COMPLETE AMINO ACID SEQUENCE OF BOVINE α_{s2}-CASEIN

Ghislaine BRIGNON, Bruno RIBADEAU DUMAS, Jean-Claude MERCIER and Jean-Pierre PELISSIER
Laboratoire de Recherches sur les Protéines, INRA, 78350 Jouy-en-Josas

and

B. C. DAS
Institut de Chimie des Substances Naturelles, CNRS, 91190 Gif-sur-Yvette, France

Received 4 February 1977

1. Introduction

In a preceding paper [1] it was shown that the so-called α_{s2}-, α_{s3}-, α_{s4}- and α_{s6}-caseins, which seem to occur in all individual bovine milks, have identical peptide chains and may only differ in their phosphate content (13–10 phosphate groups/mol). Thus, the primary structure of these proteins was studied on a mixture of the four, designated as α_{s2}-casein. Five peptides, CN1 to CN5, containing 49, 22, 17, 4 and 115 residues, respectively, were isolated from a cyanogen bromide hydrolysate of α_{s2}-casein. They accounted for the 4 methionyl and 207 amino acid residues of the peptide chain. Except for residues 186–187 (His, Gln) of CN1, peptides CN1, CN3 and CN4 were completely sequenced. The alignment of the 5 CNBr peptides, H.CN4–CN2–CN5–CN1–CN3.OH, was determined from sequences studies carried out on these peptides as well as on intact α_{s2}-casein and on some of its 25 tryptic peptides. The sequence of approximately 100 residues (1–12, 24–51, 138–207) was given [1].

The present communication reports the complete primary structure of α_{s2}-casein and preliminary data on the location of the phosphate groups.

2. Material and methods

The isolation of α_{s2}-casein and of the CNBr and tryptic peptides derived from its *S*-carboxymethylated derivative has been previously reported [1]. Most of the techniques used in the present study were either described or quoted in our first paper on α_{s2}-casein [1] and in other publications from our laboratory; peptic and tryptic hydrolysis [2], thermolysin hydrolysis [3], maleylation and demaleylation [4]. Some peptides were dephosphorylated using potato acid phosphatase [5] (Boehringer, 60 U/mg). In this case the peptide was dissolved in 0.05 M Na-acetate, pH 5.2. The commercial enzyme suspension was then added to obtain a molar ratio enzyme/substrate of 1/200 and the mixture was incubated at 32°C for 16 h. About 80% dephosphorylation was obtained. The amino acid sequence of some peptides was determined by mass spectrometry, using a AEI MS9 apparatus, after acetylation (with a 1:1 mixture of normal and deuterated acetic anhydride) and permethylation [6].

3. Results

3.1. *Sequence of peptide CN2 (fig. 1)*

The sequences of 8 and 3 residues, respectively, from the N- and C-terminal ends of this phosphopeptide were already known [1]. The Edman degradation applied to dephosphorylated peptide CN2 (CN2d) established the sequence of the first 16 residues. Peptide CN2d was digested with thermolysin. Fractionation of the hydrolysate on Dowex 50 X 2 gave 8 pure peptides, CN2dTh1 to CN2dTh8, all fitting in with the known partial sequence. The

 North-Holland Publishing Company – Amsterdam

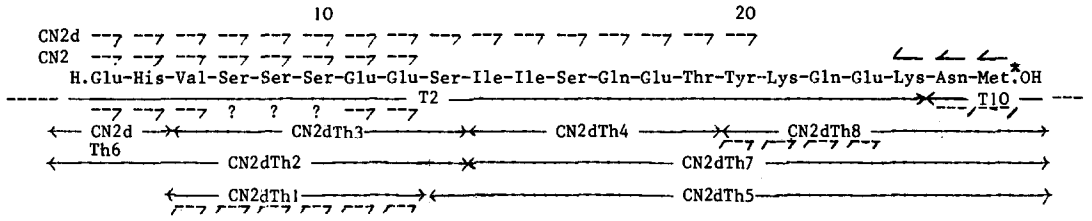

Fig.1. Determination of the primary structure of peptide CN2. Amino acid released by: ——› leucine aminopeptidase, ‹—— carboxypeptidases A and B; sequences established by: ——7 automatic Edman degradation, – – 7 manual Edman degradation, /–––7 mass spectrometry;?, PTH-amino acid not detected; Met*, Met transformed into homoserine-homoserine lactone.

sequence of peptide CN2dTh1 was confirmed by mass spectrometry. This technique applied to CN2dTh8 elucidated the order of its first 4 residues, Tyr–Lys–Gln–Glu, thus completing the study of peptide CN2.

