Blake, C. C. F., & Rice, D. W. (1981) *Philos. Trans. R. Soc. London, A 293*, 93–104.
Blostein, R., & Rutter, W. J. (1963) *J. Biol. Chem. 238*, 3280–3285.
Bradford, M. M. (1976) *Anal. Biochem. 72*, 248–254.
Buczyłko, J., Palczewski, K., & Kochman, M. (1983) *Int. J. Biochem. 15*, 453–461.
Chen, P. S., Toribara, T. Y., & Warner, H. (1956) *Anal. Chem. 28*, 1756–1758.
Crestfield, A. M., Moore, S., & Stein, W. H. (1963) *J. Biol. Chem. 238*, 622–627.
Drechsler, E. R., Boyer, P. D., & Kowalsky, A. G. (1959) *J. Biol. Chem. 234*, 2627–2634.
Eklund, M., Samama, J. P. & Jones, T. A. (1984) *Biochemistry 23*, 5982–5996.
Evans, P. R., Farrants, G. W., & Hudson, P. J. (1981) *Philos. Trans. R. Soc. London, A 293*, 53–62.
Felicioli, R., Nannicini, L., Balestreri, E., & Montagnoli, G. (1975) *Eur. J. Biochem. 51*, 467–474.
Ginsburg, A., & Mehler, A. H. (1966) *Biochemistry 5*, 2623–2634.
Gray, W. R. (1967) *Methods Enzymol. 11*, 139–151.
Holbrook, J. J., Liljas, A., Steindel, S. J., & Rossmann, M. G. (1975) *Enzymes (3rd Ed.) 11*, 191–292.
Kasprzak, A. A., & Kochman, M. (1980a) *Eur. J. Biochem. 104*, 443–450.
Kasprzak, A. A., & Kochman, M. (1980b) *Biochim. Biophys. Acta 612*, 455–459.
Kasprzak, A. A., & Kochman, M. (1981) *J. Biol. Chem. 256*, 6127–6133.
Kawahara, K., & Tanford, C. (1966) *Biochemistry 5*, 1579–1584.
Kochman, M., & Mas, M. T. (1981) *Biochim. Biophys. Acta 667*, 218–222.
Lai, C. Y. (1968) *Arch. Biochem. Biophys. 128*, 202–211.
Lai, C. Y. (1975) *Arch. Biochem. Biophys. 166*, 358–368.
Matteuzzi, M., Bellini, T., Bergamim, C. M., & Dallocchio, F. (1985) *Biochem. Int. 10*, 53–61.
Michal, G., & Beutler, H. O. (1974) in *Methods in Enzymatic Analysis* (Bergmeyer, H. U., Ed.) Vol. 3, pp 1314–1319, Verlag Chemie, Weinheim, FRG, or Academic, New York or London.
Palczewski, K. (1986) Ph.D. Thesis, Technical University of Wrocław.
Palczewski, K., Hargrave, P. A., & Kochman, M. (1983) *Eur. J. Biochem. 137*, 429–435.
Palczewski, K., Hargrave, P. A., Folta, E. J., & Kochman, M. (1985) *Eur. J. Biochem. 146*, 309–314.
Penhoet, E. E., Kochman, M., & Rutter, W. J. (1969) *Biochemistry 8*, 4391–4395.
Rossmann, M. G., Liljas, A., Branden, C.-I., & Banaszak, L. J. (1975) *Enzymes (3rd Ed.) 11*, 61–102.
Schäfer, H. J., Scheurich, P., Rathgeber, G., & Dose, K. (1978) *Nucleic Acids Res. 5*, 1345–1351.
Schettino, C. M., Lima, D. F., Leyton, J. F., El-Dorry, H. A., & Bacila, M. (1981) *Biochim. Biophys. Acta 667*, 411–420.
Shapiro, S., Enser, M., Pugh, E. & Horecker, B. L. (1968) *Arch. Biochem. Biophys. 128*, 554–562.
Spolter, P. D., Adelman, R. C., & Weinhouse, S. (1965) *J. Biol. Chem. 240*, 1327–1337.
Stellwagen, E. (1976) *J. Mol. Biol. 106*, 903–911.
Tolan, D. R., Amsden, A. B., Putney, S. D., Urdea, M. S., & Penhoet, E. E. (1984) *J. Biol. Chem. 259*, 1127–1131.

Amino Acid Sequence of Guinea Pig Prostate Kallikrein†

Joan C. Dunbar‡ and Ralph A. Bradshaw*
Department of Biological Chemistry, California College of Medicine, University of California, Irvine, California 92717
Received November 7, 1986; Revised Manuscript Received January 7, 1987

ABSTRACT: The primary structure of the major arginine esteropeptidase from guinea pig prostate has been deduced from automated Edman degradation of peptides generated by clostripain, cyanogen bromide, endoproteinase Lys-C, and *Staphylococcus aureus* V8 protease digestion of the protein. The esteropeptidase is a single polypeptide chain comprised of 239 amino acids and contains 2 apparent sites of carbohydrate attachment, Asn-78 and Asn-169. Both occur in consensus sequences for N-linked glycosylation sites. The esteropeptidase exhibits approximately 35% homology with trypsin including conservation of the catalytic residues and the aspartic acid which confers specificity toward basic amino acids. The sequence identity, however, extends to greater than 60% with the kallikrein family of serine proteases. In addition to the overall homology, the guinea pig enzyme displays a number of features characteristic of kallikreins including 10 conserved half-cystine residues, a C-terminal proline, and the "kallikrein loop". On the basis of this structural relatedness, the enzyme has been designated as guinea pig prostate kallikrein. In contrast to many of the kallikreins of other species and tissues, this enzyme does not contain any sites within the kallikrein loop sensitive to proteases that result in internal breaks in the polypeptide chain.

Glandular kallikreins are a distinct subset of the serine protease family of esteropeptidases (Schachter, 1980). They preferentially hydrolyze synthetic substrates composed of esters and amides of arginine and, to a lesser extent, lysine (Fielder, 1979). In contrast to trypsin, kallikreins have very little general protease activity but rather display a high degree of substrate selectivity (Schachter, 1969). These enzymes were originally defined as kininogenases with the capacity, in vitro, to release kinin from kininogen. However, the kininogenase activity varies considerably among members of the kallikrein family, and many esteropeptidases have now been identified as kal-

† This work was supported by USPHS Research Grants NS19964 and DK32465 (formerly AM32465).
* Correspondence should be addressed to this author.
‡ Present address: Department of Molecular Biology and Genetics, Wayne State University School of Medicine, Detroit, MI 48207.

0006-2960/87/0426-3471$01.50/0

likreins on the basis of their strict substrate preference and/or homologous amino acid sequences (Fielder, 1979). Although the physiological substrate of many of these enzymes remains to be determined, it has been proposed that some may play a specific role in the proteolytic processing of polypeptide hormones (Mason et al., 1983; Shine et al., 1983).

