225. Chemical Structure Studies of Cow $\varkappa$-Casein: Study of the Soluble Tryptic Peptides[1])

by **Jacqueline Jollès, Charles Alais** and **Pierre Jollès**

Laboratory of Biochemistry, Faculty of Sciences, 96 Bd. Raspail, Paris 6e, and Faculty of Sciences, Nancy, France

(6. VII. 70)

Dedicated to Dr. *O. Isler* on his 60th birthday

Summary. The sequences of 13 tryptic peptides of cow $\varkappa_A$-casein (accounting for about one half of the amino acid residues present in the protein) were established. The rennin sensitive linkage could be located in a large fragment (36 residues). $\varkappa$-casein consists of a hydrophilic part ($\varkappa$-caseino-glycopeptide) and of a hydrophobic moiety (para-$\varkappa$-casein); in this latter, however, several quite hydrophilic sequences were characterized. Another feature of the $\varkappa$-casein structure is the frequent duplication or triplication of certain amino acids (Pro-Pro; Phe-Phe; Gln-Gln-Gln-Asn-Glu-Glu-Glu; Pro-Pro-Lys-Lys-Asn-Gln-; etc. ...).

1. Introduction. – Among the different fractions which constitute the whole casein, $\varkappa$-casein has a special position. It plays a major role in the stabilization of the casein micelle in its natural environment; furthermore, the primary enzymic phase of the action of rennin (EC 3.4.4.3) on casein has been shown to be confined to this $\varkappa$-fraction [2]. Rennin splits a Phe-Met bond which was first characterized by our group [3] [4]: a large peptide designated $\varkappa$-macropeptide in the absence of sugars and $\varkappa$-caseino-glycopeptide in the presence of sugars is liberated and para-$\varkappa$-casein precipitates. It is now established that the peptide skeleton of $\varkappa$-casein is almost homogeneous; a few changes attributable chiefly to genetic variants and variations in the sugar composition are ultimately responsible for the heterogeneity observed on electrophoresis [3] [5].

$\varkappa$-casein contains 165 ± 5 amino acid residues calculated per monomer of molecular weight 20,000 [6], 2 phosphorous residues and variable amounts of N-acetylgalactosamine, galactose and N-acetylneuraminic acid. Preliminary data concerning a) the amino acid sequence of $\varkappa$-caseino-glycopeptide [7], which constitutes the C-terminal sequence of $\varkappa$-casein, and b) the structure of the tryptic peptide containing the rennin-sensitive linkage [8] [9] have already been published.

In order to establish the complete primary structure of $\varkappa$-casein, a systematic study of the tryptic peptides was undertaken. Trypsin (EC 3.4.4.4) gives rise to an important tryptic 'core' and to soluble peptides. The present paper deals with the establishment of the primary structure of these latter (structures, s. Tables I–VII).

2. Materials and Methods. – 2.1. *Cow $\varkappa$-casein* was prepared according to *McKenzie & Wake* [10] from pooled milk, and $\varkappa_A$ (or $\varkappa_B$) casein by the same procedure from the milk of homozygous cows.

2.2. *Reduction and aminoethylation:* 1 g $\varkappa_A$-casein was dissolved in 50 ml of a 0.1M Tris-HCl buffer of pH 8 containing 8 M urea. 1 ml mercaptoethanol was added and the solution was left for

[1]) 24th communication on caseins; 23rd communication, s. [1].

3 h at 38° under nitrogen; after addition of 2 ml ethylenimine, the solution was again left for 1 h at 20° (under nitrogen). After dialysis against water at 4°, the reduced aminoethylated $\varkappa$-casein was lyophilized.