3.2. *Sequence of peptide CN5 (fig.2)*

The sequences of 25 and 4 residues, respectively, from the N- and C-terminal ends of this phosphopeptide were presented earlier [1].

3.2.1. Tryptic digest

A tryptic hydrolysate of peptide CN5 was fractionated on Dowex 50 X 2. After further purification (Sephadex G-25, preparative paper electrophoresis) 12 pure peptides were recovered, corresponding to the whole peptide CN5. Ten of them were identical with peptides already obtained from a tryptic hydrolysate of the α_{s2}-casein (T4, T14, T1, T8, T23, T12, T25, T13, T6, T3) [1]. The last 2 peptides,

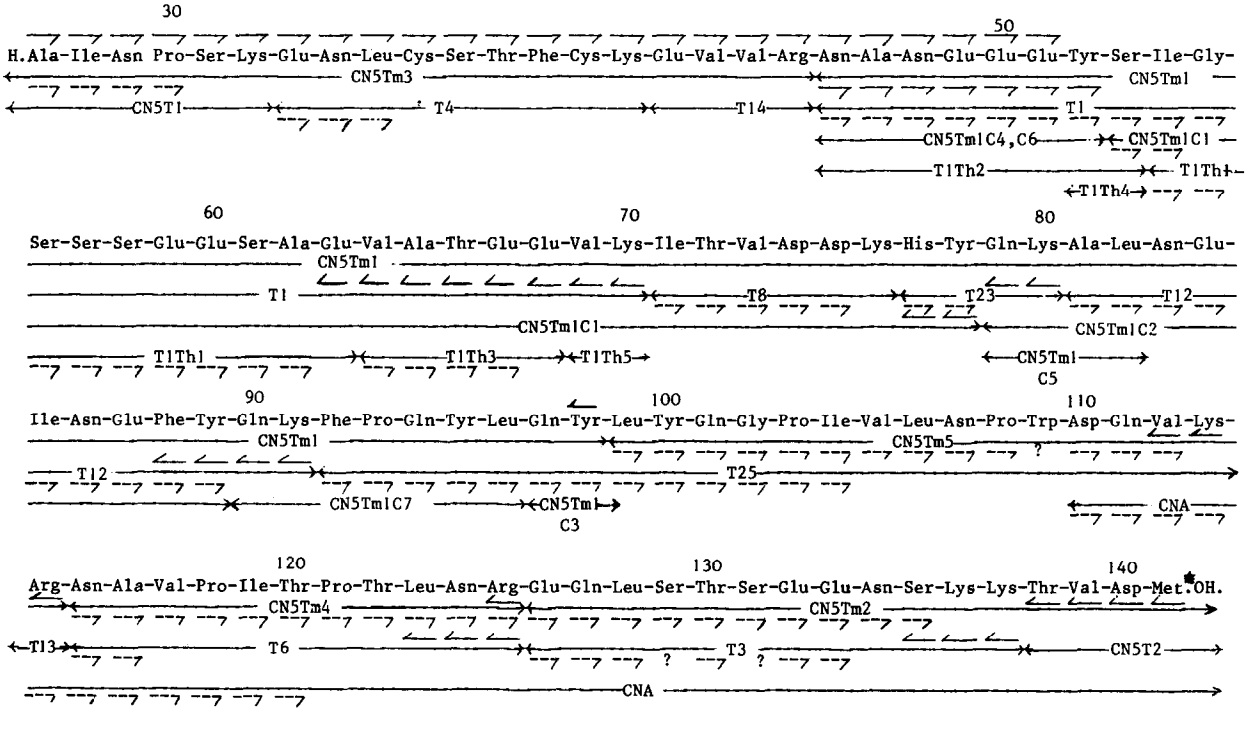

Fig.2. Determination of the primary structure of peptide CN5. Symbols as in fig.1.

