Sequence similarity between epidermal growth factor precursor and atrial natriuretic factor precursor

Hidenori Hayashida* and Takashi Miyata*+○

*Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, and +National Institute of Genetics, Mishima, Japan

Received 18 March 1985

Computer-assisted analysis for homology of the EGF precursor revealed the presence of two large duplication units, each comprising 5 non-EGF-like homologous segments each of about 40 residues length and 3 or 4 EGF-like segments. The amino acid sequences of the non-EGF-like repeats were subjected to search for homology with 2600 known protein sequences compiled in our database. An unexpected but statistically significant homology has been found, when compared with the atrial natriuretic factor precursor. The functional and evolutionary implications of the homology observed between the two different precursors are also discussed.

Repeated structure *EGF precursor* *LDL receptor* *ANF precursor* *Homology* *Evolution*

1. INTRODUCTION

The complete sequence of mouse epidermal growth factor (EGF) precursor has been determined by two groups [1,2] and shown to contain two clusters, each comprising tandemly repeated EGF-like segments [1–3]. Both clusters are flanked by regions, designated X and Y, of unknown functions which exhibit sequence homology to each other [3]. An unexpected homology has been found, when compared with the low density lipoprotein (LDL) receptor [4,5], suggesting that they share a common evolutionary origin. We report here that each of the X and Y regions of the EGF precursor also consists of tandemly repeated segments, each being about 40 residues long, which shares a significant homology with atrial natriuretic factor (ANF) precursor. Possible evolutionary relationships among the EGF and ANF precursors and the LDL receptor are also discussed.

* To whom correspondence should be addressed

2. MATERIALS AND METHODS

Sequence data were taken from [1,2] for mouse EGF precursor, [4] for bovine and [5] for human LDL receptors, and [6–8] for human, mouse and rat ANF precursors, respectively. Methods for detecting homology and that for alignment have been described previously [9].

3. RESULTS AND DISCUSSION

The EGF precursor has been shown to consist of 3 different kinds of homologous segments; 10 EGF-like segments a–j (unit j overlaps with EGF), two longer non-EGF-like segments X and Y and another kind of non-EGF-like segments q–t (see fig.1 of [3]). Computer-assisted analysis revealed the presence of large duplication units (designated I and II) of apparent homology, including segments q, X and b–e in unit I and r, Y and f–g as well as s in unit II (although the segment s was classified as a member of non-EGF-like segments [3], it exhibits an obvious homology with the EGF-like segment d). It was also demonstrated that each of the

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

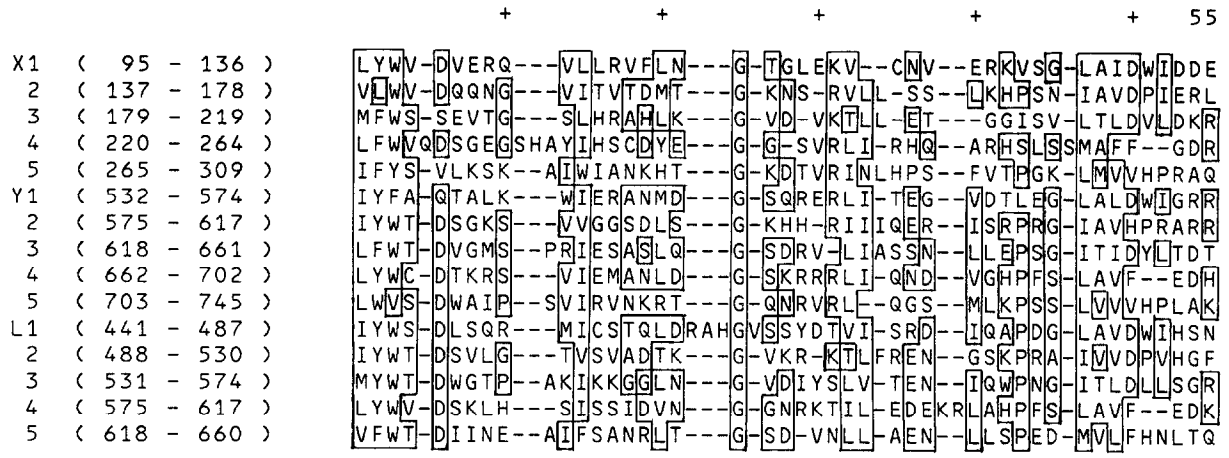

Fig.1. Alignment of non-EGF-like repeats X1–5 and Y1–5 of EGF-precursor and the corresponding repeats (L1–5) of LDL receptor. Boxed residues indicate identities or favoured amino acid substitutions among more than half of the sequences. Gaps (–) were inserted to increase sequence similarity. Amino acid positions of each repeat in EGF precursor [1,2] and that in LDL receptor [5] are shown in parentheses.