2.3. *Tryptic hydrolysis:* 700 mg of reduced $\varkappa$-casein were digested during 6 or 24 h with 14 mg trypsin at 37° and pH 7.8 (solution containing 1% collidine and 1% pyridine adjusted to pH 7.8 with acetic acid; or 0.1M ammonium hydrogen carbonate adjusted at pH 7.8). Trypsin (*Worthington*) was pretreated during 16 h with 0.0625N HCl at 37°. The tryptic digest was centrifuged; the precipitate (tryptic 'core') has not yet been submitted to further purifications.

2.4. *Preliminary separation of the soluble tryptic peptides.* The soluble tryptic peptides were chromatographed on a 140 × 2.4 cm column of Dowex 1 X 2 (200–400 mesh) equilibrated with a solution containing 1% pyridine and 1% collidine the pH of which had been adjusted to 8.4 with acetic acid. The elution of the most basic peptides occured at this pH. Gradient elution was then begun by allowing successively 0.5N and 2N acetic acid to flow into the solution of pH 8.4 in a mixing chamber of 1000 ml of capacity; by this procedure, neutral and acidic peptides were eluted. The chromatography was achieved by a direct elution of the resin with 2N acetic acid or 2N HCl; some strongly acidic peptides could only be obtained by this procedure. The peptides were characterized by a ninhydrin determination after alkaline hydrolysis.

2.5. *Purification of the tryptic peptides.* – a) By rechromatography on *Dowex* 50 × 2 of some of the peaks characterized on *Dowex* 1 X 2. *Dowex* 50 × 2 (200–400 mesh) was equilibrated with a 0.2M pyridine solution, the pH of which had been adjusted to 2.2 with acetic acid and HCl. The resin was first eluted with this buffer, then a gradient elution was begun (increase of pH from 3.1 to 5.0; increase of the content of pyridine from 0.2M to 2.0M). Basic peptides were rechromatographed on small columns (15 × 0.9 cm), neutral peptides on larger ones (50–100 × 0.9 cm).

b) By preparative paper chromatography (*Whatman* No. 1) in the solvents A: *n*-butanol-formic acid-water (75:15:10, *v/v*) or B: *n*-butanol-pyridine-acetic acid-water (15:10:3:12, *v/v*).

2.6. *Determination of the amino acid composition and of the structure of the peptides.* The different procedures used have recently been described in this review [11].

3. Results. – 3.1. *Chromatography on Dowex 1 X 2.* After chromatography of the soluble tryptic hydrolysate of $\varkappa_A$-casein on *Dowex* 1 X 2, 22 peaks were characterized (Fig. 1). The figure is identical to that previously obtained with a tryptic digest of $\varkappa$-casein (from pooled milk).

3.2. *Purification and structure of the peptides contained in peaks (1 + 2 + 3) eluted from Dowex 1 X 2 column.* Peaks 1, 2 and 3 characterized after chromatography of the tryptic hydrolysate of cow $\varkappa_A$-casein on a *Dowex* 1X 2 column were pooled and rechromatographed on a *Dowex* 50X 2 column (Fig. 2). 11 peaks (A–K) were obtained; peaks A–H were further purified by paper chromatography in solvent B. In Table I are indicated the Rf and mobility *m* values as well as the structures of all these peptides with the exception of I.

Peptide (1 + 2 + 3) A: the establishment of the structure of this nonapeptide is indicated in Table II.

Peptides (1 + 2 + 3) B and (1 + 2 + 3) G: they constitute the N-terminal sequence of the $\varkappa$-caseinoglycopeptide and have previously been studied in detail [7].

Peak (1 + 2 + 3) C consists of free lysine.

Peptide (1 + 2 + 3) D: the structure of this pentapeptide has been determined by the *Edman* technique; after each step, the new N-terminal amino acid has been characterized after dansylation.