CN5T1 and CN5T2, corresponded to the N- and C-terminal parts of CN5, respectively, which were already known [1]. The sequence information obtained directly from the tryptic peptides is indicated on fig.2.

3.2.2. Tryptic digest of maleyl peptide CN5

This digest contained 5 peptides, CN5Tm1 to CN5Tm5, which were separated on Sephadex G-50 and purified using the techniques mentioned above. They contained 53, 16, 19, 11 and 16 residues, respectively, and accounted for the whole peptide CN5. Peptide CN5Tm3 was the N-terminal of CN5, as indicated by the results of 4 steps of the Edman degradation and the presence of 2 cysteyl residues. Its complete structure was therefore already known from 25 steps of the Edman degradation carried out on peptide CN5 [1], and it obviously corresponds to the alignment CN5T1–T4–T14. The sequence of 7 N-terminal residues of peptide CN5Tm1 obtained by the Edman degradation, in addition to direct sequence information obtained on CN5, indicated that this peptide was linked to the carboxyl terminus of CN5Tm3 and that peptide T1 constituted its N-terminal end. The action of carboxypeptidase A (CPA) on CN5Tm1, which was devoid of arginine, showed that this peptide originated from a non-specific cleavage after a tyrosyl residue. Comparison of its amino acid composition with that of the tryptic peptides obtained from CN5 indicated that it contained peptides T1, T8, T23, T12 and, at its C-terminal end, a fragment of peptide T25. Peptide CN5Tm5 was the other part of peptide T25 with an additional arginyl residue (T13) at its carboxyl terminus. Peptide CN5Tm4 was identical with peptide T6. The only maleyl peptide-containing homoserine was CN5Tm2, which therefore represented the carboxyl-terminus of CN5. It contained peptide T3 and a C-terminal tetrapeptide, Thr–Val–Asp–Met*, which was already known. The complete sequences of peptides CN5Tm5, Tm4 and Tm2 were determined by the Edman procedure and by degradations carried out with CPA and carboxypeptidase B (CPB). A 32 residue peptide, CNA, resulting from non-specific cleavage after a tryptophanyl residue, was found in a CNBr hydrolystae of the α_{s2}-complex. The sequence of the first 11 residues of CNA, determined by the Edman procedure, ordered 2 maleyl peptides as Tm5–Tm4. Comparison of amino acid compositions obviously indicated that CNA was the carboxyl terminus of CN5. All these results gave complete alignment of the Tm-peptides from CN5, H.Tm3–Tm1–Tm5–Tm4–Tm2.OH, and a partial alignment of its tryptic fragments, HCN5T1–T4–T14–T1–(T8, T23, T12)–T25–T13–T6–T3–CN5T2.OH. At this stage, it remained to order peptides T8, T23, T12 which were parts of peptide CN5Tm1, and to sequence peptide T1, so as to achieve the complete sequence determination of peptide CN5.

3.2.3. Chymotryptic digest of peptide CN5Tm1

After a mild digestion of peptide CN5Tm1 with chymotrypsin, the hydrolystae was fractionated on Dowex 50 × 2. Seven pure peptides (C1 to C7) were obtained. The study of the ends of C1 and comparison of its composition with that of the tryptic peptides from CN5 showed that C1 was composed of T1 (lacking 7 N-terminal residues), T8 and T23 (lacking 2 C-terminal residues). This gave the alignment T1-T8-T23. The orderings T23–T12 and T12–T25 were confirmed by the amino acid compositions of peptides C2 and C7 respectively. From this last study the complete alignment of the tryptic peptides from CN5 was established as: H.CN5T1–T4–T14–T1–T8–T23–T12–T25–T13–T6–T3–CN5T2.OH.

