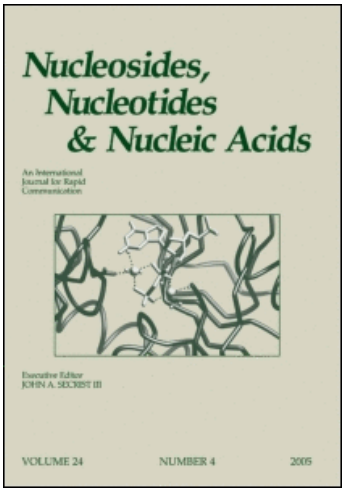

Nucleosides, Nucleotides and Nucleic Acids
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597286

Characterization of Insulin Ilpr Sequences for Their Ability to Adopt a G-Quadruplex Structure

Hanumantharao Paritala[a]; Steven M. Firestine[a]
[a] Department of Pharmaceutical Sciences, Eugene Applebaum College of Pharmacy and Health Sciences, Wayne State University, Detroit, Michigan, USA

Online publication date: 24 February 2010

To cite this Article Paritala, Hanumantharao and Firestine, Steven M.(2010) 'Characterization of Insulin Ilpr Sequences for Their Ability to Adopt a G-Quadruplex Structure', Nucleosides, Nucleotides and Nucleic Acids, 29: 2, 81 — 90
To link to this Article: DOI: 10.1080/15257771003597691
URL: http://dx.doi.org/10.1080/15257771003597691

Nucleosides, Nucleotides and Nucleic Acids, 29:81–90, 2010

ISSN: 1525-7770 print / 1532-2335 online
DOI: 10.1080/15257771003597691

Taylor & Francis
Taylor & Francis Group

CHARACTERIZATION OF INSULIN ILPR SEQUENCES FOR THEIR ABILITY TO ADOPT A G-QUADRUPLEX STRUCTURE

Hanumantharao Paritala and Steven M. Firestine

Department of Pharmaceutical Sciences, Eugene Applebaum College of Pharmacy and Health Sciences, Wayne State University, Detroit, Michigan, USA

□ *A major genetic factor linked to the progression of type 1 diabetes occurs in the insulin-linked polymorphic repeat region (ILPR) located 363 bp upstream of the human insulin gene. Genetic studies have shown that individuals with class I repeats (30–60) are predisposed to the development of type 1 diabetes while individuals with longer repeats are protected. Previous research has suggested that some sequences found within the ILPR can adopt a G-quadruplex structure, and this finding has lead to speculation that G-quadruplexes may control insulin expression in certain circumstances. Unfortunately, relatively little study has been done on whether sequences found in the ILPR can adopt a quadruplex fold. In this study, we have utilized circular dichroism, thermal difference spectroscopy and ultraviolet (UV) melting studies to examine the first seven common repeat sequences (**A-G**) found in the ILPR. We find that sequences **A-E** adopt a quadruplex fold while sequences **F** and **G** likely do not. Examination of sequence **B** and a single nucleotide variant, **B2**, revealed that both folded into a G-quadruplex. This result casts doubt on previous studies suggesting that the formation of a quadruplex was related to the ability of ILPR sequences to regulate transcription.*

Keywords insulin-linked polymorphic repeat region (ILPR); G-quadruplex; type 1 daibetes; human insulin gene; class I allele repeats

INTRODUCTION

Diabetes mellitus type 1 (type I diabetes, T1D, juvenile diabetes) is an autoimmune disease that results in destruction of the insulin producing beta cells of the pancreas. One of the genetic factors linked to T1D is the insulin linked polymorphic region (ILPR, also called INS-VNTR or IDDM2)

Received 17 July 2009; accepted 4 December 2009.

We thank Wayne State University and the Juvenile Diabetes Research Foundation (SMF) for funding. We are grateful to Dr. Andrew Feig (Department of Chemistry, WSU) for the use of his circular dichroism instrument.

