Vol. 83, No. 1, 1978 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 14, 1978 Pages 75-82

MURINE ALPHA-FETOPROTEIN: N-TERMINAL AMINO ACID SEQUENCE AND C-TERMINAL RESIDUE

Erwin H. Peters[1], Shinzo Nishi[2] and Taiki Tamaoki[3]

[1]Division of Medical Biochemistry
Faculty of Medicine, University of Calgary
Calgary, Alberta T2N 1N4

[2]La Jolla Cancer Research Foundation
La Jolla, California, USA 92037
On leave of absence from:
Department of Biochemistry
Hokkaido University
Sapporo, Hokkaido, Japan

[3]La Jolla Cancer Research Foundation
La Jolla, California, USA 92037
On leave of absence from:
University of Alberta Cancer Research Unit
Edmonton, Alberta, Canada

Received April 25,1978

Summary

We have purified to homogeneity murine alpha-fetoprotein (MAFP) and determined the amino acid sequence of the first twenty-four residues. The N-terminal sequence obtained shows a high degree of homology with human and rat AFP's, but not human or rat albumins. The C-terminal residue is the same as human and "slow" rat AFP, but different from the corresponding albumins. We conclude that the AFP's are derived from homologous genes which are at best distantly related to the ancestral gene for albumin. The single C-terminal residue and N-terminal sequence suggests that the multiple forms of MAFP observed by others are due to carbohydrate micro-heterogeneity.

Alpha-fetoprotein (AFP) is an onco-developmental glyco-protein having a molecular weight of approximately 69,000 daltons (1, 2). This protein, which is the major serum protein of the developing fetus has been isolated and physically characterized from eight mammalian species and is reported also to be present in developing chickens (3). Alpha-fetoprotein serum levels of non-pregnant adults are normally extremely low, being detectable only by radio-immunoassay in some species (4-6). However, during onset of primary liver

*The abbreviations used are: AFP - Alpha-fetoprotein
MAFP - murine alpha-fetoprotein
DMAA - dimethylallylamine
PITC - phenylisothiocynate

0006-291X/78/0831-0075$01.00/0

cancer, teratoma or certain other malignant diseases, circulating AFP levels show dramatic increases (7). Because of the potential of AFP as a clinical tool for cancer detection, several laboratories have reported initial attempts to determine the primary structure of human (8-10) and rat AFP's (11).

Because of the similarity of amino acid compositions and molecular weights of AFP and albumin, these two proteins are thought by some to be closely related (2,12). We have determined the N-terminal amino acid sequence for the first 24 residues of murine alpha-fetoprotein (MAFP) along with its C-terminal residue. Murine AFP shows a high degree of homology to both rattine and human AFP's. However, the amino acid sequence of MAFP, like that of rattine AFP (11) and human AFP (8-10) suggests that the relationship between AFP and albumin is not as close as has been postulated (2,12). Additionally, the unexpectedly early termination of the automated sequencing of these protein suggests that at least one, but not necessarily the only, carbohydrate attachment point may be located near the N-terminus of the protein.

Materials and Methods

MAFP was prepared to homogeneity, as judged by SDS-acrylamide gel electrophoresis from saline extracts of whole murine fetuses of 17 to 20 days of gestation, according to the method of Watabe (13).

Automated Edman degradation (14) was performed with Beckman 890C Sequencer (Beckman Instruments, Palo Alto, Ca.) using the fast DMAA Program (102974) supplied by Beckman for cycles 2 through 40. The first cycle was done by a modification of the same program but added the DMAA 15 minutes before the PITC to allow sufficient time for *in situ* disulphide bond reduction. The modified program also had an additional coupling step in the first cycle to ensure that all amino groups of the protein were PITC derivitized. The MAFP (7.8 mgs) was loaded into the sequencer cup with 2.0 mgs dithioerythitol and 1.0 mg Polybrene (Sigma, St. Louis) in 0.1 M ammonium formate. Polybrene is reported to decrease non-specific protein losses from the cup during sequencer runs (D. Gibson, personal communication). All reagents and solvents were obtained from the Pierce Chemical Company (Rockford, Illinois U.S.A.). All procedures used for sample handling HI back hydrolysis and data reduction were those of Smithies, *et al.* (15).

Hydrazinolyses of MAFP were essentially according to the method of Schroeder (16) except that the protein was extensively dried over P_2O_5 both before and after hydrozinolysis.

