3000 Ci/m*M*) (UTP) was then added, and the nuclear suspension was incubated at 30°C for 30 minutes, after which time 15 μl of DNase I (5 μg/ml) in 10 m*M* $CaCl_2$ (5 μg/ml) was added. After 5 minutes at 30°C, the reaction was made 1× SET (1 percent sodium dodecyl sulfate (SDS), 5 m*M* EDTA, 10 m*M* tris-HCl, *p*H 7.4), and proteinase K was added to a concentration of 200 μg/ml. After incubation at 37°C for 45 minutes, the solution was extracted with an equal volume of a mixture of phenol and chloroform, and the interphase was again extracted with 100 μl of 1× SET. Ammonium acetate (10*M*) was added to the combined aqueous phases (original plus reextraction) to a final concentration of 2.3*M*, an equal volume of isopropyl alcohol was added, and nucleic acid was precipitated (−70°C for 15 minutes). The precipitate was centrifuged in a microcentrifuge for 10 minutes, and the pellet was resuspended in 100 μl of TE (10 m*M* tris-HCl, 1 m*M* EDTA) and centrifuged through a G-50 (medium) spin column. The eluate was made 0.2*M* in NaOH and after 10 minutes on ice, HEPES was added to a concentration of 0.24*M*. Two and one-half volumes of ethanol were then added, and the solution containing the precipitate held overnight at −20°C. After centrifugation in a microcentrifuge for 5 minutes, the pellet was resuspended in hybridization buffer, which consisted of [10 m*M* TES, *p*H 7.4, 0.2 percent SDS, 10 m*M* EDTA, 0.3*M* NaCl, 1× Denhardt's, and *Escherichia coli* RNA (250 μg/ml)]. Nitrocellulose filters containing plasmid DNA's were prepared with a Schleicher & Schuell Slot Blot Apparatus under conditions suggested by S and S, except that wells were washed with 10× SSC (saline sodium citrate). These filters were first hybridized in the hybridization solution described above for a minimum of 2 hours at 65°C. After this preliminary hybridization, the filters were hybridized to the runoff products in hybridization solution for 36 hours. A typical reaction contained 2 ml of hybridization solution with 1 × 10^7 cpm/ml. After hybridization, filters were washed for 1 hour in 2× SSC at 65°C. The filters were then incubated at 37°C in 2× SSC with RNase A (10 mg/ml) for 30 minutes and were subsequently washed in 2× SSC at 37°C for 1 hour. Alternatively, after hybridization the filters were washed twice for 15 minutes in 0.1 percent SDS, 2× SSC at room temperature, and then washed at 60°C (0.1 percent SDS, 0.1× SSC) for 30 minutes. Either protocol for processing of the filters after hybridization yielded the same specificity in signal. Filters were then exposed to Kodak XAR film in cassettes containing Lightening-Plus screens at −70°C for various times.

45. C. Yanisch-Perron, J. Vierra, J. Messing, *Gene* **33**, 103 (1985).
46. S. L. McKnight, E. R. Gavis, R. Kingsbury, R. Axel, *Cell* **25**, 385 (1981).
47. M. Groudine and C. Casimir, *Nucleic Acids Res.* **12**, 1427 (1984).

48. We thank many of our colleagues for discussion and suggestions during the course of this work; Hal Weintraub, Paul Neiman, and Craig Thompson for comments on the manuscript; Craig Thompson for assistance in obtaining lymphocyte preparations; Bill Schubach for plasmid pBK25; and Kay Shiozaki for assistance with the manuscript. Supported by NIH grants CA 18282 (M.L.) and CA 28151 (M.L. and M.G.), and NSF grant PCM 82-04696 (M.G.), and a scholarship from the Leukemia Society of America (M.G.)

30 July 1985; accepted 15 October 1985

RESEARCH ARTICLE

Tyrosine Kinase Receptor with Extensive Homology to EGF Receptor Shares Chromosomal Location with *neu* Oncogene

Lisa Coussens, Teresa L. Yang-Feng, Yu-Cheng Liao
Ellson Chen, Alane Gray, John McGrath, Peter H. Seeburg
Towia A. Libermann, Joseph Schlessinger, Uta Francke
Arthur Levinson, Axel Ullrich

Growth factors and their receptors are involved in the regulation of cell proliferation, and several recent findings suggest that they also play a key role in oncogenesis (*1–4*). Of approximately 20 identified oncogenes, the three that have been correlated with known cellular proteins are each related to either a growth factor or a growth factor receptor. The B chain of platelet-derived growth factor (PDGF) is encoded by the proto-oncogene c-*sis* (*2*), the erb-B oncogene product gp68 is a truncated form of the epidermal growth factor (EGF) receptor (*3*), and the proto-oncogene c-*fms* may be related or identical to the receptor for macrophage colony-stimulating factor ($CSF\text{-}1^R$) (*4*).

The receptor-related oncogenes are members of a gene family in that each has tyrosine-specific protein kinase activity, and is associated with the plasma membrane (*5*). Such features are also shared by several other polypeptide hormone receptors, including those for insulin (*6*), PDGF (*7*), and insulin-like growth factor 1 (IGF-1) (*8*); hence more connections may be found between tyrosine kinase growth factor receptors and tyrosine kinase oncogene products.

Comparison of the complete primary structure of the human EGF receptor (*9*) with the sequence of the avian erythroblastosis virus (AEV) transforming gene, v-erbB (*10*), revealed close sequence similarity; in addition, there were amino and carboxyl terminal deletions that may reflect key structural changes in the generation of an oncogene from the gene for a normal growth factor receptor (*3, 9*). Another oncogene, termed *neu*, is also related to v-erbB and was originally identified by its activation in ethylnitrosourea-induced rat neuroblastomas (*11*). In contrast to v-erbB, which encodes a 68,000-dalton truncated EGF receptor, the *neu* oncogene product is a 185,000-dalton cell surface antigen that can be detected by cross-reaction with polyclonal antibodies against EGF receptor (*11*); *neu* may itself be a structurally altered cell surface receptor with homology to the EGF receptor and binding specificity for an unidentified ligand.

Using v-erbB as a screening probe, we isolated genomic and cDNA clones coding for an EGF receptor-related, but distinct, 138,000-dalton polypeptide having all the structural features of a cell surface receptor molecule. On the basis of its structural homology, this putative receptor is a new member of the tyrosine-specific protein kinase family. It is encoded by a 4.8-kb messenger RNA (mRNA) that is widely expressed in normal and malignant tissues. We have localized the gene for this protein to q21 of chromosome 17, which is distinct from the EGF receptor locus, but coincident with the *neu* oncogene mapping position (*12*). We therefore consider the possibility that we have isolated and characterized the normal human counterpart of the rat *neu* oncogene.

Tyrosine kinase–type receptor gene and complementary DNA. As part of our attempts to isolate and characterize the chromosomal gene coding for the human cellular homologue of the viral erbB gp68 polypeptide, AEV-ES4 erbB sequences (2.5-kb Pvu II fragment of pAEV) (*13*) were used as a ^{32}P-labeled hybridization probe for the screening of a human genomic DNA library at reduced stringency

Lisa Coussens, Yu-Cheng Liao, Ellson Chen, Alane Gray, Peter H. Seeburg, Arthur Levinson, and Axel Ullrich are in the Department of Molecular Biology, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080; John McGrath is currently with the Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142; Towia Libermann and Joseph Schlessinger are in the Department of Chemical Immunology at the Weizmann Institute of Science, Rehovot 76100, Israel; and Teresa L. Yang-Feng and Uta Francke are in the Department of Human Genetics at Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510.

