Amino Acid Sequence of the Ribonuclease Inhibitor from Porcine Liver Reveals the Presence of Leucine-Rich Repeats

Jan Hofsteenge,* Brigitte Kieffer, Renate Matthies, Brian A. Hemmings, and Stuart R. Stone
Friedrich Miescher Institut, Postfach 2543, CH-4002 Basel, Switzerland
Received May 31, 1988; Revised Manuscript Received August 12, 1988

ABSTRACT: The primary structure of the ribonuclease inhibitor from pig liver has been determined by amino acid sequence analysis. The N^{α}-acetylated polypeptide chain of 456 amino acids consists of 15 homologous leucine-rich repeats, characterized by leucyl residues at constant positions. Two types of alternating repeats occur, 29 (A) and 28 (B) residues long. The degree of identity between repeats of a given type ranged from 25 to 60%. Only one deletion in the B-repeat was necessary to perfectly align the leucyl residues between the two repeats. Leucine-rich repeats have previously been found in four membrane-bound proteins and one extracellular protein, and their amphiphilic character suggested that they could be involved in membrane binding. Ribonuclease inhibitor is the first example of a cytoplasmic protein containing this type of repeat. It seems likely, therefore, that leucine-rich repeats can have functions other than forming membrane binding structures.

Many types of ribonuclease (RNase)[1] are found in every cell type, and it is assumed that these enzymes play a role in the metabolism of the different kinds of RNA. Whereas a good deal is known about the function of different RNases in *Escherichia coli*, little is known about these enzymes in eucaryotes (Deutscher, 1988).

In the cytoplasm of many organisms latent RNases have been found that occur as a complex with a protein RNase inhibitor (Roth, 1956, 1958, 1962; Kraft & Shortman, 1970). The occurrence of this inhibitor seems to be widespread, since it has been found in insects (Aoki & Natori, 1981), birds (Dijkstra et al., 1978), and mammals (Roth, 1967), and it has been purified to homogeneity from several tissues, including placenta (Blackburn et al., 1977; Blackburn, 1979), brain (Burton et al., 1980), and liver (Burton & Fucci, 1982). The protein has an apparent molecular mass of approximately 50 kDa and forms a 1:1 complex with (pancreatic) RNase (Blackburn et al., 1977). Amino acid analysis of the inhibitor revealed an unusually high content of cysteinyl and leucyl residues (Blackburn et al., 1977), and it has been shown that reagents that modify sulfhydryl gorups inactivate both the free and complexed inhibitor (Roth, 1956, 1958; Shortman, 1961, 1962a,b; Blackburn et al., 1977). This modification results in the release of active enzyme from the complex. It has been speculated that thiol modification could be the means by which an equilibrium between free and inhibited RNase is regulated in vivo (Roth, 1958; Aoki & Natori, 1981). Evidence that this enzyme–inhibitor system could function in vivo in RNA metabolism comes from the observation that tissue which is active in protein synthesis has a high excess of inhibitor over RNAse. In contrast, in catabolically active tissues that do not accumulate RNA this ratio is low (Kraft & Shortman, 1970; Shortman, 1962a,b; Kraft et al., 1969).

The interaction of the inhibitor with pancreatic RNase has been studied in considerable detail [for a review, see Blackburn and Moore (1982)]. A number of regions on the surface of the RNase molecule that interact with the inhibitor have been determined by chemical modification and by the use of proteolytically altered RNase derivatives (Blackburn & Jailkhali, 1979; Blackburn & Gavilanes, 1980, 1982).

Recently, it has been found that angiogenin, a protein that induces blood vessel growth, is 35% identical with human pancreatic RNase (Fett et al., 1985; Strydom et al., 1985; Kurachi et al., 1985) and contains ribonucleolytic activity (Shapiro et al., 1986). The substrate specificity of angiogenin seems, however, to be more restricted than that of the pancreatic RNases. Whereas limited cleavage of 18S and 28S ribosomal RNA occurs, substrates such as C>p or homonucleotide polymers are not hydrolyzed (Shapiro et al., 1986). Interestingly, angiogenin binds tightly to the RNase inhibitor from human placenta, which results in the inhibition of both its angiogenic and ribonucleolytic activity (Shapiro & Vallee, 1987).

In this paper we report on the determination of the complete amino acid sequence of the RNase inhibitor from porcine liver and show that it has been built from an alternating pattern of two types of leucine-rich repeats.

Materials and Methods

Materials. Enzymes were obtained from the following suppliers: pancreatic RNase A, Sigma, St. Louis, MO; TPCK-treated bovine trypsin and protease from *Staphylococcus aureus* V_8, Worthington, Malvern, PA; lysine-specific protease from *Lysobacter enzymogenes* (endoproteinase Lys-C), Boehringer, Mannhein, FRG. Reagents for protein chemistry were obtained from the following sources: CNBr and iodoacetic acid, Fluka, Buchs, Switzerland; BNPS-skatole and citraconic acid anhydride, Pierce, Rockford, IL; guanidinium chloride, Schwarz/Mann, Orangeburg, NY. All other reagents were of the highest purity commercially available.

Purification of Ribonuclease Inhibitor. RNase inhibitor was purified from porcine liver by affinity chromatography on RNase–Sepharose (Burton & Fucci, 1982). The protein was further purified by ion-exchange chromatography on an FPLC Mono Q (HR 5/5) column (Pharmacia, Uppsala, Sweden), equilibrated in 20 mM Tris-HCl, pH 7.5, containing 5 mM DTT, 1 mM EDTA, and 0.1% (w/v) poly(ethylene

* To whom correspondence should be addressed.

[1] Abbreviations: BNPS-skatole, 2-[(2-nitrophenyl)sulfenyl]-3-bromo-3-methylindolenine; DABSYL-Cl, 4-(dimethylamino)azobenzene-4′-sulfonyl chloride; K_{av}, partitition coefficient; RNase, ribonuclease; TFA, trifluoroacetic acid; SDS, sodium dodecyl sulfate.

