Human tissue-type plasminogen activator is related to albumin and alpha-fetoprotein

Michael E. Baker

Department of Medicine, M-023, University of California, San Diego, La Jolla, CA 92093, USA

Received 4 December 1984

Using a computer program designed to detect evolutionary relationships between proteins, I find that residues 72–110 of the mature sequence of human tissue-type plasminogen activator (t-PA) and 39 residues at the carboxy terminus of human albumin have a comparison score that is 8.8 standard deviation units higher than that obtained with a comparison of randomized sequences of these proteins. The probability (p) of getting this score by chance is $\sim 10^{-18}$, indicating that part of t-PA and albumin are derived from a common ancestor. I also find that alpha-fetoprotein, a relative of albumin is related to t-PA. Part of this region on t-PA has been previously shown to be related to epidermal growth factor. t-PA, albumin, alpha-fetoprotein, and epidermal growth factor have diverse biological activites. The finding that these proteins are related suggests some new approaches for studying their functions.

Tissue-type plasminogen activator *Albumin* *Alpha-fetoprotein* *Epidermal growth factor* *Protein evolution*

1. INTRODUCTION

Human tissue-type plasminogen activator (t-PA) is a serine protease that is part of the enzymatic system for dissolving fibrinogen in blood clots [1–3]. Rat and mouse alpha-fetoprotein (AFP) are estrogen binding proteins (K_d ~1 nM) [4–7] that are related to serum albumin [8–13]. Because I have been interested in a possible relationship between alpha-fetoprotein and serine proteases [14–16], I compared the amino acid sequence of t-PA [17] with those of AFP and albumin using computer programs developed at the National Biomedical Research Foundation for studying relationships between proteins [18–22]. Here, I present evidence that 39 residues of t-PA are related to a segment at the carboxy terminus of albumin and AFP. These residues in t-PA overlap with an amino acid sequence that Banyai et al. [23] have shown is related to epidermal growth factor (EGF) [24–28]. These finding suggest some novel approaches for understanding the functions of these proteins.

2. METHODS

As described by Barker and Dayhoff [22], the Relate program compares all possible segments of a given length (in this instance 20 amino acids) from one sequence with all segments of the same length from a second sequence. A segment score is accumulated from the pair scores of the amino acids occupying corresponding positions with the two segments. The pair scores are specified using an empirically derived mutation data matrix [18–22]. The mean of a number of highest scores is determined for the given sequences and for 250 comparisons of random permutations of the sequences. The segment comparison score is calculated as the difference between the mean of the real sequence and the average value determined from the randomized sequences divided by the standard deviations (SD) of the values from the randomized sequences. The segment comparison score is thus expressed in SD units. It is assumed that a score >5 SD units ($p < 2.8 \times 10^{-7}$) indicates evolutionary relatedness of two proteins and scores

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30

between 3 SD ($p < 10^{-3}$) and 5 SD ($p < 2.8 \times 10^{-7}$) support relationship if there are other indications such as similarity of function.

3. RESULTS AND DISCUSSION

The tripartite structure of albumin and AFP suggested the protocol for comparing these proteins with t-PA. Albumin's ~590 amino acid residues consist of 3 homologous domains of about 190 amino acid residues each, which Brown has proposed evolved by gene duplication from a ~190 residue primordial domain [29–32]. Brown also proposed that this ~190 residue primordial domain itself appeared to have arisen by gene duplication from a primordial subdomain of about 77 amino acid residues. The structure of the genes for albumin [33] and AFP [34] supports the main points of Brown's hypothesis. Based on that information, I used the Relate computer program to compare the 3 domains of human, bovine, and rat albumin and human, rat, and mouse AFP with the entire sequence of t-PA. The analysis indicated that residues 72–110 of the mature t-PA sequence were related to albumin and AFP. Table 1 shows the results of the Relate analysis comparing residues 72–110 of t-PA and the 3 domains of the different albumins and AFPs. All 3 domains in albumin and rat AFP and 2 of the domains in mouse and human AFP showed a statistically significant comparison score (>3 SD units) with residues 72–110 of human t-PA. The highest scores are found in the comparison with the third domain of albumin and AFP. These scores, which range from 5.7 SD ($p \sim 10^{-8}$) to 8.8 SD ($p \sim 10^{-18}$), indicate that these proteins are related. The alignment between residues 72–110 of t-PA and residues 552–590 of the mature sequence of human albumin and residues 548–586 of the mature sequence of mouse alpha-fetoprotein presented in fig.1 shows that 3 cysteines are aligned in all 3 proteins and that there are several identities and conservative substitutions in these segments.