(*14*). Clone λc-erbB/1 was isolated; it contained a hybridizing 1.8-kb Bam HI fragment, which was subjected to DNA sequence analysis. The 1838-bp sequence contains three complete and one partial erbB-homologous exons separated by short intervening sequences (Fig. 1). Comparison of this human gene sequence with our complete cDNA-derived human EGF receptor protein sequence (*9*) revealed 32 differences (18.7 percent) within the 171 amino acid stretch of combined exons, suggesting that this gene fragment was not derived from the human EGF receptor gene. Since this gene may code for an unknown tyrosine kinase–type receptor that is closely related to the human EGF receptor, we named it HER2.

Northern blot analysis (*15*) with the ^{32}P-labeled 1.8-kb HER2 fragment as a hybridization probe revealed a 4.8-kb mRNA in human term placenta poly(A)$^+$ RNA, distinct from the 5.8- and 10.5-kb EGF receptor mRNA's also present at high levels in this tissue (Fig. 2a, lane 1). Thus, we had isolated a portion of an EGF receptor–erbB-related but distinct gene. To obtain its complete primary structure, two single-stranded synthetic oligonucleotide probes (*16*) were prepared from HER2 exon sequence regions that differed sufficiently (less than 60 percent nucleotide sequence homology) from EGF receptor DNA sequences (Fig. 1, 1 and 2) and used to screen a term placenta complementary DNA (cDNA) library of 2×10^6 independent recombinant clones in λgt10 (*17*). Fifty-two clones were isolated; they hybridized strongly with both synthetic probes and weakly with an EGF receptor cDNA fragment (HER64-3) (*9*) containing the homologous region within the tyrosine kinase domain. One of these, λHER2-436, had the longest cDNA insert (4.5 kb), consisting of three Eco RI fragments (1.4, 1.5, and 1.6 kb).

The complete cDNA sequence of this clone is shown in Fig. 3. The longest open reading frame starting with a methionine codon codes for a 1255 amino acid polypeptide (137,828 daltons) and contains the 171 residues encoded by the four exons in the 1.8-kp HER2 gene Bam HI fragment (Fig. 1). This 3765-bp coding sequence is flanked by 150 bp of 5′ untranslated sequence and a TGA stop codon, followed by a 627-nucleotide 3′ untranslated sequence. No stop codon is found in the 5′ untranslated region. In support of our assignment, however, the initiation codon at position 151 is flanked by sequences that follow perfectly Kozak's rule (*18*) for translation initiation. The 3′ untranslated sequence contains a potential poly(A) addition signal sequence (AATATA) 12 nucleotides upstream from a stretch of 15 adenylate residues. We are not certain if this $(A)_{15}$ stretch is part of a poly(A) tail or represents an internal poly(A) stretch of a longer 3′ untranslated sequence.

Abstract. *A novel potential cell surface receptor of the tyrosine kinase gene family has been identified and characterized by molecular cloning. Its primary sequence is very similar to that of the human epidermal growth factor receptor and the v-erbB oncogene product; the chromosomal location of the gene for this protein is coincident with the* neu *oncogene, which suggests that the two genes may be identical.*

Comparison of EGF receptor and HER2 sequence. As already indicated by the v-erbB sequence homology used to isolate HER2, the putative HER2 protein is very similar in its overall domain organization and sequence to the EGF receptor. Nevertheless, there are differences that are likely to define a specific biological role for the HER2 polypeptide.

The predicted HER2 polypeptide contains each of the domain features found in hormone receptor precursors, such as those for EGF and insulin (*9, 19*). Such features are apparent in the hydropathy profile (*20*) comparison (Fig. 4a). On the basis of this comparison, and on amino acid sequence alignment with the EGF receptor (Fig. 4b, region 1), we predict a 21 amino acid signal sequence (Fig. 4b, 1), an amino terminal serine residue, and a 632 amino acid putative extracellular ligand-binding domain; a highly hydrophobic, 22–amino acid transmembrane anchor domain separates the extracellular domain from a 580-residue-long carboxyl-terminal cytoplasmic domain, which possesses the highest homology to v-erbB and other members of the tyrosine kinase family.

The 632–amino acid, putative HER2 ligand binding domain is about 40 percent homologous with the 621-residue

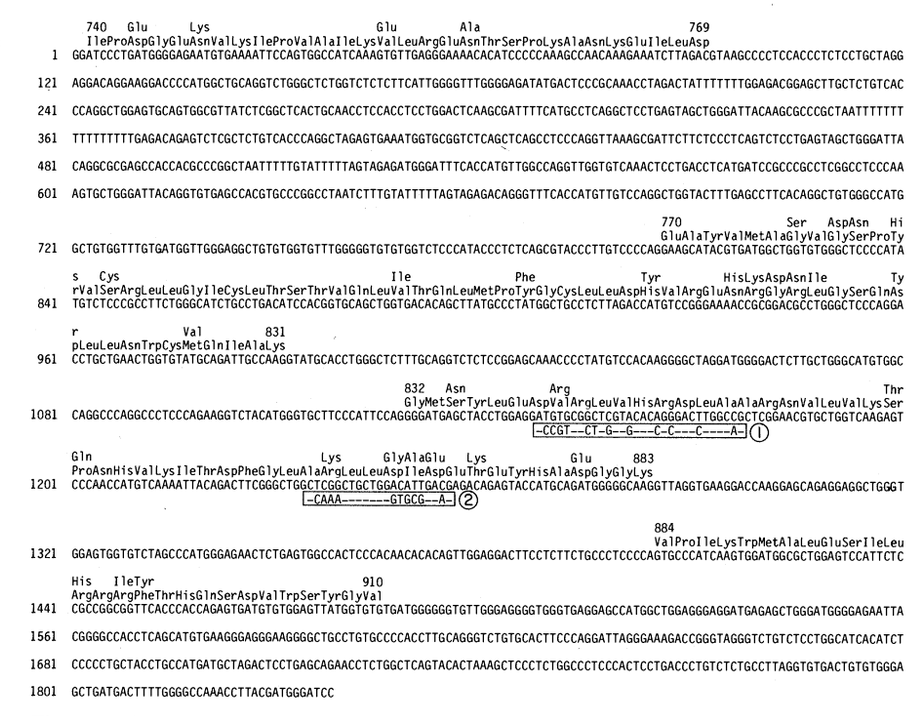


Fig. 1. Partial sequence of the HER2 gene. A partial Hae III-Alu I genomic library (*14*) of human fetal DNA in λ Charon 4A was screened using a radiolabeled 2.5-kb Pvu II fragment of pAEV (*13*) containing coding sequences for the tyrosine kinase domain. Hybridization was as described elsewhere (*31*), except that 30 percent formamide was used at 42°C. Three independent clones were isolated which shared a 1.8-kb hybridizing Bam HI fragment. This fragment and subsets thereof were isolated, subcloned into M13mp10 and M13mp11, and sequenced (*32*). The intron-exon organization was determined by comparison with v-erbB sequences (*10*). Amino acid numbering is based on the complete cDNA sequence shown in Fig. 3. Nucleotide sequence differences with the human EGF receptor sequence are shown in the regions that were used for the design of synthetic oligonucleotide probes 1 (30 nucleotides) and 2 (21 nucleotides). Amino acid sequence differences with the EGF receptor are shown above the HER2 sequence.

