

COMPARATIVE STUDY OF THE AMINO ACID SEQUENCES OF THE CASEINOMACROPEPTIDES FROM SEVEN SPECIES

Jean-Claude MERCIER and Jean-Marc CHOBERT

Laboratoire de Recherches sur les Protéines, I.N.R.A.-C.N.R.Z., 78350, Jouy-en-Josas, France

and

Francesco ADDEO

Istituto di Industrie Agrarie dell'Università di Napoli, 80055 Portici, Italy

Received 28 August 1976

1. Introduction

Caseins, the predominant proteins synthesized in the mammary gland, occur in milk as large and stable calcium phosphate complexes called micelles. When milk is treated with a low amount of a protease such as chymosin (rennin EC 3.4.23.4), the micelles become clotted. This is the result of the hydrolysis of a labile peptide bond in κ -casein, which then loses its ability to stabilize casein micelles. The soluble carbohydrate-rich C-terminal fragment, caseinomacropeptide (CMP), which is released accounts for about one third of κ -casein polypeptide chain. The N-terminal fragment, para- κ -casein, which polymerizes is insoluble.

Some data concerning the amino acid sequence of bovine κ -casein were published in 1970 by Jollès et al. [1]. In 1972, we determined the complete primary structure of this casein which was shown to be a single polypeptide chain containing 169 amino acid residues [2–4]. In the same year, Grosclaude et al. [5] reported the localization of the 2 amino acid substitutions differentiating the CMPs of bovine genetic variants A and B which later were compared to their zebu counterparts [6]. In 1974, Jollès et al. [7,8] published the complete sequence of ovine κ -casein.

We have recently examined the amino acid sequences of κ -casein from 4 other species in an attempt to determine the regions of the polypeptide chain which

might have been conserved during the course of evolution. Such regions may presumably be involved in the stabilization of casein micelles or responsible for the sensitivity of κ -casein to chymosin and other proteolytic enzymes. The primary structures of caprine κ -casein [9–10], water buffalo [11] and porcine [12] CMPs have been completely established. That of human CMP [12] has been nearly elucidated.

In this paper we compare the amino acid sequences of the CMPs from seven species, cow, zebu, water buffalo, goat, ewe, sow and woman. In addition, we present the amino acid sequences of the C-terminal portion of bovine, caprine, ovine and water buffalo para- κ -caseins. The amino acid replacements which differentiate the sequences of bovine, caprine and ovine para- κ -caseins are also reported. The evolutionary rate of change of the CMP was found to be similar to that of fibrinopeptide A, one of the most rapid observed so far. The middle and the carboxy-terminal regions of the CMP have a rather variable sequence whereas its N-terminal region as well as the C-terminal portion of para- κ -casein are remarkably similar in the species studied. The conservation during the course of evolution of the primary structure of κ -casein in the area containing the labile peptide bond, whose enzymatic cleavage releases the negatively charged CMP and triggers the milk-clotting process, is the most striking feature and its biological significance is discussed.

