
226 Ala→Gly revertant	Arg	11.8 ± 0.5	1719 ± 42	146 ± 10	11.7 ± 2.2
	Lys	118 ± 3.0	1477 ± 18	12.5 ± 1.3	
216 Gly→Ala	Arg	393 ± 15	1017 ± 13	2.6 ± 0.2	28.9 ± 6.1
	Lys	3904 ± 427	337 ± 26	0.09 ± 0.01	
226 Gly→Ala	Arg	482 ± 26	13.0 ± 0.3	0.027 ± 0.002	0.57 ± 0.12
	Lys	3665 ± 248	172 ± 6.0	0.047 ± 0.005	
216,226 Gly→Ala	Arg	215 ± 20	1.10 ± 0.02	0.0051 ± 0.0006	2.8 ± 0.8
	Lys	331 ± 18	0.61 ± 0.01	0.0018 ± 0.0002	

*Ratio of k_{cat}/K_m values.

References and Notes

1. R. M. Stroud, L. M. Kay, R. E. Dickerson, *J. Mol. Biol.* **83**, 185 (1974); R. Huber *et al.*, *ibid.* **89**, 73 (1974).
2. W. Bode, Z. Chen, K. Bartels, C. Kutzbach, G. Schmidt-Kastner, H. Bartunik, *ibid.* **164**, 237 (1983); L. T. Delbaere, W. L. Hutcheon, M. N. G. James, W. E. Thissen, *Nature (London)* **257**, 758 (1975).
3. R. M. Stroud, A. A. Kossiakoff, J. L. Chambers, *Annu. Rev. Biophys. Bioeng.* **6**, 177 (1977); W. Bode, P. Schwager, R. Huber, *J. Mol. Biol.* **118**, 99 (1978).
4. J. Kraut, *Annu. Rev. Biochem.* **46**, 331 (1977); T. A. Steitz and R. G. Shulman, *Annu. Rev. Biophys. Bioeng.* **11**, 419 (1982).
5. C. S. Craik, Q.-L. Choo, G. H. Swift, C. Quinto, R. J. MacDonald, W. J. Rutter, *J. Biol. Chem.* **259**, 14255 (1984). A cDNA clone nearly identical to that described for the trypsin II cDNA (*6*) was used for this construction. The clone was isolated from a pancreatic cDNA library by hybridization to an isotopically labeled synthetic oligonucleotide, 5′ CCACA-TAACCCTGGTGTGTACACCAAG 3′, which encodes amino acids 222 to 230 (chymotrypsin residue numbering). Discrepancies between the original clone and the clone used for the construction were third–base pair changes in the codons for amino acids Asn^{100} (ACC→AAT), Thr^{175} (ACC→ACT), and Gly^{193} (GGC→GGT) and did not alter the amino acid sequence.
6. R. J. MacDonald, S. J. Stary, G. H. Swift, *ibid.* **257**, 9724 (1982).
7. G. I. Bell, C. S. Craik, C. Quinto, W. J. Rutter, *ibid.* **259**, 14265 (1984).
8. G. H. Swift *et al.*, *ibid.*, p. 14271; R. J. MacDonald *et al.*, *Biochemistry* **21**, 1453 (1982).
9. A. J. Mason, B. A. Evans, D. R. Cox, J. Shine, R. I. Richards, *Nature (London)* **303**, 300 (1983); G. H. Swift, J.-C. Dagorn, P. L. Ashley, S. W. Cummings, R. J. MacDonald, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 7263 (1982).
10. The numbering system adopted is according to the standard chymotrypsinogen residue numbering scheme [B. S. Hartley, *Phil. Trans. R. Soc. London Ser. B* **257**, 77 (1970)].
11. W. Bode, P. Schwager, R. Huber, *Miami Winter Symp.* **11**, 43 (1976).
12. W. Bode, J. Walter, R. Huber, H. R. Wenzel, H. Tschesche, *Eur. J. Biochem.* **144**, 185 (1984).
13. W. Bode and P. Schwager, *J. Mol. Biol.* **98**, 593 (1975).
14. L. Sawyer *et al.*, *ibid.* **118**, 137 (1978).
15. D. M. Shotton and H. C. Watson, *Nature (London)* **225**, 811 (1970).
16. C. S. Craik, O. Laub, G. I. Bell, S. Sprang, R. Fletterick, W. J. Rutter, *UCLA Symp. Mol. Cell. Biol.* **26**, 35 (1982); O. Laub and W. J. Rutter, *J. Biol. Chem.* **258**, 6043 (1983).
