From DEPARTMENT OF LABORATORY MEDICINE
Karolinska Institutet, Stockholm, Sweden

Targeting double-stranded DNA with locked nucleic acid-based (LNA) oligonucleotides (ON) and DNA-binding peptide conjugates

Y. Vladimir Pabón-Martínez

Karolinska Institutet

Stockholm 2016

Published by Karolinska Institutet.
Printed by E-Print AB 2016

ISBN 978-91-7676-526-5

Targeting double-stranded DNA with locked nucleic acid-based (LNA) oligonucleotides (ON) and DNA-binding peptide conjugates

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Y. Vladimir Pabón-Martínez

Principal Supervisor:
Professor **C. I. Edvard Smith**
Karolinska Institutet
Department of Laboratory Medicine
Clinical Research Center

Co-supervisor(s):
Associate Professor **Rula Zain**
Karolinska University Hospital
Department of Clinical Genetics
Centre for Rare Diseases

PhD **Karin Lundin**
Karolinska Institutet
Department of Laboratory Medicine
Clinical Research Center

Opponent:
Professor **Poul Nielsen**
University of Southern Denmark
Department of Physics, Chemistry and Pharmacy

Examination Board:
Professor **Sam Okret**
Karolinska Institutet
Department of Biosciences and Nutrition

Professor **Anna Karlsson**
Karolinska Institutet
Department of Laboratory Medicine

Professor **Lars Wieslander**
Stockholm University
Department of Molecular Biosciences

*DNA…**"This structure has novel features which are of considerable biological interest"***
Watson JD, Crick FH. 1953

To my family,

"A mi familia terrenal y celestial"

Too…
I would like to dedicate my thesis to the victims of the Colombian Conflict. I pray that one day not too far in the future, my homeland (Colombia) and other regions in the world that currently suffer from war can live in peace. I am also hopeful that one day every child has the opportunity to go to school as was afforded to me.

"One child, one teacher, one book and one pen can change the world. Education is the only solution. Education first."
Malala Yousafzai (2014 Peace Nobel Prize)

ABSTRACT

In this thesis we have studied two different types of molecules that bind to the major groove of the DNA: (i) oligonucleotides (ONs) containing locked nucleic acid (LNA) and in combination with intercalating agents (Twisted Intercalating Nucleic Acid, TINA, and Benzoquinoquinoxaline, BQQ) and (ii) peptides mimicking the leucine zipper transcription factor (GCN4 bZIP TF). Both ONs and peptides bind to the DNA double helix aiming to serve as transcription regulators of disease related genes. In paper I, we developed a clamp type LNA ON (bisLNA) having the capacity to invade and bind double-stranded supercoiled DNA, as demonstrated by chemical probing. BisLNA has a dual mode of binding to duplex DNA, by Hoogsteen interactions forming a triplex structure and by double strand invasion (DSI) forming new Watson-Crick hydrogen bonds. Optimization of the bisLNA construct was carried out in particular regarding the number and location of LNA nucleotides and the length of the triplex-forming and the strand-invading arms, and we achieved 30% DSI using plasmid DNA. In paper II, by combining EMSA and molecular dynamics (MD) simulations we evaluated the structural features of modified ONs in stabilizing both duplex and triplex structures; we found that reduction in the LNA content at the 3'-end of a triplex forming oligonucleotide (TFO) destabilized the triplex. Moreover, we demonstrated that positioning TINA at the 3'-end of a TFO had an advantageous effect on triplex stability, and BQQ was able to stabilize the pyrimidine motif triplex containing LNA ON and TINA. MD simulation showed that LNA-substitution in the pyrimidine strand of a duplex alters the double helix structure, changing x-displacement, slide and twist allowing triplex formation through enhanced TFO major groove accommodation. Finally, we also elucidated the mechanism of bisLNA binding to a dsDNA, which basically involves a two-step process where the triplex is formed first followed by double strand invasion. Finally, in paper III, four models of the GCN4 bZIP TF were evaluated, but only one achieved high sequence specific dsDNA binding as was demonstrated by electrophoretic mobility shift assay (EMSA) and the obtained dissociation constant. The same peptide showed a better uptake when evaluated in macrophages, demonstrating the potential of peptide-steroid based TF mimics.

In summary, the results provide a basis for further development of DNA binding chemical compounds, which could potentially have future applications in medicine and biotechnology in the future.

LIST OF SCIENTIFIC PAPERS

I. **Development of bis-locked nucleic acid (bisLNA) oligonucleotides for efficient invasion of supercoiled duplex DNA**
Pedro M. D. Moreno[§], Sylvain Geny[§], **Y. Vladimir Pabon**, Helen Bergquist, Eman M. Zaghloul, Cristina S. J. Rocha, Iulian I. Oprea, Burcu Bestas, Samir EL Andaloussi, Per T. Jørgensen, Erik B. Pedersen, Karin E. Lundin, Rula Zain, Jesper Wengel and C. I. Edvard Smith
Nucleic Acids Res, 2013. 41(5): p. 3257-73 .

II. **LNA effects on DNA binding and conformation: from single strand to duplex and triplex structures**
Y. Vladimir Pabon[§], You Xu[§], Alessandra Villa, Karin E. Lundin, Sylvain Geny, Chi-Hung Nguyen, Erik B. Pedersen, Per T. Jørgensen, Jesper Wengel, Lennart Nilsson, C. I. Edvard Smith and Rula Zain
Nucleic Acids Res, 2016 (re-submitted, after review)

III. **Sequence-selective DNA recognition and enhanced cellular up-take by peptide-steroid conjugates**
Yara Ruiz García, Abhishek Iyer, Dorien Van Lysebetten, **Y. Vladimir Pabon**, Benoit Louage, Malgorzata Honcharenko, Bruno G. De Geest, C. I. Edvard Smith, Roger Strömberg and Annemieke Madder
Chem Commun (Camb). 2015. 51(99):17552-5.

Publications and manuscripts by the author not included in this thesis:

IV. **Cyclodextrin - peptide conjugates for sequence specific DNA binding**
Yara Ruiz García, Jan Zelenka, **Y. Vladimir Pabon**, Abhishek Iyer, Miloš Buděšínský, Tomáš Kraus, C. I. Edvard Smith and Annemieke Madder
Org Biomol Chem. 2015. 13(18):5273-8.

V. **Next generation bisLNAs with stacking linker and 2´-glycylamino-LNA show enhanced invasion into duplex DNA**
Sylvain Geny, Pedro M. D. Moreno[§], Tomasz Krzywkowski[§], Olof Gissberg, Nicolai K. Andersen, Abdirisaq J. Isse, Amro M. El-Madani, Chenguang Lou, **Y. Vladimir Pabon**, Brooke A. Anderson, Eman M. Zaghloul, Rula Zain, Patrick J. Hrdlicka, Per T. Jørgensen, Mats Nilsson, Karin E. Lundin, Erik B. Pedersen, Jesper Wengel and C. I. Edvard Smith
Nucleic Acids Res, 2016. 44(5):2007-19.

VI. **Specific recognition of a non-palindromic dsDNA sequence by a heterodimer transcription *factor mimic***
Yara Ruiz García, **Y. Vladimir Pabon**, C. I. Edvard Smith and Annemieke Madder
Chemical Communications, 2016 (Manuscript submitted)

§ Shared authorhip

CONTENTS

LIST OF ABBREVIATIONS

BQQ	benzoquinoquinoxaline
CAA	chloroacetaldehyde
ddNTPs	dideoxynucleotide triphosphates
DNA	deoxyribonucleic acid
DS	double strand
dsDNA	double-stranded DNA
EMSA	electrophoretic mobility shift assay
HG	Hoogsteen
INA	intercalating nucleic acid
LNA	locked nucleic acid
nt	nucleotide
MD	molecular dynamics
mer	oligonucleotide length (number of nucleotide units)
o-TINA	ortho-TINA
ON	oligonucleotide
p-TINA	para-TINA
R	polypurine strand
PAGE	polyacrylamide gel electrophoresis
PNA	peptide nucleic acid
PE	primer extension
RNA	ribonucleic acid
TFO	triplex forming oligonucleotide
TINA	twisted intercalating nucleic acid
TS	triplex structure
TTS	triple-helix target sites
WC	Watson-Crick
Y	polypyrimidine strand

1 INTRODUCTION

1.1 OVERVIEW OF DNA

In 1865 Gregor Mendel described the fundamental laws of inheritance and deduced that genes come in pairs and are inherited as distinct units, one from each parent (1). A few years later, Friedrich Miescher, was interested in the chemistry of cells and discovered a chemical substance “nuclein” isolated from the cell nuclei of white blood cells, that he described as a substance high in phosphorus and later known as nucleic acid (2,3); one of his collaborators, Kossel determined the chemical composition of the nucleic acid describing the nucleobases: adenine, cytosine, guanine, thymine and uracil (4). At the same time Phoebus Levine discovered the carbohydrate compound of RNA (ribose) and DNA (deoxyribose), and the chemical composition of a single nucleotide (phosphate-sugar-base) (5,6). In the 1940s, Avery and his colleagues based on the analysis of Griffith´s work about DNA as the "transforming principle" (7), demonstrated that the DNA was the hereditary material by using two types of Pneumococcus strains: smooth (S) that are encapsulated, and rough (R); they treated the S-strain with enzymes (protease, RNase and DNase) and mixed with R-form Pneumococcus, and observed that after treatment the S-strain appeared in all of the cultures, except in the cultures where the S-strain that was treated with DNase (8). A few years later Chargaff isolated DNA from different organisms, measured the four nitrogen bases and demonstrated that the total amount of purine and pyrimidine are usually nearly equal, but that their percentage and the amount of DNA varies between species (9). It was in the beginning of the 1950s when Rosalinda Franklin, using X-ray crystallography, determined the “signature” of the DNA and thereby provided information about the shape and structure and calculated the basic dimensions (10,11). Based on this information, in 1953 James Watson and Francis Crick proposed the structure of the DNA, a double helix with the phosphate groups located on the outside and the bases inside, with specific pairing of purine with pyrimidine nucleobases through complementary hydrogen bonds (12). This model remains valid and became accepted and set the basis for the revolution of studies aimed at understanding the activity of nucleic acids and their role in the living cells as carriers of the genetic information (13). Finally, the crystal structure of the most common form, the B-type of DNA, was solved using heavy atom X-ray crystallography in the 1980s (14,15).

1.2 DNA STRUCTURE

1.2.1 Base-pairing

1.2.1.1 Watson-Crick base-pairing (WC)

Formation of WC hydrogen bonds is the more common interaction between bases. The two strands of the double helix are connected by hydrogen bonds. In DNA Adenine binds to Thymine, and Cytosine to Guanine (13).

1.2.1.2 Hoogsteen base-pairing (HG)

Hoogsteen hydrogen bonds are responsible for the specificity and stability of the triplex structure (16,17). Due to that purines contain an available position to form hydrogen bonds, it is possible to hybridize a third strand to purine-rich sequences, forming a triplex by binding in the major groove of the DNA (18). In addition HG base pairing plays an important role in stabilizing several non-B DNA conformations (described in section 1.2.3).

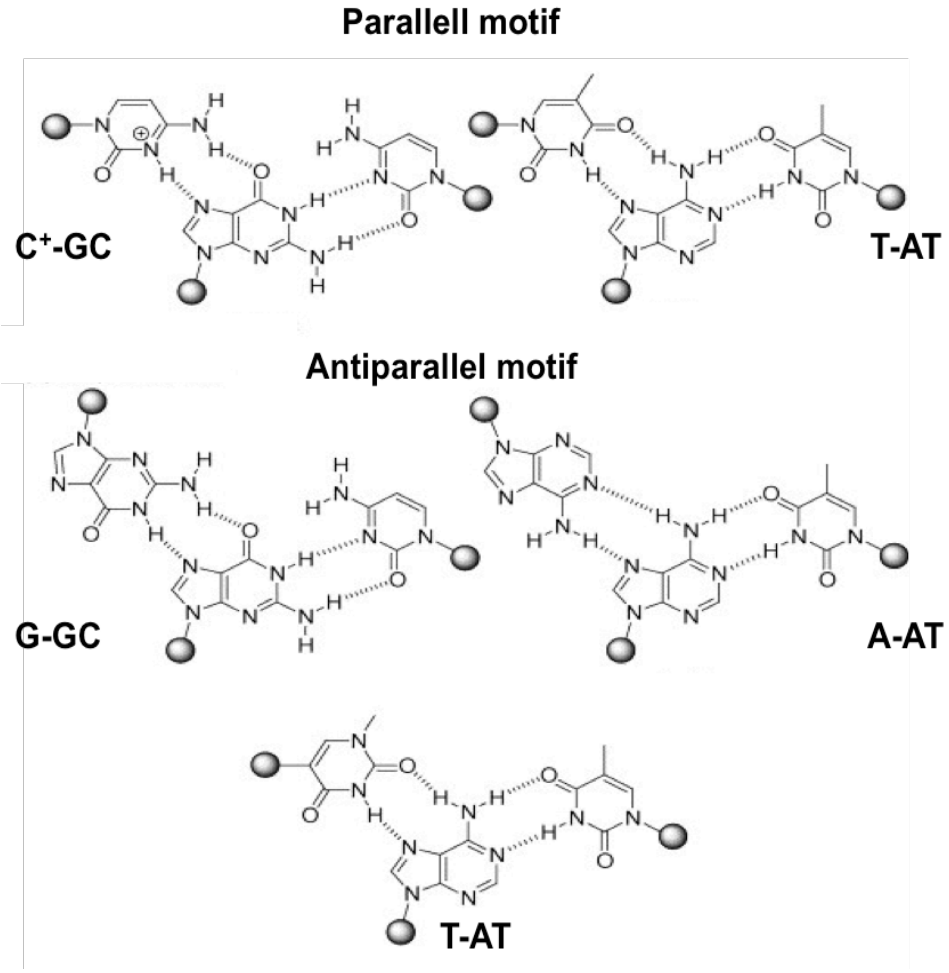


Figure 1. Hoogsteen base pairing in triplex structures.