3.2.4. Sequence of peptide T1

The first 7 residues of peptide T1 were already sequenced by Edman degradations carried out on peptides CN5(1) and CN5Tm1. The Edman procedure, applied to peptide T1, ordered 3 more residues. A partial C-terminal sequence of peptide T1, (Thr, Glu, Ala, Val)–Glu–Val–Lys, was established from study of digests obtained with both CPA and CPB at 2 different pHs (5.4 and 8.6). A thermolysin digest of peptide T1 was fractionated on Dowex 50 × 2. After further purifications on Sephadex G-15 in 30% acetic acid, 5 pure peptides T1Th1 to T1Th5, accounting for the whole peptide T1, were obtained. From its amino acid composition, the octapeptide T1Th2 was obviously the known N-terminal end of peptide T1. The dipeptide T1Th5 (Val–Lys) was its C-terminal end. The complete sequence of peptide T1Th1 was elucidated by applying the Edman procedure after dephosphorylation. The comparison of the N-terminal ends of T1 and T1Th1 showed that the latter followed T1Th2 in the sequence. Finally, peptide

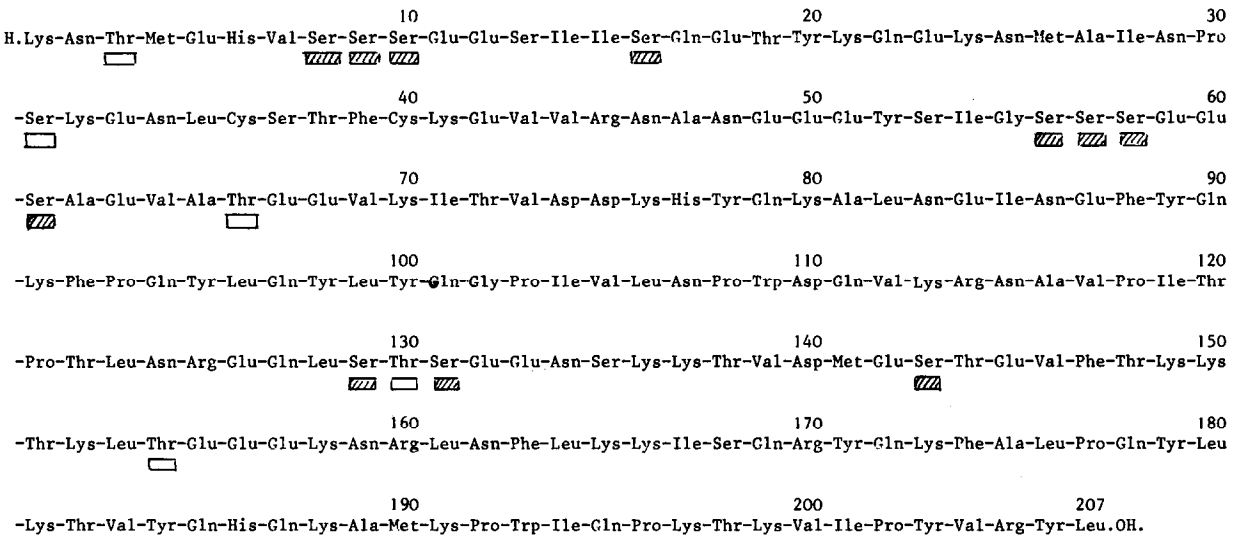

Fig.3. Complete primary structure of α_{s2}-casein. ▭, potential site of phosphorylation; ▨, phosphate groups found in this study.

T1Th3 obviously represented the penultimate fragment of peptide T1. The first 4 residues of peptide T1Th3 were ordered by the Edman degradation, thus completing the determination of the primary structure of T1, and therefore of CN5. The last thermolysin peptide, T1Th4 (Tyr–Ser), was a fragment of peptide T1Th2. Its electrophoretic mobility, at pH 1.9, showed that it was devoid of phosphate.