0006-2960/88/0427-8537$01.50/0

glycol) (M_r 6000). The inhibitor was eluted with a linear gradient of 0–300 mM NaCl in 45 min at a flow rate of 1 mL/min. The yield was 3–4 mg/kg of liver. The protein migrated as a single band on polyacrylamide–SDS gel electrophoresis (Laemmli, 1970) with an apparent molecular mass of 48 kDa. The concentration of active inhibitor was determined by titration of an accurately known amount of RNase (50.1 nM affinity purified enzyme; Wierenga et al., 1973) and determination of the remaining RNase activity with 2′,3′-c-CMP as substrate (Blackburn, 1979). Comparison of this value with the concentration determined by either the method of Bradford (1976) or by measurement of the absorbance at 280 nm (Burton & Fucci, 1982) revealed that the preparation was 97% active.

Fragmentation of the Protein and Isolation of the Peptides. RNase inhibitor was reduced with β-mercaptoethanol and carboxymethylated with iodoacetic acid as described by Crestfield et al. (1963). Excess reagent was removed by dialysis against 50 mM NH_4HCO_3, and the protein was recovered by lyophilization.

CNBr cleavage of 1-mg amounts of the reduced and carboxymethylated protein was performed by the method of Gross and Witkopf (1962). The digest was lyophilized, redissolved in 50 mM NH_4HCO_3, and fractionated by HPLC gel filtration on a Superose 12 column (Pharmacia, Uppsala, Sweden), equilibrated in the same buffer. The flow rate was 0.5 mL/min, and peptides were detected by monitoring the absorbance at 280 nm. Individual peaks (CB2 and CB3) were further purified by ion-exchange chromatography on a Mono Q column equilibrated with 50 mM NH_4HCO_3, using a linear gradient of 0–0.49 M NaCl over 45 min with a flow rate of 1.0 mL/min. CB1 (*N*-acetylhomoserine) was further purified by reversed-phase chromatography on a Vydac C_{18} column (Vydac, Hesperia, CA) employing isocratic elution with 0.1% TFA at a flow rate of 0.5 mL/min.

Cleavage with hydroxylamine at the single Asn–Gly bond present in the molecule was performed by the method of Bornstein and Balian (1977), and the C-terminal peptide was purified by HPLC gel filtration as described above.

Cleavage of 1.2 mg of the reduced and carboxymethylated inhibitor with trypsin (1% w/w), and of 50 μg with the protease from *S. aureus* V_8 (2% w/w), was performed under the conditions described previously (Vereyken et al., 1980).

Fragmentation of the CNBr Peptides and Isolation of Subpeptides. Subdigestion of the two large CNBr fragments with trypsin (T), the protease from *S. aureus* V_8 (E), and endoproteinase Lys-C (K) was performed as described previously (Vereijken et al., 1980; Hofsteenge et al., 1983). Cleavage with BNPS-skatole was done as described by Fontana and Gross (1986). To limit tryptic digestion to arginyl residues (T_R), cyanogen bromide peptide CB2 was citraconylated as described by Weng et al. (1978). Decitraconylation was achieved by incubation in 10% acetic acid for 18 h.

Small peptides resulting from enzymatic digestions were purified by reversed-phase HPLC using a C_{18} column (Vydac), or a C_8 column (Brownlee, Santa Clara, CA). The eluent was 0.1% TFA, and a linear gradient of acetonitrile in 0.09% TFA (0–56% in 90 or 300 min) was employed. If necessary, groups of peaks were rerun with a shallower gradient (0.18% acetonitrile/min). The peptides resulting from cleavage with BNPS-skatole were purified by reversed-phase HPLC on a Vydac C_4 column, using the 0.1% TFA/acetonitrile system.

Amino Acid Analysis and Sequence Determination. Peptides were hydrolyzed in the vapor of 6 N HCl, and the amino acid composition was determined according to the Pico-Tag method (Heinrikson & Meredith, 1984; Bidlingmeyer et al., 1986). Automatic Edman degradation was performed on a gas-phase sequencer (Applied Biosystems, Foster City, CA) equipped with an on-line PTH analyzer. The program 03RPTH supplied by the manufacturer was employed. In general, 50–300 pmol of peptide was sequenced.

Table I: Amino Acid Composition of the CNBr Peptides and the Entire Protein[a]

	CB1	CB2	CB3	protein A[b]	protein B[c]
Asx		25.4 (28)	10.6 (12)	31.4 (40)	44
Glx		41.1 (41)	17.4 (19)	51.6 (60)	60
CmCys		23.5 (24)	5.7 (6)	31.4 (30)	33
Ser		23.2 (25)	11.9 (13)	34.1 (38)	40
Gly		25.8 (26)	10.1 (11)	37.3 (37)	40
His		6.0 (6)	1.1 (1)	6.6 (7)	8
Arg		18.6 (17)	5.3 (5)	21.8 (22)	20
Thr		16.7 (17)	5.5 (6)	22.6 (23)	23
Ala		27.8 (27)	5.6 (5)	31.7 (32)	32
Pro		13.4 (12)	4.3 (4)	15.5 (16)	14
Tyr		2.5 (2)	2.0 (2)	4.8 (5)	4
Val		11.8 (12)	6.1 (7)	18.7 (19)	18
Met	+[d] (1)	+[d] (1)		2.4 (2)	2
Ile		5.6 (7)	1.5 (2)	5.5 (9)	8
Leu		71.0 (70)	24.9 (28)	90.0 (98)	87
Phe				(0)	2
Lys		11.5 (11)	3.1 (3)	14.0 (14)	14
Trp	ND	ND (4)	ND (1)	ND (5)	5
total	1	330	125	456	

[a] Polypeptides were hydrolyzed for 24 h. No corrections have been made for losses of Ser and Thr. The numbers obtained from the sequence are given in parentheses. [b] This study. [c] Recalculated from Burton and Fucci (1982), with a length of 456 residues for the polypeptide chain. [d] Determined as homoserine.

Analysis of Primary Structure. Searches for homology with previously determined primary structures were made by comparing the sequence in Figure 2 with the sequences in the SWISSPROT data bank (version 6; A. Bairoch, University of Geneva) with the FASTP program of Lipman and Pearson, (1985).

