Characterization of the promoter for the rat and human link protein gene

Craig Rhodes*, Pierre Savagner, Sergio Line, Makoto Sasaki, Mike Chirigos, Kurt Doege+ and Yoshihiko Yamada
Laboratory of Developmental Biology, National Institute of Dental Research, National Institutes of Health, Bethesda, MD 20892, USA

Received October 5, 1990; Revised and Accepted March 14, 1991

ABSTRACT

We have isolated the 5'-end of the gene for the rat and human link protein by screening genomic libraries with oligonucleotides corresponding to the 5'-cDNA sequence. Several overlapping clones were isolated for the human link protein gene, while only one clone was obtained for the rat. All the clones contained a single exon of which the sequence was identical to the most 5'-end of the rat and human cDNAs. Transcription initiation sites for the rat link gene were identified by primer extension and S1 protection analysis using total RNA from the rat Swarm chondrosarcoma. Transcriptional initiation sites for the human link gene were determined by specific primer extension of RNA from human fetal cartilage. Comparison of 1500 bp of 5'-flanking sequence between the rat and human link protein genes showed strong sequence conservation near the start site of transcription with 80% overall identity. Analysis of the 5'-flanking regions also revealed a large inverted repeat consisting of repeating purine-pyrimidine, which has the potential to form left-handed Z-DNA.

Transcriptional regulation of the link protein gene was studied by coupling either 7.0 kb or 0.85 kb of 5'-flanking rat DNA to the chloramphenicol acetyltransferase (CAT) gene followed by transfection into chick embryonic chondrocytes (CEC) and HeLa cells. Both constructs had considerable CAT activity in CEC cells and less activity in HeLa cells. Furthermore, inclusion of a DNA fragment from the first intron increased relative CAT activity in both of these cell types. The increased activity from the first intron was shown to be orientation independent in CEC. These results indicate the presence of positive cis-acting regulatory elements in both the promoter and first intron of the rat gene for link protein.

INTRODUCTION

Link protein (1) is a major structural component of cartilage and is present in various connective tissues. Various sizes of link protein have been described (2–8) which appear to be due to differences in the degree of glycosylation (9). Link protein heterogeneity may also be affected by alternative splicing of the link protein gene transcripts, as has been shown in chondrosarcoma tumor cells (10). Link protein and the large aggregating cartilage proteoglycan, which has recently been renamed aggrecan (11), contain similar domains (12, 13) and have an affinity for hyaluronic acid (HA). In cartilage, the interaction of these molecules produce a large stable aggregate structure which contributes to compression resistance and to shock absorption in the joint.

Developmental and tissue-specific expression of cartilage genes including type II collagen, link protein, and aggrecan have been investigated in several systems. Abundant production of these proteins is characteristic of both hyaline cartilage (15) and the fully differentiated chondrocyte which show a co-regulation of these genes. Both aggrecan and link protein are co-localized in the developing chick limb buds (16). Differential expression of collagen II, aggrecan, and link protein has also been reported. In the nanomelic chicken sternum reduced levels of aggrecan mRNA and protein are found when compared to that of a normal chicken. However, link protein levels are expressed at similar levels in both nanomelic and normal chicken (17). In normal chicken eyes, link protein but not aggrecan is expressed in the lens capsule and epithelium, whereas co-expression of aggrecan and link protein is detected in the sclera (18). The immortalized rat chondrocyte (IRC) cell line also reveals a non-coordinate pattern of expression with reduced expression of type II collagen and high levels of expression of both aggrecan and link protein (19).

We have isolated the 5'-flanking regions for both the rat and human link protein genes in order to understand the mechanisms by which the link protein gene is regulated. We have identified

* To whom correspondence should be addressed at George Washington University Genetics Program

+ Present address: Research Unit, Shriner's Hospital for Crippled Children, Portland, OR 97201, USA

start sites of transcription by S1 nuclease mapping and primer extension. We have examined the DNA regions necessary for expression of the link protein gene by transfecting chondrocytes and HeLa cells with plasmid constructs containing the promoter of the rat link protein gene coupled to a reporter gene for chloramphenicol acetyl transferase (CAT) (20).

MATERIALS AND METHODS

Construction of cDNA library

RNA was prepared from the rat Swarm chondrosarcoma (21), and poly(A)+RNA selected on oligo (dT)-cellulose. Plaques on nitrocellulose were screened as described (22) with an oligonucleotide that was located 5′ to the oligonucleotide used as a primer for the cDNA synthesis.

The human link protein cDNA sequence was determined from a cDNA clone pHL10 isolated from a human chondrocyte cDNA library (gift of T. Kimura) using the rat cDNA as probe.

Isolation of Genomic clones

A rat genomic library in charon 4A (gift of Tom Sargent, National Institute of Child Health and Human Development, NIH), or a human genomic library in EMBL3 (gift of Frank Gonzalez, National Cancer Institute, NIH), were plated and the plaques transferred to nitrocellulose filters.

