

The Complete Nucleotide Sequence of the Chicken Ovotransferrin mRNA

Jean-Marc JELTSCH and Pierre CHAMBON

Laboratoire de Génétique Moléculaire des Eucaryotes du Centre National de la Recherche Scientifique
Unité 184 de Biologie Moléculaire et de Génie Génétique de l'Institut National de la Santé et de la Recherche Médicale,
Institut de Chimie Biologique, Faculté de Médecine Strasbourg

(Received August 26, 1981)

The nucleotide sequence of an almost double-stranded cDNA copy [Cochet, M., Perrin, F., Gannon, F., Krust, A., Chambon, P., McKnight, G. S., Lee, D. C., Mayo, K. E., and Palmiter, R. D. (1979) *Nucleic Acids Res.* 6, 2435–2452] of chicken ovotransferrin (conalbumin) mRNA has been determined. Taking into account the previously reported 5'-end sequence [Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F., and Chambon, P. (1979) *Nature (Lond.)* 282, 567–574] we present the complete nucleotide sequence of the ovotransferrin mRNA from which the amino acid sequence of the protein is inferred. A computer and statistical analysis of the nucleotide sequence reveals a pattern of internal homology which confirms that the present-day chicken ovotransferrin gene (and by extrapolation the transferrin genes of other species) has evolved by duplication and gives some support to the quadruplication hypothesis of transferrin evolution.

The avian transferrin gene is interesting in several respects. It is expressed both in the liver and in the oviduct, where its protein products are known as serum transferrin and ovotransferrin (or conalbumin) respectively. The liver and the oviduct proteins differ only in the nature of their attached glycan chain [3, 4]. However, the modulation of the expression of the transferrin gene by steroid hormones and iron levels is very different in the two organs [2, 5, 6]. Determination of partial amino-acid sequences of the N-terminal and C-terminal ends of the protein, X-ray crystallography studies and partial proteolysis experiments on human, chicken and rabbit transferrins have led to the hypothesis that the present-day transferrin gene has evolved by duplication from an ancestral gene [7–10]. Moreover, it has been proposed that this ancestral gene could be itself the product of a more ancient duplication [8] and, therefore, that the present-day transferrins could be the product of a quadruplication event.

We have previously reported that the chicken transferrin gene is split in 17 exons [2]. The complete elucidation of the structure of the transferrin gene at the sequence level should reveal some of the mechanisms involved in the generation of split genes by duplication and possibly quadruplication of ancestral genes. Such a study should disclose whether the ancestral gene(s) was (were) already split and how the original exon-intron pattern has evolved since the duplication events. The presence of two homologous iron-binding domains in transferrins [7, 9, 10] will also provide an opportunity to see if the proposal that exons could correspond to functional domains applies to this gene [11].

To address these questions and to form a basis for further studies of the ovotransferrin gene structure, we have determined the nucleotide sequences of an almost full-length ovotransferrin double-stranded cDNA clone [1]. This sequence, together with the previously reported sequence of the 5'

end of the ovotransferrin gene [2], provides a complete sequence of chicken ovotransferrin mRNA and protein. Analysis of the nucleotide sequence data reveals extensive internal homologies in the ovotransferrin mRNA, which confirms that transferrins have evolved by duplication of an ancestral gene and gives some support to the quadruplication hypothesis of transferrin evolution.

MATERIALS AND METHODS

Materials

Restriction enzymes were purchased from New England Biolabs, BRL or Boehringer and used according to their recommendations. [γ - ^{32} P]ATP (> 1000 Ci $\times$ mmol $^{-1}$) was obtained from Amersham or NEN.

DNA Sequencing

pBR 32-con 1 plasmid, containing the conalbumin double-stranded cDNA (dscDNA) [1] was purified by the clear lysate technique [12], followed by CsCl equilibrium centrifugation. DNA restriction fragments used for sequencing were purified by centrifugation (SW41 rotor) through 5–20% sucrose gradients (in 20 mM Tris/HCl pH 7.9, 2 mM EDTA, 200 mM NaCl, 0.1% sarkosyl, 5 μ g/ml ethidium bromide). The fragments were detected using an ultraviolet lamp and recovered by puncturing the wall of the tubes. After butanol extraction to remove the ethidium bromide, the DNA was precipitated twice with two volumes of ethanol.