3.3. *Sequence of peptide CN1*

This sequence was known except for the order of residues 186–187 (His, Gln) [1]. A tryptic hydrolysate of maleyl α_{s2}-casein gave 5 main fractions by chromatography on Sephadex G-50. The third fraction contained a 33 residue peptide, α_{s2}-TmIII, corresponding to the fragment 171–203. The 22 N-terminal residues were sequenced by the automatic Edman degradation, thus giving the order of residues 186–187: His–Gln. This result finished the elucidation of the primary structure of α_{s2}-casein peptide chain (fig.3).

3.4. *Partial localization of phosphate groups*

The exact localization of the phosphate groups in each of the 4 α_{s2}-casein fractions will be reported in another publication. From study of the peptides used for amino acid sequence determination and from phosphorus analyses, phosphate groups were found in positions 8, 9, 10, 16, 56, 57, 58, 61, 129, 131 and 143 (fig.3). However, when peptides occurred at various degrees of phosphorylation, only one species was studied. This means that the positions mentioned above are phosphorylated in at least one of the α_{s2}-casein fractions and that order positions may be phosphorylated in some of these fractions.

4. Discussion

α_{s2}-Casein has features common to both κ- and α_{s1}-caseins. Like κ-casein it contains 2 cysteyl residues and, like α_{s1}-casein, it is insoluble in the presence of calcium. It is stabilized as micelles by κ-casein [7] and contains a fairly high number of phosphate groups. Unlike κ-casein which, in the presence of urea, is in a polydisperse state of high average molecular weight through random disulphide linkages [8], α_{s2}-casein seems to form, in the same conditions, dimers through disulphide bridge(s) [9].

The peptide chain of α_{s2}-casein contains 207 amino acid residues Asp_4, Asn_{14}, Thr_{15}, Ser_{17}, Glu_{25}, Gln_{15}, Pro_{10}, Gly_2, Ala_8, Cys_2, Val_{14}, Met_4, Ile_{11}, Leu_{13}, Tyr_{12}, Phe_6, Trp_2, Lys_{24}, His_3, Arg_6. Its molecular weight is 24 350. If the exact number of phosphate groups linked to the chain lies between 10 and 13

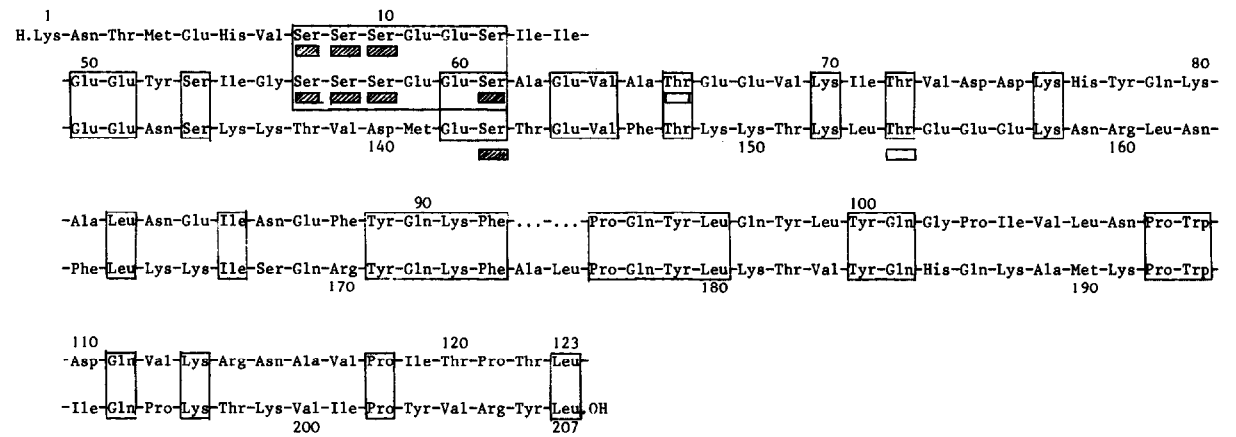

Fig.4. Sequence homologies within α_{s_2}-casein peptide chain.