Construction of Recombinant CAT Plasmids

A series of plasmids was constructed in which DNA flanking the first exon was fused to the CAT gene of E. coli. The first of this series were promoter modifications; pRL10, had a *Hind* III fragment (−7000 to +35) ligated to *Hind* III digested pSV0Cat vector (20), while pRL12 (−7000 to −850), and pRL11(−850 to +35), were made by attaching *Hind* III linkers. pRL16 was derived from pRL10 by inserting a 2.2 kb *Eco*RI fragment into the *Bam*HI site of pRL10 after blunt-ending the fragment and attaching *Bam*HI linkers. pRL18 and pRL20, which contain 1.5 kb and 0.8 kb DNA of the 3′ portion of the 2.2 kb intron segment, respectively, were prepared by attaching *Bam*HI linkers to DNA fragments isolated from asymmetric *Bal*-31 digested 2.2kb *Eco*RI fragment.

A second series of CAT expression plasmids, pRL21-pRL24, was made by subcloning specific restriction endonuclease fragments from the first intron into the *Bam*HI site of pRL10. pRL21, from a 280 bp *Acc*I-*Sty*I fragment; pRL22, from 340 bp *Sty*I-*Eco*RI fragment; and pRL23, from a 1.6 kb *Eco*RI-*Acc*I fragment (see Fig. 1 and Fig. 4).

In addition, plasmid pRL26 was identical to pRL16 but in the reverse orientation. Whereas pRL27 was made by subcloning a transcriptionally inactive 2.3 kb DNA fragment from the first intron of mouse α1 (IV) collagen chain gene (23) into the *Bam*HI site of pRL10.

DNA Sequence Analysis of the 5′ Rat and Human Genomic Clone

DNA fragments cloned into M13 phage or Bluescript vectors (Stratagene) were sequenced by the enzymatic chain termination (24) method using T7 DNA polymerase (Sequenase, U.S. Biochemicals). S1 analysis was performed as described (25).

Transient Transfection and CAT Assay

Chick embryonic chondrocytes (CEC) were prepared from sterna of 15 day-old embryos as described (26). HeLa cells were grown

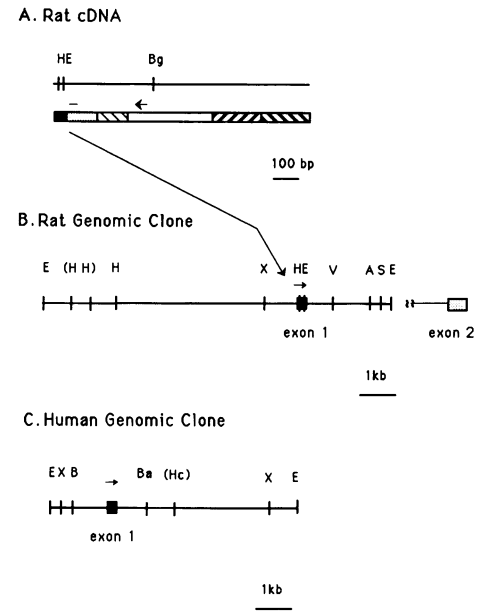

Fig. 1. Diagrammatic representation of the cDNA and genomic clones used in these studies. Restriction sites are indicated as: E, α *Eco*RI; Bg, *Bgl*II; Hc, *Hinc*II; X, *Xho*I; B, *Bam*HI; RV, *Eco*RV; H, *Hind*III;B,*Bal*I. Boxes represent exons. **A.** The complete exon structure for amino acid coding region of the rat link gene was reported previously (10). Additional cDNA clones were obtained by specific primer extension of rat chondrosarcoma poly(A)+ mRNA. The 5′ extended cDNA contained 35 bp of the first exon (designated by the 5′ solid box). The oligonucleotide used for primer extension is represented by the arrow. The oligonucleotide used for screening is represented by a short line, and the size scale is in bp. **B.** The first exon for the rat genomic clone is shown as a closed box with the small arrow above indicating direction of transcription. The second exon is represented by the stippled box. Two *Hind* III sites, whose order has not been determined, occur in the 5′ region and are designated by (H-H). Several *Hinc*II sites occur 3′ of the *Hinc*II site indicated by (Hc). The flanking *Eco*RI sites may have been introduced as synthetic linkers. **C.** The exon for the human genomic clone is shown as a closed box with a small arrow above indicating the direction of transcription.

in DMEM containing 10% fetal calf serum. The cells were transfected with 10 μg DNA from each construct, using the calcium-phosphate precipitation method (27). Precipitates were removed from the culture after 3 h (chick chondrocytes), or 15 h (HeLa cells). The cells were harvested 48 h after transfection, and CAT activity was assayed by the phase extraction method (28). Each construct was tested in duplicate in the same experiment. At least three independent assays were used in figure 4, and the constructs having a positive transcriptional influence were tested and compared in more than eight independent assays in the CEC cells. In figure 5 three transfection assays were prepared and repeated in duplicate. In addition, the plasmids used in figure 5 were prepared at the same time as a set.

