Preparation and amino acid sequence of human κ-casein

Ghislaine Brignon, Abdessatar Chtourou and Bruno Ribadeau-Dumas*

Institut National de la Recherche Agronomique, CNRZ, 78350 Jouy-en-Josas, France

Received 14 May 1985; revised version received 11 June 1985

Human κ-casein was prepared from whole casein by successive hydroxyapatite and thiol-Sepharose chromatographies. The primary structure of its 99-residue N-terminal fragment has been determined by sequencing peptides obtained by tryptic and chymotryptic digestions of the whole protein. This fragment overlaps the known sequence of the 65-residue C-terminal fragment. The 158-residue sequence of human κ-casein was compared to those of goat, ewe, cow and rat κ-caseins. Only 22% of the residues are identical in homologous positions. The rate of divergence of the 93-residue N-terminal segment (para-κ-casein) appears to be higher than that of the rest of the molecule.

κ-Casein *Primary structure* *Protein evolution* *Human*

1. INTRODUCTION

In cow milk a large proportion of the proteins occurs as a suspension of micelles containing 4 different casein species (α_{s1}, α_{s2}, β, $\varkappa$), water and minerals. Chymosin (EC 3.4.23.4) and pepsin trigger milk coagulation in the stomach of suckling calves by splitting $\varkappa$-casein (169 residues) into 2 parts: para-$\varkappa$-casein, the 105-residue N-terminal fragment, and caseinomacropeptide (CMP), the C-terminal end [1].

Mature human milk contains micelles [2] but β-casein is by far the major human casein. Its amino acid sequence has been published [3]. Some ancient data suggested the occurrence of a $\varkappa$-casein in human milk. A definite proof of its existence was obtained in our group by Chobert et al. in 1976 [4] when they isolated a human CMP, determined almost completely its amino acid sequence and observed clear homologies with CMPs of different species, including cow [5]. The sequence was completed later by Fiat et al. [6]. Nevertheless, numerous attempts at isolating human $\varkappa$-casein failed until Yamauchi et al. in 1981 [7] succeeded in obtaining a homogeneous preparation. Their results will be discussed with ours. This report presents the method we have used for preparing human $\varkappa$-casein and the amino acid sequence of para-$\varkappa$-casein, thus giving the complete primary structure of the protein.

2. MATERIALS AND METHODS

2.1. *Preparation of human $\varkappa$-casein*

Pooled mature milk was used for preparing whole human casein by acid precipitation. This casein was chromatographed on hydroxyapatite (Biogel HTP, Biorad, Richmond, CA), as in [8] with the following modifications: 2 mM K-phosphate, 100 mM KCl, 6.6 M urea, pH 5.9, instead of 5, 205, 4.5 and 6.9, respectively. The first fraction, which contained $\varkappa$-casein, was dialyzed, freeze-dried and chromatographed on thiol-Sepharose 4B (Pharmacia, Uppsala) as in [9]. The fraction which was retained and eluted with cysteine was pure $\varkappa$-casein. For sequence determination 2.5 mg $\varkappa$-casein were reduced and alkylated with iodoacetate.

2.2. *Gel electrophoresis*

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was performed as in [10] on 15%

* To whom correspondence should be addressed

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

acrylamide gels. Isoelectric focusing (IEF) was performed as in [11].

2.3. *Amino acid analysis*

Samples (1 nmol) of alkylated $\varkappa$-casein were hydrolyzed with 5.7 N HCl in evacuated and sealed tubes for 24, 48 and 96 h at 110°C. For peptides (1–2 nmol), hydrolysis times were 24 h and, when necessary, 96 h. Analyses were performed with an LC5000 amino acid analyzer (Biotronik, Münich). The program used separated glucosamine and galactosamine from the other amino acids. For the determination of these amino sugars, alkylated $\varkappa$-casein was hydrolyzed with 4 N HCl at 100°C for 6 h. The eventual occurrence of tryptophan was assessed on cellulose thin layer using Ehrlich's reagent [12]. Carboxypeptidase A (CPA) and carboxypeptidase B (CPB) digests were analyzed as such after vacuum drying.

