21. *Proto-oncogene ret Has Cadherin-like Repeats in the Extracellular Region*

By Kei-ichi KUMA, Takashi MIYATA, and Masatoshi TAKEICHI
Department of Biophysics, Faculty of Science, Kyoto University, Kyoto 606

(Communicated by Motoo KIMURA, M. J. A., Sept. 14, 1992)

Abstract: Amino acid sequences of 90 different cadherin repeats were aligned for six highly conserved segments, and the frequency of occurrence of amino acids at each amino acid position was calculated. Using the frequency of occurrence, further members of the cadherin superfamily were searched in protein database by a method developed previously. It was found that *ret* proto-oncogene product, a receptor type protein tyrosine kinase, contains cadherin repeats in the extracellular region.

Key words: *ret* oncogene; cadherin superfamily; homology; evolution.

Cadherins, a family of Ca^{2+} -dependent cell-cell adhesion molecules, are transmembrane glycoproteins that consist of an extracellular, a transmembrane and a cytoplasmic domains. The extracellular domain contains four repeats with characteristic amino acid sequences (cadherin repeat), each consisting of about 110 amino acids, in all vertebrate cadherins identified to date (Takeichi, 1991; Magee and Buxton, 1991). There are marked sequence similarities between different cadherin repeats, suggesting that they derived from a single precursor sequence by internal duplications (Hatta *et al.*, 1988). Recently the *fat* locus of *Drosophila* has been cloned and shown to be a novel member of the cadherin supergene family; the *fat* protein contains 34 tandem cadherin-like repeats in the extracellular domains (Mahoney *et al.*, 1991). The presence of cadherin-like repeats over a wide range of species suggests the remote origin of the cadherin superfamily and thus the family is expected to comprise a large number of family members with diverse functions.

To search for further members of the cadherin superfamily in protein databases, a computer-assisted search for similarity was carried out by a method developed recently (Kuma *et al.*, 1991). On the basis of the alignment of the amino acid sequences of 90 different cadherin repeats from 15 known members of the family, we selected six different segments that are strongly conserved among the sequences compared (see Fig. 1). Each amino acid position of these segments is occupied mostly by identical or conservative amino acids, but non-conservative amino acids also occur with low frequencies. On the basis of the frequency of occurrence of amino acids at each position, a scoring system was determined for the initial search of the cadherin repeats. Applying this method to NBRF database (release 30.0), we found a cadherin-related sequence in the extracellular domain of the *ret* proto-oncogene protein, a receptor type protein tyrosine kinase whose ligand has not been determined (Takahashi *et al.*, 1988, 1989).

A dot matrix comparison (Toh *et al.*, 1983) between the *ret* protein and members of the cadherin superfamily revealed a marked sequence similarity (data not shown). The amino acid sequence of the *ret* protein was aligned with those of 12 vertebrate cadherins as well as those of four homologous repeats of *Drosophila fat* protein (Fig. 1). The *ret* protein shows 20%–31% identity with the cadherins compared. These similarities are statistically significant; the probabilities $\text{Pr.}'s$ that the observed similarities occur by chance (Toh *et al.*, 1983) are in a range of 10^{-3} – 10^{-11} . The similarity with chicken T-cadherin is particularly

Fig. 1. Comparison of *ret* and members of the cadherin family from vertebrates and invertebrate. Position numbers of the first and last amino acids of the aligned regions are shown in parentheses. Amino acids that are identical with that of *ret* are boxed. %, percent identity between *ret* and each of the family members; Pr., the probability that the observed similarity is realized by chance; 1.0E-11, for example, represents 1.0×10^{-11} ; • and ○, positions that are occupied by identical and chemically similar amino acids among all the sequences compared, respectively; -, gap. Six conserved segments used in the initial search are underlined beneath the alignment (see text). Abbreviations and sequence data sources: *ret* (Takahashi *et al.*, 1988, 1989); cT, chicken T-cadherin (Ranscht and Dours-Zimmermann, 1991); cN, chicken N-cadherin (Hatta *et al.*, 1988); cLCAM, chicken L-CAM (Gallin *et al.*, 1987); cR, chicken R-cadherin (Inuzuka *et al.*, 1991); cB, chicken B-cadherin (Napolitano *et al.*, 1991); mP, mouse P-cadherin (Nose *et al.*, 1987); mE, mouse E-cadherin (Nagafuchi *et al.*, 1987); mM, mouse M-cadherin (Donalies *et al.*, 1991); bDC, bovine desmocollin (Mechanic *et al.*, 1991); hDG3, human desmosomal glycoprotein III (Parker *et al.*, 1991); hPVA, human pemphigus vulgaris antigen (Amagai *et al.*, 1991); bDG, bovine desmoglein (Goodwin *et al.*, 1990); dF, *Drosophila fat* (Mahoney *et al.*, 1991).

remarkable (31% identity and $\text{Pr.}=10^{-11}$). In the aligned region, 14 amino acid positions are identical among the cadherins and the *fat* protein compared, of which 11 are preserved when the *ret* protein is included for comparison. Furthermore, two motifs DXND and DXD, which are thought to be putative Ca^{2+} binding sites (Ringwald *et al.*, 1987; Ozawa *et al.*, 1990), are conserved in *ret*. From these results, we conclude that the *ret* protein is a novel member of the cadherin superfamily.

Acknowledgements. This work was supported in part by grants from the Ministry of Education, Science and Culture of Japan.

References

- Amagai, M., Klaus-Kovtun, V., and Stanley, J. R. (1991): Cell, **67**, 869–877.
Donalies, M. *et al.* (1991): Proc. Natl. Acad. Sci. USA, **88**, 8024–8028.
Gallin, W. J. *et al.* (1987): *ibid.*, **84**, 2808–2812.
Goodwin, L. *et al.* (1990): Biochem. Biophys. Res. Commun., **173**, 1224–1230.
Hatta, K. *et al.* (1988): J. Cell Biol., **106**, 873–881.
Inuzuka, H., Miyatani, S., and Takeichi, M. (1991): Neuron, **7**, 69–79.
Kuma, K., Iwabe, N., and Miyata, T. (1991): Proc. Japan Acad., **67B**, 188–192.
Magee, A. I., and Buxton, R. S. (1991): Current Opinion Cell Biol., **3**, 854–861.
Mahoney, P. A. *et al.* (1991): Cell, **67**, 853–868.
Mechanic, S. *et al.* (1991): Proc. Natl. Acad. Sci. USA, **88**, 4476–4480.
Nagafuchi, A. *et al.* (1987): Nature, **329**, 341–343.
Napolitano, E. W. *et al.* (1991): J. Cell Biol., **113**, 893–905.
Nose, A., Nagafuchi, A., and Takeichi, M. (1987): EMBO J., **6**, 3655–3661.
Ozawa, M., Engel, J., and Kemler, R. (1990): Cell, **63**, 1033–1038.
Parker, A. E. *et al.* (1991): J. Biol. Chem., **266**, 10438–10445.
Ranscht, B., and Dours-Zimmermann, M. T. (1991): Neuron, **7**, 391–402.
Ringwald, M. *et al.* (1987): EMBO J., **6**, 3647–3653.
Takahashi, M. (1988): Oncogene, **3**, 571–578.
Takahashi, M., Buma, Y., and Hiai, H. (1989): *ibid.*, **4**, 805–806.
Takeichi, M. (1991): Science, **251**, 1451–1455.
Toh, H., Hayashida, H., and Miyata, T. (1983): Nature, **305**, 827–829.