cDNA and protein sequence of bovine lactoferrin

P.E.Mead and J.W.Tweedie*
Department of Chemistry and Biochemistry, Massey University, Palmerston North, New Zealand

Submitted October 25, 1990 EMBL accession no. X54801

Lactoferrins belong to the transferrin family of iron binding proteins (1). This family is characterised by the capacity to reversibly bind iron with a high affinity ($K_{app} \sim 10^{20}$) (2). The primary sequence of several transferrins has been determined by protein and/or cDNA sequencing (ovotransferrin (3), mouse lactoferrin (4), human lactoferrin (5, 6), human serum transferrin (7, 8)). We report the primary sequence of bovine lactoferrin determined from a combination of cDNA and protein sequencing.

Putative cDNA clones for bovine lactoferrin were isolated from a cDNA library in pGEM, prepared from polyA$^+$ RNA isolated from involuting bovine mammary gland. The library was screened using a probe from the 3′-end of human lactoferrin (9). DNA was sequenced in m13 by the dideoxynucleotide chain termination method (10). The longest clone obtained (PM-7) extended 2152 nucleotides from the polyA tail to a position corresponding to amino acid 40 of the mature protein sequence. A second clone (PM-8) extended from nucleotide 79 to 792 of the cDNA. DNA sequence further toward the 5′-end of the mRNA was obtained by anchored polymerase chain reaction (PCR) of G-tailed first-strand cDNA using a 3′-primer derived from the cDNA sequence and oligo-dC as the 5′-primer. Clones isolated from the products of the PCR reaction extended the sequence a further 56 nucleotides to a position corresponding to amino acid 9 of the mature protein sequence.

Bovine lactoferrin was isolated from colostrum and cleaved with trypsin. The N-terminal sequences of the intact protein and of selected peptides were determined using an Applied Biosystems gas phase sequencer. The N-terminal sequence overlapped the 5′-end of the PCR product by 20 amino acids and extended to the 5′-end of PM-8. The sequence of these 20 amino acids determined by protein sequencing corresponded exactly to the sequence predicted by translation of the cDNA.

ACKNOWLEDGEMENTS

We wish to thank Dr Thomas A.Rado for providing the human cDNA clone & Dr G.G.Midwinter for assistance with protein sequencing.

