

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Replication of Bulky DNA Adducts

Replikation von sterisch anspruchsvollen DNA Addukten

Stephanie Schorr

aus

Frankfurt am Main

2010

Erklärung

Diese Dissertation wurde im Sinne von § 13 Abs. 3 bzw. 4 der Promotionsordnung vom 29. Januar 1998 von Prof. Dr. Thomas Carell betreut.

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbstständig, ohne unerlaubte Hilfe erarbeitet.

München, den 09.11.2010

A handwritten signature in black ink, reading 'Stephanie Schorr', written over a horizontal line.

(Stephanie Schorr)

Dissertation eingereicht am 09.11.2010

1. Gutachter: Prof. Dr. Thomas Carell
2. Gutachter: Prof. Dr. Karl-Peter Hopfner

Mündliche Prüfung am 16.12.2010

*Wer an der Küste bleibt,
kann keine neuen Ozeane entdecken.*

(Thor Heyerdahl)

Parts of this thesis were published or presented on conferences:

Publications

S. Schorr, S. Schneider, K. Lammens, K.-P. Hopfner, T. Carell “Mechanism of replication blocking and bypass of Y-family polymerase η by bulky acetylaminofluorene DNA adducts” *PNAS* **2010**, 107, 20720-20725.

S. Schorr, T. Carell “Mechanism of Acetylaminofluorene-dG induced Frameshifting by Polymerase η ” *ChemBioChem* **2010**, 11, 2534 – 2537.

T. Reißner, S. Schneider, S. Schorr, T. Carell “Crystal Structure of a Cisplatin-(1,3-GTG) Cross-Link within DNA Polymerase η ” *Angew. Chem. Int. Ed.* **2010**, 49, 3077 – 3080.

S. Schneider, S. Schorr, T. Carell, “Crystal structure analysis of DNA lesion repair and tolerance mechanisms” *Curr. Opin. Struct. Biol.* **2009**, 19, 87-95.

T. Reißner, S. Schorr, and T. Carell “Once Overlooked, Now Made Visible: ATL Proteins and DNA Repair” *Angew. Chem. Int. Ed.* **2009**, 48, 7293 – 7295.

conference presentations

Poster presentation: *Translesion synthesis through aromatic amine DNA adducts mediated by the eukaryotic DNA polymerase η* , FASEB conference “Nucleic Acid Enzymes”, Saxtons River, Vermont, USA, **2010**.

Poster presentation: *Translesion synthesis through aromatic amine DNA adducts mediated by the Eukaryotic DNA Polymerase η* , Days of Molecular Medicine 2010, Stockholm, Sweden, **2010**.

Oral presentation: *Lesion-polymerase complexes and frameshift mutations – 2-acetylaminofluorene-deoxyguanosine adducts and the eukaryotic polymerase η* , EMBO practical course on high throughput methods for protein production and crystallization, Oxford, UK, **2009**.

Poster presentation: *Lesion-polymerase complexes and frameshift mutations - bulky 2'-deoxyguanosine adducts and the eukaryotic DNA polymerase η* , 3rd European Conference on Chemistry for Life Sciences, Frankfurt/Main, Germany, **2009**.

Poster presentation: *2-Acetylaminofluorene-deoxyguanosine adducts and the eukaryotic DNA polymerase η* , SFB 646 meeting, Munich, Germany, **2008**.

Oral presentation: *Lesion-polymerase complexes and frameshift mutations*, regional fellowship holder meeting of the *Fonds der Chemischen Industrie*, Garching, Germany, **2008**.

Poster presentation: *Lesion-polymerase complexes and frameshift mutations – 2-acetylaminofluorene-deoxyguanosine adducts and the eukaryotic DNA polymerase η* , Bayer PhD Student Course, Cologne, Germany, **2008**.

Poster presentation: *Synthetic 2-acetylaminofluorene-deoxyguanosine adducts as a tool to investigate DNA frameshift mutations*, XII. Blue Danube Symposium on Heterocyclic Chemistry, Tihany, Hungary, **2008**.

Poster presentation: *Lesion-polymerase complexes and frameshift mutations – 2-acetylaminofluorene-deoxyguanosine adducts and the eukaryotic DNA polymerase η* , 2nd European Conference on Chemistry for Life Sciences, Wroclaw, Poland, **2008**.

Danksagung

An erster Stelle gilt mein Dank meinem Doktorvater *Prof. Dr. Thomas Carell* für die spannende und vielseitige Themenstellung, seine Begeisterungsfähigkeit, sowie für seine große Unterstützung und Förderung. Ich möchte mich zudem für die exzellenten Arbeitsbedingungen, die angenehme Atmosphäre und für die wissenschaftliche Freiheit, die er mir gewährt hat, bedanken.

Den Mitgliedern meiner Prüfungskommission danke ich für ihre Bemühungen bei der Evaluierung meiner Arbeit. Insbesondere bedanke ich mich bei *Prof. Dr. Karl-Peter Hopfner* für die Übernahme des Koreferats.

Während meiner Dissertation wurde ich vom *Fonds der Chemischen Industrie* finanziell unterstützt, wofür ich mich ebenfalls sehr bedanken möchte.

Frau Slava Gärtner danke ich für ihre Hilfe in allen organisatorischen Belangen.

Außerdem geht mein Dank an *Frau Sabine Voß* für ihre ausgezeichnete Organisation im Labor, ihre Hilfe bei der Synthese und die gute Atmosphäre im Labor. Bei den Mitarbeitern der Zentralen Analytik bedanke ich mich für das zuverlässige und schnelle Messen meiner Proben.

Mein großer Dank gilt *Dr. Sabine Schneider* für die Hilfe, Anleitung und Unterstützung in der Kristallographie und Biochemie und für ihre Fähigkeit, mich immer wieder zu motivieren. Außerdem möchte ich ihr für die zahlreichen Synchrotron-Trips in die Schweiz oder nach Grenoble danken, die durch kleine Wander/Klettereinlagen und viele Pizzen auch dann Spass gemacht haben, wenn die Messungen nicht erfolgreich waren!

Für das Korrekturlesen dieser Arbeit danke ich *Dr. Sabine Schneider* und *Martin Münzel*. Weiterhin möchte ich mich bei der gesamten Arbeitsgruppe Carell für die gute Atmosphäre bedanken. Ich habe hier gute Freunde fürs Leben gefunden.

Bei *Dr. Markus Müller* möchte ich mich für seine Hilfe in biochemischen und computertechnischen Fragen bedanken. Ebenso danke ich den Mitgliedern meines Labors für die gute Atmosphäre und ihre Hilfsbereitschaft. Hier gilt mein Dank insbesondere meinem „Boxnachbarn“ *Dr. Malte Winnacker* für seine ruhige Zusammenarbeit im Labor und für seine nicht ganz so ruhige Art bei Ausflügen und Unternehmungen. Tatkräftige Unterstützung in der Synthese und Biochemie habe ich von den F-Praktikantinnen/Bachelor-Studentinnen *Karola Gasteiger, Bianca Wöhrl, Tanja Schmid* und *Patricia Braun* erhalten.

Dr. Ralf Strasser und *Sandra Koch* danke ich für die gute Zusammenarbeit auf dem Rad14/XPA Projekt. Bei *Dr. Christian Trindler, Martin Münzel* und *Dr. Heiko Müller* möchte ich mich für ihre Hilfe in synthetischen Fragen bedanken. Hilfreiche Tipps und

Diskussionen von und mit *Dr. Melanie Maul, Dr. Philipp Gramlich, Thomas Reißner, Korbinian Heil, Tim Gehrke* und *Emine Kaya* haben ebenfalls die Entwicklung dieser Arbeit begleitet.

Für die leckere und lustige Abende mit Schnitzel, Nudeln, Computer oder Gitarre möchte ich mich bei *Christian Trindler, Thomas Reißner, Aline Gégout, Martin Münzel, Markus Müller, Christian Deiml* und *Corinna Kaul* bedanken. *Melanie Maul, Emine Kaya* und *Christian Trindler* sorgten mit zahlreichen Aktivitäten in Fels und Eis bzw. Schnee oder auf dem Wasser für den nötigen sportlichen Ausgleich neben der wissenschaftlichen Arbeit.

Ganz herzlich bedanke ich mich bei meinen Eltern, die stets für mich da sind. Vielen Dank für eure uneingeschränkte Unterstützung und euer Vertrauen in mich, was die Entstehung dieser Arbeit überhaupt erst ermöglicht hat. Zuletzt danke ich *David Kath* für sein großes Verständnis, seine Geduld und seine fortwährende Unterstützung.

TABLE OF CONTENTS

1	SUMMARY	VIII
2	ZUSAMMENFASSUNG	XIV
3	INTRODUCTION	1
3.1	CHEMICALLY INDUCED DNA DAMAGE	1
3.2	FORMATION OF BULKY DNA ADDUCTS	3
3.2.1	<i>Formation of DNA adducts by aromatic amines</i>	3
3.2.2	<i>Formation of other bulky adducts</i>	6
3.3	THE BULKY C8-dG ADDUCTS ARE MUTAGENIC	7
3.4	EUKARYOTIC LOW FIDELITY DNA POLYMERASES	9
3.4.1	<i>Polymerase switch</i>	10
3.4.2	<i>Pol η</i>	11
3.4.3	<i>Pol κ</i>	12
3.4.4	<i>Pol ι</i>	13
3.4.5	<i>Pol ζ</i>	14
3.4.6	<i>Rev1</i>	14
3.5	CRYSTALLOGRAPHIC SNAPSHOTS OF BULKY DNA ADDUCT BLOCKAGE AND BYPASS	15
3.6	REPAIR OF BULKY DNA ADDUCTS	18
4	AIMS OF THE PROJECT	20
5	SYNTHESIS OF C8-DEOXYGUANOSINE ADDUCTS	22
6	INCORPORATION OF C8- dG ADDUCTS INTO OLIGONUCLEOTIDES	29
6.1	SOLID PHASE DNA SYNTHESIS	29
6.2	PURIFICATION AND CHARACTERIZATION OF THE VARIOUS C8-dG ADDUCT CONTAINING OLIGONUCLEOTIDES	31
6.2.1	<i>HPLC purification and characterization via MALDI-TOF mass spectrometry</i>	31
6.2.2	<i>Enzymatic digestion analysis</i>	33
6.2.3	<i>Melting point analysis</i>	38
7	BYPASS OF BULKY DNA ADDUCTS BY Y-FAMILY DNA POLYMERASES	40
7.1	TRANSLATION SYNTHESIS OF BULKY DNA ADDUCTS BY POL η	40
7.1.1	<i>Purification of yeast Pol η</i>	40
7.1.2	<i>Mechanism of replication blocking and bypass of Pol η by AAF-dG adducts</i>	41
7.1.3	<i>Crystallization experiments with other bulky adducts</i>	54
7.1.4	<i>Mechanism of acetylaminofluorene-dG induced frameshifting</i>	56
7.2	BYPASS OF BULKY DNA ADDUCTS BY POL κ	64
7.2.1	<i>Cloning, expression and purification of human Pol κ</i>	64
7.2.2	<i>Primer extension studies with Pol κ</i>	65
7.2.3	<i>Preliminary crystallization screens with C8-dG containing oligonucleotides and the low fidelity polymerase κ</i>	69

8	CO-CRYSTALLIZATION OF AAF-dG CONTAINING OLIGONUCLEOTIDES WITH THE NER PROTEINS RAD14 FROM YEAST AND ITS HUMAN HOMOLOGUE XPA	70
8.1	PURIFICATION OF RAD14 AND XPA	70
8.2	CRYSTALLIZATION OF RAD14/XPA IN COMPLEX WITH AN AAF-dG CONTAINING DNA DUPLEX	71
8.3	PHASE DETERMINATION FOR STRUCTURE SOLUTION OF THE RAD14 CRYSTALS	73
8.4	DISCUSSION	75
9	CONCLUSION AND OUTLOOK	76
10	EXPERIMENTAL	79
10.1	GENERAL CHEMICAL MATERIALS AN METHODS	79
10.2	OLIGONUCLEOTIDE SYNTHESIS AND CHARACTERIZATION	81
10.2.1	<i>Automated DNA synthesis</i>	81
10.2.2	<i>Purification of bulky adduct containing oligonucleotides</i>	82
10.2.3	<i>Preparation of AF-dG containing oligonucleotides</i>	84
10.2.4	<i>DNA Quantification</i>	84
10.2.5	<i>Melting point measurements</i>	85
10.2.6	<i>Enzymatic DNA digestion</i>	85
10.3	BIOCHEMICAL MATERIALS	86
10.3.1	<i>Buffers</i>	86
10.3.2	<i>Equipment</i>	88
10.3.3	<i>Bacterial strains</i>	89
10.3.4	<i>Media</i>	89
10.3.5	<i>Solutions</i>	90
10.3.6	<i>Antibiotics and Inducers</i>	90
10.3.7	<i>Enzymes</i>	91
10.4	BIOCHEMICAL METHODS	91
10.4.1	<i>Agarose gel electrophoresis</i>	91
10.4.2	<i>Polyacrylamide gel electrophoresis</i>	91
10.4.3	<i>Polymerase chain reaction (PCR)</i>	92
10.4.4	<i>Assembly of the expression vectors for Pol κ</i>	94
10.4.5	<i>Transformation of E. coli</i>	95
10.4.6	<i>E. coli cultures</i>	95
10.4.7	<i>Isolation of plasmid DNA</i>	95
10.4.8	<i>Induction of the protein expression</i>	96
10.4.9	<i>Protein purification</i>	98
10.4.10	<i>Protein quantification</i>	100
10.4.11	<i>Primer extension studies</i>	100
10.4.12	<i>Mass spectrometry analysis of the primer extension experiments</i>	101
10.5	PROTEIN CRYSTALLIZATION	101
10.6	COLLECTION AND PROCESSING OF X-RAY DIFFRACTION DATA, PHASE DETERMINATION AND STRUCTURE REFINEMENT	104

10.7	SYNTHESIS OF BULKY DNA ADDUCTS	106
10.7.1	8-Bromo-2'-deoxyguanosine	106
10.7.2	8-Bromo-3',5'-O-bis(tert-butyltrimethylsilyl)-2'-deoxyguanosine	107
10.7.3	O ⁶ -Benzyl-8-bromo-3',5'-O-bis(tert-butyltrimethylsilyl)-2'-deoxyguanosine	108
10.7.4	O ⁶ -Benzyl-8-bromo-3',5'-O-bis(tert-butyltrimethylsilyl)-N ² -dimethoxytrityl-2'-deoxyguanosine	109
10.7.5	Coupling of the protected deoxyguanosine with the aromatic unit	110
10.7.6	N-Acylation at C8-position of the aromatic amine dG adducts	115
10.7.7	Protection of the N ² -position of the aromatic amine dG adducts	120
10.7.8	Deprotection of the hydroxyl groups of the aromatic amine dG adducts	125
10.7.9	Deprotection of the O ⁶ -position of the aromatic amine dG adducts	130
10.7.10	5'-DMT protection of the bulky dG adducts	135
10.7.11	Synthesis of the phosphoramidites	140
10.7.12	(4-Isopropyl)-phenoxy acetyl chloride	145
10.7.13	Phenoxyacetic anhydride	145
11	APPENDIX	146
11.1	CRYSTAL STRUCTURE DATA	146
11.2	MALDI-TOF MASS SPECTROMETRY ANALYSIS OF THE PRIMER EXTENSION PRODUCTS WITH AAF-dG CONTAINING OLIGONUCLEOTIDES	147
11.3	COMPOSITION OF THE LEMASTER MEDIUM ^[164]	148
11.4	LIST OF ABBREVIATIONS	150
12	REFERENCES	152

1 Summary

Aromatic amines are known to be strong carcinogens. After metabolic activation, they react as electrophilic arylnitrenium ions with nucleophilic functionalities of the DNA duplex interfering and disrupting DNA and RNA synthesis and leading to mutations. Preferred reaction sites are the amino groups of adenine and guanine and particularly the C8-position of guanine. During metabolic activation, aromatic amines are enzymatically acetylated at N^8 position.^[16, 17] The non-acetylated lesions reduce the replication efficiency, but are in general faithfully bypassed by high fidelity polymerases. In contrast, the acetylated derivatives block replicative polymerases but can be bypassed with special low-fidelity polymerases.^[51, 120, 122]

The translesion synthesis DNA polymerase η , for instance, is able to bypass C8 bulky adduct lesions such as the widely studied 2-aminofluorene-dG (AF-dG, Figure 1, **15**) and its acetylated analogue (AAF-dG, Figure 1, **14**) mainly in an error-free manner.^[56-58, 91] The distinct mutagenic properties of the acetylated and non-acetylated aromatic amine lesions are presumably caused by their different conformational preferences. While the non-acetylated lesions exist in both *syn* and *anti* conformation, the corresponding acetylated lesion seems to adopt the *syn*-conformation with high preference (Figure 1).^[52-54] The mechanism that allows low-fidelity polymerases such as Pol η to replicate past acetylated AAF-dG lesions is still unknown. Computational modeling studies of the Y-family polymerase Dpo4 suggested that the bypass reaction requires rotation of the lesion into the base pairing *anti* conformation.^[123] However, in light of the strong *syn* preference of the adducted base, this rotation is energetically demanding, which may force the enzyme to use a different, still unknown mechanism for accurate lesion bypass.

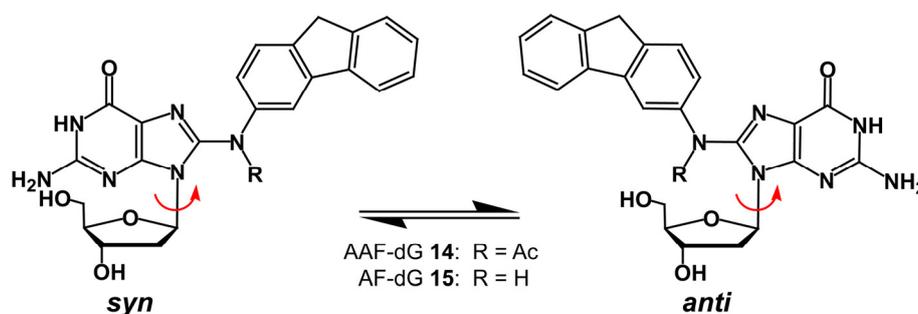


Figure 1. *Syn* and *anti* conformation of the C8-AF/AAF-dG adducts. The acetylated C8-adduct (R = Ac, AAF-dG) is restricted to *syn*-conformation.

In this thesis work, the mechanism of the error-free bypass of acetylated aromatic amine dG adducts such as the acetylaminofluorene-dG (AAF-dG) by Pol η is investigated. For that reason, AAF-dG as well as the guanine adducts of other aromatic amines (aniline, 2-aminonaphthalene, 2-aminoanthracene and 1-aminopyrene) were synthesized and converted to phosphoramidites, suitably protected for DNA synthesis (Figure 2).

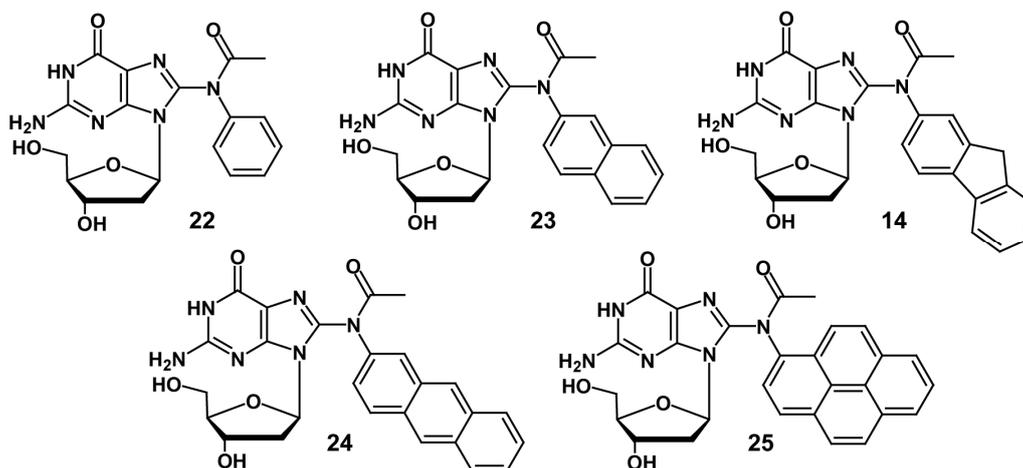


Figure 2. C8-dG adducts: AAB-dG **22**, AAN-dG **23**, AAF-dG **14**, AAA-dG **24** and AAP-dG **25**. Derivatives, suitably protected for solid-phase DNA synthesis, were prepared in this thesis work and incorporated into oligonucleotides.

The synthesized dG adducts were successfully incorporated into various oligonucleotides via automated solid-phase DNA synthesis. The oligonucleotides were purified by HPL-chromatography and characterized via MALDI-Tof mass spectrometry and enzymatic digestion analysis.

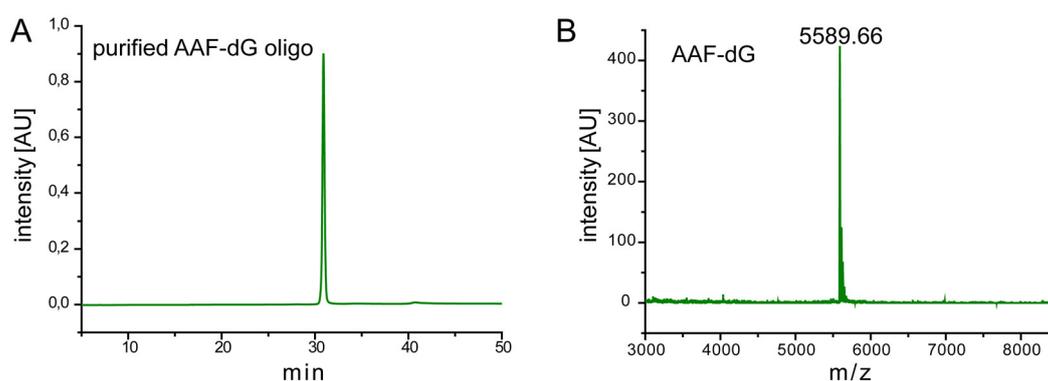


Figure 3. Example of the purification and characterization of bulky adduct containing oligonucleotides. A) HPL-chromatogram of a purified oligonucleotide containing the AAF-dG lesion. B) MALDI-Tof mass spectrum of the purified AAF-dG oligonucleotide.

The synthesized bulky adduct containing oligonucleotides were used for crystallization and for primer extension studies with the translesion synthesis (TLS) polymerase η from *S. cerevisiae* in order to get insights into the bypass mechanism of bulky adducts by this enzyme. In the present work, structural evidence is provided that yeast Pol η bypasses the bulky adducts AAF-dG and 2-acetylaminanthracene-dG (AAA-dG) by rotation of the DNA around the bulky moiety, while keeping the AAF-dG in *syn* conformation. Two crystal structures of AAF-dG and AAA-dG containing DNA in complex with Pol η were obtained. There are two Pol η molecules in the asymmetric unit cell of both crystals, showing the enzyme-DNA complexes in two different states (A and B). In complex A, the aromatic units stack on top of the Watson-Crick base pair formed by the dG:dC base pair 3' to the lesion (Figure 4). The bulky aromatic units therefore fully block the active site for the incoming dCTP.

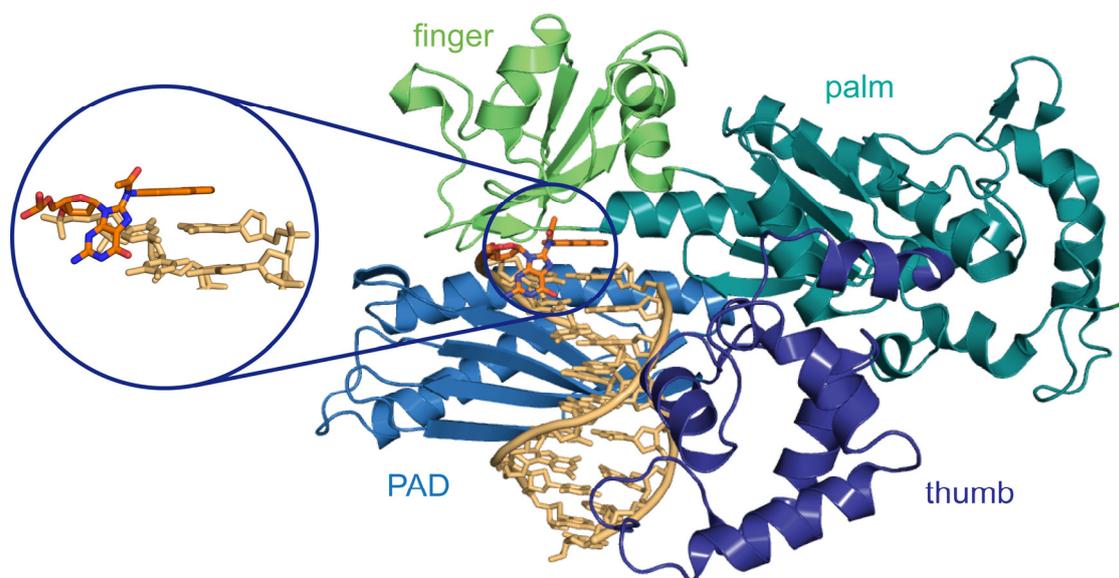


Figure 4. Crystal structure of *S. cerevisiae* Pol η with the AAF-dG oligonucleotide in complex A. The overall folding topology is depicted with the enzyme as cartoon representation and the DNA as stick model (gold) and the AAF-dG adduct in orange.

In complex B the situation is surprisingly different (Figure 5 B). First of all, a relatively small movement of the DNA template strand compared to complex A is observed. However, the primer DNA strand is rotated in complex B relative to the aromatic moiety resulting in a displacement of the primer end by 6 Å. In both structures, this primer rotation brings the dG part into the active site despite their prevalent *syn*-conformation. Thus the dG-lesion packs above the Watson-Crick base pair formed by the primer end and the template strand. The primer strand displacement results in a partial opening of the active site for the putative dCTP to bind. Hence TLS is accomplished without rotation of the lesion into the *anti* conformation

as previously thought. In this situation correct dCTP incorporation opposite the lesion can be envisioned by formation of one hydrogen bond between the dCTP and the lesion (Figure 5 C).

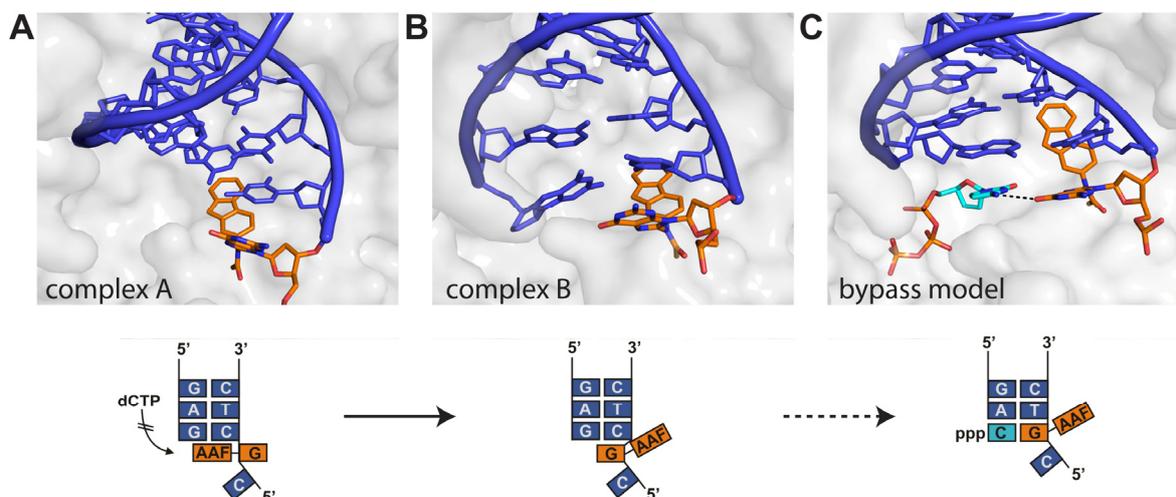


Figure 5. Model for AAF-dG bypass by Pol η . A) In complex A, the fluorene moiety of the AAF-dG adduct stacks onto the previous Watson-Crick base pair, blocking the active site for the incoming nucleotide. B) In contrast, the DNA in complex B is rotated in respect to the aromatic unit, with the adducted dG moved towards the templating position, partially freeing the active site. C) Hypothetical model of the AAF bypass by Pol η . The dG : dC basepair 3' of the lesion in the AAF-dG-Pol η complex B was replaced by the AAF-dG adduct and the dCTP triphosphate. Thus the active site for the incoming dCTP opens, which may form one critical hydrogen bond to the adducted nucleobase in *syn* conformation.

The model was further supported by biochemical studies using synthetic zebularine triphosphate which is not incorporated opposite the AAF-dG lesion due to the lacking ability to form the critical hydrogen bond. Moreover, TLS of dG adducts containing aromatic units of different sizes (benzene, naphthalene, pyrene) was investigated. It could be shown that the TLS efficiency of Pol η depends critically on the size of the bulky adduct forming the lesion.

Another part of this thesis work involves the investigation of frameshift mutations. Bulky adduct DNA lesions, such as the AAF-dG lesion, induce frameshift mutations if they are placed in special gene sequences.^[16, 53, 61-64] By inserting the AAF-dG lesion into different repetitive frameshift prone sequences and performing primer extension studies, the molecular basis of the frameshifting event was determined. Primer extension studies were analyzed using PAA-gels. The composition of the elongated primer strands was determined by MALDI-Tof mass spectrometry. It could be shown that frameshifting starts with correct dCTP incorporation opposite the lesion. Then, the DNA polymerase induces a primer-slippage followed by primer misalignment and extension of the misaligned primer-template complex.

The mechanisms of -I, -II and -III frameshifts were deciphered. An example is shown in Figure 6.

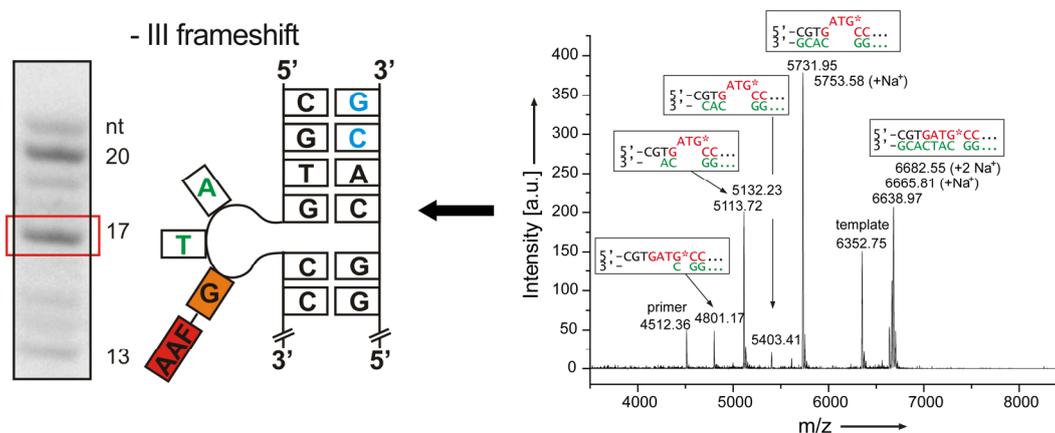


Figure 6. Schematic overview of the sequence dependent bypass products occurring during primer extension of AAF-dG adducts of a sequence allowing a -III frameshift. The MALDI-ToF mass spectrum (right) confirms the observed frameshifting products observed on the PAA gel (left).

In an additional part of this thesis, the catalytic core of the human low fidelity polymerase κ was cloned, overexpressed and purified for primer extension studies. It could be shown, that Pol κ is, like Pol η , able to replicate across the synthesized C8 bulky DNA adducts. However, bypass seems to be more error-prone for Pol κ , the incorporation of nucleotides opposite the AAF-dG lesion is not selective. First crystallization screenings were performed with Pol κ in complex with an AAF-dG oligonucleotide in order to get insights into the bypass mechanism of bulky adducts by Pol κ .

This thesis work also includes crystallization of the yeast NER protein Rad14 and its human homologue XPA together with AAF-dG containing oligonucleotides. The mechanism of lesion recognition in the NER pathway is still not well understood. It has been shown that the Rad14 and XPA bind specifically to DNA strands containing bulky adducts.^[139] In this thesis, the minimal DNA binding domains were overexpressed, purified and crystallized in complex with AAF-dG containing DNA. Crystals of Rad14 could be obtained diffracting X-rays to 3 – 4.5 Å. As no Rad14 crystal structure is known, phase information had to be obtained using heavy atom derivatized protein/DNA complex crystals. Selenomethionine labeled Rad14 was prepared and crystals were grown with either selenomethionine labeled protein or DNA containing 5-Br-dU. Some phase information was derived from SAD experiments using these crystals. However, structure solution was not successful up to now due to anisotropic diffraction and radiation damage.

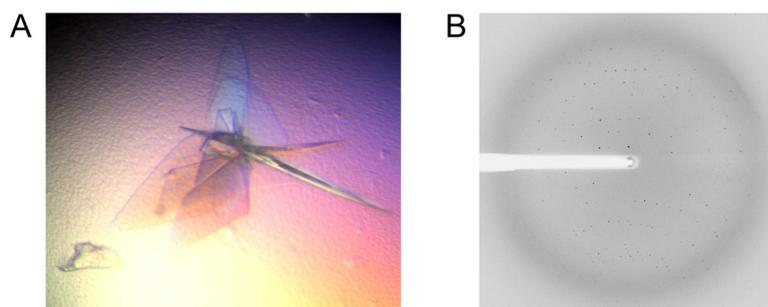


Figure 7. A) Crystals of Rad14 in complex with AAF-dG containing DNA. B) Diffraction images of a Rad14 crystal in complex with an AAF-dG containing DNA duplex taken at the microfocus beamline ID23-2 at the ESRF (France).

In summary, the combination of organic synthesis, biochemical studies and macromolecular crystallography in this thesis work provides mechanistic insights into the error-free and error-prone bypass of bulky DNA adducts during replication.

2 Zusammenfassung

Aromatische Amine sind starke Karzinogene. Nach metabolischer Aktivierung reagieren sie als elektrophile Arylnitreniumionen mit nucleophilen Gruppen der DNA Doppelhelix. Eine Störung der DNA und RNA Synthese durch so gebildete Addukte kann zu Mutationen im Genom führen. Aromatische Amine reagieren bevorzugt mit den Aminogruppen der Basen Adenin und Guanin und vor allem mit der C8-Position von Guanin. Während der metabolischen Aktivierung werden aromatische Amine enzymatisch an der N^8 Position acetyliert.^[16, 17] Nicht acetylierte Adduktschäden reduzieren die Replikationseffizienz von DNA Polymerasen, werden aber normalerweise korrekt überlesen. Im Gegensatz hierzu stellen die acetylierten Derivate eine Blockade für replikative Polymerasen dar und müssen von speziellen, sogenannten *low-fidelity* Polymerasen überlesen werden.^[51, 120, 122]

Die DNA Polymerase η , beteiligt an der Translesion-Synthese (TLS), ist in der Lage, sterisch anspruchsvolle C8-Addukte, wie zum Beispiel das gut untersuchte 2-Amino-fluoren-dG (AF-dG, Figure 1, **15**) und sein acetyliertes Derivat (AAF-dG, Figure 1, **14**), meist fehlerfrei zu überlesen.^[56-58, 91] Die unterschiedlichen mutagenen Eigenschaften der acetylierten und nicht acetylierten Adduktschäden werden vermutlich durch ihre unterschiedlichen konformellen Präferenzen hervorgerufen. Nicht acetylierte Schäden können in *syn* und *anti* Konformation auftreten, die entsprechenden acetylierten Addukte scheinen mit hoher Präferenz in *syn*-Konformation vorzuliegen (Abbildung 1).^[52-54]

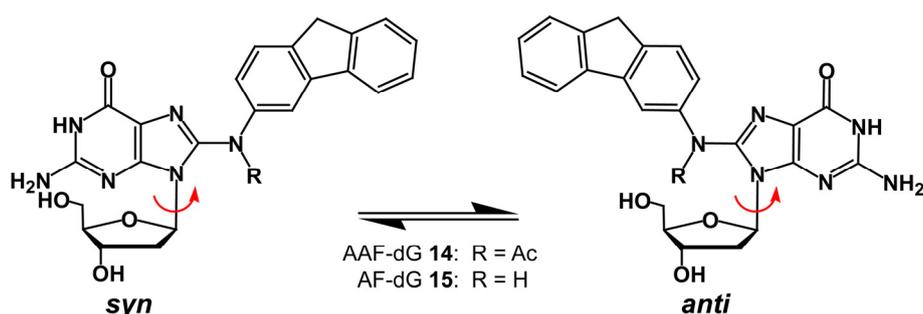


Abbildung 1. *Syn* und *anti* Konformation der C8-AF/AAF-dG Addukte. Das acetylierte C8-Addukt (R = Ac, AAF-dG) liegt nur in *syn*-Konformation vor.

Der Mechanismus, der es *low-fidelity* Polymerasen wie Pol η erlaubt, über acetylierte Adduktschäden wie AAF-dG Schäden hinwegzulesen, ist unbekannt. Gestützt auf molekulardynamische Rechnungen mit der zur Y-Familie gehörenden Polymerase Dpo4 wurde vorgeschlagen, dass für das Überlesen der Adduktschäden eine Rotation des Schadens

in die *anti*-Konformation zur Ausbildung der Watson-Crick Basenpaarung notwendig ist.^[123] Aufgrund der starken *syn* Präferenz der geschädigten Base ist dies energetisch ungünstig. Daher verwendet das Enzym vermutlich einen anderen, noch unbekanntem Mechanismus für das akkurate Überlesen der Adduktschäden.

In dieser Arbeit wird die fehlerfreie Replikation von acetylierten dG Addukten wie Acetylaminofluoren-dG (AAF-dG) durch Pol η untersucht. Hierfür wurden AAF-dG und die Guanin-Addukte anderer aromatischer Amine (Anilin, 2-Naphthylamin, 2-Aminoanthracen und 1-Aminopyren) synthetisiert und in Phosphoramidite, entsprechend geschützt für die automatisierte DNA Festphasensynthese, umgewandelt (Abbildung 2).

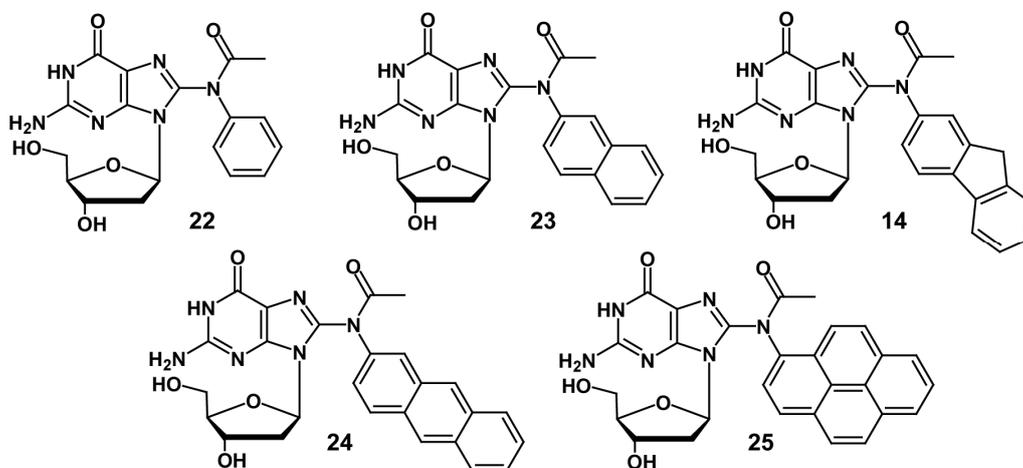


Abbildung 2. C8-dG Addukte: AAB-dG **22**, AAN-dG **23**, AAF-dG **14**, AAA-dG **24** und AAP-dG **25**. In dieser Arbeit wurden Derivate, entsprechend geschützt für die automatisierte DNA Festphasensynthese, hergestellt und in Oligonukleotide eingebaut.

Die synthetisierten dG Addukte wurden erfolgreich über automatisierte DNA Festphasensynthese in verschiedene Oligonukleotide eingebaut. Die Aufreinigung der Oligonukleotide erfolgte über HPL-Chromatographie. Anschließend wurden die Oligonukleotide mittels MALDI-Tof Massenspektrometrie und enzymatischem Verdau charakterisiert (Abbildung 3). Die synthetisierten, Adduktschäden enthaltenden Oligonukleotide wurden für die Kristallisation und für Primerverlängerungsstudien mit der TLS Polymerase η aus *S. cerevisiae* verwendet. Aus den Ergebnissen konnten Rückschlüsse auf den Mechanismus des Überlesens von Adduktschäden durch dieses Enzym gezogen werden.

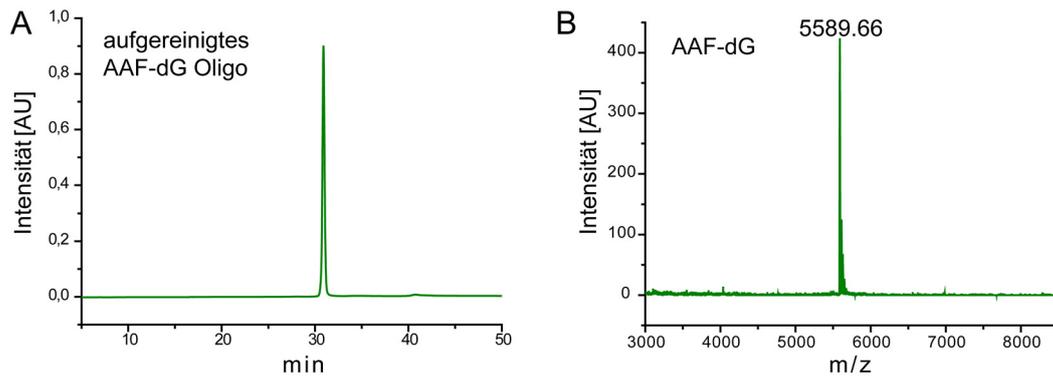


Abbildung 3. Beispiel der Aufreinigung und Charakterisierung eines Oligonukleotids mit Adduktschaden. A) HPL-Chromatogramm eines aufgereinigten AAF-dG Oligonukleotids. B) MALDI-ToF Massenspektrum eines aufgereinigten AAF-dG Oligonukleotids.

Die in dieser Arbeit enthaltenen Kristallstrukturen von Pol η in Komplex mit den Adduktschäden AAF-dG und 2-Aminoanthracen-dG (AAA-dG) lassen Rückschlüsse auf den Mechanismus der Schadensüberlesung zu. Pol η liest über den AAF-dG und den AAA-dG Schaden hinweg, indem die DNA um den Schaden rotiert wird. Der Schaden selbst verbleibt in *syn*-Konformation. Bei beiden Pol η Kristallstrukturen mit dem AAF-dG und dem AAA-dG Addukt werden zwei Pol η Moleküle pro asymmetrische Einheitszelle gefunden, welche die Enzym-DNA Komplexe in zwei verschiedenen Zuständen zeigen (A und B). In Komplex A gehen die aromatischen Reste π -Wechselwirkungen mit dem Watson-Crick Basenpaar 3' des Schadens ein und sind oberhalb des Basenpaars positioniert (Abbildung 4). Das aktive Zentrum ist somit für ein ankommendes dCTP vollständig blockiert.

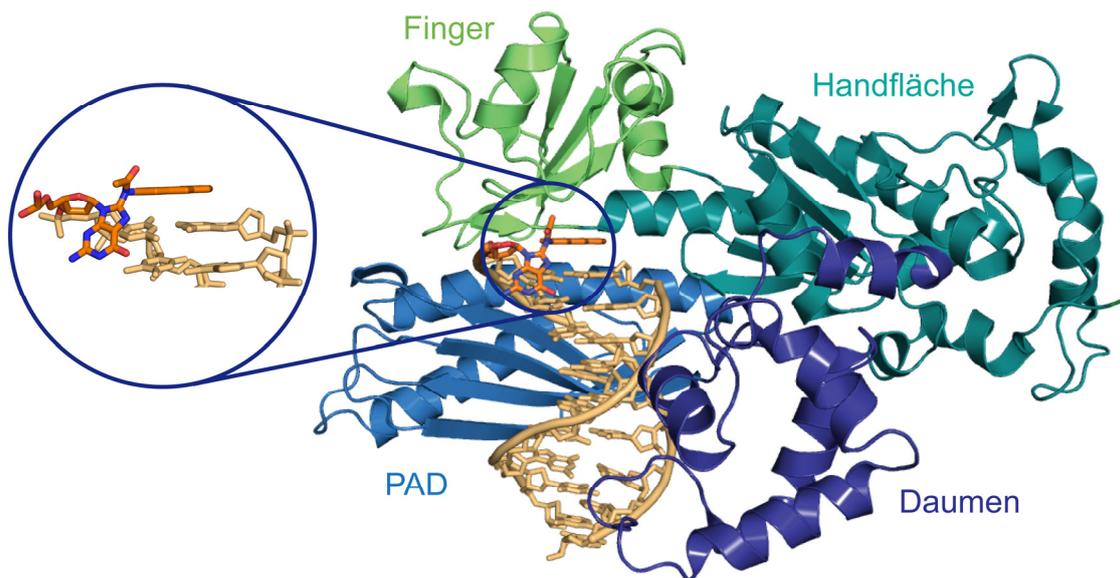


Abbildung 4. Co-Kristallstruktur von *S. cerevisiae* Pol η mit einem AAF-dG Oligonukleotid in Komplex A. Die DNA ist als goldenes Stab-Modell dargestellt, das AAF-dG ist orange eingefärbt.

In Komplex B ist die Situation deutlich unterschiedlich (Abbildung 5 B). Zum einen wird ein geringe Bewegung des DNA Templatstranges im Vergleich zu Komplex A beobachtet. Für den Primerstrang hingegen wird eine grössere Bewegung erfasst. Der Primerstrang rotiert relativ zum aromatischen Rest. Dies ergibt eine Verschiebung des Primer-Endes um 6 Å. In beiden Strukturen bringt diese Primerrotation den dG Teil des Schadens in das aktive Zentrum, obwohl sich dieser in *syn*-Konformation befindet. Der Schaden sitzt oberhalb des Watson-Crick Basenpaars, welches durch das Primerende und den Templatstrang gebildet wird. Die Bewegung des Primerstranges ergibt eine partielle Öffnung des aktiven Zentrums für ein dCTP. Aus diesen Strukturen lässt sich erkennen, dass die TLS ohne Rotation des Schadens in die *anti*-Konformation stattfinden kann. In dieser Situation kann der Einbau eines korrekten dCTPs gegenüber dem Schaden durch die Bildung einer einzelnen Wasserstoffbrückenbindung zwischen dem dCTP und dem Schaden erklärt werden (Abbildung 5 C).

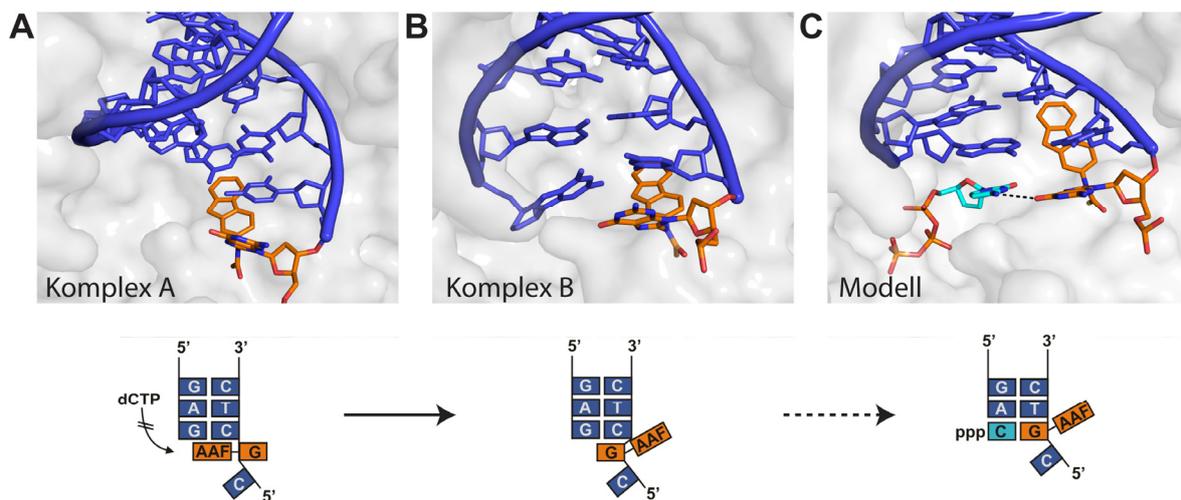


Abbildung 5. Modell für das Überlesen des AAF-dG Schadens durch Pol η . A) Im Komplex A ist der Fluorenrest oberhalb des letzten Basenpaares positioniert und geht mit diesem π -Wechselwirkungen ein. Das aktive Zentrum ist blockiert. B) Im Komplex B hingegen ist die DNA um den aromatischen Rest gedreht. Das Addukt-dG bewegt sich auf die Templatposition zu und das aktive Zentrum ist partiell geöffnet. C) Hypothetisches Modell der Replikation über den AAF-dG Schaden. Das dG : dC Basenpaar 3' des Schadens wurde durch das AAF-dG Addukt und das dCTP Triphosphat ersetzt. Das aktive Zentrum ist für das dCTP geöffnet, dieses kann eine kritische Wasserstoffbrückenbindung zur geschädigten Nukleobase in *syn*-Konformation ausbilden.

Das Modell wurde zusätzlich durch biochemische Studien mit synthetischem Zebularin-Triphosphat überprüft. Dieses kann die essentielle Wasserstoffbrücke nicht ausbilden und wird deshalb nicht gegenüber dem AAF-dG Schaden eingebaut. Zusätzlich wurde der Einfluss der Grösse des aromatischen Restes (Anilin, Naphthalin, Pyren) auf die TLS

untersucht. Es konnte gezeigt werden, dass die TLS Effizienz sehr stark von der Grösse des aromatischen Restes abhängt.

Ein weiterer Teil dieser Arbeit umfasst die Untersuchung von Leseraster-Mutationen. Sterisch anspruchsvolle Adduktschäden, wie der AAF-dG Schaden, induzieren Leseraster-Mutationen, wenn sie in speziellen Gensequenzen vorkommen.^[16, 53, 61-64] Der AAF-dG Schaden wurde in verschiedene repetitive Oligonukleotidsequenzen eingebaut. Mit diesen Sequenzen wurden Primerverlängerungsstudien durchgeführt. Die molekulare Basis der Entstehung von Leseraster-Mutationen konnte somit geklärt werden. Die Primerverlängerungsstudien wurden auf PAA-Gelen analysiert, die Zusammensetzung der verlängerten Primerstränge wurde über MALDI-Tof Massenspektrometrie untersucht. Es konnte gezeigt werden, dass alle Leserasterverschiebungen mit dem korrekten Einbau eines dCTPs gegenüber dem Schaden beginnen. Anschliessend rutscht der Primer am Templatstrang entlang, und ein oder mehrere Nukleotide des Templatstranges formen eine Blase. Verlängerung dieses falsch gepaarten Komplexes führt zur Manifestierung der Leseraster-Mutation. Die Mechanismen der Entstehung von -I, -II und -III Leserasterverschiebungen konnten aufgeklärt werden. Ein Beispiel wird in Abbildung 6 gezeigt.

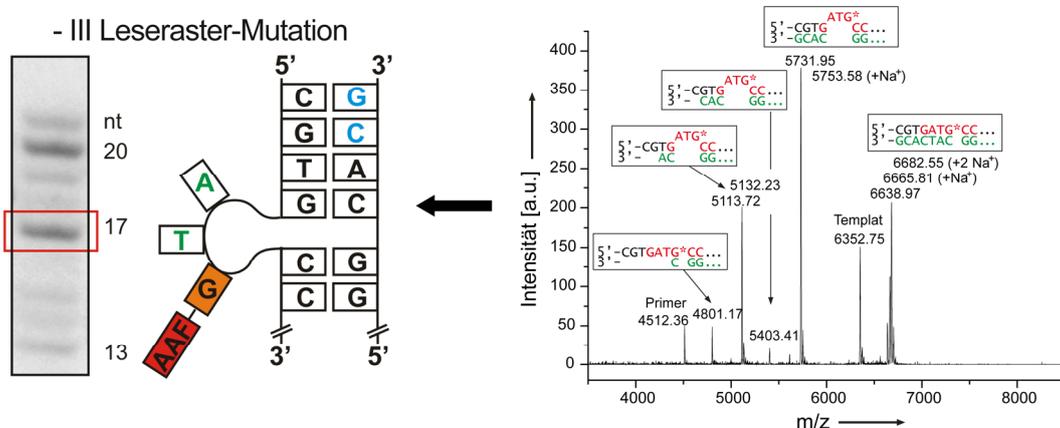


Abbildung 6. Schematische Darstellung der sequenzabhängigen Replikation des AAF-dG Schadens. Die gezeigte Sequenz erlaubt die Entstehung einer -III Leserasterverschiebung. Das MALDI-Tof Massenspektrum (rechts) bestätigt die Entstehung einer -III Leseraster-Mutation (links) während einer Primerverlängerungsstudie.

Ein weiterer Teil dieser Arbeit umfasst die Klonierung, Expression und Reinigung des katalytisch aktiven Teils der humanen *low-fidelity* Polymerase κ für Primerverlängerungsstudien. Es konnte gezeigt werden, dass Pol κ , ebenso wie Pol η , in der Lage ist, die synthetisierten C8 Adduktschäden zu überlesen. Die Replikation scheint jedoch im Falle von Pol κ fehlerhafter zu sein. Der Einbau von Nucleotiden gegenüber dem AAF-dG Schaden

ist nicht selektiv. Erste Kristallisationsversuche wurden mit Pol κ in Komplex mit einem AAF-dG Oligonukleotid durchgeführt, um den Mechanismus des Überlesen der Adduktschäden durch Pol κ genauer zu verstehen.