Preparation and sequencing of DNA fragments, ^{32} P-labelled at the 5' end, were performed according to Maxam and Gilbert [13]. Products of the chemical degradation reactions were separated on thin, denaturing 8.3 M urea polyacrylamide gels with acrylamide concentrations of 6%, 8%, 10% or 20% [13, 14].

Statistical Analysis of Sequence Data

A computer program, written by R. Fritz, was used to search for sequence homology. After a preliminary analysis the

Abbreviation. dscDNA, double-stranded complementary DNA.

Enzymes. Restriction endonucleases *Taq* I (EC 3.1.23.39), *Hinf* I (EC 3.1.23.22), *Pvu* II (EC 3.1.23.33), *Bam* HI (EC 3.1.23.6), *Eco* RI (EC 3.1.23.13), *Bst* NI (EC 3.1.23.–), *Ava* I (EC 3.1.23.3), *Acc* I (EC 3.1.23.–), *Ava* II (EC 3.1.23.4), *Sau* 3A (EC 3.1.23.27), *Bgl* II (EC 3.1.23.10), *Hinc* II (EC 3.1.23.20).

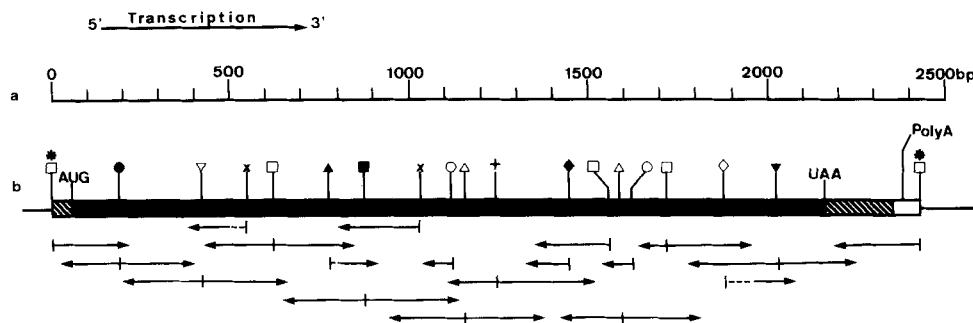


Fig. 1. Strategy for sequencing the conalbumin dscDNA. (a) Scale in base pairs (bp). (b) Restriction map of the conalbumin dscDNA inserted in the pBR 322 unique repaired *SalI* site generating two *TaqI* sites (noted $\square$ in the figure). This map shows only the sites used for 5'-end labelling and sequencing. Direction and extent of sequencing determination are indicated by the arrows. The 5' and 3' non-coding regions are hatched and the poly(A) tail is also indicated. The light lines correspond to the flanking plasmid sequences. Dashed lines correspond to sequences which were not read. It should be noted that the localization of some of the restriction enzyme sites is different from that reported in the preliminary mapping study of Cochet et al. [1]. A complete restriction enzyme map of ovotransferrin dscDNA, as deduced from the sequencing data, is available upon request. ($\square$) *TaqI*; ($\times$) *HinfI*; ($\circ$) *PvuII*; (Δ) *BamHI*; (∇) *EcoRI*; ($\diamond$) *BstNI*; ($\blacksquare$) *AvaI*; (+) *AccI*; ($\bullet$) *AvaII*; ($\blacktriangle$) *Sau3A*; ($\blacktriangledown$) *BglII*; ($\blacklozenge$) *HincII*