17. O. Laub *et al.*, *J. Virol.* **48**, 271 (1983).
18. N. W. Jessop and R. J. Hay, *In Vitro* **16**, 212 (1980).
19. W. N. Burnette, *Anal. Biochem.* **112**, 195 (1981).
20. G. W. Jameson, D. V. Roberts, R. W. Adams, S. A. Kyle, D. T. Elmore, *Biochem. J.* **131**, 107 (1973).
21. T. G. Sanders and W. J. Rutter, *J. Biol. Chem.* **249**, 3500 (1974).
22. M. Zoller and M. Smith, *Nucleic Acids Res.* **10**, 6487 (1982).
23. J. Messing and J. Vieira, *Gene* **19**, 269 (1982).
24. B. M. Alberts and L. Frey, *Nature (London)* **227**, 1313 (1970).
25. M. Hall and C. S. Craik, unpublished results.
26. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
27. J. Baratti, S. Maroux, D. Louvard, *Biochim. Biophys. Acta* **321**, 632 (1973).
28. C. Kettner, S. Springhorn, E. Shaw, *Hoppe-Seyler's Z. Physiol. Chem.* **359**, 1183–1191.
29. B. J. Davis, *Ann. N.Y. Acad. Sci.* **121**, 404 (1964).
30. S. S. Wang and F. H. Carpenter, *Biochemistry* **6**, 215 (1967); B. Keil, in *The Enzymes*, P. D. Boyer, Ed. (Academic Press, New York, ed. 3, 1971), vol. 3, pp. 250–275; M. Zimmerman, B. Ashe, E. C. Yurewicz, G. Patel, *Anal. Biochem.* **78**, 47 (1977).
31. A. A. Kossiakoff, J. L. Chambers, L. M. Kay, R. M. Stroud, *Biochemistry* **16**, 654 (1977).
32. W. S. Bennett and R. Huber, *CRC Crit. Rev. Biochem.* **15**, 291 (1984).
33. A. Fersht, *Enzyme Structure and Mechanism* (Freeman, San Francisco, 1977), pp. 95–97.
34. Stringent size limitations of the SV40 expression vector [<2.5 kilobases (kb) of foreign DNA] prevented the expression of the 7.0-kb trypsinogen gene. A Hind III–Eco RI gene fragment encoding 22 nucleotides of the 5′ flanking region, the first exon [the signal peptide; 52 base pairs (bp)], the first intron (2600 bp at amino acid −2), and 57 bp of the second exon (which includes the activation peptide and ten amino acids) was fused to a 680-bp Eco RI–Hinc II cDNA fragment encoding the rest of the trypsinogen coding sequence plus 55 nucleotides of the 3′ untranslated region. The fused fragments were inserted into the Hind III–Hinc II sites within the polylinker of bacteriophage MP8. The 2600-bp intron was reduced to 498 bp by removing the internal Bcl I fragment from the intron and then ligating the remaining 5′ portion (192 bp) to the remaining 3′ portion (306 bp) of the intron, resulting in the trypsinogen gene–cDNA–Bcl I fusion product. The oligonucleotide 5′ TGTGGGAGCTGCTGTTGCTTTTCCC-GTG 3′, which spans 14 nucleotides of the 3′ junction of exon 1 and 14 nucleotides of the 5′ junction of exon 2, was used to delete the intron from the trypsinogen gene–cDNA–Bcl I fusion product by oligonucleotide-directed, site-specific mutagenesis (see Fig. 4), resulting in the trypsinogen minigene.
