Endoproteolytic Processing of the Dibasic Cleavage Site in the Human Protein C Precursor in Transfected Mammalian Cells: Effects of Sequence Alterations on Efficiency of Cleavage†

Donald C. Foster,* Cindy A. Sprecher, Rick D. Holly, Jay E. Gambee, Kathleen M. Walker, and A. Ashok Kumar
Zymogenetics, Inc., 4225 Roosevelt Way, NE, Seattle, Washington 98105
Received February 17, 1989; Revised Manuscript Received June 29, 1989

ABSTRACT: The human protein C precursor undergoes extensive co- and posttranslational modification during its biosynthesis in the liver. These modifications include glycosylation, γ-carboxylation and β-hydroxylation of specific amino acids, and endoproteolytic processing to remove the pre- and propeptides and also to remove the pair of basic amino acids that connect the light and heavy chains in the precursor. Specific molecular signals have been elucidated which direct several of these modifications; however, the mechanism for cleavage and removal of the basic amino acid pair has not been established. In the present study, a recombinant mammalian expression system has been used to study the molecular signals that direct removal of this basic amino acid pair. Mutations were introduced by site-directed mutagenesis either to insert additional basic amino acids or to alter the sequence adjacent to the basic pair by point mutations. The mutant protein precursors were expressed and analyzed for the degree of processing to 2-chain form and also for the location of the cleavage site (by N-terminal sequencing) and subsequent removal of the basic amino acids from the newly formed C terminus of the light chain. These experiments have shown that human protein C can be readily synthesized and secreted in several mammalian cell lines. However, cell lines vary considerably in their capacity to remove the dibasic pair in the protein C precursor and, like the liver, secrete a mixed population of 1-chain and 2-chain forms of the protein. Comparison of mutant precursors with the wild-type precursor indicates that the endoprotease which performs this cleavage in human kidney 293 cells and also in BHK cells recognizes and cleaves this site much more efficiently if a basic amino acid is also present in the −4 position. This finding suggests a substrate specificity for this protease that is similar to the protease which removes the propeptide from the family of vitamin K dependent proteases [Bentley et al. (1986) *Cell 45*, 343].

Protein C is the precursor to a serine protease present in plasma that plays an important physiological role in the regulation of blood coagulation pathways (Esmon, 1987). Human protein C undergoes extensive co- and posttranslational modification during its biosynthesis in the liver. These modifications include the γ-carboxylation of nine glutamic acid residues near its amino terminus (DiScipio & Davie, 1979; Fernlund & Stenflo, 1982), β-hydroxylation of aspartic acid 71 (Drakenberg et al., 1983; McMullen et al., 1983), glycosylation of four asparagine residues, proteolytic removal of a prepeptide and a propeptide (Foster et al., 1987), and endoproteolytic internal cleavage that converts the M_r = 62 000 single-chain precursor to a light chain (M_r 21 000) and a heavy chain (M_r = 41 000) which are held together by a disulfide bond (Stenflo, 1976; Kisiel et al., 1976).

The cDNA sequence for human protein C (Foster & Davie, 1984; Beckmann et al., 1985) indicates that human protein C is initially synthesized as a single-chain precursor in which the light chain and the heavy chain are connected by a basic Lys-Arg dipeptide. Conversion of the precursor to the mature, 2-chain form must involve some combination of endo- and/or exoproteolytic processing to remove these two amino acids. However, these processing steps are incompletely performed in the liver, as indicated by the presence of small amounts of unprocessed, single-chain protein C present in plasma (Miletich et al., 1983). Human factor X also exists as a disulfide-bonded two-chain zymogen and its cDNA sequence (Leytus et al., 1984) predicts that similar processing of a basic Arg-Lys-Arg tripeptide results in conversion of a single-chain precursor to the 2-chain form (Figure 1).

Similar processing of internal pairs of basic amino acids occurs in the maturation of essentially all prohormone polypeptide precursors (Douglass et al., 1984) in mammalian cells. In the case of prohormone maturation processing, the liberation of mature peptides is generally a two-step process (Steiner et al., 1980): (i) the precursor is cleaved on the carboxy-terminal side of the pair of basic residues by a trypsin-like endopeptidase followed by (ii) the removal of the pair of basic residues from the C-terminal end of the resulting fragment by a carboxypeptidase-B-like (CP-B)[1] protease. Such processing in mammalian cells is known to be highly tissue specific. The basis of this tissue specificity is not known, but may involve differential expression of proteolytic enzymes involved in the maturation processes or differential recognition of precursor cleavage site sequences.

Expression of recombinant human protein C has provided a useful model system to study a variety of co- and posttranslational modifications involved in biosynthesis of this protein, including γ-carboxylation (Foster et al., 1987) as well as the present analysis of endoproteolytic maturation. We and others have found a significant variation in the efficiency with which different cell lines carry out the processing of the recombinant protein C precursor to two-chain form. Human kidney 293 cells process the precursor to approximately 80–90% two-chain form (Foster et al., 1987; Grinnell et al.,

† This work was supported in part by a grant from the Department of Health and Human Services Public Health Service HL 36496-03.
* To whom correspondence should be addressed.

[1] Abbreviations: BHK, baby hamster kidney; CP-B, carboxypeptidase B; DHFR, dihydrofolate reductase; POMC, preproopiomelanocortin.

0006-2960/90/0429-347$02.50/0

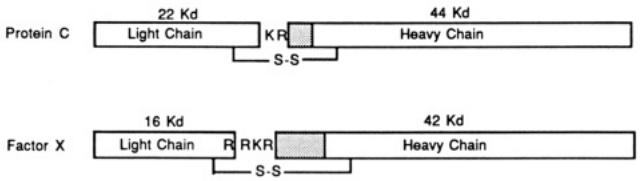

FIGURE 1: Diagram of the human protein C and factor X precursors. Both precursors are shown without prepropeptide structures. The basic amino acids connecting the two chains are predicted from the cDNA sequences of human protein C (Foster & Davie, 1984; Beckmann et al., 1985) and human factor X (Leytus et al., 1984). The activation peptide regions are shaded.

