Journal of Cellular Biochemistry 33:199–211 (1987)
Proteases in Biological Control and Biotechnology 183–195

Redesigning Trypsin Via Genetic Engineering

Charles S. Craik, Steven Roczniak, Stephen Sprang, Robert Fletterick, and William Rutter

Departments of Pharmaceutical Chemistry (C.S.C.) and Biochemistry and Biophysics (C.S.C., S.S., R.F., W.R.), and The Hormone Research Institute (S.R., W.R.), University of California, San Francisco, California 94143

Proteins can be regarded as the principal dynamic molecules in living cells owing to their unparalleled ability to perform an immense assortment of biological functions. These include indispensable roles in catalysis, transport and storage, coordinated motility, maintenance of structural integrity, immune protection, and metabolic regulation. The remarkable adaptability of proteins is further demonstrated by their capacity to mediate critical biological processes tailored to a variety of conditions such as extremes of pH or temperature. The functional diversity exhibited by proteins is based on an equally extraordinary structural diversity. The linear sequence of amino acids unique to each protein determines its characteristic three-dimensional form. Higher order structural diversity can be achieved by self-assembly of monomeric units into molecular macrostructures. It is a goal of protein chemists to explain the function of these molecules in terms of their structure and to derive a set of basic principles that will allow the design of proteins with novel structures and functions.

During the past 50 years the methods available for studying protein structure and function have increased greatly in sophistication. Ligand binding and catalysis can be characterized by stopped-flow or low-temperature kinetics using a variety of spectroscopic techniques. The linear amino acid sequence can be determined now by standard procedures or with greater facility by determining the nucleotide sequence of the corresponding gene or cDNA. X-ray diffraction analysis of proteins in single crystals yields precise images of the three-dimensional structure of a protein. Crystal structure determinations of enzymes complexed with natural inhibitors or pseudosubstrates provide a detailed view of enzyme-substrate interactions. These studies can be supplemented by high-resolution spectral analyses of the behavior of atoms (for example, protons) of the protein in real time. While these studies provide some insight into the mechanism involved in protein activity, deeper understanding requires

Received July 10, 1986; accepted October 15, 1986.

the identification of individual amino acid residues, and specifically the time-dependent function of the atoms involved in the crucial interactions. For this purpose, two strategies have emerged. The first involves selective chemical modification of residues within the protein. Although there is much useful information that has been derived from this approach, the modifications are rarely specific enough to be decisive. It is likely that several identical amino acid side chains will interact with a reagent under a given set of circumstances to produce proteins that are heterogeneous with respect to the extent of derivatization. Further, chemical modification usually results in a change in mass or charge, which can influence the active site. Finally, reagents are not available that allow the modification of all the residues within the confines of the active site.

The second strategy involves studies of the effects of amino acid replacements in the primary sequence. Historically, the biophysical analysis of naturally occurring mutant proteins has proven to be invaluable in studying the wild-type molecule. Pioneering studies on hemoglobin variants by Perutz and his coworkers were essential to unraveling the functional complexities of hemoglobin [1]. Additional critical information can be obtained from sequence-function comparisons of proteins from different species or cell types. These comparisons pinpoint conserved residues that could be required for function and variable residues that are likely to play no critical role. Recent advances in molecular biology have obviated our reliance on naturally occurring homologs, which has allowed us to deliberately modify the amino acid sequence by a genetic approach.

Using the methods of molecular genetics, it is possible to modify the genetic sequence that encodes the protein of interest. To exploit the molecular genetic approach, four conditions must be met (Fig. 1): 1) the gene or cDNA (natural or synthetic) encoding the protein must be available, and its sequence must be completely characterized; 2) a method for introducing alterations into the DNA sequences must be established; 3) an expression system must be available that can produce sufficient quantities of the protein for biochemical and structural analyses; and 4) the modified proteins must be readily purified and assayed.

It is now possible to isolate and determine the linear order of monomeric units of virtually any DNA, RNA, or protein component or product of the cell [2–5]. If the complete amino acid sequence of the protein to be studied is known, an alternative to cloning the gene or cDNA is to synthesize a DNA sequence that encodes the protein.

A single nucleotide of the DNA coding sequence can be replaced to introduce a single amino acid substitution in the protein, or the sequence encoding several amino acids can be added, subtracted, or modified. Indeed, domains of homologous molecules can be exchanged with those having similar or different structures or functions. Efficient application of this so-called "reverse-genetics" [6] in the analysis of proteins requires prior knowledge of the relative importance of a given amino acid with respect to the rest of the protein. In cases in which accurate crystallographic determinations have been made, the three-dimensional structure of a protein may suggest functional roles for individual amino acids. Computer graphics-assisted model-building studies can then suggest amino acid residue replacements that may predictably modify the activity of the protein. If direct observations fail to localize possible functional residues in the protein sequence, an approach of random mutagenesis [7] involving DNA replacements, insertions, or deletions throughout the coding region can be applied to identify important regions of the protein by determining the

