Stony Brook University

Nitrogen Mustard Reloaded: Synthesis and Structural Studies of DNA Interstrand Crosslinks

A Dissertation Presented

by

Angelo Guainazzi

to

The Graduate School

In Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Molecular and Cellular Pharmacology

Stony Brook University

August 2010

Stony Brook University

The Graduate School

Angelo Guainazzi

We, the dissertation committee for the above candidate for Doctor in Philosophy degree, hereby recommend acceptance of this dissertation.

Orlando D. Schärer, PhD – Dissertation Advisor
Associate Professor, Department of Pharmacological Sciences

Carlos De los Santos, PhD – Chairperson of Defense
Associate Professor, Department of Pharmacological Sciences

Francis Johnson, PhD
Professor, Department of Pharmacological Sciences

Carlos Simmerling, PhD
Associate Professor, Department of Chemistry

Isaac Carrico, PhD
Assistant Professor, Department of Chemistry, Stony Brook University

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Nitrogen Mustard Reloaded: Synthesis and Structural Studies of DNA Interstrand Crosslinks

by

Angelo Guainazzi

Doctor in Philosophy

in

Molecular and Cellular Pharmacology

Stony Brook University

2010

Many of the drugs used in cancer chemotherapy target DNA to kill malignant cells. Some of them form DNA interstrand crosslinks (ICLs), which are extremely cytotoxic lesions that block essential metabolic process such as replication, transcription and recombination by forming covalent bonds between opposite strands of DNA. Despite the importance of chemotherapeutic agents that rely on ICLs for their efficacy, the mechanisms by which these lesions are repaired remains poorly understood. A major impediment in studying ICLs repair has been the limited availability of well-defined substrates.

This dissertation describes the development of a new strategy for the synthesis of defined site-specific ICLs in high yields and purity. This strategy relies on the incorporation of ICL precursors bearing reactive aldehyde functionalities on

complementary strands of DNA, followed by ICL formation via double reductive amination. We were able to synthesize different crosslinks that are isosteric to the therapeutic nitrogen mustard (NM) ICLs, introducing substitution of a few atoms to make them more stable and therefore more suitable for chemical, structural and biological studies.

The synthetic substrates were validated through molecular dynamic studies, confirming that our mimic has all the essential structural features to its natural counterpart. Modeling data also demonstrate that both the natural and the synthetic ICL induce a bend in the DNA, which could play an important role in the way the lesion is repaired. Our synthetic approach furthermore allows for the synthesis of major groove ICLs with different degrees of distortion, providing unique and valuable tools for biochemical and cell biological studies of ICL repair.

Dedicated to my mother

Table of Contents

List of Figures

Chapter 1:

Chapter 2:

Chapter 3:

Acknowledgments

I would like to thank my advisor, Orlando Schärer, for giving me the opportunity to join his group and work on this project. He was always supportive and encouraging even in those moments where I thought I would never finish my PhD. I appreciate your patience and trust in my way of working independently but still always ready to help me when I need it.

I am thankful to all the people who collaborate to this project, particularly AJ Campbell with whom I shared opinions, ideas, moments of enthusiasm for some results and moments of despair for some other ones. Thanks a lot to Tanya Zaliznyak and Carlos de Los Santos for their great help with the NMR, thanks to Vinh for using my substrate for some useful experiment and to Robert Rieger for being so helpful with the MS analysis of my samples. Very importantly I must thank Todor Angelov who started this project and set up the basis for my successful PhD studies.

I thank the other members of my thesis committee Carlos Simmerling, Francis Johnson and Isaac Carrico for all the suggestions and critics that improved the quality of my project.

I would like also to thank all the people from the IMCR in Zurich where I started this tortuous path to the PhD. Particularly the people of the Schärer's group who did not came with us on this side of the ocean, Ludovic and Jawad who helped me a lot at the beginning, Rolf, Müriel and Milica.

Thanks to all the member of the Schärer's group, particularly Jung-Eun with whom I shared many opinions and ideas on the chemistry part of my project. Thanks to Lidija, Banke, Jerome, Yan, Burak, Shivam who will work on this project in the future and Julie, my rotation student that worked with me on the project for more than a year.

A very special thanks goes to a very special member of the group that was so much more than that. She was the person that I have on my side, that supported me unconditionally, that made me laugh in the moments I need it, that made my life in the

last x years (I don't want to know how long my PhD was) so much easier to live and to enjoy. Thanks a lot Babs.

Then I have to thanks my friends that makes life possible. I can't imagine a life without friends… My buddy Jose with whom I shared so many things on Long Island, running, football (played and on TV), gardening and much more, Victoria, Sergio, Pablo, Robi Juan C., Juan G. and many more. All the members the LasTeamA, all the people that played football with me.

Gigantic thanks to my very good friend in Switzerland, particularly the ones that flooded my inbox with tons of emails everyday and that although they are far far away they were always and always will be very close! You really are good friends: Wombie, Vano, Torria, Meui, Geo, Claude and of course all the others that I see at the "Guaina day".

Great special thanks to my parents, my siblings and all my family that always supported and encouraged me and particularly believed in me, sometimes more than what I do. I particularly thank my father also for all the times he came here to visit me making the distance between Europe and NY much shorter.

A great thanks to a special person like Veronica that was certainly one of the most important people of my life for a long time and will always remain in my heart, she gave me the enthusiasm and the courage to come to the US.

Very much thanks to all the member of the pharmacology department, especially Dumaine that was so kind and helpful when we first arrived into the new world.

Thanks to all the people that I should have mentioned but I forgot, to the ones that are going to read this dissertation and to the ones who are going to come to my defense.

Chapter 1

General introduction: Targeting DNA Interstrand Crosslink Repair Repair in Cancer Chemotherapy

Foreword

I wished to introduce my dissertation from a general prospective, describing the background that constituted the incentive and the rationale to investigate such a complex molecular process as interstrand crosslink repair. This introduction should make the reader aware of what the implications and the benefits of the results obtained in this project are.

This chapter is adapted from the review "Using synthetic DNA interstrand crosslinks to elucidate repair pathways and identify new therapeutic targets for cancer chemotherapy" by Angelo Guainazzi, and Orlando D. Schärer, 2010, Cellular and Molecular Life Science, in press

The killer from within

The term cancer refers to a class of diseases characterized by an uncontrolled and abnormal growth of cells and invasion of adjacent tissues. Cancer cells break the basic rules of behavior that a multicellular organism needs to survive and prosper - the sacrifice of individual cells for the good of the whole organism. Cells send, receive, interpret and elaborate signals that tell them when to replicate, grow or die depending on the need of the organism. Considering the extremely high number of cells and the complexity of cellular metabolism, mistakes in different metabolic steps are inevitable. If the damage is extensive, incorrectly behaving cells can be eliminated by apoptosis. However, mistakes can also lead to the generation of mutations in the genome and although most of the time these mutations do not affect the behavior of the cell, sometimes they involve key regulatory genes. In this case they can cause the cell to escape the social control by allowing faster and more vigorous division compared to its neighbors and to proliferate autonomously. This selfish and insubordinate cell can potentially jeopardize the entire organism because if its proliferation is out of control it will give rise to a tumor, also known as neoplasm. As long as tumor cells remain clustered together the tumor is said to be benign and can be removed by surgery, conversely if its cells invade adjacent tissues the tumor is said to be malignant. Sometimes malignant cells also spread into different parts of the body via bloodstream or lymphatic system giving rise to secondary tumors, in a process known as metastasis. A malignant tumor can be defined as cancer although some types of cancer like leukemia do not form tumors at all. Interestingly the term cancer was first defined by the Greek physician Hippocrates, considered to be “the father of medicine”. According to the legend, he noticed that a malignant tumor spreads the blood vessels on all sides, resembling the leg of a crab and he therefore called it *carcino*s (greek for crab).

The struggle against cancer

To cure cancer all malignant cells have to be killed. There are several problems with the efficacy of cancer treatment: first of all, it is very difficult to selectively kill cancer cells versus normal cells, second, there is minimal participation of the immune system in this process and third, during a treatment the symptoms disappear even before all malignant cells are removed, often leading to renewed tumor growth.

The only way to treat cancer until the end of the 19th century was the surgical removal of the tumor. This of course was often not enough to save the life of the patient, since the many cancer cells that remain in the organism are still able to proliferate and regrow. The first alternative to surgery was developed after William Röntgen discovered a new type of rays in 1895 [2]. He almost immediately realized that these rays could be used them for medical applications photographing his wife's hand and formally giving rise to radiology. The use of ionizing radiation for the treatment of cancer (radiotherapy) followed very quickly and improved with the contribution by Marie and Pierre Curie in discovering radioactivity. Radiotherapy is still widely used for the treatment of a variety of malignancies (reviewed in [3]) and was the dominant way to treat cancer until the advent of chemotherapy.

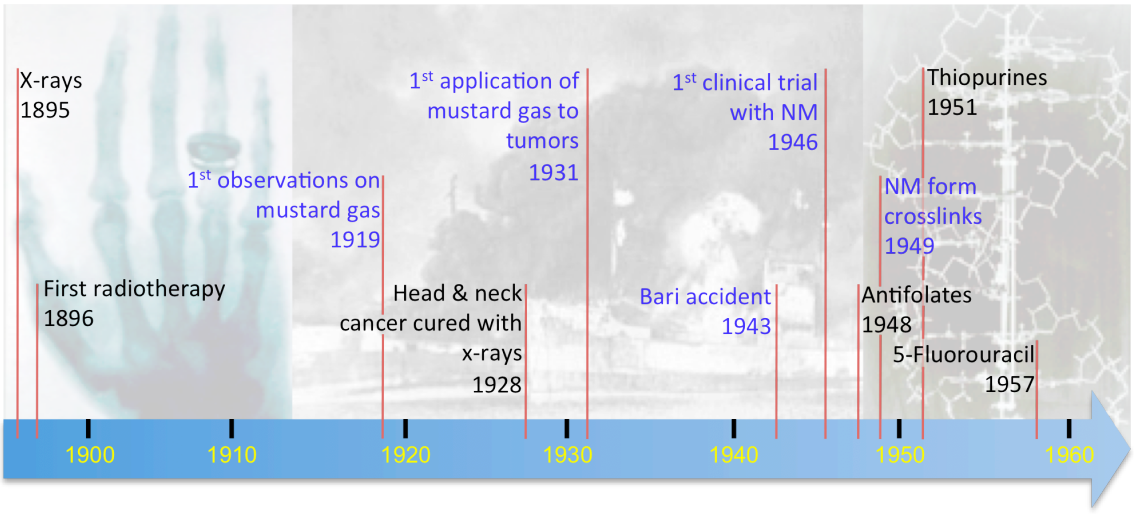


Figure 1a: Timeline showing the progress of anticancer therapy until 1960. The events related to nitrogen mustard are shown in blue. Background images show Röntgen's first X-rays picture [2], Bari's harbor after the attack and a DNA structure model by Watson and Crick [4].

The killer that became a savior

The term chemotherapy refers to the use of chemical compounds to treat disease but is popularly associated with cancer treatment. The first hint that chemicals could be used to kill malignant cells with some selectivity came during the First World War when Krumbhaar, a captain of the U.S. medical corps, reported severe leucopenia and suppressed bone marrow of soldiers who survived several days after mustard gas exposure [5]. The first application of mustard gas to treat tumors occurred in the years between the two world wars with a study reporting successful eradication of tumors in animals [6]. It was not though until Second World War that more in-depth analysis and real clinical trials were performed. In fact, the sulfur mustard gas used in WWI was substituted with the more stable nitrogen mustard gas that was used by Japan in the war against China, but never in Europe. Nonetheless, an accident in Bari, Italy is the episode that marked the switch to use nitrogen mustards as chemotherapeutic agents rather than in warfare. In 1943, Bari was a major seaport selected to supply the Allied troops; the harbor was crowded with a lot of medical personnel from the USA, UK and New Zealand. At the docks was anchored, between other ships, the Liberty ship *John Harvey*, laden with ammunition, gasoline and most importantly 2000 hundred-pound

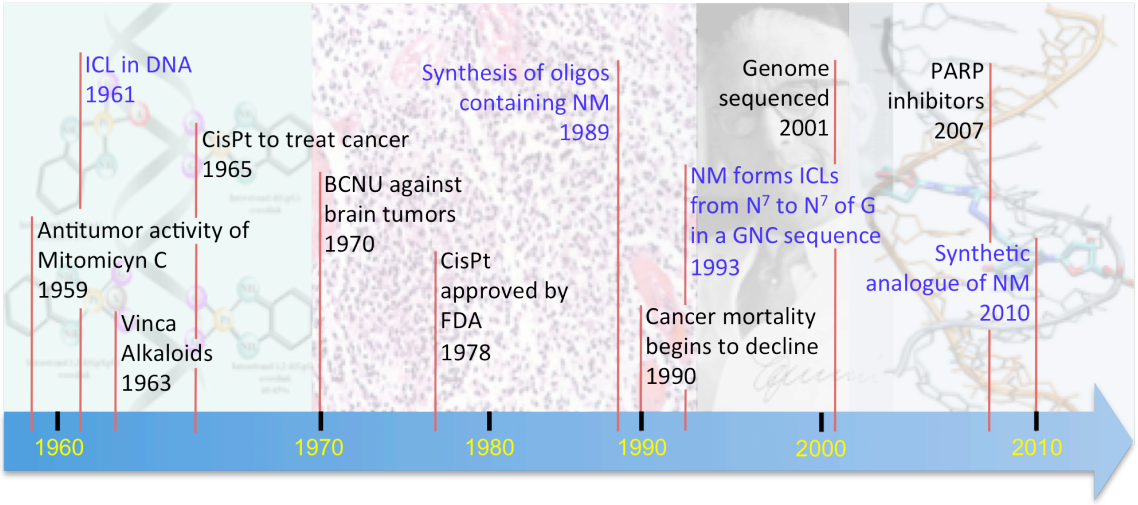


Figure 1b: Timeline showing the progress of anticancer therapy from 1960. Background images show oxaliplatin adducts to the DNA [7], a neoplastic tissue, Guido Fanconi and the model of the NM ICL analogue. Some fact on the timeline was taken from [8].

mustards bombs [9]. Unfortunately for the Allies, the evening of December 2, 1943 a group of JU88 planes from the Axis attacked and destroyed almost every ship in the Bari harbor, including the John Harvey. Military personnel and civilians alike were exposed to low doses of a toxic substance that was mixed with gasoline. Due to the secrecy of the war, nobody knew about the presence of a ship containing nitrogen mustards. Thus, in the days following the disaster, there was no particular concern about the toxic effects of chemical warfare agents. However, after few days the medical personnel, finally informed of the nitrogen mustard cargo, recorded a severe drop in white blood cell counts in many people that survived the accident and started to consider whether nitrogen mustards could have some useful medical application [10].

In reality, clinical trials of nitrogen mustards began before the accident in Bari, as reported in the results published by Goodman in 1946 [11], but everything was kept secret until the end of the war when Gilman and Philips reported on the biological action and therapeutic application of nitrogen mustards [12]. Shortly thereafter a more complete report, finally free from the secrecy of the war, described in more detail how mechlorethamine, a nitrogen mustard, was used in clinical trials for the treatment of Hodgkin's disease and leukemia [13,14]. A new line of treatment for cancer was born.

DNA as a target

After the war, many investigators focused their research on improving and perfecting antitumor therapies with nitrogen mustards. Numerous derivatives of mechlorethamine such as chlorambucil, cyclophosphamide and ifosfamide were developed that are still mainstays of the current antitumor therapy. The precise mechanism of toxicity of these drugs was not known at the time. There was evidence that they could damage the genetic material, even though the nature of genes was not yet established. The possibility of a crosslink formation emerged in a study published in 1949 by Goldacre and coworkers [15], although it was still not clear what cellular entities were being crosslinked.

Only several years later, after the elucidation of the double helical structure of DNA, was it established that nitrogen mustards form DNA interstrand crosslinks (ICLs) by reacting with the N^7 positions of two guanine residues on two complementary strands [16-18]. Nevertheless both DNA and DNA metabolism in general, were already a main target for other chemotherapeutic agents. For instance thiopurines, which target adenine metabolism and antifolates, which target thymidine biosynthesis (Figure 2) were developed shortly after WW II [8]. The identification of DNA adducts revealed that alkylating agents like nitrogen mustard interfere with essential aspects of metabolism such as DNA replication and transcription, triggering cell death. The formation of DNA ICLs is a particularly toxic event, as strand separation is made impossible by the covalent linkage of two DNA strands [19,20].

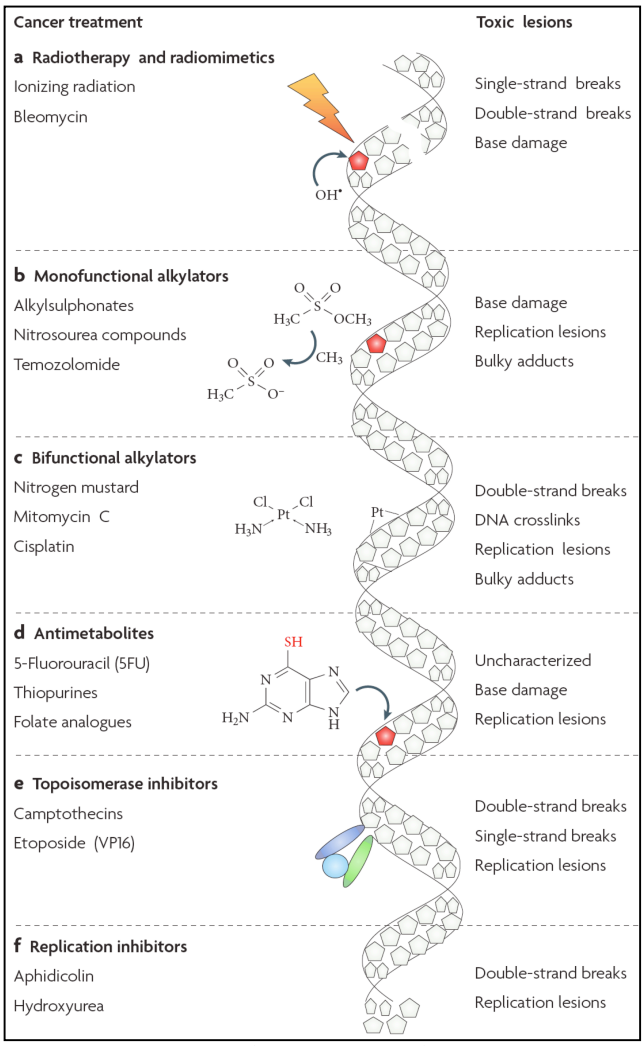


Figure 2: Overview of the variety of toxic DNA lesions caused by different cancer treatments. Adapted from [1]

ICL formation was subsequently demonstrated for a number of additional antitumor agents that were developed in the following years, including platinum complexes, mitomycin C and chloro ethyl nitroso ureas [18,21].

The problem of resistance

The formation of ICLs, like most DNA lesions, triggers cellular signaling cascades and repair events that are associated with resistance of tumor cells to ICL-forming agents [1]. Due to genetic make up of tumor cells and their generally higher mutation rates, the levels of resistance to chemotherapeutic agents vary greatly and present a serious problem in finding optimal therapies.

ICLs are however not only formed by agents used in cancer chemotherapy. A number of endogenous and environmental agents such malondialdehyde (a lipid peroxidation product), acetaldehyde and natural products like the furocoumarins (present in plants and cosmetics), can form ICLs [22]. Since such adducts constitute a serious threat to cellular survival, they have provided an evolutionary incentive for organisms to develop the ability to repair ICLs. The importance of these repair pathways is evidenced by the existence of the human cancer-prone disorder Fanconi anemia (FA). Cells from FA patients are specifically sensitive to crosslinking agents, implicating deficient ICL repair as a cause for the genomic instability that leads to carcinogenesis. While these repair pathways are vital for healthy cells, they cause resistance to ICL-forming agents in a therapeutic setting [23,24].

Generation and structures of ICL-containing DNA

One of the prerequisites for studying ICL repair is the availability of defined ICL-containing oligonucleotides. Initially, this was achieved by the reaction of duplex DNA with crosslinking agents (Figure 3A). While this approach led to the generation of DNA duplexes with ICLs formed by nitrogen mustards, cisplatin, BCNU or mitomycin C, it is very inefficient, yielding only about 1-5% of ICL adducts, with monoadducts and intrastrand crosslinks making up the majority of products [19,20]. This approach can be improved to a minor extent by generating first a monoadduct with a single-stranded oligonucleotide, followed by annealing to a complementary strand and ICL formation (Figure 3B). More targeted strategies for the synthesis of ICLs have recently been

developed that make use of site specific solid-phase DNA synthesis and introduce the ICL either in form of a crosslinked dimer (Figure 3C) or as ICL precursors that can undergo a specific coupling reaction after incorporation on complementary strands and

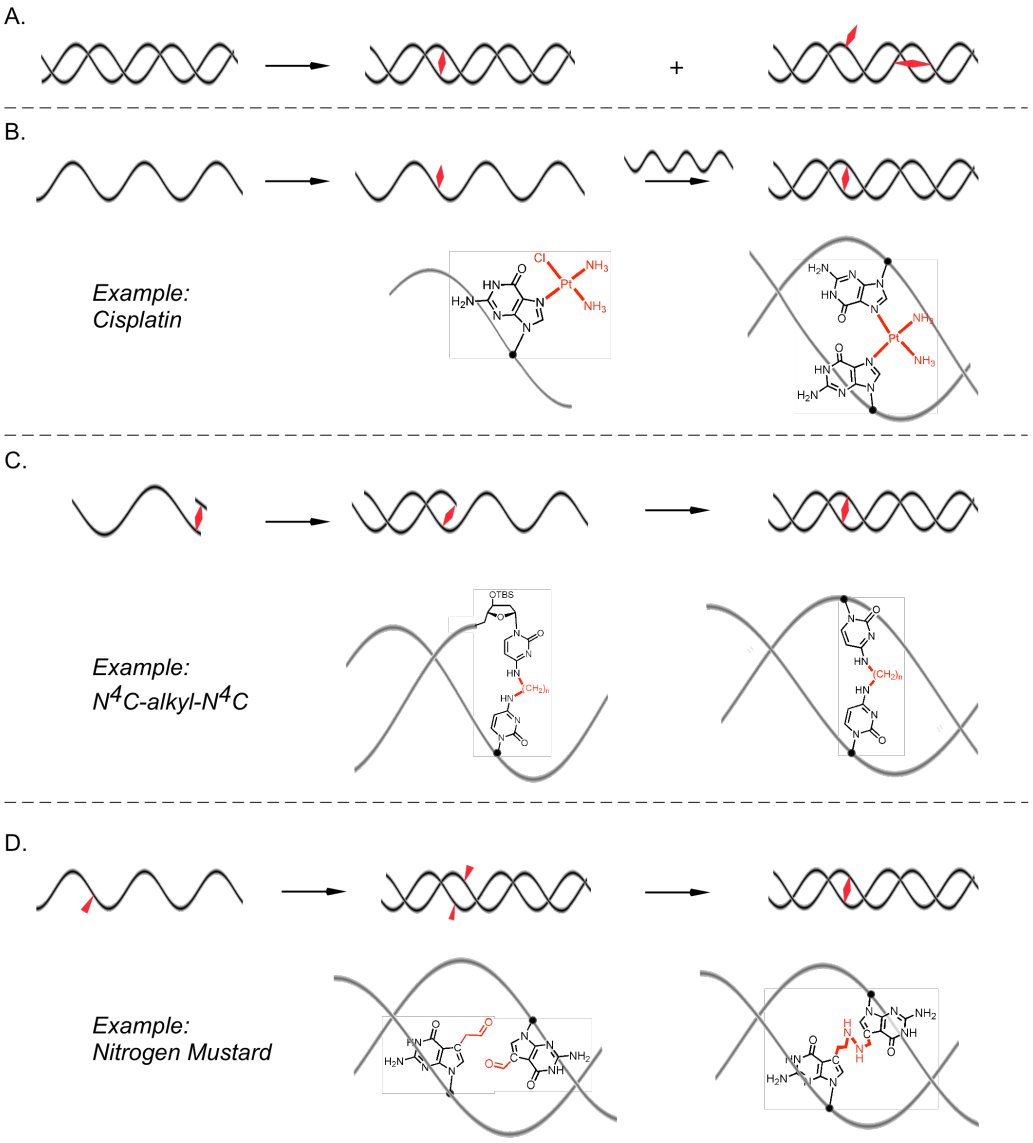


Figure 3. Methods for the preparation of DNA interstrand crosslinks. A. Treatment of an oligonuceotide with a bifunctional alkylating agent. This method gives rise to a mixture of products (*intra*- and *inter*strand crosslinks, monoaducts) and ICL typically make up less than 5% of all the products. B. Two step ICL formation. After reaction of a single strand containing one guanine residue, the monoadduct is purified, annealed with a complementary strand and activated to react with the other strand. This method is more efficient that A., but yields do not typically exceed 20%. C. Multi-step solid phase synthesis; the crosslink is chemically synthesized as a nucleotide dimer and incorporated into DNA followed by bidirectional solid-phase DNA synthesis. This approach yields highly specific ICLs. D. Two crosslink precursors are incorporated into DNA using solid-phase DNA synthesis and the ICL formed by a selective post-synthetic crosslinking reaction. This approach also yields ICLs with high specificity and in high yields. ICLs are shown as red diamonds.

annealing (Figure 3D). These approaches are discussed in more detail in conjunction with crosslinking agents and structures of ICL-containing oligonucleotides below.

Nitrogen Mustard ICLs

Nitrogen mustards (NM) react with the N^7 position of guanine residues through the reactive *N*,*N*-bis-(2-chloroethyl)amine functional group. NMs preferentially form a 1,3 ICL in a 5'-d(GNC) sequence context [25-27]. It has been shown that NM ICLs induce a slight bend in duplex DNA (~15°) to best accommodate the bridge between the two strands. In fact the maximal theoretical length of the 5 atom bridge formed by NM (~7.5 Å), is less than the average distance between the N^7-G sites in 5'-d(GNC) sequence (~8.9 Å) (Figure 4) (Guainazzi et al., submitted for publication)[28]. Initially, NM ICLs were generated by treatment of dsDNA with nitrogen mustards, yielding only small amounts of ICL in a non-specific reaction. Additionally, NM ICLs are inherently unstable on a laboratory time scale (in the order of hours at room temperature) hindering detailed characterization [29-32]. One way to generate more stable NM ICLs has been to convert them to the stable ring-opened formamidopyrimidine (FAPY) form by treatment with aqueous base enabling initial studies of the repair of NM ICLs in bacteria [25,33-35], although this approach did not address the problem of selectivity and the final structure of the lesion is significantly different from the original NM ICL.

We recently developed a strategy to synthesize stable analogs of NM ICLs site-specifically with high specificity and in high yield [36](Guainazzi et al., submitted for publication). This approach involved the incorporation of acetaldehyde functionalities at the 7-positions of G residues in complementary oligonucleotides and use of a double reductive amination reaction to generate the ICL (Figure 3D). dG was substituted with 7-deaza-2'-deoxyguanosine to counteract the inherent lability of the glycosidic bond in N^7-alkylated guanines. This approach not only allowed for the synthesis of ICLs isosteric to the drug-induced adducts, but also of ICLs that induce distinct amount of bending of the DNA helix by variation of the length of the bridge linking the two bases on

complementary strands. As discussed in more detail below, evidence is increasing that a number of steps in ICL repair, in particular the initial recognition by cellular proteins and interaction with polymerases, is dependent upon the structure of the ICL.

Cisplatin ICLs

Cisplatin was serendipitously discovered as a compound with cytotoxic activity in 1965 [37,38]. It is still widely used and is particularly successful against testicular cancer. Based on the initial success, many analogous of cisplatin (e.g. carboplatin, oxaliplatin, satraplatin, picoplatin) have been generated and others are in development to improve pharmacological properties [39]. Cisplatin is a planar coordination complex of Pt(II) with two chloride ligands in cis-position. Once the molecule enters the cell, the chlorides are displaced by two molecules of water producing an activated form of cisplatin, which is capable of forming adducts with proteins, RNA and DNA [40]. Like nitrogen mustards, cisplatin preferentially reacts with the nucleophilic N^7 position of guanine, forming a monoadduct and subsequently forming intra- and interstrand crosslinks. Among these the 1,2-intrastrand crosslinks (65% at 5'-GG and 25% at 5'-AG) are formed most frequently while 1,3-intrastrand crosslinks (up to 5% at 5'-GNG) and 1,2-interstrand crosslink (up to 8% at 5'-GC) are also formed in significant amounts [40]. The cisplatin interstrand crosslink severely distorts the double helix, due to the shortened distance between two guanines on complementary strands (Figure 4). The cytosines opposite to the connected guanosines are flipped out into an extrahelical conformation and the DNA is considerably bent [41,42]. Although no fully synthetic approach to cisplatin ICLs has been reported, these adducts have been prepared relatively efficiently by hybridization directed crosslinking (Figure 3B). Activated cisplatin (cis-$[Pt(NH_3)_2(H_2O)Cl]$) is first incubated with an oligonucleotide containing a unique G in a 5'-d(GC) site to generate a monoadduct. Following purification, annealing to a complementary strand results in ICL formation [43]. We have employed this strategy in our laboratory for the

preparation of oligonucleotides with a single defined cisplatin ICLs that were inserted into plasmids [44] to study replication-dependent repair of ICLs (see below).

Zhu and Lippard synthesized a derivative of a cisplatin ICL in which one of the amines was conjugated to a photoreactive benzophenone moiety [45]. This photoreactive substrate was used to identify proteins that bind to the ICL in cells extracts revealing important differences in which proteins bind to inter- versus intrastrand crosslinks.

Chloro ethyl nitroso urea ICLs

Chloro ethyl nitroso ureas (CENUs), in particularly carmustine (BCNU) are widely used in cancer chemotherapy and are particularly useful for the treatment of brain tumors due their ability to pass the blood brain barrier [46]. The mechanism by which CENUs form ICLs is distinct from that of NMs and cisplatin. CENUs initial alkylate the O^6 position of dG and rapidly react again to form O^6-ethanoguanine, which slowly rearranges to an ICL formed between N^3 of dC and N^1 of dG [47]. Although CENUs ICL-containing oligonucleotides have been generated by treatment of oligonucleotides with BCNU [48], strategies for the efficient site-specific generation of this adducts on a larger scale have so far remained elusive. A number of BCNU-like adducts have been synthesize either by incorporation of a precursor (a chloroethyl modified thymine) that can form an ICL with a guanine residue [49] or by incorporation of crosslinked nucleotide dimer into DNA using solid-phase synthesis (Figure 3C) [50-52]. This later strategy allowed the synthesis of bridges with different lengths and different structure. Although these synthetic adducts are distinct from the CENU ICLs, some of them have started to yield insight into how the repair of ICL that are linked through their Watson-Crick pairing surfaces may differ from those that linked through either major or minor groves [53,54].

Mitomycin C: ICLs induced by a natural product

Mitomycin C (MMC) belongs to a class of antibiotics originally isolated from *Streptomycyces caespitosis* and it is widely used in chemotherapy against gastrointestinal tumours, gastric, pancreatic, biliary tract, colorectal, and anal cancer [55]. Mitomycin C only forms adducts with DNA after reduction of its quinone ring reacting mainly with exocyclic amines of dG. After rearrangement and alkylation of a second guanine it forms ICLs in a 5'-(GC) sequence in addition to intrastrand crosslinks and monoadducts [56,57]. MMC ICLs are formed in the minor groove, where they induce a moderate widening to accommodate the MMC heterocycle [58,59]. Since ICLs are formed relatively efficiently by treatment of a duplex with MMC, this agent has been used in cellular ICL repair studies as well as for some biochemical studies [60-62].

Psoralen: ICL formation can be induced by UV light

Psoralens are linear furocumarins isolated from plants and fungi and are used in the treatment of skin disease such as vitiligo or psoriasis. Psoralens intercalate into the DNA and can be photoactivated with UV-A radiation to form covalent adducts with thymidine bases, forming interstrand crosslinks (Figure 4) [63]. This ability to control the activity of psoralen by photoactivation is why this drug is specifically useful for the treatment of skin disease [64] and has made psoralen one of the most useful agents to generate ICL-containing oligonucleotides. Treatment of oligonucleotides containing a specific 5'-AT sequence with psoralen and UV can be controlled to produce mostly ICLs without excessive formation of byproducts such as monoadducts or intrastrand crosslinks. The efficiency of ICL formation has been additionally increased by the use of high intensity lasers to specifically produce first the monoadduct and then the ICL [65] and by the chemical synthesis and incorporation into DNA by solid-phase synthesis of the monoadduct followed by the UV irradiation to obtain the ICL [66]. Psoralen ICLs have been structurally characterized by NMR [67,68] and X-ray crystallography [69]

revealing that these ICLs locally constrain and distort the DNA at the site of the intercalation but do change the overall DNA structure (Figure 4).

Due to the relative ease of preparation and stability, psoralen ICLs have been the most frequently used substrate in investigations of ICL repair in bacteria, yeast and mammalian cells. More recently, Seidman and coworkers developed a method for the induction of spatially defined ICLs in living cells by introduction of psoralen conjugates into cells and generation of ICLs using UV laser irradiation at defined position in cell nuclei [70]. In this approach, psoralen was conjugated to a digoxigenin, allowing for the detection of the psoralen ICL by immunofluorescence. This approach allows the study of the recruitment of repair proteins to sites of ICLs [71] and provides a powerful new tool to study ICL repair.

ICLs formed by endogenous and environmental compounds

While most of the current interest in crosslinking agents stems from their use in cancer chemotherapy, the many endogenous and environmental ICL-forming agents must have been the driving force behind the evolution of the responses they trigger. One large group of agents that can form ICL includes bifunctional aldehydes that are formed in cellular metabolic processes such as lipid peroxidation [22]. A prototypical bifunctional electrophilic compound is malondialdehyde, which can crosslink DNA via two exocyclic guanine amino groups [72]. A number of additional aldehydes (formaldehyde and acetaldehyde) and α,β-unsaturated aldehydes (acrolein, crotonaldehyde), can be found in food, pesticides and tobacco smoke and can also form DNA ICLs [73,74]. One characteristic of aldehyde-induced ICLs is that they are intrinsically reversible, which can complicate biochemical and cellular studies. To circumvent this issue, Harris, Rizzo and coworkers devised a way to synthesize a trimethlylene ICL connecting two G residues in the minor groove, representing a reduced and stabilized form of aldehyde ICLs [75]. The structures of these ICLs in 5’-GC and 5’-CG sequence contexts were characterized by NMR [75,76]. The 5’-GC ICL is

readily accommodated in the minor groove and causes little distortion, while the 5'-CG destabilizes the duplex and induces a bend and a twist in the double helix (Figure 4).

ICLs can be formed from additional potentially endogenous sources. For example, nitric oxide has the ability to form ICLs through diazotization of the exocyclic amine of a G and subsequent ICL formation by reaction with an adjacent G on a complementary strand. This reaction yields an adduct in which two guanine residues are directly linked through a single amine group in the minor groove [77-79].

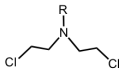
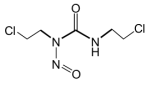
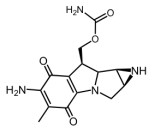
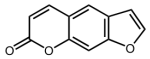
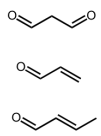
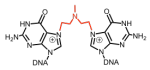
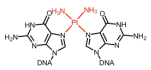
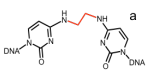
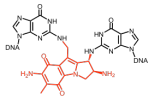
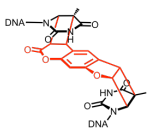
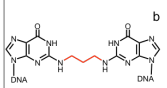
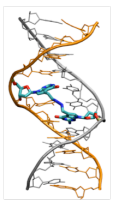
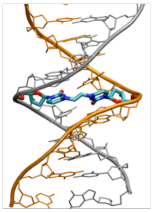
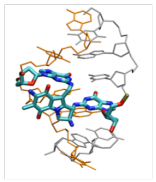
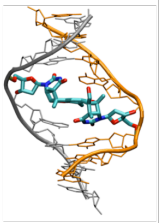
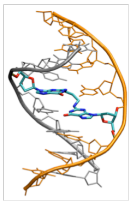
ICL-forming agent	Nitrogen Mustard	Cisplatin	Carmustine	Mitomycin C	Psoralen	Malondialdehyde
Target Sequence	5'-GNC	5'-GC	G-C basepair	5'-CG	5'-TA	5'-GC/5'-CG
Alkylating agent						
ICL structure[a]						
DNA structure						
PDB file	N/A[c]	1DDP	1N4B	Ref [50][d]	204D	1HZ2
Distortion	-15˚ bend -Minor local distortion	-Major distortion -Two Cs opposing the crosslinked Gs are extrahelical	Very minor local distortion[e]	Minor widening of MG	Minor local distortion	5'GC: Minor 5'CG: Major

Figure 4: Common ICL-forming agents and their DNA adducts. [a]) While native BCNU adducts have not yet been synthesized, various mimics have been synthesized and used to study repair. The structure of the ICL and the pdb file refer to one of these mimics. [b]) This ICL has been synthesized in two sequence contexts. This ICL is a reduced and stabilized form of the crosslinks formed by malondildehyde. [c]) The structure shown here is based on a molecular modeling study (AG, Campbell AJ, Angelov T, Simmerling C, ODS, submitted for publication). [d]) The coordinates for the mitomycin C ICL were kindly provided by Suse Broyde based on Ref [50]. [e]) The distortion refers to the mimic showed here. The exact distortion provoked by BCNU is not known and is likely to be different than the mimic.

Such NO-induced ICLs have been incorporated into oligonucleotides as dimers using solid-phase synthesis strategy [80,81]. Structural characterization of these ICLs by NMR revealed that although they do not bend the DNA they induce distortion to the double helix by flipping the cytosines paired with the crosslinked guanines into an extrahelical position in the minor groove [80].

Finally, the Greenberg laboratory has demonstrated that radicals such as the ones formed at the methyl group of thymidine can from ICLs, pointing to yet another potential endogenous source of ICLs [82,83].

ICLs pose a difficult problem for the DNA repair machinery

ICLs are uniquely complex lesions, as they need to be removed from both strands of DNA. An intact template for the regeneration of the original DNA sequence is therefore not available as it is for monoadducts that are repaired by base excision or nucleotide excision repair. It became apparent from the first studies of ICL repair in *E. Coli* by R. Cole [84,85] that this process requires the interplay of multiple pathways, including nucleotide excision repair (NER) and homologous recombination (HR). The situation is even more complex in eukaryotes, where additional layers of regulation are present and ICLs are repaired by distinct pathways in S and G0/G1 phases of the cell cycle. Thus in addition to NER and HR, repair and signaling pathways like Fanconi anemia (FA), mismatch repair (MMR) and translesion synthesis (TLS) have been implicated in ICL repair.

Replication forks stall at ICLs and trigger their repair

ICLs exhibit the highest levels of toxicity when they block strand separation and polymerase activity of the replication machinery and it is believed that the arrest of the replication fork is the trigger to initiate ICL repair [86]. Any details of replication-dependent repair have remained elusive until very recently, but several classes of

proteins have been shown to be involved in ICL repair based on the sensitivity of cell lines with deficiencies in the corresponding genes. Such studies implicated endonucleases, including ERCC1-XPF [87-89] and MUS81-EME1 [90,91], translesion synthesis polymerases, in particular Pol ζ and Rev1 [92-94] and proteins involved in homologous recombination, such as Rad54, XRCC2 and XRCC3 [95,96]. In vertebrates, cells deficient in one of the thirteen proteins associated with the cancer-prone inherited disorder Fanconi anemia (FA) pathway are specifically hypersensitive to ICL-forming agents. The mechanism by which FA pathway is involved in mediating and regulating ICL repair is slowly emerging [97-99].

A recent breakthrough in our understanding of ICL repair came from studies in cell-free *Xenopus* egg extracts, a biochemical system that supports efficient replication of plasmids [100]. To adapt this system to study ICL repair, short oligonucleotides containing site-specific cisplatin or nitrogen mustard-like ICL prepared as described above were ligated into plasmids [44]. Incubation of these ICL-containing plasmids with the extracts allows plasmid replication under physiological conditions. Upon initiation of replication, two forks approach the crosslink from opposite directions, initially stalling at ~20-40 nucleotides from the ICL (Figure 5). After a delay of ~20 minutes one of the leading strands is extended to within one nucleotide of the ICL. At this point, dual incision around the ICL of the lagging strand template leads to unhooking of the ICL, and translesion synthesis leads to bypass of the lesion and full extension of the nascent leading strand. Although it has not yet been experimentally demonstrated, the ICL-remnant may then be removed by NER and the unhooked strand reengaged in replication by homologous recombination in a way similar as for double-strand breaks formed during replication.

Regulation of ICL repair by the FA pathway

The various steps and factors involved in replication-dependent ICL repair will be discussed here in the framework provided by this study. The stalling of replication forks

can trigger many pathways and we will limit discussion here to the Fanconi Anemia pathway, which is most relevant to ICL repair. Indeed, the FA pathway is activated in S-Phase in particular after exposure to crosslinking agents and this activation is also fully recapitulated in the cell-free xenopus system. The FA pathway contains three components; the first one is the core complex, which contains eight FANC proteins (A, B, C, E, F, G, L, M) and at least five associated proteins (FAAP100, FAAP24, HES1, MHF1 and MHF2) [99,101,102]. Only two proteins of the FA core complex have known catalytic activities: FANCM is a DNA translocase and interacts tightly with MHF1-2 and FAAP24 [101-104]. It has the ability to remodel stalled replication fork structures and it is likely to play a role in loading the FA core complex at replication forks that are stalled at ICLs. FANCM also has a role in activating checkpoint signaling in the absence of the FA core complex [105].

The other protein with known catalytic activity is FANCL, which is a ubiquitin ligase [106,107] and its activity is required for the monoubiquitination of the FANCD2-FANCI proteins, the second component of the FA pathway [108,109]. Ubiquitination of FANCD2-FANCI leads to colocalization of the heterodimer with other DNA repair factors in chromatin and is essential for mediating cellular resistance to crosslinking agents. A matter of debate for a long time, the direct role of FANCD2-FANCI in ICL repair was recently demonstrated in cell-free xenopus system using a site specific cisplatin ICL [110]. In the absence of FANCD2-FANCI or in the presence of a mutant FANCD2 protein that can no longer be ubiquitinated (K562R), the replication fork was still able to approach the ICL to the -1 position, but both the unhooking and translesion synthesis steps were blocked. This result thus suggests that ubiquitinated FANCD2-FANCI complex has a direct role in recruiting nucleases and/or translesion synthesis polymerases to the sites of ICLs (see below).

The remaining FA proteins, FANCD1 (better known as BRCA2), FANCJ (a helicase, also known as BACH1 or BRIP1) and FANCN (also know as PALB2) act downstream of FANCD2-FANCI are believed to have a role in mediating recombination later in the pathway [99].

Recently, it has been suggested that the FA pathway may also function outside of replication, but the mechanisms by which this might occur are not yet clear [111,112].

Which endonucleases are involved in unhooking the ICL?

An important feature of ICL repair in eukaryotes is that double stranded breaks (DSBs) are induced at stalled replication forks [86,88,113]. They are probably induced in an unhooking step immediately before the replication machinery bypasses the lesion (Figure 5). Two structure-specific endonucleases involved in ICL repair are ERCC1-XPF and MUS81-EME1, since cells defective in these enzymes are hypersensitive to crosslinking agents [87,89,90,114]. Both of these enzymes cleave ss/dsDNA junction with 5' ssDNA overhangs and therefore have the wrong polarity to make the first cut to generate DSBs in ICL repair [115]. In line with this view, studies have shown that neither ERCC1-XPF nor MUS81-EME1 are absolutely required for the formation of DSBs in response crosslinking agents during S-phase as evidenced by the formation of γ-H2AX-foci, cellular markers of DSBs [89,114]. Although dissenting views have been voiced as well [90,116], it is most likely that the these two nucleases are instead involved in mediating the second incision in ICL repair [88,117,118] or perhaps also resolving intermediates further downstream in the pathway during the homologous recombination step [119]. Both proteins seem to be recruited to the site of ICL by interacting with the SLX4 protein, which can also form a Holliday junction resolvase with another nuclease SLX1. Knock-down of SLX4 renders cells sensitive to crosslinking agents and SLX4 may therefore have a special role in coordinating nuclease activities at ICL [120-123]. Since SLX4 also contains a ubiquitin-binding (UBZ) domain this may happen by interaction with a ubiquitinated protein such as FANCD2-FANCI.

