MOLECULAR AND CELLULAR BIOLOGY, May 1985, p. 1122–1129
0270-7306/85/051122-08$02.00/0

Vol. 5, No. 5

Human Cellular *src* Gene: Nucleotide Sequence and Derived Amino Acid Sequence of the Region Coding for the Carboxy-Terminal Two-Thirds of pp60$^{c\text{-}src}$

STEPHEN K. ANDERSON,[1] CAROL P. GIBBS,[2]† AKIO TANAKA,[1] HSING-JIEN KUNG,[2]‡ AND DONALD J. FUJITA[1]*

Cancer Research Laboratory and Department of Biochemistry, University of Western Ontario, London, Ontario, Canada N6A 5B7,[1] and Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824[2]

Received 18 December 1984/Accepted 18 February 1985

The nucleotide sequence of the 3′ two-thirds of a highly conserved, molecularly cloned human cellular *src* gene (c-*src*) has been determined. This region of the c-*src* gene encodes the tyrosine kinase domain of the cellular *src* protein (pp60$^{c\text{-}src}$) and corresponds to exons 6 through 12 of the chicken c-*src* gene, as well as nucleotides 545 to 1542 of the Rous sarcoma virus *src* gene (v-*src*). The human c-*src* sequence is very strongly conserved with respect to both the chicken c-*src* and the Rous sarcoma virus v-*src* genes, with nearly 90% nucleotide homology observed in this region. Amino acid sequence conservation in this region is even greater; 98% of the amino acids are conserved between human and chicken c-*src*. Furthermore, the exon sizes and the locations of the exon-intron boundaries are identical in the human and chicken c-*src* genes. However, sequences within the introns have not been conserved, and the introns within the human c-*src* gene are significantly larger than the corresponding introns within the chicken c-*src* gene. The strong amino acid conservation between the carboxy-terminal two-thirds of pp60$^{c\text{-}src}$ of species as divergent as humans and chickens suggests that this portion of the pp60$^{c\text{-}src}$ protein specifies one or more functional domains that are of great importance to some aspect of normal cellular growth or differentiation.

The retroviral oncogenes are derived from normal cellular sequences (proto-oncogenes) which have been acquired from host cells (5). Often these genes have undergone modification during the process of viral transduction and virus propagation, resulting in small changes in nucleotide sequence, insertions, or deletions, or more than one of the above, within the coding or control sequences of the gene. It is thought that some of these modifications might be involved in imparting an acute transforming ability to the retroviruses that carry these oncogenes (reviewed in reference 4). The Rous sarcoma virus (RSV) *src* gene (v-*src*) is one of the most extensively studied retroviral oncogenes. Nucleic acid hybridization studies have shown that a cellular counterpart of v-*src* is present in the normal cells of all vertebrate species examined (50, 51) as well as in *Drosophila melanogaster* (46). The chicken c-*src* gene is expressed at relatively low levels in most cell and tissue types, but mRNA levels have been found to be elevated in macrophages (16), and levels of pp60$^{c\text{-}src}$ were found to be elevated in neural tissue during development (8, 49). pp60$^{c\text{-}src}$ possesses a tyrosine kinase activity analogous to that of the v-*src* protein pp60$^{v\text{-}src}$ (7, 33), and recently pp60$^{v\text{-}src}$ has been shown to possess an inositol phosphorylating activity (53), which may be related to the tyrosine kinase activity.

The chicken c-*src* gene has recently been characterized by molecular cloning and DNA sequencing (35, 43, 55, 56). These studies have shown that most of the chicken c-*src* coding sequences are very homologous to those of v-*src* and are contained within 11 exons (numbered 2 through 12) that span approximately 6 kilobases (kb) of cellular DNA; in addition another exon (exon 1) codes for a 5′ untranslated region of mRNA (55). At this time, relatively little is known about the properties of human c-*src* loci and human pp60$^{c\text{-}src}$. Our laboratory has recently characterized, by restriction enzyme mapping, molecular clones of a highly conserved human c-*src* gene isolated from a lambda phage genomic library (14). This c-*src* locus spans 20 kb of cellular DNA and contains at least 11 exons homologous to v-*src* (14). This gene has been called human c-*src* locus 1 (c-*src*-1), because another somewhat less homologous c-*src*-related gene (locus 2) has been detected by others (14, 34). c-*src*-1 is transcriptionally active and codes for a polyadenylated mRNA species of ca. 5 kb (34; J. Radul and D. J. Fujita, unpublished results).

In the present communication we present our results of the nucleotide sequencing of the 3′ two-thirds of the human c-*src*-1 coding region, which is found to be highly conserved relative to the corresponding regions of chicken c-*src* and RSV v-*src* (14, 54, 55). This region is of considerable interest because it contains the coding sequences for the tyrosine kinase domain of pp60^{src}. Similar, closely related domains have been shown to be present in the products of several other retroviral oncogenes, including *yes*, *fes*, *fps*, *abl*, *fos* and *fgr*, which also possess tyrosine kinase activity (15, 17), and somewhat more distantly related domains have been identified in the products of other oncogenes (*mos*, *raf*, *mil*, *fms*, *erb*B) (reviewed in reference 22). The corresponding region of a *D. melanogaster* c-*src* (D*src*) gene product has also been deduced by nucleotide sequencing and exhibits

* Corresponding author.
† Present address: The Rockefeller University, New York, NY 10021.
‡ Present address: Department of Molecular Biology, Case Western Reserve University, Cleveland, OH 44106.