1.2.2 Double strand invasion (DSI)

Double strand invasion (DSI) can occur when an ON displaces one of the DNA strands forming new WC hydrogen bonds with the other strand. Different modes of DSI binding have been described and a summary is presented below at the section of nucleic acid analogs.

1.2.3 DNA conformations

By oversimplifying the thermodynamic factors associated with helix formation of DNA it is possible to say, that the enthalpy of interstrand hydrogen bonds, base stacking and free water molecules are favorable for helix formation. Contrary, the entropy of free bond rotations, free translation/rotation and hydrogen bonds to water are unfavorable (19). Thus, DNA can adopt different conformations depending on the base sequence and environment (20). The most studied conformations of the double helix are three: A, B and Z. Table 1 summarizes of the helical parameters for A-DNA, B-DNA and Z-DNA.

A-DNA is a right-handed antiparallel double helix that appears in x-ray diffraction studies when the relative humidity is reduced to less than about 75% and is rarely observed under physiological conditions (21).

The **B-DNA** conformation was proposed by Watson and Crick in 1953 (12,13). B-DNA is a right-handed antiparallel helical duplex and is the most prevalent form of DNA in living cells (22).

Z-DNA (unlike A- and B-DNA) is the only known left-handed form of the double helix and is called Z, because the phosphate groups make a zigzag along the backbone (23). Z-DNA was discovered in 1979 (24), and was the first DNA structure determined by crystallography. Its physiological role mainly remains elusive and it was not until many years later that its role in viral infections was deciphered (25).

Table 1. Helical parameters

	A-DNA (21)	**B-DNA (22)**	**Z-DNA (24)**
Twist	33°	36°	-30°
Rise	2.56 Å	3.38 Å	3.7 Å
Roll	6.0°	0°	0°
Inclination	21°	-6.0°	-6.2°
χ-Displacement	-4.5 Å	0.23 Å	3.0 Å
Propeller twist	-7.5	-4.4°	-4.4°

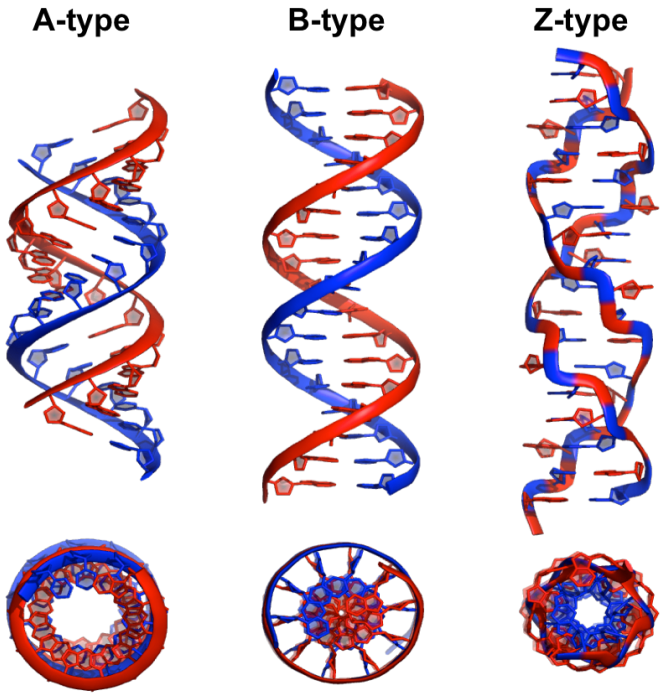


Figure 2. The duplex 3D structures. The A, B, Z type helixes with 17-mer sequences are shown in front and top views. The structures were generated using w3DNA (26). (Structures courtesy of You Xu)

Moreover, DNA has the capability to adopt a variety of non-canonical structures (Non B-form secondary structures) (27): G-quadruplex (28), cruciform, hairpin and triplex, etc. The non-B DNA structure-forming sequences can induce genetic instability and consequently be the cause of human diseases (29-33), some of these structures are present in genomic regions as in expanded number of copies of simple repeats, which are associated with neurological diseases (34-37). The formations of those structures are highly favorable by the enthalpy but highly unfavorable by the entropy in aqueous solutions (19,38,39).

1.3 DNA TRIPLE HELIX (DNA TRIPLEX)

Linus Pauling was the first to propose the formation of a triple-strand helix structure for DNA (40,41). However, it was only four years after Watson and Crick published their seminal paper in the Nature journal about the DNA structure (12), that Felsenfeld *et al*. discovered that RNA can form triple helices (42). This type of binding was sequence specific and stabilized by divalent cations. Later on, in 1968 it was shown that an RNA-dsDNA triplex can form and inhibit transcription (43). Triple helix structures of nucleic acids (DNA and RNA) can be divided into: intermolecular and intramolecular triplexes (44) (45) .

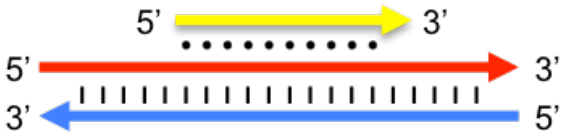
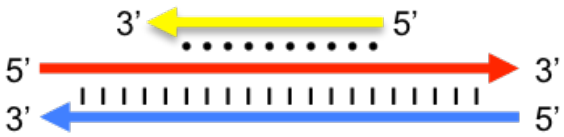
1.3.1 Intermolecular triplexes

Felsenfeld *et al*. described the first triple helix in polyribonucleotides formed in the presence of Mg^{2+} (42). Intermolecular triplexes are formed when the third strand originates from another DNA molecule (45). The third strand ON recognizes oligo-purine sequences and binds the dsDNA in the major groove forming hydrogen bonds (46,47). Those types of ONs are known today as triple helix forming oligonucleotides or triplex forming oligonucleotides (TFOs). Intermolecular triplexes form with a pyrimidine third strand (Y:RY), a purine third strand (R:RY) or mixed pyrimidine/purine third strand. In this thesis we focused on intermolecular triplexes. Table 2 summarizes the characteristics of the different triplex helix motifs.

The formation of triplexes under physiological conditions is a process with slow kinetics when compared with the formation of a duplex (48-52), but once triplexes are formed, they are stable and have half-lives of days (48,50,51).

Transcription can be inhibited by TFOs as a consequence of two different mechanisms (53,54): (i) inhibiting transcription initiation by competing with transcription factor binding (55) or (ii) arresting transcription elongation by binding further downstream in the transcribed gene (56).

Table 2. Triple helix classification

Motif	Characteristics
Py motif: **(T, C) (G, T) (T, C, G)** 	The third strand binds **parallel** (5'-3') to the purine strand by **Hoogsteen** hydrogen bounds Protonation of cytosine (N3) is required for Hoogsteen bonding with guanine (N7) Acidic conditions are required
Pu motif: **(G, A) (G, T) (G, A, T) (G, T, A)** 	Third strand binds **antiparallel** (3'-5') to the purine strand by **reverse Hoogsteen** hydrogen bonds No required protonation Essentially pH independent

Pu, purine. Py, pyrimidine. TFO (yellow), polypurine strand (red), polypyrimidine strand (blue). (•) indicates HG hydrogen bonds and (**I**) indicates WC hydrogen bonds.

1.3.2 Intramolecular triplexes

In 1987 Mirkin discovered the intramolecular triplex structure in plasmid DNA at low pH (57,58). Intramolecular triplexes are also called H-DNA (59). In intramolecular triplexes the third strand originates from one of the strands of the same double helix DNA that contains a mirror repeat sequence. These triplexes can form different structures depending on their triplex structural motifs (60), if the third strand is Py-rich (H-DNA)(59), and if the third strand is Pu-rich (*H-DNA)(61).

1.4 NUCLEIC ACID ANALOGS

The requirements for an optimal TFO are: high affinity binding to the dsDNA target at physiological pH and salt conditions, sequence specificity and resistance to degradation by endonucleases. One way to achieve these requirements is by introducing modifications to the base, the sugar or the backbone portion of the DNA or RNA. Modified ONs, including peptide nucleic acid (PNA) or locked nucleic acid (LNA) have been developed and used in the anti-gene to target dsDNA.

1.4.1 Locked nucleic acid (LNA)

Locked nucleic acid (LNA) is a nucleic acid modification developed in the late 1990s by two independent groups, one in Denmark and the other in Japan (62,63). LNA contains a methylene bridge connecting the 2′-oxygen with the 4′-carbon in the furanose ring. The bridge locks the sugar moiety in an *N*-type sugar ring conformation providing enhanced hybridization properties. Due to this conformation, the backbone of LNA is reorganized as a consequence of base stacking (64); another effect of the locked conformation is the increase in melting temperature (2-8°C) per nt (65). Moreover, NMR studies of duplexes containing LNA confirmed that LNA appears as an RNA-analogue (66,67). New LNA chemistries have provided different analogues with new properties such as, 2'-amino-LNA (68), α-L-LNA (69) and ethylene-bridged nucleic acid (ENA) (70) among others.

Various LNA chemistries have been used in TFOs (71-78), clamp ONs (79-81), invader-LNAs (82) and Zorro-LNAs (83-86), as anti-gene reagents (87) to block transcription, and in antisense approaches (87) to induce RNA degradation by RNase H, to increases the stability of small interfering RNA (siRNA) (88), to alter pre-mRNA splicing, as anti-mirRs (89).

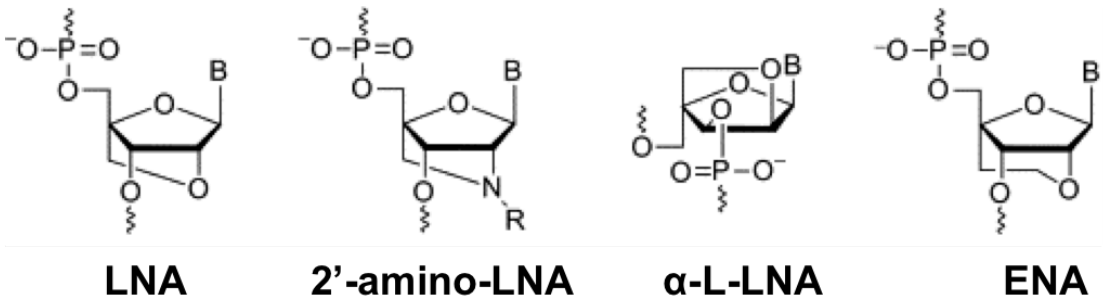


Figure 3. Chemical structure of locked nucleic acid (LNA), 2'-amino-LNA, α-L-LNA and ethylene-bridged nucleic acid (ENA)

1.4.2 Peptide nucleic acid (PNA)

Peptide nucleic acid (PNA) was described originally by Nielsen *et al*. as a DNA analogue where the phosphodiester backbone is replaced by a pseudo-peptide (90,91). PNA has an efficient discrimination of single-base mismatches, forming the basis for the development of diagnostic applications to identify polymorphisms (92-94). Originally, PNA was designed to bind DNA as a TFO aiming to control transcription (95). Table 3 summarizes the different modes of PNA interactions with dsDNA. A remarkable idea was the development of the first ON with two arms able to form both WC and HG bonding, named bisPNA (96), features later replicated, and further developed, using LNA (80). PNA was shown to form various complexes, which have higher thermal stability under moderate ionic strengths compared to unmodified nucleic acids (97). However, PNA-RNA duplexes can not be cleaved by RNaseH, which is a disadvantage when used in antisense strategies (98). Conversely, PNA has different advantages over other nucleic acid modifications, being extremely resistant to nuclease degradation, showing poor or no binding to serum proteins and also being stable against protease activity (91). PNA can pass slowly by passive diffusion through cell membranes, which can be due to the neutral charge backbone (99). An antisense effect by PNA has been reported in primary neurons (100), and PNA has been injected *in vivo* in different ways, intracerebral (101), intrathecal (102) and intraventricular (103) causing biological effects in rats.

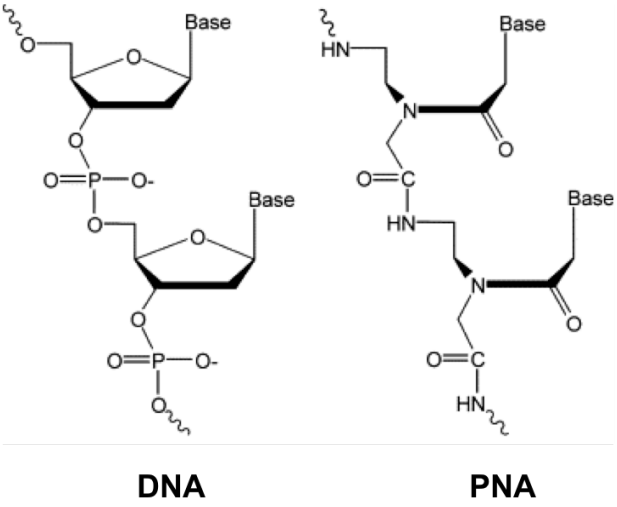
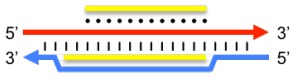
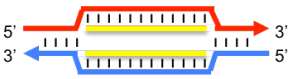
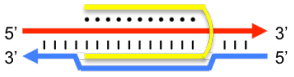
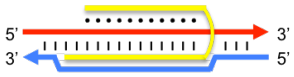
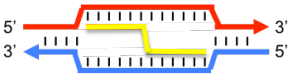


Figure 4. Chemical structures of DNA and peptide nucleic acid (PNA).

Table 3. Binding modes that generate double strand invasion

Binding mode	Schematic illustration	Mechanism of binding	Demonstrated	
			LNA	PNA
Duplex invasion		A linear ON bind to the sense-strand by WC. The antisense-strand is displaced.	(80,87)	(104)
Triplex invasion		Two separate ONs, one binding by HG and the other by WC; both ONs target the sense strand. The antisense-strand is displaced.	(80,87)	(96)
Double duplex invasion		Pair of ONs with "weak bases" that cannot hybridize to each other, binds by WC to the sense and anti-sense strand of dsDNA (also known as pseudocomplementary).	-	(105)
Clamp invasion		The two arms are linked. The TFO-arm binds by HG to the sense-strand, while the WC-arm invades the duplex forming new WC bonds with the TFO-targeted strand. The antisense-strand is displaced.	(81)	(106)
Tail clamp invasion			(80,81)	(107)
Zorro-LNA invasion		An ON binding simultaneously to both strands by WC.	(83-85)	-

Pu, purine. Py, pyrimidine. TFO (yellow), polypurine strand (red), polypyrimidine strand (blue). (•) indicates HG hydrogen bonds and (**I**) indicates WC hydrogen bonds.

