Introduction
INTRODUCTION

Lipoprotein(a) belongs to a group of spherical particles (lipoproteins), transports variable amounts of cholesterol and triglyceride in the blood circulation (Björkegren, J., 2001). Lipoprotein particles are divided into several major groups including low-density lipoprotein (LDL), high-density lipoprotein (HDL), and very low-density lipoprotein (VLDL). High levels of HDL cholesterol have been shown to compensate for some of the harmful effects of LDL cholesterol (P de Silva, P., 2005). Lp(a) exhibit a significantly lower cholesteryl ester content and similar clearance rate as those for LDL (Demant, T., 2001 and Rader, D. J., 1993). Lp(a) is heterogeneous in mass, density, and electrophoresis mobility (Superko, H.R., 2008). Lp(a) is certainly non essential, since a considerable number of individuals have no detectable Lp(a) with no apparent consequence though plays an important role in wound healing (McCormick, P. A. S., 2004). Lp(a) has an affinity for many components of the subendothelial matrix including proteoglycans, (Lundstam, U., 1999) fibrinogen and fibronectin (Nielsen, L.B., 1996). Lp(a) is concentrated in the artery wall by virtue of binding to fibrin, plasminogen receptor, matrix, and other targets (Cushing, G. L., 1989, Pepin, J. M., 1991 and Kreuzer, J., 1994). Apo(a), one of the components of Lp(a), competes for the binding and activation of plasminogen because of an inactive homology and interferes with clot lysis (Sangrar, W., 1997). Apo(a) delivers cholesterol to wound site for cell biosynthesis (Lawn, R. M., 1997) but condition becomes pathogenic at the time of elevated plasma concentration. It inhibits plasminogen activation and the prolonged presence of thrombus on the vessel wall may promote the growth and migration of smooth muscle cells and the development of atherosclerotic lesions through several intermediate pathways (Wagner, D. D., 2008).
Introduction

Atherosclerosis and Lipoprotein(a) level

Lipoprotein(a) [Lp(a)] was first described by Kare Berg in 1963 as an antigen in human plasma whose concentration is significantly correlated with the risk of atherosclerosis (Berg, 1963). Lipoprotein(a) [Lp(a)] is, one of the most atherogenic lipoprotein present only in human plasma, Old World primates, and the European Hedgehog (Utterman, G., 1989 and Lawn, R. M., 1995, 1996). Plasma concentration of Lp(a) varies over a wide range among individuals in a population but remain remarkably constant within an individual over time (Albers, J. J., 1990). The plasma Lp(a) concentration in humans ranges from 0 to >200mg/dl and there is general agreement that plasma levels of >30mg/dl are casually linked to atherosclerosis, coronary heart disease, ischemic cerebrovascular disease, chronic thromboembolic pulmonary hypertension and stroke (Depka, M. V., 2000). Although Lp(a) is an LDL-like lipoprotein consisting of lipids and apoB-100 (M_r ~ 550K) molecule, it differs from LDL in having an additional hydrophilic glycoprotein called apolipoprotein(a) [apo(a)] (Becker, L., 2001). Apo(a) is covalently linked to apoB via a disulfide bridge (Horejsi, B., 2000, and Byrne, C. D., 1994), and can also associate noncovalently with apoB-100 (Fless, G. M., 1985 and Trieu, V.N., 1991). Metabolic and genetic evidence shows that Lp(a) synthesis, rather than catabolism, is the key determinant for plasma level concentration which is not affected by diet (Chan, D. C., 2004). Lp(a) levels are virtually entirely heritable (Austin, M.A., 1992, and Boomsma, D.I., 1993) and that the apo(a) gene accounts for almost the entire heritability (Kraft, H., 1992, Boerwinkle, E., 1992, and De Meester, C., 1995).
Apolipoprotein(a)

Apolipoprotein (a) [apo(a)], a gene of human telomeric region of chromosome 6 region 6q26-27, a major component of lipoprotein (a) [Lpa (a)] has been established as one of the major risk factor for atherosclerosis (Wu, Q., 2007 and Mooser, 1995). Apolipoprotein(a) gene in human and hedgehog lineages accounting for its phylogenetic distribution in primates, hedgehogs, and at present, no other species reveals a novel and fascinating pathway for the evolution of genes (Lawn, R. M., 1995). Apo(a) is potentially athero-thrombogenic, a property which may be influenced by its size, sequence polymorphism, type of lipoprotein it is linked to and the inflammatory state of the vessel wall (Scanu, A.M., 2001). This gene contains multiple IL6 response elements (Wade, D.P., 1993) and cell culture studies have shown that apo(a) gene expression is up-regulated by IL6, leading to an accumulation of Lp(a) particles (Ramharack, R., 1998). Apo(a) fragments ranging from 60 to 220 KDa are excreted in urine and the excretion rate correlates significantly with the plasma levels of Lp(a) (Kostner, K., 2001).