Despite the implication of a number of nucleases in unhooking ICLs, it is likely that the protein(s) that makes the first incision on the lagging strand has not yet been identified.

Translesion synthesis restores the leading strand during ICL repair

Translesion synthesis is an essential step in replication-dependent (and replication-independent, see below) repair as it restores one of the two strands affected by the ICL (Figure 5). Human cells contain at least 15 polymerases, many of which have the ability to bypass various types of DNA damage [124,125]. Among these, there is a striking sensitivity of cells deficient in Polζ (Rev3-Rev7) and Rev1 to crosslinking agents [94]. Chicken DT40 cells deficient in Rev3 are in fact more sensitive than any other cell line to crosslinking agents pointing to a key role of this polymerase in ICL repair. The polymerase activity of Rev1 is limited to the insertion of a single dCTP opposite a damaged base while Polζ is generally considered to be a good extender polymerase and the two polymerases often act in concert [126]. The kinetics of ICL repair in the xenopus egg extracts system suggest that one polymerase may insert a dNTP opposite the crosslinked nucleotide, while a second polymerase may responsible for the extension past the ICL [44]. Depletion of Rev7 from the xenopus extracts leads to an additional stalling of the replication fork at the 0 position at a cisplatin ICLs, after insertion of a dNTP opposite the ICL, strongly suggesting that Polζ is responsible for the extension step past the ICL.

While these observations cumulatively suggest that Rev1 and Polζ are the key players for the TLS step in ICL repair, there is increasing evidence that additional polymerase take part in this process. For example, while the bypass of the helix-distorting cisplatin ICLs was dependent on Rev7 in xenopus egg extracts, a nitrogen-mustard like ICL that does not distort the DNA helix was bypassed with comparable efficiency in the presence and absence of Rev7. This indicates that other polymerases are able to bypass this ICL. Indeed, other TLS polymerases such as Polη, Polκ or Polι are able to bypass such non-distorting ICLs in vitro under certain conditions (Ho, TV, AG and ODS, manuscript in preparation). Lloyd and coworkers have studied the ability of minor groove trimethylene ICLs (Figure 2) or non-distorting major groove ICLs. They found that these minor groove ICLs were readily and specifically bypassed by Polκ [74], while the

major groove ICL were specifically bypassed by Polν [127]. The studies suggest that the bypass of ICLs by polymerases is highly dependent of the structure of the ICL and that a number of polymerases aside from Rev1 and Polζ may contribute to the process.

The completion of replication-dependent ICL repair involves homologous recombination and possibly nucleotide excision repair. The unhooked crosslink now resembles a "common" NER substrate, affecting only one strand, however genetic experiments showed only mild sensitivity to cells defective in NER genes other than ERCC1-XPF. On the other hand loss of HR proteins showed substantially increased sensitivity to crosslinking agents. The role of HR in ICL repair has been recently reviewed by Hinz [128].

Replication-independent ICL repair

While most evidence points to replication-dependent repair being the main mechanism of dealing with ICLs, there is now robust support for ICL repair mechanisms in G0/G1 phases of the cell cycle (reviewed in [129]). The contribution of replication-independent pathways to ICL repair has not been fully appreciated, primarily because the extreme toxicity of ICLs is most apparent when a replication fork is blocked. Therefore, cells with deficiencies in genes that contribute to ICL repair during S phase display higher toxicity to exposure to crosslinking agents.

Conclusive evidence for a replication-independent pathway stems from several experimental systems. Genetic studies in the yeast *S. cerevisiae* have revealed that NER and TLS pathways are essential for the repair of nitrogen mustard ICLs in the G0/G1 phase of the cell cycle [130]. The use of reporter assays has suggested similar pathways in mammalian cells [62,131,132]. In these studies oligonucleotides containing site-specific ICLs were ligated into the promoter region of a gene of a plasmid. Activation of gene expression required the repair of the ICLs in mammalian cells.

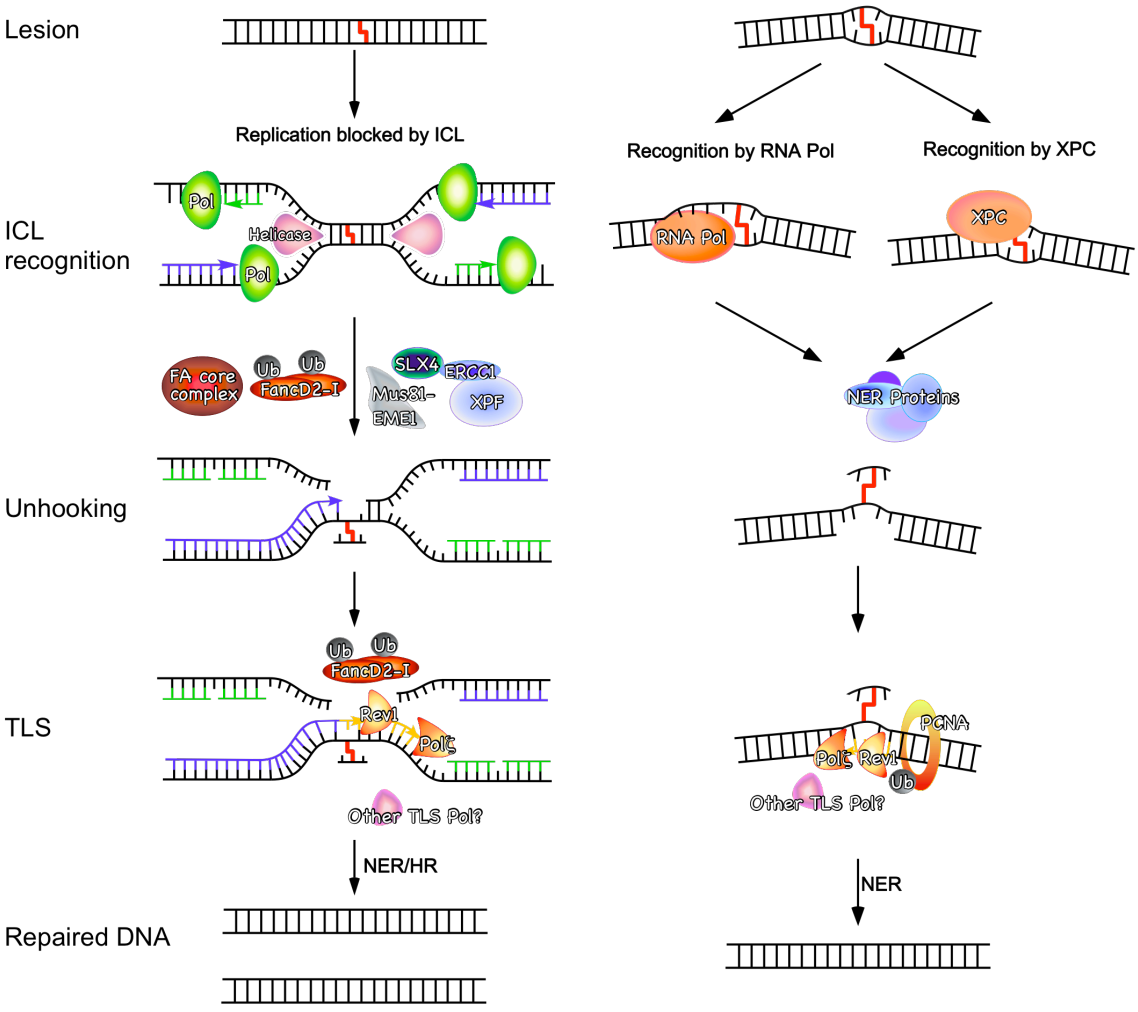


Figure 5: Models for replication-dependent (left) and -independent (right) ICL repair pathways. In the replication-dependent pathway the lesion is detected when replication forks are blocked by the ICL. The replication fork initially pauses 20-40 nucleotides from the ICLs, and then approaches to the ICL, activating the FA pathway. The FA core complex ubiquitinates the FANCD2-FANCI complex, a step required for the endonucleases to incise the lagging strand on both sides of the ICL and REV1 and Polz to extend the leading strand past the ICL. It is believed that HR and possibly NER then complete the process and restore the intact DNA seqeunce.
In G0/G1, the ICL can also be recognized by an RNA polymerase during transcription or, depending on its structure by the NER damage recognition factor XPC-RAD23B. NER proteins are thought to be responsible to unhook the ICL initiating repair synthesis and carrying out TLS past the unhooked ICL in a manner dependent on ubiquitinated PCNA and a TLS polymerase (probably Rev1 and Pol ζ) to bypass the lesion in an potentially error-prone manner. Finally, it is believed that the NER machinery eliminates the ICL remnant restoring the intact DNA.

These studies showed that, consistently with the studies in *S. cerevisiae*, ICLs were repaired by a pathway involving NER and TLS. A number of ICLs were repaired in this way, including those formed by psoralen [131], mitomycin C [62] N^4C-N^4C and N^3T-N^3T alkyl ICLs [133] and cisplatin (Enoiu M and ODS, unpublished observations). As it has been shown for replication-dependent ICL repair, the TLS polymerases Rev1 and Polζ are of particular importance for the bypass of ICLs [132], in line with the observation that deletions in Rev1 and Polζ render DT40 cells more sensitive to crosslinking agents than deletions in any other gene [134]. The TLS polymerases are recruited to bypass the ICLs in G0/G1 in dependence of ubiquitination of PCNA by Rad18 [130,132], similar to TLS of single-stranded lesions.

Interestingly, the recognition of some ICLs (psoralen, MMC) occurs by global genome NER (GG-NER) involving XPC-RAD23B [62,131], while the repair of others (alkyl ICLs, cisplatin) was found to be dependent on transcription-coupled NER (TC-NER) [133](Enoiu M and ODS, unpublished observations). The efficiency of NER of lesions affecting one strand of DNA is believed to be roughly proportional to amount of thermodynamic destabilization induced by a lesion [135]. By contrast, it is believed that all bulky adducts can stall an RNA polymerase and trigger TC-NER (in the transcribed strands of active genes) [136]. An example of a lesion that is addressed mainly by TC-NER is the cyclopyrimidine dimer (CPD), which induces only minor distortion in the DNA [62,135].

Although NER clearly plays a role in repairing ICLs, our understanding of how these lesions interact with NER proteins is limited. Although it has been shown that XPC-RAD23B is likely to be the first NER protein to arrive at psoralen ICLs in living cells [137], very little is known about how various ICLs are recognized by NER proteins. It is intriguing that cisplatin ICLs, which are highly helix-distorting [41], are not repaired by GG-NER [138], while psoralen or MMC ICLs, which are less distorting are processed by GG-NER [62,131,137]. It therefore appears that the rules of damage recognition in NER and binding by XPC-RAD23B as they have been defined for lesions on one strand of DNA do not simply apply for ICLs [139-141]. It has been suggested that the mismatch repair

proteins MutSβ (MSH2-MSH3) may play a role in facilitating the recognition of ICLs by NER and perhaps also independently of NER [142,143]. The structural features that trigger NER of ICL remain therefore to be determined.

In vitro studies of processing of ICLs by NER have revealed additional features that differ from canonical repair. NER proteins have been shown to make two incisions on one side of a psoralen or N^4C-ethyl-N^4C ICL, resulting in a gap 5' to an ICL [53,144,145]. Such a gap could in principle facilitate TLS past the ICL, perhaps in conjunction with an incision 3' to the lesion mediated by a nuclease that is not involved in NER, but the physiological relevance of this observation remains to be established.

Toward the discovery of clinically useful inhibitors of ICL repair

As it has been shown that the repair of ICLs in tumor cells leads to resistance to treatment with crosslinking agents such as cisplatin or nitrogen mustards [146,147], the inhibition of these repair pathways is an important goal for anti-tumor therapy. The recent demonstration that inhibition of ssDNA break repair protein PARP in BRCA2-deficient tumors is synthetically lethal demonstrates that DNA repair inhibitors can be effective even in the absence of DNA damaging agents [148-150].

As discussed above, the repair of ICLs is a complex affair and involves players from many different DNA repair pathways. To ensure specific and efficient inhibitory effects in conjunction with the treatment of crosslinking agents it would be desirable to target those pathways that are specifically responsible for the repair of ICLs. With the advances of our understanding of ICL repair, proteins and pathways that may be targeted to inhibit ICL repair are emerging.

So far, the FA pathway has received most of the attention as a target due to the specific sensitivity of FA-deficient cell to crosslinking agents. Indeed, preclinical model studies in mice have revealed that disruption of FA core complex proteins FANCC and FANCG in adenocarcinoma cell lines abrogates FANCD2 monoubiquitylation rendering cells sensitive to agents such oxaliplatin and melphalan [151,152]. To date, screens for

inhibitors of the FA pathway have mostly focused on the most clearly discernable step, the ubiquitination of FANCD2-FANCI and the concomitant translocation of these two proteins into cellular foci and chromatin. D'Andrea and coworkers used a cellular screen to identify compounds that inhibit the translocation of a GFP-labeled FANCD2 to nuclear foci following genotoxic treatment [153]. The Hoatlin laboratory used *Xenopus laevis* cell-free extracts to screen for small molecules that inhibit FANCD2 ubiquitination upon incubation with different DNA molecules [154]. These two studies identified inhibitors of the FA pathway, including the natural product curcumin and derivatives thereof, that render cells sensitive to treatment with crosslinking agents. Gallmeier et al. carried out a different type of screen to identify compounds that are selectively toxic to FA-deficient cells, yielding a number of interesting lead compounds [152]. These studies raise the possibility that FA inhibitors with clinically useful properties may be found, both as a chemosensitizer in cisplatin-based cancer treatment and as synthetic-lethal agents in tumors with defects in the FA genes. Further development of FA pathway inhibitors will require a more detailed understanding of the mechanisms by which the identified compounds operate. Furthermore, emerging structural studies of FA proteins [107,155] and a better understanding of the mechanisms by which FA proteins such as FANCM contribute to ICL repair will provide additional opportunities for inhibitors that target FA proteins.

As discussed above another group of proteins that have ICL repair-specific functions are the endonucleases that make incision around ICLs during replication-dependent repair. Although proteins like ERCC1-XPF have a number of functions, recent studies suggest that this protein is recruited to sites of ICL repair through specific interaction with partners such as SLX4 [120-123]. Importantly, cells expressing mutant ERCC1-XPF with a specific defect in NER are not sensitive to crosslinking agents such as cisplatin or MMC [156], suggesting that the identification and characterization of interaction sites that recruit nucleases to carry out their function in ICL repair may be fruitful targets for therapeutic intervention.

Concluding Remarks

The repair of ICLs is accomplished by an assembly of many components from different repair pathways. Our understanding of the intricate network of interaction between signaling and repair pathways triggered by ICLs is still far from complete. One development that has moved the field forward has been the development of new methods to synthesize oligonucleotides containing site-specific ICLs and their used to study ICL repair in plasmid-based assays in cell-free extract and living cells. Together with cell biological and genetic approaches this will aid the understanding of molecular basis of how ICL repair differs from repair pathways of damage that only affect one strand of DNA. These insights should provide opportunities to find new targets for drug developments to increase the therapeutic efficiency of crosslinking agents and to target tumor cells with specific defects in ICL repair.

Acknowledgements

We thank Suse Broyde (NYU) for providing the coordinates for the mitomycin C ICL.

Chapter 2

Generation of DNA Interstrand Crosslinks by Post-Synthetic Reductive Amination

Foreword

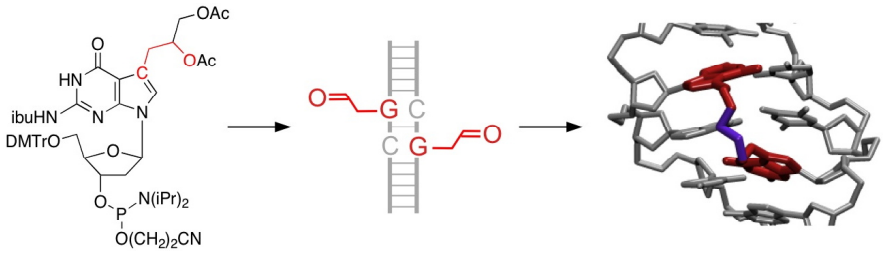
This chapter will describe the strategy and the synthesis of different DNA interstrand crosslinks that mimic the one formed by the clinically relevant Nitrogen Mustard alkylating agent.

The work presented in this chapter is the base for the rationale behind my dissertation project. The main idea and the strategy were developed by Orlando and Todor and the synthesis that led to the results presented here were developed and perfectioned by Todor in what it resulted to be his PhD thesis. Without this fertile scientific background none of my experiments and results would have been possible. I partially contributed to refine the synthesis of the Nitrogen Mustards Interstrand Crosslink described in this chapter and I use this procedure for the synthesis of the crosslink described in chapter 3. In addition I used these protocols for preparing and providing different substrates to collaborators and people in our lab to perform their experiments and possibly expand our knowledge of the DNA interstrand crosslink repair mechanism, which was actually the inspiration that brought us to work on this project.

Generation of DNA Interstrand Crosslinks by Post-Synthetic Reductive Amination

Adapted from the manuscript by Todor Angelov, Angelo Guainazzi and Orlando D. Schärer, published in “Organic Letters” in 2009, volume 11, issue 3, pages 661-664.

ABSTRACT



DNA interstrand crosslinks (ICLs) are the clinically most relevant adducts formed by many anti-tumor agents. To facilitate the study of biological responses triggered by ICLs, we developed a new approach toward the synthesis of mimics of nitrogen mustard ICLs. 7-Deazaguanine residues bearing acetaldehyde groups were incorporated into complementary strands of DNA and crosslink formation induced by double reductive amination. Our strategy enables the synthesis of major groove crosslinks in high yields and purity.

Introduction

A critical setback in the study of ICL repair has been the limited availability of defined ICL adducts [19,157]. ICLs were initially synthesized by the reaction of DNA with crosslinking agents followed by isolation and purification of the ICL [25-27,48,158]. This approach yields mixtures of products (mono adducts, *intra*- and *inter*strand crosslinks) and the desired ICLs usually make up only a small fraction (typically 1-5%) of all the products formed. Two other more efficient approaches to the chemical synthesis of ICLs have subsequently been developed. The first one was based on the crosslinking of nucleosides outside of DNA followed by incorporation of the crosslinked dimer into DNA using solid-phase synthesis [50-52,81,159]. This approach has been particularly successful in yielding ICLs that connect two bases through their Watson-Crick faces, although the solid-phase synthesis procedures are complicated by the connected nucleoside dimers. A second approach consists of the site-specific incorporation of post-synthetically modifiable crosslink precursors on one or two opposing strands of DNA, annealing of the two single strands and subsequent use of a specific coupling reaction to generate the ICL. This concept has been used in the synthesis of ICLs containing disulfide, psoralen, malondialdehyde and alkyl linkages [66,82,160-163]. With the exception of disulfide ICLs, there has been a lack of efficient synhtheses of ICLs formed in the major groove of DNA, where adducts by the clinically most important ICL-forming agents including the nitrogen mustards (NMs) are formed.

NMs preferentially form ICLs between the N7 positions of dG residues in a 5'-GNC sequence context and introduce a slight distortion into the DNA [25-27,31,164,165]. Following the initial pioneering studies of NM ICLs by the Loechler and Hopkins groups no further efforts toward high yielding synthesis of these adducts have been reported [25,27].

Synthesis of oligonucleotides containing a precursor aldehyde

We designed a strategy for the synthesis of NM ICL mimics **2** based on the incorporation of ICL precursor nucleosides on opposing strands of DNA and the use of a subsequent specific coupling reaction to establish the ICL (Figure 1).

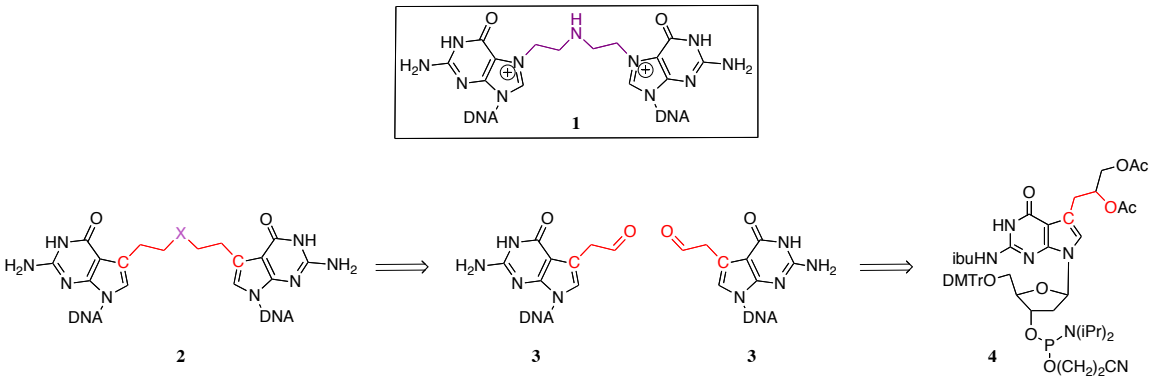
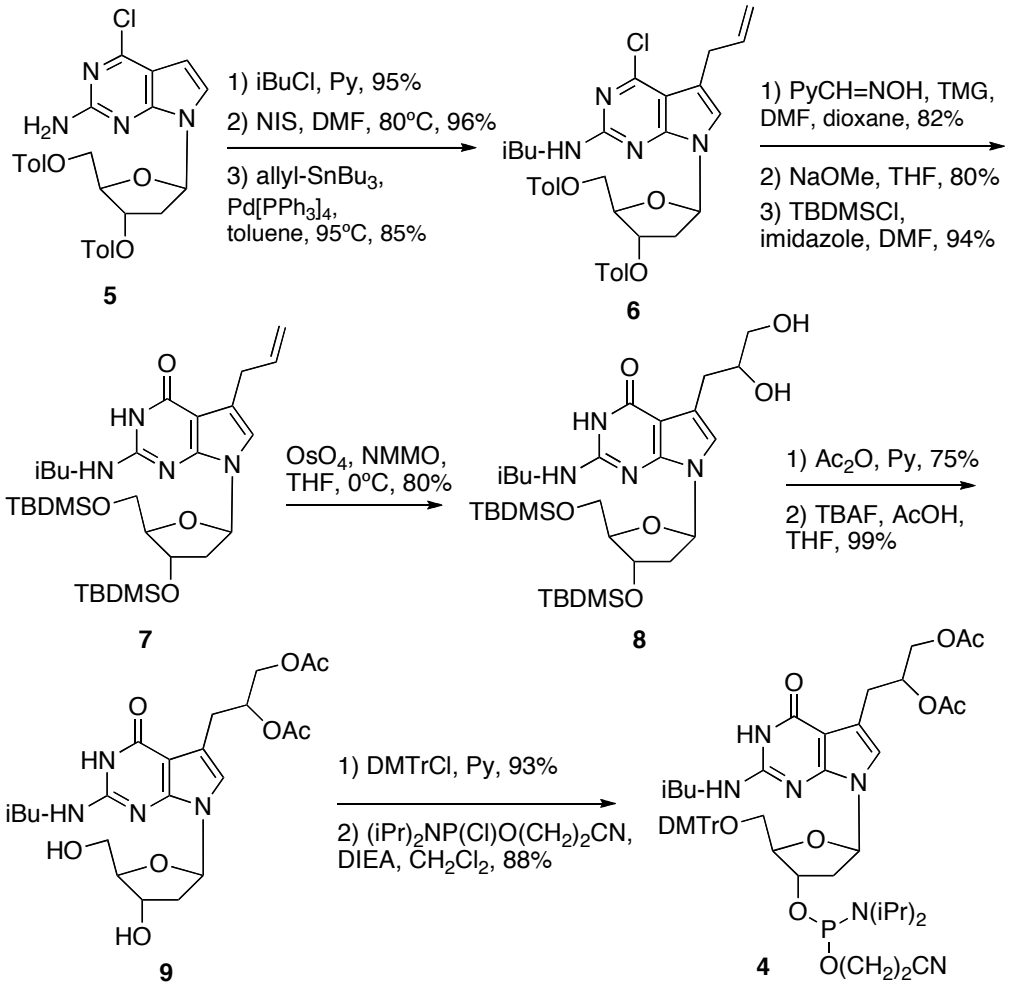


Figure 1: Structures of NM ICLs 1 and strategy for the synthesis of NM ICL analog 2. NM ICLs (1) connect two complementary DNA strands by connecting two guanine bases through the N(7) positions. Our target NM ICL mimic 2 has the nitrogen at the 7 positions replaced with carbon to render the glycosidic bond of the ICL stable for synthesis and functional studies (X = amine-containing compound). We envisioned that ICL 2 could be obtained via double reductive amination of an aldehyde 3 with an appropriate amine, XH. 3 in turn could be derived from phosphoramidite 4, where the aldehyde is masked as a protected diol.

We reasoned that alkylamine-containing crosslinks may be accessed by a double amination reaction from two aldehyde groups (**3**) on complementary DNA strands using an appropriate amine. Since guanine bases alkylated at the 7 position in the native NM ICL **1** are prone to depurination due to the positive charge on nitrogen, we decided to pursue the synthesis of the more stable, isosteric 7-deaza analogs **2**. The increased hydrolytic stability of the 7-deaza compounds should make it possible to incorporate them into DNA using solid-phase synthesis and make them attractive substrates for biological studies. We envisioned that the aldehyde would be introduced into DNA masked as a protected diol **4** [166], using standard phosphoramidite chemistry (Figure 1). The synthesis of **4** (Scheme 1) started with 6-chloro-7-deaza-3',5'-di-O-p-toluoyl-2'-deoxyguanosine **5**, which was protected as an isobutyric ester at the N(2) position and selectively iodinated at C(7)[167]. A Stille coupling reaction was then used to introduce the allyl group in **6**.



Scheme 1: Synthesis of the ICL precursor phosphoramidite 4

Treatment with pyrimidine-2-carboxaldoxyme restored the deazaguanine core and the toluoyl protecting groups in the carbohydrate moiety were replaced with TBDMS. The allyl group of **7** was then oxidized to diol **8** using osmium tetroxide and the newly generated hydroxyl groups were protected as acetate esters. Finally, the TBDMS protective groups were removed and the sugar moiety was elaborated to the phosphoramidite **4** using standard procedures.

Generation of the Interstrand Crosslink

Using solid phase synthesis, **4** was incorporated into two complementary DNA strands as a part of a 5'-d(GNC) sequence (**10a-d**, Figure 2), which has been shown to be the preferred site for NM ICL formation [26].

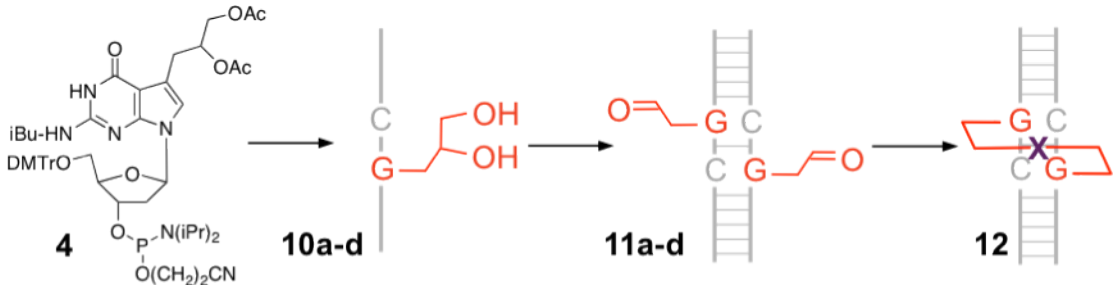


Figure 2: Phosphoramidite **4** was incorporated into complementary strands of DNA and deprotected (10a-d). Following annealing and oxidization with $NaIO_4$ to aldehydes **11a-d**, reaction with ammonium choride, methylamine, hydrazine, ethylene diamine or dimethylethylene diamine followed by reduction with sodium cyanoborohydride was attempted to generate ICLs

The two single-stranded oligonucleotides were purified by solid phase extraction using TOP-cartridges (Varian) and the incorporation and integrity of the diol was verified by ESI-MS (Table 1 in the Experimental Details section). Exploratory experiments revealed that the diol-containing DNA could be oxidized to the aldehyde and derivatized with an aldehyde-reactive semicarbazide fluorescent dye (data not shown), validating our approach for the generation of the aldehyde precursor.

With the aldehyde containing single-stranded oligonucleotides in hand, we investigated the use of these building blocks in ICL formation. To circumvent the need for lengthy manipulations of the potentially unstable aldehyde functionalities, the two diol-containing strands were first annealed and subsequently oxidized with sodium periodate. Excess oxidizing reagent was quenched by addition of sodium sulfite and the aldehyde-containing oligonucleotides were treated with a variety of amines and $NaBH_3CN$. Unexpectedly, incubation with NH_4OAc or methylamine did not lead to any detectable formation of ICLs **12a** or **12b**, as only a band corresponding to the 20-mer

single stranded oligonucleotide substrates was visible on a denaturing polyacrylamide gel (Figure 3B, lanes 2 and 3).

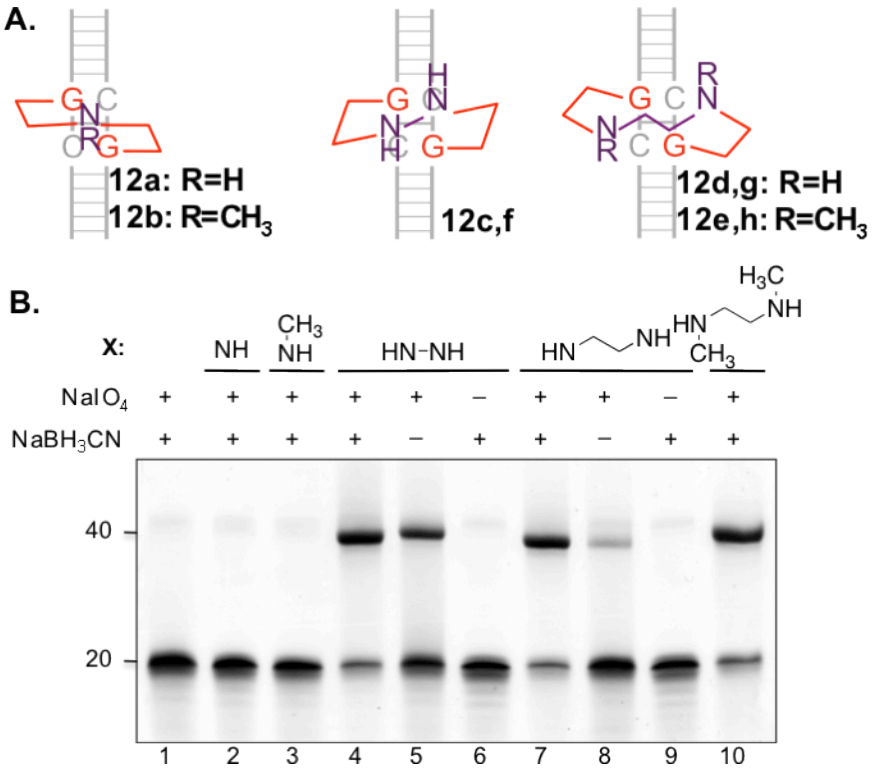


Figure 3: A. Structures of the target ICLs **12a-i**. **B.** Denaturing PAGE analysis of the reductive amination reaction of the two 20-mer aldheyde containing oliconucleotides. The amine group X resulting from the reaction, as well as presence or absence of $NaIO_4$ and $NaBH_3CN$ in the reaction mixture are indicated above the gel. The positions of marker 20-mer (for ssDNA) and 40-mer (for ICL containing DNA) oligomers are indicated. The gel was stained with methylene blue.

By contrast, treatment with hydrazine, ethylenediamine or N,N'-dimethylethylenediamine resulted in the formation of a band with mobility corresponding to a size roughly double that of the starting 20-mers, indicating that efficient formation of ICLs **12c**, **12d** and **12e** had occurred (Figure 3A, B, lanes 4, 7 and 10). Interestingly, while the formation of ICLs by the ethylenediamines was dependent on the presence of $NaBH_3CN$, hydrazine ICLs were also formed in the absence of any reducing agent, which could be explained by the formations of a stable hydrazone adduct (Figure 3B, compare lanes 5 and 8). Treatment with formaldehyde readly reversed the nonreduced form of hydrazine ICL, but not the reduced form of **12c** (data

not shown), indicating that the hydrazone linkage was indeed reduced by $NaBH_3CN$. The formation of an ICL was dependent on the presence of aldehyde reactive groups on both strands of DNA, as no slower migrating band was formed if the aldehyde functionalized guanine residues were absent or present only on one of the two strands (data not shown). This observation eliminated the possibility that a hydrazone mono-adduct would react with an unmodified nucleotide in the complementary strand. For further characterization, the ICL-containing oligonucleotides were purified by reverse-phase HPLC and subjected to MS and nucleoside composition analysis. ESI-MS analysis of the crosslinked 20-mer oligonucleotides revealed good agreement between the calculated and measured mass, but was only accurate enough in the case of the dimethylethylenediamine ICL to ascertain that the ICL was present with two reduced amine bonds (Table 1 in the Experimental Details Section). To obtain more precise mass determinations, we repeated the synthesis to obtain the shorter 11-mer ICL-containing oligonucleotides **12f-h**. In this case accurate molecular weight could be determined for the hydrazine, ethylenediamine and dimethylethylenediamine oligonucleotides (Table 1 in the Experimental Details Section) confirming that ICL formation and reduction of the imine and hydrazone linkages took place in all three cases. We further attempted to analyze the ICL-containing oligonucleotides by nucleoside composition analysis and incubated them with phosphodiesterase I, exonuclease III and calf intestine phosphatase, conditions we had previously employed in the analysis of ICLs [163]. HPLC of the digestion of the dimethylethylenediamine ICL **12e** yielded the expected peaks for the four native nucleosides dA, dC, dG and T and an additional slower eluting peak (Figure 5 in the Experimental Details Section). This peak was identified by MS as the expected crosslinked nucleoside monomer. Unexpectedly, digestion of the hydrazine or ethylenediamine ICLs did not result in the formation of a peak for the crosslinked monomers, as we observed either incomplete digestion or decomposition at the peak in the area of the dimethyl ethlylenediamine dimer (data not shown). The reason for the decomposition of the hydrazine-ICL and ethylenediamine ICL monomers is presently unknown, but the digestion of the dimethyl ethlylenediamine ICL and the MS analysis of

all the ICL-containing oligonucleotides confirms that the ICLs were formed with the expected structures **12c-h**.

Discussion

The lack of ICL formation in the reductive amination reaction with ammonia and methylamine deserves some comment. It has previously been shown that since the distance between the N7-dG sites in 5'-d(GNC) sequences (~8.9Å) is longer than the NM linkage (~7.5Å) the NM ICLs must induce a bend into DNA (Figure 4)[31].

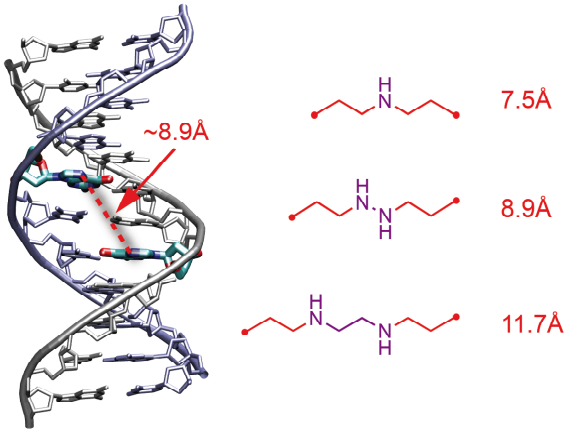


Figure 4: Structure of B-from DNA highlighting the distance linked by a NM ICL. The distance between the two crosslinked N7 atoms of guanine is marked by a red dotted line. The nominal distance of an amine, hydrazine and ethylene diamine ICLs are indicated in the right panel.

In our case, the intrinsically reversible initial imine formation step of the reductive amination reaction apparently does not provide enough strength to lead to the formation of ICLs **12a** and **12b** with these two amines. To exclude that the inability to form ICL **12a** was due to reductive amination of both aldehyde groups with ammonia prior to ICL formation, we also attempted to generate ICL **12a** by reaction between a 7-(2-oxoethyl)-7-deazaguanine and a 7-(2-aminoethyl)-7-deazaguanine residue on opposing strands in a d(GNC) sequence. This approach also did not lead to formation of ICL **12a** (T.A. and O.D.S., data not shown), suggesting that indeed distance and reactivity constraints were responsible for the inability to complete the reductive amination with

ammonia and methylamine. The formation of hydrazine ICLs (**12c**) and ethylenediamine ICLs (**12d**, **12e**) on the other hand would not require the introduction of a bend in the DNA due to the longer linkage (8.9Å for hydrazine, 11.7Å for ethylenediamine), thereby facilitating ICL formation. An open question is how much the higher reactivity of hydrazine compared to an amine contributes to successful ICL formation in the case of **12c**. Upon reaction with the aldehyde, hydrazine forms a hydrazone, which is much more stable than an imine and more reactive toward the aldehyde in the second reaction with. Our observations that the reaction of the dialdehyde can lead to efficient ICL formation with hydrazine in the absence of a reducing agent supports this notion (Figure 3B, lane 5).

In conclusion, we have developed synthetic methodology that allows for the preparation of large amounts of pure oligonucleotides containing site-specific ICLs in the major groove using a double reductive amination reaction of aldehyde functionalities in two complementary DNA strands. These defined ICLs have already been proven to be valuable tools for the study of ICL repair mechanisms [168].

Experimental Details

General Information

6-Chloro-7-deaza-2'-deoxy-3',5'-di-O-p-toluoyl-guanosine (**5**) was prepared using published procedures [169-171]. 4,4'-Dimethoxytrityl chloride was purchased from SynGen (USA). Protected 2'-deoxyribonucleoside-3'-phosphoramidites and all other reagents necessary for automated DNA synthesis, except ethylthiotetrazole, were purchased from Glen Research (USA). All other compounds and solvents were purchased from Sigma-Aldrich (Switzerland). Dry solvents (acetonitrile, dichloromethane, dioxane, DMF, pyridine, THF and toluene) were received and stored over 4Å molecular sieves in crown-capped bottles. NMR Spectra were recorded on a Brucker ARX-300 Mhz spectrometer. HR-MALDI MS of nucleosides were recorded on an Ionspec FT MS Ultima spectrometer. ESI-MS of oligonucleotides were recorded on a Perkin-Elmer/Sciex API III spectrometer.

Synthesis of Phosphoramidite 4

6-Chloro-7-deaza-N(2)-isobutyryl-3',5'-di-O-p-toluoyl-2'-deoxyguanosine

To a solution of 6-chloro-7-deaza-2'-deoxy-3',5'-di-O-p-toluoyl-guanosine (**5**) (8.19 g, 15.7 mmol) in pyridine (40 ml) was added isobutyryl chloride (1.83 ml, 17.3 mmol) at 0°C. The reaction mixture was stirred for 1 h at room temperature, quenched with H_2O (60 µl, 3.33 mmol) at 0°C and evaporated to dryness under reduced pressure. The solid residue was resuspended in ethyl acetate (250 ml) and washed with 5% HCl (100 ml), sat. $NaHCO_3$ (100 ml) and brine (100 ml), dried over Na_2SO_4 and evaporated to dryness under reduced pressure. Purification by silica gel column chromatography (dichloromethane/ethyl acetate 10:1) gave the product (8.79 g, 14.9 mmol, 95%) as a white solid.

Rf = 0.47 (hexane:EtOAc 2:1), 0.53 (CH_2Cl_2:EtOAc 10:1).

^{1}H NMR ($CDCl_3$): δ 8.04 (s, 1H, C(2)-NH), 7.98 (dt, J = 8.3, 1.7 Hz, 2H, Tol-H), 7.90 (dt, J = 8.3, 1.7 Hz, 2H, Tol-H), 7.28 (d, J = 8.7 Hz, 2H, Tol-H), 7.27 (d, J = 3.7 Hz, 1H, C(8)-H), 7.22 (d, J = 8.0 Hz, 2H, Tol-H), 6.69 (dd, J = 8.1, 6.0 Hz, 1H, C(1')-H), 6.51 (d, J = 3.8 Hz, 1H, C(7)-H), 5.78 (dt, J = 6.3, 2.3 Hz, 1H, C(3')-H), 4.76 (dd, J = 11.5, 3.8 Hz, 1H, C(5')-H), 4.63 (dd, J = 11.5, 4.4 Hz, 1H, C(5')-H), 4.59 (ddd, J = 4.2, 4.1, 2.7 Hz, 1H, C(4')-H), 3.06–2.92 (m, 1H, iBu-CH) 2.97 (ddd, J = 14.3, 8.1, 6.6Hz, 1H, C(2')-H), 2.76 (ddd, J = 14.3, 6.0, 2.4Hz, 1H, C(2')-H), 2.44 (s, 3H, Tol-CH_3), 2.41 (s, 3H, Tol-CH_3), 1.28 (d, J = 6.9 Hz, 6H, iBu-CH_3).

^{13}C NMR ($CDCl_3$): δ 175.6 (C), 166.1(C), 165.9 (C), 152.3 (C), 151.9 (C), 151.2 (C), 144.3 (C), 144.0 (C), 129.7 (CH), 129.6 (CH), 129.2 (CH), 129.1 (CH), 126.7 (CH), 126.5 (CH), 125.3 (CH), 114.5 (C), 101.2 (CH), 84.6 (CH), 82.3 (CH), 75.0 (CH), 64.1 (CH_2), 37.7 (CH_2), 35.8 (CH), 21.7 (CH_3), 21.6 (CH_3), 19.20 (CH_3), 19.18 (CH_3).

HR-MALDI: m/z calculated for $C_{31}H_{31}ClN_4O_6.Na^+$ 613.1830, found 613.1824.

6-Chloro-7-deaza-7-iodo-N(2)-isobutyryl-3',5'-di-O-p-toluoyl-2'-deoxyguanosine

A solution of 6-chloro-7-deaza-N(2)-isobutyryl--3',5'-di-O-p-toluoyl-2'-deoxy guanosine (6.5 g, 11.0 mmol) and N-iodo-succinimide (17.32 g. 77.0 mmol) in DMF (310 ml) was stirred at 80°C for 8 h under argon atmosphere. The reaction mixture was diluted with ethyl acetate (600 ml) and washed with sat. $NaHCO_3$ (600 ml), sat. $Na_2S_2O_3$ (100 ml) and brine (200 ml), dried over Na_2SO_4 and evaporated to dryness under reduced pressure. Purification by silica gel column chromatography (dichloromethane/ethyl acetate 10:1) gave the product (7.55 g, 10.5 mmol, 96%) as a white solid.

Rf = 0.67 (CH_2Cl_2:EtOAc 10:1).