Results

General Strategy for Sequence Determination. The intact protein, both unmodified or reduced and carboxymethylated, was not susceptible to Edman degradation, suggesting the presence of a blocked N-terminal residue. Amino acid analysis of the protein (Table I) indicated the occurrence of two methionyl residues, predicting the formation of three peptides upon CNBr cleavage, one of which should be blocked. In agreement with this, quantitative N-terminal analysis (Chang, 1983a) of the complete CNBr digest showed the presence of 13.1 nmol of Asn and 11.5 nmol of Leu. Fractionation of the mixture by gel filtration resulted in two major peaks (Figure 1) containing peptide CB2 (330 residues, K_{av} = 0.22) and CB3 (125 residues, K_{av} = 0.39). In addition, a blocked form of homoserine was found in the included volume of the column. The CNBr peptides were further purified as described under Materials and Methods and their amino acid compositions determined (Table I). The detailed proof of the amino acid sequence is shown in Figure 2 and consists of the sequences of the three individual CNBr peptides and those of the two overlapping fragments T and E. The latter two were obtained from a digest of the entire protein with trypsin and the protease from *S. aureus* V_8, respectively.

Peptide CB1. Amino acid analysis of CB1 yielded homoserine only. However, no amino acid was found if hydrolysis with 6 N HCl was omitted, suggesting the presence of a blocking group on the α-amino function of the homoserine.

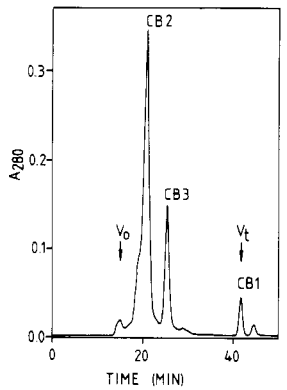


FIGURE 1: Fractionation of the CNBr peptides of the porcine ribonuclease inhibitor. One-milligram amounts of reduced and carboxymethylated RNase inhibitor were cleaved with CNBr as described under Materials and Methods. The digest (500-μg portions) was fractionated on a Superose 12 column (1.2 × 30 cm) equilibrated in 50 mM NH_4HCO_3. The flow rate was 0.5 mL/min, and the peaks were collected by hand. The peptides (CB) are labeled according to their occurrence in the polypeptide chain and were purified to homogeneity as described under Materials and Methods. V_0, void volume; V_t, included volume. No peptide material was found by amino acid analysis in the peaks eluting after the included volume and in the region between CB3 and CB1.

The results of mass spectroscopic analysis were consistent with the presence of an N^{α}-acetyl group ($m/e + H^+ = 144$, which corresponds to N^{α}-acetylhomoserine lactone).

Peptide CB2. Automatic Edman degradation of peptide CB2 showed it to be pure and yielded the sequence of this peptide up to and including residue 37. A complete set of subpeptides was obtained by tryptic digestion of unmodified CB2 and of CB2 that had been treated with citraconic acid anhydride. Overlapping peptides generated by cleavage with the protease from *S. aureus* V_8 and endoproteinase Lys-C were used to obtain most of the sequence of CB2 (Figure 2). An additional peptide (HA) was purified after hydroxylamine cleavage of the entire protein (see Materials and Methods), which provided the overlap for peptides CB2T17 and CB2T19. The amino acid compositions of the tryptic peptides, given in Table II, were in good agreement with the amino acid sequence. The sequence of the C-terminal part of this peptide was obtained by sequencing peptide E, which formed the overlap with peptide CB3. It was confirmed by sequence analysis of peptide CB2T23, which was shown to terminate in homoserine. During sequence analysis of peptide CB2K2, both Gly and Ser were found at cycle 3 (residue 44). This heterogeneity was further confirmed by the occurrence of two peptides from this area formed by cleavage with trypsin (CB2T3A and CB2T3B; Table II) or the protease from *S. aureus* V_8 (data not shown).

The amino acid composition of peptide CB2 calculated from the sequence was in good agreement with the one determined directly (Table I). In addition, all residues of this peptide were sequenced at least twice.

Peptide CB3. The amino acid composition of peptide CB3 showed that it did not contain homoserine, which positioned this peptide at the C-terminus of the protein. Automatic Edman degradation of the intact peptide showed it to be pure and established the sequence up to and including residue 378. A complete set of subpeptides was purified from a digest with the protease from *S. aureus* V_8, and overlapping peptides were obtained by subdigestion of CB3 with trypsin and BNPS-skatole (Figure 2). The amino acid composition data of the former are shown in Table III and were in good agreement with the sequence. In addition, all tryptic peptides, except the one encompassing residues 336–347, were isolated and sequenced. The C-terminal sequence from residue 437 onward was established by sequencing peptide CB3W (Figure 2). In total, three peptides obtained from CB3 by different cleavage methods (CB3W, CB3E9, and CB3T8) had the same C-terminal sequence: -Val-Ile-Ser (Figure 2). A tryptic peptide with the sequence Val-Ile-Ser was also isolated from the entire protein. Therefore, this is the C-terminus of peptide CB3 and of the protein. One overlap in the sequence of CB3 was marginal (between CB3E4 and CB3T6). Since the sequence Arg-Glu occurs only once in CB3, we consider the proof of the structure to be sufficient. These conclusions are further strengthened by the agreement between the determined amino acid composition and the one calculated from the amino acid sequence (Table I). In addition, all amino acid residues of this peptide were sequenced at least twice. Moreover, the sequence Arg-Glu fits the pattern of conserved residues in the leucine-rich repeats (see Discussion).

Overlap of Peptides CB1, CB2, and CB3. Peptide T (Figure 2) was purified from a tryptic digest of the entire protein and was found by amino acid analysis to contain one residue of methionine (Table II) in addition to the residues occurring in the N-terminal tryptic peptide of CB2 (residues 2–14). The peptide was not susceptible to Edman degradation, suggesting that it was N-terminally blocked. Treatment of peptide T with CNBr produced blocked homoserine and a peptide with a sequence identical with that of residues 2–14. These data, together with those obtained by quantitative N-terminal analysis of the complete CNBr digest (described above), showed that peptide CB1 is the N-terminus of the inhibitor.

The overlap between peptides CB2 and CB3 was established by the sequence determination of peptide E.