GENETIC MODIFICATION OF PROTEIN STRUCTURE AND FUNCTION

CHARACTERIZATION OF GENE OF INTEREST

MUTATION

Site-Specific Mutagenesis
Cassette Mutagenesis
Random Mutagenesis

EXPRESSION

Bacteria
Yeast
Mammalian Cells

ANALYSIS

Purification
Functional Characterization
Structural Characterization

Fig. 1. Genetic modification of protein structure and function: A general approach.

structural and functional consequences of the mutation(s). In principle, virtually all of the amino acids in the protein are subjects for change by mutation, potentially creating an enormous number of mutants—more than can be profitably analyzed. Selection or screening techniques are helpful and can be devised to identify mutants that increase or decrease the activity or some other property of the protein. The power of this approach is that unsuspected chemical relationships among functional groups of the protein can be discovered. However, it is difficult to obtain mutations in a given region, particularly if a specific replacement is desired.

As more knowledge about proteins accumulates, individual amino acids become potential candidates for modification. For these cases site-directed mutagenesis allows the evaluation of specific structure/function postulates. However, in cases in which it is difficult to predict which amino acid substitution will provide the most information, a more empirical approach can be adopted whereby all possible amino acid substitutions are separately introduced at the target site. An efficient method for generating multiple mutations at a defined location is "cassette mutagenesis" [8]. This method uses a pool of synthetic duplex DNA cassettes (10–25 bp) each of which contains a different codon at the same position to introduce the mutations. All 19 amino acid variations can be readily created this way, and the activity of the various single-site mutants can be compared to one another.

Oligonucleotide-directed site-specific mutagenesis offers a convenient method for introducing a particular amino acid substitution at a unique position (Fig. 2). The successful application of this method requires that the DNA sequence to be mutagenized be cloned into a vector that is capable of autonomous replication (eg, plasmids, viral genomes) and that the target site be available in a single-stranded form. The

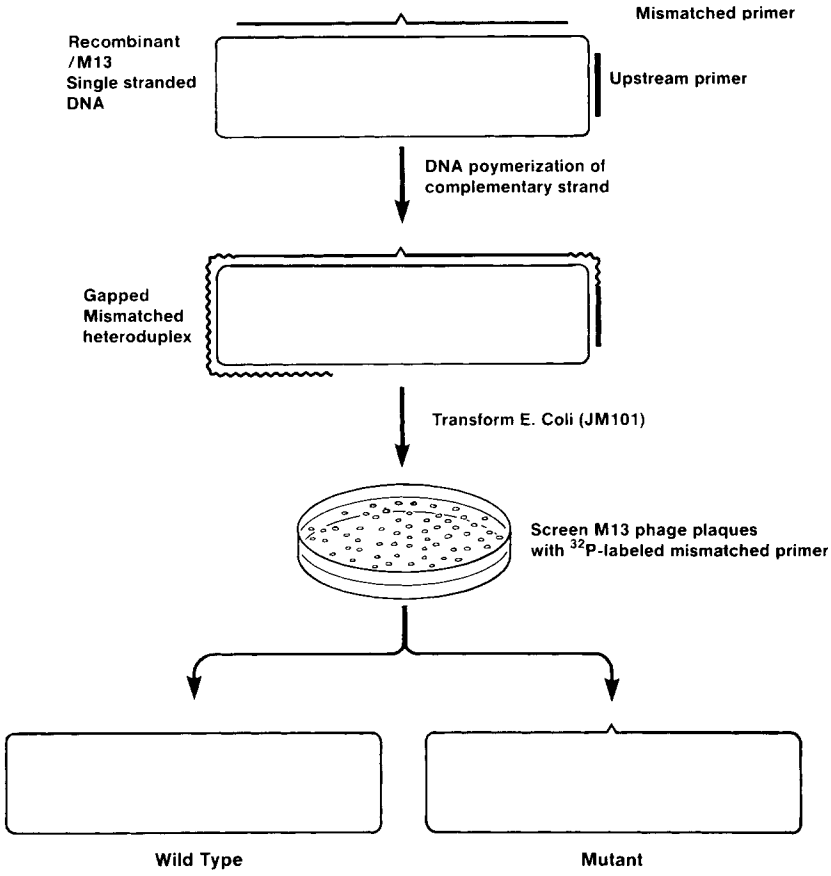

Fig. 2. General outline for oligonucleotide-directed site-specific mutagenesis. A mismatched phosphorylated oligonucleotide that contains the desired DNA mutation is annealed to a single-stranded form of the target DNA. The annealed oligonucleotide serves as a primer to initiate the synthesis of those sequences that are to remain wild type. Generation of the complete heteroduplex may be completed in vivo or in vitro. After in vivo segregation of mutant and wild-type viral replicons, the mutant template is identified, isolated, structurally characterized, and then inserted into an appropriate expression vector for analysis of the mutated gene product. These procedures have been recently reviewed [9,10].