^{1}H NMR ($CDCl_3$): δ 8.14 (s, 1H, C(2)-NH), 7.96 (dt, J = 8.2, 1.5 Hz, 2H, Tol-H), 7.90 (dt, J = 8.2, 1.5 Hz, 2H,

Tol-H), 7.41 (s, 1H, C(8)-H), 7.27 (d, J = 7.9 Hz, 2H, Tol-H), 7.24 (d, J = 7.9 Hz, 2H, Tol-H), 6.67 (dd, J = 7.8, 6.2Hz, 1H, C(1')-H), 5.76 (dt, J = 6.3, 2.5Hz, 1H, C(3')-H), 4.77 (dd, J = 12.0, 3.9 Hz, 1H, C(5')-H), 4.65 (dd, J = 12.0, 3.9 Hz, 1H, C(5')-H), 4.59 (dd, J = 6.5, 3.8 Hz, 1H, C(4')-H), 2.98 (sept, J = 6.7 Hz, 1H, iBu-H), 2.88 (ddd, J = 14.4, 7.7, 6.6 Hz, 1H, C(2')-H), 2.77 (ddd, J = 14.3, 6.1, 2.6 Hz, 1H, C(2')-H), 2.43 (s, 3H, Tol-CH_3), 2.41 (s, 3H, Tol-CH_3), 1.28 (d, J=6.9 Hz, 6H, iBu-CH_3).

^{13}C NMR ($CDCl_3$): δ 175.5 (C), 166.1 (C), 166.0 (C), 153.0 (C), 151.5 (C), 151.1 (C), 144.4 (C), 144.1 (C), 130.7 (CH), 129.8 (CH), 129.6 (CH), 129.3 (CH), 129.2 (CH), 126.6 (C), 126.4 (C), 113.6 (CH), 84.6 (CH), 82.6 (CH), 74.9 (CH), 63.9 (CH_2), 53.0 (CH), 38.1 (CH_2), 35.9 (CH), 21.7 (CH_3), 21.6 (CH_3), 19.2 (CH_3), 19.1 (CH_3).

HR-MALDI: m/z calculated for $C_{31}H_{30}ClIN_4O_6.Na^+$ 739.0796, found 739.0791.

7-Allyl-6-chloro-7-deaza-N(2)-isobutyryl-3',5'-di-O-p-toluoyl-2'-deoxyguanosine (6)

A solution of 6-chloro-7-deaza-7-iodo-N(2)-isobutyryl-3',5'-di-O-p-toluoyl-2'-deoxyguanosine (5.22 g, 7.28 mmol) and $Pd(PPh_3)_4$ (850 mg. 0.74 mmol) in toluene (100 ml) was treated with allyl trtibutyl stannane (3.3 ml, 10.6 mmol) and stirred at 95°C for 10.5 h under argon atmosphere. The solvent was removed under reduced pressure and the solid residue was purified by silica gel column chromatography (dichloromethane/ethyl acetate 10:1) to give the product (3.91 g, 6.20 mmol, 85%) as a white solid.

Rf = 0.58 (CH_2Cl_2:EtOAc 10:1).

^{1}H NMR ($CDCl_3$): δ 8.09 (s, 1H, C(2)-NH), 7.97 (dt, J = 8.2, 1.5 Hz, 2H, Tol-H), 7.92 (dt, J = 8.2, 1.5 Hz, 2H, Tol-H), 7.27 (d, J = 8.1 Hz, 2H, Tol-H), 7.24 (d, J = 8.2 Hz, 2H, Tol-H), 6.98 (s, 1H, C(8)-H), 6.71 (dd, J = 8.3, 5.9 Hz, 1H, C(1')-H), 5.86 (ddt, J = 17.0, 10.1, 6.5 Hz, 1H, H-C=CH_2), 5.77 (dt, J = 6.4, 2.1 Hz, 1H, C(3')-H), 5.00 (ddd, J = 17.0, 3.1, 1.5 Hz, 1H, H_a-CH=CHCH_2), 4.97 (ddd, J = 10.0, 2.9, 1.5 Hz, 1H, H_b-CH=CHCH_2), 4.78 (dd, J = 11.6, 3.5 Hz, 1H, C(5')-H), 4.61 (dd, J = 11.6, 4.1 Hz, 1H, C(5')-H), 4.56 (ddd, J = 4.1, 3.4, 2.1 Hz, 1H, C(4')-H), 3.47 (dt, J = 6.5, 1.1 Hz, 2H, C(7)-CH_2), 3.01 (sept, J = 6.4 Hz, 1H, iBu-CH), 2.91 (ddd, J = 14.3, 8.2, 6.7 Hz, 1H, C(2')-H), 2.72 (ddd, J = 14.2, 5.9, 2.1 Hz, 1H, C(2')-H), 2.43 (s, 3H, Tol-CH_3), 2.42 (s, 3H, Tol-CH_3), 1.28 (d, J = 6.9 Hz, 6H, iBu-CH_3).

^{13}C NMR ($CDCl_3$): δ 175.6 (C), 166.1 (C), 166.0 (C), 152.6 (C), 152.1 (C), 151.0 (C), 144.3 (C), 144.0 (C), 135.6 (CH), 129.8 (CH), 129.6 (CH), 129.2 (CH), 126.8 (C), 126.5 (C), 122.5 (C), 116.4 (CH), 115.4 (CH_2), 113.0 (C), 84.0 (CH), 82.2 (CH), 75.1 (CH), 64.1 (CH_2), 37.7 (CH_2), 35.8 (CH), 30.1 (CH_2), 21.7 (CH_3), 21.6 (CH_3), 19.20 (CH_3), 19.18 (CH_3).

HR-MALDI: m/z calculated for $C_{34}H_{35}ClN_4O_6.Na^+$ 653.2143, found 653.2137.

7-allyl-7-deaza-N(2)-isobutyryl-3', 5'-di-O-p-toluoyl-2'-deoxyguanosine

A solution of 7-allyl-6-chloro-7-deaza-N(2)-isobutyryl-3',5'-di-O-p-toluoyl-2'-deoxyguanosine (**6**) (4.31 g, 6.83 mmol), pyridine-2-carboxaldoxime (4.16 g. 34.1 mmol) and 1,1,3,3-tetramethylguanidine (4.3 ml, 34.2 mmol) in a mixture of dioxane (76 ml) and DMF (82 ml) was stirred for 48 h under argon atmosphere at ambient temperature. The reaction mixture was diluted with ethyl acetate (300 ml) and washed with 5% HCl (300 ml), sat. $NaHCO_3$ (300 ml) and brine (100 ml), dried over Na_2SO_4 and evaporated to dryness under reduced pressure. Purification of the residue by silica gel column chromatography (dichloromethane/ethyl acetate 10:1) gave the product (3.44 g, 5.61 mmol, 82%) as a white solid.

Rf = 0.26 (CH_2Cl_2:EtOAc 10:1), 0.39 (CH_2Cl_2:MeOH 20:1).

^{1}H NMR ($CDCl_3$): δ 11.64 (s, 1H, N(3)-H), 8.9 (s, 1H, C(2)-NH), 7.97 (dt, J = 8.2, 1.5 Hz, 2H, Tol-H), 7.92 (dt, J = 8.2, 1.5 Hz, 2H, Tol-H), 7.27 (d, J = 8.1 Hz, 2H, Tol-H), 7.24 (d, J = 8.2 Hz, 2H, Tol-H), 6.54 (s, 1H, C(8)-H), 6.29 (dd, J = 6.3, 7.9 Hz, 1H, C(1')-H), 5.97 (ddt, J = 17.0, 10.0, 6.8 Hz, 1H, H-C=CH_2), 5.76 (dt, J = 6.1, 3.1 Hz, 1H, C(3')-H), 5.06 (ddd, J = 17.0, 3.4, 1.5 Hz, 1H, H_a-CH=CHCH_2), 4.97 (ddd, J = 13.8, 4.2, 3.3 Hz, 1H, C(5')-H), 4.95 (ddd, J = 10.0, 2.1, 1.0 Hz, 1H, H_b-CH=CHCH_2), 4.55 (dd, J = 14.2, 4.5 Hz, 1H, C(5')-H), 4.51 (dd, J = 4.5, 2.8 Hz, 1H, C(4')-H), 3.52 (dddd, J = 16.1, 6.8, 2.2, 1.1 Hz, 1H, C(7)-CH-H_a), 3.46 (dddd, J = 16.1, 6.8, 2.2, 1.1 Hz, 1H, C(7)-CH-H_b), 2.97 (ddd, J = 14.2, 7.8, 6.4 Hz, 1H, C(2')-H_a), 2.68 (sept, J = 6.9 Hz, 1H, iBu-CH), 2.53 (ddd, J = 14.1, 6.2, 2.6 Hz, 1H, C(2')-H_b), 2.41 (s, 6H, Tol-CH_3), 1.26 (d, J = 6.9 Hz, 3H, iBu-CH_3), 1.24 (d, J = 6.9 Hz, 3H, iBu-CH_3).

^{13}C NMR ($CDCl_3$): δ 178.4 (C), 166.7 (C), 165.9 (C), 157.9 (C), 147.1 (C), 145.9 (C), 144.3 (C), 144.2 (C), 136.6 (CH), 129.7 (CH), 129.6 (CH), 129.22 Tol (CH), 129.15 (CH), 126.7 Tol (C), 126.6 Tol (C), 119.6 (C), 116.9 (CH), 115.4 (CH_2), 104.8 (C), 85.2 (CH), 81.7 (CH), 75.1 (CH), 63.7 (CH_2), 36.9 (CH_2), 36.4 (CH), 30.4 (CH_2), 21.64 (CH_3), 21.62 (CH_3), 18.90 (CH_3), 18.87 (CH_3).

HR-MALDI: m/z calculated for $C_{34}H_{36}N_4O_7.Na^+$ 635.2482, found 635.2476.

7-allyl-7-deaza-N(2)-isobutyryl-2'-deoxyguanosine

A solution of 7-allyl-7-deaza-N(2)-isobutyryl-3',5'-di-O-p-toluoyl-2'-deoxy guanosine (2.79 g, 4.55 mmol) in THF (85 ml) and methanol (1 ml) was treated with 1M NaOMe in MeOH (4.5 ml, 4.5 mmol) at 0°C. The ice bath was removed and the reaction mixture was stirred for 5 h at ambient temperature, neutralized with acetic acid (260 μl) and evaporated to dryness under reduced pressure. Purification of the residue by silica gel column chromatography (dichloromethane/methanol 20:1) gave the product (1.66 g, 4.41 mmol, 97%) as a white solid.

Rf = 0.07 (CH_2Cl_2:MeOH 20:1), 0.28 (CH_2Cl_2:MeOH 10:1).

^{1}H NMR (d_6-DMSO): δ 11.70 (s, 1H, N(3)-H), 11.44 (s, 1H, C(2)-NH), 6.95 (s, 1H, C(8)-H), 6.38 (dd, J = 8.6, 5.7 Hz, 1H, C(1')-H), 6.05 (ddt, J = 17.1, 10.0, 6.7 Hz, 1H, H-C=CH_2), 5.20 (d, J = 3.5 Hz, 1H, C(3')-OH), 5.07 (ddd, J = 17.1, 2.1, 1.5 Hz, 1H, H_a-CH=CHCH_2), 4.98 (ddd, J = 10.0, 2.1, 1.2 Hz, 1H, H_b-CH=CHCH_2), 4.84 (t, J = 5.4 Hz, 1H, C(5')-OH), 4.30 (m, 1H, C(3')-H), 3.77 (dt, J = 4.8, 2.3 Hz, 1H, C(4')-H), 3.49 (m, J = 5.3, 4.8 Hz, 2H, C(5')-H_2), 3.43 (ddd, J = 6.7, 2.1, 1.5 Hz, 1H, C(7)-CH_2), 2.75 (sept, J = 6.8 Hz, 1H, iBu-CH), 2.35 (ddd, J = 13.0, 8.6, 5.6 Hz, 1H, C(2')-H_a), 2.10 (ddd, J = 13.0, 5.7, 2.4 Hz, 1H, C(2')-H_b), 1.11 (d, J = 6.8 Hz, 6H, iBu-CH_3).

^{13}C NMR (d_6-DMSO): δ 179.7 (C), 168.2 (C), 156.8 (C), 147.4 (C), 146.5 (C), 137.3 (CH), 117.6 (C), 115.9 (CH), 114.9 (CH_2), 102.5 (C), 86.9 (CH), 82.0 (CH), 70.8 (CH), 61.8 C(CH_2), 39.4 (CH_2), 34.6 (CH), 30.1 (CH_2), 18.78 (CH_3), 18.76 (CH_3).

HR-MALDI: m/z calculated for $C_{18}H_{24}N_4O_5.H^+$ 377.1825, found 377.1820.

7-allyl-7-deaza-N(2)-isobutyryl-3',5'-di-(O-tert-butyl-dimethylsilyl)-2'-deoxyguanosine (7)

A solution of 7-allyl-7-deaza-N(2)-isobutyryl-2'-deoxyguanosine (660 mg, 1.75 mmol), tert-butyl-chlorodimethylsilane (631 mg, 4.38 mmol) and imidazole (595 mg, 8.74 mmol) in DMF (12 ml) was stirred for 16 h under argon atmosphere at ambient temperature. The reaction mixture was diluted with ethyl acetate (50 ml) and washed with 5% HCl (50 ml), sat. $NaHCO_3$ (50 ml) and brine (50 ml), dried over Na_2SO_4, and evaporated to dryness under reduced pressure. Purification of the residue by silica gel column chromatography (hexane/ethyl acetate 4:1) gave **7** (1.00 g, 1.65 mmol, 94%) as a white solid.

Rf = 0.44 (CH_2Cl_2:MeOH 20:1), 0.79 (CH_2Cl_2:MeOH 10:1).

^{1}H NMR ($CDCl_3$): δ 11.59 (s, 1H, N(3)-H), 8.34 (s, 1H, C(2)-NH), 6.69 (s, 1H, C(8)-H), 6.38 (dd, J = 7.9, 6.0 Hz, 1H, C(1')-H), 6.08 (ddt, J = 17.1, 10.0, 6.7 Hz, 1H, H-C=CH_2), 5.11 (ddd, J = 17.1, 3.4, 1.6 Hz, 1H, H_a-CH=CHCH_2), 5.02 (ddd, J = 10.0, 2.1, 1.1 Hz, 1H, H_b-CH=CHCH_2), 4.51 (dt, J = 5.8, 2.9 Hz, 1H, C(3')-H), 3.90 (m, J = 3.8, 3.4 Hz, 1H, C(4')-H), 3.74 (dd, J = 11.4, 3.4 Hz, 1H, C(5')-H), 3.70 (dd, J = 11.4, 4.1 Hz, 1H, C(5')-H), 3.60 (dddd, J = 16.0, 6.7, 2.4, 1.2 Hz, 1H, C(7)-CH_2), 3.54 (dddd, J = 17.2, 6.7, 2.3, 1.1 Hz, 1H, C(7)-CH_2), 2.58 (sept, J = 6.9 Hz, 1H, iBu-CH), 2.36 (ddd, J = 13.1, 7.9, 5.9 Hz, 1H, C(2')-H), 2.18 (ddd, J = 13.0, 6.0, 3.1 Hz, 1H, C(2')-H), 1.25 (d, J = 6.9 Hz, 3H, iBu-CH_3), 1.21 (d, J = 6.9 Hz, 3H, iBu-CH_3), 0.92 (s, 9H, t-Bu-CH_3), 0.91 (s, 9H, t-Bu-CH_3), 0.10 (s, 6H, Si-CH_3), 0.08 (s, 6H, Si-CH_3).

^{13}C NMR ($CDCl_3$): δ 177.9 (C), 171.3 (C), 157.9 (C), 147.5 (C), 145.8 (C), 137.1 (CH), 119.7 (C), 115.5 (CH),

115.2 (CH_2), 103.8 (C), 87.2 (CH), 82.6 (CH), 72.4 (CH), 63.1 (CH_2), 40.7 (CH_2), 36.5 (CH), 30.6 (CH_2), 26.0 (CH_3), 25.8 (CH_3), 19.0 (CH_3), 18.9 (CH_3), 18.4 (C), 18.0 (C), -4.7 (CH_3), -4.8 (CH_3), -5.4 (CH_3), -5.5 (CH_3).

HR-MALDI: m/z calculated for $C_{30}H_{52}N_4O_5Si_2.H^+$ 605.3554, found 605.3549.

7-deaza-7-(2,3-dihydroxy-propyl)-N(2)-isobutyryl-3',5'-di-(O-tert-butyl-dimethylsilyl)- 2'-deoxyguanosine (8)

To a solution of 7-allyl-7-deaza-N(2)-isobutyryl-3',5'-di-(O-tert-butyl-dimethylsilyl)-2'-deoxyguanosine (**7**) (900 mg, 1.49 mmol) in a mixture of THF (22 ml) and H_2O (2.2 ml) was added a 2% soln. of OsO_4 in t-BuOH (2 ml, 0.197 mmol) at 0°C. After addition of 4-methylmorpholine-4-oxide monohydrate (343 mg, 2.54 mmol) the reaction mixture was stirred for 3 h at 0°C under argon atmosphere. The reaction mixture was diluted with ethyl acetate (50 ml) and washed with 5% HCl (50 ml), sat. $NaHCO_3$ (50 ml), sat. Na_2SO_3 (10 ml) and brine (50 ml), dried over Na_2SO_4, and evaporated to dryness under reduced pressure. Purification of the residue by silica gel column chromatography (dichloromethane/methanol 20:1) gave **8** (760 mg, 1.20 mmol, 80%) as a mixture of diastereoisomers as a white solid.

Rf = 0.46 (CH_2Cl_2:MeOH 10:1), 0.24 (CH_2Cl_2:EtOAc 20:1).

^{1}H NMR ($CDCl_3$): δ 11.86 (s, 1H, N(3)-H), 8.65 (s, 1H, C(2)-NH), 6.88&6.86 (2 peaks, 1H, C(8)-H), 6.42 (dd, J = 7.5, 6.2 Hz, 1H, C(1')-H), 4.52 (m, 1H, C(3')-H), 3.93 (dd, J = 6.4, 3.4 Hz, 1H, C(4')-H), 3.80 (ddd, J = 10.8, 6.3, 4.6 Hz, 1H, CH(OH)), 3.74 (d, J = 3.6 Hz, 2H, C(5')-H_2), 3.58 (dd, J = 11.9, 4.6 Hz, 1H, CH_2(OH)), 3.54 (dd, J = 11.9, 4.6 Hz, 1H, CH_2(OH)), 3.08-2.91 (m, 2H, C(7)-CH_2), 2.60 (sept, J = 6.9 Hz, 1H, iBu-CH), 2.34 (ddd, J = 5.8, 7.6, 13.1 Hz, 1H, C(2')-H_a), 2.23 (ddd, J = 3.2, 6.1, 13.1 Hz, 1H, C(2')-H_b), 1.26 (d, J = 6.9 Hz, 3H, iBu-CH_3), 1.25 (d, J = 6.9 Hz, 3H, iBu-CH_3), 0.10 (s, 6H, Si-CH_3), 0.09 (s, 6H, Si-CH_3).

^{13}C NMR ($CDCl_3$): δ 178.3 (C), 159.5 (C), 147.8 (C), 145.8 (C), 117.8 (CH), 116.2 (C), 104.5 (C), 87.4 (CH), 82.9 (CH), 73.0 (CH), 72.3 (CH), 65.2 (CH_2), 63.1 (CH_2), 41.1 (CH_2), 36.5 (CH), 30.0 (CH_2), 25.9 (CH_3), 25.8 (CH_3), 19.0 (CH_3), 18.4 (C), 18.0 (C), -4.7 (CH_3), -4.8 (CH_3), -5.4 (CH_3), -5.5 (CH_3).

HR-MALDI: m/z calculated for $C_{30}H_{54}N_4O_7Si_2.Na^+$ 661.3429, found 661.3423.

7-deaza-7-(2,3-diacetoxy-propyl)-N(2)-isobutyryl-3',5'-di-(O-tert-butyl-dimethylsilyl)- 2'-deoxyguanosine

A solution of 7-deaza-7-(2,3-dihydroxy-propyl)-N(2)-isobutyryl-3',5'-di-(O-tert-butyl-dimethylsilyl)-2'-deoxy-guanosine (**8**) (750 mg, 1.17 mmol) in pyridine (17 ml) was treated with acetic anhydride (5 ml, 52.5

mmol) at 0°C and stirred for 1 h at room temperature. The mixture was quenched with H_2O (1 ml, 55.6 mmol) at 0°C and evaporated to dryness under reduced pressure. The solid residue was resuspended in ethyl acetate (50 ml) and washed with 5% HCl (50 ml), sat. $NaHCO_3$ (50 ml) and brine (50 ml), dried over Na_2SO_4, and evaporated to dryness under reduced pressure. Purification by silica gel column chromatography (hexane/ethyl acetate 2:1) gave the product (636 mg, 0.88 mmol, 75%) as a white solid.

Rf = 0.76 (CH_2Cl_2:MeOH 10:1).

^{1}H NMR ($CDCl_3$): δ 11.62 (s, 1H, N(3)-H), 8.31&8.30 (2 peaks, 1H, C(2)-NH), 6.78&6.77 (2 peaks, 1H, C(8)-H), 6.34 (m, 1H, C(1')-H), 5.46 (m, 1H, AcO-CH), 4.50 (m, 1H, C(3')-H), 4.38-4.28 (m, 1H, AcO-CH_2), 4.15-4.05 (m, 1H, AcO-CH_2), 3.90 (m, 1H, C(4')-H), 3.77-3.65 (m, 2H, C(5')-H_2), 3.26-3.10 (m, 1H, C(7)-CH_2), 3.10-2.91 (m, 1H, C(7)-CH_2), 2.58 (sept, J = 6.9 Hz, 1H, iBu-CH), 2.38-2.26 (m, 1H, C(2')-H_a), 2.24-2.15 (ddd, J = 3.1, 5.9, 13.0 Hz, 1H, C(2')-H_b), 2.05 (s, 3H, Ac), 2.00 (s, 3H, Ac), 1.26 (d, J = 6.9 Hz, 3H, iBu-CH_3), 1.24 (d, J = 6.9 Hz, 3H, iBu-CH_3), 0.93 (s, 9H, t-Bu-CH_3), 0.91 (s, 9H, t-Bu-CH_3), 0.11 (s, 6H, Si-CH_3), 0.09 (s, 6H, Si-CH_3).

^{13}C NMR ($CDCl_3$): δ 177.0 (C), 170.7 (C), 170.20 & 170.16 (C), 157.8 (C), 147.44 & 147.38 (C), 146.0 (C), 116.8 & 116.6 (CH), 115.5 & 115.3 (C), 104.2 (C), 87.27 & 87.25 (CH), 82.67 & 82.65 (CH), 72.43 & 72.35 (CH), 71.7 & 71.5 (CH), 64.8 & 64.7 (CH_2), 63.31 & 63.27 (CH_2), 40.7 & 40.6 (CH_2), 36.5 (CH), 27.6 & 27.4 (CH_2), 25.9 (CH_3), 25.8 (CH_3), 21.0 (CH_3), 20.8 (CH_3), 19.02 & 18.98 (CH_3), 18.94 & 18.90 (CH_3), 18.4 (C), 18.0 (C), -4.7 (CH_3), -4.8 (CH_3), -5.4 (CH_3), -5.5 (CH_3).

HR-MALDI: m/z calculated for $C_{34}H_{58}N_4O_9Si_2.Na^+$ 745.3640, found 745.3635.

7-deaza-7-(2,3-diacetoxy-propyl)-N(2)-isobutyryl-2'-deoxyguanosine (9)

A solution of 7-deaza-7-(2,3-diacetoxy-propyl)-N(2)-isobutyryl-3',5'-di-(O-tert-butyl-dimethylsilyl)-2'-deoxyguanosine (621 mg, 0.86 mmol), TBAF.$3H_2O$ (813 mg, 2.58 mmol) and acetic acid (120 μl, 2.10 mmol) in THF (12 ml) was stirred for 16 h at ambient temperature. Evaporation to dryness under reduced pressure and purification of the residue by silica gel column chromatography (dichloromethane/methanol 20:1) gave **9** (420 mg, 0.85 mmol, 99%) as a white solid.

Rf = 0.09 (CH_2Cl_2:MeOH 20:1).

^{1}H NMR (d_6-DMSO): δ 11.74 (s, 1H, N(3)-H), 11.46 (s, 1H, C(2)-NH), 7.04&7.03 (2 peaks, 1H, C(8)-H), 6.37 (dd, J = 5.8, 8.2 Hz, 1H, C(1')-H), 5.35-5.26 (m, 1H, AcO-CH), 5.21 (bs, 1H, C(3')-OH), 4.85 (t, J = ~5 Hz, 1H, C(5')-OH), 4.31 (bs, 1H, C(3')-H), 4.22&4.18 (2t, J = 2.8 Hz, 1H, AcO-CH_2), 4.07-3.99 (m, 1H, AcO-CH_2), 3.78 (ddd, J = 2.3, 4.6, 4.6 Hz, 1H, C(4')-H), 3.50 (bs, 2H, C(5')-H_2), 3.07&3.02 (2t, J = 6.0 Hz, 1H, C(7)-CH_2), 2.93-

2.82 (m, 1H, C(7)-CH_2), 2.75 (sept, J = 6.8 Hz, 1H, iBu-CH), 2.39-2.27 (m, 1H, C(2')-H_a), 2.17-2.08 (m, 1H, C(2')-H_b), 2.00 (s, 3H, Ac), 1.95 (s, 3H, Ac), 1.11 (d, J = 6.8 Hz, 6H, iBu-CH_3).

^{13}C NMR (d_6-DMSO): δ 179.8 (C), 170.0 (C), 169.6 (C), 156.9 (C), 147.31 & 147.27 (C), 146.6 (C), 117.5 & 117.4 (CH), 113.9 & 113.8 (C), 102.8 (C), 87.0 (CH), 82.2 (CH), 70.9 (CH), 70.8 (CH), 64.0 & 63.9 (CH_2), 61.87 & 61.83 (CH_2), 40.3 (CH_2), 39.5 & 39.4 (CH_2), 34.6 (CH), 26.9 (CH_2), 20.7 (CH_3), 20.4 (CH_3), 18.80 & 18.78 (CH_3), 18.76 & 18.73 (CH_3).

HR-MALDI: m/z calculated for $C_{22}H_{30}N_4O_9.H^+$ 495.2091, found 495.2087.

7-deaza-7-(2,3-diacetoxy-propyl)-5'-O-(4,4'-dimethoxytrityl)-N(2)-isobutyryl-2'-deoxyguanosine

A solution of 7-deaza-7-(2,3-diacetoxy-propyl)-N(2)-isobutyryl-2'-deoxyguanosine (**13**) (360 mg, 0.728 mmol) in pyridine (17 ml) was treated with 4,4'-dimethoxytrityl chloride (296 mg, 0.874 mmol) and stirred for 1 h at 23°C under argon atmosphere. The reaction was quenched with methanol (100 μl) and concentrated under reduced pressure. The resulting oil was dissolved in ethyl acetate (50 ml) and washed with H_2O (50 ml), sat. $NaHCO_3$ (50 ml) and brine (50 ml), dried over Na_2SO_4, and evaporated to dryness under reduced pressure. Purification of the residue by silica gel column chromatography (dichloromethane/methanol 20:1) gave the product (540 mg, 0.678 mmol, 93%) as a white solid.

Rf = 0.28 (CH_2Cl_2:MeOH 20:1).

^{1}H NMR ($CDCl_3$): δ 11.72 (s, 1H, N(3)-H), 8.98&8.92 (2 peaks, 1H, C(2)-NH), 7.44-7.38 (m, 2H, DMTr), 7.32-7.18 (m, 7H, DMTr), 6.82-6.76 (m, 4H, DMTr), 6.68&6.62 (2 peaks, 1H, C(8)-H), 6.35 (dd, J = 5.8, 8.2 Hz, 1H, C(1')-H), 5.38 (m, 1H, AcO-CH), 4.53 (m, 1H, C(3')-H), 4.30-4.21 (m, 1H, AcO-CH_2), 4.10-3.99 (m, 2H, AcO-CH_2 & C(4')-H), 3.76 (s, 6H, DMTr-OCH_3),3.29 (d, J = 4.0 Hz, 2H, C(5')-H_2), 3.13-2.81 (m, J = 6.0 Hz, 2H, C(7)-CH_2), 2.45 (m, 1H, C(2')-H_a), 2.40 (sept, J = 6.9 Hz, 1H, iBu-CH), 2.31-2.17 (m, 1H, C(2')-H_b), 1.98&1.97 (2 s, 3H, Ac), 1.89&1.86 (2 s, 3H, Ac), 1.09 (m, 6H, iBu-CH_3).

^{13}C NMR ($CDCl_3$): δ 178.73 & 178.69 (C), 171.0 & 170.8 (C), 170.4 (C), 158.6 & 158.0 (C), 147.8 & 147.7 (C), 146.3 & 146.2 (C), 144.5 (C), 135.7 (C), 135.6 (C), 130.0 (CH), 128.1 (CH), 127.8 (CH), 126.9 (CH), 117.2 & 117.1 (CH), 115.3 (C), 113.2 (CH), 104.2 (C), 86.5 (C), 85.6 & 85.5 (CH), 83.1 & 82.9 (CH), 72.6 (CH), 71.8 & 71.6 (CH), 64.7 (CH_2), 64.1 (CH_2), 55.2 (CH_3), 46.1 (CH2), 40.2 & 40.1 (CH_2), 36.2 (CH), 27.4 (CH_2), 21.0 (CH_3), 20.7 (CH_3), 18.8 (CH_3), 18.76 & 18.73 (CH_3).

HR-MALDI: m/z calculated for $C_{43}H_{48}N_4O_{11}.HNEt_3^+$ 898.4602, found 898.4557.

7-deaza-7-(2,3-diacetoxy-propyl)-5'-O-(4,4'-dimethoxytrityl-N(2)-isobutyryl)-2'-deoxyguanosine-3'-[(2-cyanoethyl)-N,N-diisopropylphosphoramidite] (4)

A solution of 7-deaza-7-(2,3-diacetoxy-propyl)-5'-O-(4,4'-dimethoxytrityl)-N(2)-isobutyryl-2'-deoxyguanosine (281 mg, 0.35 mmol) and N-ethyl-diisopropylamine (242 μl, 1.41 mmol) in dichloromethane (6 ml) was treated with 2-cyanoethyl-N,N-diisopropyl-chlorophosphoramidite (118 μl, 0.53 mmol) and stirred for 1 h at 23°C under argon atmosphere. The reaction was quenched with methanol (0.5 ml) and concentrated under reduced pressure. The resulting oil was dissolved in ethyl acetate (20 ml) and washed with H_2O (20 ml), sat. $NaHCO_3$ (20 ml) and brine (20 ml), dried over Na_2SO_4 and evaporated to dryness under reduced pressure. Purification of the residue by silica gel column chromatography (dichloromethane/methanol 20:1) gave **4** (308 mg, 0.31 mmol, 88%) as a white solid.

Rf = 0.44 (CH_2Cl_2:MeOH 20:1);

^{1}H NMR ($CDCl_3$): δ 11.58 (s, 1H, N(3)-H), 8.42&8.07 (2 peaks, 1H, C(2)-NH), 7.48-7.41 (m, 2H, DMTr), 7.36-7.17 (m, 7H, DMTr), 6.83-6.77 (m, 4H, DMTr), 6.72-6.68 (m, 1H, C(8)-H), 6.37-6.27 (m, 1H, C(1')-H), 5.40 (m, 1H, AcO-CH), 4.74-4.58 (m, 1H, C(3')-H), 4.31-4.23 (m, 1H, AcO-CH_2), 4.23-4.14 (2 m, 1H, C(4')-H), 4.10-4.01 (m, 1H, AcO-CH_2), 3.90-3.76 (m, 1H, iPr-CH), 3.78 (s, 6H, DMTr-OCH_3), 3.76-3.64 (m, 1H, iPr-CH), 3.65-3.52 (m, 2H, P-O-CH_2), 3.42-3.22 (m, 2H, C(5')-H_2), 3.17-2.88 (m, 2H, C(7)-CH_2), 2.72-2.65 (2t, J = 6.4 Hz, 1H, NC-CH_2), 2.57-2.50 (m, 1H, C(2')-H_a), 2.47 (t, J = 6.4 Hz, 1H, NC-CH_2), 2.41-2.30 (m, 1H, C(2')-H_b), 2.34 & 2.24 (2 sept, J = 6.9 Hz, 1H, iBu-CH), 2.00 & 1.99 (2 s, 3H, Ac), 1.91&1.88 (3 s, 3H, Ac), 1.20-1.06 (m, 18H, iBu-CH_3&iPr-CH_3).

^{13}C NMR ($CDCl_3$): δ 178.4 & 178.1 (C), 170.7 & 170.2 (C), 158.6 & 157.7 (C), 147.44 & 147.38 (C), 146.0 & 145.9 (C), 144.8 & 144.6 (C), 135.91 & 135.87 (C), 135.8 & 135.7 (C), 130.1 & 130.0 (CH), 128.2 & 128.1 (CH), 127.8 (CH), 126.9 (CH), 117.94 & 117.89 (CH), 117.3 (CH), 115.1 (C), 113.1 (CH), 104.9 (C), 86.4 & 86.3 (C), 85.0 (CH), 84.2 (CH), 73.9 & 73.7 (CH), 71.64 & 71.59 & 71.52 (CH), 64.7 (CH_2), 63.8 (CH_2), 58.3 (CH_2), 58.0 (CH_2), 55.2 (CH_3), 43.32 & 43.26 (CH), 43.16 & 43.09 (CH), 41.32 (C), 36.2 & 36.1 (CH), 29.0, 27.6, 27.35 & 27.31 (CH_2), 24.64 (CH_3), 24.59 (CH_3), 24.54 (CH_3), 24.49 (CH_3), 22.6 (CH_3), 20.98 & 20.94 (CH_3), 20.7 (CH_3), 20.40 (CH_2), 20.35 (CH_2), 20.3 (CH_2), 20.2 (CH_2), 18.94 & 18.88 (CH_3), 18.82 & 18.80 (CH_3).

^{31}P NMR ($CDCl_3$): δ 148.64, 148.60, 148.31, 148.27.

HR-MALDI: m/z calculated for $C_{52}H_{65}N_6O_{12}P.Na^+$ 1019.4296, found 1019.4290.

DNA synthesis & Purification

Phosphoramidite **4** was introduced at position **X** of the oligonucleotides **10a** d(GTCACTGGTA**X**ACAGCATTG), **10b** d(CAATGCT**X**TCTACCAGTGAC), **10c** d(CGTA**X**TCATGC) and **10d** d(GCAT**X**ACTACG) by automated DNA synthesis on an Expedite 8909 Nucleic Acid Synthesis System (Applied Biosystems) using 1μmol 1000Å CPG-dG and -dC column cartridges (Glen Research) and standard reagents and conditions. For **4** an extended coupling time of 15min was used. The trityl-on oligomers were deprotected by treatment with concentrated ammonia (50°C, 12h), lyophilized and then purified and detritylated using 1μmol TOP-cartridges (Varian) according to the manufacturers instructions. After lyophilization the pure oligonucleotides were redissolved in de-ionized sterile water and stored at -20°C. The concentrations of the oligonucleotide solutions were estimated by measuring their UV-absorption at 260nm wavelength. The incorporation and integrity of the protected diol functionality was verified by mass spectrometry (see Table 1).

Formation of ICLs by reductive amination

A solution of oligonucleotides **10a** (25 nmol) and **10b** (25 nmol) in 125 μl H_2O was heated to 95°C and allowed to cool to room temperature over a period of 3 hours. After addition of 15 μl 1M sodium phosphate buffer (pH 5.4) and 10 μl 50 mM $NaIO_4$ the reaction mixture was kept for 6 hours at 4°C in the dark. Excess $NaIO_4$ was quenched with 10 μl 55mM Na_2SO_3 and following the addition of 10 μl 5 mM amine (either NH_4OAc, NH_2CH_3, hydrazine, ethylenediamine or N, N'-dimethyl-ethylenediamine) solution in water and 10 μl 0.5 M $NaCNBH_3$ the reaction mixture was left overnight at room temperature in the dark. ICL formation was assessed by electrophoresis on a denaturing 20% polyacrylamide gel, containing 7M urea. ICL products were purified by HPLC on a C18 column (Nucleosil C18, 5 μm, 100Å, 4mm x 250mm, BGB Analytic AG) using the following elution gradient: linear 2.5-25% B over 30 min, linear 25-90% B till 31 min, isocratic 90% B till 35 min, linear 90-2.5% B till 36 min, isocratic 2.5% B till 40 min; flow rate: 1 ml/min; eluent A: 0.1 M TEAA (pH=7); eluent B: CH_3CN. In these conditions the elution times of the starting oligonucleotides and the ICL products were 19.5 min and 21.5 min, respectively. The identity of the purified ICL products was confirmed by ESI-MS (Table 1) and by enzymatic digestion (Figure 4).

Table 1: ESI-MS data of ss 11- and 20-mer and crosslinked oligonucleotides.

#	Sequences	oligomer	Calc m/z	Found m/z
10a	5'-GTCACTGGTAXACAGCATTG	ss 20-mer diol	6150.0	6149.6
10b	5'-CAATGCTXTCTACCAGTGAC	ss 20-mer diol	6230.1	6229.6
10c	5'-CGTAXTCATGC	ss 11-mer diol	3405.1	3405.7
10d	5'-GCATXACTACG	ss 11-mer diol	3414.1	3415.0
12c	5'-GTCACTGGTA**X**ACAGCATTG 3'-CAGTGACCATCT**X**TCGTAAC	ds 20-mer Hy ICL	12316.3	12311.3
12d	5'-GTCACTGGTA**X**ACAGCATTG 3'-CAGTGACCATCT**X**TCGTAAC	ds 20-mer EDA ICL	12344.3	12342.8
12e	5'-GTCACTGGTA**X**ACAGCATTG 3'-CAGTGACCATCT**X**TCGTAAC	ds 20-mer DMEDA ICL	12372.4	12372.1
12f	5'-CGTA**X**TCATGC 3'-GCATCA**X**TACG	ds 11-mer Hy ICL	6755.3	6754.0
12g	5'-CGTA**X**TCATGC 3'-GCATCA**X**TACG	ds 11-mer EDA ICL	6783.3	6784.9
12h	5'-CGTA**X**TCATGC 3'-GCATCA**X**TACG	ds 11-mer DMEDA ICL	6811.3	6811.5

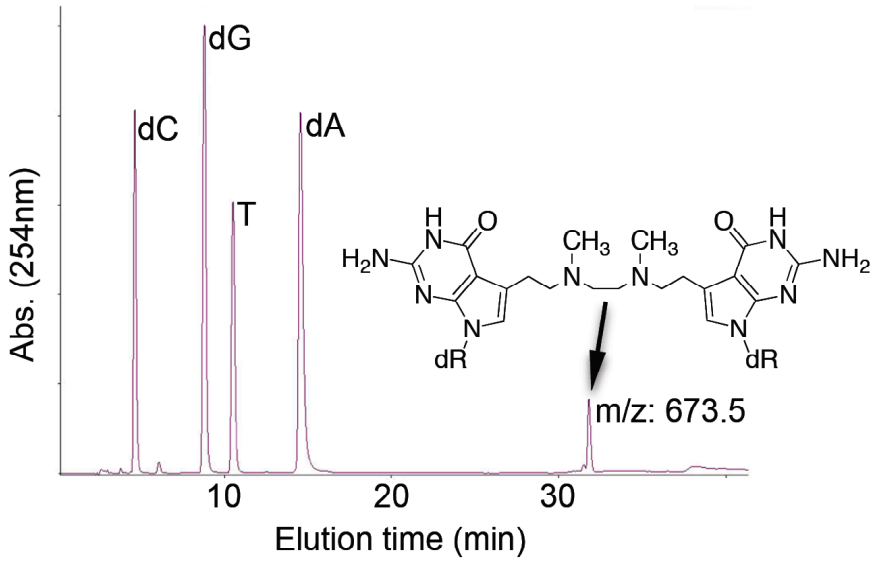
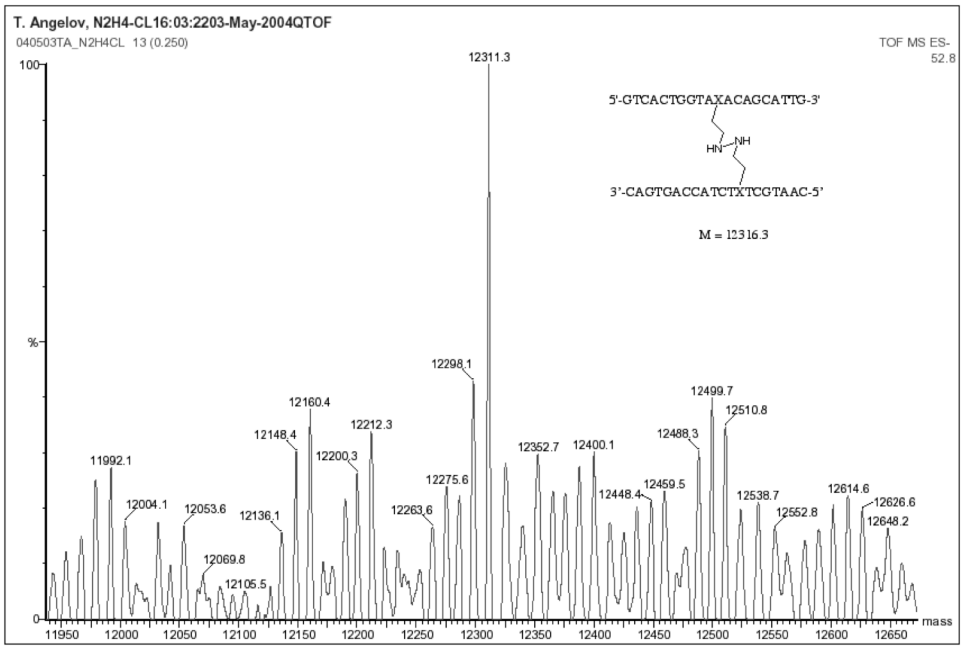


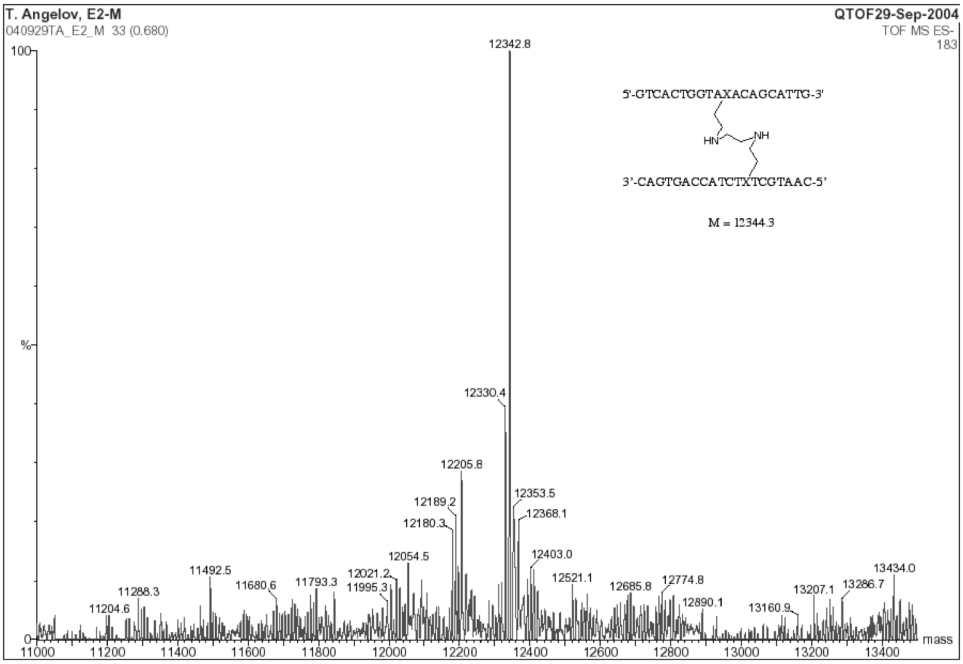
Figure 5: Enzymatic digestion of cross-linked DNA. A 20-mer dsDNA containing an DMEDA-ICL was digested with snake venom phosphodiesterase, exonuclease III, calf intestine phosphatase (CIP) and the digestion mixture analyzed by HPLC: Elution time of the five main peaks was as follows: dC (4.6 min), dG (8.7 min), T (10.5 min), dA (14.5 min), cross-linked dimer (31.8 min). The standard nucleosides were identified by co-injection with authentic standards and the identity of the cross-linked dimer was confirmed by MS analysis. M+/z: found: 673.5, calcd: 673.3