DISCUSSION

The correctness of the sequence of the inhibitor as shown in Figure 2 is confirmed by the agreement between the determined amino acid composition and the one calculated from the sequence (Table I). No obvious weaknesses occur in the presented sequence (Figure 2), and each amino acid residue has been determined at least twice. Generally, a good agreement existed also between the amino acid composition determined by Burton and Fucci (1982) and the one found here. The only major discrepancy was that no phenylalanine occurs in the sequence, whereas Burton and Fucci (1982) found 2 mol of Phe/mol of inhibitor. Our evidence consists of the complete absence of Phe in the composition of the entire protein, the CNBr peptides, and the fact that no Phe was found during sequence analysis of any of the subpeptides. This is further strengthened by the fact that we did not find phenylalanine following amino acid analysis of the entire protein determined by two additional methods: the DABSYL-Cl method of Chang (1983b) and the classical ion-exchange method using postcolumn detection with ninhydrin (Spackman et al., 1958).

The molecular mass of the polypeptide chain calculated from the sequence, 49 093 Da, is close to the apparent molecular mass of 48–50 kDa reported by other laboratories (Burton & Fucci, 1982; Blackburn et al., 1977; Brockdorf & Knowler, 1987). At neutral pH the molecule posesses a nominal net charge of −18. This is in contrast to RNases, which are generally positively charged (Beintema, 1987).

The N-terminal sequence of the protein was found to be *N*-acetyl-Met-Asn-. As pointed out by Flinta et al. (1986), Asn is one of the three amino acids most commonly found in the penultimate position when the N-terminal methionyl residue is acetylated.

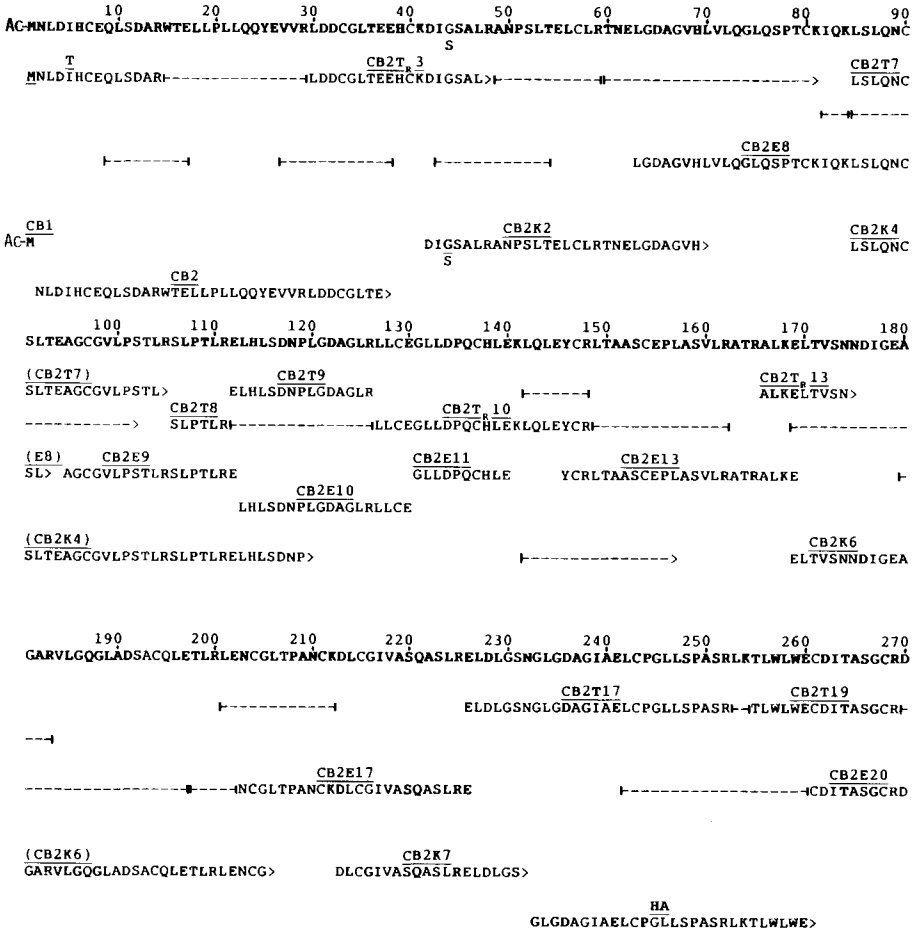


FIGURE 2: Proof of the amino acid sequence of the porcine liver ribonuclease inhibitor. The amino acid sequence is shown in boldface. Peptide names are shown above the sequences and are underlined. They are numbered according to their occurrence in the polypeptide chain. The following codes have been used: CB, cyanogen bromide; E, protease from *S. aureus* V_8; HA, hydroxylamine; K, endoproteinase Lys-C; T, trypsin; T_R, trypsin after citraconylation; W, BNPS-skatol. (>) indicates that the peptide has not been sequenced up to its C-terminus. In one case the assignment was ambiguous (Thr-333 in peptide E), which has been indicated by a lower-case letter. In peptide T the methionyl residue at position 1 has been identified by amino acid analysis, which has been indicated by underlining. Sequenced peptides that are only confirmatory have been indicated with a dashed line (their names have been omitted from this figure for sake of clarity). The amino acid compositions of peptides encompassing the entire polypeptide chain are given in Tables II and III.

Table II: Amino Acid Composition of the Peptides Obtained by (Limited) Tryptic Digestion of CB2[a]

