

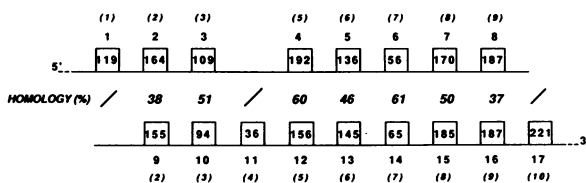
Sequence of the chicken ovotransferrin gene

Jean-Marc Jeltsch, René Hen, Luc Maroteaux, Jean-Marie Garnier and Pierre Chambon\*

INSERM/U184 and LGME/CNRS, Faculté de Médecine, 11 rue Humann, 67085 Strasbourg Cédex, France  
 Submitted July 28, 1987 Accession no. Y00407

The sequencing method of Maxam and Gilbert (1) was used to determine the 10567 bp sequence of the cloned (2) 17-exon chicken ovotransferrin (conalbumin) gene on both strands; the sequence of 266 bp upstream from the cap site and 235 bp downstream from the polyadenylation site are also given. The TATA box [-31 to -25] and polyadenylation signal [10549 to 10555] sequences are underlined. Repetitive sequences (underlined) are present in introns B [1663-1869] and C [2540-2840] (2). A comparison between the gene and cDNA(3, 4) sequences indicates that intron/exon boundaries follow the GT/AG rule (5, 6) and reveals 26 nucleotide differences. Only 6 of these differences lead to amino-acid changes [Ala 64, Val 81, Arg 135, Gln 220, Lys 221 and Ser 667 in the gene instead of Val, Ile, Trp, Leu, Asn and Asn as deduced from the cDNA (3)] and 3 are localized in the 3' untranslated region. These differences most probably reflect the known polymorphism of transferrins (7).

The existence of two homologous iron-binding domains in the transferrins led to the hypothesis that the present-day transferrin genes arose by duplication of a common ancestor [for review see (8)]. This internal homology, verified at the protein level for human lactotransferrin (9), human serotransferrin (10) and hen ovotransferrin (3, 7), is confirmed here at the gene level. The same scheme can be proposed for the origin of both human serotransferrin (11, 12) and chicken ovotransferrin genes. This scheme (intragenic duplication of exons 2-8 of an ancestor gene containing 10 exons and loss of exon 4 from the first half of the duplicated gene) is based entirely on the homology of the corresponding exons from the two halves of the gene, since no significant homology could be detected between the corresponding introns. Note also that the size differences between homologous exons are multiples of 3 bp, suggesting that the evolution of this gene proceeded by insertion/deletion of whole codons.



**Duplication scheme.** The nucleotide homologies (in %) between the homologous exons (represented as boxes, numbered from 1 to 17), the exon's size in bp (in the boxes) and the numbering of the putative exons of the ancestral gene (in parentheses) are indicated.

AGCAAGGGAACATGTCGAACCTGGGAAATTTCCATATGAGAGCCCTTGGAAAAAAGAGAGACCTTGGCTGTCAGCTCCGGCTGGCTGAAAGGTTT...
Met1 Lys Leu Ile Leu Cys Thr Val Ser Leu Glu Ile Ala A
ACCCTGCTCCGCCCTCCAAACCCAGCTCCCTGCCCAAC ATG AAG CTC ATC CTC ACC GTC CTC TCC GGC ATA GGA GGC C