RESULTS

Isolation of Rat cDNA and Genomic Clones

We previously isolated several cDNA and genomic clones for the rat link protein. Sequence analysis of these clones revealed that they coded for the entire link protein plus a short 5′-untranslated sequence (10, 29). These cDNA clones were missing the most 5′-untranslated sequence and the genomic clones did not contain the first exon. In order to obtain a full length cDNA clone, a new cDNA library was constructed by extension

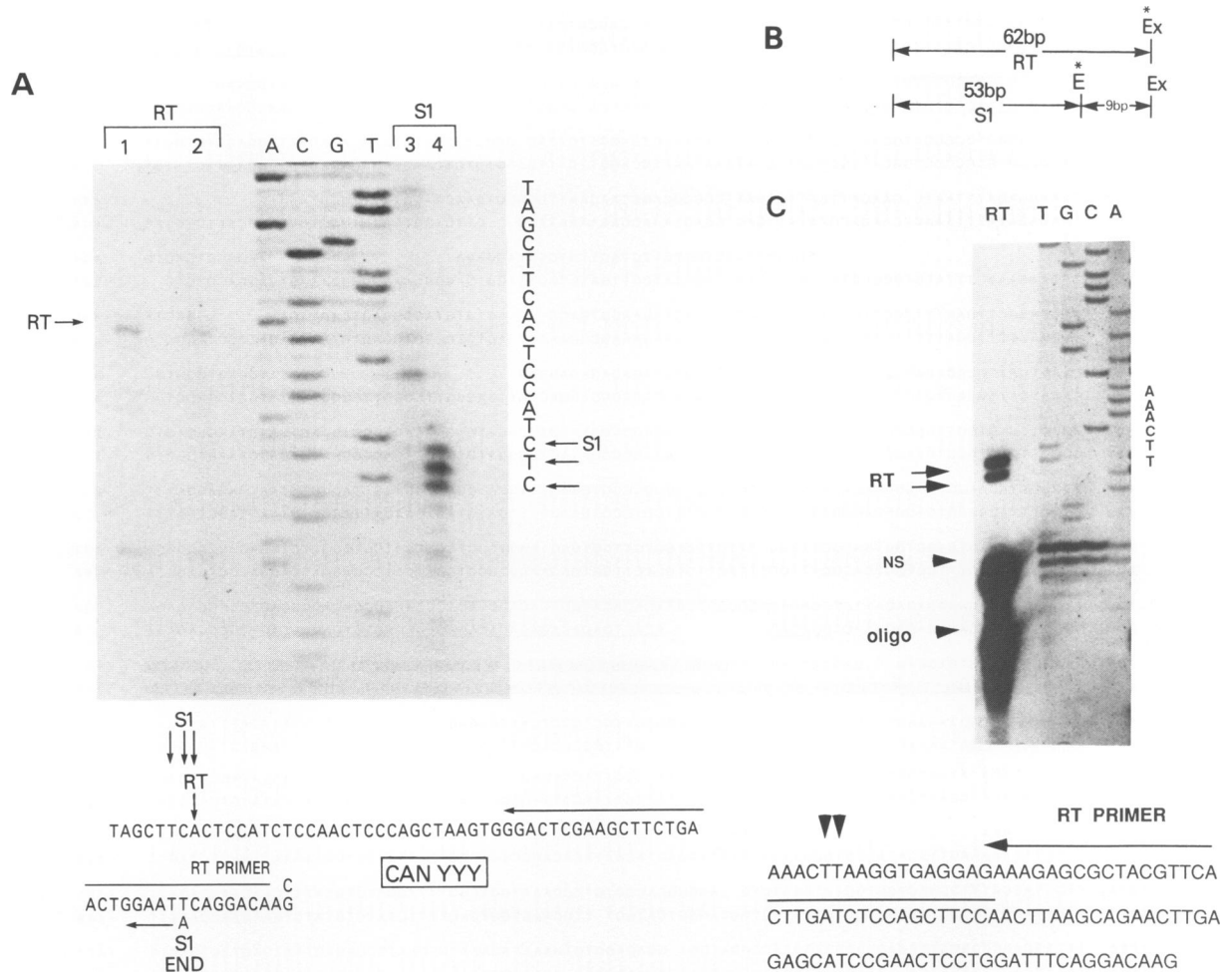

Fig. 2. S1 nuclease and primer extension analysis of the rat link gene first exon and primer extension of the human link gene first exon. **A.** Primer extension reactions (RT) were carried out using a 35-mer oligonucleotide (based on cDNA sequence), and 15 μg total RNA from rat chondrosarcoma, lanes 1 and 2 . S1 nuclease mapping, lanes 3 and 4, was done using a *Xho*I-*Eco*RI fragment labeled at the *Eco*RI site by polynucleotide kinase. The DNA was then hybridized to RNA and incubated with 1 unit/ml S1, lane 3, and 10 unit/ml S1, lane 4. A sequencing reaction was performed using as a sequencing primer the same oligo as used for the primer extension reactions. The sequencing template used was a plasmid containing the region. A 7 M urea/ 8% acrylamide sequencing gel was used for the electrophoresis. The DNA sequence of the region is shown. Arrows denote major products of the S1 nuclease protection assay or the primer extension product. Note the end of the DNA fragment used in the S1 analysis is closer to the start of transcription than is the oligonucleotide used for primer extension. The canonical cap sequence is boxed. Y represents a pyrimidine, while N represents any base. **B.** A diagram of the region showing the relationship of the S1 nuclease and reverse transcriptase products. (Ex) Denotes the end of the exon; (E), denotes the EcoRI site in the exon; (*), represents the ^{32}P labeled position. **C.** Primer extension reactions (RT) were carried out using a 35-mer oligonucleotide (based on cDNA sequence), and 15 μg total RNA from human fetal cartilage. (NS) denotes non-specific; oligo, denotes the 35 base position.

of a specific oligonucleotide primer from the link cDNA sequence (nucleotides 396−419, (10)). The library was screened using a second oligonucleotide from the most 5′ cDNA sequence and several clones were isolated, which included 10−25 nucleotides of additional 5′ sequence. The longest previous clone was terminated at an internal *Eco*RI site, of which only fifteen bases extended into the first exon. Based upon the sequence of the longest cDNA obtained, a 35-mer (nucleotides +62 to +28, Fig. 3) was prepared for screening a rat genomic library. A single clone containing two *Eco*RI inserts, 9 kb and 2.2 kb was isolated (Fig. 1, B). Because there was an *Eco*RI site within the oligonucleotide used in the screening, one *Eco*RI site was predicted to be within the first exon.