	T (1–14)	CB2T2 (15–29)	CB2TR3 (30–48)	CB2T3A (42–48)	CB2T3B (42–48)	CB2T4 (49–59)	CB2T5 (60–81)	CB2T6 (82–84)	CB2T7 (85–105)
Asx	2.3 (3)		3.2 (3)	0.9 (1)	0.9 (1)	0.7 (1)	1.9 (2)		0.8 (1)
Glx	1.9 (2)	3.6 (4)	1.6 (2)	0.4 (0)		1.0 (1)	3.0 (3)	0.8 (1)	1.5 (2)
CmCys	0.7 (1)		2.2 (2)			0.6 (1)	0.8 (1)		1.9 (2)
Ser	1.0 (1)		1.0 (1)	1.0 (1)	1.7 (2)	1.1 (1)	0.9 (1)		2.6 (3)
Gly			2.4 (2)	0.7 (1)		0.2	2.9 (3)		1.9 (2)
His	0.8 (1)		0.9 (1)				0.7 (1)		
Arg	1.1 (1)	1.1 (1)	1.0 (1)	1.0 (1)	1.0 (1)	1.1 (1)			1.0 (1)
Thr		1.0 (1)	0.9 (1)			1.0 (1)	1.8 (2)		1.8 (2)
Ala	1.1 (1)		1.1 (1)	0.9 (1)	1.0 (1)	1.1 (1)	1.0 (1)		1.0 (1)
Pro		1.3 (1)				1.1 (1)	1.0 (1)		1.0 (1)
Tyr		1.2 (1)							
Val		1.9 (2)					1.9 (2)		0.9 (1)
Met	0.8 (1)								
Ile	0.8 (1)		1.2 (1)	0.9 (1)	0.9 (1)			1.1 (1)	
Leu	2.1 (2)	4.5 (4)	2.8 (3)	0.9 (1)	1.1 (1)	3.0 (3)	3.9 (4)		5.1 (5)
Phe									
Lys			1.0 (1)				1.0 (1)	1.0 (1)	
Trp	ND	ND (1)	ND	ND	ND	ND	ND	ND	ND
total	14	15	19	7	7	11	22	3	21
yield (%)	40	1[b]	5	10	16	11	40	38	59

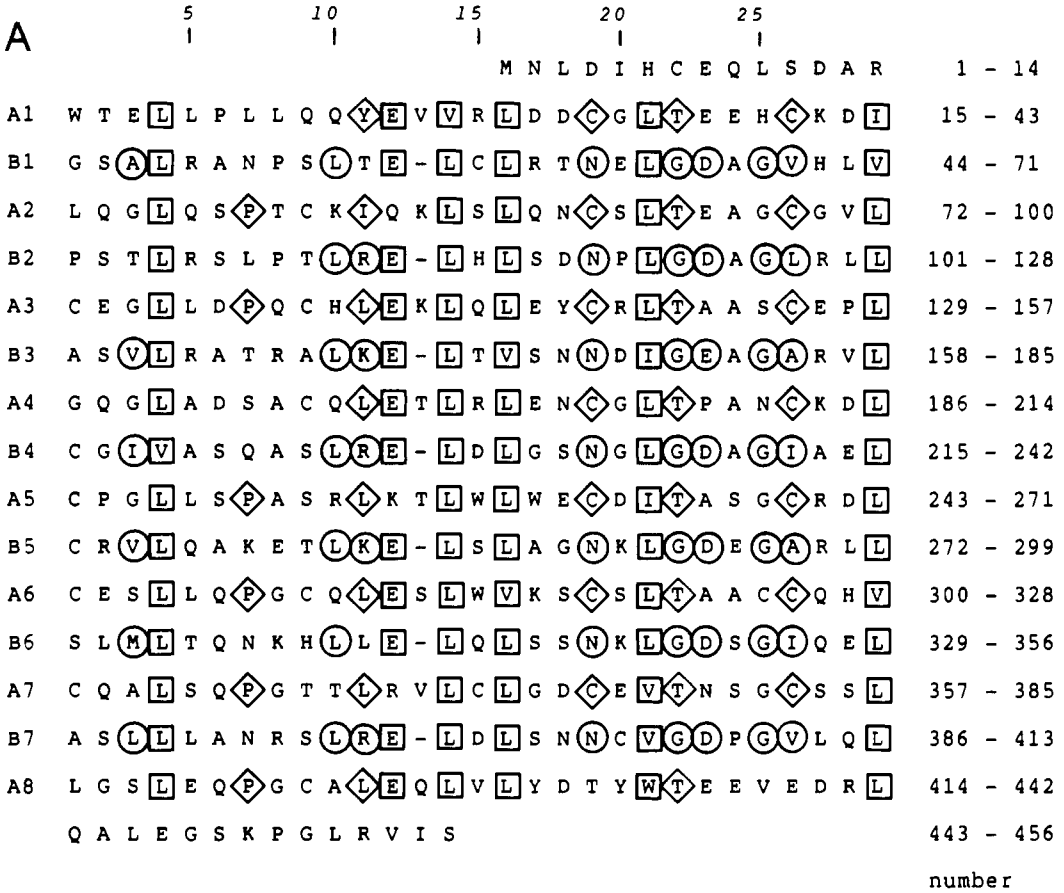
	CB2T8 (106–111)	CB2T9 (112–126)	CB2TR10 (127–148)	CB2T11 (149–162)	CB2T12 (163–165)[c]	CB2TR13 (166–183)	CB2T14 (184–200)	CB2T15 (201–212)
Asx		2.2 (3)	1.0 (1)			2.2 (3)	0.7 (1)	2.0 (2)
Glx		0.6 (1)	4.2 (5)	0.9 (1)		1.9 (2)	1.8 (3)	1.2 (1)
CmCys			2.0 (3)	0.7 (1)			0.8 (1)	1.7 (2)
Ser	0.9 (1)	0.8 (1)		1.8 (2)		1.1 (1)	0.9 (1)	
Gly		1.7 (2)	2.1 (1)	0.6		2.3 (2)	2.3 (2)	1.1 (1)
His		0.7 (1)	0.9 (1)					
Arg	1.0 (1)	0.9 (1)	1.1 (1)	1.1 (1)	ND (1)	1.2 (1)	1.1 (1)	
Thr	1.0 (1)			0.9 (1)	ND (1)	1.0 (1)	0.9 (1)	1.2 (1)
Ala		1.0 (1)		3.0 (3)	ND (1)	3.2 (3)	2.2 (2)	1.1 (1)
Pro	1.1 (1)	0.9 (1)	ND[d] (1)	1.1 (1)				1.3 (1)
Tyr			0.7 (1)					
Val				1.0 (1)		1.0 (1)	1.1 (1)	
Met								
Ile						1.0 (1)		
Leu	2.1 (2)	4.0 (4)	7.0 (7)	3.1 (3)		1.9 (2)	4.2 (4)	2.3 (2)
Phe								
Lys			1.1 (1)			0.9 (1)		1.0 (1)
Trp	ND	ND	ND	ND	ND	ND	ND	ND
total	6	15	15	14	3	18	17	12
yield (%)	21	10	1	3		3[b]	6	33

