Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 1921–1925, April 1985
Biochemistry

Epidermal growth factor precursor is related to the translation product of the Moloney sarcoma virus oncogene *mos*

(epidermal growth factor receptor/*erbB*/tyrosine kinase)

G. S. BALDWIN

Melbourne Tumour Biology Branch, Ludwig Institute for Cancer Research, P.O. Royal Melbourne Hospital, Victoria 3050, Australia

Communicated by Sir Macfarlane Burnet, November 19, 1984

ABSTRACT **Murine epidermal growth factor (EGF) is synthesized as part of a large precursor (pro-EGF), which is thought to span the cell membrane. Comparison of the published pro-EGF sequence with the sequences of the translation products of viral oncogenes reveals that pro-EGF is related to the translation product of *mos*, the oncogene of Moloney murine sarcoma virus. Similarity is greatest between the COOH-terminal region of v-mos (residues 317–360) and part of the cytoplasmic domain of pro-EGF (residues 1127–1174). Statistical comparison of these sequences indicates that the probability of the similarity arising by chance is less than 2×10^{-8}. This similarity extends to the corresponding regions of the translation products of the cellular homologues (c-*mos*) of the v-*mos* gene present in normal murine and human DNA. Similarities are also observed between two other regions of the murine c-mos sequence (residues 48–134 and 196–275) and parts of the extracellular domain of pro-EGF (residues 565–651 and 741–817, respectively). All three *mos* genes are members of the tyrosine kinase family of oncogenes, as is *erbB*, the oncogene of avian erythroblastosis virus. Since the sequences of the *erbB* translation product and the EGF receptor are closely related, the relationship between mos and pro-EGF suggests that pro-EGF and the EGF receptor have evolved from a common ancestor.**

Abundant evidence is now available that a number of the oncogenes of transforming viruses are related to normal growth factors and their receptors. The first such similarity to be reported was between a region of middle tumor (T) antigen, the transforming protein of polyoma virus, and the hormone gastrin (1). [As well as stimulating acid secretion by the stomach, gastrin acts as a growth factor for the gastric mucosa (2).] More recently, close similarity has been noted between platelet-derived growth factor and the translation product of v-*sis*, the oncogene of simian sarcoma virus (3, 4), and between the cytoplasmic domain of the epidermal growth factor (EGF) receptor and the translation product of *erbB*, the oncogene of avian erythroblastosis virus (5). The EGF receptor has intrinsic tyrosine kinase activity (6), and, although the *erbB* translation product does not appear to possess tyrosine kinase activity (7), its sequence is related to the family of oncogenes encoding tyrosine kinases (8). The observation of intrinsic tyrosine kinase activity associated with the receptors for insulin (9), somatomedin C (10), and platelet-derived growth factor (11) suggests, therefore, that similar relationships may exist between these receptors and other members of the family of oncogenes possessing tyrosine kinase activity.

The recent publication of the sequence of the EGF receptor (5) also revealed some unexpected similarities in overall structure with pro-EGF, the large precursor from which EGF is cleaved (12, 13). Both molecules are membrane-bound glycoproteins of approximately 1200 amino acids, with extensive duplications in their extracellular domains (Fig. 1). Since the cytoplasmic domain of the EGF receptor is closely related to the translation product of *erbB* (5), the cytoplasmic domain of pro-EGF was compared with the sequences of the translation products of other members of the tyrosine kinase family of oncogenes. This search revealed the similarity reported in the present paper between pro-

