Volume 10 Number 8 1982 Nucleic Acids Research

Mouse whey acidic protein is a novel member of the family of 'four-disulfide core' proteins

Lothar G.Hennighausen and Albrecht E.Sippel

Institut für Genetik, University of Cologne, Weyertal 121, 5000 Köln 41, FRG

Received 22 February 1982; Revised and Accepted 23 March 1982

ABSTRACT

Unlike in other mammalian species, the major whey protein in mouse is not α-lactalbumin, but a cysteine rich,acidic protein with a molecular weight of 14.0 kDa. We have deduced the amino acid sequence of this mouse acidic whey protein from the nucleotide sequence of cloned cDNA. The positions of the half cysteines suggest that mouse whey acidic protein (WAP) is a two domain protein, very similar in structure to the plant lectin wheat germ agglutinin and the hypothalamic carrier protein neurophysin.

INTRODUCTION

The mammary gland is a highly specialized organ synthesizing milk proteins in a hormonally controlled manner during the lactational period. Up to date the milk proteins of several species have been characterized (1-8). The milk of ruminants (1-4), guinea pig (5,6) and rat (1,7) consists of only very few major protein species. The protein pattern of mouse milk is more complex (8), indicating a rapid evolutionary change. In contrast to other species (1,2,5,6) the major whey protein in mouse is neither α-lactalbumin nor β-lactoglobulin, but an acidic, cysteine rich protein with a molecular weight of 14.0 kDa (9,10). This novel protein was called whey acidic protein (WAP) by Piletz et al. (10). We have isolated, characterized and cloned the mRNA specific for mouse WAP (8). In this communication we report its amino acid sequence as deduced from the nucleotide sequence of cloned cDNA. On the basis of X-ray diffraction studies Drenth et al. (11) proposed a new grouping of small cysteine-rich proteins. Common to these proteins is their overall folding pattern. They consist of a core of eight similarly positioned and equally linked cysteines from which peptide loops of varying length are protruding. Among those proteins are snake venom neurotoxins and wheat germ agglutinin (11). According to the cysteine pattern the same structure was predicted for snake venom cardiotoxins, hevein, ragweed pollen allergen Ra5

© IRL Press Limited, 1 Falconberg Court, London W1V 5FG, U.K.
0305-1048/82/1008-2677$ 2.00/0

(11) and neurophysin (12).

Mouse WAP contains a cysteine pattern very similar to that of wheat germ agglutinin (11) and neurophysin (12) and we therefore suggest that it is a member of the 'four-disulfide core family' of small cysteine rich proteins.

MATERIALS AND METHODS

MATERIALS

Radiochemicals were obtained from Amersham. Restriction endonucleases, E. coli DNA polymerase I (large fragment) and deoxynucleotide triphosphates were from Boehringer. Dimethylsulfate, hydraziniumhydroxide and piperidine were supplied by Merck.

Methods

Mouse WAP specific mRNA has been isolated, characterized and cloned in our laboratory as described previously (8). In order to simplify the detection of recombinant plasmids we used a host vector system which allowed to screen in one step by a combined antibiotic resistance and colour indication method (13). Recombinant plasmid DNA was isolated according to Birnboim & Doly (14). After digestion of the plasmid DNA with restriction endonucleases as outlined in Fig. 1, the DNA fragments were labelled at one end directly in the restriction buffer by a 'fill in' reaction, using a twofold molar excess of the appropriate α-^{32}P dNTP and 4U E. coli DNA polymerase I (large fragment). After 10 min incubation at 15^{o}C the DNA was extracted with chloroform:isoamylalcohol (24:1), precipitated with ethanol and separated according to size on a 5% Polyacrylamide gel (8). After electroelution the DNA was sequenced according to Maxam & Gilbert (15).