1.5 DNA MODE OF BINDING

Binding of small molecules to the DNA double helix has been studied in detail. The two common binding modes are intercalation and groove binding. Some of the methods for distinguishing between intercalation and grove binding modes were described by Suh *et al.* (108).

1.5.1 Intercalators

DNA intercalation changes the structure of the double helix, and causes lengthening, stiffening and unwinding (109-111). Zeglis *et al.* defined intercalators as small organic molecules or metal complexes that unwind DNA in order to π-stack between two base pairs (112). π-stack interactions are non covalent (113,114), and are important in chemical and biological recognition (115). The study of intercalating chemical compounds is important as they are among the most promising therapeutic agents to treat many diseases (54,116-118); some of these molecules are approved by the FDA and used for the treatment of cancer, like Mitoxantrone that inhibits DNA synthesis by intercalating in DNA (119).

The stability of triplex helix structures under physiological conditions can be potentiated by using chemically modified ONs, but also by the use of agents that bind specifically to triplex helices. Ethidium bromide is a classical duplex intercalator able also to intercalate in T.AxT (120) but not in C.GxC+ (121). However, Benzopyridoindole derivatives (BPI) can stabilize triplex C.GxC+ and bind preferentially to triplex as compared to duplex (122). Other compounds have been described with the capacity to intercalate into triplexes, agents such as coralyne (123), sanguinarine (124), benzophenazine (125), benzophenanthridine (126) and tricyclic amidoanthraquinones (127) among others. In this thesis we focused on benzoquinoquinoxaline and twisted intercalating nucleic acid compounds.

1.5.1.1 Benzoquinoquinoxaline (BQQ)

Benzoquinoquinoxaline (BQQ) is a triple helix-intercalating compound, able to bind specifically and stabilize triplex structures of purine and pyrimidine motifs (128-131). BQQ is a low-molecular weight compound and has been shown to enter cells, bind and stabilize intra-molecular triplexes (H-DNA) (132,133). When BQQ is conjugated to neomycin it also intercalates into triple-helical DNA (134). Other uses of BQQ are related to DNA labeling by using electron or optical microscopy (135,136), and for gene delivery by plasmid modification (137). BQQ inhibits the enzymatic cleavage (138), and blocks the transcription by RNA polymerase (139). In this thesis we used BQQ to probe for triplex formation of the different complexes analyzed.

Moreover, BQQ has been converted to a triplex-specific cleaving agent when conjugated to EDTA (140) or to a 1,10-phenanthroline derivate (130).

1.5.1.2 Twisted intercalating nucleic acid (TINA)

Twisted intercalating nucleic acid (TINA) is an intercalator covalently joined and inserted into ONs (141). TINA is a modification derived from an intercalating nucleic acid (INA), used as a bulge insertion to stabilize duplex DNA, but having the opposite effect in triplexes, decreasing their stability (142). TINA contains an aromatic group to stack with the bases allowing stabilization of ON binding. TINA discriminates between triplex and duplex DNA, and increases the thermal stability of parallel triplexes (143,144). There are two variants of TINA (145): (i) ortho-TINA (o-TINA) designed to stabilize WC antiparallel DNA duplexes and (ii) para-TINA (p-TINA) designed to stabilize HG triplexes and quadruplexes.

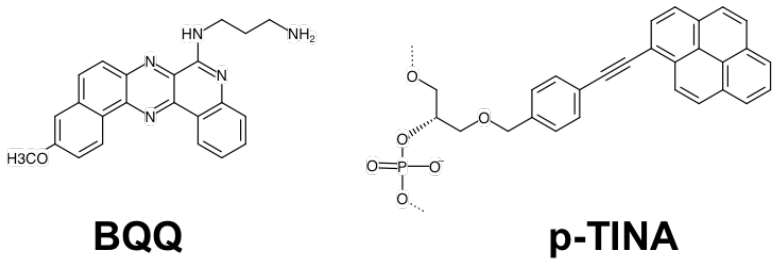


Figure 5. Chemical structures of benzoquinoquinoxaline (BQQ) and para-twisted intercalating nucleic acid (p-TINA).

1.5.2 Groove binders

In contrast to intercalation, groove binding only causes a subtle change to the double helix (109-111). Nucleic acid binding within the major groove of the DNA is rare, but is common for DNA-binding proteins (below section). Two adjacent base pairs must be separated to create a binding pocket for the ligand, this is a requirement for binding (109).

Binding to the minor groove does not require a big distortion of the DNA. The majority of minor groove binders has four structural characteristics, positive charge, curvature, flexibility and hydrogen bond donor and acceptor groups (146). A good example of minor groove binder is the antibiotic netropsin, that binds selectively to AT-rich sequences in the minor groove of DNA (147,148). The main focus in this thesis is in major groove binders.

1.6 DNA BINDING PROTEINS

The ability that specific proteins have to bind selectively to a particular DNA region in the genome is the basis for the transcriptional regulatory pathways (149). DNA-binding proteins

are very important in biology and are responsible for replicating the genome, controlling gene expression and repairing damaged DNA (150). Transcription factors (TFs) are one of the groups of DNA-binding proteins. There are approximately 12 to 15 structurally distinct DNA-binding domains known from eukaryotic transcription factors (151). Some of them are the helix-turn-helix, Zn binding domains, the leucine-zipper coiled coil, β-ribbon proteins, among others (152).

1.6.1 The leucine-zipper coiled coil

Known also as leucine-zipper bZIP, this domain consists of an α-helix with leucines on the same side of the helix, every seven amino acids. A leucine residue is found at the fourth position of four of five successive heptads in the activator protein GCN4 (General Control Nonderepressible 4) (153). The yeast transcriptional activator GCN4 is one of over 30 identified eukaryotic proteins containing the basic region leucine zipper (bZIP) DNA-binding motif (154) (bZIP GCN4). The crystal structure is available at the Protein Data Bank (PDB-ID:1YSA) 2.9 Å resolution.

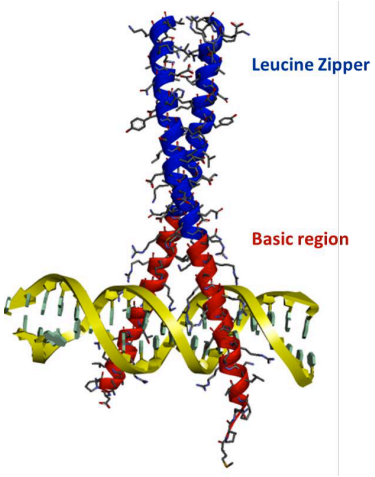


Figure 6. Crystal structure of bZIP GCN4.

1.7 GENE THERAPY

In 1972 it was suggested that "in the future" gene therapy could be used as a treatment for human genetic diseases (155). Since then significant events have occurred, as the design of a retrovirus vector to efficiently insert foreign genes into mammalian chromosomes (156), the first successful clinical trial by using a retroviral system for a severe immune system deficiency (adenosine deaminase, ADA) (157), and the use of gene transfer into hematopoietic stem cells for correcting of human severe combined immunodeficiency (SCID)-X1 disease (158). More than that, in 2003 China approved Gendicine, the world's first commercial gene therapy that delivers p53 using an adenoviral vector for the treatment

of head and neck squamous cell carcinoma (159). After that was demonstrated that a lentiviral containing a sequence complementary to a miRNA can be used to control transgene expression (160) and genetically engineered cells (peripheral blood lymphocytes (PBLs)) can be used for cancer immunotherapy (161). Currently, according to the Journal of Gene Medicine there are in the world 1357 clinical trials in phase I and 994 in other trials, what is a probe of the big progress in gene therapy approaches. For more details in clinical trials (see review (162)).

1.8 THE ANTI-GENE STRATEGY

Anti-gene strategy is the approach of gene therapy for control the gene expression at the transcriptional level based on the sequence specificity (163). TFOs are potential tools to control by inhibition of the transcription the gene expression in a sequence specific manner (46,47), what can be used as anti-gene strategy. The first successful example of in *vitro* inhibition of the transcription was reported by using a 27-base-long DNA oligonucleotide targeting the c-*MYC* gene (164). The triplex anti-gene approach has been investigated in other proto-oncogenes like *ki-ras (165)*, *bcl-2 (166)*, *her-2 (167), c-met (168)* and TFOs have been used to interfere with functions of the replicative cycle of HIV (169,170). TFOs have the possibly to be used also as an antisense strategy. By for example, blocking RNA processing (171), reverse transcription (172) and translation.

Binding sites for TFOs in the human genome, or triple-helix target sites (TTS), are abundant especially in promoter regions (173,174), which make that TFOs can potentially be exploited in anti-gene strategies for the therapy of cancers and other genetic diseases. Lately, different tools have been developed that facilitate the analysis and search of TTS in the genome, including algorithms (175) (176) and TFO databases (177), non-B DNA structures databases (178) and web-service tools (179,180) have been established. Together with the Encyclopedia of DNA Elements (181,182) these resources can help to select the best TTS to target with TFOs.

1.9 MOLECULAR DYNAMICS

In 1977 was published the first molecular dynamics (MD) simulation for proteins (183); since then MD has become an important tool to understand the structure, physics and functions of macromolecules (184). Comparison between simulation and experimental results is a way to validate the data and to improve the methodology (185,186). Simulation methods can be applied to macromolecules for determining or understanding better structures obtained from

experimental results; describe systems in equilibrium and their thermodynamical parameters; and to examine actual dynamics. In this thesis we used MD in an integrated way to suggest experimental studies and to analyze and extend the interpretations of the experimental data on ON interactions with DNA.

2 AIMS

I. To investigate and map the bisLNA binding to a supercoiled duplex DNA. (Paper I)

II. To evaluate parameters influencing triplex formation and extend these observations to elucidate the molecular mechanism of bisLNA binding to a dsDNA. (Paper II - Submitted manuscript)

III. To evaluate the binding of peptides used as dsDNA binding, synthetic transcription factor mimics. (Paper III)

3 METHODOLOGY

This section covers some relevant methods used in the present thesis.

3.1 RADIOLABELING

Radioactive end labeling can be used for visualizing and allowing the detection of nucleic acids at low concentrations. DNA, RNA and modified nucleic acids can be radiolabelled. Even when safer non-radioactive levels are available the sensitive detection of radioactive label is the "gold standard" of probe technology (187). In this thesis the primers (paper I), and the pyrimidine or purine strand of the DNA (paper II and III) was labeled using [γ-^{32}P] ATP and T4 polynucleotide kinase and purified using a Nucleotide Removal Kit. The kinase catalyzes the transfer of the ATP γ-phosphate to the 5'-hydroxyl terminal end of the DNA (188).

3.2 ELECTROPHORETIC MOBILITY SHIFT ASSAY (EMSA)

Electrophoretic mobility shift assay was originally developed as a method to detect protein-DNA interactions (189,190). Currently, EMSA has been used also to detect different types of interactions between ligands (DNA, RNA, proteins) and DNA (191). After electrophoresis, the different complexes containing nucleic acid are determined, usually by autoradiography of ^{32}P-labeled nucleic acid. In this thesis (paper II and paper III) we used EMSA to detect ONs (TFO and bisLNA) or peptide interactions with their target sequence. Due to the rate of DNA migration in the gel a shifted band can be detected; the shifted band corresponds to the complexes formed that migrate more slowly compared to the free DNA. EMSA allows the quantification and identification of complexes formed between a ligand and their target. EMSA can be used also to study binding affinities and thermodynamic and kinetics parameters (189,190). PAGE gels have a better resolution than the less toxic agarose gels.

3.3 FOOTPRINTING

Footprinting is a method that allows identifying sequence-specific interactions of drugs, proteins (192,193), nucleic acids like e.g. TFOs (194,195), and DNA intercalators (196) with a DNA target; and can be used to determine the sequence selectivity, affinity and kinetics of DNA-binding ligands (197). Originally the method was used for DNA-protein interactions (198). If a ligand such as a TFO is bound to a DNA target site, this region is then protected from cleavage by DNase I or S1 nuclease and from chemical modification

using for example chloroacetaldehyde (CAA). For clamp constructs, like bisLNA, the displaced strand in the targeted dsDNA is susceptible to CAA chemical modification, , as was described in paper (I). To examine the binding site of the ligand, primer extension reaction is carried out using the modified DNA strand as a template (199).

3.4 CHLOROACETALDEHYDE (CAA) REACTION

Chloroacetaldehyde (CAA) reacts with the exocyclic amino group of cytosine and adenine residues preferentially in single-stranded regions of nucleic acids (200). After the CAA reaction, the modified DNA single strand regions are more susceptible to alkali cleavage (201). Single strand DNA chemical modification using CAA was used in paper (I) to map, at one nucleotide resolution, the site of the displaced strand, which showed bisLNA binding to the DNA target sequence.

3.5 PRIMER EXTENSION (PE)

Primer extension is basically a PCR performed with only one primer, which was used in paper (I). By primer extension the complementary primer binds to the specific region of the single stranded template and Taq polymerase and dNTPs were used to synthesize the complementary DNA strand. The primers are radiolabeled with [γ-^{32}P] ATP and the PE products are analyzed using polyacrylamide gel electrophoresis (PAGE). Radioactivity is the more common labeling method for footprinting (202,203). Currently, other possibilities to detect the "footprint" include fluorescently labeled primers analyzed by capillary electrophoresis avoiding the risk of exposition to radiation (81,204).