extracellular EGF binding domain of the EGF receptor. This homology includes two cysteine-rich subdomains of 26 and 21 regularly organized cysteine residues (Figs. 4a and 2c, subdomains 2 and 3), all of which are conserved in the EGF receptor. The cysteine residue spacing in this region is also homologous with the single cysteine-rich domain in the insulin receptor α subunit (*19*). In contrast, HER2 contains only eight potential *N*-linked glycosylation target sites (Asn-X-Thr or Ser) as compared to 12 in the corresponding region of the EGF receptor. Only five of these are conserved with respect to their relative position in each polypeptide.

The hydrophobic, putative membrane anchor sequence located between residues 653 and 676 (Fig. 4b, region 4) is flanked at its carboxyl terminus by a stretch of amino acids of predominantly basic character (KRRQQKIRKYTMRR) (*21*), as is found in the EGF receptor sequence (*9*) (Fig. 4b, region 5). This region of the EGF receptor contains Thr^{654}, which plays a key role in protein kinase C–mediated receptor modulation (*22*). A homologous threonine residue is embedded in a basic environment in the HER2 sequence at position 685 (Fig. 4, a and b).

The region of most extensive homology (78.4 percent) between EGF receptor and HER2 (beginning at residue 687) extends over 343 amino acids and includes sequences specifying the adenosine triphosphate (ATP) binding domain (*23*) and tyrosine kinase activity (Fig. 4b, region 6) (*5*). This region is also the most conserved between v-erbB and EGF receptor (95 percent) (*9*). The collinear homology between the EGF receptor-erbB and HER2 ceases at position 1032, but introduction of gaps into the EGF receptor or HER2 sequences reveals continued, although decreased, relatedness (Fig. 4b, region 7). This sequence alignment suggests that the two genes evolved by duplication of an ancestral receptor gene, and that subsequent nucleotide sequence divergence in this carboxyl terminal domain led to diverged biological roles for the encoded polypeptides.

The carboxyl terminal domain of HER2 is characterized by an unusually high proline content (18 percent) and predominant hydrophilicity (Fig. 4a). These general features are also found in the EGF receptor carboxyl terminal domain with a 10 percent proline content. The sequences in this region that are found to be conserved are almost exclusively centered around five tyrosine residues, which include the major (Tyr^{1173}) and two minor (Tyr^{1148}, Tyr^{1068}) in vitro autophosphorylation sites in the human EGF receptor (*24*) (Fig. 4, a and b). Three of these tyrosine residues of HER2 (positions 1139, 1196, 1248) are flanked by homologous sequences PQPEYV, ENPEYL, and ENPEYL (*21*), respectively (Fig. 4b, region 7).

HER2 chromosomal location. In situ hybridization of two ^{3}H-labeled HER2 probes (legend, Fig. 5a) to human chromosomes resulted in specific labeling at bands q12→q22 of chromosome 17 (Fig. 5a). Metaphase cells (100) were analyzed for each probe; 40 percent of cells scored for HER2 probe 1 (HER2-1) had silver grains over 17q12→q22 (Fig. 5b). Of the 209 grains observed, 42 (20 percent) were found at this specific region, with no other site labeled above background. For HER2 probe 2, 36 percent of cells had silver grains over the q12→q22 bands of chromosome 17. Of all silver grains, 17 percent (42/246) were localized to this chromosomal region. A secondary site of hybridization with 3.3 percent (8/246) of silver grains was detected at bands p13→q11.2 of chromosome 7.

To test whether this secondary site represented cross-hybridization with the EGF receptor gene, in situ hybridization was carried out with ^{3}H-labeled EGF receptor subclone 64-3. Of 100 cells examined, 30 had silver grains at bands p13→q11.2 of chromosome 7 and 3 percent (5/166) of total grains were found over q12→q22 of chromosome 17. With the other variant probe (HER2-1) no grain accumulation was observed at the EGF receptor site on chromosome 7.

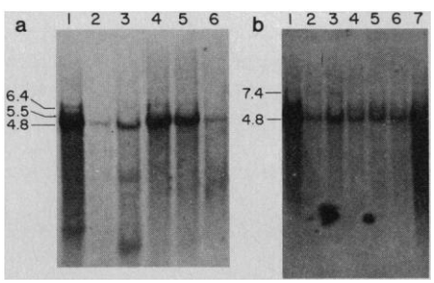


Fig. 2. Northern blot hybridization analysis of normal and malignant human tissues. (a) Fetal tissues; (lane 1) term placenta, (lane 2) 20-week placenta, (lane 3) 20-week liver, (lane 4) 20-week kidney, (lane 5) 20-week lung, (lane 6) 20-week brain. (b) Embryonic tumors; (lane 1) hepatoblastoma, (lanes 2 and 3) Ewing sarcoma, (lane 4) rhabdomyosarcoma, (lanes 5 and 6) neuroblastoma, (lane 7) Wilms' tumor. Total poly(A)$^+$ RNA was isolated as described (*33*); 4 μg per lane was analyzed on a 1 percent formaldehyde-agarose gel. ^{32}P-Labeled HER2-1 and HER-2 (legend to Fig. 5) were used as hybridization probes under high stringency conditions [50 percent formamide, 5× Denhardt's solution, 5× standard saline citrate (SSC), sonicated salmon sperm DNA (50 μg/ml), 50 μ*M* sodium phosphate buffer (*p*H 6.8), 1 m*M* sodium pyrophosphate, and 10 μ*M* ATP at 42°C for 16 hours; filters were washed three times for 15 minutes at 45°C with 0.2× SSC]. The filters were exposed at −60°C with a Cronex Lightning Plus intensifying screen (Dupont) for 7 days. Rat ribosomal RNA's were used as size standards (28*S*, 4.8 kb; 18*S*, 1.8 kb). RNA sizes are given in kilobases.

Southern blot analysis (*25*) of DNA extracted from nine somatic cell hybrids from human and rodent cells confirmed the localization of HER2 sequences to chromosome 17. ^{32}P-labeled HER2-1 and HER2-2 probes were hybridized to the same set of Eco RI–digested DNA samples. With HER2-1, a 13-kb hybridizing band was detected in human DNA (Fig. 5c, lane 1) and in DNA samples from hybrids containing human chromosome 17 (Fig. 5c, lanes 6, 8, 10, and 12). Likewise, hybridization of HER2-2 to a 6.6-kb DNA fragment was observed in human control DNA (Fig. 5c, lane 1) and in hybrids containing human chromosome 17 (Fig. 5c, lanes 6, 8, 10, and 12). Chromosome 17 was the only chromosome with perfect concordant segregation; all other chromosomes were excluded by two or more discordant hybrids.