Two growth factors of the adult male mouse submandibular gland, β-nerve growth factor (β-NGF)[1] and epidermal growth factor (EGF), occur as complexes with the kallikrein-like enzymes γ-NGF (Greene et al., 1968; Thomas et al., 1981a) and EGF binding protein (EGF-BP) (Taylor et al., 1974), respectively. Both β-NGF (Scott et al., 1983a; Ullrich et al., 1983) and EGF (Gray et al., 1983; Scott et al., 1983b) are synthesized as large precursors, and since the association of the respective kallikreins with the growth factors is highly specific, it has been proposed that γ-NGF and EGF-BP are directly involved in the processing of the growth factor precursors (Angeletti & Bradshaw, 1971; Berger & Shooter, 1977; Bothwell et al., 1979; Frey et al., 1979).

β-NGF and EGF are also found in relatively high concentrations in guinea pig prostate (Harper et al., 1979; Rubin, 1983), and we have examined this tissue for arginine esteropeptidases that could interact with guinea pig prostate β-NGF (Dunbar & Bradshaw, 1985). A single serine protease was detected by [^{3}H]DFP labeling of the soluble proteins in a crude tissue extract. The purified esterase exhibits a high degree of specificity toward arginine methyl ester substrates but very little general protease activity. In order to further investigate a role of the esteropeptidase in the processing of the growth factors in guinea pig prostate, we have determined the primary structure of the enzyme and have ascertained that it is a member of the kallikrein family.

Materials and Methods

Guinea pig prostate esterase was isolated from frozen tissue and assayed as previously described (Dunbar & Bradshaw, 1985). Clostripain and *Staphylococcus aureus* V8 protease were from Sigma Chemical Co. Endoproteinase Lys-C from *Lysobacter enzymogens* was obtained from Boehringer Mannheim Biochemicals. Cyanogen bromide and iodoacetic acid were from Sigma Chemical Co., and the iodoacetic acid was recrystallized before use. Chemicals used for the gas-phase sequencer were from Applied Biosystems, and HPLC-grade solvents were from Fisher. Sequenal grade trifluoroacetic acid was purchased from Pierce. All other chemicals were of reagent grade.

Generation of Peptides. Enzymatic and chemical cleavages were performed after reductive alkylation of the guinea pig esterase with iodoacetic acid, essentially as described by Allen (1981). At the completion of the reaction, the protein was separated from the reagents by desalting on a Pharmacia PD-10 (Sephadex G-25) column. The protein was eluted from the column with 1% acetic acid and subsequently lyophilized. Arginine peptides were obtained by cleavage of the S-carboxymethylated esterase (CM-esterase) with clostripain. Prior to the reaction, clostripain was activated by preincubation for 4 h at room temperature with 2.5 mM dithiothreitol and 1.0 mM calcium acetate after which the enzyme was diluted 10-fold into the reaction mixture. The digestion was carried out in 75 mM sodium phosphate, pH 7.8, at an enzyme to substrate ratio of 1:100 for 5 h at room temperature. At the completion of the reaction, the sample was lyophilized, and the peptides subsequently were separated by reverse-phase HPLC as described in the text.

The CM-esterase was digested with *S. aureus* V8 protease (2% w/w) in 0.1 M NH_4HCO_3, pH 7.8. After incubation for 5 h at 37 °C, the reaction was terminated by the addition of excess PMSF. Cleavage of the protein with endoproteinase Lys-C was performed for 6 h at 37 °C in 20 mM Tris-HCl, pH 7.8, containing 1 mM EDTA. The substrate concentration was 5 mg/mL, and the endoproteinase Lys-C concentration was 0.025 mg/mL.

Cleavage at methionine residues by cyanogen bromide was carried out as described by Thomas et al. (1981b).

Sequencing Procedures. The amino-terminal sequence of the CM-esterase was determined in a Beckman 890C spinning-cup sequencer using a 0.1 M Quadrol program. The peptide fragments were analyzed by automated Edman degradation in an Applied Biosystems 470A gas-phase sequencer. The phenylthiohydantoin derivatives were identified by reverse-phase HPLC with a Hewlett-Packard Model 1084B chromatograph equipped with an Altex Ultrasphere ODS column. The aqueous phase was 15 mM sodium phosphate, pH 5.5, and the PTH–amino acids were eluted with a gradient of methanol/acetonitrile (17:3 v/v).

Amino Acid Analyses. The amino acid composition was determined after hydrolysis of the protein and peptides in 6 N HCl for 24 h at 110 °C, in vacuo. The amino acid analyses were carried out on a Durrum D-500 amino acid analyzer.

Results

The strategy used for determining the primary structure of guinea pig esterase is shown in Figure 1. The amino acid sequence was derived predominantly from peptides obtained from clostripain, endoproteinase Lys-C, and cyanogen bromide cleavages of the esterase. Digestion with *S. aureus* V8 protease yielded fragments which provided essential overlaps. Overlaps were obtained for all peptides except between residues 98 and 99 where the order of the peptides was deduced from the sequence homology of the guinea pig enzyme to serine proteases (vide infra). The peptides have been designated and numbered according to the type of digest and the theoretical order in which they appear in the sequence; i.e., peptides were given a designation whether or not they were isolated (see Figure 1).

The amino acid composition calculated from the sequence of the guinea pig esterase is shown in Table I and is in good agreement with that determined from amino acid analyses of the hydrolyzed protein. The protein contains 239 amino acids with a calculated molecular weight of 26 300 which is in accordance with the molecular weight determined from SDS–polyacrylamide gel electrophoresis of the deglycosylated protein (Dunbar & Bradshaw, 1985).

Automated Edman degradation of the CM-esterase yielded the sequence of the first 36 amino acids (Figure 1). Only one N-terminal sequence was apparent which is consistent with the enzyme being comprised of a single polypeptide chain. The complete amino acid sequence is shown in Figure 2.

Clostripain Peptides. A significant proportion of the sequence was derived from peptides obtained by digestion of the CM-esterase with clostripain (Figure 1). The digestion products were fractionated directly on a Vydac C4 reverse-phase HPLC column, and the peptides were eluted with a

[1] Abbreviations: NGF, nerve growth factor; EGF, epidermal growth factor; EGF-BP, epidermal growth factor binding protein; DFP, diisopropyl fluorophosphate; PMSF, phenylmethanesulfonyl fluoride; PTH, phenylthiohydantoin; TFA, trifluoroacetic acid; HPLC, high-performance liquid chromatography; CM-esterase, S-carboxymethylated esterase; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride; EDTA, ethylenediaminetetraacetic acid; SDS, sodium dodecyl sulfate; ELC, endoproteinase Lys-C; pPK, porcine pancreatic kallikrein; gpPK, guinea pig prostate kallikrein; BAPNA, benzoylarginine-*p*-nitroanilide.

