

Molecular Cloning of Dog Mast Cell Tryptase and a Related Protease: Structural Evidence of a Unique Mode of Serine Protease Activation^{†,‡}

Peter Vanderslice,[‡] Charles S. Craik,[§] Jay. A. Nadel,^{‡,||} and George H. Caughey^{*,‡}

Cardiovascular Research Institute and Departments of Medicine, Physiology, and Pharmaceutical Chemistry, University of California, San Francisco, California 94143

Received December 29, 1988; Revised Manuscript Received February 3, 1989

ABSTRACT: Mast cell tryptase is a secretory granule associated serine protease with trypsin-like specificity released extracellularly during mast cell degranulation. To determine the full primary structure of the catalytic domain and precursor forms of tryptase and to gain insight into its mode of activation, we cloned cDNAs coding for the complete amino acid sequence of dog mast cell tryptase and a second, possibly related, serine protease. Using RNA from dog mastocytoma cells, we constructed a cDNA library in λ gt10. Screening of the library with an oligonucleotide probe based on the N-terminal sequence of tryptase purified from the same cell source allowed us to isolate and sequence overlapping clones coding for dog mast cell tryptase. The tryptase sequence includes the essential residues of the catalytic triad and an aspartic acid at the base of the putative substrate binding pocket that confers P1 Arg and Lys specificity on tryptic serine proteases. The apparent N-terminal signal/activation peptide terminates in a glycine. A glycine in this position has not been observed previously in serine proteases and suggests a novel mode of activation. Additional screening of the library with a trypsinogen cDNA led to the isolation and sequencing of a full-length clone apparently coding for the complete sequence of a second tryptic serine protease (DMP) which is only 53.4% identical with the dog tryptase sequence but which contains an apparent signal/activation peptide also terminating in a glycine. Thus, the proteases encoded by these cloned cDNAs may share a common mode of activation from N-terminally extended precursors.

Mast cell tryptase is a serine protease of mast cell secretory granules. High levels of intracellular tryptic activity were first noted as a striking characteristic of mast cells 29 years ago via enzyme histochemical staining (Glennner & Cohen, 1960). Tryptase has since been isolated from human (Schwartz et al., 1981a; Smith et al., 1984; Cromlish et al., 1987), rat (Kido et al., 1985), and dog (Caughey et al., 1987) mast cells and mast cell rich tissues; it can constitute as much as 23% of mast cell protein (Schwartz et al., 1981a). During mast cell degranulation, tryptase is released extracellularly along with other preformed mediators of mast cell degranulation (Schwartz et al., 1981b; Caughey et al., 1988a). Like trypsin, tryptase hydrolyzes peptide bonds on the C-terminal side of basic residues (Arg > Lys), but unlike trypsin, tryptase has marked subsite preferences and a correspondingly high degree of substrate selectivity (Tanaka et al., 1983).

Although the exact in vivo roles of tryptase remain to be proven, several potential biologically important actions have been identified. Tryptase rapidly hydrolyzes vasoactive intestinal peptide and reverses airway smooth muscle relaxation by this neuropeptide which is believed to modulate neural nonadrenergic relaxation in airways (Caughey et al., 1988b; Franconi et al., 1989); the enzyme also potentiates the smooth muscle contractile effects of mediators such as histamine

(Sekizawa et al., 1989), which is coreleased from mast cells with tryptase. Thus, tryptase released from mast cells may promote pathological airway narrowing by hydrolyzing a bronchodilating peptide and by augmenting the effects of bronchoconstricting agonists. In vitro, tryptase also generates the anaphylatoxin C3a (Schwartz et al., 1983), destroys fibrinogen (Schwartz et al., 1985) and high molecular weight kininogen (Maier et al., 1983), and activates procollagenase (Gruber et al., 1988). These latter findings suggest that tryptase may have a proinflammatory role and may act locally with degranulated mast cell heparin as an anticoagulant.

Mast cell tryptase differs markedly from other serine proteases in many of its biochemical features. It has two or more subunits which associate noncovalently as catalytically active tetramers and appears to be stored and released as an active enzyme rather than as a zymogen. No new tryptase activity is generated from putative inactive precursors during exocytosis (Schwartz et al., 1981b; Caughey et al., 1988a). It is not known whether tryptase is synthesized initially as a zymogen and activated before packaging in the granule. Tryptase is remarkably resistant to inactivation by circulating inhibitors of serine proteases. Dog and human tryptases remain active when incubated with plasma or with α_1 -proteinase inhibitor or soybean trypsin inhibitor (Smith et al., 1984; Caughey et al., 1987; Cromlish et al., 1987). The rat enzyme differs in that it appears to be associated with an endogenous inhibitor in mast cell granules (Kido et al., 1985). Rat tryptase also differs in its substrate preferences and does not cross-react strongly with antibodies which recognize both dog and human tryptase (Schechter et al., 1988). An additional important feature of tryptase is its interaction with sulfated glycosaminoglycans. The stability of tryptase in plasma and in low ionic strength buffers is markedly enhanced by heparin (Schwartz & Bradford, 1986; Caughey et al., 1987). Heparin, which is released with tryptase from mast cell granules, appears to be important in maintaining the enzyme as a catalytically

[†] This work supported by Grants HL-24136 (J.A.N.), HL-01736 (G.H.C.), and DMB 8608086 (C.S.C.) from the National Institutes of Health and the National Science Foundation and by the American Lung Association of California (G.H.C.).

[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02862.

^{*} Correspondence should be addressed to this author at the Cardiovascular Research Institute, Box 0130, University of California, San Francisco, CA 94143.

[‡] Cardiovascular Research Institute and Department of Medicine.

[§] Department of Pharmaceutical Chemistry.

^{||} Department of Physiology.

active tetramer (Schwartz & Bradford, 1986; Alter et al., 1987). The structural features of trypsin which give rise to its distinctive biochemical attributes are unknown.

In the present study, we used RNA isolated from dog mastocytoma cells to create a cDNA library and thus determine the full primary structure of trypsin and obtain insights into its mode of activation. Synthetic oligonucleotides based on the N-terminal sequence of dog trypsin were used to screen the library for cDNAs coding for trypsin. Because of the unique features of trypsin, we used trypsinogen cDNA as a probe to identify additional, possibly related mast cell serine proteases in the library. This screening strategy resulted in the identification of cDNAs coding for mast cell trypsin and for an additional protease. The precursor sequences deduced from the cloned cDNAs suggest that the two proteins they encode may be part of a family of serine proteases which share a unique mode of activation.