Regional localization to chromosome 17 was also confirmed by Southern blot analysis. In a mouse-human hybrid containing a rearranged human chromosome 17 with region 17q21→qter, the human HER2 restriction fragments were detected (Fig. 5c, lane 4). The HER2 gene was therefore localized to region 17q21→qter, in agreement with the localization made by in situ hybridization.

Even though a low level of hybridization with probe HER2-2 was seen at the site of the EGF receptor gene on chromosome 7, we were able to show that this finding represented cross-hybridization. In a control experiment an EGF receptor probe cross-hybridized to the same extent with the HER2 site on 17q.

Taken together, the results of the in situ and Southern blot hybridizations permit the site of the HER2 sequences to be further narrowed down to bands 17q21-q22, with the major peak of silver grains at band 17q21.

HER2 expression in normal and malignant tissues. To obtain further clues regarding the function of this receptor both in normal cells and in neoplasms, Northern hybridization analyses (*15*) were carried out with several normal human tissues and randomly collected tumors. A hybridizing 4.8-kb mRNA was detected in all human fetal tissues analyzed, including term placenta, 20-week placenta, liver, kidney, lung, and brain obtained

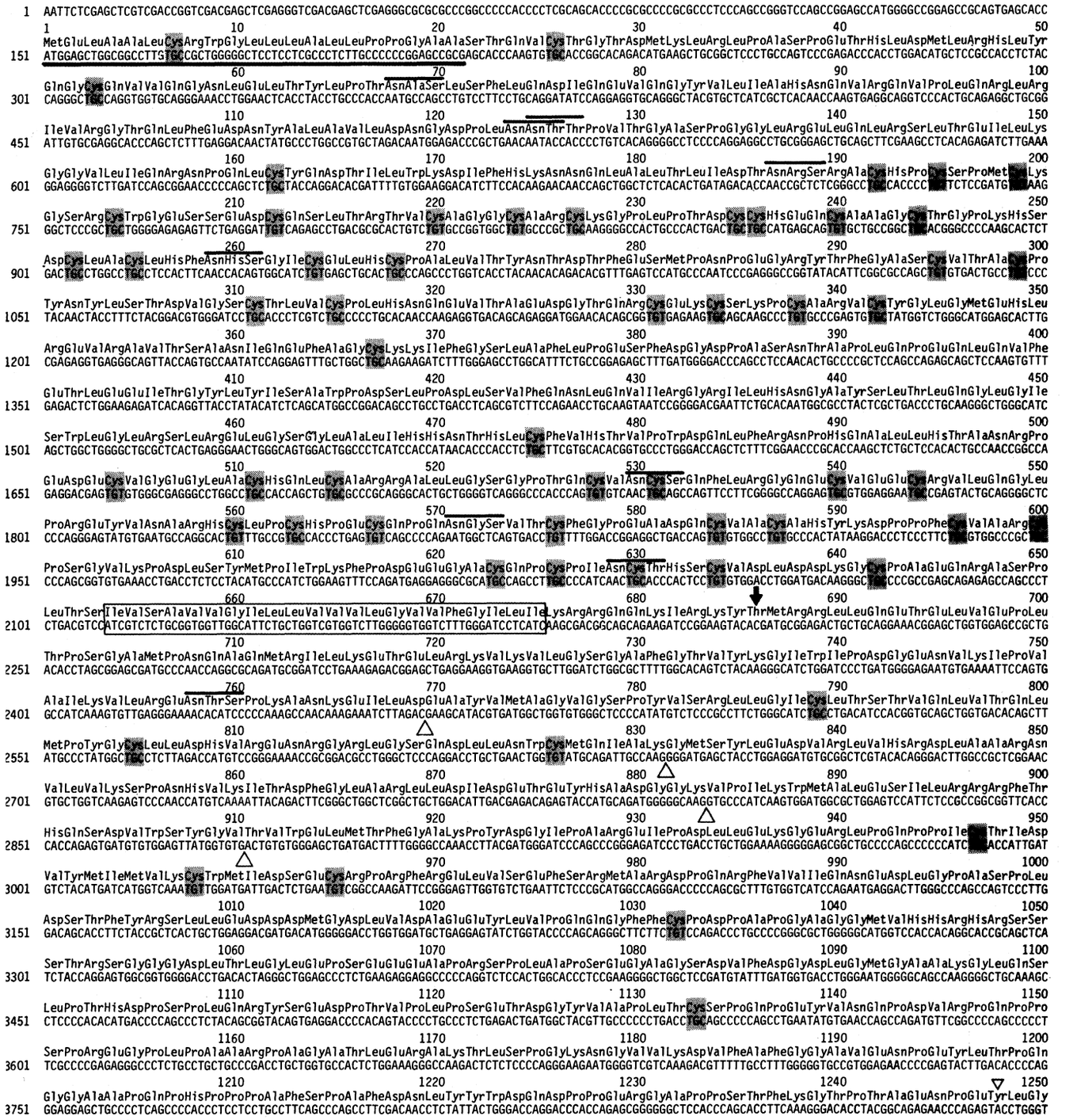
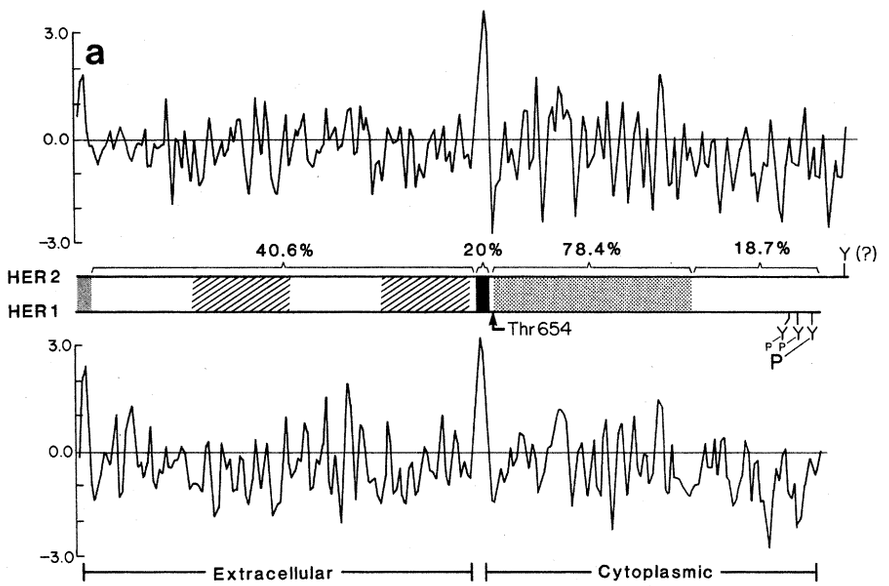