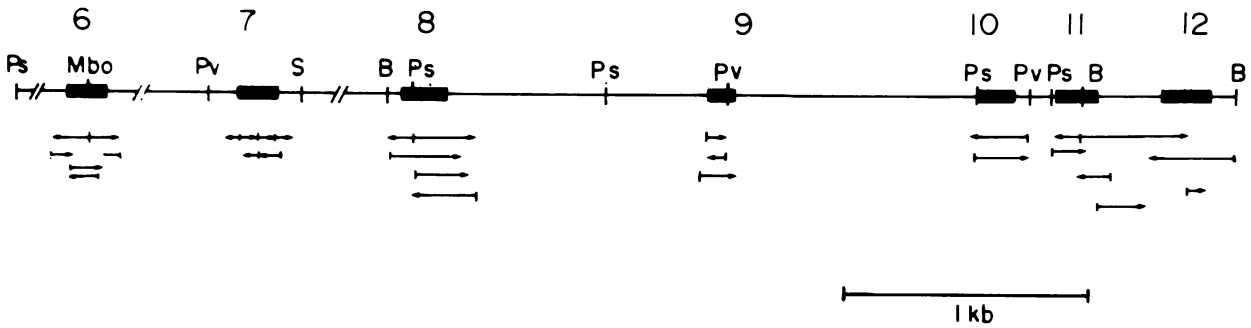

FIG. 1. Sequencing strategy for human c-*src*. The filled boxes depict exon regions, and the arrows depict the location and 5′ to 3′ polarity of fragments sequenced by the M13-dideoxy chain termination method. The DNA sequence of both strands was determined to resolve any ambiguities observed in exon sequences. Pv, *Pvu*II; S, *Sac*I; B, *Bam*HI; Ps, *Pst*I; Mbo, *Mbo*I. —/ /— indicates not drawn to scale.

54% amino acid homology to pp60$^{v\text{-}src}$ in this region (21). The nucleotide sequence of the 5′ one-third of the human c-*src* coding region will be presented in a subsequent report; because of technical reasons resulting from a lower level of nucleotide sequence homology, exon 3 has not yet been definitively identified (14).

Several cellular oncogenes have recently been implicated in a variety of cancers and leukemias (see, for example, references 4, 11, and 20). Although the human c-*src* gene has not yet been implicated in the causation or maintenance of human neoplasms, increased levels of c-*src* transcripts have been observed in certain human hematological disorders (38, 47). A precise characterization of the structure and nucleotide sequence of this gene will be of the utmost importance for further studies aimed at detecting possible c-*src* gene alterations associated with neoplastic disorders, as well as for resolving questions concerning the number and nature of pp60^{src} proteins in human cells (45), the possible occurrence of alternative RNA splicing events, and the functional roles of c-*src* in normal cell growth or development. It is also of great interest to observe how the nucleotide and amino acid sequence of c-*src* and related genes have been modified through evolution.

MATERIALS AND METHODS

Subcloning and DNA sequencing of c-*src* exons. The techniques used in the restriction enzyme cleavage and molecular cloning of the human c-*src* gene have been described elsewhere (14). The region of DNA sequenced here is contained entirely within the lambda-S3H cloned insert, which spans 16 kb and encompasses exons 4 through 12 of human c-*src* (14). A 0.6-kb *Bam*HI fragment subcloned in pBR322 was cloned directly into the *Bam*HI site of M13mp8 (29). This fragment was shown to contain the 3′ end of exon 11 and all of exon 12. A 2.8-kb *Bam*HI pAT153 subclone contained exons 8 through 10 and the 5′ end of exon 11. These *Bam*HI fragments were subjected to a limited digestion with *Hae*III plus *Alu*I and cloned directly into the *Sma*I site of M13 mp8. *Pst*I and *Pst*I-*Pvu*II digestions were also performed on the *Bam*HI fragments, and the resulting DNA fragments were cloned into appropriate sites in M13mp8. M13 clones containing exons 6 and 7 were obtained by limited digestion of a 1.9-kb *Hin*dIII pAT153 subclone with *Hae*III or *Mbo*I and cloning into the appropriate site of M13mp8. Clones harboring c-*src* sequences were selected by hybridization to v-*src* probes, as described below.

DNA sequencing was performed by using the dideoxy method and the conditions described by Sanger et al. (39) when [α-^{32}P]dATP (New England Nuclear Corp., Boston, Mass.) was used for labeling. The protocol of Biggin et al. (3) was used when [α-^{35}S]dATP (New England Nuclear) was used. Sequencing gels (40 cm long by 0.3 mm thick) of 8% acrylamide–7.7 M urea were used to determine sequences up to 200 nucleotides from a 17-nucleotide M13 primer (New England Biolabs, Beverly, Mass.). Eighty-centimeter gels were used to determine sequences up to 450 nucleotides from the priming site. A long-run buffer (1) was used for all sequencing gels. In many instances, human c-*src* coding sequences that were homologous to v-*src* sequences were detected by using a computer-assisted matching program (26).

Hybridization of M13 clones. Individual, colorless M13 plaques formed on X-Gal indicator plates (29) were picked and used to inoculate 2 ml of exponentially growing *Escherichia coli* JM103. After 6 h of growth, the bacteria were pelleted and 2 μl of supernatant was spotted onto nitrocellulose filters (Schleicher & Schuell, Inc., Keene, N.H.). The filters were treated with 0.5 M NaOH–1.5 M NaCl for 1 min to expose the viral DNA and then treated with 0.5 M Tris-hydrochloride (pH 8)–3 M NaCl for 5 min. The filters were baked and then prehybridized and hybridized at 42°C by using the conditions described by Wahl et al. (57). In some cases, v-*src* subgenomic fragments specific for certain exons were used as probes (14). Methods used for the preparation of labeled nick-translated v-*src* fragments have been described previously (14). Most of the hybridizing clones could be detected by exposure of filters to Kodak XAR 5 film for 6 h with a Dupont Cronex intensifying screen.

