

The Primary Structure of the Major Parvalbumin from Hake Muscle

Overlapping Peptides Obtained with Chemical and Enzymatic Methods

The Complete Amino-Acid Sequence

Jean-Paul CAPONY, Lars RYDÉN, Jacques DEMAILLE, and Jean-François PECHÈRE

Département de Biochimie Macromoléculaire du Centre National de la Recherche Scientifique, Montpellier

(Received August 11, 1972)

Work on the amino acid sequence of the major parvalbumin from hake muscle has been pursued with the aim of obtaining large fragments and overlapping peptides permitting the tryptic peptides sequenced previously to be ordered and the structure of the N-terminal portion of the molecule to be established.

Cyanogen-bromide cleavage at the single methionine and tryptic digestion limited to the single arginine after maleylation were found useful in this respect and, with the help of the previous data, allowed the last 44 residues of the molecule to be placed unambiguously.

Additional information was obtained from a chymotryptic digest of the protein which resulted in almost complete elucidation of the sequence of the first 24 residues and valuable knowledge on two other fragments, 12 and 29-residues long, respectively.

Application of both limited and extensive peptic-digestion, the first one supplemented with the diagonal technique for cysteine peptides, permitted the remaining gaps to be filled, yielding an unique sequence of 108 residues now in full agreement with the amino acid composition of the protein.

The location of the three amide groups of this parvalbumin is described and completes the description of its primary structure.

The preceding paper [1] described the isolation and the characterization of the tryptic peptides derived from the major parvalbumin from hake muscle. These almost completely accounted for the whole molecule, but yielded only very little information as to their order. To establish this, and to cope with the impossibility of sequencing some of these peptides, a variety of chemical and other enzymatic methods was used, which will now be described. Their application was first oriented towards the obtainment of large fragments allowing extensive overlaps to be recognized. Subsequently, in order to meet the deficiencies of this method, a more pragmatic approach was adopted which permitted the establishment of the complete amino-acid sequence of the protein.

Abbreviation. Dansyl, 1-dimethylamino-naphthalene-5-sulfonyl.

Enzymes. Aminopeptidase or amino-acyl-oligopeptide hydrolase (EC 3.4.1.2); carboxypeptidase A or peptidyl-L-amino-acid hydrolase (EC 3.4.2.1); carboxypeptidase B or peptidyl-L-lysine hydrolase (EC 3.4.2.2); chymotrypsin (EC 3.4.4.5); trypsin (EC 3.4.4.4); pepsin (EC 3.4.4.1); subtilisin or subtilopeptidase A (EC 3.4.4.16).

MATERIALS AND METHODS

Chemicals

Chymotrypsin (α -chymotrypsin, 45 U/mg) was from Worthington Biochemical Corp., pepsin (from hog stomach, $3 \times$ cryst.) from Mann Research Laboratories, aminopeptidase M from Carl Roth OHG, and pronase (B grade) from Calbiochem.

DEAE-cellulose (Whatman DE-32, 1.0 mequiv./g) was from W. and R. Balston Ltd.

Cyanogen bromide was a product from Aldrich Chemical Cy., Inc., and maleic anhydride from British Drug Houses Ltd.

Benzene (Merck), ethylene chloride (Merck) and ethylacetate (Prolabo) were redistilled before use in the Edman degradations. Standard phenylthiohydantion derivatives of the amino acids were obtained from Sigma Chemical Cy.

Silicagel for thin-layer chromatography of these derivatives was a product from Camag-AG, type DF-5. The solvents used for the separations were all redistilled.

All other chemicals were either as described previously [1] or of the highest purity available.

CNBr Cleavage

This was accomplished at room temperature over 30 h as described by Gross [2] after dissolution (1 ml/ μ mol) of the parvalbumin in 70% (v/v) aqueous formic acid and using a 150-fold molar excess of CNBr. The peptides generated were fractionated on a column (1.2 $\times$ 115 cm²) of Sephadex G-25 in 0.2 M acetic acid. The effluent was monitored both by ultraviolet absorption at 220 nm and by ninhydrin analysis after alkaline hydrolysis [3]. The fractions of interest were pooled and lyophilized after neutralization with ammonia.

Limited Tryptic Digestion after Maleylation

The parvalbumin was dissolved (3 ml/ μ mol) in 0.2 M sodium bicarbonate, 6 M guanidinium hydrochloride, and additions were made, in the pH-stat (Radiometer, type TTT I) of portions (6/ μ mol protein) of 0.050 ml maleic anhydride [4] dissolved in dioxane (1 g/ml), the pH being maintained at 9 with 4 M KOH. The reaction mixture was then desalted on a column (2 $\times$ 40 cm²) of Sephadex G-25 in 0.05 M ammonium acetate pH 9.5 and the protein peak was lyophilized. Ninhydrin and dansyl chloride tests indicated that the substitution had occurred with 95–100% yield. Cellulose-acetate electrophoresis at pH 8.9 (0.09 M barbiturate buffer) showed the presence of a single component having a higher negative mobility than the original parvalbumin, as expected.

The maleylated parvalbumin was dissolved (1 ml/10 mg) in 0.2 M *N*-ethylmorpholine buffer pH 8.0, 5 mM CaCl₂, and digested at 40 °C over 4 h with 2.5% (w/w) trypsin. At the end of the reaction, a 50-nmol aliquot was removed and subjected to analysis with carboxypeptidase B in order to ascertain the specificity and the extent of the cleavage.

The fragments resulting from the digestion were separated at 3 °C on a DEAE-cellulose column (2.5 $\times$ 30 cm²) equilibrated with 0.1 M HCl adjusted to pH 7.5 with Tris and developed with a gradient of increasing chloride concentration at pH 7.5 [5]. The fractions of interest were pooled, lyophilized, desalted on Sephadex G-25 in 0.1 M ammonia and lyophilized.

Chymotryptic Digestion and Separation of the Resulting Peptides

Chymotryptic digestion were performed on 10-mg/ml solutions of the protein after its denaturation in guanidine hydrochloride (incubation at room temperature for 5 h in 0.01 M phosphate pH 8, 6 M guanidine hydrochloride, followed by a rapid desalting on a Sephadex G-25 column (4 $\times$ 15 cm²) in 0.05 M ammonium acetate pH 9.5). The digestions

occurred in 0.1 M *N*-ethylmorpholine–acetic acid buffer pH 8.0, for 8 h at room temperature, with 5% (w/w) enzyme dissolved (5 mg/ml) in 1 mM HCl.

