Volume 301, number 1, 79–82 FEBS 10920 April 1992

A consensus repeat sequence from the human insulin gene linked polymorphic region adopts multiple quadriplex DNA structures in vitro

Michael C.U. Hammond-Kosack* and Kevin Docherty

Department of Medicine, University of Birmingham, Queen Elizabeth Hospital, Birmingham B15 2TH, UK

Received 18 February 1992

A hypervariable region consisting of repeats of a 14 base pair (bp) consensus sequence ACAGGGGT(G/C)(T/C)GGGG is located 363 bp upstream of the human insulin gene. Different repeat numbers if this oligonucleotide give rise to a polymorphism, and so this region is commonly known as the insulin gene linked polymorphic region (ILPR). Here we present evidence, based on the mobility in non-denaturing polyacrylamide gels of two dissimilarly sized oligonucleotides containing the ILPR consensus sequence, that this sequence can adopt a number of quadriplex DNA structures in vitro.

G4 DNA; Insulin gene linked polymorphic region; Insulin gene; DNA structure

1. INTRODUCTION

The insulin gene linked polymorphic region (ILPR), located 363 base pairs (bp) upstream of the human insulin gene, is composed of tandem repeats of the consensus sequence ACAGGGG(G/C)(T/C)GGGG [1]. We have previously presented evidence, based on the reactivity to bromoacetaldehyde (BAA), and homoduplex mapping, of a plasmid containing a 4.8 kb human insulin gene fragment, that the ILPR adopts an altered DNA structure in vitro [2]. This structure was present in the top, G-rich strand, with the C-rich bottom strand largely single stranded. The presence of a run of G residues together with fine mapping of the BAA reactivity to nucleotides flanking the G-rich sequences within the consensus repeat sequence, prompted us to speculate the existence of a quadriplex structure [3,4] within the G-strand of the ILPR [2].

Here we present unequivocal evidence that the ILPR can adopt a number of quadriplex DNA structures in vitro. This conclusion is based on the effect of alkali metal ions on the mobility in non-denaturing polyacrylamide gels of two dissimilarly sized oligonucleotides containing the ILPR consensus sequence.

Abbreviations: ILPR, insulin gene linked polymorphic region; BAA, bromoacetaldehyde.

Present address: Department of Molecular Genetics, JI Centre for Plant Science Research, Colney Lane, Norwich, Norfolk NR4 7UJ, UK.

Correspondence address: K. Docherty, Department of Medicine, Queen Elizabeth Hospital, Birmingham B15 2TH, UK. Fax: (44) (21) 414 1355.

2. EXPERIMENTAL

2.1. *Synthesis of oligonucleotides*

The two oligonucleotides used in this study, oligonucleotide S (5'ACAGGGGTGTGGG3') and oligonucleotide L (5'TAGTCCAGCACAGGGGTGTGGGGTCAGAT3'), were synthesised on a Pharmacia LKB Gene Assembler Plus on the 0.2 μmol scale. Following release from the support column, oligonucleotides were precipitated with isopropanol, and purified by electrophoresis through a 16% denaturing polyacrylamide gel.

2.2. *Radiolabelling of oligonucleotides*

Oligonucleotides were labelled with $\gamma[^{32}P]$ATP in a reaction catalysed by T4 polynucleotides kinase [5]. The radiolabelled oligonucleotides were then separated from unincorporated $\gamma[^{32}P]$ATP by centrifugation through Sephadex G50 spin columns.

2.3. *Non-denaturing polyacrylamide gel electrophoresis*

Non-denaturing polyacrylamide gels (8% w/v acrylamide) were run in 0.5× TBE (45 mM Tris–borate, 1 mM EDTA). Gels were dried and subjected to autoradiography using Fuji RX film.

3. RESULTS

Quadriplex (tetrastrand DNA) structures are formed only by oligomers containing one or several runs of guanine residues. These structures are stabilised by Hoogsteen base pairing, involving the N-7 positions of the contributing G residues. The exact nature of the quadriplex formed depends on the number of G runs within the oligomers and environmental conditions such as salt concentration, temperature and torsional stress. The three main quadriplex structures identified at present are: (i) tetrastrand parallel quadriplexes (G4-DNA); (ii) unimolecular antiparallel quadriplexes (g4'-DNA); and (iii) bimolecular antiparallel quadriplexes (G'2-DNA) (we adopt here the nomenclature of Sen and Gilbert [3]).

Published by Elsevier Science Publishers B.V.