mRNA sequence was divided into segments of 100 nucleotides with an increment size of 30 nucleotides (in order to remain in phase with the codon reading frame). Each of the 100 nucleotide segments was compared with all of the others. No allowance was made for gaps and the homology between 100-nucleotide blocks was expressed as a percentage. The validity of a given homology between two sequences was then evaluated as previously described by Sargent et al. [15]. Taken P (the probability that any given position will be a match) equal to 0.25 (i. e. there is no preference for any of the four nucleotides at any given position), a homology extending over 100 nucleotides with a maximum P_a of 7×10^{-4} was considered 'legitimate' [15]. (P_a is the probability that a homology equal to or greater than that being considered might arise accidentally). This value of P_a corresponds to a 43% homology between blocks of 100 nucleotides when P is equal to 0.25. In fact, there is some deviation from ideal randomness. Empirical determination of P for ovotransferrin mRNA sequences yields values falling between 0.23 and 0.29. (A P_a value of 7×10^{-4} would correspond to 40% and 48% homologies between blocks of 100 nucleotides for P values of 0.23 and 0.29 respectively.)

RESULTS AND DISCUSSION

The Ovotransferrin mRNA Sequence

The cloning and preliminary characterization of an almost full-length chicken ovotransferrin (conalbumin) dscDNA was previously described [1]. We have used the method of Maxam and Gilbert [13] to determine the sequence of this cloned conalbumin dscDNA. The strategy for sequencing is outlined in Fig. 1, where the sites used for 5'-end labelling, the direction and extent of sequence determination are indicated. All of the restriction sites used for sequencing were overlapped with the exception of the two *TaqI* sites [1] located at the extremities of the conalbumin dscDNA insert. {We are, however, confident that no *TaqI* fragment was overlooked, since one of the *TaqI* sites is located at the 3' end of the mRNA poly(A) tail, while the other is situated at the 5' end of the cloned dscDNA in a region which was previously sequenced in the cloned chromosomal ovotransferrin gene [2].} Most of the sequencing operations were performed on the two DNA strands with the exception of about 300 base pairs in a region located at the 3' end of the dscDNA, where sequencing was repeated twice from the *BglII* and *TaqI* sites. Therefore, the resulting dscDNA sequence is most likely free of errors, although we cannot exclude mu-

tations which could have occurred during the cloning. The complete ovotransferrin mRNA sequence is shown in Fig. 2, taking into account the first 20 nucleotides, which are not represented in the dscDNA but were previously determined by sequencing the 5' ends of the chromosomal gene and of the ovotransferrin mRNA [2].

The ovotransferrin mRNA contains 2376 nucleotides [excluding the cap nucleotide and the poly(A) tail]. There is a 182-nucleotide 3'-untranslated region and, as previously reported [2], the 5'-untranslated sequence is 76 nucleotides long. The sequence 5'-A-A-U-A-A-A-3', which is found in polyadenylated mRNAs 11–30 nucleotides upstream from the start of the poly(A) sequence and which forms part of the recognition site for polyadenylation, is present 18 nucleotides upstream from the poly(A) tail (for references, see [16]).

The ovotransferrin mRNA codes for a 705-amino-acid preprotein. As previously reported [2, 17], there is a 19-amino-acid protein leader sequence (or signal peptide). The present version of the leader sequence is in complete agreement with that of Thibodeau et al. [17] obtained by amino acid sequencing. This is in contrast with our previous sequencing results [2] in which an isoleucine was attributed to position 2 instead of a phenylalanine. A detailed comparison of the sequence of the 686-amino-acid secreted ovotransferrin protein, as deduced from the mRNA sequence, with the partial sequence established by amino acid sequencing is presented in the accompanying paper of Williams et al. [18] in the light of the known existence of ovotransferrin protein variants.