Figure 1 Model of Lp(a) assembly shows apo(a) interaction by disulphide bonds (McCormick, P. A. S., 2004)
A recent study by Pelligrino et al. (Pellegrino, M., 2004) has identified yet another mechanism for Lp(a)’s involvement in atherogenesis by showing that the apo(a) component of Lp(a) induces a rearrangement of actin fibres in cultured endothelial cells. This leads to a loss of cell to cell contact which may contribute to the initial damage and increased endothelial layer permeability that precedes the development of atherosclerotic lesions.

Apolipoprotein(a) Gene Polymorphism

In Caucasians, African-American, and Asian populations variation in this gene is the major determinant of the plasma concentrations of the atherogenic lipoprotein(a) which varies enormously between individuals and considerably across populations (Ordovas, J. M., 2004). Apo (a) is synthesized almost in the liver. Upon entering into the blood stream apo (a) associate with circulating LDL and forms a mature Lp(a) where apo(a) is linked to LDL by a disulphide bridge. Although apo(a) gene accounts for greater than 90% (Utermann, G.,1999) of the variation in plasma lipoprotein(a) concentrations, there is still a pronounced variation in Lp(a) concentrations among individuals carrying apo(a) isoforms of the same size (Belotserkovskii, 2007). The cDNA of the apo(a) gene exhibits variable length due to variable number of repeat sequences within the gene (McLean, J.W., 1987, Lindahl, G., 1990, and Koschinsky, M., 1990). In addition to the size variation, this gene contains other polymorphisms, including a pentanucleotide repeat (PNR) polymorphism, (TTTTA)ₙ, in the 5'-flanking region of the gene (Rogers, F. A., 2005) and three single nucleotide substitutions (+121 G>A, +93 C>T, and -772 G>A) (Trommsdorff, M., 1995 and Zysow, B. R., 1995). In contrast to the other polymorphic sites (-772 G>A), the substitution was not reported to be functional. The apo(a) gene that resides on human chromosome
6 also has homology with plasminogen gene but with various degree of nucleotide variations even at promoter region.

Three types of variation have been identified in the apo(a) gene: a size polymorphism in the coding region (KIV type 2 repeats), a pentanucleotide repeat (PNR) polymorphism in the promoter upstream region \((TTTTA)_n\) and sequence variation in coding and non-coding regions of the gene including a C/T polymorphism at +93. This creates an additional ATG start codon and also affects transcription. Three widely studied apo(a) gene polymorphisms apo(a) size, +93 C/T and pentanucleotide repeat region (PNR) could not give complete information about isoforms of apo(a) and level of lipoprotein(a) in the individual separately. The association between level of lipoprotein(a), apolipoprotein(a) size and the \((TTTTA)_n\) polymorphism, located in the 5' non-coding region of the apo(a) gene, is correlated with severe coronary heart disease.

Although the apo(a) size polymorphism is an important predictor of Lp(a) level with an inverse relationship between apo(a) size and Lp(a) concentration, the latter vary considerably among individuals carrying apo(a) isoforms of the same size, implicating the presence of other predictors (Paultre, F., 2000, Marcovina, S. M., 1996, and Geethanjali, F.S., 2003). The size variation of apo(a) gene is strongly controlled by pentanucleotide repeat (PNR) polymorphism \((TTTTA)_n\).

**Pentanucleotide repeats (PNR) polymorphism of apo(a)**

Lp(a) levels are independently associated with two repeat polymorphism at the apo(a) gene; apo(a) protein sizes reflecting the Kringle-4 domain repeat polymorphism and the genotype of the pentanucleotide \((TTTTA)_n\) repeat. The basal level of apo(a) seem to be controlled by 1.4 kb promoter upstream TTTTA repeat polymorphism. Several polymorphisms of the apo(a) gene have been
described in the 5’ untranslated region, in introns (Wade, D. P., 1993, Ichinose, A., 1995), and in the coding sequence (Vanderhoek, Y.Y., 1993, and Prins, J., 1997). Lp(a) concentration and (TTTTA)_n polymorphism are the most significant predictors of CHD, beside age, diet and environment. In obese postinfarction patients, elevated lipoprotein(a) [Lp(a)] was found to be a highly significant risk marker rather then genetic or blood marker variables (Corsetti, J.P., 2008). The published data suggested that (TTTTA)_n polymorphism strongly affects the Lp(a) expression independently of apo(a) size polymorphism in CHD. The pentanucleotide (TTTTA)_n repeat polymorphism, located in the 5’ untranslated region, has been shown to be associated with plasma Lp(a) levels in healthy Caucasians (Mooser, V., 1995, Trommsdorff, M., 1995, Suzuki, K., 1997, and Valenti, K., 1999).