Fig. 3. Complete nucleotide and amino acid sequences of HER2 (clone λHER2-436). Synthetic probes 1 and 2 (Fig. 1) were used to screen 2 × 10^6 clones of a human placental cDNA library in λgt10 (*17*) as described (*9*). Fifty-two clones were isolated and characterized by Eco RI restriction digestion and Southern blot hybridization (*25*) with the radiolabeled synthetic oligonucleotides (*34*) and, on a duplicate blot, an EGF receptor Eco RI fragment (HER64-3) as probes. Labeling of oligonucleotides was achieved with 1 unit of T4 polynucleotide kinase (New England Biolabs), 10 pmol of oligonucleotide and 30 pmol of [γ-^{32}P]ATP (Amersham) as specified by the supplier. HER64-3 was labeled to high specific activity (>10^8 cpm/μg) as described (*34*). Clone λHER2-436 was used for nucleotide sequence analysis (*32*) of overlapping subclones. Amino acids are numbered starting with the initiation methionine (1). Underlining indicates the putative signal sequence (heavy) and the potential poly(A) addition signal (fine). Lines on top of the sequence indicate potential glycosylation sites, the black arrow demarcates the EGF receptor threonine 654 analogue (Thr 686), shading indicates cysteine residues, boxing shows the putative transmembrane region, large open triangles indicate some locations of introns in the HER2 gene, and the small triangle emphasizes a possible tyrosine autophosphorylation site by homology to Tyr 1173 in the EGF receptor sequence.

from a single fetus (Fig. 2a). Two mRNA's, of 5.4 and 6.4 kb, were also detected in term placenta. No cross-hybridization with the 5.8-kb and 10.5-kb EGF receptor mRNA's in term placenta mRNA was observed under these stringent hyridization conditions (legend, Fig. 2). Normal adult human tissues, including kidney, liver, skin, lung, jejunum, uterus, stomach, and colon, contained lower but significant amounts of the same 4.8-kb mRNA. Because of the magnitude of fetal expression, we also examined several embryonic tumors (Fig. 2b); each expressed large amounts of the 4.8-kb transcript, although not more than that detected in normal fetal tissue.

Thus, it appears that the HER2 gene is widely expressed, in both normal adult tissues and in several normal fetal tissues. While detected in most embryonic tumors, the HER2 gene was not present at higher levels than in fetal tissues; thus, the particular level may reflect the state of differentiation of a given tumor.

HER2 structurally characterized as cell surface receptor. Using the transforming gene of the avian erythroblastosis virus, v-erbB, as a hybridization probe, we isolated genomic and cDNA sequences of an uncharacterized human gene. The 1255 amino acid polypeptide sequence



b

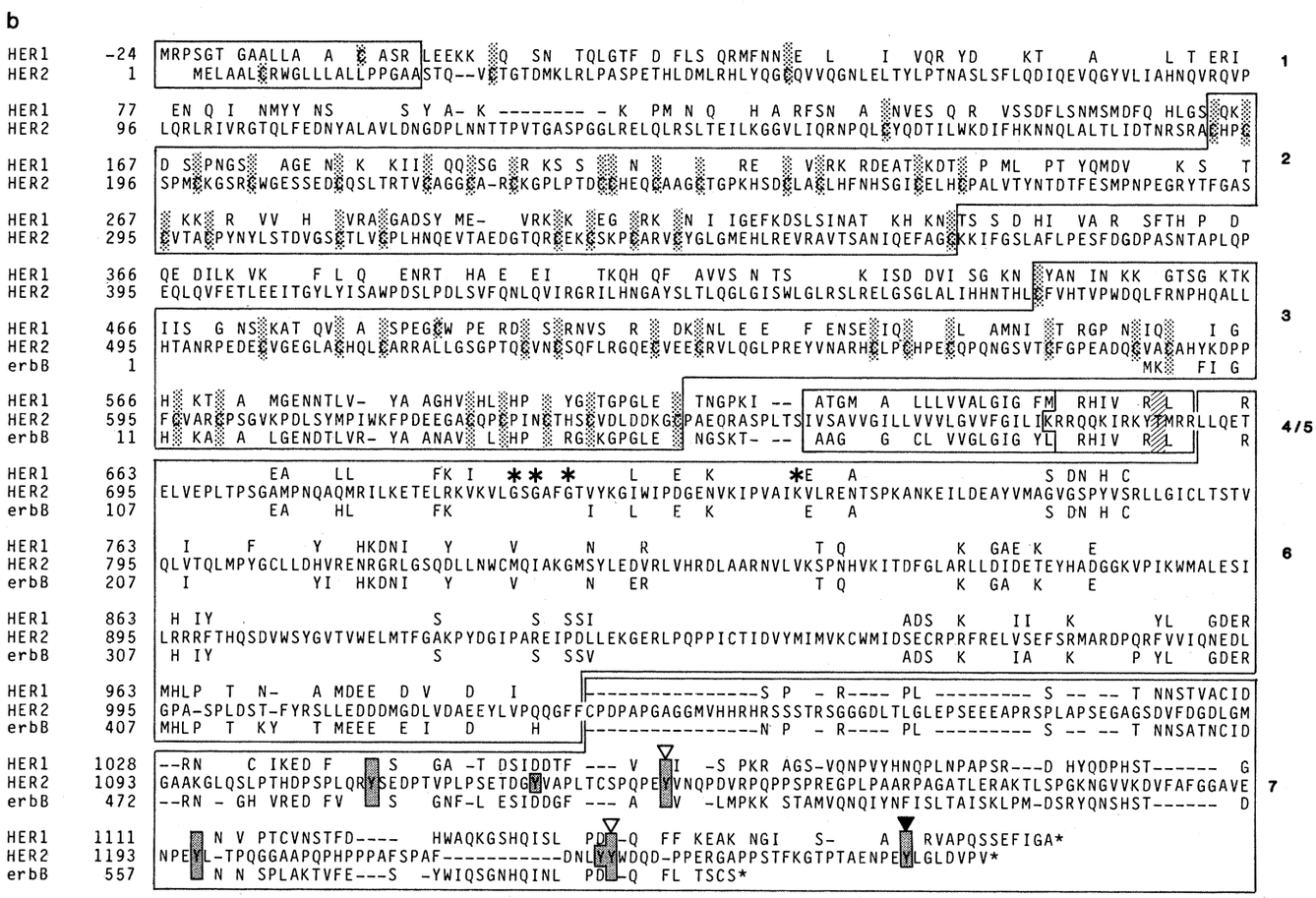


Fig. 4. (a) Hydropathy analysis (*20*) of HER2 (1255 amino acids) coding sequences and comparison with EGF receptor (HER1; 1210 amino acids). Different receptor domains and the extent of amino acid sequence homology are indicated. The autophosphorylation tyrosines in the EGF receptor sequence are shown, as is a potential analogue in HER2. The signal sequence is shown by fine shading, the cysteine-rich subdomains by hatching, the transmembrane region by a black bar, and the tyrosine kinase domain by coarse shading. (b) Comparison of HER2, EGF receptor (HER1), and v-erbB amino acid sequences. Identical residues are deleted, gaps are introduced to optimize alignment, cysteine residues are shaded, and carboxyl terminal tyrosines are in shaded boxes. The black and open triangles indicate the positions of the major and minor autophosphorylation sites in the EGF receptor, respectively. Asterisks indicate residues involved in ATP binding. Boxed regions include (i), signal sequence; (ii, iii), cysteine-rich domains; (iv), transmembrane domain; (v), protein kinase C modulation domain; (vi), tyrosine kinase domain; (vii), signal transfer domain.

derived from this cloned cDNA shows extensive homology to v-erbB and its cellular homologue, the human EGF receptor, and was therefore termed HER2. Its primary amino acid sequence displays all the structural features known to define a cell surface receptor for a polypeptide ligand. This finding represents the isolation and detailed structural characterization of a putative cell surface receptor before determination of its specific ligand, based solely on its structural homology to another receptor gene. In a similar case, the v-*fms* gene (*26*) was characterized because of its oncogenic activity, and was later determined to be identical or related to the receptor for CSF-1, a hematopoietic growth factor (*4*).