	CB2T16 (213–225)	CB2T17 (226–252)	CB2T18 (253–254)	CB2T19 (255–269)	CB2T20 (270–273)	CB2TR21 (274–297)	CB2T22 (298–316)	CB2T23 (317–331)
Asx	1.0 (1)	2.2 (3)		0.7 (1)	0.8 (1)	1.5 (2)		
Glx	1.0 (1)	1.8 (2)		0.9 (1)		3.7 (4)	3.9 (4)	1.0 (1)
CmCys	0.7 (1)	0.8 (1)		1.4 (2)	0.7 (1)		1.3 (2)	1.6 (3)
Ser	1.6 (2)	2.6 (3)		1.2 (1)		1.4 (1)	1.5 (2)	3.2 (3)
Gly	1.0 (1)	4.8 (5)		1.2 (1)		4.5 (3)	0.9 (1)	
His								1.2 (1)
Arg	0.9 (1)	1.1 (1)		1.1 (1)	0.9 (1)	1.1 (1)		
Thr				1.8 (2)		1.0 (1)		0.8 (1)
Ala	2.0 (2)	3.0 (3)		1.1 (1)		3.2 (3)		1.9 (2)
Pro		2.2 (2)					1.2 (1)	
Tyr								
Val	0.9 (1)					1.0 (1)	1.2 (1)	1.0 (1)
Met								+[e] (1)
Ile	1.2 (1)	1.0 (1)		1.0 (1)				
Leu	2.3 (2)	6.3 (6)	1.0 (1)	2.1 (2)	1.0 (1)	6.0 (5)	7.2 (6)	2.1 (2)
Phe								
Lys			1.0 (1)			3.1 (3)	0.9 (1)	
Trp	ND	ND	ND	ND (2)	ND	ND	ND (1)	ND
total	13	27	2	17	4	24	19	15
yield (%)	10	38	49	23	21	1[b]	13	ND

[a] See legend to Table I for details. [b] Only a small part of the peak was collected in these cases. [c] This polar peptide did not bind to a C_{18} column and was not recovered. [d] Not determined due to overlap between NH_3 and Pro. [e] Identified as homoserine.

A heterogeneity was observed at position 44 where both glycine and serine occur. Since the inhibitor used in these studies was purified from pooled porcine liver, it is not clear whether this is caused by allelic polymorphism or whether more than one gene for the inhibitor exists.

Between residues 15 and 442 the polypeptide chain contains a considerable amount of internal homology (Figure 3A). It was possible to recognize 15 repeats, on the basis of the dis-



B

		number of repeats	localization
1.	x x x L x x P x x x L E x L x L x x C x L T x x x C x x L	8	cytoplasm
2.	x x α L x x x x x L x E - L x L x x N x L G D x G α x x L	7	
3.	P P G L L Q G L P Q L R x L D L S G N x L E S - - - - - L	13	extracellular
4.	P x G L L x x L P x L x x L x L S x N x L T T - - - - - L	8	membrane (e.c.)
5.	P x x L F x H x x N L x x L x L x x N x L x x - - - - - L	15	membrane (e.c.)
6.	P x x x F x x L x x L x x L D L S x N x I x x - - - - - L	41	membrane (e.c.)
7.	P x x α - x x L x x L x x L x L x x N x α x x - - - - - α	26	membrane (i.c.)

FIGURE 3: Internal homology of the RNase inhibitor. (A) The amino acid sequence of the inhibitor has been optimally aligned to show its repetitive nature between residues 15 and 442. One gap, at position 13 in the type B repeat, has been introduced. The numbers at the top (italics) indicate the position in the repeat, whereas the position within the entire protein is indicated at the end of each line. Residues that are common to both types of repeats have been indicated with (□), type A specific residues with (◇), and type B-specific residues with (○). (B) The consensus sequences for the two repeats of the RNase inhibitor (A and B) were compared with consensus sequences of leucine-rich repeats found in the literature. A five amino acid residue insert was introduced, but no other changes were made to the published sequences. (1) RNase inhibitor (type A repeat) (this paper); (2) RNase inhibitor (type B repeat) (this paper); (3) leucine-rich glycoprotein (Takahashi et al., 1985); (4) glycoprotein 1bα (Lopez et al., 1987; Titani et al., 1987); (5) Toll (Hashimoto et al., 1988); (6) Chaoptin (Reinke et al., 1988); (7) adenylate cyclase (Kataoka et al., 1985). e.c., extracellular; i.c., intracellular; α, aliphatic residue.

tribution of apolar (nearly always leucine) and polar residues. Interestingly, two different types of repeats can be recognized that occur alternately. The type A repeat is 29 amino acids long, and the degree of identity between the eight repeats ranges from 24 to 52% (average ± SD: 35 ± 8%), whereas the degree of identity for the 28 residue long type B repeat ranged from 29 to 64% (average ± SD: 44 ± 7%). In contrast, the average degree of identity between the type A and B repeats was found to be only 21 ± 4.5%. Only one deletion in the type B repeat was necessary to align the leucyl residues between the two repeats, as shown in Figure 3A.

The two types of repeats are rich in leucyl residues and have a number of features in common (Figure 3A): the occurrence of a hydrophobic amino acid (Leu in 84% of the cases) at positions 4, 14, 16, 21, and 29 and a charged residue (most commonly Glu) at position 12. The two types of repeats can be distinguished by the amino acids that occupy positions 11, 19, 22, and 26. Whereas in the type A repeat a hydrophobic residue (often Leu), Cys, Thr, and Cys occur at these positions (identified by diamonds in Figure 3A), in the type B repeat a basic residue, Asn, Gly, and a hydrophobic residue are found (identified by circles in Figure 3A). Furthermore, the type A repeat most frequently contains a Pro at position 7. Characteristically, positions 3 and 10 in the type B repeat are occupied by hydrophobic residues and positions 23 and 25 by Gly and an acidic residue, most frequently Asp, respectively.