At the end of the digestion, the mixture was lyophilized, and then redissolved with the help of trifluoroacetic acid (0.1 ml/ μ mol digested protein) and separated on chromobeads P, as described [1] for the tryptic digest of the performic-acid-oxidized protein, except that the gradient was spread over 31 h. Further separations on paper, either electrophoresis at pH 6.3 or chromatography, were made as described previously [1].

Diagonal Separation of Cysteic-Acid Peptides

The procedure of Brown and Hartley [6] was applied after digestion of the parvalbumin with pepsin for 7 h under the conditions described by Ambler [7].

Peptic Digestion and Isolation of the N-Terminal End

The parvalbumin was digested for 69 h at room temperature with 2.5% (w/w) pepsin in water (10 mg/ml), after its dissolution (0.75 ml/ μ mol) in formic acid–acetic acid–water (1:4:45, v/v/v) [7]. The mixture was then applied directly to the top of a chromobeads P column (0.6 $\times$ 95 cm²) equilibrated with 0.05 M pyridine adjusted to pH 2.5 with acetic acid. Elution was effected with this same buffer only.

Sequence Methods

Both the direct Edman method as described by Iwanaga *et al.* [8] and the indirect Edman-dansyl procedure [9] were used. Amino acid phenylthiohydantoin were identified by thin-layer chromatography on silica-gel plates using the following solvent system: dichloroethane–acetic acid (90:9, v/v) [10], chloroform–methanol (90:7, v/v) (D. Eaker, personal communication), and heptane–propionic acid–dichloroethane (58:17:25, v/v/v) [11], and complementing them with the ninhydrin colour test [12]. Dansyl-amino acids were identified as described previously [1].

Carboxypeptidase A and B digestions followed the procedure indicated in the preceding paper [1].

Amide Determinations

The possible presence of asparagine and glutamine residues in the tryptic peptides described in the preceding paper [1] and in the chymotryptic peptides isolated during the present study was established by studying their electrophoretic mobility on paper at pH 6.3 as suggested by Offord [13]. Within the peptides thus selected, the identification

of amidated residues was first attempted by examining the result of their digestion with pronase and aminopeptidase as described by Bennett *et al.* [14]. When this did not yield unambiguous results, use was made of appropriate Edman degradations which permitted the amidated residues to be identified either directly under the form of the corresponding phenylthiohydantoin derivative liberated or indirectly by a study of the electrophoretic mobility, at pH 6.3, of an aliquot of the residual peptide after each cycle. In some cases, the arrest of the carboxypeptidase-A digestion at acidic residues was used to confirm the previous inferences.

Peptide Nomenclature and Yield-Calculations

Peptides designated with a symbol starting with a C represent chymotryptic peptides and those starting with a P refer to peptic peptides. The figure and occasional small letter following this indicate a fraction from the chromobeads P column and a further fractionation by paper methods, respectively. The cyanogen bromide peptide is labelled CNBr, the fragment obtained after the arginine-limited tryptic digestion is designated by FrI, and the diagonal peptide by di.

RESULTS

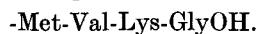
CNBr Cleavage

Fig. 1 depicts the result of the fractionation of the peptide mixture resulting from the cleavage of 0.6 μ mol major parvalbumin from hake muscle with cyanogen bromide acting at the level of its single methionine. The head peak contained unreacted protein and the fragment terminating with homoserine lactone. On the other hand, analysis of fractions 37–42 indicated that a single peptide was present, as expected. Its composition was found to be: Lys: 0.91; Gly: 1.12; Val: 0.94 (in no. of residues), and it appeared in amounts corresponding to a 50% reaction yield.

Its sequence was determined by the Edman-dansyl procedure (2 steps) which afforded



Examination of the location of the four valine residues of the molecule in the tryptic peptides sequenced earlier [1] conclusively excluded the possibility that the tripeptide above might have resulted from a partial acid hydrolysis of the chain at the level of some of its aspartyl-residues as a result of the reaction conditions [15]. It therefore must have arisen from a cleavage at the level of the single methionine residue, a situation which implies that the following sequence is present in the protein:



This result further confirms the position of glycine at the C-terminus of the chain [16, 1]. Furthermore,

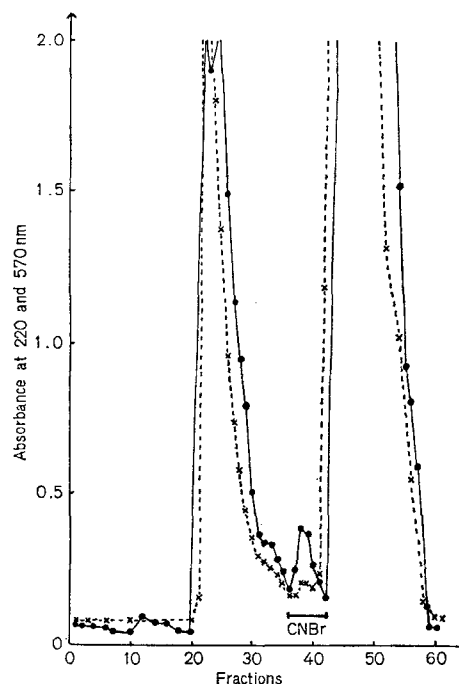
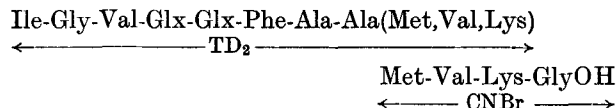


Fig. 1. Fractionation on Sephadex G-25 of the reaction medium resulting from treatment of 0.6 μ mol parvalbumin with CNBr. The column ($1.2 \times 115 \text{ cm}^2$) was run at room temperature. Elution occurred at 14.5 ml/h with 0.2 M acetic acid. Fractions of 2.7 ml were collected and analyzed by their ultraviolet absorption at 220 nm ($\times$ --- $\times$) and by reaction of an aliquot (1/6) with ninhydrin after alkaline hydrolysis (measured by absorbance at 570 nm) ($\bullet$ — $\bullet$).

it allows the unsequenced amino acids of peptide TD₂ [1] to be ordered:



as well as this peptide to be positioned close to the C-end of the molecule.

Large Fragments Resulting from Limited Tryptic Digestion

When 1.3 μ mol major parvalbumin from hake muscle was maleylated and subsequently digested with trypsin as described above (see Methods), analysis of the reaction mixture with carboxypeptidase B indicated that splitting had occurred only at the single arginine-X bond of the molecule, and in 90% yield.

