Characterization of cDNA Clones for the Human c-yes Gene

JUN SUKEGAWA, KENTARO SEMBA, YUJI YAMANASHI, MAKOTO NISHIZAWA, NOBUYUKI MIYAJIMA, TADASHI YAMAMOTO, and KUMAO TOYOSHIMA*

The Institute of Medical Science, University of Tokyo, 4-6-1, Shirokanedai, Minato-ku, Tokyo 108, Japan

Received 2 June 1986/Accepted 24 September 1986

Three c-yes cDNA clones were obtained from $poly(A)^+$ RNA of human embryo fibroblasts. Sequence analysis of the clones showed that they contained inserts corresponding to nearly full-length human c-yes mRNA, which could encode a polypeptide of 543 amino acids with a relative molecular weight (M_r) of 60,801. The predicted amino acid sequence of the protein has no apparent membrane-spanning region or suspected ligand binding domain and closely resembles pp60^{c-src}. Comparison of the sequences of c-yes and v-yes revealed that the v-yes gene contains most of the c-yes coding sequence except the region encoding its extreme carboxyl terminus. The region missing from the v-yes protein is the part that is highly conserved in cellular gene products of the protein-tyrosine kinase family.

The human cellular yes gene, c-yes-1, was initially identified as a homolog of v-yes, the oncogene of avian sarcoma virus Y73 (47), and was thought to be involved in a case of human gastric cancer (32). The product of the v-yes gene, $pp90^{gag-yes}$, exerts tyrosine-specific protein kinase activity (18). The amino acid sequence of the yes specific part of $pp90^{gag-yes}$ is very similar to that of $pp60^{src}$, and the yes gene is considered to be the most intimate cognate of the src gene (19). Although there is accumulating information about the nature of the c-src gene product, $pp60^{c-src}$ (3, 8, 11, 20, 37), not even the structure of the c-yes gene product has been elucidated.

At least two v-onc genes coding for tyrosine kinases are derived from genes encoding receptor molecules for polypeptide growth factors: the v-fms and v-erbB genes have been shown to form parts of the genes encoding feline colony-stimulating factor 1 receptor and avian epidermal growth factor receptor, respectively (9, 35). In contrast, the predicted amino acid sequence of c-src shows no membranespanning region or extracellular domain for ligand binding (39). Nevertheless, $pp60^{c-src}$ is localized on the inner surface of the plasma membrane and again is thought to be involved in as yet unknown processes of signal transduction at the plasma membrane.

We recently found that c-fgr, a cellular homolog of the oncogene of the Gardner-Rasheed strain of feline sarcoma virus (26), and two novel genetic loci named syn (33) and lyn (45a) in the human genome could encode protein-tyrosine kinases that are very similar to but distinct from the src and yes gene products. Our primary interest is to determine the functions of these genes that are similar in structure and to find out whether these kinases are involved in the key steps in the process of neoplastic transformation of cells. We have isolated c-yes cDNA clones from cultured fibroblasts of human embryo, and here we report the complete structure of these cDNA clones. This information should be useful in determining the normal functions of the c-yes gene product.

MATERIALS AND METHODS

Isolation of cDNA clones. A human embryo fibroblast cDNA library was constructed as described elsewhere by using a λ gt10 vector system (33). The library (3.7 \times 10⁵

plaques) was screened with a 0.9-kilobase-pair (kbp) DNA probe prepared from the human c-yes-2 gene, a processed pseudogene of c-yes-1 (34). The 1.2-kbp BamHI-EcoRI fragment of the c-yes-2 pseudogene was freed from repetitive elements by removing a 0.3-kbp PstI fragment containing an Alu repeat. Comparison of the nucleotide sequences of the resultant 0.9-kbp pseudogene fragment and genomic c-yes-1 DNA confirmed that this DNA fragment reproduced the human c-yes-1 coding sequence more faithfully (95% homology) than did the v-yes gene (85% homology) (Semba et al., unpublished data). Hybridization was carried out in a solution of 50% formamide, 4× SSC (1× SSC is 150 mM NaCl plus 15 mM sodium citrate), 50 mM N-2-hydroxyethylpiperazine-N'-2-Ethanesulfonic acid (pH 7.0), and $10\times$ Denhardt solution at 42°C. The filters were finally washed with a solution of $0.2 \times$ SSC-0.1% sodium dodecyl sulfate at 50°C.

Nucleotide sequence determination. Nucleotide sequencing analyses were performed by the dideoxy chain termination procedure (30). Complementary DNA inserts were recloned into M13 phage vector mp18 and unidirectionally deleted with *Escherichia coli* exonucleases III and VII (46). Some G+C-rich sequences were determined with deoxy-7deazaguanosine triphosphate in place of dGTP (24).

RESULTS

Isolation of cDNA clones. A cDNA library was constructed with the λ gt10 phage vector system and poly(A)⁺ RNA from cultured human embryo fibroblasts. By use of the human c-yes-2 pseudogene probe, we carried out the screening procedures under stringent conditions to eliminate signals caused by other yes-related genes (34). By screening about 3.7×10^5 plaques, we obtained three positive clones, designated as λ YS-7, λ YS-12, and λ YS-17. Restriction mapping showed that the clones covered approximately 4.5 kb of the 4.8-kb length of c-yes-1 mRNA (Fig. 1).

Analysis of cDNA clones. All four EcoRI fragments of the cDNA clones were subjected to unidirectional deletion procedures and subsequent nucleotide sequence analysis. The overlapped portions of each clone showed completely identical nucleotide sequences. The 5' end of clone λ YS-17 was about 50 nucleotides shorter than that of clone λ YS-12. Continuity of the sequence across the internal EcoRI site was confirmed by comparing the sequence with that of a

^{*} Corresponding author.



