Complete nucleotide sequence of human mammary gland lactoferrin

Michael W.Rey, Shelley L.Woloshuk[1], Herman A.deBoer[1] and Frank R.Pieper[1]*
Genencor International, South San Francisco, CA 94080, USA and [1]GenePharming Europe B.V., Einsteinweg 5, 2333 CC Leiden, The Netherlands

Submitted July 26, 1990 EMBL accession no. X53961

Human lactoferrin (hLF) is a single chain glycoprotein that can bind two ferric ions along with two bicarbonate ions. It is secreted by exocrine glands (1) and by cells of polymorphonuclear neutrophil granulocytes (2), and is a major whey protein in human milk. hLF functions as part of a non-specific defense system. It inhibits the growth of a diverse spectrum of bacteria by chelation of the available iron in the medium making this essential metal inaccessible to invading iron-requiring microorganisms (3). This effect is blocked if the hLF protein is saturated with ferric ions.

The complete amino acid sequence has been determined (4), a crystallographic structure analysis has been reported (5) and a partial cDNA sequence for neutrophil hLF has been published (6). Using two oligomer probes derived from the neutrophil hLF cDNA sequence, a 2320 bp clone was isolated from a human mammary gland cDNA library.

With the exception of 5 basepairs (bp) at the 5′ end, this clone encodes the entire protein. The missing sequence was obtained by cloning the hLF genomic gene from a cosmid library and sequencing the 5′ region. The complete coding sequence is presented below. The cDNA encodes a protein with a signal peptide of 19 amino acids followed by a mature protein of 692 residues. The putative TATA box (underlined) is positioned 70 bp 5′ to the ATG. There is 99.7% agreement between the partial cDNA sequence for neutrophil hLF and the overlapping cDNA sequence from human mammary gland presented here. The deduced amino acid sequence is 97% identical to the amino acid sequence of Metz-Boutigue (4) and 98% identical to that of Anderson (5). A 35 bp region including the TATA box is 84% identical to the corresponding region of the human serum transferrin gene.