Amino acid analyses were performed on a Beckman 121M Amino Acid Analyzer equipped with a Beckman Systems A.A. integrator/data reduction system (for

sequence determination and hydrazinolyses) and/or a micro-bore Beckman 121 Amino Acid Analyzer equipped with a System A.A. modified for Ortho-phthaldelyde fluorescence detection (sequencer samples only; Peters, in preparation).

Amino acid sequence homologies were examined by the computerized comparison method of McLachlan (17).

Results

Figure 1 shows a sample of the homogenous MAFP preparation used for the sequence determination.

Figure 2 is a Smithies plot ($\log_{10}$ of amino acid recoveries vs. sequence step) of the sequence determination of MAFP with the amino acid residue for each step indicated below the step number, in the single letter amino acid code, (see legend Fig. 2 for the single letter and three letter amino acid abbreviations). It is obvious there are two sequences present in this run; the major sequence (88%) has the N-terminal peptide of *Leu His* the minor sequence (12%) has the N-terminal sequence *Sac Thr Leu His*; where *Sac* is either Serine, Alanine or Cysteine. In this case the *Sac* is assumed to be *Ser* by homology with human AFP and because of its low yield. Both sequences are the same except for the additional two amino acids present on the N-terminus of the minor sequence. This phenomenon was first noticed with human AFP by Roushlati, *et al.* (8) and is attributed to limited proteolysis of the AFP N-terminus.

The C-terminal residue of MAFP was found to be valine by hydrazinolysis with no evidence of any other amino acids being present.

Discussion

We have purified murine AFP to homogeneity and determined the first twenty-four residues of its amino acid sequence. Table 1 shows a comparison of the N-terminal amino acid sequence of murine AFP, rattine AFP and human AFP. It is immediately obvious from the visual inspection of the sequence, that the murine AFP that we have prepared and sequenced is closely related to both the human and rattine AFP's. This is confirmed by computerized amino acid sequence comparisons in pairwise combinations of these proteins. By the

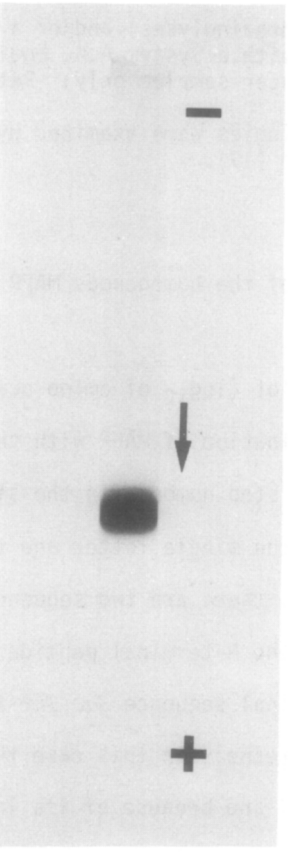

Fig. 1: Polyacrylamide gel electrophoresis of purified MAFP according to to the method of Davis (23). The 7.5% gel contained 50 μg of MAFP.

McLachlan scoring method, human vs. murine AFP has a score of 134 out of a possible maximum McLachlan score of 163; murine vs. rattine 111/137, and human vs. rattine 103/137. These values leave no doubt that these proteins over the lengths compared (20,17 and 17 residues, respectively) are indeed homologous and are therefore probably derived from closely related genes. However, it is surprising that the murine sequence is as closely related to the human sequence as to the rat. This is somewhat unexpected when one considers the various evolutionary distances between the species. It seems likely that these relationships may change when more AFP sequence data becomes available