The last 39 residues of domain III of albumin and AFP, which have the 8.8 SD comparison score with residues 72–110 of human t-PA, comprise half of exon L(albumin) [33] or exon C(AFP) [34] and all of exon M(albumin) or exon D(AFP). Domain III is thought to be most closely related to the primordial domain that duplicated twice about 500–700 million years ago to form the ~590 residue albumin structure that we see today. Furthermore, these exons contain a 6 or 9 residue sequence [12,35] that is hypothesized to have duplicated several times to form the primordial ~77 residue subdomain about 10^9 years ago.

There are also chemical similarities between albumin, AFP, and serine proteases. The most notable one comes from Shaw's studies, which were reported in 1963, that DFP reacted with a tyrosine residue (later identified as Tyr 411) on human serum albumin [36,37]. Also, several

Table 1

Segment comparison scores of albumin and alpha-fetoprotein with human tissue-type plasminogen activator

	Plasminogen activator (residues 72–110) compared with:					
	Domain I		Domain II		Domain III	
	Score (SD units)	P	Score (SD units)	P	Score (SD units)	P
Rat albumin	3.5	$\sim 10^{-3}$	4.5	$\sim 3.5 \times 10^{-6}$	5.7	$\sim 10^{-8}$
Bovine albumin	4.1	$\sim 3 \times 10^{-5}$	3.2	$\sim 10^{-3}$	6.7	$\sim 10^{-11}$
Human albumin	3.2	$\sim 10^{-3}$	3.6	$\sim 10^{-4}$	8.8	$\sim 10^{-18}$
Mouse AFP	3.9	$\sim 5 \times 10^{-5}$	1.8	–	5.9	$\sim 5 \times 10^{-9}$
Rat AFP	3.5	$\sim 10^{-4}$	3.6	$\sim 10^{-4}$	5.7	$\sim 10^{-8}$
Human AFP	4.1	$\sim 3 \times 10^{-5}$	1.1	–	6.5	$\sim 5 \times 10^{-11}$

These segment comparison scores were obtained using the Relate program (segment length 20 amino acids) and 250 random permutations of the sequences for statistical analysis

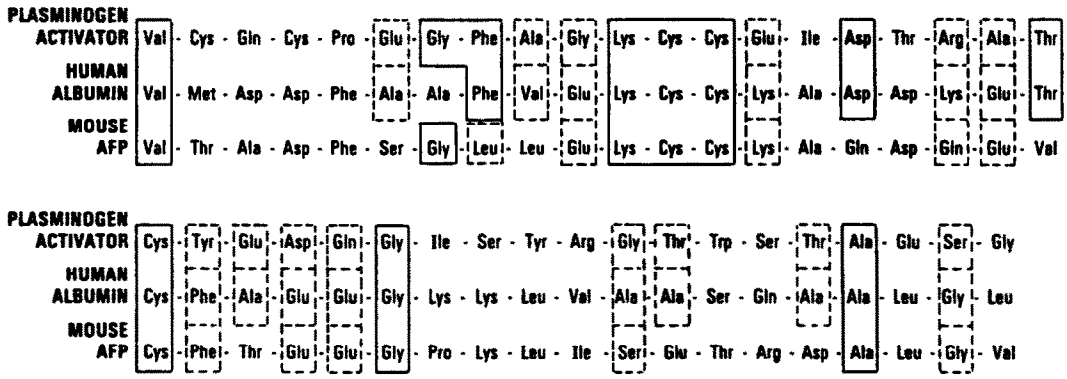

Fig.1. Alignment of residues 72–110 of human tissue-type plasminogen activator and residues 552–590 of human albumin and 548–586 of mouse alpha-fetoprotein. Residue numbering refers to the mature sequence of these proteins. Solid boxes show identities and the dotted boxes show readily accepted conservative substitutions.