REFERENCES

1. Mason,D.Y. *et al.* (1978) *Journal of Clinical Pathology* **31**, 316–327.
2. Mason,P.L. *et al.* (1969) *J. Exp. Med.* **130**, 643–658.
3. Reiter,B. *et al.* (1975) *Immunol.* **28**, 83–95.
4. Metz-Boutigue,M.H. *et al.* (1984) *Eur. J. Biochem.* **1451**, 659–676.
5. Anderson *et al.* (1989) *J. Mol. Biol.* **209**, 711–734.
6. Rado,T.A. *et al.* (1987) *Blood* **70**, 989–993.

```
   1 GACTCCTAGG GGCTTGCAGA CCTAGTGGGA GAGAAAGAAC ATCGCAGCAG CCAGGCAGAA CCAGGACAGG TGAGGTGCAG GCTGGCTTTC CTCTCGCAGC
 101 GCGGTGTGGA GTCCTGTCCT GCCTCAGGGC TTTTCGGAGC CTGGATCCTC AAGGAACAAG TAGACCTGGC CGCGGGGAGT GGGGAGGGAA GGGGTGTCTA
 201 TTGGGCAACA GGGCGGCAAA GCCCTGAATA AAGGGGCGCA GGGCAGGCGC AAGTGCAGAG CCTTCGTTTG CCAAGTCGCC TCCAGACCGC AGAC ATG AAA CTT
 -19                                                                                                    M   K   L
 304 GTC TTC CTC GTC CTG CTG TTC CTC GGG GCC CTC GGA CTG TGT CTG GCT GGC CGT AGG AGA AGG AGT GTT CAG TGG TGC GCC GTA TCC
 -16 V   F   L   V   L   L   F   L   G   A   L   G   L   C   L   A   G   R   R   R   S   V   Q   W   C   A   V   S
 391 CAA CCC GAG GCC ACA AAA TGC TTC CAA TGG CAA AGG AAT ATG AGA AAA GTG CGT GGC CCT CCT GTC AGC TGC ATA AAG AGA GAC TCC
  14 Q   P   E   A   T   K   C   F   Q   W   Q   R   N   M   R   K   V   R   G   P   P   V   S   C   I   K   R   D   S
 478 CCC ATC CAG TGT ATC CAG GCC ATT GCG GAA AAC AGG GCC GAT GCT GTG ACC CTT GAT GGT GGT TTC ATA TAC GAG GCA GGC CTG GCC
  43 P   I   Q   C   I   Q   A   I   A   E   N   R   A   D   A   V   T   L   D   G   G   F   I   Y   E   A   G   L   A
 565 CCC TAC AAA CTG CGA CCT GTA GCG GCG GAA GTC TAC GGG ACC GAA AGA CAG CCA CGA ACT CAC TAT TAT GCC GTG GCT GTG GTG AAG
  72 P   Y   K   L   R   P   V   A   A   E   V   Y   G   T   E   R   Q   P   R   T   H   Y   Y   A   V   A   V   V   K
 652 AAG GGC GGC AGC TTT CAG CTG AAC GAA CTG CAA GGT CTG AAG TCC TGC CAC ACA GGC CTT CGC AGG ACC GCT GGA TGG AAT GTC CCT
 101 K   G   G   S   F   Q   L   N   E   L   Q   G   L   K   S   C   H   T   G   L   R   R   T   A   G   W   N   V   P
 739 ACA GGG ACA CTT CGT CCA TTC TTG AAT TGG ACG GGT CCA CCT GAG CCC ATT GAG GCA GCT GTG GCC AGG TTC TTC TCA GCC AGC TGT
 130 T   G   T   L   R   P   F   L   N   W   T   G   P   P   E   P   I   E   A   A   V   A   R   F   F   S   A   S   C
 826 GTT CCC GGT GCA GAT AAA GGA CAG TTC CCC AAC CTG TGT CGC CTG TGT GCG GGG ACA GGG GAA AAC AAA TGT GCC TTC TCC TCC CAG
 159 V   P   G   A   D   K   G   Q   F   P   N   L   C   R   L   C   A   G   T   G   E   N   K   C   A   F   S   S   Q
 913 GAA CCG TAC TTC AGC TAC TCT GGT GCC TTC AAG TGT CTG AGA GAC GGG GCT GGA GAC GTG GCT TTT ATC AGA GAG AGC ACA GTG TTT
 188 E   P   Y   F   S   Y   S   G   A   F   K   C   L   R   D   G   A   G   D   V   A   F   I   R   E   S   T   V   F
1000 GAG GAC CTG TCA GAC GAG GCT GAA AGG GAC GAG TAT GAG TTA CTC TGC CCA GAC AAC ACT CGG AAG CCA GTG GAC AAG TTC AAA GAC
 217 E   D   L   S   D   E   A   E   R   D   E   Y   E   L   L   C   P   D   N   T   R   K   P   V   D   K   F   K   D
1087 TGC CAT CTG GCC CGG GTC CCT TCT CAT GCC GTT GTG GCA CGA AGT GTG AAT GGC AAG GAG GAT GCC ATC TGG AAT CTT CTC CGC CAG
 246 C   H   L   A   R   V   P   S   H   A   V   V   A   R   S   V   N   G   K   E   D   A   I   W   N   L   L   R   Q
1174 GCA CAG GAA AAG TTT GGA AAG GAC AAG TCA CCG AAA TTC CAG CTC TTT GGC TCC CCT AGT GGG CAG AAA GAT CTG CTG TTC AAG GAC
 275 A   Q   E   K   F   G   K   D   K   S   P   K   F   Q   L   F   G   S   P   S   G   Q   K   D   L   L   F   K   D