1987). Mouse epithelioid C127 cells process approximately 50% of the protein to 2-chain form (Oppenheimer & Wydro, 1988). Baby hamster kidney (BHK) cells also perform this processing step with low efficiency, with approximately 70% of the protein being secreted as a single-chain form (this paper). Human liver derived HepG2 cells also secrete a mixture of approximately 50% single-chain protein C (J. Miletich, personal communication).

In contrast, the human factor X precursor is processed to 2-chain form almost completely in both recombinant human (293) and hamster (BHK) cell systems (data not shown) and also in human HepG2 cells (J. Miletich, personal communication), suggesting that a structural difference between the protein C precursor and the factor X precursor may be responsible for the differential processing of these precursors. In particular, the presence of two additional arginine residues on the N-terminal side of the Lys-Arg pair in the factor X precursor suggested that the endoprotease responsible for efficient processing of factor X may have substrate specificity which requires basic amino acids in the −3 or −4 position for efficient recognition.

To test this hypothesis, a series of mutations have been constructed in the cDNA for human protein C which either introduce additional basic amino acids immediately N-terminal to the Lys-Arg pair or make point mutational changes in the nature of the amino acid in the −4 position (His-154) relative to the cleavage site following Arg-157. These constructions were then inserted into a mammalian cell expression vector and introduced into both BHK cells and human kidney 293 cells. The mutant proteins that were secreted were then analyzed for the degree of processing to 2-chain form, the location of the proteolytic processing events, and carboxypeptidase-like removal of residual basic amino acids from the C terminus of resultant peptide chains. These experiments have demonstrated that the efficiency with which these sites are recognized by cellular endoproteases depends on structural determinants beyond the Lys-Arg pair and is particularly sensitive to the nature of the amino acid in the −4 position.

Materials and Methods

Materials. HPLC columns used for purification of protein C and fragments were C4 (Vydac, Hesperio, CA) and Bioseries Poly F (Du Pont, Wilmington, DE). Vapor-phase hydrolysis of proteins for amino acid analysis was performed with a Pico-Tag workstation (Waters-Millipore, Milford, MA). Amino acid analyses were performed on a Beckmann 6300 analyzer (Beckmann, Palo Alto, CA), and automated Edman degradation was performed with a 470A sequenator (Applied Biosystems, Foster City, CA). The HPLC was performed with a Varian 5000 liquid chromatograph and Varian UV-5 detector.

Mutagenesis. An 840-bp *Sst*I fragment of the human protein C cDNA which corresponds to amino acids 9 of the light chain through 119 of the heavy chain was subcloned in a 3′–5′ orientation in M13mp11. Various mutations in this fragment were then prepared by the two-primer oligonucleotide-directed mutagenesis technique (Zoller & Smith, 1982). All mutants were sequenced by dideoxy chain termination technique to verify each mutation and the fidelity of the remaining sequence. RF DNA was prepared from mutant plaques, and the mutant *Sst*I fragments were used to replace the corresponding fragment of the wild-type protein C cDNA in the expression vector.

Vectors. The mutant protein C cDNAs were ligated into a mammalian cell expression vector called zEM229R. This vector permits insertion of *Eco*RI fragments into a cloning site downstream of the mouse MT-1 promoter and upstream of an SV-40 polyadenylation sequence and also carries an expression unit for the dihydrofolate reductase (DHFR) gene under control of the SV-40 early promoter. This vector was generously provided by Dr. Eileen Mulvihill.

Transfection/Selection. BHK cells were grown in Dulbecco's modified Eagle's medium containing 10% fetal calf serum and vitamin K (5 μg/mL). Cells were transfected with the protein C mutant expression plasmids by the calcium phosphate procedure (Graham & van der Eb, 1973). For selection of stable colonies, the cells were divided 48 h after transfection into media containing 250 nM methotrexate. Colonies were screened for protein C production by an immunofilter assay as previously described (Busby, 1988) and grown individually for protein analyses.

Pulse Labeling. Cell lines synthesizing the various mutant forms of protein C were characterized for their rate of protein C production per cell per day. Clonal cell lines that synthesized approximately 1 $\mu g\ mL^{-1}\ day^{-1}\ (3 \times 10^6\ cells)^{-1}$ were chosen for each mutant to eliminate potential effects of variations in protein synthesis levels on the degree of proteolytic processing observed. For metabolic labeling, confluent 10-cm plates of cells were washed with PBS, then incubated overnight with 2 mL of cysteine-free Dulbecco's MEM supplemented with 25 μCi/mL [^{35}S]cysteine, 5 μg/mL vitamin K, and 1% fetal calf serum. Following the labeling period the media were harvested, and the cell layer was washed with PBS and then lysed in 2 mL of lysis buffer (10 mM HEPES, pH 7.4, 50 mM NaCl, 2.5 mM $MgCl_2$, 300 mM sucrose, 0.5% Triton X-100). Labeled media and cell lysates were stored at −20 °C until use. Radioimmunoprecipitations of labeled samples were performed with a monoclonal antibody (PCH-1, unpublished data) that is directed against an epitope in the heavy chain of human protein C. Immunoprecipitated samples were boiled in reducing gel sample buffer prior to electrophoresis on 10% SDS–PAGE gels.

Immunopurification of Protein C. Recombinant wild-type and mutant protein C samples were purified by affinity chromatography on a Sepharose-bound anti-protein C antibody column (PCL-2). This antibody has been shown to be specific for the calcium-induced conformation of protein C (unpublished data). Culture media containing protein C were made 10 mM in $CaCl_2$, filtered through a 0.45-μm filter, and applied to the antibody column that was previously equilibrated with 50 mM Tris-HCl, pH 7.4, buffer containing 150 mM NaCl, 10 mM $CaCl_2$, and 0.02% NaN_3. The unbound protein was washed off with 1.0 M NaCl in the equilibration buffer, and bound protein C was eluted with 15 mM EDTA in 50 mM Tris-HCl, pH 7.4. Immunoaffinity pure protein C samples have been subjected to further purification on a C4 HPLC column (4.6 × 250 mm) with a water–acetonitrile gradient (15–50%, 60 min) containing 0.1% TFA at a flow rate of 1 mL/min.

