A conserved structural element in horse and mouse IGF2 genes binds a methylation sensitive factor

Kerstin Otte1,*, Devapriya Choudhury2, Marika Charalambous3, Wilhelm Engström1,4 and Björn Rozell1

1Department of Pathology, Swedish University of Agricultural Sciences, PO Box 7028, 75007 Uppsala, Sweden, 2Department of Molecular Biology, Swedish University of Agricultural Sciences, PO Box 590, 75124 Uppsala, Sweden, 3Developmental Biology Programme, Department of Biochemistry and Biology, University of Bath, Bath BA2 7AY, UK and 4Centre for Reproductive Biology, Swedish University of Agricultural Sciences, PO Box 7038, 75007 Uppsala, Sweden

Received January 2, 1998; Revised and Accepted February 18, 1998 DDBJ/EMBL/GenBank accession nos AF020598, AF020599

ABSTRACT

The equine IGF2 gene has been cloned and characterised. It spans a 9 kb region, which is substantially less than the corresponding human gene. Three coding exons and three untranslated leader exons, all highly homologous to those in other species, were identified. Downstream of the polyadenylation site in exon 6, a dinucleotide repeat sequence was identified. Three putative promoters (P1–P3) were localised in the 5′ region of the gene. RNase protection analysis revealed two active promoters in fetal tissues, P2 and P3, whereas P3 was the only promoter active in adult tissues. This represents a transcriptional pattern different from that in humans or rodents. A novel structural element, an inverted repeat, is predicted in the 3′ region of the IGF2 gene. This repeat is conserved between species and located in a region which is differentially methylated in the human and mouse genes and might therefore be involved in the imprinting mechanism. The inverted repeat acquires a stem-loop structure in vitro with a hybrid A/B-DNA conformation in the stem area. Both in horse and mouse, a methylation-sensitive protein binds this structure with a strong requirement for the loop area. Furthermore, the protein might be developmentally regulated.

INTRODUCTION

Insulin-like growth factor 2 (IGF2) is a small mitogenic peptide (1) and is one of the most ubiquitous growth factors in the mammalian embryo, where it plays an important role in regulating fetal growth. This was demonstrated when transgenic mice with a disrupted IGF2 gene showed retardation of fetal growth (2). The subsequent finding that the IGF2 gene is parentally imprinted (3), focused much interest on its genomic structure. Furthermore, relaxation of imprinting leads to altered expression of the gene and has been implicated in the pathogenesis of neoplastic and non-neoplastic overgrowth conditions (reviewed in ref. 4).

The IGF2 gene shows a complex structural organisation in all species analysed. It consists of at least nine exons in man and sheep, and six exons in rat and mouse (5–13). Its expression is regulated in a developmental and tissue-specific manner (14), involving differential promoter usage and alternative splicing (15), as well as differential usage of an RNA processing site (16,17). Translatability of the different promoter transcripts is variable and growth-dependent (18–20) and the translated product is also subject to posttranslational modification (21).

The IGF2 mRNA population originates from the use of four promoters in man and sheep (6,7,22–24) and three promoters in rodents (8,10,13). During fetal life three promoters are active both in human and rodents, with promoter P3 in humans and promoter P3 in rodents (which corresponds to P4 in human) being predominantly used. Transcription from these promoters is repressed during adult life and a fourth promoter becomes activated in human liver (22,23,25–27). No homologue to human promoter P1 has been identified in rodents, but is present in the ovine and baboon IGF2 genes (7,28). Furthermore, imprinted antisense transcripts are expressed in the mouse IGF2 gene (29).

IGF2 is subject to epigenetic modification (30–33) and parental imprinting in human, mouse and rat, which causes a predominant expression from the paternal allele (34–37). The molecular basis of parental imprinting is still unknown, but DNA methylation might be one of the mechanisms involved. Expression of imprinted genes has been altered in transgenic mice deficient in DNA methyltransferase (38) and parent-specific methylation patterns have been detected in two regions of the human and mouse IGF2 genes (30–33). Despite the main focus on methylation of cytosine residues, repeat sequences and alterations in chromatin structure have also been suggested as regulatory features in the imprinting mechanism (39,40).