M13 bacteriophage system permits facile cloning of a DNA fragment of interest and provides an ample supply of single-stranded circular recombinants [11,12]. An appropriate synthetic oligonucleotide that contains the desired mutant sequence must be of sufficient length and homology that it will specifically anneal to the single-stranded target site. The nucleotide sequence that is to remain unchanged must then be faithfully regenerated. The resultant DNA heteroduplex containing both mutant and wild-type strands can be resolved by replication in vivo and by subsequent screening of the resultant clones with the appropriate mutant oligonucleotide as a probe. The altered genetic sequence can then be expressed in heterologous cells (either bacterial, yeast, or mammalian) using an appropriate expression vector. The altered protein can then be purified to homogeneity by more or less standard procedures and characterized functionally. Finally, elucidation of the structural consequences of the site-

specific mutation can be accomplished for select mutants by determining their three-dimensional structure with X-ray crystallographic techniques.

Oligonucleotide synthesis can also be used for the de novo synthesis of a gene with unique structures, but there are substantive problems that must be resolved before one can successfully design from first principles a novel functional structure. There has been considerable progress in developing methods for predicting secondary structure from sequence, but it is still a formidable problem to predict three-dimensional conformations from the linear amino acid sequence. It is known that the same amino acid sequence can assume different conformations in different proteins [13]. Perhaps an even greater concern is our limited understanding of the relationship between the three-dimensional structure of a protein and its physiological function. Typically, form does not follow function with regard to protein structure. In particular, substrate specificity and the mechanism of catalysis at the atomic level remain hypothetical, even when the three-dimensional structure of the enzyme has been determined, the amino acid resudues in the active site are defined, the kinetic features of the catalysis are elucidated, and enzyme substrate adducts or enzyme modifications have been produced. At the present state of our knowledge, we think it prudent to study proteins for which there is substantial structural and functional information to lay the groundwork for more difficult experiments that create completely novel structures. One such enzyme is trypsin. Trypsin is a prototypic hydrolytic enzyme that has been thoroughly studied at the level of both structure and mechanism.

There are several advantages in using trypsin to study structure/function relationships in proteins. Trypsin is a secretory protein that is synthesized as a zymogen. Heterologous expression systems can therefore be devised that secrete the inactive precursor into the media, simplifying the purification procedures and also maintaining the viability of the expression host. The protease is a stable enzyme that can withstand extremes of pH, ionic strength, and organic solvents, permitting ready purification without loss of activity. Purification procedures are further simplified by using well characterized affinity columns, which afford rapid isolation of the activity of interest.

Current kinetic and structural characterizations of trypsin benefit from the extensive studies that have been previously carried out on the enzyme. The kinetics of ester and amide bond hydrolysis by trypsin are well studied and straightforward [14]. Kinetic microconstants can be determined providing an in-depth kinetic view of the enzyme. A battery of inhibitors (4-methylumbelliferyl p-guanidinobenozoate [MUGB]; tosyl-L-lysine chloromethyl ketone [TLCK], diisoprophylfluorophosphate [DIFP], bovine basic pancreatic trypsin inhibitor [BPTI], soybean trypsin inhibitor [STI]) also provide useful reagents for determining dissociation constants and active-site reactivities. Large protein crystals that are suitable for X-ray diffraction studies can be readily formed by precipitation in concentrated solutions of either magnesium sulfate, ammonium sulfate, or polyethylene glycol. The crystals can be reacted with substrates to study structures involved in binding and catalysis. Furthermore, the high degree of structural resolution and a well defined mechanism permit quantum mechanical and molecular dynamic studies to be performed to correlate energy differences with modified functional and structural characteristics [15–17].

Alteration of Substrate Specificity

Trypsin is a pancreatic protease that belongs to a large family of homologous serine proteases. These enzymes utilize the amino acid, serine, at their active site and

appear to have the same catalytic mechanism. [18,19] This biologically important class of proteins includes enzymes involved in protein degradation (trypsin, chymotrypsin, elastase, α-lytic protease, *Streptomyces griseus* protease A and B), blood coagulation (thrombin), clot dissolution (plasmin), complement fixation (C1 protease), pain sensing (kallikrein), and fertilization (acrosomal enzyme). The diverse biological functions of the structurally homologous serine proteases are presumably the result of the different constellations of amino acids utilized by each enzyme for substrate binding.

In order to use a genetic approach to investigate the role of specific amino acids in the structure and function of trypsinogen, it was necessary to develop a system that permitted the substitution, deletion, or insertion of amino acids into the coding sequences of the enzyme. Therefore, the cDNA and gene for trypsinogen were isolated from the rat, completely characterized [20,21], and then used to construct a full length copy of the trypsinogen sequence (including the signal peptide) [22].

A heterologous expression system using mammalian cells and a simian virus 40 (SV40) replacement vector [23,24] was chosen in order to maximize the probability that native disulfide bond formation and secretion of the zymogen would occur. The early gene containing the T antigen-coding region of SV40 was replaced by the trypsinogen sequence, and the recombinant viruses were propagated to high copy numbers in transformed monkey kidney cells (COS) [22].

