Nucleic Acids Research, 1994, Vol. 22, No. 25 5709–5716

A guanosine quadruplex and two stable hairpins flank a major cleavage site in insulin-like growth factor II mRNA

Jan Christiansen*, Margrethe Kofod[1] and Finn C.Nielsen[1]
Department of Biological Chemistry, Institute of Molecular Biology, University of Copenhagen, Sølvgade 83 H, DK-1307 Copenhagen K and [1]Department of Clinical Biochemistry, University Hospital Rigshospitalet, DK-2100 Copenhagen Ø, Denmark

Received August 9, 1994; Revised and Accepted November 15, 1994

ABSTRACT

Insulin-like growth factor II (IGF-II) mRNAs are cleaved by an endonucleolytic event in a conserved part of their 3′ untranslated region that is predicted to exhibit a complex higher-order RNA structure. In the present study, we have examined the putative secondary structures of *in vitro* transcripts from the conserved part of human and rat mRNAs by enzymatic and chemical probing. The results show that the cleavage site is situated between two highly structured domains. The upstream domain consists of two large hairpins, whereas the downstream domain is guanosine-rich. The guanosine-rich domain adopts a compact unimolecular conformation in Na^+ or K^+ but not in Li^+, and it completely arrests reverse transcription in K^+ but only partially in Na^+, indicating the presence of an intramolecular guanosine quadruplex. The flanking higher-order structures may ensure that the cleavage site is not sequestered in stable RNA structures, thus allowing interactions with RNA or proteins at posttranscriptional stages of IGF-II expression.

INTRODUCTION

Insulin-like growth factor II (IGF-II) belongs to the family of insulin-like peptides and plays a major role in foetal development and growth (for review see 1). The IGF-II gene comprises nine and six exons in human and rat, respectively (Figure 1). By initiating transcription at four and three promoters in human and rat, respectively, the gene generates multiple mature transcripts with different 5′ untranslated regions (5′UTR) but identical coding regions and 3′UTRs. In both human and rat, a 1.8 kb RNA has been detected at steady-state that does not correspond to any of the identified promoters. The 1.8 kb RNA is uncapped, non-polysomal, and is generated from the mature mRNAs by an endonucleolytic event in their 3′UTR (2,3). The cleavage site is located 2183 and 1210 nucleotides downstream from the translation termination codon in human and rat 3′UTR, respectively (Figure 1). The nucleotides encompassing the cleavage site in IGF-II mRNAs are conserved, and studies of IGF-II minigenes including deletions of the cleavage region have indicated that they are important for the cleavage (4). Moreover, an additional element located close to the coding region in the 3′UTR is also required for the reaction (4), but the actual mechanism of cleavage in terms of participating macromolecules is unknown.

The RNA structures and *trans*-acting factors governing mRNA turnover of eukaryotic mRNAs are beginning to emerge, although progress has been hampered by the general absence of surviving steady-state intermediates, making deductions regarding precursor-product relationships a difficult task. Studies of a group of transiently expressed protooncogenes and cytokine mRNAs have shown that their rapid turnover are conferred by an A+U-rich sequence. Several proteins bind to this element, and some may stimulate deadenylation of the transcripts thereby directing subsequent degradation (for review see 5). In the case of the transferrin receptor mRNA, multiple hairpins in the 3′UTR bind the stabilising iron-responsive-element binding protein (IRE-BP), and an endonucleolytic cleavage site has recently been described (6). In *Xenopus* oocytes, a single-stranded cleavage region has been identified with overlapping determinants for a *trans*-acting inhibitor and a putative endonuclease (7). Finally, studies in yeast have disclosed multiple pathways of decay including decapping and efficient 5′→3′ exonucleases (8). Thus the turnover of eukaryotic mRNAs appears to involve alternative pathways, where primary and higher-order RNA structures and multiple *trans*-acting factors are involved. However, our knowledge of the elusive endonucleolytic activities is still extremely limited.

In this study, we employ enzymatic and chemical probes to determine the putative secondary structures of *in vitro* generated transcripts from human and rat IGF-II 3′UTRs that contain the conserved structural domains surrounding the endonucleolytic cleavage site. Moreover, we establish a differential monocation-dependent stability and an increased gel-mobility of the guanosine-rich domain downstream from the cut site. The results show that the cleavage site is located between two large hairpins and an intramolecular guanosine quadruplex. We suggest that the complex higher-order structure ensures that the cleavage site is not sequestered in stable RNA structures, thereby providing attachment sites for RNA or proteins co- and posttran-

*To whom correspondence should be addressed

scriptionally. The structural model should provide a rational basis for the design of substrates for use in the isolation and characterisation of binding proteins and the endonucleolytic activity.

MATERIALS AND METHODS

Materials

AMV reverse transcriptase was purchased from Life Sciences (USA), and ribonucleases T1, T2 and V1 were from Sankyo Inc. (Japan) and Pharmacia (Sweden), respectively. T7 RNA polymerase was purified from overproducing *E.coli* strain BL21/pAR1219 that was kindly provided by F.W.Studier. Dimethylsulphate (DMS), kethoxal and 1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimide metho-*p*-toluenesulphonate (CMCT) were from Merck (Germany), Serva (Germany), and Sigma (USA), respectively.

Plasmid constructions and RNAs

cDNA fragments were inserted into the polylinker of the *in vitro* transcription vector pUT719, that is a pUC19 derivative with a T7 RNA polymerase promoter-element and linker inserted between the *Sph*I and P*st*I sites of the pUC19 polylinker. The human cDNA *hum*316 corresponds to positions 2045–2360 downstream from the translation termination codon of exon 9 (Figure 1). The rat cDNA fragment *rat*296 corresponds to positions 1068–1363 downstream from the translation termination codon of exon 6 (Figure 1). The triple-repeat RNA of nucleotides encompassing positions 2174–2204 of human exon 9 (Figure 1) was generated by ligation of synthetic oligonucleotides, and inserted into the pUT719 *in vitro* transcription vector. The 39 nucleotides RNA of the guanosine-rich domain correspond to positions 2204–2242 from the translation termination codon of human exon 9 (Figure 1). The pyrimidine-rich transcript used in the quadruplex mobility experiment has the sequence: GGCUUCCUCCUCCUCCU-CUUCCUCCUCCUCCUCCU.