2.4. *Enzymic digestions*

$\varkappa$-Casein was digested with chymosin (Sigma, St. Louis, MO) at 30°C in 50 mM Na citrate buffer, pH 6.2 (E/S = 1:10000, molar ratio) for various time periods. Trypsin and chymotrypsin (Sigma, TPCK- and TLCK-treated respectively) were employed for digesting $\varkappa$-casein and some peptides, in 0.1 M ammonium bicarbonate, pH 8.0, at 40°C, with E/S = 1:50 and 1:100, digestion times 4 and 5 h, respectively. A few peptides were digested either with proline-specific endopeptidase (Seikagaku, Kogyo, Tokyo) in 0.2 M *N*-ethylmorpholine acetate, pH 7.0, at 30°C for 4 h (E/S = 1:20), or with *S. aureus* V8 protease (Sigma) in 50 mM ammonium acetate, pH 4.0, at 37°C for 16 h (E/S = 1:30). $\varkappa$-Casein and its N-terminal chymotryptic peptide were treated with pyroglutamate aminopeptidase (Boehringer, Mannheim, FRG) according to Podell and Abraham [13]. Digestions of peptides with CPA or CPB (Worthington, Freehold, NJ; PMSF treated) were carried out in 0.1 M ammonium bicarbonate, pH 8.0, at 37°C for various time periods (E/S = 1:50). When acidic residues occurred near the C-terminus, CPA was used in 0.1 M pyridine acetate buffer, pH 5.4.

2.5. *Fractionation of peptide mixtures*

All separations were achieved by reverse-phase HPLC. The equipment consisted of a 720 System Controller, two 6000 A pumps, 1 U6K injector, 1 μm Bondapak C18 column (Waters, Milford, MA) and a CE 2023 variable wavelength UV monitor (Cecil, Cambridge, England). Most fractionations were carried out using the following solvent system I: A, 0.115% trifluoroacetic acid (TFA); B, 0.1% TFA in 60% CH_3CN. Some purifications were achieved with system II: A, 25 mM ammonium acetate, pH 6.0; B, 40% 50 mM ammonium acetate, pH 6.0, 60% CH_3CN. All separations were carried out at 40°C with linear gradients, at a flow rate of 1 ml/min. The absorbance was recorded at 220 nm. The manually collected fractions were dried under vacuum.

2.6. *Edman degradation*

The manual 'partitioning method for small peptides' described by Tarr [14] was used. For the identification of PTH amino acids the procedure of Bhown et al. [15] was employed with the following modifications: column C18 Novapak (Waters,

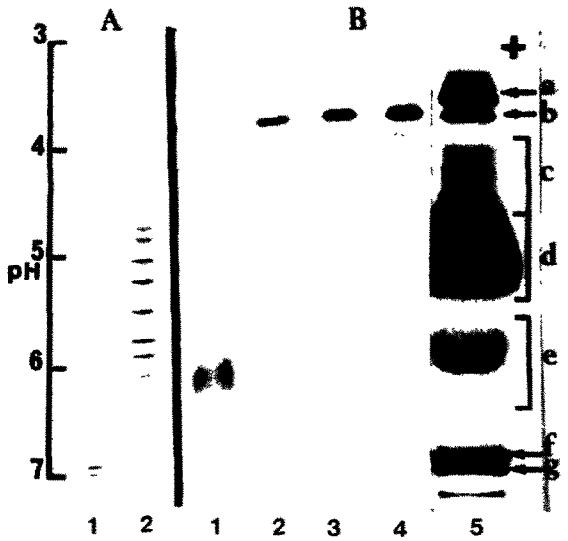

Fig.1. Isoelectric focusing (A) and SDS-polyacrylamide gel electrophoresis (B) of human $\varkappa$-casein before and after chymosin action. (A1) Chymosin-treated $\varkappa$-casein (60 min reaction); (A2, B1) $\varkappa$-casein, (B2, B3, B4) chymosin-treated $\varkappa$-casein (15, 30, 60 min reaction), (B5) human whole casein, overloaded for a clear observation of $\varkappa$-casein and whey protein contaminants. (a) α-Lactalbumin, (b) lysozyme and para-$\varkappa$-casein, (c) degradation products of β-casein, (d) β-casein, (e) $\varkappa$-casein, (f) heavy chains of sIgAs, (g) lactoferrin.