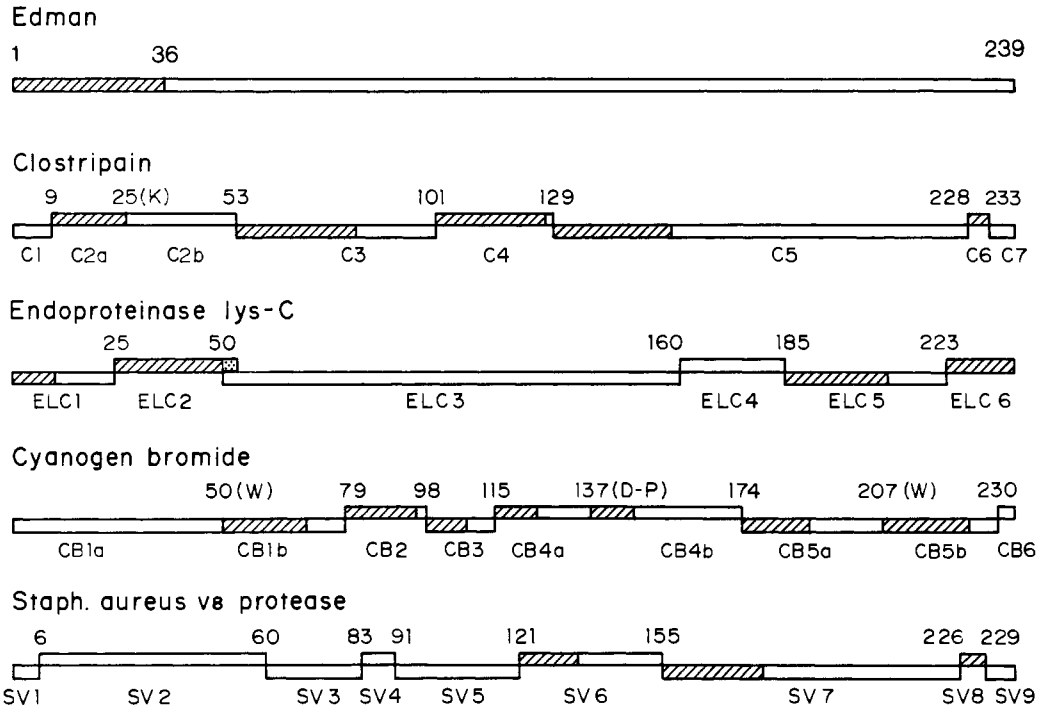

FIGURE 1: Strategy used to determine the amino acid sequence of guinea pig prostate esteropeptidase (kallikrein). The peptides expected from each of the cleavages are indicated and are numbered continuously whether they were isolated or not; the hatched areas represent the portion of the sequence determined by automated Edman degradation. The stipled area beyond residue 50 in the endoproteinase Lys-C digest represents the alternate peptide arising from the allotypic replacement at position 50 (see text). Cleavage at Lys-25 in the clostripain digest is indicated by (K); cleavages at residues other than methionine in the cyanogen bromide digest are indicated: W, tryptophan; P–D, acid cleavage at an Asp–Pro bond.

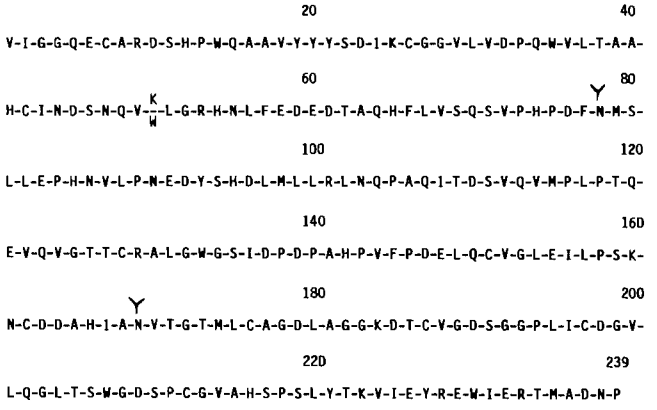

FIGURE 2: Complete amino acid sequence of guinea pig prostate esteropeptidase. The symbol Y represents potential sites of carbohydrate attachment at Asn-78 and Asn-169.

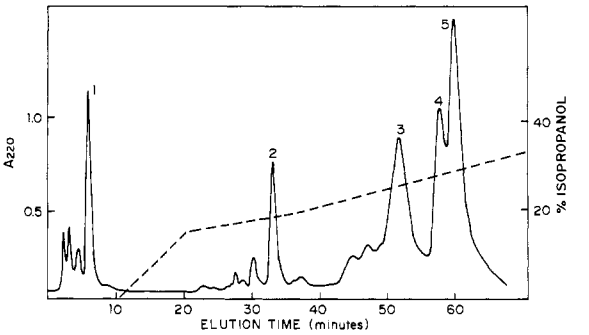

FIGURE 3: Fractionation of the peptides obtained from a clostripain digestion of CM-esteropeptidase on a Vydac C4 reverse-phase HPLC column (4.6 × 250 mm). The column was equilibrated with 0.1% TFA, and the peptides were eluted with a gradient of 2-propanol, as shown. The flow rate was 0.75 mL/min.

gradient of 2-propanol as shown in Figure 3. Five major peaks were obtained, yielding five of the seven peptides predicted from the amino acid composition (six arginines). The major peak in the flow through (peak 1) contained a small pentapeptide, C6, which was sequenced in its entirety and is derived from residues 229–233. The peptides that eluted in the fractions corresponding to peaks 2 (peptide C4), 3 (peptide C3), and 5 (peptide C5) (Figure 3) were also of sufficient purity to sequence directly without further purification. Their amino acid compositions are shown in Table I.

Peak 2 corresponded to peptide C4 (residues 102–129), and 26 of the 28 residues were sequenced. Peptide C3 appeared in peak 3, and automated Edman degradation yielded the sequence of 29 of the 48 residues. Cycle 25 was blank with no detectable PTH–amino acid while the two subsequent cycles yielded Met-Ser. This is consistent with the consensus sequences (Asn-X-Ser/Thr) for an asparagine-linked carbohydrate side chain (Waechter & Lennarz, 1976). The amino acid composition of C3 (Table I) is also indicative of nine aspartic acid residues after acid hydrolysis of the peptide whereas only eight are predicted by the sequence. Cycle 25 of this peptide, corresponding to residue 78, was therefore identified as an asparagine residue (linked to a carbohydrate side chain).

Peak 5 from the HPLC separation of the clostripain fragments contained a long peptide (C5) of ~99 residues of which 28, encompassing amino acids 130–157, were sequenced. Peak 4 (Figure 3) was comprised predominantly of peptide C2 corresponding to cleavage at Arg-9. Sixteen residues were sequenced which corresponded to residues 10–25 initially identified by Edman degradation of the intact protein. Peak 4 also contained a minor fragment with an N-terminal sequence of Cys-Gly-Gly-. This peptide was subsequently identified as corresponding to residues 26–50 (or 53). It also occurred in the endoproteinase Lys-C digest (from which it was readily purified and sequenced) and apparently arose from cleavage at Lys-25 by clostripain. Neither the N-terminal