Isolation of Protein C Light Chain. Protein C was reduced and carboxymethylated as described (Crestfield et al., 1963). Excess reagents were removed by dialysis against 0.1 M NH_4HCO_3, pH 8.5. The separation of light and heavy chains of protein C was achieved by HPLC of the reduced, carboxymethylated protein on a Du Pont BioSeries Poly F column (6.2 × 80 mm). The eluent was 0.1 M NH_4HCO_3, pH 8.5, and a linear gradient of acetonitrile (0–80% in 20 min) was employed at a flow rate of 2 mL/min. The protein peaks were collected and concentrated either by lyophilization or by Speed-Vac concentration.

CNBr Digestion of Protein C Light Chain. Protein C light chain (0.2–1.0 nmol) in 70% HCOOH was mixed with CNBr (200–500-fold excess per Met residue) and incubated at 22 °C for 24 h in the dark and under nitrogen. The digest was lyophilized, redissolved in 15% CH_3COOH, and subjected to Edman degradation without separating the peptides. In an alternative procedure, the light chain was applied to a polybrene-coated, precycled (program 03R PRE) glass fiber cartridge filter and was treated with a solution of CNBr in 70% HCOOH (30 μL; 20 mg of CNBr/100 μL). The filter was then incubated in an atmosphere of CNBr–HCOOH at 22 °C in the dark for 20 h, dried under vacuum, and subjected to N-terminal sequence analysis.

Amino Acid Analysis and Sequence Determination. Proteins were subjected to vapor-phase hydrolysis in 6 N HCl containing 2% phenol for 22 h at 110 °C. The dry hydrolysate was resuspended in sample diluent (Beckman Na–S buffer) and analyzed. Automated Edman degradation was performed on a gas-phase sequencer equipped with an on-line Applied Biosystems 120A PTH analyzer. PTH amino acids were separated and identified on a microbore PTH C-18 column employing solvents and gradient suggested by the manufacturer.

Results

The observation that factor X is processed to 2-chain form much more efficiently in BHK and human kidney 293 cells than is protein C suggested that the two additional basic amino acids in the factor X processing site may be important in recognition of the processing site by the endoprotease(s) which perform this processing step. This speculation led first to the construction of a protein C mutant in which two additional Arg residues were inserted into the protein C processing site by loop-in oligonucleotide mutagenesis to make PC962, which has a processing site essentially identical with that of factor X (Figure 2).

Following mutagenesis, the native and mutant forms of the protein C cDNA were ligated into a mammalian cell expression vector—zEM229—in which constitutive, high-level transcription is driven by the mouse metallothionine promoter and which carries the DHFR gene as a selectable marker on the same plasmid. To maximize our ability to detect changes in the degree of processing, these plasmids were then introduced into BHK cells (which process the wild-type precursor more poorly) by calcium phosphate precipitation. Methotrexate-resistant colonies were isolated and analyzed for their protein C production levels.

Analysis of the degree of endoproteolytic processing of the different forms of recombinant protein C was accomplished by SDS–PAGE fractionation of the protein C immunoprecipitated from culture media of cells that had been pulse-labeled with [^{35}S]Cys. The gels were run under reducing conditions to separate 2-chain molecules and then dried and autoradiographed for evaluation of the relative amounts of 1-chain and 2-chain forms of the protein. Since the distribution

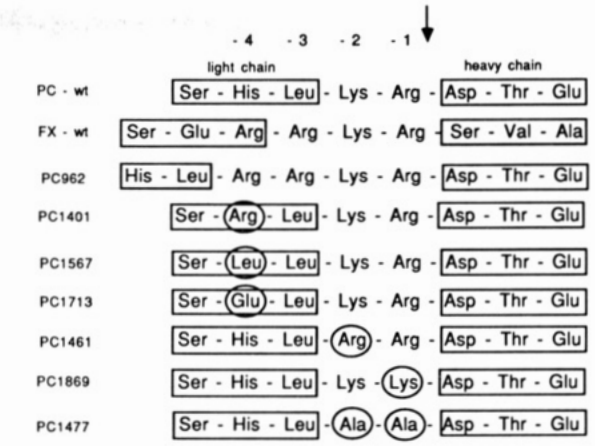

FIGURE 2: Processing-site sequences of human protein C mutants created by site-directed mutagenesis. Amino acids representing the mature C terminus of the light chain and the mature N terminus of the heavy chain are boxed. Amino acids substituted by point mutation are circled. The endoproteolytic cleavage site is indicated by an arrow. Amino acid positions upstream of the cleavage site are numbered with respect to distance from the cleavage site.

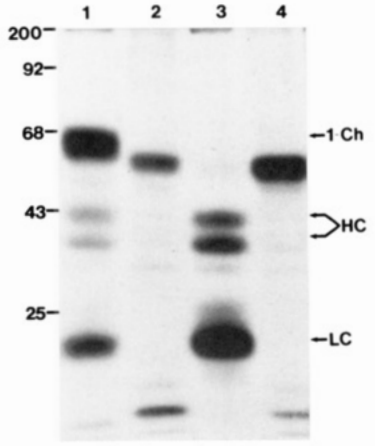

FIGURE 3: Immunoprecipitation of cellular and secreted protein C. Clonal cell lines expressing either wild-type protein C (lanes 1 and 2) or mutant PC962 (lanes 3 and 4) were labeled with [^{35}S]Cys, and the media and cell lysates were harvested as described under Materials and Methods. Samples of media fractions (lanes 1 and 3) and cell lysates (lanes 2 and 4) were immunoprecipitated with a monoclonal antibody against human protein C and run on 10% SDS–PAGE under reducing conditions. Labeled protein bands were visualized by fluorography.

of cysteines is nonuniform, the light chain shows a higher relative specific activity than the heavy chain and appears darker on the autoradiograms. The heavy chain shows molecular weight heterogeneity similar to that observed in plasma protein C which appears to be due to differential glycosylation patterns.