RESULTS

Strategy for sequencing human c-*src*. A human c-*src* gene containing coding sequences that are closely related to the chicken c-*src* gene and the v-*src* gene of RSV has recently been isolated in our laboratory. Sequences homologous to v-*src* were present within a region spanning approximately 20 kb, and at least nine exon regions were defined by combining fine-structure restriction enzyme mapping with hybridization to a series of sub-genomic v-*src* probes (14). Through these studies, we were able to identify discrete DNA fragments, generated by restriction enzyme cleavage, that contained sequences homologous to v-*src*. Putative exon-containing fragments were purified after being

subcloned into plasmid vectors and subjected to a limited digestion with the restriction enzymes *Alu*I, *Hae*III, *Mbo*I, and *Rsa*I, either singly or in combination, and the resulting DNA fragments were subcloned into M13mp8 (29). Phage with inserts containing exon sequences were identified by hybridization to ^{32}P-labeled v-*src* probes, and nucleotide sequences were determined by the dideoxy sequencing method (39). In certain cases, small restriction fragments (<700 base pairs) known to contain exon sequences were cloned directly into the appropriate cloning sites in M13mp8. The relevant restriction enzyme sites used and the regions that were sequenced are illustrated in Fig. 1.

Nucleotide sequence of the 3′ two-thirds of human c-*src* and deduced amino acid sequence. In the present study, the seven human c-*src* exons that contain nucleotides corresponding to nucleotides 545 through 1542 of v-*src* (10, 14, 51, 54) and are assumed to code for the carboxy-terminal two-thirds of human pp60$^{c\text{-}src}$ were sequenced. The resulting nucleotide sequence and deduced amino acid sequence are presented in Fig. 2. Nucleotides and amino acids that differ between the human and chicken c-*src* genes are indicated beneath the human c-*src* sequence, and the numbering system adopted for chicken c-*src* is used here (55). There is considerable c-*src* nucleotide homology between the two species, with 89% (935 out of 1,055 residues) of the nucleotides conserved. The majority of these nucleotide changes are silent, third-position codon changes resulting in no amino acid substitutions; therefore, a divergence of only 2% is seen at the amino acid level. Furthermore, all six of the amino acid changes that were observed within the region analyzed are conservative substitutions as defined by the criteria of French and Robson (12). Interestingly, both the molecular size and the relative hydrophobicity of the amino acid residues are conserved in five of the six changes. The tyrosine residue 416 which is phosphorylated in pp60$^{v\text{-}src}$ in vivo and in the chicken pp60$^{c\text{-}src}$ in vitro (33, 48) has also been conserved in this human c-*src* locus, as has lysine 295, which has been implicated as an ATP-binding site in pp60$^{v\text{-}src}$ (24).

Of particular interest is the sequence homology at the 3′ end of the human and chicken c-*src* genes. The deduced amino acid sequence of the human c-*src* carboxy terminus is identical to that of the chicken pp60$^{c\text{-}src}$ but different from that of RSV pp60$^{v\text{-}src}$. Takeya and Hanafusa (55) have shown that the predicted carboxy-terminal region of chicken pp60$^{c\text{-}src}$ differs from that of pp60$^{v\text{-}src}$; specifically, the last 19 amino acids of pp60$^{c\text{-}src}$ have been substituted by 12 unrelated amino acids in the v-*src* protein. Nucleotide sequences homologous to the v-*src*-specific 3′ terminus were found approximately 1 kb downstream from the c-*src* termination codon in chicken cellular DNA (55). The v-*src*-specific 3′ terminus is thought to have been captured by RSV through a recombinational event involving the region downstream from the normal termination codon of the chicken c-*src* gene. Our nucleotide sequencing results predicting the identical amino acid sequence at the carboxy termini of the human and chicken pp60$^{c\text{-}src}$ proteins support this model. However, at this time, we have no evidence suggesting the existence of an analogous "v-*src*-like" sequence downstream from the carboxy-terminal coding region of human c-*src*; preliminary hybridization analyses with a total v-*src* probe have failed to detect such a sequence within a region extending 2 kb downstream from exon 12.

Human c-*src* exons. The sequence data also indicate that the human c-*src* exon sizes and intron-exon junctions of exons 6 through 12 are identical to those of the chicken c-*src* gene (55) (Fig. 2; Table 1). Sequences of the predicted human c-*src* splice sites bordering each exon are listed in Table 1. All of the putative splice acceptor signals are similar to the consensus sequence YNYYYNCAG/, and the presumed splice donor signals conform to the /GTNAG consensus sequence (44), strongly suggesting that these regions are indeed splice junctions.

In contrast to the extensive nucleotide homology found within human and chicken c-*src* exons, sequences within the introns have not been conserved. This finding is supported by several lines of evidence: (i) the proposed intron regions do not form heteroduplexes that are visible by electron microscopy, even under low-stringency spreading conditions, in contrast to the exon regions (data not shown); (ii) the introns within the human c-*src* gene are all significantly larger than the equivalent chicken c-*src* introns (Table 2); (iii) the nucleotide sequence of chicken c-*src* intron 11 and the corresponding human c-*src* intron were compared, and no regions of homology were present (T. Takeya and H. Hanafusa, personal communication; S. Anderson, unpublished results).

DISCUSSION

It has been suggested that the pp60$^{v\text{-}src}$ protein contains at least two functional domains: a domain in the carboxy-terminal half, which contains tyrosine kinase activity, and one or more domains in the amino-terminal half which may affect membrane binding, cell morphology, and tumorigenicity (9, 13, 27, 37, 52). The region of the human c-*src* locus 1 described in this report encompasses the tyrosine kinase coding domain which is highly homologous to corresponding regions of chicken c-*src* and RSV v-*src*. The region of pp60$^{c\text{-}src}$ that is specified by these nucleotides also exhibits considerable homology to regions within several oncogene products, including those of the *abl*, *fes*, *fps*, and *yes* genes. These are apparently members of a gene family coding for proteins that possess a tyrosine kinase activity (15, 17, 22). In contrast, our preliminary results suggest that the region of human c-*src* coding for the amino-terminal one-third of pp60$^{c\text{-}src}$ exhibits considerably less homology to chicken c-*src* and RSV v-*src* (A. Tanaka, C. P. Gibbs, and S. Anderson, unpublished observations).

