Human Cellular src Gene: Nucleotide Sequence and Derived Amino Acid Sequence of the Region Coding for the Carboxy-Terminal Two-Thirds of pp60<sup>c-src</sup>

STEPHEN K. ANDERSON,† CAROL P. GIBBS,‡ AKIO TANAKA,† HSING-JIEN KUNG,‡ and DONALD J. FUJITA†

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

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<sup>c-src</sup>) 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<sup>c-src</sup> of species as divergent as humans and chickens suggests that this portion of the pp60<sup>c-src</sup> 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<sup>c-src</sup> were found to be elevated in neural tissue during development (8, 49). pp60<sup>c-src</sup> possesses a tyrosine kinase activity analogous to that of the v-src protein pp60<sup>v-src</sup> (7, 33), and recently pp60<sup>c-src</sup> 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<sup>c-src</sup>. 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<sup>c-src</sup>. 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, fns, erbB) (reviewed in reference 22). The corresponding region of a D. melanogaster c-src (Dsrc) gene product has also been deduced by nucleotide sequencing and exhibits...
54% amino acid homology to pp60-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 BamHI fragment subcloned in pBR322 was cloned directly into the BamHI 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 BamHI pAT153 subclone contained exons 8 through 10 and the 5′ end of exon 11. These BamHI fragments were subjected to a limited digestion with HaeIII plus Alul and cloned directly into the Smal site of M13 mp8. PstI and PstI-PvuII digestions were also performed on the BamHI 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 HindIII pAT153 subclone with HaeIII or MboI 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-32P]dATP (New England Nuclear Corp., Boston, Mass.) was used for labeling. The protocol of Biggin et al. (3) was used when [α-35S]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 made 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 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

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. Pvu II; S; SacI; B; BamHI; P; PstI; MboI. — — indicates not drawn to scale.
subcloned into plasmid vectors and subjected to a limited digestion with the restriction enzymes AluI, HaeIII, MboI, and RsaI, 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 32P-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 pp60c-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 pp60c-src in vivo and in the chicken pp60c-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 pp60v-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 pp60c-src but different from that of RSV pp60v-src. Takeya and Hanafusa (55) have shown that the predicted carboxy-terminal region of chicken pp60c-src differs from that of pp60v-src; specifically, the last 19 amino acids of pp60c-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 pp60c-src proteins support this model. However, at this time, we have no evidence suggesting the existence of an intact, v-src-specific "v-src-like" gene in the chicken genome. 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 YNNYYNCA/G, 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 longer 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 pp60v-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 pp60c-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 pp60c-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). In contrast, 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-v-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.

The 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).
Although we have not performed an exhaustive analysis of the structural features of human pp60src as predicted by our sequencing results, we have noticed that several regions thought to be of functional importance in RSV pp60src or chicken pp60src, or both, have been conserved in human pp60src. For example, the tyrosine 416 residue and adjacent amino acids are totally conserved in human pp60src; this region is known to be phosphorylated in vivo in pp60src and in vitro in chicken pp60src, although it apparently is not phosphorylated in vivo in chicken pp60src (33, 48). It is not known at this time whether this site is phosphorylated in human pp60src species. Similarly, lysine 295, a residue involved in ATP binding of pp60src (24), and neighboring amino acids have been conserved in both the human and the chicken pp60src.

Interestingly, a visual search of the amino acid sequence has shown that human pp60src 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 bacterio-

<table>
<thead>
<tr>
<th>Exon no.</th>
<th>Intron (SA)</th>
<th>Exon size (base pairs)</th>
<th>Intron (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 H C</td>
<td>GCCCGCGAG</td>
<td>. . . . . . . . 150</td>
<td>GGGGTT</td>
</tr>
<tr>
<td></td>
<td>TA T</td>
<td></td>
<td>TACTACCCA</td>
</tr>
<tr>
<td></td>
<td>GGGGGT</td>
<td></td>
<td>GTGAG</td>
</tr>
<tr>
<td>7 H C</td>
<td>CCTCTCTAG</td>
<td>. . . . . . . . 156</td>
<td>AACACGGCGA</td>
</tr>
<tr>
<td></td>
<td>TGTGCT</td>
<td></td>
<td>GGGGTT T</td>
</tr>
<tr>
<td>8 H C</td>
<td>CCTCAACAG</td>
<td>. . . . . . . . 180</td>
<td>GGACCTGGA</td>
</tr>
<tr>
<td></td>
<td>T CTC</td>
<td></td>
<td>ATGAGCAAGG</td>
</tr>
<tr>
<td>9 H C</td>
<td>TCTGCCAG</td>
<td>. . . . . . . . 77</td>
<td>GGAGTTTCTG</td>
</tr>
<tr>
<td></td>
<td>CT A</td>
<td></td>
<td>GGGGCC C</td>
</tr>
<tr>
<td>10 H C</td>
<td>CTGGAG</td>
<td>. . . . . . . . 154</td>
<td>GTGGGCGCAA</td>
</tr>
<tr>
<td></td>
<td>ATGCGCTCAG</td>
<td></td>
<td>GTGGG A</td>
</tr>
<tr>
<td>11 H C</td>
<td>TTCTGACAG</td>
<td>. . . . . . . . 132</td>
<td>GTGCCAAATT</td>
</tr>
<tr>
<td></td>
<td>GTGCCAATT</td>
<td></td>
<td>CCCTACCTTG</td>
</tr>
<tr>
<td></td>
<td>CC GCC</td>
<td></td>
<td>A A G</td>
</tr>
<tr>
<td>12 H C</td>
<td>CTGCCACAG</td>
<td>. . . . . . . . 209</td>
<td>GGAATGTTGAA</td>
</tr>
<tr>
<td></td>
<td>GTGGGTTC</td>
<td></td>
<td>(TAG)</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Intron no.</th>
<th>Intron size (base pairs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Human</td>
</tr>
<tr>
<td>6^</td>
<td>1,200</td>
</tr>
<tr>
<td>7</td>
<td>2,300</td>
</tr>
<tr>
<td>8</td>
<td>1,600</td>
</tr>
<tr>
<td>9</td>
<td>780</td>
</tr>
<tr>
<td>10</td>
<td>160</td>
</tr>
<tr>
<td>11</td>
<td>280</td>
</tr>
</tbody>
</table>

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

^ Introns are numbered here with the same number as the exon located immediately 5' of the intron.
from the American Cancer Society.

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. Varumus, 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 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


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 locus terminus (55). It is perhaps also possible that the pp60src 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 pp60src 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 pp60src is an important cellular function and possibly plays a tissue-specific role in development or differentiation (8, 16, 31, 49).

FIG. 3. Comparison of human c-src region I with DNA binding regions of prokaryotic 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 prokaryotic DNA binding domains are underlined, as are the matching residues.