MS data

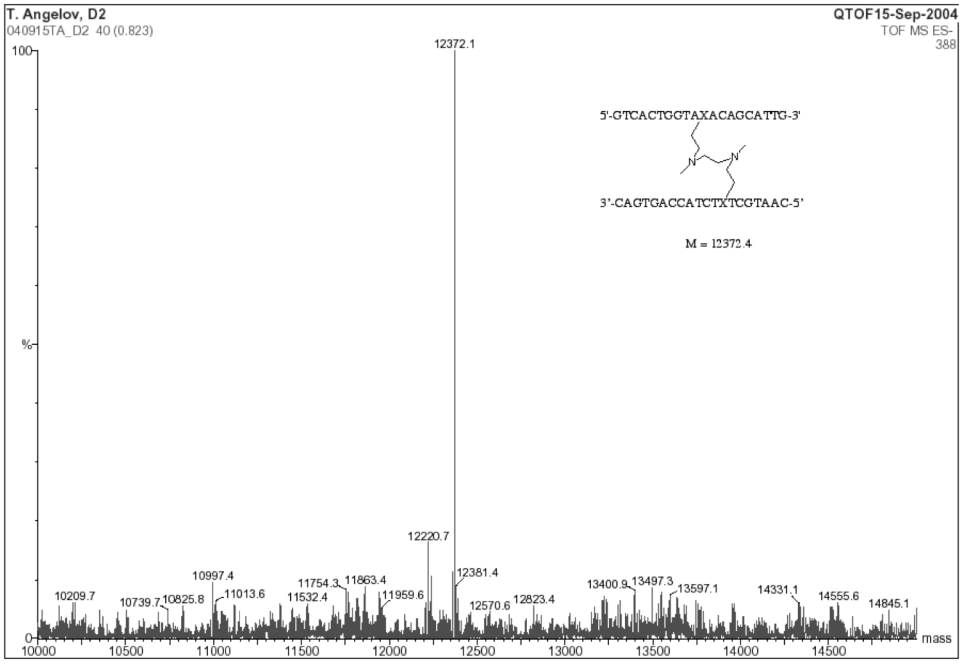
ESI-MS of hydrazine ICL 12c



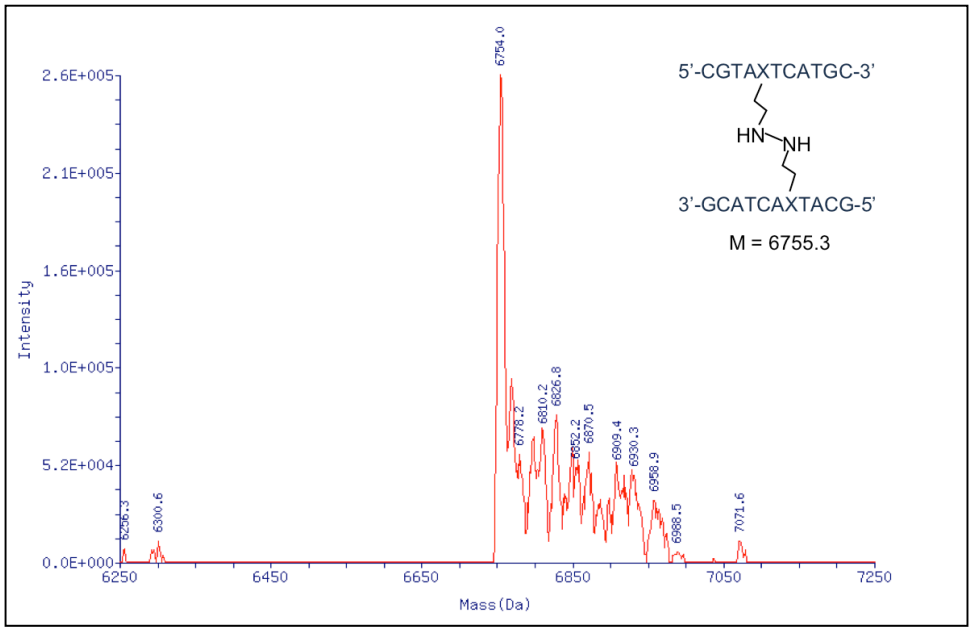
ESI-MS of ethylenediamine ICL 12d



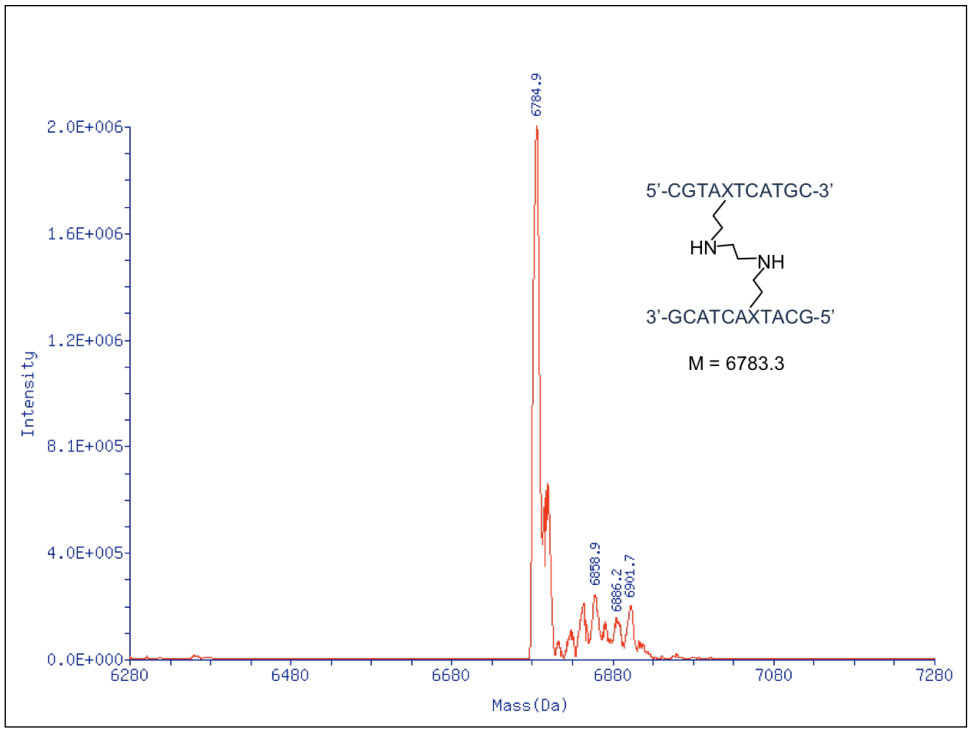
ESI-MS of N,N'-dimethyl-ethylenediamine ICL 12e



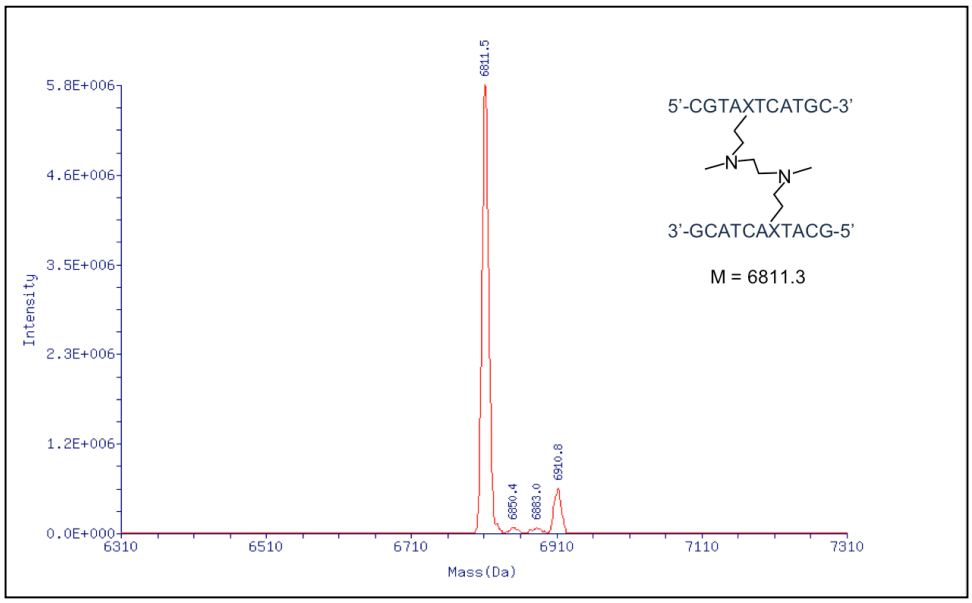
ESI-MS of hydrazine ICL 12f



ESI-MS of ethylendiamine ICL 12g



ESI-MS of N,N'-dimethyl-ethylenediamine ICL 12h



Chapter 3

Synthesis and Molecular Modeling of a Nitrogen Mustard DNA Interstrand Crosslink

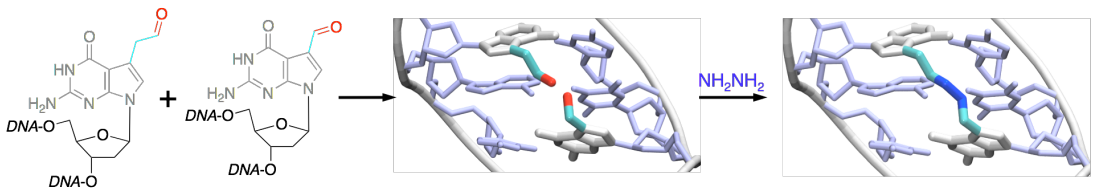
Foreword

This chapter is the central core of this dissertation. We used the strategy described in Chapter 2 to synthesize a Nitrogen Mustard DNA interstrand crosslink that is isosteric to the natural occurring one. To confirm our new substrate as a valid alternative to the original one, we established a strong collaboration with Dr. Simmerling laboratory where Arthur J. Campbell, a member of both the Schärer and the Simmerling group, performed extensive molecular modeling analysis of the new substrate. The results of the synthetic and the computational work are described in this chapter that was adapted from the manuscript submitted for publication.

Synthesis and Molecular Modeling of a Nitrogen Mustard DNA Interstrand Crosslink

Adapted from the manuscript by Angelo Guainazzi, Arthur J. Campbell, Todor Angelov, Carlos Simmerling and Orlando D. Schärer, submitted for publication

ABSTRACT



We report the synthesis of a stable NM ICL using a post-synthetic double reductive amination reaction. Our synthetic analog contains substitutions of three atoms to overcome the relative instability of NM ICLs and to expedite synthetic efficiency. We used molecular dynamics simulations to confirm that the native and synthetic NM ICLs have identical structures and dynamic properties. Thus, over 60 years after nitrogen mustards were the first agents used to treat tumors by chemotherapy, our studies provide a method to generate the main DNA adduct formed by NMs in amounts permitting extensive structural and biological studies.

Introduction

DNA interstrand cross-links (ICLs) are formed by bifunctional agents with the ability to covalently link two strands of duplex DNA. ICLs are extremely cytotoxic since they block essential processes such as DNA replication and transcription. Based on these properties, agents such as nitrogen mustards (NM) or cisplatin are widely used in cancer chemotherapy [172] but ICLs induce a number of biological responses, which counteract the therapeutic effects of crosslinking agents. The elucidation of the mechanism by which these responses cause resistance in tumor cells has been hampered by the lack of efficient methods to generate ICLs formed by antitumor agents. Current approaches toward the synthesis of site-specific ICLs have yielded mostly models of the clinically relevant lesions or ICLs formed by endogenous agents [50,52,66,82,160-163,173,174]. A number of recent studies have shown that ICLs with different structures are processed in distinct ways (see [129,133,175] for some recent reviews), emphasizing the need to generate substrates containing the ICLs formed by the clinically important drugs.

Here we report the synthesis of a stable NM ICL isostere (**2**) and use atomic detail simulations to show that it recaptures the structural and dynamic properties of the native NM ICL **1** (Figure 1).

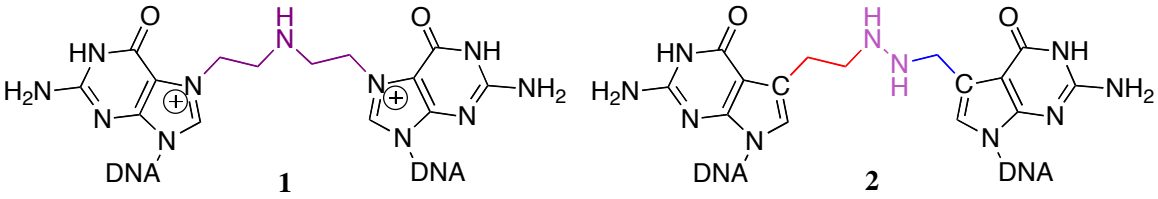


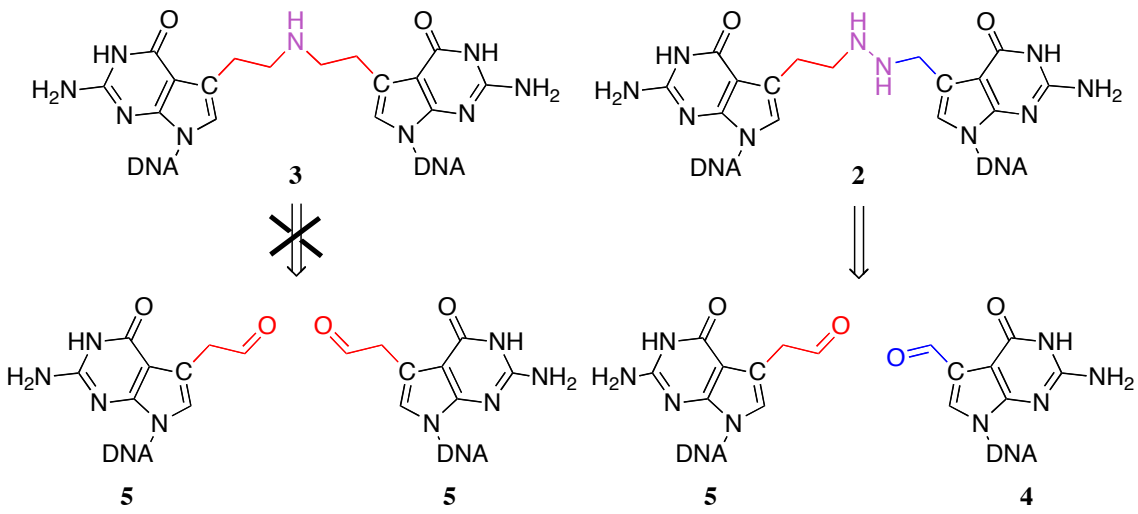
Figure 1. Structure of a NM ICL (1) and the stable analog (2).

Synthesis of the isosteric analogue

To obtain the NM ICL isostere we used a strategy that we recently developed for the synthesis of major groove ICLs [36]. This approach involves a double reductive amination reaction of two acetaldehyde functionalities (**5**) linked to the 7-position of G

residues on complementary strands of dsDNA. dG was substituted with 7-deaza-2'-deoxyguanosine to counteract the inherent lability of the glycosidic bond in N7-alkylated guanines [176].This approach allowed the generation of a six atom ICL by coupling two acetaldehyde groups with hydrazine. However, we were unsuccessful in generating ICL **3** with the 5-atom bridge found in NM ICLs [25,177,178] (Scheme 1), possibly because the reductive amination with ammonia was not powerful enough to introduce the strain in the DNA caused by the ~7.5Å bridge of the NM [28,36,165].

We reasoned that an ICL isosteric to those formed by NM might be formed by reaction of hydrazine with an acetaldehyde (**5**) and a formyl aldehyde derivative (**4**) of deazaguanine, exploiting the higher reactivity of hydrazine over ammonia (Scheme 1). [36]

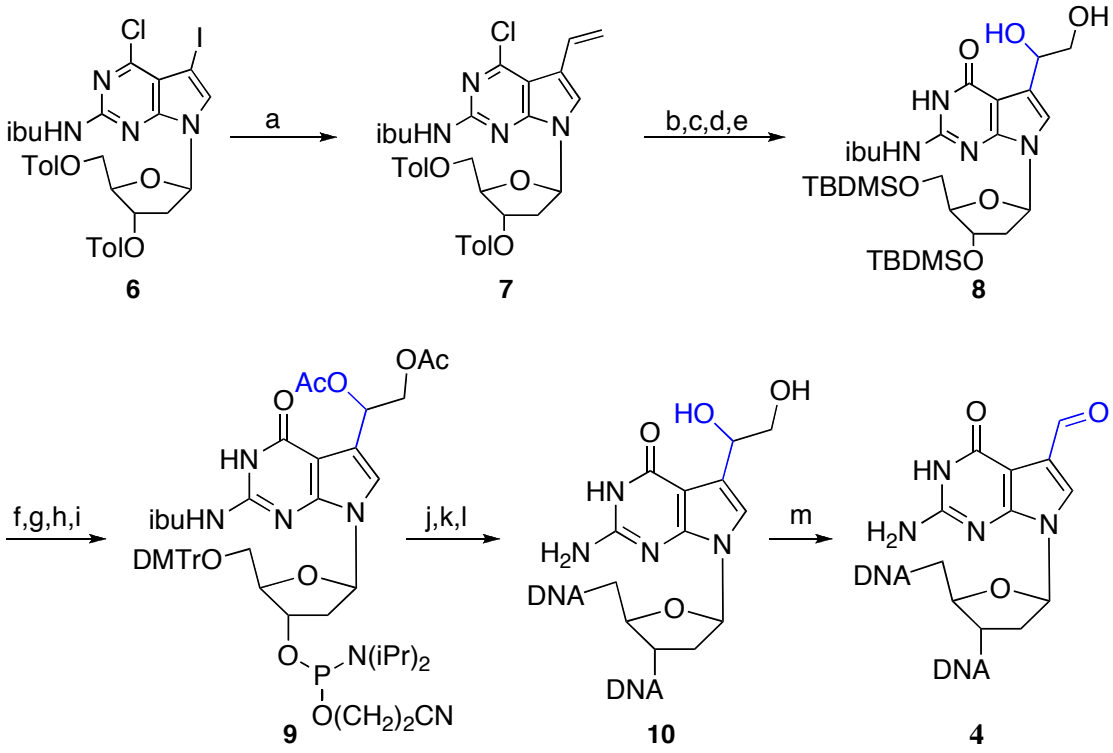


Scheme 1. Formation of NM ICLs by reductive amination was not successful using two acetaldehyde precursors 5 and NH_4Cl, prompting us to explore the generation of isostere 2 by linking precursors 4 and 5 with hydrazine.

We synthesized formyl aldehyde **4** in an analogous fashion to the previously reported synthesis of acetaldehyde **5**, masking the aldehyde as a protected diol during solid-phase DNA synthesis [36,166]. The synthesis started with vinylation of a protected 7-iodo-7-deazaguanine derivative (**6**) using a Stille coupling reaction (Scheme 2). Oxidation of the allyl group to the diol, protection and functionalization yielded

phosphoramidite **9**. The aldehyde precursors were incorporated into complimentary 20-mer oligonucleotides in a 5'-d(GNC) sequence (the preferred sequence context for NM ICL formation [25,177,178]) by solid phase synthesis, and the oligonucleotides were deprotected and annealed.

Following oxidation of the diols using periodate, the two aldehydes were coupled by double reductive amination with hydrazine and $NaBH_3CN$ [36].



Scheme 2. Synthesis of the formyl aldehyde 4: a. vinyl-$Sn(Bu)_3$, $Pd[P(Ph)_3]_4$, toluene, 90°C, 43h, 70%; b. pyridine-2-carboxaldoxime, N,N,N',N'-tetra-methylguanidine, dioxane, DMF, rt, 42h, 88%; c. NaOMe, THF, rt, 5h, 97%; d. TBDMS-Cl, imidazole, DMF, rt, 16h, 93%; e. OsO_4, NMMO, THF, 0°C, 3.5h, 63%; f. Ac_2O, pyridine, rt, 1h, 82%; g. TBAF, AcOH, THF, rt, 20h, 79%; h. DMTr-Cl, pyridine, rt, 1h, 74%; i. $iPr_2NP(Cl)OC_2H_4CN$, DIEA, CH_2Cl_2, rt, 1h, 72%; j. oligonucleotide synthesis; k. 33% NH_4OH, 50°C, 12h; l. annealing; m. $NaIO_4$, rt, 12h.

The reaction along with appropriate controls was analyzed by denaturing PAGE. As already discussed, [36] ICL formation from two acetaldehyde precursors was successful with hydrazine, but not with ammonium acetate (Figure 2, lane 2 and 3). However, reaction of a duplex containing **4** and **5** with hydrazine led to formation of the desired 5-

atom ICL **2**, evidenced by a band with the same mobility as the previously analyzed crosslink (Figure 2, lanes 3 and 4). Although the yield of the 5 atom ICL was lower than that of the 6-atom ICL (~25% vs ~75%), we were able to isolate the product by gel purification and electroelution in amounts exceeding 100 nmol. For simplicity of analysis we also synthesized ICL **2** in an 11-mer duplex and unambiguously identified it as the desired product by ESI-MS (m/z calc. 6741.23, found 6741.6, Figure 4). If either aldehyde precursor was present only on one strand of the duplex, no significant amount of the slower migrating species were observed, further demonstrating the specificity of ICL formation (Figure 2, lanes 1 and 5).

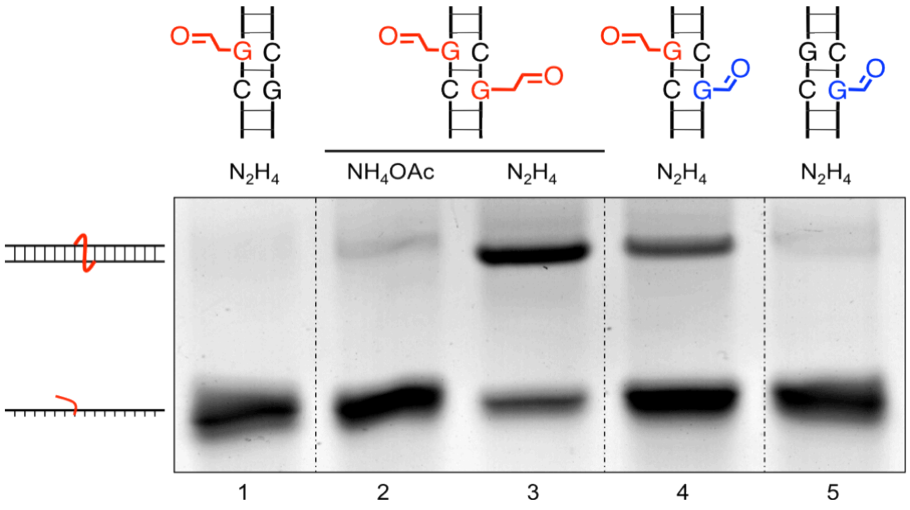


Figure 2. Analysis of ICL formation by denaturing PAGE with methylene blue staining. The duplexes, amine used and position of single-stranded or ICL-containing DNA are indicated. The sequences used were 5'-d(GTCACTGGTAXACAGCATTG) and 5'-d(CAATGCTXTCTACCAGTGAC) where X represents the modified G. The small amount of bands running as duplexes in lanes 2 and 5 are likely due to residual amount of duplex that was not denaturated during electrophoresis.

Molecular Modeling of the new substrate

Molecular modeling was used to validate ICL **2** as a model for NM **1** and to compare the structural consequences of the two ICLs with the uncrosslinked control (C) in identical 11-mer sequences. We used the Amber simulation package with atomic detail and explicit water (see experimental details) to validate that **2** gives similar amounts of distortion when compared to the NM ICL **1** and consistent differences when both are compared to C [179]. Two independent simulations of 50ns were run for each of the **1**, **2** and C systems. The duplex was stable throughout all six simulations.

The N7 to N7 distances of the NM ICL **1** overlapped well with the C7 to C7 distances of **2**; both are restricted as compared to C, which samples a broader range (Figure 3A). The decreased distance between the two crosslinked bases has a direct effect on the local distortion around the ICL but is also influenced by the sequence (Figure 3B and Table 1). One measure of distortion is to examine the buckling and propeller twist [180] of the two crosslinked base pairs (Table 1, Figure 5 for sequence). Since molecular mechanics force fields have limited accuracy, it is more important to note the trends rather than the specific values. It is well accepted that differences are more reliable in simulations than absolute values, due to cancellation of systematic error. The crosslinked guanine 16 that neighbors the large purine ring of adenine (residue 17) accommodates the crosslink by buckling of the 7:16 base pair (C: -3.5˚ ±0.2˚, **1:** -13.2˚ ±0.1˚ and **2:** -21.8˚ ±0.6˚). The crosslinked guanine 5 that neighbors the small pyrimidine ring of thymine 6 has more room to move and accommodates the ICL by an increased propeller twist of the 5:18 base pair (C: -13.2˚ ±0.1˚, **1:** -22.3˚ ±0.2˚ and **2:** -26.9˚ ±0.8˚). The addition of a covalent bridge between both strands of DNA nearly doubles residue to residue correlation in the central region of the duplex. Residues 5 and 16 have a correlation coefficient (r) of 0.29 ±0.04, 0.62 ±0.01 and 0.58 ±0.02 for C, **1** and **2** respectively (Figure 5). **1** and **2** have a substantial decrease in flexibility (Figure 3A) when compared to the uncrosslinked reference C, resulting in the formation of a slight kink in the duplex of both **1** and **2** to accommodate the ICL (Figure 3B).

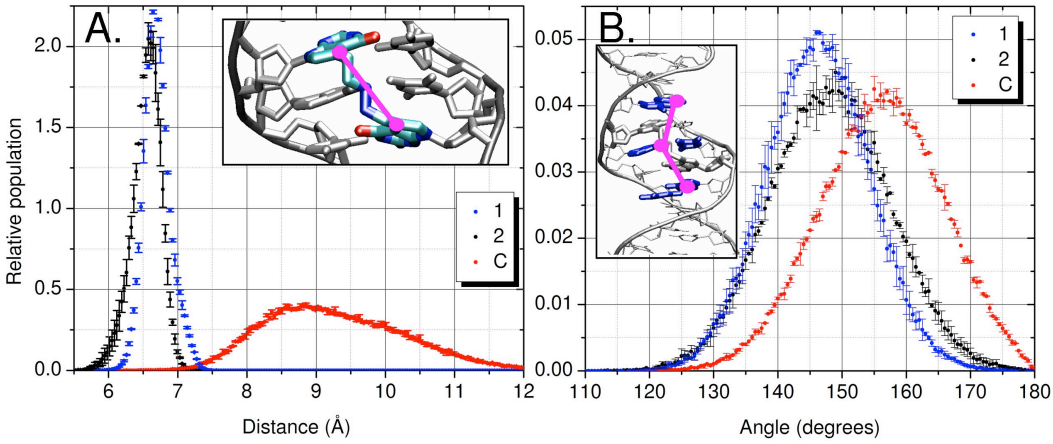


Figure 3. Data from MD simulations of 2 (black), 1 (blue) and C (red). A) Distance measurement marked in pink, C/N7 to C/N7 atoms between residues 5 and 16. B) Center of mass angle measurement, where each point is defined by the heavy atoms of the base pair (highlighted in blue in inset picture): point 1 represents 4A and 19T, point 2 represents 6T and 17A and point 3 represents 8A and 15T [181].

To test for local duplex bending at the crosslink site we measured an angle that would encompass the tightening of the 7 to 7 position of residues 5 and 16 (Figure 3A). The smaller angles sampled by **1** and **2** suggest that the analog slightly bends the duplex outside the crosslink in the same manner that the NM ICL simulations do (Figure 3B). The slight bending of the central region of the ICL simulations of **1** and **2** compared to C is qualitatively similar to past experimental work [28,165].

Circular dichroism (CD) spectra were recorded to gain experimental insight into to what extent the NM-ICLs containing oligonucleotide **2** deviates from B-form DNA. The CD spectra of **2** and a unmodified duplex of the same sequence displayed the characteristic features of B-form DNA (Figure 7), consistent with our molecular dynamics simulations indicating the NM ICL induces only a minor bend in the DNA. It has been shown that only more dramatic distortions, such as the ones induced by cisplatin ICL result in significant changes in CD spectra [43].

Discussion

We describe the synthesis of the stable NM ICL analog **2** by post-synthetic double reductive amination. Although this ICL has three atoms substituted compared to the NM ICL **1**, our molecular dynamics simulations show that the two ICLs affect DNA structure and motion in equivalent ways. In the absence of high-resolution experimental structure data for NM, our simulations are invaluable in probing the extent to which the analogs reflect NM properties. The availability of stable, site-specific NM ICLs will enable studies of the structural consequences and biological responses induced by NM ICLs, [168,182] more than sixty years after NMs were the first agents to be used in cancer chemotherapy [11]. Such studies will provide new insights into how tumors become resistant to treatment by crosslinking agents.

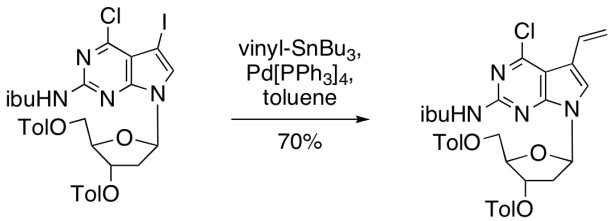
Experimental Details

General Information

6-chloro-7-deaza-7-iodo-N(2)-isobutyryl-3',5'-di-O-p-toluoyl-2'-deoxyguanosine (**6**) was prepared using published procedures [36] 4,4'-Dimethoxytrityl chloride was purchased from SynGen (USA). Protected 2'-deoxyribonucleoside-3'-phosphoramidites and all other reagents necessary for automated DNA synthesis, were purchased from Glen Research (USA). All other compounds and solvents were purchased from Sigma-Aldrich (USA), EMD (USA) or Alpha Aesar. Dry solvents (acetonitrile, dichloromethane, dioxane, DMF, pyridine, THF and toluene) were received and stored under inert gas in sealed bottles. NMR Spectra were recorded on Varian 300 MHz and 400 MHz spectrometers. HR-ESI MS of nucleosides were recorded on a LTQ Orbitrap XL (Thermo-Fisher Scientific). ESI-MS of oligonucleotides were recorded on a TSQ Quantum Access (Thermo Scientific).

Synthesis of phosphoramidite 9

6-Chloro-7-deaza-N(2)-isobutyryl-7-vinyl-3',5'-di-O-p-toluoyl-2'-deoxyguanosine (7)



6 7

A solution of 6-chloro-7-deaza-7-iodo-N(2)-isobutyryl-3',5'-di-O-p-toluoyl-2'-deoxyguanosine (**6**) (1.00 g, 1.39 mmol) and $Pd(PPh_3)_4$ (161 mg. 0.14 mmol) in toluene (20 ml) was treated with vinyl tributyl stannane (0.610 ml, 2.09 mmol) and stirred at 90°C for 43 hrs under argon atmosphere. The solution cooled down to RT and the excess of stannane was quenched with CsF for 30 minutes. The mixture was diluted with 25 ml CH_2Cl_2 and 10 ml H_2O and filtered over Celite. The organic phase was separated and dried with Na_2SO_4. The solvent was removed under reduced pressure and the solid residue was purified by silica gel column chromatography (dichloromethane/ethyl acetate 20:1) to give **7** (0.60 g, 0.97 mmol, 70%) as a white solid.

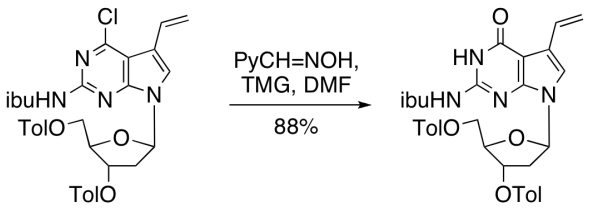
Rf = 0.60 (CH_2Cl_2:EtOAc 10:1).

^{1}H NMR ($CDCl_3$): δ 7.99 (s, 1H, C(2)-NH), 7.97 (d, J = 8.2, 2H, Tol-H), 7.92 (d, J = 8.2, 2H, Tol-H), 7.39 (s, 1H, C(8)-H), 7.27 (d, J = 8.1 Hz, 2H, Tol-H), 7.24 (d, J = 8.2 Hz, 2H, Tol-H), 7.05 (dd, J = 14.4, 8.8 Hz, 1H, C(7)CH), 6.73 (dd, J = 8.3, 5.9 Hz, 1H, C(1')-H), 5.77 (dt, J = 6.3, 2.3 Hz, 1H, C(3')-H), 5.25 (d, J = 23.6 Hz, 1H, H_a-C=CH), 5.10 (d, J = 15.6 Hz, 1H, H_b-CH=CH), 4.80 (dd, J = 11.6, 3.5 Hz, 1H, C(5')-H_a), 4.61 (dd, J = 11.6, 4.1 Hz, 1H, C(5')-H_b), 4.56 (ddd, J = 4.1, 3.4, 2.1 Hz, 1H, C(4')-H), 3.47 (dt, J = 6.5, 1.1 Hz, 2H, C(7)-CH_2), 3.01 (sept, J = 6.4 Hz, 1H, iBu-CH), 2.91 (ddd, J = 14.3, 8.2, 6.7 Hz, 1H, C(2')-H), 2.72 (ddd, J = 14.2, 5.9, 2.1 Hz, 1H, C(2')-H), 2.43 (s, 3H, Tol-CH_3), 2.42 (s, 3H, Tol-CH_3), 1.28 (d, J = 6.9 Hz, 6H, iBu-CH_3).

^{13}C NMR ($CDCl_3$): δ 175.8 (C), 166.4 (C), 166.3 (C), 152.7 (C), 152.6 (C), 151.3 (C), 144.7 (C), 144.4 (C), 130.0 (CH), 129.8 (CH), 129.6 (CH), 129.5 (CH), 127.0 (C), 126.7 (C), 121.2 (C), 116.3 (CH), 114.6 (CH_2), 112.2 (C), 84.5 (CH), 82.7 (CH), 75.4 (CH), 64.3 (CH_2), 38.3 (CH_2), 36.1 (CH), 29.9 (CH_2), 21.9 (CH_3), 21.9 (CH_3), 19.5 (CH_3), 19.4 (CH_3).

HR ESI-MS: $[M+H]^+$ m/z calculated for $C_{33}H_{34}ClN_4O_6$ 617.2161, found 617.2158

7-vinyl-7-deaza-N(2)-isobutyryl-3', 5'-di-O-p-toluoyl-2'-deoxyguanosine (10)



7 10

A solution of 7-vinyl-6-chloro-7-deaza-N(2)-isobutyryl-3',5'-di-O-p-toluoyl-2'-deoxyguanosine (**7**) (1.12 g, 1.81 mmol), pyridine-2-carboxaldoxime (1.11 g, 9.07 mmol) and 1,1,3,3-tetramethylguanidine (1.14 ml, 9.07 mmol) in a mixture of dioxane (20 ml) and DMF (21 ml) was stirred for 42 hrs under argon atmosphere at ambient temperature. The reaction mixture was diluted with ethyl acetate (130 ml) and washed with 5% HCl (130 ml), sat. $NaHCO_3$ (130 ml) and sat. NaCl (100 ml), dried over Na_2SO_4 and evaporated to dryness under reduced pressure. Purification of the residue by silica gel column chromatography (dichloromethane/ethyl acetate 20:1) gave **10** (0.95 g, 1.58 mmol, 88%) as a white solid.

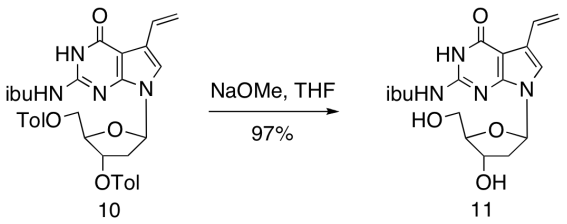
Rf = 0.28 (CH_2Cl_2:EtOAc 10:1)

^{1}H NMR ($CDCl_3$): δ 11.80 (s, 1H, N(1)-H), 9.59 (s, 1H, C(2)-NH), 7.85 (d, J = 8.0, 2H, Tol-H), 7.80 (d, J = 8.0 Hz, 2H, Tol-H), 7.16 (d, J = 8.0 Hz, 2H, Tol-H), 7.11 (d, J = 8.0 Hz, 2H, Tol-H), 6.82 (s, 1H, C(8)-H), 6.77 (dd, J = 11.8, 11.6 Hz, 1H, H-C=CH_2), 6.30 (dd, J = 7.2, 6.8 Hz, 1H, C(1')-H), 5.86 (dd, J = 17.8, 1.6 Hz, 1H, H_a-CH=CHC(7)), 5.66 (dt, J = 6.4, 3.2 Hz, 1H, C(3')-H), 5.08 (dd, J = 11.2, 1.2 Hz, 1H, H_b-CH=CHC(7)), 4.83 (dd, J = 12, 4.4, Hz, 1H, C(5')-H_a), 4.50 (dd, J = 12.0, 3.6 Hz, 1H, C(5')-H_b), 4.40 (dt, J = 4.0, 2.8 Hz, 1H, C(4')-H), 2.74 (ddd, J = 14.2, 7.2, 6.4 Hz, 1H, C(2')-H_a), 2.65 (sept, J = 6.8 Hz, 1H, iBu-CH), 2.50 (ddd, J = 14.1, 6.2, 2.6 Hz, 1H, C(2')-H_b), 2.34 (s, 3H, Tol-CH_3), 2.31 (s, 3H, Tol-CH_3), 1.16 (d, J = 6.8 Hz, 3H, iBu-CH_3), 1.11 (d, J = 6.8 Hz, 3H, iBu-CH_3).

^{13}C NMR ($CDCl_3$): δ 179.1 (C), 166.5 (C), 165.9 (C), 158.1 (C), 148.2 (C), 146.6 (C), 144.3 (C), 144.3 (C), 129.6 (CH), 129.6 (CH), 129.3 Tol (CH), 129.2 (CH), 127.8 Tol (C), 126.7 Tol (C), 126.4 (C), 120.2 (C), 116.9 (CH), 114.5 (CH_2), 103.2 (C), 84.3 (CH), 81.8 (CH), 74.9 (CH), 63.8 (CH_2), 37.4 (CH_2), 36.1 (CH), 21.64 (CH_3), 21.6 (CH_3), 19.0 (CH_3), 18.8 (CH_3).

HR ESI-MS: $[M+H]^+$ m/z calculated for $C_{33}H_{35}N_4O_7$ 599.2500, found 599.2495.

7-vinyl-7-deaza-N(2)-isobutyryl-2'-deoxyguanosine (11)



10 11

A solution of 7-vinyl-7-deaza-N(2)-isobutyryl-3',5'-di-O-p-toluoyl-2'-deoxy guanosine (**10**) (2.79 g, 4.55 mmol) in THF (85 ml) and methanol (1 ml) was treated with 1M NaOMe in MeOH (4.5 ml, 4.5 mmol) at 0°C. The ice bath was removed and the reaction mixture was stirred for 5 hrs at ambient temperature, neutralized with acetic acid (260 μl, 15 mmol) and evaporated to dryness under reduced pressure. Purification of the residue by silica gel column chromatography (dichloromethane/methanol 20:1) gave **11** (1.66 g, 4.41 mmol, 97%) as a white solid.

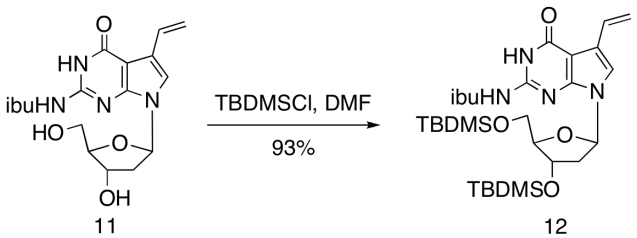
Rf = 0.20 (CH_2Cl_2:MeOH 10:1).

^{1}H NMR (d_4-Methanol): δ 7.35 (s, 1H, C(8)-H), 6.86 (dd, J = 18, 11.2 Hz, 1H, H-C=CH_2), 6.53 (dd, J = 6.8, 6.8 Hz, 1H, C(1')-H), 6.08 (d, J = 17.6 Hz, 1H, H_a-CH=CHC(7)) 5.14 , (J = 11.6 Hz, 1H, H_b-CH=CHC(7)), 4.49 (m, 1H, C(3')-H), 3.94 (m, J = 2.8, 1H, C(4')-H), 3.74 (m, 2H, C(5')-H_2), 2.75 (sept, J = 6.8 Hz, 1H, iBu-CH), 2.48 (ddd, J = 13.0, 8.6, 5.6 Hz, 1H, C(2')-H_a), 2.29 (ddd, J = 13.0, 5.7, 2.4 Hz, 1H, C(2')-H_b), 1.24 (d, J = 6.8 Hz, 6H, iBu-CH_3).

^{13}C NMR (d_4-Methanol): δ 180.6 (C), 160.2 (C), 150.5 (C), 148.2 (C), 129.5 (C), 121.2 (CH), 118.9 (CH), 114.7 (CH_2), 103.3 (C), 88.7 (CH), 84.6 (CH), 72.8 (CH), 63.5 (CH_2), 41.6 (CH_2), 37.1 (CH), 19.5 (CH_3), 19.5 (CH_3).

HR ESI-MS: $[M+H]^+$ m/z calculated for $C_{17}H_{23}N_4O_5$ 363.1663, found 363.1659.

7-vinyl-7-deaza-N(2)-isobutyryl-3',5'-di-(O-tert-butyl-dimethylsilyl)-2'-deoxyguanosine (12)



11 12

A solution of 7-vinyl-7-deaza-N(2)-isobutyryl-2'-deoxyguanosine (**11**) (1.10 g, 2.99 mmol) and imidazole (1.01 g, 14.95 mmol) was treated with a 50% solution of tert-butyl-chlorodimethylsilane in toluene (3.0 mL, 8.6 mmol) in DMF (10 ml) for 16 hrs under argon atmosphere at ambient temperature. The reaction mixture was diluted with ethyl acetate (80 ml) and washed with 5% HCl (80 ml), sat. $NaHCO_3$ (80 ml) and sat. NaCl (80 ml), dried over Na_2SO_4, and evaporated to dryness under reduced pressure. Purification of the residue by silica gel column chromatography (hexane/ethyl acetate 4:1) gave **12** (1.65 g, 2.79 mmol, 93%) as a white solid.

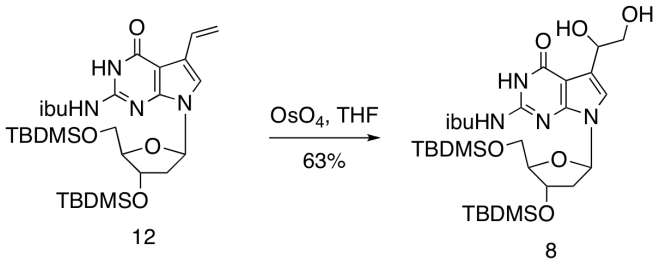
Rf = 0.89 (CH_2Cl_2:MeOH 10:1).

^{1}H NMR ($CDCl_3$): δ 11.56 (s, 1H, N(1)-H), 7.91 (s, 1H, C(2)-NH), 7.04 (s, 1H, C(8)-H), 6.89 (dd, J = 17.6, 11.2, Hz, 1H, H-C=CH_2), 6.41 (dd, J = 7.6, 6.0 Hz, 1H, C(1')-H), 6.11 (dd, J = 17.6, 1.6 Hz, 1H, H_a-CH=CHC(7)), 5.21 (dd, J = 11.2, 1.6, Hz, 1H, H_b- CH=CHC(7)), 4.54 (dt, J = 5.2, 2.8 Hz, 1H, C(3')-H), 3.93 (m, J = 2.8 Hz, 1H, C(4')-H), 3.73 (m, J = 4.0, 1.6 Hz, 2H, C(5')-H), 2.57 (sept, J = 6.8 Hz, 1H, iBu-CH), 2.36 (ddd, J = 13.2, 7.8, 6.0 Hz, 1H, C(2')-H), 2.22 (ddd, J = 13.2, 6.0, 3.2 Hz, 1H, C(2')-H), 1.28 (d, J = 6.8 Hz, 3H, iBu-CH_3), 1.27 (d, J = 6.8 Hz, 3H, iBu-CH_3), 0.92 (s, 9H, t-Bu-CH_3), 0.91 (s, 9H, t-Bu-CH_3), 0.10 (s, 6H, Si-CH_3), 0.08 (s, 6H, Si-CH_3).