FIG. 1. Restriction maps of cDNA clones. The line shown above is a scheme of total c-yes mRNA. The open box indicates the predicted coding frame. The mapped restriction sites are abbreviated as follows: Bm, BamHI; E, EcoRI; Pv, PvuII; Sp, SphI; Ss, SstI; Xh, XhoI.

genomic clone of human c-yes-1. The sequence of 4472 nucleotides covered by clones λ YS-12 and λ YS-7 contains an open reading frame specifying a 543-amino-acid polypeptide initiated with methionine at position 163 (Fig. 2). Although precise location of the 5' terminus of the c-yes mRNA was not examined, the codon seems to be a bona fide initiating codon for the polypeptide, because there is no additional ATG triplet preceding this one, and the reading frame begins just behind the termination codon at the nucleotide at position 4.

The relative molecular weight of the deduced polypeptide, 60,801, is very similar to those of the c-src and syn proteins. The amino acid sequence of the polypeptide from residue 91 to the carboxy-terminal residue 543 that includes the suspected kinase domain shows 96% identity to the corresponding region of the v-yes protein (excluding the eight carboxyterminal amino acids), 84% identity to that of the chicken c-src protein, and 82% identity to that of the syn protein. In addition, the sequence retains a high degree of homology to the v-yes protein, even in its amino terminal moiety (residues 1 to 90), except in some parts (residues 34 to 47; see below), whereas the same regions of the c-src and syn proteins show less homology to the v-yes protein (Fig. 3). These data seem to validate the conclusion that the clones represent the c-yes-1 genetic locus, the human homolog of the v-yes gene.

As in other tyrosine kinases, a well-confirmed consensus sequence, Gly-X-Gly-X-Cly (residues 284 through 289), and Lys-305 for ATP binding are conserved in the polypeptide (16, 44). Moreover, Tyr-426 corresponds to Tyr-416 of $pp60^{c-src}$, which is the target of the autophosphorylation reaction (36). These features suggest that the c-yes-1 protein is a protein-tyrosine kinase.

 $pp60^{c-src}$ was reported to be phosphorylated at Ser-17 by A kinase (6) and at Ser-12 by C kinase (13). But the serine or threonine residues at homologous positions in the c-yes protein cannot be identified because of the limited degree of homology between $pp60^{c-src}$ and the c-yes-1 protein in their amino-terminal regions.

Myristic acid is attached to Gly-2 of $pp60^{c-src}$ through an amide bond, whereas the initiating methionine of the protein is removed in the mature form (4, 17). In the predicted c-yes-1 protein, the initiating methionine is also followed by glycine. This Gly-2 therefore may be the myristylation signal in the human c-yes protein, whereas the corresponding Gly-278 could be silent in the v-yes protein. Studies are required on whether the c-yes protein is myristylated at the Gly-2 in mature form.

The newly identified sequence of 12 amino acids at the carboxyl terminus of the peptide is very similar to those of the human and chicken c-src proteins (1) and identical to that

of the human syn protein (33). The sequence contains a tyrosine residue (Tyr-537) corresponding to Tyr-527 in $pp60^{c-src}$, which is suggested to be phosphorylated for negative regulation of its kinase activity (7).

Comparison with the v-yes gene. Comparison of the cloned cDNA sequence with that of the v-yes gene revealed that the entire coding sequence of the c-yes-1 gene is included in the v-ves gene except for the sequence encoding the carboxyl terminus of the protein. Nucleotide sequence homology between v-yes and human c-yes-1 was clearly observed downstream of the second letter in the codon of the initiating methionine (Fig. 2). It appears probable that the chicken c-yes gene product is initiated from the same site with human c-yes-1, and methionine at the initiation codon seems to be replaced by valine in the v-yes gene. In this case, codon GTG for valine could be the result of a transition event from adenine to guanine nucleotide in the viral genome. Three additional stretches of homology, A, B, and C, are observed between the 5' noncoding region of human c-yes-1 cDNA and the corresponding part of the v-yes gene (Fig. 2). First, there is a homologous sequence A of eight nucleotides a little upstream of the initiating methionine codon. Interestingly, a sequence similar to the left seven nucleotides is located at the 3' border of the first exon of the chicken c-src gene (39). Since the exon-intron architectures of the c-src and c-yes-1 genes are identical so far as examined (Semba et al., unpublished data), this seven-nucleotide sequence of c-yes-1 cDNA is likely to be the exact 3' border of the first exon of the human c-yes-1 gene. Just upstream from this site, the sequences of v-yes and human c-yes are both G+C rich, and the homologous region B of the two is a G+C stretch of 10 identical nucleotides in this G+C-rich region. There is a third homologous sequence, C, of 17 nucleotides in the region corresponding to the 5' recombination site of the chicken c-yes and viral sequences, which is discussed below. Although these homologous sequences may suggest that the v-yes in this part represents the 5' noncoding sequence of chicken c-yes mRNA, we cannot tell decisively at present whether it includes some intron structures.

To our surprise, in spite of the partial conservation of the nucleotide sequence (55% homology) from position 261 to 302, the deduced amino acid sequence (residues 34 to 47) of human c-yes-1 is completely different from the corresponding sequence of v-yes (Fig. 2 and 3). To avoid possible mistakes in sequencing, we confirmed the nucleotide sequence of this region by sequencing two independent cDNA clones (λ YS-12 and λ YS-17) and a v-yes cloned DNA.

As mentioned above, nucleotide sequence analysis disclosed an array of 12 amino acids in the carboxyl terminus of the c-yes-1 protein that is very similar to those of other tyrosine kinases. Our preliminary data on the chicken c-yes cDNA sequence showed that the chicken c-yes protein had the same carboxy-terminal 12 amino acids as those of the human c-yes-1 protein, except that Glu-541 in human c-yes-1 is replaced by aspartic acid in chicken c-yes (Sukegawa et al., unpublished data). In the v-yes gene, the extreme 3' terminus of the chicken c-yes coding sequence that specifies these 12 amino acids is replaced by the viral sequence for the env protein gp37 in a frame that differs from that of the original protein. At the site of this replacement, both the nucleotide sequences of the chicken and human c-yes and that of the Prague strain of Rous sarcoma virus (Pr-RSV) share a homologous sequence of 10 nucleotides, which is discussed below. Therefore, we conclude that by this replacement the carboxy-terminal eight amino acids in the c-yes protein are replaced by three different amino acids in