Fractionation of the bulk of the reaction mixture on DEAE-cellulose resulted in the diagram depicted in Fig. 2. Examination of the contents of the first of the two major peaks (fractions 46–51, FrI) provided evidence for the presence of a single large peptide in satisfactory pure form and having the

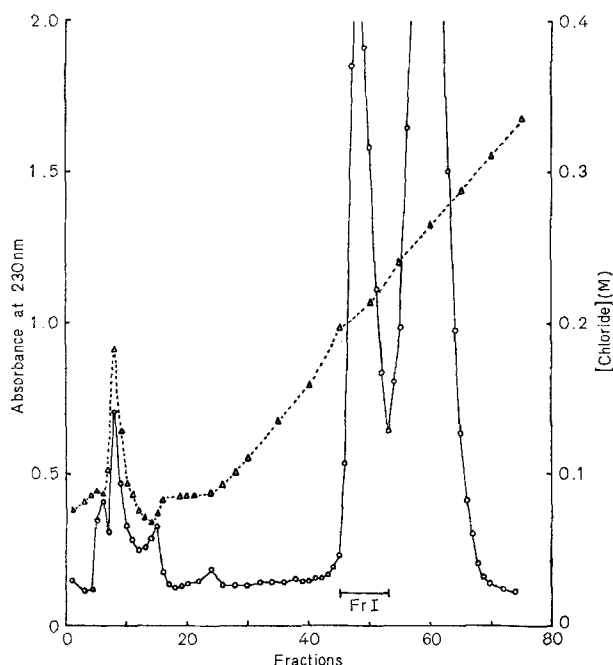
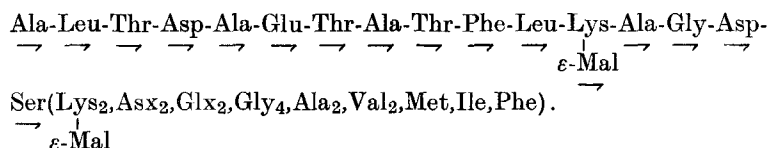


Fig. 2. Fractionation on DEAE-cellulose of the tryptic digest of 1.3 μmol maleylated parvalbumin. The column ($2.5 \times 30 \text{ cm}^2$) was run in the cold room (3°C). Elution occurred at 20 ml/h starting with 0.1 M HCl adjusted to pH 7.5 with Tris and then using a gradient reaching 0.6 M chloride in 12 h by addition of NaCl. Fractions of 2.8 ml were collected and analyzed by their ultraviolet absorption at 230 nm (O—O) and, intermittently, their chloride content, (Δ --- Δ) was determined by conductivity measurements

following amino-acid composition (in no. of residues): Lys: 3.15; Asx: 4.56; Thr: 2.93; Ser: 1.33; Glx: 3.22; Gly: 4.96; Ala: 5.73; Val: 1.81; Met: 0.63; Ile: 1.08; Leu: 2.20; Phe: 2.28. The presence of methionine as well as the absence of arginine, clearly showed that this 33-residue fragment represented the expected C-terminal part of the molecule. Its final yield amounted to 72%. On the other hand, the amino-acid analysis and the carboxypeptidase-B digestion of the second major peak (fractions 52–70) indicated that it contained all the N-terminal part of the molecule, terminating with arginine, and about 10% of the original maleylated protein, undigested. In spite of numerous attempts, these two entities could not be separated from each other.

The sequential degradation of Edman was applied directly on 0.425 μmol of maleylated FrI, with the following result:



The fact that the three lysine residues of FrI were unavailable for coupling with phenylisothiocyanate and the presence of several acidic residues, readily explains that the degradation occurred without important losses up to the 16th step where the yield of the serine derivative still amounted to 13%. At step 12, in fact, two derivatives were detected: a first one corresponding to a small amount of α -lysine derivative in the aqueous phase, and another major one in the organic phase with chromatographic behaviour in chloroform–methanol (90:7, v/v) very similar to that of the glutamate derivative, probably corresponding to α -phenylthiohydantoin, ϵ -maleyl-lysine.

The first twelve residues of this long sequence match exactly with the sequence of peptide TD₁ [1], while the next four clearly belong to peptide TF [1]. If the five other residues of peptide TF are subtracted from the unsequenced part, it is seen that the remaining peptide fits very well with, and only with, the sum of the composition of peptides TD₂ and TB [1], a situation which places these two peptides at the C-end of the molecule, in agreement with the preceding results (CNBr cleavage). As FrI must be preceded by arginine, its partial sequence thus establishes the order of the last 34 amino acids of the protein. When the overlaps described earlier [1] for peptides TA_{1b}, T₃ and TA_{2a} are also taken into account, the number of C-terminal ordered residues is raised to 44.

The success of the preceding strategy suggested that considerable progress might similarly be achieved in the ordering of the tryptic peptides [1] if the parvalbumin chain could be split specifically, by the above technique, in three pieces rather than in two, taking advantage of the presence of the single cysteine residue which, after aminoethylation, could also be made the place of a tryptic attack. The following sequence of operations was thus performed on 1.3 μmol parvalbumin: *S*-sulfonation [1], maleylation (see Methods), reduction of the *S*-sulfonate group [17], aminoethylation [18] and tryptic digestion under the conditions above. Carboxypeptidase-B analysis of an aliquot of the reaction mixture again indicated that, the protein had been split at its single arginyl-bond in 90% yield; in addition, aminoethyl-cysteine was liberated in amounts corresponding to 75% tryptic splitting at this new position also. From these results, a total of 6 different peptides, including three predominant ones, could be expected to be present in the final reaction mixture. However, only FrI,