Trypsinogens were isolated from COS cell supernatants by affinity chromatography using a column of pancreatic trypsin inhibitor-sepharose or immobilized IgG specific to rat anionic trypsin. (Fig. 3) The affinity-purified protein exhibited no activity when assayed with fluorogenic peptide substrates. However, after treatment with the natural activating enzyme, enterokinase, [25] hydrolysis of arginine- and lysine-containing peptides could be monitored. This catalytic activity was rapidly and completely blocked by Phe-Ala-Arg-chloromethylketone (FARCK), an irreversible trypsin inhibitor [26]. These results show that the zymogen form of the protease is expressed by the COS cells.

Conversion of wild-type codons to codons of choice was accomplished by oligonucleotide-directed, site-specific mutagenesis using methods adapted from Zoller and Smith [9]. To ensure the accuracy of the process, the entire mutant DNA template was sequenced by the chain termination method of Sanger et al [27] with a set of synthetic oligonucleotides (see Fig. 4) and was shown to correspond to the sequence of the wild-type enzyme except at the point of mutation.

Although the three-dimensional structure for rat pancreatic trypsin is not known, the primary structure has 74% identity with bovine trypsin, whose three-dimensional crystal structure is known [28,29]. The active site residues His57, Asp102, and Ser195 are present in both enzymes and are located in regions with sequence identity. Similarly, Asp189 at the base of the substrate binding pocket, which presumably confers the substrate specificity for arginyl and lysyl substrates, is present in the rat sequence. The glycine residues at positions 216 and 226, which ostensibly allow entry of large amino acid side chains into the pocket, are also conserved. Indeed, when the modeled structure of rat trypsin is compared with the actual structure of bovine trypsin complexed with pancreatic trypsin inhibitor [30] (with Lys15 of the inhibitor in the substrate binding pocket), there are no substitutions within 7.6 Å of the ligand side chain. We therefore used the bovine trypsin structures complexed with either the pancreatic trypsin inhibitor or benzamidine (an arginine analog [31,32]) as a model for the rat trypsin complexed with lysine and arginine substrates, respectively.

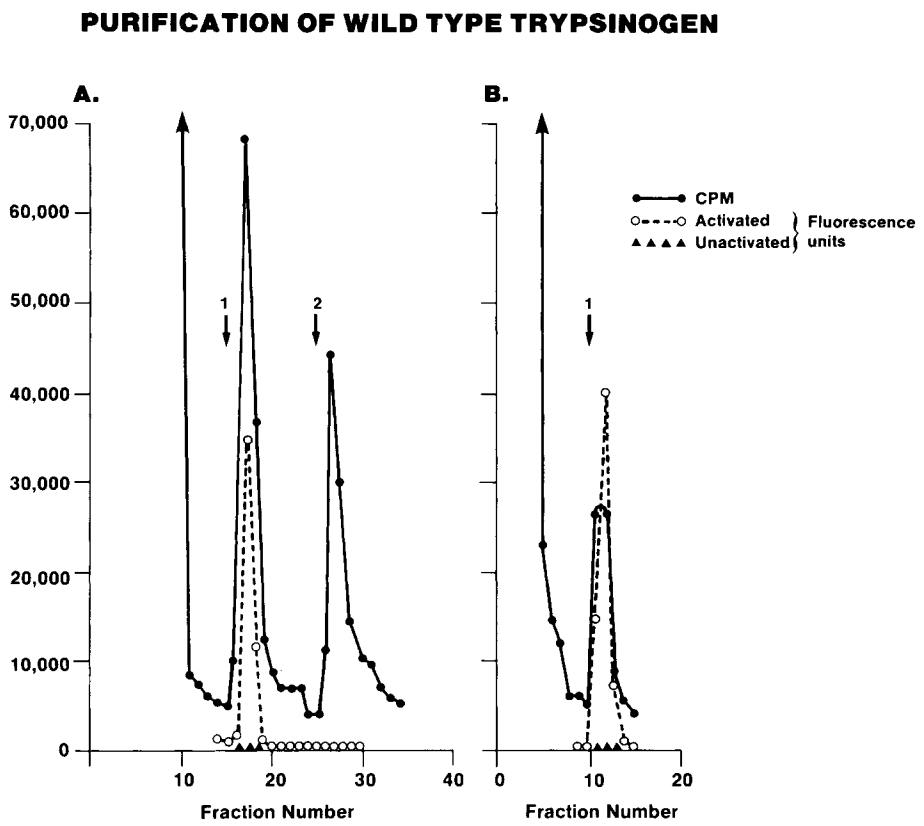

Fig. 3. A) Isolation of ^{35}S-labeled wild-type trypsinogen on pancreatic trypsin inhibitor-sepharose. Following pass-through of approximately 50 ml containing 10^7 cpm of labeled cellular proteins, trypsinogen was eluted with 0.1 M formic acid, pH 5.0 (arrow 1). Activity following enterokinase activation co-eluted with radiolabel. No activity was detected following elution with 0.1 M formic acid (pH 2.1) (arrow 2), which elutes authentic trypsin from this resin. Eluted material from the first peak migrated as trypsinogen on SDS gel electrophoresis. B) Isolation of ^{35}S-labeled wild-type trypsinogen on anti-anionic trypsinogen IgG-sepharose. Following pass-through of approximately 20 ml containing 10^7 cpm of labeled cellular proteins, trypsinogen was eluted with 0.1 M formic acid (pH 2.1). Activity following enterokinase activation coeluted with radiolabel. Eluted material migrated as trypsinogen on SDS gel electrophoresis.