11e His Asn Arg Thr Gly Thr Cys Asn Phe A 880  
 ATT CAC AAC AGA ACA GGG ACC TGC AAT TTC G GTAGCTCAAGT TGTGCATCTGCTTGTGAGATGAAATGCGTAGGAAAGCGTGT 7614  
 Intron 1  
 TCGCAGGCTATCTCTGACTTTCGAAGCTTCTTGTCCCGGACAGCTCTCACTTGCAGGTGGAATAAATGAATGCTGTACTTANGACTTCACGAGGA 7714  
 AAGTCTCAGGCTCTTGTGAGCTGTGCGACAGGATGCGACATACAGCTCTTTGCGAGCACTTGGCTTTGTTTTCAGATCTCTGCTTTCGGCTTTTG 7814  
 CGACGGCAAAAAGAGCATTCGGCAAGTGGGAGGAGCGCTGCTCTTCTGCGAAGGACAGCGGCTGGGCTTGAGAGATGCAATTACAGCTCCGTCAGGGCTG 7914  
 GAGACCGGGAGAGGCTGATTCAGGCTTCAGAGATCTTCAAGAGCTCTGATCTGACTTCGCTGAGAGCGGCTGGCTTAAAGATCTGGAGCTCAGCTCTACG 8014  
 CGACCTTGGCATCTACAGATTAACACTTCACAGCAAAAATAGGCTGCAATTAAGACCTAGGAGAAAATAGACATACGATTAATATGGCAATTAATACGC 8114  
 CAGGAGCTCTGCTGTGTTTAACTTATACAGCCGACCTCAGCCGCTCTTCTGCTTTTCTACAGTCCGCTCTCTGCGAGCTCAGATATCTGCTCC 8214  
 AACAGCGGTTGCTGGCTTGGGCTCAGGCTTGGCAAAAAGAAACCGCGTGTTCACAGAGTAGGAAAGTGAATTTCTGCTATCTGGCTGGCGATCAA 8311  
 500  
 Tyr Phe Ser Gly Cys Ala Pro Gly Ser Pro Phe Asn Ser Arg Leu Cys Gln Leu Cys Gln Gly Ser Gly Cys 500  
 TAC TTC ACC GAG GAG GGT TCT OCT OCT GGG TCC OCT OCT AAG TCC GDE CTC TCC CAG CTC TCC CAG GGC TCA GGC GCA 8386  
 11e Pro Pro Gly Lys Cys Val Ala Ser Ser His Glu Lys Tyr Phe Gly Tyr Thr Gly Ala Leu Arg 500  
 ATCTCAGGTAAGCTCGTCCAGCGCTAAAAGAAAGAGAGGCTGACAGATGAGCGCTGCTTTCTGTCTGGGCTTACAGAGAGTACAGAGCTGAGCGCT 8564  
 AATTCAGCGTGGGCTGAGATTTGGAAAGGAGAAAGAGGCTTGTTCAGCTTGAGCTTAATAGCAGAGAGCGGCTGCTGGAGCGGCTGGAGCTTCTGCA 8664  
 AGCTTGGCGTTACATGCGCTACTGAATATGCACCTCTGCTTGTACTCTGCGAGCTCTGCTGCGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8752  
 541  
 1e Gln His Ser Thr Val Glu Glu Asn Thr Gly G 500  
 TT CAG CAT TCC ACC GTT GAG GAA AAC ACT GGC G GTAGCTGTATGCTATTCGATGGGCTTGGCGGCTGGCTCCCGCAGCTGCAAT 8840  
 Intron 1  
 AGCTTTCTTGGCAAGGATTCATTTAGATATACATTTTCATATACATATATGCTTGGCCATTAACAGCGCGCGCGCTCAGCAGCAGCTCAGCAGATATGCA 8940  
 CTCCGCTGGCGATGGAGAGAGGCTGGCTTCTGAGGAGAGTGGCTGGAGCTTTTAAAGGCTCTGGGGGAAAGATGATGCAAGAAAGATGCTGAGAGAGCT 9040  
 CCGAGCGCGCTCTCAGAGTTTGAATTCAGCGCCACTGACTCCCTGAGTTTACTTGCATGGCTCTCAGTACAGCGAGGCTATCGAAGGCTGAGCTGACT 9140  
 AATCGAGGAGAAACACTCTGCTGGCTGCTGGCTGCTGACTGCTCTGCTTACAGAAAGCTCTGAACTCCCAATCAATTCCTCTACCTGCAAGTCAA 9238  
 560  
 Asn Lys Ala Asp Trp Trp Ala Lys Asn Leu Gln Met Asp Asp Phe Glu Leu Leu Cys Thr Asp Gly Arg Arg Ala Asn 9238  
 AAC AAA GGT CAT TGC GGC AAG AAT CTC CAA ATG CAT GAC TTT GAG TTG CTC TCC ACC GAT GGC ACA GGG GCA ACA 9313  
 581  
 Val Met Asp Tyr Arg Glu Cys Asn Leu Ala Glu Val Pro Thr His Ala Val Val Val Arg Pro Glu Lys Ala Asn 9388  
 CTC ATG GAT TAC AGR GAA TGC AAC CTG OCT GAA CTT OCT ACC CAG OCT CTC GTC GTA GGC CGA GAG AAA GCA AAC 9400  
 Lys Lys Arg Asp Leu Leu Glu Arg Gln Glu Arg CTTGGTGGGCTGACAGCTGAGAAATGTTCTGGGGGGTTTTTACAGCTGCTGCTGCA 9477  
 589  
 ATTTTGGGGGAGTACTGCTGCAATTCACCGGCTGGCAAAAATGCTGGCTGCTTGCATCTACTGCAAGCAAGAGCGCTGAGCGCTTAACTGCT 9577  
 TCCAGCGCTGGCTGCAATGCTTCCACTACAGCGGAGTGAAGCGGCTGCCGATGATGCTTGAGCTCTTCATGCTGAGAGAGCTCTCAGAGCGCTGTA 9677  
 620  
 CAGCTGAATCTGCTTCTTCTGCTTTTACATACACTAACCTTCTGCTGACTTATTTTCAG Lys Arg Phe Gly Val Asn Gly Ser Glu 9768  
 AAA ACC AAG TTC ATC ATC TTT CAG CTA AAT GCA AGT GAG CTT TTT GCA CTA AAT GCA AGT GAG  
 640  
 Lys Ser Lys Phe Met Met Phe Glu Ser Gln Asn Lys Asp Leu Leu Phe Lys Asp Leu Thr Lys Cys TCC CTC TTT Lys 9843  
 AAA ACC AAG TTC ATC ATC TTT CAG CTA AAT GCA AGT GAG CTT TTT AAA GAC TTA ACC AAG TCC CTC TTT Lys  
 680  
 Val Arg Glu Gly Thr Tyr Thr Glu Phe Leu Gly Asp Lys Phe Tyr Thr Val Ile Ser Ser Leu Cys Lys Ala Asn 9918  
 CAG GCA GCA GTC G GTAATATAGCGGCTCAAAAAGAGTCAATTTAGCGAGCTGCTTAGCTCAGCATCTGACAGCTTCTGCTGCCCAATTT 10014  
 Intron 1  
 CATTTCTACAGAAAGAAAGAGGCACTCATGAGGCTTATGATTTCCGACGCTTCTGAGCAAGCTGAGCGGATGGCAATGGCTTCTCTGCTGATC 10114  
 AATGCAATGAGCATTTTACTTGGCTTGAAGCGGCAATCTGAGAGATGCTGTAATAATAGCGGCTGCTCTCTCAGCTGCTGAGCTGCTGACTGAT 10214  
 GCTTAGCTGCACTTCTGGCGATCTGCTGCAATTTGGCATGCTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 10314  
 680  
 GCTCAACTTTTCTGCTGCTTTTCTGCTTTCAGATATTC CAG GGC TTC CAG GGC AAG TAA AGGGAGGAGAGGGCC 10399  
 11e His Leu Glu Met Cys Ser Phe Leu Glu Lys His 680  
 CTTTCTTGAAGGGGAGCAAGCTTGGCGCACTGCTCTGCTTGGCGCGGACGCAAGCTGATGCTGACAGGGGCTCTGGCATTTCTGCTGCTGCT 10499  
 CCTTCTGCT 10599  
 11e His Leu Glu Met Cys Ser Phe Leu Glu Lys His 680  
 TARGACATTTCTTCCGAGCAGGCTTCTGCTGAGCGAGTGGCAGGCAAGCTTGGCTTACTGCTTGGCTTGGAGAGAAATGCTGCAATTTTTCGCA 10699  
 AGCAAAAAGCATTAAGATTTGGCTTAGCTGATTTAGAGGGCTTTGATTTATCTGCTTGGCTTCCGCGAGATTAAGCAATGCTGCTGCGAGCT 10799  
 CGA