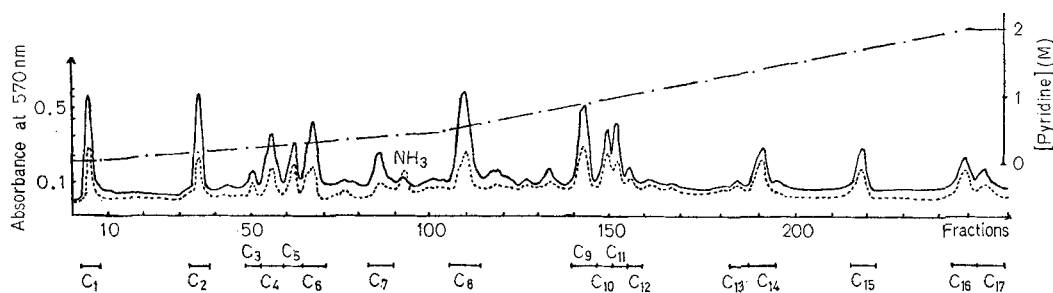


Fig. 3. Separation, on chromobeads-P cation exchanger, of the peptides resulting from the chymotryptic digestion of 2 μ mol guanidine-hydrochloride-denatured parvalbumin. The column (0.6×95 cm²) was thermostated at 50 °C. Elution occurred at 30 ml/h, starting with 0.05 M pyridine adjusted to (—) alkaline hydrolysis

pH 2.5 with acetic acid for 2 h and then using a gradient of increasing pyridine concentration according to the profile shown (---). Fractions of 3.25 ml were collected, representing 90% of the effluent. The remaining 10% was used for automatic analysis with ninhydrin, before (----) and after

corresponding to the 33-residue C-terminal peptide described above, could be isolated in pure form using a wide variety of either DEAE-cellulose or Sephadex columns. This approach had thus to be abandoned.

Chymotryptic Peptides

In the hope, therefore, to obtain the missing overlaps for ordering the tryptic peptides [1] the major parvalbumin from hake muscle was subjected to chymotryptic digestion. Fig. 3 shows the separation, on chromobeads P, of the soluble peptides obtained after digestion of 4 μ mol denatured protein. Peaks C₂, C₃, C₄, C₆, C₁₅ and C₁₇ were found to correspond to pure peptides in sufficient amounts to be studied further. Peaks C₁₁ and C₁₂ also appeared to be pure but were recovered only in trace quantities and were thus not examined further. The remaining major peaks were found to be complex, and their contents were thus subjected to fractionation by paper methods. In this way, pure peptides C_{5a} and C_{5b}, C_{7c} and C_{7e}, C_{8b}, C_{9b} and C_{9c}, C_{10e}, C_{13a}, C_{14f} and C_{16a} could be isolated in sizable quantities. C₁, which should contain the unretarded ninhydrin-negative blocked N-terminal fragment, was found to be contaminated with salt (resulting probably from an incorrect regeneration of the Chromobeads P column); it was therefore passed on a Sephadex G-10 column whereby it resolved into peptides C_{1a} and C_{1b} (Fig. 4). These results are summarized in Table 1, which also gives the amino-acid composition and the yields of the peptides recovered.

Sequence determination on these peptides (Table 2), using the Edman-dansyl procedure and carboxypeptidase A, generally did not present difficulties, in particular because a limited application of these techniques was sufficient, in many cases, either to locate the peptide in those parts of the total sequence which had already been ordered,

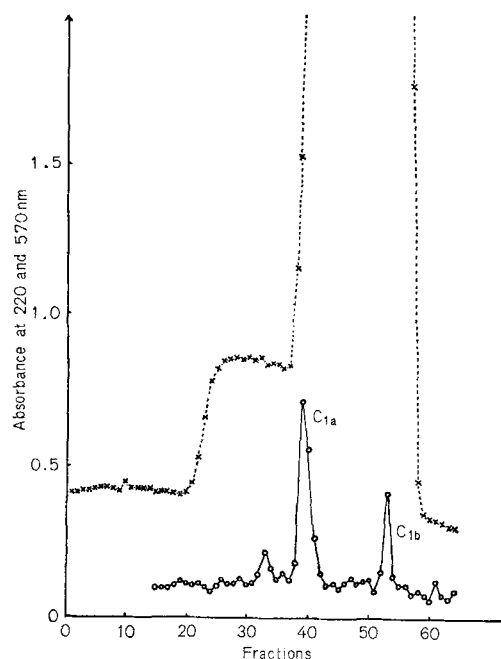


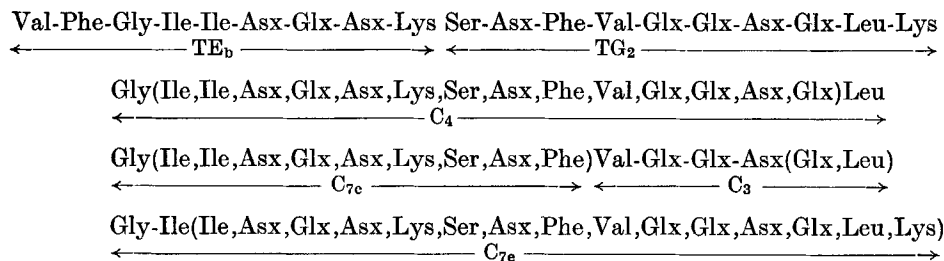
Fig. 4. Desalting of peak C₁ from the chromobeads column on Sephadex G-10. The column (0.9×140 cm²) was operated at room temperature. Elution occurred at 7.5 ml/h with 0.1 M ammonia. Fractions of 2.5 ml were collected and analyzed by their ultraviolet absorption at 220 nm (x----x) and by reaction of an aliquot (1/20) with ninhydrin after alkaline hydrolysis (o—o)

or to established unambiguously new overlaps between tryptic peptides. Such is the case for peptides C₂, C_{5a}, C_{8b}, C_{9b} and C_{10e} which are easily placed in the 44-residue C-terminal fragment described above. Also, it is clear that peptide C₄, with its eight acidic residues and its valine, and taking into account, as for peptide TE_b [1], a low recovery of isoleucine during amino acid analysis, can only

Table 1. *Amino-acid composition and final yields of chymotryptic peptides obtained from 4 μ mol parvalbumin*