In dieser Arbeit wurde ausserdem das an der Nukleotidexzisionsreparatur beteiligte Hefe Protein Rad14 und sein humanes Homolog XPA zusammen mit AAF-dG Oligonukleotiden kristallisiert. Der Mechanismus der Schadenserkenkung in der Nukleotidexzisionsreparatur ist bisher nicht bekannt. Es wurde gezeigt, dass Rad14 und XPA spezifisch an DNA Stränge, die Adduktschäden enthalten, binden.^[139] In dieser Arbeit wurden die minimalen DNA Bindedomänen überexprimiert, gereinigt und im Komplex mit AAF-dG DNA kristallisiert. Es konnten Rad14 Kristalle erhalten werden, die Röntgenstrahlen bis 3 – 4.5 Å beugten. Da bisher keine Rad14 Struktur bekannt ist, mussten Informationen über die Phasen über Schweratomderivatisierung erhalten werden. Es wurde Selenomethionin-markiertes Rad14 hergestellt und Kristalle mit entweder Selenomethionin-markiertem Protein oder DNA, welche 5-Br-dU beinhaltete, erhalten. Einige Informationen über die Phasen konnten bei SAD Experimenten mit diesen Kristallen gesammelt werden. Durch anisotrope Streuung und Strahlenschäden war die Strukturlösung bisher nicht erfolgreich.

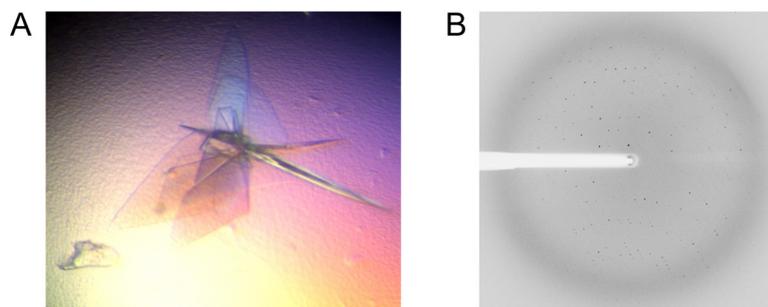


Abbildung 7. A) Rad14 Kristalle im Komplex mit einem AAF-dG DNA Duplex. B) Diffraktionsbild eines Rad14 Kristalls in Komplex mit AAF-dG DNA aufgenommen an der Mikrofocus Strahlungsquelle ID23-2 am ESRF (Frankreich).

Zusammenfassend ergibt die Kombination von organischer Synthese, biochemischen Studien und makromolekularer Kristallographie in dieser Arbeit mechanistische Einsichten in die fehlerfreie und fehlerhafte Replikation von sterisch anspruchsvollen DNA Addukten.

3 Introduction

3.1 Chemically induced DNA damage

DNA damages form by various exogenous and endogenous processes. They are a constant challenge for DNA polymerases that ensure accurate transmission of genomic information in all living organisms. Several ways lead to DNA lesions. Within the cell, spontaneous hydrolysis of the phosphodiester, *N*-glycosidic and C-NH₂ bonds within DNA can occur.^[1, 2] Moreover, DNA bases are constantly damaged by UV-light,^[3, 4] various oxygen species,^[5] free radicals and chemical pollutants.^[6] UV-light, for instance, produces crosslinks between adjacent pyrimidine bases to form cyclobutane thymine dimers (CPD).^[3, 4] In total, each cell accumulates about 10⁶ DNA damages per day.^[7] If these damages are not removed, they can lead to mutations or replication arrest and consequently to apoptosis or cancer. Dependent on the type of lesion, the cell uses multiple pathways to overcome these damages such as nucleotide excision repair (NER), base excision repair (BER), homologous recombination, non-homologous end joining or direct mutagenic or non-mutagenic bypass of the lesion during replication, entitled as translesion synthesis (TLS).

Mutagenic TLS can ultimately cause the development of cancer. A variety of chemical compounds in the environment are well known to induce cancer in humans. Already in 1775, a connection between certain chemicals and tumorigenesis has been formulated for the first time.^[6] *P. Pott* reported the increase of verrucas caused by carbon black and tar on the skin of chimney sweepers. Carcinogenic substances, which turned out to be aromatic amines, were first identified in 1895 when an increase of urinary bladder carcinomas of employees working in the production of aniline red dyes (fuchsine) was observed.^[6] The term “aniline-cancer” was used during this time. Later, several byproducts of the aniline dye production were identified as strong carcinogens such as 2-naphthylamine **2** or benzidine **3** (Figure 8).^[8, 9]

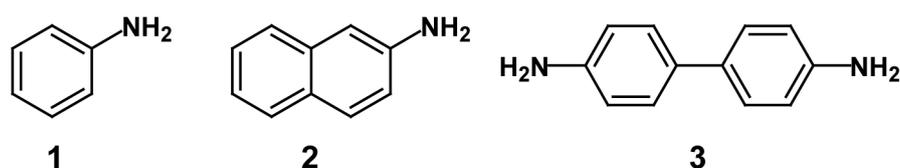


Figure 8. Prospective carcinogenic aromatic amines: **1** = aniline, **2** = 2-naphthylamine, **3** = benzidine

Since then, a wide range of chemical compounds are classified as carcinogens in addition to aromatic amines such as polycyclic aromatic hydrocarbons (PAH),^[10, 11] heterocyclic aromatic amines,^[12] *N*-nitroso compounds and other alkylation reagents.^[13] Nevertheless, these compounds play a major role in several industrial processes, for example in the production of dyes, synthetics, pharmaceuticals or explosives.^[14] They therefore represent ubiquitous environmental hazards. Nitropyrenes, for instance, can be found in urban air particles, coal fly ash and automobile exhaust.^[15] Moreover, heterocyclic aromatic amines have been detected in heat-cooked meat and fish, originated from pyrolysis of amino acids, sugars and fats (Figure 9).^[12]

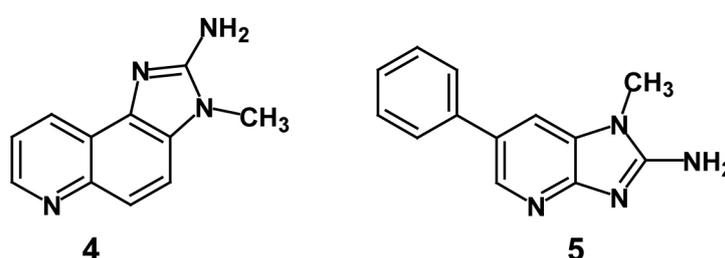


Figure 9. Food carcinogens: Heterocyclic aromatic amines found in cooked meat and fish. **4** = 2-Amino-3-methyl-imidazo[4,5-*f*]quinoline (IQ) **5** = 2-Amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine (PhIP)

Two well investigated aromatic amine compounds, serving as model substances for the investigation of carcinogenesis by aromatic amines, are 2-aminofluorene **6** (AF) and *N*-acetyl-2-aminofluorene **7** (AAF).^[16, 17]

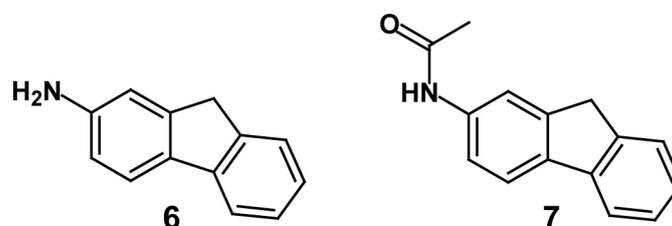


Figure 10. 2-aminofluorene (**6**, AF) and *N*-acetyl-2-aminofluorene (**7**, AAF)

These compounds were originally developed as insecticides, but they were never used for such purposes after early studies revealed the strong carcinogenic activity of AAF.^[17] Although these compounds are not environmental hazards per se, they have served as prototypes of DNA lesions formed by various aromatic amines, for example by food mutagens such as IQ **4** and PhIP **5** (Figure 9).

These bulky aromatic amine compounds do not react directly with cellular DNA, but require metabolic activation in the liver, leading to the formation of electrophilic nitrenium ions, which form covalent adducts with DNA bases.^[18, 19] If these adducts are not recognized and

repaired by cellular repair mechanisms, they induce mutations during replication and transcription which may cause the development of cancer.^[16, 20-22]

3.2 Formation of bulky DNA adducts

3.2.1 Formation of DNA adducts by aromatic amines

Procarcinogenic arylamines/amides, aminoazo dyes or heterocyclic amines form adducts with DNA bases especially at the C8-position of 2'-deoxyguanosine.^[18, 19, 23] A minor adduct formed is the C8-2'-deoxyadenosine lesion. Additionally, *N*²-2'-deoxyguanosine adducts or *N*⁶-2'-deoxyadenosine adducts can also be found.^[19, 23] The aromatic amines are metabolically converted to reactive hydroxylamine derivatives by cytochrom P450 monooxygenases (CYPs) in the liver. These species undergo further activation such as protonation, *O*-acetylation or *O*-sulfonation followed by cleavage of the created leaving group. The resulting nitrenium ion then reacts with nucleophilic functions of DNA bases forming covalent DNA adducts.^[19]

If aromatic amines are taken up in the human body, a complex detoxification mechanism ensures excretion of these substances by converting them into water-soluble derivatives. The first step involves an enzymatic oxidation at the aromatic ring to phenol derivatives (*C*-oxidation).^[24-26] The second step is mostly the esterification of the hydroxyl group to a water-soluble substance, which can be excreted via the urinary tract. It was found that the first oxidation step is also responsible for the genotoxicity of aromatic amines.^[27] Due to the unspecific nature of the CYP enzymes, oxidation can not only occur at the aromatic unit but also at the nitrogen (*N*-oxidation). At present, 57 genes encoding these enzymes have been identified in the human genome. CYP enzymes, which are most important for the activation of carcinogens, belong to the families 1-3 and have broad and overlapping substrate specificities.^[27-30] AAF **7** is converted by CYP1A2 into the proximate carcinogen *N*-hydroxy-AAF **8**. Enzymatic deacetylation of AAF and acetylation of AF can also occur at all stages of their metabolic activation resulting in acetylated and non-acetylated bulky DNA adduct lesions.

The AF and AAF derivatives **8** and **9** are capable of undergoing subsequent transferase-catalysed conjugation to polar molecules such as glutathione by glutathione *S*-transferases (GSTases), to glucuronic acid by glucuronosyltransferases,^[31] or to small residues such as sulphate by sulphotransferases (STases)^[32-34] and acetic acid by *O*-acetyltransferases

(*O*-ATases)^[35-40] (Figure 11). These reactions generate the ultimate genotoxic forms of the AF and AAF derivatives such as *N*-sulphoxy-AF/AAF or *N*-acetoxy-AF/AAF **11** and **10**.^[37, 41] These ultimate carcinogens can bind covalently to DNA and RNA via an arylnitrenium ion intermediate (**12**, **13**).

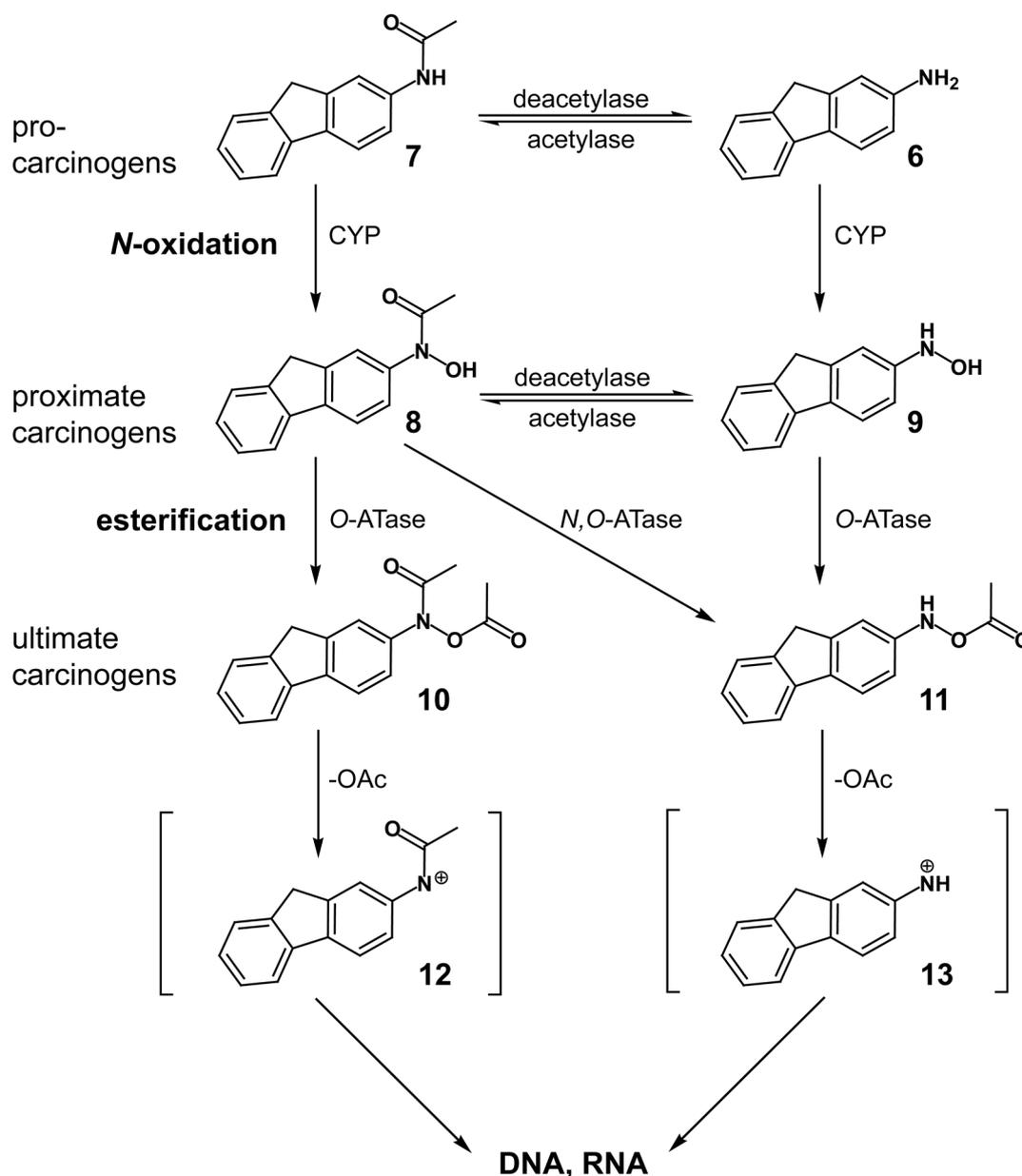


Figure 11. Metabolic activation of AF and AAF by cytochrom P450 monooxygenases (CYP) and *O*-acetyltransferases (*O*-ATases).

The major AAF adducts formed *in vivo* with DNA bases are 8-(*N*-acetyl-2-aminofluorene)-2'-deoxyguanosine (AAF-dG) **14** and the deacetylated derivative 8-(*N*-2-aminofluorene)-2'-deoxyguanosine (AF-dG) **15**, followed by the minor 3-(deoxyguanosin-*N*²-yl)-*N*-2-acetylaminofluorene (*N*²-AAF-dG) adduct **16** and the corresponding C8-dA adducts (Figure

12).^[23, 42] The ratio between the generated adducts depends strongly on the type of cells and organism.

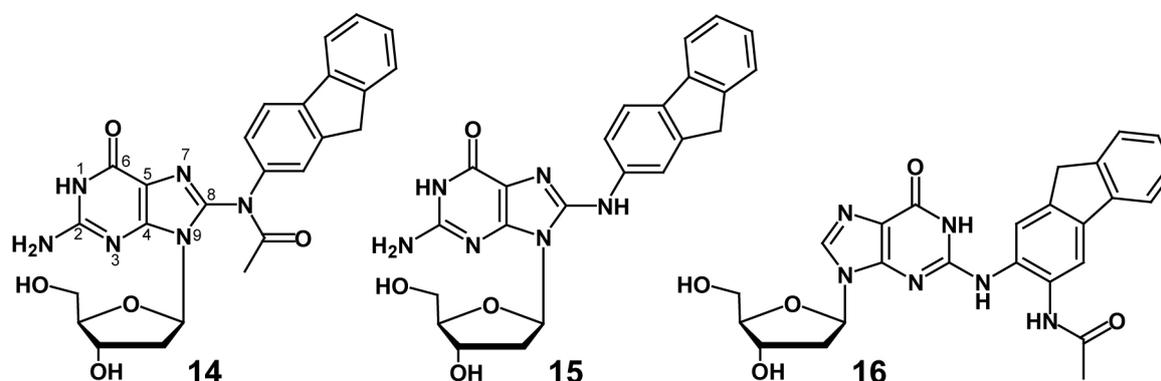


Figure 12. DNA adducts formed by AF and AAF. C8-AAF-dG adduct **14**, C8-AF-dG adduct **15** and N²-AAF-dG adduct **16**.

The mechanism of the C8-dG adduct formation is not completely clarified. The C8 position of deoxyguanosine is not the most nucleophilic function within the molecule. The most basic and nucleophilic position is the nitrogen at N7 position. Alkylation agents such as methyl iodide for instance, react with the N7 position in high yields.^[43] Thus, a mechanism was postulated which includes the preliminary attack of the electrophile at the N7 position (Figure 13).^[44, 45] The resulting positive charge is delocalized via the N7, C8 and N9 positions. Deprotonation of the C8 carbon gives an ylide allowing a Stevens rearrangement type of migration of the aryl fragment to the C8 position. Further rearomatisation results in the C8-adduct (Figure 13).

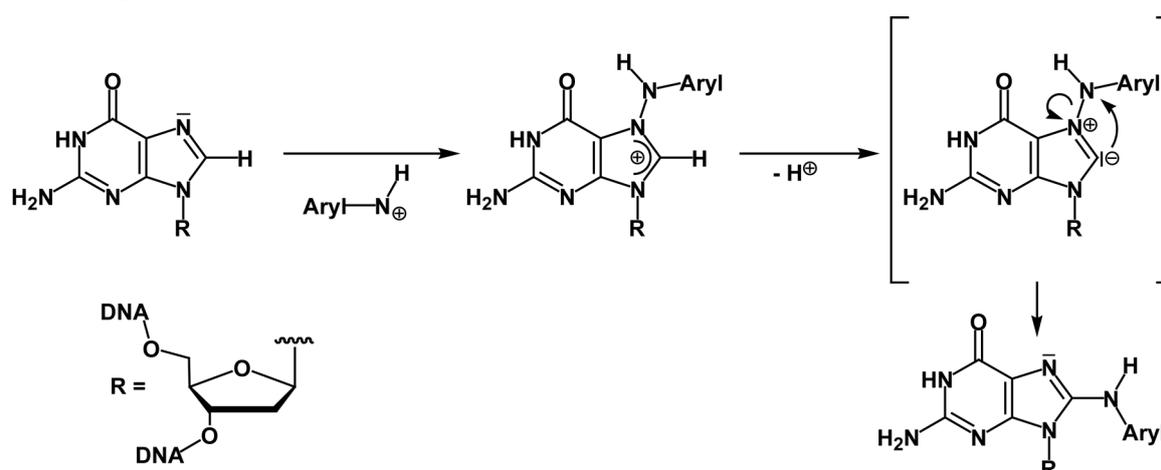


Figure 13. Postulated mechanism of the electrophilic arylamination at the C8-position of 2'-deoxyguanosine

3.2.2 Formation of other bulky adducts

Cytochrome-P450-dependent monooxygenases are also involved in the activation of other aromatic compounds such as polycyclic aromatic hydrocarbons (e.g. benzo[*a*]pyrene **19** (BP)) or the strong mycotoxin aflatoxin B₁ (AFB₁, **17**). CYP3A4 activates AFB₁ at its 8,9-bond, resulting in the AFB₁ *exo*-8,9-oxide **18** (Figure 14). The *endo*-diastereomer is not formed by CYP3A4, but might be formed in small amounts by CYP1A2. The primary DNA adduct of AFB₁ is 8,9-dihydro-8-(*N*7-guanosyl)-9-hydroxyaflatoxin B₁ (AFB₁-*N*7-dG), which is converted naturally to two secondary lesions, an apurinic site and an AFB₁-formamidopyrimidine (AFB₁-FAPY) adduct.^[46]

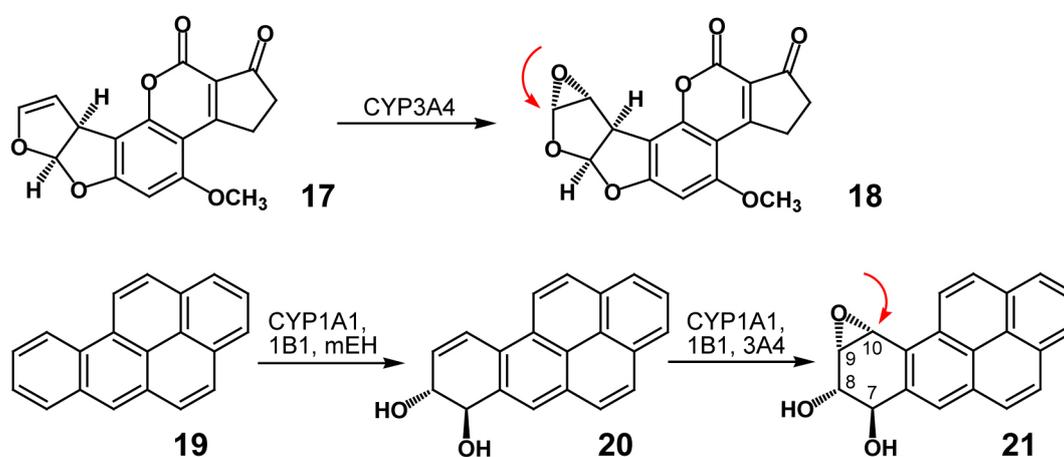


Figure 14. Metabolic activation of aflatoxin B₁ **17** and benzo[*a*]pyrene **19** by CYPs. The red arrows highlight the position of the nucleophile attack (DNA).

BP **19** is initially converted mainly by CYP1A1 or CYP1B1 into the 7,8-epoxide.^[11] This epoxide is a substrate of microsomal epoxide hydrolase (mEH), which produces the *R,R*-7,8-dihydrodiol **20** (Figure 14). Further epoxidation at the vicinal double bond catalysed by CYP1A1, CYP1B1 and CYP3A4 generates the ultimate genotoxic diol-epoxide of BP (BPDE) **21**. Of the four possible resulting diastereomers, the (+)-*anti*-BPDE **21** is formed with highest amounts. The *N*² or *N*⁶ amino groups of guanines or adenines are able to attack the epoxides to give the corresponding covalent adducts (major adducts: 10*S* (+)-*trans-anti*-BP-*N*²-dG, [BP]-dG; 10*R* (+)-*cis-anti*-BP-*N*⁶-dA, [BP]-dA).^[47]

3.3 The bulky C8-dG adducts are mutagenic

Bulky adducts are normally repaired by the cellular repair systems. Some aromatic amine DNA adducts are repaired faster than others.^[48] This can be traced back to the fact that the DNA adducts influence the DNA conformation in various ways. The aromatic system plays a major role, some aromatic units do not disturb the helical structure of the DNA, others promote conversion from B-DNA to Z-DNA or lead to denaturation of the DNA.^[48-50] A correlation between the conformational change and the persistence of the adducts *in vivo* can be found. Adducts with denaturing conformation are repaired faster than adducts which do not disturb the DNA helix. The mutagenicity of these bulky aromatic amine DNA lesions is established, when they are encountered by DNA polymerases. Due to constant damage formation there are always lesions present during replication of the genome in the S phase, which were not removed by the various repair mechanisms of the cell.

The mutagenic potential of the model compounds 8-(*N*-acetyl-2-aminofluorene)-2'-deoxyguanosine **14** (AAF-dG) and 8-(*N*-2-aminofluorene)-2'-deoxyguanosine **15** (AF-dG) are well established.^[19] Surprisingly they exhibit, despite their close chemical relationship, strongly different mutagenic characteristics. The non-acetylated AF-dG lesion is a lesion with low mutagenic potential. It is correctly replicated by high fidelity polymerases in most cases, which base pair AF-dG with dC.^[51] The basis for the low mutagenicity is probably the ability of the lesion to exist in *anti*-conformation, which allows formation of a correct Watson-Crick base pair (Figure 15).^[51] The acetylated bulky adducts in contrast, are not bypassed by high fidelity polymerases because these lesions are restricted to *syn*-conformation due to the steric hindrance caused by their acetyl group (Figure 15).^[52-54] This places the aromatic unit at C8 in the active site of the polymerase. This fact is independent of the size of the aromatic unit attached to the C8 portion of the dG base.^[55]

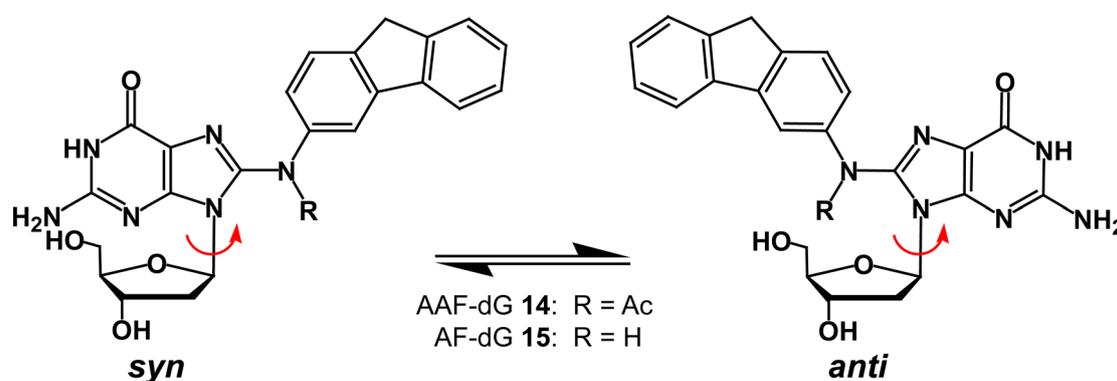


Figure 15. *Syn* and *anti* conformation of the C8-AF/AAF-dG adducts. The acetylated C8-adduct (R = Ac, AAF-dG) is restricted to *syn*-conformation.

3.4 Eukaryotic low fidelity DNA Polymerases

The bypass of bulky adducts requires specialized translesion synthesis (TLS) polymerases which enable replication through the damaged DNA structures. This mobilizes stalled replication forks, thus inhibiting the activation of apoptotic pathways.^[71] In eukaryotes, DNA polymerases belonging to the Y-family (Pol η , Pol κ , Pol ι , Rev1) and the B-family DNA polymerase ζ promote replication through DNA lesions.^[72, 73] The Y-family polymerases are divided into five subfamilies.^[74] The three polymerases Pol η , Pol ι and Rev1 are found exclusively in eukaryotes, Pol κ belongs to the DinB branch with the prokaryote homologue *E. coli* Pol IV and the archaea homologue Dpo4. The fifth branch, the UmuC branch (*E. coli* Pol V) is only found in prokaryotes. Furthermore, Pol ζ is exclusively present in eukaryotes.^[74]

In contrast to high fidelity polymerases, the TLS polymerases have much higher error rates if they operate on undamaged template DNA.^[75] Their error rate during normal DNA synthesis is 10^{-2} to 10^{-4} , which is 1 to 2 orders of magnitude higher than those of replicative polymerases (A- and B-family) even after removal of their $3' \rightarrow 5'$ exonuclease activity.^[75] But TLS polymerases are able to synthesize DNA through lesions which block high fidelity polymerases.^[76, 77] This fact can be assigned to their flexible structure and more open active site allowing mostly faithful bypass of distorted DNA structures. High fidelity polymerases use an induced fit conformational change as active mechanism to discriminate between correct and incorrect base pairs. For low fidelity polymerases, hardly any conformational changes are observed in their active sites. The active site is located at the *N*-terminal part of the Y-family polymerases, the *C*-terminus is variable and involved in localization, recruitment and protein-protein interactions. All members of the Y-family exhibit the conserved structural motif of the right-hand structure of DNA polymerases consisting of palm, thumb and finger domain. However, finger and thumb domain are smaller and stubby compared to high-fidelity polymerases. In all polymerases, the palm domain harbours the three conserved acidic residues that coordinate the two divalent metal ions. This domain also contributes to the binding of the incoming dNTP. Additionally, the Y-family polymerases own a specialized PAD (polymerase associated domain or “little finger domain”). This domain increases the DNA binding surface area of the polymerase. Nevertheless, DNA interaction is still weak in comparison to high-fidelity polymerases. In addition, Y-family polymerases lack a $3' \rightarrow 5'$ exonuclease activity, which is an integral part of all replicative polymerases.^[76, 77]

3.4.1 Polymerase switch

When replicative polymerases present in the replication fork are stalled at a lesion such as AAF-dG, TLS polymerases have to be recruited and directed to the lesion site. This process has to be tightly regulated in order to prevent accumulation of mutations caused by unrestricted activity of the error-prone low fidelity polymerases. The key regulator of this process is the replication sliding clamp proliferative cell nuclear antigen (PCNA). The trimeric PCNA binds multiple proteins involved in both normal replication and TLS.^[78] This is largely regulated by ubiquitylation and sumoylation of PCNA (Figure 17). Stalling of the replication fork by bulky DNA adducts for instance, results in monoubiquitylation of PCNA by enzymes of the Rad6/Rad18 group, which increases the affinity of TLS DNA polymerases such as Pol η , Pol κ or Pol ι to the replication clamp (Figure 8).^[79-81] These polymerases possess ubiquitin-binding domains which interact with monoubiquitin on PCNA. The small ubiquitin-like modifier (SUMO) can also target PCNA and bind to the same site as ubiquitin.^[82] The role of sumoylation in eukaryotes is still unclear and sumoylation in humans has not been observed yet. It is suggested that PCNA sumoylation favors the repair of fork blocking lesions by a mutagenic TLS pathway over an error-free recombination based pathway.^[83]

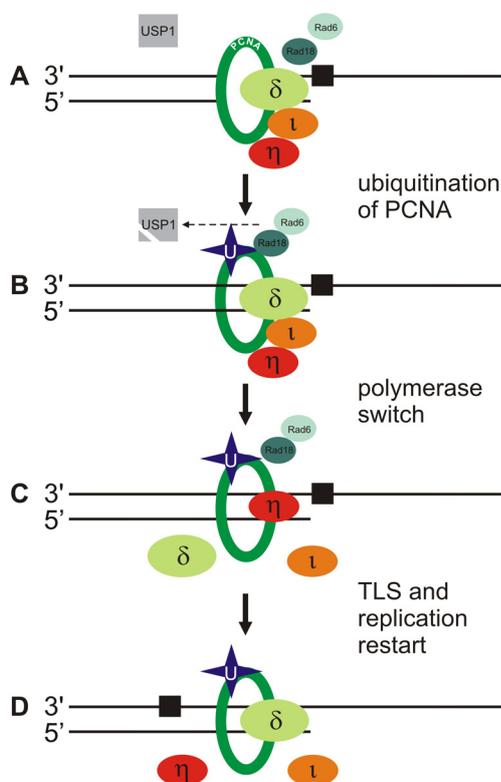


Figure 17. Simplified model for translesion DNA synthesis.^[84]

A) Replication is stalled at a lesion site (black square) and the high fidelity polymerase δ is blocked by the lesion. Rad18 is activated by single-stranded DNA regions and the de-ubiquitinating isopeptidase USP1 undergoes autocleavage. B) Rad18 binds to Rad6 and the complex ubiquitinates PCNA (only one PCNA monomer of the trimeric PCNA complex is shown ubiquitinated in this scheme, although all three monomers of the trimeric molecule are most likely ubiquitinated). C) The affinity for TLS polymerases to bind to PCNA is increased and one of them, in this case Pol η ^[79], bypasses the lesion. D) Pol η dissociates and replication restarts.

In summary, polymerase switch at sites of stalled replications forks enable bypass of replication blocking lesions. Some lesions can be replicated by a single TLS polymerase. After the lesion is bypassed, replication by Pol δ restarts. Nevertheless, depending on the lesion, more than one TLS polymerase may be needed for the bypass process.^[84] It is still debatable, how certain TLS polymerases, assigned to a specific DNA lesion, are recruited to the lesion site and if TLS polymerases compete for the bypass of other lesions. Possibly, different polymerases bind briefly to PCNA and more tightly if the clamp is ubiquitinated (Figure 17). If the polymerase is able to accommodate the lesion in its active site, it will carry out TLS synthesis. Alternatively, the polymerase might only be able to insert a nucleotide opposite the lesion but not to extend from the lesion. Now, a second switch occurs and another polymerase catalyzes the extension from the lesion.^[85] A general model suggests, that Pol η , Pol κ , Pol ι and Rev 1 insert nucleotides opposite a lesion and Pol ζ is responsible for the extension from the lesion.^[86, 87] Concluding, error-free and error-prone bypass of DNA lesions depends on the interplay of several TLS polymerases and is not entirely understood so far.

In the following section, recent literature covering the bypass of bulky DNA adducts by TLS polymerases will be discussed in detail.

3.4.2 Pol η

The Y-family DNA polymerase η is a key player in the replicative bypass of UV lesions such as CPD lesions in DNA. Inactivation of Pol η in humans results in the variant form of Xeroderma Pigmentosum (XPV) which is associated with a high risk of developing skin cancer. Lately, this bypass process was investigated on the molecular level explaining the efficient CPD lesion bypass by both human and yeast Pol η .^[88, 89] Moreover, cisplatin induced 1,2-d(GpG) adducts (Pt-GGs), which are formed in a typical cancer therapy with cisplatin, can be processed by Pol η (Figure 18).^[90] The bypass of such intrastrand crosslinks by high fidelity DNA polymerases is particularly difficult because the linked bases are strongly helix deforming. Pol η possesses an open and flexible active site. It is able to accommodate the full Pt-GG adduct in the active site where it correctly base pairs the 3' G of the Pt-GG lesion with a dCTP (Figure 18).

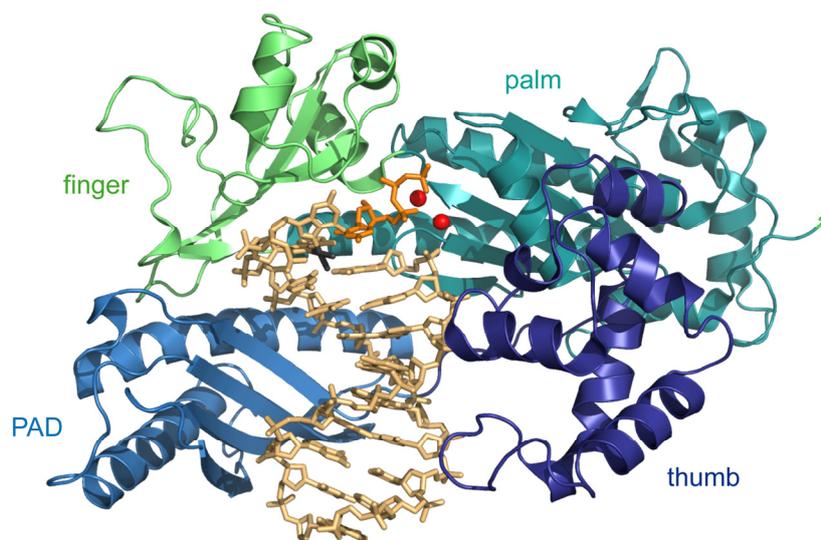


Figure 18. Crystal structure of DNA Polymerase η from *Saccharomyces cerevisiae* in complex with an intrastrand cisplatinum GG adduct.^[90] Pol η possesses the typical right-hand polymerase structure with palm, thumb and finger domain and the additional PAD domain which is unique for Y-family polymerases. The DNA helix is colored in gold, the cisplatinum adduct in grey and the incoming dCTP in orange. The catalytically active Ca^{2+} ions are depicted as red spheres.

Moreover, recent *in vivo* experiments revealed that Pol η also plays a major role in the replication through acetylated aromatic amine lesions such as AAF-dG.^[91] Data from *Saccharomyces cerevisiae* strains with a Pol η deficiency show that the enzyme participates in TLS of the AAF-dG adduct *in vivo*. In the absence of Pol η , the sequence dependent accurate and inaccurate bypass of AAF lesions is reduced up to 8-fold.^[91] Further *in vitro* studies have shown that both human and yeast Pol η base pair the AAF-dG lesion correctly with dC^[56-58, 91] and are able to catalyze the extension from the lesion but with reduced efficiency.^[57] In addition, Pol η also incorporates a correct dCTP opposite the N^2 -AAF-dG adduct.^[20]

It was also shown that Pol η is responsible for the occurrence of frameshift mutations in response to the presence of aromatic amine lesions during replication.^[91, 92] Its flexible and open active site is able to accommodate bulge structures formed by slipped mispairing of primer-template constructs. Nevertheless, the occurrence of frameshift mutations seems to be strongly sequence dependent and is not systematically investigated and understood so far.

3.4.3 Pol κ

As mentioned before, Pol κ belongs to the DinB subfamily of Y-family polymerases, which includes *E. coli* Pol IV and the archeal polymerase Dpo4. In contrast to the prokaryotic and archeal DinB proteins, human Pol κ and other eukaryotic homologs possess unique

N-terminal and *C*-terminal extension.^[74, 93] The *C*-terminus, involved in binding to other proteins, is not required for polymerase activity. The *N*-terminal region, however, is obligatory for the activity of Pol κ and removing of this extension gives a completely inactive polymerase.^[93]

Pol κ is the most faithful of all Y-family DNA polymerases with an error rate of about 10^{-3} to 10^{-4} on undamaged template DNA.^[74] However, Pol κ is, in addition to Pol η , involved in the generation of frameshift errors, particularly -I frameshift mutations.^[94] Insertion of nucleotides opposite some DNA lesions can be difficult for Pol κ , but the polymerase is also able to extend from nucleotides inserted by another DNA polymerase. For instance, Pol κ is unable to replicate through a CPD lesion. But it is capable to extend a primer after a dG was inserted opposite the 3'dT of the lesion.

Pol κ can bypass the AAF-dG lesion *in vitro* in an error-free as well as in an error-prone manner incorporating dTTP with the same ratio as the correct dCTP opposite the lesion.^[20, 95, 96] Extension from the lesion occurs with reduced efficiency compared to undamaged template DNA, nevertheless, the contribution of Pol κ to the generation of frameshift mutations during bypass of the AAF-dG adduct has not been entirely clarified yet.^[96] *In vivo* studies revealed, that Pol κ is involved in the error-prone bypass of the AAF-dG lesion. In simian kidney cells, A, G and T are misincorporated opposite the AAF-dG lesion and single two-base deletion have been observed.^[97] On the contrary, Pol κ correctly bypasses the N^2 -[BP]-dG adduct, incorporating a dCTP opposite the lesion with high selectivity.^[96]

3.4.4 Pol ι

Pol ι replicates undamaged DNA with different velocity and accuracy. Opposite purine template bases, incorporation of nucleotides occurs with much higher efficiency and fidelity than opposite pyrimidine bases.^[86, 98-100] Bypass of dC and dT is highly mutagenic and inefficient. Moreover, Pol ι is unable to replicate through a CPD lesion but is proficient in bypassing N^2 -adducted guanines. Incorporation of a correct nucleotide opposite lesions is presumably achieved via formation of a protonated Hoogsteen pair with the templating base in *syn* and the incoming nucleotide in *anti* conformation.^[101, 102] Opposite an AAF-dG, human Pol ι incorporates predominantly a correct dCTP, although it is unable to catalyze the extension from the lesion.^[99]

3.4.5 Pol ζ

The heterodimeric DNA Pol ζ consists of the Rev3 catalytic subunit and the Rev7 accessory subunit ^[103] and belongs to the B-family of polymerases, which includes the high-fidelity replicative polymerases δ , ϵ and α . Nonetheless, Pol ζ lacks, such as all TLS polymerases, the 3' \rightarrow 5' exonuclease activity. Pol ζ is in general highly inefficient in replicating through DNA lesions. It is mostly unable to incorporate nucleotides opposite the lesions. However, Pol ζ is capable of bypassing some DNA damages such as CPD lesions. *In vitro*, yeast Pol ζ is highly inefficient in incorporating nucleotides opposite an AAF-dG lesion and can only incorporate dGTP opposite the lesion to some extent. However, Pol ζ is capable of catalyzing the extension from the AAF-dG lesion, most efficiently if a dA or dC is opposite the AAF-dG lesion.^[104] It was observed, that translesion synthesis of AAF-dG adducts by Pol ζ is further stimulated by Rev1 in yeast.^[105] In yeast cells, Pol ζ is evidently required for error-prone translesion synthesis across from AAF-dG DNA adducts.^[106, 107] Thus, the enzyme appears to be specialized in extending correctly paired and mismatched template-primer constructs resulting from correct or inaccurate base insertion by another TLS polymerase.^[74, 108, 109] Therefore, Pol ζ is often entitled as the *general extender* taking over extension from a lesion which is (correctly) bypassed by another TLS polymerase.

3.4.6 Rev1

Rev1 acts together with Pol ζ in translesion synthesis. By itself, Rev1 is specific for dG templates and is restricted primarily to inserting dCTP opposite template dGs and certain DNA lesions such as abasic sites and adducted dG residues.^[108, 110, 111] This specialized dCMP transferase activity is conserved from yeast to humans^[112], but the catalytic activity of Rev1 seems not to be required for the bypass of several lesions for which Rev1 is essential *in vivo*. Rev1 possesses a variety of protein-protein interaction modules and is able to interact with many other proteins, particularly other TLS polymerases.^[109]

In vitro, AAF-dG is a strong block for nucleotide incorporation by Rev1. Nevertheless, Rev1 may contribute to the *in vivo* bypass of AAF-dG lesions together with Pol ζ .^[104, 107]

3.5 Crystallographic snapshots of bulky DNA adduct blockage and bypass

Although the biochemistry of lesion tolerance by TLS polymerases is well investigated, only limited structural data of TLS polymerases in complex with bulky adduct containing DNA are available to date. The following section provides an overview about the known crystal structures of high-fidelity and TLS polymerases in complex with bulky lesions containing DNA.

Besides benzo[a]pyrene which generates adducts with dG and dA (section 3.2.2), benzo[c]phenanthrene (B[c]Phe) is another well studied polycyclic aromatic hydrocarbon, which forms covalent adducts with DNA after metabolic activation. The most frequently formed bulky adduct is the N^2 -guanine adduct [BPhe]-dG formed from (-)-1R,2S-epoxy-3S,4R-dihydroxy-1,2,3,4-tetrahydrobenzo[c]-phenanthrene.^[113] Crystal structures of [BP]-dG, [BP]-dA and [BPhe]-dG in complex with high-fidelity and TLS polymerases are available.^[114-117] The crystal structure of the 10*S* (+)-*trans-anti*-BP- N^2 -dG lesion ([BP]-dG) in complex with the high-fidelity polymerase Bst Pol I shows that the [BP]-dG still has some base pairing capabilities with dC. Nevertheless, the adduct protrudes into the minor groove, which leads to severe distortion of the DNA duplex polymerase complex and blocks further replication.^[118] The archaeal Y-family polymerase Dpo4 in contrast is capable of bypassing both the [BP]-dG and the 10*R* (+)-*cis-anti*-BP- N^6 -dA ([BP]-dA) lesion by flipping the lesion out into the minor and major groove, respectively (Figure 19, A).^[114, 115]

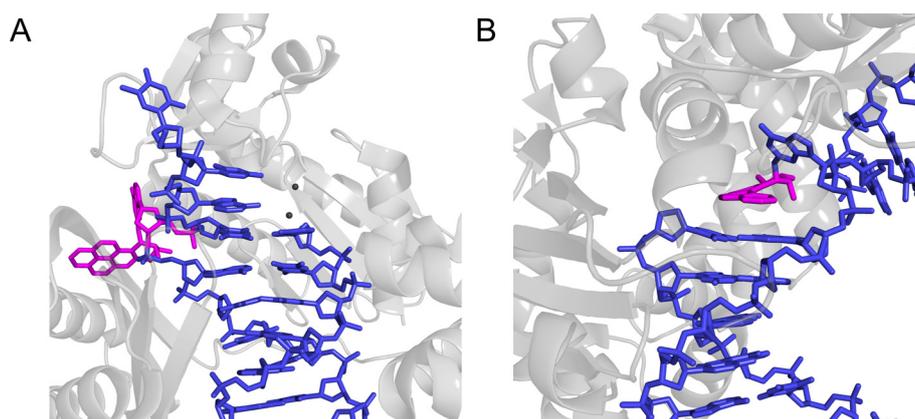


Figure 19. A) Crystal structure of the Y-family polymerase Dpo4 in complex with [BP]-dG lesion containing DNA. A –I frameshift mutation is observed. B) Crystal structure of Pol β in complex with a DNA duplex containing the [BPhe]-dG lesion. The phenanthrene ring stacks on top of the previously formed base pair occupying the binding pocket for the incoming nucleotide. The protein is shown as cartoon (grey) and the DNA as stick model (blue) with the lesions highlighted in pink.

The polymerase accommodates the huge template distortion using a sequence dependent 5' template slippage/misalignment mechanism. This results either in a base substitution mutation or, in case of template realignment, generates a -I frameshift mutation (Figure 19, A).^[114, 115] A similar template misalignment/slippage mechanism depending on the sequence context seems to allow Dpo4 to bypass also the 1,*N*²-etheno(ϵ)guanine adduct.^[119]

Furthermore, the human DNA polymerase β involved in gap filling during BER, was found to bypass the [BPhe]-dG lesion.^[116] However, this proceeds inefficiently and error-prone. The structure of Pol β in complex with [BPhe]-dG containing template revealed the mechanism of this erroneous bypass: The stacking of the phenanthrene ring on the previous correctly formed base pair occupies the binding pocket for the incoming nucleotide holding the polymerase in an open conformation (Figure 19, B). Thus, correct nucleotide insertion is prevented. Due to the established stacking interactions with the phenanthrene ring, purine (dA) incorporation is preferred opposite the lesion.^[116]

As discussed above, AF-dG is replicated by high-fidelity polymerases while AAF-dG is locked in *syn*-conformation due to the sterical restriction by its *N*⁸-acetyl group and therefore a strong block for all replicative polymerases. Crystal structures of the AAF-dG adduct in DNA in complex with the high fidelity polymerases T7 and Bst Pol I reveal that the adduct either prevents dNTP entry (T7, Figure 20, A)^[120] or prevents the modified template from entering the preinsertion site (Bst Pol I)^[121]. The crystal structure with T7 DNA polymerase shows that the AAF-dG template base is in *syn* conformation and the adduct inserts in a hydrophobic pocket of the fingers' subdomain locking the fingers in an open conformation.^[120] In Bst Pol I polymerase, the AAF-dG does not enter the active site and is therefore partially disordered and not visible in the electron density. The preinsertion site is empty and no incorporation of an incoming dNTP can be observed.^[121] Thus, replication past AAF is blocked in both cases.

In contrast, AF-dG adopts *anti*-conformation with the fluorene residue projecting into the major groove in the active site of Bst Pol I. Thus correct dCTP incorporation can take place (Figure 20, B).^[121] For T7 polymerase, electron density for the AF-dG is absent. However, the active site is empty and AF-dG does not bind in the active site.^[120] Moreover, there is no evidence that the AF-dG is bound in a hydrophobic pocket similar to AAF-dG and no bound nucleotide can be observed. Nevertheless, bypass of the AF-dG is most likely achieved by the polymerase. A recent crystal structure of archaeal Y-family polymerase Dpo4 in complex with AF-dG containing DNA shows the AF-dG in the process of being bypassed

(Figure 20, C).^[122] AF-dG adopts *anti* conformation which allows correct Watson-Crick base pairing with the 3'-terminal primer dC and positions the AF moiety in the major groove (Figure 20, C). Further extension from this complex is observed.

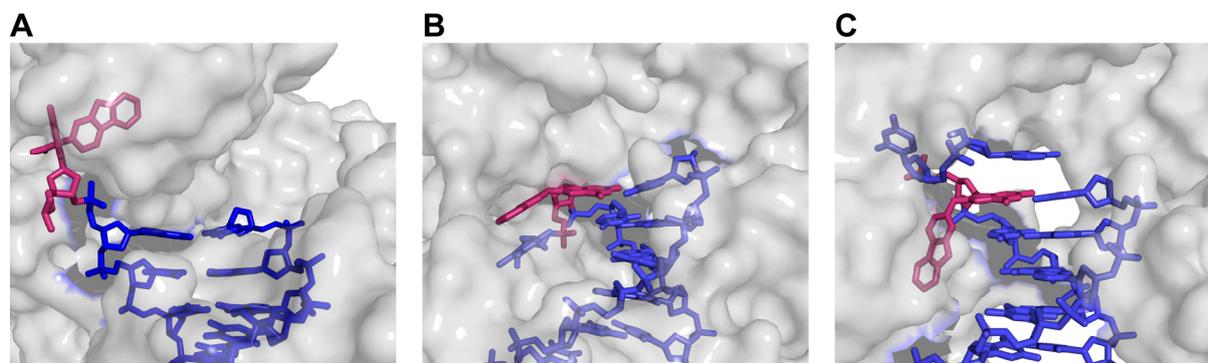


Figure 20. Crystal structures of AF-dG and AAF-dG lesions inside different polymerases. A) The high-fidelity bacteriophage T7 polymerase in complex with AAF-dG containing DNA; The AAF-dG adopts *syn*-conformation, preventing dNTP entry and holding the polymerase open. B) The high-fidelity polymerase Bst Pol I in complex with AF-dG containing DNA; The AF-dG base pairs with dC. C) The archaeal low-fidelity polymerase Dpo4 accommodating an AF-dG lesion within its active site, the AF-dG base pairs with dC. The protein is shown as semi-transparent surface representation (grey), with the DNA as stick model (blue) and the lesions highlighted in pink.

In summary, the crystal structures of *Ellenberger*, *Beese* and *Patel* displaying the non-acetylated AF-dG lesion inside different polymerases proof that the lesion is indeed bound in *anti* conformation allowing Watson-Crick base pairing with an incoming dCTP.^[51, 120, 122] However, the mechanism that allows low-fidelity polymerases such as Pol η to replicate past acetylated AAF-dG lesions is still unknown. Biochemical studies demonstrate that Pol η bypasses the lesion correctly.^[56-58, 91] Nonetheless, there is no crystal structure of a Y-family polymerase in complex with the AAF-dG lesion available so far. Computational modeling studies of the Y-family polymerase Dpo4 suggest that rotation of the lesion into the *anti* conformation is required for the bypass to allow base pairing.^[123] Due to the strong *syn* preference of the adducted base, this rotation is difficult to accomplish, which may force Pol η or other TLS polymerases to find a different mechanism for accurate lesion bypass. The bypass of acetylated bulky DNA lesions by Pol η is subject of this thesis work.

3.6 Repair of bulky DNA adducts

Bulky DNA adduct lesions are mostly removed by the nucleotide excision repair pathway (NER), which is one of the major DNA repair pathways in eukaryotic cells, able to remove a variety of DNA damages.^[124-126] To date, only limited structural insight into lesion recognition and removal by the NER machinery is available. Defects in the NER system lead to severe diseases such as xeroderma pigmentosum (XP), a hereditary disease characterized by extreme photosensitivity and a 2000-fold increased incidence of sunlight-induced skin cancer, or trichothiodystrophy (TTD).^[127]

Two ways of NER can be distinguished: Repair of lesions within the entire genome, referred to as global genome NER, and repair of transcription-blocking lesions during transcription, thus called transcription-coupled NER. Multiple proteins are involved in the NER pathway. First, the lesion is recognized, then DNA incision occurs on the 3' and 5' sides of the lesion and a short oligonucleotide segment containing the lesion is removed. Thirdly, the gap is filled by DNA synthesis and ligation occurs using the undamaged strand as template.^[128] Whereas the actual repair mechanism in NER is rather well investigated, damage recognition is still not understood. It is remarkable how a certain pool of proteins is able to recognize various DNA lesions.^[129, 130] The detection of lesions leading to the assembly of the NER system is controversially discussed. The two essential human NER proteins XPA (xeroderma pigmentosum group A) and XPC (xeroderma pigmentosum group C) are believed to be primarily involved in damage recognition.^[128] To date, there exist three models for the critical initial damage recognition step in NER with either the XPA, XPC, or both proteins acting randomly as the primary lesion sensor. The first model suggests, that XPC is responsible for the initial damage recognition step recognizing distorted DNA helix structures.^[131] A crystal structure of a truncated version of the XPC yeast homologue Rad4 in complex with a CPD lesion shows how the protein binds to the DNA, surprisingly without contacting the lesion itself.^[132] The second model proposes that XPA is the primary lesion sensor. This idea is supported by studies in which DNA repair is stimulated by pre-incubation of lesion containing DNA with a complex of XPA and the replication protein A (RPA).^[133] The third model suggests a cooperative damage recognition by RPA, XPA, and XPC and final complex formation of the four proteins XPA, RPA, XPC and the xeroderma pigmentosum group C protein complex XPC-HR23B at the lesioned site.^[134]

XPA is suggested to be involved in recognition and removal of a large variety of structurally diverse bulky DNA lesions.^[135-137] The protein is required for both global genome repair and

transcription-coupled repair, in the absence of XPA, no incision complex can form and no excision of damaged DNA damage occurs.^[127, 138] The protein, which contains a zinc finger motif, is capable of binding specifically to damaged DNA *in vitro* recognizing aromatic units attached to DNA, which do not distort the helical structure of the DNA.^[139, 140] Moreover, it interacts with RPA. The nucleic acid interaction domain of XPA has been identified and its structure was determined by NMR spectroscopy (Figure 21).^[141, 142]

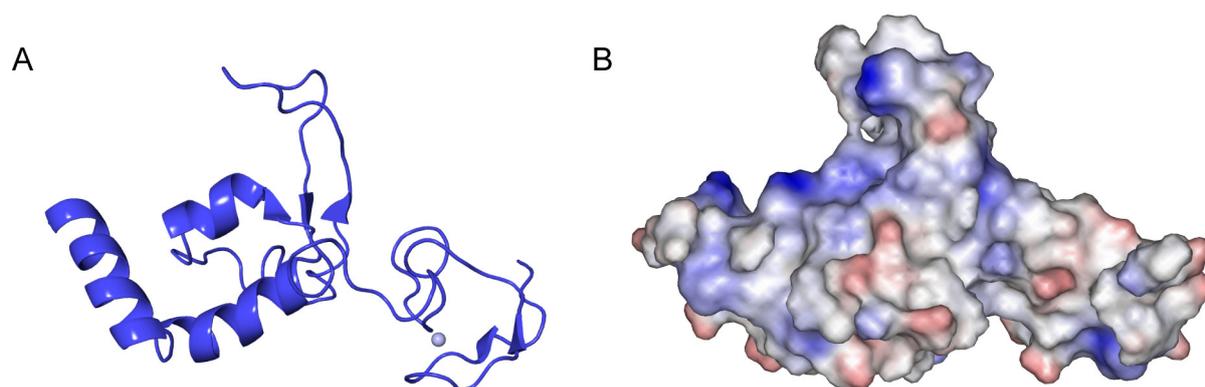


Figure 21. NMR solution structure of the minimal DNA binding domain of XPA (residues 98–210).^[141] A) Cartoon representation of the structural motifs of XPA, the Zn^{2+} is depicted as light blue sphere, B) Electrostatic potential on the solvent-accessible surface (displayed with *WebLab Viewer*). Blue corresponds to positive potential and red to negative potential.