Several studies have demonstrated an inverse relationship between PNR (pentanucleotide repeats) number and Lp(a) concentration among Caucasians; alleles containing more repeats were generally associated with lower plasma Lp(a) concentration (Zlatohlavek, L., 2008, Mooser, V., 1995, Trommsdorff, M., 1995 and Valenti, K., 1999). It is likely that a major part of this correlation is attributable to additional genetic factors (Boerwinkle, E., 1992), but these factors have not been completely characterized in detail. Studies on Blacks have been carried out mostly in South Africans and Khoi San, and data on African Americans are limited (Mooser, V., 1995). In a previous study by Brazier et al (Brazier, L., 1999), the (TTTTA)_n polymorphism was found to be significantly associated with Lp a levels in both healthy and CHD subjects. A borderline association between the (TTTTA)_n polymorphism in the 5’ region and CHD was reported in a small study (Ameniya, H., 1996). In addition, the repeats polymorphism of
pentanucleotide sequence (TTTTA)$_n$ at position -1373 before the translation initiation codon of the apo(a) gene is likely to be one of the factor associated with the plasma Lp(a) concentration. This polymorphism may account for ~10-40% of the inter-individual variation of Lp(a) level of Caucasians, Europe, Africa and Asia.

**PNR polymorphism and C/T correlation**

The PNR 9 allele is more common on a T background (67%) compared to a C background (13%), while the PNR 8 allele is more common on a C background (64%) compared to a T background (19%), indicating linkage disequilibrium between the two loci for Caucasians. PNR 10 alleles are more common among Caucasian than among African American C alleles (15.4% vs. 4.5 %). In contrast, small PNR alleles (PNR <8) are more common among African American than among Caucasian C alleles (16.7% vs. 3.8 %) (Rubin, J., 2008). Thus, after controlling for the C/T polymorphism, the PNR distribution remained significantly different in Caucasians compared to African Americans. The small PNR alleles are more common among African Americans, while large PNR alleles are more common among Caucasians. In Caucasians the C/T polymorphism is in linkage disequilibrium with both the apo(a) gene size and the PNR polymorphisms; the T allele is more commonly associated with larger apo(a) size and the PNR 9 allele. Further, the PNR polymorphism is in linkage disequilibrium with the apo(a) gene size in Caucasians but not in African Americans. Among Caucasians, the PNR 9 allele is more commonly associated with larger apo(a) sizes and the PNR 10 and 11 alleles with smaller apo(a) sizes. Mooser et al (Mooser, V., 1995) reported, using a family study, that carriers of PNR 11 repeats had exclusively small apo(a) size, in their study.
Introduction

Homologous/Heterologous Alleles of PNR

Allele variant of pentanucleotide repeats of apolipoprotein(a) has been reported to be associated with elevated plasma lipoprotein(a) [Lp(a)] and increased cardiovascular risk. Allelic frequencies of polymorphic repeats of apo(a) detected in all genotypes, are related to Lp(a) levels which differ from individual to individual depending upon homologous and heterologous repeats (Wade, D. P., 1993). Plasma Lp(a) levels were significantly higher in individuals homozygous for the 8(TTTTA) repeats allele than in those homozygous for the 9(TTTTA) repeat allele in Japanese individuals. Alleles containing eight PNRs are more common in African American and Caucasians ethnic group (Rubin, J., 2001). Although the average apo(a) protein size between the homozygotes for the 8 and 9 (TTTTA) repeats allele were not significantly different, the 9(TTTTA) repeats appeared to be associated with the narrower range of apo(a) protein size than the 8(TTTTA) repeat in Japanese. A tendency to associate with the severity of atherosclerosis in CHD for the genotype homozygous for the 8(TTTTA) repeat allele was also found. Allele frequency study in Danish population showed elevated level of Lp(a), carried apolipoprotein (a) 14-15 pentanucleotide repeats have 3 to 2-fold increased risk of myocardial infarction (MI) and ischemic heart disease (IHD) whereas 6, 7, 8, 9, 10, 11, and 12 repeats have lipoprotein(a) levels 40, 36, and 27 mg/dl, respectively (Kamstrup, P. R., 2008).

Recently Portuguese obese children and adolescents aged 4 to 16 years were studied for lipoprotein level. Genomic DNA was extracted from a buffy coat by the proteinase K/salt precipitation method (Gaffney, D., 1994 and Olerup, O., 1992). The number of pentanucleotide repeat were confirmed by automated sequencing for homozygous 8/8, 9/9, and 10/10 repeats and for heterozygous 6/8, 7/8, 8/9,
Introduction

8/10, 8/11, and 9/10 repeats. The apolipoprotein(a) allele with (TTTTA) 8-repeat sequences was most prevalent (71.1%). The pentanucleotide (TTTTA)n repeat and G/A-914 polymorphisms at 5' promoter region of the apo(a) gene studied by Dincic, D., et al (Dincic, D., 2005) in 211 Serbian, has reported that the patient having at least one allele 8 (heterologous) had significantly higher Lp(a) values than controls. These results indicate a contribution of the pentanucleotide repeat polymorphism on serum Lp(a) concentration in an obese pediatric Portuguese population, with alleles with more repeats correspond to lower Lp(a) values (Nascimento, H., 2009). The negative regulatory element β (NREβ) (Negi, S., 2004) at 5' flanking region of apo(a), present alongwith PNR (pentanucleotide repeat region), does not have any structural effect on gene regulation.