Peptides	Amino-acid composition														Yield
	Lys	His	Arg	Asx	Thr	Ser	Glx	Gly	Ala	Val	Met	Ile	Leu	Phe	
	residues/mol														%
C _{1a}				1.97	0.91			1.20	4.74			2.00	2.10		18.3
C _{1b}														1.00	10.3
C ₂				0.96	2.64		1.04		1.92					1.12	38.0
C ₃				1.38			3.26			0.94			0.94		3.6
C ₄	1.00			3.62		1.28	3.79	1.25		1.12		1.00	1.21	1.05	5.5
C _{5a}									2.08		0.66				1.4
C _{5b}								1.01	0.96			0.69	0.80		2.5
C ₆									2.06				0.94		9.5
C _{7a}															traces
C _{7b}															traces
C _{7c}	0.70			3.03		0.74	1.13	1.39				0.95		1.03	0.9
C _{7d}															traces
C _{7e}	1.03			3.64		0.94	3.55	0.89		1.00		0.75	0.86	1.08	1.2
C _{7f}															traces
C _{8a}															traces
C _{8b}	2.09			2.70		0.96	2.05	3.95	1.17	1.00		0.91		0.96	12.8
C _{8c}															traces
C _{9a}															traces
C _{9b}			0.96			0.54		1.09	3.04				1.00		6.4
C _{9c}	0.96				0.73			1.00				0.87	1.00		12.5
C _{10a}															traces
C _{10b}															traces
C _{10c}															traces
C _{10d}															traces
C _{10e}	0.81							1.00		1.06					9.5
C ₁₁															traces
C ₁₂															traces
C _{13a}	1.00									0.94				1.00	9.2
C _{13b}															traces
C _{13c}															traces
C _{13d}															traces
C _{14a}															traces
C _{14b}															traces
C _{14c}															traces
C _{14d}															traces
C _{14e}															traces
C _{14f}	0.91												1.03	0.87	4.1
C ₁₅	0.94	1.06					0.94	1.12						1.00	20.8
C _{16a}	1.30	1.00					0.79	1.13						2.00	14.0
C _{16b}															traces
C ₁₇	3.82			0.97		0.95		1.05	2.00	0.87		0.99		1.12	

overlap peptides TE_b and TG₂, a situation which is further documented by the nature of its end-groups which comply well with the usual specificity of chymotrypsin and permits the order of the two tryptic peptides to be established:



Peptides C₃, C_{7c} and C_{7e} clearly originated, as shown, from the same region. Similarly, peptide C₁₇ provides an overlap which is only compatible with peptides TA₃, TC₁ and TA₁, in that order, while its C-terminal sequence indicates that these three peptides are followed in the molecule by peptide TE_b:

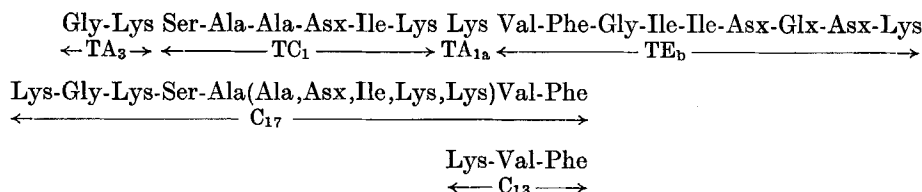
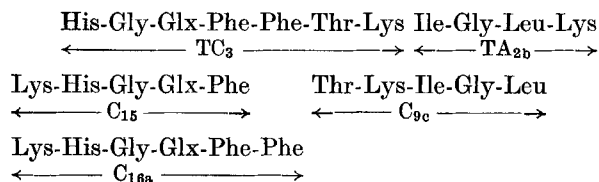


Table 2. *Sequence of the major chymotryptic-peptides obtained from the denatured parvalbumin*
 →, Edman-dansyl degradation; ←, carboxypeptidase digestion

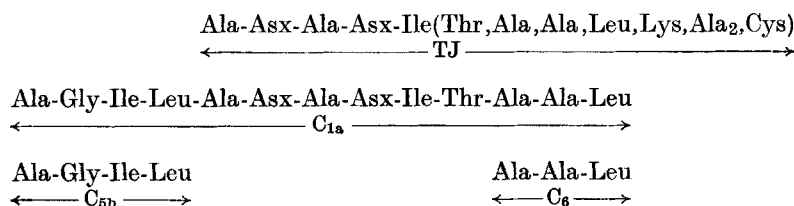
Peptide	Sequence
C _{1a}	Ala-Gly-Ile-Leu-Ala-Asx-Ala-Asx-Ile-Thr-Ala-Ala-Leu → → → → → → → → ← ← ←
C _{1b}	X-Phe ←
C ₂	Thr(Thr ₂ ,Asx,Glx,Ala ₂)Phe ←
C ₃	Val-Glx-Glx-Asx(Glx,Leu) → → →
C ₄	Gly(Lys,Asx ₄ ,Ser,Glx ₄ ,Gly,Val,Ile ₂ ,Phe)Leu → → → → → → → ←
C _{5a}	Ala-Ala-Met → → →
C _{5b}	Ala-Gly-Ile-Leu → → → →
C ₆	Ala-Ala-Leu → → →
C _{7c}	Gly(Lys,Asx ₃ ,Ser,Glx,Ile ₂ ,Phe) → → → → → →
C _{8b}	(Lys ₂ ,Asx ₃ ,Ser,Glx ₂ ,Gly ₄ ,Ala,Val,Ile)Phe → → → → → → ←
C _{9b}	Ser-Ala-Gly-Ala-Arg-Ala-Leu → → → → →
C _{9c}	Thr-Lys-Ile-Gly-Leu → → → → ←
C _{10c}	Val-Lys-Gly → → →
C ₁₃	Lys-Val-Phe → → →
C _{14r}	Lys-Leu-Phe → → →
C ₁₅	Lys-His-Gly-Glx-Phe → → → → ←
C _{16a}	Lys-His-Gly-Glx-Phe-Phe → → → → ← ←
C ₁₇	Lys-Gly-Lys-Ser-Ala(Lys ₂ ,Asx,Ala,Ile)Val-Phe → → → → → → ← ←

The sequence of peptide C₁₃ is in good agreement with these conclusions, so that the order TA₃-TC₁-TA_{1a}-TE_b-TG₂ becomes evident when all these results are taken together; however, there are no data permitting the location of this 29-residue fragment in the molecule. Finally, the sequence of peptide C_{9c} clearly shows the ordering of TC₃ and TA_{2b}, while peptides C₁₅ and C_{16a} are evidently part of this same locus:



Here again, there are no arguments to place this 12-residue fragment in the molecule. It is striking, however, that if the sum of the composition of the three large ordered fragments which have already been described (44-residue, 29-residue and 12-residue) is subtracted from the composition of the whole protein, a difference is obtained which, except again for one alanine, is identical with the composition of peptide TK, for which there is some evidence [1] that it is N-terminal. This would indicate that the 12 and 29-residue fragments are adjacent, and are located just before the 44-residue one. Unfortunately, the sequence of peptide C_{14r}, which in all probability extends from the N-terminus of the 44-residue fragment, is of no value for deciding which of the two comes last.