Comparison of the related tertiary structures of trypsin and elastase [33] 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 Gly216 and Gly226 in elastase are Val216 and Thr226 [34], 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.

Assuming 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. Although the same binding pocket is shared by cationic substrates, the amino acid residues comprised by the pocket are employed

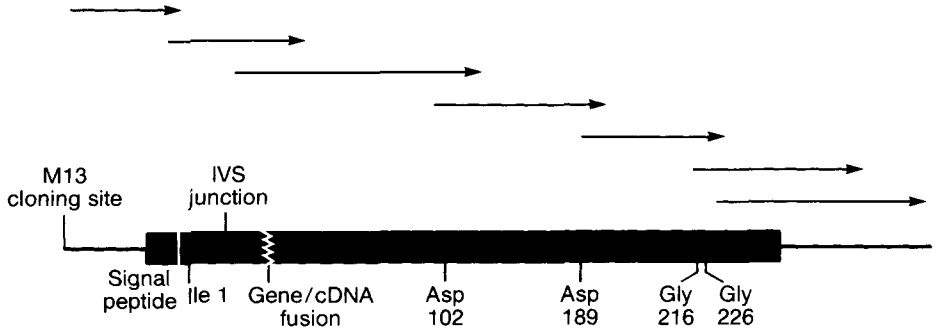

Fig. 4. Sequence confirmation of trypsin mutants. To ensure that inadvertant mutations do not occur in the trypsin sequence, a set of synthetic oligonucleotides are used to sequence the mutant DNA template. The arrows represent sequence information derived from a given oligonucleotide.

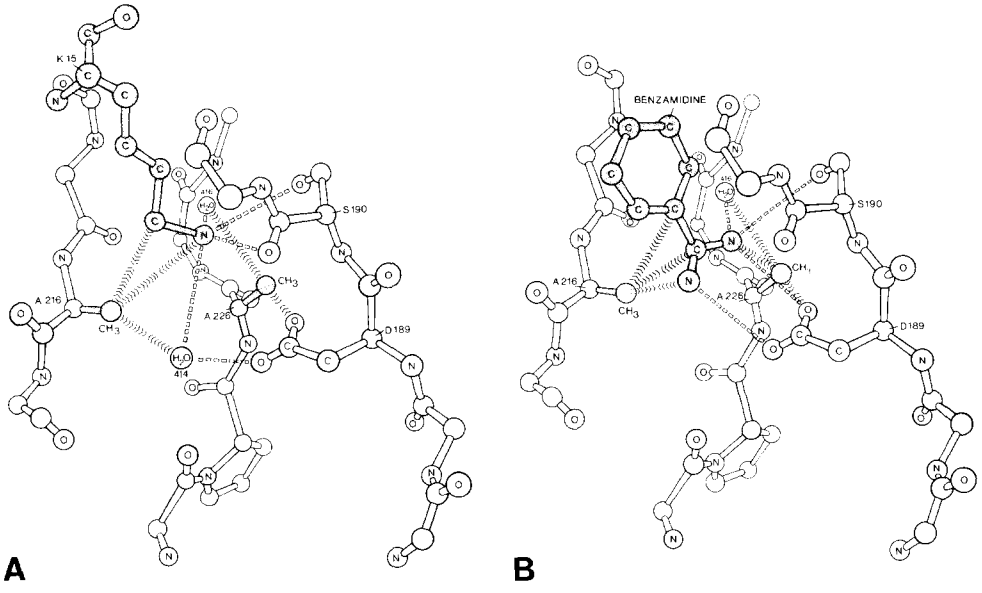

Fig. 5. Substrate-binding pocket of mutant trypsins. The backbone is shown for the chain segments, including amino acids 189–191, 214–217, and 224–227. Atomic coordinates for the drawing were obtained from the Brookhaven Protein Data Bank (entry set 3PTB). Both Ala216 and Ala226 as well as the ligands, Lys15, 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 Lys15 of pancreatic trypsin inhibitor bound in the trypsin specificity pocket. B) Benzamidine bound in the trypsin specificity pocket.

differently for lysyl substrate binding than for arginyl substrate binding (Fig. 5). Both arginyl and lysyl substrates form hydrogen bonds with a water molecule, which is bound in the specificity pocket by hydrogen bonds to the backbone carbonyl oxygens of Trp215 and Val227. 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) with Aspl89.

In contrast to lysine, the longer arginine side chain extends deeper into the binding pocket than the lysine side chain, displacing the water molecule and forming a cyclic network of direct hydrogen bonds with Asp189. 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. Since catalytic activity is dependent on the substrate alignment, which is determined in part by the specificity pocket, these changes may effect the kinetic constants Km, which relates to binding affinity, and kcat, which is a measure of catalytic activity. From modeling studies, the trypsin-216Ala mutant is predicted to show relatively better activity for arginine as compared to lysine because the water molecule that is presumably displaced by Ala216 does not take part directly in binding of arginine. On the other hand, the trypsin-226Ala mutant should show relatively enhanced lysine activity relative to arginine because there is more space to accommodate the steric conflicts of Asp189, the methyl group at position 226, and the substrate at the base of the pocket (see Fig. 5).