studies have established that albumin weakly catalyzes the hydrolysis of ester substrates [38–41]. Means and Wu [42] showed that the site on albumin that reacts with DFP also hydrolyzes *p*-nitrophenyl acetate. All of the above suggests that albumin has some chemical properties in common with serine proteases except that in albumin it is a nucleophilic hydroxyl on a tyrosine instead of on a serine that reacts with DFP. Like albumin, rat AFP forms a covalent adduct with [^{3}H]DFP (unpublished results). Moreover, AFP has high affinity for *p*-nitrophenyl esters [15].

Of relevance to the findings reported in table 1 and fig.1 is the report by Banyai et al. [23] that

Table 2

Segment comparison scores of human tissue-type plasminogen activator with epidermal growth factors

	Plasminogen activator (residues 43–95) with:	
	Score (SD units)	*P*
Human EGF	5.8	$<10^{-8}$
Mouse EGF	6.8	$<10^{-11}$
Rat transforming growth factor	2.4	$\sim 6 \times 10^{-3}$

These segment comparison scores were obtained using the Relate program (segment length 20 amino acids) and 250 random permutations of the sequences for statistical analysis

residues 41–91 of t-PA are related to epidermal growth factor. Using the Relate computer program, I have quantifid their finding (table 2). The segment on t-PA that is related to albumin and AFP overlaps the segment on t-PA that is related to EGF, which raises some interesting questions about the evolution and function(s) of these proteins. First, do these proteins derive from a common ancestor? And, second, do these proteins share some biological activities?

The evolution of proteins by successive duplications of a small primordial protein-encoding sequence, which bears on the first question, has been discussed in detail by Ohno [12,44–46]. All of these proteins contain regions that appear to have been formed by several duplications of a small building block. This process is clearly evident in albumin and AFP, where it appears that the entire 590 amino acid residue sequence of albumin and AFP formed from a building block of 77 residues, which itself is thought to have formed by successive duplications of a 6 or 9 residue building block. The ~1175 residue mouse EGF sequence contains several ~40 residue repeating units [26,27] suggesting that a substantial part of mouse EGF formed by several duplications of a 40 residue unit [28]; and the kringle containing region of t-PA (part of which overlaps residues 72–110) has duplicated in t-PA [17] as well as in plasmin and thrombin [43,47].

The computer-based evidence for relatedness between parts of albumin, AFP, human t-PA, and

mouse EGF presented here and elsewhere; the evidence that substantial parts of these proteins formed by duplication of a primordial building block; and the likelihood that the age of the primordial protein-encoding sequence for albumin is greater than 10^9 years suggests that parts of these proteins evolved from a common ancestral gene.

To obtain a biological benchmark for discussing the possibility that these proteins share biological activities, I used the Relate program to compare mouse EGF, human EGF, and 'EGF-like' rat transforming growth factor [48–50] with each other (table 3). The comparison score of 19.2 SD for mouse and human EGF is what would be expected for closely related proteins. The comparison of rat transforming growth factor with human EGF (8.2 SD) and mouse EGF (6.3 SD) indicates a more distant relationship. (These comparison scores are in agreement with those using the Align analysis [48].) The biological importance of these comparisons is that even with substantial divergence from a putative common ancestor of mouse and human EGF, rat transforming growth factor can bind to EGF receptors with high affinity [49] as well as elicit an EGF-like response in cells with EGF receptors [50]. This finding is very interesting because there are no identities between residues 21–30 of mouse EGF and rat transforming growth factor [48,49], where the receptor binding domain of mouse EGF exists [51]. Evidently, there is sufficient chemical similarity in this region and/or information in the rest of rat transforming growth factor to permit it to bind to the EGF receptor and to be biologically active. Since the Relate score of 8.8 SD for the comparison of domain III of human albumin with residues 72–110 of t-PA is equal to or greater than the Relate comparison scores for rat transforming growth factor with human and mouse EGF, it is possible that t-PA shares some biological activities with albumin and AFP.