EXPERIMENTAL PROCEDURES

Construction of a Mastocytoma cDNA Library. Total RNA was isolated from collagenase-disaggregated BR dog mastocytoma cells (Caughey et al., 1988a) using guanidine isothiocyanate (Chirgwin et al., 1979). Poly(A) RNA was then enriched by two passages on oligo(dT)-cellulose (New England Biolabs) (Aviv & Leder, 1972). A cDNA library from poly(A) RNA was constructed in λ gt10, by using the Amersham cDNA synthesis and cloning systems. Poly(A) RNA was primed with oligo(dT) and extended with reverse transcriptase to synthesize the first strand of cDNA (Okayama & Berg, 1982). The RNA was partially hydrolyzed with ribonuclease H, and the resulting nicked RNA was used to prime second-strand synthesis using DNA polymerase I (Okayama & Berg, 1982; Gubler & Hoffman, 1983). Double-stranded cDNA that was not size-selected was made blunt-ended with T4 DNA polymerase, and internal *Eco*RI restriction endonuclease sites were protected by using *Eco*RI methylase (Hyunh et al., 1985). Following attachment of *Eco*RI linkers, the cDNA was digested with *Eco*RI, ligated into λ gt10, and packaged (Hyunh et al., 1985). The library was amplified once in *Escherichia coli* strain NM514.

Oligonucleotide Synthesis. All oligonucleotides for library screening, Northern blot analysis, and DNA sequencing with the exception of the M13 forward, reverse, and KS sequencing primers (Stratagene) were synthesized by using phosphoramidite chemistry (Beaucage & Caruthers, 1981) on an Applied Biosystems 380B DNA synthesizer at the Biomolecular Resource Center, University of California, San Francisco.

Library Screening. To screen the cDNA library for clones coding for trypsin, a 47-base oligonucleotide probe based on the N-terminal 16 amino acid residues of trypsin purified from BR dog mastocytoma cells (Caughey et al., 1987) was radioactively phosphorylated using $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ and T4 polynucleotide kinase (Bethesda Research Laboratories) (Maniatis et al., 1982). Additional screening was carried out by using a 1169 base pair bovine cationic trypsinogen cDNA (J. Higaki and C. S. Craik, unpublished results) labeled with $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$ and $[\alpha\text{-}^{32}\text{P}]\text{dGTP}$ by nick translation (Maniatis et al., 1982). Both probes were radiolabeled to $\sim 2 \times 10^8$ cpm/ μg . λ phage from the library were plated in agarose on 150-mm petri dishes and incubated at 37 °C for 12 h. The bacteriophage were transferred to nitrocellulose filters, and their DNA was immobilized (Benton & Davis, 1977). The DNA was hybridized to the probes as described (Maniatis et al., 1982) except that the concentration of formamide in the hybridization solution was 20% when screening with the oligonucleotide and 30% when screening with the trypsinogen

cDNA. After hybridization, the filters were exposed to Kodak XAR-5 film with a Cronex Lightning-Plus intensifying screen at -70 °C for 16 h.

Analysis of Recombinants. Positive recombinants were plaque-purified and reprobated. Bacteriophage DNA was then isolated by the plate lysate method (Maniatis et al., 1982). DNA samples were digested with *Eco*RI (Bethesda Research Laboratories) and electrophoresed through a 1% agarose gel. The DNA was transferred to a nitrocellulose filter (Southern, 1975) and hybridized either to the radiolabeled 47-base oligonucleotide or to the bovine trypsinogen cDNA as described above. The largest cDNA inserts that hybridized to the probe were cloned into the *Eco*RI site of pBluescript KS+/- phagemids (Stratagene) and transformed into *E. coli* XL1-Blue cells (Bullock et al., 1987). Single-stranded DNA was isolated and prepared as described (Dente et al., 1983) except that VCSM13 helper phage (Stratagene) was used. The nucleotide sequence of the cDNA was determined by the dideoxy chain termination method (Sanger et al., 1977) using Sequenase (United States Biochemical Corp.) (Tabor & Richardson, 1987). To determine the sequence of both strands in full, we used initially the M13 forward, reverse, and KS primers (Stratagene), and then we used a series of synthetic oligonucleotides designed from previously determined sequence.

Northern Blotting. Five-microgram samples of total and poly(A)-enriched RNA were denatured with glyoxal and dimethyl sulfoxide (McMaster & Carmichael, 1977) and electrophoresed through a 1% agarose gel. The RNA was transferred and immobilized on a Hybond-N membrane (Amersham). The RNA was then hybridized either to a 23-base oligonucleotide complementary to bases 151-173 of the trypsin cDNA or to an 18-base oligonucleotide complementary to bases 682-699 of the dog mastocytoma protease (DMP)¹ cDNA. The oligonucleotides were chosen to minimize potential cross-hybridization between trypsin and DMP mRNA. The oligonucleotides were radioactively phosphorylated using $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ and T4 polynucleotide kinase (Maniatis et al., 1982). The specific activation of the labeled trypsin and DMP oligonucleotides were 641 and 494 $\mu\text{Ci}/\text{mmol}$, respectively. Hybridizations were carried out at 55 °C for 16 h in 10% dextran sulfate, 0.25 M sodium phosphate, pH 7.0, 7% sodium dodecyl sulfate, and 10^6 cpm of probe/mL of solution. The filters were washed once in 0.15 M NaCl/0.015 M sodium citrate (pH 7.0) at room temperature for 10 min and twice in the same buffer plus 0.1% sodium dodecyl sulfate at 55 °C for 30 min and then dried and exposed to Kodak XAR-5 film with intensifying screens at -70 °C for either 2 h (DMP oligonucleotide probe) or 16 h (trypsin oligonucleotide probe). The approximate size of RNAs hybridizing to the probes was estimated by using the 28S (5.1 kb) and 18S (2.0 kb) rRNA bands as molecular weight markers.

Protein Data Base Searches. The catalytic domains of trypsin and DMP were compared to other protein sequences contained in the National Biomedical Research Foundation protein data base using the FASTP search algorithm (Lipman & Pearson, 1985).

Deglycosylation of Dog Trypsin. Trypsin was purified from dog mastocytoma cells as previously described (Caughey et al., 1987). Six-microgram samples of dog trypsin were treated with peptide:N-glycosidase F (Genzyme) as described (Tarentino et al., 1985). Control samples were treated in an

¹ Abbreviations: DMP, dog mastocytoma protease; RMCP, rat mast cell proteinase.