Milford, MA); 30°C; no acetone in solvent A; 1 ml/min; linear gradient from A/B (85:15) to A/B (25:75) in 14 min, followed by a 1 min linear gradient to the initial conditions and 9 min equilibration in the starting mixture. The absorbance was recorded at 269 nm with 0.05 AUFS full scale. PTH dehydrothreonine was successively recorded at 269 and 312 nm.

2.7. *Identification of the amino-terminus of the N-terminal chymotryptic peptide*

As treatments with pyroglutamate aminopeptidase had failed on the intact protein, the N-terminal tryptic and chymotryptic peptides, the procedure described by Muranova et al. [16] was used on the N-terminal chymotryptic peptide. This peptide was treated with 17% methylamine at 37°C for 14 h, dried and subjected to Edman degradation.

3. RESULTS AND DISCUSSION

3.1. *Preparation and amino acid composition of human κ-casein*

Fig.1 shows the electrophoretic patterns obtained by SDS-PAGE and IEF with our prepara-

Table 1

Amino acid composition of human κ-casein

Amino acid	This work (from amino acid analysis)	This work[a] (from sequence)	From Yamauchi et al. [7][b]	Bovine κ-casein from Mercier et al. [19]
SCM-Cys	0.8	1	0.8	2
Asp	12.0	3	12.8	4
Asn		9 (12)		7 (11)
Thr	18.0[c]	18	17.8	14
Ser	8.0[c]	7	7.5	13
Glu	15.4	7	15.2	12
Gln		8 (15)		15 (27)
Pro	28.1	28	28.7	20
Gly	1.9	1	2.6	2
Ala	13.5	14	13.8	15
Val	11.1	13	12.4	11
Met	0.4	1	1.3	2
Ile	12.0	13	10.6	13
Leu	3.5	3	4.3	8
Tyr	11.3	12	10.3	9
Phe	3.1	3	3.1	4
His	4.0	4	3.0	3
Trp	–[d]	–[d]	1.1	1
Lys	5.0	5	4.9	9
Arg	7.7	8	7.7	5
Total		158		169
M_r (peptide chain)		17707		19023
GlcN	21.6			
GalN	11.7			

[a] Including the amino acid composition of CMP, according to Chobert et al. [4]
[b] Recalculated for 158 residues
[c] Values obtained by extrapolation to 0 h
[d] No reaction with Ehrlich's reagent