Address correspondence to Steven M. Firestine, Department of Pharmaceutical Sciences, Eugene Applebaum College of Pharmacy, Wayne State University, 259 Mack Avenue, Detroit, MI 48201, USA. E-mail: sfirestine@wayne.edu

located 363 bp upstream of the insulin gene. The ILPR diabetes susceptibility locus maps to a minisatellite that contains a number of G-rich repeat sequences.[1–4] Genetic studies have identified two main classes of ILPRs in individuals, those with class I alleles (30–60 repeats) and those with class III alleles (120–170 repeats). Individuals with class I alleles are predisposed to the development of type-1 diabetes, while class III individuals are protected. Research has observed that insulin expression in class I individuals is three times lower in the thymus when compared to class III individuals.[3] This is believed to result in a failure to remove insulin-targeted antibodies, which leads to the destruction of the pancreas.

It has been shown that some of the G-rich repeat sequences found in the ILPR can fold into a G-quadruplex structure and some researchers have speculated that this is required for steady-state insulin transcription.[5–7] Researchers have postulated that the lower number of the G-rich repeat sequences within the ILPR of class I individuals is correlated with a lower probability of forming G-quadruplexes, thus, leading to a decreased expression of insulin in the thymus. Lew et al. conducted transcriptional activity experiments on three ILPR repeat sequences (**A**, **B**, **C**), and found that transcription was dependent on the sequence present and whether a quadruplex was formed.[6] The same group also found that compensatory changes in the repeat sequence could control quartet formation and transcriptional activity. Thus, seemingly minor nucleotide variations in these sequences result in wide ranges of transcriptional activity.[6] Recently, research has discovered that insulin and IGF-2 display affinity toward a G-quadruplex formed by a two-repeat length of one of the ILPR sequences.[8–11] This intriguing finding raises the possibility that insulin itself may play a role in the regulation of its own gene through association with G-quadruplexes formed in the ILPR. Despite these connections, few studies have been done on individual sequences to identify which sequences within the ILPR can adopt a quadruplex fold.[5,12–15]

The work presented here focuses on the ability of seven commonly observed ILPR repeat sequences to fold into a G-quadruplex.[3,5,16] To this end, we examined oligonucleotides containing a duplication of these seven sequences (**A-G**, see Table 1 for sequences). Each oligonucleotide was examined for the ability to adopt of G-quadruplex using thermal difference spectroscopy[17] and CD spectroscopy. To understand the stability of the quadruplexes, we also determined the melting temperature in the presence and absence of LiCl. We find that sequences **A-E** can fold into a quadruplex structure, but sequences **F** and **G** cannot. We also find that, in contrast to previous results, sequence **B** and a single nucleotide variant, **B2**, can adopt a quadruplex fold. We discuss this result with regard to the hypothesis that the presence of a quadruplex fold is necessary for transcriptional activity.

TABLE 1 The first seven commonly found sequences in the ILPR region of the insulin promoter and with their melting temperatures in the presence and absence of LiCl

Name	Sequence	T_m (°C)[a] KCl	T_m (°C)[a] LiCl
A	5′-ACAGGGGTGTGGGGACAGGGGTGTGGGG-3′	82	56
B	5′-ACAGGGGTCTGGGGACAGGGGTCTGGGG-3′	62	49
B2	5′-AGAGGGGTCTGGGGAGAGGGGTCTGGGG-3′	66	56
C	5′-ACAGGGGTCCTGGGGACAGGGGTCCTGGGG-3′	92	60
D	5′-ACAGGGGTCCCGGGGACAGGGGTCCCGGGG-3′	94	64
E	5′-ATAGGGGTGTGGGGATAGGGGTGTGGGG-3′	91	59
F	5′-ATAGGGGTGTGTGGATAGGGGTGTGTGG-3′	ND	ND
G	5′-ACAGGGGTCTGAGGACAGGGGTCTGAGG -3′	ND	ND

[a]Melting temperature determined from the first derivative plot of the melting curves shown in Figure 3. N.D. = not determined.