^{13}C NMR ($CDCl_3$): δ 177.8 (C), 157.9 (C), 148.1 (C), 146.2 (C), 128.3 (C), 120.6 (CH), 116.5 (CH), 115.0 (CH_2), 103.3 (C), 87.7 (CH), 83.0 (CH), 72.6 (CH), 63.3 (CH_2), 41.1 (CH_2), 36.9 (CH), 26.0 (CH_3), 25.9 (CH_3), 19.2 (CH_3), 18.6 (CH_3), 18.4 (C), 18.3 (C), -4.4 (CH_3), -4.5 (CH_3), -5.1 (CH_3), -5.3 (CH_3).

HR ESI-MS: $[M+H]^+$ m/z calculated for $C_{29}H_{51}N_4O_5Si_2.H^+$ 591.3393, found 591.3389.

7-deaza-7-(1,2-dihydroxy-ethyl)-N(2)-isobutyryl-3',5'-di-(O-tert-butyl-dimethylsilyl)- 2'-deoxyguanosine (8)



To a solution of 7-vinyl-7-deaza-N(2)-isobutyryl-3',5'-di-(O-tert-butyl-dimethylsilyl)-2'-deoxyguanosine (**12**) (1.65 g, 2.79 mmol) in a mixture of THF (40 ml) and H_2O (4 ml) was added a 2.5% soln. of OsO_4 in t-BuOH (4.5 ml, 0.361 mmol) at 0°C. After addition of 4-methylmorpholine-4-oxide monohydrate (629 mg, 4.7 mmol) the reaction mixture was stirred for 3 ½ hrs at 0°C under argon atmosphere. The reaction mixture

was diluted with ethyl acetate (90 ml) and washed with 5% HCl (90 ml), sat. $NaHCO_3$ (90 ml), sat. Na_2SO_3 (18 ml) and sat. NaCl (90 ml), dried over Na_2SO_4, and evaporated to dryness under reduced pressure. Purification of the residue by silica gel column chromatography (dichloromethane/ethyl acetate 2:1) gave **8** (1.09 g, 1.74 mmol, 63%) as a mixture of diastereoisomers as a white solid.

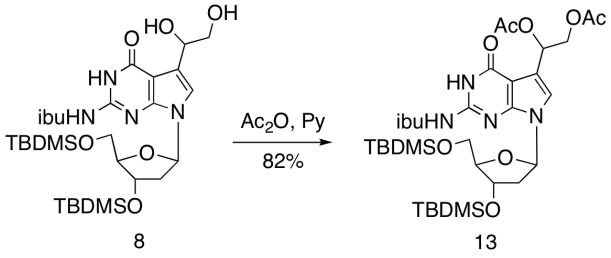
Rf = 0.49 (CH_2Cl_2:MeOH 10:1)

^{1}H NMR ($CDCl_3$): δ 12.01 (s, 1H, N(1)-H), 8.90 (s, 1H, C(2)-NH), 6.98&6.93 (2 peaks, 1H, C(8)-H), 6.41 (dd, J = 11.4, 6.4 Hz, 1H, C(1')-H), 4.81&4.75 (2 peaks, dd, J = 7.2, 4.4 Hz, 1H, CH(OH)), 4.50 (m, 1H, C(3')-H), 3.91 (m, 1H, C(4')-H), 3.79-3.70 (m, 4H, C(5')-H_2, CH_2(OH)), 2.60 (sept, J = 6.8 Hz, 1H, iBu-CH), 2.33-2.22 (m, 2H, C(2')-H_2), 1.24 (d, J = 7.2 Hz, 3H, iBu-CH_3), 1.22 (d, J = 7.2 Hz, 3H, iBu-CH_3), 0.90 (s, 9H, t-Bu-CH_3), 0.88 (s, 9H, t-Bu-CH_3), 0.08 (s, 6H, Si-CH_3), 0.07 (s, 6H, Si-CH_3).

^{13}C NMR ($CDCl_3$): δ 178.8 (C), 159.1&159.0 (C), 148.2&148.0 (C), 146.4 (C), 121.0&120.8 (CH), 115.9&115.7 (C), 103.9&103.7 (C), 87.7&87.6 (CH), 83.2&83.1 (CH), 72.5&72.2 (CH), 69.5&69.2 (CH), 68.2&67.9 (CH_2), 63.3&63.2 (CH_2), 41.6&41.3 (CH_2), 36.6 (CH), 26.1 (CH_3), 25.9 (CH_3), 19.2&19.1 (CH_3), 18.6&18.2 (CH_3), 18.0 (C), -4.5 (CH_3), -4.6 (CH_3), -5.2 (CH_3), -5.3 (CH_3).

HR ESI-MS: $[M+H]^+$ m/z calculated for $C_{29}H_{53}N_4O_7Si_2$ 625.3447, found 625.3441.

7-deaza-7-(1,2-diacetoxy-ethyl)-N(2)-isobutyryl-3',5'-di-(O-tert-butyl-dimethylsilyl)- 2'-deoxyguanosine (13)



A solution of 7-deaza-7-(1,2-dihydroxy-ethyl)-N(2)-isobutyryl-3',5'-di-(O-tert-butyl-dimethylsilyl)-2'-deoxy-guanosine (**8**) (1.09 g, 1.74 mmol) in pyridine (25 ml) was treated with acetic anhydride (7.4 ml, 78 mmol) at 0°C and stirred for 1 hr at this temperature. The mixture was quenched with H_2O (1.48 ml, 82 mmol) and evaporated to dryness under reduced pressure. The viscous residue was dissolved in ethyl acetate (70 ml) and washed with 5% HCl (70 ml), sat. $NaHCO_3$ (70 ml) and sat. NaCl (70 ml), dried over Na_2SO_4, and evaporated to dryness under reduced pressure. Purification by silica gel column chromatography (hexane/ethyl acetate 2:1) gave **13** (1.0 g, 1.41 mmol, 82%) as a white solid.

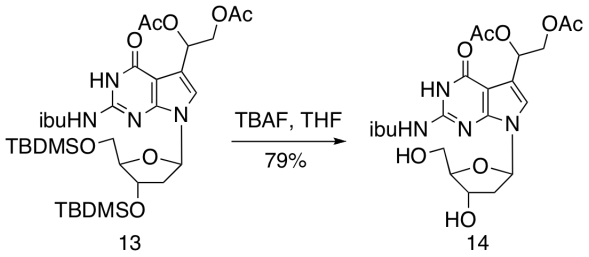
Rf = 0.82 (CH_2Cl_2:MeOH 10:1).

^{1}H NMR ($CDCl_3$): δ 11.90 & 11.66 (2 peaks, 1H, N(1)-H), 8.24-8.10 (4 peaks, 1H, C(2)-NH), 6.95, 6.92 & 6.90 (3 peaks, 1H, C(8)-H), 6.42-6.35 (m, 2H, C(1')-H & C(7)-CH), 4.63-4.57 (m, 1H, C(3')-H), 4.54-4.49 (m, 2H, AcO-CH_2), 3.93 (m, 1H, C(4')-H), 3.75-3.66 (m, 2H, C(5')-H_2), 2.57 (sept, J = 7.2 Hz, 1H, iBu-CH), 2.37-2.30 (m, 1H, C(2')-H_a), 2.22-2.18 (m, 1H, C(2')-H_b), 2.09 & 2.08 (2 peaks, 3H, Ac), 2.03 (s, 3H, Ac), 1.26 (d, J = 6.8 Hz, 3H, iBu-CH_3), 1.24 (d, J = 6.8 Hz, 3H, iBu-CH_3), 0.92 (s, 9H, t-Bu-CH_3), 0.91 (s, 9H, t-Bu-CH_3), 0.10-0.08 (12H, Si-CH_3).

^{13}C NMR ($CDCl_3$): δ 178.1 (C), 170.9 (C), 170.25 (C), 157.1 (C), 148.0 (C), 146.5 & 146.3 (C), 117.6 & 117.1 (CH), 116.3 & 116.1 (C), 103.2 (C), 87.7 (CH), 83.24 & 83.11 (CH), 72.8 (CH), 68.2 & 68.1 (CH), 65.8 (CH_2), 63.5 (CH_2), 41.0 & 40.9 (CH_2), 36.8 (CH), 26.2 (CH_3), 26.0 (CH_3), 21.4 (CH_3), 21.1 (CH_3), 19.21 (CH_3), 19.18 (CH_3), 18.6 (C), 18.2 (C), -4.4 (CH_3), -4.5 (CH_3), -5.0 (CH_3), -5.3 (CH_3).

HR ESI-MS: $[M-H]^-$ m/z calculated for $C_{33}H_{55}N_4O_9Si_2$ 707.3502, found 707.3501.

7-deaza-7-(1,2-diacetoxy-ethyl)-N(2)-isobutyryl-2'-deoxyguanosine (14)



A solution of 7-deaza-7-(1,2-diacetoxy-ethyl)-N(2)-isobutyryl-3',5'-di-(O-tert-butyl-dimethylsilyl)-2'-deoxyguanosine (**13**) (1.0 g, 1.41 mmol) and acetic acid (400 μl, 6.9 mmol) in THF (20 ml) was treated with a fresh 1M solution of $TBAF^-3H_2O$ in THF (7.0 ml, 7.0 mmol) at 0ºC. The mixture was then stirred for 20 hrs at ambient temperature. Evaporation to dryness under reduced pressure and purification of the residue by silica gel column chromatography (dichloromethane/methanol 20:1) gave **14** (534 mg, 1.11 mmol, 79%) as a white solid.

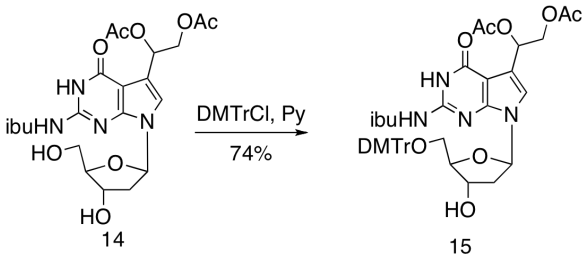
Rf = 0.32 (CH_2Cl_2:MeOH 10:1).

^{1}H NMR ($CDCl_3$): δ 8.90&8.87 (2 peaks, 1H, C(2)-NH), 6.88&6.86 (2 peaks, 1H, C(8)-H), 6.33 (dd, J = 7.6, 4.0 Hz, 1H, C(1')-H), 6.15 (dt, J = 14.4, 6.8 Hz, 1H, C(7)-CH), 4.61-4.50 (m, 3H, AcO-CH_2 and C(3')-H), 4.07 (s, 1H, C(4')-H), 3.50 (ddd, 48.4, 12.0, 2.8, 2H, C(5')-H_2), 2.68-2.63 (m, 2H, C(2')-H_2), 2.31-2.23 (m, 1H, iBu-CH), 2.08 (s, 3H, Ac), 2.03 (s, 3H, Ac), 1.24 (d, J = 7.2 Hz, 3H, iBu-CH_3), 1.22 (d, J = 7.2 Hz, 3H, iBu- CH_3).

^{13}C NMR ($CDCl_3$): δ 179.1 (C), 171.19 & 171.13 (C), 170.67 & 170.6 (C), 157.2 (C), 147.16 & 147.03 (C), 146.82 & 146.78 (C), 119.9 (CH), 115.6 (C), 104.2 (C), 87.83 & 87.79 (CH), 87.07 & 86.80 (CH), 72.8 (CH), 68.15 & 67.99 (CH_2), 65.6 (CH), 63.4 (CH_2) 40.7 (CH_2), 36.6 (CH), 21.3 (CH_3), 21.0 (CH_3), 19.23 & 19.18 (CH_3), 19.03 & 18.98 (CH_3).

HR ESI-MS: $[M-H]^-$ m/z calculated for $C_{21}H_{27}N_4O_9$ 479.1773, found 479.1766.

7-deaza-7-(1,2-diacetoxy-ethyl)-5'-O-(4,4'-dimethoxytrityl)-N(2)-isobutyryl-2'-deoxyguanosine (15)



A solution of 7-deaza-7-(1,2-diacetoxy-ethyl)-N(2)-isobutyryl-2'-deoxyguanosine (**14**) (150 mg, 0.312 mmol) in pyridine (8 ml) was treated with 4,4'-dimethoxytrityl chloride (127 mg, 0.36 mmol) and stirred for 90 min at room temperature under argon atmosphere. The reaction was quenched with methanol (80 μl) and concentrated under reduced pressure. The resulting oil was purified by silica gel column chromatography (dichloromethane/methanol 20:1, 0.5% Et_3N) gave **15** (180 mg, 0.23 mmol, 74%) as a white solid.

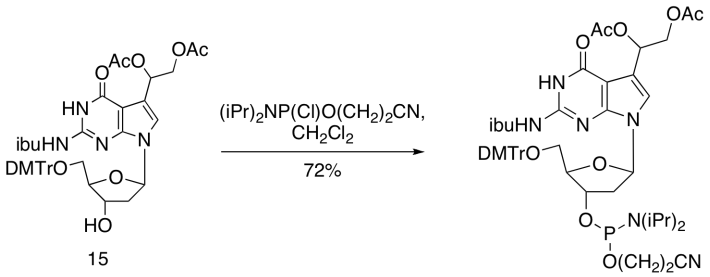
Rf = 0.35 (Hexane:EtOAc 1:5).

^{1}H NMR ($CDCl_3$): δ 7.38-7.36 (m, 2H, DMTr), 7.27-7.10 (m, 7H, DMTr), 6.98&6.95 (2 peaks, 1H, C(8)-H), 6.75-6.70 (m, 4H, DMTr), 6.53-6.46 (m, 1H, C(1')-H), 6.27 (dt, J = 14.4, 6.8 Hz, 1H, C(7)-CH), 4.50-4.39 (m, 3H, AcO-CH_2 and C(3')-H), 4.07&4.03 (2 peaks, 1H, C(4')-H), 3.69 (s, 6H, DMTr-OCH_3), 3.36-3.27 (m, 2H, C(5')-H_2), 2.38 (m, 1H, C(2')-H_a), 2.35 (sept, J = 7.2 Hz, 1H, iBu-CH), 2.31-2.27 (m, 1H, C(2')-H_b), 1.88&1.86 (2 s, 3H, Ac), 1.72&1.51 (2 s, 3H, Ac), 1.09 (m, 6H, iBu-CH_3).

^{13}C NMR ($CDCl_3$): δ 179.24 & 179.20 (C), 170.82 & 170.77 (C), 170.2 (C), 158.7 (C), 157.51 & 157.49 (C), 148.39 & 148.30 (C), 146.8 (C), 144.76 & 146.71 (C), 135.87 & 135.84 (C), 135.78 & 135.75 (C), 130.1 (CH), 128.22 & 128.01 (CH), 127.04 (CH), 117.2 & 117.1 (CH), 116.25 & 116.15 (C), 113.3 (CH), 102.7 (C), 86.5 (C), 85.5 & 85.4 (CH), 83.8 & 83.5 (CH), 72.33 & 72.07 (CH), 68.01 & 67.94 (CH_2), 65.7 (CH), 64.3 & 64.1 (CH_2), 55.3 (CH_3), 41.1 (CH_2), 36.2 (CH), 20.8 (CH_3), 20.7 (CH_3), 19.02 & 19.01 (CH_3), 18.96 & 18.89 (CH_3).

HR ESI-MS: $[M-H]^-$ m/z calculated for $C_{42}H_{45}N_4O_{11}$ 781.3079, found 781.3070.

7-deaza-7-(1,2-diacetoxy-ethyl)-5'-O-(4,4'-dimethoxytrityl-N(2)-isobutyryl)-2'-deoxyguanosine-3'-[(2-cyanoethyl)-N,N-diisopropylphosphoramidite] (9)



15 9

A solution of 7-deaza-7-(1,2-diacetoxy-ethyl)-5'-O-(4,4'-dimethoxytrityl)-N(2)-isobutyryl-2'-deoxyguanosine (**15**) (160 mg, 0.20 mmol) and N-ethyl-diisopropylamine (143 μl, 0.80 mmol) in dichloromethane (3.2 ml) was treated with 2-cyanoethyl-N,N-diisopropyl-chlorophosphoramidite (85 μl, 0.38 mmol) and stirred for 1 hr at room temperature under argon atmosphere. The reaction was quenched with methanol (0.3 ml) and concentrated under reduced pressure. Purification of the resulting oil by silica gel column chromatography (dichloromethane/methanol 20:1, 0.5% Et_3N) gave **9** (308 mg, 0.14 mmol, 72%) as a white solid.

Rf = 0.80 (Hexane:EtOAc 1:5).

^{1}H NMR ($CDCl_3$): δ 11.72 (s, 1H, N(1)-H), 8.58, 8.52, 8.14 (3 peaks, 1H, C(2)-NH), 7.47-7.42 (m, 2H, DMTr), 7.36-7.20 (m, 7H, DMTr), 6.99, 6.98, 6.93 (3 peaks, 1H, C(8)-H), 6.83-6.77 (m, 4H, DMTr), 6.38-6.35 (m, 2H, C(1')-H, AcO-CH), 4.60-4.47 (m, 2H, AcO-CH_2), 4.24 (m, 1H, C(3')-H), 4.17 (2 m, 1H, C(4')-H), 3.86-3.80 (m, 1H, iPr-CH), 3.77 (s, 6H, DMTr-OCH_3), 3.73-3.67 (m, 1H, iPr-CH), 3.62-3.53 (m, 2H, P-O-CH_2), 3.44-3.19 (m, 2H, C(5')-H_2), 2.72-2.66 (m, 1H, NC-CH_2), 2.57-2.50 (m, 1H, C(2')-H_a), 2.47 (t, J = 6.4 Hz, 1H, NC-CH_2), 2.41-2.30 (m, 1H, C(2')-H_b), 2.34 & 2.24 (2 sept, J = 6.9 Hz, 1H, iBu-CH), 1.88&1.78 (2 s, 3H, Ac), 1.71&1.64 (3 s, 3H, Ac), 1.20-1.06 (m, 18H, iBu-CH_3&iPr-CH_3).

^{31}P NMR ($CDCl_3$): δ 149.01, 148.98, 148.50, 148.48.

HR ESI-MS: $[M-H]^-$ m/z calculated for $C_{51}H_{62}N_6O_{12}P$ 981.4144, found 981.4148.

DNA synthesis & Purification

Phosphoramidite **9** was introduced at position X of the oligonucleotide **S1** with the sequence 5'-d(CGTA-X-TCATGC) and the phosphoramidite precursor of the acetaldehyde functionality (**16**)[36] was introduced at position X of the oligonucleotide **S2**, 5'-d(GCAT-X-ACTACG). Oligonucleotides synthesis was performed by automated DNA synthesis on an Expedite 8909 Nucleic Acid Synthesis System (Applied Biosystems) using 1μmol 1000Å CPG-dG and -dC column cartridges (Glen Research) and standard reagents and conditions. For **9** and **16** an extended coupling time of 15min was used. The trityl-on oligomers were deprotected by treatment with concentrated ammonia (50°C, 12h), then purified and detritylated using 1μmol TOP-cartridges (Varian) according to the manufacturers instructions. After lyophilization the pure oligonucleotides were resuspended in de-ionized sterile water and stored at -20°C. The concentrations of the oligonucleotide solutions were estimated by measuring their UV-absorption at 260nm wavelength. The incorporation and integrity of the protected diol functionality was verified by mass spectrometry.

ESI-MS: S1: m/z calcd 3391.1, found 3391.7; **S2**: m/z calcd 3414.2, found 3415.0

Formation of ICLs by reductive amination

A solution of the single strand oligonucleotides (25 nmol) in 125 μl 10 mM NaCl was heated to 95°C and allowed to cool to room temperature over a period of 4 hrs to allow for annealing of the sequences. After addition of 15 μl 1M sodium phosphate buffer (pH 5.4) and 10 μl 50 mM $NaIO_4$ the reaction mixture was kept in the dark overnight at 4°C. Excess $NaIO_4$ was removed by centrifugation through Microcon columns with a 3K cutoff (Millipore). The crosslink was formed through reductive amination by the addition of 10 μl 5 mM hydrazine solution in water and 10 μl 0.5 M $NaCNBH_3$. The reaction mixture was left overnight at room temperature in the dark. ICL formation was assessed by electrophoresis on a denaturing 20% polyacrylamide gel, containing 7M urea. The band containing the crosslinked oligonucleotide was excised from the gel and the DNA was extracted by electroelution using D-Tube™Dialyzer (Novagen) or the Elutrap™(Schleicher & Schuell) device. A small fraction was further purified for mass analysis by HPLC on a C18 column (Oligo-RP, 5 μm, 90Å, 10mm x 50mm, Clarity) using the following elution gradient: linear 2.5-25% B over 30 min, linear 25-90% B till 31 min, isocratic 90% B till 35 min, linear 90-2.5% B till 36 min, isocratic 2.5% B till 40 min; flow rate: 1 ml/min; eluent A: 0.1 M TEAA (pH=7); eluent B: CH_3CN. Under these conditions, the elution times of the starting oligonucleotides and the ICL products were 19.5 min and 21.5 min, respectively. The identity of the purified ICL products was confirmed by ESI-MS.

ESI-MS: crosslinked oligonuclotide **2**: m/z calcd 6741.2, found 6741.6 (Figure 4).

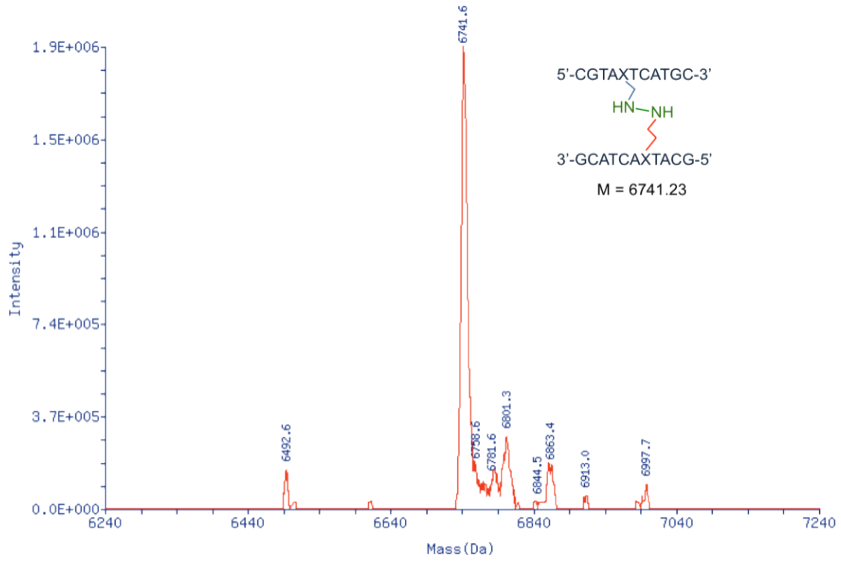


Figure 4: Mass spectrum of the crosslinked oligonucleotide (**2**). The structure and the calculated mass are indicated.

Parameterization Details

For the nitrogen mustard **1** and the analog **2** systems new parameters were generated for the crosslinked residues, in a manner previously undertaken in our laboratory [183]. The parameterization was accomplished in two steps: in step 1 the partial charges were generated by RESP fitting and in step 2, analogous atom type information was assigned from similar atom types in ff99SB with the parmbsc0 backbone additions and the GAFF forcefields to complete any unknown atom types for residues of **1** and **2** [184-186]. The partial charges and atom types on the phosphates and sugars were kept consistent with the ff99SB and parmbs0 parameters for DNA [186]

In stage 1, we created two guanine nucleotides in GaussView 3.0, bridging the two bases with the crosslink attached at the N7 position of each guanine. For **2** the N7 was replaced with a C7 for both residues (Figure 1). Three conformations for each system were generated where the dihedrals of NM and **2** were rotated 60 degrees about the C11-C12-N13-C14 and C12-N13-N14-C15 dihedrals, respectively, to obtain three different structures for RESP fitting. These structures were then optimized in Gaussian 03 using the HF/6-31G* basis set RESP fitting was then used to combine the Gaussian electrostatic surface potentials of each conformation into single point charges for the bases and crosslinks in the new systems [185]. For **2**, single point charges on the bases and crosslink were uniformly decreased after RESP fitting to make the net charge an integer (-2 for **2**, which includes the two phosphate groups, sugars, bases and

crosslink). For 1 an additional step was taken for generating the partial charges since the molecule is symmetrical. After the single point charges were generated for **1**, the accompanying symmetrical atoms were then averaged. Next the single point partial charges of the 1 were uniformly increased to bring the net charge to an integer (+1) for **1**, the two crosslinked residues, this including the phosphate groups, sugars, bases and crosslink). In some cases it was necessary to change the actual name of atom type while keeping the parameters to address any overlap in atom type parameters (discussed further below).

In the parameterization of the two crosslinks, special care was given in the assignment of atom types. When defining new bonds, angles, dihedrals and improper dihedrals in the frcmod, the parameters defined in the original forcefield for the normal nucleotides used can be overwritten [186]. To avoid this problem, several new atom types (CH, CL, HL, NL and HM) were defined to prevent the normal definitions from being overwritten.

Simulation Details

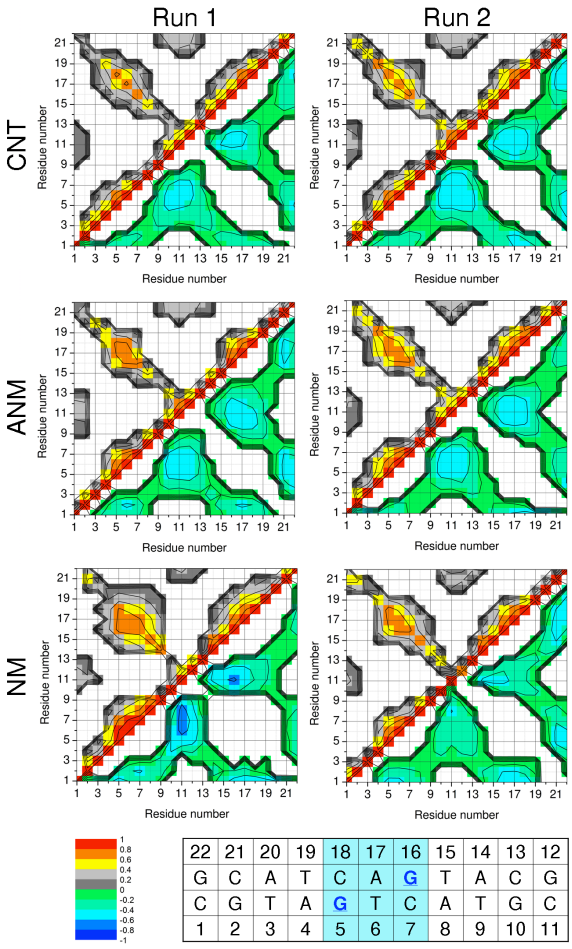
The Amber 9 suite of programs was used for all molecular mechanics calculations [179]. In this work three systems were simulated: the control (**C**), nitrogen mustard (**1**) and analog nitrogen mustard (**2**). The ff99SB forcefield with the parmbsc0 DNA backbone parameters was used for the majority of this work with additional GAFF parameters used on the crosslink to supplement regions that were not defined in the ff99SB forcefield [186]. The initial coordinates of each ICL system were generated in Nucgen using the identical sequence except for the crosslinked residues 5 and 16 (Figure 5). For each system, run-1 was built as A-form DNA and run-2 was built as B-form DNA to provide an independent measure of sampling convergence. For the crosslinked systems, after the core sequence was built in Nucgen, the atom names of residues 5 and 16 were then changed to reflect the new parameter atom names of the crosslinked residues [179]. Any atoms that were in guanine but not in the **1** or **2** systems were deleted. These structures were then loaded in Tleap and any missing atoms were added. After the systems were built the connectivity for the crosslink was verified through visual inspection in Xleap and changed to accurately represent Figure 5 [179]. The systems were then solvated with the TIP3P explicit water model in a truncated octahedron periodic box using an 8Å buffer containing ~9,000 water molecules [187]. After solvation the control was treated to a process that involved minimization, equilibration and then a production run. The two crosslinked systems were treated to two consecutive rounds of minimization, equilibration, and then unrestrained dynamics. The **1** and **2** systems were simulated in two stages since the initial structures may not have been compatible with the crosslink. To address this issue the **1** and **2** systems were created with incomplete crosslinks, where a single bond was missing on the alkyl linker, thereby giving two small separate lesions on residues 5 and 16 (see Figure 5 for sequence numbering). For

the first stage of the two crosslinked systems (**1** and **2**) the C12 to N13 bond on the crosslink was purposefully omitted from the initial structures, having an incomplete crosslink. These structures were minimized, equilibrated, and then simulated in unrestrained dynamics for 10ns. The simulations of these incomplete structures were necessary to allow the duplex to accommodate the crosslink without any initial major distortions (data not shown). For the second round of the ICL system process the 10ns trajectory from round one was analyzed and the C12 to N13 distance was measured. From this measurement, one structure towards the end of the 10ns unrestrained dynamics trajectory was extracted from each of the A/B form ICL systems (**1** and **2**) that had a low C12 to N13 distance (~3Å). The solvent was removed from these structures and the C12 to N13 bond was created. These structures were then resolvated, minimized, equilibrated, and simulated in unrestrained dynamics for 50ns. Further details of these steps are given below.

The two control systems were minimized for 1,000 steps at constant volume with a restraint force of 500 kcal/mol-$Å^2$, which was applied to all atoms of the duplex. After minimization, equilibration was performed under constant pressure, with a timestep of 2fs and with SHAKE employed to restrain the distances for bonds involving hydrogen [179,188]. The equilibration was executed in five stages, where the restraint force holding all atoms of the duplex decreased incrementally. At the first stage all atoms of the duplex were held with a restraint force of 10.0 kcal/mol-$Å^2$ and the temperature was slowly raised to 330K over 100ps. The temperature of 330K was chosen to reduce any energetic barriers in sampling. The next four stages of equilibration were then simulated at constant temperature with a restraint force of 5.0, 2.5, 1.0 and 0.5 kcal/mol $Å^2$ respectively at 100ps increments. After equilibration the **C** system was simulated for 50ns of unrestrained dynamics at constant pressure, with a timestep of 2fs and SHAKE on [188]. Temperature was held constant with the weak coupling algorithm with a coupling constant of 1 ps [189]. A time step of 0.002 ps was used.

The nitrogen mustard system (**1**) and analog nitrogen mustard system (**2**) were simulated in two stages. As mentioned above, the first stage contained an incomplete crosslink (missing the C12 to N13 bond). The second stage contained the complete crosslink. In stage 1, the system was minimized and equilibrated in the same manner as that described above for the control with the exception of the minimization, which was done in two parts. The first part was 1,000 steps with a restraint force of 500 kcal/mol $Å^2$ on all atoms of the duplex. The second minimization was 1,000 steps with a restraint force of 50 kcal/mol $Å^2$ on all atoms. After stage 1 equilibration, the system was simulated in unrestrained dynamics for 10ns at constant temperature (330K). For the 10ns simulation, the C12 to N13 distances were measured and a structure was extracted for each run. As mentioned above the structures that were extracted had a low C12 to N13 distance (~3 Å) and were extracted from a time point that was near the end of the 10ns simulation in order to allow the system to accommodate the incomplete lesion. Next, the waters from

these structures were removed in ptraj and were reloaded into Xleap where we created the final connectivity between the two atoms (making the C12 to N13 bond complete) to make the complete crosslinks [179] In stage 2, after the crosslinks were completed as shown in the top of Figure 6 the systems were solvated, minimized, equilibrated and ran a final production simulation in the same manner as stage 1 but the production simulation was extended to 50ns. The final 50ns of unrestrained dynamics generated was used for our final analysis (Figure 3, 5 and Table 1).



22	21	20	19	18	17	16	15	14	13	12
G	C	A	T	C	A	**G**	T	A	C	G
C	G	T	A	**G**	T	C	A	T	G	C
1	2	3	4	5	6	7	8	9	10	11

Figure 5. Correlation coefficient matrix plots of runs 1 and 2 (left and right columns, respectively) of uncrosslinked control (**C)**, ICL (**2)** and NM ICL (**1)** systems. The correlation coefficients were calculated using ptraj in the Amber package [179]. In these plots the correlation coefficient is the measure of each residue's movement relative to all others, as well as to itself. Each residue represents the heavy atoms of its base group. The positive and negative correlation coefficients (n) are separated into two regions. The upper left triangular region of each plot represents the positive correlation coefficients ($n \geq 0$) and the lower right triangular regions represent the negative coefficients ($n < 0$) of each plot. The bins on each plot represent the average correlation coefficient across each 50ns simulation for the two residues specified by the axes. For clarity, the respective residue numbers of the duplex sequence are shown at the bottom of the Figure. The sequence region highlighted in cyan represents the three base pairs that are highly involved in the crosslink, and the bases that are crosslinked in **1** and **2** are in blue font. From the correlation plots it is apparent that the positive correlation coefficient region where the crosslink was introduced (highlighted in blue in the sequence chart on the bottom of the figure) increases in correlation

for **1** and **2** when both are compared to the control. This observation implies that the introduction of an interstrand crosslink of the length and chemical makeup as observed in Figure 1 directly increases the amount of tandem movement of the residues around the crosslink.

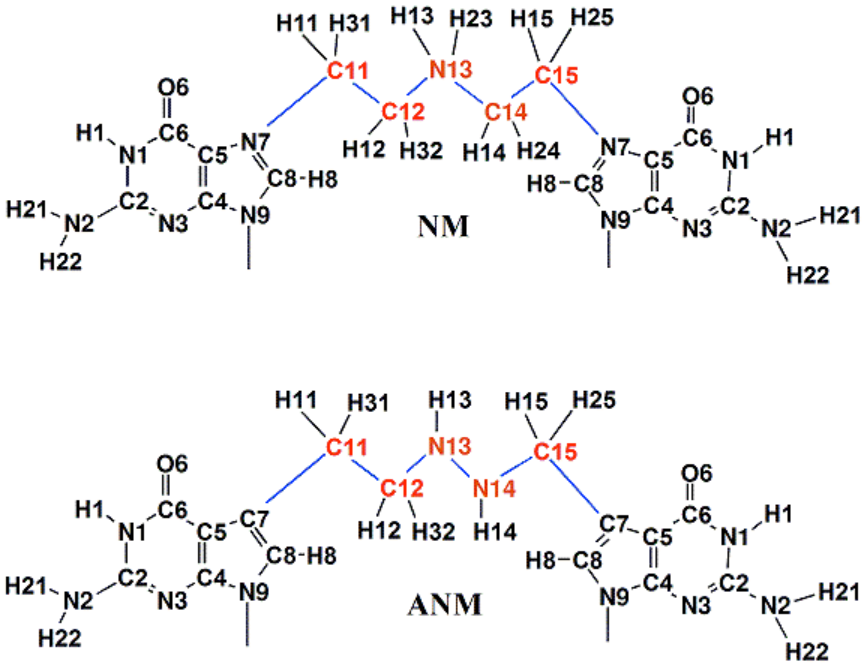


Figure 6. Atom names of **1** (top) and **2** (bottom). The heavy atoms of the crosslink are highlighted in red and their interconnecting bonds are in blue for clarity. These names match the names in the library files, which are provided as supplemental material.

CURVES analysis				
System	Base pair	Measurement	Average (degrees)	Uncertainty (degrees)
2	5:18	Buckle	1.9	±0.0
2	7:16	Buckle	-21.8	±0.6
2	5:18	Propeller twist	-26.9	±0.8
2	7:16	Propeller twist	-12.7	±0.6
C	5:18	Buckle	-1.1	±0.1
C	7:16	Buckle	-3.5	±0.2
C	5:18	Propeller twist	-13.2	±0.1
C	7:16	Propeller twist	-8.3	±0.0
1	5:18	Buckle	-2.5	±0.3
1	7:16	Buckle	-13.2	±0.1
1	5:18	Propeller twist	-22.3	±0.2
1	7:16	Propeller twist	-18.3	±0.1

Table 1: CURVES analysis measurement of the 50ns simulations for base pairs 5:18 and 7:16 of the **1**, **2** and **C** systems [180]. Corresponding residue numbers in the duplex can be seen at the bottom of Figure 5. From this data the 7:16 base pair buckling became more negative when a crosslink was introduced, while there is a buckle of ~ -3 degrees for the control. For the analysis of the 5:16 base pair it was found that the propeller twist of the two crosslinked systems deviated most from the control rather then the buckling. The main conclusion that could be drawn from these measurements is that interstrand crosslinks gave some degree of local distortion but the type of resulting distortion was dependent on other factors besides the crosslink introduced. We speculate that the local distortions such as buckling and propeller twist from the crosslinks in this work (Figure 1) are sequence dependent, and the positioning of a purine/pyrimidine between the 1,3 ICL (at residue 6 or 17, see Figure 5) may have a significant effect on the amount of distortion. The precision was calculated from the two separate simulations from each system.

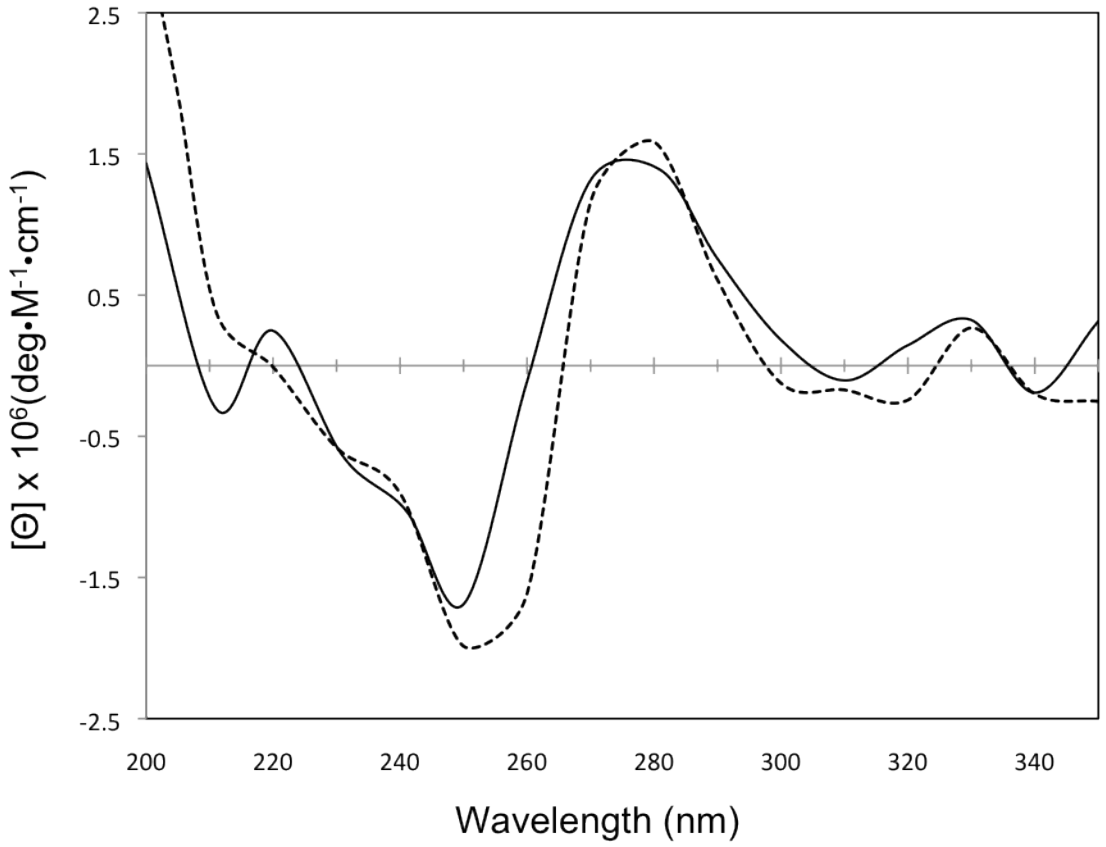
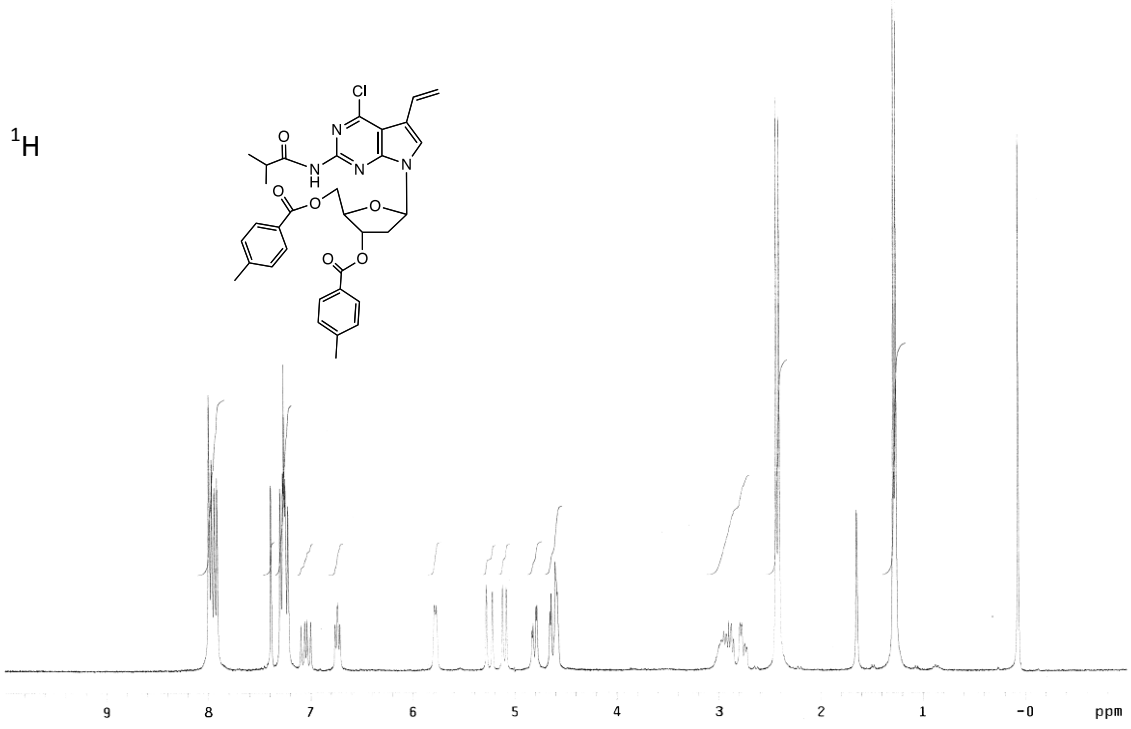


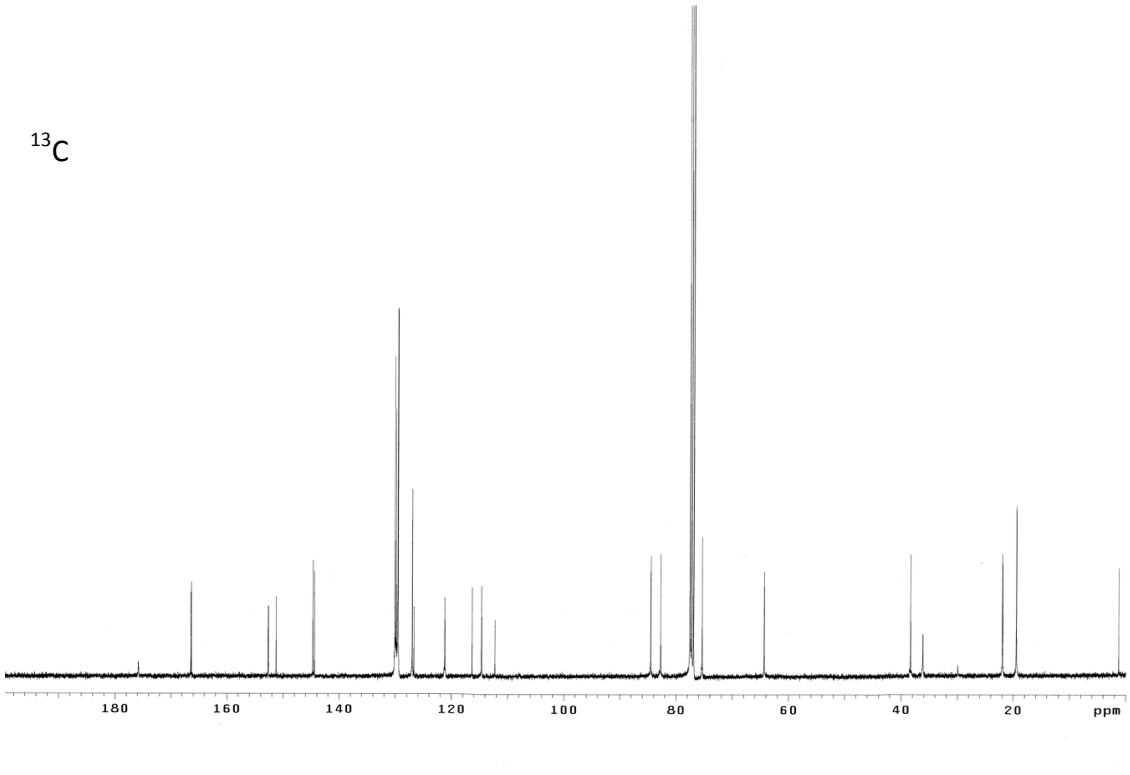
Figure 7. Circular dichroism spectra of the ICL-containing oligonucleotide 2 (dashed line) and the non-modified 11-mer control C (solid line). The samples were prepared in a 15 µM solution in a buffer containing 100 mM NaCl, 25 mM phosphate buffer (pH 7) and 0.5 mM EDTA. The measurements were performed at 10 ˚C and each spectrum is an average of three scans. Data were collected with a bandwidth of 1 nm and at a rate of 2 nm/s on an Applied Photophysics Chirascan CD spectrophotometer.