Examination of the HER2 primary sequence reveals extensive collinearity with the EGF receptor sequence. Detailed sequence comparison of these two closely related receptors, each having presumably different biological roles, provides important clues to the functional significance of the different regions of the proteins. A hydrophobic amino terminal sequence of HER2 presumably functions as a signal sequence (Fig. 4b, region 1) that, like the pre-EGF receptor, would be cleaved after alanine (*22*) to generate an amino terminal serine; however, cleavage after glycine (*19*) may occur, since signal peptidase frequently cleaves after small amino acid residues. Again, as in the case of the EGF receptor, the next 632 amino acids would represent the extracellular domain of the HER2 receptor that forms the binding pocket for a specific polypeptide ligand. Some structural features in this domain are highly conserved, while others have diverged extensively. Of the 51 cysteine residues of the EGF receptor ligand binding domain, 50 are conserved in HER2, of which 21 and 26, respectively, are clustered in subdomains 2 and 3 (Fig. 4b). Sequences between these conserved, regularly spaced cysteine residues, as well as the ones flanking subdomains 2 and 3 (Fig. 4b), have diverged to result in an overall homology of 40.6 percent (Fig. 4a).

Cysteine rich clusters (*27*) are found in receptors for EGF (*9*), low-density lipoprotein (*28*), and insulin (*19*), as well as in the EGF precursor (*29*), which is a potential receptor (Fig. 6). Since cysteine-rich domains are found in a family of cell surface receptors and potential receptors with very diverse functions, they are likely to form an essential structural backbone of the ligand binding domain, but not define ligand specificity. The ligand affinity of each of these receptors may be specified by sequences flanking these cysteine residue clusters.

The presence of regularly repeated, extracellular cysteine clusters (*27*) defines a gene family that overlaps with another gene family, the membrane-spanning proteins that use an intrinsic tyrosine-specific kinase activity in the signal transduction process (*5*) (Fig. 6). v-*fms*, the oncogenic homologue of the M-CSF-1 (colony-stimulating factor) receptor, contains a tyrosine kinase domain (*26*), but lacks cysteine-rich clusters in its extracellular domain, and thus is only a member of this second gene family.

Only 5 of the 12 potential *N*-linked glycosylation sites of the EGF receptor are conserved with respect to their approximate position in HER2, which also contains only eight such sites in this extracellular domain. The role of carbohydrate side chain differences in defining the biological function of these proteins requires further investigation.

The cytoplasmic domains of EGF receptor and HER2 also exhibit striking regional differences in sequence homology. These differences (Fig. 4, a and b) most likely distinguish regions that define receptor-specific, ligand-induced signal generation from support functions such as ATP binding and enzymatic tyrosine kinase activity.

Directly adjacent to the putative transmembrane domain, a cluster of 14 predominantly basic amino acids is found in both HER2 and EGF receptor sequences, which may interact with the membrane phospholipid head groups. The Thr^{654} residue, which is located within this region of the EGF receptor, is a phosphorylation substrate site for receptor activity down-modulation by protein kinase C (*22*). The HER2 sequence contains a threonine residue at an analogous position (686) (Figs. 3 and 4a and region 5, Fig. 4b) and may therefore also be subject to protein kinase C modulation.

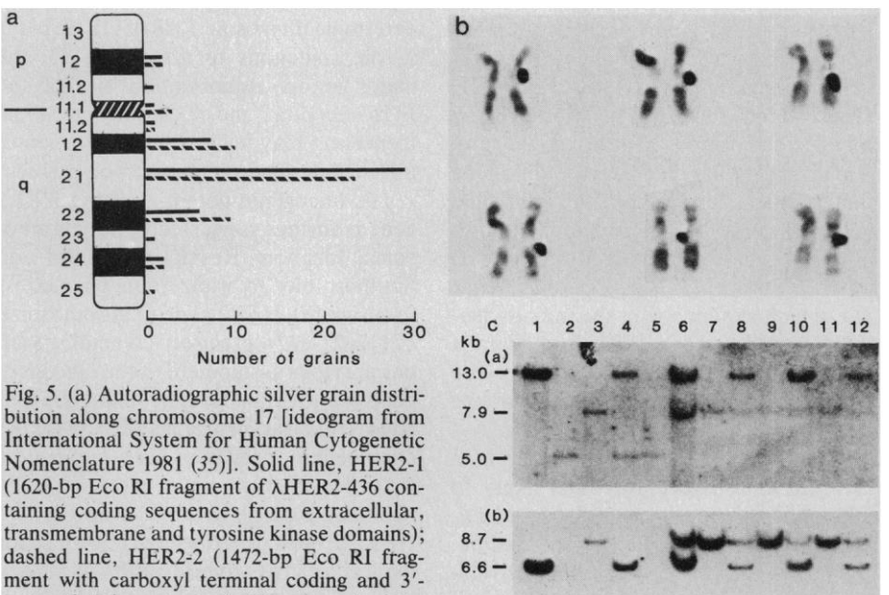


Fig. 5. (a) Autoradiographic silver grain distribution along chromosome 17 [ideogram from International System for Human Cytogenetic Nomenclature 1981 (*35*)]. Solid line, HER2-1 (1620-bp Eco RI fragment of λHER2-436 containing coding sequences from extracellular, transmembrane and tyrosine kinase domains); dashed line, HER2-2 (1472-bp Eco RI fragment with carboxyl terminal coding and 3′-untranslated sequences). Human metaphase and prometaphase chromosomes were prepared from methotrexate-synchronized peripheral lymphocyte cultures (*36*) of two normal individuals. The probes for in situ hybridization (*37*) were ^{3}H-labeled by nick translation with three labeled nucleotides to specific activities of 2.4×10^{7} cpm/μg (HER2-1), 3.4×10^{7} cpm/μg (HER2-2), and 2.5×10^{7} cpm/μg (HER1 64-3); the probes (25 or 50 ng/ml) were hybridized to chromosome preparations (16 hours at 37°C). The emulsion-coated slides were exposed for 10 to 14 days at 4°C. Chromosomes were stained with Wright's stain for G-banding, and a second photograph was taken of the previously selected cells. G-banded chromosomes were analyzed for silver grain localizations. (b) Pairs of chromosome 17 from three cells, illustrating typical labeling at bands q12→q22 for HER2-1 (top row) and HER2-2 (bottom row). (c) Southern blot hybridization of HER2-1 (a) and HER2-2 (b) to Eco RI–digested DNA from somatic cell hybrids of humans and rodents. Somatic cell hybrids of four different series were used. Their origin and characterization has been reported (*38*). Samples (10 μg) of hybrid, mouse, Chinese hamster, and human DNA were cleaved with Eco RI, subjected to electrophoresis on 0.8 percent agarose gels, transferred to nitrocellulose filters, and hybridized to ^{32}P-labeled HER2 probes according to standard procedures (*26, 39*). (Lane 1) Human lymphoblastoid control cell line; (lane 2) mouse 3T3 cells; (lane 3) Chinese hamster cell line; (lane 4) subclone of hybrid of mouse and human cells with a rearranged chromosome 17 containing only region q21→qter, retained by growth in hypoxanthine-amnopterin-thymidine (HAT) medium; (lane 5) same subclone as lane 4, except missing this rearranged chromosome after counterselection in bromodeoxyuridine medium; (lanes 6, 8, 10, and 12) clones of Chinese hamster and human hybrid cells containing human chromosome 17; (lanes 7, 9, and 11) Chinese hamster X human hybrid cell clones lacking human chromosome 17.