3.6 S1 NUCLEASE DIGESTION

S1 is an enzyme able to cleave single stranded DNA/RNA, but not double strand DNA or RNA/DNA. S1 has been used to open genomic regions like hairpins (205), inverted repeats (206), promoter regions (207) and to detect strand invasion in the case of PNA (202), bisPNA (203), Zorro-LNA (84) and for the study of triplexes *in vivo* (208-210). When bisLNA strand invades it is possible to use S1 to cleave the single strand that is displaced. It acts as a probe of the strand invasion and it is possible to determine the efficiency of the duplex invasion/hybridization of the ON by quantifying the amount of supercoiled and nicked plasmid. S1 was used to verify DSI by bisLNA in paper (I).

4 RESULTS AND DISCUSSION

PAPER 1

Development of bis-locked nucleic acid (bisLNA) oligonucleotides for efficient invasion of supercoiled duplex DNA

The main aim of this paper was to demonstrate the capacity of LNA to achieve duplex invasion of supercoiled DNA under intranuclear salt and pH conditions, by using a designed clamp-type construct.

Previous work in our group using the LNA technology was done, with the aim to develop a construct with the potential to sequence-specifically invade dsDNA and block gene expression, a construct called zorro-LNA. This construct was able to simultaneously bind to both the coding strand of DNA and to the template strand showing double-strand invasion (DSI); furthermore, Zorro-LNA was able to block both RNA-polymerase II and RNA-polymerase III (83,84).

We observed that zorro-LNAs were able to bind only to a supercoiled plasmid, but not to a linearized plasmid. We therefore decided to explore a new strategy using two modes of binding to DNA, the triplex formation and WC-mediated DSI, based on a clamp-type construct that previously demonstrated efficiency using PNA oligomers. In this paper, with the aim of developing a construct with the potential to specifically block gene expression in a genomic context, the bisLNA construct was developed.

BisLNA ONs are single strand mixed LNA-DNA ONs designed for strand-invasion of duplex DNA (DSI), with the ability to bind to the double helix via both HG and WC interactions. The bisLNA sequence is composed of nucleotides or nucleic acid analogues, and the structure consists of two arms and one flexible non-complementary linker. Both arms of the bisLNAs, that were used in this thesis, bind to DNA polypurine sequences, where one arm forms Hoogsteen hydrogen bonds to the polypurine target strand as a TFO, whereas the other arm binds through DSI of the double helix, forming Watson-Crick hydrogen bonds and displacing the opposite strand. The TFO-part subsequently forms a new triplex with the ensuing LNA-DNA duplex. This mode of binding results in an extremely stable complex, but has so far not been reported to arrest transcription *in vivo*. However, TFO binding requires special sequences with only purines in the target strand, while the DSI-part does not suffer from any restriction. Thus, extending this part outside the polypurine target site to form a

"tail-clamp" construct was an option, which possibly could become an important asset for efficient biological targeting.

As LNA, in contrast to PNA, had not been previously used in clamp constructs, the first assays were oriented to perform this evaluation. In order to do that, a 25-mer ON with 10 bases in each arm and a 5 nt linker was used (bisLNA20). A plasmid containing six bisLNA binding sites was used as a target for the bisLNA20 a,b,c,d and e. The incubation was performed in acidic pH (5.8) in order to favour triplex formation. After the incubation, the hybridization products (plasmid+bisLNA) were analysed by fluorescence. According to the results, only the clamp construct containing LNA could efficiently invade the DNA, in contrast to all the other constructs lacking LNA. This was an initial proof demonstrating that LNA is also able to invade dsDNA. This evaluation was possible thanks to that the ONs were Cy3 or Cy5 labelled.

The constructs (TFO, WC, TFO+WC and bisLNA20) were evaluated by gel electrophoresis. As was expected, the signal of the clamp construct (bisLNA20) was more intensive than the signal from the TFO+WC separately, and TFO and WC did not show any binding. These results were confirmed by using the S1 nuclease assay generating single strand DNA cleavage. Thus, the S1 cleavage is a quantitative evaluation of the DSI. The results confirmed the previous conclusion.

Next we evaluated the influence of each arm in the clamp construct and how the effect of the presence or absence of LNA. Here we studied two new constructs with different LNA content in one of the arms (TFO or WC). Our results indicate that having 50% LNA-substitution in each arm of the bisLNA is crucial for efficient DSI.

To probe for bisLNA binding of DNA in nearly physiological salt and pH conditions, the incubation reactions were carried out at pH 7.4, 37°C with intranuclear salt conditions. Different experiments were performed to optimize the bisLNA constructs under these conditions. Influence of the length of each arm on bisLNA binding to a single site was examined using different bisLNAs with short TFO arm, short WC and long WC-arm. We observed that shortening the WC-arm causes a negative effect in the DSI when as compared to a bisLNA construct that has the WC-arm extended (tail). In addition, when bisLNA that has short TFO (3nt-shorter) was compared with the bisLNA control (m44), no significant effect was observed. Here we also conclude that bisLNA tail clamp constructs were more efficient in DSI than clamp type.

Next, we aimed to evaluate the bisLNA binding to a single site and a new plasmid was used, which contained the full c-*MYC* promoter sequence and a binding site for the bisLNA. We examined plasmid DNA binding of a bisLNA (m44) with a tail (WC-arm extended, 29nt) and 15nt TFO arm, which was incubated for 72h, and we achieved nearly 80% DSI. This was compared with the results obtained in the presence of bisLNAm30 (15-mer), which showed less than 5% of DSI.

To provide evidence for the degree of DSI, footprinting was performed. This allows mapping of the DNA single strand regions that result from bisLNA invasion by the WC-arm binding. A plasmid containing a single binding site for the bisLNA in the c-*MYC* promoter region was used. The plasmid was incubated in the presence or absence of bisLNAm44 or the WC-arm separately. After 72h, the DSI was evaluated by EMSA. The samples were treated with chloroacetaldehyde (CAA). CAA can react with the cytosines, adenines and some guanines of the single strand, which was followed by a cleavage reaction under alkaline conditions. Primer extension reaction with radiolabeled primers was performed to detect the modified DNA sites. ddNTPs were used for the synthesis of the four different sequence ladders. We observed that the samples treated with CAA, showed a strong signal (as judged from the intensities of the gel bands) in the region that corresponds to the bisLNA binding site. The CAA reaction was enhanced in the DNA sample treated with bisLNA as compared to the corresponding DNA in the presence o the WC-LNA ON. The results of the described chemical probing provide a proof for the following conclusions: bisLNA invasion is highly sequence specific, both WC-LNA arm and bisLNA can invade double-strand DNA, but bisLNA is more efficient, and finally bisLNAs is able to invade dsDNA at intra-nuclear pH and salt conditions. Therefore, bisLNA is a new efficient clamp type ON, being an alternative to bisPNA, which primarily works under non-physiological, 'low salt' conditions.

PAPER 2

LNA effects on DNA binding and conformation: from single strand to duplex and triplex structures

The main aim of this paper was to better understand the underlying structural features of LNA containing ONs starting from single to duplex and triplex structures.

Here we investigated the role of the position and number of locked nucleic acid (LNA) substitutions in the ON and the effect of TINA at three different positions of the TFO (5'end, middle and 3'-end).

In the previous paper, the bisLNA, WC-arm was evaluated extensively and also the DSI. However not much information was available about the TFO-arm and triplex formation. With this curiosity we started to evaluate TFOs and triplex formation. Here we used the same target sequence and we began by using the TFO-arm (15-mer) of the bisLNA that we previously evaluated by chemical probing.

Molecular modeling (MD) simulations was used to confirm mostly all our experimental results and also provided new conformation-based information. The TFO was evaluated by EMSA using a short, radioactively labeled target sequence. All the binding experiments were performed in parallel in the presence of BQQ, as a probe for triplex formation. LNA substitution was evaluated using different TFOs. With the purpose to better understand the triplex formation, shorter and longer incubation times were used. Also different ratios TFO:dsDNA were studied to optimize triplex formation with regards to TFO concentration. First, we examined dsDNA binding of a TFO containing LNA or non-modified DNA. We showed that the TFO lacking LNA was unable to form triplex, demonstrating the impact of LNA substitution in TFOs. Second, we evaluated the importance of the Cy3 labeled and the end position (5'-end) of the TFO starting with an LNA instead of DNA. According to our results we could see that triplex formation takes at least 24h when using the lowest ratio of DNA:TFO (1:12). Also, we found that the TFO labeled with Cy3 was less efficient compared with the unlabeled; which is important to point out since the use of fluorescently labeled ONs is quite common in several cell-based assays.

Then, we decided to evaluate the importance of the LNA position in the TFO. For this, we studied a set of TFOs with less LNA at one of the ends. The 3' position was the most important for LNA substitution, which could be potentially explained by the mechanism of TFO binding.

To further improve triplex formation, we decided to introduce an intercalating agent in the TFO. The TINA intercalator was placed at the ends or in the middle of the TFO. The first observation was the dramatic effect of TINA on triplex formation. After one hour at a DNA:TFO ratio of 1:100, 100% triplex formation was achieved. This shows the effect of TINA and the importance of the 3'-end position, which was consistent when using the corresponding shorter TFOs

Moreover, we used a duplex target sequence including a DNA-LNA pyrimidine strand, thus constituting a hetero-duplex target. We compared TFO binding to the homoduplex and heteroduplex targets and we found that triplex formation was faster with the hetero-duplex sequence. We found that LNA-substitutions in the WC pyrimidine strand alter the conformation of the duplex, which forms an intermediate of A/B-type DNA. This conformation is observed after the TFO is bound to both DNA homo-duplexes and to hetero-duplexes, where one of the strands is a DNA-LNA mixmer. Such hetero-duplexes are formed when bisLNAs (80,211) invade into homo-duplex DNA, and this information helps to understand the mechanism of bisLNA invasion.

Finally, we also evaluated the mechanism of bisLNA binding. For this, binding of the TFO, WC and bisLNA to the dsDNA target were examined in parallel. Taken together we believe that different DNA complexes are formed during bisLNA binding with triplex formation as the first event followed by double strand invasion.

PAPER 3

Sequence-selective DNA recognition and enhanced cellular up-take by peptide-steroid conjugates

Currently, a strategy in development for cancer therapy is the use of sequence-specific DNA binders based on peptide-steroid conjugates. A requirement for that is that the peptides have to be able to penetrate the membrane and reach to the nucleus. In the evaluated peptides the basic DNA binding region of the GCN4 protein has a positive charge due to the six arginine residues, similar to various cell-penetrating peptides, something that might help to enter into the cell. The transcription factor protein GCN4, belongs to the bZIP Leucine Zipper family.

This paper is part of our cooperation with Gent University. The peptides were synthetized by the substitution of the Leucine zipper domain of the protein by a deoxycholic acid derivative of the two GCN4 binding region peptides. Four peptides were evaluated. For the peptides number 8 and 10 no binding was detected. However peptide 7 and 9 showed a high affinity for the CRE target sequence. Peptide 7, shows higher binding affinity compared with peptide 9. However, with the first EMSA results it was not possible to calculate the KD because not so many concentrations were available to make the calculations. This is a really good indication that the peptide has a high affinity for the target. To proceed with the analysis we decreased the concentration of the peptides and we included a DNA competitor with the

purpose to eliminate the possibility of unspecific binding. When the DNA competitor (500nM) was added a better-shifted result was observed. Even under this condition peptide 7 continued being the best compared with 8 and the specificity to the sequence is high, as can be seem for the gradual shifted band opposite to the first EMSA analysis without the DNA competitor. When the KD was calculated, peptide 7 had the lowest KD compared with peptide 9, what can be interpreted as a decent specificity. Peptide 7 has a longer and more flexible linker that allows the construct to grip the major groove of the DNA and that could explain the better binder capacity

5 CONCLUDING REMARKS AND FUTURE PERSPECTIVE

BisLNA is an ON with seemingly great potential for DSI under physiological conditions. However, a limitation could be size, due to the long tail. One alternative to potentiate and optimize the bisLNA can be the use of other modifications. Some of these modifications have been already evaluated in the bisLNA report published in 2016. In this paper was probed that bis-m44-gly-M3 that contains a stacking linker M3 and glyLNAs was better invader to DNA and with maintained specificity. In the analysis of this optimized bisLNA a footprinting assay was also performed but without the use of radiolabeling, instead of that capillary electrophoresis and fluorescent primer were used, thus avoiding the use of radioactive labeling. Introduction of more than one M3 intercalator had favorable effects on tail-less bisLNAs and future research on intercalator positioning may identify the optimal positioning of these.

Some of the TFOs that were evaluated in paper II were directed to pyrimidine regions, providing information about the design of TFOs to target these regions. Moreover, poly-purine regions are highly prevalent, especially in promoter regions. We have already started some studies, not included in this thesis. The new designs include the use of pseudoisocytidine to overcome the problem of the cytosine protonation with the aim of obtaining good TFO binding a pH 7.4.

Another, possibility to improve the bisLNA design is to study in more detail the effect of the WC-arm and one possibility is using the same model that we evaluated in paper II, where we have a hetero-duplex target sequence. In this case we also have preliminary indications that the position and amount of LNA in the WC of this type of design could be crucial to improve the bisLNA technology.

Peptides are an alternative option to target DNA. As shown in paper II, some of the designs are promising and have demonstrated effects in cell culture. We have also carried out experiments with other chemistries that we evaluated by EMSA and follow-up studies will include the evaluation of their activity in cell cultures.

6 ACKNOWLEDGEMENTS

These years have been really nice for me and I have learned science, but also i have met amazing people that have made my stay in Sweden during my PhD studies one of the biggest time in my life. Thanks to **God** and thanks to:

Professor **C.I. Edvard Smith,** the best example of an excellent supervisor and researcher. Thank you for allowing me to do my PhD in your group with all these amazing people! Thanks for always being available and really enthusiastic about my results. Thank you for your support and for being at my side when I needed you. I think you are a really good master that teach through his own examples.

My co-supervisor **Rula Zain**, "the lady of the triplexes". Thanks for a excellent co-supervision and orientation. Thanks for teaching me and taking me into the chemistry "thinking". You are the best example showing how a woman researcher should be like!

My co-supervisor Ka**rin Lundin**, you are the best example of a lab keeper! You are able to answer as many questions about molecular biology and you always know where everything is in the lab.