To test these hypotheses, glycine residues at 216 and 226 were replaced by alanine residues, resulting in three trypsin mutants (trypsin-216Ala, trypsin-226Ala and trypsin-216Ala,226Ala). Both the wild-type and the mutant enzymes were expressed in the mammalian cell expression system, the enzymes were purified, and their activities were compared. The wild type and a revertant enzyme in which the alanine at position 226 was site-specifically reverted to a glycine (to test the fidelity of the system) exhibited identical catalytic parameters. Results presented in Table I show that the addition of a methyl group (alanine vs. glycine) in either position 216 or 226 compromises the ability of the pocket to accept either lysine or arginine substrates (Kms are higher). However, the catalytic activity can be altered in a discriminatory manner. The trypsin-216Ala mutant does indeed operate more selectively on arginine substrates than wild type, and the trypsin-226Ala mutant operates more selectively on lysine substrates. The double mutant, trypsin-216Ala,226-Ala has a very low catalytic activity (1,000-fold lower than normal) owing to the greatly constricted binding pocket, and shows a modest preference for lysine substrates. In addition, unpredictably, the trypsin-226Ala and trypsin-216Ala,226Ala mutants exhibit an altered conformation, which resembles trypsinogen as detected by native gel

TABLE I. Modification of Arginine/Lysine Specificity of Trypsin*

	Arg		Lys		
	kcat	Km	kcat	Km	Arg/Lys[a]
Trypsin-216Ala	1.0	30	.3	30	3
Trypsin-226Ala	.01	40	0.1	25	.05
Trypsin-216Ala,226Ala	.001	15	.005	2	.3

*The data depicted are ratios of substrate-binding pocket mutant and wild-type trypsin kinetic parameters. The absolute values obtained are approximated for purposes of clarity.
[a]Ratio of kcat/Km values.

electrophoresis [22]. A similar analysis shows that this comformation is converted to a trypsin-like conformation upon binding of a substrate analog. Thus, the two different conformational states of trypsin and trypsinogen are not determined by the zymogen peptide alone, but may also be produced by subtle changes in the interior of the molecule, namely the addition of a methyl group at amino acid position 226.

Probing the Catalytic Mechanism of Trypsin

Ser195, His57, and Asp102, the catalytic triad, are present in the catalytic sites of all serine proteases from bacteria to humans. A serine residue was originally shown to be required for catalysis on the basis of its unique reactivity with organophosphates such as DIFP [35]. A histidine residue in trypsin was implicated in catalysis through its specific alkylation by the affinity label TLCK [36]. The high-resolution crystal structure of trypsin revealed that His57 and the β-hydroxyl group of Ser195 are within hydrogen-bonding distance of one another [28,29]. These and other data are consistent with the catalytic mechanism in which His57 abstracts a proton from Ser195 and donates a proton to the leaving group. Surprisingly, the three-dimensional structure also revealed that the carboxyl group of the buried Asp102 was within 2.8 Å of the imidazole ring of His57, suggesting that the aspartic acid through its effect on His57 and Ser195 may aid in catalytic activity. Consistent with this theory, the Asp102 residue is invariant in the crystal structures of all studied serine proteases. However, its role in catalysis has never been verified since selective chemical modification of the aspartic acid residue has not been possible. Evidence for its participation has relied on indirect methods such as nuclear magnetic resonance studies of the imidazole protons [37], neutron diffraction studies on deuterated trypsin [38], and quantum mechanical studies of the "proton relay system" [15]. These studies suggest that Asp102 shares the proton with the imidazole of His57, thereby 1) increasing the basicity of His57, 2) orienting the imidazole ring into a proper position for interacting with the substrate, and 3) modulating the nucleophilicity of Ser195.

We have directly examined the role of Asp102 in the catalytic mechanism of the serine proteases by replacing this residue in trypsin with an asparagine, an isosteric amino acid, by site-directed mutagenesis. The mutant enzyme, trypsin-102Asn was expressed to high levels (10 mg/liter) by establishing a stable eucaryotic cell line that secreted trypsinogen into the culture medium. Chinese hamster ovary cells were cotransfected with a plasmid containing either the wild-type or the mutant trypsinogen gene under transcriptional control of the early promoter from simian virus 40 [22] and a plasmid that encoded the bacterial phosphotransferase gene (neo). The neo gene confers resistance to the aminoglycoside antibiotic G418 [39] and permits the phenotypic selection of a cell line that expresses the gene. Cells that coexpressed the trypsinogen and neo genes were screened for overexpressors, and the cell line that produced the highest level of trypsinogen was expanded into mass culture (10 liters). Wild-type trypsin and trypsin-102Asn were purified to homogeneity by a combination of ion-exchange and affinity chromatography techniques. Wild-type trypsin isolated from this expression system showed identical physical and catalytic properties as the natural rat enzyme. Trypsin-102Asn, however, exhibited remarkably different catalytic properties (Table II). Kinetic analysis below pH 7 indicated that the acylation reaction follows a titration curve similar to that found with the native enzyme except the pKa of histidine is lowered about 1.5 pH units to 5.3, and the maximal rate constant is at least 5000-fold lower than that for wild-type trypsin. Thus, Asp102 is