EXPERIMENTAL

Oligonucleotides

All of the oligonucleotides presented here were synthesized and gel purified by Integrated DNA Technologies (Coralville, IA, USA). Each oligonucleotide was dissolved in 200–500 μl of TE buffer, pH 7.4, and the concentration of the oligonucleotide was determined by absorbance using the supplied extinction coefficient for the oligonucleotide. Each oligonucleotide concentration was adjusted to 400 μM, aliquoted and stored at −20°C until needed.

Circular Dichroism Spectroscopy

Circular dichroism (CD) spectra were collected on a Jasco J-715 spectropolarimeter (Jasco Products Company, Easton, MD, USA) using a 0.1 cm path length quartz cuvette at room temperature. DNA samples were prepared at 25 μM strand concentration in 10 mM sodium cacodylate, 100 mM KCl, 1 mM EDTA, pH 7.2. Prior to data collection, samples were pre equilibrated at 25°C for 10 minutes. Spectra were collected from 200 to 350 nm at a speed of 100 nm/min. The final CD spectra of each oligonucleotide were the average of 5 accumulated spectra.

Thermal Difference Studies[17,18]

All UV absorption studies were performed by placing 5 μM oligonucleotide into either 10 mM sodium cacodylate, 100 mM NaCl, 1 mM EDTA, pH 7.4 buffer or 10 mM sodium cacodylate, 75 mM KCl, 15 mM LiCl, 1 mM EDTA, pH 7.4 buffer. Thermal denaturation profiles were obtained using a

Varian spectrophotometer (Palo Alto, CA, USA). Absorbance spectra were recorded in the 220–335 nm range, with a scan speed of 600nm/min using quartz cuvettes with a 1 cm pathlength. The maximum absorbance of the samples was between 0.5–1.2. Each oligonucleotide was scanned at both 25°C and 95°C and the difference between the two spectra were plotted against wavelength to obtain thermal difference spectra (TDS).

Melting Studies

Melting studies were performed by dissolving 5μM oligonucleotide in 10 mM sodium cacodylate, 1 mM EDTA, 100 mM KCl, or 100 mM LiCl, pH 7.4 buffer. Melting curves were obtained by measuring the absorbance of the sample at 295 nm as the temperature was increased from 25°C to 100°C. The temperature was increased in 1°C increments with a 2-minute equilibration at each temperature before the absorbance was measured. Melting curves for each oligonucleotide were normalized and plotted against temperature to obtain melting profiles of the oligonucleotides. The melting temperature (T_m) was determined from the first derivative plot.

RESULTS AND DISCUSSION

Characterization of the ILPR Sequences Using Circular Dichroism

Previous researchers have shown that the ILPR region contains analogs of the repeating sequence ACAGGGGTGTGGGG.[3,7] In this study, we studied seven commonly observed sequences found in this region for their ability to adopt a quadruplex structure (Table 1). To examine these sequences, oligonucleotides containing two, tandem copies of each sequence were synthesized and gel purified. Previous researchers have shown that at least two copies of an ILPR sequence is needed for quadruplex formation.[8]

In addition to the commonly observed ILPR sequences, we also choose to examine sequence **B2**, which is derived from **B**.[6] The **B2** sequence was utilized by Lew et al. to examine the relationship between quadruplex formation and insulin expression.[6] These investigators found that a promoter containing repeat copies of the **B** sequence possessed poor transcriptional activity, while promoters containing the **B2** sequence were maximally transcriptional activity. Thus, investigating the ability of an oligonucleotide containing the **B2** sequence to adopt a quadruplex structure would allow us to correlate quadruplex formation to transcriptional activity.