I have learn a lot from all you and, I think you were the best "master-mix" as co-supervisors of my PhD studies. Thanks to all of you for sharing your knowledge. "Tack så mycket" ! "Muchas gracias"!

My collaborators from University of Southern Denmark, professors **Jesper Wengel**, **Erik Pedersen** and **Per Jørgensen**. Thanks you for all the synthesis of the oligonucleotides. From you I learned the meaning of a cooperative project. Thanks for the intensive and productive discussions in our meetings in Odense and Stockholm. "Muchas gracias"!

My collaborators from University of Ghent in Belgium professor **Annemieke Madder.** Especially to **Yara Ruiz-Garcia**. Thanks for the synthesis of the peptides and for a really productive cooperation. Gracias Yara por permitirme hacer parte de tu trabajo de Doctorado, fue muy agradable trabajar contigo!

My collaborators from Karolinska Institutet, professor **Lennart Nilsson**, and his Molecular Modeling team: **Alessandra Villa**, **You Xu**, **Yossa Dwi Hartono** and **Arzu Uyar**. Thanks for supporting my experimental data and for helping me understand the fabulous world of the structures. Thanks to all of you for sharing your knowledge.

To the past members of MCG: **Joana Viola**, **Iulian Oprea**, **Pedro Moreno** "the father of the bisLNA", **Per Lundin**, **Alamdar Hussain**, **Hossain Nawaz**, **Izzy Daud**, **Fatin Izzati**, **Yue Chen**, **Lotta Asplund**, **Sofia Stenler**, **Janne Turunen**. I am glad that I had the oportunity to share some time with you, specially when I was in the group in 2010. Thanks to Niels and Mamiko, I wish you the best luck in the future. Thanks **Sylvain Geny** for all your help and knowlegde about bisLNA**,** Merci! **Fiona Lee**, it was really nice to share time with you in the lab, your smile and funny messages in your t-shirt always make me feel happy. **Adbi (Abdalla M. Jama)** thanks for your good advices and support for my PhD grant application when I visited the lab for the first time. You are not only a good researcher; you are a person that can see into the heart of people. **Sonia Al-Qadi** for your support when I applied to my

PhD grant. Muchas gracias por tu colaboración y palabras de ánimo, la carta de Avaris fue muy efectiva! **Helen Bergquist** thanks for teach me about chemical probing and helping me to set up the techniques by myself. **Noriko Hamada** is always nice to have you in our lab ☺ and share some café. **Leonardo Vargas** que agradable fue tener un colombiano en el grupo y poder compartir nuestros pensamientos sobre la ciencia y la vida. **H. José Arteaga**, my master supervisor and **María E. Cardona** who introduce me in the world of science and opened for me the posibility to become a PhD at Karolinska Institutet. Muchas gracias!

To the current MCG members: **Beston Nore**, **Eman Zaghloul**, **Dara Mohammad**, **Burcu Bestas**, **Joel Nordin**, **Oscar Wiklander**, **Samir El-Andaloussi**, **Giulia Corso**, **Cristina Rocha**, **Manuela Gustafsson**, **Helena Sork** (my lady, it was always nice to dance with you!), **Narmeen Fathi**, **Kariem Ahmed**, **Taavi Lehto**. All of you have part of this adventure called PhD. Thanks! It was really nice to share my PhD studies with all of you!

To the people in Avaris: **Olof Gissberg** and **Abdulrahman Hamasy (Abdul)**, we start this process together, I am really glad for being friends during all my PhD. Olof, we must have a "Te" session! And everyone must be invited! **Qing Wang** "my dear peer" you are the beauty of Avaris, thanks for being always helpful and friendly. **Anna** and **Björn Berglöf,** for hosting me when I came to Sweden for my PhD studies, and for teaching me how to survive in Sweden.

To the new MCG members: **Professor Anthony Wright**, **Amir Mahani**, **Gustav Arvidsson**, **André Görgens**, **Laia Sadeghi**, **Ulrika Felldin**, **André Görgens**, **Xiuming Liang**. I wish you the best! I am sure you will have a lot of success in our group.

To new generation of MCG: **Dhanu Gupta**, **Osama Ahmed**, **Negin Mozafari**, **Narmeen Fathi**, **Tea Umek**. Good luck, I am sure you will be really good PhD students.

Thanks to Professor **Moustapha Hassan** for being at the lab, corridor or kitchen during the weekend, always having a funny thing to say!

Thanks to **Emelie Blomberg**, for helping me with the genotyping and now together with **Hanna Gador** and **Kathrin Reiser** helping me with administration and many other issues, all of you are really nice and helpful.

Thanks to other people in Novum for being always friendly in the corridor (**Pauliina Damdimopoulou**, **Sandra Petrus**, **Alvaro Plaza**, **Jp Schell**) and during the fika (**Fadwa Ben Kessou**, **Andrea Bieder**, **Towe Jakobsson**, **Brigitte Twelkmeyer**, **Nicolas Tardif**, **Eva Skop Nejman**, **Nelson Morales** y **Sara**).

Mis amigos del grupo de investigación de Inmunología y Epidemiología Molecular (GIEM) del Laboratorio Central de Investigaciones (LCI) en la Universidad Industrial de Santander, especialmente, a **Clara González**, por su apoyo constante e incondicionalidad, un ejemplo de perserverancia y dedicacion a la ciencia en Colombia.

Muy especialmente, a **Esperanza Vesga Torres**, mi querida Presidenta, mi mentora, la gestora de todo esto. Muchas gracias!

También me gustaría agradecer a todos quienes han sido mis maestros. Gracias por contribuir en mi formación y educación.

A mis amigos en Colombia, quienes a pesar de la distancia siempre me brindan su apoyo, amistad y carino: (Los bacteriólogos ☺ **Andrea Alvarez**, **Juliana Navarro**, **Mayerly Villamil**, **Laura Pacheco**, **Dana Zorro**, **Edith Acero**, **Farid Uscátegui**, **Hailyn Jaimes**, **Hugo Velarde**, **Jonathan Vargas**, **Maria Cardona**, **Rocio Abaunza**, **Ruth Gomez**, **Iris Pinto**, **Oscar Ramirez**, **Diego Tarazona**), **Camilo Guerrero**, **Ximena Mancera**, **Alberto Amaya**, **Martha Hincapie**, **Yesid Estupiñan**.

To my friends in Sweden "Alguien dijo los amigos son la familia que se escoge" thanks to all you my stay here was a really pleasure and I have the best memories! Thanks to all of you for being my family in "Sverige" (min familj i Sverige).

Lenis Alvarez-Rozo por su apoyo incondicional en Suecia y Colombia, gracias por tu amistad desde la UIS hasta ahora en el KI. A los chicos de Colombia en el KI **Paola Martinez**, **Oscar Pico**, por ser parte de aventura en Suecia! To my friends **Thomas Perrot**, **Erle Refsum, Ladan Jama**, **Daniel McCain**, **Mary Rosmarino Irina Jovel-Dalmau**, **Magda Torres**. Is always nice to meet all of you and have a good time.

To my swedish friends **Steffan**, **Otto Måhlgren**, **Martin Balke**, **Benny Hjalmarsson**, **Martin Jefflén**, **Andreas Hagelin**, **Tomas and Helena Nimreus.** Tack!

And very specially thanks to **Kristoffer Bonde-Poll,** "tack så mycket", for being my righ hand in Sweden during all this years, for showing me the funny, cultural and friendly and non-scientific face of "Sverige". "Tack" for being always there supporting me! And thanks to your lovely family, **Roland Poll**, **Kirsten Bonde**, **Jonna Poll**, for being so friendly and nice.

A mis padres Miguel y Silda (ami y api) por su amor e incondicionalidad. Algunos aöos atrás ellos decidieron dejar su lugar de origen para buscar un mejor futuro para sus hijos, hoy yo soy resultado de sus esfuerzos y su decisión. Gracias a mis hermanos (**Yorguin**, **Miguel**, **Diego**), mi cuñada Milena, sobrinos (**Viviana**, **Carmen**, **Laura** y **Angel**), mis tías **Dioselda** y **Rosalba** (checha y chica) y **toda mi familia** por su constante apoyo y cariño. Gracias por haberme ayudado hacer posible este trabajo.

To **HIV Research Trust** fundation, who granted me in 2010 to make my instership at KI. I hope you will continue supporting young researchers, you do a great work!

To "**Departamento Administrativo de Ciencia, tecnología e innovación (COLCIENCIAS)"**, for my PhD grant. I hope I can contribute to develop science in Colombia in the future.

And as Dara K. Mohammed (2015) wrote in his thesis "I would like to apologize to those that by any change I have forgotten to mention your names, please forgive me"

7 REFERENCES

1. Mendel, G. (1865) Versuche über Plflanzen-hybriden.Verhandlungen des naturforschenden Ver-eines in Brünn. *Bd. IV für das Jahr 1865, Abhandlungen*, 3-47.
2. Dahm, R. (2005) Friedrich Miescher and the discovery of DNA. *Developmental biology*, **278**, 274-288.
3. Dahm, R. (2008) Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Human genetics*, **122**, 565-581.
4. Jones, M.E. (1953) Albrecht Kossel, a biographical sketch. *The Yale journal of biology and medicine*, **26**, 80-97.
5. Levene, P.A. (1920) THE STRUCTURE OF YEAST NUCLEIC ACID: V. AMMONIA HYDROLYSIS. *Journal of Biological Chemistry*, **41**, 19-23.
6. Bass, L.W. (1940) Phoebus Aaron Theodor Levene 1869-1940. *Science*, **92**, 392-395.
7. Griffith, F. (1928) The Significance of Pneumococcal Types. *The Journal of hygiene*, **27**, 113-159.
8. Avery, O.T., Macleod, C.M. and McCarty, M. (1944) Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *The Journal of experimental medicine*, **79**, 137-158.
9. Chargaff, E. (1950) Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia*, **6**, 201-209.
10. Franklin, R.E. and Gosling, R.G. (1953) Evidence for 2-chain helix in crystalline structure of sodium deoxyribonucleate. *Nature*, **172**, 156-157.
11. Franklin, R.E. and Gosling, R.G. (1953) Molecular configuration in sodium thymonucleate. *Nature*, **171**, 740-741.
12. Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737-738.
13. Watson, J.D. and Crick, F.H. (1953) Genetical implications of the structure of deoxyribonucleic acid. *Nature*, **171**, 964-967.
14. Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R.E. (1980) Crystal structure analysis of a complete turn of B-DNA. *Nature*, **287**, 755-758.
15. Dickerson, R.E. and Drew, H.R. (1981) Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure. *Journal of molecular biology*, **149**, 761-786.
16. Hoogsteen, K. (1959) The Structure of Crystals Containing a Hydrogen-Bonded Complex of 1-Methylthymine and 9-Methyladenine. *Acta Crystallogr*, **12**, 822-823.
17. Hoogsteen, K. (1963) Crystal and Molecular Structure of a Hydrogen-Bonded Complex between 1-Methylthymine and 9-Methyladenine. *Acta Crystallogr*, **16**, 907-+.
18. Beal, P.A. and Dervan, P.B. (1991) Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation. *Science*, **251**, 1360-1363.

19. Kool, E.T. (1997) Preorganization of DNA: Design Principles for Improving Nucleic Acid Recognition by Synthetic Oligonucleotides. *Chem Rev*, **97**, 1473-1488.

20. Ghosh, A. and Bansal, M. (2003) A glossary of DNA structures from A to Z. *Acta Crystallogr D Biol Crystallogr*, **59**, 620-626.

21. Hays, F.A., Teegarden, A., Jones, Z.J., Harms, M., Raup, D., Watson, J., Cavaliere, E. and Ho, P.S. (2005) How sequence defines structure: a crystallographic map of DNA structure and conformation. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 7157-7162.

22. Prive, G.G., Yanagi, K. and Dickerson, R.E. (1991) Structure of the B-DNA decamer C-C-A-A-C-G-T-T-G-G and comparison with isomorphous decamers C-C-A-A-G-A-T-T-G-G and C-C-A-G-G-C-C-T-G-G. *Journal of molecular biology*, **217**, 177-199.

23. Rich, A., Nordheim, A. and Wang, A.H. (1984) The chemistry and biology of left-handed Z-DNA. *Annu Rev Biochem*, **53**, 791-846.

24. Wang, A.H., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G. and Rich, A. (1979) Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature*, **282**, 680-686.

25. Rich, A. and Zhang, S. (2003) Timeline: Z-DNA: the long road to biological function. *Nat Rev Genet*, **4**, 566-572.

26. Zheng, G., Lu, X.J. and Olson, W.K. (2009) Web 3DNA--a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic acids research*, **37**, W240-246.

27. Bochman, M.L., Paeschke, K. and Zakian, V.A. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet*, **13**, 770-780.

28. Gellert, M., Lipsett, M.N. and Davies, D.R. (1962) Helix formation by guanylic acid. *Proceedings of the National Academy of Sciences of the United States of America*, **48**, 2013-2018.

29. Wells, R.D., Dere, R., Hebert, M.L., Napierala, M. and Son, L.S. (2005) Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic acids research*, **33**, 3785-3798.

30. Wang, G. and Vasquez, K.M. (2006) Non-B DNA structure-induced genetic instability. *Mutat Res*, **598**, 103-119.

31. Bacolla, A. and Wells, R.D. (2009) Non-B DNA conformations as determinants of mutagenesis and human disease. *Mol Carcinog*, **48**, 273-285.

32. Zhao, J., Bacolla, A., Wang, G. and Vasquez, K.M. (2010) Non-B DNA structure-induced genetic instability and evolution. *Cell Mol Life Sci*, **67**, 43-62.

33. Wu, Y. and Brosh, R.M., Jr. (2010) G-quadruplex nucleic acids and human disease. *FEBS J*, **277**, 3470-3488.

34. Mirkin, S.M. (2006) DNA structures, repeat expansions and human hereditary disorders. *Curr Opin Struct Biol*, **16**, 351-358.

35. Mirkin, S.M. (2007) Expandable DNA repeats and human disease. *Nature*, **447**, 932-940.

36. Lopez Castel, A., Cleary, J.D. and Pearson, C.E. (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat Rev Mol Cell Biol*, **11**, 165-170.