RESULTS AND DISCUSSION

Mouse milk (8) has a more complex protein pattern than the milk of other mammalian species (1-7). Beside additional caseins it contains a major acidic whey protein which is significantly more abundant than α-lactalbumin (9,10). This whey acidic protein (WAP) with a molecular weight of 14.0 kDa has been isolated only from mouse and rat milk (9,10,16). In order to determine the primary structure of mouse WAP and to get a highly specific hybridization probe for the isolation of its gene, we have isolated, characterized and cloned the respective mRNA (8). As a prerequisite for the determination of the nucleotide sequence of the cloned cDNA corres-

ponding to mouse WAP, the restriction endonuclease cleavage maps of the plasmids pWAP1 and pWAP2 (8) were established. The cloned inserts of pWAP1 and pWAP2 cover an overlapping stretch of 600 nucleotides of mouse WAP mRNA. The mRNA size had previously been determined by RNA blot analysis to be 620 nucleotides (8). Fig. 1 shows the restriction endonuclease cleavage maps of pWAP1 and pWAP2 as well as the strategy of the sequence analysis. Including a poly(A) tail of 53 deoxyadenosine monophosphates a total of 600 bp from cloned mouse WAP specific cDNA have been determined (Fig. 2).

The amino acid sequence of the preprotein (Fig. 2) was deduced from the nucleotide sequence of the longest open reading frame. With the most likely start of translation at the first sequenced AUG, the mRNA contains a coding region for 134 amino acids. For 4 reasons we assume the first sequenced AUG being the start codon for translation of the preprotein:

1) The molecular weight of the preprotein, as calculated from the mRNA derived amino acid composition correlates well with that of the in vitro translation product (8).
2) The cloned cDNA lacks only about 20 nucleotides corresponding to the 5'-end of the mRNA. This makes it unlikely that an AUG lying further upstream might be used for translational initiation.
3) The first 19 amino acids of the deduced protein sequence form a hydrophobic signal peptide, typical for the amino terminal extensions of

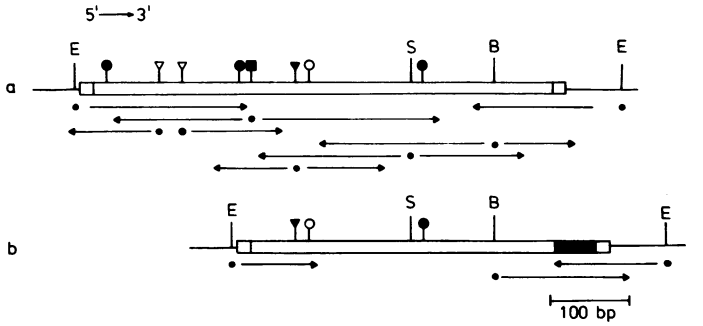

Figure 1

Restriction endonuclease cleavage maps and sequence analysis strategy of the cDNA inserts of pWAP1 (a) and pWAP2 (b). Following symbols were used; E: EcoRI, B: BamHI, S: SalI, ● : HpaII, ○ : Sau3A, ▼ : HinfI, ▽ : DdeI, ■ : Ava II. Black dots represent the labelled endonuclease restriction sites and arrows indicate the direction and length of determined sequences. The terminal open boxes represent the oligo (dG-dC)-tracks, and the black box represents the poly(A)-track of the mRNA template used in cDNA synthesis.

```
                             -19                                                                                  -1  +1
protein                      Met Arg Cys Leu Ile Ser Leu Val Leu Gly Leu Leu Ala Leu Glu Val Ala Leu Ala GlN AsN
pWAP1    CCTGACACCGGTACC AGT CGT TGC CTC ATC AGC CTT GTT CTT GGC CTG CTG GCC CTG GAG GTG GCC CTC GCT CAG AAC

                                     +10                                     +20
         Leu Glu Glu GlN Val Phe AsN Ser Val GlN Ser Met Phe Pro Lys Ala Ser Pro Ile Glu Gly Thr Glu Cys Ile
         CTA GAG GAA CAA GTC TTC AAC TCA GTT CAG TCC ATG TTC CCA AAA GCC AGC CCC ATT GAG GGC ACA GAG TGT ATC

                 +30                                     +40                                     +50
         Ile Cys GlN Thr AsN Glu Glu Cys Ala GlN AsN Ala Met Cys Cys Pro Gly Ser Cys Gly Arg Thr Arg Lys Thr
         ATC TGC CAA ACC AAC GAG GAG TGT GCC CAG AAT GCC ATG TGC TGT CCC CGT TCC TGT GGT AGG ACC CGC AAA ACT

                                     +60                                     +70
         Pro Val AsN Ile Gly Val Pro Lys Ala Gly Phe Cys Pro Trp AsN Leu Leu GlN Thr Ile Ser Ser Thr Gly Pro
         CCT GTG AAC ATT GGT GTT CCG AAA GCT GGC TTC TGC CCT TGG AAT CTA CTC CAA ACG ATC AGT AGT ACC GGG CCG

                 +80                                     +90                                     +100
         Cys Pro Met GlN Ile Glu Cys Ser Ser Asp Arg Glu Cys Ser Gly AsN Met Lys Cys Cys AsN Val Asp Cys Val
         TGC CCA ATG CAG ATA GAA TGC TCC AGC GAC CGT GAG TGT TCT GGC AAC ATG AAA TGC TGC AAC GTC GAC TGT GTC

                                     +110
         Met Thr Cys Thr Pro Pro Val Pro Val Ile Thr Leu GlN xxx
         ATG ACA TGT ACA CCC CCA GTG CCA GTG ATA ACC CTT CAG TGA GCAGCCGGCCCTGGCCATCCCTGCCTGTCGGGAGCGACCAGCC

         CAAGTCTATACAGCAAGACCCTTCACTCGTGGATCCAGAGAGAACATAATGCTTTCAATATGCTAATAAAAATCCATTTAGCTTTA (n=53)
```