To characterize these oligonucleotides, we determined the topology of each sequence by CD spectroscopy. Previous investigations have shown that a CD signature featuring a negative peak at 240 nm and a positive peak at 260 nm typically indicates the formation of a parallel quadruplex, while a CD spectra containing a negative peak near 240 nm and a positive peak at

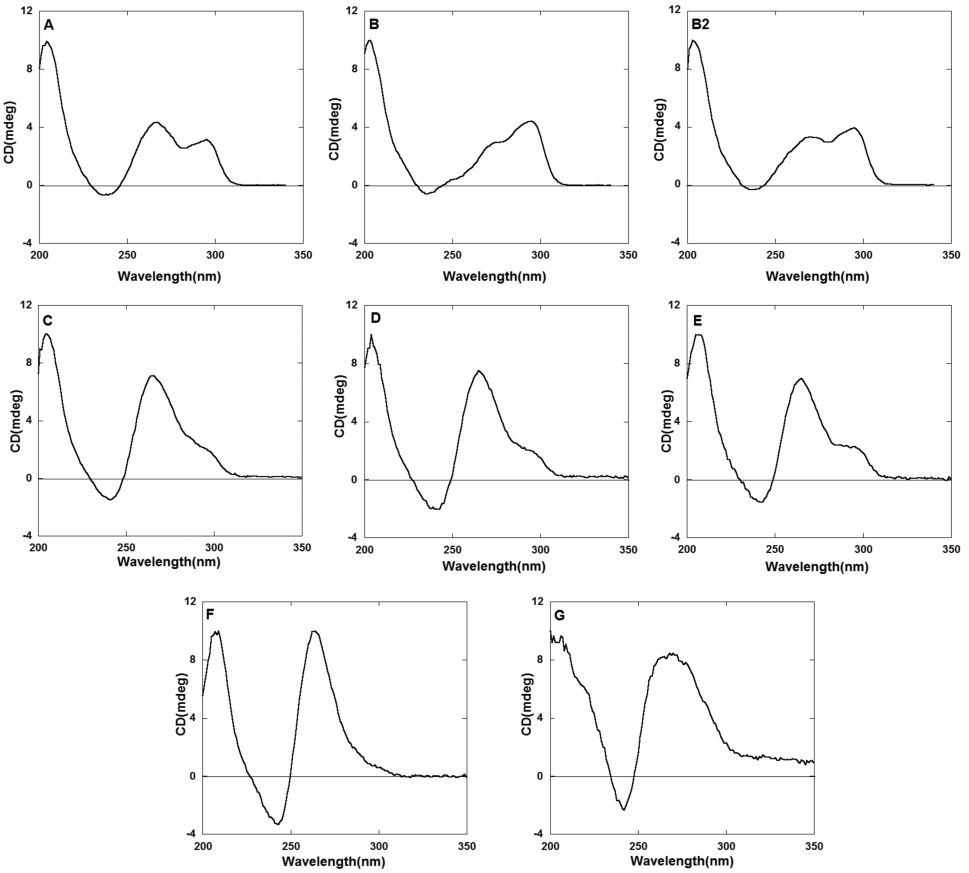

FIGURE 1 Circular dichroism (CD) spectra of oligonucleotides **A-G** and **B2.** The presence of a positive peak at 260 nm is indicative of a parallel quadruplex fold while a peak at 295 nm indicates an antiparallel quadruplex.

295 nm generally indicates an antiparallel quadruplex.[19,20] The CD spectra for each oligonucleotide in the presence of 100 mM KCl is shown in Figure 1. Oligonucleotides containing **A**, **B**, and **B2** sequences have positive peaks at both 260 and 295 nm indicating that these sequences possess a mixed (both parallel and antiparallel) quadruplex conformation under the conditions of the experiment. The observation that both **B** and **B2** exist as a quadruplex is in contrast to previous reports and suggests that quadruplex formation may not be related to transcriptional activity since both form quadruplexes yet only the **B2** sequence has full transcriptional activity.[6,7] These results are more in keeping with recent studies suggesting that insulin binding to ILPR sequences may play an important role in transcriptional regulation of the promoter.[8,10,11]

We also examined the remaining sequences for their ability to adopt a quadruplex fold. Oligonucleotides containing **C**, **D**, and **E** sequences adopt

a predominately parallel conformation, although the presence of a shoulder at 295 nm suggests that each sequence can also adopt a antiparallel conformation as well (Figure 1). An oligonucleotide containing sequence **F** showed a positive peak at 260 nm suggesting a parallel quadruplex conformation, while an oligonucleotide containing sequence **G** displayed a broad, positive peak ranging from 260 nm to 300 nm, which likely represents a mixed conformation. Taken together, the CD spectra presented here suggest that all of the oligonucleotides are capable of folding into a quadruplex.