We have focused our attention on the equine IGF2 gene and have previously demonstrated that its expression is developmentally regulated (41). In this study, we present the genomic sequence of the equine IGF2 gene and characterise its structural...
and functional properties. The examination of promoter usage of the equine IGF2 gene revealed a type of regulatory mechanism which is substantially different from that in human and rodent IGF2 genes. Furthermore, a novel structural element, an inverted repeat, was identified in a region which is differentially methylated in the human and mouse genes and therefore thought to be involved in the imprinting mechanism. This inverted repeat is conserved between species and is proposed to form a stem-loop structure. We show that both equine wild type and methylated sequences acquire a stable intramolecular duplex in vitro with a mixed A/B-DNA conformation in the stem region. Both in horse and mouse, this stem-loop is bound by a specific, methylation sensitive protein which requires presence of the loop area for binding. These findings may provide new important clues as to how structural elements and allele specific methylation might be involved in imprinting control.

MATERIALS AND METHODS

Equine IGF2 gene cloning

An equine (arabian) genomic phage library in λ vectors (FIX 11) was obtained from Stratagene and was screened with a PCR amplified equine IGF2 cDNA probe (41). The cDNA was random prime labelled with [α-32P]ATP using the Amersham Rediprime kit according to the manufacturer’s instructions. Subsequently, the library was rescreened with an equine IGF2 genomic subclone (eqIGF2-Sac03). Plaque purification and DNA preparation followed standard methods (42). The identity of the recombinant clones (eqIGFII-D and eqIGFII-I) was determined by Southern blot analysis with the equine cDNA. An restriction map was obtained by single and double digests with EcoRI, SalI and XhoI (Fig. 1). The presence of exons was detected by Southern blot analysis using human IGF2 exon specific oligonucleotides. The oligonucleotides were end-labelled with digoxigenin according to instructions provided by Boehringer Mannheim. Exon-containing fragments were sub cloned into a plasmid vector (pBluescript II KS+; Stratagene) according to standard procedures (42). The following clones were obtained: eqIGF2-SalCA, eqIGF2-Eco21, eqIGF2-Eco24, eqIGF2-EcoSal, eqIGF2-XbaSal and eqIGF2-Sac03 (Fig. 1).

Determination of nucleotide sequence

Sequencing of the different subclones was performed on an Applied Biosystems Instrument 377 using universal M13 and internal primers. The sequence data were processed with the Genetics Computer Group (Madison, WI) package of the HGMP, Cambridge, and compared with the GenBank data base. Sequences of the subclones are deposited in GenBank under the accession numbers AF020598 and AF020599.

RNA isolation and preparation of labelled RNA probes

Liver and kidney were obtained from five equine foetuses of 250 days gestation and from five adult horses. Total RNA was prepared using the TRIzol® reagent (Gibco). The concentration of RNA was determined by spectrophotometry. RNA probes were prepared by in vitro transcription using T3 and T7 RNA polymerases provided with the SP6/T7 Transcription Kit (Boehringer Mannheim) with [α-32P]UTP (Amersham) according to the manufacturer’s protocol. Probes were labelled to a specific activity of ∼5 × 10⁸ c.p.m./μg. The exon-specific antisense riboprobes were generated from pBluescript KS+ (Stratagene). The following subclones were used: probe 1, a SacI fragment covering bases 458–775 in AF020598 linearized with EcoRI; probe 2, an EaeI/NotI fragment covering bases 1340–1483 in AF020599 linearized with EcoRI; probe 3, a SacII/EcoRI fragment covering bases 1953–2879 in AF020599,

Figure 1. Physical map and gene organisation of equine IGF2. (A) Two overlapping λ clones (eqIGFII-D and eqIGFII-I) and resulting subclones are indicated at the top, recognition sites for several restriction enzymes below. (B) Exon organisation, location of promoters, poly A site, (TCC) deletion, microsatellite and the inverted repeat are shown. Coding exons are represented as filled boxes and non-coding exons by striped boxes. The sequenced region is shown as a bold line with an interruption in the 5’ region. The exon contribution to the deduced pre-pro-peptide is shown at the bottom, with the shaded area indicating the mature protein.
linearized with SacI. The 3' protruding ends generated by linearization with SacI were filled in using Klenow DNA polymerase (Promega) to avoid promoter-independent transcription from the ends of the template.