The human c-*src* gene displays the typical intron-exon structure characteristic of eucaryotic cellular genes. We have assigned intron-exon boundaries by comparison with the v-*src* gene and by observing splice donor or acceptor consensus sequences at the proposed splice junctions. Exons have been numbered by analogy to the corresponding exons of the chicken c-*src* gene, which exhibits the same general structural organization as human c-*src*. Human c-*src* exons 6 through 12 have been found to correspond precisely to the homologous chicken c-*src* exons with respect to size and location of intron-exon boundaries (Fig. 2). However, preliminary evidence indicates that the human c-*src* exon 2 contains some structural alterations relative to chicken c-*src* and RSV v-*src* (Tanaka et al., unpublished observations). Introns within the human c-*src* gene are all significantly larger than the corresponding introns within the chicken c-*src* gene (Table 2). The increased intron size in the human c-*src* gene accounts for its much larger size (20 kb) relative to that of the chicken c-*src* gene (6 kb) (14).

The coding region of the c-*src* gene presented here is very highly conserved in chickens and humans, exhibiting 89% nucleotide homology, which probably reflects strong functional constraints on pp60$^{c\text{-}src}$ evolution, as well as a similar pattern of codon usage in vertebrates (19). On the amino acid level, 98% homology exists between their products in the

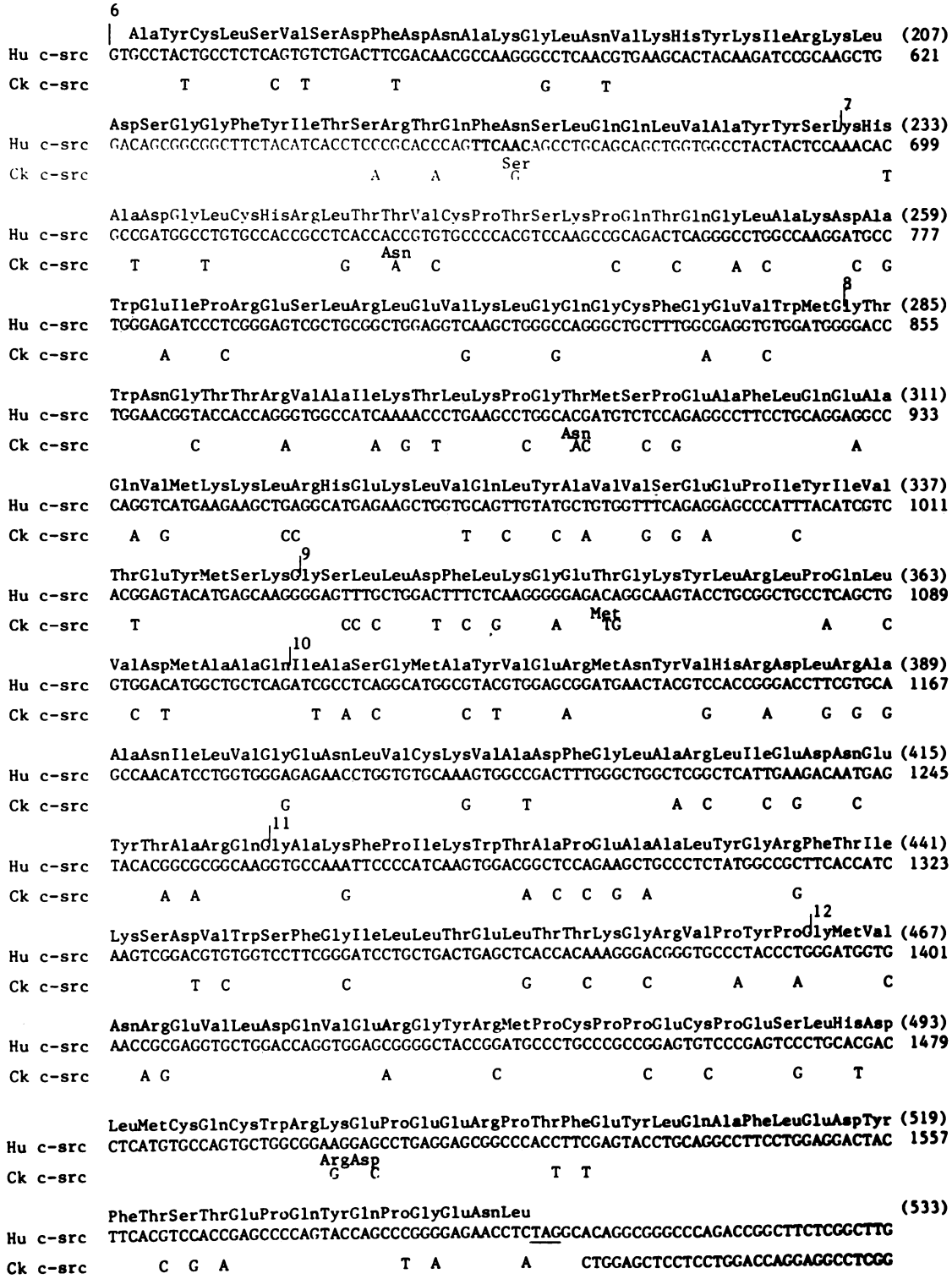

FIG. 2. Nucleotide sequence of the 3′ two-thirds of human c-*src* and the deduced amino acid sequence. The region comprising exons 6 through 12 of human c-*src* is compared with the homologous chicken exons, using the numbering system of chicken c-*src* (55). Only nucleotides or amino acids that differ from those of human c-*src* are shown for chicken c-*src*.

corresponding region encompassing amino acids 183 to 533, and the six amino acid changes observed in this region (Fig. 2) can be interpreted as conservative substitutions (12). As a point of comparison, in contrast to the highly conserved nature of the c-*src* gene, the human c-*myc* gene shares somewhat less amino acid homology (72%) with the chicken c-*myc* gene and contains several large deletions and insertions relative to chicken c-*myc* (6, 58).