Characterization of VNTR Sequences by Thermal Difference Spectroscopy

While circular dichroism is commonly used to determine the folding topology of nucleic acids, previous investigators have reported that, in some cases, the CD spectra of single-stranded G-rich oligonucleotides have features very similar to those for parallel folded quadruplexes.[21–23] This suggests that care should be taken in using CD spectra to determine the folding topology of nucleic acids. To validate or refute the structural conclusions derived from the CD spectra, we determined the TDS of each of the oligonucleotides. TDS provide a rapid and inexpensive method to determine the folding topology of nucleic acids.[17] This method relies on the fact that folded versus thermally denatured nucleic acids have different UV-vis spectra. The difference spectra are characteristic for a particular nucleic acid fold and, thus, can be used to diagnose the structure of the nucleic acid. For G-quadruplexes, a negative peak at 295 nm is indicative of this fold.

TDS of each oligonucleotide were determined using the recommended buffer (100mM NaCl, 10mM sodium cacodylate, and 1mM EDTA, pH 7.4). As seen in Figure 2, only sequences **A**, **B**, **B2**, and **E** generated a TDS indicative of a quadruplex structure. The remaining sequences gave spectra consistent with single-stranded DNA under these conditions.

The results from the TDS conflict with those obtained from the CD studies. TDS relies upon differences in UV properties between folded and unfolded states. If the quadruplexes under investigation were exceptionally stable, the percentage of unfolded DNA would be low. This would result in a TDS which would not have a significant negative peak at 295 nm. Since the stability of quadruplexes can be altered by the presence of Li^+ ion, we repeated the thermal difference experiment using a buffer containing LiCl. The TDS generated under these conditions revealed that oligonucleotides **A-E** adopt a quadruplex fold; however, oligonucleotides **F** and **G** do not (Figure 2). Modification of the experiment (i.e. changes in LiCl concentration, salt composition and concentration, etc.) failed to identify conditions in which oligonucleotides **F** and **G** generated a TDS indicative of a quadruplex. While we cannot definitively indicate whether the ILPR repeat sequences **F** and **G** fold into a G-quadruplex, the known ambiguity of CD spectra for