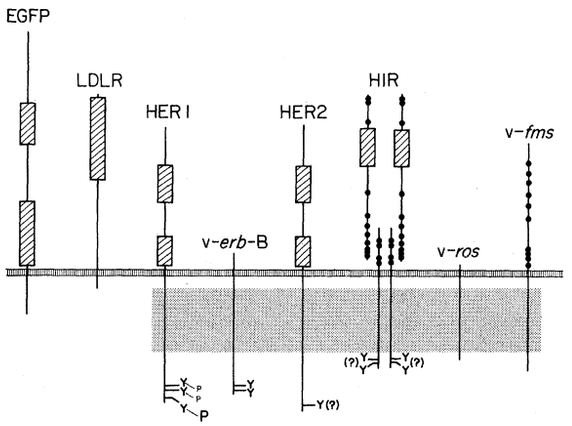


Fig. 6. Schematic comparison of HER2-related molecules. Hatched boxes indicate cysteine-rich regions within extracellular or putatively extracellular domains. The shaded region below the membrane bilayer indicates the location of the tyrosine kinase domain. Black circles within the human insulin receptor (HIR) and v-*fms* structures indicate cysteine residues that are not included in cysteine clusters and may be involved in heterotetramer (HIR) or ligand binding pocket formation (v-*fms*). EGFP, EGF precursor; LDLR, LDL receptor; HER1,

The subsequent 341 residues are not only highly homologous to the EGF receptor and the v-erbB oncogene, but also to all the other members of the *src* family of tyrosine kinase oncogenes (*22*), as well as the β subunit of the insulin receptor (*19*). This sequence similarity suggests that this HER2 region encompasses at least a portion of the kinase domain that may be activated by extracellular ligand binding. A Gly X Gly XX Gly sequence at HER2 position 727 and a lysine at position 753 are known from studies on other members of this gene family to be involved in ATP binding (*23*). Overall, this HER2 region (6 in Fig. 4b) is the most homologous to v-erbB and EGF receptor (78.4 percent; Fig. 4a), which explains why the initially isolated genomic clone (Fig. 1) contained exons coding for part of this tyrosine kinase domain (intron positions are shown in Fig. 3).

The region within the sequence of a tyrosine kinase receptor most likely to be responsible for translation of the activation signals intitiated by specific, extracellular ligand binding into physiological action is localized at the carboxyl terminus of the receptor polypeptide chain. This region contains the 32–amino acid sequence deletion that plays a crucial role in the generation of the AEV transforming gene v-erbB from the EGF receptor proto-oncogene (Fig. 6) (*30*), as well as the tyrosine residues that are major (1173) and minor (1068 and 1148) sites of ligand-induced autophosphorylation in the EGF receptor (*24*) (Fig. 4, a and b). This region contains the most extensive differences in length and primary structure between HER2 and EGF receptor/erbB. However, five of seven tyrosines are conserved in HER2, including short surrounding sequences. This suggests that, like the conserved cysteine residues in the ligand binding domain, these tyrosine residues may serve an auxiliary activity-control function in the generation of a cytoplasmic signal. Other sequences in this signal-control and transfer domain are likely to define specificity, which may include recognition of possible substrates and/or cofactors. Future work will be needed to determine if tyrosine 1248 of HER2 plays a role analogous to tyrosine 1173, the major autophosphorylation site of the EGF receptor, and if carboxyl terminal truncation may lead to the generation of a HER2-derived oncogenic polypeptide.

The oncogenic potential of the HER2 gene is further suggested by its chromosomal location. Results of in situ and Southern blot hybridizations enabled us to assign the HER2 gene to chromosome 17 (17q21-q22), precisely coinciding with our previous assignment of the *neu* oncogene (*12*), a transforming sequence isolated from rat neuroblastoma (*11*). As documented elsewhere (*12*), our results with the rat *neu* probe were virtually identical to those obtained with the two HER2 probes described here, except for the lack of cross-hybridization of rat *neu* genomic sequence with the EGF receptor locus on chromosome 7. Our mapping results support the notion that the proto-oncogene *neu* and the HER2 gene, although both are very likely not distinct from each other. Further detailed analysis will be necessary to establish whether they actually represent the same gene.

Can we obtain clues to the cellular role of this putative receptor by examination of its gene expression? Northern blot hybridization analysis of poly(A)$^+$ RNA's from a variety of normal and malignant tissues implicates a role for the HER2 receptor-ligand system in many cell types in both fetal and adult life. Highest levels of the 4.8-kb mRNA were found in several fetal tissues, which rules out a cell lineage–specific receptor for a growth or differentiation factor and is consistent with a more general role in fetal development. No clues were obtained from tumor RNA screening to suggest that increased expression was involved in tumor generation or growth. In some tissues, such as term placenta, fetal kidney, and embryonic tumors, we found additional larger (5.5, 6.4, and 7.4 kb) mRNA's hybridizing with our probe, suggesting transcriptional diversity or the presence of homologous mRNA's more closely related to HER2 than to the EGF receptor. Further experiments should establish the biological role of the HER2 gene, and its role in oncogenesis.