TABLE II. Relative Activity of Trypsin-102Asn to Wild-Type Trypsin at Neutral and Alkaline pH*

	kcat	Km
Neutral pH	.0002	2
Alkaline pH	.15	10

*Trypsin-102Asn mutants have dramatically lower catalytic activity at neutral pH but approach wild-type activity at alkaline pH. The values shown are ratios of trypsin-102Asn and wild-type trypsin. Neutral pH is 7.0 and alkaline pH is 10.5.

TABLE III. Relative Reactivity of Trypsin-102Asn to Wild-Type Trypsin With Active Site Titrants*

	Trypsin	Trypsin-102Asn
DIFP	1	.0001
MUGB	1	< .002
TLCK	1	0.2

*Trypsin-102Asn shows dramatically lowered activity with Ser195-directed active site reagents but not with a histidine-specific reagent. Values shown are relative to wild-type trypsin. DIFP, diisopropylfluorophosphate, phosphorylates Ser195; MUGB, 4-methylumbelliferyl p-guanidinobenzoate, monitors the acylation reaction; TLCK, tosyl L-lysine chloromethylketone, specifically alkylates histidine 57.

clearly important but not absolutely required for catalytic activity. The acylation of trypsin-102Asn still occurs at a rate 400-fold greater than alkaline hydrolysis of the same ester substrate. Above pH 7, the rate of the reaction increases in direct proportion to the solvent hydroxide ion concentration such that at pH 10.5 the kcat value of the trypsin-102Asn-catalyzed hydrolysis of esters is about 15% that of the wild-type enzyme (Table II).

To determine the chemical reactivity of the active site residues of trypsin-102Asn we employed the active site reagents DIFP, MUGB, and TLCK (Table III). The reactivity of trypsin-102Asn with the Ser195-specific reagent DIFP is approximately 10,000-fold less than that observed for wild-type trypsin. In addition, the mutant enzyme is at least 500-fold less reactive than wild-type trypsin with the acylating reagent MUGB. Thus, the nucleophilicity of Ser195 is greatly compromised in trypsin-102Asn. On the other hand, the reactivity of trypsin-102Asn with TLCK, a His57-specific affinity label, is only decreased fivefold. This suggests that the histidine can be properly positioned in the active site of the mutant enzyme for reaction with the affinity label. The primary effect on catalysis of the aspartic-acid-to-asparagine replacement at position 102 therefore appears to be on Ser195, implying that it is the network of hydrogen bonds involving Asp102, His57, and Ser195 that leads to the strong nucleophilic character of Ser195 that ultimately accounts for its role in catalysis. This supports the importance of the catalytic triad: all the residues are required for optimal function.

Surprisingly, trypsin-102Asn displays a unique set of catalytic characteristics that qualitatively differentiates it from the parent enzyme. Trypsin-102Asn is a good

catalyst at basic pH because of its ability to utilize solvent hydroxyl in the catalytic reaction. It is not yet known whether the hydroxyl interacts directly with the histidine residue or the serine residue or facilitates the reaction in some other manner. This is clearly an example of the production via genetic engineering of an enzyme with a qualitatively different reaction mechanism. It illustrates the possibility of producing new proteins that show reduced activity under physiological conditions but are efficient catalysts in different environments.

Although the studies described above involving enzyme kinetics provide substantial information on the functional consequences of the site-specifically mutated enzymes, a three-dimensional structural analysis is required to fully understand the modified functional state. Therefore, we crystallized trypsin-102Asn by vapor diffusion against polyethylene glycol at pH 6.0 in the presence of benzamidine. The crystals attained dimensions of 0.5–1.0 mm and are suitable for study by X-ray crystallography. The packing of the molecules in the rat trypsin crystal differs from that of the bovine crystal. This difference is probably not a consequence of the engineered mutation at position 102. The three-dimensional structure of the rat structure has been determined by molecular replacement methods using the atomic coordinates of the bovine enzyme as a starting model.

The three-dimensional structure of rat trypsin-102Asn is currently determined at 2.3 Å resolution. This permits the visualization of side chain positions and hydrogen-bonding patterns. From the similarity in primary structure between the rat and bovine enzymes and the overall tertiary structure similarity between the serine proteases with known structures, we assumed that rat trypsin would be structurally similar to bovine trypsin. The X-ray structure results confirm that there is nearly exact structural identity between the rat and bovine enzymes. There is 0.5 Å deviation between similar atoms of the polypeptide backbone in the two structures. Although subtle differences were found between the two structures at the active site region, the asparagine residue in trypsin-102Asn occupies the same site as Asp102 in the bovine structure. Surprisingly, however, the rat His57 and Ser195 residues are in different conformational states compared with the same residues in the bovine enzyme. An analysis of the conformational differences of these and other residues of the enzyme structure is presently in progress.