The minimal DNA binding domain (122 amino acids, residues 98-219) consists of an acidic subdomain containing the zinc finger motif and a C-terminal subdomain (residues 138-209) which forms a positively charged cleft on the protein surface which is presumably involved in DNA binding (Figure 21, B).^[141, 143] Additionally, chemical shift perturbation experiments carried out in the presence of DNA or a short RPA peptide sequence revealed that the zinc finger motif is required for the interaction with RPA but not involved in DNA binding.^[141]

However, structural data of XPA-DNA complexes, or its yeast homologue Rad14, are still not available. Thus, the recognition of bulky DNA lesions by XPA proteins needs to be elucidated.

4 Aims of the project

Heterocyclic aromatic amines produce bulky C8 guanine lesions *in vivo*, which interfere and disrupt DNA and RNA synthesis. These lesions are consequently strong replication blocks. In addition they give rise to point and frameshift mutations. The translesion synthesis DNA polymerase η is able to slowly bypass C8 bulky adduct lesions such as the widely studied 2-aminofluorene-dG (AF-dG) and its acetylated analogue (AAF-dG) mainly in an error-free manner.^[56-58, 91]

Aim of this thesis work was to decipher the bypass mechanism of the AAF-dG lesion by translesion synthesis polymerases by crystallizing AAF-dG containing DNA in complex with the TLS polymerase η . For crystallographic and biochemical studies, bulky lesion containing DNA should be used. This requires the synthesis of the lesions, protection and activation as phosphoramidites and site-specifically insertion into oligonucleotides. AAF-dG could be obtained via a published procedure, derivatives with other aromatic units may possibly be synthesized in the same way. Figure 22 gives the structure of the lesions to be synthesized.

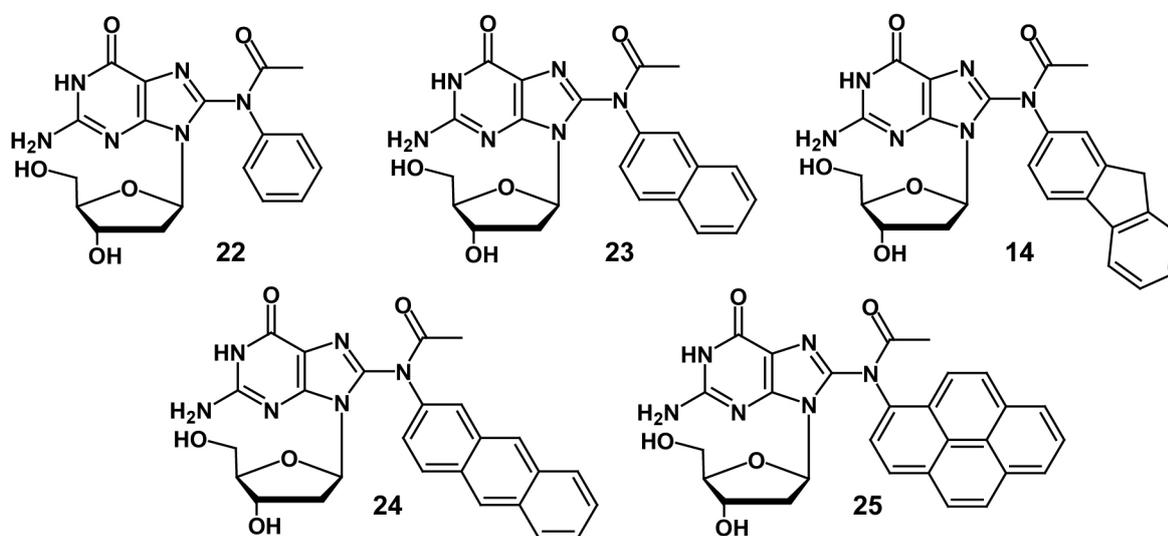


Figure 22. Bulky C8-adducts which are planned to be synthesized and used for crystallization as well as biochemical studies in this thesis work.

The mechanism of the AAF-dG bypass by Pol η could be investigated using biochemical methods such as primer extension studies to determine the base pairing partner of the AAF-dG and to examine the sequence dependency of the bypass process by Pol η . Moreover, by using different bulky dG adducts, the influence of the aromatic unit on the replication ability of Pol η across these lesions should be investigated.

A second part of this thesis work includes the investigation of frameshift mutations by AAF-dG adducts. Frameshift mutations induce a shift of the genetic reading frame. If the shift involves one or two bases, the corresponding frameshift mutations cause the loss of the information content of the whole gene. The mechanisms that lead to frameshift mutations are not well understood but it is known that frameshifts are produced in consequence of the mutagenic effect of aromatic amines.^[16, 53, 61-64] The occurrence of frameshift mutations is highly sequence specific. Probably, insights into the mechanism of frameshift mutations induced by AAF-dG could be obtained by placing the AAF-dG lesion in different oligonucleotide sequences and studying the bypass products of Pol η . Furthermore, the TLS polymerase κ is supposed to be cloned, expressed and purified in order to investigate the bypass of bulky adducts by this enzyme in comparison to Pol η .

An additional project was to start crystallization experiments of AAF-dG containing DNA in complex with the yeast NER protein Rad14 and its human homologue XPA. It has been shown that Rad14 binds DNA containing bulky adducts such as AAF-dG. To date, only a solution structure of the minimal DNA binding domain of XPA is available but no structural data about the lesion recognition step. A crystal structure together with AAF-dG would give valuable insights into the mechanism of lesion recognition.

5 Synthesis of C8-deoxyguanosine adducts

In order to investigate translesion synthesis and repair of bulky DNA adducts, oligonucleotides containing several bulky C8-adducts were prepared. In the past, oligonucleotides containing C8-arylamine and acetylarylamine adducts of dG have been synthesized using a post-synthetic modification strategy.^[144, 145] Therefore, oligonucleotides were treated with the corresponding *N*-hydroxy or *N*-acetoxy arylamine derivatives.^[144] This reaction is severely limited in yield and scope and the utilized arylamine derivatives are extremely toxic (carcinogenic). In particular, a single site-specific modification cannot be introduced into oligonucleotides that contain multiple guanines, therefore the sequence context in which adducted dG can be introduced is severely restricted (Figure 23, A).

Thus, the corresponding phosphoramidites, suitably protected for automated solid-phase DNA synthesis, were synthesized according to a reported strategy.^[146-148] By using the phosphoramidite approach, site specific incorporation of the acetylarylamine adducts into oligonucleotides is possible and consequently, the influence of the sequence context around the lesion can be investigated.

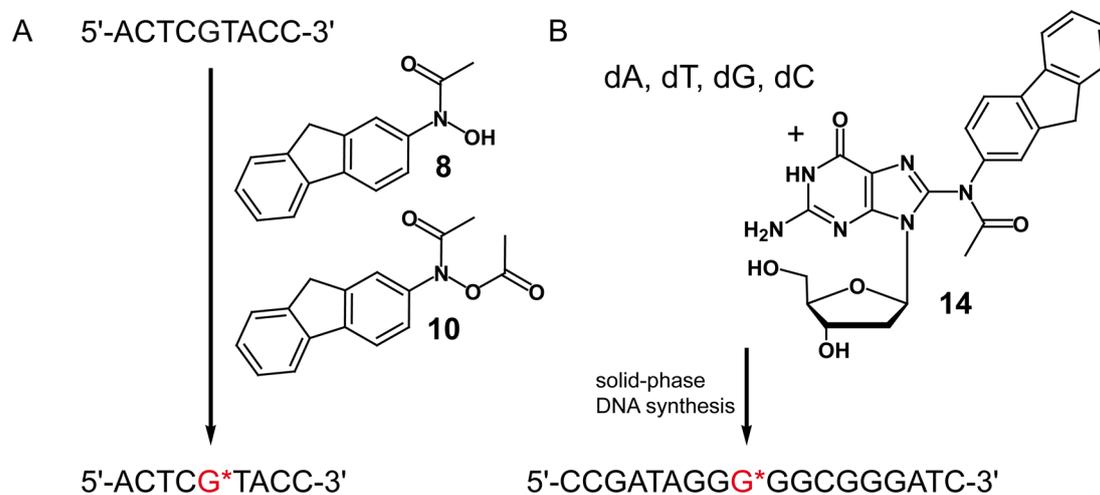


Figure 23. Preparation of AAF-dG containing oligonucleotides. A) Post-synthetic approach: Short DNA strands are reacted with *N*-hydroxy-AAF **8** or *N*-acetoxy-AAF **10**. B) Synthesis of an AAF-dG nucleoside **14** via Buchwald-Hartwig coupling of 2-aminofluorene to a protected 8-bromo-2'-dG (8-Br-dG) and site specific incorporation into oligonucleotides using solid-phase DNA synthesis.

The synthetic strategy for these adducts is based on Buchwald-Hartwig coupling reaction of protected 8-bromo-2'-deoxyguanosine (Br-dG) derivatives with aromatic amines (Figure 23, B). This method was first reported by *Rizzo*^[146] and later applied by *Schärer* to a synthetic route for AAF-dG **14**, suitably protected for automated solid-phase DNA synthesis.^[147, 148]

The synthesis poses several challenges. First, a mild deprotection step after oligonucleotide synthesis, which does not affect the base labile N^8 -acetyl group of AAF-dG, is obligatory. Thus, a protecting group for the exocyclic N^2 amino group of guanine which is cleavable under conditions maintaining the acetyl group, needs to be used. To this end, an *isopropylphenoxyacetyl* (*iPrPac*) protecting group was introduced at the N^2 position.

Moreover, acetylation of the N^8 position needs to be selective in the presence of the NH at position N^2 . The *iPrPac* group is not sterically demanding enough and increases the acidity of the NH, in consequence, the reaction is preferred at this site. Thus, the transient, sterically more demanding dimethoxytrityl (DMT) protecting group was used for the N^2 position of Br-dG prior to introduction of the *iPrPac* protecting group. The DMT group allows efficient coupling of several amines and further acetylation at the N^8 -position without affecting the N^2 position.

In this thesis, the aromatic amines aniline, 2-naphthylamine, 2-aminoanthracene, 2-aminofluorene and 1-aminopyrene were coupled to the C8 position of Br-dG using the strategy of Rizzo^[146] and Schärer^[147] and the corresponding phosphoramidites of **14**, **22-25**, suitably protected for automated solid-phase DNA synthesis, were prepared (Figure 24).

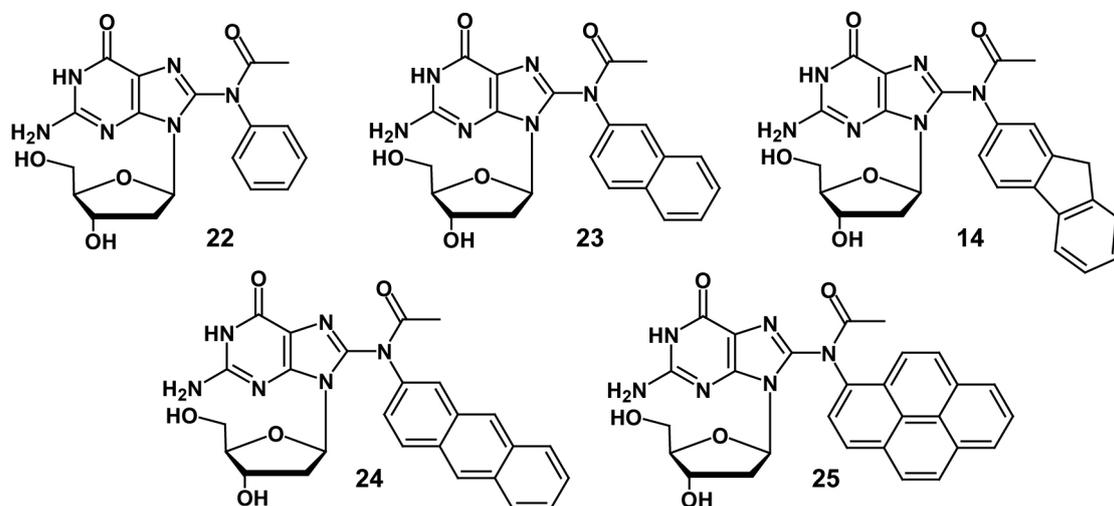


Figure 24. C8-dG adducts 8-(*N*-acetyl-amino-benzene)-2'-deoxyguanosine **22** (AAB-dG), 8-(*N*-acetyl-2-amino-naphthalene)-2'-deoxyguanosine **23** (AAN-dG), 8-(*N*-acetyl-2-amino-fluorene)-2'-deoxyguanosine **14** (AAF-dG), 8-(*N*-acetyl-2-amino-anthracene)-2'-deoxyguanosine **24** (AAA-dG) and 8-(*N*-acetyl-1-amino-pyrene)-2'-deoxyguanosine **25** (AAP-dG). Derivatives, suitably protected for solid-phase DNA synthesis, were prepared.

The main purpose of all syntheses presented in the following was to supply sufficient amounts of material for automated solid-phase DNA synthesis. Thus, as soon as this purpose was

accomplished no further optimization of the reaction conditions was performed. In Figure 25, the synthetic pathway to the acetylarlyamine dG adducts is depicted.

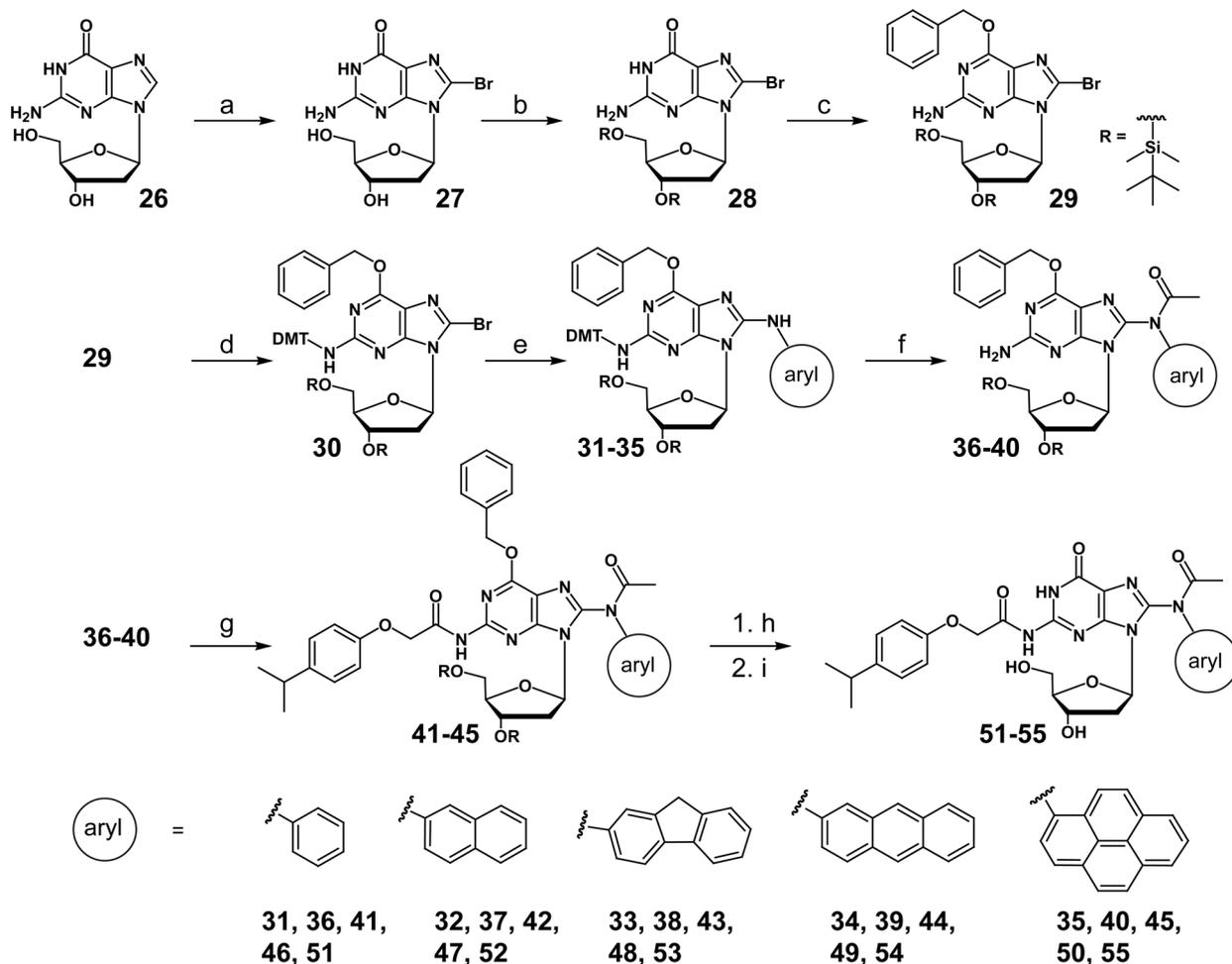


Figure 25. Synthetic pathway towards acetylarlyamine dG adducts bearing an *i*PrPac protecting group at N^2 -position for solid-phase DNA synthesis. Reaction conditions are: a) NBS, water, acetonitrile (**26** \rightarrow **27**); b) TBDMS-Cl, imidazole, DMF (**27** \rightarrow **28**); c) Bn-OH, PPh₃, DIAD, dioxane (**28** \rightarrow **29**); d) DMT-Cl, pyridine (**29** \rightarrow **30**); e) Pd₂(dba)₃, BINAP, toluene, NaOtBu, arylamines aniline / 2-naphthylamine / 2-aminofluorene / 2-aminoanthracene or 1-aminopyrene (**30** \rightarrow **31-35**); f) 1.) Ac₂O, Et₃N, DMAP, pyridine, 2.) 0.01 M HCl in MeOH (**31-35** \rightarrow **36-40**); g) *i*PrPac-Cl, pyridine (**36-40** \rightarrow **41-45**); h) CH₃COOH, TBAF, THF (**41-45** \rightarrow **46-50**); i) Pd/C, cyclohexene, EtOH (**46-50** \rightarrow **51-55**).

The synthesis started by bromination of dG **26** with *N*-bromosuccinimide (NBS) at C8 position and precipitation of the product in acetone as the sole purification step. Silylation of the 5' and 3' hydroxy groups with *tert*-butyldimethylsilyl chloride (TBDMS-Cl) resulted in compound **28**. The O^6 position was protected as a benzyl ether (**29**) under Mitsunobu conditions. The exocyclic N^2 amino group was further protected with dimethoxytrityl chloride (DMT-Cl), which is stable under basic conditions and thus survives the subsequent Buchwald-Hartwig coupling reaction (**30**). In the Buchwald-Hartwig coupling reaction, the

arylamines aniline, 2-naphthylamine, 2-aminoanthracene, 2-aminofluorene or 1-aminopyrene were coupled to **30** using tris-(dibenzylideneacetone)-dipalladium ($\text{Pd}_2(\text{dba})_3$) and rac-2,2'-bis(diphenylphosphino)-1,1'-binaphthyl (BINAP) as a catalyst, sodium *tert*-butoxide (NaOtBu) as a base, and toluene as a solvent. The reaction gave the adducted nucleosides **31-35** (Figure 26).

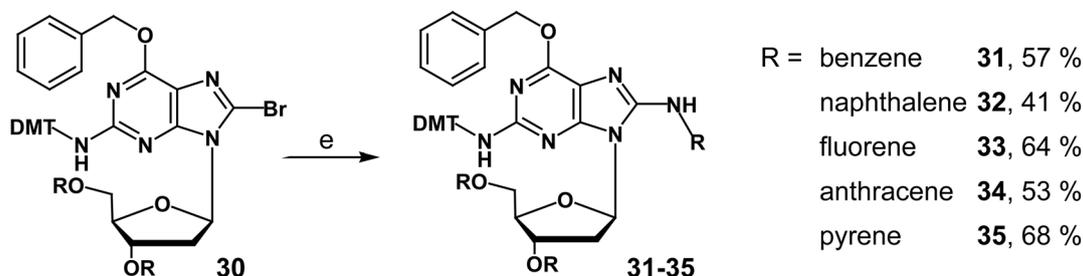


Figure 26. Buchwald-Hartwig coupling reaction of the protected Br-dG **30** with several arylamines (aniline, 2-naphthylamine, 2-aminoanthracene, 2-aminofluorene or 1-aminopyrene) provided the adducted nucleosides **31-35** (yields are shown next to the compound number). Reaction conditions are $\text{Pd}_2(\text{dba})_3$, BINAP, toluene, NaOtBu , arylamines: aniline / 2-naphthylamine / 2-aminofluorene / 2-aminoanthracene or 1-aminopyrene.

Protection of both the N^2 and O^6 position is absolutely required for the successful coupling reactions in order to avoid side reactions. Not only 2-acetylaminofluorene and 1-aminopyrene (literature known)^[147] but also aniline, 2-aminonaphthalene and 2-aminoanthracene could be successfully coupled to **30** in reasonable yields (Figure 26). Direct coupling of **30** with 2-acetylaminofluorene is difficult and does not work using the applied coupling conditions.^[147] However, the acetylated arylamine dG adducts **36-40** were obtained by selective acetylation of the N^8 position with acetic anhydride adding triethylamine and 4-(dimethylamino)-pyridine (DMAP) in pyridine (Figure 27).

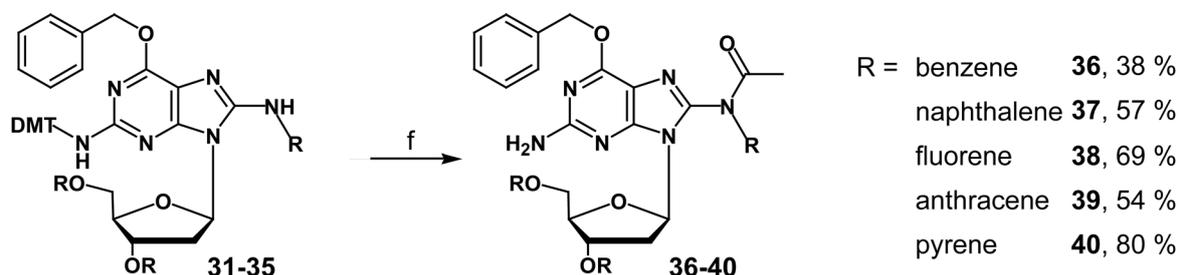


Figure 27. Acetylation of the N^8 position of the arylamine dG adducts **31-35** resulting in the nucleosides **36-40**. Yields are given next to the compound number. Reaction conditions are 1.) Ac_2O , Et_3N , DMAP, pyridine, 2.) 0.01 M HCl in MeOH.

The large, sterically demanding DMT group at N^2 position prevents acetylation at this position. The product of the acetylation is rather acid- and base-sensitive, therefore, the DMT group had to be removed under extremely mild conditions using dilute HCl in methanol

without previous purification of the acetylation reaction product. This two-step reaction known for AF-dG^[147] **33** was applied to the other bulky adducts **31**, **32**, **34** and **35** as well, giving the corresponding acetylated nucleosides **36**, **37**, **39** and **40** in good yields. The *N*² position was subsequently protected with *i*propylphenoxyacetyl chloride (*i*PrPac-Cl, synthesized from (4-*Isopropyl*)-phenoxy acetic acid) in pyridine providing **41-45**.

The silyl protecting groups of **41-45** were finally removed using tetrabutylammonium fluoride (TBAF) buffered with acetic acid to maintain the base-sensitive *N*⁸ acetyl group. This gave the nucleosides **46-50**. Removal of the benzyl protection group was carried out by hydrogenolysis using palladium on charcoal and cyclohexene as hydrogen donor yielding the adducted nucleoside building blocks ready for protection and activation for solid-phase DNA synthesis (**51-55**).

Synthesis of the phosphoramidites

The *i*PrPac-protected acetylarylamine-2'-deoxyguanosine adducts **51-55** were selectively DMT-protected at their 5' hydroxyl groups using dimethoxytrityl chloride (DMT-Cl) in pyridine (Figure 28, a).

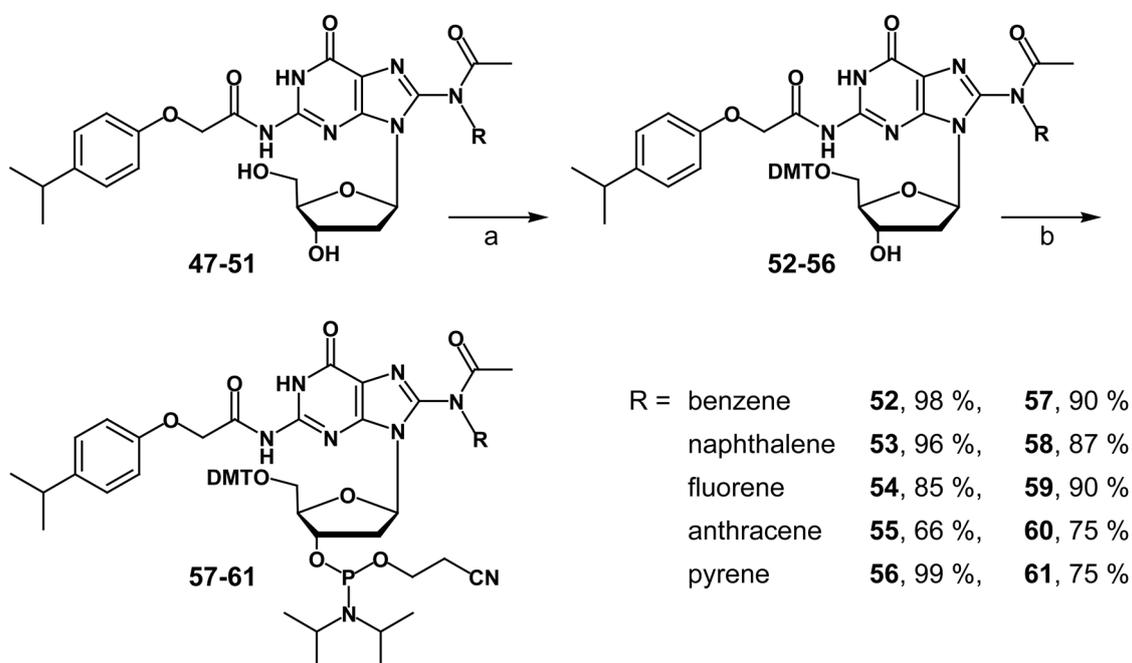


Figure 28. DMT-protection and synthesis of the phosphoramidites **61-65** for solid-phase DNA synthesis. Reaction conditions are a) DMT-Cl, pyridine and b) Bis-(diisopropylamino)-(2-cyanoethoxy)-phosphine, DIAT, CH₂Cl₂.

Last step of the synthesis was the introduction of the 3' phosphoramidite function. Therefore, **56-60** were reacted with bis-(diisopropylamino)-(2-cyanoethoxy)-phosphine using

diisopropylammonium tetrazolide (DIAT) as activator to yield the acetylamine dG phosphoramidites **61-65** which could be used for solid-phase DNA synthesis (Figure 28).

Table 1 gives an overview about the yields obtained for the synthetic routes to the various 2'-deoxyguanosine phosphoramidites containing acetylated C8-adducts of aniline, 2-naphthylamine, 2-aminoanthracene, 2-aminofluorene and 1-aminopyrene.

synthetic step	no.		yield	
bromination	27		85 %	
TBDMS protection	28		91 %	
O ⁶ -benzyl protection	29		45 %	
DMT protection	30		99 %	

	benzene		naphthalene		fluorene		anthracene		pyrene	
	AAB-dG		AAN-dG		AAF-dG		AAA-dG		AAP-dG	
synthetic step	no.	yield	no.	yield	no.	yield	no.	yield	no.	yield
coupling	31	57 %	32	41 %	33	64 %	34	53 %	35	68 %
acetylation	36	38 %	37	57 %	38	69 %	39	54 %	40	80 %
iPrPac protection	41	86 %	42	91 %	43	92 %	44	84 %	45	99 %
OH deprotection	46	63 %	47	95 %	48	57 %	49	74 %	50	62 %
O ⁶ deprotection	51	63 %	52	68 %	53	77 %	53	36 %	55	82 %
DMT protection	56	98 %	57	96 %	58	85 %	59	66 %	60	99 %
phosphitylation	61	90 %	62	87 %	63	90 %	64	75 %	65	75 %

Table 1. Yields obtained for the synthetic steps to the arylamine dG phosphoramidites.

Even though the individual yield (Table 1) are quite acceptable, the overall yields are low and careful handling in each of the 11 steps is required. Critical transformations are still the Buchwald-Hartwig coupling and the acetylation step with yields of approximately 50%. Figure 29 summarizes the synthesized acetylamine-dG-phosphoramidites which are subsequently used for solid-phase DNA synthesis.

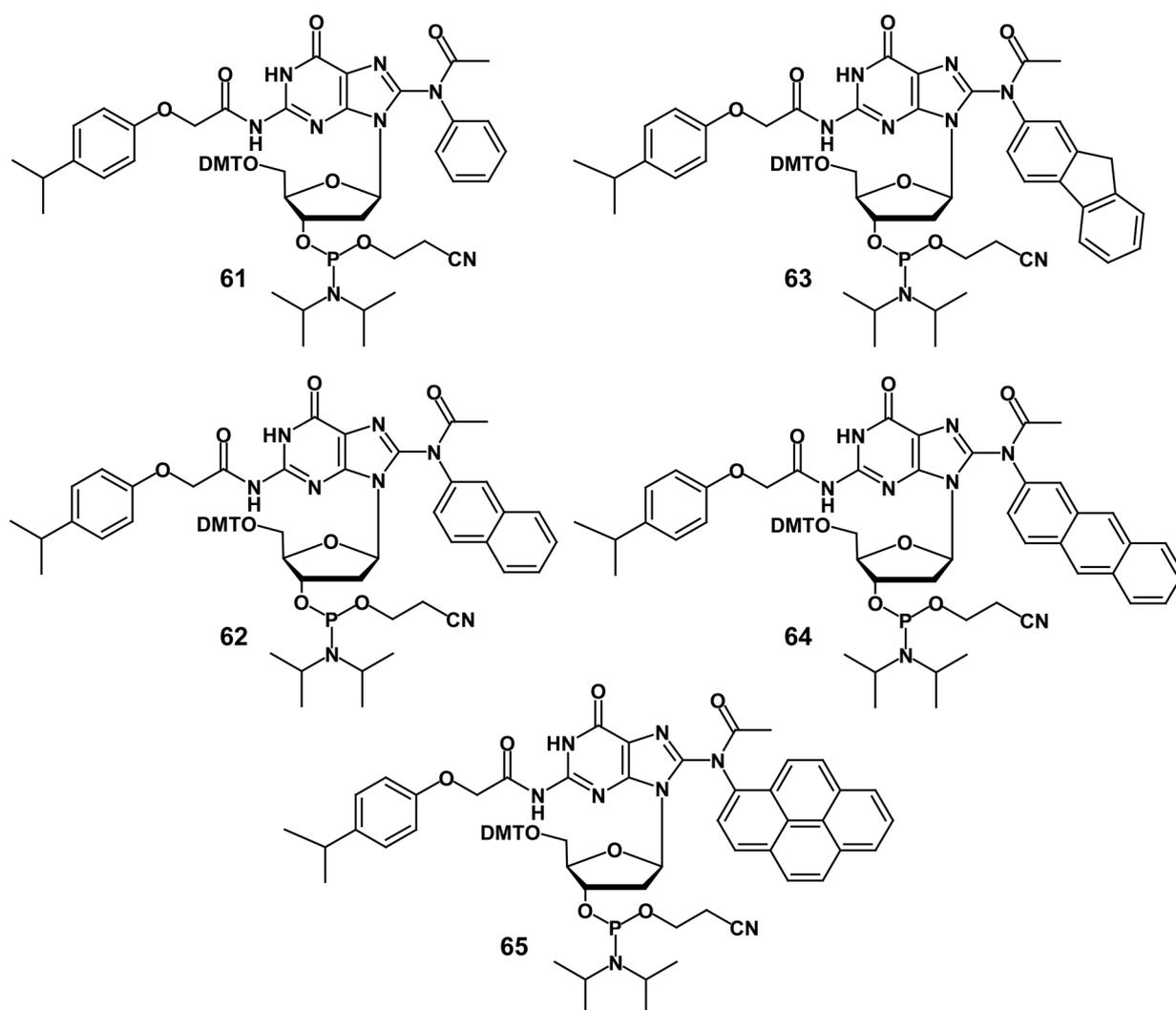


Figure 29. Synthesized acetylated arylamine-dG phosphoramidites further used for solid-phase DNA synthesis: 8-(*N*-acetyl-amino-benzene)-3'-*O*-(2'-cyanoethoxydiisopropyl-amino-phosphino)-(5'-*O*-dimethoxytrityl)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine **61**, 8-(*N*-acetyl-2-amino-naphthalene)-3'-*O*-(2'-cyanoethoxydiisopropyl-amino-phosphino)-(5'-*O*-dimethoxytrityl)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine **62**, 8-(*N*-acetyl-2-amino-fluorene)-3'-*O*-(2'-cyanoethoxydiisopropyl-amino-phosphino)-(5'-*O*-dimethoxytrityl)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine **63**, 8-(*N*-acetyl-2-amino-anthracene)-3'-*O*-(2'-cyanoethoxydiisopropyl-amino-phosphino)-(5'-*O*-dimethoxytrityl)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine **64** and 8-(*N*-acetyl-1-amino-pyrene)-3'-*O*-(2'-cyanoethoxydiisopropyl-amino-phosphino)-(5'-*O*-dimethoxytrityl)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine **65**.

6 Incorporation of C8-dG adducts into oligonucleotides

6.1 Solid phase DNA synthesis

The DNA oligonucleotides were synthesized on a *PerSeptive Biosystems Expedite 8900* Synthesizer and an *Äkta Oligopilot 10 (GE)* following standard phosphoramidite protocols for the solid phase oligonucleotide synthesis.^[149, 150] The used amounts of the phosphoramidites of the bulky adducts were equal to those of the standard bases but coupling times were extended to 10 minutes. Ultramild protected phosphoramidites were used, which could be deprotected under mild conditions without cleaving the base-labile acetyl group of AAF-dG.^[148] These ultramild phosphoramidites contain an *isopropylphenoxyacetyl* (*iPrPac*) group at N^2 position of dG, an acetyl group at N^4 position of dC and a phenoxyacetyl group at N^6 position of dA (Figure 30, B). A special linker, the Q-linker (hydroquinone-*O,O'*-diacetate linker with CPG resin) was used, which can be cleaved faster and under milder conditions compared to the standard succinate linker (Figure 30, B).^[148]

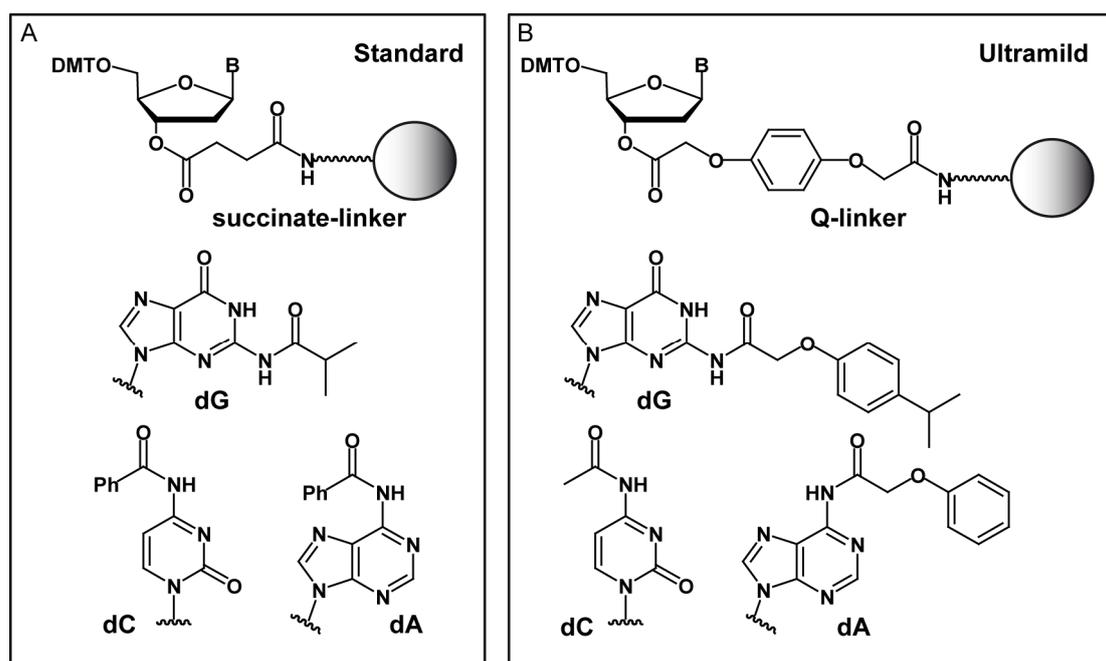


Figure 30. Protecting groups used for the exocyclic amino functions of the nucleobases and linker employed to attach the DNA to the solid support for DNA synthesis. A) Standard protection groups. The N^2 position of dG is protected with an *isobutyryl*, the N^4 position of dC and the N^6 position of dA with a benzoyl group. The first base is coupled via a succinate linkage to the solid support. B) Protection groups and linker for ultramild DNA synthesis. The N^2 position of dG is protected with an *isopropylphenoxyacetyl*, the N^4 position of dC with an acetyl and the N^6 position of dA with a phenoxyacetyl group. Attachment to the solid support was achieved via a hydroquinone-*O,O'*-diacetate linker (Q-linker).

In addition, a different, more bulky capping reagent had to be used compared to standard methods (acetic anhydride) to avoid exchange of the *i*PrPac protecting group on the dG by acetate from the typically employed acetic anhydride capping solution. For all syntheses, phenoxyacetic anhydride, synthesized from phenoxyacetic acid, was consequently used as the capping reagent.

The AAF-dG as well as the other C8-dG adducts were stable under the conditions of the solid-phase DNA synthesis, including the deblocking step with dichloroacetic acid and the oxidation step with iodine in water/acetonitrile.

The oligonucleotide synthesis was monitored online by measurement of the UV-Vis absorption of the reagents leaving the solid support cartridge. Since every consecutive coupling step finishes with the acidic cleavage of the 5'-DMT protecting group (“detritylation”), this process can be used to quantify coupling efficiencies by monitoring the cleavage of the 4,4'-dimethoxytrityl cation at a wavelength of 500 nm. For this work, various modified oligonucleotides carrying the acetylarylamine-dG adducts (Figure 29) were synthesized. Coupling of the acetylarylamine dG phosphoramidites was usually as efficient as the coupling of the standard ultramild phosphoramidites. A representative chart of the trityl-values obtained during the synthesis of the oligonucleotide ODN **33** on the *Expedite 8900* DNA synthesizer is depicted in Figure 31.

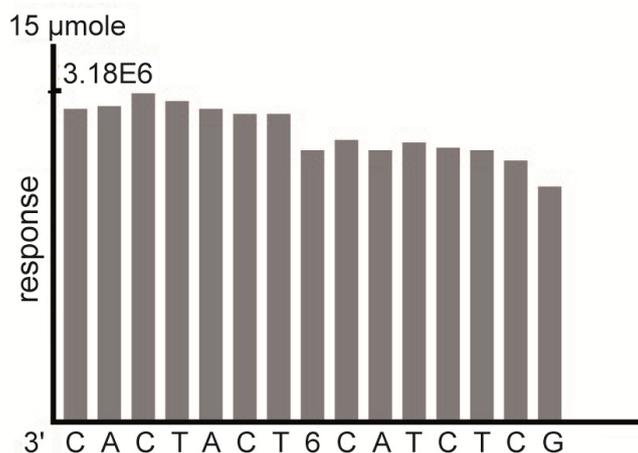


Figure 31. Trityl values monitored during DNA synthesis of ODN **33** on an *Expedite 8900* DNA synthesizer. Position 6 marks the incorporation of the AAP-dG. Coupling of the modified phosphoramidite is nearly as efficient as the coupling of the standard ultramild phosphoramidites.

Following the automated oligonucleotide synthesis, the DNA strands were cleaved from the solid support and all protecting groups were removed from the bases and the backbone. The standard method for both processes is the treatment of the solid-phase bound oligonucleotides

with aqueous ammonia for several hours. This method cannot be used for the acetylated arylamine containing oligonucleotides due to their labile N^8 -acetyl groups. Particularly mild deprotection conditions involving 10 % diisopropylamine and 0.25 M β -mercaptoethanol in methanol at 55°C for 20 hours were effective in removing the protecting groups from the oligonucleotides and in releasing them from the Q-solid support without cleaving the base-labile acetyl group at N^8 of the acetylarlyamine dG adducts.^[148] β -Mercaptoethanol was added to avoid oxidation of the acetylarlyamine dG adducts during the long incubation at high temperatures.^[151, 152]

Not only the literature known AAF-dG could be successfully incorporated into oligonucleotides, but, in this way, also the other acetylarlyamine adducts AAB-dG, AAN-dG, AAA-dG and AAP-dG were successfully inserted. A list of all DNA strands containing the bulky adducts synthesized for this thesis can be found in the experimental section (10.2.2, Table 3).

6.2 Purification and characterization of the various C8-dG adduct containing oligonucleotides

6.2.1 HPLC purification and characterization via MALDI-TOF mass spectrometry

After deprotection and cleavage from the solid support, the oligonucleotide solutions were concentrated *in vacuo*, redissolved in ddH₂O and the oligonucleotides were purified via preparative RP-HPL-chromatography. The preparative *Waters* HPLC system allows monitoring a second wavelength during purification. Thus, the oligonucleotides containing aromatic adducts could be easily detected at wavelengths between 300 and 350 nm depending on the aromatic unit attached. Thus separation from shorter truncated fragments was in this way facilitated. Moreover, due to the aromatic adducts, the retention time of the modified oligonucleotides was delayed by 5-10 minutes compared to unmodified oligonucleotides.

Figure 32 shows an exemplary HPL-chromatogram of a crude, AAP-dG containing oligonucleotide directly after DNA synthesis, as well as examples of purified, the various bulky adduct containing oligonucleotides analyzed via analytical HPL-chromatography with a purity greater than 98 %. All oligonucleotides gave crude HPLC data similar to these reported.

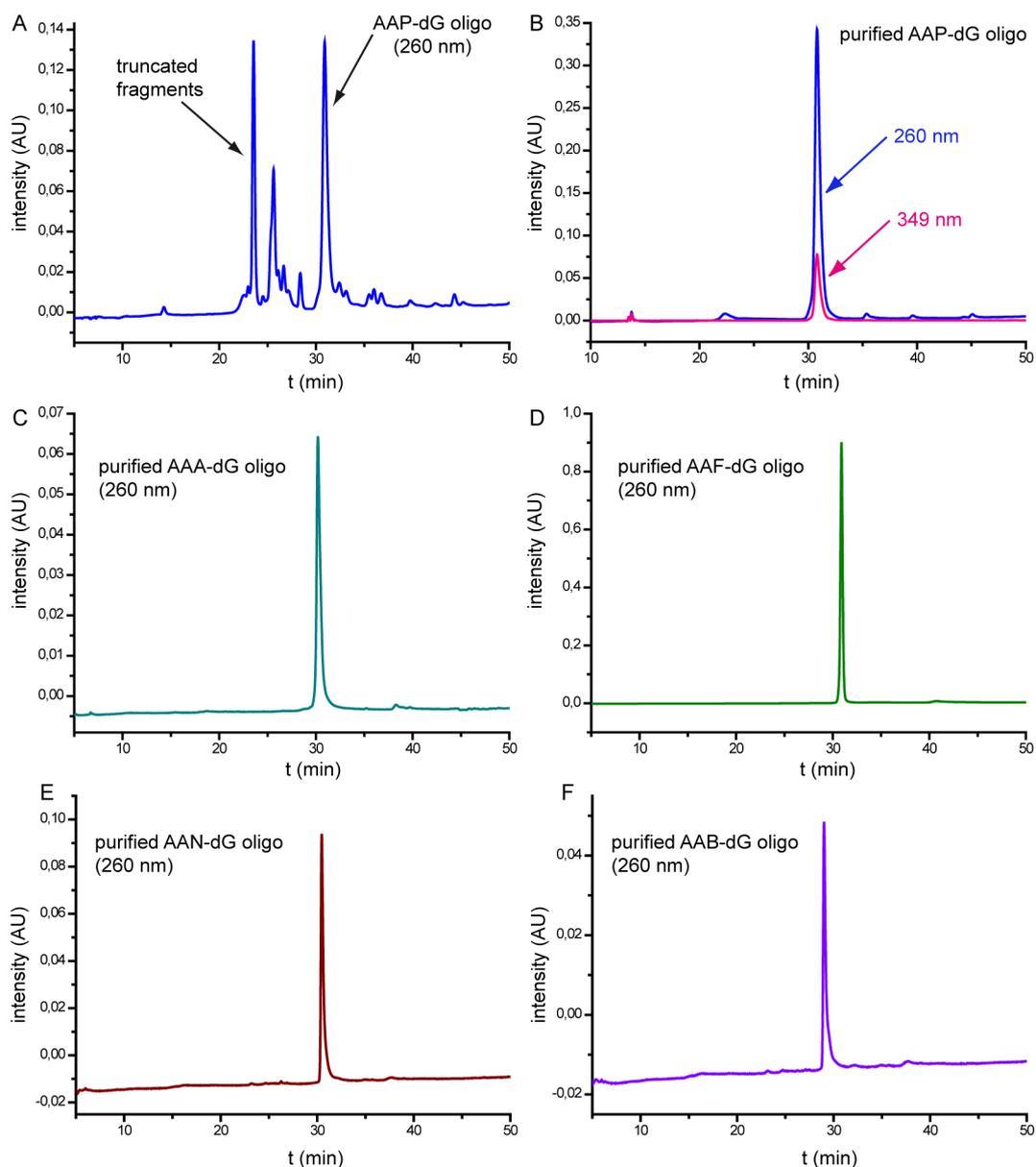


Figure 32. HPL-chromatograms of the synthesized acetylarylamine containing oligonucleotides. A) Crude HPL-chromatogram of the AAP-dG containing **ODN 32*** directly after cleavage from the solid support and deprotection. Several shorter, truncated oligonucleotides are visible. B) **ODN 32** after purification. The oligonucleotide can be detected at 260 and 349 nm due to the large aromatic system attached. C) HPL-chromatogram of the purified AAA-dG containing **ODN 28**. D) HPL-chromatogram of the purified AAF-dG containing **ODN 16**. E) HPL-chromatogram of the purified AAN-dG containing **ODN 25**. F) HPL-chromatogram of the purified AAB-dG containing **ODN 22**.

The identity of the C8-adduct containing oligonucleotides was confirmed by MALDI-Tof mass spectrometry. In Figure 33, mass spectra of different oligonucleotides containing the various synthesized lesions (AAB-dG, AAN-dG, AAF-dG, AAA-dG and AAP-dG) are displayed as examples.

* The synthesized strands are summarized in the experimental section and continuously numbered from 1-33.

To obtain deacetylated AF-dG oligonucleotides needed for further biochemical studies, they were prepared from the corresponding AAF-dG containing oligonucleotides by simple deacetylation. To this end, the oligonucleotide was incubated in sodium hydroxide solution in the presence of β -mercaptoethanol. Total cleavage of the acetyl group was achieved after three hours at 37 °C. Afterwards, the oligonucleotides were again purified by reversed phase HPLC (Figure 33, B).

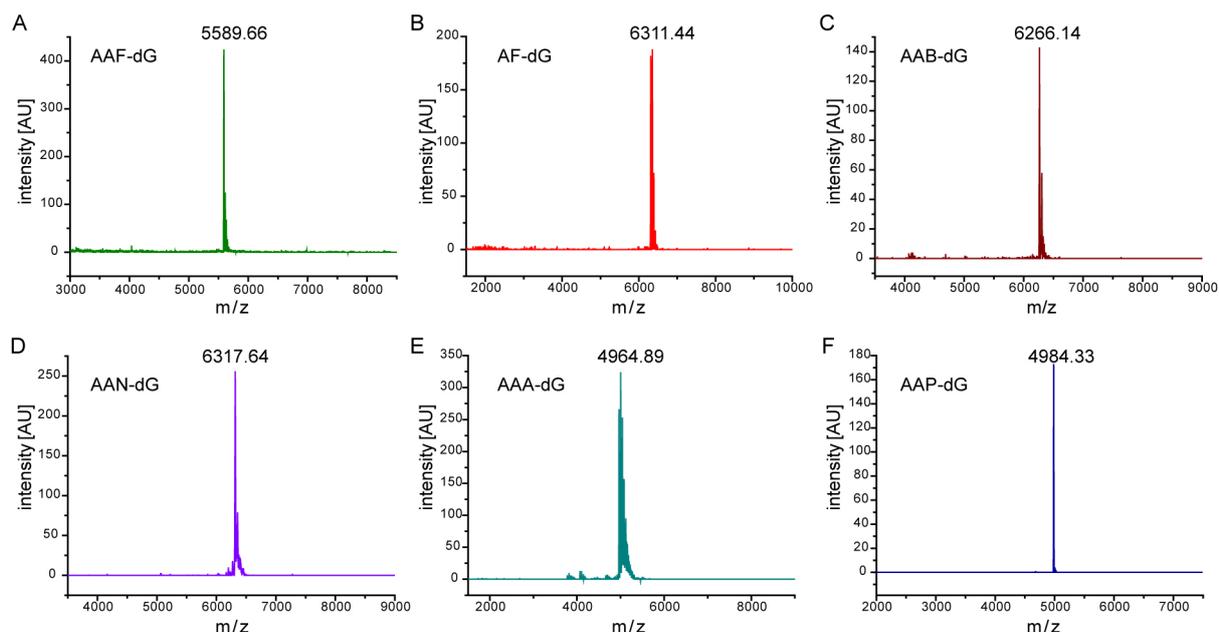


Figure 33. MALDI-ToF mass spectra of purified bulky adduct containing oligonucleotides. A) **ODN 17**, $m/z_{calc} = 5592.6$; B) **ODN 1** after cleavage of its acetyl group, $m/z_{calc} = 6313.3$; C) **ODN 21**, $m/z_{calc} = 6267.9$; D) **ODN 24**, $m/z_{calc} = 6317.7$; E) **ODN 29**, $m/z_{calc} = 4964.8$; F) **ODN 32**, $m/z_{calc} = 4987.3$.

6.2.2 Enzymatic digestion analysis

Analysis of the prepared oligonucleotides by enzymatic digestion was performed to further prove the presence of the bulky adduct lesions and to ensure that the additional acetyl group found in mass spectrometry of the intact oligonucleotides was correctly located on the N^8 of the adducted dG and not on a dC residue due to incomplete deprotection of an acetyl-dC. The modified oligonucleotides were incubated with a large excess of snake venom phosphodiesterase following incubation with calf intestine phosphatase. The resulting nucleosides were analyzed by HPLC-MS (Figure 34).

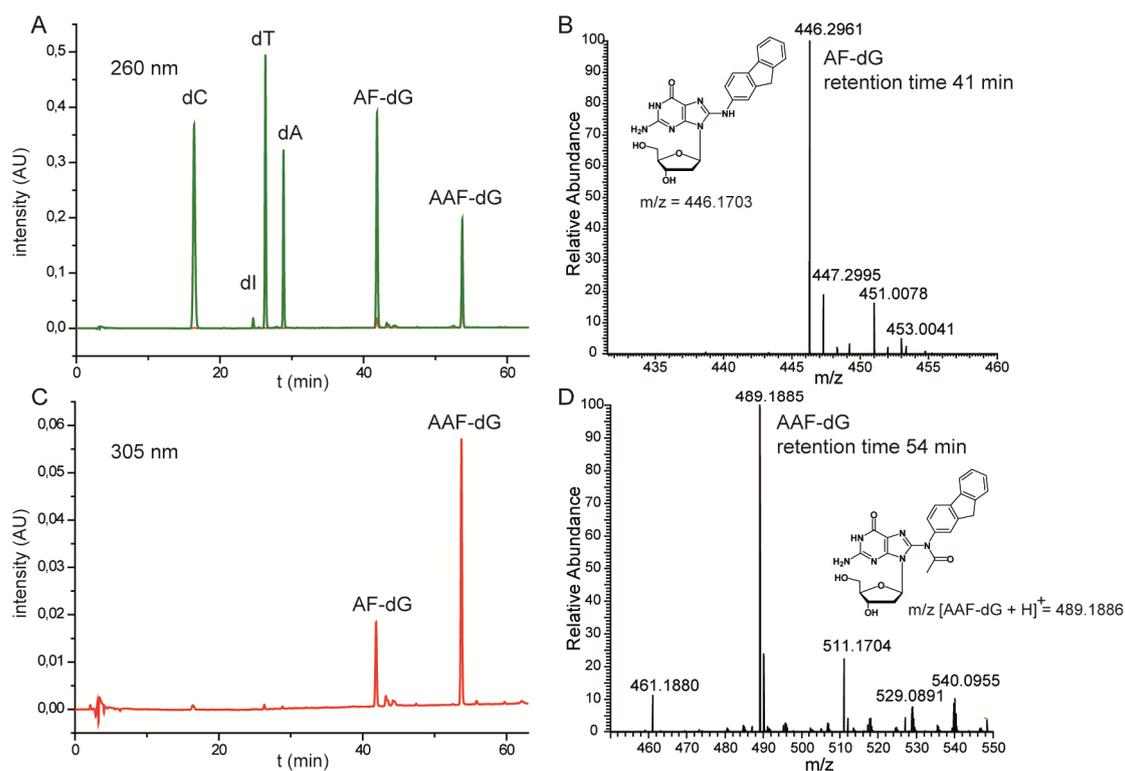


Figure 34. A) HPL-chromatogram of the enzymatic digest of the AAF-dG containing **ODN 3** at 260 nm. B) ESI-MS spectrum of the peak eluting at a retention time of 41 min. The signal corresponds to the deacetylated AF-dG. C) HPL-chromatogram of the digest of **ODN 3** at 305 nm. D) ESI-MS spectrum of the peak eluting at a retention time of 54 min. The signal corresponds to the AAF-dG nucleoside.