In vitro transcription

When *hum*316 RNA was prepared for probing in parallel with *rat*296 RNA, the transcription vector was linearized with *Eco*RI that cleaves in the polylinker. In addition to the nucleotides corresponding to the IGF-II insert, the transcripts exhibit 32 extraneous nucleotides (AGGUCGACUCUAGAGG-AUCCCCCGGCGAGCUCG) at their 3′ termini. Moreover, two extraneous guanosine residues are present in their 5′ termini. For experiments that only involved *hum*316 RNA the vector was linearized with *Sma*I to avoid extraneous nucleotides. The *in vitro* transcripts were generated by the addition of T7 RNA polymerase as described (9). Transcripts containing inosine instead of guanosine were produced by including 1.4 mM GMP and 0.7 mM ITP and omitting GTP in the transcription mixture. The RNAs were purified by electrophoresis in a 8% polyacrylamide, 7 M urea gel, after which the full-length transcripts were visualised by UV-shadowing, excised, and eluted from the gel with 250 mM sodium acetate, pH 6.0, 1 mM EDTA. After extraction with phenol/chloroform and precipitation with ethanol, the transcripts were dissolved in water.

The α-[^{32}P]UTP labelled 39 nucleotides RNA corresponding to the human guanosine-rich domain and the pyrimidine-rich transcript of 35 nucleotides were synthesised as described (9). The RNA was purified on a 15% polyacrylamide, 7 M urea gel and eluted as above. Following extractions with phenol/chloroform and precipitation, approximately 10^{-7} M RNA was renatured (see below) and applied to either a 15% polyacrylamide gel with 8 M urea in 45 mM Tris–HCl, 45 mM boric acid, 0.5 mM EDTA, pH 8.3 (0.5×TBE buffer) at 60°, or a 15% non-denaturing polyacrylamide gel in 1×TBE buffer with 25 mM KCl at room temperature. Folding experiments in the presence of various ions were carried out in a similar manner but non-denaturing gel electrophoresis was in the absence of KCl.

Structural probing with ribonucleases and chemicals

RNAs were always renatured or folded in the appropriate buffer by heating the samples to 95°C for 1 min and 56°C for 10 min, followed by slow cooling to 30°C, prior to digestion or modification. The *hum*316, *rat*296, and triple-repeat RNAs were partially digested with RNases T1, T2, and V1 or modified with dimethylsulphate (DMS), kethoxal, and 1-cyclohexyl-3-(2-morpholinoethyl)-carbodiimide metho-*p*-toluenesulphonate (CMCT) essentially as described (10). Briefly, enzymatic probings were performed by incubating 4 μg renatured RNA in 20 μl 30 mM Tris–HCl, 5 mM $MgCl_2$, 100 mM KCl at pH 7.8 with 0.01 unit RNase T1, 0.005 unit RNase T2 or 0.15 unit RNase V1 at 0°C. The digestion was terminated by addition of 130 μl 0.3 M sodium acetate pH 6.0, before the RNA was phenol/chloroform extracted. After precipitation the RNA was dissolved in 10 mM Tris–HCl, 0.1 mM EDTA at pH 7.5. Chemical modifications were performed by incubating 4 μg RNA in 100 μl 70 mM Hepes–KOH, 10 mM $MgCl_2$, 270 mM KCl, 2 mM DTT at pH 7.8 with 0.5 μl DMS, 200 μg kethoxal, or 800 μg CMCT. Incubations were performed at 30°C and lasted 10 min for DMS and kethoxal, and 20 min for CMCT. After termination of incubation, RNA was ethanol precipitated twice and phenol/chloroform extracted before it was reprecipitated and dissolved in 10 mM Tris–HCl, 0.1 mM EDTA at pH 7.5.

Primer extension

0.5 pmol of γ-[^{32}P]ATP labelled primers #1 or #2 (Figure 1) and 0.5 pmol of the enzymatically digested or chemically modified *hum*316 or *rat*296 RNA were heated to 95°C for 1 min and annealed at 50°C for 20 min in 12 μl 10 mM Tris–HCl, 40 mM KCl, 0.5 mM EDTA. Extensions were carried out for 30 min. at 45°C by adding 8 μl of a mixture containing 125 mM Tris–HCl, pH 8.4, 25 mM MgCl, 5 mM dithiotreitol, 2.5 mM dNTPs and 1 unit avian myeloblastosis virus reverse transcriptase. Sequencing tracks were obtained by including 350 μM ddGTP, 500 μM ddATP, 500 μM ddTTP, or 150 μM ddCTP in the final extension solution.

5′ end-labelling and enzymatic RNA sequencing

5′ end-labelled transcripts were generated by including 50 μCi γ-[^{32}P]GTP in 20 μl *in vitro* transcription buffer containing 1 μg linearised template, and the end-labelled transcripts were purified by denaturing polyacrylamide gel electrophoresis. Enzymatic RNA sequencing of 5′ end-labelled transcripts was carried out as described (11).

RESULTS

Structural probing with ribonucleases and chemicals

In order to gain insight into the putative secondary structure of the cleavage site and the surrounding regions in the 3′UTR of both human and rat IGF-II mRNAs, a 316 bp fragment spanning

Table 1. Structural specificities of ribonuclease and chemical probes

Ribonuclease	Specificity
RNase T1	Gp↓
RNase T2	Ap↓ > Np↓
RNase V1	N↓p double strand
Chemical	**Specificity**
DMS	G (N-7) > A (N-1) > C (N-3)
CMCT	U (N-3) > G (N-1)
Kethoxal	G (N-1, N-2)

positions 2045–2360 in human exon 9 and a 296 bp fragment encompassing positions 1078–1363 in rat exon 6 were inserted in an *in vitro* transcription vector containing the T7 RNA polymerase promoter. Following linearisation, all *in vitro* transcripts contain two extraneous guanosines at the 5′ end and varying numbers of extraneous nucleotides at the 3′ end (see Materials and methods for further details). The generated transcripts were invariably renatured before ribonucleases or chemicals were used to probe the structures. The employed ribonucleases were RNase T1, RNase T2, and RNase V1, and the chemicals were dimethyl sulphate (DMS), carbodiimide