Nucleic Acids and its Structures

The discovery of double-helical structure of the DNA (deoxyribonucleic acid) in 1953, by Watson, Crick, Wilkins and Franklin was an important milestone in modern science (Franklin, R. E, 1953, Watson, J. D., 1953, and Wilkins, M. H., 1953). The basic unit of life consists of two complementary DNA double strands made up of four bases (A, T, C, and G) linked through hydrogen bonding. Nucleic acids are polymorphic macromolecules and can exist in different forms such as left handed Z helices, cruciform (Lilley, D. M. J., 1980), heteronomous DNA (Greaves, D. R., 1985), bent DNA (Gartenberg, M.R., 1988), parallel DNA (van de Sande, 1988), quadruples DNA, etc., which may provide important signals for the control of gene expression. The torsional stress of duplex DNA that results from negative supercoiling is relieved through the formation of non-B-form DNA structures. The first example of triple helix was observed with ribonucleotide polymers (Felsenfeld, G., 1957). Formation of triple helical structures of nucleic
Introduction

acids, DNA as well as RNA, can occur inter- or intramolecularly. DNA secondary structure, one of the intrinsic components that have impact on regional genomic stability, associated with human diseases, can be affected by polypurine-polypyrimidine mirror repeat sequences (Valenti, K., 1999). Sequences containing overlapping Pu·Py tracts require significantly less energy to form triplexes than individual Pu·Py tracts. Pyrimidine-purine (pyr-pur) sequences are abundant in the eukaryotic genome and may represent up to 1% of total DNA (Birnboim, H. C., 1979, Hoffman-Liebermann, B., 1986, and Wong, A. K. C., 1990). Mirror-repeated Pu·Py tracts that form H-DNA occur as frequently as 1 in 50,000 bp of human genomic DNA (Schroth, G. P., 1995). Biological role of Pu·Py repeats have been supported by a statistical overrepresentation in the genome of eukaryotes and prokaryotes (Birnboim, H. C., 1978, Bucher, P., 1991, and Behe, M. J., 1995). The studies presented here demonstrate that mirror repeat Pu·Py tracts in DNA sequence of apo(a) regulatory region form intramolecular and intermolecular triplex structures under modest energy and environmental conditions and regulate transcription.

Triplex structure of DNA

Human genome, sequenced in the year 2001 (IHGSC, 2001), have provided us with hundreds of regions to think about the structure of DNA due to stretch of Pu·Py repeats. The non-B DNA structure not only changes the phenotype of structure, also imparts an effect on replication, transcription, recombination and induced genomic instability in vivo (Wang, G., 2008). An intramolecular DNA triplex is formed when pyrimidine or purine bases occupy major groove of the DNA double Helix forming Hoogsten pairs with purines/pyrimidine, already involved with hydrogen bonds to its complementary in the duplex DNA (Figure 2).
Introduction

Intermolecular triplexes are formed between artificially synthesised oligonucleotides, like triplex forming oligonucleotides (TFO) and target sequences on duplex DNA with polypurine/poly嘧midine repeats (Figure 3). Moreover, such triplex structures inhibits DNA polymerase elongation in vitro when triplex formation takes place at upstream.
Introduction

(Guieysse, A. L., 1995) or downstream (Diviacco, S., 2001, and Nagatsugi, F., 2003) of the initiation site. Homopurine or homopyrimidine mirror repeats can form triple helical DNA structures by interacting in a sequence specific manner (characterized by the base triplets T*A-T, G*G-C, A*A-T, and C*G-C) with the homopurine strand of duplex DNA (Frank-Kamenetskii, 1995). Within H-DNA, the triplex is stabilized by either Hoogsteen (YRY triplex) or reverse Hoogsteen (RRY triplex) interactions, depending upon whether the pyrimidine or the purine-rich strand participates in the triplex formation. The YRY triplex-containing structure is referred to as H-DNA or H-y-DNA, and the RRY triplex-containing structure is referred to as H*-DNA or H-r-DNA.