REFERENCES

1. Brock,J.H. (1985) in *Metalloproteins Part II* (Harrison,P.M., ed) pp 183–262, Macmillan, London.
2. Aisen,P. and Listowsky,I. (1980) *Annu. Rev. Biochem.* **49**, 357–393.
3. Jeltsch,J.M. and Chambon,P. (1982) *Eur. J. Biochem.* **122**, 291–293.
4. Pentecost,B.T. and Teng,C.T. (1987) *J. Biol. Chem.* **262**, 10134–10139.
5. Metz-Boutigue,M.-H. *et al.* (1984) *Eur. J. Biochem.* **145**, 659–676.
6. Powell,M.J. and Ogden,J.E. (1990) *Nucl. Acids Res.* **18**, 4013.
7. MacGillivray,R.T.A. *et al.* (1983) *J. Biol. Chem.* **258**, 3543–3553.
8. Huggenvik,J.I. *et al.* (1987) *Endocrinology* **120**, 332–340.
9. Rado,T.A. *et al.* (1987) *Blood* **70**, 989–993.
10. Sanger,F. *et al.* (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.

```
1                           GGTGTACCATCTCCCAACCTGACTCCTTCAAATGCCCG
     A  P  R  K  N  V  R  W  C  T  I  S  Q  P  D  S  F  K  C  P
61   CGATGGCAGTGGAGGATGAAGAAGCTGGGTGCTCCCTCTATCACCTGTGTGAGGAGGGCC
     R  W  Q  W  R  M  K  K  L  G  A  P  S  I  T  C  V  R  R  A
121  TTTGCCTTGGAATGTATCCGGGCCATCGCGGAGAAAAAGGCGGATGCTGTGACCCTGGAT
     F  A  L  E  C  I  R  A  I  A  E  K  K  A  D  A  V  T  L  D
181  GGTGGCATGGTGTTTGAGGCGGGCCGGGACCCCTACAAACTGCGGCCAGTAGCAGCAGAG
     G  G  M  V  F  E  A  G  R  D  P  Y  K  L  R  P  V  A  A  E
241  ATCTATGGGACGAAAGAGTCTCCCCAAACCCACTATTATGCTGTGGCCGTCGTGAAGAAG
     I  Y  G  T  K  E  S  P  Q  T  H  Y  Y  A  V  A  V  V  K  K
301  GGCAGCAACTTTCAGCTGGACCAGCTGCAAGGCCGGAAGTCCTGCCATACGGGCCTTGGC
     G  S  N  F  Q  L  D  Q  L  Q  G  R  K  S  C  H  T  G  L  G
361  AGGTCCGCTGGGTGGATCATCCCTATGGGAATCCTTCGCCCGTACTTGAGCTGGACAGAG
     R  S  A  G  W  I  I  P  M  G  I  L  R  P  Y  L  S  W  T  E
421  TCACTCGAGCCCCTCCAGGGAGCTGTGGCTAAATTCTTCTCTGCCAGCTGTGTTCCCTGC
     S  L  E  P  L  Q  G  A  V  A  K  F  F  S  A  S  C  V  P  C
481  ATTGATAGACAAGCATACCCCAACCTGTGTCAACTGTGCAAGGGGGAGGGGGAGAACCAG
     I  D  R  Q  A  Y  P  N  L  C  Q  L  C  K  G  E  G  E  N  Q
541  TGTGCCTGCTCCTCCCGGGAACCATACTTCGGTTATTCTGGTGCCTTCAAGTGTCTGCAG
     C  A  C  S  S  R  E  P  Y  F  G  Y  S  G  A  F  K  C  L  Q
601  GACGGGGCTGGAGACGTGGCTTTTGTTAAAGAGACGACAGTGTTTGAGAACTTGCCAGAG
     D  G  A  G  D  V  A  F  V  K  E  T  T  V  F  E  N  L  P  E
661  AAGGCTGACAGGGACCAGTATGAGCTTCTCTGCCTGAACAACAGTCGGGCGCCAGTGGAT
     K  A  D  R  D  Q  Y  E  L  L  C  L  N  N  S  R  A  P  V  D
721  GCGTTCAAGGAGTGCCACCTGGCCCAGGTCCCTTCTCATGCTGTCGTGGCCCGAAGTGTG
     A  F  K  E  C  H  L  A  Q  V  P  S  H  A  V  V  A  R  S  V
781  GATGGCAAGGAAGACTTGATCTGGAAGCTTCTCAGCAAGGCGCAGGAGAAATTTGGAAAA
     D  G  K  E  D  L  I  W  K  L  L  S  K  A  Q  E  K  F  G  K
841  AACAAGTCTCGGAGCTTCCAGCTCTTTGGCTCTCCACCCGGCCAGAGGGACCTGCTGTTC
     N  K  S  R  S  F  Q  L  F  G  S  P  P  G  Q  R  D  L  L  F
901  AAAGACTCTGCTCTTGGGTTTTTGAGGATCCCCTCGAAGGTAGATTCGGCGCTGTACCTG
     K  D  S  A  L  G  F  L  R  I  P  S  K  V  D  S  A  L  Y  L
961  GGCTCCCGCTACTTGACCACCTTGAAGAACCTCAGGGAAACTGCGGAGGAGGTGAAGGCG
     G  S  R  Y  L  T  T  L  K  N  L  R  E  T  A  E  E  V  K  A
1021 CGGTACACCAGGGTCGTGTGGTGTGCCGTGGGACCTGAGGAGCAGAAGAAGTGCCAGCAG
     R  Y  T  R  V  V  W  C  A  V  G  P  E  E  Q  K  K  C  Q  Q