Table III: Amino Acid Composition of the Peptides Obtained by Digestion of CB3 with the Protease from *S. aureus* V_8[a]

	CB3E1 (332–340)	CB3E2 (341–355)	CB3E3 (356–376)	CB3E4 (377–397)	CB3E5 (398–425)	CB3E6 (426–431)	CB3E7 (432–437)	CB3E8 (438–446)	CB3E9 (447–456)
Asx	1.1 (1)	1.9 (2)	1.2 (1)	1.6 (2)	3.7 (4)	0.8 (1)		0.8 (1)	
Glx	2.2 (2)	3.4 (3)	2.6 (3)	0.9 (1)	4.0 (4)	0.9 (1)	1.8 (2)	2.8 (3)	
CmCys			2.4 (3)	0.8 (1)	1.8 (2)				
Ser		2.5 (3)	1.0 (1)	4.5 (5)	2.0 (2)				1.6 (2)
Gly		2.1 (2)	2.0 (2)	1.1 (1)	4.1 (4)				2.0 (2)
His	1.0 (1)								
Arg			1.0 (1)	2.1 (2)				1.0 (1)	1.1 (1)
Thr	1.0 (1)		1.8 (2)	0.8 (1)			2.1 (2)		
Ala			1.0 (1)	1.9 (2)	1.1 (1)			1.0 (1)	
Pro			1.0 (1)		2.0 (2)				1.0 (1)
Tyr						0.9 (1)	1.1 (1)		
Val			1.0 (1)	1.0 (1)	2.0 (2)	1.0 (1)		1.1 (1)	0.8 (1)
Met									
Ile		0.8 (1)							0.8 (1)
Leu	3.2 (3)	3.0 (3)	4.9 (5)	5.0 (5)	7.3 (7)	2.3 (2)		2.5 (2)	1.1 (1)
Phe									
Lys	1.0 (1)	0.9 (1)							0.9 (1)
Trp	ND	ND	ND	ND	ND	ND	ND (1)	ND	ND
total	9	15	21	21	28	6	6	9	10
yield (%)	33	14	45	14	17	11	18	42	47

[a] See the legend of Table I for details.

The activity of the inhibitor is critically dependent on the presence of thiol in the medium (Roth, 1962; Gribnau et al., 1970; Blackburn et al., 1977), and modification of cysteine residues has been found to release active RNase from the enzyme–inhibitor complex (Roth, 1956, 1958; Shortman, 1961, 1962a,b). The distribution of cysteinyl residues appears to be nonrandom. Of the 30 half-cystine residues, 25 occur at only four positions of the repeating units (positions 1, 9, 19, and 26). At present, it is not clear whether these residues occur as free thiols or whether disulfide bridges are present as well. It should be noted, however, that disulfide bridges are rare in intracellular proteins (Schulz & Schirmer, 1979).

Leucine-rich repeats have been found in a number of other proteins: leucine-rich glycoprotein, a protein from human plasma with an unknown function (Takahashi et al., 1985); adenylate cyclase from yeast, a membrane-bound enzyme that converts ATP into cAMP (Kataoka et al., 1985); glycoprotein 1bα, a membrane glycoprotein from human platelets involved in the interaction with von Willebrand factor and platelet adhesion to the subendothelium (Titani et al., 1987; Lopez et al., 1987); chaoptin from *Drosophila*, an integral membrane protein involved in cellular adhesion between R-cells in the compound eye (Reinke et al., 1988); *Toll*, a putatively membrane-bound protein from *Drosophila* that plays a role in embryonic differentiation (Hashimoto et al., 1988). Interestingly, both the type A and the type B repeat of the RNase inhibitor fit the consensus sequences found in these proteins. The RNase inhibitor seems to be unique in that it is the first cytoplasmic protein with such a repeat and that it contains two clearly distinct, alternating repeats. It has been noted before (Takahashi et al., 1985) that the constant apolar residues are interspersed by residues that are often polar, resulting in repeats with an amphiphilic character. In none of the proteins shown in Figure 3B, however, is it clear what exactly the spatial structure of this repeat is. In contrast to other proteins containing amphiphilic structures (e.g., apolipoproteins and mellitin; Eisenberg, 1984), the residues in the repeats in Figure 3B do not cluster when projected onto an α-helical wheel. Since adenylate cyclase and chaoptin do not contain a classical hydrophobic membrane spanning segment for anchoring (Kataoka et al., 1985; Reinke et al., 1988), it has been proposed that in these proteins the amphiphilic leucine-rich repeats interact with the membrane. However, a stretch of hydrophobic amino acids that could anchor the protein to membranes does occur in both glycoprotein 1bα and *Toll*, so that membrane binding does not necessarily take place through the leucine-rich repeat in these proteins. Studies by Roth (1956) have shown that little, if any, RNase inhibitor activity is associated with membrane fractions. Thus, it appears likely that the leucine-rich repeat can form structures that have other functions, e.g., protein binding. Adenylate cyclase (at least in vertebrates), glycoprotein 1bα, and RNase inhibitor are all known to interact with other proteins. Conceivably, the conserved amino acid residues in the leucine-rich repeats of these proteins could form the basis for a common scaffold, whereas the residues in between determine the specificity of the protein–protein interaction. In this context it is interesting to note that the RNase inhibitor binds RNase with a 1:1 stoichiometry, which suggests that the RNase binding site does not correspond to a single leucine-rich repeat but rather that it is formed by a structure resulting from the collection of repeats in the protein.

The availability of sufficient quantities of a soluble protein, in the form of the RNase inhibitor, opens the possibility of studing the three-dimensional structure of leucine-rich repeats by X-ray crystallography and other biophysical techniques.

Added in Proof

Since submitting this paper, we have isolated and sequenced a cDNA clone from LLC-PK_1 porcine kidney cells that codes for residues 71–456 of the ribonuclease inhibitor. The deduced amino sequence was in complete agreement with the one presented in this paper.

Acknowledgments

We thank F. Raschdorf (Ciba-Geigy Ltd., Basel) for performing the mass spectroscopic analysis, Dr. J.-Y. Chang (Ciba-Geigy Ltd., Basel) for performing the quantitative N-terminal analyses, and Drs. Ch. Nager and E. Shaw for carefully reading the manuscript.

Registry No. RNase inhibitor (pig liver), 116926-06-8; nuclease inhibitor, 39369-21-6; ribonuclease inhibitor (pig liver-44-Ser reduced), 116926-07-9.