Internal Duplication of the Ovotransferrin mRNA

Several lines of evidence have suggested that the present-day transferrin genes have evolved by duplication (for references, see introduction an [19]). The mRNA sequence was analyzed as described in Materials and Methods and in the legend to Fig. 3 to search for extensive internal homology. We looked for sequence homologies starting from position 134, which corresponds to the first nucleotide coding for the secreted protein, since a preliminary analysis did not reveal any significant homology to the 5' untranslated region or to the sequence coding for the leader sequence. As shown in Fig. 3, two blocks of sequences (134–934 and 1134–1934) exhibit a striking homology (as shown by the discontinuous diagonal series of squares), which strongly supports the duplication hypothesis. Each block of homology appears to be divided into two sub-blocks of similar lengths separated by a region of lower

5'-AUCACAGACCCAGAGGGGACGGUCUGUGACCAACACCGCUGCUCUCCCCUCCUACACCCAGCUCUCCGCCCCAACAUGAAGCUCAUCCUC
MetLeuLeuLeu
120 160
UGCACCGUGCUGUCCUUGGGGAUAGCGGCUGUGUGUUCGCCGUCUCCCCAAGUCAGUCAUAGAUUGGUGACCAUUAUCCUCCAGAGAGAAGAAA
CysThrValLeuSerLeuGlyIleAlaAlaValCysPheAlaAlaProProLysSerValIleArgTrpCysThrIleSerSerProGluGluLysLys
200 240 280
UGCAACAACCCAGGGACCCACACAACAAGAGAGGAUUCUUGACCGUGUGCAGAAAGCAACGUACCUUGACUGCAUUAAGCCAUUGCGGAUAAU
CysAsnAsnLeuArgAspLeuThrGlnGlnGluArgIleSerLeuThrCysValGlnLysAlaThrTyrLeuAspCysIleLysAlaIleAlaAsnAsn
320 360
GAGGCGAUGCCAUUAGCUUGGACGGCGGCAAGUGUUGAGGCGAGCCUUGCCCCAUACAAGCUGAAGCCCAUUGCUGCUGAGAUUUAUGAACAUACU
GluAlaAspAlaIleSerLeuAspGlyGlyGlnValPheGluAlaGlyLeuAlaProTyrLysLeuLysProIleAlaAlaGluIleTyrGluHisThr
400 440 480
GAAGGCUCCACCAACAGCUACUUGCUGUGGCGGUGUGAAAAAGGAACCAUUAUCCGUGAACGACUUGCAGGGCAAGAACUCCUGCCACACGGGG
GluGlySerThrThrSerTyrTyrAlaValAlaValValLysLysGlyThrGluPheThrValAsnAspLeuGlnGlyLysThrSerCysHisThrGly
520 560
CUGGGCAGGUCUGCGGGCUGGAACUCCCAUUGGGACACUCCUCCACUGGGGGCCAUUGAGUGGGAAGGCAUAGAACUGGGCUCAGUGGAACAAGCG
LeuGlyArgSerAlaGlyTrpAsnIleProIleGlyThrLeuLeuHisTrpGlyAlaIleGluTrpGluGlyIleGluSerGlySerValGluGlnAla
600 640 680
GUGGCCAAGUUUUUCUGCCAGCUGCGUGCCCGGUGCCACCAUCGAGCAAAACUGUGCCGUCAGUGCAAAGGAGACCCCAAAACCAAGUGUGCCCGC
ValAlaLysPhePheSerAlaSerCysValProGlyAlaThrIleGluGlnLysLeuLysCysArgGlnCysLysGlyAspProLysThrLysCysAlaArg
720 760
AAUGCACCUAUUCUGGAUUAUUGGAGCUUCCAUUGUUUGAAAGGAUGGAAAGGAGACGUGGCUUUGUGAAGCACACAACUGUUAUAGAAAUGCC
AsnAlaProTyrSerGlyTyrSerGlyAlaPheHisCysLeuLysAspGlyLysGlyAspValAlaPheValLysHisThrThrValAsnGluAsnAla
800 840
CCGGAUCUGAUGAUGAGUAUGAGCUGCUGUGCCUGGAUGGACGCGGCGAGCCUGUGGACAACUACAAGACUGUAACUGGGCCAGAGUGGCGUCAC
ProAspLeuAsnAspGluTyrGluLeuLeuLysLeuAspGlySerArgGlnProValAspAsnTyrLysThrCysAsnTrpAlaArgValAlaAlaHis
880 920
GCCGUCUGGUCUGGGAGACAAAGGUUGAAGAUUUCUGGAGCUUCCUCUCAAAGACACAGAGUGACUUGGCGUGGACCAAGAGUGACUCCAC