**RNase protection analysis**

The promoter usage was analysed by annealing exon-specific, antisense oligonucleotides. The promoter usage was analysed by annealing exon-specific, antisense oligonucleotides. Each spectrum was the average of 16 accumulations.

**Thermal melting studies of oligonucleotides**

Two oligonucleotides corresponding to the equine inverted repeat sequence 5'-GGGGAGACCGGAGAAGAGGAAGGATGGTTTTTG-3' and 5'-GGGTGCCTCACTCTCACAAGCTGAC-3' were synthesised (CyberGene). In one of these, 5-methyl cytosine was incorporated at positions 9, 25 and 59. Thermal melting studies were carried out on these two oligonucleotides in a Perkin-Elmer Lambda 40, UV-Vis spectrophotometer, with 1.5 ml cuvettes. The samples were dissolved in 10 mM PIPES buffer (pH 7.0) and were first heated to 94°C and gradually cooled to room temperature in order to facilitate secondary structure formation. During melting experiments the samples (0.15 mM) were heated from 15 to 80°C at a heating rate of 1°C/min and with a lag time of 1 min between each heating step. Absorbance at 260 nm was monitored as a function of temperature throughout the experiments. The melting temperature (T_m) was determined from the peak value of derivative plots. Both the forward and the reverse transitions were monitored and were not found to differ significantly.

**Circular dichroism spectroscopy (CD)**

CD spectra were recorded using a JASCO J-720 instrument. The samples were dissolved in sodium cacodylate buffer (0.01 M, pH 7.0). A 0.5 cm quartz cuvette was used, temperature was maintained at 20°C using a thermostated cell holder. Spectral data were collected between 200 and 320 nm, using a step resolution of 0.2 nm, response time of 2 s, and a bandwidth of 1 nm. Each spectrum was the average of 16 accumulations.

**Preparation of protein extracts**

Nuclei were prepared from whole liver and brain from 3 day old mice (of a mixed C57BL/CBA strain background) using a modification of the method used by Sasaki et al. (30). Organs were dissected into PBSA on ice, and rinsed in PBSA several times to remove blood clots. Crude preparations of nuclei were obtained by homogenisation of single liver samples in 200 μl ice-cold nuclear buffer [15 mM NaCl, 5 mM MgCl₂, 60 mM KCl, 0.1 mM EGTA, 15 mM Tris-HCl (pH 7.4), 15 mM DTT, 0.1 mM PMSF] containing 0.3 M sucrose, 10% glycerol; using an Ultraturrax T8 homogeniser. The homogenate was overlaid on a 1 ml sucrose cushion (1.2 M sucrose/5% glycerol in a microfuge tube). Samples were then spun for 20 min in a microcentrifuge at 4000 r.p.m. at 4°C. Pelleted nuclei were resuspended in nuclear buffer/0.3 M sucrose/10% glycerol and stored at −70°C. Protein extracts from equine fetal liver, kidney, adult liver (same animals as previously described) and mouse nuclei were prepared by homogenisation of tissue in 20 mM HEPES (pH 7.5), 1 mM PMSF, 2 mM DTT, 400 mM KCl, 20% glycerol and centrifugation at 20 000 g for 1 h. Protein concentrations were determined by spectrophotometry (43).