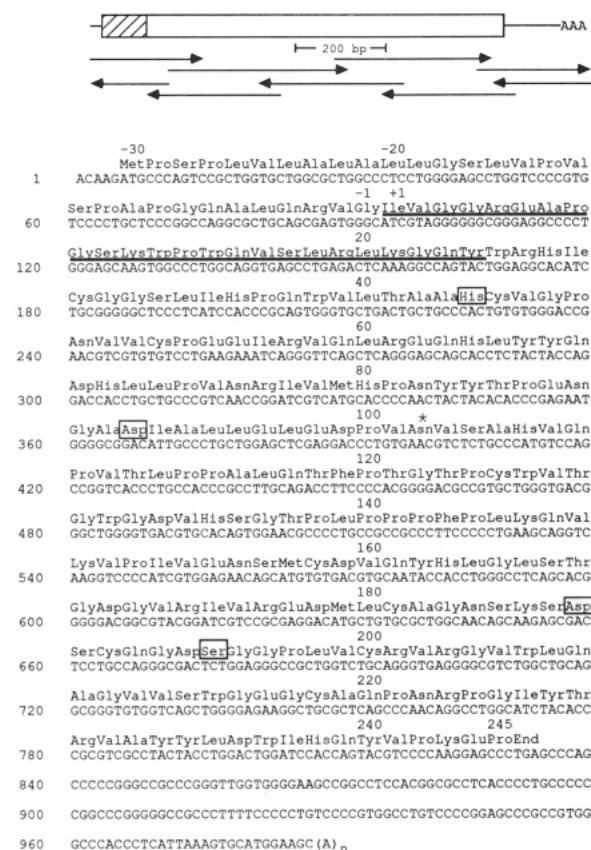


FIGURE 1: Tryptase nucleotide and deduced amino acid sequence. The upper portion of the diagram depicts the sequencing scheme for tryptase clone 1. The rectangles show the open reading frame starting with the first ATG. The hatched area represents the nucleotides coding for the putative preprosequence; arrows indicate the length and direction of sequence obtained by dideoxy sequencing using oligonucleotide primers. The lower portion of the diagram shows the composite coding sequence of tryptase derived from two overlapping cDNAs, clones 1 and 2. The nucleotides are numbered to the left and right of the figure. The composite sequence omits a stretch of 69 nucleotides (GTGGGACCCCCCGCCCTGCCCTGGCC-CCCGCCGCCCCGGCCCCGGCCGCCCTGACCCTCCCGCT-CCCCAG) inserted between the codons for Gln-191 and Gly-192. This sequence, thought to represent an intron, is present in clone 1 but not in clone 2 (see text). The deduced amino acids shown above the nucleotide sequence are numbered at intervals above the amino acid sequence. The His, Asp, and Ser of the catalytic triad are boxed. A shaded box shows the "trypsin-specific" Asp. The consensus N-linked glycosylation site is marked with an asterisk. The heavily underlined amino acids show the N-terminal sequence of purified dog tryptase which is identical with the deduced sequence of the cDNAs starting with Ile-1. The lightly underlined nucleotides represent the probable polyadenylation signal.

identical fashion except that no peptide:N-glycosidase F was added to the reaction mix. Samples were electrophoresed through a 12.5% sodium dodecyl sulfate-polyacrylamide gel and visualized with Coomassie blue.

RESULTS

Cloning of Dog Mast Cell Tryptase. The BR dog mastocytoma cDNA library contains 2×10^5 unique recombinants. Approximately 10^6 plaques from the amplified library were screened with an oligonucleotide probe based on the N-terminal amino acid sequence of dog tryptase. Of 27 hybridizing clones, 10 were plaque-purified, and their DNA was isolated. The phage DNA was digested with *EcoRI* and electrophoresed through a 1% agarose gel to determine insert size. The largest cDNA insert was chosen for sequence analysis. An 1120 base pair cDNA insert (clone 1) was cloned into pBluescript which enabled both strands to be sequenced in their entirety using

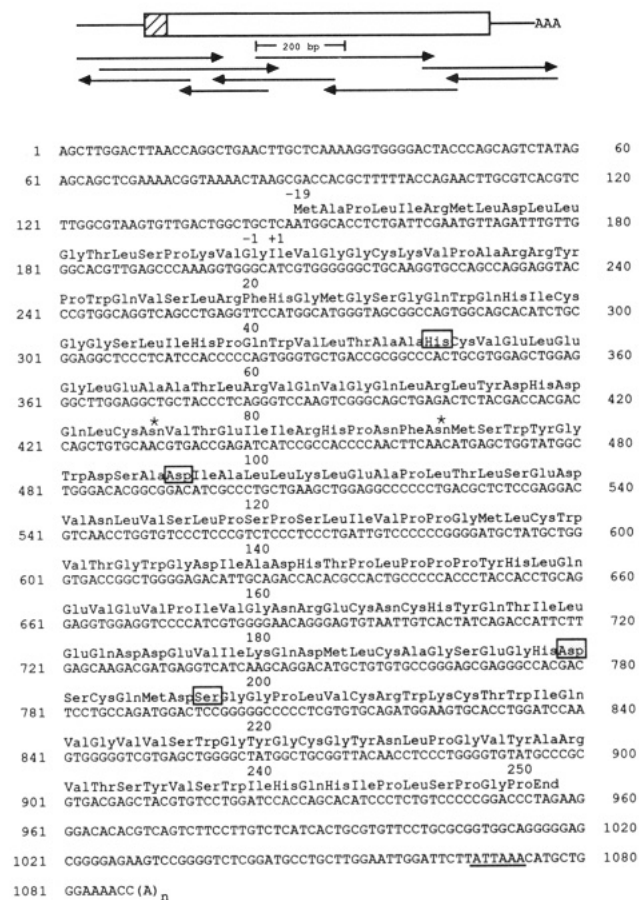


FIGURE 2: Dog mastocytoma protease (DMP) nucleotide and deduced amino acid sequence. The upper portion of the diagram depicts the sequencing scheme for DMP. The lower portion of the diagram shows the nucleotide sequence obtained from DMP cDNA. The deduced amino acids are shown above the nucleotide sequence. See Figure 1 legend for explanation of boxes, arrows, and labels.

M13 primers and oligonucleotide primers designed from previously determined sequence (Figure 1). Clone 1 cDNA contains an open reading frame of 894 base pairs terminating with the codon TGA and is followed by a 157 base pair 3' untranslated region containing the presumed polyadenylation signal ATTTAAA beginning 18 bases 5' to the poly(A) tail. The amino acid sequence deduced from the open reading frame indicates that the cDNA codes for dog tryptase. Residues 1-24 (Figure 1) of the deduced sequence are identical with the N-terminal 24 residues determined for dog mast cell tryptase purified from BR dog mastocytoma (Caughey et al., 1987). His-44, Asp-91, and Ser-194 of the tryptase sequence (Figure 1) correspond to the catalytic triad residues His-57, Asp-102, and Ser-195 (chymotrypsinogen numbering) common to all serine proteases (Figure 3). Asp-188 of tryptase corresponds to the substrate-specific Asp-189 characteristic of trypsins and other serine proteases of tryptic specificity. The tryptase cDNA codes for a protein with 9 cysteines, 1 consensus N-linked glycosylation site, and a distinctive 30-residue signal/activation peptide containing a C-terminal glycine. The average size of tryptase mRNA estimated from Northern blots (Figure 4) is 1260 base pairs, suggesting that the cDNA is nearly full length, lacking only a portion of the 5' noncoding region.

Clone 1 cDNA contains an insert of 69 base pairs between the codons for Gln-191 and Gly-192 (see legend of Figure 1). Five of 8 bases surrounding the 5' end of this insert and 14 of 15 bases surrounding the 3' end are identical with the consensus sequence of intron splice sites (Padgett et al., 1986).