(A)	1	8	10	19	46	65	73	80	82	90	94	95	100	105									
Bovine*	Gln	Pro	Arg	Ser	Lys	Ala	Ile	Ser	Thr	Ala	Thr	Met	Ala	Arg	His	Pro	His	Pro	His	Leu	Ser	Phe	
Water buffalo*													Met	Thr	Arg	His	Pro	His	Pro	His	Leu	Ser	Phe
Goat*	Gln	Pro	Cys	Asp	Arg	Val	Thr	Pro	Thr	Asp	Thr	Leu	Ala	Arg	His	Pro	His	Pro	His	Leu	Ser	Phe	
Sheep	Gln	Arg	Cys	Asp	Arg	Val	Thr	Pro	Ala	Asp	Ala	Met	Ala	Arg	His	Pro	His	Pro	His	Leu	Ser	Phe	
(B)	106	110	115	120	125	130																	
Bovine*	-Met-Ala-Ile-Pro-Pro-Lys-Lys-Asn-Gln-Asp-Lys-Thr-Glu-Ile-Pro-Thr-Ile-Asn-Thr-Ile-Ala-Ser-Gly-Glu-Pro-																						
Zebu*	-Met-Ala-Ile-Pro-Pro-Lys-Lys-Asx-Glx-Asx-Lys-Thr-Glx-Ile-Pro-Thr-Ile-Asx-Thr-Ile-Ala-Ser-Gly-Glx-Pro-																						
Water buffalo*	-Met-Ala-Ile-Pro-Pro-Lys-Lys-Asn-Gln-Asp-Lys-Thr-Glu-Ile-Pro-Thr-Ile-Asn-Thr-Ile-Val-Ser-Val-Glu-Pro-																						
Goat*	-Met-Ala-Ile-Pro-Pro-Lys-Lys-Asn-Gln-Asp-Lys-Thr-Glu-Ile-Pro-Ala-Ile-Asn-Thr-Ile-Ala-Ser-Ala-Glu-Pro-																						
Sheep	-Met-Ala-Ile-Pro-Pro-Lys-Lys-Asp-Gln-Asp-Lys-Thr-Glu-Ile-Pro-Ala-Ile-Asn-Thr-Ile-Ala-Ser-Ala-Glu-Pro-																						
Pig*	-Ile-Ala-Ile-Pro-Pro-Lys-Lys-Asn-Gln-Asp-Lys-Thr-Ala-Ile-Pro-Ala-Ile-Asn-Ser-Ile-Ala-Thr-Val-Glu-Pro-																						
Human*	-Ile-Ala-Ile-Pro-Pro-Lys-Lys-Ile-Gln-Asp-Lys-Ile-Ile-Ile-Pro-Thr-Ile-Asn-Thr-Ile-Ala-Thr-Val-Glu-Pro-																						
	135	140	145																				
Bovine	-Thr-	-Ser-Thr-Pro-Thr-Ile ^b -Glu-Ala-Val-Glu-Ser-Thr-	-Val-Ala-Thr-Leu-Glu-																				
Zebu	,Thr,	,Ser, Thr, Pro-Thr-Ile ^b -Glu-Ala/Val, Glx, Ser, Thr,	,Val, Ala, Thr/Leu-Glu-																				
Water buffalo	-Thr-	-Ser-Thr-Pro-Thr-Ile ^a -Glu-Ala-Ile-Glu-Asn-Thr-	-Val-Ala-Thr-Leu-Glu-																				
Goat	-Thr-Val-His-Ser-Thr-Pro-Thr-Thr-Glu-Ala-Ile-Val-Asn-Thr-	-Val-Asp-Asn-Pro-Glu-																					
Sheep	-Thr-Val-His-Ser-Thr-Pro-Thr-Thr-Glu-Ala-Val-Val-Asn-Ala-	-Val-Asp-Asn-Pro-Glu-																					
Pig	-Thr-	-Ile-Val-Pro-Ala-Thr-Glu-Pro-Ile-Val-Asn-Ala-Glu-Pro-Ile-Val-Asn-Ala-Val-Val-Thr-Pro-Glu-																					
Human	-Thr-	-Pro-Ala-Pro-Ala-Thr-Glu-Pro-Thr-Val-Asp-Ser-	-Val-Val-Thr-Pro-Glu-																				
	150	155	160	165	169																		
Bovine	-Ala ^b -SerP-Pro-Glu-Val-	-Ile-Glu-Ser-Pro-Pro-Glu-Ile-Asn-Thr-Val-Gln-Val-Thr-Ser-Thr-Ala-Val.OH																					
Zebu	-Ala ^a -SerP-Pro-Glu-Val,	,Ile, Glx, Ser, Pro, Pro, Glx, Ile, Asx, Thr/Val, Gln/Val, Thr, Ser, Thr/Ala-Val.OH																					
Water buffalo	-Ala-SerP-Ser-Glu-Val-	-Ile-Glu-Ser-Val-Pro-Glu-Thr-Asn-Thr-Ala-Gln-Val-Thr-Ser-Thr-Val-Val.OH																					
Goat	-Ala-SerP-Ser-Glu-Ser-	-Ile-Ala-Ser-Ala-Ser-Glu-Thr-Asn-Thr-Ala-Gln-Val-Thr-SerP-Thr-Glu-Val.OH																					
Sheep	-Ala-Ser*-Ser-Glu-Ser-	-Ile-Ala-Ser-Ala-Pro-Glu-Thr-Asn-Thr-Ala-Gln-Val-Thr-Ser*-Thr-Glu-Val.OH																					
Pig	-Ala-Ser-Ser-Glu-Phe-Leu-Ile-Thr-Ser-Ala-Pro-Glu-Thr-Thr-Thr-Val-Gln-Val-Thr-Ser-Pro-Val-Val.OH																						
Human	-Ala-Phe-Thr-Glu-Ser-Ile-Ile-Ser-Thr-Thr-Pro-Glu-Thr-Pro-Thr-Val-Ala, Val, Thr, Ser, Thr, Pro-Ala.OH																						

