Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 5360–5363, September 1984
Biochemistry

Neoplastic human fibroblast proteins are related to epidermal growth factor precursor

(oncogenes/amino acid composition/polypeptide p788/polypeptide p789)

STEPHEN BURBECK, GERALD LATTER, EDWARD METZ, AND JOHN LEAVITT

The Armand Hammer Cancer Research Center, Linus Pauling Institute of Science and Medicine, 440 Page Mill Road, Palo Alto, CA 94306

Communicated by Linus Pauling, May 14, 1984

ABSTRACT **We report the amino acid composition of two polypeptides, p788 and p789. These polypeptides are reliable markers for neoplastic transformation in human fibroblasts. Their compositions are unusually rich in cysteine and serine. Because the recently reported amino acid sequence of mouse epidermal growth factor precursor (prepro-EGF) is also rich in those two amino acids and because the role of p788 and p789 as markers for neoplastic transformation is consistent with the fact that epidermal growth factor has been shown to play some role in transformation, we investigated the hypothesis that p788 and p789 are related to prepro-EGF. We compared the amino acid composition of p788 with that of all possible interior domains of prepro-EGF of appropriate length. We found that the composition of p788 is remarkably similar to that of residues 630–880 of prepro-EGF. The similarity is sufficiently strong to support the conclusion that it reflects amino acid sequence homology.**

Two polypeptides, p788 and p789 (M_r, 26,000 and 27,000, respectively) (1, 2), are reliable markers for neoplastic transformation in human fibroblasts. After carcinogen-induced neoplastic transformation, their rates of synthesis are elevated 30-fold; this is the most dramatic example of quantitatively enhanced gene expression among the 700 most abundant polypeptides in the human KD diploid fibroblast strain (1). Similarly elevated levels of synthesis of p788 and p789 [previously called A2 and A3 (3)] are found in a simian virus 40-transformed derivative of WI-26 embryonic human lung fibroblasts (3) that fails to produce tumors in athymic mice (4) and in the highly tumorigenic (4) HT-1080 fibroblast strain derived from a human osteosarcoma (5). By contrast, 10 diploid human fibroblast strains show little or no synthesis of p788 and p789 (1).

We report here the amino acid composition of p788. It shows strong similarity to the composition of residues 630–880 of mouse epidermal growth factor precursor (prepro-EGF) (6, 7). The similarity is sufficiently strong (8) to demonstrate homology with that portion of prepro-EGF. This finding is consistent with evidence from a variety of sources (9–19) for a connection between epidermal growth factor (EGF) and neoplastic transformation.

MATERIALS AND METHODS

Amino acid compositions give surprisingly useful information for the purpose of identifying polypeptides (8, 20–22). The method used here for amino acid analysis has been described in detail elsewhere (23). It is based on computerized microdensitometry of autoradiograms of two-dimensional gels from 20 separately radiolabeled cultures of the HuT-14 cell line. Each culture is labeled with one of the 20 amino acids and the total proteins are then separated and resolved by two-dimensional polyacrylamide gel electrophoresis (Fig. 1). From these 20 gels one can, in principle, obtain the amino acid composition of any or all resolved polypeptides visible on the gel. The amount of each radiolabeled amino acid incorporated into each polypeptide of interest is determined relative to the amount simultaneously incorporated into a "reference" protein that has a precisely known amino acid composition. In this case the reference protein was mutant β-actin (24).

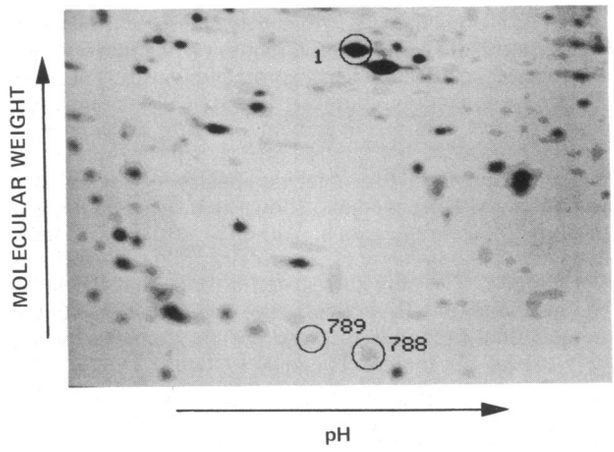

FIG. 1. Computer image of an autoradiogram labeled with [^{14}C]asparagine. Spot 1 is the mutant β-actin standard. p788 and p789 exhibit apparent molecular weights of 26,000 and 27,000, respectively, and isoelectric points of 5.3 and 5.2, respectively, in the presence of reducing reagent (1).