**DNA binding assays**

The following oligonucleotides were used: equine inverted repeat sequence (see section Thermal melting studies of the oligonucleotides); mouse inverted repeat sequence, 5'-GGGGAGACCGGAGAAGAGGAAGGATGGTTTTTG-3' and 5'-GGGTGCCTCACTCTCACAAGCTGAC-3'; mouse stem, 5'-GGGGAGACCGGAGAAGAGGAAGGAGGAG-3' and 5'-CTCACTTCCTCCCTCCCGTGTTGCCG-3'; mouse 5' half-site, 5'-GGGGAGACCGGAGAAGAGGAAGGAGGAG-3' and 5'-CTCACTTCCTTCCTTCCCTGCCCTTTCCC-3'; mouse 3' half-site, 5'-CTCACTTCCTCCCTCCCGTGTTGCCG-3' and 5'-GGACAAGACCGGAGAAGAGGAGTGGAG-3'. Oligonucleotides were annealed in 10 mM HEPES by heating at 94°C for 5 min and cooling to room temperature over a period of 3 h. Radioactive labelling was performed using polynucleotide kinase and [γ-32P]ATP (Amersham). Protein extract (10 μg total cell protein and 5 μg nuclear protein) was incubated together with 30 fmol labelled oligonucleotide and 200 ng dIdC (Pharmacia Upjohn) for 15 min at room temperature in a buffer containing 10 mM HEPES (pH 7.5), (0.2 mM EDTA, 10% glycerol, 2.5 mM MgCl₂, 100 mM KCl and 2 mM DTT. Competition was carried out by adding 100× molar excess of native or methylated unlabelled oligonucleotide, or unrelated annealed oligonucleotides of the sequence 5'-GGGGAGACCGGAGAAGAGGAAGGATGGTTTTTG-3' and 5'-GGGTGCCTCACTCTCACAAGCTGAC-3'. Oligonucleotides were separated from free DNA on a non-denaturing 5% polyacrylamide gel at 4°C. Gels were run at 200 V for 45 min, dried and exposed to autoradiographic film overnight.

**RESULTS**

**Screening of a genomic library**

Screening of ~5 × 10⁵ colonies of a non-amplified, equine genomic phage library with an equine IGF2 cDNA yielded two independent clones, eqIGFII-D and eqIGFII-I (Fig. 1). Southern blot analysis using human exon-specific oligonucleotide probes detected exon-containing fragments which were subcloned into phagemid vectors. Both phage clones represent largely the same genomic region and hence served as controls for each other. The cloned region spans an ~20 kb stretch of chromosomal DNA of which the equine IGF2 gene only represents 9 kb in length. In order to isolate additional 5' regions of the equine IGF2 gene, the library was rescreened with the 5' genomic subclone eqIGF2-Sac03 (Fig. 1). However, Southern blot analysis of these additional clones revealed similarity to the previously isolated ones (data not shown).

**Sequence and structure of the equine IGF2 gene**

Exon and intron sequences of the genomic subclones covering the equine IGF2 gene were determined and the overall exon/intron structure compared with that of other species. Sequence comparison with an equine IGF2 cDNA identified three coding exons in the 3' region of the gene (exons 4–6).
Comparison with transcribed IGF2 sequences from a variety of species located three additional non-coding exons in the more 5′ region (exons 1–3) (Fig. 1). No exons corresponding to human 5′ exons 1–3 could be identified. The coding as well as non-coding exons display a high degree of homology between a variety of species. Similarity of the coding exons is between 94 and 89%, while the 5′ exons are less well conserved with nucleotide identities between 90% and 75.

In the horse, the IGF2 intron regions seem to be generally shorter than in other species. Equine exon 3 is the shortest of the exons and consists of only 81 bp. The (TCC)_{10} repeat present in human and ovine corresponding exons is partly deleted, with only a (TCC)_{2} repeat remaining. A putative polyadenylation site flanks the coding region in the last exon. Furthermore, a microsatellite repeat with a length of ~600 bp is located downstream of the polyadenylation site. Sequence data are not available beyond this point and it is possible that a more downstream polyadenylation site exists as seen in the human IGF2 gene. Therefore, it cannot be excluded that the repeat area is present within the transcribed region as in other species.