It is worth mentioning that this tryptic peptide has an N-terminal proline residue; usually trypsin does not split an Arg-Pro or Lys-Pro linkage. Peptide (1 + 2 + 3) D

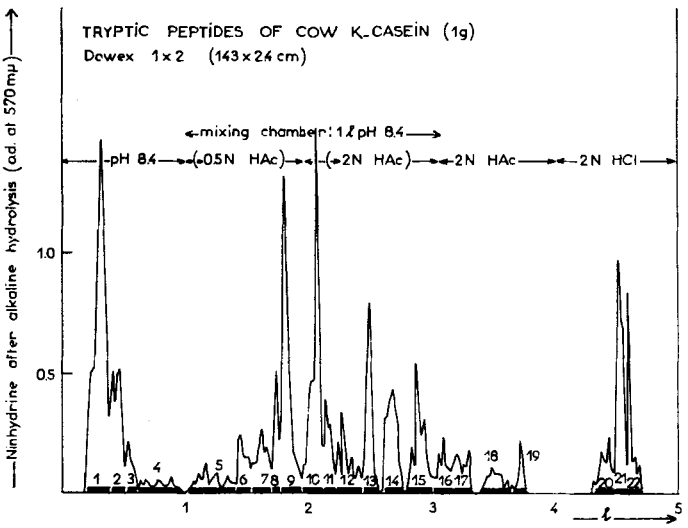


Fig. 1. *Chromatography on* Dowex *1 X 2* (143 × 2.4 cm) *of the soluble tryptic hydrolysate of reduced aminoethylated cow* $\varkappa_A$*-casein*
(1 g protein; 6 h enzymic hydrolysis)

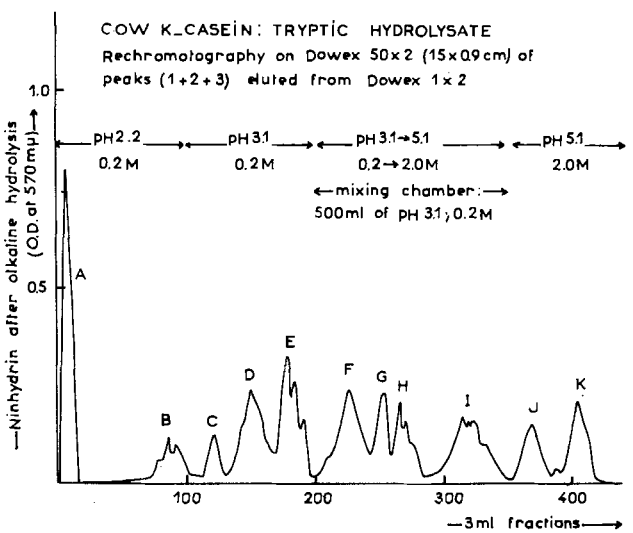


Fig. 2. *Rechromatography on* Dowex *50 X 2* (15 × 0.9 cm) *of the pooled fractions 1 + 2 + 3 obtained after chromatography on* Dowex *1 X 2 of a soluble tryptic hydrolysate of 1 g cow* $\varkappa_A$*-casein* (see Fig. 1)

also gives a positive test with the isatine reagent [12]; this reaction proves again that proline is in a N-terminal position.

Peptide (1 + 2 + 3) E: the peptide was submitted to the action of aminopeptidase; after 10 min only isoleucine was released, whereas equal amounts of the three amino acids appeared after 30 min.

Peptide (1 + 2 + 3) F: this peptide was also digested by aminopeptidase; after 10 min, only valine (20%) could be detected; after 20 min valine (50%) and leucine (40%) were characterized.

Peptide (1 + 2 + 3) H: the establishment of the structure of this pentapeptide is indicated in Table III.