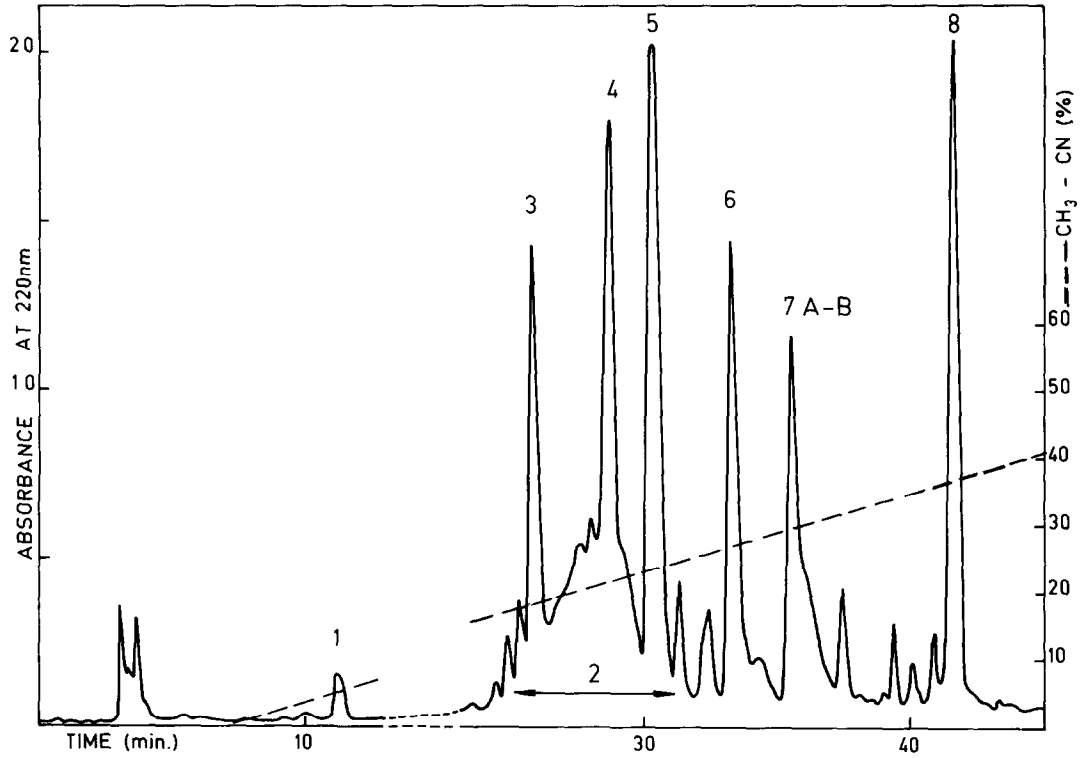

Fig.2. Separation of tryptic peptides of human SCM-$\varkappa$-casein by HPLC. Elution with system I. Conditions in section 2.

tion of human $\varkappa$-casein. The location of the corresponding fraction in whole human casein is also shown in SDS-PAGE. The protein displays at least 11 bands in IEF. These arise from microheterogeneity of the carbohydrate moieties, since they disappear after chymosin treatment which only leaves 2 para-$\varkappa$-derivatives. Indeed it is known that these carbohydrates are all attached to CMP [7], which is not stained by Coomassie blue. SDS-PAGE gives an apparent M_r of ~14000 for the major para-$\varkappa$-derivative. The reason for the appearance of the minor one is not known.

The amino acid and hexosamine composition is shown on table 1. The assumption was first made that the protein had the same number of amino acid residues as its bovine counterpart. Thereafter calculations were made from the actual number of residues deduced from the sequence. The absence →

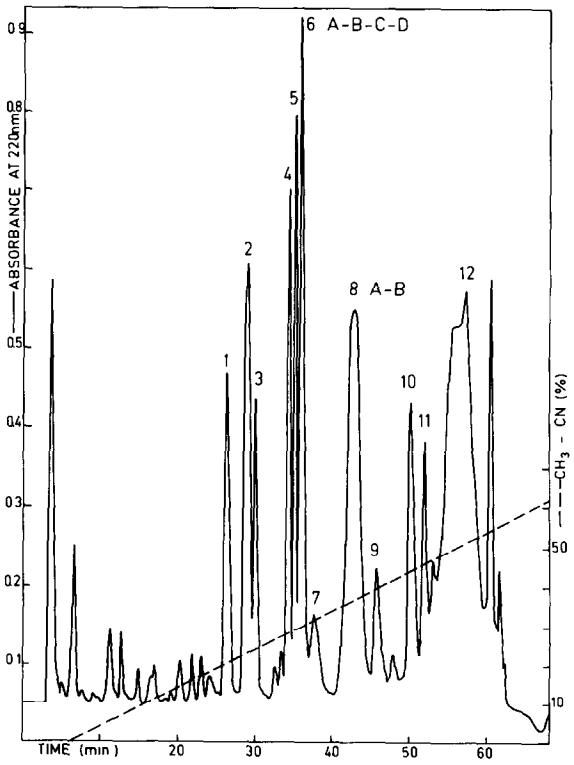

Fig.3. Separation of chymotryptic peptides of human SCM-$\varkappa$-casein by HPLC. Elution with system I. See conditions in section 2.