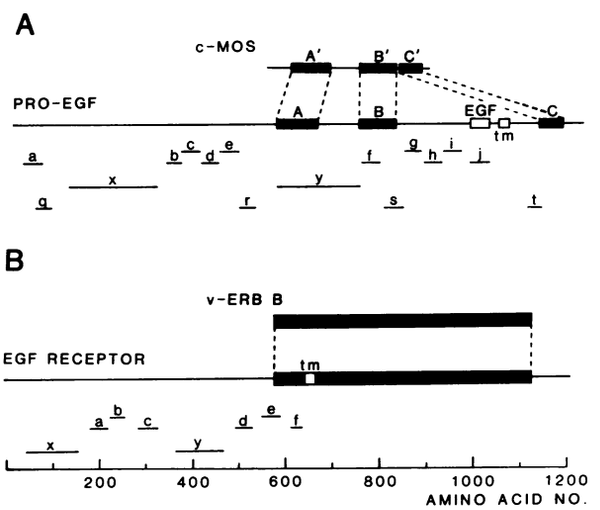

FIG. 1. (*A*) Comparison of the structures of pro-EGF and the c-*mos* translation product. The schematic representations of the amino acid sequences of pro-EGF (12, 13) and c-mos (14) are drawn to the same scale. The three regions (A, B, and C) of pro-EGF similar in sequence to the three regions (A', B', and C') of the c-*mos* translation product are shown by solid bars. The positions of the cysteine-rich (*a–j*) and long (*x* and *y*) repeating units of pro-EGF are given by Doolittle *et al.* (15). The same authors also noted four short repeats (*q–t*) at positions 49–83, 485–520, 795–837, and 1099–1134, respectively, of pro-EGF (15). The positions of mature EGF (cysteine-rich repeat *j*) and the putative transmembrane (tm) segment of pro-EGF are indicated by open boxes. (*B*) Comparison of the structures of the EGF receptor and the v-*erbB* translation product. The schematic representations of the amino acid sequences of the EGF receptor (5) and v-erbB (8) are drawn to the same scale. The single region of the EGF receptor similar in sequence to the v-*erbB* translation product is shown by a solid bar. Cysteine-rich (*a–f*) and long (*x* and *y*) repeating units are present at positions 187–227, 229–264, 291–337, 499–539, 555–595, and 617–637 and at 51–164 and 372–477, respectively, of the EGF receptor. These repeating units appear to be unrelated to the cysteine-rich and long repeating units of pro-EGF. The putative transmembrane (tm) segment of the EGF receptor is indicated by an open box.

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; LDL, low density lipoprotein.

EGF and the translation product of the Moloney sarcoma virus oncogene, *mos* (14, 16, 17).

METHODS

Sequences were first compared with the programs RELATE and SEARCH developed by Dayhoff and her co-workers (18). Optimal alignments of similar sequences were then obtained with the program ALIGN (18) using the mutation data matrix, a matrix bias of +6 and a break penalty of 6 (18). The program calculates an alignment score by comparison of the score for the optimal alignments of the given sequences with the mean score for the optimal alignments of 100 random permutations of the given sequences. The score for each pair of amino acids in the alignments is obtained from the mutation data matrix (18), the elements of which give the probability for the replacement of one amino acid by another in two proteins separated by an evolutionary distance of 250 accepted point mutations per 100 residues. This distance corresponds to approximately 20% residual similarity in sequence.

RESULTS

Sequence Similarities Between mos and Pro-EGF. Three regions of murine pro-EGF (12, 13) are related to the translation products of v-*mos*, the oncogene of Moloney murine sarcoma virus (16), and its cellular progenitor, the murine c-*mos* gene (14, 17). These regions are denoted A–C for pro-EGF and A′–C′ for c-mos in Fig. 1*A*. The greatest similarity was observed between region C of pro-EGF (residues 1127–1174) and region C′ of v-mos (residues 317–360) (Table 1). The alignment score of 5.75 standard deviations above random is well above the value of 3.0 standard deviations, which is taken customarily as the lower limit for related proteins (18) and corresponds to a probability of $<1.9 \times 10^{-8}$ that such a match should occur by chance. Similar scores were reached when the same region of pro-EGF was compared with the corresponding regions of murine and human c-mos (Table 1 and Fig. 2). In particular the sequence Gln-Ser-Cys-Trp (single letter code QSCW; residues 1160–1163 of pro-EGF) was found in none of the remaining 2508 sequences in version 8.1 of the Dayhoff protein data base (18), whereas the Cys-Trp pair (single letter code CW) is a recurring feature of the tyrosine kinase family. In the remainder of this region, *mos* diverged significantly from the other members of the tyrosine kinase family (19), so that it is not surprising that the optimal alignments between pro-EGF and the other members achieved lower scores (Table 1).