AlaValValAlaArgAspAsnLysValGluAspIleTrpSerPheLeuSerLysAlaGlnSerAspPheGlyValAspThrLysSerAspPheHis
960 1000
CUCUUGGGCCACCUGGAAAGAGGACCGGUCUCAAAGACUUCUUAUCAAAGACUUGCCAUAAUGCUGAAGCGCGUCCUCCGUGGAGGUAUCC
LeuPheGlyProProGlyLysLysAspProValLeuLysAspLeuLeuLysLysAspSerAlaIleMetLeuLysArgValProSerLeuMetAspSer
1040 1080
CAGCUCUACCUGGGCUUUGAGUACUACAGUGCCAUCCAGGAUGCGGAAAGAUACGUGACUCCAGCCCGAGAGAAACAGGAUCCAUGGUGUGCA
GlnLeuTyrLeuGlyPheGluTyrTyrSerAlaIleGlnSerMetArgLysAspGlnLeuThrProSerProArgGluAsnArgIleGlnTrpCysAla
1120 1160
GUAGGCAAGGAUGAGAAGAGCAAGUGGACCGUGGAGUGUGGAGCAACGGGGACGUGGAGUGCACCUGGUAGACGAGACAAGGACUGCAUCAU
ValGlyLysAspGluLysSerLysCysAspArgTrpSerValValSerAsnGlyAspValGluCysThrValValAspGluThrLysAspCysIleIle
1200 1240 1280
AAGAUCAGAAAGGUGAAGCAGUUGCUGGUGGAGGUGGUGCUUACUGCUGGUGUGUGGCGUGGUGCAGUGAUGGAGAACGAUUAU
LysIleMetLysGlyGluAlaAspAlaValAlaLeuAspGlyGlyLeuValTyrThrAlaGlyValCysGlyLeuValProValMetAlaGluArgTyr
1320 1360
GACGAUGAAAGCACAUGCAGCAAAACAGAUAGACGACGACCAUACUUGCUGUGGUGUGCGGAGGAAAGACAGCAACGUAACUGGAACAACUG
AspAspGluSerGlnCysSerLysThrAspGluArgProAlaSerTyrPheAlaValAlaValAlaArgLysAspSerAsnValAsnTrpAsnAsnLeu
1400 1440
AAGGGCAAGAAUUGGUGCCACACCGCUGUGGGGAGGACUGGUGGUGUACUCCCAUGGGCUUGAUUACAACAGAAGGACUGCAUUAUUGAU
LysGlyLysLysSerCysHisThrAlaValGlyArgThrAlaGlyTrpValIleProMetGlyLeuIleHisAsnArgThrGlyThrCysAsnPheAsp
1480 1520
GAAUACUUCAGCGAGGGUUGUGCUCUGGAUCCUCCUUAACUCCGCCUUGCCAGCUGGCCAGGGGUCAGGGGGAUCCACCGGAGAAGUGCGUC
GluTyrPheSerGluGlyCysAlaProGlySerProProAsnSerArgLeuLysGlnLeuLysGlnGlySerGlyGlyIleProProGluLysCysVal
1560 1600
GCCAGCAGCCAUAGAGAAUACUUGGAUUAUCCGAGCUCUACGGUGUCUGGUGGAGAGGGUGAUGUGGCGUUAUUCAGCAUUCACCGUUGAGGAA
AlaSerSerHisGluLysThrPheGlyTyrThrGlyAlaLeuArgCysLeuValGluLysGlyAspValAlaPheIleGlnHisSerThrValGluGlu
1640 1680
AACACUGGCGGCAAAACAAAGCUGAUUGGGCAGAAUUCUGCAAUGGAUGACUUGAGUUGCUGGACCGAGGGAGGCGGGCAACGUCUAGGAU
AsnThrGlyGlyLysAsnLysAlaAspTrpAlaLysAsnLeuGlnMetAspAspPheGluLeuLeuLysThrAspGlyArgArgAlaAsnValMetAsp
1720 1760
UACAGGGAUUGCAACUGGUGAAGUUCUACCCACGUGUGGUGGUGCGCCGAGAAAGCAAAACAAAUCCUGAUCUGGAGAGACAGGAGAAA
TyrArgGluCysAsnLeuAlaGluValProThrHisAlaValValValArgProGluLysAlaAsnLysIleArgAspLeuLeuGluArgGlnGluLys
1800 1840
CGGUUGGAGUAAUUGAAGCGAGAAAGCAAGUUAUGAUGUUGAGUCUAAAACAAAGAUUUCUGUUUAAAGAUUAAACCAAGUGCCUGUUUAAA
ArgPheGlyValAsnGlySerGluLysSerLysPheMetMetPheGluSerGlnAsnLysAspLeuLeuPheLysAspLeuThrLysLeuPheLys
1880 1920
GUCCGAGAAGGAACAACAUACAAGGAGUUCUUGGAGAUAAUUUUUUAUCUGUGAUUUCCAACCUCAAACUGCAACCAUCAGAUUCCUCCAGAU
ValArgGluGlyThrThrTyrLysGluPheLeuGlyAspLysPheTyrThrValIleSerAsnLeuLysThrCysAsnProSerAspIleLeuGlnMet
1960 2000
UGCAGCUUCUUGAGGGCAAGUAAAGGGAGGGAAGGGCCUUCUGAAGGGGAGGAACUUCGCGCAUGACUCCUCCUGGCCCCCGACGGAACAC
CysSerPheLeuGluGlyLys
2040 2080
UGAUGUGCAGAGGGCCUCCUGCAUUGCUGUCCUCCUGCCUCCUCCUGACUCUGAAUGGGUUCUUCGCUACUGCCACAGCAAGAAUAAAUUCU
CAACAUU-POLY(A)_{N=71}