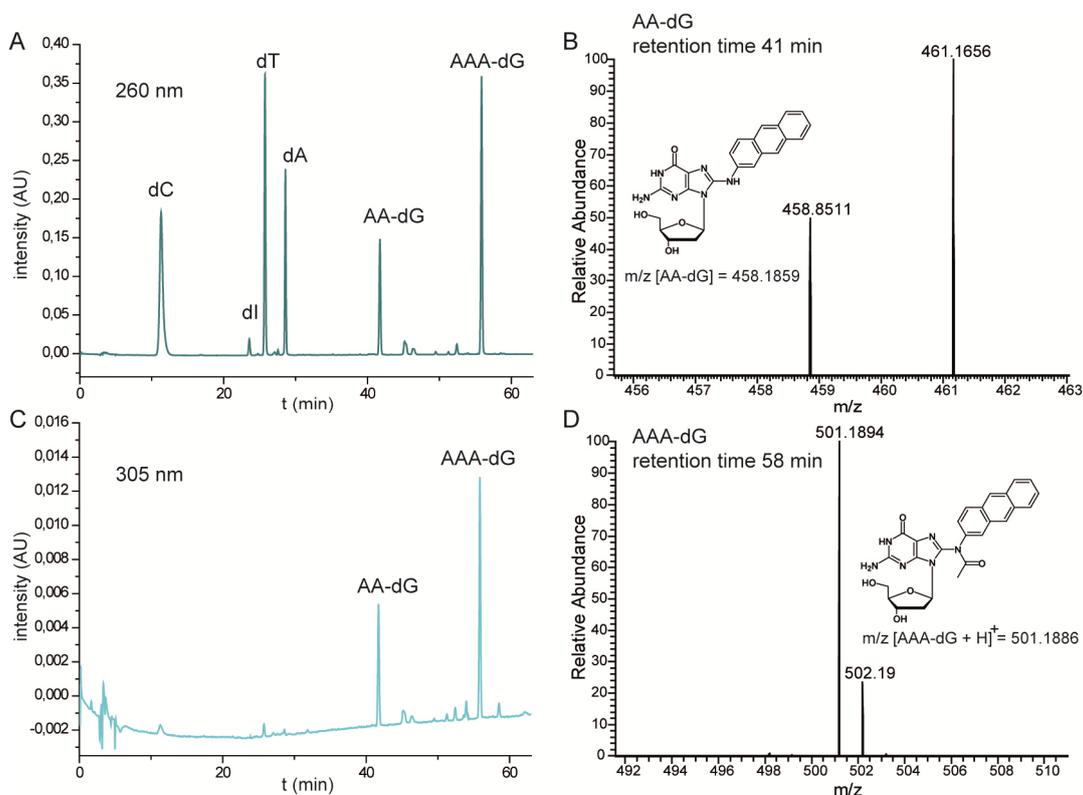


Figure 35. A) HPL-chromatogram of the enzymatic digest of the AAA-dG containing **ODN 29** at 260 nm. B) ESI-MS spectrum of the peak eluting at a retention time of 41 min (AA-dG). C) HPL-chromatogram of the digest of **ODN 29** at 305 nm. D) ESI-MS spectrum of the peak eluting at a retention time of 58 min. The signal corresponds to the AAA-dG nucleoside.

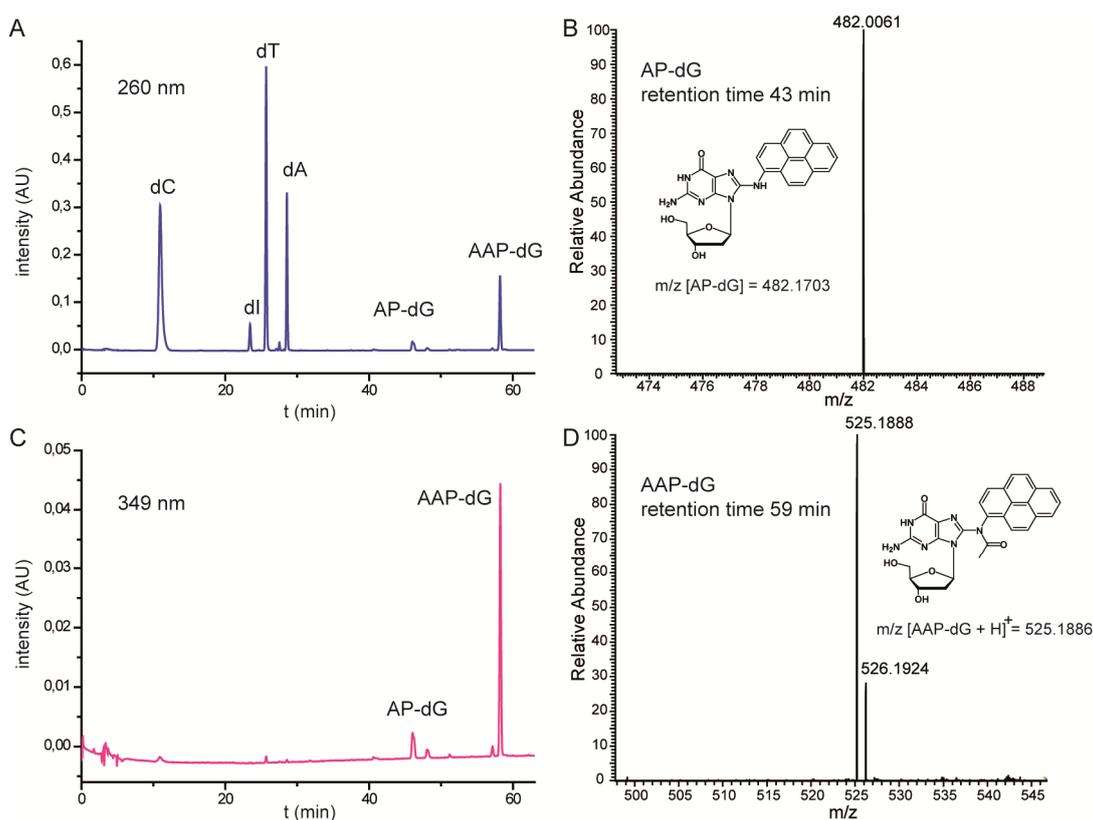


Figure 36. A) HPL-chromatogram of the enzymatic digest of the AAP-dG containing **ODN 32** at 260 nm. B) ESI-MS spectrum of the peak eluting at a retention time of 43 min (AP-dG). C) HPL-chromatogram of the digest of **ODN 32** at 349 nm. D) ESI-MS spectrum of the peak eluting at a retention time of 59 min. The signal corresponds to the AAP-dG nucleoside.

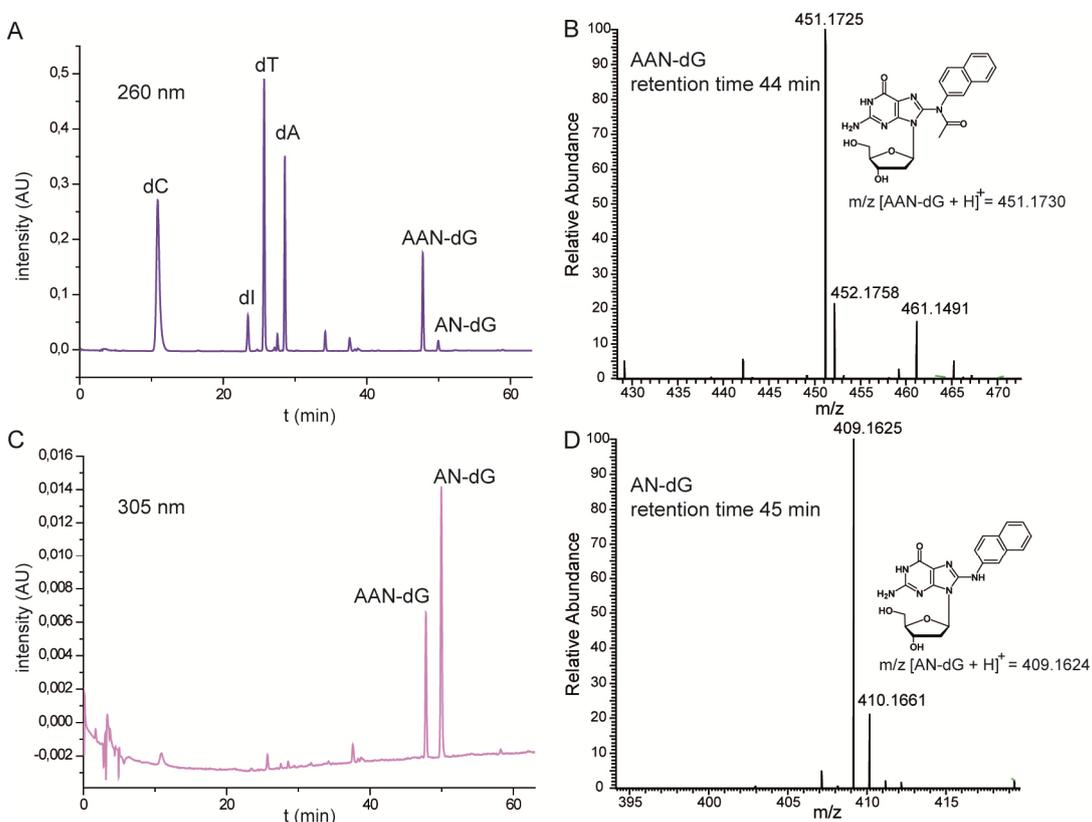


Figure 37. A) HPL-chromatogram of the enzymatic digest of the AAN-dG containing **ODN 26** at 260 nm. B) ESI-MS spectrum of the peak eluting at a retention time of 44 min. The signal corresponds to

the AAN-dG nucleoside. C) HPL-chromatogram of the digest of **ODN 26** at 305 nm. D) ESI-MS spectrum of the peak eluting at a retention time of 45 min. The signal corresponds to the deacetylated AN-dG nucleoside.

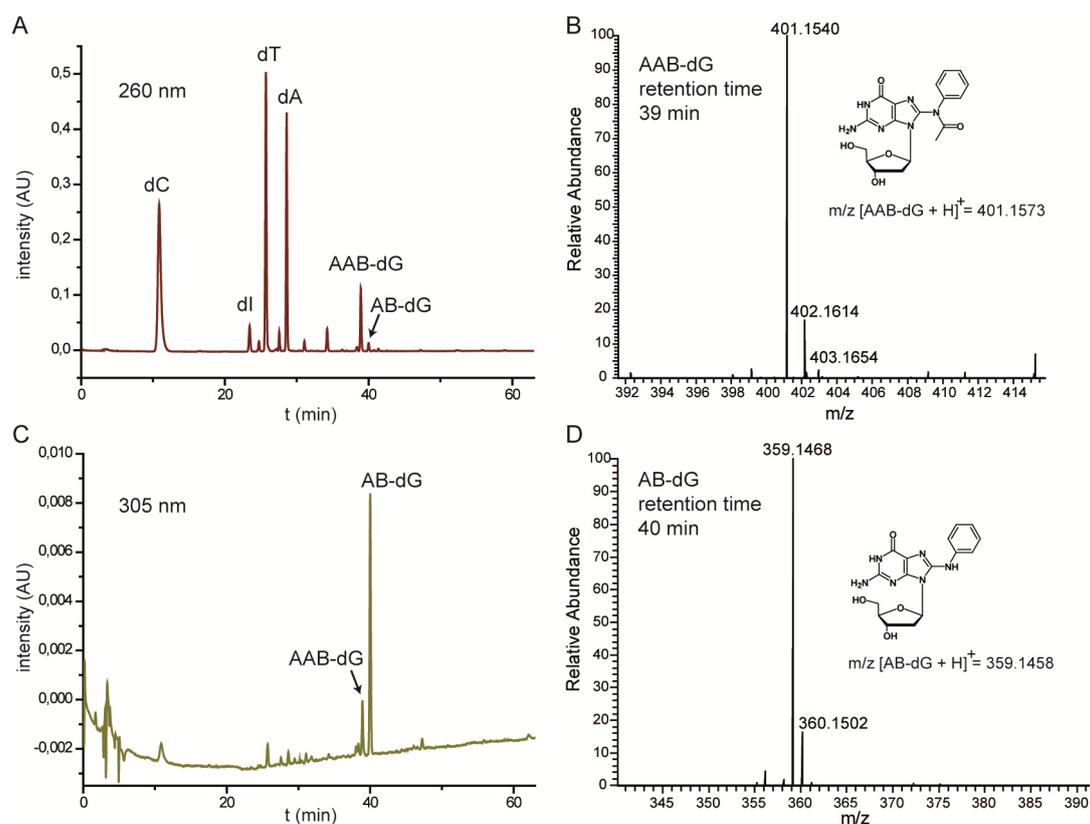


Figure 38. A) HPLC-chromatogram of the enzymatic digest of the AAB-dG containing **ODN 23** at 260 nm. B) ESI-MS spectrum of the peak eluting at a retention time of 39 min. The signal corresponds to the AAB-dG nucleoside. C) HPLC-chromatogram of the digest of **ODN 23** at 305 nm. D) ESI-MS spectrum of the peak eluting at a retention time of 40 min. The signal corresponds to the deacetylated AB-dG nucleoside.

The canonical bases dC, dT and dA as well as dI, which is the deamination product of dA, were detected in all HPLC-spectra. The nucleosides were assigned via their ESI mass spectra to their calculated molecular masses. Moreover, the peaks corresponding to the adducted dG's were identified. Additionally, a second peak corresponding to the deacetylated adduct is visible for all adducts. The deacetylation of the adducted dG's occurs during the enzymatic digest, which is performed at pH 8.5 and 37°C for several hours. The pH optimum for the snake venom phosphodiesterase is between pH 9.8-10.4.^[153] A compromise had to be found between deacetylation of the adducted dG and maintaining the activity of the enzyme. However, the HPLC and mass spectra of the total oligonucleotides show only one peak corresponding to the acetylated adducts (Figure 33). Thus, deacetylation occurs only under the conditions used for the enzymatic digestion and the acetylated adducts are stable at physiological pH.

Furthermore, several small peaks are detected with retention times similar to the deacetylated and acetylated adducts. These peaks correspond to mass values which could be decomposition products of the deacetylated adducted dG's. *Kadlubar et al.* found a degradation product in the alkaline enzymatic digests (without addition of an antioxidant) of DNA containing a C8-naphthalene-dG adduct.^[154] This was assumed to be an alternate ring-opened structure. Currently, there is not enough spectroscopic evidence to support this theory. *Kriek et al.* found that AF-dG decomposed under alkaline conditions into three new compounds which were assigned to the structures **67**–**68a, b** (Figure 39).^[155] All these structures are still speculative.^[152] In the present thesis, no reduction agent such as β -mercapthoethanol was added to the digest solutions. Thus oxidative decomposition of the adducted dG's may occur.

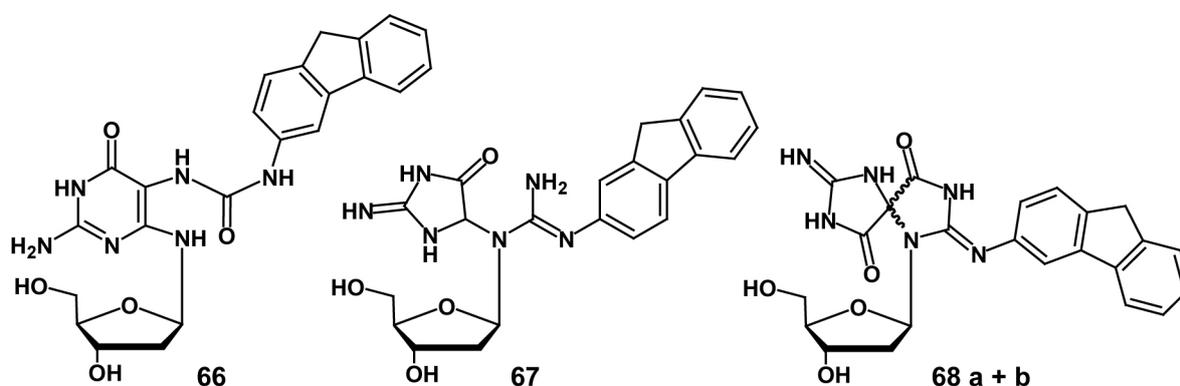


Figure 39. Speculative degradation products of AF-dG: Ring-opened structures **66** and **67** and spirodiastereomers **68a** and **68b**.

Particularly after enzymatic digestion of the AAN, AAP and AAB-dG containing oligonucleotides, additional small peaks were detected in the HPL-chromatograms at 260 nm. Analysis of these peaks gave mass values which are in agreement with ring-opened and spiro-structures. An example of the data obtained for ODN **26** is shown below (Figure 40).

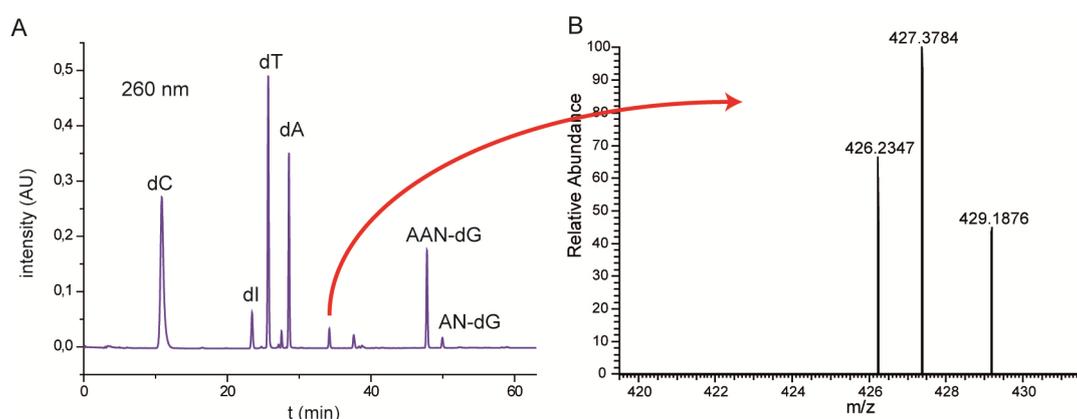


Figure 40. A) HPL-chromatogram of the enzymatic digestion of the AAN-dG containing ODN **26** at 260 nm. B) ESI-MS spectrum of the signal with a retention time of 34 min. Calculated m/z mass values of the ring-opened structure (analogous to **66**) and the spiro-structure (analogous to **68 a** and **b**) are 426.1652 and 424.1496, respectively.

The assignment of the found mass values to the corresponding spiro-compounds is highly speculative. Nevertheless, due to the used digestion conditions, some degradation products of the modified nucleosides occur.

In summary, the digestion analysis of the acetylarylamine-dG containing oligonucleotides proves that the modified phosphoramidites were successfully incorporated into oligonucleotides and that the acetyl groups found by mass spectrometry are located at the adducted dG's and not at a dC within the DNA strand due to incomplete deprotection. However, due to the used basic digestion conditions and a lack of reduction agents in the digest solution, several decomposition products of the modified nucleosides were additionally found. These degradation products occurred only under digestion conditions and were not found in the HPLC and mass spectra of the synthesized oligonucleotides. The lesions are hence stable under physiological conditions and could be used for biochemical investigations.

6.2.3 Melting point analysis

In order to determine the thermal stability of DNA duplexes containing acetylarylamine dG lesions, melting point measurements were performed. The optical density of oligonucleotide solutions at a wavelength of 260 nm is different for single strands and double strands. Double strands have a lower extinction coefficient at 260 nm than the corresponding mixture of single strands because the absorption of the aromatic systems of the bases is reduced by the stacking interactions in the double helix. Heating of hybridized DNA duplexes results, above a certain temperature depending on the length, the oligonucleotide sequence and the buffer conditions, in dissociation of the duplexes into single strands. Thus, the height of the UV absorption maximum increases (hyperchromicity). The melting point (T_m) of an oligonucleotide duplex is defined as the transition point of the absorption at 260 nm plotted against the temperature. The sequence 5'-TCTCTCG*CTCATCCAC-3' containing either the AAF-dG, AF-dG or the AAP-dG adduct was annealed to the corresponding counter strand in tris buffer at pH 7.4. Several cycles of heating and cooling between 0 to 85 °C were performed and the average value was used for calculating of the melting points. In Figure 41, the measured melting curves of AAF-, AF- and AAP-dG containing DNA duplexes are displayed.

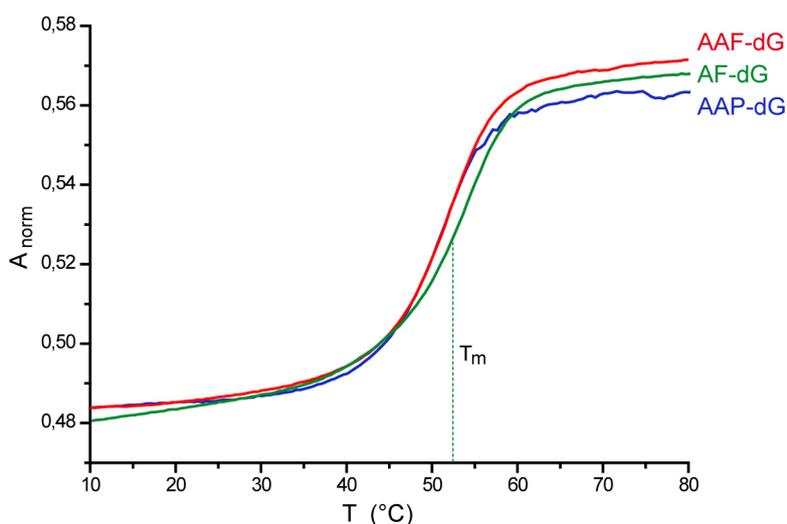


Figure 41. Graphical comparison of the melting curves of the sequence 5'-TCTCTCG*CTCATCCAC-3' containing different modified G's (either AAF-dG, AF-dG or AAP-dG). Melting points of the oligonucleotides are: AAF-dG = 49.8 °C (red curve), AAP-dG 51.4 °C (blue curve) and AF-dG 52.5 °C (green curve). The calculated melting point of the undamaged oligonucleotide is 55 °C.

The UV-based melting curve studies with AAF-dG, AF-dG and AAP-dG containing oligonucleotides show that all lesion containing oligonucleotides form stable duplexes with melting temperatures around 50 °C (Figure 41). The undamaged oligonucleotide duplex has a slightly higher melting temperature of 55 °C. Even the acetylated AAF-dG and AAP-dG lesions reduce the melting point of the duplex by only a small amount (around 4-5 °C) showing that the adducts cause only small overall structural perturbations of the duplex structure.

7 Bypass of bulky DNA adducts by Y-family DNA polymerases

The Y-family DNA polymerases Pol η and Pol κ were analyzed with oligonucleotides containing the synthesized C8 bulky adducts described in section 5 and 6. Primer extension experiments as well as co-crystallization with Pol η were performed in order to give insights into the blockage/bypass mechanism of these adducts by Y-family polymerases.

7.1 Translesion synthesis of bulky DNA adducts by Pol η

7.1.1 Purification of yeast Pol η

The *S. cerevisiae Rad30* gene encodes the 632 residues of full length Pol η . A plasmid (pDEST007-pol η_{1-513}) containing the *Rad30* gene coding for a shortened, but still active version of 513 amino acids, truncated at the C-terminus was cloned and transformed into *E. coli* Rosetta gami by Dr. Claudia Chiocchini.^[90] The missing C terminal residues of the yeast Pol η catalytic core are of regulatory use, but do not affect the catalytic properties of the DNA polymerase and its translesion synthesis properties. The expression and purification protocol was also established by Dr. Claudia Chiocchini.^[90] Pol η_{1-513} , containing an N-terminal Strep-tag, was purified on two columns. First, a Strep-Tactin column was employed. Finally, a Heparin ion exchange column was used to remove residual DNA (yield: 0.5 mg protein / L culture). In Figure 42, a SDS-gel of the purified Pol η as well as the UV chromatogram of the elution profile of the heparin column purification are displayed.

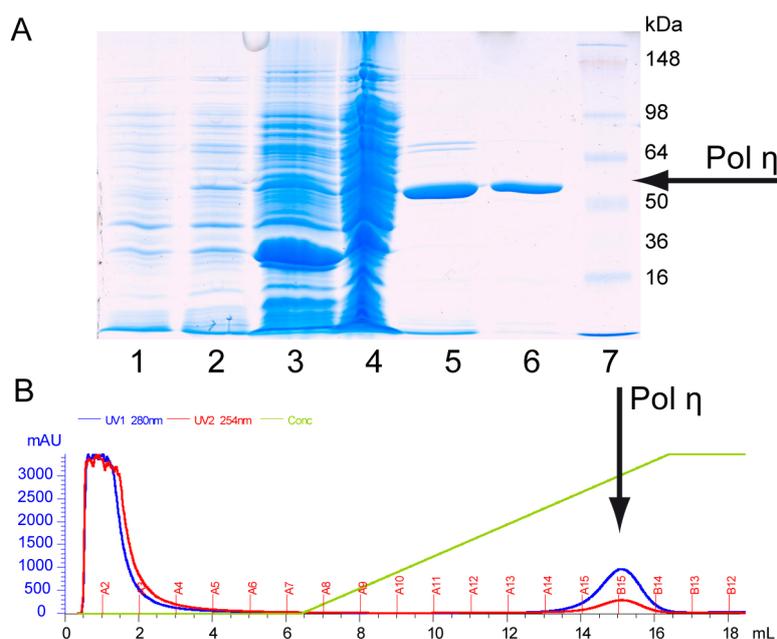


Figure 42. Expression and purification of yeast Pol η . A) SDS-PAGE: 1: Cells before induction; 2: Cells after 4 h induction; 3: Cell pellet after lysis; 4: Crude cell lysate; 5: Pol η after Strep-tag purification;

6: Pol η after heparin purification; 7: Protein marker. B) Chromatogram of the Heparin affinity purification. DNA and impurities are found in the flow-through. The red curve corresponds to the detection at 254 nm, the blue line shows the elution profile at 280 nm. The green line marks the NaCl-gradient for elution of the protein (Pol η size: 58.1 kDa + 1.25 kDa Strep-tag, 513 amino acids).

7.1.2 Mechanism of replication blocking and bypass of Pol η by AAF-dG adducts

The translesion synthesis DNA polymerase η is able to replicate through C8 bulky adduct lesions such as AF-dG and its acetylated analogue (AAF-dG). Bypass occurs mainly in an error-free manner, but at a reduced rate.^[56-58, 91] Replicative polymerases are in contrast fully blocked by the acetylated lesion **14** (see section 3.3), probably due to its restriction to the *syn*-conformation (Figure 43).^[52-54] It is not understood so far, how Pol η is able to incorporate a dCTP opposite the lesion in order to bypass the AAF-dG lesion correctly despite the *syn*-conformation of the lesion. Co-crystallization of yeast Pol η in complex with DNA containing the two bulky adducts AAF-dG and AAA-dG as well as *in vitro* primer extension studies were performed in order to give insights into the bypass mechanism of Pol η .

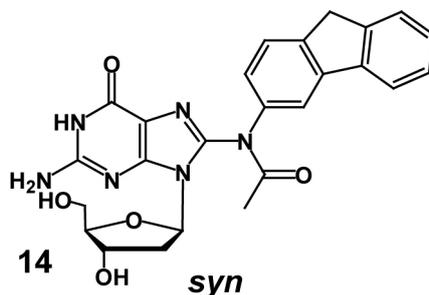


Figure 43. *Syn*-conformation of the 2-acetylaminofluorene-dG lesion (AAF-dG)

7.1.2.1 *In vitro* bypass of AAF-dG and AAA-dG lesions by Pol η

20mer DNA strands containing either the AAF-dG (ODN **2**) or the AAA-dG (ODN **28**) lesion at position 14 and, as a control, the undamaged DNA strand with identical sequence were annealed to shorter 5' fluorescence-labeled DNA counter strands (13mer primers). The DNA constructs were incubated with the purified yeast Pol η at concentrations ranging from 100-300 nM in the presence of one or all four nucleotide triphosphates and the primer extension reactions were analyzed by gel electrophoresis (Figure 44).

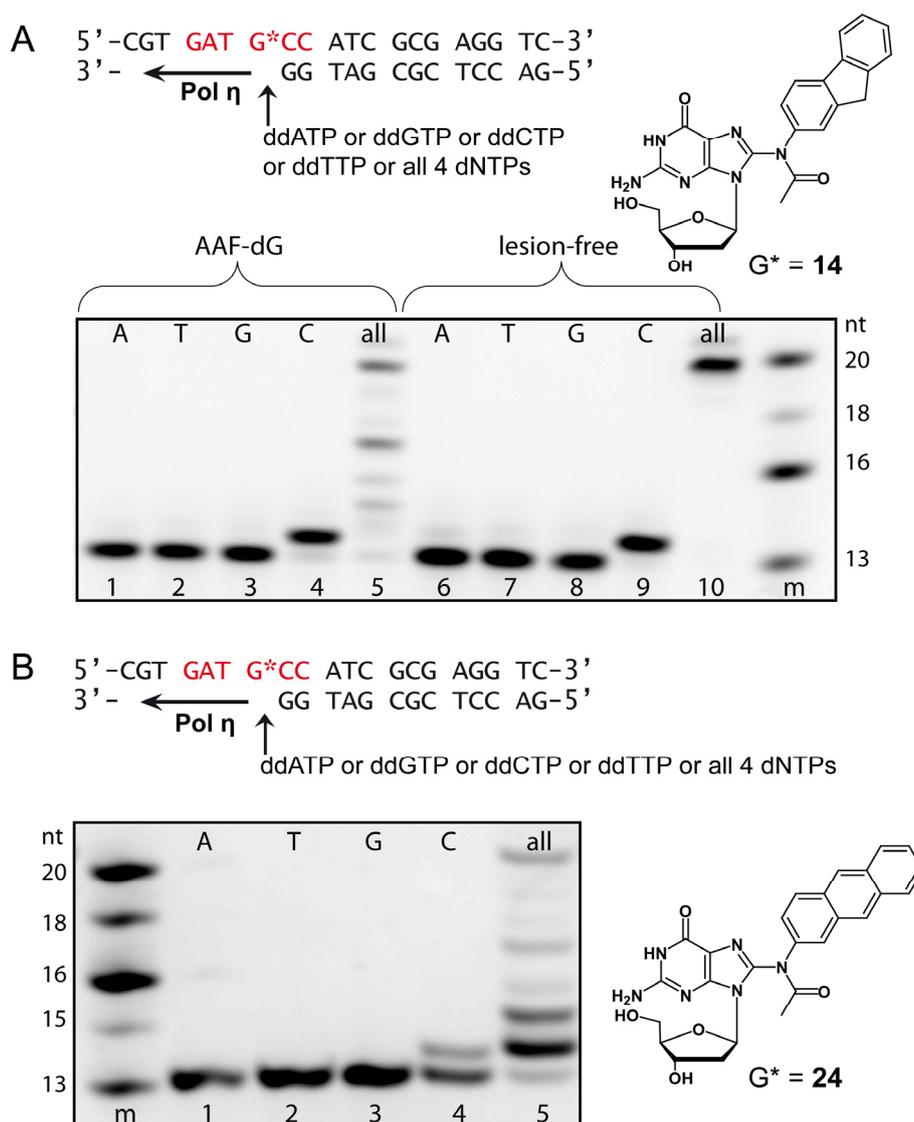


Figure 44. Translesion synthesis through different bulky dG DNA adducts catalyzed by Pol η . The primer extension reactions were carried out with fluorescence labeled primers and the extension products were resolved on a denaturing PAA-gel. A) Incorporation of individual nucleotides and of all four nucleotides with AAF-dG containing oligonucleotides and lesion-free oligonucleotides (200 nM Pol η , 1 μ M template DNA, 200 μ M of each dideoxynucleotide or all four dNTPs; 1, 6 = ddATP, 2, 7 = ddTTP, 3, 8 = ddGTP, 4, 9 = ddCTP, 5, 10 = all four dNTPs). B) Incorporation of individual nucleotides and all four nucleotides with AAA-dG containing oligonucleotides (1 = ddATP, 2 = ddTTP, 3 = ddGTP, 4 = ddCTP, 5 = all four dNTPs).

The enzyme fully extends undamaged template DNA (Figure 44 A, lane 10). In the presence of AAF-dG and AAA-dG, bypass is indeed possible, although with strongly reduced efficiency (Figure 44 A, lane 5 and Figure 44 B, lane 5). In both cases fully extended primer strands are however clearly observed, showing that Pol η is able to read through both lesions. To determine the base pairing partner of the adducted guanine during bypass, the primer extension experiments were performed adding 2',3'-dideoxynucleotides individually to the

reaction mixture. As depicted in Figure 44 A and B (lanes 4), Pol η incorporates, correctly and with high specificity, a dC opposite the AAF-dG and AAA-dG lesions despite the presence of the aromatic unit attached to C8. No incorporation of other nucleotides is observed opposite both lesions.

7.1.2.2 Structures of AAF and AAA-dG containing oligonucleotides in complex with Pol η

To elucidate the mechanism of the observed faithful replication across the AAF/AAA-dG adducts by Pol η , primer-template constructs of AAF-dG and AAA-dG were crystallized with Pol η from *S. cerevisiae* adding dCTP. Nucleotide insertion through the enzyme was prevented by using a primer strand ending with a dideoxy nucleotide, which pairs with the nucleobase located directly 3' of the lesion (for details, see section 10.5). The following primer-template construct was used:



Ditetrahedral-bipyramidal shaped crystals were obtained of both lesion-polymerase complexes and diffracted X-rays to about 2.7Å spacing (Figure 45).



Figure 45. Crystals obtained with AAF-dG (A) and AAA-dG (B) containing oligonucleotides in complex with Pol η from *S. cerevisiae* (residues 1-513).

The structures were solved by molecular replacement. For details on data collection and structure refinement see section 10.6 and appendix 11.1. As previously observed for the enzyme, there are two molecules in the asymmetric unit cell of both crystals^[90, 156, 157] showing the enzyme-DNA complexes in two different states (A and B, Figure 47). An overall structure of the protein and the DNA duplex in complex A is depicted in Figure 46. In the crystal structure, the AAF and AAA moieties of the AAF-dG lesion are well defined in both complexes in the electron density (Figure 47).

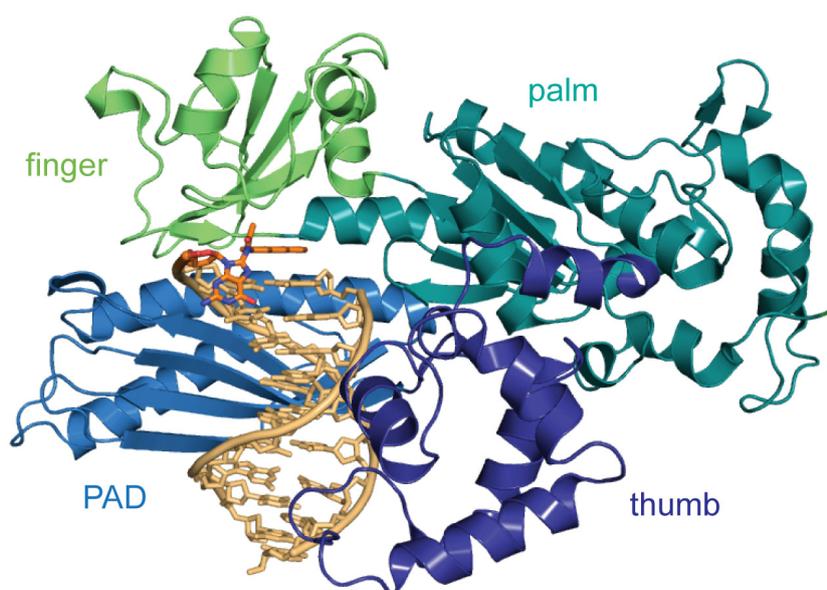


Figure 46. Crystal structure of *S. cerevisiae* Pol η with the AAF-dG oligonucleotide in complex A. The overall folding topology is depicted with the enzyme as cartoon representation, the DNA as stick model (gold) and the AAF-dG adduct in orange. The different domains typical for Y-family DNA polymerases are highlighted by special colors.

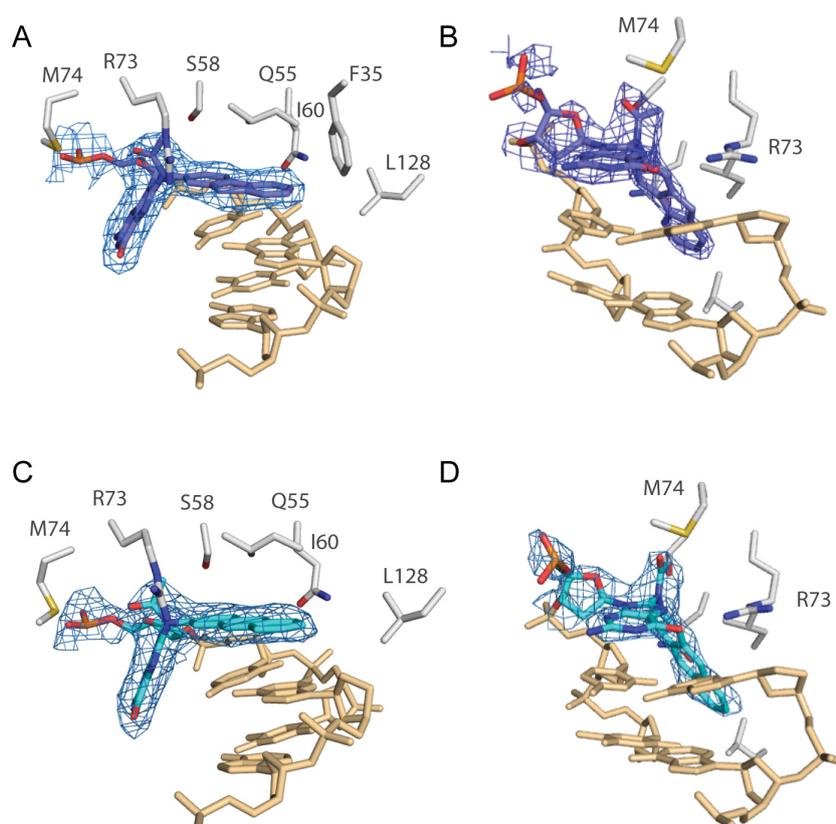


Figure 47. Difference-omit electron density map ($F_{\text{obs}} - DF_{\text{calc}}$) of the AAF-dG (A, B) and the AAA-dG (C, D) lesion with the surrounding protein residues. Figures A and C show the AAF-dG and the AAA-dG, respectively in complex A, Figures B and D both lesions in complex B. The aromatic C8-adduct of the fluorene as well as the anthracene are found at structural equivalent positions in respect to the active site of the protein.

In the two complexes A, the aromatic amine units stack on top of the Watson-Crick base pair formed by the primer ddC and the -1 dG. For both lesions, the bulky aromatic units therefore fully block the active site for the incoming dCTP (Figure 47 A, C and Figure 48 A). In complex B the situation is surprisingly different (Figure 47 B, D and Figure 48 B). First of all, a relatively small movement of the DNA template strand compared to complex A is observed. In both structures, the aromatic C8-adduct of the fluorene as well as the anthracene are found at structural equivalent positions in respect to the active site of the protein (Figure 47). However, the primer DNA strand is rotated in complex B relative to the aromatic moiety by about 6 Å. Due to this strand rotation, the location of the primer end is strongly different in both complexes A and B. This primer rotation brings the dG part of both lesions despite their prevalent *syn*-conformation closer into the active site, destacking the aromatic unit. In consequence of the primer motion, the dG-lesion packs above the Watson-Crick base pair formed by the primer end and the template strand. The primer strand displacement results in a partial opening of the active site, for the putative dCTP to bind (Figure 47 B, D and Figure 48 B). However, no bound dCTP could be observed in the crystal structures.

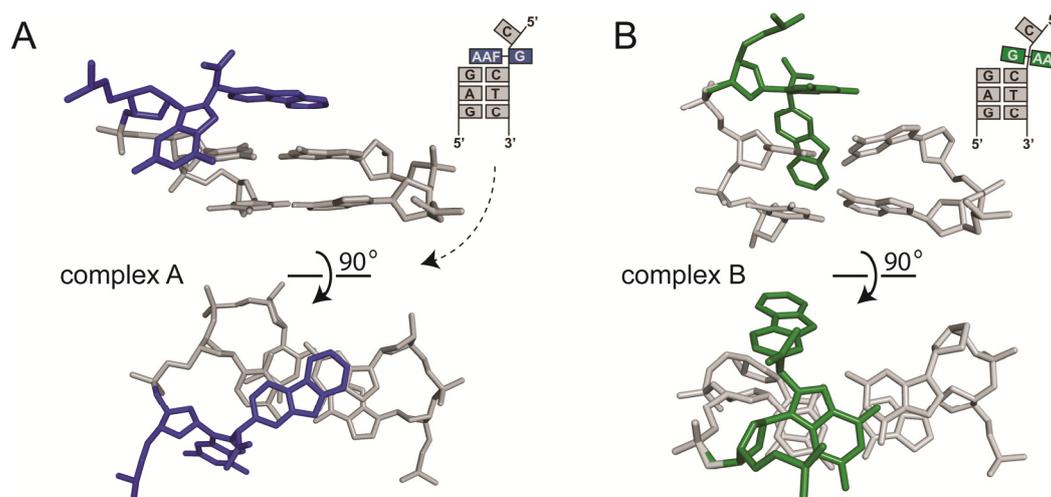


Figure 48. Stick model of the DNA in both AAF-complexes A and B containing the AAF-dG lesion A) In complex A, the fluorene moiety of the AAF-dG adduct stacks on the previous Watson-Crick base pair, blocking the active site for the incoming nucleotide. B) In contrast, the DNA in complex B is rotated in respect to the aromatic unit, with the adducted dG moved towards templating position, partially freeing the active site. Whereas in complex A the AAF and AAA unit stack on top of the base pair, this stacking interaction is fully sacrificed in complex B.

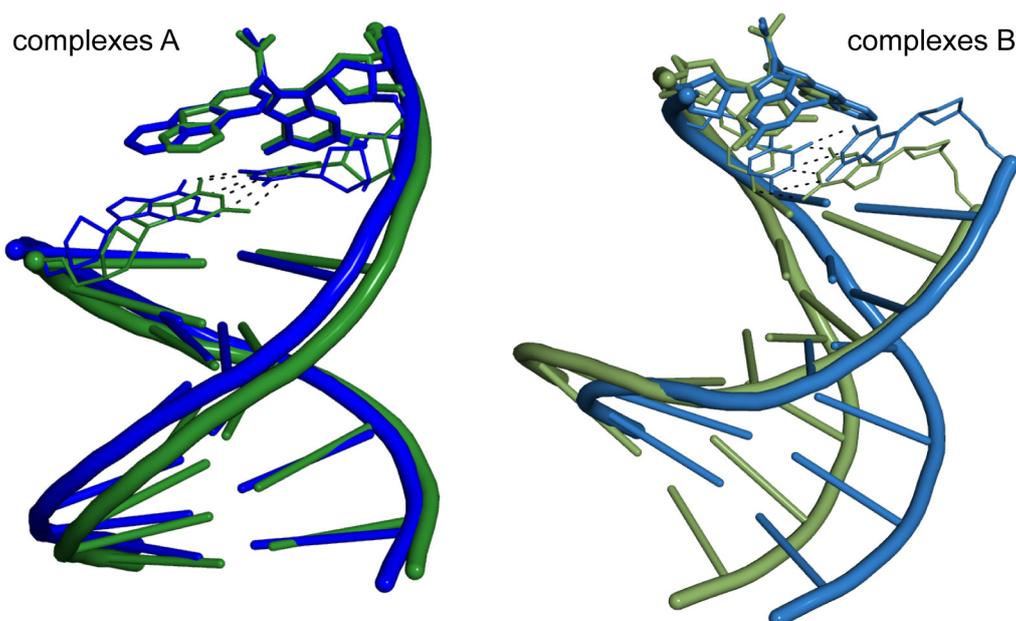


Figure 49. Overlay of the AAF-dG and AAA-dG containing DNA duplexes in complex A and B. The AAF-dG strands are depicted in green (complex A) and light green (complex B). The AAA-dG are shown in blue (complex A) and light blue (complex B).

In summary, a small primer movement is detected in going from complex A to complex B, which opens up the active site for the incoming dCTP. A rather small further movement of the DNA along this trajectory would result in a situation in which the triphosphate can finally bind and nucleotide incorporation can occur.

7.1.2.3 Model of the bulky adduct bypass

In order to visualize how Pol η copies through the AAF/AAA-dG lesions, the complex was modeled with the bound dCTP (Figure 51). To this end, the primer was shortened by one base and the dC-triphosphate was inserted. In all so far obtained crystal structure of Pol η with a bound triphosphate, the position of the triphosphate is invariant (Figure 50).^[88, 90, 157] Thus, the dCTP was placed at exactly the same position relatively to the protein where it is positioned in the known Pol η structures. Figure 50 shows a structural superposition of known yeast Pol η structures, the close-up view of the active site with the incoming dCTP is depicted in Figure 50 C.

Minor changes of the overall protein conformation were discovered, when the presented bulky adduct structures and structures of Pol η in complex with the cisplatin adducts^[90, 157] were compared with a recent structure of Pol η reading through a CPD lesion (Figure 50 A, B).^[88] The crystallographic results of this thesis work as well as of the cisplatin lesion structures were obtained with the wild type Pol η enzyme, which crystallizes with an open PAD domain.

The CPD structure however was obtained with a mutated Pol η . Here, particular crystal contacts were removed by site directed mutagenesis resulting in a further closing of the PAD domain. This however does not change the situation in the active site, particularly not of the base pairing motifs and the positioning of the incoming dCTP (Figure 50 C).

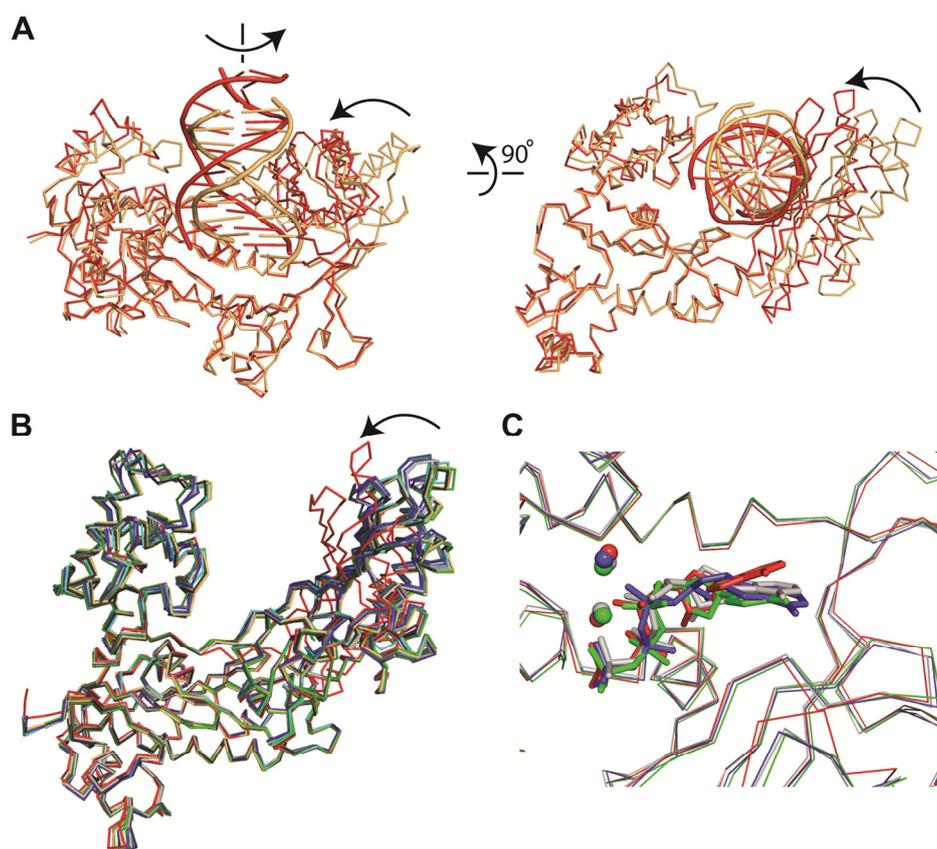


Figure 50. Structural superposition of known yeast Pol η structures. The proteins were least-square fitted using the main-chain atoms of residues number 1-230, which were chosen as reference frame. A) Closing of the PAD domain and slight rotation of the DNA helix in the Pol η structure in complex with the CPD lesion (PDB code 3MFI, red) compared to the AAF (this thesis work, golden). In the 90° rotated view the finger domain (residues 1-122) was removed for clarity. The polypeptide chain is shown as ribbon and the DNA as cartoon representation. B) Superposition of all known yeast Pol η - DNA complex structures: the cisplatin GG adduct (PDB codes 2R8J and 2R8K, green and blue); the cisplatin-GTG adduct (PDB code 2WTF, grey); and CPD lesion (PDB code 3MHI, red), AAF (PDB code 2XGP, gold). C) Close-up in the active site with the same color scheme as in B. The respect triphosphates are drawn as stick model and the ions as spheres. For clarification, the DNA is not shown. The superposed structures are Pol η bound to lesion containing DNA: the cisplatin GG adduct (PDB codes 2R8J and 2R8K, green and blue); the cisplatin-GTG adduct (PDB code 2WTF, grey); and CPD lesion (PDB code 3MHI, red). In all structures with a bound triphosphate, the ions in the active site are in structural similar positions, regardless if there are catalytic competent magnesium (Pol η - CPD) or calcium ions (all other Pol η - DNA complex structures). Despite closing of the PAD domain, no structural changes occur in the active site.

After shortening of the primer by one base, the triphosphate fits perfectly into the active site. The AAF-dG lesion was then moved into the -I position. In this way, Pol η moving one base pair forward could be simulated (Figure 51 C). After these manipulations, the mechanism of the difficult bypass step can be visualized.

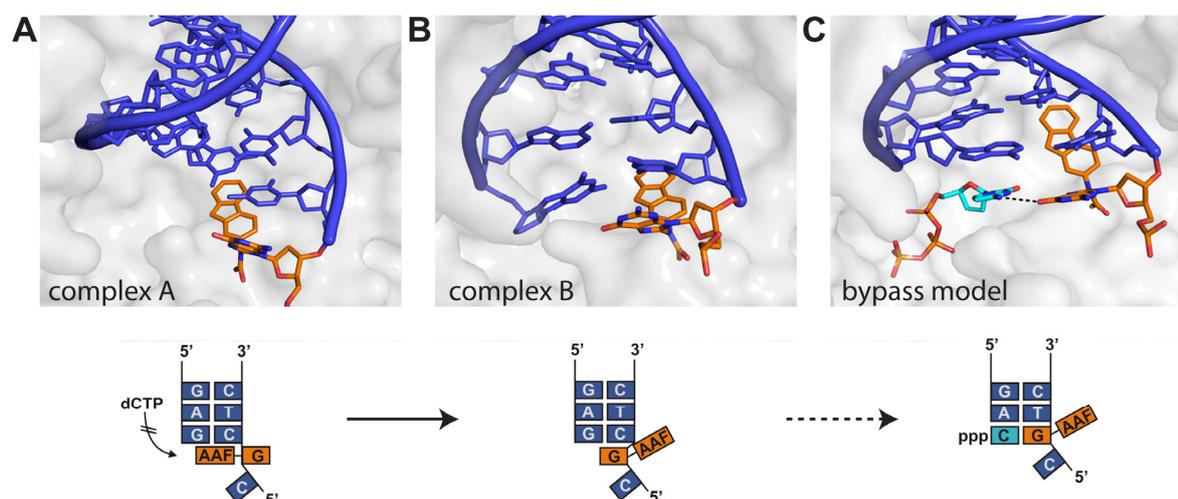


Figure 51. Model for AAF-dG bypass by Pol η . A) In complex A, the fluorene moiety of the AAF-dG adduct stacks on the previous Watson-Crick base pair, blocking the active site for the incoming nucleotide. B) In contrast, the DNA in complex B is rotated in respect to the aromatic unit, with the adducted dG moved towards the templating position, partially freeing the active site. C) Hypothetical model of the AAF bypass by Pol η . The dG : dC basepair 3' of the lesion in the AAF-dG-Pol η complex B was replaced by the AAF-dG adduct and the dCTP triphosphate. The dCTP was placed in the active site according to the structure of Pol η in complex with the cisplatin GG lesion (PDB code 2R8J). Thus the active site for the incoming dCTP opens, which forms one critical hydrogen bond to the adducted nucleobase in *syn* conformation.

In the proposed model, the extension of the template rotation as described above is sufficient to fully open the active site for the incoming dCTP. The dCTP placed in the active is able to form one hydrogen bond with the dG of the lesion. Thus, the crystal structure (Figure 51 A and B) in combination with the model (Figure 51 C) allows to decipher the basis for faithful bypass of acetylated bulky adduct lesions. Here, the acetylated bulky adduct lesions do not have to rotate into the *anti*-conformation to enable Watson-Crick base pairing. Instead, the DNA primer strand rotates, which allows Pol η to decode the bulky dG adduct with dCTP. One critical hydrogen bond is established between the N^4 of the triphosphate and the O^6 of the dG part of the lesion (Figure 51 C and 52). A similar single H-bond based decoding was also observed to allow Pol η to bypass the cisplatin GG lesion.^[90]

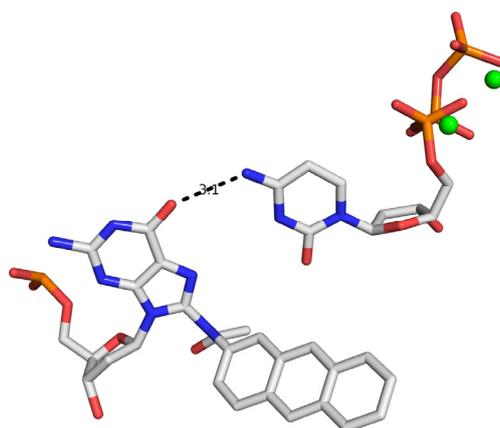


Figure 52. Close-up view of the bypass model for AAF/AAA-dG lesions by Pol η . A critical hydrogen bond is established between the N^4 of the dCTP and the O^6 of the dG part of the AAA-dG lesion.