	3CGGAGCCAAGGCACACGGGTCTGACCCTTGGGCCGGAGC
1	
151	1 10 20 30 40 MetGlyCysIleLysSerLysGluAsnLysSerProAlaIleLysTyrArgProGluAsnThrProGluProValSerThrSerValSerHisTyrGlyAlaGluProThrThrValSerProCysProSerSerSerSerSerSerSerSerSerSerSerSerSerS
301	50 60 70 80 90 AlaLysGlyThrAlaValAsnPheSerSerLeuSerMetThrProPheGlyGlySerSerGlyValThrProPheGlyGlyAlaSerSerSerPheSerValValProSerSerTyrProAlaGlyLeuThrGlyGlyValThrIlePhe 90 GCAAAGGGAACAGCAGTTAATTTCAGCAGTCTTICCATGACACCACTTTGGAGGATCCTCAGGGGTAACGCCTTTTGGAGGTGCCACTTTCCACTGTGTGCCAAGTTCATATCCTGCTGGTTTTAACAGGTGGTGGTTAACTATATTT 90 AT T A T T GC AA G C C A G T CT C AG AC G G G
451	100 110 120 130 140 ValAlaLeuTyrAspTyrGluAlaArgThrThrGluAspLeuSerPheLysLysGlyGluArgPheGinIleIleAsnAsnThrGluGlyAspTrpTrpGluAlaArgSerIleAlaThrGlyLysAsnGlyTyrIleProSerAsnTyr GTGGCCTTATATGATTATGAAGCTAGAACTACAGAAGAACCTTTCATTTAAGAAGGGGGAAAGAATTCAAATACGAAAGAATGGGTGGG
601	150 160 170 180 190 ValAlaProAlaAspSerIleGInAlaGluGluTrpTyrPheGlyLysMetGlyArgLysAspAlaGluArgLeuLeuLeuAsnProGlyAsnGlnArgGlyIlePheLeuValArgGluSerGluThrThrLysGlyAlaTyrSerLeu GTAGCGCCTGCAGAATTCCATTCAGGCAGAAATGGTATTTTGGCAAAATGGGGGGGAAAAGATGCTGAAAGATTACTTTTGAAATCATCGGGAGAATCAACGAGGTATTTCTTAGTAAGAGAGGGGGTAAACAACGATGGTCTATTCCTT T C A G C T C C C C
751	200 210 220 230 240 SerIleArgAspTrpAspGluIleArgGlyAspAsnValLysHisTyrLysIleArgLysLeuAspAsnGlyGlyTyrTyrIleThrThrTArgAlaGlnPheAspThrLeuGlnLysLeuValLysHisTyrThrGluHisAlaAspGly 210 240 ICTATIGGTGATTGGGATGAGATAAGGGGTGACAATGGTGAAACCTACAATGGTGGAAACTTGGTGAAACCATGCTGATGGTGAAACCATGCTGATGGTGAAACCATGCTGATGGTGAAACCATGCTGATGGTGAAACCATGCTGATGGTGAAACCATGCTGATGGTGAAACCATGCTGACAATGGTGGAACCATGCTGATGGTGGATACCAATGGTGGATACCATGCTGATGGTGAAACCATGCTGATGGTGAAACCATGCTGATGGTGAAACCATGCTGATGGTGAAACCATGCTGATGGTGAAACCATGCTGATGGTGAAACCATGCTGATGGTGAAACCATGCTGATGGTGAAACCATGCTGATGGTGAAACCATGCTGATGGTGAAACCATGCTGATGGTGAAACCATGCTGATGGTGAACCATGCTGATGGTGGATGCAATGGTGGATGCTGATGGTGAAACCATGCTGATGGTGAAACCATGCTGATGGTGAAACTGGTGAAGCATGGTGGATGCTGATGGTGAACCATGCTGATGGTGGATGCTGATGGTGGATGCTGATGGTGGATGCTGATGGTGGATGCTGATGGTGGATGCTGATGGTGGATGCTGATGGTGGATGCTGATGGTGGATGCTGATGGTGGATGCTGATGGTGGATGCTGATGGTGGATGCTGATGGTGGATGCTGATGGTGGATGCTGATGGTGGATGCTGATGGTGGATGCTGATGGTGGATGCTGATGGTGGATGGTGGATGGTGGATGGTGGATGGTGGATGGTGG
901	250 260 270 280 290 LeuCysHisLysLeuThrThrValCysProThrValLysProGlnThrGlnGlyLeuAlaLysAspAlaTrpGluIleProArgGluSerLeuArgLeuGluValLysLeuGlyGlnGlyCysPheGlyGluValTrpMetGlyThrTrp TTATGCCACAAGTTGACAACTGTGGTGTCCAACTGTGAAACTCAAGGTCTAAGGAAAATGCTTGGGAAATCCTCTGCGACTGAGGATGATGGTGGAAGGTGGGAAGTGGTGGGAACATGG C G T T C A A C G A A A G A C T A G G A G G C T T A C