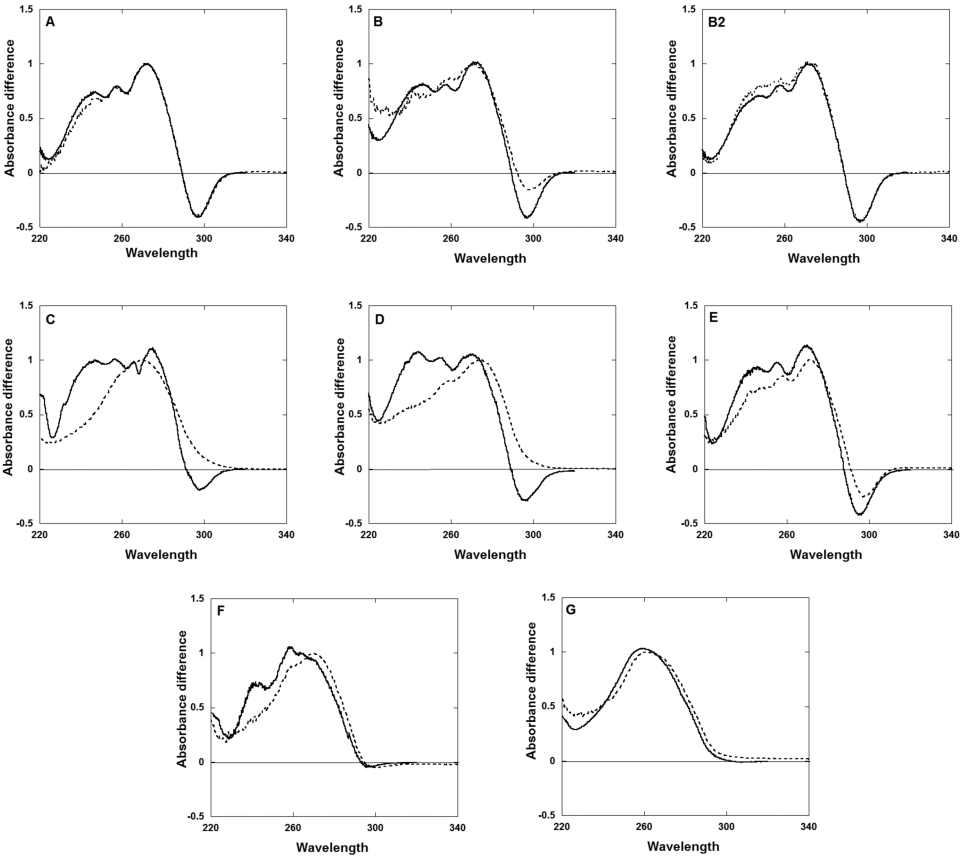

FIGURE 2 Thermal difference spectra of oligonucleotides (TDS) **A-G** and **B2**. The dashed line is the TDS in the presence of 100 mM NaCl while the solid line is the TDS in the presence of 15 mM LiCl. A negative absorbance difference at 295 nm is indicative of quadruplex formation.

G-rich sequences suggests that it is likely that **F** and **G** are not quadruplexes.[21–23] Thus, the CD spectra for these sequences likely represent other topologies in which guanines are stacked.

Thermal Melting Profiles of Sequences A-E

Our results from the TDS suggested that quadruplexes formed from these oligonucleotides may be very stable. To examine this and to verify the formation of a quadruplex from oligonucleotides **A-E**, we conducted UV melting studies in the presence and absence of LiCl. The unfolding of the G-quadruplexes was measured by a decrease in absorbance at 295 nm (Table 1, Figure 3).[24] The melting temperature (T_m) of the oligonucleotides, in the presence of K^+ ion, ranged from 62°C to 94°C demonstrating that these sequences fold into remarkably stable quadruplexes. When the concentration of KCl was dropped to 75 mM and 15 mM LiCl was included

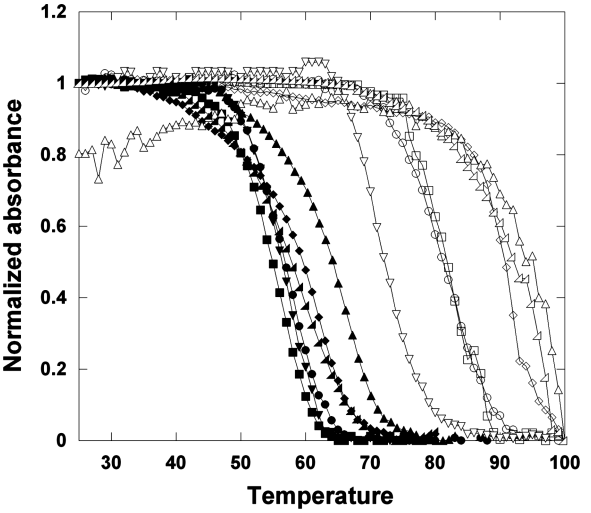

FIGURE 3 Ultraviolet (UV) melting studies for ILPR sequences in the presence (dark symbols) and absence (open symbols) of Li^+ ion. The melting curves correspond to the following sequences: (●) A with Li^+; (○) A without Li^+; (■) B with Li^+; (□) B without Li^+; (▼) B2 with Li^+; (▽) B2 without Li^+; (⧫) C with Li; (◇) C without Li^+; (▲) D with Li^+; (△) D without Li^+; (◢) E with Li^+; (◿) E without Li^+.

in the buffer, there was a dramatic shift in the melting curve to lower temperatures. In general, the melting temperature changed by 10–30°C when Li^+ ion was present. This decrease in the presence of LiCl is characteristic for G-quadruplexes and thus validates the conclusions derived from the CD and TDS.