Table I: Amino Acid Composition of Guinea Pig Esteropeptidase and Derived Peptides[a]

residue	esteropeptidase	ELC1 (01–25)[b]	T4 (26–53)[b]	C3 (54–101)[b]	CB2 (80–98)[b]	C4 (102–129)[b]	C5 (130–228)[b]	SV7 (156–226)[b]	ELC5 (185–223)[b]	ELC6 (224–239)[b]
aspartic acid	31.9 (32)	2.1 (2)	3.7 (4)	8.6 (9)	3.6 (4)	1.9 (2)	12.4 (12)	8.8 (9)	3.7 (4)	2.0 (2)
threonine	12.3[b] (12)	0.2 (0)	1.1 (1)	1.3 (1)	0.1 (0)	3.2 (4)	5.4 (5)	4.7 (5)	2.9 (3)	0.9 (1)
serine	14.1[b] (15)	2.1 (2)	1.5 (1)	4.0 (4)	1.9 (2)	1.0 (1)	6.2 (7)	5.7 (6)	4.2 (5)	0.1 (0)
glutamic acid	23.9 (24)	3.2 (3)	2.4 (2)	6.3 (6)	2.1 (2)	5.4 (6)	5.1 (5)	2.4 (2)	1.4 (1)	3.2 (3)
proline	17.6 (18)	0.9 (1)	1.2 (1)	3.5 (4)	1.8 (2)	2.7 (3)	7.9 (8)	3.7 (4)	3.1 (3[c])	0.8 (1)
glycine	19.4 (19)	2.4 (2)	2.6 (3)			1.0 (1)	14.1 (14)	11.4 (11)	7.1 (7)	0.1 (0)
alanine	14.3 (15)	2.6 (3)	1.5 (2)	1.1 (1)		1.1 (1)	6.7 (7)	4.8 (5)	1.1 (1)	1.1 (1)
half-cystine	9.2[c] (10)	g (1)	g (2)			g (1)	g (6)	g (5)	g (3)	
valine	19.5[d] (20)	1.7 (2)	2.7 (4)	3.1 (3)	0.8 (1)	3.2 (4)	6.4 (7)	4.1 (5)	2.6 (3)	0.6 (1)
methionine	4.2[e] (5)			1.4 (2)	ND (1)	0.6 (1)	0.4 (1)	ND (1)		0.5 (1)
isoleucine	9.4[d] (10)	1.7 (2)	1.4 (1)		0.2 (0)	0.9 (1)	5.9 (5)	3.2 (4)	0.9 (1)	1.7 (2)
leucine	23.4[d] (23)	0.3 (0)	3.2 (0)	7.6 (8)	3.9 (4)	1.7 (2)	9.8 (10)	6.9 (7)	3.5 (4)	
tyrosine	4.1 (6)	2.1 (3)		0.5 (1)	0.6 (1)		1.3 (2)	0.7 (1)	0.6 (1)	0.7 (1)
phenylalanine	3.4 (4)			2.5 (3)			0.5 (1)			
lysine	4.1 (4)	0.7 (1)	0.4 (1)				2.6 (3)	2.9 (3)	0.8 (1)	
histidine	9.5 (10)	0.9 (1)	0.9 (1)	4.9 (5)			3.1 (3)	2.1 (2)	0.8 (1)	
arginine	5.5 (6)	0.9 (1)	0.7 (1)	0.8 (1)	1.7 (2)	0.7 (1)	1.2 (1)	0.2 (0)		1.9 (2)
tryptophan	5.4[f] (6)	ND (1)	ND (1)				ND (2)	ND (1)	ND (1)	ND (1)
no. of residues	239	25	28	48	19	28	99	71	39	16

[a] Determined from acid hydrolysates as described in the text. Peptides represented are as follows: C, clostripain; ELC, endoproteinase Lys-C; CB, cyanogen bromide; T, tryptic; SV, *Staph. aureus* V8 protease; ND, not determined. Values in parentheses are from the final sequence. [b] Serine and threonine determined by extrapolation to zero hydrolysis time. [c] Determined as cysteic acid after performic acid oxidation (Hirs, 1967). [d] Valine, isoleucine, and leucine values were obtained after 96-h hydrolysis. [e] Methionine was determined as methionine sulfone. [f] Tryptophan was estimated spectrophotometrically by the method of Edelhoch (1967). [g] Detected, but not quantitated, as *S*-(carboxymethyl)cysteine. [h] Residue numbers from sequence (Figure 2).

peptide (C1) nor the C-terminal peptide (C7) was recovered.

Endoproteinase Lys-C Digest. Cleavage at lysine residues was accomplished by incubation of the CM-esterase with endoproteinase Lys-C (ELC). Gel filtration of the digestion products on a column of Sephadex G-75, followed by reverse-phase HPLC of the partially fractionated peptides, resulted in the purification of four ELC peptides. The peptide corresponding to ELC2 initially eluted very close to the void volume of the Sephadex G75 column. After further purification by chromatography on a reverse-phase C4 HPLC column at approximately 35% 2-propanol, the peptide was sequenced for 26 residues. The N-terminus of the peptide corresponded to Cys-26, and the sequence included His-41 which is analogous to the histidine residue implicated as catalytically important in serine proteases (Blow et al., 1969). The amino acid in this peptide, corresponding to residue 50 in the sequence, was identified as tryptophan. However, a similar peptide encompassing residues 26–53 was isolated from a tryptic digest of the esterase (data not shown). The entire peptide was sequenced, as judged by comparison with the amino acid analysis, and in this case, residue 50 was identified as a lysine, albeit one which was not cleaved by trypsin. There appears, then, to be an allotypic variation in the sequence at this position. The ELC2 peptide may then correspond to residues 26–160 and represent an ELC2/ELC3 overlap. A fragment of this size would be consistent with the position at which the peptide eluted from the gel filtration column.

Peptide ELC5 which contained the active-site serine, analogous to that of other serine proteases, was composed of 39 residues; 25 amino acids were sequenced, encompassing residues 185–210.

Peptide ELC6 represents the C-terminal peptide, and 16 residues were sequenced. The amino acid composition calculated from the sequence is consistent with that obtained from amino acid analyses of the peptide (Table I), suggesting that the peptide was completely sequenced. The identification of ELC6 as the C-terminal peptide was based on homologous sequences in the serine proteases. Attempts to verify that this peptide represented the C-terminus by carboxypeptidase digestion were unsuccessful; no amino acids were released from the guinea pig esterase during incubation with carboxypeptidase Y or carboxypeptidase P (despite prior denaturation of the protein by performic acid oxidation, reductive alkylation, urea, or SDS). These exopeptidases were similarly inactive toward the ELC6 peptide, suggesting that this Asn-Pro sequence is resistant to hydrolysis by these proteases.

The remaining peptide purified from the ELC digest corresponded to ELC1, the N-terminal peptide. Its sequence confirmed the assignments from Edman degradation of the intact protein. Neither peptide ELC3 (residues 51–160) nor peptide ELC4 (residues 161–185) was recovered from the digest.