1051	300 310 320 330 340 AsnGlyThrThrLysValAlaIleLysThrLeuLysProGlyThrMetMetProGluAlaPheLeuGInGluAlaGinIleMetLysLysLeuArgHisAspLysLeuValProLeuTyrAlaValValSerGluGluProIleTyrIle AATGGAACCACGAAAGTAGCAATCAAAAACCAGGTACAATGATGAGTGCCAGGAGAAGCTTCCTCTAGAGGAGCCAGTTAATGAAAAAATTAAGACATGATAAACCTGTTCCCACTATATGGTGTTTCTGAAGAACCAATTTAACATC A C G T TI C G G C A
1201	350 360 370 380 390 ValThrGluPheMetSerLysGlySerLeuLeuAspPheleuLysGluGlyAspGlyLysTyrLeuLysLeuProGlnLeuValAspMetAlaAlaGlnIleAlaAspGlyMetAlaTyrIleGluArgMetAsnTyrIleHisArgAsp GTCACTGAATTIATGTCAAAAGGAAGGTTATTAGATTICCTTAAGGAAGGAAGGAAGGTATTGCACGCTGGTTGATATGGCTGCTCAGATTGCTGATGGCTATTATTGAAAGAAA
1351	400 410 420 430 440 LeuArgAlaAlaAssIleLeuValGlyGluAssLeuValCysLysIleAlaAspPheGlyLeuAlaArgLeuIleGluAspAssGluTyrThrAlaArgGlnGlyAlaLysPheProIleLysTrpThrAlaProGluAlaAlaLeuTyr CTTCGGGCTGCTAATATTCTTGTAGGAGAAAATCTTGGTGCAAAATAGCAGGACTTGGTTTAGCAAGGTTAATTGAAGACAATGAATACACAGCAAGGCCAAGGTGCAAAATTTCCAATGAAGCAGCTCCTGAAGCTGCACGCTGCTAAT C A C C C C A G T T A T
1501	450 460 470 480 490 GlyArgPheThrIleLysSerAspValTrpSerPheGlyIleLeuGlnThrGluLeuValThrLysGlyArgValProTyrProGlyMetValAsnArgGluValLeuGluGlnValGluArgGlyTyrArgMetProCysProGlnGly GGTCGGTTTACAATAAAGTCTGATGTCTGGTATTGGAATTCTGGAACTAGGATGACCAGGGCCCAAGGGCCCATATCCAGGGTGCAGTGGCAGTGACCGTGAAGTACTAGGATGACGAGGATGACGAGGCCGGGGCCCAGGGCCCAGGGCCCAGGGCCCATATCCAGGATGCCGTGAACCAGGATGCCGGGAGGATACAGGATGCCGGGCCGGGCCCAGGGCCCGGGGCCCGAGGGCCCGAGGGCCCGAGGCCCGGGCCCGAGGCCCGGGCCCGAGGCCCGAGGCCCGGGCCCGAGGCCCGGGCCCGAGGCCCGGGCCCGAGGCCCGAGGCCCGAGGCCGGCGCCGAGGCCGAGGCCGGGCCGAGGCCGGCGCGGCCGAGGCCGGCGCGGCCGGGCCGAGGCCGGGCCGAGGCCGGGCGCGGGCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGCGGGCGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGG
1651	500 510 520 530 540 543 CysProGluSerLeuHisGluLeuHetAsnLeuCysTrpLysLysAspProAspGluArgProThrPheGluTyrIleGInSerPheLeuGluAspTyrPheThrAlaThrGluProGInTyrGInProGlyGluAsnLeuEnd TGTCCAGAATCCCTCCATGAATCTGTGGATGGAAGAAGGACCCTGATGAAAGACCAACATTTGAATATTCAGTCCTICTTGGAAGACTACTTCAAGG <u>CCC</u> ACGAGGCCACGAGGAAAAATTTATAATTCAAG C G T C G A A A G G A C G C G C C G C C G C C C C
1801 1951 2101 2251 2401 2551 2701 2851 3001 3151 3301 3451 3601 4051 4201 4351	AGCCTATTITTATATGCACAAAATCTGCCAAAATATAAAAGAACTTGTGTGAGATTITCTACAGGAATCAAAAGAAGAAAATCTTCTTTACTCTGCATGTTTTAATGGCAAAACGGAATCCAAGGATGGTAAACGAAAACCACTTTTTTTT