CONCLUSION

Previously, researchers studying the relationship between sequences found in the VNTR and transcription found that the presence of sequence **B** resulted in low transcription, whereas a single nucleotide change to sequence **B2** resulted in enhanced transcription.[8] These researchers speculated that the difference in transcriptional activity was related to their ability to adopt a quadruplex structure. However, the ability of **B** and **B2** to fold into a quadruplex was based upon changes in band migration in a nondenaturing polyacrylamide gel.[6] To gain additional information on the ability of these sequences to adopt a quadruplex structure, we conducted circular dichroism spectroscopy, thermal difference spectroscopy and UV melting studies. These studies reveal that sequences **A-E** and **B2** adopt a quadruplex structure, while oligonucleotides containing sequences **F** or **G** likely do not. In contrast to previously reported results, our data indicate that both **B** and **B2** fold into a quadruplex.[6] This finding indicates that the observed differences in transcription must be due to other factors besides the adoption of a folded structure.

How does the ILPR region effect the formation of T1D? Recent studies have shown that insulin can bind ILPR quadruplexes suggesting that insulin expression may be mediated by a complex feedback loop.[8–11] Additional studies have also shown that transcriptional proteins are also capable of binding to oligonucleotides containing sequences from the ILPR.[6] This suggests that the control of insulin expression by the ILPR region may be mediated by the ability of these sequences to recruit transcriptional proteins to the promoter. Recently, we have shown that a unique set of quadruplex binding compounds can bind sequences within the ILPR.[25] If these agents have the ability to modulate protein binding to ILPR quadruplexes, they may be useful in unraveling the complexities of insulin regulation by this upstream region. Additional studies aimed at these questions will be reported in due course.

REFERENCES

1. Bennett, S.T.; Lucassen, A.M.; Gough, S.C.L.; Powell, E.E.; Undlien, D.E.; et al. Susceptibility to human type-1 diabetes at Iddm2 is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nature Genetics* **1995**, 9(3), 284–292.
2. Bennett, S.T.; Todd, J.A. Human type 1 diabetes and the insulin gene: Principles of mapping polygenes. *Ann. Rev. Genet.* **1996**, 30, 343–370.
3. Vafiadis, P.; Ounissi-Benkalha, H.; Palumbo, M.; Grabs, R.; Rousseau, M.; et al. Class III alleles of the variable number of tandem repeat insulin polymorphism associated with silencing of thymic insulin predispose to type 1 diabetes. *J. Clin. Endocrinol. Metab.* **2001**, 86(8), 3705–3710.
4. Bennett, S.T.; Wilson, A.J.; Cucca, F.; Nerup, J.; Pociot, F.; et al. IDDM2-VNTR-encoded susceptibility to type 1 diabetes: Dominant protection and parental transmission of alleles of the insulin gene-linked minisatellite locus. *Journal of Autoimmunity* **1996**, 9(3), 415–421.
5. Hammond-Kosack, M.C.; Kilpatrick, M.W.; Docherty, K. The human insulin gene-linked polymorphic region adopts a G-quartet structure in chromatin assembled in vitro. *J. Mol. Endocrinol.* **1993**, 10(2), 121–126.
6. Lew, A.; Rutter, W.J.; Kennedy, G.C. Unusual DNA structure of the diabetes susceptibility locus IDDM2 and its effect on transcription by the insulin promoter factor Pur-1/MAZ. *Proc. Natl. Acad. Sci. USA* **2000**, 97(23), 12508–12512.
7. Catasti, P.; Chen, X.; Moyzis, R.K.; Bradbury, E.M.; Gupta, G. Structure-function correlations of the insulin-linked polymorphic region. *J. Mol. Biol.* **1996**, 264(3), 534–545.
8. Connor, A.C.; Frederick, K.A.; Morgan, E.J.; McGown, L.B. Insulin capture by an insulin-linked polymorphic region G-quadruplex DNA oligonucleotide. *J. Am. Chem. Soc.* **2006**, 128(15), 4986–4991.
9. Wang, Y.; Zhang, H.; Ligon, L.A.; McGown, L.B. Association of insulin-like growth factor 2 with the insulin-linked polymorphic region in cultured fetal thymus cells. *Biochemistry* **2009**, 48(34), 8189–8194.
10. Xiao, J.; Carter, J.A.; Frederick, K.A.; McGown, L.B. A genome-inspired DNA ligand for the affinity capture of insulin and insulin-like growth factor-2. *J. Sep. Sci.* **2009**, 32(10), 1654–64.
11. Xiao, J.; McGown, L.B. Mass Spectrometric Determination of ILPR G-Quadruplex Binding Sites in Insulin and IGF-2. *J. Am. Soc. Mass Spectrom.* **2009**, 20(11), 1974–1982.
12. Hammond-Kosack, M.C.; Dobrinski, B.; Lurz, R.; Docherty, K.; Kilpatrick, M.W. The human insulin gene linked polymorphic region exhibits an altered DNA structure. *Nucleic Acids Res.* **1992**, 20(2), 231–236.
13. Hammond-Kosack, M.C.; Docherty, K. A consensus repeat sequence from the human insulin gene linked polymorphic region adopts multiple quadriplex DNA structures in vitro. *FEBS Lett.* **1992**, 301(1), 79–82.