Figure 3 (lanes 1 and 3 media fractions) shows the results of this analysis for wild-type protein C and the insertion mutant PC962, respectively. The secreted wild-type protein C (lane 1) is cleaved to 2-chain form poorly in BHK cells, with most of the radioactivity appearing in the band associated with intact, single-chain protein. In contrast, the secreted mutant PC962 (lane 3) is cleaved with high efficiency, with nearly all of the radioactivity migrating in association with the light-chain and heavy-chain bands of the 2-chain form. This result indicates that the insertion of two additional Arg residues in PC962 has a significant impact on the recognition of this processing site by the processing endoprotease. Immunoprecipitation of intracellular fractions (lanes 2 and 4, Figure 3) shows that, for both PC-wt (lane 2) and PC962 (lane 4) the

Table I: Amino-Terminal Sequence of the Light Chain and Heavy Chain of PC962[a]

	light chain		heavy chain	
cycle	expected	observed	expected	observed
1	A	A (893)[b]	D	D (699)[b]
2	N	N (556)	T	T (101)
3	S	S (156)	E	E (403)
4	F	F (236)	D	D (284)
5	L	L (182)	Q	Q (339)
6	E[c]		E	E (274)
7	E[c]		D	D (222)
8	L	L (45)	Q	Q (234)
9	R	R (53)	V	V (206)
10	H	H (36)	D	D (174)
11	S	S (39)	P	P (53)
12	S	S (46)	R	R (100)

[a] Reduced and carboxymethylated protein (1500 pmol) was used for sequence analysis. [b] Numbers in parentheses represent the yield in picomoles at each cycle. [c] γ-Carboxyglutamic acid residues.

intracellular form of protein C is almost completely nonprocessed and migrates as a single-chain form, whereas the form of PC962 that is secreted into the media (lane 3) is nearly completely 2-chain form. Radiolabeled protein C can be exposed to media for extended lengths of time at 37 °C without noticeable conversion of 1-chain form to 2-chain form (data not shown), excluding the possibility that this cleavage occurs subsequent to secretion. The lack of detectable amounts of cleaved protein C within cellular fractions would indicate that this cleavage probably occurs very late in the secretory pathway, probably in the Golgi or in secretory vesicles, and that the transit time between the compartment where cleavage events occur and the plasma membrane is extremely short. The single-chain wild-type protein C secreted into the media (lane 1) is somewhat higher in molecular weight than the single-chain form present in intracellular fractions. This difference is presumably attributable to additional carbohydrate on the secreted form.

The pulse-label analysis indicated that PC962 was being cleaved by a cellular endoprotease in a position which generated proteolytic products very similar in size to the normal light and heavy chains of protein C. However, since the normal processing site had been altered by introduction of additional amino acids, it was important to determine the precise location and nature of processing events responsible for generating the 2-chain form of the mutant protein. Of particular significance were the questions of whether cleavage still occurred following the Lys-Arg, yielding the normal N terminus for the heavy chain, and whether the presence of additional amino acids would lead to a modified C-terminal light-chain structure.

To address these structural questions, the mutant PC962 protein was purified by affinity chromatography using a calcium-induced conformation-specific monoclonal antibody conjugated to Sepharose. Following reduction and carboxymethylation, the light chain and heavy chain of the purified protein were separated by HPLC chromatography on a Poly F column.

N-Terminal sequence analysis of the two chains is shown in Table I. Both the light chain and the heavy chain showed the amino acid sequence predicted from the cDNA sequence (Foster & Davie, 1984; Beckmann et al., 1985), indicating the complete proteolytic removal of the propeptide from the light chain and also complete proteolytic cleavage following the Lys-Arg sequence that precedes the heavy chain in the protein C precursor.

Analysis of the C-terminal processing of the isolated light chain of PC962 was carried out in two ways. Purified PC962 protein as well as the isolated light chain of PC962 was analyzed for amino acid composition. The results, shown in Table II, indicate that both the intact protein and the isolated light chain have compositions consistent with the amino acid composition predicted from the cDNA sequence and ending with C-terminal Leu-155. Of particular significance is the lack of any additional observed Arg or Lys component which might indicate incomplete carboxypeptidase-B-like trimming of these residues from the C terminus of the light chain.

Another more direct analysis of the light-chain C terminus involved digestion of the isolated light chain with CNBr, followed by N-terminal sequence analysis of the digestion products. CNBr should cleave the light chain following a unique Met residue at position 148 to yield two peptides corresponding to amino acids 1–148 and 149–155. One would then expect two sequences from Edman degradation of the mixture, one from the mature N terminus and one starting with Glu-149. The results of this sequence analysis are shown in Table III. Two amino acids were identified at each cycle for the first seven cycles, corresponding to residues 1–7 of the N terminus and residues 149–155 from the C terminus. Only one amino acid was identified in cycles 8–12, corresponding to residues 8–12 of the N-terminal sequence. These data also are consistent with C-terminal Leu-155 for the light chain of PC962.

Table II: Amino Acid Composition of PC962 and the Isolated Light Chain[a]

	light chain		PC962	
amino acid	expected	observed	expected	observed
Asp	14	13.0	42	40.8
Thr	3	3.2	15	13.9
Ser	12	9.9	30	23.8
Glu	19	19.5	46	46.6
Pro	6	6.3	18	18.3
Gly	13	13.9	33	33.6
Ala	6	6.5	21	21.4
Cys	17	*b*	24	7.8[c]
Val	6	6.2	26	24.8
Met	1	1.0	7	6.8
Ile	5	5.1	16	15.3
Leu	15	14.9	43	43.0
Tyr	2	2.1	8	8.6
Phe	8	7.9	13	13.8
His	7	6.6	17	16.0
Lys	7	7.1	22	22.2
Arg	10	9.7	23	22.8

[a] Compositions were calculated by using the composition-conscious amino acid analysis method (Brandstrom & Ranby, 1987). [b] CM-Cys was not quantitated. [c] Intact protein was used for analysis.