FIG. 2. Complete sequence of the human c-yes-1 cDNA and corresponding portion of the v-yes gene. The sequences of the human c-yes-1 cDNA and the v-yes gene are aligned in accordance with the sequence homology between them. Nucleotides of the v-yes coding sequence are shown only where they are different from those of the cDNA. The sequence of the v-yes are shown a little longer than that of the 5' end of the cDNA to exhibit the additional sequence homology 'C' between the two (the single line of sequence at the top of the figure). Nucleotides of the cDNA are numbered on the left, and deduced amino acids are also numbered from the initiating methionine. Three homologous sequences between the 5' noncoding region of human c-yes-1 gene and the corresponding part of v-yes and the viral sequence described previously and the site corresponding to the 5' recombination in the human c-yes is also indicated by an arrow at the box C (19). A 10-nucleotide stretch of homology between human c-yes-1 and v-yes at the 3' recombination site is underlined (see the text). The sequence AATAAA in the 3' non-coding region is underlined.

c-yes	1	MGC1KSKENKSPAIKYRPENT-PEPVSTSVSHYGAEPTTVSPCPSSSAKGTAVNFSSLSMTPFGGSSGVTPFGGASSSFS
v-yes	277	VGCIKSKEDKORPAMKYRTDNT-PEPIISSHVSHYGSDSSQATQSPAIKGSAVNFNSHSMTPFGGPSGMTPFGGASSSFS
syn	1	MGCVQCK-DKEATKLTEERDGSL-NOSS-GYRYCTDPT-POHYP3FCVTSIPNYNNFHAAGC-OCLIMFGCVNSS-S
c-src	1	MGSSKS-K-P-KDPSQRRRSLERPDSTH-HGGFPASQTPNKTAAPDTHRTP-SRSFGTVATEPKLFGGFNTSDT
c-yes	80	VVPSSYPAG-LTGGVTIFVALYDYEARTTEDLSFKKGERFQIINNTEGDWWEARSIATGKNGYIPSNYVAPADSIQAEEW
v-yes	354	AVPSPYPST-LTGGTTMFVALYDYEARTTDDLSFKKGERFOIINNTEGDWWEARSIATGKTGYIPSNYVAPADSIQAEEW
syn	72	HTGTLRTRG-GT-GVTLLFVALYDYEARTEDDLSFHKGEKFQILDNSSEGDWWEARSLTTGETGYIPSNYVAPVDSIQAEEW
c-src	69	-WISPORAGALAGOVITEVALYDYESRTETDLSFKKGERLOIVNNTEGDWWLAHSLTTOOTGYIPSNYVARSDSIOAEEW
c-fgr		EGDWWEARSLSGCHTGCIPSNYVAPVDSIOAEEW
c-yes	159	YFGKMGRKDAERLLLNPGNQRGIFLVRESETTKGAYSLSIRDWDE <u>I</u> RGDNVKHYKIRKLDNGGYYITTRAQF <u>DT</u> LQKLVK
v-yes	433	YFGK <u>M</u> GRKDAERLLL <u>NP</u> GN <u>O</u> RG <u>I</u> FL <u>V</u> RESETTKGAYSLSIRDWD <u>H</u> VRGD <u>N</u> VKHYKIRKLDNGGYYITTRAQFE <u>S</u> LQ <u>K</u> LV <u>K</u>
syn	150	ŸFGK <mark>UGRKDAERQLUSFGNPROTFUI</mark> RESETTKGAYSLSIRDWDDMRGDHVKHYKIRKLDNGGYYIT <u>TRAOFETLOQLVO</u>
c-src	149	ŸFGKI <u>TRRES</u> ERLLL <u>NPENPR</u> OTFLVRESETTKGAYOLSVSDFDNAKCUNVKHYKIRKLOSCOFYITSBRTQFSSLOOLVA
c-fgr	35	<u>YFGKIGRKDAERQLLSPGNPQQAFUIRESETTKGAYSLSIRDWDQTRGDHVKHYKIRKLDMGGYYITTRVQF</u> NSVQBLVQ
		* * * **
c-yes	239	HYTEHADGLCHKLTTVCPTV-KPQTQGLAKDAWEIPRESLRLEVKLGQGCFGEVWMGTWNGTTKVAIKTLKPGTMMPE
v-yes	513	HYREHADGLCHKLTTVCPTV-KPQTQGLAKDAWEIPRESLRLEVKLGQGCFGEVWMGTWNGTTKVAIKTLKUGTMMPE
syn	230	<u>HYSERAAGLOCRLVVPCHKGMPRUTDLSVKT</u> KDVWEIPRESLQLIKRLQNGQFGEVWMGTWNQNT <u>K</u> VAIKTLKPG <u>T</u> MSPE
c-src	229	<u>Y</u> YS <u>KHA</u> DGLCHRLTNVCPTS-KPQT- <u>QGLAKDAWEIPRESLR</u> LE <u>VK</u> LGQGCFGEVWMGTWNGTTRVAIKTLKPGMMSPE
c-fgr	115	<u>HYMEVNDGLO</u> NUUIAPO-TIMKPQT-L-GLAKDAWEISBESITLERRLOTGCFODVWLGTWNOSTKVAVKTLKPGTMSEK
c-yes	316	AFLQEAQIMKKLRHDKLVPLYAVVSEEPIYIVTEFMSKGSLLDFLKEGDGKYLKLPQLVDMAAQIADGMAYIERMNYIHR
v-yes	590	AFLQEAQIMKKLRHDKLVPLYAVVSEEPIYIVTEFMIKGSLLDFLKEGBGKFLKLPQLVDMAAQIADGMAYIERMNYIHR
syn	310	SFLBEADIMKKLKHDKLVQLYAVVSEEPIYIVTEYMNKGSLLDFLADGEGRALKLPNLVDMAAQMAAGMAYIERMNYIHR
c-src	306	AFLQEAQYMKKLRHEKLVQLYAVVSEEPIYIVTEYMSKGSLLDFLKGEMGKYLRLPQLVDMAAQIASGMAYVERMNYMHR
c-fgr	192	<u>AFUBEAGVMBULRHDKLVQLYAVVSEEPIYIVTEFM</u> CHGSLLDFLBNPBGQDURLPQLVDMAAGVABGMAYMERMNYIHR
		\checkmark
	206	DI DA NULL VCENI VCELA DECIA DE LEDNEVINA DOCA VEDI VUINA DEA AL VCDEM L'UCDUDUCECTI AMEL VIII VCDUDUDO
c-yes	390	DL RAAN I LVGENLVCKI AD FGLARLIED NEVI TAROGAKEPI KWTAPEAALIGRETIKSDVWSFGI LOTELVTKGRVPIPG
v-yes	200	DE RANTE VOLGE VOLGE LA DE LE DE LE DE LE TARGARE PER VITAPERALE GRE TERSOVOS EGENTE UN KORVETEG
syn c-syn	200	DE DANIEU GENIUCKIADE GLARLIEDNEI TARGAKEPIKWAADEAALIGKEI IKSDWSEGILLIELVIKGKVPIPG
c-far	272	DIDAANILUGENLUCNYADEGLARDIEUNEINANGAREEINWAADEANIGKEINSDUWSEGLIHUUUUNKAVEIPG
C-LAL	212	PREMITER OF COMPANY CONTRACTOR OF COMPANY CONTRACTOR OF CONT
c-ves	476	MVNREVLEOVERGYRMPCPOGCPESLHELMNLCWKKDPDERPTFEYIOSFLEDYFTATEPOYOPGENL
v-yes	750	MVNREVLEOVERGYRMPCPOGCPESLHELMKLCWKKDPDERPTFEYIOSFLEDYFTAAEPSGY
syn	470	MINREVLEOVERGYRMPCPODCPISLHELMIHCWKKDPERENTFEYDOSFLEDYFTATEPOYOPGENL
c-src	466	MVNREVLDOVERGYRMPCPPECPESLHDLMCOCWRRDPBERPTFEYLOAFLEDYFTSTEPOYOPGENL
- far	252	MUZD BY LEONED VIEW DC DDCC DDCC DDCC DDCC DDCC DDCC DDCC

c-fgr 352 <u>M</u>NK<u>REVLEQVEQCGYHMPCPPGCPASL</u>YEBAMEQTWARLDPEERPTFEYLQSFLEDYFTSALEPOYOPCDQT

FIG. 3. Comparison of the amino acid sequences of human c-yes-1, v-yes (19), syn (33), chicken c-src (39), and human c-fgr (26, 28). Amino acids identical to those of human c-yes-1 are boxed. Amino acids of each protein are numbered on the left. The amino-terminal sequence of the human c-fgr protein is not available. Asterisks indicate the conserved amino acids supposed for nucleotide binding. The arrowhead indicates the target residue for the autophosphorylation reaction.

the v-yes protein, and that Thr-533 in the c-yes protein is replaced by alanine in the v-yes protein (Fig. 3 and 4).