14. Schonhoft, J.D.; Bajracharya, R.; Dhakal, S.; Yu, Z.; Mao, H.; et al. Direct experimental evidence for quadruplex-quadruplex interaction within the human ILPR. *Nucleic Acids Res.* **2009**, 37(10), 3310–3320.
15. Schonhoft, J.D.; Das, A.; Achamyeleh, F.; Samdani, S.; Sewell, A.; et al. ILPR repeats adopt diverse G-quadruplex conformations that determine insulin binding. *Biopolymers* **2009**, 93(1), 21–31.
16. Bennett, S.T.; Todd, J.A. Human type 1 diabetes and the insulin gene: principles of mapping polygenes. *Annu. Rev. Genet.* **1996**, 30, 343–370.
17. Mergny, J.L.; Li, J.; Lacroix, L.; Amrane, S.; Chaires, J.B. Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.* **2005**, 33(16), e138.
18. Mergny, J.L.; Phan, A.T.; Lacroix, L. Following G-quartet formation by UV-spectroscopy. *FEBS Letters* **1998**, 435(1), 74–78.
19. Kumar, N.; Maiti, S. A thermodynamic overview of naturally occurring intramolecular DNA quadruplexes. *Nucleic Acids Res.* **2008**, 36(17), 5610–22.
20. Paramasivan, S.; Rujan, I.; Bolton, P.H. Circular dichroism of quadruplex DNAs: applications to structure, cation effects and ligand binding. *Methods* **2007**, 43(4), 324–331.
21. Lane, A.N.; Chaires, J.B.; Gray, R.D.; Trent, J.O. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **2008**, 36(17), 5482–515.
22. Kypr, J.; Fialova, M.; Chladkova, J.; Tumova, M.; Vorlickova, M. Conserved guanine-guanine stacking in tetraplex and duplex DNA. *Eur. Biophys. J.* **2001**, 30(7), 555–558.
23. Kypr, J.; Vorlickova, M. Circular dichroism spectroscopy reveals invariant conformation of guanine runs in DNA. *Biopolymers* **2002**, 67(4–5), 275–277.
24. Mergny, J.L.; Phan, A.T.; Lacroix, L. Following G-quartet formation by UV-spectroscopy. *FEBS Lett.* **1998**, 435(1), 74–78.
25. Paritala, H.; Firestine, S.M. Benzo(h)quinoline derivatives as G-quadruplex binding agents. *Bioorg. Med. Chem. Lett.* **2009**, 19(6), 1584–1587.