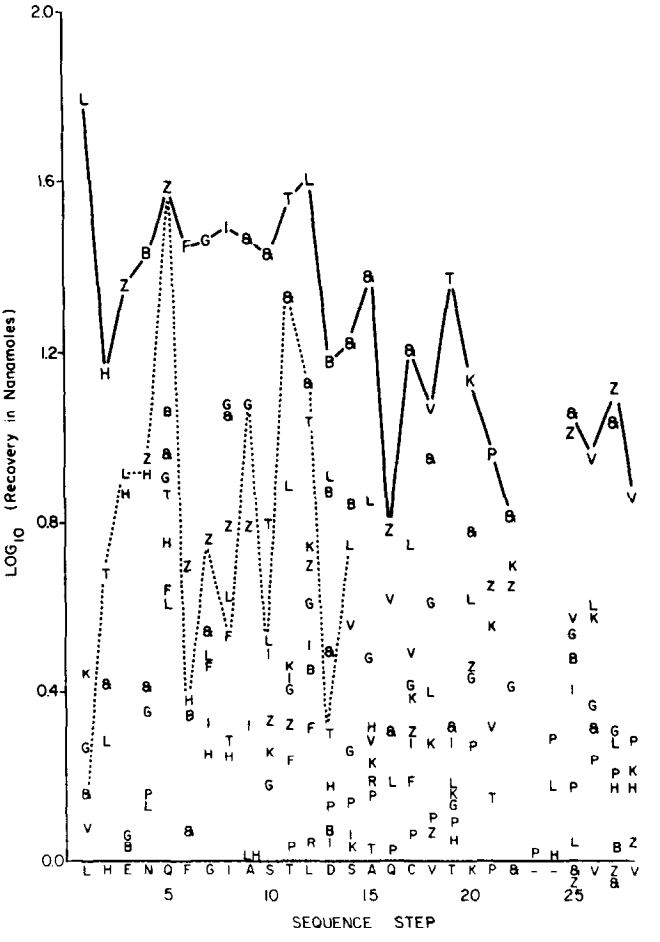

Fig. 2: A Smithies plot of the sequence determination of MAFP with the deduced sequence under the abscissa. The initial yield of the N-terminal was 57% of the 113 n moles applied. The amide status of residues 3, 4, 5, 13 and 16 was determined by quantitation of ammonia at those positions. Residues 11 and 14 were determined to be serines and residues 10 and 15 were determined to be alanines because of their low and high yields, respectively, and along with residue 17 by homology with either the rat or human AFP sequences. The heavy solid line connects the main sequence; the dashed line connects the minor sequence.

The three letter and single letter amino acid abbreviations used are: Ala, A, alanine; Arg, R, arginine; Asn, N, asparagine; Asp, D, aspartic acid; Asx, B, aspartic acid or asparagine; Cys, C, cysteine; Gln, Q, glutamine; Glu, E, glutamic acid, Glx, Z, glutamine or glutamic acid; Gly, G, glycine, Hir, H, histidine; Ile, I, isoleucine; Leu, L, leucine; Lys, K, lysine; Phe, F, phenylalanine; Pro, P, proline; Sac, & serine alanine or cysteine; Ser, S, serine; Thr, T, threonine, Tyr, Y, tyrosine, Val, V, valine.

Table 1

Amino acid sequences of murine, rattine and human alphafetoprotein.

	1				5					10		12
Murine	Ser	Thr	Leu	His	Glu	Asn	Gln	Phe	Gly	Ile	Ala	Ser
Rattine (11)	Arg	Val*			Thr						Glx	
Human (9)					Arg		Glu	Tyr				

	13		15					20				24
Murine	Thr	Leu	Asp	Ser	Ala	Gln	Cys	Val	Thr	Lys	Pro	Sac
Rattine												
Human	Ala				Tyr			Ala				

*Only those residues differing from the murine sequence are shown.

as we are only comparing the N-terminal 20 residues out of an estimated 583 amino acids in the entire protein. It is of interest to note in this connection that antisera to human AFP react with AFP from a number of mammalian species including the mouse and the rat, but antisera to rat AFP do not react with human AFP, although they react with mouse AFP (18). Apparently, the amino acid residues near the amino terminus are not involved in the determination of that portion of the tertiary structure which determines the antigenic specificity.

Similar computerized comparisons of the N-terminal amino acid sequences of human, murine or rattine AFP's vs. human, murine or rattine albumin revealed no N-terminal sequence homologies. This result, along with the lack of nucleic acid hybrydization between murine AFPcDNA and murine albumin polysomal RNA (19,20), supports the concept put fourth by Peters, *et al.* (11), that the AFP and albumin genes are at best only very distantly related and not derived from an immediate common ancestoral gene as suggested by Roushlati (21).

The single C-terminal valine residue seen in murine AFP is the same as is seen in human AFP (8,9), and "slow" rat AFP (11) but different from that of "fast" rat AFP (11). Little can be said, however regarding C-terminal homology because of the paucity of data. The fact that only one C-terminal

residue and related N-terminal sequences are observed supports the concepts that the micro-heterogeneity observed by Gustine and Zimmerman (22) in MAFP is due to different degrees of carbohydrate attachment to a single molecular species of AFP.