Besides the alterations described above, there are many other amino acid differences between the human c-yes-1 protein and the corresponding portions of the chickenderived v-yes protein. Our limited degree of data on chicken c-yes cDNA so far available have demonstrated that the amino acid differences in the part downward from His-513 in the v-yes protein, other than at Lys-583 and at the C terminus, are due to species differences and therefore irrelevant to the transforming ability of the v-yes gene. The Pro-312 in the human c-yes-1 protein, which corresponds to Lys-583 of v-yes, is conserved in the chicken c-yes protein (Sukegawa et al., unpublished data). Complete nucleotide sequencing of chicken c-yes is necessary to determine whether the rest of the amino acid differences are important for the transforming ability of v-yes.

Sequences at recombination junctions in the v-yes gene. Avian sarcoma virus Y73 has incorporated the chicken c-yes gene into its genome by at least two successive events of recombination between the c-yes gene and viral sequences. The 5' recombination has taken place in the coding sequence for the gag precursor protein, and the 3' recombination has occurred in the viral env sequence. Both recombination sites on the virus genome were suggested in our previous study from a comparison of the v-yes sequences with those of Pr-RSV (19, 31). Comparison of the nucleotide sequence of human c-yes-1 with that of Pr-RSV revealed that the two genes shared short stretches of nucleotide sequences in regions corresponding to the 5' and 3' recombination sites (Fig. 2 and 4). In the region corresponding to the 5' recombination junction, which seems to have occurred in the 5' untranslated region of the c-yes gene, the human c-yes-1 and v-yes sequences have a homologous sequence of 17 nucleotides (15 of 17 nucleotides are identical). The gag sequence of Pr-RSV also shows homology with the 5' half of this 17-nucleotide stretch (8 of 9 nucleotides are identical). This stretch of homology extends six nucleotides further upstream from the point at which the 5' recombination is thought to have occurred. Downstream from the point of the tentative 3' recombination junction there is a 10-nucleotide stretch which shows partial homology in human c-yes-1 and Pr-RSV (8 of 10 nucleotides are identical). The 10-nucleotide sequence of Y73 in this region completely matches that of Pr-RSV, which is a part of the env gene specifying gp37. Preliminary sequencing data on chicken c-yes cDNA showed that 9 of 10 nucleotides in this region of the chicken sequence are identical with those of Pr-RSV (Fig. 4).

DISCUSSION

Cellular protein-tyrosine kinases are classified into two groups. One group includes receptor molecules for polypeptide growth factors. To date, five receptor molecules of this type are known: the receptors for insulin (17, 40), somatomedin C (29), epidermal growth factor (5), platelet-derived growth factor (10), and colony-stimulating factor 1 (35). In addition, the trk (oncD) (23), erbB2 (2, 45) and c-ros (25) genes are suggested to encode receptor-type proteins. In contrast, $pp60^{c-src}$ and $pp56^{lsk-tck}$ (22, 43) have no structures characteristic of transmembrane receptor molecules; they have no stretch of hydrophobic amino acids for spanning the membrane and no extracellular domain for ligand binding. Both molecules are acylated by myristic acid and anchored to the inner surface of the plasma membrane (12, 21, 42). Although they are not receptor molecules, these nonreceptor-type kinases, constituting the second group of cellular protein-tyrosine kinases, are believed to be essential as effectors or modulators of some signal transduction processes occurring at the plasma membrane. The predicted products of the syn and lyn genes are both classified in this group. We have isolated and characterized cDNA clones for the human c-yes-1 gene and demonstrated that the product of the c-yes-1 gene also belongs to this nonreceptor type of protein-tyrosine kinase family.

The abilities of many viral oncogenes to transform cells have been ascribed to their protein-tyrosine kinase activities. These viral tyrosine kinase genes are activated versions of normal cellular counterparts incorporated into the virus genome. The transduction of the cellular gene into a viral genome inevitably causes structural alterations of the cellular gene product. Subsequent cycles of virus replication may well allow additional mutations in the coding sequence of the incorporated cellular gene. These alterations are considered to account for the acquisition of transforming ability of the viral oncogenes. The nucleotide sequence of the human c-yes-1 cDNA revealed major alterations that the chicken c-yes gene had undergone during the processes of its incorporation into the viral genome. The carboxy-terminal sequence of eight amino acids of the c-yes-1 protein, which includes the tyrosine residue (Tyr-537) that is thought to be involved in negative regulation of kinase activities (7), is replaced by different amino acids in the v-yes protein. The 5' recombination event resulted in joining of the viral gag sequence and the suspected 5' noncoding exon or intron sequence of the chicken c-yes gene. Similar recombinations are suggested in the cases of v-fps and v-myc, in which 5' untranslated or intron sequences are used to generate the viral oncogenes (14, 27). The extraneous portion of the fused



FIG. 4. Comparison of the nucleotide sequences of c-yes, Y73, and Pr-RSV around the recombination junctions. Nucleotides identical to those of Y73 are boxed. The two arrows indicate the sites of 5' and 3' recombination junctions between chicken c-yes and the viral sequence described previously (19).

protein might alter the biological specificities of the authentic cellular gene product, such as its substrate specificity and subcellular localization. Besides these changes, there are many differences between the amino acid sequences of the human c-yes-1 and v-yes proteins. Some of these alterations could be responsible for the transforming ability of the v-yes protein, but the exact mode of activation of chicken c-yes cannot be defined until the entire structure of chicken c-yes is determined.

The mechanism by which the cellular sequence is incorporated into the viral genome is not fully understood, although several models have been suggested (38, 41). Comparison of the nucleotide sequence of human c-yes-1 cDNA with that of Pr-RSV revealed remarkable homologies between the two at both of the recombination sites postulated for the v-yes gene. Van Beveren et al. reported that the sequence homologies between the incorporated cellular gene and the viral genome were also observed in the case of fos in Finkel-Biskis-Jinkins strain of murine sarcoma virus (41). They proposed three possible modes of incorporation of the cellular sequence into the viral genome: (i) double crossover between viral and cellular DNAs, (ii) 5' recombination at the DNA level and 3' recombination at the RNA level during reverse transcription, and (iii) both the 5' and 3' recombinations at the RNA level during reverse transcription. The second mechanism, which appears to be widely applicable to many other cases of retroviral transduction of the cellular sequence, could be facilitated by the sequence homologies described here, although the homologies cannot exclude either of the other mechanisms. We have confirmed that the chicken c-yes sequence retains the homology at the 3' recombination site. We are now trying to isolate a full-length chicken c-yes cDNA clone to examine whether the 5' sequence homology described here is also conserved in the chicken genome.