Isolation of the Human Link Genomic Clones

Human link protein genomic clones encoding the first exon and flanking sequence were obtained by screening a human genomic library with a synthetic oligonucleotide constructed from the most 5′ sequence of human link protein cDNA. A cDNA sequence for human link protein has been determined by us (data not shown) and others as previously reported (30). Three distinct overlapping genomic clones were obtained and were characterized by both restriction enzyme analysis and Southern blot hybridization. Using the oligonucleotide as probe, blot hybridization analysis allowed the identification of a 9 kb *Eco*RI fragment. This was subcloned into Bluescript and a restriction endonuclease map was prepared (Fig 1, C).

Determination of the Transcriptional Initiation Sites

S1 nuclease protection and primer extension were performed to determine the transcriptional start site for the first exon of the rat link protein gene. Total cellular RNA from rat chondrosarcoma was used to protect the *Xho*I-*Eco*RI fragment ^{32}P-labeled at the *Eco*RI site (Fig. 2, B). Three major protection

Fig. 3. Nucleotide sequence comparison of the 5'-end of the rat and human link genes. DNA sequence from the human and rat link genes was compared using the University of Wisconsin DNA analysis programs (32) and displayed using DNA:Draw (33). The top and bottom sequence rows correspond to sequences for the rat and human, respectively. Numbering on the right column is relative to the start site of transcription (+1). Left column numbering is relative to the beginning of the sequence shown. Vertical lines indicate nucleotides of identity between rat and human. Periods in the sequence represent insertion/deletion bases. The underlined regions near the beginning of the sequences are long inverted repeats repeated before the start site of transcription (see text). The inverted repeats are present in both rat and human as shown. The rat and human first exons are in reverse font. A putative TATAA box homolog for the rat gene is boxed.

products were obtained, (Fig. 2, lane 4), corresponding to lengths of 53, 54, and 55 bases (Fig. 2, B). Reverse transcriptase extension with an oligonucleotide primer produced a major cDNA of 62 b (Fig. 2, lanes 1 and 2). The difference in size between the primer-extension products and S1-protected fragments products was due to the utilization of different labeled ends; the *Eco*RI site labeled in the S1 protection assay is 9 bases closer to the start site of transcription than the oligonucleotide used in the primer extension (Fig 2, B). The site defined as the first transcribed base is an A following a C in close agreement the most frequent eukaryotic CAP site CANYYY (31). The agreement in size of the primer-extended product and the S1 nuclease protected fragment product, excluded both the possibility of S1 nuclease detecting an intron boundary and the possibility of an additional exon.