Figure 2

Nucleotide sequence of the cloned cDNA specific for mouse whey acidic protein (WAP) and its deduced amino acid sequence.

secretory proteins (17-19).

4) The start codon on the mRNA for whey acidic preprotein of the rat is located at a homologous position (J.M. Rosen, pesonal communication, 22).

As we have shown earlier (8), the in vitro synthesized mouse whey acidic preprotein has an apparent molecular weight of 15.2 kDa which is shifted in the presence of dog pancreatic microsomal membranes to a molecular weight of 13.7 kDa, indicating the cotranslational processing of a signal peptide sequence. The primary structure of the signal peptides of ovine caseins (20) and of mouse ε-casein (21) has been determined. We observed a strong conservation of the signal peptides even between species (21). The 4 amino acids proximal to the clipping point of the signal peptide in the ovine calcium sensitive caseins and in mouse ε-caseins are identical to amino acid -1 to -4 in mouse whey acidic preprotein. Therefore we assume that the peptide bond following Ala No. -1 (Fig. 2) is clipped during cotranslational membrane transfer. For the rat whey acidic preprotein, which has a very similar sequence to the corresponding mouse protein (22), the clipping point of the signal peptide recently has been localized distal to Ala No. -1 by determination of the N-terminal sequence of the mature protein (K. E. Ebner, personal communication).

Table 1 shows that the amino acid composition of mouse WAP as derived from the nucleotide sequence of cloned cDNA (Fig. 2), correlates well with the amino acid composition of WAP-A and WAP-B, as determined from two allelic

Table 1

Amino acid composition of mouse whey acidic protein

	MOUSE WAP[1]	MOUSE WAP-A[2]	MOUSE WAP-B[2]
	mol amino acids/mol WAP		
Gly	7	7.8	7.8
Ala	4	4.5	4.5
Leu	4	4.4	4.5
Ile	7	6.5	6.5
Val	8	8.3	7.8
Met	5	7.8	7.9
Ser	9	8.7	9.1
Thr	9	8.2	8.1
Tyr	0	0.0	0.0
Trp	1	not done	not done
Phe	3	3.2	3.2
Glu	6	16.2	16.1
GlN	8	-	-
Asp	2	10.9	10.8
AsN	8	-	-
Cys	14	7.5	6.8
Pro	11	11.8	11.7
His	0	0.0	0.0
Lys	4	4.7	4.5
Arg	3	3.0	3.9

[1] as calculated from the amino acid sequence of the whey acidic protein (WAP) which was deduced from the nucleotide sequence of cloned cDNA;

[2] as determined from authentic WAP by Piletz et al. (10). The figures are corrected for 115 amino acids per molecule WAP.

whey proteins by Piletz et al. (10,23). WAP-B has one less cysteine and one more arginine than WAP-A. Due to deamination occurring during total amino acid analysis of the protein, all glutamine and asparagine residues had been determined as glutamic and aspartic amino acids respectively (10). The discrepancy in the cysteine content can be explained since in the total amino acid analysis a stable derivative of cysteine was not prepared by carboxymethylation or performic acid oxidation (10).