ACKNOWLEDGMENTS

We thank S. Nishimura for kindly supplying deoxy-7deazaguanosine triphosphate and M. Ide for typing the manuscript.

This work was supported in part by a Grant-in-Aid for Special Project Research, Cancer-Bioscience from the Ministry of Education, Science and Culture of Japan.

LITERATURE CITED

- Anderson, S. K., C. P. Gibbs, A. Tanaka, H.-J. Kung, and D. J. Fujita. 1985. Human cellular *src* gene: nucleotide sequence and derived amino acid sequence of the region coding for the carboxy-terminal two-thirds of pp60^{c-src}. Mol. Cell. Biol. 5:1122–1129.
- 2. Bargmann, C. I., M.-C. Hung, and R. A. Weinberg. 1986. The *neu* oncogene encodes an epidermal growth factor receptor-related protein. Nature (London) **319**:226–230.
- Brugge, J. S., P. C. Cotton, A. E. Queral, J. N. Barrett, D. Nonner, and R. W. Keane. 1985. Neurones express high levels of a structurally modified, activated form of pp60^{c-src}. Nature (London) 316:554–559.
- 4. Buss, J. E., and B. M. Sefton. 1985. Myristic acid, a rare fatty acid, is the lipid attached to the transforming protein of Rous sarcoma virus and its cellular homolog. J. Virol. 53:7-12.
- Cohen, S., H. Ushiro, C. Stoscheck, and M. Chinkers. 1982. A native 170,000 epidermal growth factor receptor-kinase complex from shed plasma membrane vesicles. J. Biol. Chem. 257: 1523–1531.
- 6. Collett, M. S., E. Erikson, and R. L. Erikson. 1979. Structural analysis of the avian sarcoma virus transforming protein: sites of phosphorylation. J. Virol. 29:770–781.
- Cooper, J. A., K. L. Gould, C. A. Cartwright, and T. Hunter. 1986. Tyr⁵²⁷ is phosphorylated in pp60^{e-src}: implications for regulation. Science 231:1431–1434.
- Cotton, P. C., and J. S. Brugge. 1983. Neural tissues express high levels of cellular *src* gene product pp60^{c-src}. Mol. Cell. Biol. 3:1157–1162.
- Downward, J., Y. Yarden, E. Mayes, G. Scrace, N. Totty, P. Stockwell, A. Ullrich, J. Schlessinger, and M. D. Waterfield. 1984. Close similarity of epidermal growth factor receptor and v-erb-B oncogene protein sequences. Nature (London) 307:521-527.
- Ek, B., B. Westermark, A. Wasteson, and C.-H. Heldin. 1982. Stimulation of tyrosine-specific phosphorylation by plateletderived growth factor. Nature 295:419–420.
- Fults, D. W., A. C. Towle, J. M. Lauder, and P. F. Maness. 1985. pp60^{c-src} in the developing cerebellum. Mol. Cell. Biol. 5:27–32.
- Garber, E. A., F. R. Cross, and H. Hanafusa. 1985. Processing of p60^{v-src} to its myristylated membrane-bound form. Mol. Cell. Biol. 5:2781-2788.
- Gould, K. L., J. R. Woodgett, J. A. Cooper, J. E. Buss, D. Shalloway, and T. Hunter. 1985. Protein kinase C phosphorylates pp60^{src} at a novel site. Cell 42:849–857.
- Huang, C.-C., C. Hammond, and J. M. Bishop. 1985. Nucleotide sequence and topography of chicken c-fps; genesis of a retroviral oncogene encoding a tyrosine-specific protein kinase. J. Mol. Biol. 181:175-186.
- Iba, H., F. R. Cross, E. A. Garber, and H. Hanafusa. 1985. Low level of cellular protein phosphorylation by nontransforming overproduced p60^{c-src}. Mol. Cell. Biol. 5:1058–1066.
- Kamps, M. P., S. S. Taylor, and B. M. Sefton. 1984. Direct evidence that oncogenic tyrosine kinases and cyclic AMPdependent protein kinase have homologous ATP-binding sites. Nature (London) 310:589–592.
- Kasuga, M., Y. Zick, D. L. Blith, F. A. Karlsson, H. U. Häring, and C. R. Kahn. 1982. Insulin stimulation of phosphorylation of the β subunit of the insulin receptor. J. Biol. Chem. 257:9891-9894.
- Kawai, S., M. Yoshida, K. Segawa, H. Sugiyama, R. Ishizaki, and K. Toyoshima. 1980. Characterization of Y73, an avian sarcoma virus: a unique transforming gene and its product, a phosphopolyprotein with protein kinase activity. Proc. Natl. Acad. Sci. USA 77:6199–6203.
- 19. Kitamura, N., A. Kitamura, K. Toyoshima, Y. Hirayama, and M. Yoshida. Avian sarcoma virus Y73 genome sequence and structural similarity of its transforming gene product to that of Rous sarcoma virus. Nature (London) 297:205-208.
- Levy, B. T., L. K. Sorge, A. Meymandi, and P. F. Maness. 1984. pp60^{c-src} kinase in embryonic tissues of chick and human. Dev. Biol. 104:9–17.
- 21. Marchildon, G. A., J. E. Casnellie, K. A. Walsh, and E. G.

Krebs. 1984. Covalently bound myristate in a lymphoma tyrosine protein kinase. Proc. Natl. Acad. Sci. USA 81:7679–7682.