Table III: Amino-Terminal Sequence Analysis of CNBr-Digested PC962[a]

	expected		observed	
cycle	A	B	A	B
1	A	E	A (866)	E (570)
2	N	K	N (493)	K (320)
3	S	K	S (379)	K (356)
4	F	R	F (304)	R (74)
5	L	S	L (220)	S (75)
6	(E)	H	E (19)	H (28)
7	(E)	L	E 16)	L (+)[b]
8	L	?	L (17)	
9	R	?	R (35)	
10	H	?	H (25)	
11	S	?	S (62)	
12	S	?	S (57)	

[a] CNBr-digested PC962 protein (2200 pmol) was used for sequence analysis. [b] Leucine was detected but not quantitated.

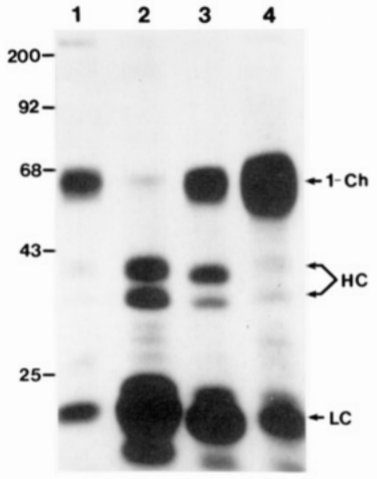

FIGURE 4: Immunoprecipitation of protein C mutants substituted in the −4-position. Clonal cell lines expressing either wild-type protein C or a mutant in the −4-position (relative to the endoproteolytic cleavage site) were labeled with [^{35}S]Cys, and the culture media were harvested as described under Materials and Methods. Samples of media from wild-type protein C [His-153 (lane 1)], PC1401 [Arg-153 (lane 2)], PC1567 [Leu-153 (lane 3)], and PC1713 [Glu-153 (lane 4)] were immunoprecipitated with a monoclonal antibody against human protein C and run on 10% SDS–PAGE under reducing conditions. Labeled protein bands were visualized by fluorography.

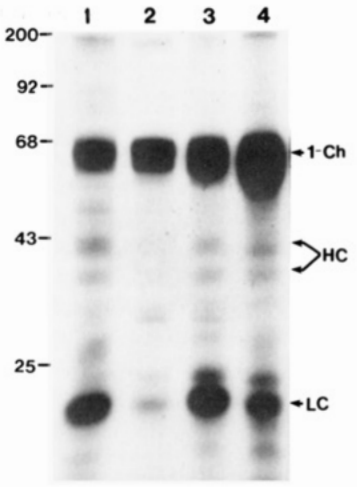

FIGURE 5: Immunoprecipitation of protein C mutants substituted in the −1 and −2 basic amino acid pair. Clonal cell lines expressing either wild-type protein C or a mutant in the −1- or −2-position (relative to the endoproteolytic cleavage site) were labeled with [^{35}S]Cys, and the culture media were harvested as described under Materials and Methods. Samples of media from wild-type protein C [Lys-Arg (lane 1)], PC1477 [Ala-Ala (lane 2)], PC1461 [Arg-Arg (lane 3)], and PC1869 [Lys-Lys (lane 4)] were immunoprecipitated with a monoclonal antibody against human protein C and run on 10% SDS–PAGE under reducing conditions. Labeled protein bands were visualized by fluorography.

The observation that PC962 and factor X are cleaved efficiently in BHK cells and in human 293 cells, whereas PC-wt is not, suggested that the endoprotease responsible for these cleavages has substrate specificity more restrictive than simple pairs of basic amino acids. This is further supported by the observation that several other basic amino acid pairs within protein C, including two other Lys-Arg pairs four and nine amino acids upstream of the processing site, are not detectably cleaved by endoproteases in the cell.

To test the hypothesis that the nature of the amino acid in the −4 position may have a significant influence on the cleavage efficiency, we constructed another series of PC mutants in which point mutations were introduced at this position by oligonucleotide-directed mutagenesis to introduce amino acids of different chemical nature into the −4 position in place of His-154. These mutations, shown in Figure 2, replace His with Arg, Leu, or Glu.

The effect of replacements at this position was examined by reducing SDS–PAGE fractionation of the protein C products immunoprecipitated from culture media of cells pulse labeled with [^{35}S]Cys. The results of this experiment, shown in Figure 4, indicated that the nature of the amino acid in the −4 position has a significant influence on the efficiency of precursor cleavage. Substitution of His-154 with Arg (PC1401, lane 2) causes nearly complete conversion to 2-chain form. Substitution at this position with Leu (PC1567, lane 3) functions similarly to His, with partial conversion to 2-chain form. Substitution with Glu (PC1713, lane 4) causes nearly complete loss of recognition of the processing site with only minor levels of conversion to 2-chain form. The potentiation of site recognition by arginine cannot be explained by creation of a new proteolytic site (following the newly introduced Arg), since N-terminal sequence of purified 2-chain PC1401 shows the normal N-terminal sequence of the heavy chain (data not shown).

We also investigated the specificity of the endoprotease for the nature of the basic amino acids in positions −1 and −2. The effect of replacements at these positions was also examined by reducing SDS–PAGE fractionation of the protein C immunoprecipitated from culture media of cells pulse labeled with [^{35}S]Cys and is shown in Figure 5. Wild-type protein C is shown in lane 1. In mutant PC1477 (lane 2), both of these positions have been converted to Ala. The nearly complete lack of processing of this mutant confirms the trypsin-like specificity of the protease responsible for basic amino acids and also confirms the lack of alternative processing sites elsewhere in the molecule. In mutant PC1869 (lane 3) the −1 position has been converted to Lys, which has only minor effect on the degree of processing compared with the wild-type sequence. In mutant PC1461 (lane 4) the −2 position has been converted to Arg, creating an Arg-Arg pair. This mutation also has only very slight effect on the degree of processing. These data suggest that the endoprotease responsible for this cleavage has only slight preferential cleavage of Lys-Arg pairs relative to other basic pairs in the same position.