Cleavage at Methionine. Cyanogen bromide fragmentation of the CM-esterase yielded a number of peptides which were initially separated on a column of Sephadex G-75 (Figure 4a). SDS–polyacrylamide gel electrophoresis of pool I revealed the presence of three peptides with molecular weights ranging from 20000 to 30000, indicative of partial cleavage products. These may have arisen in part due to the Met-Ser sequence at positions 79–80; they were not further analyzed. The major peptides eluting in pools II and III were readily purified by reverse-phase HPLC on Vydac C4 and Supercosil LC-8 columns, respectively. Peak IV from the gel filtration column was very heterogeneous. Partial fractionation was achieved by chromatography on a C8 column (Figure 4b) which yielded

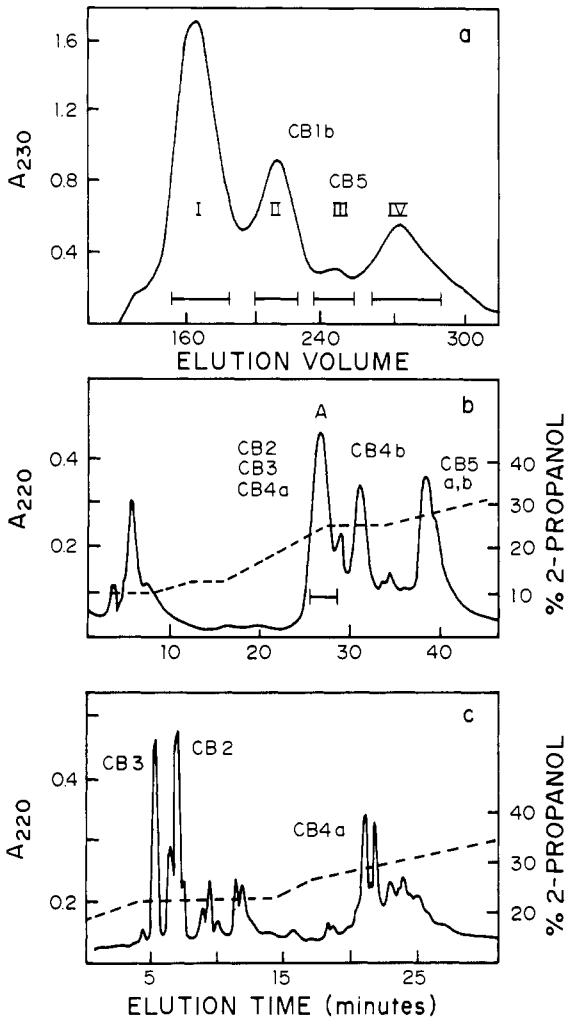

FIGURE 4: Purification of the cyanogen bromide peptides derived from guinea pig CM-esterase. (a) Fractionation of the digest on a column of Sephadex G-75 (2.5 × 90 cm). The column was equilibrated with 50 mM NH_4HCO_3, pH 7.8, and the peptides were eluted at a flow rate of 25 mL/h. The bars represent the fractions which were pooled from each peak for further chromatography. (b) Reverse-phase HPLC of the pool IV peptides on a column of Supercosil (Supelco) LC-8. The aqueous phase was 0.1% TFA in H_2O, and the column was initially equilibrated with 10% 2-propanol, and the peptides were subseqeuntly eluted with a gradient as indicated. The flow rate was 0.75 mL/min. CB4a and CB4b correspond to peptides arising from cleavage at Met-115 and Asp-137, respectively. CB5a represents residues 175–207, and CB5b corresponds to the peptide arising from cleavage at Trp-207. (c) Pool A from the C8 column was rechromatographed on a Vydac C18 column. The sample was applied at 15% 2-propanol (aqueous-phase 0.1% TFA) and eluted with a gradient of 2-propanol, as shown, at a flow rate of 0.75 mL/min.

peptides CB4b, CB5a, and CB5b. The fragments eluting in peak A from the C8 column were subsequently purified by chromatography on a C18 column (Figure 4c), ultimately producing peptides CB2, CB3, and CB4a, as indicated.

Cleavage at Met-79 and Met-98 gave rise to peptide CB2 of which 17 of the 19 amino acids were sequenced. The identity of the remaining two residues was determined from amino acid analyses of the peptide (Table I) while the sequence of these residues, Leu-Met, was deduced from homologous serine protease sequences. This region of homology was also used to juxtapose CB3 at the C-terminus of CB2. Ten residues of peptide CB3 (residues 99–115) were sequenced which provided a significant overlap with the clostripain peptide C4. A cyanogen bromide "sensitive" site also occurs within the CB4 peptide. This peptide contains the sequence Asp-Pro-Asp-Pro-Ala (residues 137–141) which is susceptible to acid hydrolysis by the 70% formic acid solvent during cyanogen bromide digestion. Cleavage occurs at both Asp–Pro bonds as judged by the copurification of peptides containing N-terminal sequences of Pro-Asp-Pro-Ala and Pro-Ala. (Only the cleavage at Asp-137 is shown in Figure 1.) CNBr fragment CB5 represents cleavage at Met-174. Seventeen residues of this peptide were sequenced which was useful as an overlap with the N-terminus of the ELC5 peptide (Figure 1).

Cleavage with CNBr also generated an additional peptide, arising from CB1, that corresponds to residues 51–79 (CB1b). This fragment was formed by cleavage at Trp-50, such cleavages being frequent occurrences during CNBr fragmentation of proteins (Blumenthal et al., 1975). In this case, however, it was fortuitous in that it provided a necessary overlap between the ELC2 and C3 peptides. Cleavage at Trp-207 also occurred during digestion with CNBr. It was difficult, however, to completely separate the resulting peptide (CB5b) from CB5a. Nonetheless, it was possible to subtract the minor sequence of CB5a at each cycle and thus generate the sequence corresponding to residues 208–228.

The esterase contains five methionines (Table I) which should yield six peptides on CNBr cleavage. However, the cleavages at 2 tryptophan residues (at 50 and 207) and 2 Asp-Pro sequences (located adjacent to each other and treated as one cleavage in Figure 1) results in 12 possible peptides (CB1, CB1a, CB1b, CB2, CB3, CB4, CB4a, CB4b, CB5, CB5a, CB5b, and CB6). Sequences were obtained for all except CB1, CB1a, and CB6.

S. aureus V8 Protease Peptides. The products of *S. aureus* V8 protease digestion were fractionated by chromatography on a C4 reverse-phase HPLC column. Three major peaks of absorbance were detected which contained peptides of sufficient purity for direct sequencing. A hexapeptide, SV8, encompassing residues 227–232 eluted at approximately 20% 2-isopropanol. Six residues were sequenced; the peptide contains a Glu–Trp bond which was not cleaved by the protease. The second major peak, eluting at approximately 28% 2-propanol, contained the SV6 fragment covering residues 122–155. Fifteen residues of this peptide were sequenced. Peptide SV7 was the third major *S. aureus* peptide isolated. The N-terminal sequence positioned this fragment within the clostripain peptide C5. The first 23 amino acids were sequenced. Cycle 14 (residue 169) also appeared as a blank cycle, while residues 170 and 171 were identified as Val-Thr. These observations are consistent with a second potential site of carbohydrate attachment and the identification of residue 169 as asparagine.