Fig. 2. The nucleotide sequence of conalbumin mRNA and the inferred protein sequence. The first 20 nucleotides were previously determined by M. Cochet et al. [2]. The AUG initiator codon and the UAA terminator codon are boxed. The common eukaryotic mRNA polyadenylation signal 5'-A-A-U-A-A-3' is underlined. The nucleotide and the protein sequences are numbered separately. For the nucleotide sequence, numbering is above the nucleotide line (the 0 of each number is placed above the corresponding nucleotide and there is a dot every 20 nucleotides). For the amino acid sequence numbering is below the amino acid line. The signal (leader) peptide extends from position 19 (Met) to 1 (Ala). * + 1 indicates the first NH₂-terminal amino acid of the secreted protein; ** indicates the amino acid bearing the glycan chain [4]

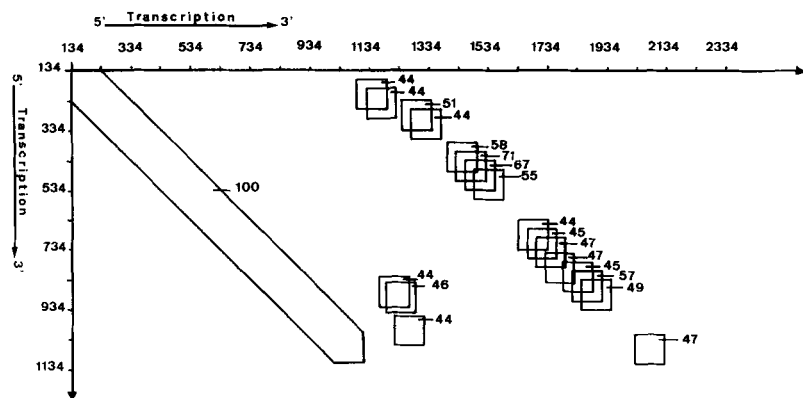


Fig. 3. Internal homology diagram of the conalbumin mRNA obtained by computer analysis (Materials and Methods). Each side of the squares corresponds to 100 nucleotides. The numbers associated to squares are the homology scores in percent. Only statistically significant homologies > 44 % are represented (see Materials and Methods). The homology search starts at the first nucleotide of the codon of Ala⁺ (the N-terminal amino acid of the secreted protein), i.e. nucleotide number 134

A.