Table I. *Rf values* (in solvents A and B), *mobility m* (at pH 6.5) *and chemical structure of the tryptic peptides (1 + 2 + 3) A–K* (see Figures 1 and 2)

Peptide	*m* [a] (pH 6.5)	Rf(A)	Rf(B)	Structure
A	+0.3	0.02	0.30	Gln-Ala-Gln-Pro-Thr-Thr-Met-Ala-Arg
B	+0.35	0.27	0.53	Met-Ala-Ile-Pro-Pro-Lys
C	+1.1	0.05	0.19	Lys
D	+0.45	0.12	0.36	Pro-Ala-Ala-Val-Arg
E	+0.58	0.12	0.30	Ile-Ala-Lys
F	+0.50	0.16	0.55	Val-Leu-Ser-Arg
G	+0.75	0.05	0.28	Met-Ala-Ile-Pro-Pro-Lys-Lys
H	+0.47	0	0.06	Lys-Asn-Gln-Asp-Lys
J	+0.25	0	0.55	His-Pro-Pro-His-Leu-Ser-Phe-Met-Ala-Ile-Pro-Pro-Lys
K	+0.40	0	0.53	His-Pro-Pro-His-Leu-Ser-Phe-Met-Ala-Ile-Pro-Pro-Lys-Lys

[a] $m = 0$ for Gly; $m = +1$ for Arg.

Table II. *Establishment of the structure of tryptic peptide (1 + 2 + 3) A*
Amino acid composition: Thr_2, Glu_2, Pro_1, Ala_2, Met_1, Arg_1

Method	Isolated substances
Edman	Glx - Ala - Glx - Pro
Aminopeptidase (24h) + dansylation	Gln - Ala
Carboxypeptidases B (24h) + A (1h)	Thr - Met - Ala - Arg (59%) (63%) (71%) (81%)
Chymotryptic digestion (24 h)	(Glx, Ala, Glx, Pro, Thr, Thr) - Met Ala - Arg
Structure	Gln[a]-Ala - Gln[a]-Pro - Thr - Thr - Met - Ala - Arg

[a] As the mobility *m* (at pH 6.5) of the peptide is +0.3, both Glx residues must be Gln residues.

Table III. *Establishment of the structure of tryptic peptide (1 + 2 + 3) H*
Amino acid composition: Asp_2, Glu_1, Lys_2

Method	Isolated substances
Edman	Lys - Asx - Glx - Asx
Aminopeptidase (6 h)	Lys (65%) - Asn (65%)
Carboxypeptidases A + B (24 h)	Asp (29%) - Lys (70%)
Structure	Lys - Asn - Gln[a] - Asp - Lys

[a] As the mobility *m* (at pH 6.5) of the peptide is +0.47, this residue must be amidated.

Peptides (1 + 2 + 3) J and (1 + 2 + 3) K: they represent the tryptic sequence with the rennin-sensitive Phe-Met bond and have been studied in detail in a previous paper [8].

3.3. *Purification and structure of soluble tryptic peptides contained in peaks 4–14 eluted from the Dowex 1 X 2 column.* In Table IV are indicated the Rf and *m* values as well as the structures of different peptides contained in peaks 4–14 eluted from the *Dowex* 1*X* 2 column (Fig. 1).

Table IV. *Rf* (in solvents A and B), *m* (at pH 6.5) *and chemical structure of various tryptic peptides contained in peaks 4–14* (see Fig. 1)

Peptide in peak	*m* [a]) (pH 6.5)	Rf(A)	Rf(B)	Structure
6	+0.80	0.02	0.15	Ser-AECys [b])
8	+0.40	0	0.12	AECys-Glu-Lys
9A	0	0	0	AECys-Glu-Lys-Asp-Glu-Arg
9B	+0.27	0	0.70	His-Pro-Pro-His-Leu-Ser-Phe
10	0	0.27	0.53	Phe-Phe-Ser-Asp-Lys
11A	−0.44	0.02	0.12	Asp-Glu-Arg
11B	−0.30	0	0.04	Glu-Lys-Asp-Glu-Arg
13	−0.46	0	0.15	Gln-Gln-Gln-Asn-Glu-Glu-Glu-Pro-Ile-Arg
14	0	0.70	0.80	Tyr-Ile-Pro-Ile-Tyr-Gln

[a]) $m = 0$ for Gly; $+1$ for Arg; -1 for $CySO_3H$.
[b]) AECys = aminoethylcysteine.