Triplex formation in eukaryotic cells can be recognised by antibodies against triple-helical DNA, which interact with eukaryotic chromosomes. The biological effects of Pu·Py tracts have not been fully elucidated yet. It has been studied that human c-MYC (Firulli, A. B., 1992) promoter which transcribe T7 RNA polymerase (Wang, G., 2004) has two mirror-symmetric homopurine-homopyrimidine block that has a tendency to form either DNA triplex (H-DNA) or quadruplex structures. H-DNA forms within homopurine-homopyrimidine mirror symmetrical repeats or H-palindromes has one-half of the H-palindrome denatured into two single strands, of which one complementary strand folds back to form a DNA triplex with the non-denatured half of H-palindrome, and the other strand remains denatured. The presence of these denatured regions makes this structure hypersensitive to single strand-specific endonucleases like S1 which are commonly used for detecting H-DNA. Because of this nuclease sensitive element, truncated transcription products are formed, an implication of partial transcription arrest.
Introduction

(Belotserkovskii, B.P., 2007). It is one of the best examples of DNA structure sensitivity and its effect on gene expression.

Studies by Albino Bacolla, et.al. (Albino Bacolla, 1995) on structures formed at the 5'-flanking region of the human γ-globin genes, both in the wild type as well as the HPFH (hereditary persistence of fetal hemoglobin) point mutations, report that change in a single nucleotide disturb the critical hoogsteen hydrogen pairing in the triplex structure formed by GGGCCC motif of the flanking region. Human haemoglobin is synthesized from two sets of clustered genes designated as α-cluster located on chromosome 16 and the β-cluster on chromosome 11. During early embryonic life in the placental yolk sac it develops in a tissue specific manner to regulate program that allow ζ and ε to be transcribed.

At subsequent stages of development, globin expression shifts to the α and γ genes in the red cells of hepatic origin. At the time of birth, β-chains are predominantly expressed, and erythropoiesis shifts to the bone marrow (Karlsson, S., 1985, and Collins, F. S., 1984). This pattern of expression can be altered by mutations affecting any of the transcribed genes (Karlsson, S., 1985, Collins, F. S., 1984, and Huisman, T. H. J., 1992), but a condition that has attracted particular attention is the hereditary persistence of fetal hemoglobin (HPFH) caused by point mutations at the 5'-flanking region of either one of the γ-genes. A single nucleotide mutation at the 5' flanking region of γ-globin gene cause the affected allele to permanently express high level of γ-chain in adult red cells.

Triplex structure in chromatin

The sequence of the human genome (IHGSC. 2001, and Venter, J.C., 2001) provides information about the nucleic acid sequence that defines the genetic code, and also suggests formation of multi-stranded DNA structures at specific sites in
the genome which can influence many cellular functions. Duplex DNA in the chromatin can form multi-stranded helical structures through folding of one of the two strands or association of two, three, or four strands of DNA. In follicular lymphomas, chromosomal translocation t(14;18) occurs at the bcl-2 major breakpoint region (Mbr) and in 150-bp of bcl-2 Mbr it contains three breakage hotspots (peaks) with a single-stranded character and non-B DNA (triplex DNA) confirmation (Raghavan, S. C., 2005). A well established multi-stranded DNA structure such as intramolecular (triplex helical DNA) and intermolecular (TFO targeted triplex) triplexes define the function of gene in vitro as well in vivo (Jain, A., 2008). Triplex formation in chromatin (a typically packed DNA histone structure in the nucleus) under chromosomal environment has been reported (Shahid, K.A., 2006, Besch, R., 2004, Brunet, E., 2006 and Ye, Z., 2007). The Py·Pu tracts of the human Na, K-ATPase α-3 gene, adjacent to a TATA box contains a 26-bp quasi-mirror repeat sequence, has potential to form intermolecular triplex (Potaman, V. N., 1995). Furthermore, Pu·Py tracts from PKD1 intron 21 form DNA intramolecular triplex structures under modest energy and inhibit replication (Bacolla, A., 2001). These tracts inhibit replication in eukaryotic replication systems as well. Such interference with DNA replication can lead to mutation (Friedberg, E. C., 1995). In autosomal dominant polycystic kidney disease (ADPKD), 85% patients have mutation in PKD1 gene. Intron 21 of the PKD1 gene contains the largest polypurine: polypyrimidine tract (2.5kb) identified in the human genome. Such Pu.Py mirror repeats are known to form intramolecular DNA triplexes in vitro (Lyamichev, V. I., 1985) and in vivo (Lee, J. S., 1989, Kohwi, Y., 1992, Kohwi, Y., 1993 and Ussery, D., 1993) and involved in genetic recombination (Kohwi, Y., 1993, Collier, D. A., 1988, Weinreb, A., 1990,
Kim, R. H., 1994, and Rooney, S. M., 1995), gene regulation (Firulli, A. B., 1994) mutagenesis (Kohwi, Y., 1993) and termination of DNA replication (Collier, D. A., 1988). It has been hypothesized that long polypyrimidine tract can form intramolecular triplex structures which predispose the normal allele to mutation and loss of function (Blaszak, R. T., 1999).