References and Notes

1. G. Guroff, Ed., *Growth and Maturation Factors* (Wiley, New York, 1983), vol. 1.
2. M. D. Waterfield *et al.*, *Nature (London)* **304**, 35 (1983); R. F. Doolittle *et al.*, *Science* **221**, 275 (1983).
3. J. Downward *et al.*, *Nature (London)* **307**, 521 (1984).
4. C. J. Sherr *et al.*, *Cell* **41**, 665 (1985).
5. T. Hunter and J. A. Cooper, *Annu. Rev. Biochem.* **54**, 897 (1985).
6. M. Kasuga, F. A. Karlsson, C. R. Kahn, *Science* **215**, 185 (1982); M. Kasuga *et al.*, *J. Biol. Chem.* **257**, 9891 (1982); M. Kasuga, Y. Zick, D. L. Blithe, M. Crettaz, C. R. Kahn, *Nature (London)* **298**, 667 (1982); Y. Zick, M. Kasuga, C. R. Kahn, J. Roth, *J. Biol. Chem.* **258**, 75 (1983); L. M. Petruzzeli *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6792 (1982); M. A. Shia and P. F. Pilch, *Biochemistry* **22**, 717 (1983); R. A. Roth and D. J. Cassell, *Science* **219**, 299 (1983); E. Van Obberghen, B. Rossi, A. Kowalski, H. Gazzano, G. Ponzio, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 945 (1983); J. Avruch, R. A. Nemenoff, P. J. Blackshear, M. W. Pierce, R. Osathanondh, *J. Biol. Chem.* **257**, 15162 (1982).
7. B. Ek, B. Westermark, A. Wasteson, C.-H. Heldin, *Nature (London)* **295**, 419 (1982).
8. S. Jacobs *et al.*, *J. Biol. Chem.* **258**, 9581 (1983); J. B. Rubin, M. A. Shia, P. F. Pilch, *Nature (London)* **305**, 438 (1983).
9. W. Weber *et al.*, *Science* **224**, 294 (1984); G. T. Merlino *et al.*, *ibid.*, p. 417; A. Ullrich *et al.*, *Nature (London)* **309**, 418 (1984).
10. T. Yamamoto *et al.*, *Cell* **35**, 71 (1983).
11. A. L. Schechter *et al.*, *Nature (London)* **312**, 513 (1984); C. Shih, L. C. Padhy, M. Murray, R. A. Weinberg, *ibid.* **290**, 261 (1981).
12. A. L. Schechter *et al.*, *Science* **229**, 976 (1985).
13. B. Vennstrom, L. Fanshier, C. Moscovici, J. M. Bishop, *J. Virol.* **36**, 575 (1980).
14. R. M. Lawn, E. F. Fritsch, R. C. Parker, G. B. Lake, T. Maniatis, *Cell* **15**, 1157 (1978).
15. H. Lehrach, D. Diamond, J. M. Wozney, H. Boedtker, *Biochemistry* **16**, 4743 (1977).
16. R. Crea and T. Horn, *Nucleic Acids Res.* **8**, 2331 (1980).
17. T. Huynh, R. Young, R. Davis, in *Practical Approaches in Biochemistry*, D. Grover, Ed. (IRL, Oxford, 1984).
18. M. Kozak, *Nucleic Acids Res.* **9**, 5233 (1981).
19. A. Ullrich *et al.*, *Nature (London)* **313**, 756 (1985).
20. J. Kyte and R. F. Doolittle, *J. Mol. Biol.* **157**, 105 (1982).
21. M. O. Dayhoff, *Atlas of Protein Sequence and Structure 5*, supplement 3 (National Biomedical Research Foundation, Washington, D.C., 1978). The one-letter abbreviations recommended by the IUPAC-IUB Commission on Biochemical Nomenclature are used: E, glutamic acid; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; T, threonine; V, valine; and Y, tyrosine.
22. T. Hunter, N. Ling, J. A. Cooper, *Nature (London)* **311**, 480 (1984).
23. M. L. Privalsky, R. Ralston, J. M. Bishop, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 704 (1984); W. C. Barker and M. O. Dayhoff, *ibid.* **79**, 2836 (1982); W. Moller and R. Amons, *FEBS Letts.* **186**, 1 (1985); M. J. E. Sternberg and W. R. Taylor, *ibid.* **175**, 387 (1984).

24. J. Downward, P. Parker, M. D. Waterfield, *Nature (London)* **311**, 483 (1984).
25. E. Southern, *Methods Enzymol.* **68**, 152 (1979).
26. S. J. Anderson *et al.*, *J. Virol.* **44**, 696 (1982); R. Manger *et al.*, *Cell* **39**, 327 (1984); C. W. Rettenmier *et al.*, *ibid.* **40**, 971 (1985); M. F. Roussel *et al.*, *Mol. Cell. Biol.* **4**, 1999 (1984).
27. S. Pfeffer and A. Ullrich, *Nature (London)* **313**, 184 (1985).
28. T. Yamamoto *et al.*, *Cell* **39**, 27 (1984).
29. A. Gray, T. J. Dull, A. Ullrich, *Nature (London)* **303**, 722 (1983); J. Scott *et al.*, *Science* **221**, 236 (1983).
30. H. Riedel and A. Ullrich, unpublished data.
31. A. Ullrich, C. H. Berman, T. J. Dull, A. Gray, J. M. Lee, *EMBO J.* **3**, 361 (1984).
32. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977); J. Messing and J. Vieira, *Gene* **19**, 269 (1982); J. Messing, R. Crea, P. H. Seeburg, *Nucleic Acids Res.* **9**, 309 (1981).
33. G. Cathala *et al.*, *DNA* **2**, 329 (1983).
34. J. M. Taylor, R. Illmensee, J. Summers, *Biochim. Biophys. Acta* **4**, 324 (1976).
35. ISCN, "An international system for human cytogenetic nomenclature—high resolution banding," *Cytogenet. Cell Genet.* **31**, 1 (1981).
36. J. J. Yunis, *Science* **191**, 1268 (1976).
37. M. E. Harper and G. F. Saunders, *Chromosoma* **83**, 431 (1981).
38. U. Francke and B. de Martinville, *Banbury Report* **14**, 175 (1983); U. Francke and M. A. Pellegrino, *Proc. Natl. Acad. Sci. U.S.A.* **7**, 1147 (1977); U. Francke, *Cytogenet. Cell Genet.* **38**, 298 (1984); _____ and B. Francke, *Somat. Cell Genet.* **7**, 171 (1981).
39. G. M. Wahl *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3683 (1979).
40. C. R. King, M. H. Kraus, S. A. Aaronson, *Science* **229**, 974 (1985).
41. We thank D. Smith, who was involved in the early stages of cDNA sequencing, W. Cavenee and X. Konfos for providing embryonic tumors and X. Breakefield for providing fetal tissues. We also thank M. Wechser and A. Douglas for technical assistance, J. Arch for preparation of the manuscript, and S. Pfeffer for discussions and help during the preparation of the manuscript. After completion of this work, King *et al.* (*40*) published a partial sequence of an EGF receptor-related gene that matches our HER2 sequence perfectly.

20 September 1985; accepted 30 October 1985

AAAS–Newcomb Cleveland Prize
To Be Awarded for an Article or a Report Published in *Science*

The AAAS–Newcomb Cleveland Prize is awarded annually to the author of an outstanding paper published in *Science*. The 1985 competition starts with the 4 January 1985 issue of *Science* and ends with the issue of 27 December 1985. The value of the prize is $5000; the winner also receives a bronze medal.

Reports and Articles that include original research data, theories, or syntheses and are fundamental contributions to basic knowledge or technical achievements of far-reaching consequence are eligible for consideration for the prize. The paper must be a first-time publication of the author's own work. Reference to pertinent earlier work by the author may be included to give perspective.

Throughout the year, readers are invited to nominate papers appearing in the Reports or Articles sections. Nominations must be typed, and the following information provided: the title of the paper, issue in which it was published, author's name, and a brief statement of justification for nomination. Nominations should be submitted to the AAAS–Newcomb Cleveland Prize, AAAS, 1333 H Street, NW, Washington, D.C. 20005. Final selection will rest with a panel of distinguished scientists appointed by the editor of *Science*.

The award will be presented at a session of the AAAS annual meeting. In cases of multiple authorship, the prize will be divided equally between or among the authors.