35. Plasmid pLSV trypsinogen was constructed by isolating the Hind III–Bam HI fragment [the Bam HI site is provided by the MP8 polylinker (*23*)] of either the trypsinogen gene–cDNA–Bcl I fusion product or the trypsinogen minigene and ligating it into the Hind III–Bcl I fragment of pSV40 as described for the expression of human insulin (*16*). The resulting plasmid contained the entire pBR322 plasmid, the SV40 origin of replication, and the coding information for the SV40 late genes. Nucleotides encoding most of the SV40 early genes (nucleotides 5107 to 2706) were deleted. The vector retained the early SV40 promoters, 105 noncoding nucleotides downstream from the 5′ end, and 136 nucleotides before the polyadenylation site at the 3′ end of the early gene region. The trypsinogen sequences were inserted in the sense direction relative to the early SV40 promoter and polyadenylation sites. Before transfection the pLSV trypsinogen was digested with Bam HI to remove the pBR322 sequences, and the Bam HI fragments containing the LSV-trypsinogen hybrids were self-ligated to form circular DNA (LSV-trypsinogen).
36. Y. Gluzman, *Cell* **23**, 175 (1981).
37. O. Laub and Y. Aloni, *J. Virol.* **16**, 1171 (1975).
38. P. W. J. Rigby, M. Dieckmann, C. Rhodes, P. Berg, *J. Mol. Biol.* **113**, 237 (1977).
39. J. M. Chirgwin, A. E. Przybyla, R. J. MacDonald, W. J. Rutter, *Biochemistry* **18**, 5294 (1979).
40. U. K. Laemmli, *Nature (London)* **227**, 680 (1970).
41. The 5′-phosphorylated primer was annealed to the trypsinogen-M13 single-stranded template at 55°C in 50 m*M* tris-HCl (*p*H 8.0), 20 m*M* KCl, 7 m*M* $MgCl_2$, and 0.1 m*M* EDTA.
42. The mismatched heteroduplex was made in a 100-μl reaction mixture containing a DNA duplex (50 ng/μl), deoxyguanosine triphosphate (150 μ*M*), deoxycytidine triphosphate (150 μ*M*), deoxythymidine triphosphate (150 μ*M*), adenosine triphosphate (50 μ*M*), T4 DNA polymerase (10 μg/ml) (*47*), T4 gene 32 protein (100 μg/ml) (*24*, *47*), T4 DNA ligase (5 units, Boehringer Mannheim), tris-acetate (33 m*M*, *p*H 7.8), potassium acetate (66 m*M*), magnesium acetate (10 m*M*), and dithiothreitol (5 m*M*). The reaction was stopped by addition of ethylene diamine tetraacetic acid (final concentration, 10 m*M*) and SDS (final concentration, 0.2 percent). Serial dilutions of the reaction mixture were used to transform competent JM101 cells.
43. C.-C. Huang and J. E. Hearst, *Anal. Biochem.* **103**, 127 (1980).
44. J. W. Brodrick, C. Largman, M. W. Hsiang, J. H. Johnson, M. C. Geokas, *J. Biol. Chem.* **253**, 2737 (1978).
45. D. D. Schroeder and E. Shaw, *ibid.* **243**, 2943 (1968).
46. D. V. Roberts, *Enzyme Kinetics* (Cambridge Univ. Press, Cambridge, 1977), pp. 299–306.
47. C.-C. Huang, J. E. Hearst, B. M. Alberts, *J. Biol. Chem.* **256**, 4087 (1981).

48. We thank V. Jongeneel and B. Alberts for T4 gene 32 protein and helpful discussions concerning its use; M. Hall and O. Laub for discussions on homologous recombination and expression, respectively; S. Gardell, S. Sprang, D. Agard, and R. Stroud for helpful suggestions; and A. Tramontano and L. Spector for technical assistance. Supported by NSF grant PCM830610 (to W.J.R.), NIH grants GM28520 (to W.J.R.) and AM26081 (to R.J.F.), and a Research Service Grant of the Veterans Administration (to C.L.). C.S.C. was a recipient of NIH postdoctoral fellowship GM07216.

26 November 1984; accepted 7 February 1985