X and Y regions comprises 5 homologous segments of about 40 residues long each (figs 1 and 2).

The LDL receptor was shown to be homologous to the EGF precursor in their extracellular regions [4,5]. Yamamoto et al. [5] also noted the presence of 5 repeats in the LDL receptor for a region of residues 409–610. This homologous region corresponds to the Y segment in the EGF precursor. Fig.1 shows the alignment of the homologous segments in region X and Y, including the corresponding segments in the LDL receptor. The modified version of the structural organization of the EGF precursor, together with that of the LDL receptor was shown in fig.2. As shown in fig.2, the EGF precursor comprises two large duplication units I and II flanked by 3 EGF-like segments h–j at the end of the unit II (an EGF-like semgent 'a' is not shown in fig.2, because it appears to share no apparent homology with other EGF-like segments). Each large duplication unit consists of 5 non-EGF-like segments and 3 or 4 EGF-like segments, together with a different kind of non-EGF-like segment.

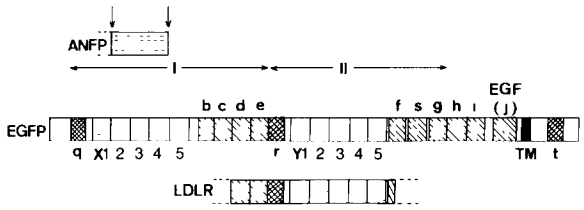

Fig.2. Structural organization of EGF precursor. EGFP, EGF precursor; LDLR, LDL receptor; ANFP, ANF precursor; I, II, large duplication units in EGF precursor; b–j, s, EGF-like homologous segments; X1–5, Y1–5, non-EGF-like homologous segments; q, r, t, another kind of non-EGF-like homologous segment. Homologous regions of ANF precursor and LDL receptor are also shown. Arrows indicate positions where introns exist in ANF precursor genes.

To suggest a possible function involved in the non-EGF-like segments X and Y, the sequences were subjected to search for homology with about 2600 known sequences compiled in our database. An unexpected homology has been detected between the sequence of X and that encoded by the second exon of atrial natriuretic factor (ANF) precursor, known to play an important role in regulating blood pressure as well as extracellular fluid volume [10–14]. The alignment is shown in fig.3. The sequence similarity is statistically significant with the probabilities of occurrence by chance of 1.1×10^{-5}, 1.9×10^{-4} and 4.8×10^{-5} for comparisons of the EGF precursor with human, mouse and rat ANF precursors, respectively. Similar but weak homology was also observed, when compared with the Y segment. Interestingly the intron positions of the ANF precursor approximately cor-

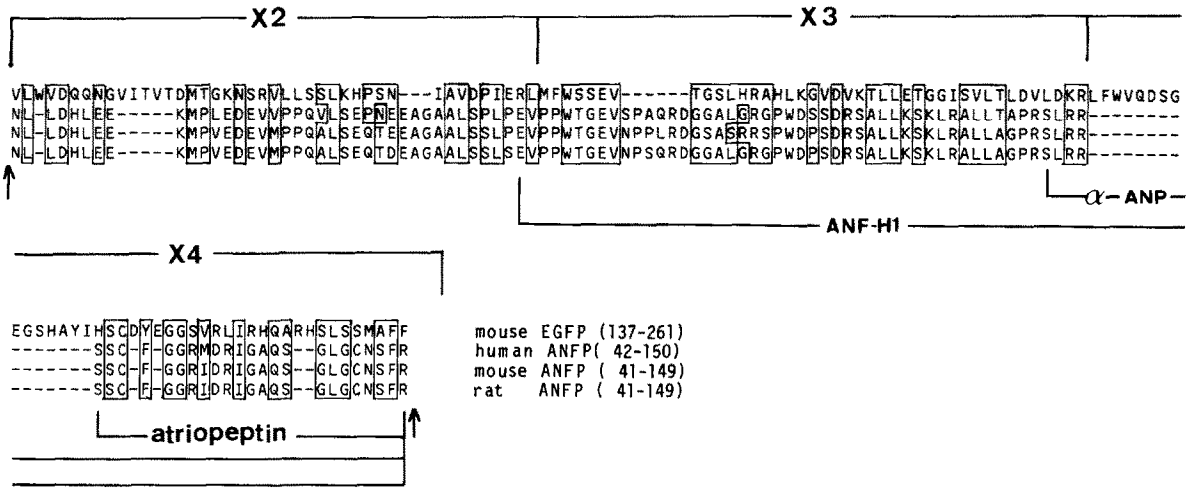

Fig.3. Alignment of the amino acid sequence of EGF precursor with those of human, mouse and rat ANF precursors. Boxed residues indicate identities or favoured amino acid substitutions between EGF-precursor and ANF-precursors [9]. Gaps (−) were inserted to increase sequence similarity. Amino acid positions of the aligned regions were shown in parentheses. X2–4, non-EGF-like repeats. Cleavage products of ANF precursor are also shown below the rat sequence. Vertical arrows indicate intron positions in ANF precursor.