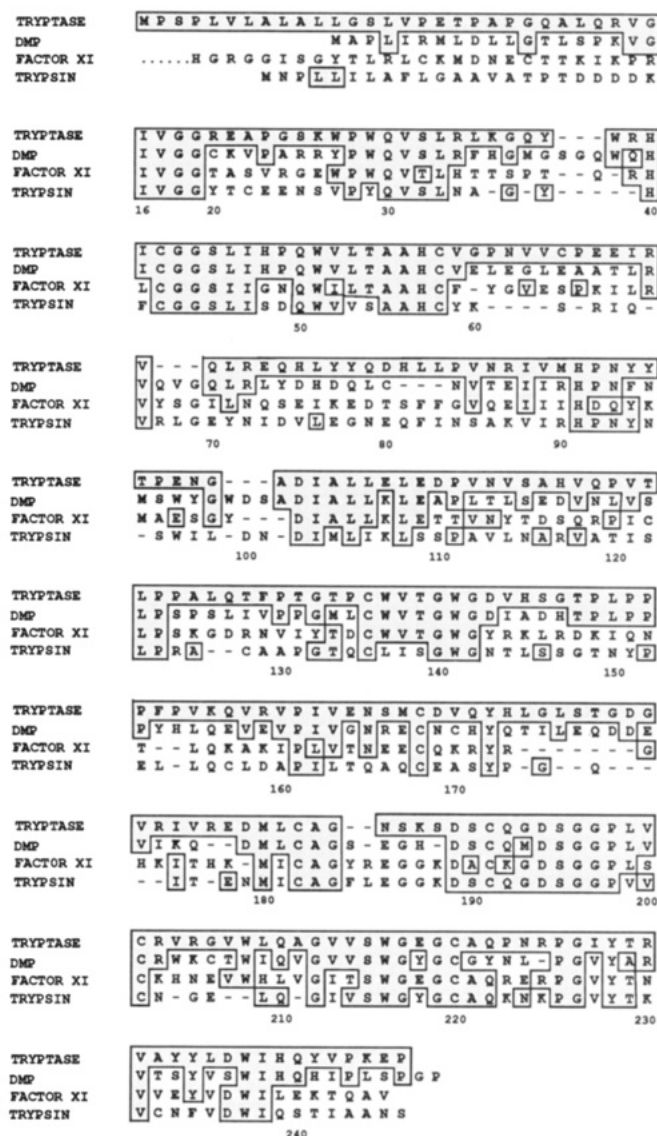


FIGURE 3: Alignment of tryptic protease primary structures. The amino acid sequences of dog trypsin, dog mastocytoma protease (DMP), human factor XI (Fujikawa et al., 1986), and dog anionic trypsin (Pinsky et al., 1985) were aligned manually with introduction of gaps to maximize homology. The numbering follows standard chymotrypsinogen notation. Shaded residues are identical with those of trypsin.

No known eukaryotic serine protease contains a peptide insert at this site which is adjacent to the absolutely conserved amino acids surrounding the catalytically active serine. The sequences of the genes coding for several serine proteases, including trypsin (Craik et al., 1984), kallikrein (Mason et al., 1983), and mast cell chymase (Benfey et al., 1987), contain an intron mapping exactly to this site. Thus, it seems highly likely that clone 1 contains an insertion corresponding to an intron from an incompletely processed mRNA that was the template for the cDNA. Therefore, we determined the nucleotide sequence of a second cDNA (clone 2) of 900 base pairs. The sequence of clone 2, which begins with base 115 and extends to the poly(A) tail, is identical with that of clone 1 except for the lack of the 69 base pair insert. Thus, we believe that mature mRNA coding for trypsin lacks the intronlike insert found in clone 1. The deduced trypsin sequence shown in Figure 1 is based on the composite sequence of clones 1 and 2, with omission of the insert.

The deduced amino acid sequence suggests that the catalytic domain of unmodified dog trypsin is 245 amino acids in length

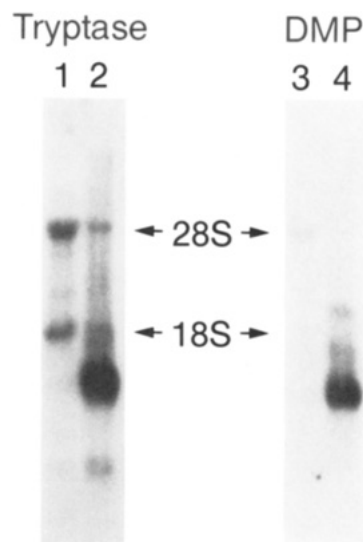


FIGURE 4: Northern blotting of dog mastocytoma RNA. Lanes 1-4 show autoradiograms of blots of dog mastocytoma RNA after electrophoretic separation, transfer to Hybond-N membrane, and incubation with radiolabeled synthetic oligonucleotides specific for trypsin (lanes 1 and 2) and for DMP (lanes 3 and 4). Lanes 1 and 3 contain 5 μ g of total RNA; lanes 2 and 4 contain 5 μ g of RNA enriched in mRNA. The locations of bands from 28S and 18S ribosomal RNA are identified.

with a calculated molecular weight of 27 157. When the catalytic domain of dog trypsin was compared to other sequences in the National Biomedical Research Foundation protein data base, the sequence with the highest similarity score was that of human factor XI. Sequence identity between dog trypsin and human factor XI in the regions of overlap is 42.9% (Figure 3).

Cloning of a Second Mastocytoma Protease. Approximately 10^6 plaques from the amplified mastocytoma cDNA library were screened with the bovine cationic trypsinogen cDNA. Analysis of seven of the most strongly hybridizing clones led to the isolation of a 1174 base pair cDNA insert which was chosen for further characterization. The sequencing strategy shown in Figure 2 was similar to that used for the trypsin clones. The cDNA contains a 147 base pair 5' noncoding region followed by an 807 base pair open reading frame and a 134 base pair 3' noncoding region. The presumed polyadenylation signal ATTAAA begins 21 bases 5' to the poly(A) tail.