Similarity is also apparent between region B of pro-EGF (residues 741–817), and region B′ of c-mos (residues 196–275) (Fig. 2), although in this case the alignment score of 3.06 standard deviations is lower (Table 1). Interestingly, although this region includes one of the EGF-like repeats of pro-EGF [Fig. 1*A*; residues 747–788, repeat *f* in Doolittle's nomenclature (15)], none of the three regions of c-mos similar to pro-EGF overlap the sequence of mature EGF (residues 977–1020, repeat *j*). In addition, region A of pro-EGF (residues 565–651) is similar to region A′ of c-mos (residues 48–134) (Fig. 2). An alignment score of 3.88 standard deviations was calculated for these regions (Table 1). Region A of pro-EGF encompasses the first half of repeat *y* (15) (Fig. 1*A*); comparison of region A′ of c-mos with the first half of the related repeat *x* of pro-EGF (15) yields a slightly lower score of 2.25 standard deviations.

No other similarities were observed between the remainder of the c-mos sequence and pro-EGF. Comparison of the c-mos sequence between regions A′ and B′ (residues 135–195) with the remainder of repeats *x* and *y* of pro-EGF (residues 214–304 and 652–740, respectively) does not provide any evidence of relatedness. Similarly, the NH_2 terminus of the c-mos sequence (residues 1–47) does not appear to be related either to the regions of pro-EGF immediately preceding repeats *x* and *y* or to the NH_2 terminus of pro-EGF. Nevertheless, the above sequence similarities between three regions of pro-EGF and the translation product of the c-*mos* oncogene, together with the fact that the three regions appear in the same order in the two proteins, leave little doubt that the two genes are related.

DISCUSSION

Sequence Similarities Between mos and Pro-EGF. Statistically significant similarities have been observed between three regions of the sequences of murine pro-EGF and the translation product of the murine c-*mos* oncogene. The three regions together comprise 17% of the pro-EGF and 58% of the c-mos sequence. The percentage identities for comparison of regions A–A′, B–B′, and C–C′ (Fig. 1*A*) are 23%, 26%, and 38%, respectively.

Convergent Versus Divergent Evolution. The similarities in sequence between the *mos* translation product and pro-EGF could have arisen either by divergent evolution from a common ancestor or by convergent evolution from unrelated ancestors. These two possibilities can be distinguished by comparison of the nucleotide sequences because convergent

Table 1. Scores for the optimal alignment of the sequences of pro-EGF and members of the tyrosine kinase family

	Pro-EGF region A (residues 565–651)			Pro-EGF region B (residues 741–817)			Pro-EGF region C (residues 1127–1174)		
Aligned sequence	Aligned residues	Score (SD)	*P*	Aligned residues	Score (SD)	*P*	Aligned residues	Score (SD)	*P*
EGF receptor	697–792	1.32	0.093	835–914	0.10	0.460	925–962	2.12	0.017
v-erbB	116–212	1.86	0.031	255–334	0.28	0.390	345–382	1.90	0.029
hu-c-mos	45–131	2.61	$<4.5 \times 10^{-3}$	199–278	2.05	0.020	289–332	5.52	$<1.9 \times 10^{-8}$
mu-c-mos	48–134	3.88	$<1.4 \times 10^{-4}$	196–275	3.06	1.1×10^{-3}	286–329	5.61	$<1.9 \times 10^{-8}$
v-mos	79–165	3.48	$<6.7 \times 10^{-4}$	227–306	1.99	0.023	317–360	5.75	$<1.9 \times 10^{-8}$
v-raf	41–132	1.02	0.154	178–259	1.30	0.097	272–312	1.01	0.156
v-rel	58–148	0.09	0.464	269–341	1.49	0.068	350–397	1.52	0.064
v-src	254–334	0.89	0.187	384–462	1.59	0.056	473–510	1.25	0.106