TABLE 1. Splice donor and splice acceptor sequences of human c-*src*[a]

Exon no.	Intron (SA)	Exon size (base pairs)[b]	Intron (SD)
6 H	GCCCCGCAG GTGCCTACTG	. . .150. . .	TACTACTCCA GTGAG
C	TA T GGGGGT		GTT
7 H	CCTCCTCAG AACACGCCGA	. . .156. . .	GTGTGGATGG GTAAG
C	TGTGCT GGGGT T		C
8 H	CCTCAACAG GGACCTGGAA	. . .180. . .	ATGAGCAAGG GTGAG
C	T CTC		
9 H	TCTGCCCAG GGAGTTTGCT	. . . 77. . .	GGCTGCTCAG GTGAG
C	CT A GGGGCC C		G
10 H	CTGCAG ATCGCCTCAG	. . .154. . .	GCGCGGCAAG GTGGG
C	CA GGT A C		A T
11 H	TTCCTGCAG GTGCCAAATT	. . .132. . .	CCCTACCCTG GTAAG
C	CC GCC GGGGGTTG		A A G
12 H	CTGCCACAG GGATGGTGAA	. . .(209. . .TAG)	
C	T T GGGGGTTC		

[a] The nucleotide sequences of human (H) splice acceptors (SA) and splice donors (SD) are compared with the chicken (C) splice junctions. Only nucleotide differences are shown for chicken c-*src*.

[b] Exon sequences at the splice junctions are shown. The size in base pairs of each exon is shown.

Although we have not performed an exhaustive analysis of the structural features of human pp60^{src} as predicted by our sequencing results, we have noticed that several regions thought to be of functional importance in RSV pp60$^{v\text{-}src}$ or chicken pp60^{src}, or both, have been conserved in human pp60^{src}. For example, the tyrosine 416 residue and adjacent amino acids are totally conserved in human pp60^{src}; this region is known to be phosphorylated in vivo in pp60$^{v\text{-}src}$ and in vitro in chicken pp60^{src}, although it apparently is not phosphorylated in vivo in chicken pp60^{src} (33, 48). It is not known at this time whether this site is phosphorylated in human pp60^{src} species. Similarly, lysine 295, a residue involved in ATP binding of pp60$^{v\text{-}src}$ (24), and neighboring amino acids have been conserved in both the human and the chicken pp60^{src}.

Interestingly, a visual search of the amino acid sequence has shown that human pp60^{src} possesses a region that shares strong homology to regions that are highly conserved in a number of regulatory proteins that have DNA binding activity, including the repressor and *cro* proteins of bacteriophage lambda and 434, and the Mu phage B transposition protein (30, 32, 40). This region is centered around glycine 288 in human pp60^{src} and includes the consensus sequence residues Gly(284)-X-X-X-Gly(288)-X-X-X-X-X-Ile(294); for the purposes of this report, we have called this region "region I." A characteristic structural motif shared by several known procaryotic DNA binding proteins consists of a DNA binding domain containing two alpha-helical regions separated by a reverse turn occurring at the central glycine residue of the consensus sequence (32, 40); a computerized secondary structure prediction indicates that the sequence of amino acids in region I is consistent with these features. For illustrative purposes, Fig. 3 provides a comparison between region I and DNA binding domains of several known DNA binding proteins. Region I is also conserved in chicken pp60^{src} and RSV pp60$^{v\text{-}src}$, according to published sequence information (10, 41, 55). We also have detected similar consensus sequences in several other oncogene products, and computer-assisted structural and statistical analyses are in progress. We do not know the significance of these results, nor do we know whether region I can indeed function as a DNA binding region. Such an activity is highly speculative at this time, especially since it is known that the majority (ca. 70 to 90%) of pp60$^{v\text{-}src}$ molecules are localized at the inner surface of the cellular plasma membrane (5, 59), except in rat cells for which both a nuclear membrane and cytoplasmic membrane localization have been reported for pp60$^{v\text{-}src}$ (25). In addition, a DNA binding activity has not been directly demonstrated for pp60^{src} species. A putative DNA binding consensus sequence has also been reported in the avian myeloblastosis virus oncogene product p48amv (*myb*), which is known to be localized predominately in cell nuclei (28). It is also possible that the presence of a DNA binding consensus sequence in this region of pp60^{src} is merely a reflection of an ATP-binding domain which has been shown to overlap with region I in RSV pp60$^{v\text{-}src}$ (24). However, it should be noted that the related bovine cAMP dependent protein kinase does not possess the DNA binding consensus sequence, although it possesses conserved residues thought to be important for ATP binding (2, 24).

The last 19 amino acids of the predicted human pp60$^{c\text{-}src}$ protein are identical to those of the chicken pp60$^{c\text{-}src}$ carboxy terminus. In contrast, these 19 amino acids are replaced by 12 unrelated amino acids in the RSV pp60$^{v\text{-}src}$ protein. This

TABLE 2. Size comparison of homologous human and chicken introns

Intron no.	Intron size (base pairs)[a]	
	Human	Chicken
6[b]	1,200	350
7	2,300	85
8	1,600	78
9	780	61
10	160	118
11	280	79

[a] Human intron sizes are estimates from restriction enzyme mapping and DNA sequencing information. Chicken intron sizes are from reference 55.