Like the trypsin clones, the cDNA codes for amino acids corresponding to the residues of the catalytic triad and to the "trypsin-specific" Asp of tryptic serine proteases (Figures 2 and 3). The cDNA appears to code for the full sequence of a second serine protease, which we have termed dog mastocytoma protease (DMP), with a catalytic domain of 250 amino acids and a calculated molecular weight of 27 807. The protein contains 12 cysteines, 2 consensus N-linked glycosylation sites, and a 19 amino acid sequence corresponding to an apparent signal/activation peptide, which, like that of trypsin, ends in Gly immediately before the start of the presumed catalytic domain. The 1174 base pair DMP cDNA correlates closely with the 1180 base pair average size of DMP mRNA estimated from Northern blots (Figure 4), suggesting that the cDNA is nearly full length. In a protein data base search, no highly similar serine proteases were identified. The most favorable similarity score was reached when the DMP catalytic domain was compared with the sequence of human plasma kallikrein, with which it shares 40.8% sequence identity in areas of overlap. The extent of identity between the deduced

Table I: Comparison of Activation Sequence Hydrolysis Sites

protease	activation sequence ^a	ref
mast cell tryptase	...APQALQRVG IVGG...	
DMP	...LLGTLSPKVG IVGG...	
mast cell chymase (RMCP-II)	...LLPSGAGAE IIGG...	Benfey et al. (1987)
neutrophil elastase	...LLGGTALASE IVGG...	Takahashi et al. (1988)
neutrophil cathepsin G	...LLPTGAEAGE IIGG...	Salveson et al. (1987)
lymphocyte CCP-I	...SLASRTKAGE IIGG...	Lobe et al. (1986)
lymphocyte granzyme C	...LLPLRAGAE IIGG...	Jenne et al. (1988b)
lymphocyte granzyme D	...LLPLRAGAE IIGG...	Jenne et al. (1988a)
lymphocyte granzyme E	...LLPLGAGAE IIGG...	Jenne et al. (1988a)
lymphocyte granzyme F	...LLPLRAGAE IIGG...	Jenne et al. (1988a)
pancreatic kallikrein	...RNDAAPPQSR VVGG	Swift et al. (1982)
pancreatic elastase	...TQDFPETNAR VVGG...	Swift et al. (1984)
chymotrypsin B	...IQPVLTGLSR IVNG...	Bell et al. (1984)
trypsin	...VATPTDDDDK IVGG	Pinsky et al. (1985)
lymphocyte HF	...LLIPEDVCEK IIGG...	Gershensfeld et al. (1988)

^a The suggested C-terminal amino acid of activation peptides is shown in bold type; the hydrolysis site is indicated by a space; the first four residues of catalytic domains are shown to the right of the site of hydrolysis.

sequence of the catalytic domains of DMP and of dog tryptase (Figure 3) is greater (53.4%) than that between DMP and other known proteases. Thus, DMP cDNA appears to code for a unique serine protease which is not closely related in primary structure to previously reported serine proteases and which has some potentially significant similarities to dog tryptase.

DISCUSSION

We have determined the complete amino acid sequence of dog mast cell tryptase and of a potentially related serine protease, DMP. The sequence of dog tryptase and DMP reveals features shared with other members of the family of serine proteases and also novel differences. In particular, the apparent preprosequences of tryptase and DMP suggest that the two enzymes may form a new subgroup of serine proteases related by their modes of activation. Although tryptase and DMP appear to be more closely related to each other than to other published serine protease sequences and may be activated in a similar fashion, their catalytic domains are sufficiently dissimilar to suggest that the mature enzymes differ in their biological roles.

Signal and Activation Peptides. Translation of tryptase mRNA most likely starts at the ATG beginning 90 bases 5' to the codon for Ile-1 of the catalytic domain. This isoleucine has been shown to be the N-terminal residue of mature tryptase by amino acid sequence analysis (Caughey et al., 1987) and is analogous to Ile-16 at the N-terminus of chymotrypsin. The presumed initiation codon is preceded by the nucleotide sequence AAG, which is identical with the sequence preceding the presumptive start sites of several other granule-associated serine proteases, including human cathepsin G (Salveson et al., 1987), human lymphocyte protease 1-3E (Schmid & Weissman, 1987), and the mouse T-lymphocyte "granzymes" B (Lobe et al., 1986), C (Jenne et al., 1988a), and D and F (Jenne et al., 1988b).

By analogy to other mammalian serine proteases without large N-terminal regulatory domains, the 30 residues of tryptase preceding Ile-1 of the catalytic domain are expected to form a signal peptide and an activation peptide. The candidate preprosequence contains a preponderance of hydrophobic residues concentrated in its N-terminal half and terminates in a novel fashion with a C-terminal glycine. The hydrophobic residues are strongly suggestive of a signal sequence. However, application of a weight-matrix method (von Heijne, 1986) for prediction of the signal peptidase cleavage site does not select one dominant residue as the site of hy-

drolisis. One potential site of signal peptidase cleavage is Gly -1. Hydrolysis of the peptide bond at this residue suggests that tryptase with a mature catalytic domain is generated in the endoplasmic reticulum and is consistent with the finding that tryptase is active within mast cell granules (Caughey et al., 1988c) and is not further activated from precursors during stimulated mast cell degranulation (Caughey et al., 1988a). Although an artificial system has been designed for the direct activation of trypsinogen by bacterial signal peptidase (Vasquez et al., 1989), direct activation of a mammalian serine protease by signal peptidase is not known to occur naturally. Furthermore, cleavage at Gly -1 violates the "-3,-1 rule" by placing an arginine three residues N-terminal to the site of cleavage (von Heijne, 1986). Therefore, it seems more likely that the site of signal peptidase cleavage is closer to the N-terminus of the tryptase precursor, perhaps at Pro -11 or Gln -4, leaving a prosequence terminating in Gly -1 to be removed by a subsequent step or steps.

As illustrated in Table I, the amino acid in the -1-position of the tryptase sequence is neither the basic residue (Lys or Arg) found in the zymogen forms of pancreatic serine proteases nor the acidic residue (Glu) found in the short prosequences of granule-associated serine proteases such as neutrophil elastase, cathepsin G, the T-lymphocyte "granzymes", and the rat mast cell chymase, RMCP-II. Indeed, the difference between the preprosequence of tryptase and of mast cell chymase suggests that the two enzymes differ in their mode of processing and activation, although studies in dog cells suggest that tryptase is released with chymase during mast cell stimulation (Caughey et al., 1988a) and immunogold studies in human tissues suggest that the two enzymes are packaged in the same granules (Craig et al., 1988). Further investigation is required to define the site of cleavage of the tryptase precursor by signal peptidase and the identity of the residues, if any, constituting the activation peptide.

The translation of DMP probably starts with base 148, at the first ATG from the 5' end (Figure 1). The size of the mRNA estimated from Northern blots suggests that DMP contains no large N-terminal domains. An open reading frame 3' to the ATG codes for an apparent 19 amino acid leader sequence preceding the Ile which corresponds to the first residue of the catalytic domain of tryptase and other serine proteases. The leader sequence presumably codes for a signal/activation peptide. Its most notable feature is the C-terminal Lys-Val-Gly which is similar to the Arg-Val-Gly tripeptide which terminates the preprosequence of tryptase. The similarities between these DMP and tryptase tripeptides

raises the possibility that the prosequences of the two enzymes may be removed by a shared mechanism, presumably by a protease which cleaves the Gly-Ile bond in the sequence X-Val-Gly-Ile, where X in the P3 position is a basic residue (Table I).