37. McMurray, C.T. (2010) Mechanisms of trinucleotide repeat instability during human development. *Nat Rev Genet*, **11**, 786-799.

38. Breslauer, K.J., Frank, R., Blocker, H. and Marky, L.A. (1986) Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences of the United States of America*, **83**, 3746-3750.

39. SantaLucia, J., Jr., Allawi, H.T. and Seneviratne, P.A. (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, **35**, 3555-3562.

40. Pauling, L. and Corey, R.B. (1953) A Proposed Structure For The Nucleic Acids. *Proceedings of the National Academy of Sciences of the United States of America*, **39**, 84-97.

41. Pauling, L. and Corey, R.B. (1953) Structure of the nucleic acids. *Nature*, **171**, 346.

42. G. Felsenfeld, D.R.D., and Alexander Rich. (1957) Formation of a three-stranded polynucleotide molecule. *Journal of the American Chemical Society*, **79**, 2023-2024.

43. Morgan, A.R. and Wells, R.D. (1968) Specificity of the three-stranded complex formation between double-stranded DNA and single-stranded RNA containing repeating nucleotide sequences. *Journal of molecular biology*, **37**, 63-80.

44. Escudé, C. and Sun, J.-S. (2005) In Waring, M. J. and Chaires, J. B. (eds.), *DNA Binders and Related Subjects: -/-*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 109-148.

45. Jain, A., Wang, G. and Vasquez, K.M. (2008) DNA triple helices: biological consequences and therapeutic potential. *Biochimie*, **90**, 1117-1130.

46. Le Doan, T., Perrouault, L., Praseuth, D., Habhoub, N., Decout, J.L., Thuong, N.T., Lhomme, J. and Helene, C. (1987) Sequence-specific recognition, photocrosslinking and cleavage of the DNA double helix by an oligo-[alpha]-thymidylate covalently linked to an azidoproflavine derivative. *Nucleic acids research*, **15**, 7749-7760.

47. Moser, H.E. and Dervan, P.B. (1987) Sequence-specific cleavage of double helical DNA by triple helix formation. *Science*, **238**, 645-650.

48. Vasquez, K.M., Wensel, T.G., Hogan, M.E. and Wilson, J.H. (1995) High-affinity triple helix formation by synthetic oligonucleotides at a site within a selectable mammalian gene. *Biochemistry*, **34**, 7243-7251.

49. Paes, H.M. and Fox, K.R. (1997) Kinetic studies on the formation of intermolecular triple helices. *Nucleic acids research*, **25**, 3269-3274.

50. Alberti, P., Arimondo, P.B., Mergny, J.L., Garestier, T., Helene, C. and Sun, J.S. (2002) A directional nucleation-zipping mechanism for triple helix formation. *Nucleic acids research*, **30**, 5407-5415.

51. James, P.L., Brown, T. and Fox, K.R. (2003) Thermodynamic and kinetic stability of intermolecular triple helices containing different proportions of C+*GC and T*AT triplets. *Nucleic acids research*, **31**, 5598-5606.

52. Rusling, D.A., Broughton-Head, V.J., Tuck, A., Khairallah, H., Osborne, S.D., Brown, T. and Fox, K.R. (2008) Kinetic studies on the formation of DNA triplexes containing the nucleoside analogue 2'-O-(2-aminoethyl)-5-(3-amino-1-propynyl)uridine. *Org Biomol Chem*, **6**, 122-129.

53. Giovannangeli, C. and Helene, C. (2000) Triplex technology takes off. *Nature biotechnology*, **18**, 1245-1246.

54. Hurley, L.H. (2002) DNA and its associated processes as targets for cancer therapy. *Nat Rev Cancer*, **2**, 188-200.

55. Vasquez, K.M., Narayanan, L. and Glazer, P.M. (2000) Specific mutations induced by triplex-forming oligonucleotides in mice. *Science*, **290**, 530-533.

56. Giovannangeli, C. and Helene, C. (2000) Triplex-forming molecules for modulation of DNA information processing. *Curr Opin Mol Ther*, **2**, 288-296.

57. Mirkin, S.M., Lyamichev, V.I., Drushlyak, K.N., Dobrynin, V.N., Filippov, S.A. and Frank-Kamenetskii, M.D. (1987) DNA H form requires a homopurine-homopyrimidine mirror repeat. *Nature*, **330**, 495-497.

58. Mirkin, S.M. and Frank-Kamenetskii, M.D. (1994) H-DNA and related structures. *Annu Rev Biophys Biomol Struct*, **23**, 541-576.

59. Lyamichev, V.I., Mirkin, S.M. and Frank-Kamenetskii, M.D. (1985) A pH-dependent structural transition in the homopurine-homopyrimidine tract in superhelical DNA. *J Biomol Struct Dyn*, **3**, 327-338.

60. Zain, R. and Sun, J.S. (2003) Do natural DNA triple-helical structures occur and function in vivo? *Cell Mol Life Sci*, **60**, 862-870.

61. Kohwi, Y. and Kohwi-Shigematsu, T. (1988) Magnesium ion-dependent triple-helix structure formed by homopurine-homopyrimidine sequences in supercoiled plasmid DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 3781-3785.

62. Obika, S., Nanbu, D., Hari, Y., Morio, K.-i., In, Y., Ishida, T. and Imanishi, T. (1997) Synthesis of 2′-O,4′-C-methyleneuridine and -cytidine. Novel bicyclic nucleosides having a fixed C3, -endo sugar puckering. *Tetrahedron Letters*, **38**, 8735-8738.

63. K. Singh, S., A. Koshkin, A., Wengel, J. and Nielsen, P. (1998) LNA (locked nucleic acids): synthesis and high-affinity nucleic acid recognition. *Chemical Communications*, 455-456.

64. Petersen, M., Nielsen, C.B., Nielsen, K.E., Jensen, G.A., Bondensgaard, K., Singh, S.K., Rajwanshi, V.K., Koshkin, A.A., Dahl, B.M., Wengel, J. *et al.* (2000) The conformations of locked nucleic acids (LNA). *J Mol Recognit*, **13**, 44-53.

65. Veedu, R.N. and Wengel, J. (2010) Locked nucleic acids: promising nucleic acid analogs for therapeutic applications. *Chem Biodivers*, **7**, 536-542.

66. Petersen, M., Bondensgaard, K., Wengel, J. and Jacobsen, J.P. (2002) Locked nucleic acid (LNA) recognition of RNA: NMR solution structures of LNA:RNA hybrids. *J Am Chem Soc*, **124**, 5974-5982.

67. Nielsen, K.E., Rasmussen, J., Kumar, R., Wengel, J., Jacobsen, J.P. and Petersen, M. (2004) NMR studies of fully modified locked nucleic acid (LNA) hybrids: solution structure of an LNA:RNA hybrid and characterization of an LNA:DNA hybrid. *Bioconjug Chem*, **15**, 449-457.

68. Singh, S.K., Kumar, R. and Wengel, J. (1998) Synthesis of 2‘-Amino-LNA: A Novel Conformationally Restricted High-Affinity Oligonucleotide Analogue with a Handle. *The Journal of Organic Chemistry*, **63**, 10035-10039.

69. Nielsen, J.T., Stein, P.C. and Petersen, M. (2003) NMR structure of an alpha-L-LNA:RNA hybrid: structural implications for RNase H recognition. *Nucleic acids research*, **31**, 5858-5867.

70. Morita, K., Hasegawa, C., Kaneko, M., Tsutsumi, S., Sone, J., Ishikawa, T., Imanishi, T. and Koizumi, M. (2002) 2'-O,4'-C-ethylene-bridged nucleic acids (ENA): highly nuclease-resistant and thermodynamically stable oligonucleotides for antisense drug. *Bioorg Med Chem Lett*, **12**, 73-76.

71. Singh, S.K., Kumar, R. and Wengel, J. (1998) Synthesis of Novel Bicyclo[2.2.1] Ribonucleosides: 2'-Amino- and 2'-Thio-LNA Monomeric Nucleosides. *J Org Chem*, **63**, 6078-6079.

72. Obika, S., Uneda, T., Sugimoto, T., Nanbu, D., Minami, T., Doi, T. and Imanishi, T. (2001) 2'-O,4'-C-Methylene bridged nucleic acid (2',4'-BNA): synthesis and triplex-forming properties. *Bioorg Med Chem*, **9**, 1001-1011.

73. Torigoe, H., Hari, Y., Sekiguchi, M., Obika, S. and Imanishi, T. (2001) 2′-O,4′-C-Methylene Bridged Nucleic Acid Modification Promotes Pyrimidine Motif Triplex DNA Formation at Physiological pH: THERMODYNAMIC AND KINETIC STUDIES. *Journal of Biological Chemistry*, **276**, 2354-2360.

74. Sun, B.W., Babu, B.R., Sorensen, M.D., Zakrzewska, K., Wengel, J. and Sun, J.S. (2004) Sequence and pH effects of LNA-containing triple helix-forming oligonucleotides: physical chemistry, biochemistry, and modeling studies. *Biochemistry*, **43**, 4160-4169.

75. Brunet, E., Alberti, P., Perrouault, L., Babu, R., Wengel, J. and Giovannangeli, C. (2005) Exploring cellular activity of locked nucleic acid-modified triplex-forming oligonucleotides and defining its molecular basis. *J Biol Chem*, **280**, 20076-20085.

76. Hojland, T., Babu, B.R., Bryld, T. and Wengel, J. (2007) Triplex-forming ability of modified oligonucleotides. *Nucleosides Nucleotides Nucleic Acids*, **26**, 1411-1414.

77. Hojland, T., Kumar, S., Babu, B.R., Umemoto, T., Albaek, N., Sharma, P.K., Nielsen, P. and Wengel, J. (2007) LNA (locked nucleic acid) and analogs as triplex-forming oligonucleotides. *Organic & Biomolecular Chemistry*, **5**, 2375-2379.

78. Torigoe, H., Sato, N. and Nagasawa, N. (2012) 2'-O,4'-C-ethylene bridged nucleic acid modification enhances pyrimidine motif triplex-forming ability under physiological condition. *J Biochem*, **152**, 17-26.

79. Hertoghs, K.M., Ellis, J.H. and Catchpole, I.R. (2003) Use of locked nucleic acid oligonucleotides to add functionality to plasmid DNA. *Nucleic acids research*, **31**, 5817-5830.

80. Moreno, P.M., Geny, S., Pabon, Y.V., Bergquist, H., Zaghloul, E.M., Rocha, C.S., Oprea, II, Bestas, B., Andaloussi, S.E., Jorgensen, P.T. *et al.* (2013) Development of bis-locked nucleic acid (bisLNA) oligonucleotides for efficient invasion of supercoiled duplex DNA. *Nucleic acids research*, **41**, 3257-3273.

81. Geny, S., Moreno, P.M., Krzywkowski, T., Gissberg, O., Andersen, N.K., Isse, A.J., El-Madani, A.M., Lou, C., Pabon, Y.V., Anderson, B.A. *et al.* (2016) Next-generation bis-locked nucleic acids with stacking linker and 2'-glycylamino-LNA

show enhanced DNA invasion into supercoiled duplexes. *Nucleic acids research*, **44**, 2007-2019.

82. Sau, S.P., Kumar, T.S. and Hrdlicka, P.J. (2010) Invader LNA: efficient targeting of short double stranded DNA. *Org Biomol Chem*, **8**, 2028-2036.

83. Ge, R., Heinonen, J.E., Svahn, M.G., Mohamed, A.J., Lundin, K.E. and Smith, C.I. (2007) Zorro locked nucleic acid induces sequence-specific gene silencing. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, **21**, 1902-1914.

84. Zaghloul, E.M., Madsen, A.S., Moreno, P.M., Oprea, II, El-Andaloussi, S., Bestas, B., Gupta, P., Pedersen, E.B., Lundin, K.E., Wengel, J. *et al.* (2011) Optimizing anti-gene oligonucleotide 'Zorro-LNA' for improved strand invasion into duplex DNA. *Nucleic acids research*, **39**, 1142-1154.

85. Ling, J.Q., Hou, A. and Hoffman, A.R. (2011) Long-range DNA interactions are specifically altered by locked nucleic acid-targeting of a CTCF binding site. *Biochimica et biophysica acta*, **1809**, 24-33.

86. Zaghloul, E.M., Moreno, P.M., Mohamed, A.J., Wengel, J., Lundin, K.E. and Smith, C.I. (2011) Zorro-LNA, delivered by a cationic lipid, down-regulates Huntingtin gene expression in mammalian cells. *Hum. Gene Ther.*, **22**, A95-A95.

87. Lundin, K.E., Hojland, T., Hansen, B.R., Persson, R., Bramsen, J.B., Kjems, J., Koch, T., Wengel, J. and Smith, C.I. (2013) Biological activity and biotechnological aspects of locked nucleic acids. *Adv Genet*, **82**, 47-107.

88. Gao, S., Dagnaes-Hansen, F., Nielsen, E.J., Wengel, J., Besenbacher, F., Howard, K.A. and Kjems, J. (2009) The effect of chemical modification and nanoparticle formulation on stability and biodistribution of siRNA in mice. *Mol Ther*, **17**, 1225-1233.

89. Thorsen, S.B., Obad, S., Jensen, N.F., Stenvang, J. and Kauppinen, S. (2012) The therapeutic potential of microRNAs in cancer. *Cancer J*, **18**, 275-284.

90. Nielsen, P.E., Egholm, M., Berg, R.H. and Buchardt, O. (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science*, **254**, 1497-1500.

91. Demidov, V.V., Potaman, V.N., Frank-Kamenetskii, M.D., Egholm, M., Buchard, O., Sonnichsen, S.H. and Nielsen, P.E. (1994) Stability of peptide nucleic acids in human serum and cellular extracts. *Biochem Pharmacol*, **48**, 1310-1313.

92. Seitz, O. (2000) Solid-Phase Synthesis of Doubly Labeled Peptide Nucleic Acids as Probes for the Real-Time Detection of Hybridization This work was supported by the Fonds der Chemischen Industrie. *Angew Chem Int Ed Engl*, **39**, 3249-3252.