These studies involving genetics, recombinant DNA technology, enzymology, protein chemistry, and crystallography reveal the importance of a coordinated multidisciplinary approach to structure/function studies. Results from these studies will provide a deeper understanding of the mechanism for designing new proteins with different structures and functions for a variety of purposes.

ACKNOWLEDGMENTS

We gratefully acknowledge support from NSF grant PCM 830610 to W.J.R., NSF grant DMB 8608086 to C.S.C., NIH grant AM 26081 to R.J.F., and a Cancer Research Coordinating Committee grant to C.S.C.

REFERENCES

1. Dickerson R, Geis I: "Hemoglobin Structure, Function, Evolution and Pathology." Menlo Park, CA: Benjamin/Cummings Publ. Inc., 1983.

2. Hunkpiller M, Kent S, Caruthers M, Dreyer JF, Giffin C, Horvath S, Hunkapiller T, Tempst P, Hood L: Nature 310:105–111, 1984.
3. Maxam A, Gilbert W: Methods Enzymol 65:499–560, 1980.
4. Sanger F, Nicklen S, Coulson AR: Proc Natl Acad Sci USA 74:5463–5467, 1977.
5. Zimmern D, Kaesberg P: Proc Natl Acad Sci USA 75:4257–4261, 1978.
6. Shortle D, DiMaio D, Nathans D: Annu Rev Genet 15:265–294, 1981.
7. Botstein D, Shortle D: Science 229:1193–1201, 1985.
8. Wells JA, Vasser M, Powers DD: Gene 34:315–323, 1985.
9. Zoller M, Smith M: Methods Enzymol 100:468–500, 1983.
10. Craik CS: Biotechniques 3:12–19, 1985.
11. Messing JB, Gronenborn B, Muller-Hill B, Hofschneider PH: Proc Natl Acad Sci USA 74:3642–3646, 1977.
12. Messing J: Methods Enzymol 101:20–78, 1983.
13. Kabsch W, Sander C: Proc Natl Acad Sci USA 81:1075–1078, 1984.
14. Fersht A: "Enzyme Structure and Mechanism," 2nd edition. New York: W.H. Freeman, 1985.
15. Umeyama H, Hirono S, Nakagawa S: Proc Natl Acad Sci USA 81:6266–6270, 1984.
16. Scheiner S, Lipscomb WN: Proc Natl Acad Sci USA 73:432, 1976.
17. Weiner SJ, Seibel GL, Kollman PA: Proc Natl Acad Sci USA 83:649–653, 1986.
18. Neurath H, Walsh KA, Winter WP: Science 158:1638–1644, 1967.
19. Kraut J: Annu Rev Biochem 46:331–358, 1977.
20. MacDonald RJ, Stary SJ, Swift GH: J Biol Chem 257:9724–9732, 1982.
21. Craik CS, Choo Q-L, Swift GH, Quinto C, MacDonald RJ, Rutter WJ: J Biol Chem 259:14255–14264, 1984.
22. Craik CS, Largman C, Fletcher T, Roczniak S, Barr PJ, Fletterick R, Rutter WJ: Science 228:291–297, 1985.
23. Gluzman Y: Cell 23:175, 1981.
24. Laub O, Rutter WJ: J Biol Chem 258:6043, 1983.
25. Baratti J, Maroux S, Louvard D: Biochim Biophys Acta 321:632, 1973.
26. Kettner C, Springhorn S, Shaw E: Hoppe-Seylers Z Physiol Chem 359:1183–1191, 1978.
27. Sanger F, Nicklen S, Coulson AR: Proc Natl Acad Sci USA 74:5463–5467, 1977.
28. Stroud RM, Kay LM, Dickerson RE: J Mol Biol 83:185–208, 1974.
29. Huber R, Kukla D, Bode W, Schwager P, Bartel K, Diesenhofer J, Steigemann W: J Mol Biol 89:73–101, 1974.
30. Bode W, Schwager P, Huber R In "Proteolysis and Physiological Regulation, Miami Winter Symp." New York: Academic Press, 1976, Vol 11, pp 43–76.
31. Bode W, Walter J, Huber R, Wenzel HR, Tschesche H: Eur J Biochem 144:185–190, 1984.
32. Bode W, Schwager P: J Mol Biol 98:693–717, 1975.
33. Sawyer L, Shotton DM, Campbell JW, Wendel PL, Muirhead H, Watson HC, Diamond R, Ladner RC: J Mol Biol 118:137–208, 1978.
34. Shotton DM, Watson HC: Nature 225:811–816, 1970.
35. Dixon GH, Go S, Neurath H: Biochim Biophys Acta 19:193–195, 1956.
36. Shaw E, Mares-Guia M, Cohen W: Biochemistry 4:2219–2224, 1965.
37. Robillard G, Shulman RG: J Mol Biol 86:541–558, 1974.
38. Kossiakoff AA, Spencer SA: Nature 288:414–416, 1980.
39. Davies J, Smith DI: Annu Rev Microbiol 32:469–518, 1978.