Regions of the pro-EGF sequence (12, 13) were aligned with regions of the sequences of selected members of the tyrosine kinase family using the program ALIGN, with the mutation data matrix, a matrix bias of +6 and a break penalty of 6 (18). Alignment scores greater than 3.0 standard deviations are usually taken as an indication of relatedness (18). The probabilities (*P*) of attaining an alignment score higher than the indicated score are shown. The sequences of the translation products of the human EGF receptor cDNA (5), the viral *erbB* (8), *mos* (16), *raf* (19), *rel* (20) and *src* (21) oncogenes, and the human and murine homologues of the v-*mos* oncogene [hu-c-*mos* (22) and mu-c-*mos* (14), respectively] were taken from the references indicated.

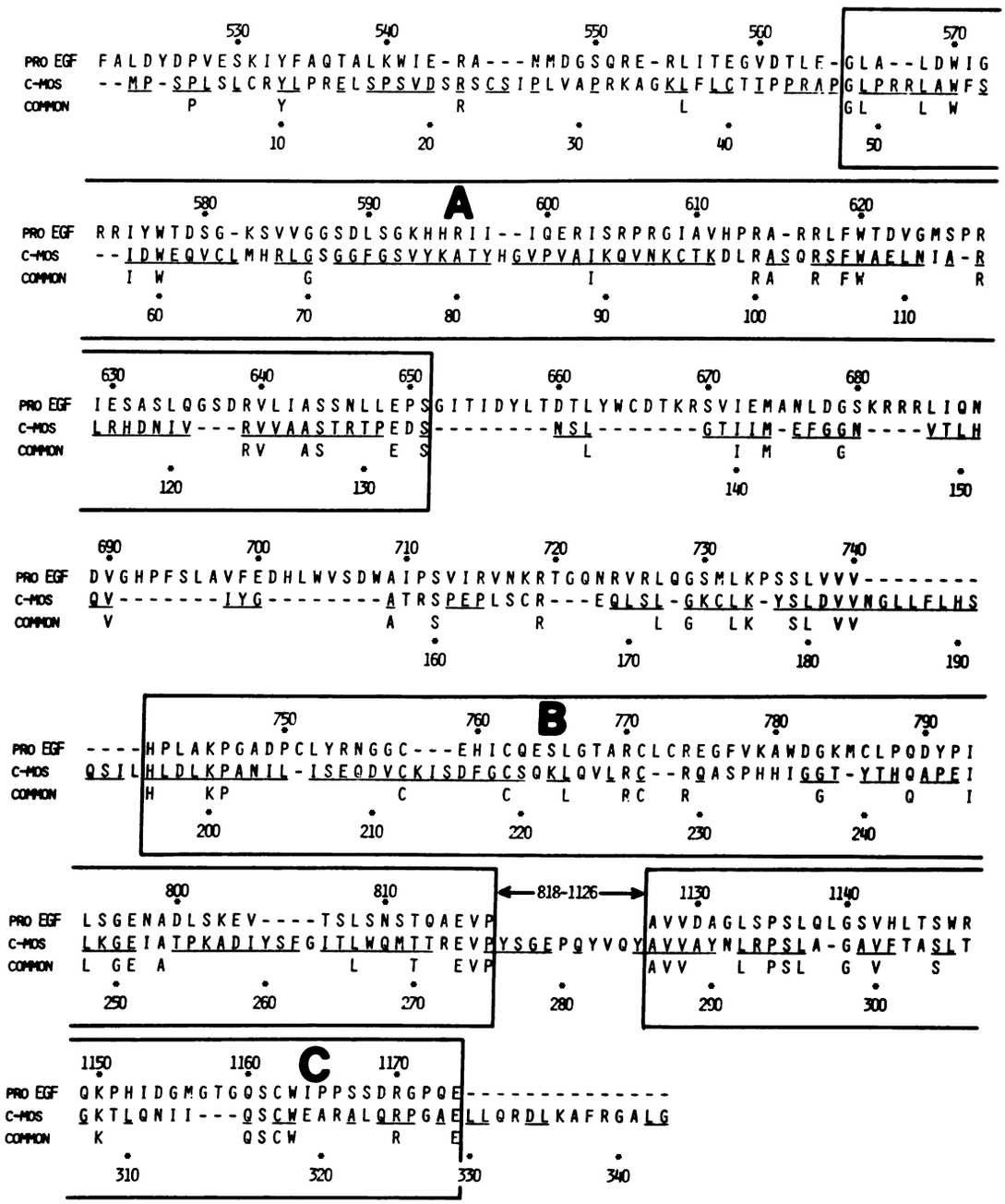

FIG. 2. Comparison of the sequences of pro-EGF and the murine c-*mos* translation product. The sequence of pro-EGF was taken from Scott *et al.* (13). [The sequence of Gray *et al.* (12) contains an extra thymine at position 3707, which alters the reading frame at Leu-1133 and results in premature termination. Correction of this error yields the sequence of Scott *et al.* except for the substitution of Lys-1217 by glutamine (A. Ullrich, personal communication).] The sequence of the murine c-*mos* translation product was taken from Blair *et al.* (17). Those residues that are the same in the human c-*mos* translation product are underlined. The three boxed segments (A, B, and C) of the two sequences were first aligned by using the Dayhoff program ALIGN as described in *Methods*. Gaps are represented by hyphens. The remaining residues 1–47 and 135–183 of the c-*mos* product were then aligned with residues 521–564 (score −0.14 SD) and 652–740 (score 0.02 SD), respectively, of pro-EGF. [Note that the other members of the tyrosine kinase family lack residues corresponding to residues 184–195 of the c-*mos* product (19), which therefore were omitted from the comparison. Deletion of these residues extends alignment B by 12 amino acids.] Residues 818–1126 of pro-EGF have no counterpart in the c-*mos* product and are presumed to have arisen by internal duplication and subsequent divergence. The standard one-letter amino acid notation (23) is used.