Glycosylation Sites. The major bands of purified dog mastocytoma trypsin on polyacrylamide gel electrophoresis decrease in apparent molecular weight after treatment with peptide:N-glycosidase F, although some microheterogeneity persists (data not shown). Thus, it seems likely that some of the size heterogeneity of dog trypsin, like that of the human enzyme (Cromlish et al., 1987), is due to variable N-glycosylation. The consensus N-linked glycosylation site at Asn-102 of dog trypsin is also present at the homologous residue in human plasma kallikrein (Chung et al., 1986) and factor XI (Fujikawa et al., 1986). The potential influence of N-linked glycans in determining the stability, substrate affinity, and organelle targeting of trypsin remains to be investigated.

The consensus N-linked glycosylation sites of DMP predict that the mature protein is more heavily glycosylated than trypsin, assuming that the DMP signal sequence directs the nascent protein into the endoplasmic reticulum as discussed above.

Cysteines. Eight of the nine cysteines of trypsin align with the cysteines of trypsin (see Figure 3) and presumably form disulfide linkages. The extra residue (Cys-52) appears likely to map to the surface of trypsin, based on the position of surface loops of other serine proteases of known tertiary structure (Craik et al., 1983). However, Cys-52 of trypsin does not appear to be involved in tetramer formation because trypsin migrates as a monomer on polyacrylamide gel electrophoresis under either reducing or nonreducing conditions (Caughey et al., 1987). Few serine proteases are known to contain an unpaired cysteine. However, two of the serine proteases of cytotoxic T-lymphocyte granules, granzymes B (CCPI) (Lobe et al., 1986) and F (Jenne et al., 1988b), contain an extra cysteine which does not appear to be involved in dimer formation. The position of the unpaired cysteines is not conserved among trypsin and the T-lymphocyte proteases, and their roles remain uncertain.

The DMP deduced sequence contains 12 cysteines, 8 of which correspond to those found in trypsin (Figure 3). Presumably, these eight cysteines are involved in disulfide pairs as in trypsin.

Catalytic Domains. The catalytic domains of trypsin and DMP are 245 and 250 amino acids, respectively. Both are large when compared to other serine proteases such as trypsin (223 amino acids) (Pinsky et al., 1985) and chymotrypsin (230 amino acids) (Bell et al., 1984). It seems unlikely that trypsin and DMP are C-terminally processed like RMCP II (Benfey et al., 1987) and neutrophil elastase (Takahashi et al., 1988). Rather, the difference appears to be due to the length of the surface loops in trypsin and DMP, particularly those mapping to residues 61-64 and 174-177 (chymotrypsinogen numbering) (see Figure 3). Although trypsin and DMP both contain large surface loops, these share little sequence identity with one another with the exception of a proline-rich hexapeptide segment, Thr-Pro-Leu-Pro-Pro-Pro (residues 148-153 in Figure 3). Neither trypsin nor DMP contains the conserved octapeptide (residues 24-31, chymotrypsinogen numbering) found in mast cell chymase and in most other granule-associated proteases of leukocytes (Caughey et al., 1988d), including the protease of tryptic specificity, HF/granzyme A (Gershenfeld et al., 1988), of cytolytic T-lymphocytes. The lack of this octapeptide provides further evidence that trypsin

is less closely related in evolution and, perhaps, function than are other leukocyte granule proteases described to date. It is of particular interest that the putative intron identified in clone 1 maps to an intron/exon boundary that would reside on the surface of the enzyme. This may be an example of a sliding junction that could evolve or lead to a functional length polymorphism in the serine protease family (Craik et al., 1983). However, it is unlikely that the product of clone 1 would lead to a catalytically active enzyme in its current state.

Because the dog mastocytoma from which the cDNA library was created was propagated in the skin of nude mice, there is a possibility that DMP is a serine protease originating from mouse cells excised with the tumor. Further studies are required to establish conclusively the cell and species of origin of DMP. However, with only 53% sequence identity to dog trypsin, DMP is not likely to be the species homologue of trypsin in the mouse.

Conformation of the Active Site. Trypsin has a characteristic pattern of substrate and inhibitor binding preferences, suggesting the presence of an extended substrate binding region. Residues in the P2, P3, and P4 positions of synthetic peptide substrates have a much greater influence on the kinetics of substrate hydrolysis by trypsin than by trypsin (Tanaka et al., 1983). Trypsin substrate preferences and the general resistance of the enzyme to circulating inhibitors of serine proteases presumably are due to a unique topography of surface residues near the active site. In this regard, it is interesting to consider the possible structural features of trypsin responsible for resistance to inactivation by soybean trypsin inhibitor, which is effective against trypsin and many other serine proteases. The crystal structure of porcine trypsin complexed with this inhibitor reveals 24 residues of the enzyme which appear to contact residues of the inhibitor (Sweet et al., 1974). Of the 24 residues, 18 are identical in trypsin. Lys-60 of trypsin, which with Asp-1 of soybean trypsin inhibitor forms the only ion pair outside of the primary substrate binding pocket, is Gly in trypsin (Figure 3). The phenolic group of trypsin's Tyr-151, which is thought to form a charge transfer complex with the guanidinium group of Arg-65 of the inhibitor, is Pro in trypsin. These differences may help to explain why soybean trypsin inhibitor is a poor inhibitor of trypsin.

Heparin Binding. The trypsin cDNAs code for a catalytic domain with 18 basic residues (5 lysines and 13 arginines). Some of these residues may bind to heparin which, in turn, plays an important role in maintaining the enzyme as a catalytically active tetramer (Schwartz & Bradford, 1986; Alter et al., 1987). The net cumulative charge of trypsin, however, is -3. This stands in striking contrast to RMCP I (Le Trong et al., 1987), the rat mast cell chymase which also binds strongly to heparin in the mast cell granule and is characterized by 34 basic residues (22 lysines and 12 arginines) and a net cumulative charge of +18. Thus, trypsin is much less cationic than RMCP I. In a model based on the crystal structure of RMCP II, the basic residues of RMCP I are distributed in two large patches on opposite sides of the molecule (Remington et al., 1988). In trypsin, the smaller number of basic residues and the net negative charge suggest that binding to heparin is more localized, perhaps to a few critical residues. A clear picture of the specific conformation of the trypsin active site, and of the location of basic residues on the surface of trypsin and their relation to tetramer formation, must await X-ray diffraction analysis of crystalline trypsin.

In summary, we have cloned and sequenced cDNAs coding for the complete amino acid sequence of dog mast cell trypsin and a second tryptic serine protease, DMP. The two sequences

suggest that these enzymes are related by a previously unidentified pathway of activation. The availability of these clones will facilitate further studies of the mechanism of activation and structure-function relationships of tryptase and related unique enzymes of biomedical interest.

ACKNOWLEDGMENTS

We thank Stephen C. Lazarus and Warren A. Gold for providing dog mastocytoma cells and Wilfred W. Raymond and Susan Ballinger for their excellent technical assistance.