On the other hand, the remaining chymotryptic peptides C_{1a}, C_{1b}, C_{5b} and C₆ permit a partial sequence for the N-terminal part of the molecule to be derived. Indeed, the sequence of residues 5–9 of peptide C_{1b} is identical with the unique N-terminal sequence of peptide-TJ [1], and its remaining composition is exactly compatible with the portion of the unsequenced part of the tryptic peptide which would result from a chymotryptic attack:



The unique sequences of C_{5b} and C₆ also fit easily in place within this scheme.

As to peptide C_{1b}, special comment is needed. This peptide did not stain with ninhydrin on paper, and neither did it stain with chlorine [19]. This

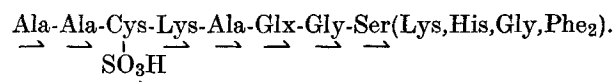
second observation is unexpected, but the first one would be understandable, since it does not contain lysine, if C_{1b} represented a short peptide originating from the blocked N-terminus of the molecule. Its very early appearance on the cation-exchange chromatogram (Fig. 3) is in line with such a situation, and the finding that it did not react with dansyl chloride (as indicated by X in Table 2) seems to confirm such an assignment. Carboxypeptidase A liberated 0.62 equivalent of phenylalanine, but no other amino acid, in agreement with its amino acid analysis. All these observations therefore suggest that indeed C_{1b} should be placed next to, and just before C_{1a}, a situation which furthermore exactly fits with the amino-acid composition found earlier [1] for peptides TK and T₁.

The first guess concerning the nature of the blocking-group X would be for an acetyl group. This is indeed the case for many muscle proteins [20] and, in particular, Gerday and Rao [21] have shown that it is also true for several parvalbumins. Furthermore, direct evidence had been obtained at that time [22] for the presence of at least one covalently bound acetyl group in the major parvalbumin from hake muscle. However, the electrophoretic mobility on paper of peptide C_{1b} at pH 6.3 (which could be ascertained using the staining procedure of Kirchenmayer and Kuffner [23]) was found to be distinctly different from that of an authentic sample of acetylphenylalanine, and also from that of formylphenylalanine. In addition, the rapidity with which phenylalanine had been liberated during the carboxypeptidase-A digestion of C_{1b} was hardly compatible with this peptide being only acetylphenylalanine. Unfortunately no more peptide C_{1b} was available after these experiments. As it appeared, in addition, that a completely satisfactory purification of that peptide might be difficult to achieve, it was decided to resort to other ways (see below) in order to clarify the ambiguities found at the N-terminus of the molecule.

Sequence Around the Single Cysteine and Complete Ordering of Tryptic Peptides

After application of the preceding techniques, moreover, a small sequence, around the single cysteine residue, as well as the order of the two

central fragments (12-residue and 29-residue) were still left undetermined. It was fortunate that these two questions were resolved simultaneously. Indeed, the diagonal procedure of Brown and Hartley [6] for isolation of cysteine peptides, when used on 0.5 μ mol pepsin-digested parvalbumin, gave rise to a single peptide migrating off the diagonal, with the following composition (in no. of residues): Lys: 1.96; His: 0.65; CysSO₃H: 0.93; Ser: 1.08; Glx: 2.12; Gly: 2.16; Ala: 3.10; Phe: 1.97, and which was recovered in 28% final yield. Its sequence could be determined, up to the eighth residue, by the Edman-dansyl technique:

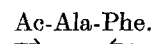


The sequence around the cysteine residue is thus completed. On the other hand, the nine last residues of this peptide clearly overlap with peptides TC₂ and TC₃, indicating that the N-terminal tryptic peptide T₁ is followed by TC₂ and TC₃ in that order; this situation, in turn, unequivocally shows that the 12-residue fragment, which contains the single histidine, precedes the 29-residue fragment. In fact, a sequence of 107 amino acid is thus derived in the protein [24] which, if the amidation of three acidic residues is momentarily disregarded, fully accounts for the amino-acid composition of the whole protein [16], except for one alanine residue; and confirming that of the two possible values (12—13) for the number of aspartic/asparagine residues, the lower value is the correct one.

Further Examination of the N-Terminal End

A critical examination of all the information gained up to this point did not suggest that the missing alanine residue could have escaped detection in the sequence of 107 amino acids derived. The inference, therefore, was made that this resulted from an error in the amino acid analysis of the N-terminal tryptic peptides [1] and that there had been something wrong with the characterization of peptide C_{1b}, as already suspected above. In the hope of obtaining a clearer insight as to the nature of the N-terminus of the molecule, the following techniques were then applied: (a) detection of formyl and acetyl groups in the whole protein by the technique of Schmer and Kreil [25]; (b) application of the method of Gray [26] on both the whole protein and peptide T₁, using in each case either subtilisin or thermolysin as digesting enzyme. However, none of those techniques yielded any useful information. The possibility of obtaining peptide C_{1b}, or an analogue of it, in a satisfactory condition was therefore reconsidered, and, in view of the success encountered during the diagonal operation described above, a long digestion (69 h) of the parvalbumin was

performed with pepsin; the resulting mixture was subsequently examined on a column of chromobeads P (see Methods). The front peak was found to consist of a clean and single peptide, designated P₁, with the following composition (in no. of residues): Ala: 1.00; Phe: 1.00 and recovered in 75% yield. It did not react with dansyl chloride, and liberated 0.73 equivalents of phenylalanine upon carboxypeptidase-A digestion. Taking the precautions recommended by Laurent *et al.* [27], hydrazinolysis and dansylation [25] now yielded 0.98 equivalents of easily characterizable acetyl hydrazide, indicating that the structure of P₁ was:



This result was fully confirmed by a study of the mass spectrum of peptide P₁ after permethylation and its comparison with that obtained under the same conditions with an authentic sample of synthetic Ac-Ala-Phe.

The nature of the N-terminal blocking group was thereby clearly established, while the results of the sequence analysis became fully concordant with those of the amino-acid analysis.

Location of Amide Groups

The analyses performed on the whole protein [16] had indicated that three out of its 23 acidic residues were amidated. A study, by the method of Offord [28], of the electrophoretic mobility of all the tryptic and chymotryptic peptides described here and before [1] showed that amides could be present in the peptides considered in Table 3; all the other ones, indeed, displayed a mobility agreeing very well with expectation. The anomalous mobility of TC₃ being readily explained by the presence of histidine in this peptide, the following residues were left as possible amidated candidates (see Fig. 5 for the numbering): one out of either Asx-51, Glx-52 or Asx-53, one of either Glx-59, Glx-60, Asx-61 or Glx-62 and one of either Glx-68 or Asx-69.