evolution would be expected to produce apparently related amino acid sequences from unrelated nucleotide sequences, whereas in the case of divergent evolution, both amino acid and nucleotide sequences would be related. Comparison of the nucleotide sequences encoding region C of murine pro-EGF and region C′ of murine c-mos (residues 1127–1174 and 286–329, respectively; alignment not shown) yields an alignment score of 3.79 standard deviations above random, suggesting that the above regions of the two proteins in fact have arisen by divergent evolution from a common precursor. Comparison of the nucleotide sequences encoding regions A or B of pro-EGF with the corresponding regions A′ or B′ of c-mos yields alignment scores of less than 3 and, therefore, provides no further evidence in support of this conclusion.

Pathway of Divergence. Further comparison of the mos and pro-EGF sequences suggests a possible pathway for the divergence of the two genes (Fig. 1). First, the many internal

duplications in pro-EGF (15), together with the absence of any obvious repeating units in the c-mos sequence, suggest that the internal duplications all arose after the divergence of the genes for pro-EGF and c-mos and, thus, that c-*mos* more closely resembles the common ancestral gene. Since Doolittle and co-workers (15) have shown that a dissimilar pair of EGF-like repeats (*b* and *g*) of pro-EGF diverged approximately 500 million years ago, the divergence of pro-EGF from c-mos can be dated to >500 million years ago, prior to the separation of fishes from the vertebrate line (18). Second, since repeat *y* of pro-EGF is more like c-mos than is repeat *x*, residues 83–520 have presumably diverged after duplication of residues 521–837 of pro-EGF. Third, the close continuity in sequence between regions B′ and C′ of c-mos suggests that residues 818–1127 are a comparatively recent insertion in the pro-EGF sequence. The occurrence of short repeats [Fig. 1*A*; *q*, *r*, *s*, and *t* in Doolittle's nomenclature (15)] in the pro-EGF sequence at both ends of all of the above regions may mean that these repeats had some role in the process of duplication.

