Amino Acid Sequence Homology between Rat α-Fetoprotein and Albumin at the COOH-terminal Regions*

(Received for publication, April 23, 1980, and in revised form, June 6, 1980)

Warren S. L. Liao, Robert W. Hamilton, and John M. Taylor

From the Department of Microbiology and Specialized Cancer Research Center, Hershey Medical Center, The Pennsylvania State University, Hershey, Pennsylvania 17033

The nucleotide sequence of the 3′-terminal untranslated region and a portion of the coding region of rat α-fetoprotein mRNA has been determined from a cloned double-stranded cDNA. The amino acid sequence of the COOH-terminal portion of α-fetoprotein was inferred from the nucleotide sequence and compared to the amino acid sequence of the corresponding portion of rat, bovine, and human albumin. A striking homology in amino acid sequence between α-fetoprotein and albumin was observed. These results confirm earlier suggestions that the two proteins are closely related in structure and probably arose from a common ancestral gene.

α-Fetoprotein (AFP) is the major protein in fetal plasma and amniotic fluid (1). Following birth, the plasma concentration of AFP decreases dramatically, and albumin becomes the major circulating protein. Rat AFP is a single chain polypeptide with a molecular weight of approximately 72,000, containing about 4% carbohydrate (2, 3). Rat serum albumin is a Mr = 65,000 single chain peptide with no prosthetic groups (4). Additional biological, physical, and chemical aspects of AFP (5) and albumin (4) have been reviewed recently.

The coordinated reciprocal regulation of the production of these proteins and their similar biological, physical, and chemical properties have suggested to many investigators that AFP is the fetal analog of albumin. This possibility was considerably strengthened by the finding that two internal peptides of human AFP, generated by cyanogen bromide cleavage, show significant amino acid sequence homology to albumin (6). However, the NH2-terminal portions of human, mouse, and rat AFP have no amino acid sequence homology with the corresponding albumin termini (7). Furthermore, cDNAs to rat AFP and albumin do not cross-hybridize to their heterologous mRNAs (8) under moderately stringent conditions. Thus, the structural relationship of these two important proteins has remained somewhat unclear.

In order to investigate this controversial problem, we have examined the nucleotide sequence corresponding to the 3′-terminal region of rat AFP mRNA. We have employed a double-stranded cDNA (ds-cDNA) copy of rat AFP mRNA cloned in the plasmid pBR322 for this purpose. This communication describes the striking homology of the COOH-terminal region of AFP with the corresponding amino acid sequences of albumin from different species.

EXPERIMENTAL PROCEDURES

Cloned AFP ds-cDNA, corresponding to the 3′-terminal region of AFP mRNA, was prepared as briefly described below. Rat yolk sac poly(A)-containing RNA was sedimented through isokinetic sucrose gradients, and the 15 S fraction corresponding to AFP mRNA was collected and further purified by a second sucrose gradient sedimentation. Translation of this material in a nuclease-treated reticulocyte lysate showed that greater than 70% of the protein synthetic product was a single peptide with a molecular weight of about 70,000 which was immunoprecipitated by specific antibodies to rat AFP (anti-albumin). Avian myeloblastosis virus reverse transcriptase was employed to synthesize single-stranded cDNA (ss-cDNA) and the cDNA was then used as template for second strand synthesis using Escherichia coli DNA polymerase I. The ds-cDNA was made blunt ended by digestion with S1 nuclease, and about 30 DCMP residues were added to the 3′ termini with terminal transferase. Tail'd cDNA was annealed with equimolar amounts of pBR322 that had been previously digested with Pst I and extended with about 15 DGMp residues at the 3′ termini. E. coli X1776 was transformed with the hybrid plasmid according to National Institutes of Health Guidelines for Research Involving Recombinant DNA Molecules (1979). Plasmids containing AFP sequences were identified by hybridization to purified AFP cDNA, hybrid arrest of translation, and comparison of the unique restriction endonuclease fragments from full length AFP ds-cDNA with cloned ds-cDNA inserts. These plasmids hybridized to cDNA prepared from mRNA isolated by immunoprecipitation of AFP-synthesizing polysomes from an AFP-producing rat hepatoma (8) but not to cDNA prepared from normal rat liver mRNA or purified rat albumin mRNA. Clones containing AFP inserts were screened with 32P-labeled poly(dT); and a plasmid (termed pAFP-3) having a PAP-3 sequence was selected for additional restriction endonuclease mapping and for sequencing. The pAFP-3 plasmid containing rat AFP ds-cDNA insert was also found to cross-hybridize to a cloned mouse AFP ds-cDNA insert (clone pBR322-AFP2 described in Ref. 9), graciously provided by Drs. Dimitris Kioussis and Shirley Tilghman.