REFERENCES

- Alter, S. C., Metcalfe, D. D., Bradford, T. R., & Schwartz, L. B. (1987) *Biochem. J.* 248, 821-827.
- Aviv, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A.* 69, 1408-1412.
- Beaucage, S. L., & Caruther, M. H. (1981) *Tetrahedron Lett.* 22, 1859-1862.
- Bell, G. I., Quinto, C., Quiroga, M., Valenzuela, P., Craik, C. S., & Rutter, W. J. (1984) *J. Biol. Chem.* 259, 14265-14270.
- Benfey, P. N., Yin, F. H., & Leder, P. (1987) *J. Biol. Chem.* 262, 5377-5384.
- Benton, W. D., & Davis, R. W. (1977) *Science* 196, 180-182.
- Bullock, W. O., Fernandez, J. M., & Short, J. M. (1987) *Biotechniques* 5, 376-379.
- Caughey, G. H., Viro, N. F., Ramachandran, J., Lazarus, S. C., Borson, D. B., & Nadel, J. A. (1987) *Arch. Biochem. Biophys.* 258, 555-563.
- Caughey, G. H., Lazarus, S. C., Viro, N. F., Gold, W. M., & Nadel, J. A. (1988a) *Immunology* 63, 339-344.
- Caughey, G. H., Leidig, F., Viro, N. F., & Nadel, J. A. (1988b) *J. Pharmacol. Exp. Ther.* 244, 133-137.
- Caughey, G. H., Viro, N. F., Calónico, L. D., McDonald, D. M., Lazarus, S. C., & Gold, W. M. (1988c) *J. Histochem. Cytochem.* 36, 1053-1060.
- Caughey, G. H., Viro, N. F., Lazarus, S. C., & Nadel, J. A. (1988d) *Biochim. Biophys. Acta* 952, 142-149.
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., & Rutter, W. J. (1979) *Biochemistry* 18, 5294-5299.
- Chung, D. W., Fujikawa, K., McMullen, B. A., & Davie, E. W. (1986) *Biochemistry* 25, 2410-2417.
- Craig, S. S., Schechter, N. M., & Schwartz, L. B. (1988) *Lab. Invest.* 58, 682-691.
- Craik, C. S., Rutter, W. J., & Fletterick, R. (1983) *Science* 220, 1125-1129.
- Craik, C. S., Choo, Q.-L., Swift, G. H., Quinto, C., MacDonald, R. J., & Rutter, W. J. (1984) *J. Biol. Chem.* 259, 14255-14264.
- Cromlish, J. A., Seidah, N. G., Marcinkiewicz, M., Hamelin, J., Johnson, D. A., & Chretien, M. (1987) *J. Biol. Chem.* 262, 1363-1373.
- Dente, L., Cesareni, G., & Cortese, R. (1983) *Nucleic Acids Res.* 11, 1645-1655.
- Franconi, G., Graf, P. D., Lazarus, S. C., Nadel, J. A., & Caughey, G. H. (1989) *J. Pharmacol. Exp. Ther.* (in press).
- Fujikawa, K., Chung, D. W., Hendrickson, L. E., & Davie, E. W. (1986) *Biochemistry* 25, 2417-2424.
- Gershenfeld, H. K., & Weissman, I. L. (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85, 1184-1188.
- Glennier, G. G., & Cohen, L. A. (1960) *Nature* 185, 846-847.
- Gruber, B. L., Schwartz, L. B., Ramamurthy, N. S., Irani, A.-M., & Marchese, M. J. (1988) *J. Immunol.* 140, 3936-3942.
- Gubler, U., & Hoffman, B. J. (1983) *Gene* 25, 263-269.
- Hyunh, T. V., Young, R. A., & Davis, R. W. (1985) in *DNA Cloning Techniques, A Practical Approach* (Glover, D., Ed.) pp 49-78, IRL Press, Oxford.
- Jenne, D., Rey, C., Haefliger, J.-A., Quao, B.-Y., Groscurth, P., & Tschopp, J. (1988a) *Proc. Natl. Acad. Sci. U.S.A.* 85, 4814-4818.
- Jenne, D., Rey, C., Masson, D., Stanley, K. K., Herz, J., Plaetinck, G., & Tschopp, J. (1988b) *J. Immunol.* 140, 318-323.
- Kido, H., Fukusen, N., & Katunuma, N. (1985) *Arch. Biochem. Biophys.* 239, 436-443.
- Le Trong, H., Parmelee, D. C., Walsh, K. A., Neurath, H., & Woodbury, B. G. (1987) *Biochemistry* 26, 6988-6994.
- Lipman, D. J., & Pearson, W. R. (1985) *Science* 227, 1435-1441.
- Lobe, C. G., Finlay, B. B., Paranchych, W., Paetkau, V. H., & Bleackley, R. C. (1986) *Science* 232, 858-861.
- Maier, M., Spragg, J., & Schwartz, L. B. (1983) *J. Immunol.* 130, 2352-2356.
- Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular cloning: A laboratory Manual*, 1st ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Mason, A. J., Evans, B. A., Cox, D. R., Shine, J., & Richards, R. I. (1983) *Nature* 303, 300-307.
- McMaster, G. K., & Carmichael, G. G. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74, 4835-4838.
- Okayama, H., & Berg, P. (1982) *Mol. Cell. Biol.* 2, 161-170.
- Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., & Sharp, P. A. (1986) *Annu. Rev. Biochem.* 55, 1119-1150.
- Pinsky, S. D., LaForge, K. S., & Scheele, G. (1985) *Mol. Cell. Biol.* 5, 2669-2676.
- Remington, S. J., Woodbury, R. G., Reynolds, R. A., Matthews, B. W., & Neurath, H. (1988) *Biochemistry* 27, 8097-8105.
- Salveson, G., Farley, D., Shuman, J., Przybyla, A., Reilly, C., & Travis, J. (1987) *Biochemistry* 26, 2289-2293.
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463-5467.
- Schechter, N. M., Slavin, D., Fetter, D., Lazarus, G. S., & Fraki, J. E. (1988) *Arch. Biochem. Biophys.* 262, 232-244.
- Schmid, J., & Weissman, C. (1987) *J. Immunol.* 139, 250-256.
- Schwartz, L. B., & Bradford, T. R. (1986) *J. Biol. Chem.* 261, 7372-7379.
- Schwartz, L. B., Lewis, R. A., & Austen, K. F. (1981a) *J. Biol. Chem.* 256, 11939-11943.
- Schwartz, L. B., Lewis, R. A., Seldin, D., & Austen, K. F. (1981b) *J. Immunol.* 126, 1290-1294.
- Schwartz, L. B., Kawahara, M. S., Hugli, T. E., Vik, D., Fearon, D. T., & Austen, K. F. (1983) *J. Immunol.* 130, 1891-1895.
- Schwartz, L. B., Bradford, T. R., Littman, B. H., & Wintroub, B. U. (1985) *J. Immunol.* 135, 2762-2767.
- Sekizawa, K., Caughey, G. H., Lazarus, S. C., Gold, W. M., & Nadel, J. A. (1989) *J. Clin. Invest.* 83, 175-179.
- Smith, T. J., Hougland, M. W., & Johnson, D. A. (1984) *J. Biol. Chem.* 259, 11046-11051.
- Southern, E. M. (1975) *J. Mol. Biol.* 98, 503-517.
- Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H., & Blow, D. M. (1974) *Biochemistry* 13, 4212-4228.
- Swift, G. H., Dagorn, J.-C., Ashley, P. L., Cummings, S. W., & MacDonald, R. J. (1982) *Proc. Natl. Acad. Sci. U.S.A.* 79, 7263-7267.