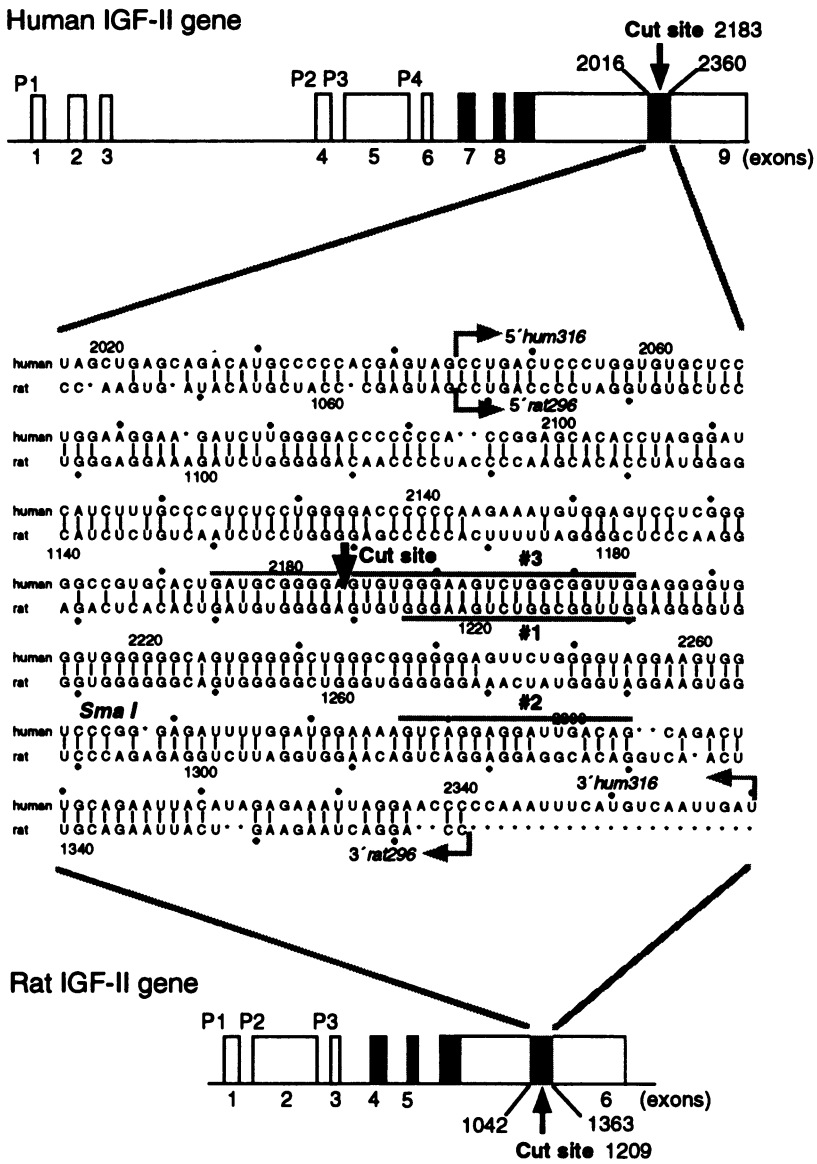

Figure 1. Genomic organisation of the human and rat IGF-II genes, and sequences of the conserved region encompassing the endonucleolytic cleavage site in the 3′ untranslated region of the mRNAs. Exons are indicated by boxes and numbered according to Holthuizen *et al.* (32). Filled boxes designate the coding regions for preproIGF-II and the sites of transcription initiation are indicated (P1, P2, P3 and P4). The human and rat IGF-II genes comprise 9 and 6 exons, respectively. Exons 7–8 and 234 nucleotides of exon 9 provide the coding region for the human prepropeptide, whereas, exons 4–5 and 234 nucleotides of exons 6 provide the rat prepropeptide. In human, exons 1–3, 4, 5, and 6 encode alternative 5′ untranslated regions, and exons 1, 2, and 3 encode different 5′ untranslated regions in the rat. Exons 9 and 6 in human and rat, respectively, encode the 3′ untranslated region. The endonucleolytic cleavage site is located 2183 and 1209 nucleotides downstream from the translation termination codon in human and rat, respectively. The transcripts examined in this study are indicated by the grey boxes. The blow-out shows the sequence of the conserved region encompassing the cleavage site (vertical arrow) in the human and rat RNAs. The bent arrows demarcate the borders of transcripts *hum*316 and *rat*296. Primers #1 and #2 are used in the primer extension analyses, whereas the unit in the triple-repeat RNA is shown by line #3. The *Sma*I site is used for linearisation in the production of *hum*316 RNA.

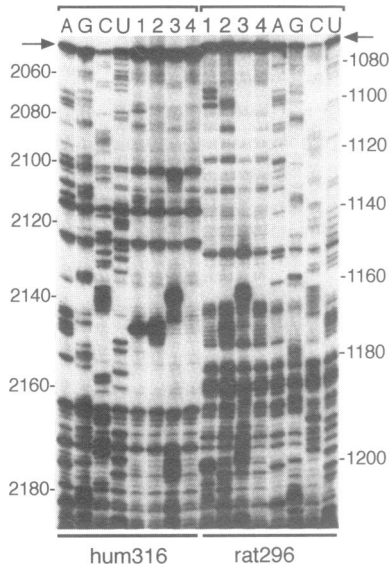

Figure 2. Enzymatic probing of hum316 and rat296 RNAs. The *hum*316 and *rat*296 RNAs were partially digested with ribonucleases T1 (track 1), T2 (track 2), V1 (track 3) or undigested (track 4), and the products examined by primer extension analysis with primer # 1. The full-length reverse transcripts are indicated by arrows. Dideoxynucleoside triphosphate sequencing tracks (A, G, C, U) of *hum*316 and *rat*296 RNAs are shows to the left and right, respectively. Numbers refer to positions downstream from the translation termination codon. A digested position gives rise to a reverse transcript that is one nucleotide shorter than the corresponding band in the sequencing track.