Isolated cloned AFP cDNA fragments were sequenced by the revised Maxam and Gilbert technique (10). Fragments were labeled (10) at the 5′ termini with [γ-32P]ATP (New England Nuclear, 3000 Ci/mmol) by polynucleotide kinase (Bethesda Research Laboratories). Reaction samples were analyzed by electrophoresis in 8% polyacrylamide gels, 0.4 mm thick, containing 8 M urea (11), using about 2 × 105 cpm/lane. Autoradiography was carried out at -70°C with Kodak X-Omatic R film.

RESULTS AND DISCUSSION

The AFP recombinant plasmid, pAFP-3, and purified ds-cDNA insert were digested with various restriction endonucleases. DNA cleavage fragments were analyzed by electrophoresis in horizontal agarose gels and vertical polyacrylamide gels and compared with DNA fragments of known length. The ds-cDNA insert was found to have a length of 640 base
Sequence Homology between Rat α-Fetoprotein and Albumin

8047

pairs (bp). A restriction endonuclease map of the pAFP-3 ds-cDNA insert is shown in Fig. 1.

For the initial nucleotide sequence determination, the two fragments generated by Sal I digestion of the ds-cDNA insert were isolated and labeled with \([γ-^32P]ATP\) by polynucleotide kinase. A continuous run of about 100 adenine residues adjacent to the 3' terminal GC tail was found in the 224-bp Sal I fragment. This length of poly(A) was unexpected since an apparent 20-fold excess of oligo(dT) primer had been used to prepare the AFP cDNA for cloning, according to standard reaction conditions (12). However, the location of the poly(A) segment identified the strand of the ds-cDNA insert coding for the amino acid sequence.

The AFP ds-cDNA insert was digested with the restriction endonuclease Bst NI and end labeled, and the fragments were isolated. Subsequent sequencing of the 248-bp Bst NI fragment overlapped the Sal I site and also confirmed the sequence previously determined for the 224-bp Sal I fragment. The 82-bp and 90-bp Bst NI fragments were end labeled, individually isolated, separated into complementary strands by gel electrophoresis (10), and sequenced.

The nucleotide sequence of the 3'-terminal portion of the coding strand of the AFP ds-cDNA insert is shown in Fig. 2. The protein synthesis termination codon for AFP mRNA is UAA. There are 168 nucleotides (56 codons) shown in the translated portion and 124 nucleotides in the noncoding portion adjacent to the poly(A) segment (about 100 nucleotides in this case). No unusual features in the coding portion are evident. The noncoding sequence, however, is relatively AT-rich. A distinctive region containing 19 purines in 20 nucleotides is observed beginning at residue 9 in the noncoding segment. This region is adenine-rich (14 adenines in 20 nucleotides). The sequence AAUAAA, characteristically found near the terminal region of the 3' untranslated segment of eukaryotic mRNAs (reviewed in Ref. 13), is located in the AFP sequence ending 19 nucleotides upstream from poly(A).

In several eukaryotic mRNAs, Benoist et al. (13) have identified a 10 nucleotide sequence family that is closely related to the model sequence UUUUACUGC. It is located at the distal region of the 3' terminus. A similar sequence, UUUUACUGU, can be found in the same region of the untranslated portion of the AFP mRNA. No function has been associated with this sequence. Further analysis of the AFP nucleotide sequence for possible secondary structures of interest with the aid of a modified computer program, originally obtained from Staden (14), did not reveal any distinctive features.

The amino acid sequence of the COOH-terminal region of AFP, inferred from the nucleotide sequence, is shown in Fig. 2. Valine has been shown previously to be the COOH-terminal residue for rat and human AFP by carboxypeptidase digestion (7, 13). In addition, the 5' terminal residues of human AFP can be inferred from the carboxypeptidase data, showing preservation of the alanine and leucine positions in this region. Comparison of the rat AFP amino acid sequence with the corresponding sequence of rat albumin (15) indicates that the two proteins are closely related (Fig. 3). In both proteins, 45% of the amino acids are identical (Fig. 3, upper section). Further comparison of the rat AFP sequence with the albumin amino acid sequence from three other species, for which sequence information is currently available in the COOH-terminal region (16-18), shows a striking homology. Considering identical amino acids, in addition to conservative amino acid substitutions (19) where the side chains have similar properties (i.e. polarity), there is a 57% amino acid sequence homology for all four proteins (Fig. 3, lower section). In this regard, rat AFP has greater amino acid sequence homology with human albumin (36/56 amino acids) than with rat albumin (32/56 amino acids). Most of the amino acid variations found between rat AFP and any of the albumins in the COOH-terminal region could be explained by a single base change in the corresponding codons. However, the extent of the homology in amino acid sequences may not directly correlate with the relative amount of nucleotide sequence homology in the mRNAs. Nevertheless, the amino acid sequence homology suggests that AFP and albumin share and conserve the same structural features in the COOH-terminal region.