Naturally occurring DNA structures which are obstacle to transcription could be hot spots for mutations. Among such sequences that can be hinder transcription are those, that are able to form unusual (i.e. non-B-form) DNA structures. These structures are known to interfere with a number of enzymes that operate on DNA, and they cause genomic instability in many systems (Wang, G., 2006, Mirkin, S. M., 2006 and Bacolla, A., 2006). Transcription through a particular DNA region can also increase the rate of mutation in this region (transcription-assisted mutagenesis) or create a hot spot for homologous recombination (transcription-assisted recombination) (Aguilera, A., 2002). As a consequence, DNA can be attacked during transcription. This is because single-stranded DNA is more sensitive than double-stranded DNA to attack from a number of agents, including some DNA-modifying enzymes. Mutagenesis induced by triplex-forming DNA oligonucleotides, known to arrest transcription, appear to be dependent upon TCR (transcription-coupled DNA repair) (Duquette, M. L., 2005) factors (Wang, G., 1996).

Triplex DNA and gene regulation

Triplex H-DNA control gene regulation either directly or inhibiting the binding of regulating factors on the target site by disturbing the structure of DNA. The diseases inherited by non-Mendelian genetic mechanisms have been associated with the expansion of tandemly repeated DNA motifs which have
Introduction

Structural differences because of repetition (Koide, R., 1994, and Nagafuchi, S., 1994). Triplex formation due to Pu-Py repeats has been shown to present a potent block for DNA replication and transcription (Dayn, A., 1992, Krasilnikov, A.S., 1997, Potaman, V.N, 1999 and Sambrook, J., 1989). However, the features of Pu-Py tracts which favor the inhibition of replication and transcription are less clear. Those DNA strands containing more than four closely localized guanine blocks have three or more guanines each and can fold into another unusual DNA structure a DNA quadruplex, stabilized by G-quartets (Davis, J. T., 2004). DNA quadruplexes are known to be present in many biologically important sequences, including telomeres (Davis, J. T., 2004) and regulatory elements of the c-MYC promoter. (Simonsson, T., 1998 and Yang, D., 2006). The primer extension reactions following chloroacetaldehyde and potassium permanganate modification establish that a mirror repeat Pu-Py tract from the PKD1 gene can form triplex DNA (Blaszak, R.T., 1999). Tiner and coworkers captured the images of intramolecular triplex structure by atomic force microscopy (Tiner, W.J., 2001). Transcription inhibition due to triplex has been described for plasmid harbored genes (Faria, M., 2001, Arimondo, P .B.,2006, and Bailey, C., 2000) for foreign sequences in the cellular genome (Faria,M.,2002), and in several endogenous genes including c-myc (Hacia, J. G., 1994, and Napoli, S., 2006), ets2 (Carbone, G. M., 2003), tie1 (Hewett, P. W., 2006), HER2/ neu (Ebbinghaus, S.W., 1999), bcr/abl (Rapozzi, V., 2002), the inflammatory mediators TNF-α (Aggarwal, B. B., 1996), MCP-1 (Marchand, P., 2000), or GMF/CSF (Kochetkova, M., 1997) and the cell adhesion molecule ICAM-1 (Besch, R., 2002).
Introduction

Triplex DNA Structural variation and Diseases

DNA sequence variation involved in disease or other biological traits of interest are continuously being identified. Both exploring and exploiting DNA sequence data, therefore, create a huge need for highly efficient methods of genome manipulation (Kraft, H., 1998). For example, cancer is a genetic disease, the revolutionary discoveries in the identification of altered genes in human cancers have made it clear that specific targets can be identified in tumor cells, whose functions are known to cause or contribute to the initiation of disease progression (Rogers, F. A., 2005). These targets might be specifically blocked by molecules designed to alter or inhibit their function (gene expression). In fact, triplex formation can modify the gene structure and functions both in vitro and in vivo leading to site-specific modulation of gene expression, modulation of protein binding, targeting of DNA damage, mutagenesis and enhancement of homologous recombination (Rogers, F. A., 2005, Besch, R., 2004, Faria, M., 2001, Vasquez, K. M, 2002 and Maurisse, R., 2002). It has been described that cell cycle could also influence triplex formation (Majumdar, A., 2003). The activation of transcription has also been demonstrated in vivo with hairpin TFO in Saccharomyces cerevisiae (Ghosh, M. K., 2005).