Table 3

Epidermal growth factor family comparisons

	Human EGF	Mouse EGF	Rat transforming growth factor
Human EGF	–	19.2 SD	8.2 SD
Mouse EGF		–	6.3 SD
Rat transforming growth factor			–

These segment comparison scores were obtained using the Relate program (segment length 20 amino acids) and 250 random permutations of the sequences for statistical analysis

Albumin and AFP bind and/or transport fatty acids, bilirubin, steroids, and a variety of small ligands [30,31,52–54]. It will be interesting to know if these ligands affect the actions of t-PA, and also the actions of mouse EGF in view of the amino acid sequence relatedness of t-PA with this protein.

The biological functions of t-PA either in dissolving clots or in influencing the growth of transformed cells [55–57] depend on its proteolytic activity. At this time there is no evidence for a biological role for the esterolytic activity of albumin. However, because of albumin and AFP's high concentration in serum, it is possible that even with low catalytic activity, they could have an enzymatic function. Of course, the assays used thus far may not be measuring albumin and AFP's true catalytic activity. Higher catalytic activity may require different substrates, the presence of metals ions, limited proteolysis, chemical modification (e.g., phosphorylation), cofactors such as pyridoxal phosphate, etc.

Alternatively, albumin and AFP may have functions that depend on the binding of, but not the hydrolysis of proteins [58], more like EGF and the kringle region of plasmin [47,59,60]. There is support for this possibility, first from the report by Weisiger et al. [61] of an albumin receptor on liver cell membranes, which may mediate the uptake of fatty acids, and second from the report by Villacampa et al. [62] of a membrane receptor for AFP on human MCF-7 breast cancer cells.

Thus, the computer-based evidence for a common ancestor for albumin, alpha-fetoprotein, and human tissue-type plasminogen activator presented here, combined with other studies indicating that tissue-type plasminogen activator is related to mouse and human epidermal growth factors [23] suggests some new avenues of research for understanding the functioning of these proteins.

ACKNOWLEDGEMENT

This research was supported by NIH grant HD 14968.