7.1.2.4 Further biochemical investigation of the bypass model

The bypass hypothesis proposes, that Pol η keeps the bulky adduct of the lesion at the same position, but rotates the primer strand around the lesion in order to bring the dG of the lesion into a templating position. The major energetic penalty of this rotational process is the required destacking of the aromatic bulky adduct unit and the last formed Watson-Crick base pair between the primer and the template strand. In complex A, the fluorene and anthracene units stack almost perfectly on top of the base pair (Figure 47, A and C). This stacking interaction is fully sacrificed in complex B (Figure 47, B and D). Biochemical support for the mechanistic proposal could therefore be obtained by measuring the translesion synthesis process depending on the size of the aromatic unit, which determines the amount of the stacking interaction. The required destacking forces should increase with increasing size of the flat aromatic adduct, thus slowing down the bypass process. Thus, primer extensions were also performed with the small benzene-, the larger naphthyl- or the largest pyrene-dG adduct (see sections 5 and 6). The results are depicted in Figure 53 and Figure 54 (lanes 1-5). The experiment shows that TLS efficiency is strongly reduced if the size of the aromatic adduct is increased. Whereas bypass is surprisingly efficient for the small benzene unit, bypass is almost impossible when the benzene was replaced by a large pyrene unit.

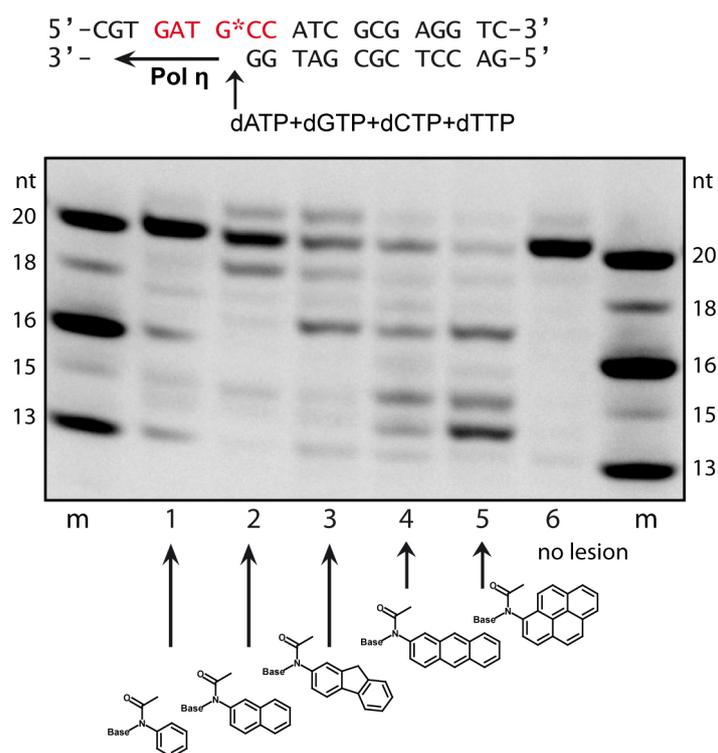


Figure 53. Translesion synthesis of different bulky dG DNA adducts in the same sequence context catalyzed by Pol η. The primer extension reactions were carried out with fluorescence labelled primers and extension products resolved on a denaturing PAA-gel. Reaction conditions are: 300 nM Pol η, 1 μM template DNA, 200 μM dNTPs each, 1 = AAB-dG, 2 = AAN-dG, 3 = AAF-dG, 4 = AAA-dG, 5 = AAP-dG, 6 = undamaged template strand, m = marker.

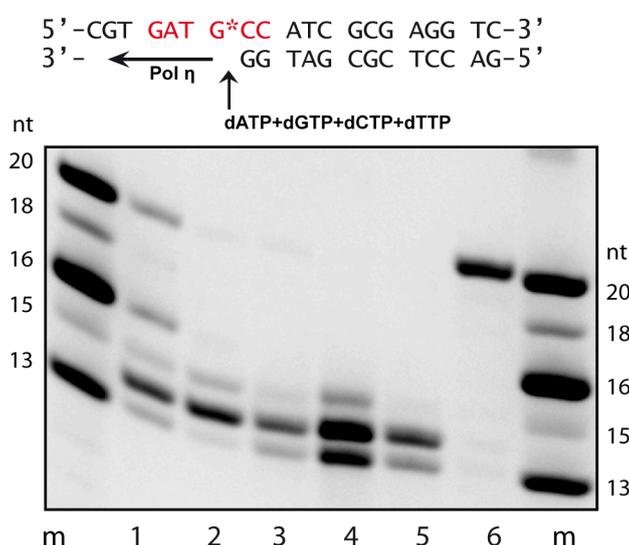


Figure 54. Primer extension studies with different bulky dG DNA adducts and lower Pol η concentrations (150 nM) compared to the experiments presented in Figure 53. 1 = AAB-dG, 2 = AAN-dG, 3 = AAF-dG, 4 = AAA-dG, 5 = AAP-dG, 6 = undamaged template strand, m = marker.

To further support the proposed TLS mechanism, additional experiments with zebularine triphosphate (dZTP) were performed (Figure 55). Since the bypass process was found to rely on the formation of one hydrogen bond between the incoming dCTP and the dG part of the

lesion, lesion bypass should be reduced with dZTP where this H-bond cannot form. In agreement with the postulated mechanism, dZTP indeed cannot be incorporated by Pol η opposite the lesion (Figure 55, lane 5).

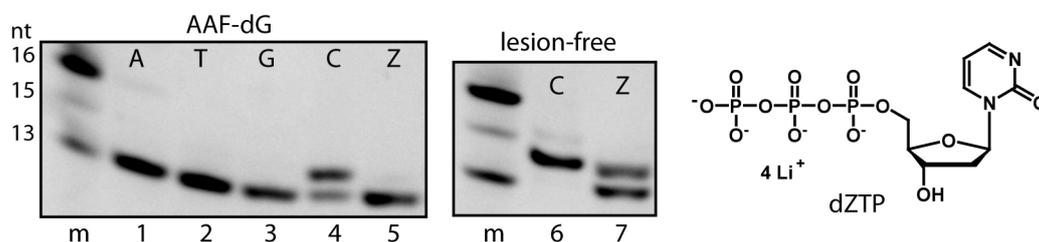


Figure 55. Left: Primer extension reaction with 2'-Deoxyzebarine-5'-Triphosphate (dZTP) and AAF-dG containing oligonucleotides (100 nM Pol η , 1 μ M template DNA, 200 μ M of each deoxynucleotide of all four dNTPs and dZTP; 1 = dATP, 2 = dTTP, 3 = dGTP, 4 = dCTP, 5 = dZTP); Right: Primer extension reaction with undamaged template DNA, 6 = dCTP, 7 = dZTP (200 nM Pol η)

In a control experiment with undamaged template DNA, dZTP incorporation opposite dG is observed, albeit with reduced efficiency compared to a dCTP (Figure 55, lane 7).

7.1.2.5 Discussion

The acetylated C8-AAF-dG and the non-acetylated C8-AF-dG adducts have distinct mutagenic properties. The AF-dG lesion strongly reduces the efficiency of replicative bypass, but read through by high fidelity polymerases is accomplished because these lesions can rotate around the glycosidic bond into the *anti*-conformation.^[51, 120] Thus, the AF-dG unit can correctly pair with an incoming dCTP in the active site by forming a Watson-Crick base pair.^[158, 159] The acetylated bulky adducts in contrast, are not bypassed by high fidelity polymerases because these lesions are fixed in the *syn*-conformation.^[52-54] Faithful bypass of the acetylated lesions *in vivo* is achieved by low-fidelity TLS polymerases, with Pol η playing a major role.^[56, 57, 60] Two crystal structures of DNA containing the two different acetylated bulky adduct lesions AAF-dG and AAA-dG in complex with Pol η were obtained in this thesis work. Both structures show the primer-template constructs in two different positions (complex A: Figure 56 A and B, complex B: Figure 56, C and D). Complex A features the bulky adduct in the active site, blocking the dCTP from entering. In this complex, the bulky adduct fully stacks on top of the last primer-template base pair. Complex B in contrast shows the bulky adduct in the process of being bypassed.

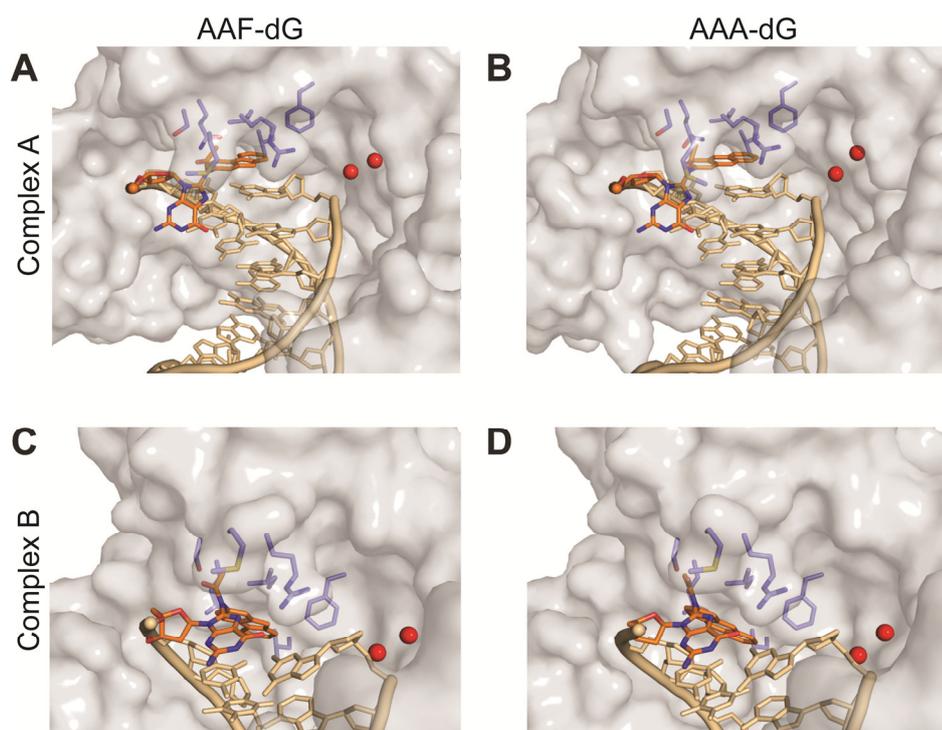


Figure 56. Schematic representation of Pol η binding AAF-dG and AAA-dG containing DNA. In the complexes A, the aromatic moieties of AAF-dG (A) and AAA-dG (B) stack on top of the previous G-C base pair, adopting structural equivalent positions. In the complexes B, the enzyme rotates the DNA in respect to the fluorene moiety in order to free its active site for an incoming dCTP (AAF-dG (C) and AAA-dG (D)). The protein is shown as semi-transparent surface representation (grey) with the DNA as stick model (gold) and the lesion highlighted in orange. Active site residues are also depicted as stick model (blue). The catalytic calcium ions in the active site are displayed as red spheres.

In complex B, the enzyme has rotated the DNA primer strand, while keeping the aromatic C8-adduct at a structural equivalent position. This rotational motion partly unblocks the active site but a dCTP can still not enter, which explains why bypass of these lesions is so difficult even for Pol η . A modeling study, in which further movement along this trajectory was performed, shows how the active site might open to enable the nucleotide triphosphate binding in the active site (Figure 51 C). The model shows that a bound dCTP might be able to form one H-bond between the C4-NH₂ group of the dCTP and the C6=O group of the AAF/AAA-dG. The starting point of the model is the fact that the triphosphate occupies in all known structures exactly the same position (Figure 50 C), which allows to simulate the dCTP bound structure when the primer and the template have moved one step further on the elongation pathway. The formed critical H-bond could provide the basis for Pol η to correctly insert a dC opposite the lesion. Purine triphosphates can not bind due to their larger size. The alternative pyrimidine base dTTP would fit sterically and can indeed be modeled into the active site, but dTTP would be unable to form the critical H-bond. The importance of this H-bond was proven with the artificial triphosphate dZTP lacking the C4-amino group. The

dZTP triphosphate is not inserted opposite the lesion, thus supporting the proposed bypass model.

Further support comes from a chemical biology study, in which the size of the aromatic adduct was systematically changed, thereby modulating the stacking interactions between the lesion and the last primer template base pair. The experiments show that smaller sized adducts which establish smaller π -stacking forces are more efficiently bypassed, even though the energy potential for the *syn-anti* rotation around the glycosidic bond is unchanged.^[55] Bypass of the large pyrene unit is in consequence hindered because of the large size of the stacking interaction.^[160, 161] In summary, efficient destacking of the aromatic unit on top of the primer template base pair is a key prerequisite for the efficient bypass reaction by the low fidelity polymerase Pol η .

7.1.3 Crystallization experiments with other bulky adducts

Bypass of the large pyrene adduct is almost impossible for Pol η . The smaller benzene and naphthalene adducts are replicated more easily. Crystal structures of oligonucleotides containing these adducts in complex with Pol η could give further insight into the bypass process, especially in the blockage by the AAP-dG adduct. Thus, additional crystallization experiments were performed with AAN-dG, AAB-dG and AAP-dG containing oligonucleotides in complex with Pol η using the same DNA sequence as before (section 7.1.2.2). For the AAN-dG and AAB-dG adducts, only crystals of the apo-enzyme exhibiting the typical, rhombic shape were obtained (Figure 57 A). AAP-dG crystals had a similar morphology as the AAF-dG crystals (Figure 57 B).

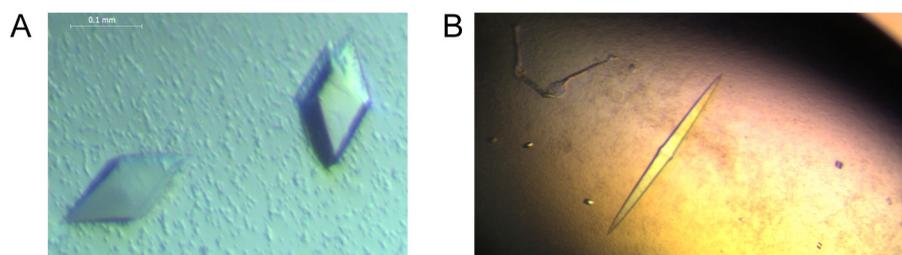


Figure 57. Crystals obtained with bulky adduct containing oligonucleotides in complex with Pol η from *S. cerevisiae* (residues 1-513). A) The crystals obtained with addition of AAB-dG DNA (ODN **23**) contained no DNA. B) AAP-dG (ODN **32**) crystals diffracted X-rays to 2.7 Å spacing and contained DNA.

The AAP-dG crystals diffracted X-rays to about 2.7 Å spacing and the structure was provisionally solved by molecular replacement. Again, two Pol η molecules were found in the asymmetric unit cell of the crystals. The protein is well defined in electron density. Unfortunately, the lesion is flexible and does not enter the active site of the enzyme, thus the AAP-dG is not defined in the electron density. The observation that the AAP-dG is unable to access the polymerases active site in the crystal may also explain why bypass of the lesion is so difficult for the enzyme. Figure 58 shows a close-up view of the active site of the enzyme. The C:G base pair 3' to the lesion can still form but is slightly out of plane. There is no electron density found for either the guanine, the deoxyribose or the pyrene unit.

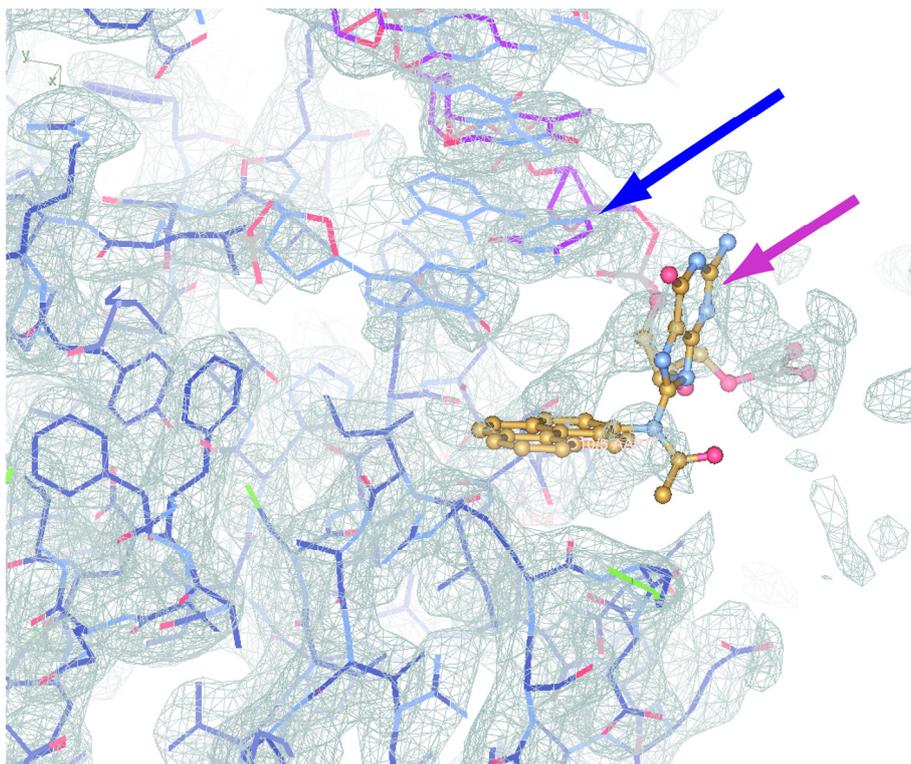


Figure 58. Provisional crystal structure of AAP-dG containing DNA in complex with Pol η . The modeled position of the AAP-dG (golden stick model) is marked by a pink arrow. The G:C base pair 3' to the lesion is highlighted by a blue arrow.

7.1.4 Mechanism of acetylaminofluorene-dG induced frameshifting

7.1.4.1 Primer extension studies in combination with mass spectrometry measurements give insights into the occurrence of frameshift mutations

Frameshift mutations induce a shift of the genetic reading frame, which can be fatal for the cell. If the shift involves one or two bases, the corresponding frameshift mutations cause the loss of the information content of the whole gene. The mechanisms that lead to frameshift mutations are not well understood but, as discussed in section 3.3, frameshifts are produced in consequence of the mutagenic effect of aromatic amines. Bypass of the acetylated AAF-dG lesion is often accompanied by a frameshift event. If the lesion is located in special repetitive sequences such as the *NarI* sequence, it induces a –I or –II frameshift mutation.^[16, 53, 61-64] This process involves a primer-template misalignment by one or two bases during replication and it is believed that the misaligned primer-template duplex is stabilized by the AAF (see section 3.3).^[21, 62, 65-68] In order to study the mechanisms of the various frameshift processes several oligonucleotides containing the AAF-dG lesion in different sequence contexts were synthesized (**ODN 1, 2, 13-15 and 18-20**). Primer extension studies with Pol η were performed and the *NarI* base sequence was systematically changed to study which part of the sequence is needed to induce the frameshift reaction.

For the primer extension reaction, fluorescent labeled primers, ending one nucleotide 3' of the lesion or with a dC directly opposite the lesion, were annealed to the corresponding template strand. Primer extension was carried out using 300 nM Pol η and 200 μ M of each nucleoside triphosphate at 30°C for 5 min. The primer extension assay products were investigated by gel electrophoresis and also directly by MALDI-Tof mass spectrometry. This allows gaining information not only about the lengths of the synthesized primers but also about the base composition of the primer product. For the mass spectrometry analysis, the DNA of the primer extension reactions containing 30 pmol of the primer/template construct was extracted by phenol/chloroform extraction. The aqueous phase was concentrated and desalted using a ZipTip_{C18} (Millipore).

Figure 59 contains the complete gel of the primer extension reaction performed with the various AAF-dG containing oligonucleotides.

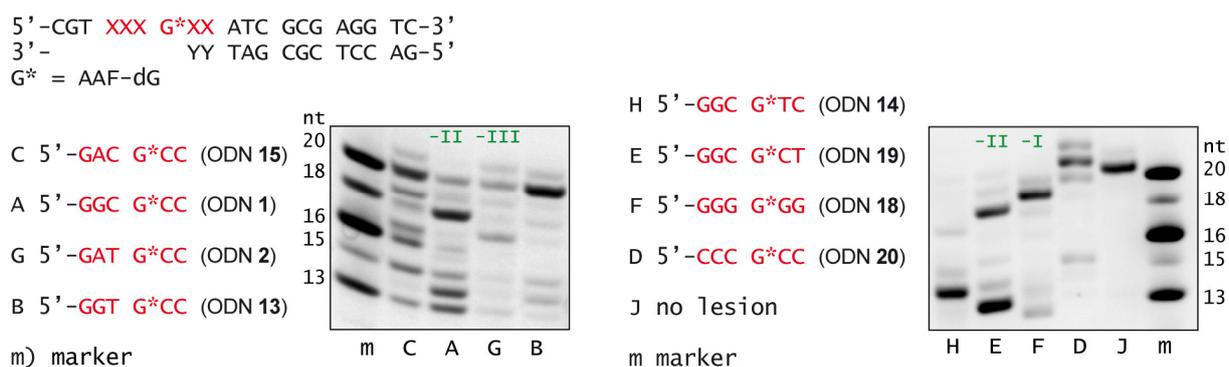


Figure 59. Primer extension reaction catalyzed by Pol η with the AAF-dG in various sequences and complementary primers ending 3' before the AAF-dG lesion at 5 min incubation time. Reaction conditions are 300 nM Pol η , 1 μ M template DNA, 200 μ M deoxynucleotides. Individual lanes of both gels are depicted in Figure 59.

The results of the study are compiled in the Figure 60. For this, single lanes from the gel were individually discussed. If the AAF-dG lesion is placed in the *NarI* sequence (Figure 60, A) formation of only a small amount of the fully elongated primer is observed. Most of the product is a primer, which is two bases shorter in agreement with data from literature.^[61] The MALDI-ToF mass spectrum provided for the truncated primer an m/z -value of 6021, which is in agreement with a -II-frameshift product, in which the AAF-dG lesion itself and one of the cytosines next to the lesion are not correctly translated (see appendix, section 11.2). If one changes the base sequence 5' to the AAF-dG lesion (Figure 60, B, C and D) a normal full length primer product is formed comparable to what is obtained in the absence of any lesion (Figure 60, J). Since in sequence B, the 5'-CG^{AAF}-3' is replaced by 5'-TG^{AAF}-3', misalignment of the primer end would involve formation of a 5'-TG^{AAF}-3' bulge. This is apparently not occurring, instead full extension is observed. These results suggest, that it is specifically the dC base located 3' to the AAF-dG lesion, which is responsible for the frameshifting process. In all these primer extension reactions we used a primer that ended directly in front of the AAF-dG lesion was used. The dC base in position -1, located 3' to the lesion, is consequently the last template base that is correctly base paired with a dG base in the primer. The obtained primer extension results allow therefore postulating directly a mechanism for the -II-frameshift process which is shown in Figure 61, A. Pol η correctly inserts a dC into the primer opposite the AAF-dG lesion despite the *syn*-conformation of the lesion. Then, however, primer slippage occurs, which is accompanied by an extrusion of the 3'dC and the AAF-dG lesion. The primer slips over this dinucleotide sequence and misaligns with the -CG- sequence present in the template in 5' direction relative to the AAF-dG lesion. This misaligned primer is subsequently elongated, which manifests the -II frameshift product.

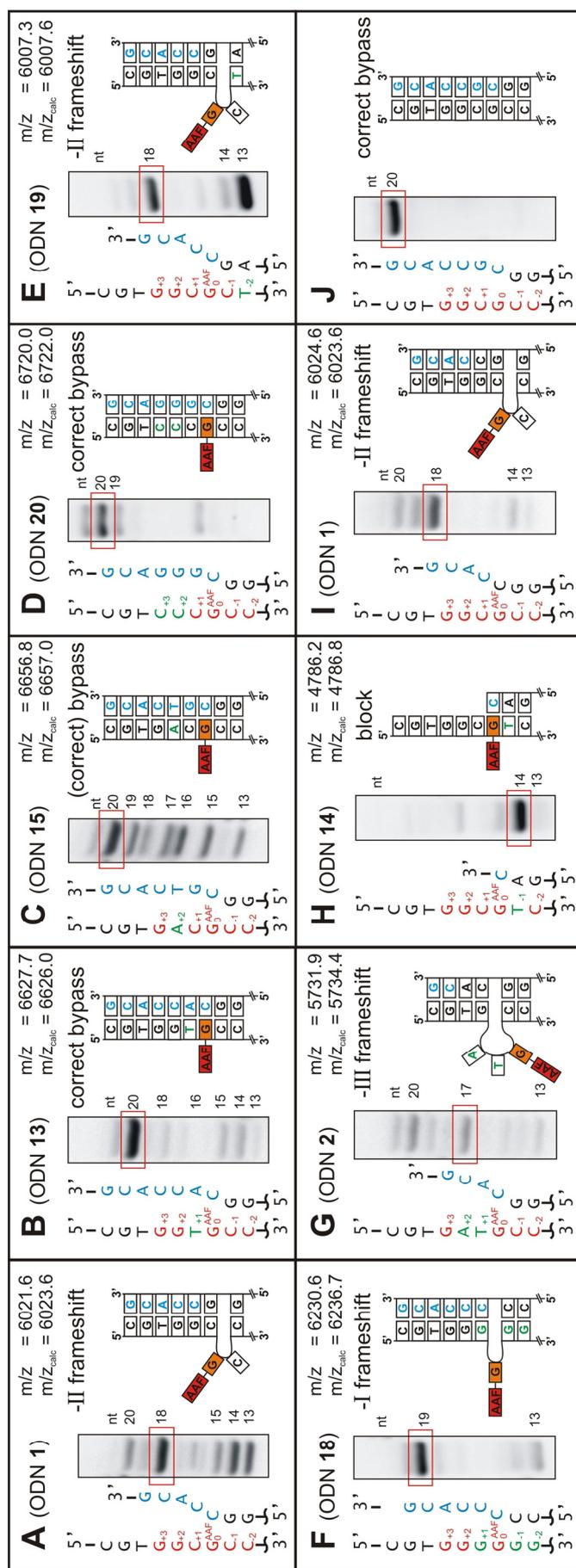


Figure 60. Schematic overview of the sequence dependent bypass products occurring during primer extension of AAF-dG adducts. For the different sequences A-I, the corresponding gel lane, a scheme of the bypass process and the mass values are depicted. Initially, a dCTP is incorporated opposite the lesion in all cases. Bypass of the *NarI* sequence (A) and a variation at -2 position (E) lead to a -II frameshift event. B, C and D lead to correct bypass since no slippage is possible. F and G show other frameshift products (-I and -III) depending on the sequence around the lesion. H shows an experiment in which the polymerase is blocked.



Figure 61. A) Schematic description of the slippage process leading to a –II frameshift, B) to a –III frameshift and C) to a –I frameshift.

Thus, a dC-base in 3' direction to the AAF-dG lesion and a CG sequence 5' to the AAF-dG lesion are needed for the frameshift process. This idea was examined with a template sequence in which we replaced the dC base at position -2 relative to the AAF-dG by a dT. The result is depicted in Figure 60, E. This dC to dT change did not affect formation of the –II frameshift showing that indeed only the dC in –I position is involved in the frameshift process. Even more informative is experiment F shown in Figure 60. Here all dC bases in close vicinity of the lesion were replaced by dG. This gives a tightly closed and stable duplex in which primer-template misalignment should be energetically costly. Now, slippage and misalignment is indeed limited to just one base. It is presumably the lesion itself, which is bulged out to provide the basis for a –I frameshift (Figure 61, C). Astonishing is the high yield by which this -I frameshift is formed. On the gel shown in Figure 60, F it is the only formed reaction product. MALDI-Tof analysis* shows the presence of only two oligonucleotides in the assay solution (Figure 62, A). Next to the primer strand with a $m/z = 4431$, the only other oligonucleotide detected has a value of $m/z = 6231$, which is the correct molecular weight for the –I frameshifted primer.

In a further experiment (Figure 60, G), the sequence 5' to the AAF-dG lesion was changed so that misalignment of the primer can only occur if in total three bases are extruded. Now, next to full length extended primer, an additional band is detectable on the gel, which is formed by an extended primer indeed three bases too short. Analysis of the primer extension reaction by MALDI-Tof mass spectrometry shows in fact formation of the full length primer with $m/z = 6639$ (Figure 62, B). All other signals, however, are derived from primers with feature the –III

* MALDI-Tof spectra of all primer extension reactions can be found in the appendix (section 11.2)

frameshift (Figure 62, B). Consequently, Pol η indeed first extends the primer with a dC base inserted opposite the lesion. After this, a substantial fraction of the primer misaligns to pair with the alternative dG in the template two bases further away (Figure 61, B). This large misalignment process, which is in full agreement with the formulated frameshift mechanism, is surprisingly efficient.

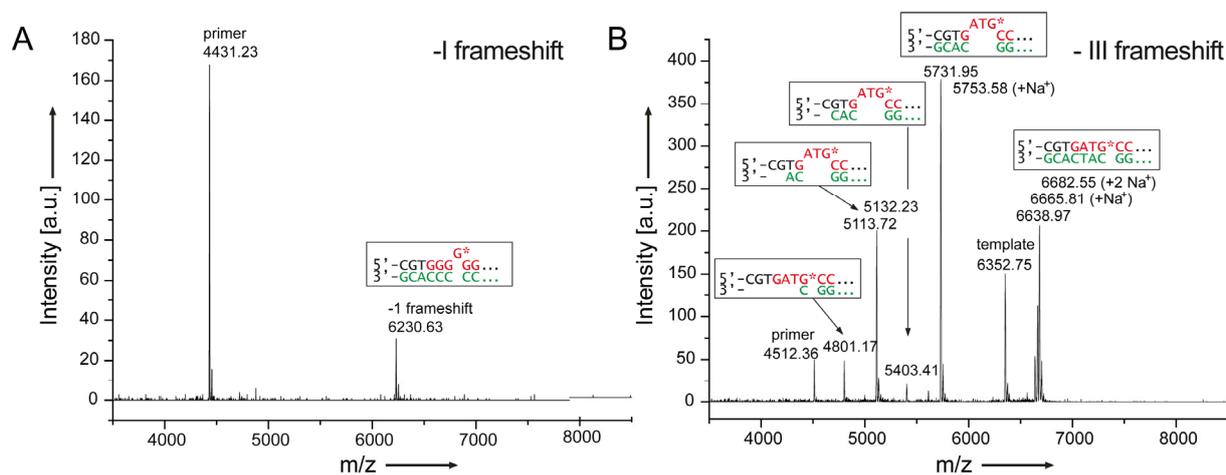


Figure 62. A) MALDI-ToF mass spectrum of experiment F. The –I frameshift is detected at $m/z_{\text{calc}} = 6236.7$. B) MALDI-ToF of experiment G. Next to full extension at $m/z_{\text{calc}} = 6639.6$, the –III frameshift is detected at $m/z_{\text{calc}} = 5734.7$ together with truncated –III frameshift products.

The results show that the template-primer stability in close proximity to the lesion is a key parameter which determines the kind of frameshift established. Formation of a –III frameshift is limited to sequences in which the AAF-dG lesion is located in weakly paired A:T sequences. If the lesion is embedded in G:C sequences, only a –I or –II frameshift is observed. This idea is further supported by experiment H in Figure 60. When the dC 3' to the lesion is replaced by a dT, slippage and misalignment is not possible any more. Now, the AAF-dG lesion establishes a replication block even for the low fidelity polymerase Pol η . It seems that for primer elongation to occur through the lesion, a stable primer-template duplex with a GC base pair below the lesion is required.

The experiments show that acetylated C8 bulky adduct lesions such as AAF-dG can, when they are bypassed by Pol η , induce –I, –II, and –III frameshift mutations depending on the base sequence around the lesion. The critical event is in all cases the insertion of a correct dC base opposite the bulky adduct. This seems to require a G:C base pair 3' to the lesion, possibly to ensure proper alignment of the primer opposite the lesion in order to enable attacking of the triphosphate. If the G:C base pair is replaced by an A:T base pair, replication past the lesion is not longer possible. In this case, the AAF-dG lesion establishes a replication block. After

insertion of the dC base opposite the lesion, the primer starts to probe the base sequence behind (5') the lesion. If a complementary sequence is available, the primer end will misalign, bulging the lesion itself and if necessary up to two additional bases out. In the majority of cases and in particular in the *NarI* sequence, the dinucleotide 3'- dC-AAF-dG -5' is bulged out. In G:C rich sequences, in which bulging out of bases is energetically more costly, the bulging is limited to a single base inducing a -I frameshift.

To further prove that correct dCTP incorporation opposite the lesion is indeed always preceding mutagenic extension slippage, a 14mer primer instead of the 13mer was employed, containing additionally the correct initial dC incorporated opposite the AAF-dG. This primer was annealed to the *NarI* sequence and extended under the same conditions. Again, the -II frameshift can be clearly detected (Figure 60, I and 63, A and E), thus confirming that the -II frameshift originates after correct 13mer primer extension with a dC, followed by slippage and realignment of the primer template construct. Moreover, using the proper 14mer primers in all the other primer extension reactions, the -I and the -III frameshifts as well as the replication blocking events could be reproduced, showing that correct pairing of the lesion with dC precedes all frameshifting events (Figure 63).

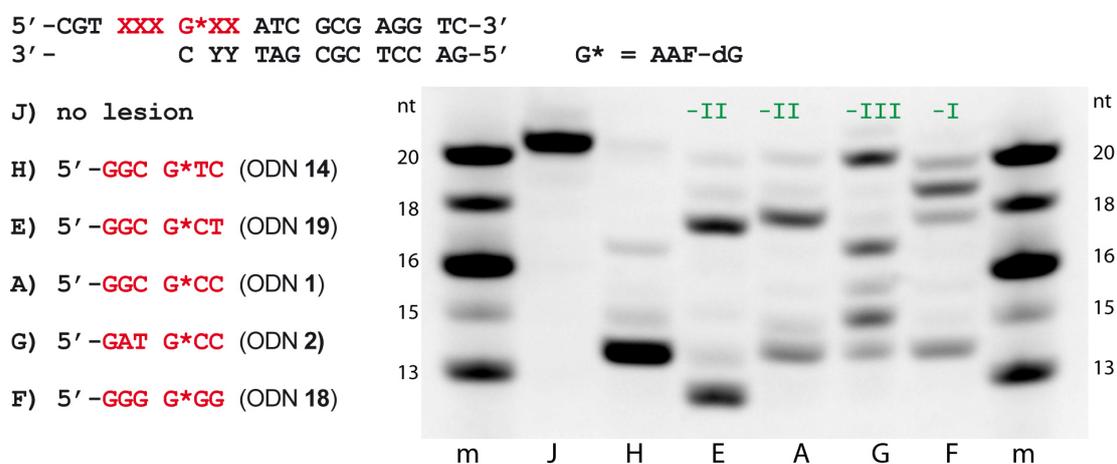


Figure 63. Primer extension reaction catalyzed by Pol η with the AAF-dG in various sequences and complementary primers ending with a dC opposite the AAF-dG lesion at 5 min incubation time. Reaction conditions are 300 nM Pol η , 1 μ M template DNA, 200 μ M deoxynucleotides. The -II frameshift is clearly visible (lanes E and A), in lane J, the correct extension of the 14mer primer annealed to an undamaged template is shown. Sequence H leads to a total block of the polymerase, no extension of the primer takes place. Sequence E shows a different mobility compared to the other 14mer primer sequences. Gel electrophoresis and mass spectrometry analysis confirmed the purity of the primer strand E. Formation of secondary structures might lead to a different mobility of this oligonucleotide. The experiment demonstrates that the various frameshift events (-I, -II, -III) and blockage of the polymerase (sequence H) also occur if an elongated primer with a dC opposite the

lesion is used. This supports the proposed model of the AAF-dG bypass in which dCTP incorporation opposite the AAF-dG precedes mutagenic or error-free extension from the lesion.

The general scheme that emerges from the study is that frameshift formation is triggered by a special instability of the AAF-dG:dC base pair. Once formed, extension of the dC-ending primer seems to be so difficult that it is easier for the enzyme to slip the primer end forward until the primer dC end misaligns with a complementary dG base upstream on the template strand. If the formed bulged out structure is stable enough, elongation occurs and the frameshift is manifested. The instability of the AAF-dG:dC base pair and not a particular stabilizing effect of the aromatic adduct seems to determine the frameshift process.

7.1.4.2 Towards crystallization of the -II frameshift

In order to get a better insight into the formation of the -II frameshift caused by AAF-dG in a repetitive sequence context, oligonucleotides were designed for the co-crystallization of the two (and one) nucleotide bulge with Pol η . The AAF-dG containing DNA primer/template constructs used for crystallization are summarized in Table 2.

ODN	primer/template construct	dNTP/ ddNTP	crystals	diffraction
ODN 8	5'-CTCTGGCG*CCTCACAC-3' 3'- ddCG - - GAGTGTG-5'	dCTP	small crystals	8 Å
ODN 8	5'-CTCTGGCG*CCTCACAC-3' 3'- dCG - - GAGTGTG-5'	ddCTP	large crystals, tetrahedron-shaped	2.9 Å, DNA disordered
ODN 8	5'-CTCTGGCG*CCTCACAC-3' 3'- ddCCG - - GAGTGTG-5'	dATP	no crystals	
ODN 8	5'-CTCTGGCG*CCTCACAC-3' 3'- dCCG - - GAGTGTG-5'	ddATP	large crystals, tetrahedron-shaped	3.4 Å, DNA disordered
ODN 10	5'-CTCTGGCG*CCTCACACC-3' 3'- ddCG - - GAGTGG-5'	dCTP	no crystals	
ODN 10	5'-CTCTGGCG*CCTCACACC-3' 3'- dCG - - GAGTGG-5'	ddCTP	large crystals	2.9 Å, DNA disordered
ODN 12	5'-CTCTGGCG*CCTCACACC-3' 3'- ddCG - - GAGTGTGG-5'	dCTP	no crystals	
ODN 9	5'-GACGTCTCG*CTCATCC-3' 3'- dG - GAGTAGG-5'	ddATP	crystals	4.5 Å
ODN 11	5'-GACGTCTCG*CTCATCAC-3' 3'- dG - GAGTAGTG-5'	ddATP	deformed crystals	

Table 2. AAF-dG containing oligonucleotides used for co-crystallization with Pol η . Results of the crystallization experiments as well as resolution of the diffraction data are presented.

The primer used for crystallization possessed either a 2'-deoxy base or a 2',3'-dideoxy base at the 3' end. For deoxy primer, a 2',3'-dideoxynucleoside triphosphate was used for crystallization. Incorporation of the incoming nucleotide was therefore possible, but afterwards, addition of further nucleotides was prevented due to the lacking 3'-OH group. For dideoxy primer, standard deoxy nucleoside triphosphates were utilized. Thus, no incorporation of the incoming nucleotide could take place. Crystallization conditions were similar to the conditions used for the crystallization experiments in section 7.1.2 (10–20% (w/v) PEG 3350, 150–200 mM CaCl₂, 5 mM MgSO₄, 1 mM dNTP).

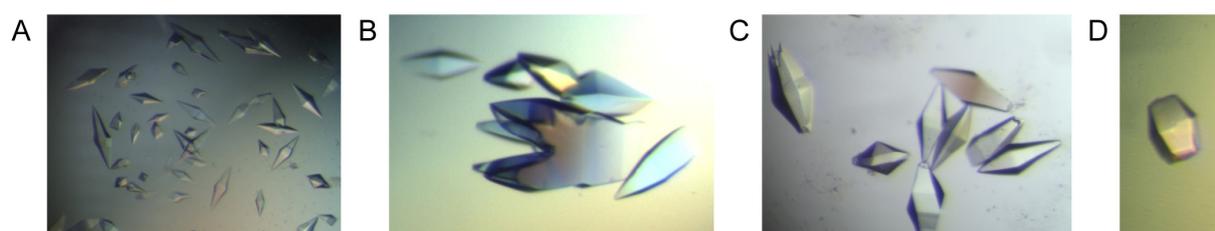


Figure 64. Crystals obtained with several oligonucleotides containing the AAF-dG lesion in the *NarI* frameshift sequence. A) **ODN 8** with 9mer dideoxyprimer and dCTP (lane 1 in Table 2). B) **ODN 8** with 9mer deoxyprimer and ddCTP (lane 2 in Table 2). C) **ODN 10** with 8mer deoxyprimer and ddCTP (lane 6 in Table 2). D) **ODN 9** with 8mer deoxyprimer and ddATP (lane 8 in Table 2).

Table 2 summarizes the results obtained with the crystallization experiments of the various AAF-dG containing oligonucleotides with the *NarI* sequence or a sequence allowing formation of a 1 nucleotide bulge. Only small, poorly diffracting crystals could be obtained, when a dideoxy primer was used. Crystallization with deoxy primer and ddNTPs resulted in bigger, mostly sufficiently diffracting crystals. In all cases however, electron density for the DNA was hardly visible. Often, only electron density for the phosphate backbone could be seen, the nucleobases were disordered. Several explanations are possible. Either, the deoxy primer produce a mixture of DNA-polymerase complexes. Some complexes contain incorporated nucleotides and some might not contain incorporated nucleotides, which leads to the poor electron density for the DNA. Or, the bulge formed in the extended frameshift sequences **ODN 8** and **10** can be bound by the protein in different positions, interchanging the relative positions of the nucleobases but not of the phosphate backbone.

7.2 Bypass of bulky DNA adducts by Pol κ

It was previously shown that Pol κ is able to bypass the AAF-dG lesion *in vitro* and *in vivo* in an error-free as well as in an error-prone manner by incorporating dTTP with the same ratio as the correct dCTP opposite the lesion.^[20, 95, 96, 97] Moreover, Pol κ is, in addition to Pol η , involved in the generation of frameshift errors, particularly -I frameshift mutations.^[94] The contribution of Pol κ to the generation of frameshift mutations during bypass of bulky adducts is not fully understood so far.^[96] The bypass of several bulky DNA adducts by Pol κ , the investigation of frameshift errors occurring during Pol κ mediated lesion bypass, and first crystallization experiments with Pol κ and AAF-dG containing oligonucleotides are described in the following section.

7.2.1 Cloning, expression and purification of human Pol κ

Two truncated Pol κ constructs, coding for the residues 50-530 and 1-530, were generated by PCR amplification of the human Pol κ gene (full gene coding for the residues 1-870). Each of the constructs was cloned in two steps into expression vectors using *StarGate*® cloning technology from *IBA*. The constructs were confirmed by sequence analysis, and expressed as C-terminal Strep-tag proteins in *E. coli* BL21(DE3)pLysS cells. The longer Pol κ construct (residues 1-530, Pol κ_L), truncated only at the C-terminus, should exhibit full polymerase activity according to the Pol κ construct prepared by *Uljon et al.*^[93] The shorter construct (residues 50-530, Pol κ_S), truncated at both N- and C-terminus, does not possess complete polymerase activity (comparable to a known Pol κ version of residues 68-526), but this version crystallizes more easily.^[162] Expression was performed in standard LB medium, however, the protein yield obtained for both Pol κ versions was not sufficient for further crystallization studies. Therefore, an autoinduction medium from *Novagen* was tested for the expression of Pol κ_L . Autoinduction media allow regulated protein expression from lac promoters in *E. coli* without monitoring the culture or adding inducers during cell growth. The medium contains a mixture of carbon sources including lactose. Initially, glucose is metabolized by the *E. coli* cells. When the glucose is depleted, uptake of lactose occurs which induces protein expression. With this method, high cell densities are obtained and, in this case, much higher yields of Pol κ were obtained.

Both Pol κ versions (Pol κ_S and Pol κ_L) were purified with the help of two columns. First, a Strep-Tactin column was employed followed by a Heparin ion exchange column to remove residual DNA. The yield using standard LB medium was around 0.2 mg protein per litre of *E. coli* expression culture. As mentioned before, the use of the autoinduction medium led to much higher yields of around 30 mg protein per litre of *E. coli* culture. In Figure 65, SDS-gels of both purified Pol κ versions as well as UV chromatograms of the heparin column purification procedures are displayed.

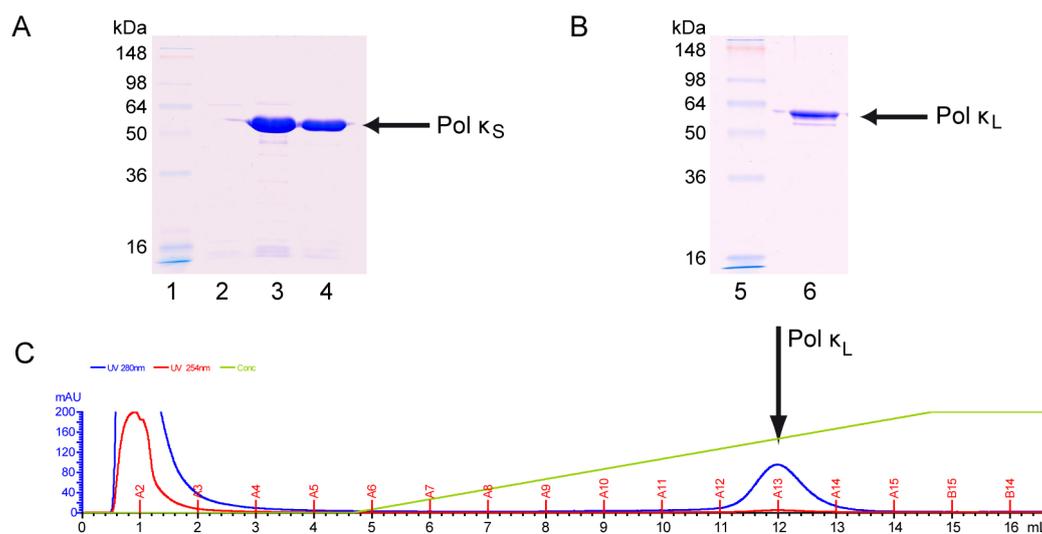


Figure 65. Purification of human Pol κ .

A) SDS-PAGE of Pol κ_S : 1: Protein marker; 2: Flow-through of the heparin affinity column; 3: 3 μ g Pol κ_S after heparin affinity purification; 4: 1 μ g Pol κ_S after heparin affinity purification. (Pol κ_S size: 55.01 kDa + 1.25 kDa Strep-tag, 481 amino acids) B) SDS-PAGE of Pol κ_L : 1: Protein marker; 2: 1 μ g Pol κ_L after heparin affinity purification (Pol κ_L size: 60.6 kDa + 1.25 kDa Strep-tag, 530 amino acids) C) Chromatogram of the Heparin affinity purification of Pol κ_L . DNA and impurities are found in the flow-through. The red curve corresponds to the detection at 254 nm, the blue line shows the elution profile at 280 nm. The green line marks the NaCl-gradient for elution of the protein.

7.2.2 Primer extension studies with Pol κ

In order to compare the activity of both purified Pol κ versions, Pol κ_S and Pol κ_L , primer extensions with undamaged template DNA as well as with an AAF-dG lesion containing oligonucleotide were performed. **ODN 1** and an undamaged template strand with the same sequence were annealed to a fluorescence labeled primer ending one nucleotide 3' to the lesion in Pol κ primer extension buffer (section 10.3.1). Primer extension was carried out with high concentrations of Pol κ_S or Pol κ_L (2 μ M) and simultaneous addition of all four deoxynucleoside triphosphates at 37 °C for 30 min (Figure 66).

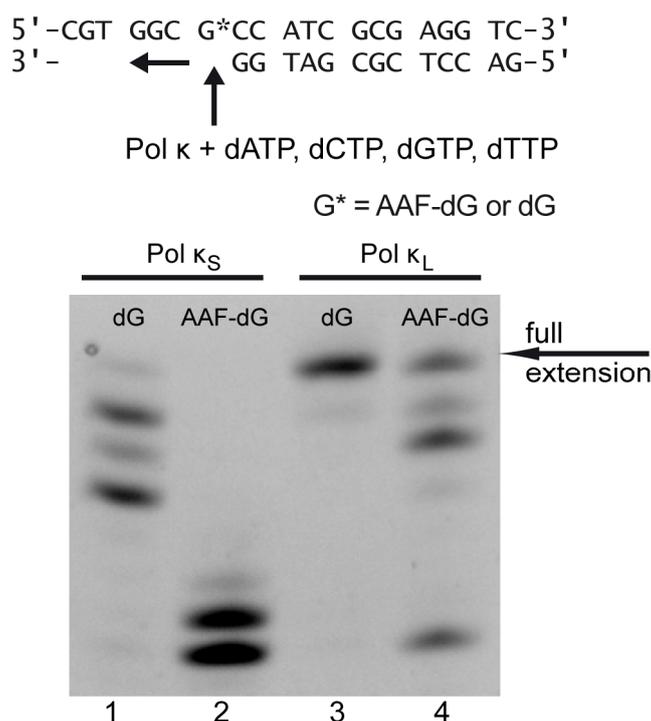


Figure 66. Translesion synthesis of AAF-dG containing oligonucleotides catalyzed by Pol κ_S and Pol κ_L (2 μ M Pol κ , 1 μ M template DNA, 200 μ M of all four dNTPs). Lane 1: Extension of undamaged template DNA by Pol κ_S ; Lane 2: Extension of and AAF-dG containing oligonucleotide (**ODN 1**) Pol κ_S ; Lane 3: Extension of undamaged template DNA by Pol κ_L ; Lane 4: Extension of and AAF-dG containing oligonucleotide (**ODN 1**) Pol κ_L .

Pol κ_S , truncated at the N- and the C-terminus, is not even able to extend undamaged template DNA completely at the high enzyme concentrations used in this study (Figure 66, lane 1). Accordingly, AAF-dG damaged DNA presents a strong block for the N-terminal truncated Pol κ (lane 2). Pol κ_L fully extends undamaged template DNA (lane 3) and also the AAF-dG containing ODN **1** (lane 4). At these high polymerase concentrations, Pol κ is able to bypass the AAF-dG lesion. Additionally, shorter bands are visible, which might be lesion induced frameshifting product or not fully extended primers. For all following primer extension studies, Pol κ_L , truncated only at the C-terminus, was chosen to ensure the presence of active enzyme. To analyze the results further, single nucleotides were inserted in primer extension studies with AF-dG and AAF-dG containing oligonucleotides using 200 nM Pol κ_L at 37 °C for 30 min. The result of this experiment is displayed in Figure 67.

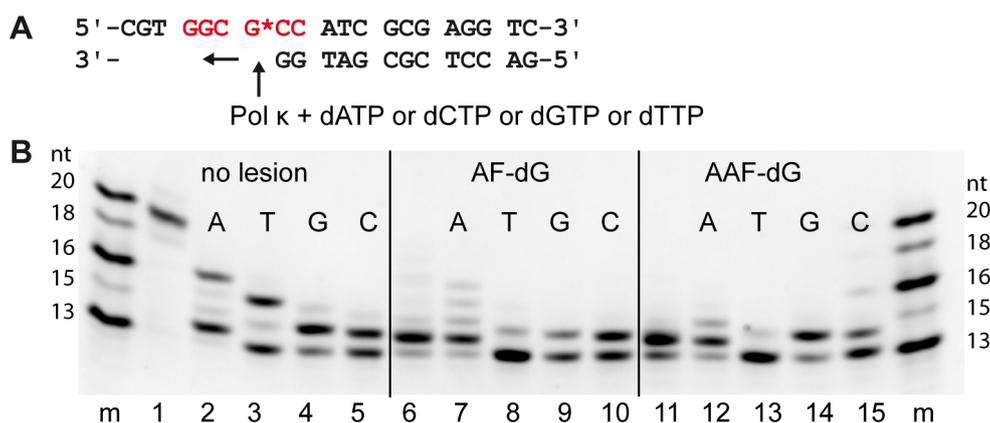


Figure 67. Incorporation of individual nucleotides with AF-dG, AAF-dG containing oligonucleotides and lesion-free oligonucleotides (200 nM Pol κ , 1 μ M template DNA, 200 μ M of each deoxynucleotide or all four dNTPs). A) Sequence used in this study (**ODN 1**); B) PAA-gel of the primer extension experiment; M: marker; Lanes 1, 6 and 11: all four dNTPs; Lanes 2, 7 and 12: dATP; Lanes 3, 8 and 13: dTTP; Lanes 4, 9 and 14: dGTP; Lanes 5, 10 and 15: dCTP. AF-dG containing DNA was used in lanes 6-10 and AAF-dG containing templates in lanes 11-15.

First of all, the relatively low fidelity of Pol κ on undamaged template DNA is obvious (Figure 67, lanes 2-5). Misincorporation of dA, dT and dG opposite a dG in the template strand occurs, if there is no other nucleotide present. For AF-dG and AAF-dG containing oligonucleotides, dC is well incorporated opposite the lesion. However, dA and dG are also inserted opposite the lesions with high efficiency. Misincorporation of dT opposite the AF-dG occurs to some extent as well. At these polymerase concentrations (200 nM), Pol κ is not able to bypass the AF-dG and AAF-dG lesions. Thus, Pol κ is able to insert nucleotides opposite both lesions and also to extent from the AAF-dG at higher enzyme concentrations (Figure 66). The polymerase does not possess the high selectivity for the incorporation of dCTP opposite the AAF-dG lesion found using Pol η .

Pol κ was now tested with a variety of bulky adduct lesions. For this, 500 nM Pol κ were used at 37 °C and 30 min reaction time (Figure 68).