The *in vitro* transformed cell line HuT-14 has been described (3). High-density subconfluent monolayer cultures in multiwell plates were incubated in 2 ml of radiolabeling medium for 24 hr. A modified Eagle's medium Selectamine kit (GIBCO) was used to prepare the 20 radiolabeling media, each containing a different ^{14}C-labeled amino acid at 10–50 μCi/ml (excepting [^{35}S]methionine) (1 Ci = 37 GBq) from New England Nuclear. Radiolabeled cell monolayers were rinsed three times with cold phosphate-buffered saline, drained, and then lysed directly in O'Farrell's "lysis buffer A" (25). Protein samples were stored at −76°C prior to electrophoresis. Two replicate two-dimensional polyacrylamide electrophoresis gels were made from each protein sample with the Iso-Dalt system from Electro-Nucleonics (Oak Ridge, TN). However, the second-dimension gels were thinner (1 mm as opposed to 1.5 mm) to diminish the quenching effect of the gel. Kodak XAR-2 film was used for autoradiography. Exposure time varied from 3 to 8 weeks in order that the actin standard be well exposed in every case. The film was developed in a Kodak X-Omat model 5 processor and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: EGF, epidermal growth factor; prepro-EGF, EGF precursor.

digitized on an Optronics P-1000 film scanner at a pixel size of 200 μm. Measurement of the apparent abundance of each polypeptide of interest for each of the amino acids was done with an interactive computer graphics system as described elsewhere (23). The calculation of amino acid composition was as before (23) with two exceptions. First, we used mutant β-actin as the standard. Second, for each spot we summed the integrated counts for the two replicate gels and divided that sum by the similarly summed counts for the reference spot. In the previously published method the integrated counts for each spot were divided by the counts in the reference spot on the same gel and the resulting ratios for the two replicates were averaged.

RESULTS AND DISCUSSION

The amino acid compositions of p788 and p789 are given in Table 1. These two polypeptides were known to be closely related; their tryptic peptide fingerprints are similar (3), and they have equally short half-lives (<2 hr; ref. 26). Their similar amino acid compositions are further confirmation of that fact. However, subsequent quantitative analysis will be confined to p788 because the weaker p789 is so near the lower limit of film sensitivity with some of the labeled amino acids that measurement accuracy was unacceptable. Measurement error for the more abundant p788 is on the order of 15%.

p788 and p789 are unusually rich in cysteine and serine, a property shared by the recently sequenced mouse prepro-EGF. Also, the apparent molecular weights of p788 and p789 are near that of the high molecular weight form of human EGF [approximately 28,000 (27)]. Those observations suggest the possibility that p788 may be related to high molecular weight human EGF. A test of that hypothesis is made more complicated by the fact that high molecular weight human EGF derives from an unknown region of prepro-EGF. In particular, it is not known whether high molecular weight EGF contains the short EGF sequence. We therefore compared our data with all possible regions of the 1217-amino acid sequence of mouse prepro-EGF (6). The result of comparison of the composition of p788 with successive 251-residue subsequences of prepro-EGF is shown in Fig. 2. The

Table 1. Amino acid compositions of transformation markers versus a subsequence of prepro-EGF

	p789, mol %	p788		prepro-EGF 630–880, no. of residues
		mol %	Residues,* no.	
Ala	0.034	0.055	13.8	14
Arg	0.044	0.049	12.3	14
Asn	0.044	0.038	9.5	10
Asp	0.091	0.086	21.6	21
Cys	0.052	0.050	12.6	13
Gln	0.039	0.034	8.4	8
Glu	0.079	0.083	20.7	16
Gly	0.133	0.118	29.6	22
His	0.006	0.010	2.4	5
Ile	0.028	0.036	9.0	11
Leu	0.096	0.105	26.4	23
Lys	0.033	0.022	5.5	9
Met	0.055	0.030	7.5	5
Phe	0.001	0.007	1.7	4
Pro	0.005	0.026	6.6	11
Ser	0.164	0.112	28.1	26
Thr	0.053	0.044	11.0	11
Trp	0.004	0.007	1.7	4
Tyr	0.008	0.033	8.2	5
Val	0.031	0.058	14.5	17

*Assuming a length of 251 residues.