^{1}H and ^{13}C NMR Spectra of 7-vinyl-6-chloro-7-deaza-N(2)-isobutyryl-3',5'-di-O-p-toluoyl-2'-deoxyguanosine (8)

^{1}H

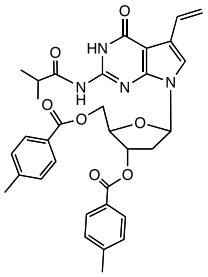


^{13}C

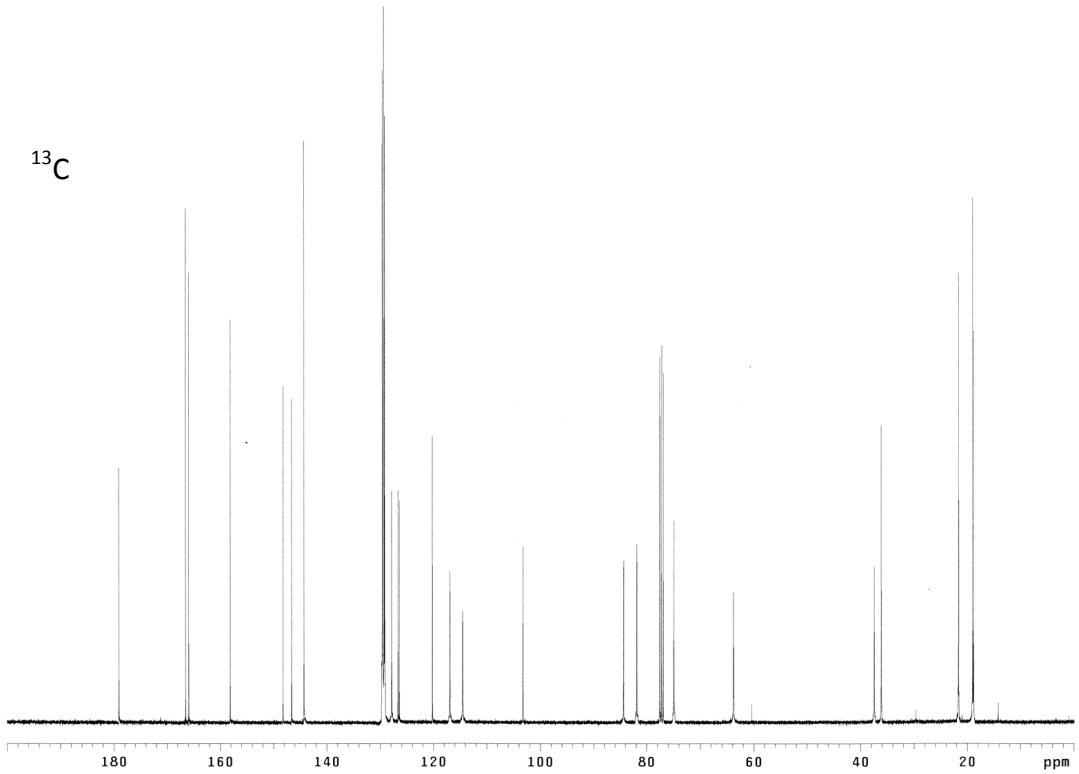


^{1}H and ^{13}C NMR Spectra of 7-vinyl-7-deaza-N(2)-isobutyryl-3', 5'-di-O-p-toluoyl-2'-deoxyguanosine (9)

^{1}H

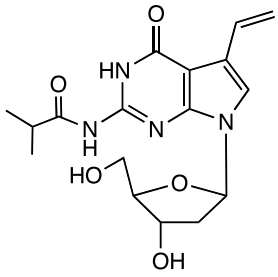


^{13}C

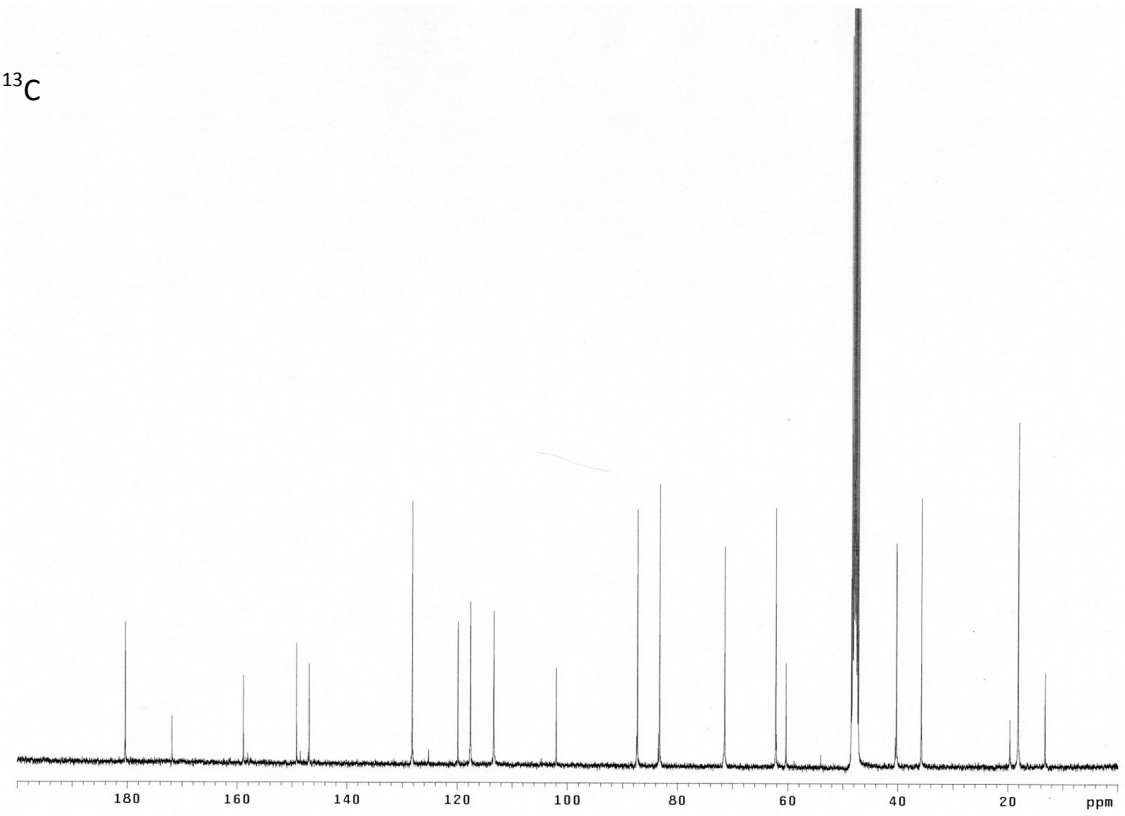


^{1}H and ^{13}C NMR Spectra of 7-vinyl-7-deaza-N(2)-isobutyryl-2'-deoxyguanosine (10)

^{1}H

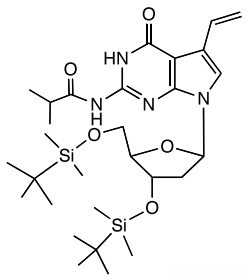


^{13}C

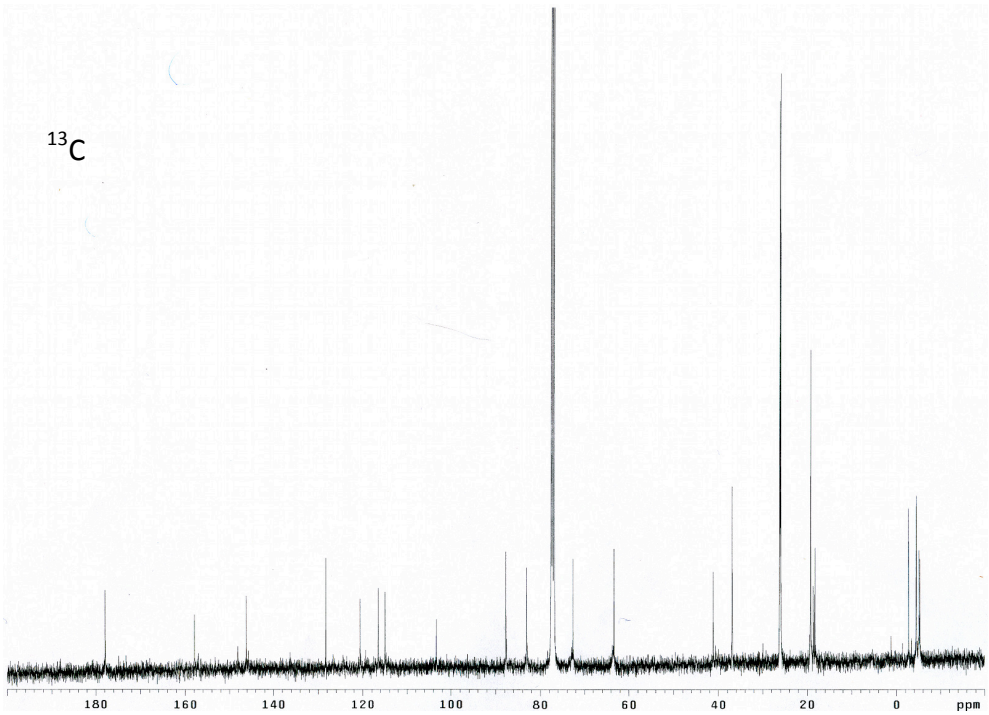


^{1}H and ^{13}C NMR Spectra of 7-vinyl-7-deaza-N(2)-isobutyryl-3',5'-di-(O-tert-butyl-dimethylsilyl)-2'-deoxyguanosine (11)

^{1}H

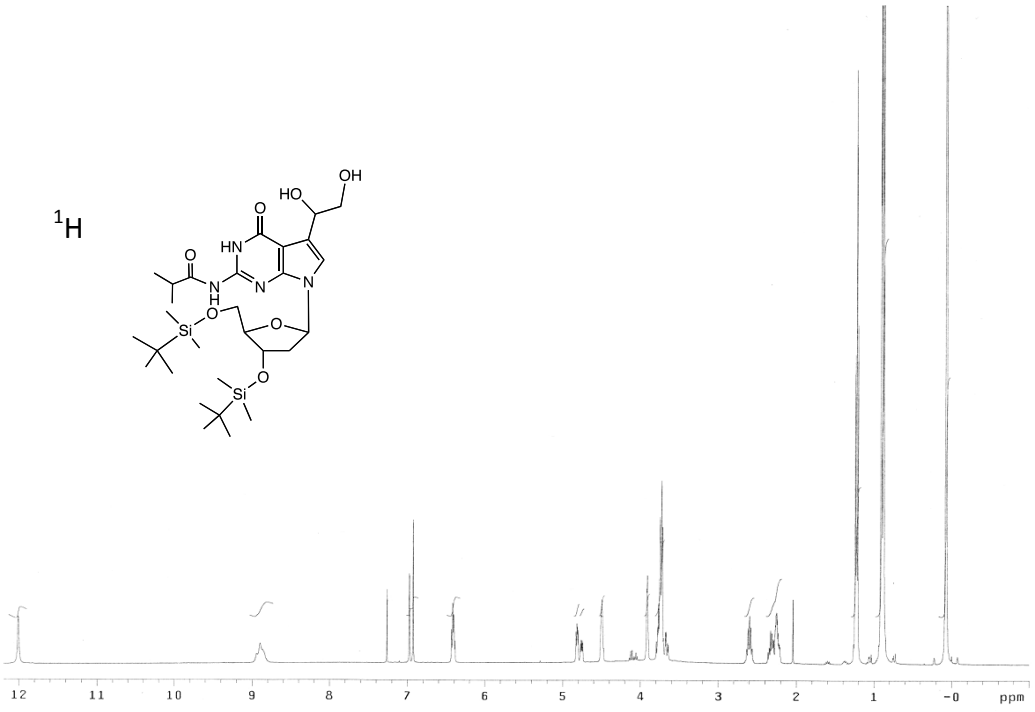


^{13}C

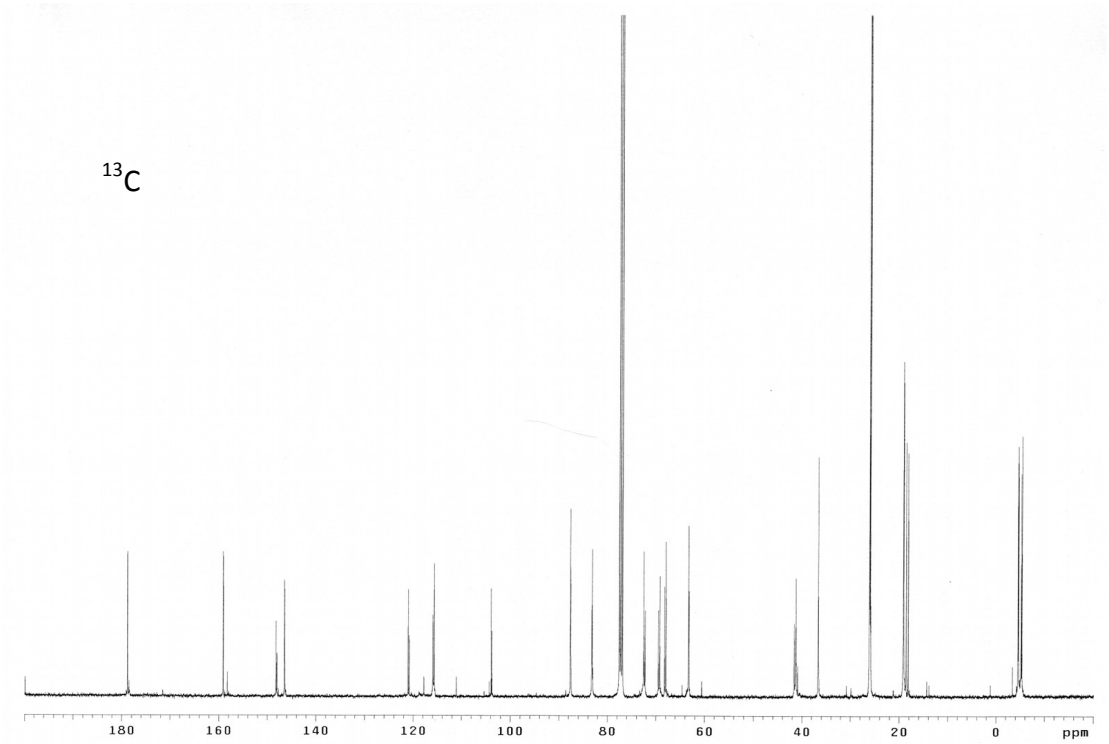


^{1}H and ^{13}C NMR Spectra of 7-deaza-7-(1,2-dihydroxy-ethyl)-N(2)-isobutyryl-3',5'-di-(O-tert-butyl-dimethylsilyl)- 2'-deoxyguanosine (12)

^{1}H

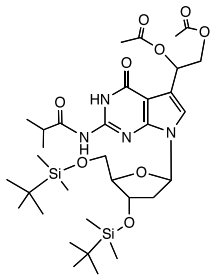


^{13}C

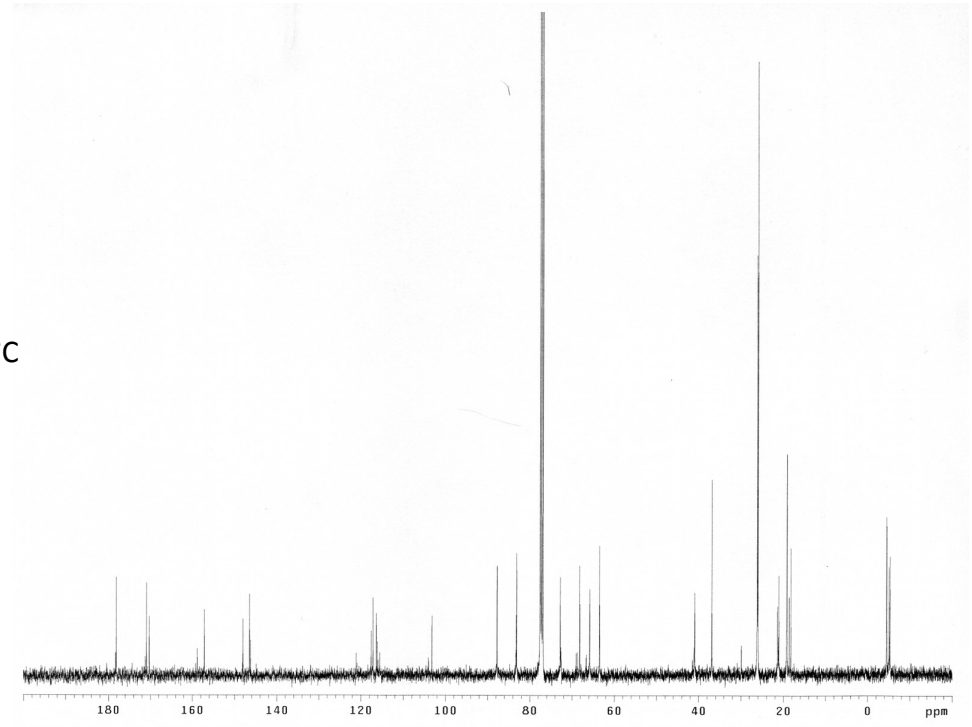


^{1}H and ^{13}C NMR Spectra of 7-deaza-7-(1,2-diacetoxy-ethyl)-N(2)-isobutyryl-3',5'-di-(O-tert-butyl-dimethylsilyl)- 2'-deoxyguanosine (13)

^{1}H

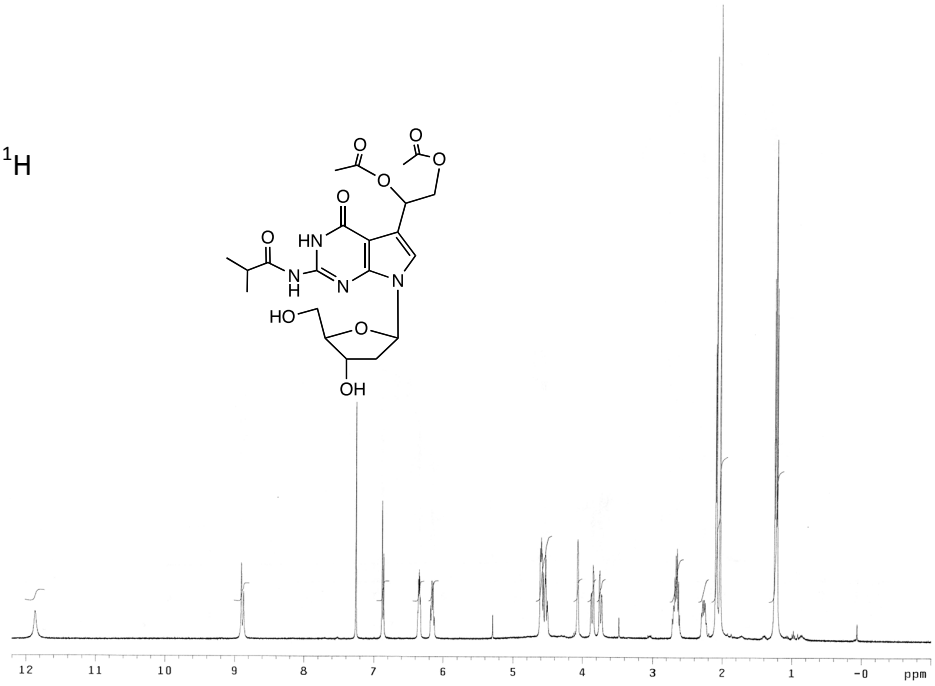


^{13}C

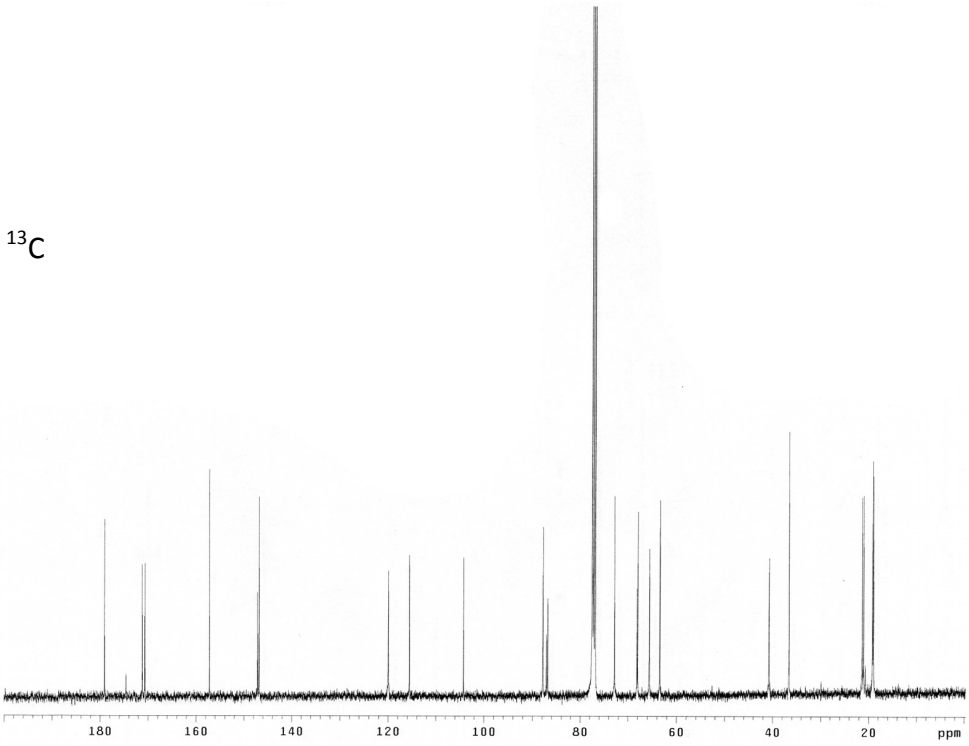


^{1}H and ^{13}C NMR Spectra of 7-deaza-7-(1,2-diacetoxy-ethyl)-N(2)-isobutyryl-2'-deoxyguanosine (14)

^{1}H

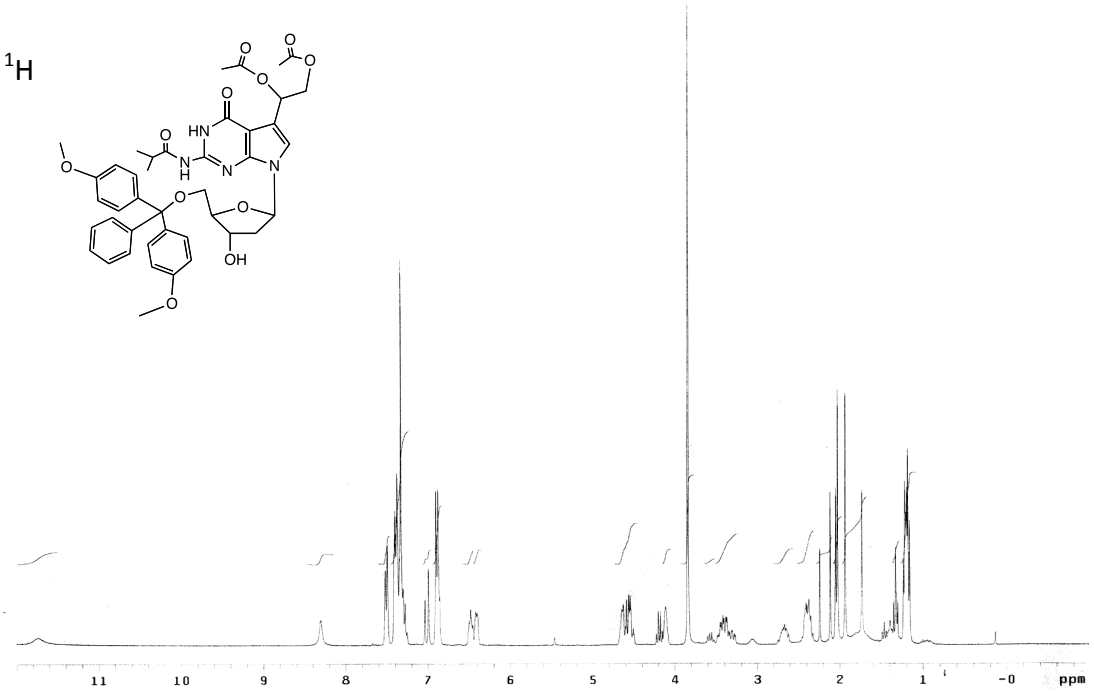


^{13}C

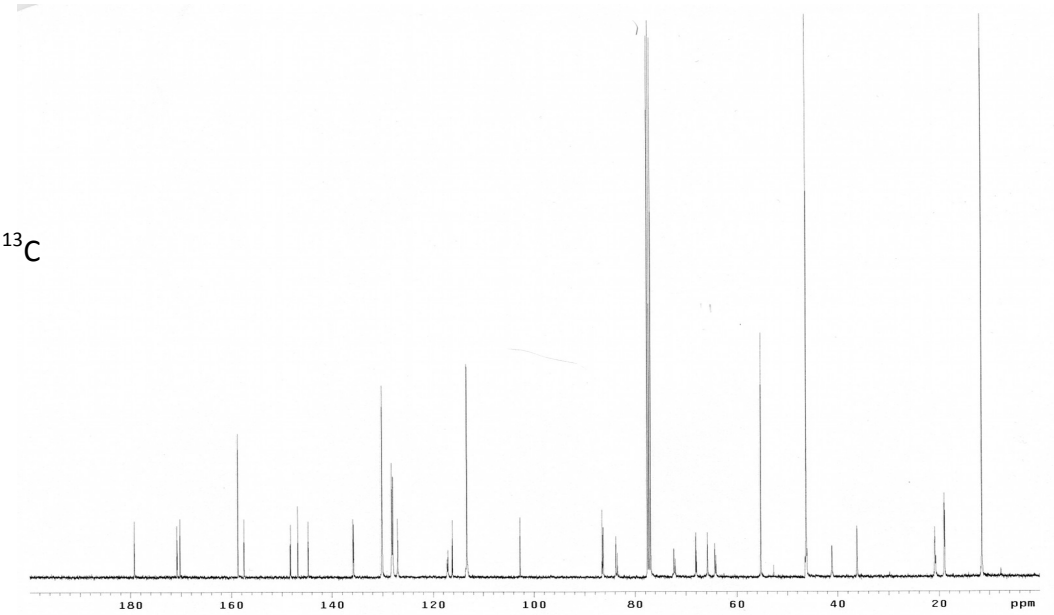


^{1}H and ^{13}C NMR Spectra of 7-deaza-7-(1,2-diacetoxy-ethyl)-5'-O-(4,4'-dimethoxytrityl)-N(2)-isobutyryl-2'-deoxyguanosine (15)

^{1}H

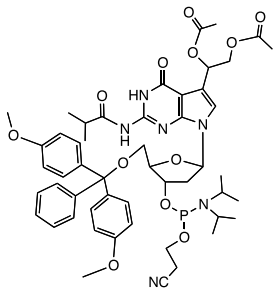


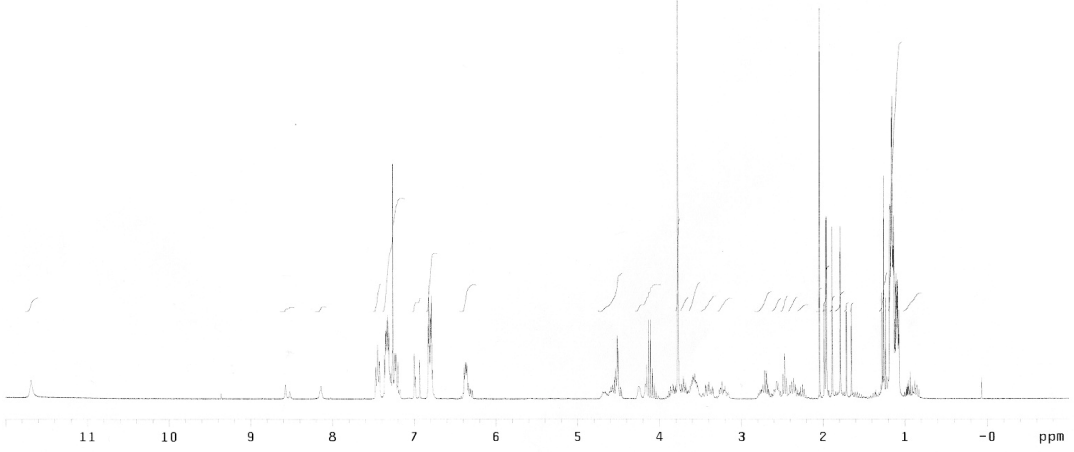
^{13}C



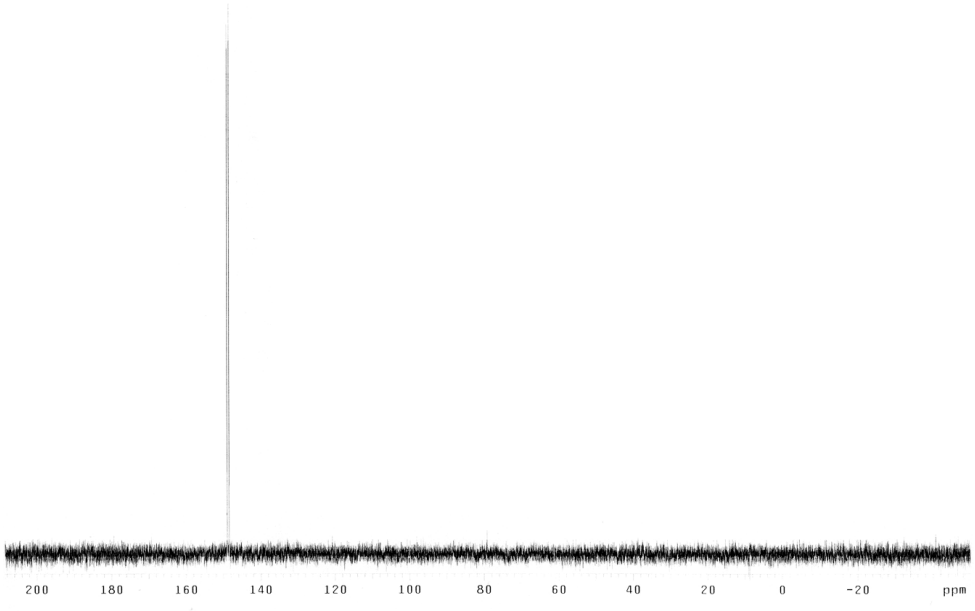
^{1}H and ^{31}P NMR Spectra of 7-deaza-7-(1,2-diacetoxy-ethyl)-5'-O-(4,4'-dimethoxytrityl-N(2)-isobutyryl)-2'-deoxyguanosine-3'-[(2-cyanoethyl)-N,N-diisopropylphosphoramidite] (4)

^{1}H

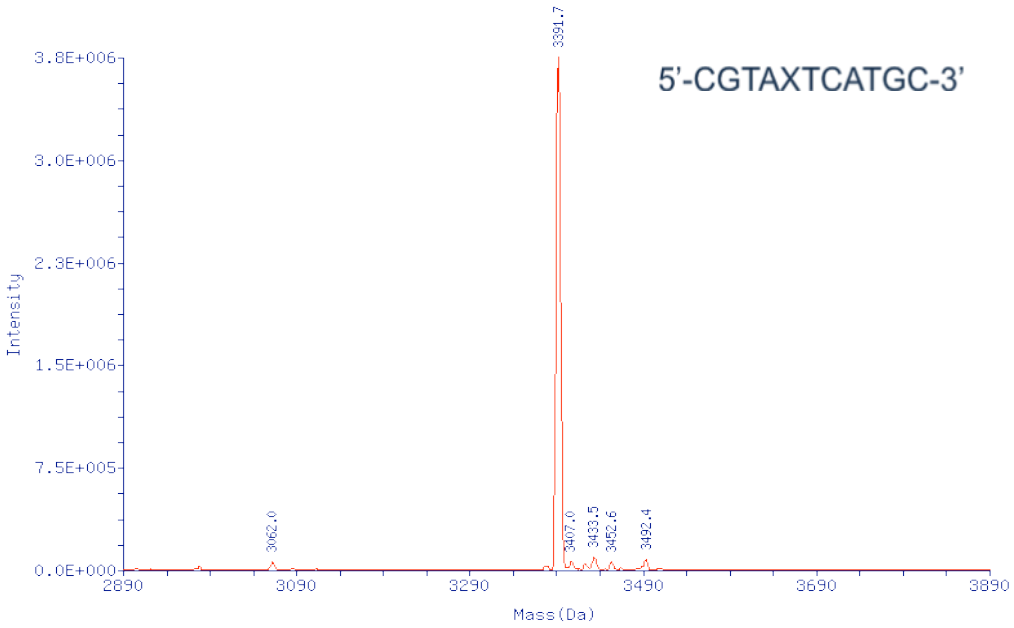




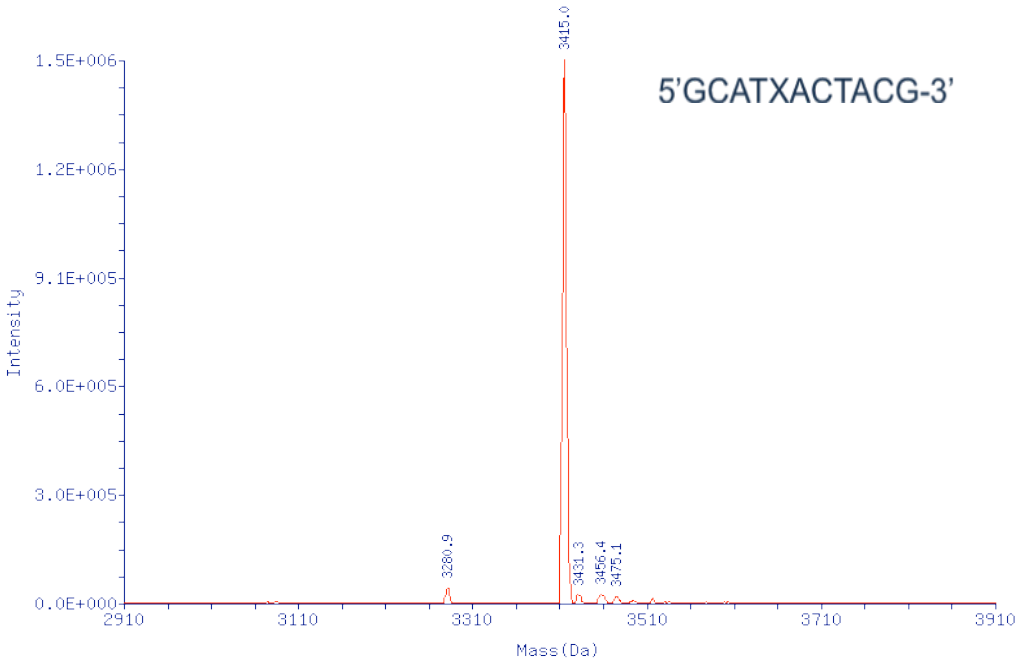
^{31}P



ESI-MS of diol oligonucleotide S1



ESI-MS of diol oligonucleotide S2

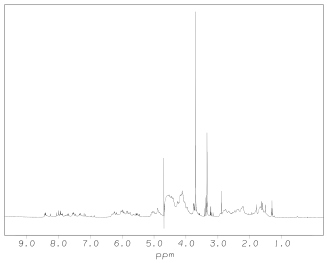


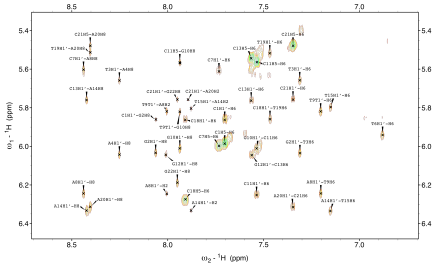
Chapter 4

NMR structural studies of a Nitrogen Mustard DNA Interstrand Crosslink Analogue

NMR Structural Studies of a Nitrogen Mustard Interstrand Crosslink Analogue

ABSTRACT





We used the previously synthesized interstrand crosslink to perform NMR structural studies. Thanks to our newly developed strategy we were able to synthesize the ICL in a suitable amount. The ICL was analyzed using 2D-NOESY, COSY and TOCSY experiments. We were not able to fully determine the 3-D structure of the oligonucleotide due to the inconclusiveness of the spectra.

Why a structural characterization?

The relationship between structure and function is a crucial principle of biochemistry. Different structural characteristics allow for different types of interactions between proteins, nucleic acids, membranes and all the biological entities inside and outside the cell. Very small structural differences can regulate a particular metabolic process like import/export through protein channels across the membrane. Small molecules, including drugs, often rely on a very tailored design to properly perform their task. On the other hand, proteins that need to remove and metabolize those drugs also rely on accurate structural features to complete their job.

A very compelling example for the importance of the structure-function relationship in DNA repair is the recognition of the lesions by the NER recognition factor XPC [190]. The canonical structure of DNA is often perturbed by lesions. Since XPC has to identify a considerable variety of DNA adducts, it recognizes structural alterations of the DNA rather than the chemical features of the adduct itself, which can have various shapes and different chemical properties.

Interstrand crosslinks (ICLs) are mainly recognized by stalled replication forks and the exact structure of the lesion is therefore less important for the recognition step (see Chapter 1). Nonetheless, in later steps of the repair the structure of the ICL is indeed important, since interactions with specific proteins such as translesion synthesis polymerases are unavoidable. The structure of an ICL lesion is critical in the minor repair pathway where the lesion is probably recognized by XPC and subsequently also involves translesion synthesis polymerases. Structural studies of the XPC yeast orthologue Rad4 have shown that the protein recognizes the distortion on the DNA and the loss of base stacking that provokes a local thermodynamic destabilization [139]. Additionally, genetic experiments in yeast have shown an increased sensitivity of cells depleted of Rad4 to ICL forming agents [191]. Taken together, these data suggest that XPC can recognize ICLs due to the perturbation of the canonical B-form DNA.

More clear evidences of the importance of the ICL structure were revealed in later steps of the repair pathways. Particularly, the unhooking of the lesion seems to be

dependent on the distortion provoked by the ICL [53] and the efficiency of the bypass of the lesion during TLS vary depending on the type of crosslink [74,127,192].

Knowing the exact structure of the lesion and how it affects the overall shape of the DNA can help to shed light on the relationship between structural perturbation and efficiency of repair. In the literature, several example of structural characterization of ICLs can be found. An oligonucleotide containing a cisplatin adduct was solved by NMR spectroscopy [41] and x-ray crystallography [42], showing a quite severe distortion of the double helix due to an extrahelical conformation of the two C residues pairing the crosslinked Gs, while psoralen ICLs [67-69] revealed only locally constrains and distort the DNA at the site of the intercalation without inducing overall distortion or bend in the DNA (see Chapter 1 for more details).

Due to the inherent instability of nitrogen mustard ICLs, a detailed structural analysis has not yet been achieved. In the past, biochemical experiments have provided important hints on the extent of DNA binding by this lesion [28] but the molecular details have remained obscure. The availability of a stable, well-defined analogue of NM ICLs and the possibility of producing the required amount needed for structural studies, gave us the opportunity to finally have a much clearer picture of how this ICL affects the structure of DNA. In addition, having performed extensive molecular dynamic simulations of both the NM ICL analogue and its natural counterpart, we have a good indication on how the two structures differ from one other. A complete structural characterization of the analogue lesion could validate the molecular dynamic studies and indirectly provide accurate structural insights into the original NM ICL.

NMR Spectroscopy

Structural biology started with the discovery that X-rays could be used to determine the structure of very small biological entities and its impact has increased exponentially in the past decades, making a tremendous contribution in the understanding molecular biology and biochemistry. Although X-rays crystallography

remains the main source of three-dimensional information with 86% of the coordinates deposited in the Protein Data Bank, about 13% of the almost 63,000 total structures were resolved using NMR spectroscopy. Both techniques have advantages and disadvantages mainly depending on the properties of the samples under analysis. In general is rather difficult to obtain crystals with a good diffraction of short oligonucleotides, mostly due to the intrinsic "polymorphic" nature of nucleic acids that lacks a globular shape and have a more undifferentiated molecular structure [193]. The structure of damaged DNA has therefore generally been solved using established ^{1}H-NMR techniques that usually provide the necessary information to determine the perturbations caused by lesions. Most exchangeable and non-exchangeable protons are assigned from the analysis of 2D-NOESY spectra and confirmed by analysis of COSY, DQFCOSY and TOCSY data [194,195]. The Watson-Crick alignment is evaluated by the chemical shift of thymine and guanine imino protons in 10% D_2O. In the case of our ICL-containing oligonucleotide we should also be able to identify the amino protons of the NM bridge. COSY and TOCSY experiments should be able to provide information on the protons of the bridge and their interaction with the modified guanines.

Results

The synthesis of the substrate to be analyzed by NMR was performed as described in Chapter 3 with few adjustments to perform the reaction in a higher scale (see experimental details). The ICL was incorporated into a sequence similar to the one that was used in previous studies by the group of Dr. De Los Santos [196] (Sequence 1).

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	
5'-	C	G	T	A	**X**	T	C	A	T	G	C	
	G	C	A	T	C	A	**X**	T	A	C	G	-5'
	22	*21*	*20*	*19*	*18*	*17*	*16*	*15*	*14*	*13*	*12*	

Sequence 1: Sequence of the duplex that was used for the NMR experiments. **X** represents the modified G

1D ^{1}H-NMR: The sample was first subject to one-dimensional ^{1}H-NMR varying the temperature from 20 ˚C to 35 ˚C with 5 ˚C intervals to establish the temperature at which the duplex showed the best behavior. The experiment at 35 ˚C displayed the best resolution, particularly for the peaks that we later identify to be T6(H-6) and G5(H-8) that were shifted upfield compared to the other H-6 and H-8 (Figure 1).