Fig.1. Comparison of the amino acid sequences (A) of bovine, caprine, and ovine para- κ -caseins (the C-terminal sequence of water buffalo para- κ -casein [11] is also shown), (B) of the caseinomacropetides from 7 species. Amino acid residues underlined are those which are different in homologous positions. Residue numbers are those of bovine κ -casein [4] whose sequence has been arbitrarily chosen as reference. From residues 1-94, only the alterations found in the primary structure are indicated. Original references for these sequences are listed in the text. *Notation:* (*) Sequences which were determined in our laboratory, ^a and ^b refer to both bovine [5] and zebu [6] genetic variants A and B respectively. * Probable location of phosphate groups in ovine CMP according to caprine sequence data [9]. (↓) Peptide bond split by proteases such as chymosin. (-) Indicates connected residues as determined experimentally. (.) Indicates that residues were positioned by homology. (/) Indicates fragments connected by homology.

2. Discussion

The amino acid sequences of the 7 CMPs and the amino acid replacements differentiating bovine, ovine and caprine para- κ -caseins are presented in fig.1 which also includes the C-terminal sequence of water buffalo para- κ -casein. For ease of comparison, the amino acid residues of the κ -casein polypeptide chain of the other species have been numbered similarly to those of bovine κ -casein [4].

When compared to their bovine B counterpart, which is made up of 64 amino acid residues, water buffalo, ovine, caprine, porcine and human CMPs differ by 10, 16, 18, 25 and 28 amino acid substitutions respectively, most of them being non conservative. Furthermore, all but water buffalo CMPs contain additional amino acid residues.

Ovine and caprine CMPs have in common the extra dipeptide Val-His inserted between positions 131 and 132. In this connection, it is noteworthy that in both species [13,14] the β -casein polypeptide chain lacks a dipeptide, either Pro179-Tyr180 or Tyr180-Pro181, in relation to its bovine counterpart [15]. Since the 3 structural genes responsible for the synthesis of the 3 main bovine caseins are known to be grouped, most likely in the order α_{s1} - β - κ as proposed by Grosclaude et al. [16], it is possible that the deletion and insertion of a dipeptide in ovine and caprine β - and κ -caseins respectively are the result of a translocation. A block of 6 base pairs from the structural gene of β -casein may have been mistakenly shifted to that of κ -casein during crossing over, before the divergence of sheep and goat.

Assuming that human CMP very likely contains 65 amino acid residues [12], the extra residue may be located (as well as one of the 7 additional residues of its porcine counterpart) between positions 152 and 153 according to the chosen alignment of the sequences. The remaining 6 extra residues of porcine CMP are obviously inserted between positions 142 and 143 and have very likely arisen from the duplication of the DNA fragment coding for the 6 preceding amino acid residues 137 to 142.

Variation of the primary structure of the CMP has been estimated by calculating the percent differences between the available amino acid sequences of the 7 species as shown in table 1A. The results indicate a rather high evolutionary rate of the primary structure.

There are differences between the CMPs of even closely related species such as goat and sheep. It was interesting to compare the CMPs data with those related to peptides released in the clotting of blood. Table 1B presents the percent differences between the amino acid sequences of fibrinopeptides A and B, the negatively charged N-terminal fragments released by thrombin from the α and β chains of fibrinogen respectively, during the terminal stage of the blood-clotting process. Comparison of tables 1A and 1B reveals that the percent differences between CMPs and fibrinopeptides A respectively, are very similar.

Like fibrinopeptides whose rate of change is among the most rapid so far observed [17], the CMP seems to have evolved very rapidly. In contrast, the primary structure of para- κ -casein appears to be less variable according to the data of table 1C. The percent difference between bovine and ovine CMPs is for example about three times higher than that between para- κ -caseins from the same species. However, this conclusion is a tentative one since the para- κ -casein sequences of only 3 species have been determined.

Comparison of amino acid sequences presented in fig.1 reveals the uneven distribution of the mutation sites. This is best illustrated by a schematic representation of the para- κ -casein C-terminus and the CMP region of the κ -casein molecule from several species (fig.2). The number of different amino acid residues as well as the gaps delineate a rather variable area roughly corresponding to the last two thirds of the CMP polypeptide chain whereas its N-terminal third and the C-terminal portion of para- κ -casein are almost invariable.