Discussion

The major soluble arginine esteropeptidase isolated from guinea pig prostate is comprised of a single polypeptide chain of 239 amino acids. The amino acid sequence of the first 36 amino acids was obtained by automated Edman degradation of the intact protein. Peptide ELC2, commencing at Cys-26, extends this sequence to residue 52. Residue 50 displayed an allotypic variation, appearing as either lysine or a tryptophan. Cyanogen bromide cleavage, occurring at Trp-50, yielded a peptide which provided an overlap between ELC2 and C3. Peptide C3 was sequenced through Leu-82 which provides the overlap to peptide CB2. As noted earlier, peptide CB3 was juxtaposed to CB2 on the basis of homology with other serine proteases; no overlap between these peptides was obtained. Peptide CB3 extends into the clostripain peptide C4 which is connected to C5 by the *S. aureus* peptide SV6. Peptide C5 was sequenced to Leu-157, while SV7 contributed the sequence of residues 156–179. Peptide CB5 provided the intervening

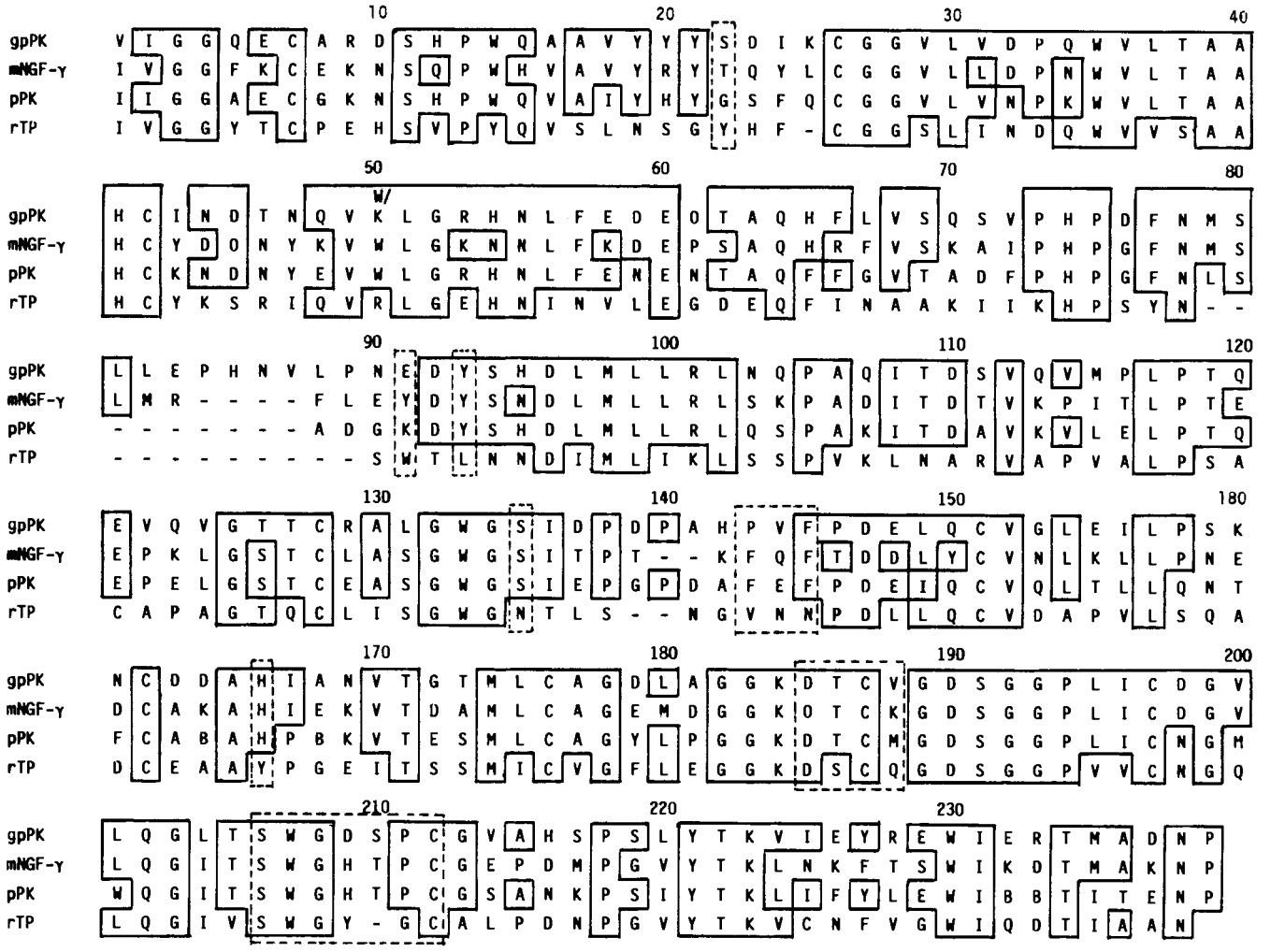

FIGURE 5: Comparison of amino acid sequences of guinea pig prostate kallikrein (gpPK), mouse NGF-γ (mNGF-γ) (Thomas et al., 1981b), porcine pancreatic kallikrein (pPK) (Tschesche et al., 1979), and rat trypsin (rTP) (Swift et al., 1982). Homologous residues are enclosed in solid boxes, and dashed boxes represent regions of the polypeptide implicated in defining substrate specificity (Mason et al., 1983).

sequence between 174 and 190, including the overlap to peptide ELC5, which was sequenced to residue 209. The peptide (CB5b) arising from cyanogen bromide cleavage at Trp-207 extended the sequence to Tyr-227. This latter peptide also provided the final overlap with the C-terminal peptide ELC6.

The primary structure of the protein is shown in Figure 5 aligned with that of mouse γ-NGF, porcine pancratic kallikrein (pPK), and rat trypsin. Gaps have been inserted in the latter sequences in order to maximize homology with the guinea pig enzyme. The guinea pig esteropeptidase displays several features characteristic of serine proteases including the N-terminal tetrapeptide comprised of two β-branched amino acids followed by a pair of glycine residues. The amino acids corresponding to those proposed to comprise the catalytic "charge-relay" system (Blow et al., 1969) are all conserved, i.e., His-41, Asp-96, and Ser-191. There is also extensive conservation of sequences adjacent to these active-site residues including, for example, residues 37–42, which border His-41, and residues 189–199, which surround the active-site serine. The aspartic acid which has been implicated in conferring specificity toward basic amino acids (Kreiger et al., 1974) is also conserved (Asp-185) together with residues 206–208 which occur at the entrance of the substrate binding cavity.

The guinea pig esteropeptidase exhibits approximately 35% identity to rat trypsin. However, this value rises to greater than 60% with mouse γ-NGF and porcine pancreatic kallikrein. This sequence homology to the glandular kallikreins suggests that esteropeptidase should be designated as guinea pig prostate kallikrein (gpPK). This is also in accordance with the catalytic properties of the enzyme which have previously been reported (Dunbar & Bradshaw, 1985). In addition, gpPK also possesses a number of specific structural features common to the kallikreins. It contains 10 of the 12 half-cystines present in trypsin which occur at homologous positions within the sequence. These residues apparently all occur as disulfides, and although the pairing of the half-cystines was not examined, they may be inferred from the X-ray crystal structure of porcine pancreatic kallikrein (Bode et al., 1983). All of the kallikreins which have been sequenced to date contain an additional C-terminal residue compared to trypsin. γ-NGF (Thomas et al., 1981b), mGK-1 (Mason et al., 1983), porcine pancreatic kallikrein (Tschesche et al., 1979), tonin (Lazure et al., 1984), and rat pancreatic kallikrein (Swift et al., 1985) and gpPK contain a C-terminal Asn-Pro dipeptide while mGK-2 (Mason et al., 1983) and pMK-1 (Richards et al., 1982) both exhibit C-terminal Asn-Ala residues. Pro-Ala is, however, a conservative substitution requiring only a single base change in the DNA sequence. Human pancreatic kallikrein differs by having a C-terminal serine (Fukushima et al., 1985).