REFERENCES

[1] Christman, J.K., Silverstein, S.G. and Acs, G. (1977) in: Proteinases in Mammalian Cells and Tissues (Barrett, A.J. ed.) pp.91–149, Elsevier, Amsterdam.
[2] Wiman, B. and Collen, D. (1978) Nature 272, 549–550.
[3] Doolittle, R.F. (1981) Scientific American 245, 126–135.
[4] Aussel, C., Uriel, J. and Mercier-Bodard, C. (1973) Biochimie 55, 1431–1437.
[5] Aussel, C. and Masseyeff, R. (1978) J. Steroid Biochem. 9, 547–551.
[6] Payne, D.W. and Katzenellenbogen, J.A. (1979) Endocrinology 105, 743–753.
[7] Savu, L., Benassayag, C., Vallette, G., Christeff, N. and Nunez, E. (1981) J. Biol. Chem. 256, 9414–9418.
[8] Ruoslahti, E. and Terry, W.D. (1976) Nature 260, 804–805.
[9] Gorin, M.B., Couper, D.L., Eiferman, F., van de Rijn, P. and Tilghman, S.M. (1981) J. Biol. Chem. 256, 1954–1959.
[10] Law, S.W. and Dugaiczyk, A. (1982) Nature 291, 201–205.
[11] Jagodzinski, L.L., Sargent, T.D., Yang, M., Glackin, C. and Bonner, J. (1981) Proc. Natl. Acad. Sci. USA 78, 3521–3525.
[12] Ohno, S. (1981) Proc. Natl. Acad. Sci. USA 78, 7657–7661.
[13] Morinaga, T., Sakai, M., Wegmann, T.G. and Tamaoki, T. (1983) Proc. Natl. Acad. Sci. USA 80, 4604–4608.
[14] Baker, M.E., Vaughn, D.A. and Fanestil, D.D. (1978) J. Supra Molec. Struct. 9, 421–426.
[15] Baker, M.E., Frecker, D.G.N. and Fanestil, D.D. (1982) J. Steroid Biochem. 16, 503–507.
[16] Baker, M.E. (1984) FEBS Lett. 175, 41–44.
[17] Pennica, D., Holmes, W.E., Kohr, W.J., Harkins, R.N., Vehar, G.A., Ward, G.A., Bennett, W.F., Yelverton, E., Seeburg, P.H., Heyneker, H.L. and Goeddel, D.V. (1983) Nature 301, 214–222.
[18] Dayhoff, M.O., Eck, R.V. and Park, C.V. (1972) in: Atlas of Protein Sequence and Structure (Dayhoff, M.O. ed.) vol.5, pp.89–99, National Biomedical Research Foundation, Washington, DC.
[19] Barker, W.C. and Dayhoff, M.O. (1972) in: Atlas of Protein Sequence and Structure (Dayhoff, M.O. ed.) vol.5, pp.1–110, National Biomedical Research Foundation, Washington, DC.
[20] Dayhoff, M.O. (1978) in: Atlas of Protein Sequence and Structure (Dayhoff, M.O. ed.) vol.5, supp.3, pp.1–8, National Biomedical Research Foundation, Washington, DC.
[21] Schwartz, R.M. and Dayhoff, M.O. (1978) in: Atlas of Protein Sequence and Structure (Dayhoff, M.O. ed.) vol.5, pp.353–358, National Biomedical Research Foundation, Washington, DC.
[22] Barker, W.C. and Dayhoff, M.O. (1982) Proc. Natl. Acad. Sci. USA 79, 2836–2839.
[23] Banyai, L., Varadi, A. and Patthy, L. (1983) FEBS Lett. 163, 37–41.
[24] Carpenter, G. and Cohen, S. (1979) Annu. Rev. Biochem. 48, 193–216.
[25] Savage, C.R. jr, Inagami, T. and Cohen, S. (1972) J. Biol. Chem. 247, 7612–7621.
[26] Gray, A., Dull, T.J. and Ullrich, D. (1983) Nature 303, 722–725.
[27] Scott, J., Urdea, M., Quiroga, M., Sanchez-Pescado, R., Fong, N., Selby, M., Rutter, W.J. and Bell, G.I. (1983) Science 221, 236–240.
[28] Doolittle, R.F., Feng, D.F. and Johnson, M.S. (1984) Nature 307, 558–560.
[29] Brown, J.R. (1976) Fed. Proc. 35, 2141–2144.
[30] Brown, J.R. (1978) in: Albumin: Structure, Biosynthesis and Function (Peters, T. and Sjoholm, I. eds) pp.1–10, Pergamon Press, New York.
[31] Brown, J.R. and Shockley, P. (1982) in: Lipid-Protein Interactions, vol.1 (Jost, P. and Griffith, O.H. eds) pp.25–68, J. Wiley and Sons, New York.
[32] McLachlan, A.D. and Walker, J.E. (1977) J. Mol. Biol. 112, 543–558.
[33] Sargent, T.D., Jagodzinski, L.L., Yang, M. and Bonner, J. (1981) Mol. Cell. Biol. 1,871–883.
[34] Eiferman, F.A., Young, P.R., Scott, R.W. and Tilghman, S.M. (1981) Nature 294, 713–718.
[35] Alexander, F., Young, P.R. and Tilghman, S.M. (1981) J. Mol. Biol. 173, 159–174.
[36] Sanger, F. (1963) Proc. Chem. Soc. 76–83.
[37] Shaw, D.C. (1965) Aust. J. Sci. 28, 11–18.
[38] Casida, J.E. and Augustinsson, K.-B. (1959) Biochim. Biophys. Acta 36, 411–426.
[39] Tove, S.B. (1962) Biochim. Biophys. Acta 57, 230–235.
[40] Tildon, J.T. and Ogilvie, J.B. (1972) J. Biol. Chem. 247, 1265–1271.
[41] Means, G.E. and Bender, M.L. (1975) Biochemistry 14, 4989–4994.
[42] Means, G.E. and Wu, H.-L. (1979) Arch. Biochem. Biophys. 194, 526–530.