- 22. Marth, J. D., R. Peet, E. G. Krebs, and R. M. Perlmutter. 1985. A lymphocyte-specific protein-tyrosine kinase gene is rearranged and overexpressed in the murine T cell lymphoma LSTRA. Cell 43:393–404.
- 23. Martin-Zanca, D., S. H. Hughes, and M. Barbacid. 1986. A human oncogene formed by the fusion of truncated tropomyosin and protein tyrosine kinase sequences. Nature (London) 319:743-748.
- 24. Mizusawa, S., S. Nishimura, and F. Seela. 1986. Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. Nucleic Acids Res. 14:1319–1324.
- 25. Neckameyer, W. S., and L.-H. Wang. 1985. Nucleotide sequence of avian sarcoma virus UR2 and comparison of its transforming gene with other members of the tyrosine protein kinase oncogene family. J. Virol. 53:879–884.
- Nishizawa, M., K. Semba, M. C. Yoshida, T. Yamamoto, M. Sasaki, and K. Toyoshima. 1986. Structure, expression, and chromosomal location of the human c-fgr gene. Mol. Cell. Biol. 6:511-517.
- Papas, T. S., N. K. Kan, D. K. Watson, C. S. Flordellis, M. C. Psallidopoulos, J. Lautenberger, P. Duesberg, and K. P. Samuel. 1984. myc-related genes in viruses and cells. Cancer Cells 2:153–163.
- Parker, R. C., G. Mardon, R. V. Lebo, H. E. Varmus, and J. M. Bishop. 1985. Isolation of duplicated human c-src genes located on chromosome 1 and 20. Mol. Cell. Biol. 5:831-838.
- Rubin, J. B., M. A. Shia, and P. F. Pilch. 1983. Stimulation of tyrosine-specific phosphorylation *in vitro* by insulin-like growth factor I. Nature (London) 305:438-440.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.
- 31. Schwarz, D. E., R. Tizard, and W. Gilbert. 1983. Nucleotide sequence of Rous sarcoma virus. Cell 32:853-869.
- Seki, T., G. Fujii, S. Mori, N. Tamaoki, and M. Shibuya. 1985. Amplification of c-yes-1 protooncogene in a primary human gastric cancer. Jpn. J. Cancer Res. (Gann) 76:907–910.
- 33. Semba, K., M. Nishizawa, N. Miyajima, M. C. Yoshida, J. Sukegawa, Y. Yamanashi, M. Sasaki, T. Yamamoto, and K. Toyoshima. 1986. *yes*-related protooncogene, *syn*, belongs to the protein-tyrosine kinase family. Proc. Natl. Acad. Sci. USA 83:5459-5463.
- 34. Semba, K., Y. Yamanashi, M. Nishizawa, J. Sukegawa, M. Yoshida, M. Sasaki, T. Yamamoto, and Kumao Toyoshima. 1985. Location of the c-yes gene on the human chromosome and its expression in various tissues. Science 227:1038–1040.
- 35. Sherr, C. J., C. W. Rettenmier, R. Sacca, M. F. Roussel, A. T. Look, and E. R. Stanley. 1985. The c-fms proto-oncogene product is related to the receptor for the mononuclear phagocyte growth factor, CSF-1. Cell 41:665–676.
- 36. Smart, J. E., H. Oppermann, A. P. Czernilofsky, A. F. Purchio, R. L. Erikson, and J. M. Bishop. 1981. Characterization of sites for tyrosine phosphorylation in the transforming protein of Rous sarcoma virus (pp60^{v-src}) and its normal cellular homologue (pp60^{e-src}). Proc. Natl. Acad. Sci. USA 78:6013–6017.
- Sorge, L. K., B. T. Leby, and P. F. Maness. 1984. pp60^{c-src} is developmentally regulated in the neural retina. Cell 36:249-257.
- Swanstrom, R., R. C. Parker, H. E. Varmus, and J. M. Bishop. 1983. Transduction of a cellular oncogene: the genesis of Rous sarcoma virus. Proc. Natl. Acad. Sci. USA 80:2519–2523.
- Takeya, T., and H. Hanafusa. 1983. Structure and sequence of the cellular gene homologous to the RSV src gene and the mechanism for generating the transforming virus. Cell 32:881-890.
- 40. Ullrich, A., J. R. Bell, E. Y. Chen, R. Herrera, L. M. Petruzzelli, T. J. Dull, A. Gray, L. Coussens, Y.-C. Liao, M. Tsubokawa, A. Mason, P. H. Seeburg, C. Grunfeld, O. M. Rosen, and J. Ramachandran. 1985. Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes. Nature (London) 313:756-761.

- 41. Van Beveren, C., F. V. Straaten, T. Curran, R. Müller, and I. M. Verma. 1983. Analysis of FBJ-MuSV provirus and c-fos (mouse) gene reveals that viral and cellular fos gene products have different carboxy termini. Cell 32:1241-1255.
- 42. Voronova, A. F., J. E. Buss, T. Patschinsky, T. Hunter, and B. M. Sefton. 1984. Characterization of the protein apparently responsible for the elevated tyrosine protein kinase activity in LSTRA cells. Mol. Cell. Biol. 4:2705–2713.
- 43. Voronova, A. F., and B. M. Sefton. 1986. Expression of a new tyrosine protein kinase is stimulated by retrovirus promoter insertion. Nature (London) 319:682-685.
- 44. Wierenga, R. K., and W. G. J. Hol. 1983. Predicted nucleotidebinding properties of p21 protein and its cancer-associated variant. Nature (London) **302**:842–844.
- 45. Yamamoto, T., S. Ikawa, T. Akiyama, K. Semba, N. Nomura, N. Miyajima, T. Saito, and K. Toyoshima. 1986. Similarity of protein encoded by the human c-*erb*-B-2 gene to epidermal growth factor receptor. Nature (London) **319**:230–234.
- 45a. Yamanashi, Y., S. Fukushige, K. Semba, J. Sukegawa, N. Miyajima, K. Matsubara, T. Yamamoto, and K. Toyoshima. 1987. The yes-related cellular gene lyn encodes a possible tyrosine kinase similar to p56^{lck}. Mol. Cell. Biol. 7:237-243.
- 46. Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119.
- 47. Yoshida, M., S. Kawai, and K. Toyoshima. 1980. Uninfected avian cells contain structurally unrelated progenitors of viral sarcoma genes. Nature (London) 287:653-654.