Triplex Forming Oligonucleotide (TFO)

In the genomic era, TFOs (Triplex Forming Oligonucleotides) have acquired critical importance in the field of biotechnology and have been used as biotechnological tools in various assays (Pingoud, A., 2007), for example site directed mutagenesis, translocation of proteins on DNA (Levy, O., 2005), topoisomerase activity (Maxwell, A., 2006) and to target cleaving or cross-linking agents, transcription factors or nucleases for a chosen site on DNA. Structural
requirements influence the design of TFOs and their classification into different subtypes: TFOs containing C and T nucleotides, also named pyrimidine TFOs, bind in parallel orientation via Hoogsteen hydrogen bonds [(T, C) - motif] and TFOs containing G and A or T nucleotides [(G, A) - and (G, T)-motif] mainly bind in antiparallel orientation via reverse Hoogsteen bonds. Chromatin structure precludes access of TFOs to target sequences in the chromosomes. The locked nucleic acid-containing TFOs (TFO/LNAs) have been introduced to target chromosomal DNA in cells, for probing chromatin structure and function (Brunet, E., 2006).

The binding of TFO to the target duplex generally results in a thermodynamically weaker interaction than the one observed between two strands of the duplex itself (Shafer, R. H. 1998). TFOs backbone modification,, such as PS and PN, have been conceived in order to change the electrostatic properties of the negative phosphodiester backbone of natural DNA molecules and achieve greater degree of nuclease resistance and cell membrane permeability (Buchini, S., 2003). These cationic backbone modifications can produce both zwitterionic and fully modified cationic TFOs, thus resulting in favourable electrostatic interactions (Michel, T., 2005 and Michel, T., 2003). The practical problem (stability, sequence restrictions, susceptibility to nucleases and delivery into cellular nucleus) with TFOs to be used in anti-gene therapy was poor uptake due to polyanionic nature of oligonucleotides. This problem was looked upon by using cationic lipids, polycations (such as polyethylenimine), and dendrimers (highly branched 3D molecules) to increase uptake by cells in culture (Boletta, A., 1997, Vijayanathan, V., 2004, Santhakumaran, L. M., 2004 and Zinselmeyer, B. H., 2002). Lipid, cholesterol, and alkyl chain in conjugation with TFOs can penetrate efficiently the
cellular and the nuclear membranes increase uptake *in vitro* and *in vivo* (Cheng, K., Ye, Z., 2006).

**Triple-helix Target Sites (TTS)**

The specific binding of a single-stranded polyuridine oligonucleotide to a polyuridine/polyadenosine duplex to form a triple helical nucleic acid structure was first observed by Felsenfeld et al. (Felsenfeld G, 1957). Oligonucleotides directed triplex helix formation can be designed to form triplex with double-stranded genomic DNA in a sequence-specific manner by binding in the major groove of the double strand via Hoogsteen hydrogen bonds between ODN bases and purine bases in the DNA targets that are already engaged in Watson-Crick hydrogen bonds (Suzuki, 1997, Valenti, K., 1999, and Puckey, 1997). The presences of TFOs in the major groove of a duplex DNA interfere in the recognition of proteins to its target site and modulate the functionality of the DNA (Giovannangeli, C, 1997, Maher, L. J. III, 1996, Praseuth, D., 1999, and Vasquez, K.M., 1998). TFO not only alter gene expression during transcription process, but also arrests transcription elongation by binding to the transcribed position of the targeted gene (Giovannangeli, C., 2000). TFOs do not form triplexes on sequences already covered by nucleosomes, except at sites located towards the extremities of the nucleosomal DNA fragments (Faria, M., 2004). TFOs can alter gene expression by interfering either with the binding of transcription factors (Hewett, P.W., 2006, and Svinarchuk, F., 1997) or with the formation of the initiation complex (Karympalis, V., 2004) during transcription process. It can also arrest transcription elongation by binding to the transcribed position of the targeted gene (Giovannangeli, C., 2000). It has greater stability after binding to the target site than the complex formed by DNA and the transcription factors.
Introduction

Triple-helix target sites (TTS) are over-represented in the human genome and especially at promoter regions (Goni, J. R., 2004). The population of TTS is large throughout the genome, without major differences between chromosomes. Although TTS are not directly targeted by transcription factors, they may be important for gene functionality by acting as spacing fragment to help the correct positioning of transcription factors (Goni, J. R., 2006). Even if it has not been applied in therapy yet, its great potential has been widely demonstrated, both for therapeutic and biotechnological aspects. The largest relative concentration of TTS is found in regulatory regions, especially in promoter regions (Goni, J. R., 2004, and Goni, J. R., 2006). Softwares able to identify high affinity TTS have also been developed (Wu, Q., 2007).