[43] Young, C.L., Barker, W.C., Tomaselli, C.M. and Dayhoff, M.O. (1978) in: Atlas of Protein Sequence and Structure (Dayhoff, M.O. ed.) vol.5, suppl.3, pp.73–93, National Biomedical Research Foundation, Washington, DC.
[44] Ohno, S. (1970) Evolution by Gene Duplication, Springer, Berlin.
[45] Ohno, S. (1982) Perspect. Biol. Med. 25, 559–572.
[46] Ohno, S. and Epplen, J.T. (1983) Proc. Natl. Acad. Sci. USA 80, 3391–3395.
[47] Magnusson, S., Petersen, T.E., Sottrup-Jensen, L. and Claeys, H. (1975) in: Proteases and Biological Control (Reich, E. et al. eds) pp.123–149, Cold Spring Harbor Laboratory, New York.
[48] Marquardt, H., Hunkapiller, M.W., Hood, L.E., Twardzik, D.R., DeLarco, J.E., Stephenson, J.R. and Todaro, G.J. (1983) Proc. Natl. Acad. Sci. USA 80, 4684–4688.
[49] Marquardt, H., Hunkapiller, M.W., Hood, L.E. and Todaro, G.J. (1984) Science 223, 1079–1082.
[50] Tam, J.P., Marquardt, H., Rosberger, D.F., Wong, T.N. and Todaro, G.J. (1984) Nature 309, 376–378.
[51] Komoriya, A., Hortsch, M., Meyers, C., Smith, M., Kanety, H. and Schlessinger, J. (1984) Proc. Natl. Acad. Sci. USA 81, 1351–1355.
[52] Peters, T. jr (1977) Clin. Chem. 23, 5–12.
[53] Fchske, K.J., Muller, W.E. and Wollert, U. (1981) Biochem. Pharmac. 30, 687–692.
[54] Kragh-Hansen, U. (1981) Pharmacol. Rev. 33, 17–53.
[55] Quigley, J.P., Ossowksi, L. and Reich, E. (1984) J. Biol. Chem. 249, 4306–4311.
[56] Reich, E. (1975) in: Proteases and Biological Control (Reich, E. et al. eds) pp.333–342, Cold Spring Harbor Laboratory, New York.
[57] Quigley, J.P. (1979) Cell 17, 131–141.
[58] Baker, M.E. (1985) in: Alpha-Fetoprotein in Congenital Disorders (Porter, I.H. and Mizeijewski, G. eds) pp. 79–105, Academic Press, NY.
[59] Thorsen, S. (1975) Biochim. Biophys. Acta 393, 55–65.
[60] Morris, J.P., Blatt, S., Powell, J.R., Strickland, D.K. and Castellino, F.J. (1981) Biochemistry 20, 4811–4816.
[61] Weisiger, R., Gollan, J. and Ockner, R. (1981) Science 211, 1048–1051.
[62] Villacampa, M.J., Moro, R., Naval, J., Failly-Crepin, C., Lampreave, F. and Uriel, J. (1984) Biochem. Biophys. Res. Commun. 122, 1322–1327.