Whether c-*mos* is a pseudogene is a relevant question in this context, since the rate of divergence of c-mos from pro-EGF would have been more rapid if c-*mos* were a pseudogene, with the unrestricted freedom to mutate that is characteristic of nonfunctional genes. Unfortunately, the available evidence is inconclusive. The failure to detect c-*mos* transcripts in mouse embryos and fetuses (24), in adult mouse brain, kidney, liver, and spleen (25), or in a variety of mouse and rat cell lines (25) may only reflect a restricted tissue distribution or a restricted timing of transcription. [The only known transcription from the c-*mos* gene occurs in murine plasmacytomas such as XRPC24 after insertion of sequences similar to those of endogenous intracisternal A particle genes within the c-*mos* sequence (26–28).] The absence of introns from the c-*mos* gene is also inconclusive evidence, since the positions of intron/exon junctions in the related c-*abl* (29) and c-*src* (30) genes are not conserved. The recent observation of further sequence similarities between the human and mouse c-*mos* genes ≈350 base pairs upstream from the initiating ATG codon (17) is perhaps the best evidence that the *mos* gene is part of a functional locus and not a pseudogene. If the *mos* gene is transcribed, then the similarities in sequence between mos and pro-EGF may reflect a common function for the two translation products.

Proteins Related to mos or Pro-EGF. Sequence similarities have been reported between several other proteins and either mos or pro-EGF. For example, Russell and co-workers (31) recently reported that regions of the bovine low density lipoprotein (LDL) receptor are related to murine pro-EGF. The greatest similarity (38% identity) was observed between residues 28–133 of the LDL receptor and residues 692–787 of pro-EGF. Although this region overlaps region B of the pro-EGF/c-mos alignment, no significant similarities were observed when the LDL receptor was compared with c-mos. The alignment score for comparison of residues 1–76 and 77–163 of the LDL receptor (31) with residues 135–195 and 196–275 (region B′) of c-mos was 0.30 and 0.48 standard deviations, respectively. Similarities between the sequence of c-mos and both the bovine cyclic AMP-dependent protein kinase [21% identity (32)] and the product of the yeast cell division control gene *CDC28* [22% identity (33)] also have been reported. Again no significant similarities were observed when the appropriate regions of pro-EGF were compared with the latter two proteins. Thus, although sequence comparisons provide no direct evidence of relationship either between the LDL receptor and the c-*mos* translation product or between pro-EGF and bovine cyclic AMP-dependent kinase or the yeast *CDC28* gene product, this may only be a reflection of the fact that pro-EGF and c-mos are central and the other three proteins are outlying branches on the evolutionary tree of the tyrosine kinase family.

The relationship of pro-EGF and the EGF receptor to the translation products of different members of the tyrosine kinase family of oncogenes (mos and erbB, respectively) implies that pro-EGF is related to the EGF receptor. There are several superficial similarities between the structures of the two molecules (Fig. 1). Both are large transmembrane glycoproteins with extracellular domains that have evolved by extensive duplication. However, direct comparison of the sequences of pro-EGF and the EGF receptor does not provide conclusive evidence for a common origin for the two molecules (Table 1). For example, the score for the optimal alignment of residues 1127–1174 of pro-EGF with residues 925–962 of the EGF receptor is only 2.12 standard deviations above random, a value lower than the 3.0 standard deviations that is usually taken as the limit for related proteins. Nevertheless, the indirect (via mos and erbB) relationship between pro-EGF and the EGF receptor strongly suggests that the two proteins share a common ancestor and may even share kinase activity.

Are mos and Pro-EGF Kinases? Although the sequence of the *mos* gene is related to the family of oncogenes encoding tyrosine kinases, the available evidence suggests that the *mos* translation product may possess the ability to phosphorylate serine and threonine rather than tyrosine residues. Thus, cells infected with Moloney sarcoma virus do not contain elevated levels of phosphotyrosine, and immunoprecipitates of the *mos* translation product do not contain tyrosine kinase activity (34). In contrast immunoprecipitates of a gag-mos fusion protein from a Moloney sarcoma virus mutant temperature-sensitive for transformation contain a temperature-sensitive serine and threonine kinase activity (35). Moreover, the *mos* genes encode both a lysine residue (position 90 in c-mos, Fig. 2) in a position equivalent to the $pp60^{src}$ lysine residue that reacts with an ATP analogue, and an adjacent cluster of glycines (G-X-G-X-X-G, G = glycine and X = other; residues 70–75 in c-mos, Fig. 2) thought to be involved in ATP binding (36). These residues are also present in the cAMP-dependent kinase, which is related to mos (32) and which phosphorylates predominantly serine and occasionally threonine residues (37).