The regions of the polypeptide which are proposed to line the binding pocket and confer substrate specificity are indicated in Figure 5. The variability which is apparent in these residues is consistent with the high degree of substrate specificity which characterizes the kallikreins. Mason et al. (1983) have also observed that the limited variability in the

coding regions of the mouse kallikrein genes predominantly involves amino acid residues implicated in substrate specificity.

gpPK also exhibits several unique properties. Most of the kallikreins are glycosylated at the residue corresponding to Asn-78, and gpPK contains an amino sugar in the protein hydrolysate. The guinea pig enzyme contains a further potential site of N-glycosylation at residue 169. Although no direct evidence for, or identification of, the carbohydrate was pursued, the appearance of a blank cycle during automated Edman degradation, followed by the Val-Thr sequence, is typical of glycosylated sequences (Waechter & Lennarz, 1976). Furthermore, chemical deglycosylation of gpPK with trifluoromethanesulfonic acid reduces the apparent molecular weight of the protein from 36 000 to 26 000, while *N*-glycanase yields a species of approximately 30 000 molecular weight (J. C. Dunbar, unpublished observations). These observations are consistent with two sites of carbohydrate attachment of which only one is susceptible to cleavage by *N*-glycanase. By analogy with the structure of porcine pancreatic kallikrein (Bode et al., 1983), this second site of glycosylation at Asn-169 occurs immediately adjacent to the "intermediate helix" (an additional helix encompassing residues 158–166 in kallikreins which is not apparent in trypsin) and appears to reside on the surface of the molecule. A carbohydrate site at this position is therefore not inconsistent with the general conformational features attributed to the kallikreins. The cDNA sequence which was recently determined for the S2 kallikrein from rat submaxillary gland (Ashley & MacDonald, 1985) also contains an Asn-Val-Thr sequence and thus a potential glycosylation site at the analogous position.

The sequence of amino acids immediately following the common site of carbohydrate attachment as Asn-78 has been designated as the "kallikrein loop". This segment of the polypeptide in kallikreins contains a minimum of four amino acid insertions compared to the same region in trypsin. In gpPK, there are 11 additional amino acids. This region of the kallikrein structure is also the site at which many of the proteins undergo proteolytic processing. The processing event frequently involves not only cleavage but also excision of several residues, resulting in kallikreins comprised of two polypeptide chains held together by disulfide bridges. Mouse γ-NGF, for example, undergoes cleavage at Arg-83, and by comparison of the protein and cDNA sequences (Isackson et al., 1985), it is apparent that four amino acids are excised. In contrast, gpPK is comprised of only a single polypeptide chain, with no comparable proteolysis within the kallikrein loop. The amino acid sequence of gpPK within this region is also devoid of any arginine residues which presumably explains the resistance to proteolysis.

Alignment of the kallikrein sequences reveals that gpPK, like porcine pancreatic kallikrein, contains additional amino acids in the segment of the polypeptide encompassing residues 138–142. This region of the protein appears to form part of the substrate binding pocket, and it is of interest that, in contrast to many kallikreins, gpPK and pPK fail to hydrolyze BAPNA (benzoylarginine-*p*-nitroanilide). The insertion of additional residues in both gpPK and pPK may account for the lack of reactivity of these proteins toward the synthetic *p*-nitroanilide substrates.

gpPK is the major arginine esteropeptidase in the guinea pig prostate. The function of this protein, however, remains to be determined. Kallikreins have been implicated in growth factor processing, and gpPK bears considerable homology to mouse γ-NGF and EGF-BP, enzymes that have been suggested to be involved in the maturation of mouse β-NGF and EGF. Guinea pig prostates contain both β-NGF and EGF (Harper et al., 1979; Rubin, 1983), but gpPK does not appear to form stable complexes, analogous to those which occur in the mouse submandibular gland, with either of these growth factors (Dunbar & Bradshaw, 1985). Furthermore, the concentration of the kallikrein is far in excess of the concentration of either of the growth factors, indicating that, unlike γ-NGF (Isackson et al., 1987), its synthesis is not coordinately regulated with either hormone. While these observations do not preclude a processing role for the enzyme, it is suggestive that the formation of stable complexes between growth factors and kallikreins may be a unique feature of the male mouse submandibular gland. The structural data determined in this study provide the base from which to determine the role of guinea pig prostate kallikrein in processing the precursors of NGF and EGF in a homologous system.

ACKNOWLEDGMENTS

We thank Steven S. Disper and Owen J. Bates of the Protein/Nucleic Acid Analysis Laboratory, Department of Biological Chemistry, University of California, Irvine, for their assistance with the protein and peptide sequencing and Marja Uskali for the expert preparation of the manuscript.

Registry No. Kallikrein, 9001-01-8; kallikrein (guinea pig prostate gland protein moiety reduced), 108007-32-5.

REFERENCES

Allen, G. (1981) in *Sequencing of Proteins and Peptides* (Work, T. S., & Burdon, R. H., Eds.) pp 30–31, Elsevier, Amsterdam.

Angeletti, R. H., & Bradshaw, R. A (1971) *Proc. Natl. Acad. Sci. U.S.A. 68*, 2417–2420.

Ashley, P. L., & MacDonald, R. J. (1985) *Biochemistry 24*, 4512–4519.

Berger, E., & Shooter, E. M. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 3647–3651.

Blow, D. M., Birktoft, J. J., & Hartley, B. S. (1969) *Nature (London) 221*, 337–340.

Blumenthal, K. M., Moon, K., & Smith, E. L. (1975) *J. Biol. Chem. 250*, 3644–3656.

Bode, W., Chen, Z., Bartels, K., Kutzbach, C., Schmidt-Kastner, G., & Bartunik, H. (1983) *J. Mol. Biol. 164*, 237–282.

Bothwell, M., Wilson, W. H., & Shooter, E. M. (1979) *J. Biol. Chem. 254*, 7287–7294.

Dunbar, J. C., & Bradshaw, R. A. (1985) *J. Cell. Biochem. 29*, 309–319.

Edelhoch, H. (1967) *Biochemistry 6*, 1948–1954.

Fielder, F. (1979) *Handb. Exp. Pharmacol. 25*, 103–161.

Frey, P., Forand, R., Maciag, T., & Shooter, E. M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 6294–6298.

Fukushima, D., Kitamura, N., & Nakanishi, S. (1985) *Biochemistry 24*, 8037–8043.

Gray, A., Dull, T., & Ullrich, A. (1983) *Nature (London) 303*, 722–725.

Greene, L. A., Shooter, E. M., & Varon, S. (1968) *Proc. Natl. Acad. Sci. U.S.A. 60*, 1383–1388.

Harper, G. P., Barde, Y. A., Burnstock, G., Carstairs, J. R., Dennison, M. E., Suda, K., & Vernon, C. A. (1979) *Nature (London) 279*, 160–162.