93. Kuhn, H., Demidov, V.V., Gildea, B.D., Fiandaca, M.J., Coull, J.C. and Frank-Kamenetskii, M.D. (2001) PNA beacons for duplex DNA. *Antisense Nucleic Acid Drug Dev*, **11**, 265-270.

94. Kuhn, H., Demidov, V.V., Coull, J.M., Fiandaca, M.J., Gildea, B.D. and Frank-Kamenetskii, M.D. (2002) Hybridization of DNA and PNA molecular beacons to single-stranded and double-stranded DNA targets. *J Am Chem Soc*, **124**, 1097-1103.

95. Lundin, K.E., Good, L., Stromberg, R., Graslund, A. and Smith, C.I. (2006) Biological activity and biotechnological aspects of peptide nucleic acid. *Adv Genet*, **56**, 1-51.

96. Nielsen, P.E., Egholm, M. and Buchardt, O. (1994) Evidence for (PNA)2/DNA triplex structure upon binding of PNA to dsDNA by strand displacement. *J Mol Recognit*, **7**, 165-170.

97. Dueholm, K.L., Egholm, M., Behrens, C., Christensen, L., Hansen, H.F., Vulpius, T., Petersen, K.H., Berg, R.H., Nielsen, P.E. and Buchardt, O. (1994) Synthesis of Peptide Nucleic Acid Monomers Containing the Four Natural Nucleobases: Thymine, Cytosine, Adenine, and Guanine and Their Oligomerization. *The Journal of Organic Chemistry*, **59**, 5767-5773.

98. Knudsen, H. and Nielsen, P.E. (1996) Antisense properties of duplex- and triplex-forming PNAs. *Nucleic acids research*, **24**, 494-500.

99. Wittung, P., Kajanus, J., Edwards, K., Haaima, G., Nielsen, P.E., Norden, B. and Malmstrom, B.G. (1995) Phospholipid membrane permeability of peptide nucleic acid. *FEBS Lett*, **375**, 27-29.

100. Adlerz, L., Soomets, U., Holmlund, L., Viirlaid, S., Langel, U. and Iverfeldt, K. (2003) Down-regulation of amyloid precursor protein by peptide nucleic acid oligomer in cultured rat primary neurons and astrocytes. *Neurosci Lett*, **336**, 55-59.

101. Tyler, B.M., McCormick, D.J., Hoshall, C.V., Douglas, C.L., Jansen, K., Lacy, B.W., Cusack, B. and Richelson, E. (1998) Specific gene blockade shows that peptide nucleic acids readily enter neuronal cells in vivo. *FEBS Lett*, **421**, 280-284.

102. Rezaei, K., Xu, I.S., Wu, W.P., Shi, T.J., Soomets, U., Land, T., Xu, X.J., Wiesenfeld-Hallin, Z., Hokfelt, T., Bartfai, T. *et al.* (2001) Intrathecal administration of PNA targeting galanin receptor reduces galanin-mediated inhibitory effect in the rat spinal cord. *Neuroreport*, **12**, 317-320.

103. Fraser, G.L., Holmgren, J., Clarke, P.B. and Wahlestedt, C. (2000) Antisense inhibition of delta-opioid receptor gene function in vivo by peptide nucleic acids. *Mol Pharmacol*, **57**, 725-731.

104. Nielsen, P.E. and Christensen, L. (1996) Strand displacement binding of a duplex-forming homopurine PNA to a homopyrimidine duplex DNA target. *Journal of the American Chemical Society*, **118**, 2287-2288.

105. Lohse, J., Dahl, O. and Nielsen, P.E. (1999) Double duplex invasion by peptide nucleic acid: a general principle for sequence-specific targeting of double-stranded DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 11804-11808.

106. Egholm, M., Christensen, L., Dueholm, K.L., Buchardt, O., Coull, J. and Nielsen, P.E. (1995) Efficient pH-independent sequence-specific DNA binding by pseudoisocytosine-containing bis-PNA. *Nucleic acids research*, **23**, 217-222.

107. Kaihatsu, K., Shah, R.H., Zhao, X. and Corey, D.R. (2003) Extending Recognition by Peptide Nucleic Acids (PNAs): Binding to Duplex DNA and Inhibition of Transcription by Tail-Clamp PNA−Peptide Conjugates. *Biochemistry*, **42**, 13996-14003.

108. Suh, D. and Chaires, J.B. (1995) Criteria for the mode of binding of DNA binding agents. *Bioorg Med Chem*, **3**, 723-728.

109. Lerman, L.S. (1961) Structural considerations in the interaction of DNA and acridines. *Journal of molecular biology*, **3**, 18-30.

110. Lerman, L.S. (1963) The structure of the DNA-acridine complex. *Proceedings of the National Academy of Sciences of the United States of America*, **49**, 94-102.

111. Satyanarayana, S., Dabrowiak, J.C. and Chaires, J.B. (1993) Tris(phenanthroline)ruthenium(II) enantiomer interactions with DNA: mode and specificity of binding. *Biochemistry*, **32**, 2573-2584.

112. Zeglis, B.M., Pierre, V.C. and Barton, J.K. (2007) Metallo-intercalators and metallo-insertors. *Chem Commun (Camb)*, 4565-4579.

113. Hunter, C.A. and Sanders, J.K.M. (1990) The nature of .pi.-.pi. interactions. *Journal of the American Chemical Society*, **112**, 5525-5534.

114. Müller-Dethlefs, K. and Hobza, P. (2000) Noncovalent Interactions: A Challenge for Experiment and Theory. *Chemical Reviews*, **100**, 143-168.

115. Meyer, E.A., Castellano, R.K. and Diederich, F. (2003) Interactions with aromatic rings in chemical and biological recognition. *Angew Chem Int Ed Engl*, **42**, 1210-1250.

116. Brana, M.F., Cacho, M., Gradillas, A., de Pascual-Teresa, B. and Ramos, A. (2001) Intercalators as anticancer drugs. *Curr Pharm Des*, **7**, 1745-1780.

117. Martinez, R. and Chacon-Garcia, L. (2005) The search of DNA-intercalators as antitumoral drugs: what it worked and what did not work. *Curr Med Chem*, **12**, 127-151.

118. Palchaudhuri, R. and Hergenrother, P.J. (2007) DNA as a target for anticancer compounds: methods to determine the mode of binding and the mechanism of action. *Curr Opin Biotechnol*, **18**, 497-503.

119. Alberts, D.S., Peng, Y.M., Bowden, G.T., Dalton, W.S. and Mackel, C. (1985) Pharmacology of mitoxantrone: mode of action and pharmacokinetics. *Invest New Drugs*, **3**, 101-107.

120. Scaria, P.V. and Shafer, R.H. (1991) Binding of ethidium bromide to a DNA triple helix. Evidence for intercalation. *Journal of Biological Chemistry*, **266**, 5417-5423.

121. Mergny, J.L., Collier, D., Rougee, M., Montenay-Garestier, T. and Helene, C. (1991) Intercalation of ethidium bromide into a triple-stranded oligonucleotide. *Nucleic acids research*, **19**, 1521-1526.

122. Mergny, J.L., Duval-Valentin, G., Nguyen, C.H., Perrouault, L., Faucon, B., Rougee, M., Montenay-Garestier, T., Bisagni, E. and Helene, C. (1992) Triple helix-specific ligands. *Science*, **256**, 1681-1684.

123. Lee, J.S., Latimer, L.J. and Hampel, K.J. (1993) Coralyne binds tightly to both T.A.T- and C.G.C(+)-containing DNA triplexes. *Biochemistry*, **32**, 5591-5597.

124. Latimer, L.J., Payton, N., Forsyth, G. and Lee, J.S. (1995) The binding of analogues of coralyne and related heterocyclics to DNA triplexes. *Biochem Cell Biol*, **73**, 11-18.

125. Tarui, M., Doi, M., Ishida, T., Inoue, M., Nakaike, S. and Kitamura, K. (1994) DNA-binding characterization of a novel anti-tumour benzo[a]phenazine derivative NC-182: spectroscopic and viscometric studies. *Biochem J*, **304 (Pt 1)**, 271-279.

126. Fox, K.R., Thurston, D.E., Jenkins, T.C., Varvaresou, A., Tsotinis, A. and Siatra-Papastaikoudi, T. (1996) A novel series of DNA triple helix-binding ligands. *Biochem Biophys Res Commun*, **224**, 717-720.

127. Fox, K.R., Polucci, P., Jenkins, T.C. and Neidle, S. (1995) A molecular anchor for stabilizing triple-helical DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 7887-7891.

128. Escude, C., Nguyen, C.H., Kukreti, S., Janin, Y., Sun, J.S., Bisagni, E., Garestier, T. and Helene, C. (1998) Rational design of a triple helix-specific intercalating ligand. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 3591-3596.

129. Zain, R., Polverari, D., Nguyen, C.H., Blouquit, Y., Bisagni, E., Garestier, T., Grierson, D.S. and Sun, J.S. (2003) Optimization of triple-helix-directed DNA cleavage by benzoquinoquinoxaline-ethylenediaminetetraacetic acid conjugates. *Chembiochem*, **4**, 856-862.

130. Zaid, A., Sun, J.S., Nguyen, C.H., Bisagni, E., Garestier, T., Grierson, D.S. and Zain, R. (2004) Triple-helix directed cleavage of double-stranded DNA by benzoquinoquinoxaline-1,10-phenanthroline conjugates. *Chembiochem*, **5**, 1550-1557.

131. Bergquist, H., Nikravesh, A., Fernandez, R.D., Larsson, V., Nguyen, C.H., Good, L. and Zain, R. (2009) Structure-specific recognition of Friedreich's ataxia (GAA)n repeats by benzoquinoquinoxaline derivatives. *Chembiochem*, **10**, 2629-2637.

132. Amiri, H., Nekhotiaeva, N., Sun, J.S., Nguyen, C.H., Grierson, D.S., Good, L. and Zain, R. (2005) Benzoquinoquinoxaline derivatives stabilize and cleave H-DNA and repress transcription downstream of a triplex-forming sequence. *Journal of molecular biology*, **351**, 776-783.

133. Guo, M., Hundseth, K., Ding, H., Vidhyasagar, V., Inoue, A., Nguyen, C.H., Zain, R., Lee, J.S. and Wu, Y. (2015) A distinct triplex DNA unwinding activity of ChlR1 helicase. *J Biol Chem*, **290**, 5174-5189.

134. Arya, D.P., Xue, L. and Tennant, P. (2003) Combining the best in triplex recognition: synthesis and nucleic acid binding of a BQQ-neomycin conjugate. *J Am Chem Soc*, **125**, 8070-8071.

135. Roulon, T., Coulaud, D., Delain, E., Le Cam, E., Helene, C. and Escude, C. (2002) Padlock oligonucleotides as a tool for labeling superhelical DNA. *Nucleic acids research*, **30**, E12.

136. Geron-Landre, B., Roulon, T., Desbiolles, P. and Escude, C. (2003) Sequence-specific fluorescent labeling of double-stranded DNA observed at the single molecule level. *Nucleic acids research*, **31**, e125.

137. Roulon, T., Helene, C. and Escude, C. (2002) Coupling of a targeting peptide to plasmid DNA using a new type of padlock oligonucleotide. *Bioconjug Chem*, **13**, 1134-1139.

138. Roulon, T., Helene, C. and Escude, C. (2001) A Ligand-Modulated Padlock Oligonucleotide for Supercoiled Plasmids T.R. was the recipient of a Bourse de Docteur Ingenieur from the CNRS. We are grateful to Prof. T. Garestier and J. S. Sun for helpful discussions, and to Dr. C. H. Nguyen for a gift of BQQ. *Angew Chem Int Ed Engl*, **40**, 1523-1526.

139. Bello-Roufai, M., Roulon, T. and Escude, C. (2004) Ligand-mediated transcription elongation control using triplex-based padlock oligonucleotides. *Chem Biol*, **11**, 509-516.

140. Zain, R., Marchand, C., Sun, J., Nguyen, C.H., Bisagni, E., Garestier, T. and Helene, C. (1999) Design of a triple-helix-specific cleaving reagent. *Chem Biol*, **6**, 771-777.

141. Filichev, V.V. and Pedersen, E.B. (2005) Stable and selective formation of hoogsteen-type triplexes and duplexes using twisted intercalating nucleic acids (TINA) prepared via postsynthetic Sonogashira solid-phase coupling reactions. *J Am Chem Soc*, **127**, 14849-14858.

142. Christensen, U.B., Wamberg, M., El-Essawy, F.A., Ismail Ael, H., Nielsen, C.B., Filichev, V.V., Jessen, C.H., Petersen, M. and Pedersen, E.B. (2004) Intercalating nucleic acids: the influence of linker length and intercalator type on their duplex stabilities. *Nucleosides Nucleotides Nucleic Acids*, **23**, 207-225.

143. Filichev, V.V., Gaber, H., Olsen, T. R., Jørgensen, P. T., Jessen, C. H. and Pedersen, E. B. (2006) Twisted Intercalating Nucleic Acids – Intercalator Influence on Parallel Triplex Stabilities. *Eur. J. Org. Chem.*, **2006**, 3960–3968.

144. Geci, I., Filichev, V.V. and Pedersen, E.B. (2006) Synthesis of twisted intercalating nucleic acids possessing acridine derivatives. Thermal stability studies. *Bioconjug Chem*, **17**, 950-957.

145. Schneider, U.V., Geci, I., Johnk, N., Mikkelsen, N.D., Pedersen, E.B. and Lisby, G. (2011) Increasing the analytical sensitivity by oligonucleotides modified with para- and ortho-twisted intercalating nucleic acids--TINA. *PLoS One*, **6**, e20565.

146. Armitage, B.A. (2005) In Waring, M. J. and Chaires, J. B. (eds.), *DNA Binders and Related Subjects: -/-*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 55-76.

147. Wartell, R.M., Larson, J.E. and Wells, R.D. (1974) Netropsin. A specific probe for A-T regions of duplex deoxyribonucleic acid. *J Biol Chem*, **249**, 6719-6731.