The nucleotide sequence of the equine IGF2 gene predicts a growth factor precursor pre-pro-peptide of 181 amino acids, which consists of three parts; a 24 amino acid leader peptide, a 67 amino acid mature peptide and a 90 amino acid E-peptide (Fig. 1). The amino acid sequence of the pre-pro-peptide is 75% identical to its human counterpart (44) and the mature protein is even more conserved with 97% homology.

Promoter usage of equine IGF2 differs from that in humans and rodents

Sequence comparison with human and mouse genes revealed two putative promoter regions preceding equine exons 2 (P2) and 3 (P3). These two promoters are highly conserved in their sequences as compared with corresponding regions in human, rodent and ovine genes. An additional promoter might be located upstream of equine exon 1 (P1), as determined in human and rodent genes. This promoter, however, fails to show sequence homology between species.

Previously we have analysed equine IGF2 transcription by northern blotting (41). In the present study we have used the RNase protection assay as a more sensitive method to identify transcripts as well as to determine promoter usage during fetal and adult life. Promoter-specific probes were constructed (Fig. 2A) and used in RNase protection analysis on fetal and adult hepatic and non-hepatic tissues. In fetal liver and kidney, P2 and P3 are transcriptionally active, whereas transcripts derived from P1 could not be detected by RNase protection analysis. Apparently the most abundant transcript in fetal liver and kidney is derived from P3. Transcripts derived from P3 could be observed as two protected fragments of only slightly different size. Promoter usage in adult liver and kidney changes markedly; only transcripts from P3 are present. Therefore it seems that P3 is the main promoter in both fetal and adult tissues and that P2 is downregulated postnatally.
A conserved inverted repeat forms a stable intramolecular duplex

A novel element was identified in the equine IGF2 gene. It is located upstream of equine exon 5 and consists of two half-sites of 26 bp, separated by a 16 bp spacer, forming an inverted repeat (Fig. 3A). This structure displays a highly atypical distribution of nucleotides, one half-site consisting of a large excess of pyrimidines and the other dominated by purines. The inverted repeat is present in all species from which sequence information is available (human, rat and mouse) and its location and sequence is conserved (Fig. 3A), with only the spacer region displaying gross difference in length and sequence. Furthermore, it has the capacity to form a stem-loop structure as indicated in Figure 3B.

To investigate if the inverted repeat could form a secondary structure in vitro, thermal melting studies were carried out on an oligonucleotide corresponding to the equine element. Since the sequence contains three CpG dinucleotides, two of which are conserved (Fig. 3A), we also investigated whether methylation at these positions 6, 26 and 68, could influence the physicochemical properties. The melting curves for both oligonucleotides were monophasic (Fig. 4A). The Tm for the native sequence was found to be 67.5°C and that of the methylated counterpart 65.3°C at 10 mM PIPES. The high Tm and the monophasic melting curves are indicative of a highly stable intramolecular duplex structure, as any higher order intermolecular complexes and triplex structures are likely to show multiple melting transitions.

The stem-loop structure acquires an A/B hybrid form

Figure 4B shows the CD spectrum of the native and methylated equine oligonucleotides in the 200–320 nm region. The CD spectrum of both the oligonucleotides is almost superimposable onto each other. This indicates that methylation does not affect the structure of the oligonucleotide under the experimental conditions used. However the spectra are not conservative and show a very strong positive Cotton effect centred around 279 nm (molar ellipticity, 160 000 at 279 nm) and a somewhat weaker negative Cotton effect centred around 242 nm (molar ellipticity, −70 000 at 242 nm). This indicates that the structure has hybrid features of both A- and B-type conformations. The polypurine–polypyrimidine sequence of the oligonucleotides would suggest the possibility of intermolecular triplex formation (45). However, absence of a strong negative Cotton effect around 212 nm which is characteristic of Hoogsteen base pairing (46,47) rules out this possibility.
The stem-loop structure is bound by methylation sensitive protein factors