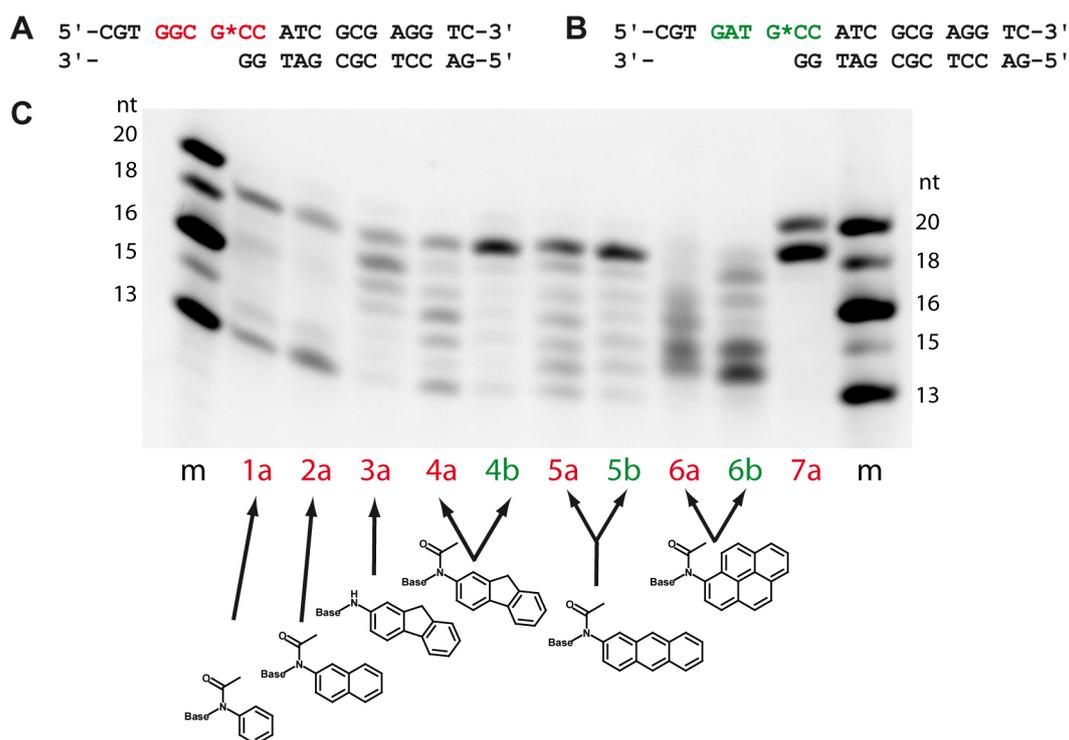


Figure 68. Translesion synthesis of different bulky dG DNA adducts catalyzed by Pol κ_L . The primer extension reactions were carried out with fluorescence labeled primers and extension products were resolved on a denaturing PAA-gel. A) and B) Sequence contexts used for the primer extension studies. C) PAA-gel of the bypass reaction of various bulky adducts. Reaction conditions are: 500 nM Pol κ , 1 μ M template DNA, 200 μ M dNTPs each. The different sequences A and B are marked with a (red) and b (green). 1a: AAB-dG (ODN 21), 2: AAN-dG (ODN 24), 3a: AF-dG (deacetylated ODN 1), 4a: AAF-dG (ODN 1), 4b: AAF-dG (ODN 2), 5a: AAA-dG (ODN 27), 5b: AAA-dG (ODN 28), 6a: AAP-dG (ODN 30), 6b: AAP-dG (ODN 31) 7a: Undamaged template strand (A), m = marker.

Pol κ is able to bypass all bulky lesions except the huge pyrene dG adduct (Figure 68). TLS efficiency is reduced if the size of the aromatic adduct is increased. Whereas bypass is surprisingly efficient for the small benzene and naphthalene units, bypass is nearly impossible when the benzene was replaced by the large pyrene unit. Moreover, during bypass of AAF-dG and AAF-dG in the *NarI* frameshift sequence (marked with “a” in Figure 68), additional shorter extension products are observed, possibly due to the occurrence of frameshifts. However, no clear –II frameshift product can be found as in the case of Pol η bypassing AAF-dG or AAA-dG the *NarI* sequence. Instead, truncated products from incorporation of one to six nucleotides were observed. Possibly, Pol κ does bypass lesions in such repetitive sequences by a frameshift mechanism, but the generated frameshift mutations are not as sequence specific as in the case of Pol η . The extension of the primer stopped in all cases one nucleotide before full extension (19 nt). Only for undamaged template DNA, some fully extended product could be found. This observation was made before in *in vitro* bypass studies of DNA lesions by Pol κ .^[96] *In vitro*, complete extension of primers is difficult for Pol κ and

only achieved at high enzyme concentrations. Mostly, Pol κ dissociates from the template DNA without incorporating the last nucleotide.

In summary, Pol κ is, like Pol η , able to bypass the C8 bulky DNA adducts synthesized in this thesis work. However, bypass seems to be more error-prone. The incorporation of nucleotides opposite the AAF-dG lesion is less selective.

7.2.3 Preliminary crystallization screens with C8-dG containing oligonucleotides and the low fidelity polymerase κ

In order to clarify the bypass mechanism of bulky DNA adducts, first crystallization experiments with Pol κ and AAF-dG containing oligonucleotides were performed. Pol κ_L was chosen for these experiments, because this Pol κ version is able to bypass the lesion whereas the shorter version (Pol κ_S) is blocked (see section 7.2.2). Crystallization of the polymerase containing the complete N-terminal part would be significant because the encirclement of the DNA by its unique “N-clasp” at the N-terminus could further be studied.^[162] Thus, Pol κ_L was incubated with double stranded DNA containing the template with the AAF-dG lesion annealed to primer with a 2',3'-dideoxybase at the 3' end in a 1 to 2 molar ratio in crystallization buffer for co-crystallization. First, a screening with a 16mer AAF-dG strand (**ODN 3**), 5 mg/ml Pol κ and addition of dCTP was performed using commercial available screening kits from *Qiagen*. A second screening was performed with a 18mer AAF-dG strand (**ODN 17**), higher enzyme concentrations (11 mg/ml) and again addition of dCTP using commercial available screening kits from *Hampton*. This work was done at the *Oxford Protein Production Facility* in Oxford, UK. The DNA sequence containing the AAF-dG (**ODN 17**) was designed according to the sequence used for crystallization of undamaged DNA in complex with Pol κ .^[162] Additionally, a manual screening using the same DNA sequence was carried out in 24 well format using conditions previously published for the Pol κ in complex with undamaged DNA.^[162] No diffracting crystals could be obtained so far.

8 Co-crystallization of AAF-dG containing oligonucleotides with the NER proteins Rad14 from yeast and its human homologue XPA

In addition to translesion synthesis, bulky DNA adduct lesions are usually repaired by the nucleotide excision repair pathway (NER).^[124-126] The mechanism how the repair system is able to recognize a variety of lesions is unclear and, as discussed in section 3.6, several models are proposed for the initial lesion recognition step. It has been shown that the NER protein Rad14 from yeast and its human homologue XPA bind DNA containing bulky DNA adducts such as AAF-dG.^[139] To date, only a solution structure of the minimal DNA binding domain of XPA is available, but no structural data of the lesion recognition. In order to shed light on the mechanism of lesion recognition by these NER proteins, co-crystallization experiments of Rad14 and XPA with AAF-dG containing oligonucleotides were performed.

8.1 Purification of Rad14 and XPA

A plasmid containing the DNA sequence coding for the DNA binding domain of yeast Rad14 from *S. cerevisiae* (residues 185-306) was cloned and transformed into *E. coli* BL21 in frame with a C-terminal Strep-tag by *Dr. Ralf Strasser*. A plasmid containing the DNA binding domain of the human XPA protein (residues 98-219) was prepared analogous by *Sandra Koch*. The expression and purification protocol was also established by *Dr. Ralf Strasser*.^[139]

Initially, both proteins were purified using a Strep-Tactin column, which was followed by ion exchange chromatography using a MonoQ anion exchange column (*GE*) for Rad14 and a Heparin affinity column for XPA. In order to solve the phase problem inherent to macromolecular crystallography, selenomethionine labeled protein was also prepared. To this end, the plasmids were transformed into a methionine auxotrophic *E. coli* strain and expression was carried out in a selenomethionine-containing medium. Purification of the proteins was performed analogous to the non-labeled proteins with degassed buffers and in the presence of DTT as reduction agent. In Figure 69, SDS-gels of the purified selenomethionine labeled Rad14 and XPA proteins as well as the UV chromatograms of their final purification steps are displayed.

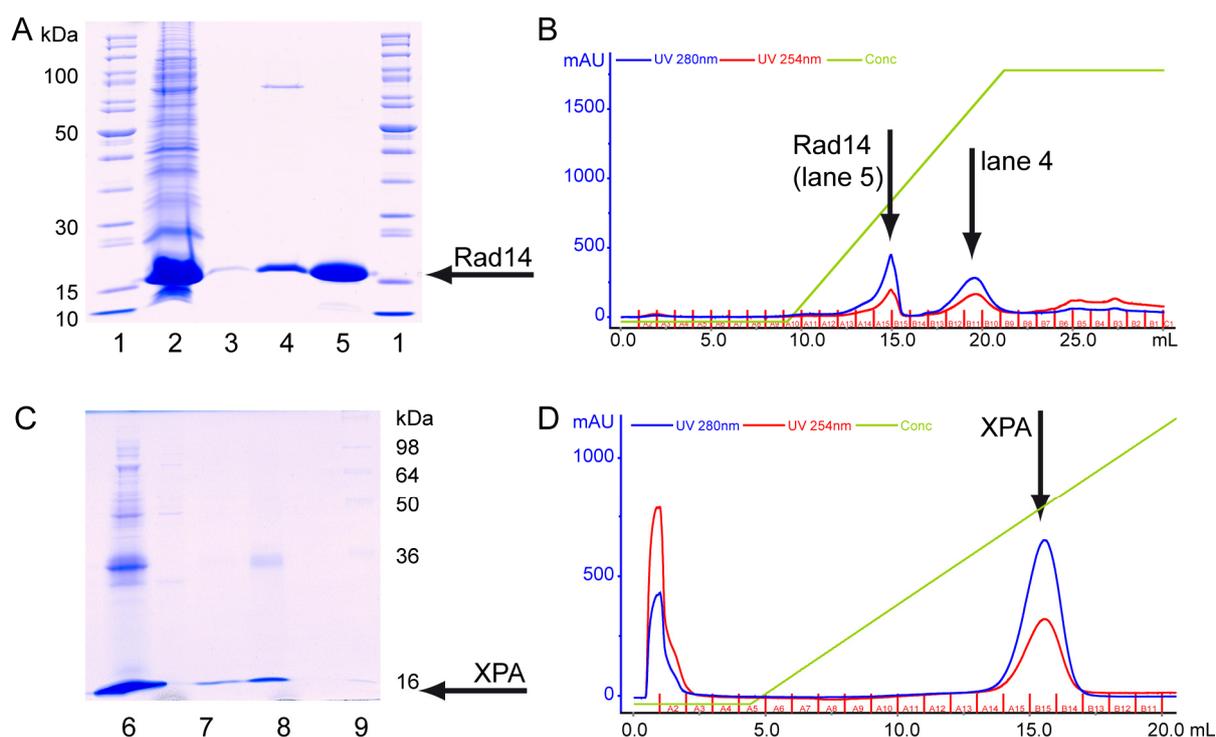


Figure 69. Purification of selenomethionine labeled yeast Rad14 and human XPA

A) SDS-PAGE of Rad14 (14.2 kDa + 1.25 kDa Strep-tag): 1: Protein marker; 2: Rad14 after Strep-tag purification; 3: Flow through of the MonoQ anion exchange column; 4: Second peak of the MonoQ anion exchange purification at 19 mL elution volume (see UV-chromatogram in B) The additional band at around 90 kDa might be a multimeric complex of the protein. 5: Rad14 after MonoQ anion exchange purification (at 15 mL elution volume, see UV-chromatogram in B) B) Chromatogram of the MonoQ anion exchange purification of Rad14. The red curve corresponds to absorption at 254 nm (DNA contamination), the blue line at 280 nm. The green line marks the NaCl-gradient for elution of the protein.

C) SDS-PAGE of XPA (14.6 kDa + 1.25 kDa Strep-tag): 6: XPA after Strep-tag purification; 7: Flow through of the heparin affinity column; 8: XPA after heparin affinity purification D) Chromatogram of the Heparin affinity purification of XPA. The red curve corresponds to absorption at 254 nm, the blue line at 280 nm. The green line marks the NaCl-gradient for elution of the protein.

8.2 Crystallization of Rad14/XPA in complex with an AAF-dG containing DNA duplex

Initial crystallization experiments with the DNA binding domain of Rad14 were made together with *Dr. Ralf Strasser*. The first Rad14 crystals in complex with AAF-dG containing DNA were obtained with 2.8- 3.3 M ammonium sulfate in 0.08 – 0.11 M Tris pH 8 buffer. They diffracted X-rays to 7 Å spacing.^[139]

8. Co-crystallization of AAF-dG containing oligonucleotides with Rad14 and XPA

Screening for new crystallization conditions was performed as described in section 10.5. Therefore, **ODN 16** containing an AAF-dG lesion (5'-GCTCTACG^{AAF}TCATCAC-3') was annealed to its counter strand and incubated with either the purified Rad14 or XPA protein. Initial high-throughput crystallization screening with the unlabeled proteins was carried out using commercial available screening buffers from *Qiagen* and robotics. For Rad14, an initial hit was obtained with a sodium cacodylate buffer at pH 6.5 and calcium acetate/PEG 8000. These conditions were manually optimized by varying the protein concentration, concentration of the components of the crystallization buffer and drop size in 24 well formats resulting in larger, but thin crystal plates (Figure 70, A).

For XPA, an initial hit was obtained with PEG 3350 / MgSO₄ (Figure 70 A). Conditions were also refined in 24 well formats, however only very small crystals were obtained (Figure 70 B) which could not be measured. The two crystals shown in Figure 70 A were frozen in artificial mother liquor containing ethylene glycol as cryoprotectant but did not diffract X-rays at a synchrotron radiation source (PX I, Swiss Light Source (SLS), Switzerland).

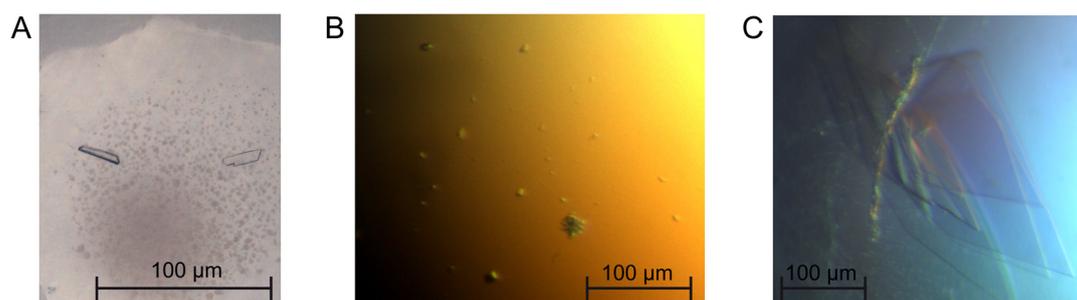


Figure 70. Crystallization of Rad14 and XPA in complex with AAF-dG containing DNA. A) Crystals obtained in high-throughput screening experiment with XPA in complex with AAF-dG containing DNA (drop size 100 nL + 100 nL). B) Crystals obtained with XPA and AAF-dG containing DNA after optimization of the initial screening conditions (1 μ L + 1 μ L drop size). C) Crystals obtained with Rad14 and AAF-dG containing DNA after optimization of the initial screening conditions.

The Rad14/AAF-dG complex crystals (Figure 70 C) were transferred to a cryoprotectant of artificial mother liquor containing ethylene glycol prior to flash freezing. Data collection was carried out at the microfocus beamline ID23-2 at the European Synchrotron Radiation Facility (ESRF, France). Due to the plate like morphology of the crystals, very anisotropic diffraction was observed, ranging from 3 to 4.5 \AA . Indexing of the data revealed that the crystals belong to the orthorhombic space group $C222_1$ with a unit cell of $a=54.1 \text{ \AA}$, $b=154.2 \text{ \AA}$, $c=99.3 \text{ \AA}$, $\alpha=\beta=\gamma=90^\circ$. By collecting data over the whole area of the crystal, a complete data set with a resolution of 3.2 \AA was obtained.

8.3 Phase determination for structure solution of the Rad14 crystals

In order to determine the structure of the Rad14/AAF-dG complex, two different approaches were chosen to obtain experimental phase information by carrying out a SAD experiment. Thus, on the one side, all three dT's in the counter strand were replaced by 5-Bromo-deoxyuridine (5-Br-dU) instead of dT. On the other side, selenomethionine labeled Rad14 was used for crystallization together with the same DNA duplex used for production of the native crystals. Crystals were obtained with conditions similar to conditions used for the native Rad14 crystals. However, the crystals were again plate-shaped and even slightly smaller in size compared to the native crystals (Figure 71).

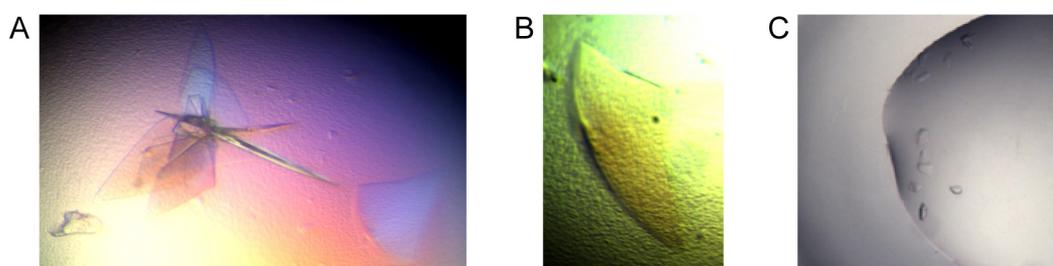


Figure 71. Crystals of selenomethionine labeled Rad14 in complex with AAF-dG containing DNA. A) and B) Crystals obtained with selenomethionine labeled Rad14 at conditions similar to the conditions used for the native crystals (drop size 1 μ L + 1 μ L). C) Crystals obtained with selenomethionine labeled Rad14 and AAF-dG containing DNA in a new crystallization screening testing different buffer conditions (drop size 100 nL + 100 nL).

The 5-Br-dU containing crystals diffracted X-rays to 3.5 Å spacing but no complete data set could be obtained due to radiation damage during the measurement. High beam intensities had to be used to get sufficient diffraction, which caused severe damages to the crystal. The space group was found to be identical to the native crystals. Nevertheless, an X-ray fluorescence scan at the Br-edge confirmed successful co-crystallization of Rad14 with the DNA duplex.

Selenomethionine-labeled Rad14 crystals were again plate-shaped and varied in size depending on the screening conditions. Unfortunately the crystals continuously grew properly into two dimensions, but not in the third dimension (Figure 71, A and B). Micro- and macro-seeding was performed in order to get thicker crystals but the crystal shape did not improve significantly.

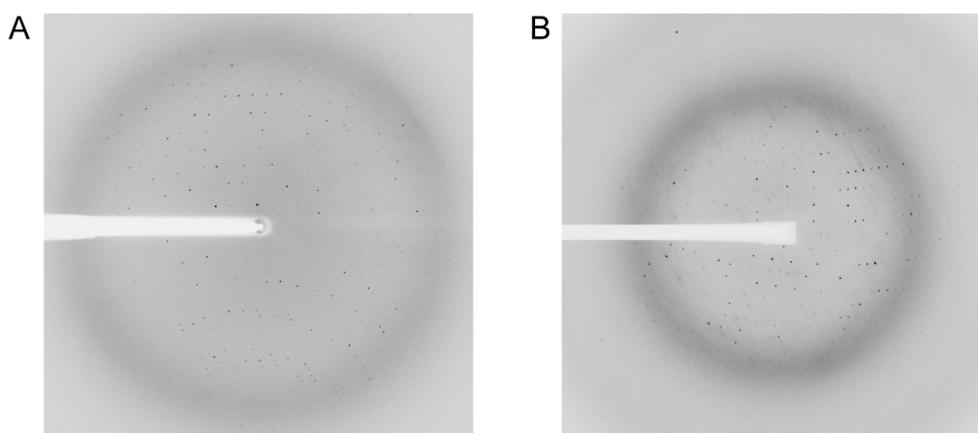


Figure 72. Diffraction images of A) a native Rad14 crystal and B) a selenomethionine labeled Rad14 crystal in complex with an AAF-dG containing DNA duplex taken at the microfocus beamline ID23-2 at the ESRF (France) and the PXI beam line at the SLS (Switzerland). In both images, high anisotropy of the crystals can be observed.

A further problem arose from the cacodylate buffer used in the crystallization setup of the selenomethionine labeled Rad14/AAF-dG complex. Arsenic in the buffer solution has a theoretical absorption edge of 11.87 eV. This absorption edge is very close to the Se-edge with a theoretical value of 12.66 eV, thus interfering with the measurement. Therefore, a buffer screening at constant calcium acetate and PEG 8000 concentrations was performed and Rad14 crystals with Tris buffer at pH 7 could be obtained (Figure 71, C). These conditions were again optimized in 24 well formats. Slightly larger and thicker crystal plates were obtained compared to the initial crystals.

The peak and inflection points of the anomalous scatterers were determined by an X-ray fluorescence scan around the respect resonance edge (Figure 73 and Figure 74).

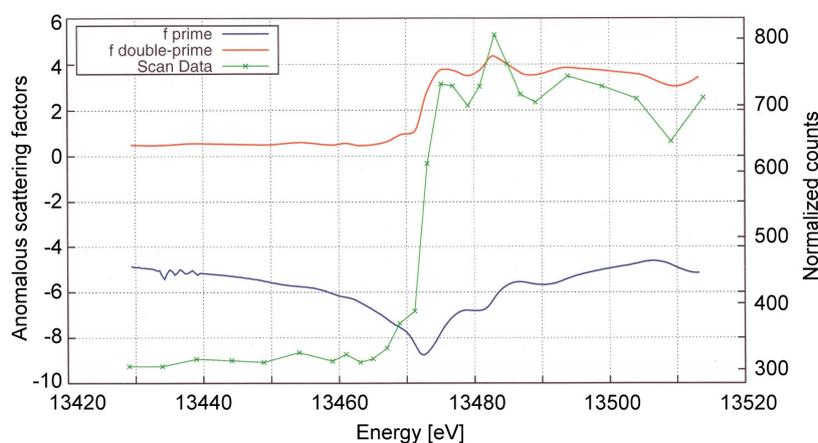


Figure 73. X-ray fluorescence scan near the bromine K - absorption edge of the 5-Br-dU containing Rad14 crystals. The inflection point is at 13472 eV with the anomalous scattering factors $f' = 8.78$ and $f'' = 1.98$. The peak is at 13483 eV with $f' = 6.33$ and $f'' = 4.36$.

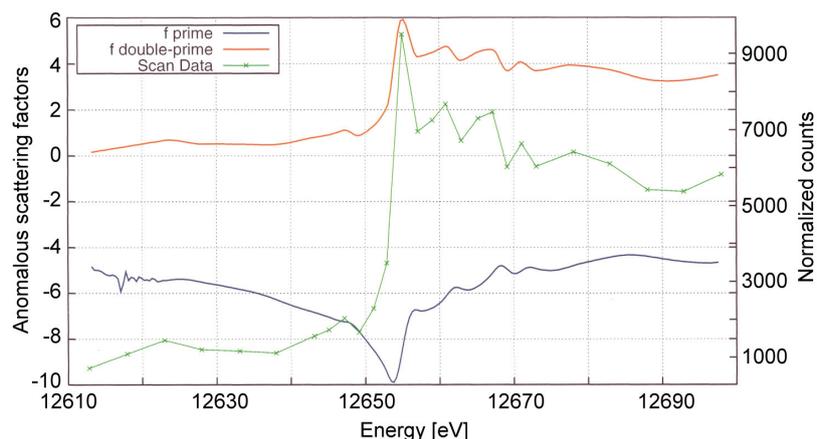


Figure 74. X-ray fluorescence scan near the selenium K - absorption edge of the selenomethionine labeled Rad14 crystals. The inflection point is at 12654 eV with the anomalous scattering factors $f' = 9.89$ and $f'' = 3.71$. The peak is at 12655 eV with $f' = 8.00$ and $f'' = 5.92$.

Several data sets were taken at the Se absorption edge showing that the protein crystallized in a different space group. The crystals were monoclinic. Again the derivatized complex crystals diffracted anisotropically and were very sensitive to radiation damage (Figure 72). Due to the lack of redundancy, radiation damage and the anomalous signal only extending to 4.5 - 5 Å in 5-Br-dU and selenomethionine containing Rad14 crystals, structure solution so far remains elusive.

8.4 Discussion

The mechanism of lesion recognition in the NER pathway is still not well understood. It has been shown that the yeast Rad14 and its human homologue XPA bind specifically to DNA strands containing bulky adducts.^[139] Here, the minimal DNA binding domains were overexpressed, purified and crystallized in complex with AAF-dG containing DNA. Diffracting crystals of Rad14 could be obtained and some phase information was derived from SAD experiments using complex crystals with 5-Br-dU containing DNA or selenomethionine labeled protein. Nevertheless, due to the anisotropic nature of the diffraction and the sensitivity to radiation damage, structure solution remained elusive. Thus, further screening by varying the DNA strands is necessary to obtain better quality crystals and so to be able to solve the structures of these so important proteins in order to shed light on their mechanism of action.

9 Conclusion and outlook

Aromatic amines are known to be strong carcinogens forming DNA adducts after metabolic activation. Preferred reaction sites are the amino groups of adenine and guanine and particularly the C8-position of guanine.^[16, 17] The latter reaction gives rise to the so called C8-bulky adduct lesions, which interfere with the replication process leading to mutations. Two types of aromatic amine lesions are known. The non-acetylated lesions reduce the replication efficiency, but are in general faithfully bypassed by high fidelity polymerases.^[51, 120] In contrast, the N^8 -acetylated derivatives block replicative polymerases but can be bypassed with special low-fidelity polymerases.^[52-54] The low fidelity Y-family polymerase η plays a major role in the replication through acetylated aromatic amine lesion such as AAF-dG *in vivo*.^[56, 57, 60]

In this thesis work, lesions derived from aminofluorene (AAF-dG) and aminoanthracene (AAA-dG) served as model compounds to investigate the mechanism of the error-free bypass by Pol η . Crystal structures of Pol η in complex with oligonucleotides containing either the AAF-dG or the AAA-dG lesion show why the polymerase is blocked if it encounters the lesion but they also allow to construct a model that explains how Pol η finally achieves the bypass (section 7.1.2). Based on the crystal structure, correct bypass is achieved by the enzyme through rotation of the DNA around the bulky moiety while keeping the damaged dG in *syn*-conformation. Thus, TLS is accomplished without rotation of the lesion into the *anti*-conformation as previously thought.^[123] In this situation, correct dCTP incorporation opposite the lesion can be envisioned by formation of one hydrogen bond between the dCTP and the lesion (Figure 75).

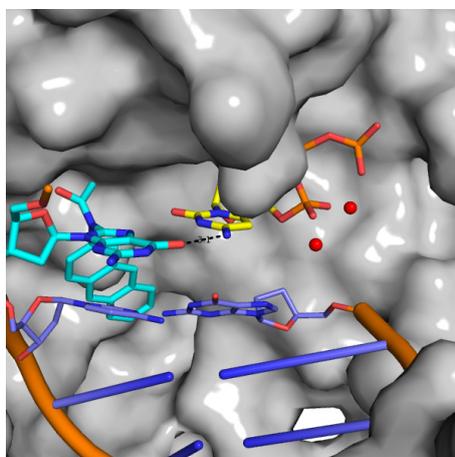


Figure 75. Model of the AAA-dG bypass by Pol η . The dG:dC basepair 3' of the lesion in the AAA-dG-Pol η complex B was replaced by the AAA-dG adduct and the dCTP triphosphate. The dCTP was placed in the active site according to the structure of Pol η in complex with the cisplatin GG lesion

(PDB code 2R8J). Thus the active site for the incoming dCTP opens, which forms one critical hydrogen bond to the adducted nucleobase in *syn* conformation.

The model was further supported by biochemical studies using synthetic zebularine triphosphate which is not incorporated opposite the AAF-dG lesion due to the lacking ability to form the critical hydrogen bond. Moreover, TLS efficiency of Pol η depends critically on the size of the bulky adduct forming the lesion.

Future work could include additional crystallization screening of Pol η in complex with various AAF-dG oligonucleotides of different length. Perhaps, dCTP incorporation could then be observed crystallographically and the model of the bypass could be confirmed. Recently, it was proposed that the crystallization ability of Pol η – DNA adducts is improved by site-directed mutagenesis of amino acid residues involved in crystal contacts.^[88] In this crystal structure, the PAD domain of the protein closes, however, no structural changes occur in the active site. Crystallization with this mutated Pol η could result in more stable protein-DNA complexes, thus facilitating crystal growth.

Another part of this thesis work involved the investigation of frameshift mutations. Bulky adduct DNA lesions such as AAF-dG lesion induce frameshift mutations if they are placed in special gene sequences.^[16, 53, 61-64] With the help of synthetic AAF-dG lesions inserted into different repetitive frameshift prone sequences, the molecular basis of the frameshifting event was determined (section 7.1.4). It could be shown that frameshifting requires DNA polymerase induced primer-slippage followed by primer misalignment and extension of the misaligned primer-template complex. The mechanisms of -I, -II and -III frameshifts were deciphered. An example (-III frameshift) is shown in Figure 76.

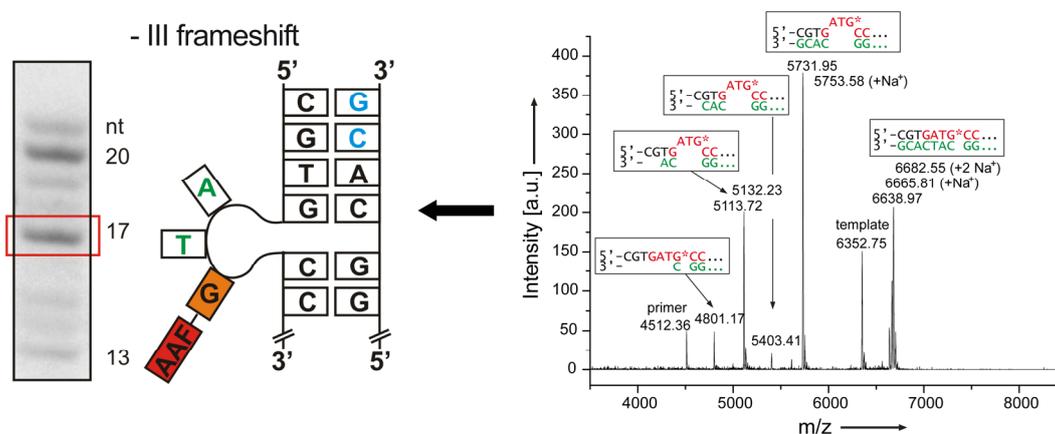


Figure 76. Slippage process leading to a -III frameshift during bypass of the AAF-dG lesion by Pol η . The polymerase first extends the primer with a dC base inserted opposite the lesion. After this, a substantial fraction of the primer misaligns to pair with the alternative dG in the template two bases

further away. MALDI-Tof mass spectrometry analysis proved the formation of a bulge consisting of three bases.

However, crystallization of Pol η in complex with an AAF-dG containing, misaligned primer/template construct forming a bulge was not successful so far (section 7.1.4.2). Possibly, crystallization tendency will be improved when the mutated Pol η mentioned above is used.

Moreover, the catalytic core of human Pol κ was cloned, overexpressed and purified for primer extension studies in this thesis work (section 7.2). It could be shown, that Pol κ is, like Pol η , able to bypass the synthesized C8 bulky DNA adducts. However, bypass seems to be more error-prone for Pol κ , the incorporation of nucleotides opposite the AAF-dG lesion is not selective. First crystallization screenings were performed with Pol κ in complex with an AAF-dG oligonucleotide. No diffracting crystals were obtained so far. Further experiments could involve screening with AAF-dG oligonucleotides with different lengths. Moreover, other truncated Pol κ versions could be prepared, for example from residues 19 to 526. This enzyme was recently used for crystallization in complex with 8-oxo-dG containing DNA.^[163] Possibly, shortening of the flexible N-terminus could also facilitate crystallization with AAF-dG containing oligonucleotides.

This thesis work also includes crystallization of the yeast NER protein Rad14 and its human homologue XPA together with AAF-dG containing oligonucleotides (section 8). The mechanism of lesion recognition in the NER pathway is still not well understood. It has been shown that the Rad14 and XPA bind specifically to DNA strands containing bulky adducts.^[139] In this thesis, the minimal DNA binding domains were overexpressed, purified and crystallized in complex with AAF-dG containing DNA. Diffracting crystals of Rad14 could be obtained and some phase information was derived from SAD experiments using complex crystals with 5-Br-dU containing DNA or selenomethionine labeled protein. However, structure solution was not successful so far due to anisotropic diffraction and radiation damage. Future work could involve further screening varying the DNA strands to obtain better quality crystals and so to be able to solve the structures.

10 Experimental

10.1 General chemical materials and methods

Chemicals and solvents were used as commercial available in the qualities *puriss.*, *p.a.* or *purum* and purchased from the companies *Fluka*, *Acros*, *Aldrich*, *ABCR* or *Sigma*. Bidistilled water was produced by a *Milli-Q Plus* unit (*Millipore*). Dry solvents (< 50 ppm H₂O) were purchased from *Fluka*, *Acros* or *Aldrich*, solvents for extraction and chromatography had technical grade and were distilled before usage. All reactions employing dry solvents were performed under inert atmosphere (N₂). Technical grade solvents were distilled prior to use for column chromatography using a *Laborota 4001-efficient* rotary evaporator (*Heidolph*) which was equipped with a *Vario PC2001* membrane pump (*Vacuubrand*). A *RZ-4* oil pump (*Vacuubrand*) was used for further drying of substances (final vacuum 10⁻²-10⁻³ mbar).

Small amounts of aqueous and buffer containing solutions were removed by a *SpeedVac Plus SC110A* and *SpeedVac SPD 111V* (*Savant*) or a *RVC 2-25* and *RVC 2-33 IR* (*Christ*) all equipped with a *CT 62-50-cooling* trap (*Christ*). An *Alpha 2-4 plus* lyophilizer (*Christ*) was used for the lyophilisation of aqueous solutions. Molecular sieve was washed with water prior to use, pre-dried in a cabinet-drier and further dried under high vacuum in a sand bath (350 °C) for 12 h.

Column chromatography was performed with silica gel 60 from Merck.

Thin layer chromatography (TLC) was performed with aluminum plates (silica gel 60 F₂₅₄, 10 × 5 cm). Substances were visualized by illumination with UV-light (λ = 254 nm, 365 nm) or by staining with subsequent heating. The staining was performed using anisaldehyde solution (2.2 g anisaldehyde, 2.0 mL conc. H₂SO₄, in 100 mL acetic acid).

HPL-Chromatography was performed on a *Merck-Hitachi* system or on a *Waters* system. In detail, the systems contain the following units: *Merck_{analytical}*: *L-7400* UV detector, *L-7480* fluorescence detector, *L-7100* pump; *Merck_{preparative}*: *L-7150* pump, *Rheodyne P/N 77 25i* injection valve, *L-7420* UV-Vis, *ERC-3415* solvent degasser; *Waters_{analytical}*: *alliance 2695 Separation Module* with *PDA 2996* or *996 Diode Array detector* and fluorescence detector (*2475*); *Waters_{preparative}*: *1525* HPLC Pump, *2487 Dual Wavelength UV Detector*. *CC 250/4 Nucleosil 120-3 C18-* and *VP 250/4 Nucleodur 100-3 C18 ec-* columns (*Macherey-Nagel*) were used for analytical reversed-phase separations. *VP 250/10 Nucleodur 100-5 C18 ec-* and

VP 250/10 Nucleosil 100-7 C18-columns were used for preparative separations. For enzymatic digests, a *MS Uptisphere 3HDO #15MS (150 x 2.0 mm)* column from *Interchim* was used.

Mass spectrometry data for ESI-MS were collected on a Finnigan LTQ FTICR (*Thermo Finnigan*, Bremen) and on a Thermo Finnigan LTQ Orbitrap XL (*Thermo Scientific*). MALDI-Tof was performed on a Bruker Autoflex II spectrometer with 6-aza-2-thiothymine (ATT) or 3-hydroxypicolinic acid (HPA) as matrix (see section 10.3.5, *Solutions*) and a *MTP AnchorChipTM var/384 (Bruker)* as target. Salt-containing samples were desalted by dialysis on *MF-MilliporeTM* membrane filters (pore size 0.025 μm) or by *Zip-Tips C18 (Millipore)*.

HPLC-MS High resolution mass spectra of digested oligonucleotides (section 10.2.6) were recorded on a Finnigan LTQ FTICR and on a Thermo Finnigan LTQ Orbitrap XL. Resolution was set to 100000 at $m/z = 400$. The mass range was set to 2000 mass units. A voltage of 3 kV and a temperature of 300°C were applied to the spray capillary. The HPLC system was operated with a flow of 150 $\mu\text{l}/\text{min}$. Detailed information to the columns and buffers used can be found in section 10.1 and 10.3.1. The mass values of the detected ions in correlation to their ion charge are indicated as m/z and the relative signal intensities are stated in percent referring to the maximum signal.

Melting points were measured with a Büchi Melting Point B-540.

Infrared spectroscopy was performed on a *Perkin Elmer Spectrum BX FT-IR-System*. The detection ranged from 400 to 4000 cm^{-1} . The following abbreviations were used for the characterization of the bands: *s (strong)*, *m (medium)*, *w (weak)*, *br (broad)*.

NMR spectra were recorded on the following spectrometers: *Varian Oxford 200*, *Bruker AC 300*, *Varian XL 400* and *Bruker AMX 600*. The chemical shifts (δ) are given in ppm. Coupling constants (J) are given in Hz, correspond to the observed frequency differences and were not corrected in non first order situations.

Extinction coefficients of the oligonucleotides at 260 nm were calculated by addition of the extinction coefficients of the individual nucleobases. These are dA 15.0 L/mmol·cm, dC 7.1 L/mmol·cm, dG 12.0 L/mmol·cm and dT 8.4 L/mmol·cm. For the bulky dG adducts, the dG value was used.

10.2 Oligonucleotide synthesis and characterization

10.2.1 Automated DNA synthesis

DNA oligonucleotides were synthesized on a *PerSeptive Biosystems Expedite 8900* Synthesizer and an *Åkta Oligopilot 10 (GE)* using *ultramild* bases and reagents and following standard phosphoramidite protocols (10 equivalents of phosphoramidite (0.1 M), 4 min recycling time and 1-2 μmol scale synthesis). The phosphoramidite amounts for the bulky adduct dG lesions were equal to those of the standard bases but coupling times were elongated to 10 minutes. *Ultramild* protected phosphoramidites were used with the following protecting groups:

dC: *N*-acetyl (Ac-dC)

dA: *N*-Phenoxyacetyl (Pac-dA)

dG: *N*-isopropylphenoxyacetyl (iPrPac-dG)

dT: no protection group necessary

The *ultramild* Q-CPG-linker (controlled pore glass, preloaded with Ac-dG or iPrPAC-dG) as solid support was purchased from *Glen Research*. Phosphoramidites and the activator were solved in acetonitrile from *Riedel de Haen* ($\text{H}_2\text{O} < 10$ ppm). All other reagents were prepared using HPLC grade acetonitrile from *Roth*. The reagents were dried over 3 Å molecular sieve for 12 h prior to synthesis. The following solutions were used for deblocking, capping and oxidation:

Deblocking: 3 % dichloroacetic acid in toluene

Cap A: 0.5 M phenoxyacetic anhydride[‡] in MeCN/2,6-lutidine 9:1

Cap B: 20 % *N*-methylimidazole in MeCN

Oxidation: 0.025 M iodine in 2,6-lutidine/MeCN/ddH₂O 6:65:30

Activator: 0.25 M 5-benzylthiotetrazole (BTT) in MeCN

In most cases, the terminal 5'-DMT group was removed after the synthesis. Only for long DNA strands (30mer oligonucleotides), the 5'-DMT group was kept to facilitate HPLC purification.

[‡] for synthesis see section 10.7.13

Deprotection of the oligonucleotides and cleavage from the solid support were carried out under mild alkaline conditions adding β -mercapthoethanol to prevent the AAF-dG from being oxidized.^[148, 151] In detail, deprotection was achieved with 10 % diisopropylamine and 0.25 M β -mercapthoethanol in methanol at 55 °C for 12 h. The solvents were removed in a SpeedVac concentrator and the resulting pellet was dissolved in ddH₂O. The sample was shortly centrifuged and the supernatant was transferred to a fresh tube. Subsequent filtration through a 0.2 μ m filter removed residual CPG before to HPLC purification.

Lesion free oligonucleotides as well as 3' ddC or 5-Br-dU containing oligonucleotides were purchased RP-HPLC purified and lyophilized from *Metabion* (Martinsried).

10.2.2 Purification of bulky adduct containing oligonucleotides

Analytics and purification were performed on *Merck* and *Waters* HPLC systems using reversed-phase *Nucleosil* or *Nucleodur* C18 columns (*Macherey-Nagel*, section 10.1). As buffer system, 0.1 M triethylammonium acetate in ddH₂O (buffer A) and 0.1 M triethylammonium acetate in 80 % MeCN (buffer B) were used. A gradient of 0-40 % buffer B over 45 min was used for separation of synthesized 15 to 30mer bulky adduct containing oligonucleotides. The preparative *Waters* HPLC system allowed monitoring a second wavelength during purification. The aromatic adduct containing oligonucleotides were detected at wavelengths between 300 and 350 nm depending on the aromatic unit attached. The HPLC fractions were lyophilized and analyzed by MALDI-Tof and analytical HPL-chromatography. The fractions containing the pure oligonucleotide were dissolved in 1 ml ddH₂O, desalted on *Waters* Sep-Pak® C18 cartridges and lyophilized again. If the terminal 5'-DMT-group was kept for easier HPLC purification, the DMT group was cleaved with TFA during the desalting process on the Sep-Pak®. Therefore, a 1.5 % TFA solution was flushed through the column with the bound oligonucleotides for 30 s and was afterward directly neutralized with HPLC buffer A. Then, the standard Sep-Pak® protocol was continued. The oligonucleotides were stored lyophilized at -20 °C prior to usage.

An overview about the synthesized oligonucleotides containing different bulky DNA lesions is found in Table 3.

ODN	description	lesion	sequence	length	m/z_{calc}	m/z_{found}
1	AAF-PE1	AAF	5'-CGT GGC G*CC ATC GCG AGG TC-3'	20mer	6356.3	6351.0
2	AAF-PE2	AAF	5'-CGT GAT G*CC ATC GCG AGG TC-3'	20mer	6355.3	6350.9
3	AAF-C1	AAF	5'-TCT CTC G*CT CAT CCAC-3'	16mer	4949.4	4949.8
4	AAF-C2	AAF	5'-TCT TCT CG*C TCA CCAC-3'	16mer	4949.4	4949.2
5	Liv1	AAF	5'-CCT CG*G CGC CTC-3'	12mer	3777.6	3780.8
6	Liv2	AAF	5'-TCT CTA G*AC CTC-3'	12mer	3775.6	3776.0
7	Liv3	AAF	5'-GTG ATT CCG GCA G*GC GTG CGA CCT GGC TAC-3'	30mer	9456.6	9451.8
8	FS1	AAF	5'-CTC TGG CG*C CTC ACA C-3'	16mer	4999.3	5001.0
9	AS1	AAF	5'-GAC GTC TCG* CTC ATC C-3'	16mer	5014.3	5015.8
10	FS15mer	AAF	5'-CTC TGG CG*C CTC ACC-3'	15mer	4686.1	4686.1
11	AS17mer	AAF	5'-GAC GTC TCG* CTC ATC AC-3'	17mer	5327.5	5327.5
12	FS17mer	AAF	5'-CTC TGG CG*C CTC ACA CC-3'	17mer	5288.4	5287.4
13	Mod2	AAF	5'-CGT GGT G*CC ATC GCG AGG TC-3'	20mer	6371.1	6674.6
14	Mod1	AAF	5'-CGT GGC G*TC ATC GCG AGG TC-3'	20mer	6371.1	6373.3
15	Mod3	AAF	5'-CGT GAC G*CC ATC GCG AGG TC-3'	20mer	6340.1	6342.5
16	ella1	AAF	5'-GCT CTA CG*T CAT CAC-3'	15mer	4709.1	4707.4
17	ella2	AAF	5'-CCT AG*GAGT CCT TCC CCC-3'	18mer	5592.6	5589.7
18	G_runs	AAF	5'-CGT GGG G*GG ATC GCG AGG TC-3'	20mer	6476.1	6474.3
19	XCT	AAF	5'-CGT GGC G*CT ATC GCG AGG TC-3'	20mer	6371.1	6371.9
20	C_runs	AAF	5'-CGT CCC G*CC ATC GCG AGG TC-3'	20mer	6276.0	6271.4
21	AAB-PE1	AAB	5'-CGT GGC G*CC ATC GCG AGG TC-3'	20mer	6267.9	6267.8
22	AAB-PE2	AAB	5'-CGT GAT G*CC ATC GCG AGG TC-3'	20mer	6268.9	6266.1
23	AAB-C1	AAB	5'-TCT CTC G*CT CAT CCAC-3'	16mer	4861.0	4860.5
24	AAN-PE1	AAN	5'-CGT GGC G*CC ATC GCG AGG TC-3'	20mer	6317.7	6316.4
25	AAN-PE2	AAN	5'-CGT GAT G*CC ATC GCG AGG TC-3'	20mer	6318.7	6317.6
26	AAN-C1	AAN	5'-TCT CTC G*CT CAT CCAC-3'	16mer	4911.1	4908.9
27	AAA-PE1	AAA	5'-CGT GGC G*CC ATC GCG AGG TC-3'	20mer	6368.1	6367.7
28	AAA-PE2	AAA	5'-CGT GAT G*CC ATC GCG AGG TC-3'	20mer	6367.1	6366.5
29	AAA-C1	AAA	5'-TCT CTC G*CT CAT CCAC-3'	16mer	4964.8	4964.9

30	AAP-PE1	AAP	5'-CGT GGC G*CC ATC GCG AGG TC-3'	20mer	6391.9	6391.3
31	AAP-PE2	AAP	5'-CGT GAT G*CC ATC GCG AGG TC-3'	20mer	6390.9	6390.5
32	AAP-C1	AAP	5'-TCT CTC G*CT CAT CCAC-3'	16mer	4987.3	4984.3
33	AAP-C15G	AAP	5'-TCT CTC G*GT CAC CAC-3'	15mer	4968.7	4723.6

Table 3. Sequence and MALDI-Tof mass values of the synthesized oligonucleotides.

AAB, AAN, AAF, AAA and AAP represent the corresponding C8-dG lesions: AAB-dG = *N*-(2'-deoxyguanosin-8-yl)-acetylaminobenzol, AAN-dG = *N*-(2'-deoxyguanosin-8-yl)-2-acetyl-aminonaphthalene, AAF-dG: *N*-(2'-deoxyguanosin-8-yl)-2-acetylaminofluorene, AAA-dG = *N*-(2'-deoxyguanosin-8-yl)-2-acetylaminanthracene, AAP-dG: *N*-(2'-deoxyguanosin-8-yl)-1-acetylaminopyrene. The strands can be found in the laboratory journal with the label of the column "description".

10.2.3 Preparation of AF-dG containing oligonucleotides

AF-dG containing oligonucleotides were prepared by incubating AAF-dG containing oligonucleotides with 1 M NaOH solution containing 0.25 M β -mercaptoethanol for 3 h at 37 °C according to a published procedure^[151] followed by HPLC purification.

10.2.4 DNA Quantification

DNA concentration was determined using a *Nanodrop* UV-spectrometer *ND-1000* (*Peqlab*). The concentration of nucleic acids was calculated as follows:

$$1.0 A_{260}\text{-unit} = 50 \text{ ng} \cdot \mu\text{L}^{-1} \text{ dsDNA} = 33 \text{ ng} \cdot \mu\text{L}^{-1} \text{ ssDNA} = 40 \text{ ng} \cdot \mu\text{L}^{-1} \text{ RNA}$$

For chemically synthesized oligonucleotides, the DNA concentration was calculated using the corresponding extinction coefficient for the oligonucleotide measured and its A_{260} absorption. In detail, 1.5 μL of the sample was pipetted directly onto the pedestal without previous dilution. The 260/280 nm ratio should be between 1.8 and 2.0 for uncontaminated DNA preparations.

10.2.5 Melting point measurements

Melting profiles were measured on a *Cary 100* UV-Vis and *Jasco V-650* spectrophotometer using quartz glass cuvettes with 1 cm path length. Samples were prepared as shown in Table 4.

1 μ M (lesion containing) oligonucleotide
1 μ M complementary strand
150 mM NaCl
10 mM Tris pH 7.4
935 μ L ddH ₂ O (total volume 1 mL)

Table 4. Pipetting scheme for melting point measurements of DNA duplexes

The samples were covered with 2 mm of dimethylpolysiloxane and tightly plugged. The measurements were repeated several times with independent sample preparations. First, the oligonucleotides were hybridized by slowly cooling the samples down from 85 °C to 0 °C. The melting profiles started with a denaturing run (0 °C to 85 °C) with a slope of 0.5 °C/min. At least two denaturing and two renaturing ramps were performed and averaged for evaluation of the melting point. (T_M = zero-crossing of second derivative of the 450 nm-background corrected change in hyperchromicity at 260 nm). For analysis of the data, the program *Origin (Microcal)* was used.

10.2.6 Enzymatic DNA digestion

To characterize DNA strands containing a bulky DNA lesion, the oligonucleotides were enzymatically digested and the digest products were analyzed via HPLC-MS. 1 nmol of DNA was dissolved in 50 μ l 1x *digest buffer* (20 mM Tris, 10 mM MgCl₂, 0.1M NaCl, pH 8.5). 5 μ l (0.1 U/ μ l) snake venom phosphodiesterase were added and the solution was incubated at 37°C for 3 h. Addition of 0.5 μ l (10 U/ μ l) calf intestine phosphatase and further incubation at 37°C for 15 min completed digestion. The sample was centrifuged (*Eppendorf MiniSpin* centrifuge, 13400 rpm, 20 min). The supernatant was removed and ddH₂O was added to a final volume of 100 μ l. 45 μ l were analyzed via HPLC using a digest HPLC buffer (2 mM ammonium acetate in ddH₂O (buffer A) and 2 mM ammonium acetate in 80 % MeCN (buffer B)) and a *MS Uptisphere 3HDO (Interchim)* column with a flow of 0.2 mL/min. Furthermore, 50 μ l of the digest were analyzed via HPLC-MS (section 10.1).

10.3 Biochemical materials

10.3.1 Buffers

If not further declared, all buffers, solutions and media were prepared with ddH₂O obtained from the *Millipore System*. Adjustment of the pH-value was done by addition of NaOH_(aq) or HCl_(aq).