(CMCT), and kethoxal (12). The sites of strand cleavage or modification can, with the exception of the guanosine N-7 position, be determined by reverse transcription from a 5′ end-labelled primer. The specificities of the probes are listed in Table 1.

Since pilot experiments had shown that reverse transcriptase had great difficulties traversing the guanosine-rich motif, primer # 1 (see Figure 1) was used to examine the accessibility of the domain upstream from the endonucleolytic cleavage site. The result from a ribonuclease probing experiment of this region in both human and rat mRNAs is shown on the autoradiograph in Figure 2, and the involved positions and cleavage intensities are compounded in Table 2 and Figure 8. The most accessible region of the upstream domain is between positions 2137 and 2147 in *hum*316 and between positions 1162 and 1173 in *rat*296. Moreover, positions 2173−2176 in *hum*316 and similar positions in *rat*296 are reactive to RNase V1. These arrays are phylogenetically equivalent, and the pattern of reactivity of the single-stranded ribonucleases versus ribonuclease V1 suggests the presence of a large hairpin, provided by positions 2113−2176 in *hum*316 and by positions 1139−1203 in *rat*296. In contrast, positions 1091−1100 are mainly reactive to the single-stranded specific ribonucleases in *rat*296. We infer that the latter region is part of a loop.

In an attempt to substantiate and expand on these findings *hum*316 and *rat*296 were probed with chemicals. Figure 3 shows an autoradiograph from the reverse transcription analysis of *hum*316, and the data from the analysis of both transcripts are summarised in Table 2 and Figure 8. The 2142−2147 positions are very reactive towards DMS and kethoxal and exhibits a weak CMCT reactivity, whereas the corresponding positions in the rat are strongly reactive to CMCT. The high reactivity supports the

Table 2. Reactivity of the upstream domain of *hum*316 and *rat*296 RNA to ribonucleases and chemicals

Human		T1		T2		V1		DMS		CMCT		Kethoxal		Rat	
Pos.	Nucl.	Hum 316	Rat 296	Hum 316	Rat 296	Hum 316	Rat 296	Hum 316	Rat 296	Hum 316	Rat 296	Hum 316	Rat 296	Nucl.	Pos.
2066	U									++	+			U	1089
			++										+	G	1091
2069	A		++					++					+	G	1092
2070	A							++	+					A	1093
2071	G											+	++	G	1094
2072	G		++									+	++	G	1095
2073	A			+				++	++					A	1096
2074	A			+				+	++					A	1097
				+					++					A	1098
2075	G	++	+++									+	++	G	1099
				+					+					A	1100
2079	U									+++	+			U	1101
2080	U									++	+			U	1103
2081	G	+													
				+					+					A	1109
									+					A	1111
2093	A							++							
2102	C					+									
2106	U									+++	++			U	1134
2136	C						++							G	1162
2137	C					+++	+++							C	1163
2138	C					++++	+++							C	1164
2139	C					+++	+++							C	1165
2140	C					++									
2142	A			++				++							
2143	A			++++	+++			++						C	1169
2144	G	++++		+	++++					+	++	++++		U	1170
2145	A			+++	+++			+++			++			U	1171
2146	A			++	++			++			+++			U	1172
2147	A			+	++			++			++			U	1173
2148	U									+					
2150	U								+	+				A	1175
2153	A						++	+						C	1192
							++							C	1198
2172	C					++									
2173	U		++			++						+		G	1200
2174	G					++									
2175	A					++									

The degree of cleavages and modifications are indicated as (+) weak, (++) medium, (+++) strong and (++++) very strong.

presence of a hairpin loop in both organisms, and the difference in specificity reflects a sequence difference between the two organisms. The other reactive region is the 2069−2080 array which is far more accessible to the chemicals than to the ribonucleases, and the equivalent region in the rat is also accessible to the chemicals. This reactivity pattern is suggestive of a loosely structured region without sharp apical bends. Uridine-2106 is situated outside the major reactive regions but it is reactive towards CMCT. However, it is not flanked by any reactive positions with the exception of a weak RNase V1 cleavage following cytidine-2102. We infer that position 2106 is an unpaired uridine in a helix.

Enzymatic probing of the cleavage region

Since the cleavage region and the guanosine-rich stretch are difficult to analyse by a primer extension approach due to extensive termination of reverse transcriptase, the *hum*316 transcript was 5′ end-labelled, renatured, and probed by RNases T1, T2, and V1. Figure 4A is an autoradiograph depicting the result from this experiment. The cleavage region (positions 2178−2204) exhibits three accessible positions to RNases T1, T2, and V1, that are indicated by arrows b, c, and a, respectively. Otherwise, the region is inaccessible. A similar experiment was carried out in the presence of 10 mM EDTA instead of 5 mM Mg^{2+}, but the probing pattern for T1 and T2 was similar (results not shown). Moreover, the experiment in the presence of Mg^{2+} was also carried out at 65°C with RNase T1 instead of at 0°C, and the results were virtually identical (results not shown). The low reactivity is in contrast to the reactivity observed under the RNA sequencing conditions of 8 M urea, pH 5.0, 50°C (track G) where guanosines in the cleavage region becomes highly reactive. In contrast the downstream guanosine-rich stretch from position 2204 to 2242 remained fairly unreactive to RNase T1 under the extreme denaturing conditons. Cleavage is only

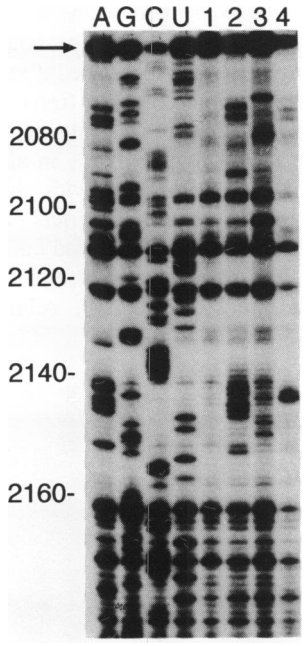

Figure 3. Chemical probing of hum316 RNA. *Hum*316 RNA was unmodified (track 1) or modified with DMS (track 2), CMCT (track 3), and kethoxal (track 4), and the products analysed by primer extension analysis with primer # 1. The full-length reverse transcript is indicated by the arrow, and dideoynucleoside triphosphate sequencing tracks (A, G, C, U) are shown to the left. Numbers refer to positions downstream from the translation termination codon. A modified position gives rise to a reverse transcript that is one nucleotide shorter than the corresponding band in the sequencing track.