- Swift, G. H., Craik, C. S., Stary, S. J., Quinto, C., Lahaie, R. G., Rutter, W. J., & MacDonald, R. J. (1984) *J. Biol. Chem.* 259, 14271-14278.
- Tabor, S., & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 4767-4771.
- Takahashi, H., Nukiwa, T., Yoshimura, K., Quick, C. D., States, D. J., Holmes, M. D., Whang-Peng, J., Knutsen, T., & Crystal, R. G. (1988) *J. Biol. Chem.* 263, 14739-14747.
- Tanaka, T., McRae, B. J., Cho, K., Cook, R., Fraki, J. E., Johnson, D. A., & Powers, J. C. (1983) *J. Biol. Chem.* 258, 13552-13557.
- Tarentino, A. L., Gomez, C. M., & Plummer, T. H., Jr. (1985) *Biochemistry* 24, 4665-4671.
- Vasquez, J., Evnin, L., Higaki, J., & Craik, C. S. (1989) *J. Cell. Biochem.* 39, 265-276.
- Von Heijne, G. (1986) *Nucleic Acids Res.* 14, 4683-4690.

pH Dependence of the Kinetic Parameters for the Pyrophosphate-Dependent Phosphofructokinase Reaction Supports a Proton-Shuttle Mechanism[†]

Yong-Kweon Cho and Paul F. Cook*

Department of Biochemistry and Department of Microbiology and Immunology, Texas College of Osteopathic Medicine/University of North Texas, Fort Worth, Texas 76107

Received October 3, 1988; Revised Manuscript Received December 14, 1988

ABSTRACT: The pH dependence of kinetic parameters for the pyrophosphate-dependent phosphofructokinase from *Propionibacterium freudenreichii* suggests that the enzyme catalyzes its reaction via general acid-base catalysis with the use of a proton shuttle. The base is required unprotonated in both reaction directions. In the direction of fructose 6-phosphate phosphorylation the base accepts a proton from the hydroxyl at C-1 of F6P and then donates it to protonate the leaving phosphate. Whether this occurs in one or two steps cannot be deduced from the present data. The maximum velocity of the reaction is pH independent in both reaction directions while V/K profiles exhibit pK_s for binding groups (including enzyme and reactant functional groups) as well as pK_s for enzyme catalytic groups. These data suggest that reactants bind only when correctly protonated and only to the correctly protonated form of the enzyme. Specifically, the requirement for two enzyme ϵ -amino groups in the protonated form for reactant binding was detected as was the requirement for the ionized phosphates of fructose 6-phosphate, fructose 1,6-bisphosphate, $MgPP_i$ and HPO_4^{2-} . The protonation state of enzyme and reactant binding groups is in agreement with data obtained previously [Cho, Y.-K., & Cook, P. F. (1988) *J. Biol. Chem.* 263, 5135].

The pyrophosphate-dependent phosphofructokinase from *Propionibacterium freudenreichii* catalyzes the phosphorylation of the hydroxyl at C-1 of β -D-fructose 6-phosphate to produce β -D-fructose 1,6-bisphosphate (Reeves et al., 1974). This enzyme is dimeric and is not regulated by either substrates or allosteric effectors (O'Brien et al., 1975).

It has been shown by initial velocity studies (Bertagnoli & Cook, 1984) that the enzyme has a rapid equilibrium random kinetic mechanism. More recently, Cho et al. (1988) have shown by using isotope exchange techniques that the phosphoryl-transfer step is likely the rate-determining step in the overall mechanism.

Little is known of the chemical interconversion of reactants and products for the pyrophosphate-dependent enzyme. The pH dependence of the enzyme-reactant dissociation constants has been obtained (Cho & Cook, 1988) by making use of protection by reactants against the modification of two active-site lysine residues by pyridoxal 5'-phosphate. One of the lysines is apparently in the vicinity of the 6-phosphate of the sugar reactants, while the second is in the vicinity of the phosphate to be transferred from $MgPP_i^1$ to the C-1 hydroxyl of F6P. For optimum binding of reactants, both lysines must be ionized as well as $MgPP_i$ and the phosphates at C-1 and

C-6 of the sugar phosphates. Inorganic phosphate apparently bind as HPO_4^{2-} .

With the above background on the optimum protonation state of active-site lysines and reactants available, it is an easier task to determine the acid-base chemistry catalyzed by the enzyme via pH-rate profiles. There is no evidence for nucleophilic catalysis by enzyme, and thus two possible mechanisms are suggested. The first involves general base-general acid catalysis in which the general base accepts a proton from the hydroxyl of C-1 concomitant with nucleophilic attack of the hydroxyl oxygen on pyrophosphate while the general acid would protonate the leaving group phosphate (Scheme IA). A second possibility is general acid-base catalysis in which a proton shuttle accepts a proton from the hydroxyl at C-1 and donates it to the leaving group phosphate either in the same step (Scheme IB) or in a subsequent step (Scheme IC). In Scheme IB, the general base could be pictured as, for example, an imidazole, accepting a proton on one imidazole nitrogen and donating the proton via the other in the same step. The mechanism depicted in Scheme IA sug-

[†]This work was supported by grants to P.F.C. from NIH (GM 36799) and the Robert A. Welch Foundation (B-1031) and by Grant BRSG SO7 RR07195 awarded by the Biomedical Research Grant Program, Division of Research Resources, National Institutes of Health, and a Faculty Research Grant from the University of North Texas.

¹ Abbreviations: Ches, 2-(N-cyclohexylamino)ethanesulfonic acid; Mes, 2-(N-morpholino)ethanesulfonic acid; Pipes, piperazine-N,N'-bis-(2-ethanesulfonic acid); Taps, 3-[[tris(hydroxymethyl)methyl]amino]-propanesulfonic acid; PP_i -PFK, pyrophosphate-dependent phosphofructokinase; F6P, fructose 6-phosphate; FBP, fructose 1,6-bisphosphate; PP_i , inorganic pyrophosphate; P_i , inorganic phosphate; PCP, methylene diphosphonate; PLP, pyridoxal 5'-phosphate; HBP, 1,6-hexanediol bisphosphate; T6P, tagatose 6-phosphate.