148. Dervan, P.B. (2001) Molecular recognition of DNA by small molecules. *Bioorg Med Chem*, **9**, 2215-2235.

149. Levine, M. and Tjian, R. (2003) Transcription regulation and animal diversity. *Nature*, **424**, 147-151.

150. Pabo, C.O. and Sauer, R.T. (1992) Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem*, **61**, 1053-1095.

151. Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V. and Romano, L.A. (2003) The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol*, **20**, 1377-1419.

152. Harrison, S.C. (1991) A structural taxonomy of DNA-binding domains. *Nature*, **353**, 715-719.

153. Lucchini, G., Hinnebusch, A.G., Chen, C. and Fink, G.R. (1984) Positive regulatory interactions of the HIS4 gene of Saccharomyces cerevisiae. *Mol Cell Biol*, **4**, 1326-1333.

154. Ellenberger, T.E., Brandl, C.J., Struhl, K. and Harrison, S.C. (1992) The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. *Cell*, **71**, 1223-1237.

155. Friedmann, T. and Roblin, R. (1972) Gene therapy for human genetic disease? *Science*, **175**, 949-955.

156. Cepko, C.L., Roberts, B.E. and Mulligan, R.C. (1984) Construction and applications of a highly transmissible murine retrovirus shuttle vector. *Cell*, **37**, 1053-1062.

157. Blaese, R.M., Culver, K.W., Miller, A.D., Carter, C.S., Fleisher, T., Clerici, M., Shearer, G., Chang, L., Chiang, Y., Tolstoshev, P. *et al.* (1995) T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years. *Science*, **270**, 475-480.

158. Cavazzana-Calvo, M., Hacein-Bey, S., de Saint Basile, G., Gross, F., Yvon, E., Nusbaum, P., Selz, F., Hue, C., Certain, S., Casanova, J.L. *et al.* (2000) Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease. *Science*, **288**, 669-672.

159. Pearson, S., Jia, H. and Kandachi, K. (2004) China approves first gene therapy. *Nature biotechnology*, **22**, 3-4.

160. Brown, B.D., Venneri, M.A., Zingale, A., Sergi Sergi, L. and Naldini, L. (2006) Endogenous microRNA regulation suppresses transgene expression in hematopoietic lineages and enables stable gene transfer. *Nature medicine*, **12**, 585-591.

161. Morgan, R.A., Dudley, M.E., Wunderlich, J.R., Hughes, M.S., Yang, J.C., Sherry, R.M., Royal, R.E., Topalian, S.L., Kammula, U.S., Restifo, N.P. *et al.* (2006) Cancer regression in patients after transfer of genetically engineered lymphocytes. *Science*, **314**, 126-129.

162. Ginn, S.L., Alexander, I.E., Edelstein, M.L., Abedi, M.R. and Wixon, J. (2013) Gene therapy clinical trials worldwide to 2012 - an update. *The journal of gene medicine*, **15**, 65-77.

163. Helene, C. (1991) The anti-gene strategy: control of gene expression by triplex-forming-oligonucleotides. *Anti-cancer drug design*, **6**, 569-584.

164. Cooney, M., Czernuszewicz, G., Postel, E.H., Flint, S.J. and Hogan, M.E. (1988) Site-specific oligonucleotide binding represses transcription of the human c-myc gene in vitro. *Science*, **241**, 456-459.

165. Cogoi, S., Ballico, M., Bonora, G.M. and Xodo, L.E. (2004) Antiproliferative activity of a triplex-forming oligonucleotide recognizing a Ki-ras polypurine/polypyrimidine motif correlates with protein binding. *Cancer Gene Ther*, **11**, 465-476.

166. Shen, C., Rattat, D., Buck, A., Mehrke, G., Polat, B., Ribbert, H., Schirrmeister, H., Mahren, B., Matuschek, C. and Reske, S.N. (2003) Targeting bcl-2 by triplex-forming oligonucleotide--a promising carrier for gene-radiotherapy. *Cancer Biother Radiopharm*, **18**, 17-26.

167. Ebbinghaus, S.W., Gee, J.E., Rodu, B., Mayfield, C.A., Sanders, G. and Miller, D.M. (1993) Triplex formation inhibits HER-2/neu transcription in vitro. *J Clin Invest*, **92**, 2433-2439.

168. Singhal, G., Akhter, M.Z., Stern, D.F., Gupta, S.D., Ahuja, A., Sharma, U., Jagannathan, N.R. and Rajeswari, M.R. (2011) DNA triplex-mediated inhibition of MET leads to cell death and tumor regression in hepatoma. *Cancer Gene Ther*, **18**, 520-530.

169. Jendis, J., Strack, B., Volkmann, S., Boni, J. and Molling, K. (1996) Inhibition of replication of fresh HIV type 1 patient isolates by a polypurine tract-specific self-complementary oligodeoxynucleotide. *AIDS Res Hum Retroviruses*, **12**, 1161-1168.

170. Jendis, J., Strack, B. and Moelling, K. (1998) Inhibition of replication of drug-resistant HIV type 1 isolates by polypurine tract-specific oligodeoxynucleotide TFO A. *AIDS Res Hum Retroviruses*, **14**, 999-1005.

171. Brossalina, E., Pascolo, E. and Toulme, J.J. (1993) The binding of an antisense oligonucleotide to a hairpin structure via triplex formation inhibits chemical and biological reactions. *Nucleic acids research*, **21**, 5616-5622.

172. Volkmann, S., Jendis, J., Frauendorf, A. and Moelling, K. (1995) Inhibition of HIV-1 reverse transcription by triple-helix forming oligonucleotides with viral RNA. *Nucleic acids research*, **23**, 1204-1212.

173. Goni, J.R., de la Cruz, X. and Orozco, M. (2004) Triplex-forming oligonucleotide target sequences in the human genome. *Nucleic acids research*, **32**, 354-360.

174. Goni, J.R., Vaquerizas, J.M., Dopazo, J. and Orozco, M. (2006) Exploring the reasons for the large density of triplex-forming oligonucleotide target sequences in the human regulatory regions. *BMC Genomics*, **7**, 63.

175. Buske, F.A., Bauer, D.C., Mattick, J.S. and Bailey, T.L. (2012) Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res*, **22**, 1372-1381.

176. Buske, F.A., Bauer, D.C., Mattick, J.S. and Bailey, T.L. (2013) Triplex-Inspector: an analysis tool for triplex-mediated targeting of genomic loci. *Bioinformatics*, **29**, 1895-1897.

177. Jenjaroenpun, P., Chew, C.S., Yong, T.P., Choowongkomon, K., Thammasorn, W. and Kuznetsov, V.A. (2015) The TTSMI database: a catalog of triplex target DNA sites associated with genes and regulatory elements in the human genome. *Nucleic acids research*, **43**, D110-116.

178. Cer, R.Z., Donohue, D.E., Mudunuri, U.S., Temiz, N.A., Loss, M.A., Starner, N.J., Halusa, G.N., Volfovsky, N., Yi, M., Luke, B.T. *et al.* (2013) Non-B DB v2.0: a database of predicted non-B DNA-forming motifs and its associated tools. *Nucleic acids research*, **41**, D94-D100.

179. Gaddis, S.S., Wu, Q., Thames, H.D., DiGiovanni, J., Walborg, E.F., MacLeod, M.C. and Vasquez, K.M. (2006) A web-based search engine for triplex-forming oligonucleotide target sequences. *Oligonucleotides*, **16**, 196-201.

180. Jenjaroenpun, P. and Kuznetsov, V.A. (2009) TTS mapping: integrative WEB tool for analysis of triplex formation target DNA sequences, G-quadruplets and non-protein coding regulatory DNA elements in the human genome. *BMC Genomics*, **10 Suppl 3**, S9.

181. Consortium, E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57-74.

182. Speir, M.L., Zweig, A.S., Rosenbloom, K.R., Raney, B.J., Paten, B., Nejad, P., Lee, B.T., Learned, K., Karolchik, D., Hinrichs, A.S. *et al.* (2016) The UCSC Genome Browser database: 2016 update. *Nucleic acids research*, **44**, D717-725.

183. McCammon, J.A., Gelin, B.R. and Karplus, M. (1977) Dynamics of folded proteins. *Nature*, **267**, 585-590.

184. Karplus, M. and Kuriyan, J. (2005) Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 6679-6685.

185. Brunger, A.T., Kuriyan, J. and Karplus, M. (1987) Crystallographic R factor refinement by molecular dynamics. *Science*, **235**, 458-460.

186. Karplus, M. and McCammon, J.A. (2002) Molecular dynamics simulations of biomolecules. *Nat Struct Biol*, **9**, 646-652.

187. Landegren, U., Kaiser, R., Caskey, C.T. and Hood, L. (1988) DNA diagnostics--molecular techniques and automation. *Science*, **242**, 229-237.

188. Chaconas, G. and van de Sande, J.H. (1980) 5'-32P labeling of RNA and DNA restriction fragments. *Methods Enzymol*, **65**, 75-85.

189. Fried, M. and Crothers, D.M. (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic acids research*, **9**, 6505-6525.

190. Garner, M.M. and Revzin, A. (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic acids research*, **9**, 3047-3060.

191. Hellman, L.M. and Fried, M.G. (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc*, **2**, 1849-1861.

192. Jain, S.S. and Tullius, T.D. (2008) Footprinting protein-DNA complexes using the hydroxyl radical. *Nat Protoc*, **3**, 1092-1100.

193. Jagannathan, I. and Hayes, J.J. (2009) Hydroxyl radical footprinting of protein-DNA complexes. *Methods Mol Biol*, **543**, 57-71.

194. Stonehouse, T.J. and Fox, K.R. (1994) DNase I footprinting of triple helix formation at polypurine tracts by acridine-linked oligopyrimidines: stringency, structural changes and interaction with minor groove binding ligands. *Biochimica et biophysica acta*, **1218**, 322-330.

195. Brown, P.M., Madden, C.A. and Fox, K.R. (1998) Triple-helix formation at different positions on nucleosomal DNA. *Biochemistry*, **37**, 16139-16151.

196. Fox, K.R. and Waring, M.J. (1986) Footprinting reveals that nogalamycin and actinomycin shuffle between DNA binding sites. *Nucleic acids research*, **14**, 2001-2014.

197. Hampshire, A.J., Rusling, D.A., Broughton-Head, V.J. and Fox, K.R. (2007) Footprinting: a method for determining the sequence selectivity, affinity and kinetics of DNA-binding ligands. *Methods*, **42**, 128-140.

198. Galas, D.J. and Schmitz, A. (1978) DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic acids research*, **5**, 3157-3170.

199. Sanger, F., Nicklen, S. and Coulson, A.R. (1992) DNA sequencing with chain-terminating inhibitors. 1977. *Biotechnology*, **24**, 104-108.

200. Kusmierek, J.T. and Singer, B. (1982) Chloroacetaldehyde-treated ribo- and deoxyribopolynucleotides. 1. Reaction products. *Biochemistry*, **21**, 5717-5722.

201. Kohwi-Shigematsu, T. and Kohwi, Y. (1992) Detection of non-B-DNA structures at specific sites in supercoiled plasmid DNA and chromatin with haloacetaldehyde and diethyl pyrocarbonate. *Methods Enzymol*, **212**, 155-180.

202. Demidov, V., Frank-Kamenetskii, M.D., Egholm, M., Buchardt, O. and Nielsen, P.E. (1993) Sequence selective double strand DNA cleavage by peptide nucleic acid (PNA) targeting using nuclease S1. *Nucleic acids research*, **21**, 2103-2107.

203. Armitage, B., Koch, T., Frydenlund, H., Orum, H., Batz, H.G. and Schuster, G.B. (1997) Peptide nucleic acid-anthraquinone conjugates: strand invasion and photoinduced cleavage of duplex DNA. *Nucleic acids research*, **25**, 4674-4678.

204. Zianni, M., Tessanne, K., Merighi, M., Laguna, R. and Tabita, F.R. (2006) Identification of the DNA bases of a DNase I footprint by the use of dye primer sequencing on an automated capillary DNA analysis instrument. *J Biomol Tech*, **17**, 103-113.

205. Kabotyanski, E.B., Zhu, C., Kallick, D.A. and Roth, D.B. (1995) Hairpin opening by single-strand-specific nucleases. *Nucleic acids research*, **23**, 3872-3881.

206. Fujimoto, M., Kalinski, A., Pritchard, A.E., Kowalski, D. and Laskowski, M., Sr. (1979) Accessibility of some regions of DNA in chromatin (chicken erythrocytes) to single strand-specific nucleases. *J Biol Chem*, **254**, 7405-7410.

207. Hoffman, E.K., Trusko, S.P., Murphy, M. and George, D.L. (1990) An S1 nuclease-sensitive homopurine/homopyrimidine domain in the c-Ki-ras promoter interacts with a nuclear factor. *Proceedings of the National Academy of Sciences of the United States of America*, **87**, 2705-2709.

208. Larsen, A. and Weintraub, H. (1982) An altered DNA conformation detected by S1 nuclease occurs at specific regions in active chick globin chromatin. *Cell*, **29**, 609-622.

209. Lee, J.S., Woodsworth, M.L., Latimer, L.J. and Morgan, A.R. (1984) Poly(pyrimidine) . poly(purine) synthetic DNAs containing 5-methylcytosine form stable triplexes at neutral pH. *Nucleic acids research*, **12**, 6603-6614.

210. Christophe, D., Cabrer, B., Bacolla, A., Targovnik, H., Pohl, V. and Vassart, G. (1985) An unusually long poly(purine)-poly(pyrimidine) sequence is located upstream from the human thyroglobulin gene. *Nucleic acids research*, **13**, 5127-5144.

211. Geny, S., Moreno, P.M., Krzywkowski, T., Gissberg, O., Andersen, N.K., Isse, A.J., El-Madani, A.M., Lou, C., Pabon, Y.V., Anderson, B.A. *et al.* (2016) Next-generation bis-locked nucleic acids with stacking linker and 2'-glycylamino-LNA show enhanced DNA invasion into supercoiled duplexes. *Nucleic Acids Res*.