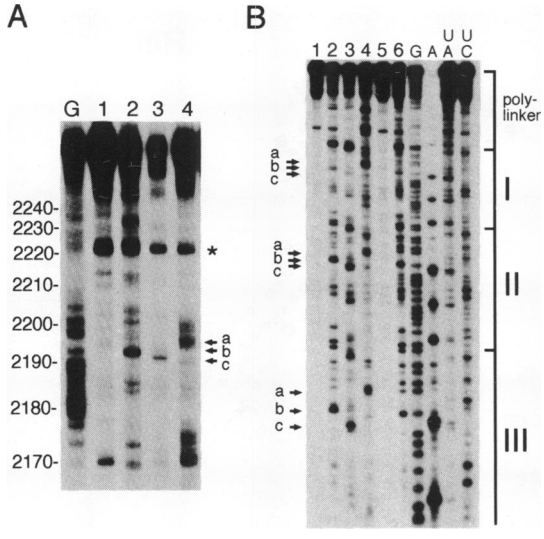

Figure 4. Enzymatic probing of 5′ end-labelled *hum*316 RNA and a triple-repeat RNA. *Hum*316 RNA (**A**) and a triple-repeat RNA (**B**) were 5′ end-labelled by the inclusion of γ-[^{32}P]GTP during synthesis, and partially digested with ribonucleases T1 (track 2), T2 (track 3), and RNase V1 (track 4). Tracks 1 and 5 show undigested RNA at 0°C and 65°C, respectively, and track 6 is a ribonuclease T1 digest at 65°C. The RNA sequencing tracks were obtained by enzymatic sequencing (11) with T1 (track G), U2 (track A), *Phy M* (track U+A), and *B. cereus* (track U+C). Arrows a, b, and c designate digested positions described in the text, and brackets I, II, and III delineate the three units of the cleavage region in the triple-repeat transcript. The asterisk (*) indicate a major control cut at C_{2221}. Numbers on the left refer to positions downstream from the translation termination codon.

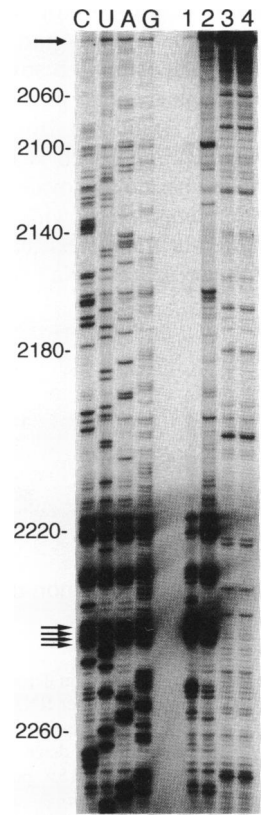

Figure 5. Ion- and nucleoside-dependent termination of reverse transcription at the guanosine-rich domain. Primer extension analysis, employing primer #2, of *hum*316 RNA in the presence of K^+ (track 1) or Na^+ (track 2) and of a corresponding in vitro transcript, where guanosine is substituted with inosine except at the initiating position (track 3, K^+ and track 4, Na^+). C, U, A and G are dideoxynucleoside triphosphate sequencing tracks. The numbers on the left indicate the positions downstream from the translation termination codon. The arrows designate the major points of premature termination and the full-length transcript (top).

observed at $G_{2212,2213}$ and G_{2233} (track 2). In an attempt to elucidate whether the low reactivity of the cleavage region is due to a local structure or due to higher-order interactions with the flanking domains, an *in vitro* transcript containing three tandem copies of the cleavage region was generated, 5′ end-labelled, and probed with the ribonucleases at 0°C and also at 65°C with RNase T1. Figure 4B is an autoradiograph showing that the ribonucleases preferentially attack the same positions (indicated by arrows a, b, and c) in each tandem copy of the cleavage region (brackets I, II, and III) as those attacked in the *hum*316 transcript. Moreover, at 65 °C the tandem copies are still structured (compare the reactivities of tracks 6 and G to RNase T1), although the main difference is in the behaviour of guanosines at positions 2179−2182 and 2188−2190. We infer that the cleavage region adopts a local structure in naked RNA that can be fully displayed by 8 M urea.

Ion- and guanosine-dependence of reverse transcriptase termination

The guanosine-rich stretch (positions 2204−2242) downstream from the cleavage region is virtually inaccessible to RNase T1 even in the presence of 8 M urea at 50°C (Figure 4A). Therefore, a primer extension analysis of transcripts containing either

guanosines or inosines in the presence of either Na^+ or K^+ was carried out. The rationale behind this particular approach was to establish whether the guanosine-rich stretch is able to adopt an intramolecular quadruplex, since planar layers of guanine quartets are preferentially stabilized by K^+, in contrast to guanine basepairs (13). The approach is feasible because the AMV reverse transcriptase has no monovalent cation requirement (14). Figure 5 is an autoradiograph showing the primer extension analysis of the two different transcripts in the presence of K^+ (tracks 1 and 3) or in the presence of Na^+ (tracks 2 and 4) when primer #2 is used. The experiment shows that virtually complete termination occurs at $G_{2242-2239}$ in the presence of K^+ (track 1) whereas full-length cDNA can be obtained if reverse transcription is carried out in the presence of Na^+ (track 2). We infer that the guanosine-rich region adopts a local higher-order structure that is more stable in K^+ than in Na^+, in agreement with the optimal radius of K^+ in a complex with eight oxygen atoms provided by two layers of guanine quartets (13). In contrast, a transcript containing inosines can be copied into full-length cDNA without extensive termination regardless of the identity of the monovalent cation (tracks 3 and 4), reinforcing that it is