I	116	His	Thr	Gly	Leu	Gly	Arg	Ser	Ala	Gly	Trp	Asn	Ile	Pro	Ile	Gly	Thr	Leu	Leu	133
II	250	His	Ala	Val	Val	Ala	Arg	Asp	Asp	Asn	Lys	Val	Glu	Asp	Ile	Trp	Ser	Phe	Leu	267
III	455	His	Thr	Ala	Val	Gly	Arg	Thr	Ala	Gly	Trp	Val	Ile	Pro	Met	Gly	---	Leu	Ile	471
IV	592	His	Ala	Val	Val	Val	Arg	Pro	Glu	Lys	Ala	Asn	Lys	---	Ile	Arg	Asp	Leu	Leu	608

B.

I	479	CAC	ACG	GGG	CUG	GGC	CAG	GUC	GGG	CUG	GAA	CAU	CCC	CAU	UGG	GAC	ACU	CCU	C	532
II	881	CAC	GGC	GUC	GUG	CUG	GGA	CAU	CAU	AGG	UUA	GAG	AUA	UUC	GAG	AGC	UUC	U	C	934
III	1496	CAC	ACG	GGG	CUG	GGC	CAG	GUC	GGG	CUG	GAA	CAU	CCC	CAU	UGG	GAC	---	UUG	AUU	1546
IV	1907	CAC	GCU	GUG	GUC	GCG	CCG	GAG	AAG	CAA	CAA	---	AUC	CGU	GAG	UUC	GUG	---	---	1957

C.

	COMPARISON OF THE FOUR REGIONS					
	I/II	I/III	I/IV	II/III	II/IV	III/IV
LENGTH IN NUCLEOTIDES	54/54	54/51	54/51	54/51	54/51	51/51
NUMBER OF CODON DELETIONS	-/-	-/1	-/1	-/1	-/1	1/1
NUMBER OF HOMOLOGOUS CODONS	2	6	2	2	2	1
NUMBER OF HOMOLOGOUS NUCLEOTIDES	24	36	24	24	25	20
% OF NUCLEOTIDE HOMOLOGY	44	67	44	44	46	37

Fig. 4. Comparison of sequences which suggests a fourfold homology in chicken ovotransferrin. (A) Alignment of the relevant amino acid sequences (see text); I, II, III and IV correspond to amino acids 116–133, 250–267, 455–471 and 592–608 of the secreted protein (Fig. 2). Gaps have been introduced in III and IV in order to maximize the homologies. (B) Alignment of the corresponding codons. (C) Homologies between the nucleotide sequences I, II, III and IV. Amino acids or nucleotides present at least three times at homologous positions in sequences I, II, III and IV, are underlined in (A) and (B)

homology. It is noteworthy that there is only weak evidence for duplication of the region between nucleotides 935 and 1100. Sequencing of the chromosomal gene (in progress) should reveal whether the corresponding exons have not been duplicated or whether they have diverged more rapidly than the others since the duplication events occurred. Two other blocks of homology (834–1034 to 1100–1300) are also present in Fig. 3. Their occurrence in these positions cannot be readily explained by a simple duplication scheme of ancestral gene and it should be noted that similar homologies are not found in these regions when the corresponding amino-acid sequences are compared.

MacGillivray et al. [8] have suggested that the human transferrin may have evolved by quadruplication, on the basis

of a weak fourfold homology present in the protein sequence. Some amino acid residues, which may belong to the iron-binding sites, are found four times at the same distance from each other. These residues are the His and Arg represented in Fig. 4A. When amino acid sequences in this region are compared, additional homologies are apparent. Comparison of the corresponding nucleotide sequences (Fig. 4B, C) gives further support to the quadruplication hypothesis, since regions I and II, II and III and I and IV exhibit 44 % homology, much above the value observed for 50 nucleotide blocks taken at random. We hope that the complete elucidation at the sequence level of the exon organization of the chick transferrin chromosomal gene, presently in progress in our laboratory, will provide a more valid approach to evaluate the quadruplication hypothesis.