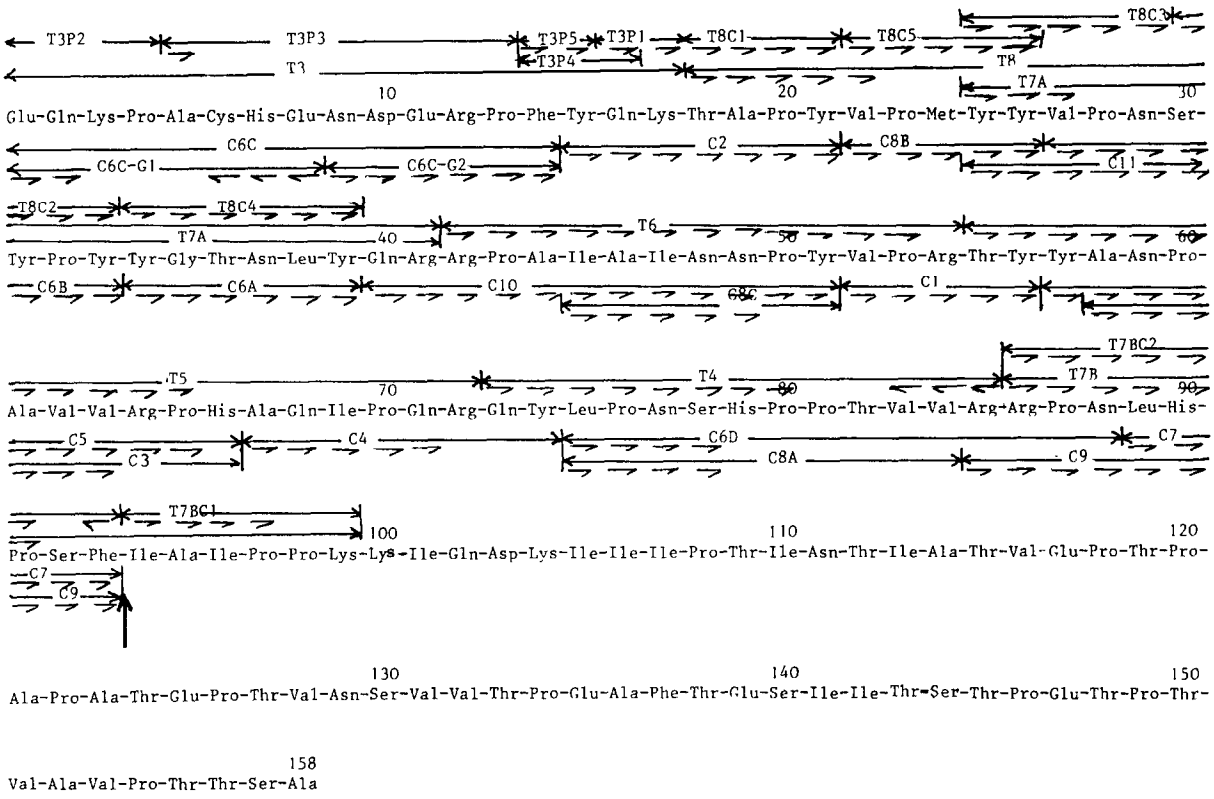

Fig.4. Alignment of the peptides used for sequence determination. The vertical arrow indicates the chymosin-sensitive bond. (T) Tryptic peptides, (C) chymotryptic peptides, (P) peptides obtained by using proline-specific endopeptidase, (G) peptides obtained by using *S. aureus* V8 protease. Sequence of CMP (94–158) from [4]; exact assignment of residues 143, 144, 153–157 from [6]. Amino acids released by: (⟶) Edman degradation, (⟵) carboxypeptidases.

of tryptophan was deduced from the negative reaction obtained with Ehrlich's reagent on the whole protein and its tryptic fragments. The Edman degradation on the intact molecule did not release any PTH-amino acid.