It must be noted that the constant region of the κ -casein molecule (residues 97-116) is the primary target of proteolytic enzymes in the initial stage of the milk-clotting process. It contains the peptide bond 105-106 which is extremely labile to physical, chemical and enzymatic attack and whose breakage initiates the clotting of milk. This bond is readily accessible to chemical probes and enzymes [18] but is not intrinsically sensitive to them. Its unusual lability is dictated by the nature and the sequence of the neighbouring amino acid residues as shown by studies of synthetic substrates of chymosin [19-26] simulating the amino acid sequence around the sensitive bond.

The integrity of the region containing the peptide

bond 105–106 is clearly responsible for its properties of accessibility and high susceptibility to proteolytic enzymes. The remarkable conservation of the primary structure of this area during the course of evolution strongly suggests that the limited proteolysis of κ -casein and hence the coagulation of milk are of prime biological importance. The primary function of the milk-clotting process, triggered by proteolytic

enzymes in the stomach of young mammals, is likely to be nutritional. Aggregation of micelles increases the retention time of caseins in the digestive tract, allowing for better assimilation of these proteins. In addition either the CMP or a fragment of it, say the N-terminal portion, may, in itself, be physiologically active. An anti-gastrin effect of the CMP has been reported by Chernikov et al. [27]. Furthermore,

Table 2
Amino acid composition, phosphate content and average hydrophobicity of bovine, water buffalo, caprine, ovine, porcine and human caseinomacropeptides

Amino acid	Bovine A / B mole amino acid / mole peptide	Water buffalo	Goat	Sheep	Pig	Human
Aspartic acid	2 / 1	1	2	3	1	2
Asparagine	3 ^a	4	5	4	4	1
Threonine	12 / 11	13 / 12	11	10	10	14
Serine	6	6	8	7	5	4
Glutamic acid	8	8	7	7	6	5
Glutamine	2	2	2	2	2	1
Proline	8	6	6	7	10	11
Glycine	1	—	—	—	—	—
Alanine	5 / 6	5	9	10	9	7
Valine	6	8	5 / 6	6	10	6
Methionine	1	1	1	1	—	—
Isoleucine	6 / 7	6 / 7	6 / 5	5	9	10
Leucine	1	1	—	—	1	—
Phenylalanine	—	—	—	—	1	1
Histidine	—	—	1	1	—	—
Lysine	3	3	3	3	3	3
Total	64	64	66	66	71	65
Phosphate group	1	1	2	2		
Carbohydrate (g/100)	11.6 ^a		9.8 ^b	5.5 ^b		55.2 ^a
H ϕ (kJ/residue)	4.31/4.52	4.23/4.39	3.68	3.85	5.32	5.60
Percent ^c						
Hydroxy amino acids	28 / 27	30 / 28	29	26	21	28
Proline	12	9	9	11	14	17
Dicarboxylic amino acids	16 / 14	14	14	15	10	11

^a Malpress and Seid-Akhavan [32]

^b Alais and Jollès [33]

^c The numbers have been rounded off to the nearest integer

A and B refer to genetic variants [5]

The amino acid compositions were deduced from the sequence data shown in fig. 1. The average hydrophobicity (H ϕ) was calculated according to Bigelow's scale [31].

whole casein seems to stimulate either the biosynthesis or the secretion of chymosin in the preruminant calf velle [28]. It is attractive to speculate that the CMP or one of its fragments may be involved in the stimulation of the cells producing chymosin. If this were the case, since chymosin is the main enzyme which is responsible for the release of the CMP from κ -casein in the calf abomasum, we have an example of positive feed-back.

Despite the differences observed between the amino acid sequences, the CMPs studied present some common features as shown in table 2. They are relatively rich in proline (9–17%) and dicarboxylic amino acids (11–16%) and their content in hydroxyamino acids is very high (21–30%). Dicarboxylic amino acids together with hydroxyamino acids linked to either carbohydrate moieties including sialic acid or phosphate groups are responsible for the high negative net charge of the CMP. In this connection, it must be pointed out that all the phosphorylated hydroxyamino acid residues are located in a tripeptide sequence Ser–X–Glu (X being any amino acid residue) which may be the 'code sequence' recognized by a specific phosphoryl kinase of the mammary gland as we postulated earlier [29,30]. The conservation of the acidic character of the CMP during the course of evolution suggests that electrical repulsive forces may play a major role in the stabilizing properties of κ -casein.