[1], the total molecular weight of the 4 α_{s_2}-casein fractions varies from 25 150–25 390, assuming that none of them contain any carbohydrate. This may not be true. Preliminary results indicate that some of them may contain some bound carbohydrates.

The peptide chain of α_{s_2}-casein, compared to that of the other caseins, carries a large number of positive charges located all along the chain, especially in the 47 C-terminal residues which contain 12 basic residues and no acidic one. The phosphate groups which have been located are situated in 3 regions of the peptide chain. In any case, every phosphorylated residue has a glutamic acid or a phosphoserine, two residues to the right in the sequence. Therefore, it seems that this requirement, observed by our group in the other caseins [10,11], is also necessary for the phosphorylation of α_{s_2}-casein.

According to this rule, 16 potential phosphorylation sites exist in α_{s_2}-casein. The preliminary results mentioned above show that at least 11 of these sites can indeed be phosphorylated. The average hydrophobicity of the peptide chain, determined according to Bigelow [12], is the lowest among the caseins: α_{s_2} 4.64, α_{s_1} 4.89, κ 5.37, β 5.56, kJ/residue. Likewise, its proline content is the lowest: α_{s_2} 4.8, α_{s_1} 8.5, κ 11.8, β 16.7, %. Nevertheless the high number and even distribution of charged groups hould prevent the formation of long periodic structures. A bitter peptide isolated in 1969 by Matoba et al. from a tryptic digest of whole casein [13] obviously corresponds to segment 174–181 of α_{s_2}-casein.

Figure 4 shows interesting sequence homologies observed namely between 2 segments of 76 residues from α_{s_2}-casein chain. They contain 38% identical residues. This feature may indicate that α_{s_2}-casein originates from the duplication of a primitive gene.

A polymorphism affecting simultaneously the 4 α_{s_2}-casein fractions has been recently discovered. Genetic studies of this polymorphism indicate that the 4 fractions are synthesized by the same structural gene and that this gene is closely linked to the genes coding for α_{s_1}, β- and κ-caseins [14].

References

[1] Brignon, G., Ribadeau Dumas, B. and Mercier, J-C. (1976) FEBS Lett. 71, 111–116.
[2] Mercier, J-C., Grosclaude, F. and Ribadeau Dumas, B. (1970) Eur. J. Biochem. 14, 108–119.
[3] Ribadeau Dumas, B., Grosclaude, F. and Mercier, J-C. (1971) Eur. J. Biochem. 18, 252–257.
[4] Grosclaude, F., Mercier, J.-C. and Ribadeau Dumas, B. (1970) Eur. J. Biochem. 14, 98–107.
[5] Alvarez, E. F. (1962) Biochim. Biophys. Acta 59, 663–672.
[6] Das, B. C. and Lederer, E. (1971) in: New Techniques in Amino acid, Peptide and Protein Analysis (Niederwieser, A. and Pataki, G. ed) pp. 175–225, Ann. Arbor Science.
[7] Toma, S. J. and Nakai, S. (1973) J. Dairy Sci. 56, 1559–1562.
[8] Swaisgood, H. E., Brunner, J. R. and Lillevik, H. A. (1964) Biochemistry 3, 1616.

[9] Hoagland, P. D., Thompson, M. P. and Kalan, E. B. (1971) J. Dairy Sci. 54, 1103–1110.
[10] Mercier, J-C., Grosclaude, F. and Ribadeau Dumas, B. (1971) Eur. J. Biochem. 23, 41–51.
[11] Mercier, J-C, Grosclaude, F., Brignon, G. and Ribadeau Dumas, B. (1977) to be published.
[12] Bigelow, C. C. (1967) J. Theor. Biol. 16, 187.
[13] Matoba, T., Nagayasu, C., Hayashi, R. and Hata, T. (1969) Agr. Biol. Chem. 33, 1662–1663.
[14] Grosclaude, F., Mahé, M-F., Mercier, J-C., Bonnemaire, J. and Teissier, J. H. (1977) Ann. Gén. Sel Anim., in press.