3.2. *Amino acid sequence of tryptic peptides*

The HPLC elution profile of the SCM-$\varkappa$-casein tryptic digest is shown in fig.2. From the amino acid sequence of the human CMP [4] it could be inferred that the tryptic hydrolysate of the whole protein should contain 2 peptides originating from the CMP: a pentapeptide and the 54-residue C-terminal peptide which should contain all the carbohydrates. Furthermore the junction peptide overlapping para-$\varkappa$-casein and 6 residues of CMP should also be present. The 54-residue C-terminal peptide was found as a spreading zone (fraction T2, fig.2) which was identified by amino acid analyses of the fractions located between peaks 3 and 4, 4 and 5. Three fractions, T1, T6, T8, corresponded to pure peptides. Fractions T3, T4, T5, all contaminated by T2, were purified by HPLC. Likewise fraction 7 was resolved into 2 pure peptides, T7A and T7B. T1 and T2 were not studied any further since they obviously corresponded to the 2 expected peptides from CMP. T3, from which no PTH amino acid could be obtained, was treated with proline-specific endopeptidase. Five peptides (T3P1–T3P5) were obtained from the digest by HPLC. Chymotryptic digests of T7B and T8 were resolved by HPLC into 2 and 5 peptides, respectively. The results of Edman degradations on all tryptic peptides are shown in fig.4. T3P2 did not give any PTH derivative.

3.3. *Amino acid sequence of chymotryptic peptides*

From a chymotryptic digest of SCM-$\varkappa$-casein, 12 main peptide fractions were obtained by HPLC