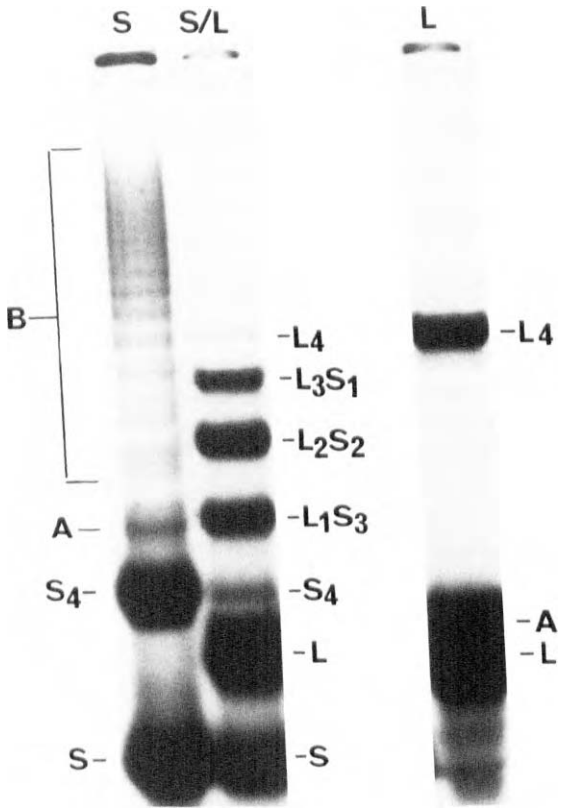

Fig. 1. The electrophoretic mobility of oligonucleotides S and L in non-denaturing polyacrylamide gel electrophoresis. ^{32}P-Labelled oligonucleotide S, L or S and L (S/L) in 20 μl 10 mM Tris–HCl, pH 8.0, 1 M EDTA and 0.5 M NaCl were evaporated to dryness at 60°C over 18 h. The samples were then dissolved in 10 mM Tris–HCl, pH 8.0, 1 mM EDTA and 0.1 M NaCl and subjected to electrophoresis on an 8% polyacrylamide gel. The position of the various single- and tetra-strand products are indicated (see text).

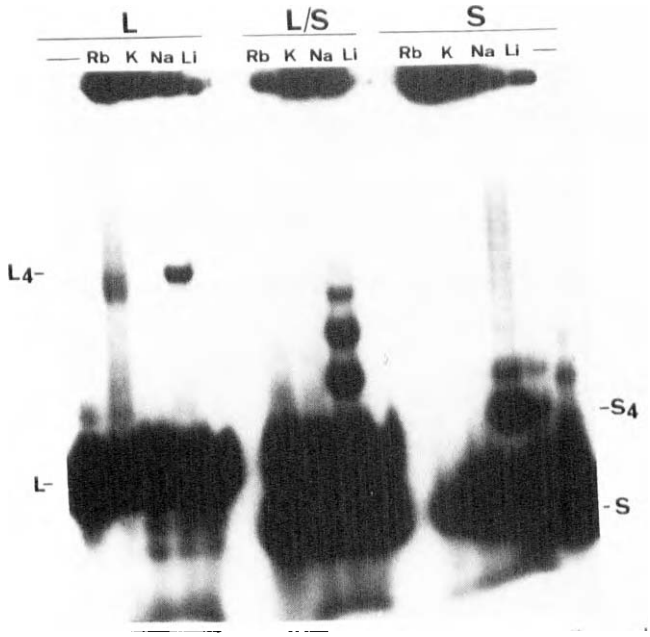

Fig. 2. The effect of alkali metal ions on the mobility of oligonucleotides S and L in non-denaturing polyacrylamide gel electrophoresis. ^{32}P-Labelled oligonucleotide L, S or S and L (S/L) in 20 μl 10 mM Tris–HCl, pH 8.0, 1 M EDTA and 0.5 M of the indicated salt, i.e. RbCl (Rb), KCl (K), NaCl (Na) and LiCl (Li), were evaporated to dryness at 60°C over 18 h. The samples were then dissolved in 10 mM Tris–HCl, pH 8.0, 1 mM EDTA and 0.1 M of the appropriate salt, and subjected to electrophoresis on an 8% polyacrylamide gel. The position of the various single- and tetra-strand products are indicated.

The consensus ILPR repeat sequence, ACAGGGGTGTGGGG, is very similar to repeat sequences in telomeric DNA [6], and to the sequence of the oligonucleotides used originally to confirm the existence of quadriplexes [3,7]. Thus we expected the single stranded ILPR consensus oligomer to behave similarly and form G4-DNA in vitro. Indeed, we have shown earlier [2] that, although the ILPR exists as double stranded DNA within the genome, under torsional stress in vitro the ILPR exists largely as a single stranded region with the top strand exhibiting a complex secondary structure, which we believe to be G4'-DNA.