An amino acid sequence homology between internal fragments of human AFP and human or bovine albumin has been reported by Ruo slahti and Terry (6). Two cyanogen bromide cleavage fragments of human AFP, derived from an internal region corresponding to the second major domain (Loops 2A and 2B) of albumin (according to Brown's model (20) and nomenclature), were analyzed and compared to the homolo-
Sequence Homology between Rat α-Fetoprotein and Albumin

Fig. 3. Amino acid sequence of the COOH-terminal regions of rat AFP and three different albumin species. The boxes in the upper section enclose amino acids that are identical in both rat AFP and rat albumin. The boxes in the lower section enclose amino acids that are either identical or closely conserved in all four protein species. Amino acid sequence data for rat, bovine, and human albumin were taken from the work of Ikenaka et al. (16), Brown (17), and Behrens et al. (18), respectively. The one-letter notation for amino acids is according to the recommended convention of Dayhoff (30).

Fig. 4. Amino acid sequence of the NH2-terminal region of different species of AFP and albumin. The boxes in both sections enclose amino acids that are identical in either all three AFP species or all three albumin species. Amino acid sequence data for the three AFP species were taken from Peters et al. (7). The albumin sequences are taken from the work of previously cited investigators (16–18).

The NH2-terminal amino acid sequences of rat, mouse, and human AFP (7), as well as rat (16), bovine (17), and human (18) albumin have been determined (Fig. 4). Within the AFP sequences, 61% of the amino acid positions are identical. An even greater homology is found within the albumin sequences, where 74% of the amino acid positions are identical. However, there is essentially no significant cross-homology between AFP and albumin NH2-terminal sequences for any species combinations. The amino acid sequence of the rat pre-pro-albumin precursor leader sequence (21) also shows no homology with any AFP sequence. Thus, a total divergence in amino acid sequence apparently has occurred between AFP and albumin in the region of the first major domain. An alternative possibility is that an additional coding segment exists at the 5'-terminal region of AFP mRNA which is not found or recognized in albumin mRNA. The mRNAs for both AFP and albumin are equivalent in size (8), but the AFP mRNA translation product is about 2000 daltons larger than that of the albumin mRNA translation product in cell-free systems. Furthermore, a copper ion binding site located at the extreme NH2-terminus of albumin (22) appears to be retained in AFP (23).

The data presented and reviewed in this communication strongly support earlier suggestions that AFP and albumin may be structurally homologous and may have arisen from a common ancestral gene. Our results, together with earlier findings (6), show an amino acid sequence homology of nearly 60% in two different regions of these proteins. The observation that antibodies prepared against denatured derivatives of AFP and albumin can cross-react (24) supports the amino acid sequence results. However, important differences between the two proteins have been characterized, with the complete lack of amino acid sequence homology at the NH2-termini being especially noteworthy. The observations that antibodies to native AFP and albumin do not cross-react (24) and that cDNAs to their mRNAs do not cross-hybridize (8) suggest that homologous regions of the corresponding molecules may be extensively interrupted by nonhomologous regions. Finally, it is of particular interest that these two structurally and functionally related proteins have acquired different regulatory controls with respect to gene expression in development.

Several reports have indicated that multiple forms of AFP may be found in amniotic fluid and in plasma (reviewed in Ref. 5). Thus, AFP could consist of a homologous family of sequences, distinct from its relationship to albumin. In most species, at least two major forms of AFP can be identified by polyacrylamide gel electrophoresis in nonnaturating conditions or by lectin binding. Although much of this heterogeneity can be explained by variation in the content of sialic acid and other carbohydrates in the prosthetic group(s), peptide differences may occur. For example, the COOH-terminal residue of rat AFP (25) can be valine or glycine (the penultimate amino acid, shown in Fig. 3). However, the differences between AFP variants in general appear to be minor. Limited amino acid sequence analyses at the NH2- and COOH-terminal regions of the two major forms of human AFP show no differences (15). Crossed immunoelectrophoresis indicates that all forms of AFP appear to have the same antigenic determinants (26). Translation of mRNA from rat yolk sac, fetal liver, and Morris Hepatoma 7777 in cell-free systems yields a single peptide with a molecular weight of about 70,000 that is immunoprecipitated by anti-AFP (8, 27–29). Melting profiles of hybrids between AFP cDNA and mRNAs prepared from rat2 and mouse (28) fetal liver and yolk sac show essentially identical single component curves. Restriction endonuclease analysis of an additional rat AFP ds-cDNA clone, also corresponding to the 3'-terminal region, shows no differences with the AFP ds-cDNA prepared from the yolk sac 18 S mRNA. Thus, AFP sequence heterogeneity, if it exists, is likely to be quite limited and might only be readily detected by detailed analysis of the genome DNA. Therefore, it is reasonable to conclude that the AFP sequence described in
this communication is a representative AFP gene product that is closely related to, but distinctly different from, albumin sequences.

REFERENCES