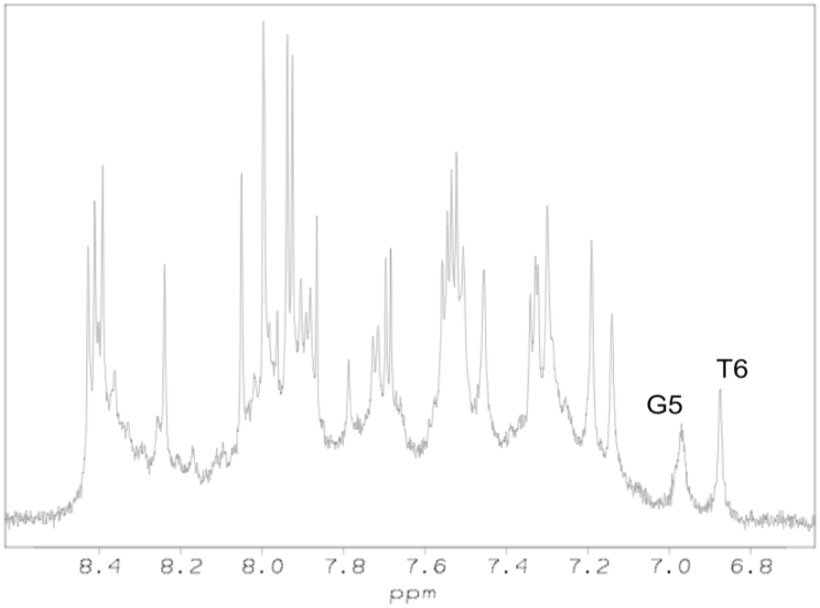


Figure 1: Base region of a one-dimensional ^{1}H-NMR spectrum recorded in 100% D_2O buffer at 35 ˚C. The signal of T6(H-6) and G5(H-8) are labeled.

Interestingly also the methyl group of T6 is shifted upfield compared to the other methyl groups (not shown).

2D-NOESY: Two-dimensional NOESY experiments allow for the assignment of proton signals. Particularly, analysis of the off-diagonal element reveals interaction between protons separated by less than 5 Å. We looked at the so-called “fingerprint region” to examine the interaction between base protons. In this region, in a canonical B-form DNA, the interactions between H-8 of purines or H-6 of pyrimidine and the H-1’ sugar protons should give a recognizable peak. Each base (H-6 or H-8) should have an interaction with its own sugar proton and a second interaction with the H1’ proton of

the residue positioned 5’ to itself (Figure 2). In this way in a canonical B-form DNA it is possible to “walk” through the entire strand.

In our spectrum we were able to assign all the peaks in the region but unfortunately we could not “walk” through both strands uninterruptedly. On both strands the region around the crosslink was disturbed in a way that some interaction were not strong enough to give a clear peak in the NOESY spectrum. On one strand we were unable to identify the interaction between G5 and A4 and also not the one between T6 and C7 (Figure 3).

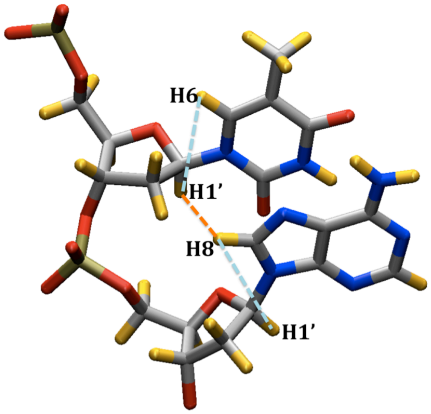


Figure 2: Interactions seen in the NOESY “fingerprint” region. H8 of purines (or H6 of pyrimidine) and the H1’ proton of the subsequent base (orange) and the intraresidual interaction (light blue).

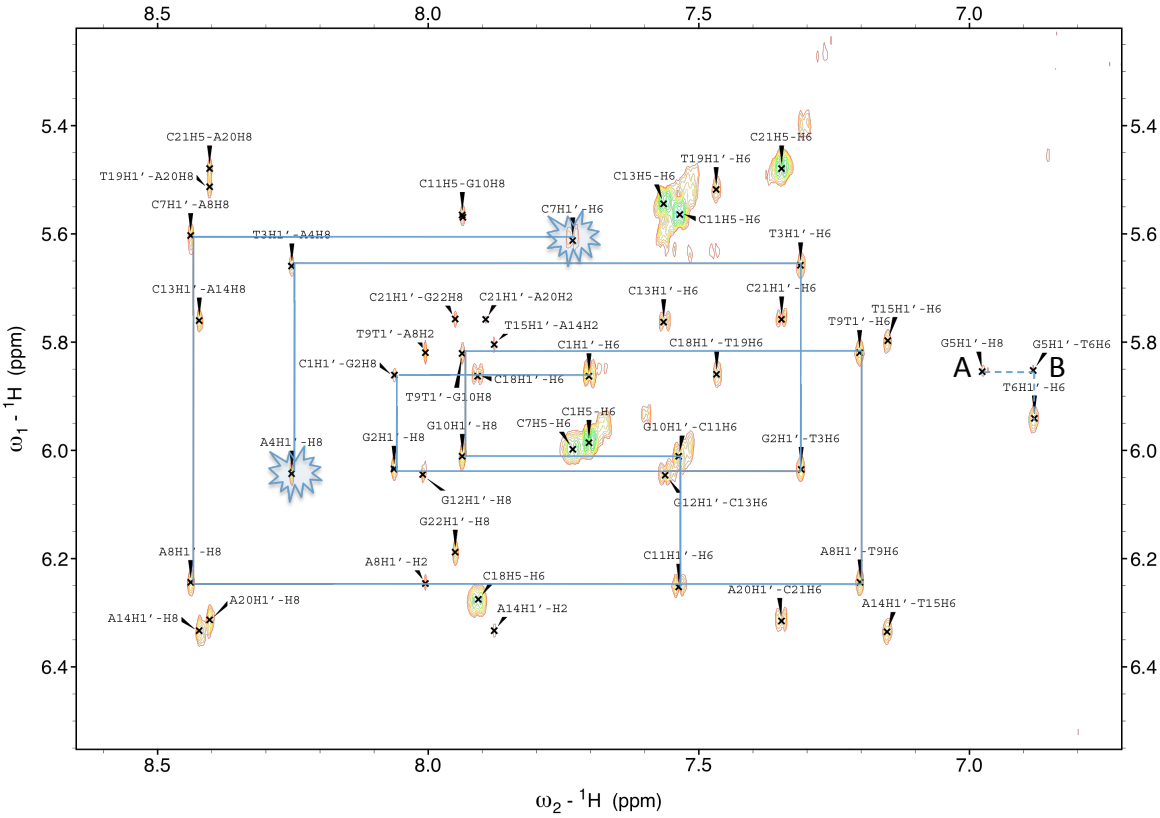


Figure 3: Fingerprint region of the 2D-NOESY spectrum recorded in 100% D_2O buffer at 35 ˚C showing the walk along the strand containing residue 1 to 11. The walk gets interrupted at the crosspeak (A4H1’-H8) and starts again at C7(H1’-H6).

Moreover, the crosspeak of the interaction between G5(H-1') and T6(H-6) was very weak (at the right center of Figure 3, labeled A). Even the peak representing the interaction between G5(H-1') and G5(H-8) does not have an intensity observed for the other Gs in the sequence (labeld B). The interaction between T6(H-1') and C7(H-6) is missing altogether. On the other strand we observe a similar effect. We identified all the interactions from G12 to T15 and from C18 to G22 but we were missing the ones between of G16 and A17 (Figure 4).

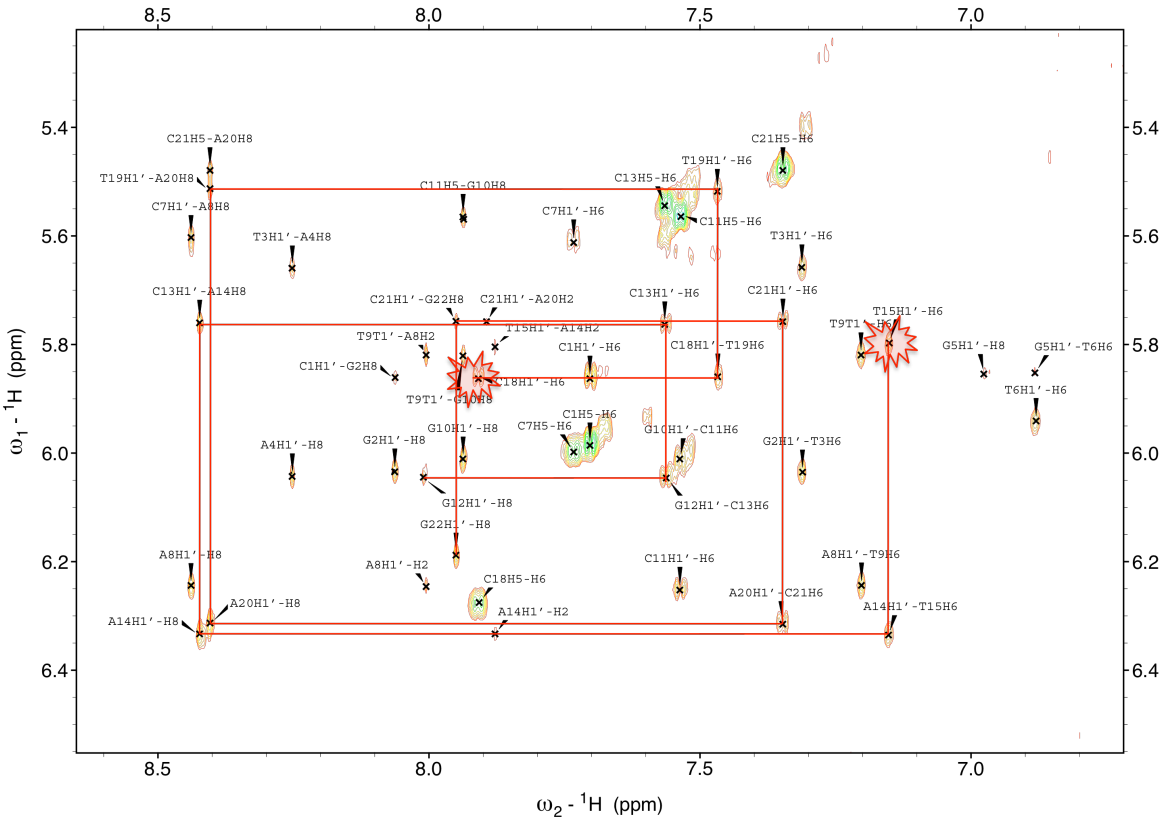


Figure 4: Fingerprint region of the 2D-NOESY spectrum recorded in 100% D_2O buffer at 35 ˚C showing the walk along the strand containing residue 12 to 22. The walk gets interrupted at the crosspeak T15(H1'-H6) and starts again at C18(H1'-H6).

Another interesting finding was the unusual downfield shift of the interaction peaks (H5-H6) of both C7 and C18; their chemical shift is similar to the terminal Cs that are more exposed to the solvent. An upfield shift is also observed for the methyl group of T6 as already observed in the 1D spectrum.

2D-COSY: unfortunately the COSY experiments did not produce any insightful spectra. The signal in general was weak and many cross peaks were missing likely either due to the low concentration of the sample or the broadening of the signals. We expected to get some information on the through-bond interactions between the protons of the bridge and the two H8 protons of the crosslinked guanosines G5 and G16. Unfortunately we were not able to identify any signals that could correspond to either one of those interactions. In the case of G16 we were not even able to identify the intraresidual interactions.

TOCSY: The TOCSY measurements confirmed what we have seen on the NOESY and COSY experiments not providing any additional information.

1D ^{1}H-NMR in 10% D_2O: We lyophilized the sample and resuspended it in 10% D_2O to observe the exchangeable protons and determine whether all the Watson-Crick base pairs were intact. The spectrum showed the imino groups of G and T but some of those signals were not vey intense. There is a broad signal that we tentatively assigned as the bridge amino-protons, but further experiments would be required to fully establish this assignment. On the other hand this signal could also belong to a non-hydrogen bonded imino proton of the C that are opposed to the modified G if the perturbation in the central region is strong enough to disrupt the H-bond. Alternatively it could be the result of breathing of the terminal bases. We decided that due to weak intensity of the signal and the broadness of the peaks it was not worth to acquire a 2D spectrum in 10% D_2O.

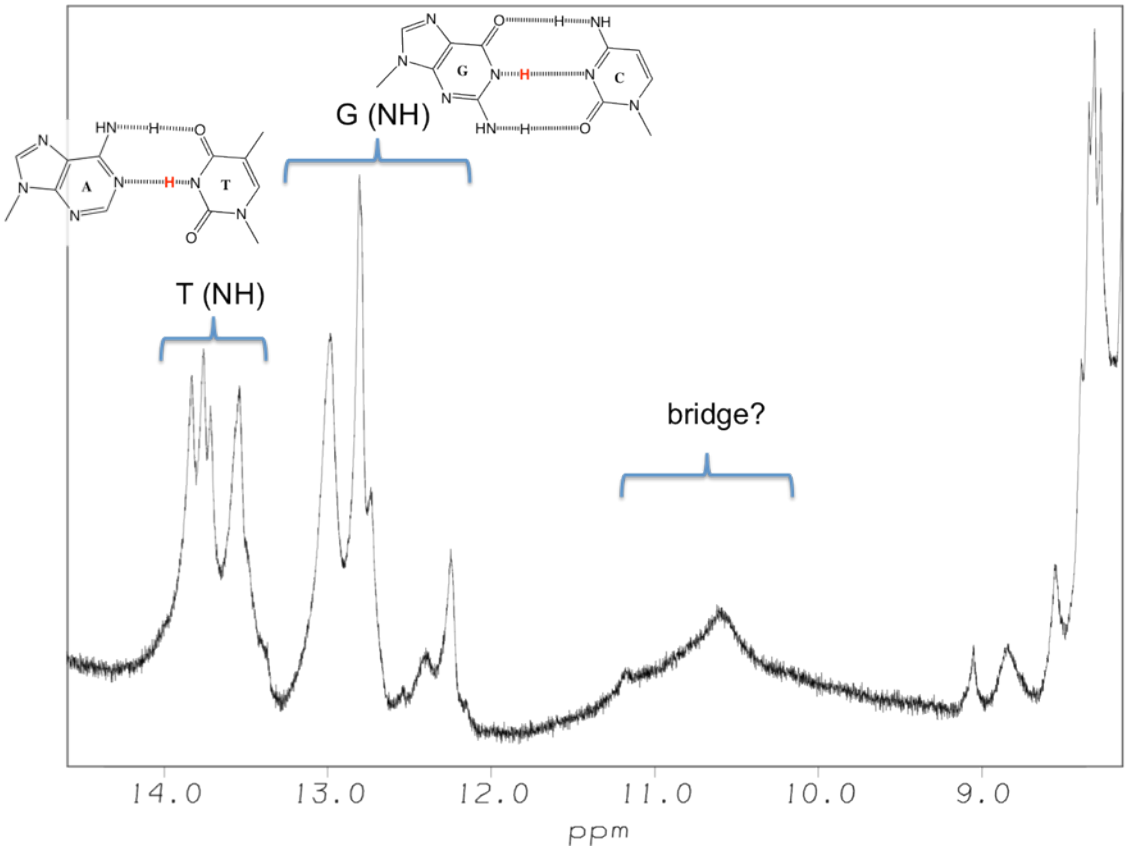


Figure 5: 1D-spectrum recorded in 10% D_2O buffer showing the region of the exchangeable protons of thymidine and guanine. The protons involved in Watson-Crick base pairs have a chemical shift between 12 and 14 ppm approximately. The broad peak could represent the protons of the bridge that are not involved in any Watson-Crick base pairs.

Discussion

We were able to synthesized 600 nmols of the isosteric analogue of the NM ICL oligonucleotide and used for NMR studies. The amount was at the lower limit to obtain good spectra, and the overall signal was not very intense making it necessary to use very long acquisition times. Nonetheless the results were sufficient to gain a considerable amount of information. The most significant data came from the 2D-NOESY experiment. In the fingerprint region we were not able to find the cross-peak for several of the interactions expected in the region of the crosslink. Intuitively, and more importantly from the free MD calculation, we expected the bases involved in the crosslinked to have

little freedom of movement, since we inserted a covalent bond that keep two bases on opposite strands together. The absence of these crosspeaks prevented us from making a clear assessment on the structural features of the crosslink and the bases around it. There are several possible explanation for these observations. One possibility is that the crosslink assumes more than one defined conformation, splitting the signal into two or more peaks, making them less intese and therfore not detectable. It is possible that multiple conformations are not detected during the MD experiments due to the very short time scale of the used in these experiments (50 ns). The data obtained are however not conclusive enough to establish the cause of the missing interaction peaks.

The disturbance of the region of the crosslink is also detected by the unusual downfield shift of the signals for the interactions C7(H5-H6) and C18(H5-H6). This type of shift is more usually an indication that the bases have extensive contact with the solvent, for example if they assume an extrahelical positioning.

One other striking feature of the NOESY spectrum is the peak of the methyl group of T6. First of all is notably shifted upfield (around 1 ppm) and secondly, from the 1D spectrum, we noticed that it appears to be split into two minor peaks (data not shown). We hypothesize that the methyl group of T6 is sort of trapped into a position where it is unable to freely rotate leading to a split in the signal of the methyl protons. In the MD simulations the average distance between the carbon of the methyl group of T6 and one of the nitrogen of the bridge is 4.2 Å, possibly causing a steric clash between the bridge and the methyl group that prevents it from rotating. Such an interactions could also partially explain the broadening of the peak in the 1D spectrum in 10% D_2O that we speculated might belong to a proton of the bridge. The methyl group could therefore be influencing the conformation of the bridge and the all region around it. Interestingly in the MD simulations also the average distance between O4 of T6 and one of the nitrogen of the bridge is only 3.4 Å, what could allow a partial H-bonding.

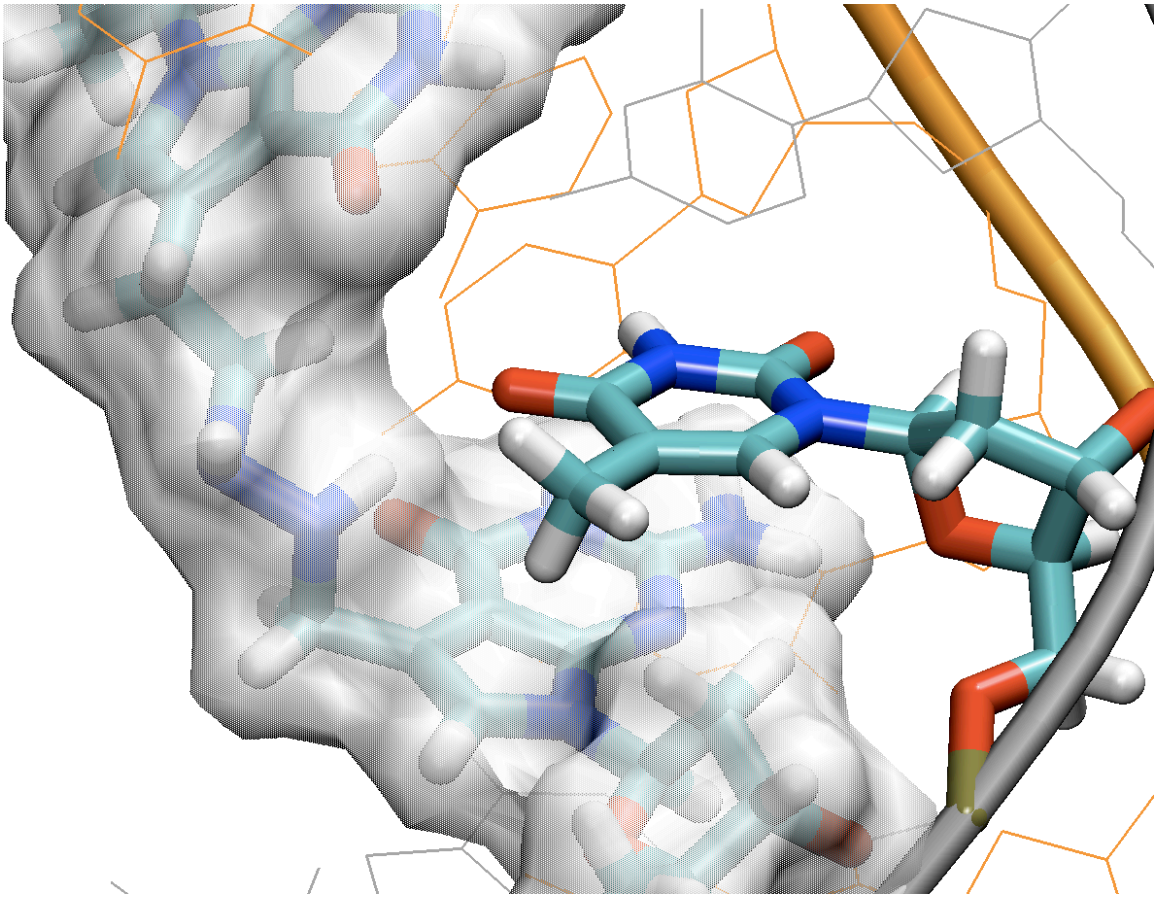


Figure 6: picture of the most populated structure of the MD simulation described in chapter 3. The picture shows the thymidine 6 and his methyl group located between the sugar of G5 and the bridge of the crosslink.

Overall we do not have sufficient data to make any conclusive statements about the structure of the oligonucleotide containing the NM ICL. A direct validation of the the MD simulation data was therefore not possible. We have indications that the time frame in which the MD simulations were carried out is too short to reveal some relatively slow change of conformation that indeed affects the NMR measurements.

Future experiments could envision the utilization of a different sequence around the crosslink, particularly sustituting T5 to avoid any possible disturbance of the crosslink by the methyl group as discussed above. Another possibility is to perform NMR analysis of an ICL with a different length of the bridge that may induce less distortion and have a more stable structure therefore being easier to characterize.

Experimental Details

Preparation of the ICL sample

The single stranded oligonucleotides were prepared as described in Chapter 3. A solution of the single strand oligonucleotides (400 nmol) in 800 μl 10 mM NaCl was heated to 95°C and allowed to cool to room temperature over a period of 4 hrs to allow for annealing of the two oligos. After addition of 100 μl 1M sodium phosphate buffer (pH 5.4) and 100 μl 50 mM $NaIO_4$ the reaction mixture was kept in the dark overnight at 4°C. Excess $NaIO_4$ was removed by centrifugation through Centricon columns with a 3K cutoff (Millipore). The volume was adjusted to 1200 μl with 0.1 M phosphate buffer. The crosslink was formed through reductive amination by addition of 150 μl 5 mM hydrazine solution in water and 150 μl 0.5 M $NaCNBH_3$. The reaction mixture was left overnight at room temperature in the dark and then concentrated to a final volume of approx 300 μl and loaded on a denaturing 20cm x 20cm x 3mm 20% polyacrylamide gel, containing 7M urea. The band containing the crosslinked oligonucleotide was excised from the gel and the DNA was extracted by electroelution the Elutrap™(Schleicher & Schuell) device. Typically the final yield is about 20%. The procedure was repeated until the final amount of product reached 600 nmols. The identity of the purified ICL products was confirmed by ESI-MS as described in chapter 3.

The sample was then desalted using a Sephadex G-25 column and they converted to the sodium form with a Dowex 50W cation exchange column. The duplex was dissolved in 0.5 ml of 10-25 mM phosphate buffer, pH 6.9, containing 50 mM NaCl and 1 mM EDTA according to published protocols [196,197]. The sample was lyophilized and dissolved in either 99.96% D_2O or 90% H_2O-10% D_2O for the collection of NMR data.

NMR methods

NMR experiments were conducted on a Varian (Inova) spectrometer at 600 MHz field strength.

One-dimensional ^{1}H-NMR: the sample was dissolved in 100% D_2O. Measurements were taken at 20, 25, 30 and 35 ˚C after suppression of the residual water signal by presaturation.

For the two-dimensional measurement the sample was allowed to equilibrate at 35 ˚C. NOESY spectra were recorded using mixing times of 300, 150 and 90 ms. TOCSY experiments were recorded using isotropic mixing times of 70 and 120 ms.

NMR spectra were processed and analyzed using FELIX98 (Accelrys, San Diego, CA) and Sparky [198] softwares. NOESY and TOCSY time domain data were multiplied by a 90˚-shifted sine bell window function while COSY data sets were multiplied by a sine bell function and an exponential function with 4 Hz of line broadening.

Chapter 5

Synthesis of nitrogen mustard ICLs with different bridge lengths in various sequence contexts for biochemical experiments

Synthesis of nitrogen mustard ICLs with different bridge lengths in various sequence contexts for biochemical experiments

ABSTRACT

We used our newly developed strategy to synthesize a series of substrates to use in a number of biological experiments. With our strategy we are able to synthesize ICLs that differ in the length of the bridge and therefore in the amount of distortion that they induce in the DNA duplex. Our strategy allows for the incorporation of the crosslink in any sequence context with a central 5'-GNC permitting the use of the substrate in a great variety of experiments. The substrates were used and are currently being used for *in vitro* experiments in our lab and by many collaborators.

Using the substrate for biological studies

We have developed a new strategy that enables the synthesis of defined site-specific ICLs in high yields and purity enabling detailed studies of ICL repair. This strategy is a tremendous advantage compared to the direct use of alkylating agents to generate ICL-containing duplexes, where many side products are formed as already described in Chapter 1. Our strategy allows the preparation of ICLs with different bridge length using a variety of precursors and amines for the reductive amination reaction [36](Figure 1).

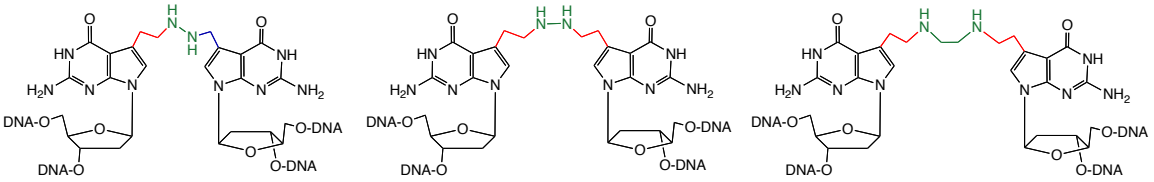


Figure 1: Three different mimics of nitrogen mustard ICL that can be synthesized using different aldehyde precursors and amines.

The difference in length of the bridge is important because it results in different degrees of distortion induced in the oligonucleotide duplex in which it is incorporated. As already described in Chapter 2 and 3 the ~7.5 Å mimic reflects the length of the original nitrogen mustard bridge, whereas the ~9.0 Å mimic spans the distance between the two bases it links in regular B-form DNA. The ~11.7 Å mimic accommodates its longer bridge in the major groove and it probably does not provoke any distortion.

Structure-dependent bypass of ICLs by TLS polymerases

These substrates were used in our laboratory for a series of experiments to investigate the ability of translesion synthesis (TLS) polymerases to bypass ICLs. The experiments were conducted using primer extension templates containing either the 1,2 cisplatin or 1,3 NM ICLs. In case of NM ICLs all three different length of the bridge were used.

We found that TLS polymerases such as Polη, Polκ or Polι are able to bypass all ICLs but with different efficiencies. These polymerases were able to bypass lesions more effectively way after recession of the non-template strand to reduce the length of the double stranded DNA around the ICL. The recession was made possible by the incorporation of uracil bases close to the ICL and subsequent treatment with uracil DNA glycosylase (UDG) and treatment with sodium hydroxide to cleave the abasic site formed by UDG. The resulting crosslink substrate contains a 39-mer oligonucleotide linked to a short 6-mer strand. The bypass was also facilitated by increasing the length of the bridge resulting in alleviation the spatial constrains induced by the ICL [192]. For instance all the TLS polymerases tested were able to bypass the 11.7 Å mimic, while only Pol η could bypass the 7.5 Å mimic (Figure 2).

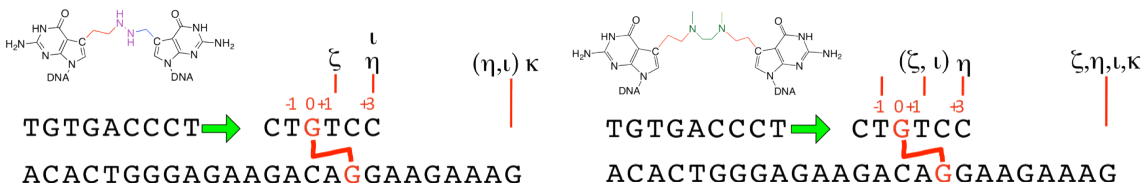


Figure 2: Substrates used for the TLS polymerase bypass experiments. On the left the crosslink inserted in the sequence was the 7.5 Å mimic, whereas the sequence on the right contains the 11.7 Å mimic. The sites where the TLS polymerases are blocked or paused and whether they can fully bypass the lesion to reach the end of the oligonucleotides are indicated.

Replication-dependent ICL repair studies

Our laboratory also prepared several substrates that were used in group of Johannes Walter (Harvard Medical School) to study replication-dependent ICL repair in cell-free *Xenopus laevis* egg extracts [44] (see chapter 1). They used the 11.7 Å mimic in a series of experiments along with a cisplatin ICL that was also prepared in our laboratory. For the study to be successful they needed to have a single ICL inserted into a specific position of an oligonucleotide that was then ligated into a plasmid.

We recently synthesized another set of ICLs that contain an additional roadblock for polymerases (Figure 3). With our strategy we can decide to incorporate other modified bases at a certain distance from the crosslink, enabling more sophisticated approaches to study ICL repair. In this case we inserted one or two biotinylated thymidine on one or both strands. The biotin can be later coupled to streptavidin-containig beads that will block the polymerase before it reaches the crosslink. This substrate will be used to address the important question whether the repair of ICLs occurs by different mechanisms whether one or two replication forks approach the ICL.

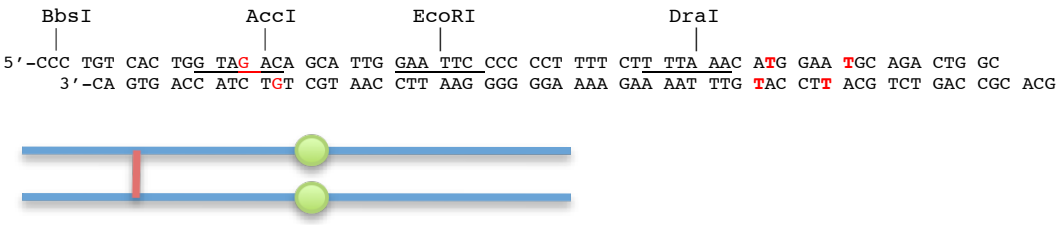


Figure 3: Sequence and schematic representation of the oligonuclotide containing an ICL (red G and red bar) and biotinylated thymidines (red bold T and green circles). Restriction sites for the enzymes specified above the sequence are underlined.

Substrate containing phosphorothioate linkages

We were also able to synthesize an ICL-containing oligonucleotide with few bases linked by a phosphorothioate bond instead of a regular phosphodiester bond. Replacing a non-bridging oxygen atom with a sulfur atom makes the linkage resistant to nuclease degradation in a regular buffer containing Mg^{2+}. The enzymatic activity can be restored however by using Mn^{2+} as a metal cofactor, making it an effective method to control the cleavage at a particular site of the oligonucleotide [199].

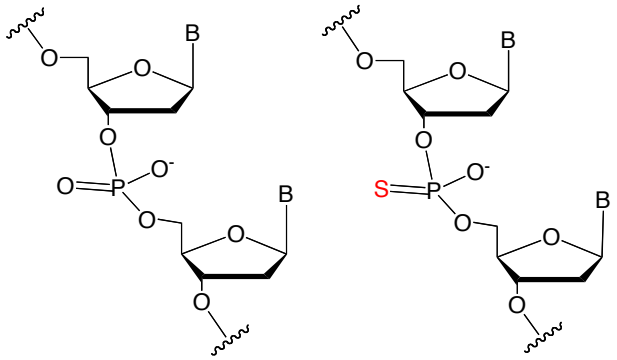


Figure 4: Comparison between a regular phosphodiester bond (left) and a phosphorothioate bond (right).

A substrate containing three phosphorotihioate linkages was prepared for a study of the exonuclease activity of Polε in the laboratory of Erik Johannson (Umea University, Sweden).

Other substrates

We also performed the synthesis of NM ICLs in different sequence contexts for the study of the DNA translocase activity of FANCM (Patrick Sung, Yale University) and for the study of the enzymatic activity of DNA toposiomerase I (Yves Pommier, NIH) as described in the experimental section.

Experimental Details

TLS polymerase substrate

The precursor phosphoramidites were prepared as described in chapters 2 and 3.

The modified phosphoramidite **4** described in Chapter 2 was introduced at position X of the oligonucleotide **NT1** and the complementary sequences **N1, N2**, containing one uracil 5' or 3'of the modified G respectively and **N3**, which contains two uracil on both sides of the modified G**.** To obtain the 7.5 Å mimic the phosphoramidite **9** described in Chapter 3 was incorporated into **NT1.** Oligonucleotides synthesis was performed by automated DNA synthesis on an Expedite 8909 Nucleic Acid Synthesis System (Applied Biosystems) using 1μmol 1000Å CPG-dG and -dC column cartridges (Glen Research) and standard reagents and conditions. For the modified phosphoramidites extended coupling time of 15min was used.

The oligomers were deprotected and purified as described in chapter 2 and 3. Crosslink formation was also performed as described in chapter 2 and 3 using either hydrazine or N,N,dimethylethlendiamine to very the length of the bridge.

Sequences: NT1: 5'-d(GAAAGAAGXACAGGAGAGGGTACCATCATAGAGTCAGTG)

N1 5'-d(CCCTCTUCTXTCCTTCTTTC)

N2 5'-d(CCCTCTTCTXTCCUTCTTTC)

N3 5'-d(CCCTCTUCTXTCCUTCTTTC)

All the biochemical experiments are described in [192]. All the crosslinks shown in Figure 1 were prepared.

NM-ICL oligonucleotides containing biotin dT residues

The modified phosphoramidite **4** described in Chapter 2 was introduced at position X of the oligonucleotide **M1** and the complementary sequences **M2**. Biotin dT was inserted at the indicated positions in the sequence. Both M1 and M2 were synthesized containing one, two or no biotin dT. Oligonucleotides synthesis was performed by automated DNA synthesis as described above. The oligomers were deprotected as described in chapter 2 and 3. Due to the extensive length, oligonucleotides were purified by 15% PAGE, the band containing the oligonucleotide was excised from the gel and the DNA was extracted by electroelution the Elutrap™ (Schleicher & Schuell) device. Crosslink formation was performed as described in chapter 2 and 3 using N,N,dimethylethlendiamine.

Sequences:

M1: 5'-d(CCCTGTCACTGGTAXACAGCATTGGAATTCCCCCCTTTTCTTTTAAACATGGAATGCAGACTGGC)

M2: 3'-d(CAGTGACCATCTXTCGTAACCTTAAGGGGGGAAAAGAAAATTTGTACCTTACGTCTGACCGCACG)

Biotin dT (Glen Research, cat #: 10-1038-xx) is indicated in violet. The different products contain one or two biotin either only on one strand or on both strands.

NM-ICL oligonucleotides containing terminal phosphorothioate linkages

The modified phosphoramidite **4** described in Chapter 2 was introduced at position X of the oligonucleotide **E1** and the complementary sequences **E2**. Oligonucleotides synthesis was performed by automated DNA synthesis as described above with the exception of the first three bases of E2 where the phosphorothioate linkages were inserted. After coupling of the phosphamidite the resulting phosphite triester was treated with sulfurizing reagent (3-((Dimethylamino-methylidene)amino)-3H-1,2,4-dithiazole-3-thione, DDTT, Glen Research, cat # 40-4037-xx).

The oligomers were deprotected as described in chapter 2 and 3. Due to the extensive length, oligonucleotides were purified by 15% PAGE, the band containing the oligonucleotide was excised from

the gel and the DNA was extracted by electroelution the Elutrap™(Schleicher & Schuell) device. Crosslink formation was performed as described in chapter 2 and 3.

Sequences:

E1: 5'-d(GGAACTAGGTCACGATGCGGCATAGCCTGCATAGACTTCACGATCACTAAC**X**GCAATCTTACC)

E2: 3'-d(TCCAGTGCTACGCCGTATCGGACGTATCTGAAGTGCTAGTGATTGCC**X**TTAGAATGGACAGTGTCA TTGTGTTACDMTr)

The phosphorothioate diester linkages are indicated in blue and the DMTr group was not removed from the 5'-OH group of E2 (in green) to allow for selective labeling of E1.

Other substrates

The modified phosphoramidite **4** described in Chapter 2 was introduced at position X of the oligonucleotide **P1** and the complementary sequences **P2.** Oligonucleotides synthesis was performed by automated DNA synthesis as described above. The oligomers were deprotected and purified as described in chapter 2 and 3. Crosslink formation was also performed as described in chapter 2 and 3.

Sequences: P1: 5'-d(GCTCACTAAATTCCTGAGGCT**X**ACTACTGTCCTGTCTTAATTAACG)

P2: 3'-d(CGAGTGATTTAAGGACTCCGACT**X**ATGACAGGACAGAATTAATTGC)

The modified phosphoramidite **4** described in Chapter 2 was introduced at position X of the oligonucleotide **Y1** and **Y3** and the complementary sequences **Y2** and **Y4**. To obtain the 7.5 Å mimic the phosphoramidite **9** described in Chapter 3 was incorporated in **Y1** and **Y3**. Oligonucleotides synthesis was performed by automated DNA synthesis as described above. The oligomers were deprotected and purified and ICLs formed as described in Chapters 2 and 3.

Sequences: Y1: 5'-d(AAAAAGACTTGGAAAAATTTTT)

Y2: 3'-d(TTTTTCTGAACCTTTTTAAAAA)

Y3: 5'-d(AAAAAGACTTGGCAAAATTTTT)

Y4: 5'-d(TTTTTCTGAACCGTTTTAAAAA)

References

1. Helleday, T., Petermann, E., Lundin, C., Hodgson, B. and Sharma, R.A. (**2008**). DNA repair pathways as targets for cancer therapy. *Nat Rev Cancer* 8, 193-204.

2. Röntgen, W.C. (**1896**). On a new kind of rays. *Science* 3, 227-231.

3. Connell, P.P. and Hellman, S. (**2009**). Advances in radiotherapy and implications for the next century: a historical perspective. *Cancer Res* 69, 383-392.

4. Watson, J.D. and Crick, F.H. (**1953**). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* 171, 737-738.

5. Krumbhaar, E. and Krumbhaar, H. (**1919**). The blood and bone marrow in yellow cross gas (mustard gas) poisoning. Changes produced in the bone marrow of fatal cases. *J Med Res* 40, 497-508.

6. Adair, F.E. and Bagg, H. (**1931**). Experimental and clinical studies on the treatment of cancer by dichlorethylsulphide (Mustard Gas). *Annals of Surgery* 93, 190-199.

7. Boulikas, T., Pantos, A., Bellis, E. and Christofis, P. (**2007**). Designing platinum compounds in cancer: structures and mechanisms. *Cancer Therapy* 5, 537-583.

8. DeVita, V.T. and Chu, E. (**2008**). A history of cancer chemotherapy. *Cancer Res* 68, 8643-8653.

9. Hirsch, J. (**2006**). An Anniversary for Cancer Chemotherapy. *JAMA*

10. Kohn, K.W. (**1996**). Beyond DNA cross-linking: history and prospects of DNA-targeted cancer treatment--fifteenth Bruce F. Cain Memorial Award Lecture. *Cancer Res* 56, 5533-5546.

11. Goodman, L., Wintrobe, M., Dameshek, W., Goodman, M. and Gilman, A. (**1946**). Nitrogen Mustard therapy - Use of methyl-bis(beta-chloroethyl)amine hydrochloride and tris(beta-chloroethyl)amine hydrochloride for hodgkins disease, lymphosarcoma, leukemia and certain allied and miscellaneous disorders. *Jama-J Am Med Assoc* 132, 126-132.

12. Gilman, A. and Philips, F. (**1946**). The Biological Actions and Therapeutic Applications of the B-Chloroethyl Amines and Sulfides. *Science* 103, 409-436.

13. Rhoads, C.P. (**1946**). Nitrogen mustards in the treatment of neoplastic disease. *Jama-J Am Med Assoc* 131, 656-658.

14. Goodman, L.S., Wintrobe, M.M., Dameshek, W., Goodman, M.J. and Gilman, A. (1946) Nitrogen mustard therapy - use of methyl-bis(beta-chloroethyl)amine hydrochloride and tris(beta-chloroethyl)amine hydrochloride for hodgkins disease, lymphosarcoma, leukemia and certain allied and miscellaneous disorders. In Jama-J Am Med Assoc ed.^eds), pp. 126-132

15. Goldacre, R.J., Loveless, A. and Ross, W.C.J. (**1949**). Mode of production of chromosome abnormalities by the nitrogen mustards; the possible role of cross-linking. *Nature* 163, 667-669.

16. Geiduschek, E.P. (**1961**). "Reversible" DNA. *Proc Natl Acad Sci USA* 47, 950-955.

17. Brookes, P. and Lawley, P.D. (**1961**). The reaction of mono- and di-functional alkylating agents with nucleic acids. *Biochem J* 80, 496-503.

18. Kohn, K.W., Spears, C.L. and Doty, P. (**1966**). Inter-strand crosslinking of DNA by nitrogen mustard. *J Mol Biol* 19, 266-288.

19. Schärer, O.D. (**2005**). DNA interstrand crosslinks: natural and drug-induced DNA adducts that induce unique cellular responses. *Chembiochem* 6, 27-32.

20. Noll, D.M., Mason, T.M. and Miller, P.S. (**2006**). Formation and repair of interstrand cross-links in DNA. *Chem Rev* 106, 277-301.

21. Rajski, S.R. and Williams, R.M. (**1998**). DNA Cross-Linking Agents as Antitumor Drugs. *Chem Rev* 98, 2723-2796.

22. Stone, M.P., Cho, Y.-J., Huang, H., Kim, H.-Y., Kozekov, I.D., Kozekova, A., Wang, H., Minko, I.G., Lloyd, R.S., Harris, T.M. and Rizzo, C.J. (**2008**). Interstrand DNA cross-links induced by alpha,beta-unsaturated aldehydes derived from lipid peroxidation and environmental sources. *Accounts Chem Res* 41, 793-804.

23. Panasci, L., Xu, Z.-Y., Bello, V. and Aloyz, R. (**2002**). The role of DNA repair in nitrogen mustard drug resistance. *Anticancer Drugs* 13, 211-220.

24. McHugh, P.J., Spanswick, V.J. and Hartley, J.A. (**2001**). Repair of DNA interstrand crosslinks: molecular mechanisms and clinical relevance. *Lancet Oncol* 2, 483-490.

25. Ojwang, J.O., Grueneberg, D.A. and Loechler, E.L. (**1989**). Synthesis of a duplex oligonucleotide containing a nitrogen mustard interstrand DNA-DNA cross-link. *Cancer Res* 49, 6529-6537.

26. Rink, S., Solomon, M., Taylor, M., Rajur, S., McLaughlin, L. and Hopkins, P. (**1993**). Covalent Structure of a Nitrogen Mustard-Induced DNA Interstrand Cross-Link - an N7-to-N7 Linkage of Deoxyguanosine Residues at the Duplex Sequence 5'-d(GNC). *J Am Chem Soc* 115, 2551-2557.