Whether pro-EGF is also a kinase remains an open question. Although the pro-EGF sequence in region A (Fig. 2) does not have a lysine residue corresponding to Lys-90 of c-mos, there are several arginine residues nearby that could fulfill a requirement for a basic residue for ATP binding. However, the central glycine of the ATP-binding cluster G-X-G-X-X-G present in the tyrosine kinase family (36) is missing from pro-EGF residues 584–592. Perhaps the strongest argument against pro-EGF retaining kinase activity is that the above residues are found in the extracellular domain of the molecule.

What Is the Function of Pro-EGF? If pro-EGF is not a kinase, what function does it perform? It seems unlikely that such a large molecule (1200 amino acids) would be synthesized solely to generate a 53-amino acid growth factor extracellularly. (The abundant production of EGF in the submaxillary gland may only reflect the high concentration of proteases in that tissue.) The common origin for pro-EGF and the EGF receptor suggested by the pro-EGF/mos sequence similarity may reflect the following common function for the two molecules. If the ability of EGF to bind to the EGF receptor is also a property of pro-EGF, then perhaps binding of cell-surface pro-EGF to the EGF receptor expressed on the surface of a different cell type may contribute to normal cell–cell interactions. EGF appearing transiently because of proteolytic degradation of pro-EGF could compete with cell-surface pro-EGF for binding to the EGF receptor and, thus, interfere with such interactions. Presumably if this arrangement were reciprocal, then EGF-binding fragments corre-

sponding to the extracellular domain of the EGF receptor (38) could bind to cell-surface pro-EGF and interfere with cell–cell interactions in a similar manner.

I thank Tony Kyne for assistance with computing and numerous colleagues for their critical comments on this manuscript.