According to data of table 2, the CMPs of the 7 species could be classified in 2 groups, group 1 being characterized by a higher content of dicarboxylic amino acids and a relatively low 'average hydrophobicity' and carbohydrate content. Bovine, zebu, water buffalo, ovine and caprine CMPs fall into this group. On the other hand, members of group 2, porcine

and human CMPs, contain more proline but less dicarboxylic amino acids and have a much higher 'average hydrophobicity' and carbohydrate content. It would therefore be interesting to investigate if there is any difference in the behaviour of κ -caseins from the 2 groups. Differences in phosphate content have not been taken into account: since as the porcine and human CMPs were prepared from a κ -casein fraction [12], the lack of phosphate group is not necessarily a feature of whole κ -casein.

In a previous paper [11], we have discussed the relationship which exists between the genetic 'electrophoretic' variants A and B found in bovine, zebu and yak κ -caseins, the 3 species of the *Bos* genus. Two apparently related substitutions Thr136 (κ A) / Ile (κ B) and Asp148 (κ A) / Ala (κ B) have been identified in both bovine [5] and zebu [6] CMPs and may exist in yak CMP. The comparison of the CMP sequences from the 7 species studied so far suggests that both variants A and B might have been derived from an inferred ancestral variant Thr136–Ala148 by one point mutation, as shown in fig.3. This might have occurred before the divergence of the species of the *Bos* genus. This being the case, it is likely that the wild variant Thr136–Ala148 will be found in the populations of the *Bos* genus, among individuals of the electrophoretic type B.

3. Conclusion

During the course of evolution, κ -casein, the casein micelle stabilizer, appears to have undergone extensive sequence changes including either insertions or deletions. Mutations have mostly affected the C-terminal third of the polypeptide chain (CMP)

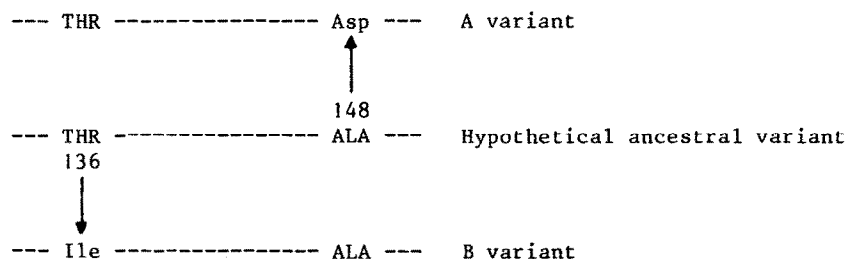


Fig.3. Hypothetical mutational pathways accounting for both bovine and zebu A and B genetic variants of κ -casein.

whose removal, through enzymatic cleavage, induces the milk-clotting process.

The rate of change of the CMP, similar to that of fibrinopeptide A, is among the most rapid so far observed, making this glycopeptide especially suitable for phylogenetic studies among mammals. Nevertheless, the overall acidic character of the CMP is preserved, suggesting that negative electrical forces are involved in the stabilizing properties of κ -casein. It must be pointed out however that this is not due to the presence of the almost invariable N-terminal portion.

The striking conservation of the primary structure around the protease-sensitive peptide bond 105–106, whose unusual lability depends on the integrity of the nearby sequence, stresses the biological importance of the release of the CMP and the subsequent clotting of micelles which normally occurs in the digestive tract. The biological significance of this process, which has previously given rise to controversy, should be worth extensively reinvestigating.

The full implications of the preceding observations are still limited owing to the restricted number of species studied so far, some of them being closely related. Added to this, only fragmentary sequence data were available. It is clear that the study of the chemical evolution of κ -casein in various mammals is a promising way to gain valuable information concerning this protein. In particular, it would be interesting to study κ -caseins from primitive mammals such as monotremes and marsupials.

Acknowledgements

We are grateful to R. E. Thompson for revision of the English text.