[b] Introns are numbered here with the same number as the exon located immediately 5′ of the intron.

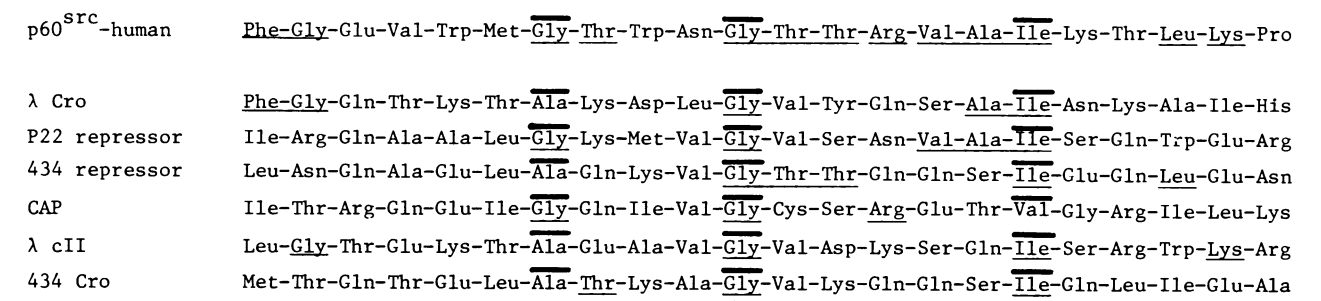

FIG. 3. Comparison of human c-*src* region I with DNA binding regions of procaryotic regulatory proteins (32, 40). Residues marked by a heavy bar are consensus residues found in a large number of DNA binding regions examined (32, 40). Human c-*src* sequences or residues that match those in corresponding regions of certain procaryotic DNA binding domains are underlined, as are the matching residues.

finding lends further support to the hypothesis that the 3′ terminus of v-*src* was acquired by a recombinational event involving the capture of a region downstream from the normal chicken c-*src* terminus (55). It is perhaps also possible that the pp60$^{v\text{-}src}$ carboxy-terminal region was derived through an alternative splicing mechanism. However, this possibility seems less likely, since a consensus splice donor sequence is not present in this region and the nucleotide homology between the human and chicken c-*src* genes extends up to, but not beyond, the proposed termination codon for each gene, suggesting that this region signals an authentic pp60$^{c\text{-}src}$ carboxy terminus (55) (Fig. 2).

Experiments designed to test the transforming ability of chicken c-*src* indicate that c-*src* is not capable of cell transformation (18, 23, 36, 42). However, chimeric forms of chicken c-*src* and RSV v-*src* that contain the v-*src* carboxy-terminal coding region are capable of causing efficient cell transformation (18, 23, 42). In addition, recent evidence suggests that alterations in other chicken c-*src* regions may result in transforming activity (23). These results suggest, by analogy, that certain alterations of the human c-*src* gene might affect its functional activity or its ability to transform appropriate cells, or both. At this time, the involvement of c-*src* in the development or progression of specific human neoplastic disorders has not been clearly documented. However, expression of human c-*src* mRNA has been found to be elevated in some human leukemias (38, 47).

In conclusion, we have demonstrated that the nucleotide sequence and exon structure of the 3′ two-thirds of the human c-*src* gene are highly conserved relative to that of chicken c-*src*. This finding is consistent with the suggestion that pp60$^{c\text{-}src}$ has an important cellular function and possibly plays a tissue-specific role in development or differentiation (8, 16, 31, 49).

ACKNOWLEDGMENTS

We thank T. Takeya and H. Hanafusa for unpublished information about a chicken c-*src* intron sequence; G. Chaconas for advice about the detection of possible DNA binding consensus sequences and assistance in computerized structural analyses of protein subregions; R. Parker, G. Mardon, H. E. Varmus, and J. M. Bishop for information about a second c-*src* locus; G. Mackie for advice; and D. Marsh and L. Bonis for help in manuscript preparation.

This work was supported by grants from the Medical Research Council of Canada, the National Cancer Institute of Canada, the Leukemia Research Fund (Toronto), and the U.S. National Science Foundation. S.K.A. was supported by a predoctoral studentship from the National Cancer Institute of Canada. H.J.K. gratefully acknowledges a Faculty Research Award from the American Cancer Society.