Discussion

Human protein C undergoes extensive co- and posttranslational modification in the liver to form the mature, 2-chain protein circulating in plasma. these processing steps include removal of the pre- and propeptides, γ-carboxylation of the first nine glutamic acid residues near the amino terminus of the light chain, β-hydroxylation of aspartic acid 71 in the light chain, addition of carbohydrate to several N-linked glycosylation sites, and proteolytic processing and removal of the basic Lys-Arg connecting dipeptide. The molecular signals that direct these modifications have been characterized to varying extents. The sequences that are recognized by the signal peptidase for cleavage of the hydrophobic prepeptide have been well established (von Heijne, 1984) and are predictable (von Heijne, 1986). Similarly, the recognition of Asn-X-Ser/Thr sequences for addition of N-linked carbohydrate has been well established. More recently, the propeptide sequences for protein C (Foster et al., 1987) and for factor IX (Jorgensen et al., 1987) have been established as recognition signals for the enzymatic γ-carboxylation of these proteins. Also, a conserved putative recognition signal for β-hydroxylation of certain proteins has been proposed (Stenflo, 1988). The recognition of molecular signals that direct processing steps

involved in cleavage and removal of the dibasic Lys-Arg peptide which connects the light and heavy chains of the human protein C precursor has not previously been investigated.

The cleavage of such a Lys-Arg sequence to yield mature peptide products bears striking similarity to both the processing of polypeptide precursors to mature peptides in specialized endocrine mammalian cells and also to the processing of the mating-type α-factor precursor to yield mature α-factor peptides in yeast. In both cases, the removal of basic dipeptides is thought to occur via a two-step process wherein the precursor is cleaved on the carboxyl side of the pair of basic residues by a trypsin-like endopeptidase, followed by removal of the doublet of basic residues from the carboxyl-terminal end of the resultant peptide by a carboxypeptidase with specificity for basic amino acids.

In yeast, the two enzymes responsible for these consecutive processing steps have been identified, respectively, as the protein products of the *KEX2* (endoprotease) and *KEX1* (carboxypeptidase) genes (Julius et al., 1984; Dmochowska et al., 1987). The requirement of fidelity of processing of the α-factor precursor to maintain the killer toxin phenotype allowed identification and cloning of these essential processing genes by complementation of their activities in deficient mutant strains. Characterization of the processing enzymes responsible for analogous events in mammalian cells is much less complete. This effort is hampered by the fact that mammalian cells, by virtue of their diversity and diploid genetic complement, are generally not amenable to the strategy of generation and complementation of mutant phenotypes that allowed characterization of these genes in yeast.

Mammalian cell processing of prohormone precursors is highly tissue specific. A given precursor protein, such as preproopiomelanocortin (POMC), can be processed into alternative and distinct sets of peptide hormones in the anterior or neurointermediate lobes of the pituitary (Krieger & Liotta, 1979). Alternatively, there are mammalian cell lines that apparently cannot process this precursor at all (Thomas et al., 1988). The basis for the tissue specificity of these processing steps is not understood. Factors that may contribute to the observed specificity may include differential expression of the required processing protease(s) responsible, differential modification of the precursor structure resulting in differential cleavage site accessibility, or different substrate specificities of distinct processing proteases present in different cell types. The lack of identification of any of these proteases precludes evaluation of the first possibility. Differential processing of POMC by the yeast *KEX2* product in different mammalian cell types (Thomas et al., 1988) indicates that differential modification of the precursor may be a contributing factor in some cases. Analysis of precursors with modified cleavage-site sequences may be a useful approach toward evaluation of the contribution of substrate specificity.

One approach toward understanding the factors that influence these processes is to compare precursor proteins which are known to undergo very similar processing steps. The observation that protein C is processed incompletely in human kidney 293 cells (~80%) and BHK cells (~30%) was in contrast to the higher degree of processing in both cell types of factor X (>90% in both). This difference suggested that a structural difference between the two protein precursors was responsible for the difference in processing levels observed, rather than differential expression of the appropriate protease(s).

In the present experiments, mutations were introduced into the protein C precursor sequence to test the effect of substrate

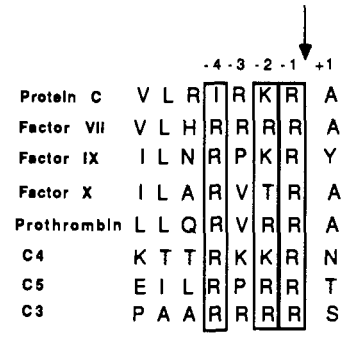

FIGURE 6: Propeptide proteolytic processing-site sequences of several processed proteins secreted from human hepatocytes. The homologous basic residues in positions −1, −2, and −4 are boxed. The endoproteolytic cleavage site is indicated by an arrow. The sequences are taken from the following references: protein C (Foster et al., 1985); factor VII (Hagen et al., 1986); factor IX (Kurachi & Davie, 1982); factor X (Fung et al., 1985); prothrombin (Degen et al., 1986); C4 (Belt et al., 1984); C5 (Lundwall et al., 1985); C3 (DeBruijn & Fey, 1985).

sequence structure on recognition and processing by these enzymes. The effect of the mutations was analyzed by expression of the mutant cDNAs in BHK cells, followed by reducing SDS–PAGE analysis of the 1-chain/2-chain form of the secreted protein C to evaluate the degree of cleavage and also N-terminal and C-terminal sequence analysis of the purified protein to confirm the identity of the processing site. BHK cells were chosen for evaluation of the effects of mutations because of their relatively poor processing of the wild-type sequence, which maximized ability to detect effects of sequence mutations.

Introduction of additional basic amino acids into the protein C processing site resulted in a processing-site sequence which resembled that of factor X (Figure 2). This mutant, PC962, in contrast to native protein C, is processed to 2-chain form efficiently in BHK cells (Figure 3). This observation suggested that the endoprotease responsible for this cleavage in BHK cells may have substrate specificity which depends on the nature of amino acids in a region that extends beyond the basic pair in positions −1 and −2 (relative to the cleavage site). Expression of mutant PC962 and wild-type PC in human 293 cells in control experiments (not shown) demonstrated that this mutation caused a similar increase in processing in the human cells, from approximately 20% single-chain wild-type protein C to nearly undetectable levels of single-chain with the mutant protein. Thus, the endoproteases from both species exhibited a similar preference for substrates with additional basic amino acids at the −3 and −4 positions. We cannot rule out the possibility that the sequence alternations in the protein C precursor have caused targeting to another endoprotease within the cell, which in turn could be responsible for the alterations in degree of cleavage. However, the fact that the intracellular mutant precursor (Figure 3, lane 4) is nearly entirely single chain would support the interpretation that such an alternative endoprotease would have to be similarly localized at a very late point in the constitutive secretory pathway.