1. Baldwin, G. S. (1982) *FEBS Lett.* **137,** 1–5.
2. Johnson, L. R. (1976) *Gastroenterology* **70,** 278–288.
3. Doolittle, R. F., Hunkapiller, M. W., Hood, L. E., Devare, S. G., Robbins, K. C., Aaronson, S. A. & Antoniades, H. N. (1983) *Science* **221,** 275–277.
4. Waterfield, M. D., Scrace, G. T., Whittle, N., Stroobant, P., Johnsson, A., Wasteson, A., Westermark, B., Heldin, C.-H., Huang, J. S. & Deuel, T. F. (1983) *Nature (London)* **304,** 35–39.
5. Ullrich, A., Coussens, L., Hayflick, J. S., Dull, T. J., Gray, A., Tam, A. W., Lee, J., Yarden, Y., Libermann, T. A., Schlessinger, J., Downward, J., Mayes, E. L. V., Whittle, N., Waterfield, M. D. & Seeburg, P. H. (1984) *Nature (London)* **309,** 418–425.
6. Cohen, S., Carpenter, G. & King, L. (1980) *J. Biol. Chem.* **255,** 4834–4842.
7. Hayman, M. J., Ramsay, G. M., Savin, K., Kitchener, G., Graf, T. & Beug, H. (1983) *Cell* **32,** 579–588.
8. Yamamoto, T., Nishida, T., Miyajima, N., Kawai, S., Ooi, T. & Toyoshima, K. (1983) *Cell* **35,** 71–78.
9. Petruzelli, L., Herrera, R. & Rosen, O. M. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 3327–3331.
10. Jacobs, S., Kull, F. C., Earp, H. S., Svoboda, M. E., Van Wyk, J. J. & Cuatrecasas, P. (1983) *J. Biol. Chem.* **258,** 9581–9584.
11. Ek, B., Westermark, B., Wasteson, A. & Heldin, C. (1982) *Nature (London)* **295,** 419–420.
12. Gray, A., Dull, T. J. & Ullrich, A. (1983) *Nature (London)* **303,** 722–725.
13. Scott, J., Urdea, M., Quiroga, M., Sanchez-Pescador, R., Fong, N., Selby, M., Rutter, W. J. & Bell, G. I. (1983) *Science* **221,** 236–240.
14. Van Beveren, C., Van Straaten, F., Galleshaw, J. A. & Verma, I. M. (1981) *Cell* **27,** 97–108.
15. Doolittle, R. F., Feng, D. F. & Johnson, M. S. (1984) *Nature (London)* **307,** 558–560.
16. Van Beveren, C., Galleshaw, J. A., Jonas, V., Berns, A. J. M., Doolittle, R. F., Donoghue, D. J. & Verma, I. M. (1981) *Nature (London)* **289,** 258–262.
17. Blair, D. G., Wood, T. G., Woodworth, A. M., McGeady, M. L., Oskarsson, M. K., Propst, F., Tainsky, M. A., Cooper, C. S., Watson, R., Baroudy, B. M. & Vande Woude, G. F. (1984) in *Oncogenes and Viral Genes*, eds. Vande Woude, G. F., Levine, A. J., Topp, W. C. & Watson, J. D. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 281–289.
18. Dayhoff, M. O. (1979) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (National Biomedical Research Foundation, Washington, DC), pp. 1–8.
19. Mark, G. E. & Rapp, U. R. (1984) *Science* **224,** 285–289.
20. Stephens, R. M., Rice, N. R., Hiebsch, R. R., Bose, H. R. & Gilden, R. V. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 6229–6233.
21. Czernilofsky, A. P., Levinson, A. D., Varmus, H. E., Bishop, J. M., Tischer, E. & Goodman, H. (1983) *Nature (London)* **301,** 736–738.
22. Watson, R., Oskarsson, M. & Vande Woude, G. F. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4078–4082.
23. IUPAC-IUB Commission on Biochemical Nomenclature (1968) *Eur. J. Biochem.* **5,** 151–153.
24. Muller, R., Slamon, D. J., Tremblay, J. M., Cline, M. J. & Verma, I. M. (1982) *Nature (London)* **299,** 640–644.
25. Gattoni, S., Kirschmeier, P., Weinstein, I. B., Escobedo, J. & Dina, D. (1982) *Mol. Cell. Biol.* **2,** 42–51.
26. Kuff, E. L., Feenstra, A., Lueders, K., Rechavi, G., Givol, D. & Canaani, E. (1983) *Nature (London)* **302,** 547–548.
27. Rechavi, G., Givol, D. & Canaani, E. (1982) *Nature (London)* **300,** 607–611.
28. Canaani, E., Dreazen, O., Klar, A., Rechavi, G., Ram, D., Cohen, J. B. & Givol, D. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 7118–7122.
29. Wang, J. Y. J., Ledley, F., Goff, S., Lee, R., Groner, Y. & Baltimore, D. (1984) *Cell* **36,** 349–356.
30. Takeya, T. & Hanafusa, H. (1983) *Cell* **32,** 881–890.
31. Russell, D. W., Schneider, W. J., Yamamoto, T., Luskey, K. L., Brown, M. S. & Goldstein, J. L. (1984) *Cell* **37,** 577–585.
32. Barker, W. C. & Dayhoff, M. O. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 2836–2839.
33. Lorincz, A. T. & Reed, S. I. (1984) *Nature (London)* **307,** 183–185.
34. Papkoff, J., Verma, I. & Hunter, T. (1982) *Cell* **29,** 417–426.
35. Kloetzer, W. S., Maxwell, S. A. & Arlinghaus, R. B. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 412–416.
36. Kamps, M. P., Taylor, S. S. & Sefton, B. M. (1984) *Nature (London)* **310,** 589–592.
37. Cohen, P., Rylatt, D. B. & Nimmo, G. A. (1977) *FEBS Lett.* **76,** 182–186.
38. Weber, W., Gill, G. N. & Spiess, J. (1984) *Science* **224,** 294–297.