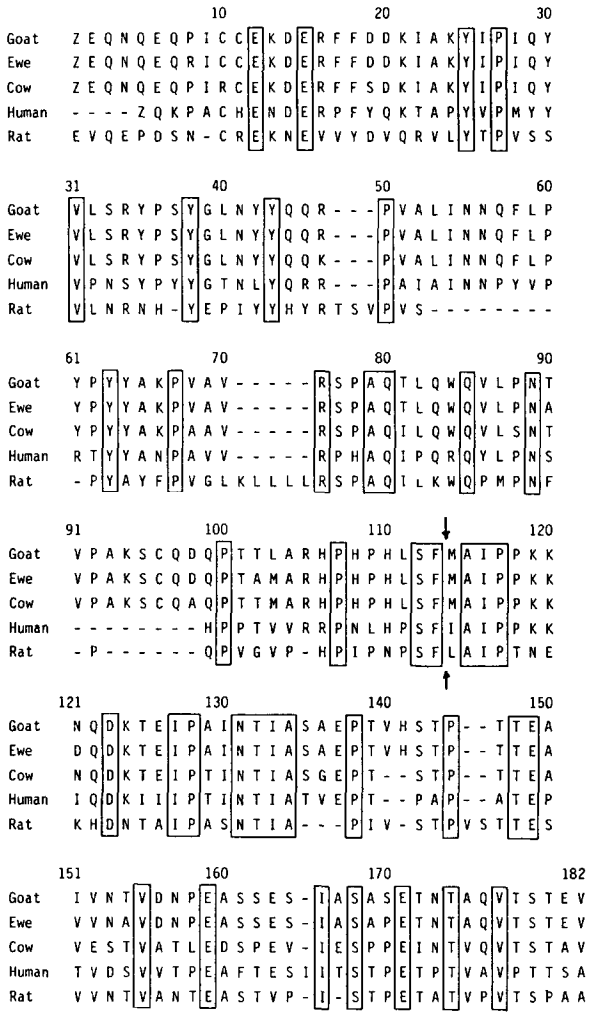

Fig.5. Alignment of the sequences of goat [5], ewe [22], cow variant A [19], human and rat [18] $\varkappa$-casein. Standard one-letter symbols for amino acids; (□) identical residue; (↓) chymosin/pepsin-sensitive bond; phosphates are attached to S 161 in goat, ewe and cow, S 179 in goat and ewe. In the pre-proteins, the N-terminal amino acid of ewe [20] and cow [21] $\varkappa$-caseins is Q. For the ewe protein, E and Q have been placed in position 2 and 7, respectively, as determined on the pre-protein [20]. For the cow protein, N occurs at position 89 [21] instead of E [19]. The assignment of residues 143, 144, 153–157 in the human protein has been taken from [6]. The N-terminal sequence of the goat protein, Z-Z-Z-N-Z-Z-Z- [5] is considered to be identical with those of the ewe and cow proteins.

(fig.3). Fractions C1–C5, C7, C9–C11 each contained a single peptide. The large trailing peak C12 contained all the carbohydrates. It was a mixture of 3–4 large peptides starting at Leu-75 or Phe-93 and ending at Phe-137 or at the carboxyl terminus (fig.4). Refractionation of fractions C6 and C8 by HPLC gave peptides C6A–C6D and C8A–C8C. Fraction C12 was not studied any further. Peptide C6C, which did not react with PITC, was digested with *S. aureus* V8 protease. The digest was resolved by HPLC into 2 peptides, C6C-G1 and C6C-G2. The former did not react with PITC. Peptide C6C was treated with methylamine [16] to unblock a putative pyroglutamyl residue and convert it into N^5-methylglutamine. Two steps of Edman degradation were carried out on the treated peptide and the synthetic peptide Pyr-Glu-Ala which had been similarly treated. The same patterns were obtained by HPLC with the PTH derivatives at the first step. PTH-Gln was detected in second position with C6C. The results of Edman degradation on all chymotryptic peptides are shown on table 2.

The sequence data obtained from the tryptic and chymotryptic peptides (fig.4) give the complete sequence of human para-$\varkappa$-casein without ambiguity. Taking into account peptide T7B, which overlaps para-$\varkappa$-casein and CMP, they give the sequence of the whole human $\varkappa$-casein molecule (fig.4), which is compared to those of the other known $\varkappa$-caseins in fig.5. The protein is composed of 158 residues. From determinations of fucose, galactose, glucosamine, galactosamine and sialic acid (given in wt%), Yamauchi et al. [7] deduced a carbohydrate content of ~40% for their $\varkappa$-casein preparation. The values we present here for the glucosamine and galactosamine contents of our preparation (table 1) have been obtained on a molar basis by carrying out amino acid composition and amino sugar determination on the same solution of $\varkappa$-casein. This gave values approximately twice higher than those reported in [7] but in the same ratio. Assuming, as a first approximation, that the other carbohydrates occur in our preparation in the proportions of the figures given in [7], we have recalculated them on the basis of our results for the amino sugars. This gives an M_r of 19400 for the whole carbohydrate moiety. The M_r of the peptide chain, deduced from the sequence, being 17707, this leads to a value of ~37100 for the M_r of the whole $\varkappa$-casein, a value close to that of 38000 recently found by ultracentrifugation by Azuma et al. [17]. Human $\varkappa$-casein

	CMP / PARA-κ	1	2	3	4	5
1	GOAT		5	27	48	47
2	EWE	5		29	46	46
3	COW	8	10		46	49
4	HUMAN	54	57	58		57
5	RAT	65	65	68	76	

Fig.6. Percent differences between para-κ-caseins (below diagonal) and CMPs (above diagonal) from goat, ewe, cow, human and rat. All positions where a residue occurred in one or both sequences were considered.

could then contain ~52% carbohydrates.