We thank Remy Fritz, who wrote the computer program, and René Hen for advice concerning DNA sequencing. The excellent technical assistance of J. M. Garnier is acknowledged. This investigation was supported by grants from the *Centre National de la Recherche Scientifique* (ATP 006520/50), the *Institut National de la Santé et de la Recherche Médicale* (CRL 80.1029), from the *Association pour le Développement de la Recherche sur le Cancer*, from the *Fondation pour la Recherche Médicale Française* and from the *Fondation Simone et Cino del Duca*. J. M. Jeltsch was supported by an *Ingénieur-Docteur* fellowship from the *Centre National de la Recherche Scientifique*.

REFERENCES

1. Cochet, M., Perrin, F., Gannon, F., Krust, A., Chambon, P., McKnight, G. S., Lee, D. C., Mayo, K. E. & Palmiter, R. D. (1979) *Nucleic Acids Res.* 6, 2435–2452.
2. Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F. & Chambon, P. (1979) *Nature (Lond.)* 282, 567–574.
3. Lee, D. C., McKnight, G. S. & Palmiter, R. D. (1980) *J. Biol. Chem.* 255, 1442–1450.
4. Dorland, L., Haverkamp, J., Vliegthart, J. F. G., Spik, G., Fournet, B. & Montreuil, J. (1979) *Eur. Biochem. J.* 100, 569–574.
5. Perrin, F., Cochet, M., Gerlinger, P., Cami, B., Le Pennec, J. P. & Chambon, P. (1979) *Nucleic Acids Res.* 6, 2731–2748.
6. McKnight, G. S. (1978) *Cell*, 14, 403–413.
7. Williams, J., Evans, R. W. & Moreton, K. (1978) *Biochem. J.* 173, 535–542.
8. MacGillivray, R. T. A., Mendez, E. & Brew, K. (1977) in *Protein of Iron Metabolism* (Brown, E. B., Aisen, P., Fielding, J. & Crichton, R. R., eds) pp 133–141, Grune and Stratton, New York.

9. Gorinsky, B., Horsburgh, C., Lindley, P. F., Moss, D. S., Parkar, M. & Watson, J. L. (1979) *Nature (Lond.)* 281, 157–158.
10. Aisen, P. & Listowsky, I. (1980) *Annu. Rev. Biochem.* 49, 357–393.
11. Crick, F. (1979) *Science (Wash. DC)* 204, 264–271.
12. Clewell, D. B. & Helinski, D. R. (1969) *Proc. Natl Acad. Sci. USA*, 62, 1159–1166.
13. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* 65, 499–560.
14. Sanger, F. & Coulson, A. R. (1977) *FEBS Lett.* 87, 107–110.
15. Sargent, T. D., Yang, M. & Bonner, J. (1981) *Proc. Natl Acad. Sci. USA*, 78, 243–246.
16. Fitzgerald, M. & Shenk, V. (1981) *Cell*, 24, 251–260.
17. Thibodeau, S. N., Lee, D. C. & Palmiter, R. D. (1978) *J. Biol. Chem.* 253, 3771–3774.
18. Williams, J., Elleman, T. C., Kingston, I. B., Wilkins, A. G. & Kuhn, K. A. (1982) *Eur. J. Biochem.* 122, 297–303.
19. Wilson, A. C., Carlson, S. S. & White, T. J. (1977) *Annu. Rev. Biochem.* 46, 573–639.

J.-M. Jeltsch, Laboratoire de Génétique Moléculaire des Eukaryotes du Centre National de la Recherche Scientifique,
Faculté de Médecine de l'Université Louis-Pasteur,
11 Rue Humann, F-67085 Strasbourg-Cedex, Bas-Rhin, France

P. Chambon, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'Institut National de la Santé et de la Recherche Médicale,
Institut de Chimie Biologique, Faculté de Médecine de l'Université Louis Pasteur
11 Rue Humann, F-67085 Strasbourg-Cedex, Bas-Rhin, France