The cysteine pattern of mouse WAP (Fig. 3) resembles that of other small cysteine rich proteins (11,12) and suggests a two domain protein structure. In mouse WAP there are 5 cysteines in each half of the polypeptide chain which match exactely in their distances to each other. The sixth cysteine in the C-terminal domain matches an arginine in the N-terminal domain. On the mRNA

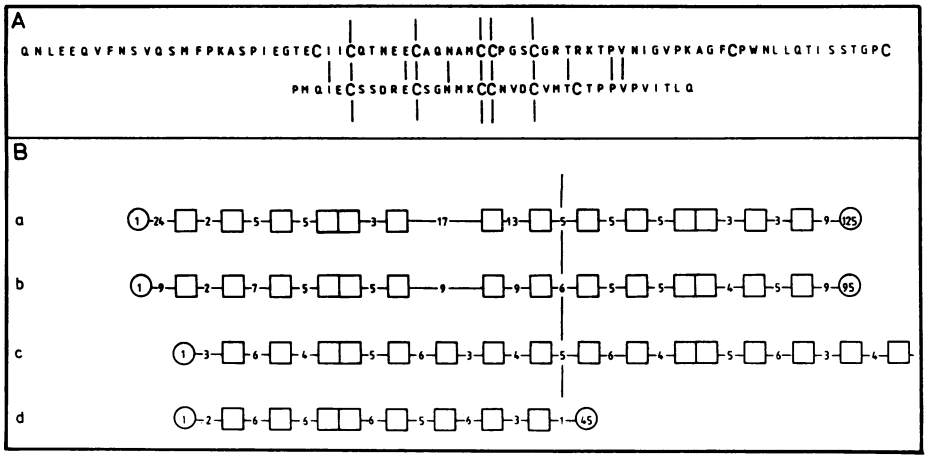

Figure 3

A) Primary structure comparison of the first and second part of mouse WAP. B) Cysteine pattern of (a) mouse WAP, (b) neurophysin, (c) two domains of wheat germ agglutinin and (d) ragweed pollen allergen Ra5. Boxes indicate the positions of the half-cysteines, numbers indicate the amino acid distances between them and numbers in the circles indicate the first and last amino acid in the chain.

level this amino acid change could be the result of a replacement substitution in the first position of the codon. According to its cysteine pattern mouse WAP very well fits into the family of proteins which contain a 'four-disulfide core' (11,12,Fig. 3B). Among those proteins are snake venom neurotoxins, wheat germ agglutinin, ragweed pollen allergen Ra5 and neurophysin (Fig. 3B). Within this protein family mouse WAP is more closely related to the neurophysins (12) than to the small plant proteins (11,Fig. 3B). There are two major differences between the cysteine clusters of mouse WAP and neurophysin (12) on the one hand and the plant proteins on the other hand (11,Fig. 3B). First, in mouse WAP and in neurophysin one 'standard' cysteine in the first cysteine cluster is missing and one additional cysteine residue has been introduced in the N-terminal part of the protein at position 26 and at position 11 respectively (Fig. 3B). This nonequivalence in the cysteine pattern does not need to cause a significant deviation from the toxin-agglutinin fold (11,12), but needs a reorganization of the disulfide bonds. Second, the C-terminal domains of mouse WAP and neurophysin lack two cysteine residues. In wheat germ agglutinin those two neighbouring cysteines form a disulfide bond which leads to the formation of a small closed peptide loop (11). The

cysteine pattern of mouse WAP and neurophysin (12) suggests a very similar domain structure in the two proteins. The distances between comparable cysteine residues in the second domain of neurophysin and mouse WAP are almost identical. The distances between the last three cysteines in the first domain, however, are more different. Under the assumption of an equal disulfide structure in both proteins, this would indicate that in mouse WAP two loops protruding from the 'four-disulfide core' are larger than the respective loops in neurophysin. These variations in loop size are common in the proteins of the 'four-disulfide core' family as can be seen by comparison of the 3D-structures of wheat germ agglutinin (24) and the snake neurotoxin erabutoxin (25). Both proteins have an almost identical cystine-core, but the protruding loops are of different size (11).

Neurophysins are hypothalamic carrier proteins for the nonapeptides oxytocin and vasopressin (26). The mammary gland is one of the main target organs for oxytocin. The presence of material crossreacting with antineurophysin immune sera has been demonstrated in oxytocin target tissues (mammary gland, uterus) of pig, guinea pig and human (27,28), but thus far only for the uterus it has been shown, that these proteins are synthesized within the tissue rather than transported to it (29). Mouse WAP and neurophysin have an almost identical cysteine pattern, and could therefore be related in 3D structure. On the level of amino acid sequence however, there is no homology between the two proteins. This makes it unlikely that WAP is identical to the immunologically detected neurophysin-like material found in the lactating mammary gland.