One hypothesis consistent with the observed specificity is that the endoprotease responsible for this cleavage may have substrate specificity similar to or identical with that of the protease which removes the propeptide from precursors to many plasma proteases. The propeptides of several of the precursors are shown in Figure 6. These propeptides are postulated to be removed by a trypsin-like endoprotease that cleaves after the Arg in position −1. With few exceptions, these propeptide sequences have basic amino acids in positions −1, −2, and −4. The −1-position is invariant with Arg, while the

-2-position is either Lys or Arg. The significance of the Arg in the -4-position was shown by molecular characterization of a naturally occurring mutant factor IX (Bentley et al., 1986) in which this Arg was mutated to Gln. This substitution prevented cleavage of the propeptide and resulted in loss of activity of the mutant factor IX protein. Nearly all of these propeptide sequences have Arg at position -4, the single exception being protein C, which has Arg at position -3. In addition, studies with naturally occurring mutants of factor IX (Diuguid et al., 1986) and artificially introduced mutations in the propeptide of factor VII (Busby et al., 1987) have demonstrated the requirement for basic amino acids in the -1-position for proper cleavage of these propeptides. The nearly universal occurrence of Arg at position -4, together with the impact of mutation in this position on processing, has led to the proposal (Bentley et al,, 1986) that the endoprotease responsible has specificity for cleavage of sites that contain Arg at positions -1 *and* -4.

The present data on cleavage of recombinant protein C are consistent with a similar specificity for the BHK processing endoprotease. In wild-type protein C precursor, the corresponding -4-position is His-154, whereas in wild-type factor X precursor this position corresponds to Arg-139. Insertion of two additional arginines in the efficiently cleaved mutant PC962 caused the -4-position to become Arg instead of His. To test the hypothesis that the nature of the amino acid in the -4-position may have a significant influence on the cleavage efficiency, we constructed another series of PC mutants in which point mutations were made at this position by oligonucleotide-directed mutagenesis to introduce amino acids of different chemical nature into the -4-position in place of His (Figure 4). When the His was replaced with Arg (PC1401), the efficiency of cleavage was substantially improved, with nearly all of the protein being secreted in 2-chain form. Substitution with Leu (PC1567) caused very little change in the cleavage efficiency compared with that of wild type, whereas substitution with an acidic Glu residue (PC1713) caused a reduction in cleavage efficiency.

Alteration of the processing-site sequence in the -1- and -2-positions caused little effect on processing efficiency when the -1 Arg was replaced with Lys (PC1869) or when the -2 Lys was replaced with Arg (PC1461, Figure 5). However, when the Lys at position -2 was replaced with neutral Ala (PC1870), processing was nearly completely abolished (data not shown). These results are consistent with a specificity requiring basic amino acids in both the -1- and -2-positions and suggest that the efficiency with which basic pairs of amino acids are recognized by the endoprotease may depend more on the extended exposure of the site on the protein surface and the nature of the amino acid in the -4-position than on a preference for Lys-Arg pairs over other basic pair combinations.

These mutations serve to demonstrate a substrate sequence specificity for this processing endoprotease which is consistent with a specificity similar to or identical with that of the propeptide processing enzyme. Such an extended substrate specificity may partially serve to explain the lack of cleavage at several other Lys-Arg pairs within the protein C precursor, since none of the other pairs have an appropriate basic amino acid in the -4-position. Similar extended substrate sequence specificities may also help explain differential processing of prohormone precursors at alternative sites in different tissues.

By contrast, there is no evidence of specificity beyond the Lys-Arg pair for the single processing endoprotease identified to date—the KEX2 protease from yeast. Neither studies with small synthetic substrates (Julius et al., 1984) nor examination of the sequences cleaved by the protease in vivo has indicated a requirement for additional structural determinants in the substrates. Since not all Lys-Arg pairs are cleaved by KEX2 in processed precursors secreted from yeast (unpublished data), this selectivity is likely due to exposure of some cleavage sites to the protease on the surface of the protein due to folding constraints (Thim et al., 1986). To compare the sequence specificity of the KEX2 protease with that of the BHK endoprotease studied here, we have introduced the yeast *KEX2* cDNA in a mammalian cell expression vector into BHK cells producing native protein C. Consistent with the expected specificity of KEX2 for accessible Lys-Arg pairs, we have found that coexpression of the *KEX2* gene together with the cDNA for native protein C in BHK cells results in secretion of protein C which is nearly completely converted to 2-chain form (unpublished results).

The physiological role of endoproteolytic maturation of human protein C to 2-chain form remains unclear. Apparently, the unprocessed single-chain form from human plasma can be activated normally by thrombin as well as alternative activators from venom and has amidolytic and anticoagulant activity equivalent to that of the activated 2-chain form (Marlar, 1985). This is consistent with reports (Oppenheimer & Wydro, 1987) of full biological activity of recombinant preparations of protein C that are 50% 1-chain form. Our preparations of PC962 (100% 2-chain) and PC1477 (100% 1-chain) can both be fully activated in an amidolytic assay with nearly identical kinetics (data not shown), further supporting the accessibility of the activation site in both molecular forms. The biological significance of the maturation from 1-chain to 2-chain form may involve release of the highly acidic activation peptide, which would remain connected to the light chain upon activation of the 1-chain form. It is possible that the activation peptide participates in formation of a calcium-binding site important in overall conformation (Laurell et al., 1985) which would be eliminated by activation of the 2-chain form but not by activation of the 1-chain form. Alternatively, the 1-chain and 2-chain forms of either the zymogen or the activated form of protein C may have significantly different half-lives which could impart physiological significance to the processing events. Resolution of such hypotheses will require detailed analysis of the biochemical properties of the two forms of the protein.