Primer extension was also performed to determine the transcriptional start site of the human link protein gene (Fig. 2, C). The oligonucleotide, a 35-mer, and a non-specific product (which was detected in a reaction without RNA) are indicated with the two specific primer extension products. The same oligonucleotide was used in the sequencing reaction (Fig 2. lanes T,G,C,A).

Nucleotide Sequence of the Rat and Human Link Protein Promoter and First Exon

The nucleotide sequence of the 5'-end of the rat and human link protein genes was determined and is shown in figure 3. The first exon in both genes coded for the untranslated sequence, while the translation initiation codon was in exon 2. No consensus TATAA box (34) was observed while a sequence, ACTTAA, occurred at −30 of the rat sequence. ACTTAA is similar to the sequence, CCTTAA, which has been described in the EIIa transcription unit of adenovirus as a TATAA sequence homolog (35). However, analysis of human fetal cartilage RNA indicates

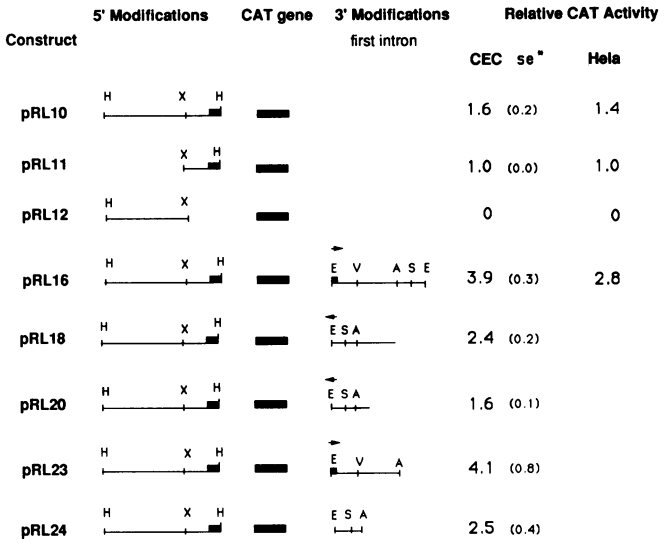

Fig 4. Transcriptional analysis of the rat link promoter and first intron. As diagramed, several fragments from the 5'-flanking region of the link protein gene were cloned into the unique *Hin*dIII site of pSV0CAT. Intron fragments were cloned into the unique *Bam*HI site, distal to the CAT gene. Constructs were transfected into Hela cells and chick embryonic chondrocytes (CEC). Some constructs were tested only in CEC. CAT assays were performed after 48 h incubation. The activity was expressed in all cases as a percentage of the activity obtained with a 850 bp active promoter construct, pRL11. Transfections were performed in duplicate, in three separate experiments. se*, indicates the standard error. Constructs pRL10, PRL11, and pRL16 were transfected into CEC more than eight times.

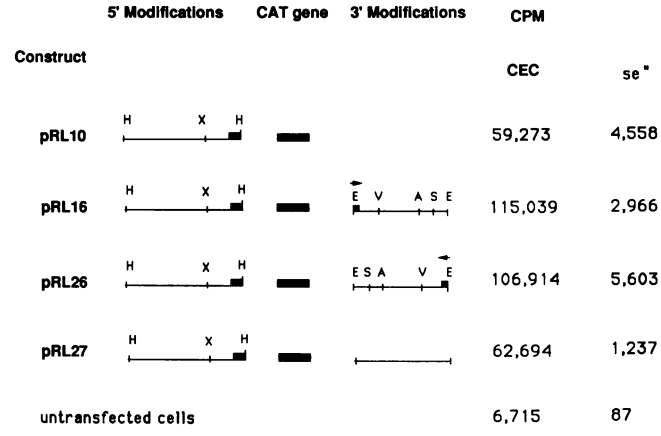

Fig 5. The first intron of the link gene increases activity of the promoter in an orientation independent manner in CEC. As diagramed plasmids pRL16 and pRL26 containing both orientations of the rat link first intron, were compared with the promoter alone construct and with a plasmid pRL27. pRL27 containing a similar in size 3' insert from the first intron of the mouse α1 (IV) collagen chain gene (2.2 kb for pRL16 and pRL26 and 2.3 kb for pRL27). Each number for pRl6, pRL26, and pRL27 is the mean of three transfections in duplicate (6 values) while pRL10 is the mean of 4 values. Statistical analysis by Scheffe F-test indicates a higher than 99% significance to the differences between pRL16 or pRL26 and pRL10 or pRL27. The differences for pRL16 and pRL26 or pRL10 and pRL27 are not significant. se*, indicates the standard error.