Peptide 6 is the dipeptide Ser-AECys: its structure was determined after dansylation.

Trypsin is thus capable of splitting the peptide chain after an aminoethylcysteine residue.

In the tryptic peptide of oxidized $\varkappa_A$-casein, we characterized the peptide Ser-$CySO_3H$-Gln-Ala-Gln-Pro-Thr-Thr-Met-Ala-Arg (*m* at pH 6.5 = 0; Rf (A) = 0; Rf (B) = 0.10). After digestion with aminopeptidase (24 h) and dansylation, Ser, $CySO_3H$, Gln and Ala were detected; carboxypeptidases A and B (2,5 h) released Arg (100%), Ala (100%), Met (80%) and Thr (70%); after chymotryptic digestion (24 h), the dipeptide Ala-Arg was isolated. Thus it was possible to demonstrate that tryptic peptide (1 + 2 + 3) A follows tryptic peptide 6.

Peptides 8, 9A, 11A and 11B: peptide 9A is composed of peptides 8 + 11A or of AECys + peptide 11B. The tryptic splits after the AECys and Lys residues were only very partial ones, probably on account of the presence of free carboxylic groups. The establisment of the structure of these four peptides is summarized in Table V.

Peptide 9B: the establishment of the structure has previously been described in detail [8].

Peptide 10: amino acid composition: Asp_1, Ser_1, Phe_2, Lys. The N-terminal sequence was determined by the Edman technique (Phe-Phe-Ser) and by the action of aminopeptidase (24 h): Phe (1.95 residues), Ser (0.5). Thus the sequence of the pentapeptide is: Phe-Phe-Ser-Asp-Lys. The presence of Asp is justified by the value of the mobility of the peptide of pH 6.5 which is 0.

Table V. *Establishment of the structure of tryptic peptides 8, 9A, 11A and 11B*

Peptide 9A: a) *Amino acid composition:* Asp_1, Glu_2, $AECys_1$, Lys_1, Arg_2. b) *Structure*

Method	Isolated substances					
Aminopeptidase (24 h)	AECys					
carboxypeptidase B (24 h)						Arg
Edman:						
a) + dansylation	AECys -	Glx -	Lys -	Asx -	Glx	
b) + analysis of the residual peptide (residues) after: 1st step	0.37	1.0	0.69	0.87	1.0	1.0
2nd step	tr.	0.1	0.40	0.90	1.0	1.0
3rd step	tr.	0.1	tr.	1.0	1.0	1.0
4th step	tr.	tr.	tr.	0.45	0.90	1.0
Structure:	AECys -	Glx -	Lys -	Asx -	Glx -	Arg
Mobilities at pH 6.5 of peptides 8, 9A, 11A, 11B	← 9A; m[a]) = 0 →					
		← 11B; m[a]) = −0.30 →				
	← 8; m[a]) = +0.40 →					
				← 11A; m[a]) = −0.44 →		
Complete structure of peptide 9A	AECys -	Glu -	Lys -	Asp -	Glu -	Arg[b])

[a]) m (at pH 6.5) = 0 (for Gly); +1 (for Arg); +1.1 (for Lys); +1.15 (for AECys); −1 (for $CySO_3H$).

[b]) The complete structures of peptides 8, 11A and 11B were established independently by similar methods.