1081 TGGAGCCAGCAGAGCGGCCAGAACGTGACCTGTGCCACGGCGTCCACCACTGACGACTGC
     W  S  Q  Q  S  G  Q  N  V  T  C  A  T  A  S  T  T  D  D  C
1141 ATCGTCCTGGTGCTGAAAGGGGAAGCAGATGCCCTGAACTTGGATGGAGGATATATCTAC
     I  V  L  V  L  K  G  E  A  D  A  L  N  L  D  G  G  Y  I  Y
1201 ACTGCGGGCAAGTGTGGCCTGGTGCCTGTCCTGGCAGAGAACCGGAAATCCTCCAAACAC
     T  A  G  K  C  G  L  V  P  V  L  A  E  N  R  K  S  S  K  H
1261 AGTAGCCTAGATTGTGTGCTGAGACCAACGGAAGGGTACCTTGCCGTGGCAGTTGTCAAG
     S  S  L  D  C  V  L  R  P  T  E  G  Y  L  A  V  A  V  V  K
1321 AAAGCAAATGAGGGGCTCACATGGAATTCTCTGAAAGACAAGAAGTCGTGCCACACCGCC
     K  A  N  E  G  L  T  W  N  S  L  K  D  K  K  S  C  H  T  A
1381 GTGGACAGGACTGCAGGCTGGAACATCCCCATGGGCCTGATCGTCAACCAGACAGGCTCC
     V  D  R  T  A  G  W  N  I  P  M  G  L  I  V  N  Q  T  G  S
1441 TGCGCATTTGATGAATTCTTTAGTCAGAGCTGTGCCCCTGGGGCTGACCCGAAATCCAGA
     C  A  F  D  E  F  F  S  Q  S  C  A  P  G  A  D  P  K  S  R
1501 CTCTGTGCCTTGTGTGCTGGCGATGACCAGGGCCTGGACAAGTGTGTGCCCAACTCTAAG
     L  C  A  L  C  A  G  D  D  Q  G  L  D  K  C  V  P  N  S  K
1561 GAGAAGTACTATGGCTATACCGGGGCTTTCAGGTGCCTGGCTGAGGACGTTGGGGACGTT
     E  K  Y  Y  G  Y  T  G  A  F  R  C  L  A  E  D  V  G  D  V
1621 GCCTTTGTGAAAAACGACACAGTCTGGGAGAACACGAATGGAGAGAGCACTGCAGACTGG
     A  F  V  K  N  D  T  V  W  E  N  T  N  G  E  S  T  A  D  W
1681 GCTAAGAACTTGAATCGTGAGGACTTCAGGTTGCTCTGCCTCGATGGCACCAGGAAGCCT
     A  K  N  L  N  R  E  D  F  R  L  L  C  L  D  G  T  R  K  P
1741 GTGACGGAGGCTCAGAGCTGCCACCTGGCGGTGGCCCCGAATCACGCTGTGGTGTCTCGG
     V  T  E  A  Q  S  C  H  L  A  V  A  P  N  H  A  V  V  S  R
1801 AGCGATAGGGCAGCACACGTGAAACAGGTGCTGCTCCACCAGCAGGCTCTGTTTGGGAAA
     S  D  R  A  A  H  V  K  Q  V  L  L  H  Q  Q  A  L  F  G  K
1861 AATGGAAAAAACTGCCCGGACAAGTTTTGTTTGTTCAAATCTGAAACCAAAAACCTTCTG
     N  G  K  N  C  P  D  K  F  C  L  F  K  S  E  T  K  N  L  L
1921 TTCAATGACAACACTGAGTGTCTGGCCAAACTTGGAGGCAGACCAACGTATGAAGAATAT
     F  N  D  N  T  E  C  L  A  K  L  G  G  R  P  T  Y  E  E  Y
1981 TTGGGGACAGAGTATGTCACGGCCATTGCCAACCTGAAAAAATGCTCAACCTCCCCGCTT
     L  G  T  E  Y  V  T  A  I  A  N  L  K  K  C  S  T  S  P  L
2041 CTGGAAGCCTGCGCCTTCCTGACGAGGTAAAGCCTGCAAAGAAGCTAGCCTGCCTCCCTG
     L  E  A  C  A  F  L  T  R  Stop
2101 GGCCTCAGCTCCTCCCTGCTCTCAGCCCCAATCTCCAGGCGCGAGGACCTTCCTCTCCC
2161 TTCCTGAAGTCGGATTTTTGCCAAGCTCATCAGTATTTACAATTCCCTGCTGTCATTTTA
2221 GCAAGAAATAAAATTACAAATGCTGTTGATTTTCATTCCCTAAAAAAAA  2269
```

Figure 1. Nucleotide sequence of the cDNA for bovine lactoferrin from nucleotide 23 to 2269 together with the protein sequence from amino acid 1 of the mature protein.

Numbering is from the position in the cDNA corresponding to the first nucleotide in the codon for amino acid 1 in the mature protein sequence. Nucleotide sequences determined by anchored PCR (23 to 190) are indicated by a dotted underline. Clone PM-7 extends from nucleotide 117 to 2269. PM-8 extends from nucleotide 79 to 792.

Protein sequences determined by sequencing the N terminus of the mature protein (amino acids 1–28) and of the N-termini of isolated tryptic peptides are shown by underlining the appropriate amino acids in the protein sequence.

* To whom correspondence should be addressed