	Sequence	Gene	Location
A			
	ccctcccttcctc	Human link promoter	-267
	ccctcccttcccc	Rat link promoter	-258
	ccatcccttccc	Cartilage matrix protein	63
	ccctcccttctaa	Type II collagen promoter	-341
	ccatcccttcccc	Fibronectin	12
B			
	ggagagctgtaaa	Rat link first intron	+422
	ggagagctgtgaa	Type II collagen enhancer	346
	ggaaaggtgagga	Human link first intron	+540

Fig 6. Regions of similarity within the 5'-flanking regions of several cartilage associated proteins. A. 5' flanking regions for type II collagen (43), cartilage matrix protein (44), and fibronectin (45); genes were compared using the University of Wisconsin bestfit program. Sequence numbering is the same as in the listed references. **B.** Short regions of similarity within the collagen II enhancer, the rat first intron, and the human first intron.

this region as the start site of transcription in humans (Fig. 2, C). A GC-boxes (36) was present in the human at −548 that was not present in the rat. An Ap1-like binding site, AGTCATCT (37), and a c-AMP responsive element (CRE), TGACG (38), occurred at rat sequences −315 and −192, respectively, and were conserved in the human sequence. This AP1-like site was shown to bind to a protein, *in vitro*, that interacts with a Fos specific antibody (unpublished results).

Two regions in the human link promoter and one region in the rat had the potential to form left-handed Z-DNA which is characterized by alternating purine and pyrimidine bases. G-C repeats, in particular, have been shown *in vitro* to form Z-DNA (39). Short G-C repeats were observed in the rat DNA sequence at −810, and in the human sequence at −900 after 26 bp of $d(G\text{-}T)_{13}$:$d(C\text{-}A)_{13}$. Computer stem-loop analysis indicated the repeating purine-pyrimidine bases occurred in two large inverted repeats of approximately 40 bp in both the rat and human sequences (underlined sequences in Fig. 3). Pairing of distantly placed inverted repeats might form large cruciform structures, however, these forms may be thermodynamically unstable (40). Within the rat second inverted repeat (−80 to −45), an imperfect repeat sequence (CACACCC), of the −885 $d(G\text{-}T)_{13}$:$d(C\text{-}A)_{13}$, occurred several times. The human sequence maintained a more perfect inverted repeat, consisting of repeating CACA in this same region. A sequence similar to CACACCC, GCCACACCC is an important cis-acting transcriptional element of globin promoters (41, 42).

Comparison of the human and rat promoter sequences revealed two regions which had greater than 90% identity: the first 70 bases of figure 3 (−913 to −844), and between bases −90 and −320 of the rat DNA sequence. The region between −30 and the start site of transcription, as determined for the rat link protein gene, contained many base changes between the two species. The untranslated sequences of exon 1 (exon 1 is completely untranslated) appeared only moderately conserved between rat and human genes. In addition, the human gene contained several insertions in the 5' flanking region, which made it approximately 100 bp longer than the corresponding region of the rat gene.

Transcriptional Activity

Activity of the promoter of the rat link protein gene was analyzed by transient transfection of CAT constructs (Fig. 4 and Fig. 5). A construct containing 7 kb of the 5'-flanking sequence, pRL10, contained a medium strength promoter in CEC cells. The activity in HeLa cells, on a per cell basis, was approximately 15% of the activity obtainable in CEC. A small construct pRL11 (−850

to +35), also revealed substantial CAT activity while a negative control CAT plasmid pRL12 (−7000 to −850, missing −850 to +35 the transcriptional start site) had no CAT activity above untransfected cells alone. Because several positive cis-acting transcriptional elements have been localized to first intron DNA, a 2.2 kb DNA fragment from the first intron fragment of rat genomic DNA was inserted into pRL10 generating pRL16. CAT activity of pRL16 was increased approximately two-fold in each cell type.

Transfections and CAT analyses were also performed with constructs including other restriction endonuclease fragments from the first intron region. pRL23, containing a 1.3 kb 5′-intron sequence, maintained the CAT activity of pRL16. pRL18, pRL20, and pRL24, which contain various lengths of the 3′ portion of the 2.2 kb intron fragment, did not show a significantly higher CAT activity when compared to the promoter-alone CAT construct pRL10; however, pRL18 produced an intermediate CAT activity through many assays (Fig 4, pRL18). Constructs pRL21and pRL22 (see Materials and Methods), which contained fragments from the most 3′-region of the rat first intron, were also tested but did not contain detectable positive elements as the CAT activity observed was again similar to pRL10 (data not shown).

A plasmid with the first intron fragment in reverse orientation (pRL26), remained similar in activity to pRl6 (Figure 5). A control plasmid, pRL27, having a 2.3 kb DNA fragment from the first intron of mouse α1 (IV) collagen chain gene inserted into the BamHI site of pRL10 did not reveal any increase in activity and remained similar in activity to the promoter alone construct.