Table VI. *Establishment of the structure of tryptic peptide 13*

Amino acid composition: Asp_1, Glu_6, Pro_1, Ile_1, Arg_1

Method	Isolated substances								
Aminopeptidase (24 h)	No reaction								
Carboxypeptidase B (24 h)									Arg (85%)
Partial hydrolysis (24 h; 100°; HCl 0.03 N)							Pro -	Ile -	Arg
Edman:									
a) + dansylation	Glx -	Glx -	Glx -	Asx -	Glx -	Glx -	Glx -	Pro -	Ile
b) mobility[a]) at pH 6.5 of the residual peptides	→	→	→	→	→	→	→	→	→
before: 1st step	←						m = −0.46		→
after: 1st step		←					no change		→
2nd step			←				in the		→
3rd step				←			mobility		→
4th step					←				→
5th step						←	m = −0.15		→
6th step							← m = 0		→
Structure	Gln -	Gln -	Gln -	Asn -	Glu -	Glu -	Glu -	Pro -	Ile - Arg

[a]) m (at pH 6.5) = −1 (for $CySO_3H$); 0 (for Gly).

Peptide 13: the establishment of the structure of this peptide is summarized in Table VI. The N-terminal Gln residue is gradually transformed into a PCA (pyrrolidone carboxylic acid) residue; the peptide thus becomes insensitive to ninhydrin on paper, and can be submitted to the *Edman* procedure only after a pretreatment (1.0N NaOH; 3 days; 20°).

Peptide 14: the establishment of the structure of this peptide is summarized in Table VII. Peptide 14 is devoid of a basic amino acid residue; its appearance is probably due to the action of some non-inactivated traces of chymotrypsin (EC 3.4.4.5) present in the sample of trypsin. This peptide does not represent the C-terminal sequence of $\varkappa$-casein which is constituted by the $\varkappa$-caseinoglycopeptide [3] [7] [9].

Table VII. *Establishment of the structure of tryptic peptide 14*
Amino acid composition: Glu_1, Pro_1, Ile_2, Tyr_2

Method	Isolated substances											
Aminopeptidase (24 h)	Tyr (100%)	-	Ile (66%)									
Carboxypeptidase (24 h) + dansylation									(Tyr,			Gln)
Edman:												
a) + dansylation	Tyr	-	Ile	-	Pro	-	Ile	-	Tyr	-		Gln
b) + analysis of residual peptide (residues)	→		→		→		→		→			→
after: 2nd step					0.6		0.75		0.45			1.0
3rd step					0.12		0.8		0.45			1.0
4th step					0.15		0.26		0.45			1.0
5th step					tr.		tr.		tr.			1.0
Structure	Tyr	-	Ile	-	Pro	-	Ile	-	Tyr	-		Gln