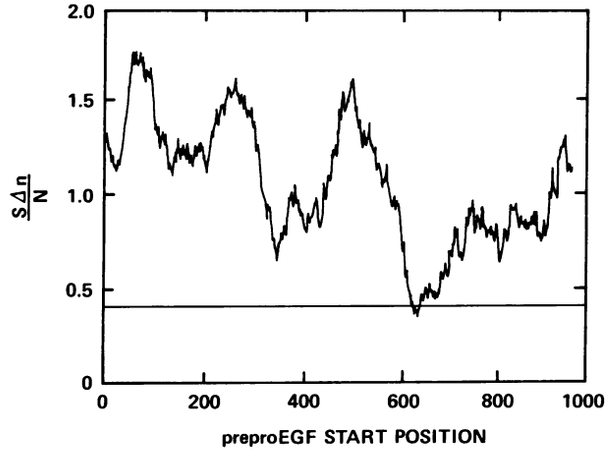

FIG. 2. Estimated sequence similarity, based on amino acid composition similarity, between p788 and all possible contiguous subsequences of length 251 from mouse prepro-EGF. $S\Delta n/N$ values of <0.42 (i.e., below the horizontal line) are significant at better than the 0.05 level (8).

vertical axis is a measure of similarity, the $S\Delta n/N$ statistic (6, 20–22), which is a probabilistic estimate of the proportion of *sequence* mismatch based only on *composition* data. For any two compositions obtained from polypeptides of nearly the same length, the smaller the $S\Delta n/N$ statistic, the more likely one is to find that the amino acid sequences are similar. The method has both theoretical and empirical justification. On a theoretical basis, given certain assumptions about the independence and randomness of the occurrence of amino acids in polypeptides, a value of $S\Delta n/N$ of <0.42 is significant at the 0.05 level (8). A practical test of the statistic by Cornish-Bowden indicates that the 0.42 cutoff may be conservative; in practice, a value of <0.42 seems to be significant at the 0.01 level (8).

In Fig. 2, $S\Delta n/N$ falls below 0.42 in only one small region (start positions 622–634) and reaches a minimum of 0.35 at start position 630. The assumption that p788 has 251 residues is not critical; other lengths from 236 to 254 give similar results. In all cases, the $S\Delta n/N$ measure reaches a minimum of <0.42 at or very near start position 630.

An analogous comparison between the known 53-residue human EGF (28) with mouse prepro-EGF (Fig. 3) shows the sensitivity as well as the validity of our search procedure. In addition to the region occupied by mouse EGF (977–1029), seven other EGF-like peptide regions, identified and aligned with EGF on the basis of a common Cys-X-Cys sequence (6), show some sequence similarity to mouse EGF. Thus we can compare results obtained based on amino acid composition alone with results that depend on knowledge of the amino acid sequences. The smallest $S\Delta n/N$ value corresponds well with the location of the true mouse EGF; it occurs 12 residues past the starting position of mouse EGF. However, the $S\Delta n/N$ estimate of sequence mismatch (0.23) underestimates the true degree of sequence mismatch between mouse and human EGF (0.30). A second region (745–758) in which $S\Delta n/N$ is <0.42 begins at the start point of the fourth EGF-like peptide region (residue 745), and $S\Delta n/N$ approaches but does not cross 0.42 at three other positions: 320, 831, and 861. In each of these cases, the resulting 53-residue subsequence of prepro-EGF overlaps one of the EGF-like regions. But the $S\Delta n/N$ measure does not locate the second, third, and seventh EGF-like regions, showing that a search for sequence homology on the basis of composition alone is not as sensitive as sequence-based methods. This is to be expected inasmuch as compositions do not provide as much information about polypeptides as do sequences. Nonethe-

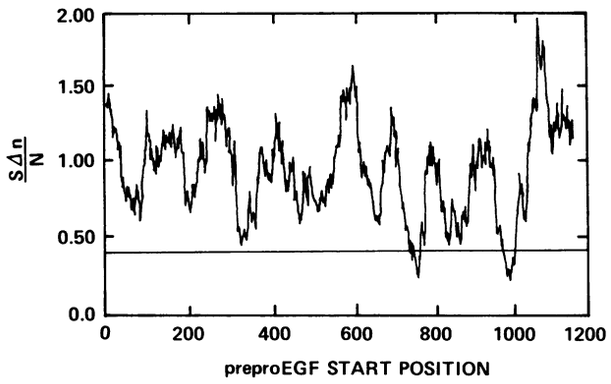

FIG. 3. Estimated sequence similarity between human EGF and mouse prepro-EGF. Mouse EGF occupies residues 977–1029 and differs from human EGF in 30% of its residues (28). The seven mouse EGF-like regions (6) begin at residues 357, 400, 441, 745, 832, 886, and 977. The two regions where $S\Delta n/N$ is <0.42 correspond to the fourth EGF-like region and the true mouse EGF.

less, $S\Delta n/N$ values of less than, or even close to, 0.42 reliably identify regions of sequence homology. Thus we are safe in concluding that the equally strong composition similarity between p788 and the 630–880 region of prepro-EGF also indicates sequence homology.