DISCUSSION

Using oligonucleotide probes from the 5′-portion of the link protein cDNA we obtained one rat and three human genomic clones. Primer extension and S1 mapping identified the transcription initiation site of the rat link protein gene in rat chondrosarcoma. The first exon of the rat clone encoded an untranslated sequence of 62 bp. Although the first exon of the human gene was predicted to be 67 bp long (by comparison to the rat sequence), our sequence analysis of the human cDNA indicated at least a 85 bp exon and specific primer extension a 95 or 96 bp exon. DNA sequence analyses revealed that the 5′-flanking sequence of the rat and human link protein genes are highly conserved and this similarity extends 5′ past base −1000 and 3′ past +600.

Computer comparison (32) of the 5′-flanking regions for the rat link protein, type II collagen gene (43), and the cartilage matrix protein gene (44), indicated no large stretches of DNA identity. However, a 12 bp sequence from the link promoter was found in the promoter region of type II collagen, the fibronectin gene (45), and the cartilage matrix protein 5′ regions (Fig. 6). Gel-shift analyses of this sequence indicated binding of a sequence-specific protein that is unaffected by retinoic acid and occurs in cell types other than chondrocytes (data not shown).

Two large inverted repeat sequences consisting of alternating purine-pyrimidine occurred in the human and rat genes. Regions of repeating purine-pyrimidine, potential left-handed Z-DNA, occur frequently in the mammalian genome (46), and long inverted repeats are observed in bi-directional promoters. The exact purine-pyrimidine repeat was not maintained in the second DNA repeat sequence of rat, and diverged from the human sequence. Z-DNA has also been reported in the rat prolactin gene and in subsequent CAT analysis the Z-DNA functioned to down regulate transcription (47).

Transcriptional activity of the rat link protein promoter was analyzed by expression of chimeric CAT plasmids in transfected cells. These experiments indicated that a 0.85 kb as well as a 7 kb 5′-flanking fragment was transcriptionally most active in CEC which produce link protein. The CAT activity of these constructs in HeLa cells, which do not synthesize link protein, was approximately 15% of the CAT activity obtained in CEC. When DNA from the first intron was included with the 7 kb link protein promoter construct, CAT activity increased approximately two-fold. This transcriptional effect was orientation independent and was not merely due to DNA size or a spacing effect because a construct containing an unrelated 2.3 kb insert had no increase in activity.

ACKNOWLEDGEMENTS

We thank Drs. Peter Burbelo, Hynda Kleinman, Leslie Bruggeman, Maura Kibbey, and Vincent Hascall for helpful discussion concerning the manuscript. We thank Dr. Victoria Porter for computer assistance and Dr. Jeff Kopp for assistance with statistical analysis. We thank Binh Nguyen for assistance during the isolation of the genomic clones. S.L. is supported by Fundacao De Amparo A PesQuisa Do Estado De Sao Paulo, Proceso No. 89/0148-7.

REFERENCES

1. Keiser, H., Shulman, H.J., and Sandson, J.I. (1972) *Biochem. J.* **126**, 163−169
2. Hascall, V.C., and Heinegard, D. (1974) *J. Biol. Chem.* **249**, 4232−4241.
3. Perin, J.P., Bonnet, F., and Jolles, P. *FEBS Lett.* **94**, 257−260
4. Perin, J.P., Bonnet, F., Pizon, V., Jolles, J., and Jolles, P. (1980) *FEBS Lett.* **119**, 333−336
5. Pal, S., Strider, W., Margolis, R., Gallo, G., Lee-Huang, S., and Rosenberg, L. (1978) *J. Biol. Chem.* **253**, 1279−1289
6. Roughley, P.J., Poole, A.R., and Mort, J.S. (1982) *J. Biol. Chem.* **257**, 11908−11914
7. Caterson, B., Baker, J.R., Christner, J.E., Lee, Y., and Lentz, M. (1985) *J. Biol. Chem.* **260**, 11348−11356
8. De Luca, S., Heinegard, D., Hascall, V.C., Kimura, J.H., and Caplan, A.I. (1977) *J. Biol. Chem.* **252**, 6600−6608.
9. Lohmander, L.S., Fellini, S.A., Kimura, J.H., Stevens, R.L., and Hascall, V.C. (1983) *J. Biol. Chem.* **258**, 12280−12286
10. Rhodes, C., Doege, K., Sasaki, M., and Yamada, Y. (1988) *J. Biol. Chem.*, **257**, 11908−11914
11. Doege, K., Rhodes, C., Sasaki, M., Hassell, J.R., and Yamada, Y. (1990) in *Matrix Genes* , Academic Press
12. Neame, P.J., Perin, J.-P., Bonnet, F., Christner, J.E., Jolles, P., and Baker, J.R. (1986) J. Biol. Chem. **260**, 12402−1240
13. Doege, K., Sasaki, M., Horigan, E., Hassell, J.R., and Yamada, Y. (1987) *J. Biol. Chem.* **262**, 17757−17767
14. Heinegard, D., and Hascall, V.C. (1974) *J. Biol. Chem.* **249**, 4250−4256
15. Hascall, V.C. (1988) ISI Atlas of Science:Biochemistry **1**,189−198
16. Mallein-Gerin, F., Kosher, R.A., Upholt, W.B., and Tanzer, M.L. (1988) *Dev Biol* **126**, 337−345
17. McKeown-Longo, P.J., Velleman, S.G., and Goetinck, P.F. (1983) *J. Biol. Chem.* 258, 10779−10785
18. Tsonis, P.A., and Goetinck, P.F. (1988) *Exp. Eye Res.*, **46**, 753−764
19. Horton, W. E., Cleveland, J., Rapp, U., Nemeth, G., Bolander, M., Doege, K., Yamada, and Hassell, J.R. (1988). *Exp. Cell Res.* **178**, 457−468
20. Gorman, C.M., Moffet, L.F., and Howard, B. (1982) *Mol. cell. Biol.* **2**, 1044−1051
21. Smith, B.D., Martin, G.R., Miller, E.J., Dorfman, A., and Swarm, R. (1975) *Arch. Biochem. Biophys.* **166**, 181−186
22. Benton, W.D., and Davis, R.W. (1977) *Science* **196**, 180−182