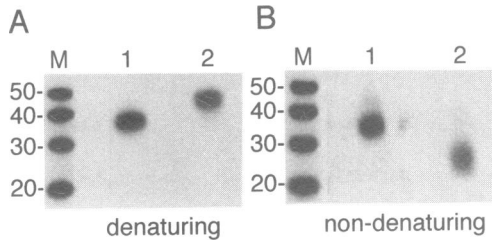

Figure 6. Migration of the guanosine-rich RNA in denaturing and non-denaturing polyacrylamide gels. α-[^{32}P]UTP labelled 39mer RNA encompassing positions 2204−2242 and a similarly labelled pyrimidine-rich 35mer (see Materials and methods) were purified by electrophoresis in a denaturing polyacrylamide gel. The RNAs were eluted and applied to either a 15% polyacrylamide, 8 M urea, gel in 0.5×TBE buffer at 60°C (**A**, *denaturing*) or a 15% non-denaturing polyacrylamide gel in 1×TBE buffer with 25 mM KCl at room temperature (**B**, *non-denaturing*) The RNAs were co-electrophoresed with 5′ end-labelled oligo-dT size markers (track M). Track 1 is the pyrimidine-rich 35mer, and track 2 is the guanosine-rich 39mer.

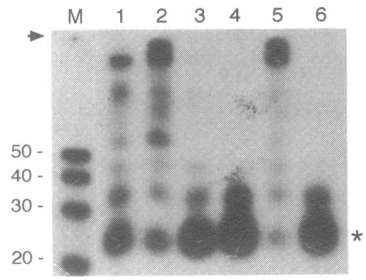

Figure 7. Ion-dependent folding of the guanosine-rich RNA. α-[^{32}P]UTP labelled 39mer RNA encompassing positions 2204−2242 was folded in the presence of 1 mM EDTA (track 1), 100 mM LiCl (track 2), 100 mM NaCl (track 3), 100 mM KCl (track 4), 5 mM $MgCl_2$ (track 5), or 100 mM KCl and 5 mM $MgCl_2$. The RNA was applied to a 15% non-denaturing polyacrylamide gel in 1×TBE buffer and co-electrophoresed with 5′ end-labelled oligo-dT size markers (track M). The arrow shows the origin, and the asterisk designates the fast-migrating species.

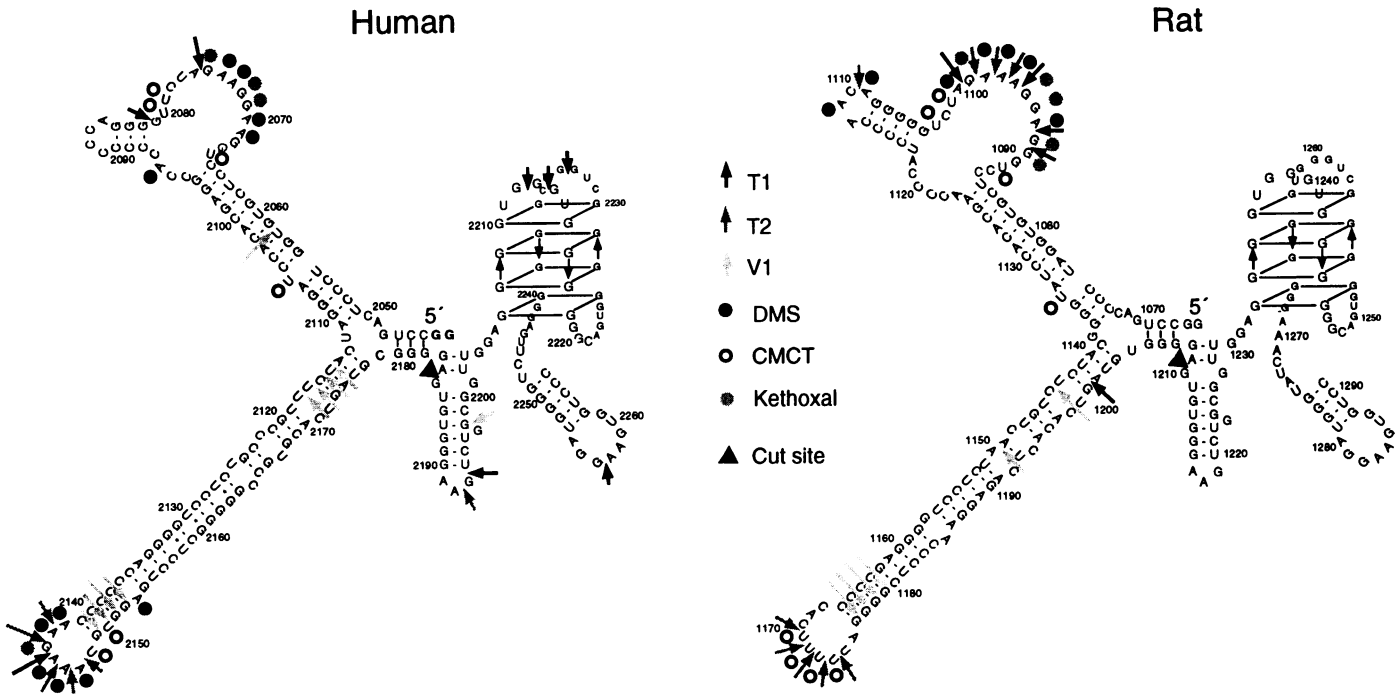

Figure 8. Putative secondary structures of the endonucleolytic cleavage region and the flanking domains. Arrows designate enzymatic attacks, and dots show chemically modified bases. Small arrows inside the quadruplex indicate the general polarity of the RNA chain. Numbers refer to positions downstream from the translation termination codons.

interactions between guanosine residues that are preferentially stabilised by K^+.