To investigate if specific proteins recognise and bind the postulated stem-loop structure, protein–DNA interactions were studied by gel mobility shift assays. This analysis revealed two distinct shifted bands when using self-annealed native oligonucleotide corresponding to the equine inverted repeat (Fig. 5, lanes 2 and 9). In adult liver, however, binding activity was reduced (lane 13). Competition experiments verified the specificity of the interaction (lanes 3, 4, 10 and 14). In order to determine whether methylation of CpG residues influences protein binding, additional experiments were carried out with the previously described modified oligonucleotide, and showed reduced binding in all tissues examined (lanes 5–7, 11, 12, 15 and 16).

To investigate if the interaction between the stem-loop and proteins is phylogenetically conserved, we extended these studies to the well characterised mouse IGF2 gene. Three distinct bands were revealed using neonatal liver nuclear extracts together with the mouse stem-loop (Fig. 6, lanes 2 and 3). Protein binding was reduced upon methylation of the CpGs in the stem area (lanes 4 and 5). Furthermore, protein binding is crucially dependent on presence of the loop as the stem only is weakly bound (lane 6). Only weak binding to the isolated half sites was observed (lanes 7 and 8). In addition, interspecies comparison revealed three shifted bands when mouse proteins were combined with the equine stem-loop, while equine proteins produced two shifted bands using the mouse sequence (lanes 9 and 10).
DISCUSSION

This study has focused on the structural properties of the equine IGF2 gene. By comparing the equine genomic IGF2 sequence with that of other animals including mouse, rat, sheep and man it was found that the homology in the coding region was very high. When the non-coding exons were studied the homology was slightly below that of the coding exons, but nevertheless significant. Intronic stretches were compared with those of other species, and while most of the intron sequences bore no significant homology there were some notable exceptions.

Based on sequence homology, this study has revealed two distinct promoter regions in introns 2 and 3 of the equine IGF2 gene (Fig. 1). The homologues to equine P2 and P3 show extensive conservation in structure and sequence in all mammals examined (7–9,13,22,23,25). An additional promoter region (P1) might be situated upstream of the first exon as seen in human and rodent genes, although as expected no sequence conservation could be identified in this region. The human and rodent counterparts to this putative equine P1 contain no known consensus promoter elements and show great heterogeneity in transcription start sites (6,10).

In all equine tissues examined by RNase protection, the putative P1 promoter appeared to be inactive. In equine fetal liver and kidney the P2 and P3 promoters were active and accounted for two different transcripts of which the P3 derived mRNA seems to be the most abundant. Transcripts derived from P3 could be detected as two species of slightly different length, suggesting that different transcription start sites are used. Furthermore, P3 was continuously active in adult tissues, whereas P2 was silenced and the corresponding mRNA undetectable. This is in agreement with our previous demonstration of two main fetal and one adult hepatic transcript (41).

The alternate promoter usage in horse is clearly different from that in mouse and man. All three murine promoters are used during embryonic and fetal life. After parturition all promoters are downregulated in most tissues and not reactivated at any stage of adult life. In man, there are three established promoters accounting for transcription during fetal life. These promoters continue to be active albeit at a much reduced level after birth. However, one promoter located at the 5′ end of the human gene is exclusively active in adult liver. The continued use of a fetal promoter after birth places the horse IGF2 gene in an intermediate functional position between human and mouse genes.

A novel structural element consisting of an inverted repeat was identified upstream of exon 5 in the equine IGF2 gene. When this element was subjected to interspecies sequence comparison, it was found to be very conserved among all species from which sequence information is available, namely mouse, rat, man and horse; differences are mainly seen in the intervening spacer region. Based on the high homology, a similar secondary structure can be proposed for all species. This element has all properties to acquire a stem-loop structure, which was confirmed by UV and CD analysis. The high Tm obtained even at relatively low salt concentration suggests highly stable structures at room temperature.