Acknowledgements

We wish to thank Drs. J.P. LePenec and R. Krishnamoorthy for helpful discussions and C. Werlé and B. Boulay for help in the preparation of the figures. This work was supported by grants from the CNRS, the INSERM, the Ministère de la Recherche et de la Technologie, the Fondation pour la Recherche Médicale Française and the Association pour la Recherche sur le Cancer.

\*To whom correspondence should be addressed

References

- Maxam, A.M. and Gilbert, W. (1980) Methods in Enzymology 65, 499-566.
- Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F. and Chambon, P. (1979) Nature 282, 567-574.
- Jeltsch, J.M. and Chambon, P. (1982) Eur. J. Biochem. 122, 291-295.
- Jeltsch, J.M. (1982) Thesis (Université Louis Pasteur, Strasbourg, France).
- Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978) Proc. Natl. Acad. Sci. USA 75, 4853-4857.
- Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.
- Williams, J., Elleman, T.C., Kingston, I.B., Wilkins, A.G. and Kuhn, K.A. (1982) Eur. J. Biochem. 122, 297-303.
- Williams, J. (1982) Trends Biochem. Sci. 7, 394-397.
- Metz-Boutigue, M.H., Jollès, J., Mazurier, J., Schoentgen, F., Legrand, D., Spik, G., Montreuil, J. and Jollès, P. (1984). Eur. J. Biochem. 145, 659-676.
- MacGillivray, R.T.A., Mendes, E., Schewale, J.G., Sinka, S.K., Lineback-Zins, J. and Brew, K. (1983). J. Biol. Chem. 258, 3543-3546.
- Park, I., Schaeffer, F., Sidoli, A., Baralle, F.F., Cohen, G.N. and Zakin, M.M. (1985) Proc. Natl. Acad. Sci. USA 82, 3149-3153.
- Schaeffer, E., Lucero, M.A., Jeltsch, J.M., Py, M.C., Levin, M.J., Chambon, P., Cohen, G.N. and Zakin, M.M. (1987). Gene, in press.