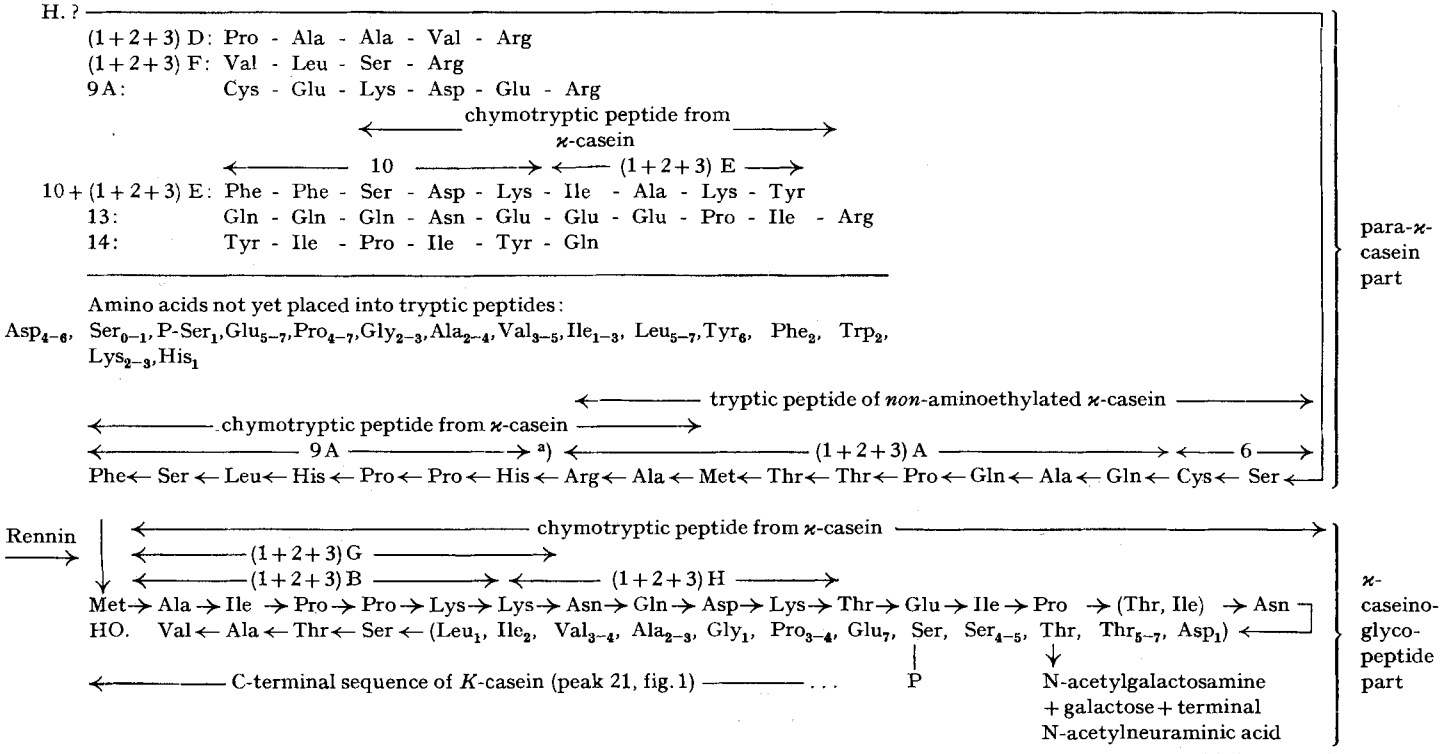
4. Discussion. – 4.1. *Present knowledge of the amino acid sequence of $\varkappa_A$-casein.* Table VIII summarizes our present knowledge of the amino acid sequence of cow $\varkappa_A$-casein. This latter is composed of (1) the para-$\varkappa$-casein part; the structures of 10 tryptic peptides containing 58 amino acids (out of 107 ± 8) have been determined; 2) the $\varkappa$-caseino-glycopeptide part; the structures of the three tryptic peptides with a C-terminal lysine residue have been established and a partial structure of the long C-terminal peptide, devoid of a basic amino acid residue, is indicated; this latter contains a phosphoserine residue and the sugars; an O-glycosidic linkage between a threonine residue and N-acetylgalactosamine has previously been characterized by *Fiat, Alais & Jollès* [13].

4.2. *Lengthening of the sequence of two tryptic peptides.* A first series of results obtained with a chymotryptic digest of $\varkappa_A$-casein allows some of the tryptic peptides to be joined.

$\varkappa_A$-casein contains 5 arginine residues which have been located in the tryptic peptides (1 + 2 + 3) A, D and F; 11A (part of 11B) and 13, the structures of which have been determined. In the chymotryptic digest of cow $\varkappa_A$-casein, we have characterized the peptide:

Ala-Arg-His-Pro-Pro-His-Leu-Ser-Phe

Table VIII. *Present knowledge concerning the amino acid sequence of cow K_A-casein*
(residues per monomer of molecular weight 20000 [6])



a) For the succession of sequences (1+2+3) A+9A, see the Discussion.

containing the C-terminal sequence of para-$\varkappa$-casein which was thus extended by the dipeptide Ala-Arg. This result suggests that the tryptic peptide (1 + 2 + 3) A (Table II) precedes the C-terminal peptide of para-$\varkappa$-casein (Table VIII). Similarly, tryptic peptides 10 and (1 + 2 + 3) E were linked, as in the chymotryptic digest of $\varkappa_A$-casein the overlapping peptide Ser-Asp-Lys-Ile-Ala-Lys-Tyr- was characterized.