The unusual polypurine–polypyrimidine sequence of the inverted repeat suggests the possibility of a triple helical conformation (45). However, this appears to be unlikely since the 212 nm negative CD band characteristic of triple helices and other Hoogsteen base paired duplexes (46) was absent. Furthermore, the monophasic UV-melting profiles argue against the presence of such higher order structures. The CD spectrum of both methylated and unmethylated oligonucleotides are almost completely superimposable onto each other. This indicates that methylation of the three CpG sites does not significantly perturb the overall conformation of the stem-loop. The CD spectrum further shows that the structure of the duplex portion of the stem-loop is somewhat unusual in having features of both the canonical A and B conformations of DNA (48).

Both horse and mouse stem-loops produced a multiple banding pattern as a result of interaction with cellular or nuclear proteins. One of these bands was predominant in both species. At present we can only speculate on the character of these interactions, which may represent monomer or dimer binding. Another explanation may be different isoforms of the same binding protein or that results represent co-operative binding by several proteins.

To investigate the importance of the loop for protein-DNA interactions, oligonucleotides corresponding to only the stem were used and only weak protein binding was observed. Still weaker binding was seen with the individual half-sites. Thus, formation of the stem-loop structure is necessary for efficient protein binding and this interaction is crucially dependent upon presence of the loop area. This was also recently shown for DNA binding by DAX-1, a DNA-binding protein involved in steroidogenesis (49). The equine and mouse stem-loops differ mainly within the loop, although the more 5′ sequence of the area is conserved. To examine if the loop sequence is important for the interaction, we made interspecies combinations of DNA and protein extracts. The binding characteristics were retained in both systems. This might indicate a dependence of the protein on the loop structure or the conserved sequence in the 5′ part of the loop.

Furthermore, protein extracts from equine adult liver showed diminished binding activity, and therefore it appears that the protein factor is under developmental control.

Since the stem-loop is located in the vicinity of a differentially methylated region in mouse and human IGF2 genes, we examined the effect of methylation at CpG dinucleotides within the repeat element. Both in horse and mouse, protein binding was reduced upon methylation of CpG residues, two of which are in base-paired form in our predicted conformation (Fig. 3B). Since the overall structure of both oligonucleotides appears to be very similar, the protein might be sensitive to local conformational changes caused by methylation or by direct steric hindrance due to the presence of methyl groups.

The formation of a stem-loop or cruciform structure of double stranded DNA, which is bound by a protein component has interesting implications, particularly as it lies in a differentially methylated region of an imprinted gene. The transcribed paternal allele is methylated while the silenced maternal allele unmethylated on specific CpG residues (30–33). Mouse IGF2 is one of the most extensively characterised imprinted genes. In the DMR2 region the transcribed paternal allele is methylated while the maternal allele is unmethylated on specific CpG residues. Furthermore, this methylation pattern is tissue-specific and correlates directly with expression. Therefore the presence of a silencer element under epigenetic control has been suggested (33). Inverted repeats and cruciform structures have previously been associated with regulation of gene expression (49–52), as well as termination of transcription and attenuation (53). One possible role of the stem-loop structure presented in this study could therefore be to act as a structural silencer element which is...
recognised by specific protein factors depending on its methylation status. Proteins interacting with cruciforms have previously been reported, as, for example, the HMG box proteins CBP and DAX-1 (49,54). All these proteins would be potential candidate peptides to recognise the presented stem-loop structure in human, rodent and horse IGF2 genes.

ACKNOWLEDGEMENTS

The authors thank Prof. Astrid Graslund for making the CD instrument available to us. Åsa Bergström and Marie Sjölund are acknowledged for their participation in the early stages of this study. Dr Andrew Ward is gratefully acknowledged for discussions and for critically reading the manuscript. This study was made possible by grants from the Swedish Council for Forestry and Agricultural Research, the Swedish Cancer Society, Barn cancerfondens and the Structural Biology Network of the Swedish Foundation of Strategic Research.

REFERENCES