Finally, in all sequence determinations of AFP's (8,11) one is immediately struck by the spectacular lack of success in sequencing this protein from the N-terminus as is evidenced in all studies. Peters *et al.* (11), reported similar observations with rattine AFP although they were able to determine unambiguously the first 35 residues of rattine albumin, a protein of similar size and amino acid composition. They suggested that this lack of success could be due to a carbohydrate attachment site at a asparagine residue (position 15) of the rat sequence. Although we report an aspartic acid at the homologous position of murine AFP, it is possible that this residue has been deglycosylated and deamidated during the acid hydrolysis and that the fall in sequencing yield that we observe is due to this residue (residue 13 of major sequence) being a carbohydrate attachment point on the AFP molecules. The only other possibilities for this lack of success are a highly unusual primary structure in this region or an unknown and unsequenceable amino acid at this position. Both of these possibilities are highly unlikely.

Acknowledgements

This work was supported by grants to T.T. from the National Cancer Institute of Canada and the Medical Research Council of Canada and a grant to W.H. Fishman from the National Cancer Institute, U.S. Public Health Service (CA 21967-01). The gift of computing time by the Computer Services Department of the University of Calgary is gratefully acknowledged.

We thank Drs. G. H. Dixon and F. L. Lorscheider and Mr. P.C.W. Lai for their helpful comments during the preparation of this manuscript.

References

1. Bergstrand, C. G. and Czar, B. (1956) Scand. J. Clin. Lab. Invest. 8, 174.
2. Rouslahti, E. and Seppala, M. (1971) Int. J. Cancer 7, 218-225.

3. For an excellent review see: Rouslahti, E., Belanger, L., Kjessler, B., Kohn, J., Masseyeff, R., Nishi, S., Norgaard-Pedersen, B., Nunez, E., Seppala, M., Talermon, A., Uriel, J. and Hirai, H. (1977) *Proceedings of the 5th meeting of IRGCP, Copenhagen.* Scand. J. Immun. Suppl. (in press).
4. Rouslahti, E. and Seppala, M. (1972) Nature 235, 161-162.
5. Sell, S. and Gord, D. (1973) Immunochemistry 10, 439-442.
6. Lai, P.C.W., Hay, D.M. and Lorscheider, F.L. (1978) J. Immunol. Methods (in press).
7. Sell, S. (1976) in Onco-developmental Gene Expression (Fishman, W.H. and Sell, S., eds.) pp. 373-374. Academic Press, New York.
8. Rouslahti, E., Pihko, H., Vaheri, A., Seppala, M., Virolainen, M. and Konttinen, A. (1974) Johns Hopkins Med. J. Suppl. 3, 249-255.
9. Aoyagi, Y., Ikenaka, R. and Ichida, F. (1977) Cancer Res. 37, 3663-3667.
10. Yachnin, W., Hsu, R., Heinrikson, R.L. and Miller, J.B. (1977) Biochem. Biophys. Acta 493, 418-428.
11. Peters, E.H., Lai, P.C.W., Hay, D.M. and Lorscheider, F.L. (1977) *Proceedings of the 5th meeting of IRGCP, Copenhagen.* Scand. J. Immun. Suppl. (in press).
12. Rouslahti, E. and Engvall, E. (1976) Proc. Nat. Acad. Sci. USA 73, 4641-4644.
13. Watabe, H. (1974) Int. J. Cancer 13, 377-388.
14. Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80-91.
15. Smithies, O., Gibson, D.M., Fanning, E.M., Goodfliesh, R.M., Gilman, J.G. and Ballantyne, D.L. (1971) Biochem. 10, 4912-4921.
16. Schroeder, W. A. (1972) Methods in Enzymology XXVB, 138-142.
17. McLachlan, A.D. (1971) J. Mol. Biol. 61, 409-424.
18. Nishi, S., Watabe, H. and Hirai, H. (1975) Annal, N.Y. Acad. Sci. 259, 109-118.
19. Innis, M. and Miller, D. (1977) J. Biol. Chem. 252, 8469-8475.
20. Tamaoki, T., Miura, L. and Law, S. (1977) *Proceedings of the 5th meeting of IRGCP, Copenhagen.* Scand. J. Immun. Suppl. (in press).
21. Rouslahti, E. and Terry, W.D. (1976) Nature 260, 804-805.
22. Gustine, D.L. and Zimmerman, E.F. (1973) Biochem. J. 132, 541-551.
23. Davis, B.J. (1964) Ann. N.Y. Acad. Sci. 121, 404-429.