Finally the tryptic peptide Ser-AECys (6) was lengthened by peptide (1 + 2 + 3) A, as indicated above (see also Table VIII). Thus the two cysteine residues of $\varkappa_A$-casein could be located in larger tryptic peptides (6 and 9A).

4.3. *Amphiphile nature of $\varkappa_A$-casein and of para-$\varkappa$-casein; duplication or triplication of the same amino acids in the studied sequences. Hill & Wake* [14] suggested that the basis for the micelle stabilizing property of cow $\varkappa$-casein was its amphiphile nature, for the N-terminal two thirds of the molecule (para-$\varkappa$-casein part) were hydrophobic and the C-terminal third hydrophilic ($\varkappa$-caseino-glycopeptide part with high proportions of Ser, Thr, Glu, Gln; presence of sugars; absence of aromatic amino acids). The statements of these authors were chiefly based on results of our group published some time ago [3] [7]. The data presented in this paper allow one to specify that if quite hydrophobic sequences occur in the para-$\varkappa$-casein part (Tyr-Ile-Pro-Ile-Tyr; Pro-Ala-Ala-Val; etc.), this latter also contains large hydrophilic sequences (example peptide 13, Table VI).

Another unusual feature is the frequent duplication or triplication of the same or very similar amino acids in the studied sequences: particularly Gln-Gln-Gln-Asn-Glu-Glu-Glu (in peptide 13), Pro-Pro (twice around the rennin-sensitive linkage), Pro-Pro-Lys-Lys-Asn-Gln, Ala-Ala, Phe-Phe, Thr-Thr.

The chemical structure studies reported in this research were carried out by *J. Jollès* and *P. Jollès*, the purification of $\varkappa_A$-casein by *C. Alais*.

The authors wish to express their appreciation to Misses *Barbara Hiltpold* and *Françoise Schoentgen* for skilful technical assistance.

BIBLIOGRAPHY

[1] *A. M. Fiat, C. Alais & J. Jollès*, Chimia *24*, 220 (1970).
[2] *D. F. Waugh & P. H. von Hippel*, J. Amer. chem. Soc. *78*, 4576 (1956).
[3] *P. Jollès*, Angew. Chem. intern. Ed. *5*, 558 (1966).
[4] *A. Delfour, J. Jollès, C. Alais & P. Jollès*, Biochem. biophysic. Res. Commun. *19*, 452 (1965).
[5] *A. G. Mackinlay & R. G. Wake*, Biochim. biophysic. Acta *104*, 167 (1965); *J. H. Woychik, E. B. Kalan & M. E. Noelken*, Biochemistry, *5*, 2276 (1966); *D. G. Schmidt, P. Both & J. P. de Koning*, J. Dairy Sci. *49*, 776 (1966).
[6] *H. A. McKenzie*, Adv. Prot. Chemistry *22*, 122 (1967).
[7] *A. Delfour, C. Alais & P. Jollès*, Chimia *20*, 148 (1966).
[8] *J. Jollès, C. Alais & P. Jollès*, Biochim. biophysic. Acta *168*, 591 (1968).
[9] *J. Jollès, P. Jollès & C. Alais*, Nature *222*, 668 (1969).
[10] *H. A. McKenzie & R. G. Wake*, Biochim. biophysic. Acta *47*, 240 (1961).
[11] *J. Jollès & P. Jollès*, Helv. *52*, 2671 (1969).
[12] *R. Acher, C. Fromageot & M. Jutisz*, Biochim. biophysic. Acta *5*, 81 (1950).
[13] *A. M. Fiat, C. Alais & P. Jollès*, Chimia *22*, 137 (1968).
[14] *R. J. Hill & R. G. Wake*, Nature *221*, 635 (1969).