To explore the possibility that some other polypeptide is even more closely related to p788 than is prepro-EGF, the composition of p788 was compared with those of the 101 polypeptides in the 1983 Protein Sequence Database* having lengths of 220–260 residues. The mole percentage data for p788 were converted to amino acid composition (in residues) by assuming in each case that the length of p788 equaled that of the polypeptide with which it was being compared. In no case did the $S\Delta n/N$ value indicate homology (the smallest value found was 1.02). Nor does p788 match the 226-residue v-*sis* (29) ($S\Delta n/N = 2.15$) or human platelet-derived growth factor I or II (30), both of which give $S\Delta n/N > 2.6$.

The apparent homology between p788 and a region of prepro-EGF suggests that p788 may be the high molecular weight variety of human EGF. If so, then the biological activity of the high molecular weight EGF (27) is not due to EGF itself, which is disjoint from the portion of prepro-EGF implicated here. The region we implicate (630–880) contains two of the EGF-like regions (6, 31): the region beginning at residue 745 and the region beginning at residue 832. Either or both of those regions may account for EGF-like biological activity. But the apparent homology does not necessarily imply that p788 derives from prepro-EGF. Transforming growth factors from rodent and human fibroblasts show amino acid sequence homology with human and mouse EGF (9). Doolittle *et al.* (31) have suggested that these transforming growth factors derive from a precursor that shares a common ancestor with prepro-EGF. In that case, p788 may be a fragment of the transforming growth factor precursor. In either case, p788 may not be active in the causation of transformation; instead, it may merely be an inactive residue that marks the elevated synthesis of its precursor. Other fragments of the precursor (i.e., epidermal or transforming growth factor) may be the active elements of this transformation process.

Our result does, however, show that p788 is related in some manner to the EGF precursor. Thus the reliability of p788 as a marker for transformation in human fibroblasts is consistent with the relationship between neoplastic transformation and a variety of oncogenes and growth factors that have an intimate relationship to EGF. Several small regions of the amino acid sequence of the v-*erbB* oncogene are homologous to portions of human EGF receptors (19). The v-*sis* oncogene product is homologous to platelet-derived growth factor (10, 11), which is known to modulate the binding of EGF (12). And a variety of transforming and sarcoma growth factors either compete for EGF receptors (13–15) or are strongly potentiated by EGF (15). Moreover, the *src* oncogene product phosphorylates tyrosine (16), much as the epidermal and platelet-derived growth factor receptor complexes (17, 18), and therefore may function as a growth factor.

However elevated synthesis of p788 and p789 appears to be unrelated to the degree of tumorigenicity of transformed cell lines. This is shown most clearly in four KD-derived HuT substrains (HuT-11, -12, -13, and -14) (1, 3). Although all four of these separate strains synthesize p788 and p789 at nearly the same rate, the HuT-11, -12, and -13 substrains rarely produce tumors in athymic mice at 2×10^7 cells per inoculum (1) whereas the HuT-14 substrain and its subclonal derivatives always produce tumors at high frequency at 10^5–10^6 cells per inoculum (5). Hence, p788 and p789, the EGF-related growth factors, and growth factor-related oncogenes may account for only a part of the transformation process. Other differences such as enhanced proteolytic activity (32) and the expression of cytoskeletal mutations (2) may govern the different degrees of tumorigenicity exhibited by these related strains of transformed human fibroblasts.

We thank James Fleming and D. Chandra for technical assistance and Dr. Emile Zuckerkandl for comments on earlier drafts of the paper. The project received financial support from the Rippel Foundation and the Koret Foundation.