27. Millard, J., Raucher, S. and Hopkins, P. (**1990**). Mechlorethamine Cross-Links Deoxyguanosine Residues at 5'-GNC Sequences in Duplex DNA Fragments. *J Am Chem Soc* 112, 2459-2460.

28. Rink, S.M. and Hopkins, P.B. (**1995**). A mechlorethamine-induced DNA interstrand cross-link bends duplex DNA. *Biochemistry* 34, 1439-1445.

29. Kallama, S. and Hemminki, K. (**1984**). Alkylation of guanosine by phosphoramide mustard, chloromethine hydrochloride and chlorambucil. *Acta Pharmacol Toxicol (Copenh)* 54, 214-220.

30. Kallama, S. and Hemminki, K. (**1986**). Stabilities of 7-alkylguanosines and 7-deoxyguanosines formed by phosphoramide mustard and nitrogen mustard. *Chem Biol Interact* 57, 85-96.

31. Fan, Y. and Gold, B. (**1999**). Sequence-Specificity for DNA Interstrand Cross-Linking by alpha,ω-Alkanediol Dimethylsulfonate Esters: Evidence for DNA Distortion by the Initial Monofunctional Lesion. *J Am Chem Soc* 121, 11942-11946.

32. Mehta, J.R., Przybylski, M. and Ludlum, D.B. (**1980**). Alkylation of guanosine and deoxyguanosine by phosphoramide mustard. *Cancer Res* 40, 4183-4186.

33. Grueneberg, D., Ojwang, J., Benasutti, M., Hartman, S. and Loechler, E. (**1991**). Construction of a human shuttle vector containing a single nitrogen mustard interstrand, DNA-DNA cross-link at a unique plasmid location. *Cancer Res* 51, 2268-2272.

34. Berardini, M., Mackay, W. and Loechler, E. (**1997**). Evidence for a recombination-independent pathway for the repair of DNA interstrand cross-links based on a site-specific study with nitrogen mustard. *Biochemistry* 36, 3506-3513.

35. Berardini, M., Foster, P. and Loechler, E. (**1999**). DNA polymerase II (polB) is involved in a new DNA repair pathway for DNA interstrand cross-links in Escherichia coli. *J Bacteriol* 181, 2878-2882.

36. Angelov, T., Guainazzi, A. and Schärer, O.D. (**2009**). Generation of DNA Interstrand Cross-Links by Post-Synthetic Reductive Amination. *Org Lett* 11, 661-664.

37. Rosenberg, B., VanCamp, L., Trosko, J.E. and Mansour, V.H. (**1969**). Platinum compounds: a new class of potent antitumour agents. *Nature* 222, 385-386.

38. Rosenberg, B., VanCamp, L. and Krigas, T. (**1965**). Inhibition of Cell Division in *Escherichia Coli* by Electrolysis Products from a Platinum Electrode. *Nature* 205, 698-699.

39. Kelland, L. (**2007**). The resurgence of platinum-based cancer chemotherapy. *Nat Rev Cancer* 7, 573-584.

40. Jamieson, E.R. and Lippard, S.J. (**1999**). Structure, Recognition, and Processing of Cisplatin-DNA Adducts. *Chem Rev* 99, 2467-2498.

41. Huang, H., Zhu, L., Reid, B.R., Drobny, G.P. and Hopkins, P.B. (**1995**). Solution structure of a cisplatin-induced DNA interstrand cross-link. *Science* 270, 1842-1845.

42. Coste, F., Malinge, J.M., Serre, L., Shepard, W., Roth, M., Leng, M. and Zelwer, C. (**1999**). Crystal structure of a double-stranded DNA containing a cisplatin interstrand cross-link at 1.63 A resolution: hydration at the platinated site. *Nucleic Acids Research* 27, 1837-1846.

43. Hofr, C. and Brabec, V. (**2001**). Thermal and thermodynamic properties of duplex DNA containing site-specific interstrand cross-link of antitumor cisplatin or its clinically ineffective trans isomer. *J Biol Chem* 276, 9655-9661.

44. Räschle, M., Knipscheer, P., Knipsheer, P., Enoiu, M., Angelov, T., Sun, J., Griffith, J.D., Ellenberger, T.E., Schärer, O.D. and Walter, J.C. (**2008**). Mechanism of replication-coupled DNA interstrand crosslink repair. *Cell* 134, 969-980.

45. Zhu, G. and Lippard, S.J. (**2009**). Photoaffinity labeling reveals nuclear proteins that uniquely recognize cisplatin-DNA interstrand cross-links. *Biochemistry* 48, 4916-4925.

46. Johannessen, T.-C.A., Bjerkvig, R. and Tysnes, B.B. (**2008**). DNA repair and cancer stem-like cells--potential partners in glioma drug resistance? *Cancer Treat Rev* 34, 558-567.

47. Ludlum, D.B. (**1997**). The chloroethylnitrosoureas: sensitivity and resistance to cancer chemotherapy at the molecular level. *Cancer Invest* 15, 588-598.

48. Fischhaber, P.L., Gall, A.S., Duncan, J.A. and Hopkins, P.B. (**1999**). Direct demonstration in synthetic oligonucleotides that N,N'-bis(2-chloroethyl)-nitrosourea cross links N1 of deoxyguanosine to N3 of deoxycytidine on opposite strands of duplex DNA. *Cancer Res* 59, 4363-4368.

49. Alzeer, J. and Schärer, O.D. (**2006**). A modified thymine for the synthesis of site-specific thymine-guanine DNA interstrand crosslinks. *Nucleic Acids Research* 34, 4458-4466.

50. Noll, D.M., Noronha, A.M. and Miller, P.S. (**2001**). Synthesis and characterization of DNA duplexes containing an N(4)C-ethyl-N(4)C interstrand cross-link. *J Am Chem Soc* 123, 3405-3411.

51. Wilds, C.J., Noronha, A.M., Robidoux, S. and Miller, P.S. (**2004**). Mispair-aligned N3T-alkyl-N3T interstrand cross-linked DNA: synthesis and characterization of duplexes with interstrand cross-links of variable lengths. *J Am Chem Soc* 126, 9257-9265.

52. Wilds, C.J., Xu, F. and Noronha, A.M. (**2008**). Synthesis and characterization of DNA containing an N-1-2 '-deoxyinosine-ethyl-N-3-thymidine interstrand cross-link: A structural mimic of the cross-link formed by 1,3-bis-(2-chloroethyl)-1-nitrosourea. *Chem Res Toxicol* 21, 686-695.

53. Smeaton, M.B., Hlavin, E.M., McGregor Mason, T., Noronha, A.M., Wilds, C.J. and Miller, P.S. (**2008**). Distortion-dependent unhooking of interstrand cross-links in mammalian cell extracts. *Biochemistry* 47, 9920-9930.

54. Smeaton, M.B., Hlavin, E.M., Noronha, A.M., Murphy, S.P., Wilds, C.J. and Miller, P.S. (**2009**). Effect of cross-link structure on DNA interstrand cross-link repair synthesis. *Chem Res Toxicol* 22, 1285-1297.

55. Hofheinz, R.-D., Beyer, U., Al-Batran, S.-E. and Hartmann, J.T. (**2008**). Mitomycin C in the treatment of gastrointestinal tumours: recent data and perspectives. *Onkologie* 31, 271-281.

56. Tomasz, M., Lipman, R., Chowdary, D., Pawlak, J., Verdine, G.L. and Nakanishi, K. (**1987**). Isolation and structure of a covalent cross-link adduct between mitomycin C and DNA. *Science* 235, 1204-1208.

57. Tomasz, M., Chowdary, D., Lipman, R., Shimotakahara, S., Veiro, D., Walker, V. and Verdine, G.L. (**1986**). Reaction of DNA with chemically or enzymatically activated mitomycin C: isolation and structure of the major covalent adduct. *Proc Natl Acad Sci USA* 83, 6702-6706.

58. Rink, S.M., Lipman, R., Alley, S.C., Hopkins, P.B. and Tomasz, M. (**1996**). Bending of DNA by the mitomycin C-induced, GpG intrastrand cross-link. *Chem Res Toxicol* 9, 382-389.

59. Norman, D., Live, D., Sastry, M., Lipman, R., Hingerty, B.E., Tomasz, M., Broyde, S. and Patel, D.J. (**1990**). NMR and computational characterization of mitomycin cross-linked to adjacent deoxyguanosines in the minor groove of the d(T-A-C-G-T-A).d(T-A-C-G-T-A) duplex. *Biochemistry* 29, 2861-2875.

60. Warren, A.J., Ihnat, M.A., Ogdon, S.E., Rowell, E.E. and Hamilton, J.W. (**1998**). Binding of nuclear proteins associated with mammalian DNA repair to the mitomycin C-DNA interstrand crosslink. *Environ Mol Mutagen* 31, 70-81.

61. Mustra, D.J., Warren, A.J. and Hamilton, J.W. (**2001**). Preferential binding of human full-length XPA and the minimal DNA binding domain (XPA-MF122) with the mitomycin C-DNA interstrand cross-link. *Biochemistry* 40, 7158-7164.

62. Zheng, H., Wang, X., Warren, A., Legerski, R., Nairn, R., Hamilton, J. and Li, L. (**2003**). Nucleotide excision repair- and polymerase eta-mediated error-prone removal of mitomycin C interstrand cross-links. *Mol Cell Biol* 23, 754-761.

63. Cimino, G., Gamper, H., Isaacs, S. and Hearst, J. (**1985**). Psoralens as photoactive probes of nucleic acid structure and function: organic chemistry, photochemistry, and biochemistry. *Annu Rev Biochem* 54, 1151-1193.

64. Stern, R.S. (**2007**). Psoralen and ultraviolet a light therapy for psoriasis. *N Engl J Med* 357, 682-690.

65. Spielmann, H.P., Sastry, S.S. and Hearst, J.E. (**1992**). Methods for the large-scale synthesis of psoralen furan-side monoadducts and diadducts. *Proc Natl Acad Sci USA* 89, 4514-4518.

66. Kobertz, W. and Essigmann, J.M. (**1997**). Solid-phase synthesis of oligonucleotides containing a site-specific psoralen derivative. *J Am Chem Soc* 119, 5960-5961.

67. Spielmann, H.P., Dwyer, T.J., Hearst, J.E. and Wemmer, D.E. (**1995**). Solution structures of psoralen monoadducted and cross-linked DNA oligomers by NMR spectroscopy and restrained molecular dynamics. *Biochemistry* 34, 12937-12953.

68. Hwang, G.S., Kim, J.K. and Choi, B.S. (**1996**). The solution structure of a psoralen cross-linked DNA duplex by NMR and relaxation matrix refinement. *Biochem Biophys Res Commun* 219, 191-197.

69. Sinden, R.R. and Hagerman, P.J. (**1984**). Interstrand psoralen cross-links do not introduce appreciable bends in DNA. *Biochemistry* 23, 6299-6303.

70. Thazhathveetil, A.K., Liu, S.T., Indig, F.E. and Seidman, M.M. (**2007**). Psoralen conjugates for visualization of genomic interstrand cross-links localized by laser photoactivation. *Bioconjug Chem* 18, 431-437.

71. Majumdar, A., Muniandy, P.A., Liu, J., Liu, J.L., Liu, S.T., Cuenoud, B. and Seidman, M.M. (**2008**). Targeted gene knock in and sequence modulation mediated by a psoralen-linked triplex-forming oligonucleotide. *J Biol Chem* 283, 11244-11252.

72. Niedernhofer, L.J., Daniels, J.S., Rouzer, C.A., Greene, R.E. and Marnett, L.J. (**2003**). Malondialdehyde, a product of lipid peroxidation, is mutagenic in human cells. *Journal of Biological Chemistry* 278, 31426-31433.

73. Cheng, G., Shi, Y., Sturla, S.J., Jalas, J.R., McIntee, E.J., Villalta, P.W., Wang, M. and Hecht, S.S. (**2003**). Reactions of formaldehyde plus acetaldehyde with deoxyguanosine and DNA: formation of cyclic deoxyguanosine adducts and formaldehyde cross-links. *Chem Res Toxicol* 16, 145-152.

74. Minko, I.G., Harbut, M.B., Kozekov, I.D., Kozekova, A., Jakobs, P.M., Olson, S.B., Moses, R.E., Harris, T.M., Rizzo, C.J. and Lloyd, R.S. (**2008**). Role for DNA polymerase kappa in the processing of N2-N2-guanine interstrand cross-links. *J Biol Chem* 283, 17075-17082.

75. Dooley, P.A., Tsarouhtsis, D., Korbel, G.A., Nechev, L.V., Shearer, J., Zegar, I.S., Harris, C.M., Stone, M.P. and Harris, T.M. (**2001**). Structural studies of an oligodeoxynucleotide containing a trimethylene interstrand cross-link in a 5'-(CpG) motif: model of a malondialdehyde cross-link. *J Am Chem Soc* 123, 1730-1739.

76. Dooley, P.A., Zhang, M., Korbel, G.A., Nechev, L.V., Harris, C.M., Stone, M.P. and Harris, T.M.

(**2003**). NMR determination of the conformation of a trimethylene interstrand cross-link in an oligodeoxynucleotide duplex containing a 5'-d(GpC) motif. *J Am Chem Soc* 125, 62-72.

77. Shapiro, R., Dubelman, S., Feinberg, A.M., Crain, P.F. and McCloskey, J.A. (**1977**). Isolation and identification of cross-linked nucleosides from nitrous acid treated deoxyribonucleic acid. *J Am Chem Soc* 99, 302-303.

78. Kirchner, J. and Hopkins, P. (**1991**). Nitrous-Acid Cross-Links Duplex DNA Fragments Through Deoxyguanosine Residues At The Sequence 5'-CG. *J Am Chem Soc* 113, 4681-4682.

79. Caulfield, J.L., Wishnok, J.S. and Tannenbaum, S.R. (**2003**). Nitric oxide-induced interstrand cross-links in DNA. *Chem Res Toxicol* 16, 571-574.

80. Edfeldt, N.B.F., Harwood, E.A., Sigurdsson, S.T., Hopkins, P.B. and Reid, B.R. (**2004**). Solution structure of a nitrous acid induced DNA interstrand cross-link. *Nucleic Acids Research* 32, 2785-2794.

81. Harwood, E., Sigurdsson, S., Edfeldt, N., Reid, B. and Hopkins, P.B. (**1999**). Chemical synthesis and preliminary structural characterization of a nitrous acid interstrand cross-linked duplex DNA. *J Am Chem Soc* 121, 5081-5082.

82. Hong, I.S. and Greenberg, M.M. (**2005**). Efficient DNA interstrand cross-link formation from a nucleotide radical. *J Am Chem Soc* 127, 3692-3693.

83. Hong, I.S., Ding, H. and Greenberg, M.M. (**2006**). Oxygen independent DNA interstrand cross-link formation by a nucleotide radical. *J Am Chem Soc* 128, 485-491.

84. Cole, R. (**1973**). Repair of DNA containing interstrand crosslinks in Escherichia coli: sequential excision and recombination. *Proc Natl Acad Sci USA* 70, 1064-1068.

85. Cole, R., Levitan, D. and Sinden, R. (**1976**). Removal of psoralen interstrand cross-links from DNA of Escherichia coli: mechanism and genetic control. *J Mol Biol* 103, 39-59.

86. Akkari, Y., Bateman, R., Reifsteck, C., Olson, S. and Grompe, M. (**2000**). DNA replication is required To elicit cellular responses to psoralen-induced DNA interstrand cross-links. *Mol Cell Biol* 20, 8283-8289.

87. Hoy, C.A., Thompson, L.H., Mooney, C.L. and Salazar, E.P. (**1985**). Defective DNA cross-link removal in Chinese hamster cell mutants hypersensitive to bifunctional alkylating agents. *Cancer Res* 45, 1737-1743.

88. De Silva, I.U., McHugh, P.J., Clingen, P.H. and Hartley, J.A. (**2000**). Defining the roles of nucleotide excision repair and recombination in the repair of DNA interstrand cross-links in mammalian cells. *Mol Cell Biol* 20, 7980-7990.

89. Niedernhofer, L.J., Odijk, H., Budzowska, M., van Drunen, E., Maas, A., Theil, A.F., de Wit, J., Jaspers, N.G.J., Beverloo, H.B., Hoeijmakers, J.H.J. and Kanaar, R. (**2004**). The structure-specific endonuclease Ercc1-Xpf is required to resolve DNA interstrand cross-link-induced double-strand breaks. *Mol Cell Biol* 24, 5776-5787.

90. Hanada, K., Budzowska, M., Modesti, M., Maas, A., Wyman, C., Essers, J. and Kanaar, R. (**2006**). The structure-specific endonuclease Mus81-Eme1 promotes conversion of interstrand DNA crosslinks into double-strands breaks. *Embo J* 25, 4921-4932.

91. Hanada, K., Budzowska, M., Davies, S., van Drunen, E., Onizawa, H., Beverloo, H.B., Maas, A., Essers, J., Hickson, I. and Kanaar, R. (**2007**). The structure-specific endonuclease Mus81 contributes to replication restart by generating double-strand DNA breaks. *Nat Struct Mol Biol* 14, 1096-1104.

92. Simpson, L.J. and Sale, J.E. (**2003**). Rev1 is essential for DNA damage tolerance and non-templated immunoglobulin gene mutation in a vertebrate cell line. *Embo J* 22, 1654-1664.

93. Niedzwiedz, W., Mosedale, G., Johnson, M., Ong, C., Pace, P. and Patel, K.J. (**2004**). The Fanconi anaemia gene FANCC promotes homologous recombination and error-prone DNA repair. *Mol Cell* 15, 607-620.

94. Gan, G.N., Wittschieben, J.P., Wittschieben, B.Ø. and Wood, R.D. (**2008**). DNA polymerase zeta (pol zeta) in higher eukaryotes. *Cell Res* 18, 174-183.

95. Essers, J., Hendriks, R.W., Swagemakers, S.M., Troelstra, C., de Wit, J., Bootsma, D., Hoeijmakers, J.H. and Kanaar, R. (**1997**). Disruption of mouse RAD54 reduces ionizing radiation resistance and homologous recombination. *Cell* 89, 195-204.

96. Liu, N., Lamerdin, J.E., Tebbs, R.S., Schild, D., Tucker, J.D., Shen, M.R., Brookman, K.W., Siciliano, M.J., Walter, C.A., Fan, W., Narayana, L.S., Zhou, Z.Q., Adamson, A.W., Sorensen, K.J., Chen, D.J., Jones, N.J. and Thompson, L.H. (**1998**). XRCC2 and XRCC3, new human Rad51-family members, promote chromosome stability and protect against DNA cross-links and other damages. *Mol Cell* 1, 783-793.

97. Niedernhofer, L.J., Lalai, A.S. and Hoeijmakers, J.H.J. (**2005**). Fanconi anemia (cross)linked to DNA repair. *Cell* 123, 1191-1198.

98. Wang, W. (**2007**). Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. *Nat Rev Genet* 8, 735-748.

99. Moldovan, G.-L. and D'Andrea, A.D. (**2009**). How the fanconi anemia pathway guards the genome. *Annu Rev Genet* 43, 223-249.

100. Walter, J., Sun, L. and Newport, J. (**1998**). Regulated chromosomal DNA replication in the absence of a nucleus. *Mol Cell* 1, 519-529.

101. Yan, Z., Delannoy, M., Ling, C., Daee, D., Osman, F., Muniandy, P.A., Shen, X., Oostra, A.B., Du, H., Steltenpool, J., Lin, T., Schuster, B., Décaillet, C., Stasiak, A., Stasiak, A.Z., Stone, S., Hoatlin, M.E., Schindler, D., Woodcock, C.L., Joenje, H., Sen, R., de Winter, J.P., Li, L., Seidman, M.M., Whitby, M.C., Myung, K., Constantinou, A. and Wang, W. (**2010**). A histone-fold complex and FANCM form a conserved DNA-remodeling complex to maintain genome stability. *Mol Cell* 37, 865-878.

102. Singh, T.R., Saro, D., Ali, A.M., Zheng, X.-F., Du, C.-h., Killen, M.W., Sachpatzidis, A., Wahengbam, K., Pierce, A.J., Xiong, Y., Sung, P. and Meetei, A.R. (**2010**). MHF1-MHF2, a histone-fold-containing protein complex, participates in the Fanconi anemia pathway via FANCM. *Mol Cell* 37, 879-886.

103. Meetei, A.R., Medhurst, A.L., Ling, C., Xue, Y., Singh, T.R., Bier, P., Steltenpool, J., Stone, S., Dokal, I., Mathew, C.G., Hoatlin, M., Joenje, H., de Winter, J.P. and Wang, W. (**2005**). A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nat Genet* 37, 958-963.

104. Ciccia, A., Ling, C., Coulthard, R., Yan, Z., Xue, Y., Meetei, A.R., Laghmani, E.H., Joenje, H.,

McDonald, N., de Winter, J.P., Wang, W. and West, S.C. (**2007**). Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. *Mol Cell* 25, 331-343.

105. Collis, S.J., Ciccia, A., Deans, A.J., Horejsí, Z., Martin, J.S., Maslen, S.L., Skehel, J.M., Elledge, S.J., West, S.C. and Boulton, S.J. (**2008**). FANCM and FAAP24 function in ATR-mediated checkpoint signaling independently of the fanconi anemia core complex. *Mol Cell* 32, 313-324.

106. Meetei, A., de Winter, J.P., Medhurst, A., Wallisch, M., Waisfisz, Q., van de Vrugt, H., Oostra, A., Yan, Z., Ling, C., Bishop, C., Hoatlin, M., Joenje, H. and Wang, W. (**2003**). A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat Genet* 35, 165-170.

107. Cole, A.R., Lewis, L.P.C. and Walden, H. (**2010**). The structure of the catalytic subunit FANCL of the Fanconi anemia core complex. *Nat Struct Mol Biol* 17, 294-298.

108. Garcia-Higuera, I., Taniguchi, T., Ganesan, S., Meyn, M., Timmers, C., Hejna, J., Grompe, M. and D'Andrea, A.D. (**2001**). Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Mol Cell* 7, 249-262.

109. Smogorzewska, A., Matsuoka, S., Vinciguerra, P., McDonald, E., Hurov, K., Luo, J., Ballif, B., Gygi, S., Hofmann, K., D'Andrea, A.D. and Elledge, S. (**2007**). Identification of the FANCI Protein, a Monoubiquitinated FANCD2 Paralog Required for DNA Repair. *Cell* 129, 289-301.

110. Knipscheer, P., Räschle, M., Smogorzewska, A., Enoiu, M., Ho, T.V., Schärer, O.D., Elledge, S.J. and Walter, J.C. (**2009**). The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. *Science* 326, 1698-1701.

111. Ben-Yehoyada, M., Wang, L.C., Kozekov, I.D., Rizzo, C.J., Gottesman, M.E. and Gautier, J. (**2009**). Checkpoint signaling from a single DNA interstrand crosslink. *Mol Cell* 35, 704-715.

112. Shen, X., Do, H., Li, Y., Chung, W.-H., Tomasz, M., de Winter, J.P., Xia, B., Elledge, S.J., Wang, W. and Li, L. (**2009**). Recruitment of fanconi anemia and breast cancer proteins to DNA damage sites is differentially governed by replication. *Mol Cell* 35, 716-723.

113. McHugh, P., Sones, W. and Hartley, J. (**2000**). Repair of intermediate structures produced at DNA interstrand cross-links in Saccharomyces cerevisiae. *Mol Cell Biol* 20, 3425-3433.

114. Dendouga, N., Gao, H., Moechars, D., Janicot, M., Vialard, J. and McGowan, C.H. (**2005**). Disruption of murine Mus81 increases genomic instability and DNA damage sensitivity but does not promote tumorigenesis. *Mol Cell Biol* 25, 7569-7579.

115. Ciccia, A., McDonald, N. and West, S. (**2008**). Structural and Functional Relationships of the XPF/MUS81 Family of Proteins. *Annu Rev Biochem* 77, 259-287.

116. Rothfuss, A. and Grompe, M. (**2004**). Repair kinetics of genomic interstrand DNA cross-links: evidence for DNA double-strand break-dependent activation of the Fanconi anemia/BRCA pathway. *Mol Cell Biol* 24, 123-134.

117. Bhagwat, N., Olsen, A.L., Wang, A.T., Hanada, K., Stuckert, P., Kanaar, R., D'Andrea, A., Niedernhofer, L.J. and McHugh, P.J. (**2009**). XPF-ERCC1 participates in the Fanconi anemia pathway of cross-link repair. *Mol Cell Biol* 29, 6427-6437.

118. McCabe, K.M., Hemphill, A., Akkari, Y., Jakobs, P.M., Pauw, D., Olson, S.B., Moses, R.E. and Grompe, M. (**2008**). ERCC1 is required for FANCD2 focus formation. *Mol Genet Metab* 95, 66-73.

119. Bergstralh, D.T. and Sekelsky, J. (**2008**). Interstrand crosslink repair: can XPF-ERCC1 be let off the hook? *Trends Genet* 24, 70-76.

120. Fekairi, S., Scaglione, S., Chahwan, C., Taylor, E.R., Tissier, A., Coulon, S., Dong, M.-Q., Ruse, C., Yates, J.R., Russell, P., Fuchs, R.P., McGowan, C.H. and Gaillard, P.-H.L. (**2009**). Human SLX4 is a Holliday junction resolvase subunit that binds multiple DNA repair/recombination endonucleases. *Cell* 138, 78-89.

121. Svendsen, J.M., Smogorzewska, A., Sowa, M.E., O'Connell, B.C., Gygi, S.P., Elledge, S.J. and Harper, J.W. (**2009**). Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair. *Cell* 138, 63-77.

122. Muñoz, I.M., Hain, K., Déclais, A.-C., Gardiner, M., Toh, G.W., Sanchez-Pulido, L., Heuckmann, J.M., Toth, R., Macartney, T., Eppink, B., Kanaar, R., Ponting, C.P., Lilley, D.M.J. and Rouse, J. (**2009**). Coordination of structure-specific nucleases by human SLX4/BTBD12 is required for DNA repair. *Mol Cell* 35, 116-127.

123. Andersen, S.L., Bergstralh, D.T., Kohl, K.P., LaRocque, J.R., Moore, C.B. and Sekelsky, J. (**2009**). Drosophila MUS312 and the vertebrate ortholog BTBD12 interact with DNA structure-specific endonucleases in DNA repair and recombination. *Mol Cell* 35, 128-135.

124. Loeb, L.A. and Monnat, R.J. (**2008**). DNA polymerases and human disease. *Nat Rev Genet* 9, 594-604.

125. Yang, W. and Woodgate, R. (**2007**). What a difference a decade makes: insights into translesion DNA synthesis. *Proc Natl Acad Sci USA* 104, 15591-15598.

126. Prakash, S., Johnson, R.E. and Prakash, L. (**2005**). Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. *Annu Rev Biochem* 74, 317-353.

127. Yamanaka, K., Minko, I.G., Takata, K.-I., Kolbanovskiy, A., Kozekov, I.D., Wood, R.D., Rizzo, C.J. and Lloyd, R.S. (**2010**). Novel enzymatic function of DNA polymerase nu in translesion DNA synthesis past major groove DNA-peptide and DNA-DNA cross-links. *Chem Res Toxicol* 23, 689-695.

128. Hinz, J.M. (**2010**). Role of Homologous Recombination in DNA Interstrand Crosslink Repair. *Environmental and molecular mutagenesis* in press

129. Muniandy, P.A., Liu, J., Majumdar, A., Liu, S.-t. and Seidman, M.M. (**2010**). DNA interstrand crosslink repair in mammalian cells: step by step. *Critical reviews in biochemistry and molecular biology* 45, 23-49.

130. Sarkar, S., Davies, A.A., Ulrich, H.D. and McHugh, P.J. (**2006**). DNA interstrand crosslink repair during G1 involves nucleotide excision repair and DNA polymerase zeta. *Embo J* 25, 1285-1294.

131. Wang, X., Peterson, C., Zheng, H., Nairn, R., Legerski, R. and Li, L. (**2001**). Involvement of nucleotide excision repair in a recombination-independent and error-prone pathway of DNA interstrand cross-link repair. *Mol Cell Biol* 21, 713-720.

132. Shen, X., Jun, S., O'Neal L, E., Sonoda, E., Bemark, M., Sale, J. and Li, L. (**2006**). REV3 and REV1 Play Major Roles in Recombination-independent Repair of DNA Interstrand Cross-links Mediated by Monoubiquitinated Proliferating Cell Nuclear Antigen (PCNA). *J Biol Chem* 281, 13869-13872.

133. Hlavin, E.M., Smeaton, M.B. and Miller, P.S. (**2010**). Initiation of DNA Interstrand Cross-link Repair in Mammalian Cells. *Environmental and molecular mutagenesis* in press

134. Nojima, K., Hochegger, H., Saberi, A., Fukushima, T., Kikuchi, K., Yoshimura, M., Orelli, B., Bishop, D., Hirano, S., Ohzeki, M., Ishiai, M., Yamamoto, K., Takata, M., Arakawa, H., Buerstedde, J., Yamazoe, M., Kawamoto, T., Araki, K., Takahashi, J., Hashimoto, N., Takeda, S. and Sonoda, E. (**2005**). Multiple repair pathways mediate tolerance to chemotherapeutic cross-linking agents in vertebrate cells. *Cancer Res* 65, 11704-11711.

135. Gillet, L.C.J. and Schärer, O.D. (**2006**). Molecular mechanisms of mammalian global genome nucleotide excision repair. *Chem Rev* 106, 253-276.

136. Hanawalt, P. and Spivak, G. (**2008**). Transcription-coupled DNA repair: two decades of progress and surprises. *Nat Rev Mol Cell Biol* 9, 958-970.

137. Muniandy, P.A., Thapa, D., Thazhathveetil, A.K., Liu, S.-t. and Seidman, M.M. (**2009**). Repair of laser-localized DNA interstrand cross-links in G1 phase mammalian cells. *J Biol Chem* 284, 27908-27917.

138. Zamble, D.B., Mu, D., Reardon, J.T., Sancar, A. and Lippard, S.J. (**1996**). Repair of cisplatin--DNA adducts by the mammalian excision nuclease. *Biochemistry* 35, 10004-10013.

139. Min, J.-H. and Pavletich, N.P. (**2007**). Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* 449, 570-575.

140. Schärer, O.D. (**2007**). Achieving broad substrate specificity in damage recognition by binding accessible nondamaged DNA. *Mol Cell* 28, 184-186.

141. Maillard, O., Camenisch, U., Blagoev, K.B. and Naegeli, H. (**2008**). Versatile protection from mutagenic DNA lesions conferred by bipartite recognition in nucleotide excision repair. *Mutat Res* 658, 271-286.

142. Zhao, J., Jain, A., Iyer, R.R., Modrich, P.L. and Vasquez, K.M. (**2009**). Mismatch repair and nucleotide excision repair proteins cooperate in the recognition of DNA interstrand crosslinks. *Nucleic Acids Res* 37, 4420-4429.

143. Zhang, Y., Wu, X., Guo, D., Rechkoblit, O., Geacintov, N.E. and Wang, Z. (**2002**). Two-step error-prone bypass of the (+)- and (-)-trans-anti-BPDE-N2-dG adducts by human DNA polymerases eta and kappa. *Mutat Res* 510, 23-35.

144. Bessho, T., Mu, D. and Sancar, A. (**1997**). Initiation of DNA interstrand cross-link repair in humans: the nucleotide excision repair system makes dual incisions 5' to the cross-linked base and removes a 22- to 28-nucleotide-long damage-free strand. *Mol Cell Biol* 17, 6822-6830.

145. Mu, D., Bessho, T., Nechev, L.V., Chen, D.J., Harris, T.M., Hearst, J.E. and Sancar, A. (**2000**). DNA interstrand cross-links induce futile repair synthesis in mammalian cell extracts. *Mol Cell Biol* 20, 2446-2454.

146. Martin, L.P., Hamilton, T.C. and Schilder, R.J. (**2008**). Platinum resistance: the role of DNA repair pathways. *Clin Cancer Res* 14, 1291-1295.

147. Spanswick, V.J., Hartley, J.M. and Hartley, J.A. (**2010**). Measurement of DNA interstrand crosslinking in individual cells using the Single Cell Gel Electrophoresis (Comet) assay. *Methods*

Mol Biol 613, 267-282.

148. Farmer, H., McCabe, N., Lord, C.J., Tutt, A.N.J., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C., Martin, N.M.B., Jackson, S.P., Smith, G.C.M. and Ashworth, A. (**2005**). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* 434, 917-921.

149. Bryant, H.E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N.J. and Helleday, T. (**2005**). Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* 434, 913-917.

150. Fong, P.C., Boss, D.S., Yap, T.A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O'Connor, M.J., Ashworth, A., Carmichael, J., Kaye, S.B., Schellens, J.H.M. and de Bono, J.S. (**2009**). Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N Engl J Med* 361, 123-134.

151. Gallmeier, E., Calhoun, E.S., Rago, C., Brody, J.R., Cunningham, S.C., Hucl, T., Gorospe, M., Kohli, M., Lengauer, C. and Kern, S.E. (**2006**). Targeted disruption of FANCC and FANCG in human cancer provides a preclinical model for specific therapeutic options. *Gastroenterology* 130, 2145-2154.

152. Gallmeier, E. and Kern, S.E. (**2007**). Targeting Fanconi anemia/BRCA2 pathway defects in cancer: the significance of preclinical pharmacogenomic models. *Clin Cancer Res* 13, 4-10.

153. Chirnomas, D., Taniguchi, T., de la Vega, M., Vaidya, A.P., Vasserman, M., Hartman, A.-R., Kennedy, R., Foster, R., Mahoney, J., Seiden, M.V. and D'Andrea, A.D. (**2006**). Chemosensitization to cisplatin by inhibitors of the Fanconi anemia/BRCA pathway. *Mol Cancer Ther* 5, 952-961.

154. Landais, I., Sobeck, A., Stone, S., LaChapelle, A. and Hoatlin, M.E. (**2009**). A novel cell-free screen identifies a potent inhibitor of the Fanconi anemia pathway. *Int J Cancer* 124, 783-792.

155. Kowal, P., Gurtan, A.M., Stuckert, P., D'Andrea, A.D. and Ellenberger, T. (**2007**). Structural determinants of human FANCF protein that function in the assembly of a DNA damage signaling complex. *J Biol Chem* 282, 2047-2055.

156. Orelli, B., McClendon, T.B., Tsodikov, O.V., Ellenberger, T.E., Niedernhofer, L.J. and Schärer, O.D. (**2010**). The XPA-binding domain of ERCC1 Is Required for Nucleotide Excision Repair but Not Other DNA Repair Pathways. *J Biol Chem* 285, 3705-3712.

157. Noll, D.M., Noronha, A.M., Wilds, C.J. and Miller, P.S. (**2004**). Preparation of interstrand cross-linked DNA oligonucleotide duplexes. *Front Biosci* 9, 421-437.

158. Warren, A.J. and Hamilton, J.W. (**1996**). Synthesis and structural characterization of the N2G-mitomycin C-N2G interstrand cross-link in a model synthetic 23 base pair oligonucleotide DNA duplex. *Chem Res Toxicol* 9, 1063-1071.

159. Akkari, Y., Bateman, R., Reifsteck, C., D'Andrea, A., Olson, S. and Grompe, M. (2001) The 4N cell cycle delay in Fanconi anemia reflects growth arrest in late S phase. In Mol Genet Metab ed.^eds), pp. 403-412

160. Ferentz, A., Keating, T. and Verdine, G. (**1993**). Synthesis and Characterization of Disulfide Cross-Linked Oligonucleotides. *J Am Chem Soc* 115, 9006-9014.

161. Erlanson, D.A., Chen, L. and Verdine, G.L. (**1993**). DNA methylation through a locally unpaired intermediate. *Journal of the American Chemical Society* 115, 12583–12584.

162. Dooley, P.A., Tsarouhtsis, D., Korbel, G.A., Nechev, L.V., Shearer, J., Zegar, I.S., Harris, C.M., Stone, M.P. and Harris, T.M. (**2001**). Structural studies of an oligodeoxynucleotide containing a trimethylene interstrand cross-link in a 5?-(CpG) motif: Model of a malondialdehyde cross-link. *Journal of the American Chemical Society* 123, 1730-1739.

163. Alzeer, J. and Schärer, O.D. (**2006**). A modified thymine for the synthesis of site-specific thymine-guanine DNA interstrand crosslinks. *Nucleic Acids Res*

164. Rink, S.M. and Hopkins, P.B. (**1995**). A mechlorethamine-induced DNA interstrand cross-link bends duplex DNA. *Biochemistry* 34, 1439-1445.

165. Dong, Q., Barskt, D., Colvin, M.E., Melius, C.F., Ludeman, S.M., Moravek, J.F., Colvin, O.M., Bigner, D.D., Modrich, P. and Friedman, H.S. (**1995**). A structural basis for a phosphoramide mustard-induced DNA interstrand cross-link at 5'-d(GAC). *Proceedings of the National Academy of Sciences of the United States of America* 92, 12170-12174.

166. Khullar, S., Varaprasad, C.V. and Johnson, F. (**1999**). Postsynthetic generation of a major acrolein adduct of 2'-deoxyguanosine in oligomeric DNA. *J Med Chem* 42, 947-950.

167. Ramzaeva, N. and Seela, F. (**1995**). 7-Substituted 7-deaza-2'-deoxyguanosines: Regioselective halogenation of pyrrolo[2,3-d]pyrimidine nucleosides. *Helvetica Chimica Acta* 78, 1083-1090.

168. Räschle, M., Knipscheer, P., Enoiu, M., Angelov, T., Sun, J., Griffith, J.D., Ellenberger, T.E., Schärer, O.D. and Walter, J.C. (**2008**). Mechanism of replication-coupled DNA interstrand crosslink repair. *Cell* 134, 969-980.

169. Hoffer, M. (**1960**). Alpha-Thymidin. *Chemische Berichte-Recueil* 93, 2777-2781.

170. Davoll, J. (**1960**). Pyrrolo[2,3-D]Pyrimidines. *Journal of the Chemical Society*, 131-138.

171. Storek, M.J., Ernst, A. and Verdine, G.L. (**2002**). High-resolution footprinting of sequence-specific protein-DNA contacts. *Nat Biotechnol* 20, 183-186.

172. Schärer, O.D. (**2005**). DNA interstrand crosslinks: natural and drug-induced DNA adducts that induce unique cellular responses. *Chembiochem* 6, 27-32.

173. Harwood, E., Sigurdsson, S., Edfeldt, N., Reid, B. and Hopkins, P. (**1999**). Chemical synthesis and preliminary structural characterization of a nitrous acid interstrand cross-linked duplex DNA. *J Am Chem Soc* 121, 5081-5082.

174. Wilds, C.J., Noronha, A.M., Robidoux, S. and Miller, P.S. (**2004**). Mispair-aligned N3T-alkyl-N3T interstrand cross-linked DNA: Synthesis and characterization of duplexes with interstrand cross-links of variable lengths. *Journal of the American Chemical Society* 126, 9257-9265.

175. Ho, T.V. and Schärer, O.D. (**2010**). Translesion DNA synthesis polymerases in DNA interstrand crosslink repair. *Environmental and molecular mutagenesis*, NA-NA.

176. Lee, S., Bowman, B.R., Ueno, Y., Wang, S. and Verdine, G.L. (**2008**). Synthesis and structure of duplex DNA containing the genotoxic nucleobase lesion N7-methylguanine. *J Am Chem Soc* 130, 11570-11571.

177. Millard, J.T., Raucher, S. and Hopkins, P.B. (**1990**). Mechlorethamine cross-links deoxyguanosine residues at 5'-GNC sequences in duplex DNA fragments. *Journal of the American Chemical Society* 112, 2459-2460.

178. Rink, S.M., Solomon, M.S., Taylor, M.J., Rajur, S.B., Mclaughlin, L.W. and Hopkins, P.B. (**1993**). Covalent Structure of a Nitrogen Mustard-Induced DNA Interstrand Cross-Link - an N7-to-N7 Linkage of Deoxyguanosine Residues at the Duplex Sequence 5'-d(GNC). *Journal of the American Chemical Society* 115, 2551-2557.

179. Case, D.A., Cheatham, T.E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M., Jr., Onufriev, A., Simmerling, C., Wang, B. and Woods, R.J. (**2005**). The Amber biomolecular simulation programs. *J Comput Chem* 26, 1668-1688.

180. Lavery, R. and Sklenar, H. (**1989**). Defining the structure of irregular nucleic acids: conventions and principles. *J Biomol Struct Dyn* 6, 655-667.

181. Humphrey, W., Dalke, A. and Schulten, K. (**1996**). VMD: visual molecular dynamics. *J Mol Graph* 14, 33-38, 27-38.

182. Knipscheer, P., Räschle, M., Smogorzewska, A., Enoiu, M., Ho, T.V., Schärer, O.D., Elledge, S.J. and Walter, J.C. (**2009**). The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. *Science* 326, 1698-1701.

183. Song, K., Hornak, V., De Los Santos, C., Grollman, A.P. and Simmerling, C. (**2008**). Molecular mechanics parameters for the FapydG DNA lesion. *J Comput Chem* 29, 17-23.

184. Wang, J., Cieplak, P. and Kollman, P. (**2000**). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem* 21, 1049-1074.

185. Cieplak, P., Cornell, W., Bayly, C. and Kollman, P. (**1995**). Application of the multimolecule and multiconformational resp methodology to biopolymers - charge derivation for dna, rna, and proteins. *J Comput Chem* 16, 1357-1377.

186. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. and Simmerling, C. (**2006**). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65, 712-725.

187. Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R. and Klein, M. (**1983**). Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79, 926-935.

188. Ryckaert, J., Ciccotti, G. and Berendsen, H. (**1977**). Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of n-alkanes. *J Comput Phys* 23, 327-341.

189. Berendsen, H., Postma, J., VanGunsteren, W., Dinola, A. and Haak, J. (**1984**). Molecular-dynamics with coupling to an external bath. *J Chem Phys* 81, 3684-3690.

190. Sugasawa, K., Ng, J.M., Masutani, C., Iwai, S., van der Spek, P.J., Eker, A.P., Hanaoka, F., Bootsma, D. and Hoeijmakers, J.H. (**1998**). Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Mol Cell* 2, 223-232.

191. Wu, H.I., Brown, J.A., Dorie, M.J., Lazzeroni, L. and Brown, J.M. (**2004**). Genome-wide

identification of genes conferring resistance to the anticancer agents cisplatin, oxaliplatin, and mitomycin C. *Cancer Res* 64, 3940-3948.

192. Ho, T.V., Guainazzi, A., Enoiu, M., Derkunt, B.S. and Schärer, O.D. Structure-dependent bypass of DNA interstrand crosslinks by tranlesion synthesis polymerases. Manuscript in preparation.

193. Lukin, M. and De Los Santos, C. (**2006**). NMR structures of damaged DNA. *Chem Rev* 106, 607-686.

194. De los Santos, C. (**1999**). Probing DNA structure by NMR spectroscopy. *Comprehensive Natural Products Chemistry* 7, 55-80.

195. Wütrich, K. (1986) NMR of Proteins and NucleicAcids, Wiley-Interscience, Wiley & Sons. New York.

196. Bohon, J. and de los Santos, C.R. (**2003**). Structural effect of the anticancer agent 6-thioguanine on duplex DNA. *Nucleic Acids Res* 31, 1331-1338.

197. Cullinan, D., Johnson, F., Grollman, A.P., Eisenberg, M. and De Los Santos, C. (**1997**). Solution structure of a DNA duplex containing the exocyclic lesion 3,N4-etheno-2'-deoxycytidine opposite 2'-deoxyguanosine. *Biochemistry* 36, 11933-11943.

198. Goddard, T.D. and Kneller, D.G. SPARKY 3. *University of California, San Francisco*

199. Nielsen, J., Brill, W. and Caruthers, M. (**1988**). Synthesis and characterization of dinucleoside phosphorodithioates. *Tetrahedron Lett* 29, 2911-2914.