Hirs, H. W. (1967) *Methods Enzymol. 11*, 59–62.

Isackson, P. J., Dunbar, J. C., Bradshaw, R. A., & Ullrich, A. (1985) *Int. J. Neurosci. 26*, 95–108.

Isackson, P. J., Silverman, R. E., Blaber, M., Server, A. C., Nichols, R. A., Shooter, E. M., & Bradshaw, R. A. (1987) *Biochemistry* (in press).

Kreiger, M., Kay, L. M., & Stroud, R. M. (1974) *J. Mol. Biol. 83*, 209–230.
Lazure, C., Leduc, R., Seidah, N., Thibaul, G., Genest, J., & Chretien, M. (1984) *Nature (London) 307*, 555–558.
Mason, A. J., Evans, B. A., Cox, D. R., Shine, J., & Richards, R. I. (1983) *Nature (London) 303*, 300–307.
Richards, R. I., Catanzaro, D. F., Mason, A. J., Morris, B. J., Baxter, J. D., & Shine, J. (1982) *J. Biol. Chem. 257*, 2758–2761.
Rubin, J. (1983) Ph.D. Thesis, Washington University, St. Louis, MO.
Schachter, M. (1969) *Physiol. Rev. 49*, 519–547.
Schachter, M. (1980) *Pharmacol. Rev. 31*, 1–17.
Scott, J., Selby, M., Urdea, M., Quiroga, M., Bell, G., & Rutter, W. J. (1983a) *Nature (London) 302*, 538–540.
Scott, J., Urdea, M., Quiroga, M., Sanchez-Pescador, R., Fong, N., Selby, M., Rutter, W. J., & Bell, G. I. (1983b) *Science (Washington, D.C.) 221*, 236–240.
Shine, J., Mason, A. J., Evans, B. A., & Richards, R. I. (1983) *Cold Spring Harbor Symp. Quant. Biol. 48*, 419–426.
Swift, G. H., Dagorn, J. C., Ashley, P. L., Cumming, S. W., & MacDonald, R. J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 7263–7267.
Taylor, J. M., Mitchell, W. M., & Cohen, S. (1974) *J. Biol. Chem. 249*, 2188–2194.
Thomas, K. A., Silverman, R. E., Jeng, I., Baglan, N. C., & Bradshaw, R. A. (1981a) *J. Biol. Chem. 256*, 9147–9155.
Thomas, K. A., Baglan, N. C., & Bradshaw, R. A. (1981b) *J. Biol. Chem. 256*, 9156–9166.
Tschesche, H., Mair, G., Godec, G., Fielder, F., Ehret, W., Hirschauer, C., Lemon, M., & Fritz, H. (1979) *Adv. Exp. Med. Biol. 120*, 245–260.
Ullrich, A., Gray, A., Berman, C., & Dull, T. J. (1983) *Nature (London) 303*, 821–825.
Waechter, C. J., & Lennarz, W. J. (1976) *Annu. Rev. Biochem. 45*, 95–112.

Correlation of Photolabeling with Occupancy of cAMP Binding Sites in the Regulatory Subunit of cAMP-Dependent Protein Kinase I†

José Bubis and Susan S. Taylor*
Department of Chemistry, University of California, San Diego, La Jolla, California 92093
Received September 3, 1986; Revised Manuscript Received December 26, 1986

ABSTRACT: Each regulatory subunit of the cAMP-dependent protein kinase contains two in-tandem cAMP binding sites. Photolabeling of holoenzyme I with 8-azidoadenosine 3′,5′-monophosphate (8-N_3-cAMP) leads to the covalent modification of two residues, Trp-260 and Tyr-371. In order to correlate photolabeling of these two residues with occupancy of each specific cAMP binding site, photolabeling was carried out in the presence of various analogues of cAMP that bind preferentially to one site. Photolabeling of holoenzyme I after dissociation of 60% of 8-N_3-[^{3}H]cAMP with an excess of N^6-monobutyryl-cAMP nearly abolished the incorporation of 8-N_3-cAMP into Trp-260, whereas the modification of Tyr-371 was reduced by 49%. When 8-N_3-[^{32}P]cAMP was bound under equilibrium conditions in the presence of various cAMP analogues, N^6-monobutyryl-cAMP also selectively abolished incorporation of radioactivity into Trp-260, whereas 8-(methylamino)-cAMP preferentially reduced the covalent modification of Tyr-371. Photolabeling with trace amounts of 8-N_3-[^{32}P]cAMP in the presence of saturating amounts of N^6-monobutyryl-cAMP led to the covalent modification of only Tyr-371. In addition, photolabeling of Tyr-371 was enhanced synergistically in the presence of N^6-monobutyryl-cAMP. MgATP reduced the covalent modification of both Trp-260 and Tyr-371 but showed no selectivity for either site. These studies support a model that correlates photolabeling of Trp-260 with occupancy of cAMP binding site A and photolabeling of Tyr-371 with occupancy of cAMP binding site B. Thus, Trp-260, although it lies at the boundary between domain A and domain B, must be in close contact with the cyclic nucleotide that is bound to domain A. The results also establish unambiguously that N^6-substituted analogues of cAMP which are selective for the fast dissociation site preferentially bind to the first cAMP binding site in the linear sequence (site A), whereas C-8-substituted analogues which are selective for the slow dissociation site preferentially bind to the second site (site B). These two sites are correlated with other features that are known to distinguish the two cAMP binding sites.

Two general classes of cAMP-dependent protein kinases have been characterized, type I and type II, on the basis of elution from (diethylaminoethyl)cellulose (Corbin et al., 1975). Both isoforms exist as inactive tetramers containing two regulatory subunits (R)[1] and two catalytic subunits (C). The primary differences in the holoenzymes can be attributed to differences in the regulatory subunits (Hofmann et al., 1975; Zoller et al., 1979), and there are at least three unique genes that code for distinct regulatory subunits (Lee et al., 1983; Weldon et al., 1985; Stein & Rubin, 1985; Jahnsen et al., 1986). Despite

† Supported by U.S. Public Health Service Grant GM-19301. J.B. was supported by the Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICIT), Caracas, Venezuela.

[1] Abbreviations: 8-N_3-cAMP, 8-azidoadenosine 3′,5′-monophosphate; R, regulatory subunit of cAMP-dependent protein kinase; C, catalytic subunit of cAMP-dependent protein kinase; CAP, *Escherichia coli* catabolite gene activator protein; TPCK, L-1-(tosylamido)-2-phenylethyl chloromethyl ketone; HPLC, high-performance liquid chromatography; EDTA, ethylenediaminetetraacetic acid; SDS, sodium dodecyl sulfate; BSA, bovine serum albumin; TCA, trichloroacetic acid; TFA, trifluoroacetic acid; PTH, phenylthiohydantoin; Trp, tryptophan; Tyr, tyrosine; K_d, apparent equilibrium dissociation constant; K_i, apparent equilibrium inhibition constant; M_r, molecular weight; MES, 2-(*N*-morpholino)-ethanesulfonic acid.

0006-2960/87/0426-3478$01.50/0