1. Leavitt, J., Goldman, D., Merril, C. & Kakunaga, T. (1982) *Carcinogenesis* **3**, 61–70.
2. Leavitt, J., Goldman, D., Merril, C. & Kakunaga, T. (1982) *Clin. Chem.* **28**, 850–860.
3. Leavitt, J. & Kakunaga, T. (1980) *J. Biol. Chem.* **255**, 1650–1660.
4. Noguchi, P. D., Johnson, J. B., O'Donnell, R. & Petricciani, J. (1978) *Science* **199**, 980–983.
5. Rasheed, S., Nelson-Rees, W. A., Toth, E. M., Arnstein, P. & Gardner, M. B. (1974) *Cancer Res.* **33**, 1027–1033.
6. Scott, J., Urdea, M., Quiroga, M., Sanchez-Pescador, R., Fong, N., Selby, M., Rutter, W. J. & Bell, G. I. (1983) *Science* **221**, 236–240.
7. Gray, A., Dull, T. J. & Ullrich, A. (1983) *Nature (London)* **303**, 722–725.
8. Cornish-Bowden, A. (1980) *Anal. Biochem.* **105**, 233–238.
9. Marquardt, H., Hunkapiller, M. W., Hood, L. E., Twardzik, D. R., De Larco, J. E., Stephenson, J. R. & Todaro, G. J. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4684–4688.
10. Waterfield, M. D., Scrace, G. T., Whittle, N., Stroobant, P., Johnsson, A., Wasteson, A., Westermark, B., Helden, C., Huang, J. S. & Deuel, T. F. (1983) *Nature (London)* **304**, 35–39.
11. Doolittle, R. F., Hunkapiller, M. W., Devare, S. G., Robbins, K. C., Aaronson, S. A. & Antoniades, H. N. (1983) *Science* **221**, 275–277.
12. Bowen-Pope, D. F., Di Corlteo, P. E. & Ross, R. J. (1983) *J. Cell Biol.* **96**, 679–683.
13. De Larco, J. E., Reynolds, R., Carlberg, K., Engle, C. & Todaro, G. J. (1980) *J. Biol. Chem.* **255**, 3685–3690.
14. Anzano, M. A., Roberts, A. B., Meyers, C. A., Komoriya, A., Lamb, L. C., Smith, J. M. & Sporn, M. B. (1982) *Cancer Res.* **42**, 4776–4778.
15. Roberts, A. B., Anzano, M. A., Lamb, L. C., Smith, J. M., Frolik, C. A., Marquardt, H., Todaro, G. J. & Sporn, M. B. (1982) *Nature (London)* **295**, 417–419.
16. Hunter, T. & Sefton, B. M. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1311–1315.

*Dayhoff, M. O., Hunt, L. T., Barker, W. C., Orcutt, B. C., Yeh, L. S., Chen, H. R., George, D. G., Blomquist, M. C. & Johnson, G. C., Protein Sequence Database, National Biomedical Research Foundation, Washington, DC, 1983.

17. Cohen, S., Carpenter, G. & King, L., Jr. (1980) *J. Biol. Chem.* **255,** 4834–4842.
18. Ek, B., Westermark, B., Wasteson, A. & Heldin, C.-H. (1982) *Nature (London)* **295,** 419–420.
19. Downward, J., Yarden, Y., Mayes, E., Scrace, G., Totty, N., Stockwell, P., Ullrich, A., Schlessinger, J. & Waterfield, M. D. (1984) *Nature (London)* **307,** 521–527.
20. Cornish-Bowden, A. (1977) *J. Theor. Biol.* **65,** 735–742.
21. Cornish-Bowden, A. (1979) *J. Theor. Biol.* **76,** 369–386.
22. Cornish-Bowden, A. (1983) *Methods Enzymol.* **91,** 60–75.
23. Latter, G. I., Metz, E., Burbeck, S. & Leavitt, J. (1983) *Electrophoresis* **4,** 122–126.
24. Vanderkerchkhove, J., Leavitt, J., Kakunaga, T. & Weber, K. (1980) *Cell* **22,** 893–899.
25. O'Farrell, P. H. (1975) *J. Biol. Chem.* **254,** 7961–7977.
26. Leavitt, J., Bushar, G., Kakunaga, T., Hamada, H., Hirakawa, T., Goldman, E. & Merril, C. (1982) *Cell* **28,** 259–268.
27. Hirata, Y. & Orth, D. N. (1979) *J. Clin. Endocrinol. Metab.* **48,** 673–679.
28. Gregory, H. (1975) *Nature (London)* **257,** 325–327.
29. Devare, S. G., Reddy, P., Law, J. D., Robbins, K. & Aaronson, S. A. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 731–735.
30. Deuel, T. F., Huang, J. S., Proffitt, R. T., Baenziger, J. U., Chang, D. & Kennedy, B. B. (1981) *J. Biol. Chem.* **256,** 8896–8899.
31. Doolittle, R. F., Feng, D. F. & Johnson, M. S. (1984) *Nature (London)* **307,** 558–560.
32. Leavitt, J. (1983) *Carcinogenesis* **4,** 1229–1237.