Gel-mobility of the guanosine-rich domain

Guanosine-rich stretches in both DNA and RNA have the potential to form inter- and intramolecular quadruplexes. To distinguish between these possibilities the gel-mobility of an *in vitro* transcript of 39 nucleotides encompassing the guanosine-rich stretch was examined. The 39mer RNA and a pyrimidine-rich transcript of 35 nucleotides, that was included for comparative purposes, were synthesised by the approach of Milligan and Uhlenbeck (9). The two transcripts were generated in the presence of α-[^{32}P]UTP, and the full-length transcripts purified by denaturing polyacrylamide gel electrophoresis. The excised transcripts were folded in the presence of 5 mM Mg^{2+} and 100 mM K^+, and one aliquot was applied to a denaturing polyacrylamide gel (Figure 6A) and another aliquot applied to a non-denaturing polyacrylamide gel (Figure 6B). Oligo-dT markers were co-electrophoresed on both gels. The pyrimidine-rich transcript migrates as a 35mer on both gels, whereas the guanosine-rich 39mer migrates between the $(dT)_{40}$ and $(dT)_{50}$ markers on the denaturing gel and between the $(dT)_{20}$ and $(dT)_{30}$ markers on the non-denaturing gel. Moreover, the fast-migrating guanosine-rich transcript was excised from the non-denaturing polyacrylamide gel, eluted, and applied to a denaturing polyacrylamide gel, where it migrated between the $(dT)_{40}$ and $(dT)_{50}$ markers (result not shown). We conclude, that an *in vitro* transcript of 39 nucleotides containing the guanosine-rich domain does not oligomerize at the concentration and time of incubation employed in this study, but is able to adopt a compact intramolecular conformation.

Since the folding experiment depicted in Figure 6 does not discriminate between a quadruplex and other fast-migrating species, the 39mer was folded in the presence of various ions and analysed in a non-denaturing gel. Figure 7 shows that Na^+, K^+, or the combination of Mg^{2+} and K^+, favours the formation of a fast-migrating species. In the presence of Li^+, Mg^{2+}, or Tris-HCl and EDTA only, the formation of the fast-migrating monomer is decreased and multimers are formed. The differential effect of Li^+ versus Na^+ or K^+ on the formation of the fast-migrating species is compatible with the requirement for Na^+ or K^+ in the generation of guanosine quadruplexes.

DISCUSSION

In this study, we have subjected the conserved region of the 3'UTR, encompassing the major endonucleolytic cleavage site, in human and rat IGF-II mRNAs to a structural analysis with ribonucleases and chemicals. The cleavage domain is highly structured and includes a guanosine-rich stretch that is able to adopt the conformation of a guanosine quadruplex. The results are compounded in Figure 8 that shows the putative secondary structural models of *in vitro* transcripts derived from this region in human and rat.

The striking features of these models are the highly structured domains surrounding the endonucleolytic cleavage site; namely the two large hairpins upstream and the intramolecular quadruplex downstream from the cleavage site. Although the region overall is highly conserved at the sequence level, the sequences in the hairpin loops are variable implying that it is structure *per se* rather than particular loop sequences that is of importance, at least for the hairpins. This is in contrast to the situation with 'functional' RNAs such as rRNAs where the loop sequences are conserved and the stem sequences often vary in a co-ordinated fashion (15). The structural argument can also be extended to the guanosine-rich domain, because here a high degree of conservation at the sequence level is mandatory for the formation of an intramolecular quadruplex. Although we have not proven the presence of an intramolecular quadruplex by high-resolution structural analysis, the inaccessibility of the suggested Hoogsteen paired guanosines to RNase T1, the monocation-selective arrest of reverse transcription, and the folding of the guanosine-rich region facilitated by Na^+ or K^+, are diagnostic of this type of structure. The putative structural models depicted in Figure 8 are based entirely on the data obtained in this study. The quadruplex is drawn in a way that is compatible with the positions of reverse transcription arrest, but alternative conformers are possible. If the examined transcripts are subjected to computer analysis (16) entirely different, although appealing, structures emerge. The computer-generated structures are incompatible with the probing data with the exception of the 5' hairpin stem which is predicted. This failure is mainly due to the presence of the quadruplex which current programmes cannot predict. However, if the computer analysis is restricted to the hairpin domain upstream from the cleavage region the predicted structures are virtually identical with those suggested in this study, the main difference being the loop regions in the 5' hairpin.

What is the functional significance of the highly structured regions surrounding the cleavage site? One possibility is that they merely have a passive role ensuring that the cleavage region is accessible to interacting macromolecules and not sequestered in stable RNA structures. This proposal may seem contradictory in terms, since the cleavage region appears to be tied up in a local structure. In contrast to the large hairpins and the quadruplex it is, however, feasible to unfold the local hairpin in urea, indicating that the cleavage region is displayable. The putative cleavage site contains the sequence element GGGGAGUGUGGG that exhibits a high degree of similarity with the DSEF-1 binding site GGGGGAGGUGUGGG (17). It is therefore tempting to envisage a role of the 50 kDa DSEF-1, or a similar RNA binding protein, in interacting with the cleavage region. In fact, a protein from HeLa nuclear extract with a molecular mass of about 50 kDa can be UV-crosslinked to the cleavage region in both the *hum*316 transcript and in the triple-repeat transcript used in Figure 4B (unpublished results). So the mRNA entering the cytosol may contain a displayed cleavage region provided by a combination of flanking RNA structures and an RNA binding protein. Additional *trans*-acting factors may act as inhibitors and ensure that the mRNA is not degraded indiscriminately.