Peptides TE_b, TG₂ and TA_{1b} were thus subjected to aminopeptidase digestion according to Bennett *et al.* [14]. With TA_{1b}, the result was

Table 3. Anomalous mobility of peptides derived from the major parvalbumin from hake muscle

Peptide	Charge expected if no amide present	Charge found
TA _{1b}	-1	+1/2 to 1
TC ₃	+1/2	+1/2
TE _b	-2	-1
TG ₂	-4	-3
C ₃	-4	-3
C ₄	-7	-5
C _{7c}	-3	-2
C _{7c}	-6	-4

very clear: glutamic acid and asparagine were liberated in the expected amounts with respect to the other amino acids of the peptide. With TE_b and TG₂, the degradation did not proceed adequately, probably because of the accumulation of acidic residues which are known to be unfavorable to the action of exopeptidases.

The mobility of those two peptides was thus examined around the critical steps of their sequential degradation with phenylisothiocyanate. Thus, the study of the mobility of peptide TE_b after 5, 6 and 7 cycles of degradation clearly indicated Glx-52 to be glutamine. The results were less clear with TG₂; however, they suggested that Asx-61 was the amidated residue.

Edman degradation (with direct identification of the liberated amino acids derivatives) of peptides TG₂ and C₃ confirmed that Asx-56, Glx-59 and Glx-60 were in the acid form, while attack of the last peptide with carboxypeptidase A did not yield anything beyond the C-terminal leucine, indicating that in all probability Glx-62 was in the acid form too, a situation which supports the probable assignment of the amide to Asx-61. Carboxypeptidases A and B acting together on peptide TE_b similarly only liberated lysine, confirming also that Asx-53 is in the acid form.

Thus the three expected amides were located at positions 52, 61 and 69 [24]. Subsequently, kind information from Dr R. A. Bradshaw, who had independently been studying the primary structure of one of the major parvalbumins (pI 4.25) from carp muscle [29], raised doubts about the correctness of the two last assignments, as his sequence (also 108-residues long) included an acidic residue at the position corresponding to our Asx-61 and a very labile amide at the position corresponding to our Glx-68. The state of amidation at these two positions in the hake sequence was thus reinvestigated.

The method used successfully in St. Louis (carboxypeptidase B digestion at pH 7.5 for 10 min followed by carboxypeptidase A digestion at pH 5.2 for 60 h) was applied to about 0.1 μ mol of the hake peptides TG₂ and TA_{1b}. In the first case, 1.15 equivalents aspartic acid and 3.24 equivalents glutamic acid were recovered in the digest; if, as expected, the N-terminal dipeptide seryl-aspartate is indeed not split by carboxypeptidase A, this strongly suggests that Asx-61 is aspartic acid and not asparagine. In the second case, however, the results were inconclusive, probably because the stretch which separates Glx-68 and Asx-69 from the C-end in TA_{1b} includes glycine which is hardly removed by carboxypeptidase-A; in addition, the presence of glutamic acid in the digest of TA_{1b} might not reflect the original situation in the protein if a deamidation at position 68 is liable to take place, for instance during the column procedure (50 °C, pH 2.5) which led to the isolation of the peptide.

About 1 μ mol hake parvalbumin was therefore subjected to chymotryptic-peptide mapping at pH 6.3 under the conditions described earlier [30], except that the denaturation of the protein occurred with guanidine hydrochloride as described above (see Methods) and that the passage through Dowex-50 was thus avoided. The map clearly showed the presence of peptide 7 which is known to correspond to the sequence Leu-Glx-Asx-Phe and is equivalent to part of the non-specific tryptic-peptide T₃ (Demaille and Pechère, unpublished results). The position of this peptide on the map unequivocally demonstrates that it is neutral and hydrophobic, which is consistent only with both Glx and Asx being amides. If there are merely three amidated groups in the molecule, it necessarily follows that Asx-61 indeed is aspartic acid, not asparagine.

It is thus concluded that the three amide residues are Gln-52, Gln-68 and Asn-69.

DISCUSSION

All the data which have been described earlier [16,1] and above are compatible with, and only with the primary structure presented in Fig. 5. The number of overlaps and cross checks make it very probable that this indeed represents the primary structure of the major parvalbumin from hake muscle. The most convincing proof for its correctness, however, stems from the structural work of Nockolds *et al.* [29]. The present primary structure, indeed, is fully compatible with the tertiary structure deduced by these authors for one of the homologous carp parvalbumins (component pI 4.25).

Most of the sequence determination was carried out without major difficulties. Three phases of the work, nevertheless, required special attention. The first one refers to the very labile and hydrophobic region including residues 65–70. Trypsin, even minimally contaminated with chymotryptic activity, always split the bond between Phe-70 and Ser-71 bond very rapidly (to a smaller extent, this was also found to be the case for the bonds between Phe-24 and Lys-25 and between Phe-85 and Leu-86). Acidity also disrupted this region in several ways (deamidation, splitting of asparaginyl and other bonds). Both these situations made it rather hard to obtain adequate overlaps in this region. Furthermore, when the appropriate fragments might have been produced, their hydrophobic character often impeded their isolation and purification.

A second region which was found difficult to characterize was the N-terminal one. Here again its hydrophobicity contributed largely to the vicissitudes, and the strategy adopted initially for the separation of the tryptic fragments, while useful for other matters, was unfortunate in this respect. This could be resolved by working with the per-

formic-acid-oxidized derivative of the protein. At this point, however, the blockage at the N-terminus introduced further complications, in particular because it was often associated with the unexpected lability of one alanine residue. It is striking that, when the whole protein was acid-hydrolyzed, the correct number of alanine residues (19) was liberated in 24 h [16], while the acid hydrolysis of peptides T_J, T_{K1}, T_I and C_{1b} was always associated with the loss of one alanine residue. The fact that the composition of peptides C_{1b} and P₁ was correctly deduced makes it difficult to rationalize the process, unless it is admitted that this is due to some impurity being present occasionally. It is noteworthy that, as can be judged from the initial data of Kretsinger *et al.* [31], a similar impediment seems to have been encountered in the determination of the structure of the carp parvalbumin.