Further studies are required to answer the question whether the members of the family containing this 'four-disulfide core' are evolutionaryly related or else have arrived independently at a similar structure, possibly in order to stabilize domains otherwise too small to form a sufficient hydrophobic core.

ACKNOWLEDGEMENTS

We thank Kurt Stüber for help with computer searches and Jan Drenth for valuable discussions. The work was supported by a grant of the Deutsche Forschungsgemeinschaft (SFB74) to A.E.S. and by a fellowship of the Studienstiftung des Deutschen Volkes to L.G.H. This paper was submitted in partial fulfillment of the graduate requirements for L.G.H.

REFERENCES

1) Jeness, R. (1974) in Lactation, Larson, B.L. & Smith, V.R. Eds., Vol. 3, pp. 3-107, Academic Press, London und New York.
2) McKenzie, H.A. (1971) "Milk proteins: Chemistry and Molecular Biology" Vol. 2, Academic Press, New York.
3) Taborski, G. (1974) Adv. Prot. Chem. 28, 1-210.
4) Brignon, G., Ribadeau-Dumas, B., Mercier, J-C. & Pellisier, J-P. (1977) FEBS Lett. 76, 274-279.
5) Craig, R.K., Brown, P.A., Harrison, O.S., McIlreavy, D. & Campbell, P.N. (1976) Biochem. J., 160, 57-74.
6) Craig, R.K., McIlreavy, D. & Hall, R.L. (1978) Biochem. J., 173, 633-641.
7) Rosen, J.M., Woo, S.L.C. & Comstock, J.P. (1975) Biochemistry 14, 2895-2903.
8) Hennighausen, L.G. & Sippel, A.E., Eur. J. Biochem., in press.
9) Zamierowski, M.M. & Ebner, K.E. (1980) J. Immun. Methods 36, 211-220.
10) Piletz, J.E., Heinlen, M. & Ganshow, R.E. (1981) J. Biol. Chem. 256, 11509-11516.
11) Drenth, J., Low, B.W., Richardson, J.S. & Wright, C.S. (1980) J. Biol. Chem. 255, 2652-2655.
12) Drenth, J. (1981) J. Biol. Chem. 256, 2601-2602.
13) Rüther, U. (1980) Molec. Gen. Genet. 178, 475-477.
14) Birnboim, H.C. & Doly, J. (1976) Nucleic Acid Res. 76, 1513-1523.
15) Maxam, A.M. & Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
16) McKenzie, R.M. & Larson, B.L. (1978) J. Dairy Sci. 61, 723-728.
17) Blobel, G. & Dobberstein, B. (1975) J. Cell Biol. 67, 835-851.
18) Blobel, G. & Dobberstein, B. 61975) J. Cell Biol. 67, 852-862.
19) Davis, B.D. & Tai, P-C. (1980) Nature 283, 433-439.
20) Gaye, P., Gautron, J-P., Mercier, J-C. & Haze, G. (1977) Biochem. Biophys. Res. Commun. 79, 903-911.
21) Hennighausen, L.G., Steudle, A. & Sippel, A.E., submitted.
22) Hennighausen, L.G., Sippel, A.E., Hobbs, A.A., and Rosen, J.M., submitted.
23) Piletz, J.E. & Ganshow, R.E. (1981) Biochem. Genet. 19, 1023-1030.
24) Schubert-Wright, C. (1977) J. Mol. Biol. 111, 439-457.
25) Low, B.W., Preston, H.S., Sato, A., Rosen, L.S., Searl, J.E., Rudko, A.D. & Richardson, J.S. (1976) Proc. Natl. Acad. Sci. USA 73, 2991-2994.
26) Breslow, E. (1979) Ann. Rev. Biochem. 48, 251-274.
27) Ginsburg, M. & Jayasena, K. (1968) J. Physiol. 197, 65-76.
28) Legros, J-J., Louis, F., Gröschel-Stewart, U. & Franchimont, P. (1975) Ann. N.Y. Acad. Sci. 248, 157-171.
29) Sinding, C. & Bobinson, A.G. (1977) Metabolism 26, 1355-1370.