TFO as biological tool

Site-specific DNA-binding molecules offer the potential for genetic manipulation of mammalian cells. The efficient and site specific recognition of duplex DNA by TFOs offer a useful approach in regulating gene expression. Triplex technology has wide applications in molecular biology and biochemistry as a tool for delivering DNA damaging drugs to a specific site, for recognition and purification of DNA and for the study of complex DNA-protein interactions. TFOs are promising gene-drugs, which can be used in an anti-gene strategy that attempt to modulate gene activity in vivo (Uil, T. G., 2003). TFOs are useful in site directed mutagenesis and site-specific recombination. Once these DNA damage agents are conjugated to TFO, they are able to induce site-specific mutagenesis and recombination (Rogers, F. A., 2005). Such induction of site-specific DNA damage by using TFOs (Vasquez, K. M., 2002, Strobel, S.A., 1991, Takasugi, M., 1991, and Vasquez, K. M., 1996), thereby enhance the frequencies of mutation (Barre, F.
X., 2000, Christensen, L. A., 2004 and Majumdar, A., 1998) and recombination (Datta, H. J., 2001, Faruqi, A. F., 2000, and Vasquez, K. M., 2001) in vitro and in vivo. Diverse damaging agents have been coupled to TFOs: (i) photoactivatable agents (Perrouault, L., 1990), (ii) metal complexes, such as Fe–EDTA (Strobel, S. A., 1990), orthophenantroline (Francois, J. C., 1989) or metalloporphirines (Bigey, P., 1995) and (iii) enzymes such as nuclease (Landgraf, R., 1994, and Pei, D., 1990). Recently, it has been shown that site-specific DNA cleavage can be obtained by attaching a restriction enzyme to a TFO, such as PvuII (Carpenter, M., 2006). Carbone et al. demonstrated the efficacy of TFO–daunomycin (DNM) conjugates in inhibiting transcription of a gene involved in tumor growth: c-myc gene (Napoli, S., 2006, and Carbone, G. M., 2004). Furthermore, if these conjugates are employed in DNM-resistant cell lines, the uptake of DNM itself is permitted by the presence of the oligonucleotide, thus showing a mutual action of the partners of these conjugates DNM and TFO. It has been reported that TFO conjugates of the anti-tumor camptothecin, a potent inhibitor of human topoisomerase IB, are able to induce topoisomerase I-mediated DNA cleavage in cells and to inhibit specifically the expression of a transient reporter gene in cells (Arimondo, P. B., 2006). TFO could selectively inhibit endogenous MCP-1 (Monocyte chemoattractant protein-1) gene expression in cultured human embryonic kidney cells (Marchand, P., 1999). Genetic diseases having low level expression, such as β-globin or sickle cell anemia, can be treated practically by TFO. Song et al (Song, J., 2004) used psoralen-TFO to activate gene expression up to 4-fold when targeted to site upstream the promoter so that recruit a transcription factor and activate transcription. In analogy, Xu et al (Xu, X. S., 2000) showed
increased transcription of γ-globin gene upto 4-fold by introducing mutation through posralen-TFO in the transcription binding site and also in cellular studies.

**Triplex forming potential of PNR upstream region of apo(a) promoter**

Although polymorphism of apo(a) gene regulates gene expression, the reason for the differences in expression level in homologous and heterologous alleles is unknown. The effect of homologous and heterologous allele’s effect on regulation of apolipoprotein(a) promoter is yet to be studied. Although it is a hereditary character, its effects changes in subsequent generation which may be due to environmental factors at the time of transcription. But polymorphic nature of pentanucleotide repetitive sequence of upstream 1.4 kb of apo(a) gene is not enough to predict its expression level in human because of its structural variation, dominant nature and environmental effect on its structure. Moreover, the effects of such homologous and heterologous pentanucleotide repetitive alleles on expression have not been studied till date. Hence, we thought to focus on structural variation in DNA of Pu-Py repetitive sequence upstream of apo(a) promoter gene’s DNA structural variation and its control on gene regulation. To elicit the role of repetitive sequences observed in apo(a) promoter vicinity with a potential to adopt different unusual structures in the transcription regulation, we have used triplex forming oligonucleotide (TFO) method for generating and introducing triplex within a gene as well as intramolecular triplex formation in acidic medium to study polymorphic nature of repetitive sequences. Triplex-forming oligonucleotides (TFO) bind in the major groove of duplex DNA a polypurine/polypyrimidine stretch in a sequence-specific manner. The binding specificity of TFOs makes them potential candidates for being used in direct genome modification. This approach was used to investigate the role of such defined sequences with unusual
secondary structure motifs, if any, in triplex formation and transcription control in living cells in culture. Using the DNA-protein interaction elaboration approach, we have identified the transcription factors binding on those sequences and tried to purify the particular factor which have potential to bind triplex structure and might be play an important role in controlling the regulation of transcription in vivo.
AIMS AND OBJECTIVES

1. To study triplex formation at polymorphic PNR regulatory region of apo(a) promoter.

2. Intermolecular triplex formation by TFO (Triplex Forming Oligonucleotide) to study its effect on gene expression.

3. Identification and purification of PNR binding proteins.

4. Characterization the role of factors, binding on PNR regulatory region, on gene expression.