References

Aoki, Y., & Natori, S. (1981) *Biochem. J. 196*, 699–703.
Beintema, J. J. (1987) *Life Chem. 4*, 333–389.

Bidlingmeyer, B. A., Cohen, S. A., & Taruin, T. L. (1984) *J. Chromatogr. 336*, 93–104.
Blackburn, P. (1979) *J. Biol. Chem. 254*, 12484–12487.
Blackburn, P., & Jailkhali, B. L. (1979) *J. Biol. Chem. 254*, 12488–12493.
Blackburn, P., & Gavilanes, J. G. (1980) *J. Biol. Chem. 255*, 10959–10965.
Blackburn, P., & Gavilanes, J. G. (1982) *J. Biol. Chem. 257*, 316–321.
Blackburn, P., & Moore, S. (1982) *Enzymes (3rd Ed.) 15*, 317–433.
Blackburn, P., Wilson, G., & Moore, S. (1977) *J. Biol. Chem. 252*, 5904–5910.
Bornstein, P. & Balian, G. (1977) *Methods Enzymol. 47*, 132–145.
Bradford, M. M. (1976) *Anal. Biochem. 72*, 248–254.
Brockdorff, N. A., & Knowler, J. T. (1987) *Eur. J. Biochem. 163*, 89–95.
Burton, L. E., Blackburn, P., & Moore, S. (1980) *Int. J. Pept. Protein Res. 16*, 359–364.
Burton, L. E., & Fucci, N. P. (1982) *Int. J. Pept. Protein Res. 18*, 372–379.
Chang, J.-Y. (1983a) *Methods Enzymol. 91*, 79–84.
Chang, J.-Y. (1983b) *Methods Enzymol. 91*, 455–456.
Crestfield, A. M., Moore, S., & Stein, W. H. (1963) *J. Biol. Chem. 238*, 622–627.
Deutscher, M. P. (1988) *Trends Biochem. Sci. (Pers. Ed.) 13*, 136–139.
Dijkstra, J., Touw, J., Halsema, I., Gruber, M., & AB, G. (1978) *Biochim. Biophys. Acta 521*, 363–373.
Eisenberg, D. (1984) *Annu. Rev. Biochem. 53*, 595–623.
Fett, J. W., Strydom, D. J., Lobb, R. R., Alderman, E. M., Bethune, J. L., Riordan, J. F., & Vallee, B. L. (1985) *Biochemistry 24*, 5480–5486.
Flinta, C., Persson, B., Joernvall, H., & von Heyne, G. (1986) *Eur. J. Biochem. 154*, 193–196.
Fontana, A., & Gross, E. (1986) in *Practical Protein Chemistry* (Darbre, A., Ed.) pp 67–120, Wiley, Chicester, U.K.
Gribnau, A. A. M., Schoenmakers, J. G. G., van Kraaikamp, M., Hilak, M., & Bloemendal, H. (1970) *Biochim. Biophys. Acta 224*, 55–62.
Gross, E., & Witkop, B. (1962) *J. Biol. Chem. 237*, 1856–1860.
Hashimoto, C., Hudson, K. L., & Anderson, K. A. (1988) *Cell (Cambridge, Mass.) 52*, 269–279.
Heinrikson, R. L., & Meredith, S. C. (1984) *Anal. Biochem. 136*, 65–74.
Hofsteenge, J., Weyer, W. J., Jekel, P., & Beintema, J. J. (1983) *Eur. J. Biochem. 133*, 91–108.
Kataoka, T., Broek, D., & Wigler, M. (1985) *Cell (Cambridge, Mass.) 43*, 493–505.
Kraft, N., & Shortman, K. (1970) *Aust. J. Biol. Sci. 23*, 175–184.
Kraft, N., Shortman, K., & Jamieson, D. (1969) *Radiat. Res. 39*, 655–668.
Kurachi, K., Davie, E. W., Strydom, D. J., Riordan, J. F., & Vallee, B. L. (1985) *Biochemistry 24*, 5494–5499.
Laemmli, U. K. (1970) *Nature (London) 227*, 680–685.
Lipman, D. J., & Pearson, W. R. (1985) *Science (Washington, D.C.) 227*, 1435–1441.
Lopez, J. A., Chung, D. W., Fujikawa, K., Hagen, F. S., Papayamopoulou, T., & Roth, G. J. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 5615–5619.
Reinke, R., Krantz, D. E., Yen, D., & Zipursky, S. L. (1988) *Cell (Cambridge, Mass.) 52*, 291–301.
Roth, J. S. (1956) *Biochim. Biophys. Acta 21*, 34–43.
Roth, J. S. (1958) *J. Biol. Chem. 231*, 1085–1095.
Roth, J. S. (1962) *Biochim. Biophys. Acta 61*, 903–915.
Roth, J. S. (1967) *Methods Cancer Res. 3*, 151–243.
Schulz, G. E., & Schirmer, R. H. (1979) *Principles of Protein Structure*, Springer-Verlag, Heidelberg.
Shapiro, R., & Vallee, B. L. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 2238–2241.
Shapiro, R., Riordan, J. F., & Vallee, B. L. (1986) *Biochemistry 25*, 3527–3532.
Shortman, K. (1961) *Biochim. Biophys. Acta 51*, 37–49.
Shortman, K. (1962a) *Biochim. Biophys. Acta 55*, 88–96.
Shortman, K. (1962b) *Biochim. Biophys. Acta 61*, 50–55.
Spackman, D. H., Stein, W. H., & Moore, S. (1958) *Anal. Chem. 30*, 1190–1206.
Takahashi, N., Takahashi, Y., & Putnam, F. W. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1906–1910.
Titani, K., Takio, K., Handa, M., & Ruggeri, Z. M. (1987) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5610–5614.
Vereijken, J. M., Hofsteenge, J., Bak, H. J., & Beintema, J. J. (1980) *Eur. J. Biochem. 113*, 151–157.
Weng, L., Russel, J., & Heinrikson, R. L. (1978) *J. Biol. Chem. 253*, 8093–8101.
Wierenga, R. K., Huizinga, J. D., Gaastra, W., Welling, G. W., & Beintema, J. J. (1973) *FEBS Lett. 31*, 181–185.