respond to the boundaries of the non-EGF-like repeats. Furthermore, cleavage points for generating atriopeptin, α-ANP and ANF-H1 [14–16] from the ANF precursor are located near the boundaries of the repeats (fig.3). These results suggest some functional and evolutionary implications of the EGF precursor. First, the non-EGF-like repeats may involve activities related to those of peptides that are derived from the ANF precursor. Second, introns may exist near the boundaries of repeats of the X and Y regions. Third, the two precursor genes share a common evolutionary origin.

The EGF precursor shows a strong homology with the LDL receptor in the region that involves the whole Y segment and the EGF-like segments of unit I in part and a weak homology in the X segment. This homology pattern suggests that the homologous portion of the LDL receptor was derived from an ancestral EGF precursor which already had two large duplicated units. On the other hand, the ANF precursor shares significant homology with the X segment, but only weakly with the Y segment. This result, together with evidence that no obvious repeated structure was detected, suggests that the second exon of ANF precursor originated from the non-EGF-like repeats, followed by accumulation of many mutational changes which resulted in the disappearance of repeated structure.

ACKNOWLEDGEMENTS

We thank Miss K. Mitsuyasu for technical assistance. This work was supported in part by grants from the Ministry of Education, Science and Culture of Japan.

REFERENCES

[1] Gray, A., Dull, T.J. and Ullrich, A. (1983) Nature 303, 722–725.
[2] Scott, J., Urdea, M., Quiroga, M., Sanchez-Pescador, R., Fong, N., Selby, M., Rutter, W.J. and Bell, G.I. (1983) Science 221, 236–240.
[3] Doolittle, R.F., Feng, D.F. and Johnson, M.S. (1984) Nature 307, 558–560.
[4] Russell, D.W., Schneider, W.J., Yamamoto, T., Luskey, K.L., Brown, M.S. and Goldstein, J.L. (1984) Cell 37, 577–585.
[5] Yamamoto, T., Davis, C.G., Brown, M.S., Schneider, W.J., Casey, M.L., Goldstein, J.L. and Russell, D.W. (1984) Cell 39, 27–38.

[6] Nemer, M., Chamberland, M., Sirois, D., Argentin, S., Drouin, J., Dixon, R.A.F., Zivin, R.A. and Condra, J.H. (1984) Nature 312, 654–656.
[7] Seidman, C.E., Bloch, K.D., Klein, K.A., Smith, J.A. and Seidman, J.G. (1984) Science 226, 1206–1209.
[8] Yamanaka, M., Greenberg, B., Johnson, L., Seilhamer, J., Brewer, M., Friedemann, T., Miller, J., Atlas, S., Laragh, J., Lewicki, J. and Fiddes, J. (1984) Nature 309, 719–722.
[9] Toh, H., Hayashida, H. and Miyata, T. (1983) Nature 305, 827–829.
[10] De Bold, A.J. (1982) Proc. Soc. exp. Biol. Med. 170, 133–138.
[11] Currie, M.G., Geller, D.M., Cole, B.R., Boylan, J.G., Yusheng, W., Holmberg, S.W. and Needleman, P. (1983) Science 221, 71–73.
[12] Flynn, T.G., De Bold, M.L. and De Bold, A.J. (1983) Biochem. Biophys. Res. Commun. 117, 859–865.
[13] Atlas, S.A., Kleinert, H.D., Camargo, M.J., Januszewicz, A., Sealey, J.E., Laragh, J.H., Schilling, J.W., Lewicki, J.A., Johnson, L.K. and Maack, T. (1984) Nature 309, 717–719.
[14] Currie, M.G., Geller, D.M., Cole, B.R., Siegel, N.R., Fok, K.F., Adams, S.P., Eubanks, S.R., Galluppi, G.R. and Needleman, P. (1984) Science 223, 67–69.
[15] Kangawa, K. and Matsuo, H. (1984) Biochem. Biophys. Res. Commun. 118, 131–139.
[16] Thibault, G., Garcia, R., Cantin, M., Genest, J., Lazure, C., Seidah, N.G. and Chretien, M. (1984) FEBS Lett. 167, 352–356.