The presence of inter- and intramolecular DNA quadruplexes has been described in several instances (13,18,19,20,21) whereas only intermolecular RNA quadruplexes have been reported (22,23). This study suggests that an intramolecular quadruplex is present in the 3'UTR of IGF-II mRNA. The functional significance of RNA quadruplexes is unclear, although it has been suggested that *HIV-1* RNA dimerises through quadruplex formation (23). Quadruplexes may be thermodynamic cul-de-sacs – incompatible with dynamic interactions such as unfolding (24). Therefore, the possible *in vivo* function of these structures is likely to be involved with obtaining stable nucleic acid structures that are not required to unfold at a later stage. It may be this particular function of providing stability that the IGF-II quadruplex fulfils in the present system, because the 3' cleavage product, although uncapped, is stable *in vivo* . In this vein, it should be noted that

studies of the decay pathways of *PGK1* mRNA in yeast are often carried out with artificial transcripts containing poly(G) stretches (18 nucleotides) to trap short-lived intermediates. In these cases, the poly(G) stretch inhibits further progression by 5′→3′ exonucleases (25). A recent report describes the purification of a yeast nuclease that recognizes a tetrastranded guanosine quadruplex in DNA and cuts a single-stranded region 5′ to the quadruplex (26). Besides being the first piece of evidence of the presence of quadruplexes *in vivo*, the scenario appears similar to the endonucleolytic event in the turnover of IGF-II mRNA. Whether the 1.8 kb tail fragment has a biological function is hypothetical, since the RNA exhibits no conserved reading frames, but untranslated, yet functional, RNAs have been suggested in cases such as the α-tropomyosin 3′UTR (27), the *lin4* product (28,29), the enigmatic *H19* RNA (30), and recently the *nanos* 3′UTR (31).

The present study emphasises the growing realisation that 3′UTRs harbour more information than the polyadenylation signal and the poly(A) tail. The structural analysis demonstrates that the endonucleolytic cleavage site is located between two large hairpins and an intramolecular guanosine quadruplex. We suggest that the complex higher-order structure ensures that the cleavage site is not sequestered in stable RNA structures, thereby providing attachment sites for interacting RNA or proteins. The examined region is the most conserved part of the entire IGF-II gene, so it is likely that this elaborate structure is crucial for the overall regulation of IGF-II expression

ACKNOWLEDGEMENTS

We thank Mette Simons and Nette Larsen for technical assistance, Jens F.Rehfeld for his support, and Roger A.Garrett for critically reading the manuscript. The study was supported by the Novo Nordisk Foundation, the Carlsberg Foundation, and the Danish Natural Science and Medical Research Councils.

REFERENCES

1. Nielsen, F.C. (1993) *Prog. Growth Factor Res.* 4, 257–290.
2. Meinsma, D., Holthuisen, P.E., Van den Brande, J.L. and Sussenbach, J.S. (1991) *Biochem. Biophys. Res. Com.* 179, 1509–1516.
3. Nielsen, F.C. and Christiansen, J. (1992) *J. Biol. Chem.* 267, 19404–19411.
4. Meinsma, D., Scheper, W., Holthiusen, P.E., Van den Brande, J.L. and Sussenbach, J.S. (1992) *Nucleic Acids Res.* 20, 5003–5009.
5. Sachs, A.B. (1993) *Cell* 74, 413–421.
6. Binder, R., Horowitz, J.A., Basilion, J.P., Koeller, D.M., Klausner, R.D. and Harford, J.B. (1994) *EMBO J.* 13, 1969–1980.
7. Brown, B.D., Zipkin, I.D. and Harland, R.M. (1993) *Genes Dev.* 7, 1620–1631.
8. Decker, C.J. and Parker, R. (1993) *Genes Dev.* 7, 1632–1643.
9. Milligan, J.F. and Uhlenbeck, O.C. (1989) *Methods Enzymol.* 180, 51–62.
10. Christiansen, J., Egebjerg, J., Larsen, N. and Garrett, R.A. (1990) In Spedding, G.(ed.), Ribosomes and Protein Synthesis - A Practical Approach, IRL Press, Oxford, pp. 229–252.
11. Donis-Keller, H. (1980) *Nucleic. Acids Res.* 8, 3133–3142.
12. Ehresmann, C., Baudin, F., Mougel, M., Ebel, J.-P. and Ehresmann, B. (1987) *Nucleic Acids Res.* 15, 9109–9129.
13. Sundquist, W. and Klug, A. (1989) *Nature* 342, 825–829.
14. Fujinaga, K., Parsons, J.T., Beard, J.W., Beard, D. and Green, M. (1970) *Proc. Natl. Acad. Sci. USA* 67, 1432–1439.
15. Noller, H.F. (1991) *Annu. Rev. Biochem.* 60, 191–227.
16. Zuker, M. (1989) *Methods Enzymol.* 180, 262–288.
17. Qian, Z. and Wilusz, J. (1991) *Mol. Cell Biol.* 11, 5312–5320.
18. Williamson, J.R., Raghuraman, M.K. and Cech, T.R. (1989) *Cell* 59, 871–880.19.
19. Sen, D. and Gilbert, W. (1990) *Nature* 344, 410–414.
20. Kang, C., Zhang, X., Ratliff, R., Moyzis, R. and Rich, A. (1992) *Nature* 356, 126–131.
21. Smith, F.W. and Feigon, J. (1992) *Nature* 356, 164–168.
22. Cheong, C. and Moore, P.B. (1992) *Biochemistry* 331, 8406–8414.
23. Sundquist, W.I. and Heaphy, S. (1993) *Proc. Natl. Acad. Sci. USA.* 90, 3393–3397.
24. Zahler, A.M., Williamson, J.R., Cech, T.R. and Prescott, D.M. (1991) *Nature* 350, 718–720.
25. Vrenken, P. and Raué, H.A. (1992) *Mol. Cell Biol.* 12, 2986–2996.
26. Liu, Z. and Gilbert, W. (1994) *Cell* 77, 1083–1092.
27. Rastinejad, F., Conboy, M.J., Rando, T.A. and Blau, H.M. (1993) *Cell* 75, 1107–1117.
28. Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) *Cell* 75, 843–854.
29. Wightman, B., Ha, I. and Ruvkun, G. (1993) *Cell* 75, 855–862.
30. Hao, Y., Crenshaw, T., Moulton, T., Newcomb, E. and Tycko, B. (1993) *Nature* 365, 764–767.
31. Gavis, E. and Lehmann, R. (1994) *Nature* 369, 315–318.
32. Holthuizen, P., Van der Lee, F.M., Ikejiri, K., Yamamoto, M. and Sussenbach, J.S. (1990) *Biochim. Biophys. Acta.* 1087, 341–343.