Sequence similarity between EGF receptor and α_1-acid glycoprotein

THE human epidermal growth factor (EGF) receptor precursor consists of three domains, an extracellular EGF binding domain, a single transmembrane domain and a cytoplasmic domain. Ullrich *et al.*[1] recently determined the complete amino acid sequence of the molecule and showed that the latter two domains exhibit close similarity in sequence to the entire v-*erb-B* oncogene product, a member of the *src* family. By aligning cysteine residues, they also found repeated structures of ~170 residues within the extracellular domain[1]. A similarity matrix comparison[2], however, revealed the presence of further extensive homology between regions of ~300 residues located at positions 1-274 and 294-615, which cover almost the entire extracellular domain (Fig. 1). Of all the positions compared in the aligned sequences, 28% are occupied by identical residues (gaps were considered as substitutions between different amino acids regardless of their length). Furthermore, a statistical test[2] showed that the similarity between the aligned sequences is highly significant, with a probability of occurrence by chance of $<10^{-9}$. Thus, the extracellular EGF binding domain consists of tandemly repeated domains which may have evolved by internal duplication. Alignment of their sequences also revealed that each repeat is composed of an N-terminal conserved region and a C-terminal cysteine-rich region.

We subjected the amino acid sequence of the EGF receptor to a computer-assisted search for homology with ~2,500 known sequences compiled in our database and found apparent sequence similarity between the C-terminal half of the EGF binding domain and α_1-acid glycoprotein (α_1-AGP), known to be an acute-phase protein[3] (Fig. 1). A statistical test shows the similarity to be highly significant, with $P = 3 \times 10^{-7}$.

Interestingly, the α_1-AGP molecule has been shown to exhibit weak sequence homology to immunoglobulin[3] as well as to part of the variable domain of the mouse T-cell receptor ($P < 3 \times 10^{-4}$). Thus, it seems likely that the EGF binding domain is indirectly related to the immunoglobulin superfamily. Lack of significant similarity between the N-terminal half of the EGF binding domain and α_1-AGP may be a result of rapid divergence in the N-terminal half, relative to the C-terminal half, presumably during the earlier evolutionary stage after the separation of the two domains by internal duplication. In addition, the transmembrane segment of this receptor bears a close resemblance to that of the HLA class II β-chain ($P < 5 \times 10^{-4}$).

These results, together with evidence that the cytoplasmic domain contains an *src*-related sequence[1], suggest that the extracellular and cytoplasmic domains may have been derived from evolutionarily independent origins and thus that this large receptor gene may have evolved by gene shuffling and partial duplication. We cannot exclude the possibility, however, that an ancestral α_1-AGP gene was derived from a part of the EGF binding domain.

HIROYUKI TOH*
HIDENORI HAYASHIDA*
REIKO KIKUNO*
TERUO YASUNAGA†
TAKASHI MIYATA*

* *Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan*
† *Institute of Physical and Chemical Research, Wako-si, Saitama 351, Japan*

1. Ullrich, A. *et al. Nature* **309**, 418-425 (1984).
2. Toh, H., Hayashida, H. & Miyata, T. *Nature* **305**, 827-829 (1983).
3. Schmid, K. *et al. Biochemistry* **12**, 2711-2724 (1973).
4. Larhammar, D., Hyldig-Nielsen, J. J., Andersson, G., Rask, L. & Peterson, P. A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7313-7317 (1983).

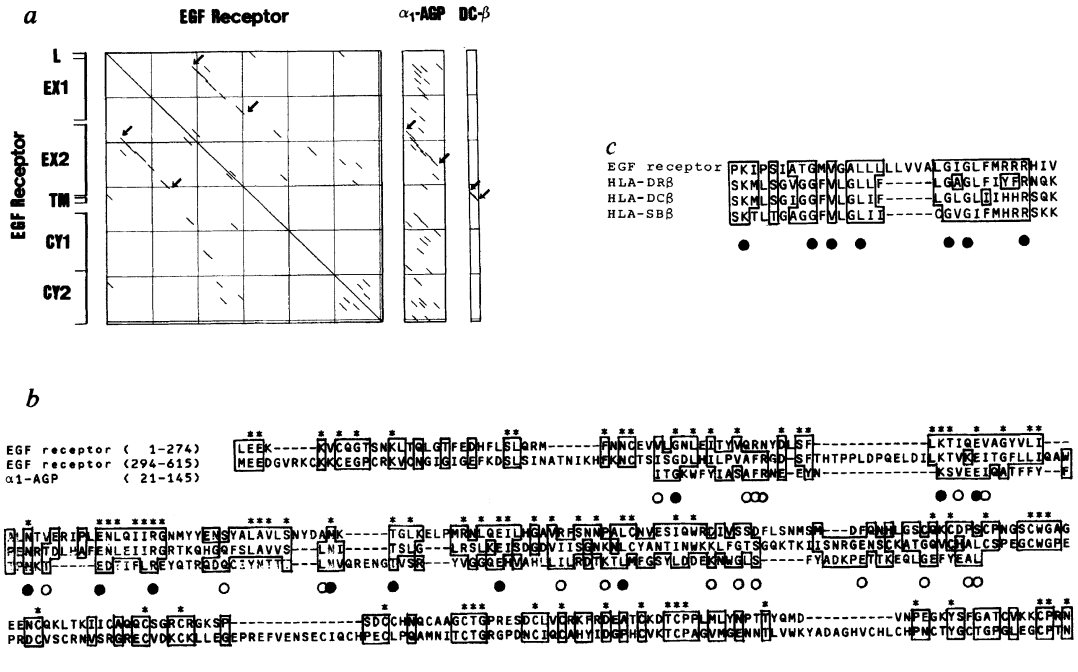

Fig. 1 *a*, Similarity matrix comparisons of amino acid sequence of EGF receptor (ordinate) with the sequences of itself, α_1-AGP and the transmembrane segment of HLA-DC β-chain (abscissa). A computer program was used to generate diagonal lines indicating segments of 25 residues which show sequence similarity (see ref. 2). L, signal peptide (positions −24 to −1); EX1, EX2, duplicated regions (1-274 and 294-615) of the extracellular EGF binding domain, respectively; TM, the transmembrane segment (622-644); CY1, CY2, *src*-homologous and -non-homologous regions (690-936 and 967-1,186) of the cytoplasmic domain, respectively. Arrows indicate regions where extensive similarities were detected. *b*, Alignment of amino-acid sequences of N-terminal (EX1) and C-terminal (EX2) halves of the EGF binding domain and α_1-AGP. Boxed residues indicate identities or favoured amino-acid substitutions (see ref. 2). Positions where residues are identical between EX1 and EX2 and between EX2 and α_1-AGP are marked by * and ○, respectively. Invariant residues among the three are indicated by ●. Gaps (—) were introduced to maximize similarity. *c*, Alignment of transmembrane segments of EGF receptor and HLA class II β-chains. Regions where sequences were aligned correspond to positions 617-650 and 197-225 of the EGF receptor and HLA-DCβ, respectively. Boxed residues indicate identities (●) or favoured amino-acid substitutions between EGF receptor and HLA chains.