Strep-tag buffers A and B:

Buffer A: 100 mM Tris, 150 mM NaCl, 1mM EDTA; pH 7.5

Buffer B: 0.5 mg/ml D-Desthiobiotin in buffer A

(+ 10 mM DTT for SeMet labeled proteins)

Heparin buffers A1 and B1 (Pol η):

50 mM Tris, 1 mM EDTA, 5 mM DTT and 5% (v/v) glycerol; pH 7.5

Buffer A: 100 mM NaCl

Buffer B: 800 mM NaCl

Heparin buffers A2 and B2 (XPA):

25 mM Tris-HCl, 1 mM DTT, 10 μ M ZnCl₂, 10 % glycerol, pH 7.5

Buffer A: 50 mM NaCl

Buffer B: 1 M NaCl

(+ 10 mM DTT for SeMet labeled proteins)

MonoQ buffer

50 mM Tris, 5 mM β -mercapthoethanol, 10 μ M ZnCl₂, 10 % glycerol, pH 7.5

Buffer A: 100 mM NaCl

Buffer B: 600 mM NaCl

(+ 10 mM DTT for SeMet labeled proteins)

Crystallization buffer for Pol η (1x):

20 mM Tris, 50 mM KCl, 10 mM β -mercaptoethanol and 10% (v/v) glycerol; pH 7.5

Crystallization buffer for Pol κ (1x):

20 mM Tris, 50 mM KCl, 10 mM β -mercaptoethanol and 10% (v/v) glycerol; pH 7.5

General primer extension buffer (1x):

10 mM Tris, 50 mM NaCl, 10 mM MgCl₂, 1 mM DTT, pH 7.9

Pol κ primer extension buffer (1x):

25 mM Tris-HCl, 5 mM MgCl₂, 0.05 mg/ml BSA, 1 mM DTT, 10 % glycerol, pH 7.5

RP-HPLC buffers A and B:

Buffer A: 0.1 M NH₄Et₃OAc_(aq)

Buffer B: 0.1 M NH₄Et₃OAc in 80% MeCN_(aq)

HPLC buffers A and B for enzymatic digest

Buffer A: 2 mM NH₄CHOO

Buffer B: 2 mM NH₄CHOO in 80% MeCN

Agarose gel loading buffer:

50% (v/v) glycerol, 0.2% (w/v) SDS, 0.05% (w/v) bromophenol blue, 0.05% (w/v)

Xylencyanol FF in TAE- buffer

2 x PAA-denaturing loading buffer:

12% (v/v) Ficoll, 0.01% (w/v) Bromophenol blue, 0.02% (w/v) Xylencyanol FF, 7 M urea in

TBE- buffer

SDS-loading buffer:

62.5 mM Tris, 4% (w/v) SDS, 20% (w/v) Glycerol, 5% (w/v) β-mercaptoethanol, pH 6.8

SDS buffer (10x):

250 mM Tris, 1.92 M Glycerol, 1% (w/v) SDS

TE- buffer:

10 mM Tris-HCl, 1 mM EDTA; pH 8.0

TAE-running buffer (agarose gels): 40 mM Tris, 1 mM EDTA, pH 7.5

TAE-Gelpuffer (agarose gels):

20 mM Tris, 0.5 mM EDTA, pH 7.5

10 x TBE buffer:

89 mM Tris, 89 mM boric acid, 20 mM EDTA; pH 8.0

resolving gel buffer (SDS-gels):

3.0 M Tris, pH 8.8

stacking gel buffer (SDS-gels):

0.5 M Tris, pH 6.8

digest buffer 10x:

200 mM Tris-HCl, 100 mM MgCl₂, 1M NaCl, pH 8

10.3.2 Equipment

Equipment	Supplier
<i>Äkta purifier chromatography system</i>	GE, Munich
Agarose gel electrophoresis chamber	Biorad, Munich
Autoclave <i>Vakulab S3000</i>	<i>Systec</i> , Gießen
<i>Biofuge pico</i>	<i>Heraeus</i> , Hanau
<i>BioPhotometer 6131</i>	<i>Eppendorf</i> , Hamburg
<i>Centrifuge 5810R</i>	<i>Eppendorf</i> , Hamburg
French pressure cell press	<i>Thermo</i> , Dreieich
Gel imaging system <i>LAS 3000</i>	<i>Raytest</i> , Straubenhardt
Incubator <i>IS</i>	<i>Noctua</i> , Wiesloch
Incubator <i>44R</i>	<i>New Brunswick</i> , Wesseling-Berzdorf
<i>Mini Protean 3 Cell</i>	<i>Biorad</i> , Munich
Multicaster	<i>Biorad</i> , Munich
<i>Nanodrop UV-spectrometer ND-1000</i>	<i>Peqlab</i> , Erlangen
pH meter <i>MP220</i>	<i>Mettler Toledo</i> , Gießen
<i>Sorvall</i> centrifuge, Evolution RC	<i>Kendro</i> , Dreieich
Thermomixer Comfort	<i>Eppendorf</i> , Hamburg
Gel chamber Protean II xi Cell	<i>Biorad</i> , Munich
<i>IDA</i> gelscanner	<i>Raytest</i> , Straubenhardt
<i>Mastercycler Personal</i>	<i>Eppendorf</i> , Hamburg
<i>PCR Realplex</i>	<i>Eppendorf</i> , Hamburg
Deep freezer Vip Series -86 °C	<i>Sanyo</i> , Bad Nenndorf
Desktop centrifuge <i>5415R</i>	<i>Eppendorf</i> , Hamburg
<i>Carey UV-Spectrometer Bio100</i>	<i>Varian</i> , Darmstadt
High pressure homogenizer	<i>Avestin</i> , Mannheim
Electrotransformator, <i>Micropuls</i>	<i>Biorad</i> , München
Crystallization robot <i>Hydra II</i>	<i>Thermo</i> , Dreieich
Vortexer	<i>VWR</i> , Darmstadt
Water bath	<i>Labora</i> , Mannheim
Lyophilizer <i>Alpha 2-4 LD plus</i>	<i>Christ</i> , Osterode am Harz
MALDI-TOF Autoflex II-Spectrometer	<i>Bruker Daltonics</i> , Bremen

<i>Expedite 8900 Nucleic Acid Synthesis System</i>	<i>PerSeptive Biosystems, Framingham, MA, USA</i>
<i>Äkta Oligopilot 10</i>	<i>GE, Munich</i>
HPLC System	<i>Waters, Eschborn</i>
Speedvac Cryostat <i>CT 04-50 SR</i>	<i>Christ, Osterode am Harz</i>
Speedvac <i>RVC 2-33 IR</i>	<i>Christ, Osterode am Harz</i>
Speedvac Plus	<i>Thermo Life Sciences, Dreieich</i>
<i>Waters Millipore System</i>	<i>Millipore, Schwalbach</i>

10.3.3 Bacterial strains

Strain	Genotype	Supplier
TOP10	F- <i>mcrA</i> Δ (<i>mrr-hsdRMS-mcrBC</i>) ϕ 80 <i>lacZ</i> Δ M15 Δ <i>lacX74 recA1 araD139 galU galK</i> Δ (<i>ara-leu</i>)7697 <i>rpsL</i> (StrR) <i>endA1 nupG</i>	<i>Invitrogen, Karlsruhe</i>
BL21 (DE3)	F- <i>ompT hsdS_B</i> (<i>r_B⁻ m_B⁻) <i>gal dcm</i> (DE3)</i>	<i>Novagene, Schwalbach</i>
BL21(DE3)pLysS	F ⁻ , <i>ompT, hsdS_B</i> (<i>r_B⁻ m_B⁻), gal, dcm, (DE3), pLysS, (Cam^R)</i>	<i>Novagene, Schwalbach</i>
B834 (DE3)pLysS	F- <i>ompT hsdS_B</i> (<i>r_B⁻ m_B⁻) <i>gal dcm met</i> (DE3), pLysS, (Cam^R)</i>	<i>Novagene, Schwalbach</i>
DH5 α	F- ϕ 80 <i>lacZ</i> Δ M15 Δ (<i>lacZYA-argF</i>) U169 <i>recA1 endA1 hsdR17</i> (<i>r_k⁻, m_k⁺</i>) <i>phoA supE44</i> λ <i>thi1 gyrA96 relA1</i>	<i>Invitrogen, Karlsruhe</i>
Rosetta™ (DE3)	F-, <i>ompT, hsdSB</i> (<i>r_B⁻ m_B⁻) <i>gal, dcm, (DE3), pRARE, (Cam^R)</i></i>	<i>Novagene, Schwalbach</i>

10.3.4 Media

LB-Medium:

1% (w/v) peptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl; pH 7.5

SOC-Medium:

0.5% (w/v) peptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl; pH 7.2
after autoclaving: 5.0 mM MgCl₂, 5.0 mM MgSO₄

Medium-agar: Medium with 1.5% (w/v) agar

Autoinduction medium: *Overnight Express*TM Instant TB Medium (*Novagen*)

LeMaster medium^[164] (*selenomethionine-containing*):

For the composition of the medium see Appendix 11.2

10.3.5 Solutions

MALDI matrices:

HPA: 3-hydroxypicolinic acid (50 mg), 15-Crown-5 (10 μ l), ammonium hydrogencitrate (10 mg) in ddH₂O

HPA Crown: 3-hydroxypicolinic acid (50 mg), 15-Crown-5 (10 μ l), ammonium hydrogencitrate (10 mg) in ddH₂O (1ml)

ATT: 6-aza-2-thiothymine (10 mg) in 1 mL H₂O

Staining solution for SDS-gels:

0.25% (v/v) Coomassie Brilliant Blue R 250, 10% (v/v) acetic acid, 40% (v/v) ethanol in ddH₂O

Destaining solution for SDS-gels:

30% (v/v) ethanol, 10% (v/v) acetic acid in ddH₂O

Rotiphorese sequencing gel concentrate:

25% (w/v) acrylamide, 1.32% (w/v) methylenbisacrylamide, 50% (w/v) urea

Rotiphorese Gel 30:

30% (w/v) acrylamide, 0.8% (w/v) methylenbisacrylamide

10.3.6 Antibiotics and Inducers

Anhydrotetracycline 2 mg/ml in DMF

Carbenicillin 100 mg/ml in ddH₂O

Chloramphenicol 34 mg/ml in ethanol

IPTG 238 mg/ml isopropyl- β -D-thiogalactopyranoside in ddH₂O

10.3.7 Enzymes

Snake venom phosphodiesterase I (*Crotalus adamanteus venom*) was obtained from *USB corporation*, calf intestine phosphatase (CIP), XbaI and HindIII from *New England BioLabs*. Klenow fragment exo^- and exo^+ was also obtained from *New England BioLabs*.

10.4 Biochemical methods

10.4.1 Agarose gel electrophoresis

Plasmid DNA was analyzed on 0.8 or 1 % TAE agarose gels. These were run with a horizontal cell (Sub-Cell *Bio-Rad*) at 100 V. The DNA was stained with ethidium bromide. The bands were visualized using a *Raytest IDA* gelscanner.

Gel buffer: tris base (40 mM, pH 7.5), EDTA (1 mM), pH adjusted with sodium acetate

Running buffer: tris base (20 mM, pH 7.5), EDTA (0.5 mM),

pH adjusted with sodium acetate

staining: 1 mg/mL EtBr (0.1 μL EtBr/1 mL agarose gel)

DNA-ladder: 1kB DNA ladder, *New England BioLabs*

10.4.2 Polyacrylamide gel electrophoresis

Denaturing PAGE

Denaturing polyacrylamide gel electrophoresis (PAGE) was used for the analytical separation of short DNA fragments of the primer extension experiments (section 10.4.11). A 20% polyacrylamide gel was used in a *Protean II xi Cell (Biorad)* gel chamber allowing resolution of one base on the gel. The gels were run at 35 mA constant current with a maximum voltage of 800 V in 1 x TBE-buffer. The 20 % polyacrylamide gels were cast according to the table below and used directly after a polymerization time of 1 h.

	Rotiphorese sequencing gel concentrate	8.3 M Urea	10x TBE	10% APS	TEMED
20% PAA-gel	32 mL	4 mL	4 mL	200 μL	20 μL

Table 5. Composition of the 20 % denaturing polyacrylamide gels.

10 μ L of a primer extension experiment mixed with 2x PAA-loading buffer (section 10.3.1) was loaded onto the gel. A self-made DNA ladder consisting of fluorescein-labeled DNA fragments of sizes ranging from 13 to 20 bases was used. The fluorescein-labeled DNA fragments were detected with a *LAS-3000 gel imaging system* (Raytest).

SDS gel electrophoresis

SDS gel electrophoresis was used to check the size and the purity of a protein after purification. 0.1 – 5 μ g protein in 10 μ L buffer (storage buffer depending on the protein) were mixed with 5 μ L SDS-loading buffer. The sample was centrifuged shortly with a *Eppendorf Minispin* centrifuge and subsequently heated to 95 °C for 10 min to destroy secondary and tertiary structures. The composition of the gels can be found in Table 6.

	Rotiphorese Gel 30	ddH ₂ O	Tris (1.5 M, pH 8.8)	Tris (0.5 M, pH 6.8)	10% SDS	10% APS	TEMED
resolving gel 12%	8 mL	6.6 mL	5 mL	-	200 μ L	200 μ L	20 μ L
stacking gel 5%	1.3 mL	5.5 mL	-	1 mL	80 μ L	80 μ L	8 μ L

Table 6. Composition of the stacking and resolving gels for SDS gel electrophoresis

The gels were cast in a *Mini Protean 3 Cell* (Biorad). The resolving gel was covered with a layer of isopropanol in order to obtain a straight edge on the top. After a polymerization time of 1 h, the stacking gel was cast on top of the resolving gel. If not directly used, the gels were stored at 4°C (covered in wet tissues). 10 μ L sample volume and 4 μ L protein standard (*SeeBlue® Plus2 Pre-Stained Protein Standard*, Invitrogen) were loaded onto the gel. Gels were run at 120 V for approximately 1 h and further stained with a Coomassie solution for 10 min. Destaining was performed with a mixture of acetic acid and ethanol in water. Details of the composition of all buffers and solutions can be found in section 10.3.1 and 10.3.5.

10.4.3 Polymerase chain reaction (PCR)

The catalytic domain of the gene coding for the *human* DNA polymerase κ was amplified by PCR. A plasmid containing the complete Pol κ gene was purchased from *Origene* (USA). The following primer were used to amplify DNA regions coding for residues 50-530 and 1-530 of Pol κ .

Pol κ_S (50-530)

Forward primer 5'-Pho-AAT GGG AAA TGA GCT CAA GAA AGA AAA GC-3'

Reverse primer 5'-Pho-TCC CAA TAA TGC TCC TTT GTT GGT GTT TC-3'

Pol κ_L (1-530)

Forward primer 5'-Pho-AAT GGA TAG CAC AAA GGA GAA GTG TGA G-3'

Reverse primer 5'-Pho-TCC CCA TAA TGC TCC TTT GTT GGT GTT TG-3'

The PCR was performed using a temperature gradient and a constant DNA template concentration.

Mastermix (11x):	44 μ L	5x buffer (<i>Phusion buffer</i>)
	11 μ L	dNTP's (10 mM, each)
	11 μ L	Primer sense (10 mM)
	11 μ L	Primer antisense (10 mM)
	6.6 μ L	DMSO (100 %)
	2.2 μ L	<i>Phusion</i> (2 U/ μ L)
	= 85.8 μ L	

The 11 samples were prepared as described below and added onto a 96 well plate.

1 μ L	cDNA template (10 ng/ μ L)
11.2 μ L	ddH ₂ O
7.8 μ L	Mastermix

PCR - program (performed on a *PCR Realplex* device, *Eppendorf*):

1.	98 °C	5 min

2.	98 °C	10 s
3.	53-55 °C	10 s temperature gradient (sample 1-10)
4.	72 °C	1 min
5.	Goto 2.	30x

6.	72°C	7 min
7.	4°C	hold

The PCR-products were analyzed on a 1% agarose gels, the corresponding bands were cut out from the gel and isolated with a *MiniElute-Gel-Extraction-Kit* (*Quiagen*).

10.4.4 Assembly of the expression vectors for Pol κ

Entry reaction

The PCR products of both Pol κ versions (chapter 10.4.3) were cloned into the *pENTRY-IBA 10* donor vector using *StarGate*[®] cloning technology (*IBA*) according to the protocol from *IBA* and were subsequently sequenced (performed by *Dr. Melanie Maul*).

Transfer reaction

After sequence confirmation, the Pol κ gene was subcloned into several acceptor vectors according to the *StarGate*[®] protocol. The following destination vectors containing different tags for protein purification were assembled:

Pol κ_S (50-530) in pPSG-IBA3 (C-terminal *Strep*-tag II, T7 promotor)

Pol κ_L (1-530) in pPSG-IBA3 (C-terminal *Strep*-tag II, T7 promotor)

Pol κ_L (1-530) in pPSG-IBA3 (C-terminal *Strep*-tag II, T7 promotor)

Pol κ_L (1-530) in pPSG-IBA33 (C-terminal 6x Histidine (His)-tag, T7 promotor)

The resulting destination vectors were then transformed into the corresponding host cells (*E. coli* Top 10 or Rosetta) for further experiments (section 10.4.5).

Verification of the transfer reaction by enzymatic restriction digestion

100 ng plasmid DNA (pPSG-IBA3 and pPSG-IBA33 containing the Pol κ gene) were digested with the restriction enzymes *XbaI* and *HindIII* (*New England BioLabs*).

Assay: 100 ng DNA
 1 μ L 10 x NEB2 buffer (New England BioLabs)
 0.1 μ L BSA (10 mg/ml)
 0.2 μ L *XbaI* (20 U/ μ L)
 0.2 μ L *HindIII* (20 U/ μ L)
 filled up to 10 μ l with ddH₂O

The sample was incubated at 37 °C for 1.5 h and then analyzed via agarose gel electrophoresis.

10.4.5 Transformation of *E. coli*

Electrocompetent *E. coli*

An electroporation cuvette was cooled on ice prior to usage. 1 mL SOC- or LB-medium was heated to 37 °C. The competent cells (stored at -80 °C) were thawed on ice. 10 ng to 100 ng DNA was added (maximum volume 2 µL) and incubated for 1 min on ice. The sample was transferred into the electroporation cuvette. Afterwards, a voltage pulse of 2.5 V was applied to the cells, they were directly transferred into the preheated medium and incubated at 37 °C for 1 h (shaking). The cells were pelleted at 4000 rpm for 3 min in a desktop centrifuge *5415 R (Eppendorf)* and plated on LB agar containing the corresponding antibiotics. Cells were grown over night at 37 °C and stored at 4 °C.

Chemically competent *E. coli*

One aliquot chemically competent cells was thawed on ice and 10 ng to 100 ng DNA was added (maximum volume 2 µL). The sample was incubated on ice for 30 min and then heat shocked at 42 °C for 30 min. Next, the sample was incubated on ice for 2 min. 250 µL SOC-medium was added and the cells were incubated at 37 °C for 1 h (shaking). The cells were pelleted at 4000 rpm for 3 min in a *5415 R* desktop centrifuge (*Eppendorf*) and plated on LB agar containing the corresponding antibiotics. Cells were grown over night at 37 °C and stored at 4 °C.

10.4.6 *E. coli* cultures

E. coli cultures were stored as glycerol-stocks at -80 °C or on agar plates. Cells from agar plates were transferred to fresh plates every second or third week. Liquid cultures were grown in a shaker incubator (*New Brunswick*) at 200 rpm and 37 °C.

10.4.7 Isolation of plasmid DNA

Plasmid-DNA was isolated with a *QIAprep Spin Miniprep Kit* from *Qiagen*. Plasmid constructs obtained from the *StarGate®* cloning (section 10.4.4) were amplified in *E. coli* Top10. Therefore, 5 mL LB-medium containing the corresponding antibiotics was inoculated with a clone and incubated at 37 °C and 200 rpm over night. The cells were pelleted at 4000 rpm for 10 min in a *5415 R* desktop centrifuge (*Eppendorf*). The *QIAprep Spin Miniprep Kit* protocol was then used to isolate the plasmid DNA. Isolated plasmid DNA was stored at -20 °C.

10.4.8 Induction of the protein expression

Pol η

Previously, the DNA sequence coding for residues 1-513 of the yeast DNA Polymerase η from *Saccharomyces cerevisiae* was PCR amplified (forward primer 5' GGG GAC AAG TTT GTA CAA AAA AGC AGG CTT CAT GTC AAA ATT TAC TTG GAA GGA G-3', reverse primer 5'-GGG GAC GAC TTT GTA CAA GAA AGC TGG GTC TCA TTT TTG TAA ATC TAT AAT ATC GAA ATT AG-3') and cloned in a two-step Gateway® reaction via the vector pDONR201 (*Invitrogen*) into the expression vector pDEST007, in frame with an N-terminal Strep-Tag II (IBA) (performed by *Dr. Claudia Chiocchini*). The plasmid containing the Pol η gene was transformed into *E. coli* Rosetta.

60 mL LB medium containing 100 $\mu\text{g}/\text{mL}$ carbenicillin and 5 $\mu\text{g}/\text{mL}$ chloramphenicol were inoculated with *E. coli* Rosetta cells containing the pDEST007-Pol η plasmid from the glycerol stock. Cells were grown overnight in 250 ml Erlenmeyer flasks at 250 rpm and 37°C. 10 mL of this culture were added to 2 l Erlenmeyer flasks containing 1 L LB medium supplemented with 100 $\mu\text{g}/\text{mL}$ carbenicillin and 5 $\mu\text{g}/\text{mL}$ chloramphenicol. Cells were grown at 37 °C until an $\text{OD}_{600}=0.8$ was reached. Expression was induced by adding anhydrotetracycline to a final concentration of 200 ng/ml and the incubation was continued at 16 °C for 4–5 h. Cells were harvested by centrifugation in a *Sorvall SLA-3000* rotor (10 min, 8000 rpm) and, if not directly lysed, stored at -20 °C.

Pol κ

The plasmids pPSG-IBA3-Pol κ_{L} and pPSG-IBA3-Pol κ_{S} were transformed into *E. coli* BL21(DE3)pLysS.

LB Medium: Human Pol κ_{S} and Pol κ_{L} were expressed according to the protocol described for Pol η differing only in the induction of the protein expression. The Pol κ gene is under control of the T7 promotor, therefore, IPTG to a final concentration of 1 mM was used for induction.

Autoinduction medium: Human Pol κ_{L} (pPSG-IBA3-Pol κ_{L}) was also expressed using the *Overnight Express*™ Instant TB Medium from *Novagen* which resulted in much higher yields. 15 mL LB medium containing 100 $\mu\text{g}/\text{mL}$ carbenicillin and 5 $\mu\text{g}/\text{mL}$ chloramphenicol were inoculated with BL21 *E. coli* cells containing the pPSG-IBA3-Pol κ_{long} plasmid at 37 °C and 225 rpm overnight. 10 mL of the overnight culture were added to 1 L of *Overnight Express*™ Instant TB medium containing 100 $\mu\text{g}/\text{mL}$ carbenicillin and 5 $\mu\text{g}/\text{mL}$

chloramphenicol. The 1 L culture was split into 2 x 2 L Erlenmeyer flasks and incubated at 37°C and 225 rpm. After 4 h, the temperature was reduced to 25 °C and incubated for further 20 h. Cells were harvested by centrifugation in a *Sorvall SLA-3000* rotor (10 min, 8000 rpm) and, if not directly lysed, stored at -20 °C.

Rad14

Previously, The DNA sequence coding for the DNA binding domain (residues 185-306) of the yeast Rad14 from *Saccharomyces cerevisiae* was PCR amplified (forward primer 5' Pho-AAT GGC GCC GAA ATG TAT TGA ATG T-3', reverse primer 5'-Pho-TCC CGT CTT TTT TCT CCT TCT GTG-3') and cloned in a two-step StarGate® reaction via the vector pENTRY-IBA10 (*IBA*) into the expression vector pPSG-IBA3 (*IBA*), in frame with an C-terminal Strep-Tag II. The plasmid containing the Rad14 gene was transformed into *E. coli* BL21 (DE3). This work was performed by *Dr. Ralf Strasser*.^[139]

60 mL LB medium containing 100 µg/mL carbenicillin was inoculated with *E. coli* BL 21 cells containing the pPSG-IBA3-Rad14 plasmid from a glycerol stock. Cells were grown overnight in 250 ml Erlenmeyer flasks at 250 rpm and 37°C.

10 mL of this culture were added to 2 l Erlenmeyer flasks containing 1 L LB medium supplemented with 100 µg/mL carbenicillin. Cells were grown at 37 °C until an OD₆₀₀=0.8 was reached. Expression was induced by adding IPTG to a final concentration of 1 mM. Additionally, 10 µL ZnCl_{2(aq)} (1M) were added and the incubation was continued at 28 °C for 3 h. Cells were harvested by centrifugation in a *Sorvall SLA-3000* rotor (10 min, 8000 rpm) and, if not directly lysed, stored at -20 °C.

XPA

Previously, The DNA sequence coding for the DNA binding domain (residues 98-219) of the human XPA protein was PCR amplified (forward primer 5'-AAT GGA ATT TGA TTA TGT AAT ATG CG-3'; reverse primer 5'-TCC CAA ATT TCT TCT GTT TCA TTT TTT CTC-3') and cloned in a two-step StarGate® reaction via the vector pENTRY-IBA10 (*IBA*) into the expression vector pPSG-IBA3 (*IBA*), in frame with an C-terminal Strep-Tag II. The plasmid containing the XPA gene was transformed into *E. coli* BL21 (DE3) pLysS. This work was performed by *Sandra Koch*. XPA was expressed according to the protocol described for Rad14.

Selenomethionine labeled Rad14 and XPA

For crystallographic phase determination via selenomethionine labeling, the plasmid containing the gene for either the Rad14 or the XPA protein, was transformed into the methionine auxotrophic *E.coli* strain B834(DE3)pLysS. Expression was performed in selenomethionine-containing LeMaster medium (section 11.2). Therefore, 60 mL LB medium containing 100 µg/mL carbenicillin and 5 µg/mL chloramphenicol was inoculated with *E. coli* B834 cells containing the pPSG-IBA3-Rad14 or XPA plasmid. Cells were grown overnight in 250 ml Erlenmeyer flasks at 250 rpm and 37°C. 5 mL of this culture were added to 2 l Erlenmeyer flasks containing 1 L LeMaster medium supplemented with 50mg/L DL-selenomethionine, 100 µg/mL carbenicillin and 5 µg/mL chloramphenicol. Cells were grown at 37 °C until an OD₆₀₀=0.6 was reached. Due to the minimum medium, cells were grown for 8-10 h until the desired density was reached. Expression was induced by adding IPTG to a final concentration of 1 mM. Additionally, 10 µL ZnCl_{2(aq)} (1M) were added and the incubation was continued at 20 °C overnight. Cells were harvested by centrifugation in a *Sorvall SLA-3000* rotor (10 min, 8000 rpm) and, if not directly lysed, stored at -20 °C.

10.4.9 Protein purification

Pol η

The cell pellet (section 10.4.8) was resuspended in cold Strep-tag buffer A with a cocktail of protease inhibitors (*Roche*) added, and lysed using a high pressure homogenizer (*Avestin Europe GmbH*). The cell lysate was cleared by centrifugation in a *Sorvall SS-34* rotor at 18000 rpm and 4°C for 30 min, filtered through a 0.2 µm filter and applied onto a Strep-Tactin column (*IBA*, 5 ml column volume) at 4°C. The protein was eluted from the column with the Strep-tag buffer B containing 2.5 mM desthiobiotin (flow 1 ml/min). The fractions were pooled and concentrated in 50 mM Tris-HCl pH 7.6, 100 mM NaCl, 1 mM EDTA, 5 mM DTT, 5% glycerol (Heparin buffer A1) and loaded onto a HiTrap Heparin column (*GE Healthcare*, 1 ml column volume). The protein was eluted with a linear gradient of the same buffer containing 800 mM NaCl (Heparin buffer B1, flow 1 ml/min). The buffer was exchanged to 20 mM Tris-HCl pH 7.6, 50 mM KCl, 10 mM β-mercaptoethanol, 10% glycerol (crystallization buffer for Pol η) using an *Amicon Ultra* centrifuge filter device (30,000 MWCO) (*Millipore*) and an *Eppendorf 5810R* centrifuge. The enzyme was stored at a concentration of 10 mg/ml at -80 °C.

Pol κ

The protocol used for the purification of Pol κ was similar to the protocol used for the Pol η purification. Both Pol κ versions, Pol κ_S and Pol κ_L were purified as C-terminally tagged *Strep-Tag II* fusion proteins using a Streptactin affinity column (*IBA*) and applying the Pol η protocol. This was followed by a heparin affinity column (for details see Pol η purification). After purification, the enzyme was stored in 20 mM Tris-HCl pH 7.6, 50 mM KCl, 10 mM β -mercaptoethanol, 10% glycerol at -80 °C.

Rad14

If selenomethionine-labeled Rad 14 was purified, 10 mM DTT was added to all buffers and all buffers were degassed prior to usage to prevent oxidation of selenomethionine.

The cell pellet (section 10.4.8) was resuspended in cold Strep-tag buffer A with a cocktail of protease inhibitors (*Roche*) added, and lysed using a high pressure homogenizer (*Avestin Europe GmbH*). The cell lysate was cleared by centrifugation in a *Sorvall SS-34* rotor at 18000 rpm and 4°C for 30 min, filtered through a 0.2 μ m filter and applied onto a Strep-Tactin column (*IBA*, 5 ml column volume) at 4°C. The protein was eluted from the column with the Strep-tag buffer B containing 2.5 mM desthiobiotin (flow 1 ml/min). The fractions were pooled and concentrated in 50 mM Tris, 5 mM β -mercapthoethanol, 10 μ M ZnCl₂, 100 mM NaCl, 10 % glycerol, pH 7.5 (MonoQ buffer A) and loaded onto a *MonoQ 5/50GL* column (*GE Healthcare*). The protein was eluted with a linear gradient of the same buffer containing 600 mM NaCl (MonoQ buffer B, flow 0.4 ml/min). The buffer was exchanged to 50 mM Tris, 5 mM β -mercapthoethanol, 10 μ M ZnCl₂, 100 mM NaCl, 10 % glycerol, pH 7.5 using an *Amicon Ultra* centrifuge filter device (10,000 MWCO) (*Millipore*) and an *Eppendorf 5810R* centrifuge. The enzyme was stored at a concentration of 10 mg/ml at -80 °C.

XPA

If selenomethionine-labeled XPA was purified, 10 mM DTT was added to all buffers and all buffers were degassed prior to usage to prevent oxidation of selenomethionine.

Cell lysis and Strep-tag purification was performed according to the protocol described for the Rad 14 purification. Subsequent to elution from the Strep-tag column, the fractions were pooled and concentrated in 25 mM Tris-HCl, 1 mM DTT, 50 mM NaCl, 10 μ M ZnCl₂, 10 % glycerol, pH 7.5 (Heparin buffer A2) and loaded onto a *HiTrap* Heparin column (*GE Healthcare*, 1 ml column volume). The protein was eluted with a linear gradient of the same buffer containing 1 M NaCl (Heparin buffer B2, flow 0.4 ml/min). The buffer was exchanged

to 25 mM Tris-HCl, 50 mM NaCl, 1 mM DTT, 10 μ M ZnCl₂, 10 % glycerol, pH 7.5 using an *Amicon Ultra* centrifuge filter device (10,000 MWCO) (*Millipore*) and an *Eppendorf 5810R* centrifuge. The enzyme was stored at a concentration of 10 mg/ml at -80 °C.

10.4.10 Protein quantification

Absorption at 280 nm (A_{280}): In most cases, the protein concentration was measured by the absorption at 280 nm using the *Nanodrop* UV-spectrometer *ND-1000* (*Peqlab*). Aromatic amino acids have their absorption maxima at this wavelength. Statistically, one absorption unit at 280 nm corresponds to a protein concentration of 1 mg/mL. The exact protein concentration was determined using the individual molar extinction coefficient of the protein (calculated from the amino acid sequence). For the consideration of nucleic acid contamination, the Warburg equation is used:

$$\text{protein concentration} = (1.55 \times A_{280}) - (0.76 \times A_{260})$$

Bradford-Assay: The Bradford solution (*Biorad*) was diluted 1:5 with ddH₂O. For determination of the reference value, 950 μ L Bradford solution were mixed with 50 μ L of the protein buffer. For the sample, 950 μ L Bradford solution, 1-5 μ L protein solution (depending on the concentration) and buffer were mixed to a total volume of 1 mL. Protein concentration was measured by the absorption at 595 nm using a BSA calibration curve.

10.4.11 Primer extension studies

A fluorescent labeled 13 or 14mer primer (*Metabion*, Germany) was annealed (95 °C for 4 min followed by cooling down to 4 °C over a period of 45 min) to the unlabeled 20mer template DNA strand, either undamaged or containing the bulky DNA lesion in a 1:1.5 molar ratio in the Pol η reaction buffer (10 mM Tris-HCl, 50 mM NaCl, 10 mM MgCl₂, 1 mM DTT, pH 7.9. For Pol κ , the following buffer was used: 25 mM Tris-HCl, 5 mM MgCl₂, 0.05 mg/ml BSA, 1 mM DTT, 10 % glycerol, pH 7.5). The final concentration of the primer in the primer extension studies was 1 μ M. Primer extensions studies using 50 – 300 nM Pol η or 100-500 nM Pol κ were initiated by addition of one or all four nucleotides (final concentration 200 μ M) and incubated at 30 °C. After 5 min, one half of the assay volume were transferred into a fresh tube and a solution sodium EDTA (100 mM, 1:1 volume ratio) was added to terminate these reactions. After 30 min, reactions of remaining samples were also terminated by addition of a solution sodium EDTA. The primer extension products were resolved on 22%

polyacrylamide-urea gels and visualized using a *LAS-3000* imaging system (*Raytest*). Table 7 describes the composition of a typical primer extension assay in detail:

1 μL DNA (primer/template, 10 μM)
1 μL dNTP/dNTPs (2 mM)
1 μL 10x primer extension buffer (see 10.3.1)
x μL DNA polymerase
y μL ddH ₂ O (total volume 10 μL)

Table 7. Pipetting scheme for primer extension assays

10.4.12 Mass spectrometry analysis of the primer extension experiments

Three primer extension assays (3x 10 μL , containing 1 μM template and extended primer DNA strands, see section 10.4.11) were combined. 20 μL ddH₂O were added and the DNA was extracted from the crude primer extension products via phenol/chloroform (*Roti-Phenol*, *Roth*, Germany) extraction. Therefore, 50 μL phenol were added and mixed by vortexing. The aqueous phase was transferred to a fresh tube and extracted twice with 2 x 50 μL chloroform. The aqueous phase was concentrated to 10 μL (Speedvac) and desalted via ZipTip_{C18} (*Millipore*). Therefore, the ZipTip_{C18} were equilibrated with 3 x 10 μL MeCN, washed with 3 x 10 μL ddH₂O and the sample was pumped through the ZipTip_{C18} 5 times. The tip was then washed with 10 x 10 μL ddH₂O and the DNA was eluted using 5 x 10 μL 50 % MeCN in ddH₂O. The desalted samples were concentrated to 1 μL (Speedvac). 0.4 μL of the sample were mixed with 0.4 μL HPA matrix (section 10.3.5). Mass spectra were recorded on a *Bruker Autoflex II* (MALDI-Tof).

10.5 Protein crystallization

Pol η

For co-crystallization, Pol η (4.5 mg/ml) was incubated with double stranded DNA containing the template with the AAF-dG, AAA-dG or the AAP-dG lesion annealed to primer with a 2',3'-dideoxybase at the 3' end (5'-GTG GAT GAG^{dd}) in a 1 to 4 molar ratio in crystallization buffer (20 mM Tris, 50 mM KCl, 10 mM β -mercaptoethanol and 10% (v/v) glycerol; pH 7.5). Crystals were grown by mixing an equal volume of protein-DNA complex with 11–15% (w/v) PEG 3350, 150–200 mM CaCl₂, 5 mM MgSO₄, 1 mM dCTP using the hanging-drop vapour diffusion method (drop size 1 μL + 1 μL). The crystallization plates

were incubated at 4 or 18 °C and crystals appeared after 10 to 60 days. Crystals were frozen in artificial mother liquor containing 20% (v/v) butanediol and stored in liquid nitrogen until data collection. Table 8 describes the composition of a typical crystallization experiment:

6 μ L DNA (primer/template, 1 nmol/ μ L)
1.25 μ L dCTP (20 mM)
0.5 μ L MgSO ₄ (250 mM)
x μ L Pol η (4.5 mg/ml final concentration)
y μ L crystallization buffer (total volume 25 μ L)

Table 8. Pipetting scheme for crystallization experiments with Pol η and bulky adduct containing DNA.

Pol κ

For co-crystallization, Pol κ_L (5 mg/ml) was incubated with double stranded DNA containing the template with the AAF-dG lesion annealed to primer with a 2',3'-dideoxybase at the 3' end in a 1 to 2 molar ratio in crystallization buffer (20 mM Tris, 50 mM KCl, 10 mM β -mercaptoethanol and 10% (v/v) glycerol; pH 7.5) adding 5 mM MgSO₄ and 1 mM dCTP. A crystallization screen was performed using the *Hydra II* crystallization robot (*Thermo*) with a drop size of 0.2 μ L protein and 0.2 μ L precipitant. The commercial available *PEG I* and *PEG II Suites* (*Qiagen*) were utilized. Additionally, a manual screening with 11 mg/ml Pol κ_{long} and an AAF-dG containing DNA duplex (1:1.3 molar ratio) was carried out in 24 well format using 11-14 % PEG 5000, 200 mM NaCl and 0.19 -0.22 M ammonium acetate at pH 6.2 and 6.5 (conditions previously published for the Pol κ in complex with undamaged DNA).^[162]

A second robot screen was done at the *Oxford Protein Production Facility* (OPPF, Oxford, UK). Pol κ_{long} was used at a concentration of 11 m/ml in 20 mM Tris pH 7.5, 200 mM NaCl, 1 mM dCTP and 5 mM MgSO₄ (1:1.3 molar ratio protein/DNA). The following commercially available screening kits from *Hampton Research* were used: *Hampton Crystal Screen 1 + 2*, *Hampton PEG/Ion Screen*, *Hampton Grid Screen PEG 6000*, *Hampton Grid Screen Ammonium Sulphate*, *Hampton Index*. The plates were incubated at 4 °C for 12 month. No diffracting crystals could be obtained in all cases.

Rad14

For co-crystallization, Rad14 (4 - 6 mg/ml) was incubated with double stranded DNA containing the template with the AAF-dG and a counter strand in a 1 to 1.2 molar ratio in crystallization buffer (50 mM Tris, 5 mM β -mercapthoethanol, 10 μ M ZnCl₂, 100 mM NaCl,

10 mM DTT, 10 % glycerol, pH 7.5). First native Rad14 crystals in complex with AAF-dG containing DNA were obtained with 2.8- 3.3 M ammonium sulfate in 0.08 – 0.11 M Tris pH 8 buffer (performed by *Dr. Ralf Strasser*).^[139] These crystals diffracted X-rays only to 7 Å spacing. Therefore, a new crystallization screen was performed with a *Phoenix nanodispenser robot* (*Art Robbins Instruments*, drop size: 0.1 µL protein + 0.1 µL precipitant) in the group of *E. Conti* at the *Max-Planck-Institute* (Martinsried). The following commercial available screening kits (*Qiagen*) were used: *Qiagen AmSO₄ Kit*, *Qiagen Classics Kit*, *Qiagen JCSG+ Kit*, *Qiagen PEGs* and *CP-PEGS-Salt screen (Pegs salt II) Kit*. After obtaining an initial hit with 0.2 M Calcium acetate, 0.1 M sodium cacodylate pH 6.5 and 18 % PEG 8000 at a protein concentration of 5.5 mg/ml and a 1 to 1.2 molar ration of protein to DNA (ODN **16**) at 4°C, the conditions of the initial hit were refined by varying the protein concentration, concentration of the components of the crystallization buffer and drop size in 24 well formats (hanging drop *Crystalgen SuperClear™ Plates*, *Jena Biosciences*). For selenomethionine-labeled Rad14 a concentration of 5 mg/ml and for crystallization with 5-Br-dU labeled DNA (*Metabion*), a concentration of 5.5 mg/ml was used. Crystals reached their maximum size after 1 week at 4 °C.

After a buffer screening at constant calcium acetate and PEG 8000 concentrations, Rad14 crystals with 0.1 M Tris pH 7 were obtained. Crystals were shock-frozen in cryoprotectant solution (see Table 9) and stored in liquid nitrogen (not longer than 1-2 weeks).

100 µM Ca acetate (1 M)
180 µL PEG 8000 (50 %)
50 µL Tris (1 M, pH 7)
10 µL DTT (100 mM)
125 µL ethylene glycol (25 %)
95 µL ddH ₂ O

Table 9. Composition of the cryoprotectant solution for Rad14 crystals.

XPA

For co-crystallization, XPA (12 mg/ml) was incubated with double stranded DNA containing the template with the AAF-dG (ODN **16**) and a counter strand in a 1 to 1.2 molar ratio in crystallization buffer (25 mM Tris-HCl, 1 mM DTT, 10 µM ZnCl₂, 50 mM NaCl, 10 % glycerol, pH 7.5).

First crystallization screening was performed with a *Phoenix nanodispenser robot* (*Art Robbins Instruments*, drop size: 0.1 μL protein + 0.1 μL precipitant) in the group of *E. Conti* at the *Max-Planck-Institute* (Martinsried). The following commercial available screening kits (*Qiagen*) were used: *Qiagen AmSO₄ Kit*, *Qiagen Classics Kit*, *Qiagen JCSG+ Kit*, *Qiagen PEGs* and *CP-PEGS-Salt screen (Pegs salt II) Kit*. The plates were incubated at 18 °C. A second *Qiagen Classics Kit* plate was also incubated at 4 °C. After obtaining an initial hit with 20 % PEG 3350 and 0.2 M MgSO_4 at 18 °C, conditions of the initial hit were refined in 24 well formats. For selenomethionine-labeled XPA, a concentration of 8-10 mg/ml was used. No large crystals could be obtained. Two small native crystals, frozen in artificial mother liquor containing 20% (v/v) ethylene glycol and measured at the PXI beam line at the Swiss Light Source (SLS, Villigen, Switzerland), did not diffract X-rays.

10.6 Collection and processing of X-ray diffraction data, phase determination and structure refinement

Pol η crystals

Diffraction data were measured at the PXI beam line at the Swiss Light Source (SLS), Villigen, Switzerland. The crystals diffracted X-rays to 2.7 Å spacing. Structure solution was performed by *Dr. Sabine Schneider*. The data were processed with the programs XDS and XSCALE^[165] and the crystals belong to the same space group, with comparable unit cell constants as the apo enzyme^[156] and co-crystals containing the cisplatin lesion^[90, 157]. The structures were solved by molecular replacement (PHASER^[166]) using the coordinates of the apo-enzyme as search model. In order to reduce model bias, the temperature factors were reset to the Wilson B-factor. Prior to model building in COOT^[167] a simulated annealing omit map, removing the area around the lesion, was calculated with PHENIX.^[168] Restrained refinement was carried out with PHENIX and REFMAC^[169], excluding the same set of free reflections as for the previously reported Pol η -DNA complex structures.^[90, 157] Data processing and refinement statistics are summarized in Table 10. Structural figures were prepared with PyMol (*Delano Scientific*). The atomic coordinates and structure factors (PDB codes 2XGP and 2XGQ) have been deposited in the PDB at the EBI Macromolecular Structure Database (<http://www.ebi.ac.uk/pdbe>). For details of the structure solution see section 11.1.

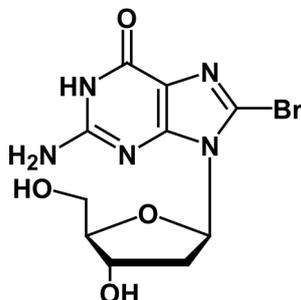
Rad14 crystals

Diffraction data were measured at the PXI beam line at the Swiss Light Source (SLS), Villigen, Switzerland and the microfocus beamline ID23-2 at *European Synchrotron Radiation Facility* (ESRF), Grenoble, France.

Single-wavelength anomalous dispersion (SAD) data at the seleno and bromine absorption edge for either selenomethionine-labeled Rad14 co-crystals with an AAF-dG containing DNA duplex or Rad14 crystals in complex with 5-Br-dU and an AAF-dG containing DNA duplex were collected. The peak and the inflection points of the anomalous scatterers were determined by an X-ray fluorescence scan on each crystal. Labeled Rad14 crystals diffracted X-rays to 3.5 Å spacing but were highly anisotropic. Additionally, a data set of a native Rad14 crystal in complex with AAF-dG with a resolution of 3.3 Å was collected.

10.7 Synthesis of bulky DNA adducts

10.7.1 8-Bromo-2'-deoxyguanosine 27



2'-deoxyguanosine monohydrate (10.00 g, 35.06 mmol) was suspended in a mixture of acetonitrile (370 mL) and water (95 mL). *N*-bromosuccinimide (9.36 g, 52.59 mmol, 1.5 eq.) was added in three portions over 15 min. The yellow reaction mixture was stirred for 40 min at room temperature. The solvents were removed *in vacuo* and the precipitate was suspended in acetone (200 mL), stirred for 2 hours at room temperature and cooled overnight at -20°C . The precipitate was collected by filtration, extensively washed with 200 mL cold acetone and dried under vacuum to provide (10.26 g, 29.65 mmol, 85 %, lit.: 80 %) of a slightly yellow powder.

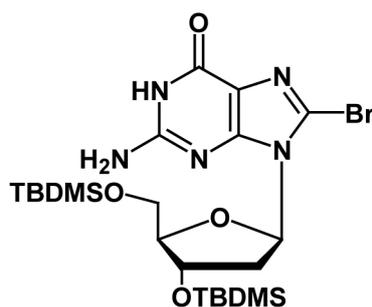
$R_f = 0.44$ ($\text{CH}_2\text{Cl}_2:\text{MeOH}$ 1:1)

$^1\text{H-NMR}$ (400 MHz, DMSO-d_6): δ (ppm) = 10.79 (s, 1H, $\text{N}^1\text{-H}$), 6.48 (s, 2H, NH_2), 6.13 (dd, $^2J = 7.8$, $^3J = 6.8$ Hz, 1H, $\text{C}_1\text{-H}$), 5.29 – 5.21 (m, 1H, 5' OH), 4.85 (br d, 1H, 3' OH), 4.37 (ddd, $^3J = 2.8$, 2.8, 5.9 Hz, 1H, $\text{C}_3\text{-H}$), 3.77 (ddd, $^3J = 3.0$, 5.6, 5.6 Hz, $\text{C}_4\text{-H}$), 3.65 – 3.55 (m, 1H, $\text{C}_5\text{-H}_a$), 3.5 – 3.42 (m, 1H, $\text{C}_5\text{-H}_b$), 3.14 (ddd, $^2J = 13.2$, $^3J = 6.4$, 8.0 Hz, 1H, $\text{C}_2\text{-H}_a$), 2.08 (m, 1H $\text{C}_2\text{-H}_b$).

$^{13}\text{C-NMR}$ (101 MHz, DMSO-d_6): δ (ppm) = 155.89, 153.76, 152.43, 121.03, 117.92, 88.33, 85.50, 71.46, 62.48, 36.86.

IR ν (cm^{-1}) 3354 (*s*), 3152 (*s*), 2952 (*w*), 2906 (*w*), 1677 (*s*), 1662 (*s*), 1605 (*s*), 1582 (*s*), 1514 (*w*), 1459 (*m*), 1362 (*m*), 1290 (*m*), 1272 (*m*), 1254 (*m*), 1111 (*w*), 1080 (*w*), 1063 (*m*), 1024 (*m*), 999 (*m*), 789 (*w*), 625 (*w*).

HR-MS (ESI^-) m/z for $[\text{C}_{10}\text{H}_{11}\text{N}_5\text{O}_4^{79}\text{Br}]^-$ ($[\text{M-H}]^-$): calc.: 343.9994, found: 343.9996.

10.7.2 8-Bromo-3',5'-*O*-bis(*tert*-butyldimethylsilyl)-2'-deoxyguanosine **28**

8-bromo-2'-deoxyguanosine **27** (8.5 g, 24.50 mmol) and imidazole (7.96 g, 166.92 mmol) were suspended in DMF (55 ml) and treated with a 1 M solution of *tert*-butyldimethylsilyl chloride in THF (67 ml, 67 mmol). The reaction mixture was stirred for 5h at room temperature and the solvents were removed *in vacuo*. 270 mL water were added and extracted with EtOAc (2 x 400 mL). The organic layer was dried over MgSO₄, concentrated *in vacuo* and purified by chromatography on silica gel (CH₂Cl₂:MeOH 20:1) to provide 12.81 g (22.30 mmol, 91 %, lit. 98 %) of a slightly yellow powder.

R_f = 0.43 (CH₂Cl₂:MeOH 10:1)

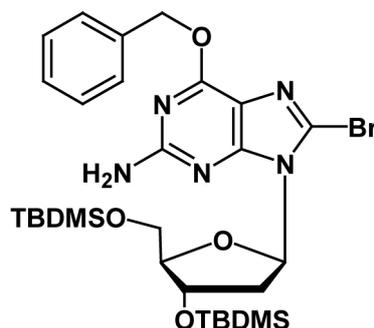
¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 10.83 (s, 1H, N¹-H), 6.41 (s, 2H, NH₂), 6.12 (dd, ³J = 7.1, 7.1 Hz, 1H, C₁'-H), 4.56 (ddd, ³J = 3.3, 3.3, 6.3 Hz, 1H, C₃'-H), 3.72 (m, 2H, C₅'-H_a+H_b), 3.63 (m, 1H, C₄'-H), 3.35 (ddd, ²J = 13.7, ³J = 6.7, 6.9 Hz, 1H, C₂'-H_a), 2.12 (ddd, ³J = 3.6, 7.0, ²J = 13.2, 1H, C₂'-H_b), 0.87 (s, 9H, *t*Bu-H), 0.80 (s, 9H, *t*Bu-H), 0.09 (s, 6H, 2xSi-CH₃), -0.03 (s, 3H, Si-CH₃), -0.05 (s, 3H, Si-CH₃).

¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 155.97, 153.80, 152.55, 121.10, 117.86, 87.57, 85.17, 72.84, 63.27, 36.44, 26.24, 26.18, 26.14, 18.42, 18.40, 18.16, -4.25, -4.43, -4.71, -4.98, -5.00.

IR ν (cm⁻¹) 3399 (*w*), 3308 (*m*), 3244 (*m*), 3199 (*m*), 2952 (*m*), 2929 (*m*), 2857 (*m*), 1691 (*s*), 1633 (*m*), 1608 (*m*), 1595 (*s*), 1557 (*m*), 1471 (*w*), 1463 (*w*), 1337 (*w*), 1285 (*w*), 1256 (*w*), 1084 (*m*), 1038 (*w*), 834 (*s*), 777 (*s*), 665 (*w*).

HR-MS (ESI⁻) *m/z* for [C₂₂H₃₉N₅O₄Si₂⁷⁹Br]⁻ ([M-H])⁻: calc.: 572.1724, found: 572.1721.

10.7.3 *O*⁶-Benzyl-8-bromo-3',5'-*O*-bis(*tert*-butyldimethylsilyl)-2'-deoxyguanosine 29



8-bromo-3',5'-*O*-bis(*tert*-butyldimethylsilyl)-2'-deoxyguanosine **28** (10.63 g, 18.50 mmol) and triphenylphosphine (7.30 g, 27.75 mmol, 1.5 eq.) were suspended in dioxane (220 ml). Benzylalcohol (2.94 mL, 27.75 mmol) and diisopropyl azodicarboxylate (4.38 mL, 27.75 mmol) were added and the reaction mixture was stirred for 7 h at room temperature under argon. After concentration, the crude product was purified by chromatography on silica gel (*iso*-hexane:EtOAc 20:1) to provide 5.59 g (8.40 mmol, 45 %, lit. 78 %) of a white foam.

$R_f = 0.52$ (*iso*-hexane:EtOAc 4:1)

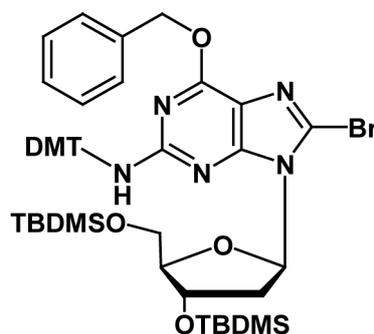
¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 7.47 – 7.45 (m, 2H, Bz-H), 7.41 – 7.26 (m, 3H, Bz-H), 6.45 (s, 2H, NH₂), 6.16 (dd, ³*J* = 6.9, 6.9 Hz, 1H, C₁-H), 5.45 (AB, ²*J* = 12.2, 2H, Bz-CH₂), 4.66 (ddd, ³*J* = 3.6, 3.7, 6.2 Hz, 1H, C₃-H), 3.75 (m, 2H, C₅-H_a+H_b), 3.61 (m, 1H, C₄-H), 3.50 (m, 1H, C₂'-H_a), 2.17 (ddd, ³*J* = 3.9, 7.0, ²*J* = 13.1, 1H, C₂'-H_b), 0.88 (s, 9H, tBu-H), 0.77 (s, 1H, tBu-H), 0.10 (s, 6H, 2xSi-CH₃), -0.07 (s, 3H, Si-CH₃), -0.09 (s, 3H, Si-CH₃).

¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 159.67, 159.42, 155.09, 136.78, 128.84, 128.44, 125.25, 114.87, 87.47, 85.43, 72.73, 66.79, 63.15, 36.17, 26.13, 17.85, -4.25, -4.27, -4.44, -5.05, -5.08.

IR ν (cm⁻¹) 3506 (*w*), 3354 (*br w*), 2954 (*s*), 2929 (*s*), 2885 (*m*), 2856 (*s*), 1710 (*br m*), 1612 (*s*), 1572 (*s*), 1520 (*w*), 1471 (*m*), 1455 (*m*), 1250 (*s*), 1217 (*m*), 1108 (*s*), 1028 (*m*), 833 (*s*), 776 (*s*), 696 (*m*).

HR-MS (ESI⁺) m/z for [C₂₉H₄₇N₅O₄Si₂⁷⁹Br]⁺ ([M+H]⁺): calc.: 664.2350, found: 664.2348.

10.7.4 *O*⁶-Benzyl-8-bromo-3',5'-*O*-bis(*tert*-butyldimethylsilyl)-*N*²-dimethoxytrityl-2'-deoxyguanosine 30



*O*⁶-Benzyl-8-bromo-3',5'-*O*-bis(*tert*-butyldimethylsilyl)-2'-deoxyguanosine 30 (5.39 g, 8.10 mmol) was dissolved in pyridine (85 ml). 4,4'-Dimethoxytrityl chloride (4.67 g, 13.78 mmol, 1.7 eq.) was added and the reaction mixture was stirred overnight at room temperature. The reaction was quenched with methanol (2 mL), the solvents were evaporated *in vacuo* and the crude product was purified by chromatography on silica gel (*iso*-hexane:EtOAc 15:1, 1 % pyridine) to yield 7.77 g (8.03 mmol, 99 %, lit. 94 %) of the product as a white foam.

$R_f=0.58$ (*iso*-hexanes:EtOAc 4:1)

¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 7.31 – 7.11 (m, 15H, Bz-5H + DMT-9H + *N*²-H), 6.82 – 6.77 (m, 4H, DMT-4H), 6.04 (m, 1H, C₁-H), 4.72 (br s, 2H, Bz-CH₂), 4.47 (m, 1H, C₃-H), 3.75 – 3.60 (m, 8H, 2xOMe + C₅-H_a+H_b), 3.52 (m, 2H, C₄-H + C₂-H_a), 2.06 (m, 1H, C₂-H_b), 0.85 (s, 9H, *t*Bu-H), 0.75 (s, 9H, *t*Bu-H), 0.08 (s, 6H, 2xSi-CH₃), -0.08 (s, 3H, Si-CH₃), -0.13 (s, 3H, Si-CH₃).

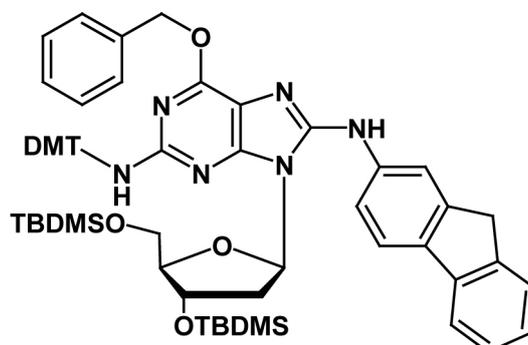
¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 158.07, 158.06, 157.37, 154.40, 146.18, 138.10, 138.07, 136.43, 130.12, 130.09, 128.81, 128.51, 128.48, 128.02, 126.81, 115.30, 113.31, 87.07, 85.87, 83.40, 80.91, 69.92, 67.34, 55.40, 35.47, 26.11, 26.08, 18.32, 18.20, 17.87, 15.07, -4.27, -4.42, -5.03, -5.17.

IR ν (cm⁻¹) 3347 (*w*), 2953 (*s*), 2929 (*s*), 2856 (*m*), 2710 (*w*), 2552 (*w*), 1608 (*s*), 1584 (*s*), 1509 (*s*), 1459 (*m*), 1442 (*m*), 1359 (*m*), 1249 (*s*), 1178 (*m*), 1078 (*w*), 1031 (*m*), 829 (*s*), 775 (*s*), 699 (*m*).

HR-MS (ESI⁺) m/z for [C₅₀H₆₃O₆N₅Si₂⁷⁹Br]⁺ ([M+H]⁺): calc.: 964.3500, found: 964.3492.