LITERATURE CITED

1. **Anderson, S.** 1981. Shotgun DNA sequencing using cloned DNaseI-generated fragments. Nucleic Acids Res. **9**:3015–3025.
2. **Barker, W. C., and M. O. Dayhoff.** 1982. Viral *src* gene products are related to the catalytic chain of mammalian cAMP-dependent protein kinase. Proc. Natl. Acad. Sci. U.S.A. **79**:2836–2839.
3. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and 35-S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. U.S.A. **80**:3963–3965.
4. **Bishop, J. M.** 1983. Cellular oncogenes and retroviruses. Annu. Rev. Biochem. **52**:301–354.
5. **Bishop, J. M., and H. Varmus.** 1982. Functions and origins of retroviral transforming genes, p. 999–1108. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
6. **Colby, W. W., D. H. Smith, and A. D. Levinson.** 1983. Identification and nucleotide sequence of a human locus homologous to the v-myc oncogene of avian myelocytomatosis virus MC29. Nature (London) **301**:722–725.
7. **Collett, M. S., J. S. Brugge, and R. L. Erikson.** 1978. Characterization of a normal avian cell protein related to the avian sarcoma virus transforming gene product. Cell **15**:1363–1369.
8. **Cotton, P. C., and J. S. Brugge.** 1983. Neural tissues express high levels of the cellular *src* gene product pp60$^{c\text{-}src}$. Mol. Cell. Biol. **3**:1157–1162.
9. **Cross, F. R., E. A. Garber, D. Pellman, and H. Hanafusa.** 1984. A short sequence in the p60^{src} N terminus is required for p60^{src} myristylation and membrane association and for cell transformation. Mol. Cell. Biol. **4**:1834–1842.
10. **Czernilofsky, A. P., A. D. Levinson, H. E. Varmus, J. M. Bishop, E. Tischer, and H. M. Goodman.** 1980. Nucleotide sequence of an avian sarcoma virus oncogene (src) and proposed amino acid sequence for gene product. Nature (London) **287**:198–203. (Revised in Nature **301**:736–738, 1983.)
11. **Dalla-Favera, R., S. Martinotti, R. Gallo, J. Erikson, and C. Croce.** 1983. Translocation and rearrangements of the c-myc oncogene locus in human undifferentiated B-cell lymphomas. Science **219**:963–967.
12. **French, S., and B. Robson.** 1983. What is a conservative substitution? J. Mol. Evol. **11**:171–175.
13. **Fujita, D. J., J. B. Bechberger, and I. Nedic.** 1981. Four Rous sarcoma virus mutants which affect transformed cell morphology exhibit altered *src* gene products. Virology **114**:256–260.
14. **Gibbs, C. P., A. Tanaka, S. K. Anderson, J. Radul, J. Baar, A. Ridgway, H.-J. Kung, and D. J. Fujita.** 1985. Isolation and structural mapping of a human c-*src* gene homologous to the transforming gene (v-*src*) of Rous sarcoma virus. J. Virol. **53**:19–24.
15. **Goffen, J., N. Heisterkamp, F. Reynolds, and J. Stephenson.** 1983. Homology between phosphotyrosine acceptor site of human c-*abl* and viral oncogene products. Nature (London) **304**:167–169.
16. **Gonda, T., D. Sheiness, and J. M. Bishop.** 1982. Transcripts from the cellular homologs of retroviral oncogenes: distribution