ACKNOWLEDGMENTS

We thank Drs. Barbara Schach, Steve Leytus, and Earl Davie for the factor X cDNA and preliminary results on factor X cleavage and Dr. Eileen Mulvihill for the Zem229 vector. We thank Dr. Joseph Miletich for helpful discussions regarding proteolytic maturation in HepG2 cells. We also thank Drs. John Forstrom and Gary McKnight for critical reading of the manuscript.

Registry No. Protein C, pre-, 98530-77-9; protein C precursor endoprotease, 124041-97-0; protein C, prepro-, 98530-79-1.

REFERENCES

Beckmann, R. J., Schmidt, R. J., Santerre, R. F., Plutzky, J., Crabtree, G. R., & Long, G. L. (1985) *Nucleic Acids Res. 13*, 841.

Belt, K. T., Carroll, M. C., & Porter, R. R. (1984) *Cell 36*, 907.

Bentley, A. K., Rees, D. J. G., Rizza, C., & Brownlee, G. G. (1986) *Cell 45*, 343.

Brandstrom, A., & Ranby, M. (1987) *Thromb. Haemostasis 58*, Abstract 1609.

Busby, S. J., Berkner, K. L., Halfpap, L. M., Gambee, J. E., & Kumar, A. A. (1988) *Current Advances in Vitamin K Research* (Suttie, J., Ed.) pp 173–181, Elsevier Science Publishing, Amsterdam.

Crestfield, A. M., Morre, S., & Stein, W. H. (1963) *J. Biol. Chem. 238*, 622.

DeBruijn, M. H. L., & Fey, G. H. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 708.

Degen, S. J. F., & Davie, E. W. (1986) *Ann. N.Y. Acad. Sci. 485*, 66.

DiScipio, R. G., & Davie, E. W. (1979) *Biochemistry 18*, 899.

Diuguid, D. L., Rabiet, M. J., Furie, B. C., Liebman, H. A., & Furie, B. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5803.

Dmochowska, A., Dignard, D., Henning, D., Thomas, D. Y., & Bussey, H. (1987) *Cell 50*, 573.

Douglass, J., Civelli, O., & Herbert, E. (1984) *Annu. Rev. Biochem. 53*, 665.

Drakenburg, T., Fernlund, P., Roepstorff, P., & Stenflo, J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1802.

Esmon, C. T. (1987) *Science 235*, 1348.

Fernlund, P., & Stenflo, J. (1982) *J. Biol. Chem. 257*, 12170.

Foster, D., & Davie, E. W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4766.

Foster, D., Yoshitake, S., & Davie, E. W. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 4673.

Foster, D., Rudinsky, M. S., Schach, B. G., Berkner, K. L., Kumar, A. A., Hagen, F. S., Sprecher, C. A., Insley, M. Y., & Davie, E. W. (1987) *Biochemistry 26*, 7003.

Fung, M. R., Hay, C. W., & MacGillivray, R. T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 3591.

Graham, F. L., & van der Eb, A. J. (1973) *J. Virol. 52*, 456.

Grinnell, B. W., Berg, D. T., Walls, J., & Yan, S. B. (1987) *Biotechnology 5*, 1189.

Hagen, F. S., Gray, C. L., O'Hara, P. J., Grant, F. J., Saari, G. c., Woodbury, R. G., Hart, C. E., Insley, M., Kisiel, W., Kurachi, K., & Davie, E. W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 2412.

Jorgensen, M. J., Cantor, A. B., Furie, B. C., Brown, C. L., Shoemaker, C. B., & Furie, B. (1987) *Cell 48*, 185.

Julius, D., Brake, A., Blair, L., Kunisawa, R., & Thorner, J. (1984) *Cell 37*, 1075.

Kisiel, W., Ericsson, L. H., & Davie, E. W. (1976) *Biochemistry 15*, 4893.

Krieger, D. T., & Liotta, A. S. (1979) *Science 205*, 366.

Kurachi, K., & Davie, E. W. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6461.

Laurell, M., Ikeda, K., Lindgren, S., & Stenflo, J. (1985) *FEBS Lett. 191*, 75.

Leytus, S. P., Chung, D. W., Kisiel, W., Kurachi, K., & Davie, E. W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 3699.

Lundwall, A. B., Wetsel, R. A., Kristensen, T., Whitehead, A. S., Woods, D. E., Ogden, R. C., Colten, H. R., & Tack, B. F. (1985) *J. Biol. Chem. 260*, 2108.

Marlar, R. A. (1985) *Thromb. Haemostasis 54*, 216, Abstract 01275.

McMullen, B., Fujikawa, K., & Kisiel, W. (1983) *Biochem. Biophys. Res. Commun. 115*, 8.

Miletich, J. P., Leykam, F. J., & Broze, G. J. (1983) *Blood Suppl. 1, 62*, 306a.

Oppenheimer, C., & Wydro, R. (1988) *Current Advances in Vitamin K Research* (Suttie, J., Ed.) p 165, Elsevier Science Publishing, Amsterdam.

Steiner, D. F., Quinn, P. S., Chan, S. J., Marsh, J., & Tager, H. S. (1980) *Ann. N.Y. Acad. Sci. 343*, 1.

Stenflo, J. (1976) *J. Biol. Chem. 251*, 355.

Stenflo, J., Ohlin, A., Owen, W. G., & Schneider, W. J. (1988) *J. Biol. Chem. 263*, 21.

Subramani, S., Mulligan, R. C., & Berg, P. (1981) *Mol. Cell. Biol. 1*, 854.

Thim, L., Hansen, M. T., Norris, K., Hoegh, I., Boel, E., Forstrom, J., Ammerer, G., & Fiil, N. P. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6766.

Thomas, G., Thorne, B. A., Thomas, L., Allen, R. G., Hruby, D. E., Fuller, R., & Thorner, J. (1988) *Science 241*, 226.

von Heijne, G. (1984) *J. Mol. Biol. 173*, 243.

von Heijne, G. (1986) *Nucleic Acids Res. 14*, 4683.

Zoller, M., & Smith, M. (1982) *Nucleic Acids Res. 10*, 6487.