1261 TCT GCC ATT GGG TTT TCG AGG GTG CCC CCG AGG ATA GAT TCT GGG CTG TAC CTT GGC TCC GGC TAC TTC ACT GCC ATC CAG AAC TTG
 304 S   A   I   G   F   S   R   V   P   P   R   I   D   S   G   L   Y   L   G   S   G   Y   F   T   A   I   Q   N   L
1348 AGG AAA AGT GAG GAG GAA GTG GCT GCC CGG CGT GCG CGG GTC GTG TGG TGT GCG GTG GGC GAG CAG GAG CTG CGC AAG TGT AAC CAG
 333 R   K   S   E   E   E   V   A   A   R   R   A   R   V   V   W   C   A   V   G   E   Q   E   L   R   K   C   N   Q
1435 TGG AGT GGC TTG AGC GAA GGC AGC GTG ACC TGC TCC TCG GCC TCC ACC ACA GAG GAC TGC ATC GCC CTG GTG CTG AAA GGA GAA GCT
 362 W   S   G   L   S   E   G   S   V   T   C   S   S   A   S   T   T   E   D   C   I   A   L   V   L   K   G   E   A
1522 GAT GCC ATG AGT TTG GAT GGA GGA TAT GTG TAC ACT GCA GGC AAA TGT GGT TTG GTG CCT GTC CTG GCA GAG AAC TAC AAA TCC CAA
 391 D   A   M   S   L   D   G   G   Y   V   Y   T   A   G   K   C   G   L   V   P   V   L   A   E   N   Y   K   S   Q
1609 CAA AGC AGT GAC CCT GAT CCT AAC TGT GTG GAT AGA CCT GTG GAA GGA TAT CTT GCT GTG GCG GTG GTT AGG AGA TCA GAC ACT AGC
 420 Q   S   S   D   P   D   P   N   C   V   D   R   P   V   E   G   Y   L   A   V   A   V   V   R   R   S   D   T   S
1696 CTT ACC TGG AAC TCT GTG AAA GGC AAG AAG TCC TGC CAC ACC GCC GTG GAC AGG ACT GCA GGC TGG AAT ATC CCC ATG GGC CTG CTC
 449 L   T   W   N   S   V   K   G   K   K   S   C   H   T   A   V   D   R   T   A   G   W   N   I   P   M   G   L   L
1783 TTC AAC CAG ACG GGC TCC TGC AAA TTT GAT GAA TAT TTC AGT CAA AGC TGT GCC CCT GGG TCT GAC CCG AGA TCT AAT CTC TGT GCT
 478 F   N   Q   T   G   S   C   K   F   D   E   Y   F   S   Q   S   C   A   P   G   S   D   P   R   S   N   L   C   A
1870 CTG TGT ATT GGC GAC GAG CAG GGT GAG AAT AAG TGC GTG CCC AAC AGC AAC GAG AGA TAC GGC TAC ACT GGG GCT TTC CGG TGC
 507 L   C   I   G   D   E   Q   G   E   N   K   C   V   P   N   S   N   E   R   Y   Y   G   Y   T   G   A   F   R   C
1957 CTG GCT GAG AAT GCT GGA GAC GTT GCA TTT GTG AAA GAT GTC ACT GTC TTG CAG AAC ACT GAT GGA AAT AAC AAT GAG GCA TGG GCT
 536 L   A   E   N   A   G   D   V   A   F   V   K   D   V   T   V   L   Q   N   T   D   G   N   N   N   E   A   W   A
2044 AAG GAT TTG AAG CTG GCA GAC TTT GCG CTG CTG TGC CTC GAT GGC AAA CGG AAG CCT GTG ACT GAG GCT AGA AGC TGC CAT CTT GCC
 565 K   D   L   K   L   A   D   F   A   L   L   C   L   D   G   K   R   K   P   V   T   E   A   R   S   C   H   L   A
2131 ATG GCC CCG AAT CAT GCC GTG GTG TCT CGG ATG GAT AAG GTG GAA CGC CTG AAA CAG GTG CTG CTC CAC CAA CAG GCT AAA TTT GGG
 594 M   A   P   N   H   A   V   V   S   R   M   D   K   V   E   R   L   K   Q   V   L   L   H   Q   Q   A   K   F   G
2218 AGA AAT GGA TCT GAC TGC CCG GAC AAG TTT TGC TTA TTC CAG TCT GAA ACC AAA AAC CTT CTG TTC AAT GAC AAC ACT GAG TGT CTG
 623 R   N   G   S   D   C   P   D   K   F   C   L   F   Q   S   E   T   K   N   L   L   F   N   D   N   T   E   C   L
2305 GCC AGA CTC CAT GGC AAA ACA ACA TAT GAA AAA TAT TTG GGA CCA CAG TAT GTC GCA GGC ATT ACT AAT CTG AAA AAG TGC TCA ACC
 652 A   R   L   H   G   K   T   T   Y   E   K   Y   L   G   P   Q   Y   V   A   G   I   T   N   L   K   K   C   S   T
2392 TCC CCC CTC CTG GAA GCC TGT GAA TTC CTC AGG AAG TAA  AACCGAAGAA GATGGCCCAG CTCCCCAAGA AAGCCTCAGC CATTCACTGC CCCCAGCTCT
 681 S   P   L   L   E   A   C   E   F   L   R   K   O
2491 TCTCCCCAGG TGTGTTGGGG CCTTGGCTCC CCTGCTGAAG GTGGGGATTG CCCATCCATC TGCTTACAAT TCCCTGCTGT CGTCTTAGCA AGAAGTAAAA
2591 TGAGAAATTT TGTTGATATT CAAAAAAAA
```

* To whom correspondence should be addressed