among chicken tissues. Mol. Cell. Biol. **2**:617–624.
17. **Hampe, A., I. Laprevotte, F. Gailibert, L. Fedele, and C. Sherr.** 1982. Nucleotide sequences of feline retroviral oncogenes (v-*fes*) provide evidence for a family of tyrosine-specific protein kinase genes. Cell **30**:775–785.
18. **Hanafusa, H., H. Iba, T. Takeya, and F. Cross.** 1984. Transforming activity of the c-*src* gene, p. 1–7. *In* G. F. Vande Woude, A. J. Levine, W. C. Topp, and J. D. Watson (ed.), Cancer cells, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Hastings, K., and C. Emerson.** 1983. Codon usage in muscle genes and liver genes. J. Mol. Evol. **19**:214–218.
20. **Heisterkamp, N., J. Stephenson, J. Groffen, P. Hansen, A. de Klein, C. Bartram, and G. Grosveld.** 1983. Localization of the c-*abl* oncogene adjacent to a translocation break point in chronic myelocytic leukaemia. Nature (London) **306**:239–242.
21. **Hoffman, F. M., L. D. Fresco, H. Hoffman-Falk, and B. Shilo.** 1983. Nucleotide sequences of the Drosophila *src* and *abl* homologs: conservation and variability in the src family oncogenes. Cell **35**:393–401.
22. **Hunter, T.** 1984. The proteins of oncogenes. Sci. Am. **251**:70–79.
23. **Iba, H., T. Takeya, F. R. Cross, T. Hanafusa, and H. Hanafusa.** 1984. Rous sarcoma virus variants that carry the cellular *src* gene instead of the viral *src* gene cannot transform chicken embryo fibroblasts. Proc. Natl. Acad. Sci. U.S.A. **81**:4424–4428.
24. **Kamps, M. P., S. S. Taylor, and B. M. Sefton.** 1984. Direct evidence that oncogenic tyrosine kinases and cyclic AMP-dependent protein kinase have homologous ATP-binding sites. Nature (London) **310**:589–592.
25. **Krueger, J. G., E. Wang, E. A. Garber, and A. R. Goldberg.** 1980. Differences in intracellular location of pp60^{src} in rat and chicken cells transformed by Rous sarcoma virus. Proc. Natl. Acad. Sci. U.S.A. **77**:4142–4146.
26. **Larson, R., and J. Messing.** 1982. Apple II software for M13 shotgun sequencing. Nucleic Acids Res. **10**:39–49.
27. **Levinson, A. D., S. A. Courtneidge, and J. M. Bishop.** 1981. Structural and functional domains of the Rous sarcoma virus transforming protein (pp60^{src}). Proc. Natl. Acad. Sci. U.S.A. **78**:1624–1628.
28. **Lipsick, J. S., W. J. Boyle, M. A. Lampert, and M. A. Baluda.** 1984. The oncogene of avian myeloblastosis virus is an altered proto-oncogene, p. 153–151. *In* G. F. Vande Woude, A. J. Levine, W. C. Topp, and J. D. Watson (ed.), Cancer cells, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**:269–276.
30. **Miller, J. L., S. K. Anderson, D. J. Fujita, G. Chaconas, D. L. Baldwin, and R. M. Harshey.** 1984. The nucleotide sequence of the *B* gene of bacteriophage Mu. Nucleic Acids Res. **12**:8627–8638.
31. **Mueller, R., D. Slamon, J. Tremblay, M. Cline, and I. Verma.** 1982. Differential expression of cellular oncogenes during pre- and post-natal development of the mouse. Nature (London) **299**:640–644.
32. **Ohlendorf, D. H., W. F. Anderson, and B. W. Matthews.** 1983. Many gene-regulatory proteins appear to have a similar alpha-helical fold that binds DNA and evolved from a common precursor. J. Mol. Evol. **11**:109–114.
33. **Oppermann, H., A. D. Levinson, L. Levintow, and J. M. Bishop.** 1979. Uninfected vertebrate cells contain a protein that is closely related to the product of the avian sarcoma virus transforming gene. Proc. Natl. Acad. Sci. U.S.A. **76**:1804–1808.
34. **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 chromosomes 1 and 20. Mol. Cell. Biol. **5**:831–838.
35. **Parker, R. C., H. E. Varmus, and J. M. Bishop.** 1981. Cellular homologue (c-*src*) of the transforming gene of Rous sarcoma virus: isolation, mapping, and transcriptional analysis of c-*src* and flanking regions. Proc. Natl. Acad. Sci. U.S.A. **78**:5842–5846.
36. **Parker, R. C., H. E. Varmus, and J. M. Bishop.** 1984. Expression of v-*src* and chicken c-*src* in rat cells demonstrates qualitative differences between pp60$^{v\text{-}src}$ and pp60$^{c\text{-}src}$. Cell **37**:131–139.
37. **Parsons, J. T., D. Bryant, V. Wilkerson, G. Gilmartin, and S. J. Parsons.** 1984. Site-directed mutagenesis of Rous sarcoma virus pp60^{src}: identification of functional domains required for transformation, p. 37–42. *In* G. F. Vande Woude, A. J. Levine, W. C. Topp, and J. D. Watson (ed.). Cancer cells, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
38. **Rosson, D., and A. Tereba.** 1983. Transcription of hematopoietic-associated oncogenes in childhood leukemia. Cancer Res. **43**:3912–3918.
39. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
40. **Sauer, R. T., R. R. Yocum, R. F. Doolittle, M. Lewis, and C. O. Pabo.** 1982. Homology among DNA-binding proteins suggests use of a conserved super-secondary structure. Nature (London) **298**:447–451.
41. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
42. **Shalloway, D., P. M. Coussens, and P. Yaciuk.** 1984. C-*src* and v-*src* homolog overexpression in mouse cells, p. 9–17. *In* G. F. Vande Woude, A. J. Levine, W. C. Topp, and J. D. Watson (ed.), Cancer cells, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
43. **Shalloway, D., A. D. Zelenetz, and G. M. Cooper.** 1981. Molecular cloning and characterization of the chicken gene homologous to the transforming gene of Rous sarcoma virus. Cell **24**:531–541.
44. **Sharp, P. A.** 1981. Speculations on RNA splicing. Cell **23**:643–646.
45. **Shealy, D. J., and R. L. Erikson.** 1981. Human cells contain two forms of pp60$^{c\text{-}src}$. Nature (London) **293**:666–669.
46. **Shilo, B. and R. A. Weinberg.** 1981. DNA sequences homologous to vertebrate oncogenes are conserved in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U.S.A. **78**:6789–6792.
47. **Slamon, D., J. deKerion, I. Verma, and M. Cline.** 1984. Expression of cellular oncogenes in human malignancies. Nature (London) **224**:256–262.
48. **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$^{v\text{-}src}$) and its normal cellular homologue (pp60$^{c\text{-}src}$). Proc. Natl. Acad. Sci. U.S.A. **78**:6013–6017.
49. **Sorge, L., B. Levy, and P. Maness.** 1984. pp60$^{c\text{-}src}$ is developmentally regulated in the neural retina. Cell **36**:249–257.
50. **Spector, D., H. E. Varmus, and J. M. Bishop.** 1978. Nucleotide sequences related to the transforming gene of avian sarcoma virus are present in DNA of uninfected vertebrates. Proc. Natl. Acad. Sci. U.S.A. **75**:4102–4106.
51. **Stehelin, D., H. Varmus, J. M. Bishop, and P. Vogt.** 1976. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. Nature (London) **260**:170–173.
52. **Stoker, A., P. J. Enrietto, and J. A. Wyke.** 1984. Functional domains of the pp60$^{v\text{-}src}$ protein as revealed by analysis of temperature-sensitive Rous sarcoma virus mutants. Mol. Cell. Biol. **4**:1508–1514.
53. **Sugimoto, Y., M. Whitman, L. Cantley, and R. L. Erikson.** 1984. Evidence that the Rous sarcoma virus transforming gene product phosphorylates phosphatidylinositol and diacylglycerol. Proc. Natl. Acad. Sci. U.S.A. **81**:2117–2121.
54. **Takeya, T., R. A. Feldman, and H. Hanafusa.** 1982. DNA sequence of the viral and cellular *src* gene of chickens. I. Complete nucleotide sequence of an *Eco*RI fragment of recovered avian sarcoma virus which codes for gp37 and pp60^{src}. J. Virol. **44**:1–11.
55. **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.
56. **Takeya, T., H. Hanafusa, R. P. Junghans, G. Ju, and A. M.**

Skalka. 1981. Comparison between the viral transforming gene (*src*) of recovered avian sarcoma virus and its cellular homolog. Mol. Cell. Biol. **1**:1024–1037.

57. **Wahl, G. M., M. Stern, and G. R. Stark.** 1979. Efficient transfer of large DNA fragments from agarose gels to diazobenzloxymethal-paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. U.S.A. **76**:3683–3687.
58. **Watson, D. K., M. C. Psallidopoulos, K. P. Samuel, R. Dalla-Favera, and T. S. Papas.** 1983. Nucleotide sequence analysis of the human c-*myc* locus, chicken homologue, and myelocytomatosis virus MC29 transforming gene reveals a highly conserved gene product. Proc. Natl. Acad. Sci. U.S.A. **80**:3642–3645.
59. **Willingham, M. C., G. Jay, and I. Pastan.** 1979. Localization of the ASV *src* gene product to the plasma membrane of transformed cells by electron microscopic immunocytochemistry. Cell **18**:125–134.