Finally, some pitfalls were associated with the location of the three amide residues of the protein. This was due essentially to the extreme lability of the amide of Glx-68, which was not suspected until the data for the apparently less unfavourable case of the carp parvalbumin were available. It is also true that the exact state of Asx-61 had only been inferred, and not determined directly. The explanation for the former, erroneous assignment probably lies in the fact that the mobility of peptides TG₂ and C₃, because of the four adjacent acidic groups, is inferior to what would be expected on the basis of the Offord equation [27]. It is quite conceivable, however, that once two of the acidic groups have been removed, the peptides resume a normal behaviour, a situation which gives the false impression that, at the last degradation step, a neutral residue has been removed.

Most of these difficulties actually stemmed from the fact that nothing was known about parvalbumin sequences when this work was started. It can reasonably be hoped, therefore, now that several of their characteristics have been outlined, that future sequences can be established with confidence in a more efficient and rapid way.

The authors are grateful to Mr E. Audemard and to Mr R. Bertrand for their help with the amino-acid analyses and the column chromatography of peptides, respectively. They wish to thank Dr B. C. Das, in Prof. Lederer's laboratory (Institut de Chimie des Substances Naturelles, CNRS, Gif-sur-Yvette) for his kind assistance with the mass spectroscopic experiments which were so helpful in confirming the nature of the N-terminal blocking group. They also wish to acknowledge the friendly cooperation of Dr R. A. Bradshaw (Dept. of Biological Chemistry, Washington University School of Medicine, St. Louis) particularly in regard to the location of amide groups. A short stay in his laboratory to clarify these matters was made possible during part of an EMBO short-term fellowship granted to one of us (J.-F. P.); this generosity is gratefully acknowledged. Gratitude for the interest of Dr E. Zuckerkandl is also expressed. This investigation was supported in part by contract no. 6801207 with the *Délégation Générale à la Recherche Scientifique et Technique* and is published as contribution no. 78 from the Department.

REFERENCES

1. Capony, J.-P. & Pechère, J.-F. (1972) *Eur. J. Biochem.* **32**, 88–96.
2. Gross, E. (1967) *Methods Enzymol.* **11**, 238–255.
3. Fruchter, R. G. & Crestfield, A. M. (1965) *J. Biol. Chem.* **240**, 3868–3874.
4. Butler, P. J. G., Harris, J. I., Hartley, B. S. & Lebermann, R. (1969) *Biochem. J.* **112**, 679–689.
5. Davidson, B. E., Sajgo, M., Noller, H. F. & Harris, J. I. (1967) *Nature (London)* **216**, 1181–1185.
6. Brown, J. R. & Hartley, B. S. (1966) *Biochem. J.* **101**, 214–228.
7. Ambler, R. P. (1963) *Biochem. J.* **89**, 349–378.
8. Iwanaga, S., Wallén, P., Gröndahl, N. J., Henschen, A. & Blombäck, B. (1969) *Eur. J. Biochem.* **8**, 189–199.
9. Gray, W. R. (1967) *Methods Enzymol.* **11**, 469–475.
10. Edman, P. (1970) in *Protein Sequence Determination* (Needleman, S. B., ed.) pp. 211–255, Springer-Verlag, Berlin-Heidelberg-New York.
11. Jeppson, J. O. & Sjöquist, J. (1967) *Anal. Biochem.* **18**, 264–269.
12. Roseau, G. & Pantel, P. (1969) *J. Chromatogr.* **44**, 392–395.
13. Offord, R. E. (1966) *Nature (London)* **211**, 591–593.
14. Bennett, C., Konigsberg, W. H. & Edelman, G. M. (1970) *Biochemistry*, **9**, 3181–3188.
15. Piszkievicz, D., Landon, M. & Smith, E. L. (1970) *Biochem. Biophys. Res. Commun.* **40**, 1173–1178.
16. Pechère, J.-F., Capony, J.-P. & Rydén, L. (1971) *Eur. J. Biochem.* **23**, 421–428.
17. Chan, W. W. C. (1968) *Biochemistry*, **7**, 4247–4254.
18. Cole, R. D. (1967) *Methods Enzymol.* **11**, 315–317.
19. Nitecki, D. E. & Goodman, J. W. (1966) *Biochemistry*, **5**, 665–672.
20. Fischer, E. H., Cohen, P., Fosset, M., Muir, L. W. & Saari, J. C., in *Metabolic Interconversion of Enzymes* (Wieland, O., Helmreich, E. & Holzer, H., eds.) Springer-Verlag, Berlin-Heidelberg-New York, in press.
21. Gerday, C. & Bushana Rao, K. S. P. (1970) *Comp. Biochem. Physiol.* **36**, 229–240.
22. Parello, J., Cavé, A., Puigdomenech, P., Capony, J.-P. & Pechère, J.-F., *Biochim.*, in press.
23. Kirchenmayer, T. & Kuffner, F. (1962) *Monatsh. Chem.* **93**, 1237–1241.
24. Pechère, J.-F., Capony, J.-P., Rydén, L. & Demaille, J. (1971) *Biochem. Biophys. Res. Commun.* **43**, 1106–1111.
25. Schmer, G. & Kreil, G. (1969) *Anal. Biochem.* **29**, 186–192.
26. Gray, W. R. & Del Valle, M. E. (1970) *Biochem. J.* **9**, 2134–2137.
27. Laurent, G., Marriq, D., Garçon, D., Luccioni, F. & Derrien, Y. (1967) *Bull. Soc. Chim. Biol.* **49**, 1035–1058.
28. Offord, R. E. (1966) *Nature (London)* **211**, 591–593.
29. Nockolds, C. E., Kretsinger, R. H., Coffee, C. J. & Bradshaw, R. A. (1972) *Proc. Nat. Acad. Sci. U. S. A.* **69**, 581–584.
30. Pechère, J.-F. & Capony, J.-P. (1969) *Comp. Biochem. Physiol.* **28**, 1089–1102.
31. Kretsinger, R. H., Nockolds, E., Coffee, C. J. & Bradshaw, R. A. (1971) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 217–220.

J.-P. Capony and J.-F. Pechère

Département de Biochimie Macromoléculaire du C.N.R.S.
Boîte postale 5051, F-34033 Montpellier CEDEX, France

L. Rydén's present address:
Biokemiska Institutionen, Uppsala Universitet
Box 531, S-751 21 Uppsala 1, Sweden

J. Demaille's present address:
Laboratoire de Biochimie Médicale de l'Université
Boîte postale 5067, Dakar-Fann, Sénégal