The hypothesis that the ILPR can exist as a quadriplex structure was tested using the method of Sen and Gilbert [3]. Sequences capable of forming G4-DNA exhibit a characteristic slower mobility in non-denaturing polyacrylamide gels than unstructured DNA. Furthermore, when two oligonucleotides (L, for long and S, for short) containing the same run of guanines but dissimilar in length are mixed and incubated under conditions which favour quadriplex formation, then five distinct four-stranded products, corresponding to L4, L3S, L2S2, LS3 and S4, should be resolved by non-denaturing polyacrylamide gel electrophoresis.

Two oligonucleotides were therefore synthesised. Oligonucleotide S contained the 14 bp ILPR consensus sequence 5'ACAGGGGTGTGGGG 3', while oligonucleotide L, a 29-mer with the sequence 5'TAGTCCAGCACAGGGGTGTGGGGTCAGAT 3', contained the consensus sequence (bold) flanked by 9 unrelated nucleotides on the 5' side and 6 unrelated nucleotides on the 3' side. When each oligonucleotide was incubated individually and subjected to non-denaturing polyacrylamide gel electrophoresis, both oligonucleotide S and oligonucleotide L migrated as two components (Fig. 1). However, importantly, only oligonucleotide S gave rise to a number of additional, albeit less abundant, bands (Fig. 1, complex B). When oligonucleotide S and oligonucleotide L were incubated together 5 additional bands were observed corresponding to L4, L3S, L2S2, LS3, and S4. This showed that the ILPR consensus sequence can indeed form quadriplex structures, the most abundant of which is G4-DNA (Fig. 1).

We next examined the salt dependence for the formation of G4-DNA in oligonucleotide L and oligonucleotide S. The results show that the quadriplex structure formed in the presence of sodium ions, but not in the presence of rubidium, lithium, or potassium (Fig. 2).

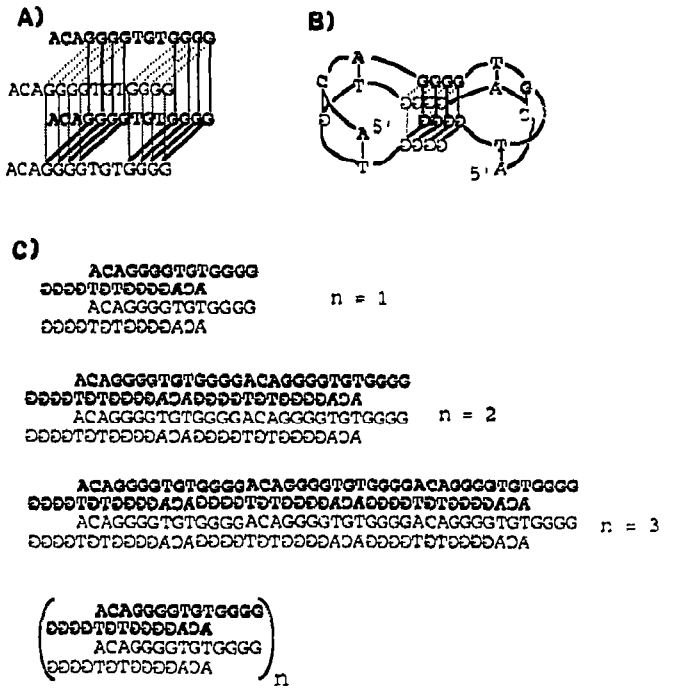

Fig. 3. Schematic representation of the most likely structures adopted by the ILPR consensus oligonucleotide S. (A) Tetramolecular, parallel G4-DNA. The two bold sequences of oligonucleotides are above the plane of the paper. The hydrogen bonding between the G residues is indicated by lines. (B) Bimolecular antiparallel K-type DNA [3] as in complex A of Fig. 1. This antiparallel structure would be further stabilised by Watson–Crick base pairing between the ACA and TGT trinucleotides. (C) Tetramolecular antiparallel (complex B of Fig.1). The antiparallel oligonucleotides (bold sequence above the plane of the paper) are stacked and stabilised by Watson–Crick base pairing between the trinucleotides. The overhanging G tetrads form the 'sticky ends' necessary for the formation of multimers of this structure.

These findings are compatible with the known properties of G4-DNA [3,4].