References

- [1] Jollès, J., Alais, C. and Jollès, P. (1970) *Helv. Chim. Acta* 53, 1918–1926.
- [2] Mercier, J-C., Uro, J., Ribadeau-Dumas, B. and Grosclaude, F. (1972) *Eur. J. Biochem.* 27, 535–547.
- [3] Brignon, G., Mercier, J-C., Ribadeau-Dumas, B. and Das, B. C. (1972) *FEBS Lett.* 27, 301–305.
- [4] Mercier, J-C., Brignon, G. and Ribadeau-Dumas, B. (1973) *Eur. J. Biochem.* 35, 222–235.
- [5] Grosclaude, F., Mahé, M-F., Mercier, J-C. and Ribadeau-Dumas, B. (1972) *Ann. Génét. Sél. Anim.* 4 (4) 515–521.
- [6] Grosclaude, F., Mahé, M-F. and Mercier, J-C. (1974) *Ann. Génét. Sél. Anim.* 6 (3) 305–329.
- [7] Jollès, J., Schoentgen, F., Hermann, J., Alais, C. and Jollès, P. *Eur. J. Biochem.* (1974) 46, 127–132.
- [8] Jollès, J., Fiat, A-M., Schoentgen, F., Alais, C. and Jollès, P. *Biochim. Biophys. Acta* (1974) 365, 335–343.
- [9] Mercier, J-C., Addeo, F. and Pélissier, J-P. (1976) *Biochimie*, for publication.
- [10] Mercier, J-C., Addeo, F. and Pélissier, J-P. unpublished results.
- [11] Addeo, F. and Mercier, J-C. (1976) *FEBS Lett.* 63, 255–259.
- [12] Chobert, J-M., Mercier, J-C. Bahy, C. and Hazé, G. (1976) *FEBS Lett.* 72, 173–178.
- [13] Richardson, B. et al. unpublished results.
- [14] Pélissier, J-P., Uro, J. et al. unpublished results.
- [15] Ribadeau-Dumas, B., Brignon, G., Grosclaude, F. and Mercier, J-C. (1972) *Eur. J. Biochem.* 25, 505–514.
- [16] Grosclaude, F., Mercier, J-C. and Ribadeau-Dumas, B. (1973) *Neth. Milk Dairy J.* 27, 328–340.
- [17] Dayhoff, M. O. (1972) in: *Atlas of Protein Sequence and Structure*, Vol 5, pp. D87–97, National Biomedical Research Foundation, Post Office Box 629 Silver Spring, Maryland 20901, USA.
- [18] Thomas, M. A. W. (1973) *Neth. Milk Dairy J.* 27, 273.
- [19] Hill, R. D. (1968) *Biochem. Biophys. Res. Commun.* 33, 659–663.
- [20] Hill, R. D. (1969) *J. Dairy Res.* 32, 193–201.
- [21] Schattenkerk, C., Holtkamp, I., Helsing, J. G. M., Kerling, K. F. T. and Havinga, F. (1971) *Rec. Trav. Chim. Pays-Bas* 90, 1320–1322.
- [22] Polzhofer, K. P. (1972) *Tetrahedron* 28, 855–865.
- [23] Raymond, M. N., Bricas, E., Garnier, J., Cilliano, S., Blasnic, M., Chaix, A. and Lefrancier, P. (1972) *Biochimie* 54, 145–154.
- [24] Raymond, M. N., Bricas, E. and Mercier, J-C. (1973) *Neth. Milk Dairy J.* 27, 298–303.
- [25] Schattenkerk, C. and Kerling, K. F. T. (1973) *Neth. Milk Dairy J.* 27, 286–287.
- [26] Wissner, S., Van Rooijen, P. J., Schattenkerk, C. and Kerling, K. E. T. (1976) *Biochem. Biophys. Acta* 438, 265–272.
- [27] Chernikov, M. P., Nikolskaya, G. V., Stan, E. Ya., Shlygin, G. K. and Vasilevskaya, L. S. (1974) *Voprosy Pitania* 2, 21–25.
- [28] Garnot, P., Vallès, E., Thapon, J-L., Toullec, R., Tomassone, R. and Ribadeau-Dumas, B. (1973) *J. Dairy Res.* 41, 19–23.
- [29] Mercier, J-C., Grosclaude, F. and Ribadeau-Dumas, B. (1971) *Eur. J. Biochem.* 23, 41–51.
- [30] Mercier, J-C., Grosclaude, F. and Ribadeau-Dumas, B. (1972) *Milchwissenschaft* 27 (7), 402–408.
- [31] Bigelow, C. C. (1967) *J. Theor. Biol.* 16, 187.
- [32] Malpress, F. H. and Seid-Akhavan, M. (1966) *Biochem. J.* 101, 764–773.
- [33] Alais, C. and Jollès, P. (1961) *Biochim. Biophys. Acta* 51, 315–322.