23. Killen, P.K., Burbelo, P.D., Martin, G.R., and Yamada, Y. (1988) *J. Biol. Chem.* 263: 12310–12314
24. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci., U.S.A.* **74**, 5463- 5467
25. Berk, A. J., and Sharp, P.A. (1978) *Proc. Natl. Acad. Sci., U.S.A.* 75, 1274–1278
26. Horton, W., E., Myashita, T., Kohno, K., Hassell, J.R., and Yamada, Y. (1987) Proc. Natl. Acad. Sci., U.S.A. 84, 8864–8868
27. Graham, F.L., and Van der Eb, A.J. (1973) *Virol* 52, 456–467
28. Neumann, J. R., Morency C. A. and Russian K. O. (1987) *Biotechniques* 5, 444–447
29. Kiss, I., Deak, F., Mestric, S., Delius, H., Soos, J., Dekany, K., Argraves, W. S., Sparks, K. J., and P. F. Goetinck (1987) *Proc. Natl. Acad. Sci., U.S.A.* **84**, 6399–6403
30. Dudhia, J., and Hardingham, T. E. (1988) *J. Mol. Biol.* **206**, 749–753
31. Corden, J., Wasylyk, B., Buchualder, A., Sassone-Corsi, P., Kedinger, C., and Chambon, P.(1980) *Science* **209**, 1406–1414
32. Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucl. Acid. res.* 12, 387–395
33. Shapiro, M. (1990) *Binary Computing in Microbiology* **2**, 187–190
34. Goldberg, M.L. (1979) Ph.D. Thesis Stanford University, Stanford, California
35. Zajchowski, D.A., Boeuf, H., and Kedinger, C. (1985) *EMBO J.* **4**, 1293–1300
36. Gidoni, D., Dyan, W. S., and Tijian, R. (1984) *Nature* **290**, 304–310
37. Montminy, M.R., Serarino, K. A., Wagner, J. A., Mandel, G. and Goodman, R. H. (1986) *Proc. Nat. Acad. Sci, U. S. A.* **83**, 6682–668
38. Lee, W., Mitchell, P., and Tijian, R. (1987) *Cell* **49**, 741–752
39. Haniford, D. B., and Pulleyblank. (1983) *Nature* **302**, 632–634
40. Jaworski, A., Wang-Ting, H., Blaho, J. A., Larson, J. E., and Wells, R.D. (1987) *Science* **238**, 773–777
41. Dierks, P., Ooyen, A.V., Cochran, M.D., Dobkin, C., Reiser, J., and Weissman, C. (1983) *Cell* 32, 695–706
42. Myers, R.M., Tilly, K., and Maniatis, T. (1986) *Science* **232**, 613–618
43. Kohno, K., Sullivan, M., and Yamada, Y. (1985) *J. Biol. Chem.* **260**, 4441–4447
44. Kiss, I., Deak, F., Holloway, R. G., Delius, H., Mebust, K. A., Frimberger, E., Argraves, W. S., Tsonis, P. A., Winterbottom, N., and Goetinck, P. F. (1989) *J. Biol. Chem.* **266**, 8126–8134
45. Dean, D. C., Bowlus, C.L., and Bourgeois, S. (1987) *Proc. Nat. Acad. Sci, U.S.A.*, **84**, 1876–1880
46. Hamada, H. M., Petrino, M. G. and Kakunaga, T. (1982) *Proc. Natl Acad. Sci. U.S.A.* **79**, 6465–6469
47. Naylor, L. H., Clark, E. (1990) *Nucl. Acids Res.* **15**, 1595–1601