4. DISCUSSION

We have shown direct evidence for the ability of the consensus oligonucleotide of the ILPR to form quadriplex DNA of the G4 type in vitro. In G4-DNA, which is an intermolecular complex consisting of 4 oligonucleotides, the four strands are aligned in a parallel fashion (Fig. 3A). Whilst this structure is the predominant one formed under the conditions used here, within the ILPR in its natural context, because all oligonucleotides are present within one continuous DNA strand, the quadriplex structure must involve antiparallel orientation of strands, and thus be of the G4′ type [3].

Both oligonucleotide S and L gave rise to a second less intense band than the G43 complex (complex A in Fig. 1). For oligonucleotide S this is clearly visible above the G4 complex, whereas for oligonucleotide L, because it ran very close to the uncomplex linear oligonucleotide, it was only visible in shorter exposures of the autoradiograph (data not shown). Complex A could represent G′2 DNA, where two oligonucleotides fold back on themselves and then pair in a parallel fashion to form a quadriplex structure containing only two oligonucleotides. However, we favour the interpretation that complex A is similar to product K observed by Sen and Gilbert [3], where the two oligonucleotides pair in an antiparallel manner. This latter structure would be further stabilised by base pairing between the ACA and TGT nucleotides between the runs of G's (Fig. 3B).

In addition, oligonucleotide S forms a series of complexes with slower mobility than G4 or complex A (complex B in Fig. 1). We believe these to be examples of tetramolecular antiparallel structures not previously reported. Apart from forming a quadriplex structure, complex B is able to multimerise, as clearly evidenced by the appearance of a ladder. This structure would be stabilised by base pairing between all the nucleotides between the runs of G's, and therefore must involve antiparallel orientation of strands. The most likely structure adopted by complex B is shown in Fig. 3C. Further support for the existence of this structure comes from the fact that oligonucleotide L does not form a complex B. This would be expected because the unrelated nucleotides on each side of the ILPR consensus oligonucleotide would not allow the formation of multimers.

The data presented here show that the ILPR can adopt quadriplex structures. Taken with our previous findings that a plasmid harbouring the human insulin gene and the ILPR exhibited an altered DNA structure under torsional stress [2], these results show that quadriplex structures exist within the context of the human insulin gene in the ILPR. Moreover, our further demonstration that the G4′ structure is exhibited in plasmid DNA assembled into chromatin in vitro (Hammond-Kosack, unpublished observation) suggests that this quadriplex structure could be exhibited by the ILPR in vivo. It will be of particular interest to determine whether this structure is exhibited constitutively or whether its presence is dependent on (or a prerequisite for) gene expression or the cell cycle stage. So far there is no evidence to suggest that the ILPR serves as a *cis*-acting element for insulin gene expression [8]. It is, however, feasible to envisage that the bending induced within the ILPR by formation of the G4′ structure could bring *cis* regulatory elements of the proximal and distal promoter into close proximity.

Quadriplex structures were first detected in oligonucleotides from the telomeric ends of chromosomes. However, their biological function is not known. It is possible that these quadriplex complexes could form during chromosome pairing involving telomeric ends of both chromosomes, thereby anchoring the chromosome to facilitate homologous recombination or similar processes. The ILPR could therefore serve as a further anchoring point. It is interesting to note that while long

tandem repeat stretches of quadriplex-forming oligonucleotides have only been reported for telomeres and the ILPR, many genes carry an oligonucleotide with close homology to these within their proximal 5′ regions. It would be of interest to investigate whether more tandem repeats of the consensus oligonucleotide are present within each chromosome.

REFERENCES

[1] Bell, G.I., Selby, M.J. and Rutter, W.J. (1982) Nature 295, 31–35.
[2] Hammond-Kosack, M.C.U., Dobrinski, B., Lurz, R., Docherty, K. and Kilpatrick, M.W. (1992) Nucleic Acids Res. 20, 231–236.
[3] Sen, D. and Gilbert, W. (1990) Nature 344, 410–414.
[4] Williamson, J.R., Raghuraman, M.K. and Cech, T.R. (1990) Cell 59, 871–880.
[5] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning, A Laboratory Manual, 2nd edition, Cold Spring Harbor Laboratory Press, New York.
[6] Blackburn, E.H. (1991) Nature 350, 569–570.
[7] Sen, D. and Gilbert, W. (1991) Curr. Opin. Struct. Biol. 1, 435–438.
[8] Clark, A.R. and Docherty, K. (1992) in: Insulin: Molecular Biology to Pathology (F.M. Ashcroft and S.J. Ashcroft, eds). Oxford University Press, Oxford (in press).