It was postulated earlier that CMP could have evolved at a much faster rate than para-κ-casein. This assumption was based on a comparison between the rates of evolution of CMPs from cow, water buffalo, goat, ewe, pig and human on the one hand, and those of para-κ-caseins of 3 closely related species (bovine, goat, sheep) on the other [5]. Fig.6 compares percent differences of CMPs and para-κ-caseins from cow, goat, ewe, human and rat. The rat κ-casein sequence was recently deduced from that of the corresponding cDNA [18]. It shows that the rate of evolution of para-κ-casein is higher than that of CMP. This is mainly due to the high homology which can be observed in fig.5 between residues 112 and 134 (55 and 22% identity in the 5 species for this fragment and the whole molecule, respectively). This region contains bond 113–114 whose cleavage by pepsin or chymosin leads to milk coagulation in the stomach. It thus appears that, during evolution, κ-casein retained only the minimal features that seem necessary for its role, which also include a high proportion and even distribution of proline residues, a high hydrophobicity and positive net charge of para-κ-casein and a negative net charge of CMP.

REFERENCES

[1] Dalgleish, D.G. (1982) in: Developments in Dairy Chemistry (Fox, P.F. ed.) vol.1, pp.157–187, Applied Science, London.

[2] Knoop, E. and Wortmann, A. (1967) Milchwissenschaft 22, 198–205.

[3] Greenberg, R., Groves, M.L. and Dower, H.J. (1984) J. Biol. Chem. 259, 5132–5138.

[4] Chobert, J.M., Mercier, J.C., Bahy, C. and Hazé, G. (1976) FEBS Lett. 72, 173–178.

[5] Mercier, J.C., Chobert, J.M. and Addeo, F. (1976) FEBS Lett. 72, 208–214.

[6] Fiat, A.M., Jollès, J., Aubert, J.P., Loucheux-Lefebvre, M.H. and Jollès, P. (1980) Eur. J. Biochem. 111, 333–339.

[7] Yamauchi, K., Azuma, N., Kobayashi, H. and Kaminogawa, S. (1981) J. Biochem. 90, 1005–1012.

[8] Chobert, J.M., Addeo, F. and Ribadeau-Dumas, B. (1981) J. Dairy Res. 48, 429–436.

[9] Nijhuis, H. and Klostermeyer, H. (1975) Milchwissenschaft 30, 530–531.

[10] Laemmli, U.K. (1970) Nature 227, 680–685.

[11] Trieu-Cuot, P. and Gripon, J.C. (1981) J. Dairy Res. 48, 303–310.

[12] Canfield, R.E. and Anfinsen, C.B. (1963) in: The Proteins (Neurath, H. ed.) p.319, Academic Press, New York.

[13] Podell, D.N. and Abraham, G.N. (1978) Biochem. Biophys. Res. Commun. 81, 176–185.

[14] Tarr, G.E. (1982) in: Methods in Protein Sequence Analysis (Elzinga, M. ed.) pp.223–232, Humana Press, Clifton, NJ.

[15] Bhown, A.S., Mole, J.E., Weissinger, A. and Bennett, J.C. (1978) J. Chromatogr. 148, 532–535.

[16] Muranova, T.A., Muranov, A.V., Markova, L.F. and Ovchinnikov, Y.A. (1980) in: Methods in Peptide and Protein Sequence Analysis (Birr, C. ed.) pp.459–466, Elsevier, Amsterdam, New York.

[17] Azuma, N., Kaminogawa, S. and Yamauchi, K. (1984) Agric. Biol. Chem. 48, 771–776.

[18] Nakhasi, H.L., Grantham, F.H. and Gullino, P.M. (1984) J. Biol. Chem. 259, 14894–14898.

[19] Mercier, J.C., Brignon, G. and Ribadeau-Dumas, B. (1973) Eur. J. Biochem. 35, 222–235.

[20] Mercier, J.C. and Gaye, P. (1983) in: Biochemistry of Lactation (Mepham, T.B. ed.) pp.177–227, Elsevier, Amsterdam, New York.

[21] Stewart, A.F., Willis, I.M. and McKinlay, A.G. (1984) Nucleic Acids Res 12, 3895–3907.

[22] Jollès, J., Fiat, A.M., Schoentgen, F., Alais, C. and Jollès, P. (1974) Biochim. Biophys. Acta 365, 335–343.