10.7.5 Coupling of the protected deoxyguanosine with the aromatic unit

10.7.5.1 8-(*N*-2-Amino-fluorene)-*O*⁶-benzyl-3',5'-*O*-bis(*tert*-butyldimethylsilyl)-*N*²-dimethoxytrityl-2'-deoxyguanosine **33**



*O*⁶-Benzyl-8-bromo-3',5'-*O*-bis(*tert*-butyldimethylsilyl)-*N*²-dimethoxytrityl-2'-deoxyguanosine **30** (741 mg, 0.77 mmol), 2-aminofluorene **6** (259 mg, 1.38 mmol, 1.8 eq.), tris(dibenzylideneacetone)-dipalladium (Pd₂(dba)₃, 81 mg, 0.088 mmol, 10 mol%) and rac-2,2'-Bis(diphenylphosphino)-1,1'-binaphthyl (BINAP, 147 mg, 0.236 mmol, 30 mol%) were degassed and flushed with argon in a flame dried and degassed apparatus. The entire reaction was carried out under argon atmosphere. Degassed toluene (7.5 mL) was added and the mixture was heated to 90°C. After 40 min, a suspension of NaOtBu (136 mg, 1.38 mmol, 1.8 eq.) in degassed toluene (7.5 mL) was added and the reaction mixture was stirred 1-2 h at 110 °C, until no more starting material **30** was detected by TLC. The reaction mixture was subsequently cooled, concentrated and purified by chromatography on aluminium oxide (*n*-hexane:EtOAc 6:1) to provide 528 mg (0.50 mmol, 64 %, lit. 72%) of the product as a slightly yellow powder. *R*_f = 0.51 (*n*-hexane:EtOAc 3:1)

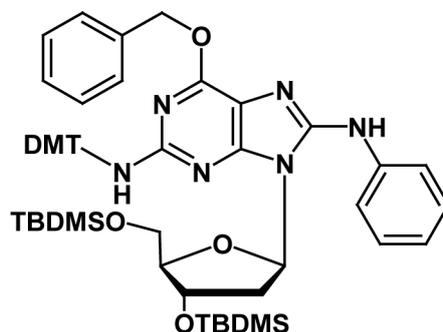
¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 8.85 (br s, 1H, *N*⁸-H), 7.82 (m, 1H, AF-C1-H), 7.72 – 7.74 (m, 2H, AF-2H), 7.57 (m, 1H, AF-H), 7.55 (m, 1H, AF-H), 7.30-7.15 (m, 16 H, AF-2H + Bz-5H + DMT-9H), 6.78 (m, 4H, DMT-4H), 6.56 (s, 1H, *N*²-H), 6.20 (m, 1H, C₁-H), 4.83 (s, 2H, Bz-CH₂), 4.57 (m, 1H, C_{3'}-H), 3.88 – 3.78 (m, 4H, AF-C9-2H + C_{5'}-H_a+H_b), 3.68 (s, 6H, 2xOMe), 3.60 (m, 1H, C_{4'}-H), 3.29 (m, 1H, C_{2'}-H_a), 2.00 (m, 1H, C_{2'}-H_b), 0.85 (s, 9H, tBu-H), 0.77 (s, 9H, tBu-H), 0.07 (s, 6H, 2xSi-CH₃), -0.06 (s, 3H, Si-CH₃), -0.07 (s, 3H, Si-CH₃).

¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 157.96, 156.58, 155.77, 153.69, 147.28, 144.17, 142.93, 141.67, 140.22, 137.14, 135.00, 130.19, 129.43, 129.00, 128.89, 128.77, 128.76, 128.56, 128.53, 128.23, 127.92, 127.09, 126.64, 126.07, 125.32, 119.50, 117.4, 115.03, 113.21, 112.34, 87.21, 83.50, 73.05, 69.84, 66.73, 63.10, 36.92, 36.28, 26.16, 26.07, 18.44, 18.28, -4.25, -4.38, -5.03, -5.07.

IR ν (cm⁻¹) 3746 (w), 2650 (w), 2447 (br w), 3369 (w), 3058 (w), 2951 (s), 2928 (s), 2855 (m), 1607 (s), 1590 (s), 1565 (m), 1507 (s), 1458 (s), 1418 (m), 1391 (m), 1359 (w), 1250 (s), 1178 (m), 1108 (br m), 1079 (br m), 1032 (m), 831 (s), 777 (m), 699 (m).

HR-MS (ESI⁺) *m/z* for [C₆₃H₇₅N₆O₆Si₂]⁺ ([M+H]⁺): calc.: 1067.5287, found: 1067.5291.

10.7.5.2 8-(*N*-Amino-benzene)-*O*⁶-benzyl-3',5'-*O*-bis(*tert*-butyldimethylsilyl)-*N*²-dimethoxytrityl-2'-deoxyguanosine **31**



The coupling of aniline to the protected bromo-guanosine **30** was carried out according to the procedure described for the coupling with 2-aminofluorene (section 10.7.5.1).

formulation

1.54 g (1.59 mmol) **30**
 0.30 g (0.48 mmol) BINAP
 0.17 g (0.16 mmol) Pd₂(dba)₃
 0.3 mL (2.87 mmol, 2 eq.) aniline
 0.27 g (2.85 mmol) NaOtBu in 18 mL toluene
 18 mL toluene

yield: 880 mg, 0.90 mmol, 57 %

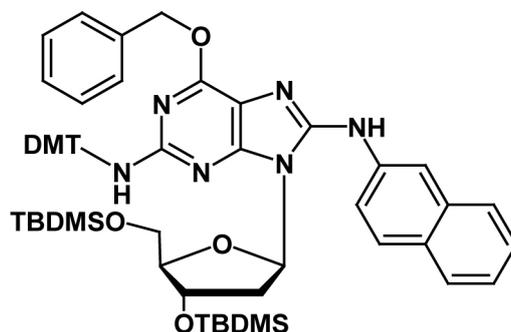
R_f = 0.52 (*n*-hexane:EtOAc 3:1)

¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 8.78 (s, 1H, *N*⁸-H), 7.56 (d, ³*J* = 7.7, 2H, Phe-H_{2,6}), 7.33 – 7.09 (m, 16H, Phe-H_{3,5} + Bz-5H + DMT-9H), 6.81 (t, ³*J* = 7.4, 1H, Phe-H₄), 6.78 (m, 4H, DMT-4H), 6.52 (s, 1H, *N*²-H), 6.18 (m, 1H, C₁-H), 4.80 (s, 2H, Bz-CH₂), 4.57 (m, 1H, C₃-H), 3.80 – 3.73 (m, 2H, C₅-H_a + H_b), 3.67 (s, 6H, 2xOMe), 3.58 – 3.56 (m, 1H, C₄-H), 3.29 (m, 1H, C₂-H_a), 2.43 (m, 1H, C₂-H_b), 0.85 (s, 9H, *t*Bu-H), 0.76 (s, 9H, *t*Bu-H), 0.07 (s, 6H, 2xSi-CH₃), -0.07 (s, 3H, Si-CH₃), -0.09 (s, 3H, Si-CH₃).

¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 157.94, 156.55, 155.72, 153.64, 147.34, 146.54, 140.99, 137.11, 130.14, 129.00, 128.87, 128.76, 128.57, 128.31, 127.90, 126.63, 121.59, 118.48, 113.20, 112.28, 87.18, 83.41, 73.02, 69.82, 66.72, 63.07, 55.38, 26.17, 26.14, 18.43, 18.28, -4.24, -4.38, -5.04, -5.09.

HR-MS (ESI⁻) *m/z* for [C₅₆H₆₉N₆O₆Si₂]⁻ ([*M*-H]⁻): calc.: 977.4817, found: 977.4890.

10.7.5.3 8-(*N*-2-Amino-naphthalene)-*O*⁶-benzyl-3',5'-*O*-bis(tert-butyl-dimethyl-silyl)-*N*²-dimethoxytrityl-2'-deoxyguanosine **32**



The coupling of 2-naphthylamine to the protected bromo-guanosine **30** was carried out according to the procedure described for the coupling with 2-aminofluorene (section 10.7.5.1).

formulation

1.00 g (1.03 mmol) **30**
 0.35 g (0.56 mmol) BINAP
 0.19 g (0.21 mmol) Pd₂(dba)₃
 0.27 g (1.86 mmol, 1.8 eq.) 2-naphthylamine
 0.18 g (1.86 mmol, 1.8 eq.) NaOtBu in 12 mL toluene
 12 mL toluene

yield: 431 mg, 0.42 mmol, 41 %

R_f = 0.58 (*n*-hexane:EtOAc 3:1)

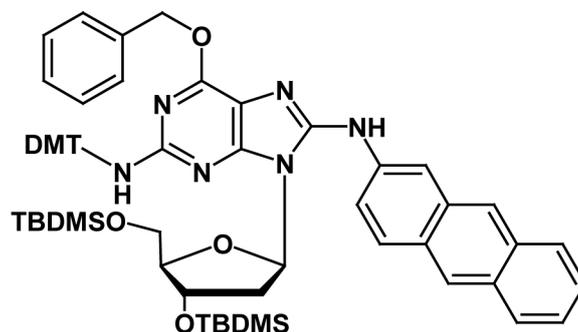
¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 9.02 (s, 1H, *N*⁸-H), 7.81 – 7.64 (m, 4H, Nap-4H), 7.47 – 7.19 (m, 17 H, Nap-3H + Bz-5H + DMT-9H), 6.79 (m, 4H, DMT-4H), 6.57 (s, 1H, *N*²-H), 6.22 (m, 1H, C₁-H), 4.84 (s, 2H, Bz-CH₂), 4.59 (m, 1H, C_{3'}-H), 3.77 (m, 2H, C_{5'}-H_a+H_b), 3.68 (s, 6H, 2xOMe), 3.59 (m, 1H, C_{1'}-H), 3.46 (br m, 1H, C_{2'}-H_a), 2.01 (m, 1H, C_{2'}-H_b), 0.84 (s, 9H, tBu-H), 0.76 (s, 9H, tBu-H), 0.07 (s, 3H, Si-CH₃), 0.06 (s, 3H, Si-CH₃), -0.07 (s, 3H, Si-CH₃), -0.09 (s, 3H, Si-CH₃).

¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 157.96, 156.69, 155.87, 153.72, 147.26, 146.53, 138.88, 138.44, 137.13, 134.06, 130.15, 129.24, 128.87, 128.75, 128.62, 128.45, 128.28, 127.91, 127.86, 127.10, 126.71, 126.64, 124.09, 120.38, 113.20, 112.36, 87.23, 83.54, 73.07, 69.85, 66.73, 63.09, 55.38, 36.22, 26.15, 26.14, 18.43, 18.27, -4.26, -4.39, -5.04, -5.08.

IR ν (cm⁻¹) 3367 (*br m*), 3059 (*br m*), 2952 (*s*), 2928 (*s*), 2896 (*m*), 2855 (*m*), 1605 (*br s*), 1566 (*s*), 1505 (*s*), 1262 (*m*), 1419 (*w*), 1391 (*m*), 1359 (*m*), 1294 (*w*), 1249 (*s*), 1178 (*m*), 1105 (*br m*), 1031 (*m*), 950 (*br w*), 831 (*s*), 778 (*m*), 698 (*m*).

HR-MS (ESI⁻) *m/z* for [C₆₀H₇₁N₆O₆Si₂]⁻ ([M-H]⁻): calc.: 1027.4974, found: 1027.4937.

10.7.5.4 8-(*N*-2-Amino-anthracene)-*O*⁶-benzyl-3',5'-*O*-bis(*tert*-butyldimethylsilyl)-*N*²-dimethoxytrityl-2'-deoxyguanosine **34**



The coupling of 2-aminoanthracene to the protected bromo-guanosine **30** was carried out according to the procedure described for the coupling with 2-aminofluorene (section 10.7.5.1).

formulation

1.30 g (1.34 mmol) **30**
 257 mg (0.41 mmol) BINAP
 142 mg (0.16 mmol) Pd₂(dba)₃
 0.40 g (2.07 mmol, 1.5 eq.) 2-aminoanthracene
 0.24 g (2.50 mmol, 1.8 eq.) NaOtBu in 16 mL toluene
 16 mL toluene

yield: 773 mg, 0.72 mmol, 53 %

R_f = 0.49 (*n*-hexane:EtOAc 3:1)

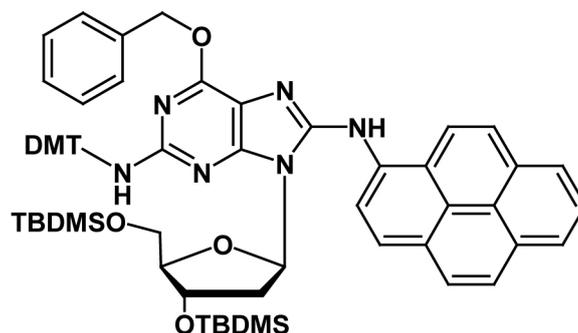
¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 9.13 (s, 1H, *N*⁸-H), 8.41 (s, 1H, Ant-H_{9/10}), 8.29 (s, 1H, Ant-H_{10/9}), 8.23 (d, *J* = 2.1, 1H, Ant-H₁), 7.97 (m, 3H, Ant-H₄, -H₅, -H₈), 7.66 (dd, *J* = 9.2, 2.1, 1H, Ant-H₃), 7.47 – 7.34 (m, 2H, Ant-H₆, -H₇), 7.33 – 7.10 (m, 14H, Bz-5H + DMT-9H), 6.83 – 6.76 (m, 4H, DMT-4H), 6.59 (br s, 1H, *N*²-H), 6.26 (br m, 1H, C₁'-H), 4.87 (br s, 2H, Bz-CH₂), 4.60 (m, 1H, C₃'-H), 3.85 – 3.74 (m, 2H, C₅'-H_a + C₄'-H), 3.69 (s, 6H, 2xOMe), 3.64 – 3.59 (m, 1H, C₅'-H_b), 3.49 (br m, 1H, C₂'-H_a), 2.02 (m, 1H, C₂'-H_b), 0.85 (s, 9H, tBu-H), 0.77 (s, 9H, tBu-H), 0.08 (s, 3H, Si-CH₃), 0.07 (s, 3H, Si-CH₃), -0.05 (s, 3H, Si-CH₃), -0.08 (s, 3H, Si-CH₃).

¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 157.97, 156.81, 155.93, 153.74, 147.10, 146.54, 138.43, 138.23, 137.18, 132.40, 132.20, 130.36, 130.16, 129.04, 128.88, 128.77, 128.44, 128.29, 128.00, 127.93, 127.59, 126.67, 126.20, 125.89, 124.91, 124.40, 121.78, 118.43, 113.23, 112.44, 111.29, 87.26, 83.58, 73.08, 69.86, 66.72, 63.10, 55.40, 36.25, 26.16, 26.15, 18.43, 18.28, -4.25, -4.37, -5.03, -5.07.

IR ν (cm⁻¹) 3358 (*br m*), 3056 (*w*), 3031 (*w*), 2951 (*s*), 2928 (*s*), 2855 (*m*), 1599 (*s*), 1565 (*s*), 1507 (*s*), 1291 (*m*), 1461 (*s*), 1414 (*m*), 1359 (*m*), 1295 (*w*), 1250 (*s*), 1220 (*m*), 1178 (*m*), 1105 (*br m*), 1020 (*m*), 828 (*s*), 777 (*m*), 739 (*m*), 698 (*m*), 663 (*w*).

HR-MS (ESI⁺) *m/z* for [C₆₄H₇₃N₆O₆Si₂]⁺ ([*M*+*H*]⁺): calc.: 1079.5287, found: 1079.5285.

10.7.5.5 8-(*N*-1-Amino-pyrene)-*O*⁶-benzyl-3',5'-*O*-bis(*tert*-butyldimethylsilyl)-*N*²-dimethoxytrityl-2'-deoxyguanosine **35**



The coupling of 1-aminopyrene to the protected bromo-guanosine **30** was carried out according to the procedure described for the coupling with 2-aminofluorene (section 10.7.5.1).

formulation

2.10 g (2.17 mmol) **30**
 416 mg (0.67 mmol) BINAP
 229 mg (0.25 mmol) Pd₂(dba)₃
 1.06 g (4.87 mmol, 2.2 eq.) 1-aminopyrene
 385 mg (4.00 mmol, 1.8 eq.) NaOtBu in 24 mL toluene
 24 mL toluene

yield: 1.63 g, 1.48 mmol, 68 %, lit. 51 %

R_f = 0.45 (*n*-hexane:EtOAc 3:1)

¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 9.05 (s, 1H, *N*⁸-H), 8.22 – 7.98 (m, 5H, Pyr-5H), 7.78 (m, 1H, Pyr-1H), 7.43 (m, 1H, Pyr-1H), 7.28– 7.17 (m, 13H, Bz-3H + DMT-9H + Pyr-H), 7.03 (m, 2H, Bz-2H), 6.80 – 6.76 (m, 4H, DMT-4H), 6.65 (br s, 1H, *N*²-H), 6.30 (br m, 1H, C₁-H), 4.76 (br s, 2H, Bz-CH₂), 4.56 (m, 1H, C₃-H), 3.84 (m, 2H, C₅-H_a + C₄-H), 3.68 (s, 6H, 2xOMe), 3.68 – 3.63 (m, 1H, C₅-H_b), 3.39 (br m, 1H, C₂-H_a), 2.05 (m, 1H, C₂-H_b), 0.80 (s, 9H, tBu-H), 0.68 (s, 9H, tBu-H), 0.06 (s, 3H, Si-CH₃), 0.05 (s, 3H, Si-CH₃), -0.13 (s, 3H, Si-CH₃), -0.17 (s, 3H, Si-CH₃).

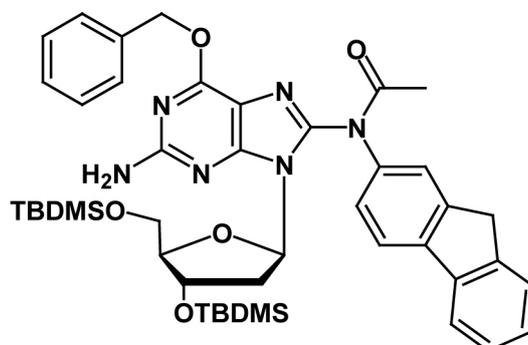
¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 157.93, 156.62, 155.79, 154.06, 149.54, 146.56, 143.23, 138.46, 136.98, 135.15, 131.45, 131.08, 130.96, 130.17, 129.42, 128.99, 128.88, 128.64, 128.42, 128.20, 127.89, 127.73, 127.15, 126.74, 126.60, 126.30, 126.08, 125.80, 125.32, 125.15, 124.98, 124.52, 123.59, 122.55, 122.10, 113.18, 112.30, 87.36, 84.09, 73.06, 69.76, 66.66, 63.15, 55.37, 26.08, 26.05, 18.37, 18.16, -4.31, -4.40, -5.12, -5.23.

IR ν (cm⁻¹) 3368 (*br m*), 3034 (*m*), 2951 (*s*), 2928 (*s*), 2855 (*m*), 1737 (*m*), 1297 (*s*), 1564 (*m*), 1507 (*s*), 1486 (*s*), 1461 (*m*), 1413 (*m*), 1248 (*s*), 1178 (*m*), 1102 (*m*), 1032 (*s*), 951 (*br w*), 830 (*s*), 776 (*m*), 698 (*m*), 677 (*w*).

HR-MS (ESI⁺) *m/z* for [C₆₆H₇₅N₆O₆Si₂]⁺ ([M-H]⁺): calc.: 1103.5287, found: 1103.5283.

10.7.6 N-Acetylation at C8-position of the aromatic amine dG adducts

10.7.6.1 8-(N-acetyl-2-amino-fluorene)-O⁶-benzyl-3',5'-O-bis(*tert*-butyldimethylsilyl)-2'-deoxyguanosine 38



8-(N-2-amino-fluorene)-O⁶-benzyl-3',5'-O-bis(*tert*-butyldimethylsilyl)-N²-dimethoxytrityl-2'-deoxyguanosine **33** (908 mg, 0.85 mmol) was dissolved in pyridine (17 mL) and triethylamine (588 μ L, 4.3 mmol), 4-dimethylaminopyridine (50 mg, 0.43 mmol) and acetic anhydride (403 μ L, 4.3 mmol) were added. After stirring for 5 hours at room temperature, the solvents were removed *in vacuo*, the reaction mixture was diluted in EtOAc (150 mL) and extracted with water (3 x100 mL). The organic layer was dried over MgSO₄, evaporated to dryness and treated with a 0.01 M solution of HCl in MeOH (4.5 mL, 45 μ mol). The mixture was stirred overnight at room temperature, concentrated and purified by chromatography on silica gel (*n*-hexane: EtOAc 6:1) to provide 474 mg (0.59 mmol, 69% overall, lit. 78 %) of the product as a slightly yellow powder.

R_f = 0.21 (*n*-hexane:EtOAc 3:1)

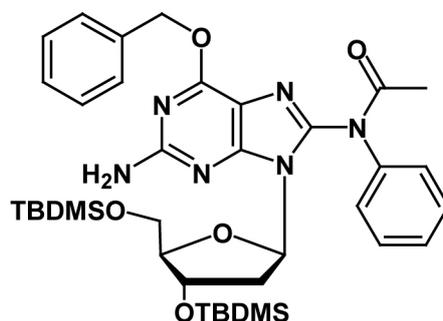
¹H-NMR (400 MHz, DMSO-d₆): δ (ppm) = 7.86 (m, 2H, AF-2H), 7.56 – 7.42 (m, 4H, AF-4H), 7.40 – 7.24 (m, 6 H, AF-H + Bz-5H), 6.44 (s, 2H, N²-H₂), 6.08 (m, 1H, C_{1'}-H), 5.45 (s, 2H, Bz-CH₂), 4.61 (m, 1H, C_{3'}-H), 3.88 – 3.72 (m, 4H, AF-C9-2H + C_{5'}-H_a+H_b), 3.64 (m, 1H, C_{4'}-H), 3.22 (m, 1H, C_{2'}-H_a), 2.02 (s, 3H, N⁸-Ac), 1.85 (m, 1H, C_{2'}-H_b), 0.84 (s, 9H, tBu-H), 0.73 (s, 9H, tBu-H), 0.07 (s, 6H, 2xSi-CH₃), -0.10 (s, 6H, Si-CH₃).

¹³C-NMR (101 MHz, DMSO-d₆): δ (ppm) = 160.57, 159.98, 154.15, 153.91, 143.76, 140.55, 138.37, 136.76, 129.17, 128.86, 128.61, 127.31, 125.58, 120.69, 113.27, 112.48, 87.71, 83.80, 72.59, 67.48, 60.19, 36.88, 26.09, 26.05, 18.29, 18.05, -4.35, -4.41, -5.02, -5.07.

IR ν (cm⁻¹) 3406 (*br m*), 3035 (*w*), 2955 (*s*), 2928 (*s*), 2856 (*m*), 1696 (*m*), 1610 (*s*), 1589 (*m*), 1508 (*s*), 1456 (*m*), 1428 (*w*), 1404 (*w*), 1361 (*m*), 1342 (*m*), 1293 (*w*), 1248 (*br s*), 1227 (*br s*), 1178 (*m*), 1100 (*br w*), 1078 (*w*), 829 (*s*), 775 (*m*), 735 (*m*), 698 (*w*), 667 (*w*).

HR-MS (ESI⁺) *m/z* for [C₄₄H₅₉N₆O₅Si₂]⁺ ([M+H]⁺): calc.: 807.4085, found 807.4071.

10.7.6.2 8-(*N*-acetyl-amino-benzene)-*O*⁶-benzyl-3',5'-*O*-bis(*tert*-butyldimethylsilyl)-2'-deoxyguanosine **36**



The acetylation at the C8 nitrogen of the aminobenzene-dG **31** was carried out according to the procedure described for the acetylation of 2-aminofluorene-dG (section 10.7.6.1).

formulation

0.88 g (0.90 mmol) 31
420 μ L (4.51 mmol) triethylamine
60 mg (0.45 mmol) 4-dimethylaminopyridine
420 μ L (4.51 mmol) acetic anhydride
17 mL pyridine
4.00 mL 0.01 M methanolic HCl solution

yield: 260 mg, 0.35 mmol, 38 %

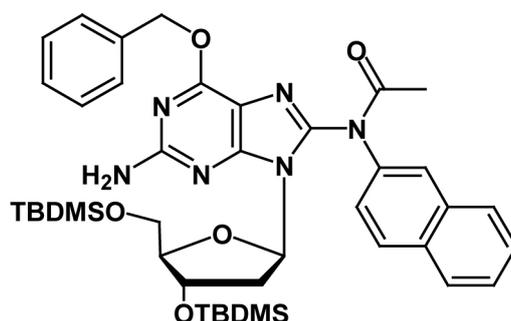
R_f = 0.33 (*n*-hexane:EtOAc 3:1)

¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 7.47 (m, 2H, Phe-H_{2,6}), 7.39 – 7.20 (m, 8H, Phe-H_{3,4,5} + Bz-5H), 6.45 (s, 2H, N²-H₂), 6.02 (m, 1H, C₁-H), 5.45 (s, 2H, Bz-CH₂), 4.62 (m, 1H, C_{3'}-H), 3.80 – 3.72 (m, 2H, C_{5'}-H_a+H_b), 3.64 (m, 1H, C_{4'}-H), 3.21 (m, 1H, C_{2'}-H_a), 1.98 (s, 3H, N⁸-Ac), 1.82 (m, 1H, C_{2'}-H_b), 0.85 (s, 9H, tBu-H), 0.74 (s, 9H, tBu-H), 0.07 (s, 3H, Si-CH₃), 0.05 (s, 3H, Si-CH₃), -0.09 (s, 3H, Si-CH₃), -0.10 (s, 3H, Si-CH₃).

¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 160.62, 159.91, 153.77, 136.75, 129.77, 129.13, 128.86, 128.60, 112.44, 87.69, 83.75, 72.56, 67.47, 63.25, 57.88, 36.46, 26.10, 26.07, 23.20, 18.31, 18.06, -4.31, -4.42, -5.00, -5.06.

HR-MS (ESI⁺) m/z for [C₃₇H₅₅N₆O₅Si₂]⁺ ([M+H]⁺): calc.: 719.3772, found: 719.3758.

10.7.6.38-(N-acetyl-2-amino-naphthalene)-O⁶-benzyl-3',5'-O-bis(tert-butyl-dimethylsilyl)-2'-deoxyguanosine 37



The acetylation at the C8 nitrogen of the aminonaphthalene-dG **32** was carried out according to the procedure described for the acetylation of 2-aminofluorene-dG (section 10.7.6.1).

formulation

0.43 g (0.42 mmol) 32
220 μ L (2.10 mmol) triethylamine
30 mg (0.21 mmol) 4-dimethylaminopyridine
200 μ L (2.10 mmol) acetic anhydride
8 mL pyridine
1.90 mL 0.01 M methanolic HCl solution

yield: 190 mg, 0.24 mmol, 57 %

R_f = 0.27 (*n*-hexane:EtOAc 3:1)

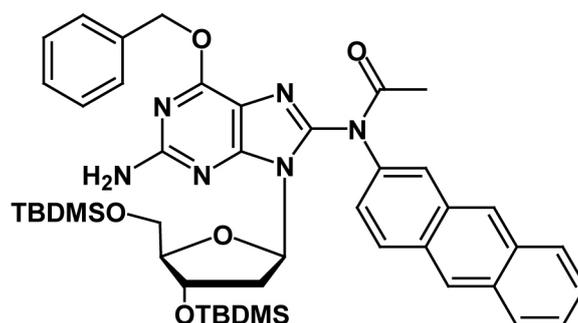
¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 8.02 – 7.81 (br m, 4H, Nap-4H), 7.58 – 7.27 (m, 8 H, Nap-3H + Bz-5H), 6.45 (s, 2 H, N²H₂), 6.04 (br m, 1H, C_{1'}-H), 5.47 (br d, 2H, Bz-CH₂), 4.61 (br m, 1H, C_{3'}-H), 3.83– 3.55 (m, 3H, C_{5'}-2H + C_{4'}-H), 3.22 (br m, 1H, C_{2'}-H_a), 2.05 (br s, 4H, N₈-Ac + C_{2'}-H_b), 0.82 (br s, 9H, tBu-H), 0.72 (s, 9H, tBu-H), 0.05 (br s, 6H, 2Si-CH₃), -0.12 (br s, 6H, 2Si-CH₃).

¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 159.75, 153.99, 136.61, 133.10, 129.14, 128.85, 128.58, 128.24, 127.93, 127.00, 112.46, 87.63, 83.77, 79.61, 72.47, 67.53, 26.07, 18.28, -4.34, -4.39, -5.04, -5.10.

IR ν (cm⁻¹) 3366 (*br m*), 2952 (*m*) 2928 (*s*), 2855 (*s*), 1933 (*w*), 1699 (*m*), 1611 (*s*), 1581 (*s*), 1508 (*w*), 1469 (*m*), 1429 (*m*), 1367 (*m*), 1336 (*m*), 1281 (*m*), 1231 (*s*), 1108 (*w*), 1068 (*m*), 1028 (*w*), 954 (*br w*), 834 (*s*), 776 (*s*), 746 (*m*), 695 (*w*), 669 (*w*).

HR-MS (ESI⁺) *m/z* for [C₄₁H₅₇N₆O₅Si₂]⁺ ([M+H]⁺): calc.: 769.3911, found: 769.3929.

10.7.6.4 8-(*N*-acetyl-2-amino-anthracene)-*O*⁶-benzyl-3',5'-*O*-bis(*tert*-butyl-dimethylsilyl)-2'-deoxyguanosine **39**



The acetylation at the C8 nitrogen of the aminoanthracene-dG **34** was carried out according to the procedure described for the acetylation of 2-aminofluorene-dG (section 10.7.6.1).

formulation

0.77 g (0.75 mmol) **34**
 530 μ L (5.06 mmol) triethylamine
 45 mg (0.31 mmol) 4-dimethylaminopyridine
 360 μ L (3.8 mmol) acetic anhydride
 15 mL pyridine

3.90 mL 0.01 M methanolic HCl solution

yield: 333 mg, 0.41 mmol, 54 %

R_f = 0.28 (*n*-hexane:EtOAc 3:1)

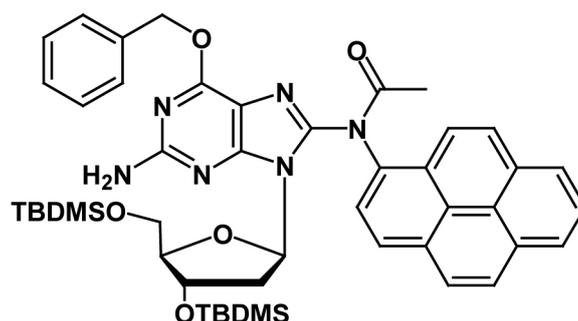
¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 8.55 (s, 1H, Ant-H_{9/10}), 8.50 (s, 1H, Ant-H_{10/9}), 8.11 – 7.81 (br m, 4H, Ant-4H), 7.57 – 7.26 (m, 8 H, Ant-3H + Bz-5H), 6.42 (s, 2 H, *N*²H₂), 6.08 (br m, 1H, C_{1'}-H), 5.47 (br s, 2H, Bz-CH₂), 4.59 (br m, 1H, C_{3'}-H), 3.79– 3.56 (m, 3H, C_{5'}-2H + C_{4'}-H), 3.23 (br m, 1H, C_{2'}-H_a), 2.08 (br s, 4H, *N*₈-Ac + C_{2'}-H_b), 0.81 (br s, 9H, tBu-H), 0.69 (s, 9H, tBu-H), 0.03 (br s, 6H, 2Si-CH₃), -0.14 (br s, 6H, 2Si-CH₃).

¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 160.50, 160.05, 153.89, 143.14, 137.21, 136.76, 131.95, 131.91, 131.13, 130.06, 129.98, 129.72, 129.12, 128.85, 128.59, 128.53, 128.29, 126.82, 126.51, 126.44, 126.28, 124.55, 124.33, 112.59, 87.70, 83.86, 79.60, 72.53, 67.53, 63.22, 26.05, 23.20, 18.24, 18.01, -4.44, -5.07, -5.12.

IR ν (cm⁻¹) 3366 (*w*), 2926 (*s*), 2855 (*s*), 1977 (*w*), 1699 (*m*), 1611 (*s*), 1582 (*s*), 1471 (*m*), 1419 (*w*), 1366 (*m*), 1338 (*m*), 1250 (*s*), 1066 (*m*), 1028 (*w*), 953 (*w*), 834 (*s*), 775 (*s*).

HR-MS (ESI⁺) *m/z* for [C₄₅H₅₉N₆O₅Si₂]⁺ ([M+H]⁺): calc.: 819.4085, found: 819.4090.

10.7.6.5 8-(*N*-acetyl-1-amino-pyrene)-*O*⁶-benzyl-3',5'-*O*-bis(*tert*-butyldimethylsilyl)-2'-deoxyguanosine **40**



The acetylation at the C8 nitrogen of the aminopyrene-dG **35** was carried out according to the procedure described for the acetylation of 2-aminofluorene-dG (section 10.7.6.1).

formulation

1.08 g (0.98 mmol) **35**
 684 μ L (6.53 mmol) triethylamine
 60 mg (0.42 mmol) 4-dimethylaminopyridine
 468 μ L (4.91 mmol) acetic anhydride
 22 mL pyridine

5.30 mL 0.01 M methanolic HCl solution

yield: 660 mg, 0.78 mmol, 80 %

R_f = 0.30 (*n*-hexane:EtOAc 3:1)

¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 8.35 – 8.08 (m, 5H, Pyr-5H), 7.92 (m, 1H, Pyr-1H), 7.5 – 7.09 (m, 7H, Bz-5H, Pyr-2H), 6.37 (br s, 2H, *N*²-H₂), 6.34 (m, 1H, C_{1'}-H), 5.53 (br s, 2H, Bz-CH₂), 4.59 (m, 1H, C_{3'}-H), 3.92 – 3.85 (m, 2H, C_{5'}-H_a + C_{4'}-H), 3.81 – 3.72 (m, 1H, C_{5'}-H_b), 3.41 (m, 1H, C_{2'}-H_a), 2.25 (s, 3H, *N*⁸-Ac), 1.93 (m, 1H, C_{2'}-H_b), 0.89 (s, 9H, tBu-H), 0.74 (s, 9H, tBu-H), 0.13 (s, 6H, 2xSi-CH₃), -0.06 (s, 3H, Si-CH₃), -0.10 (s, 3H, Si-CH₃).

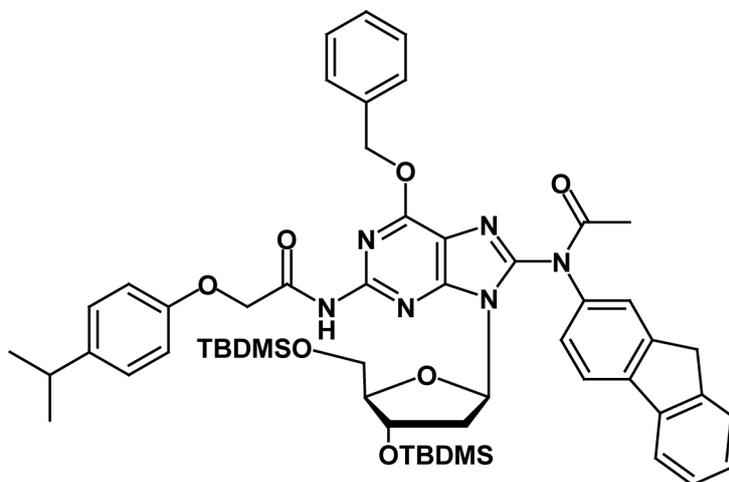
¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 158.08, 136.73, 136.14, 131.01, 130.10, 128.82, 128.04, 127.45, 126.37, 125.10, 123.92, 113.33, 87.91, 84.20, 72.81, 69.83, 67.33, 62.92, 55.39, 26.06, 18.28, -4.36, -5.05.

IR ν (cm⁻¹) 3510 (*br w*), 3368 (*br w*), 3034 (*w*), 2951 (*s*), 2928 (*s*), 2884 (*m*), 2855 (*s*), 1735 (*w*), 1694 (*m*), 1611 (*s*), 1582 (*s*), 1507 (*m*), 1462 (*m*), 1420 (*w*), 1367 (*m*), 1339 (*m*), 1248 (*s*), 1179 (*w*), 1108 (*m*), 1067 (*br m*), 1029 (*m*), 833 (*s*), 775 (*s*), 696 (*w*).

HR-MS (ESI⁺) m/z for [C₄₇H₅₉N₆O₅Si₂]⁺ ([M+H]⁺): calc.: 843.4085, found: 843.4085.

10.7.7 Protection of the *N*²-position of the aromatic amine dG adducts

10.7.7.1 8-(*N*-acetyl-2-amino-fluorene)-*O*⁶-benzyl-3',5'-*O*-bis(*tert*-butyldimethylsilyl)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine **43**



8-(*N*-acetyl-2-amino-fluorene)-*O*⁶-benzyl-3',5'-*O*-bis(*tert*-butyldimethylsilyl)-2'-deoxyguanosine (519 mg, 0.64 mmol) **38** was dissolved in pyridine (25 mL). 4-Isopropylphenoxyacetyl chloride (575 mg, 2.74 mmol, for preparation see section 10.7.12) was added and the dark red reaction mixture was stirred overnight at room temperature. The solvent was subsequently removed *in vacuo* and the crude product was diluted in EtOAc (100 ml) and extracted with water (2 x 75 ml). The organic layer was dried over MgSO₄, concentrated and purified by chromatography on silica gel (*n*-hexane:EtOAc 4:1) to provide 582 mg (0.91 mmol, 92%, lit. 97 %) of the product as a slightly yellow powder.

$R_f = 0.13$ (*n*-hexane:EtOAc 3:1)

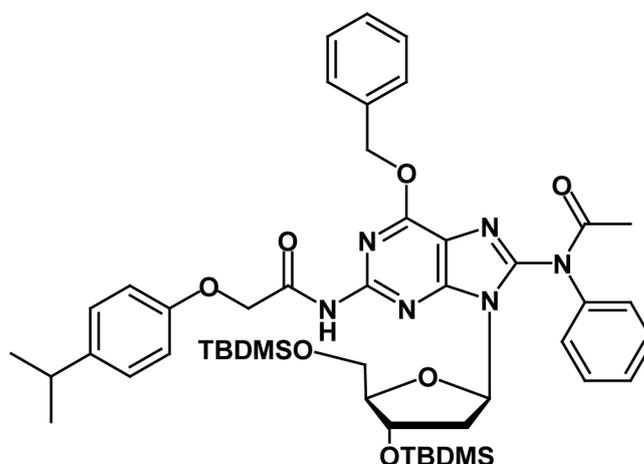
¹H-NMR (600 MHz, DMSO-*d*₆): δ (ppm) = 7.91–7.87 (m, 2H, AF-2H), 7.56–7.51 (m, 4H, AF-4H), 7.37–7.29 (m, 6 H, AF-H + Bz-5H), 7.10 (d, ³*J* = 8.6, 2H, Pac-H2), 6.81 (d, ³*J* = 8.6, 2H, Pac-H2), 6.22 (m, 1H, C_{1'}-H), 5.59 (s, 2H, Bz-CH₂), 4.84 (m, 3H, Pac-CH₂, C_{3'}-H), 3.92 (s, 2H, AF-C9-H₂), 3.77–3.70 (m, 3H, C_{5'}-H_a+H_b + C_{4'}-H), 3.23 (m, 1H, C_{2'}-H_a), 2.79 (sept, ³*J* = 6.9, 1 H, *i*Pr-H), 2.06 (s, 3H, *N*⁸-Ac), 1.95 (m, 1H, C_{2'}-H_b), 1.13–1.11 (m, 6H, 2*x**i*Pr-CH₃), 0.80 (s, 9H, *t*Bu-H), 0.69 (s, 9H, *t*Bu-H), -0.01 (br s, 6H, 2*x*Si-CH₃), -0.18 (br s, 6H, 2*x*Si-CH₃).

¹³C-NMR (151 MHz, DMSO-*d*₆): δ (ppm) = 156.43, 143.81, 141.25, 136.26, 129.23, 128.94, 128.81, 127.45, 127.32, 125.58, 116.31, 114.68, 88.17, 72.17, 68.52, 67.51, 36.89, 33.03, 26.05, 24.52, 18.26, 17.93, -4.53, -4.63, -5.07, -5.14.

IR ν (cm⁻¹) 3650 (*w*), 3414 (*br w*), 3032 (*w*), 2953 (*s*), 2928 (*s*), 2891 (*w*), 2856 (*m*), 1696 (*s*), 1610 (*s*), 1584 (*s*), 1508 (*s*), 1456 (*s*), 1426 (*m*), 1403 (*w*), 1342 (*m*), 1293 (*w*), 1249 (*s*), 1227 (*s*), 1077 (*m*), 1030 (*m*), 954 (*w*), 910 (*w*), 832 (*s*), 775 (*m*), 734 (*m*), 698 (*w*), 668 (*w*).

HR-MS (ESI⁺) *m/z* for [C₅₅H₇₁N₆O₇Si₂]⁺ ([*M*+*H*]⁺): calc.: 983.4923, found: 983.4926.

10.7.7.2 8-(*N*-acetyl-amino-benzene)-*O*⁶-benzyl-3',5'-*O*-bis(*tert*-butyldimethylsilyl)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine 41



The protection of the *N*²-position of acetylaminobenzene-dG **38** was carried out according to the procedure described for the protection of acetylaminofluorene-dG (section 10.7.7.1).

formulation

0.26 g (0.35 mmol) **38**
 310 μ L (1.50 mmol) 4-Isopropylphenoxyacetyl chloride
 14 mL pyridine

yield: 270 mg, 0.30 mmol, 86 %

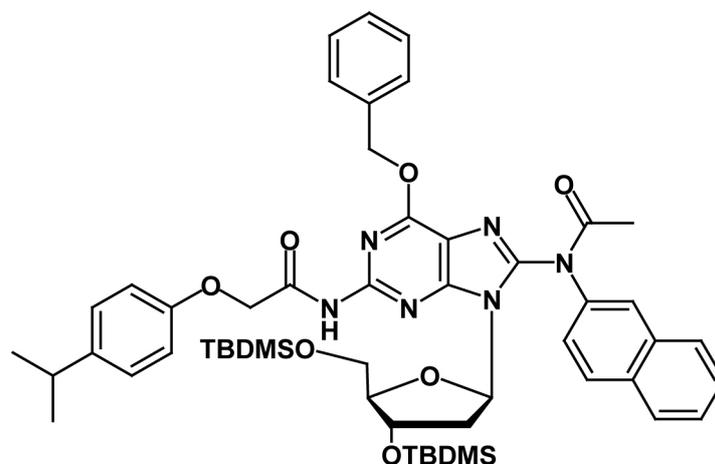
R_f = 0.15 (*n*-hexane:EtOAc 4:1)

¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 10.61 (s, 1H, *N*²-H), 7.51 – 7.32 (m, 10H, Phe-5H + Bz-5H), 7.10 (d, ³*J* = 8.4, 2H, Pac-H₂), 6.81 (d, ³*J* = 8.7, 2H, Pac-H₂), 6.17 (m, 1H, C_{1'}-H), 5.57 (s, 2H, Bz-CH₂), 4.84 (m, 3H, Pac-CH₂, C_{3'}-H), 3.80 – 3.65 (m, 3H, C_{5'}-H_a+H_b + C_{4'}-H), 3.15 (m, 1H, C_{2'}-H_a), 2.79 (sept, ³*J* = 7.3, 1 H, *i*Pr-H), 2.10 (br m, 1H, C_{2'}-H_b), 2.01 (s, 3H, *N*⁸-Ac), 1.14 (d, ³*J* = 7.2 Hz, 3H, *i*Pr-CH₃), 1.13 (d, ³*J* = 6.9 Hz, 3H, *i*Pr-CH₃), 0.80 (s, 9H, *t*Bu-H), 0.69 (s, 9H, *t*Bu-H), -0.01 (s, 3H, Si-CH₃), -0.03 (s, 3H, Si-CH₃), -0.16 (s, 3H, Si-CH₃), -0.17 (s, 3H, Si-CH₃).

¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 167.02, 156.42, 152.16, 150.03, 141.23, 136.24, 129.92, 129.24, 128.94, 128.83, 127.46, 127.41, 124.35, 116.26, 115.36, 114.64, 88.13, 84.01, 72.12, 68.50, 67.44, 33.03, 26.06, 26.04, 24.67, 24.53, 18.29, 17.94, -4.52, -4.65, -5.06, -5.13.

HR-MS (ESI⁻) *m/z* for [C₄₈H₆₅N₆O₇Si₂]⁻ ([M-H]⁻): calc.: 893.4453, found 893.4503.

10.7.7.3 8-(*N*-acetyl-2-amino-naphthalene)-*O*⁶-benzyl-3',5'-*O*-bis(*tert*-butyl-dimethylsilyl)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine **42**



The protection of the *N*²-position of acetylaminonaphthalene-dG **37** was carried out according to the procedure described for the protection of acetylaminofluorene-dG (section 10.7.7.1).

formulation

0.26 g (0.86 mmol) **37**
 760 μ L (3.50 mmol) 4-Isopropylphenoxyacetyl chloride
 35 mL pyridine

yield: 740 mg, 0.79 mmol, 91 %

R_f = 0.21 (*n*-hexane:EtOAc 4:1)

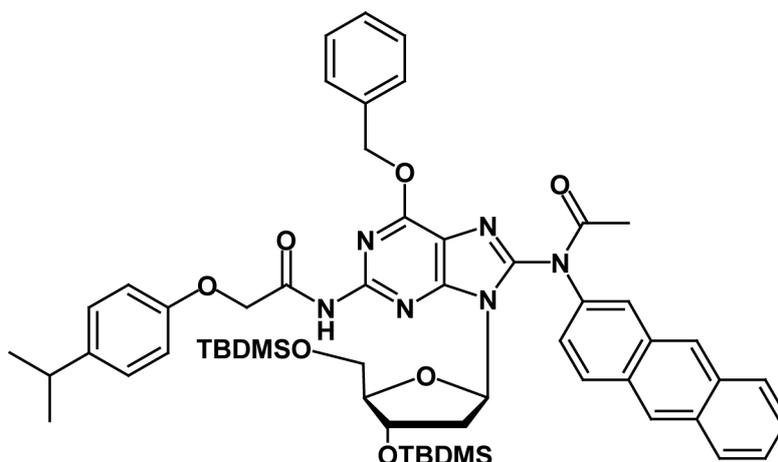
¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 10.61 (s, 1H, *N*²-H), 8.01 – 7.78 (m, 4H, Nap-4H), 7.59 – 7.31 (m, 8 H, Nap-3H + Bz-5H), 7.13 – 7.06 (m, 2 H, Pac-H_{3,5}), 6.84 – 6.77 (m, 2 H, Pac-H_{2,6}), 6.23 (br m, 1H, C_{1'}-H), 5.58 (br s, 2H, Bz-CH₂), 4.95 – 4.75 (br m, 3H, Pac-CH₂ + C_{3'}-H), 3.84 – 3.59 (m, 3H, C_{5'}-2H + C_{4'}-H), 3.16 (br m, 1H, C_{2'}-H_a), 2.78 (sept, *J* = 6.9, 1H, *i*Pr-H), 2.07 (br s, 4H, *N*₈-Ac + C_{2'}-H_b), 1.12 (d, *J* = 6.9, 6H, *i*Pr-CH₃), 0.79 (br s, 9H, *t*Bu-H), 0.67 (s, 9H, *t*Bu-H), -0.03 (s, 6H, 2Si-CH₃), -0.21 (br s, 6H, 2Si-CH₃).

¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = ¹³C NMR (101 MHz, dms) δ = 167.03, 160.21, 156.42, 152.23, 146.18, 141.23, 136.24, 133.37, 129.25, 128.94, 128.83, 128.34, 128.05, 127.46, 127.23, 116.34, 114.65, 88.16, 84.11, 72.10, 68.52, 67.46, 63.74, 33.03, 26.03, 24.53, 23.10, 18.25, 17.93, -4.55, -4.66, -5.10, -5.17.

IR ν (cm⁻¹) 3746 (*w*), 3671 (*w*), 3406 (*w*), 3058 (*w*), 2954 (*s*), 2927 (*s*), 2884 (*m*), 2856 (*s*), 1669 (*s*), 1612 (*s*), 1589 (*s*), 1508 (*s*), 1462 (*s*), 1358 (*s*), 1340 (*s*), 1282 (*w*), 1248 (*br m*), 1225 (*s*), 1178 (*w*), 1075 (*m*), 961 (*w*), 910 (*w*), 835 (*s*), 776 (*s*), 746 (*m*), 696 (*w*), 668 (*w*).

HR-MS (ESI⁺) *m/z* for [C₅₂H₆₉N₆O₇Si₂]⁺ ([*M*+*H*]⁺): calc.: 945.4766, found 945.4736.

10.7.7.4 8-(*N*-acetyl-2-amino-anthracene)-*O*⁶-benzyl-3',5'-*O*-bis(*tert*-butyldimethyl-silyl)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine 44



The protection of the *N*²-position of acetylaminoanthracene-dG **39** was carried out according to the procedure described for the protection of acetylaminofluorene-dG (section 10.7.7.1).

formulation

0.60 g (0.73 mmol) **39**
 645 μ L (2.97 mmol) 4-Isopropylphenoxyacetyl chloride
 27 mL pyridine

yield: 596 mg, 0.61 mmol, 84%

R_f = 0.17 (*n*-hexane:EtOAc 3:1)

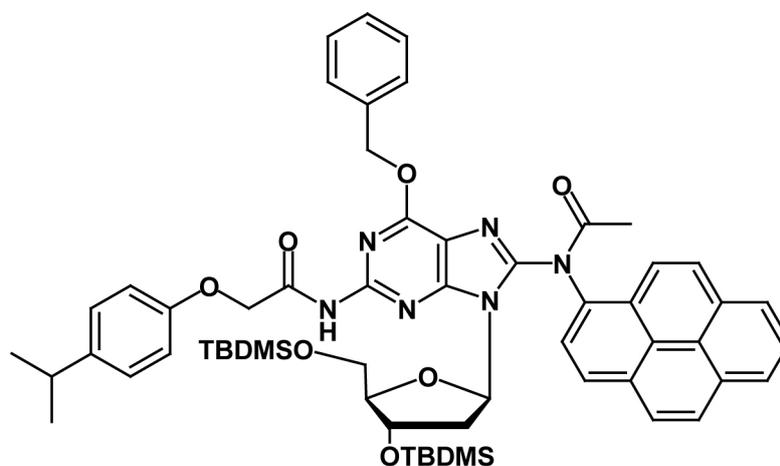
¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 10.57 (s, 1H, *N*²-H), 8.57 (s, 1H, Ant-H_{9/10}), 8.51 (s, 1H, Ant-H_{10/9}), 8.13 – 7.96 (m, 4H, Ant-4H), 7.58 – 7.29 (m, 8 H, Ant-3H + Bz-5H), 7.13 – 7.06 (m, 2 H, Pac-H_{3,5}), 6.84 – 6.77 (m, 2 H, Pac-H_{2,6}), 6.24 (br m, 1H, C₁-H), 5.59 (br s, 2H, Bz-CH₂), 4.90 – 4.73 (br m, 3H, Pac-CH₂ + C₃-H), 3.85 – 3.59 (m, 3H, C_{5'}-2H + C₄-H), 3.18 (br m, 1H, C₂-H_a), 2.78 (sept, *J* = 6.9, 1H, *i*Pr-H), 2.12 (br s, 4H, *N*₈-Ac + C₂-H_b), 1.12 (d, *J* = 6.9, 6H, *i*Pr-CH₃), 0.77 (br s, 9H, *t*Bu-H), 0.64 (br s, 9H, *t*Bu-H), 0.08 (s, 3H, Si-CH₃), 0.07 (s, 6H, Si-CH₃), -0.04 (br s, 3H, Si-CH₃), -0.22 (br s, 3H, Si-CH₃).

¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 167.02, 160.27, 156.42, 152.31, 151.86, 146.16, 141.27, 137.02, 136.26, 132.01, 131.97, 131.11, 130.13, 129.23, 128.94, 128.81, 128.54, 128.31, 127.45, 126.93, 126.73, 126.60, 126.50, 126.40, 125.30, 125.14, 124.99, 124.70, 116.40, 115.37, 114.69, 88.19, 84.17, 72.15, 68.55, 67.54, 63.72, 33.01, 26.24, 26.01, 24.50, 23.20, 18.22, 17.91, -4.56, -4.65, -5.12, -5.18.

IR ν (cm⁻¹) 3752 (*w*), 3412 (*m*), 3032 (*w*), 2954 (*s*), 2928 (*s*), 2887 (*w*), 2885 (*m*), 1701 (*br s*), 1609 (*s*), 1589 (*s*), 1508 (*s*), 1461 (*s*), 1426 (*s*), 1342 (*s*), 1284 (*w*), 1249 (*br m*), 1224 (*s*), 1074 (*br s*), 1029 (*m*), 958 (*w*), 833 (*s*), 776 (*s*), 696 (*w*) 671 (*w*).

HR-MS (ESI⁻) *m/z* for [C₅₆H₆₉N₆O₇Si₂]⁻ ([*M*-H]⁻): calc.: 993.4766 found 993.4752.

10.7.7.5 8-(*N*-acetyl-1-amino-pyrene)-*O*⁶-benzyl-3',5'-*O*-bis(*tert*-butyldimethylsilyl)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine **45**



The protection of the *N*²-position of acetylaminopyrene-dG **40** was carried out according to the procedure described for the protection of acetylaminofluorene-dG (section 10.7.7.1).

formulation

0.78 g (0.92 mmol) **40**
 830 μ L (3.82 mmol) 4-Isopropylphenoxyacetyl chloride
 34 mL pyridine

yield: 991 mg, 0.97 mmol, 99%

$R_f = 0.15$ (*n*-hexane:EtOAc 3:1)

¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 8.37 – 8.07 (m, 5H, Pyr-5H), 7.59 – 7.14 (m, 8H, Bz-5H, Pyr-3H), 7.10 (d, ³*J* = 8.6, 2H, Pac-H₂), 6.80 (d, ³*J* = 8.4, 2H, Pac-H₂), 6.51 (m, 1H, C₁-H), 5.51 (br s, 2H, Bz-CH₂), 4.95 (m, 1H, C₃-H), 4.83 (m, 2H, Pac-CH₂), 3.80 – 3.84 (m, 3H, C₅-H_a+H_b + C₄-H), 3.57 (m, 1H, C₂-H_a), 2.78 (sept, ³*J* = 6.8, 1 H, *i*Pr-H), 2.26 (br s, 3H, *N*⁸-Ac) 1.88 (m, 1H, C₂-H_b), 1.13 (m, 6H, 2*xi*Pr-CH₃), 0.81 (s, 9H, *t*Bu-H), 0.71 (s, 9H, *t*Bu-H), 0.08 (s, 3H, Si-CH₃), 0.07 (s, 3H, Si-CH₃), -0.14 (s, 3H, Si-CH₃), -0.17 (s, 3H, Si-CH₃).

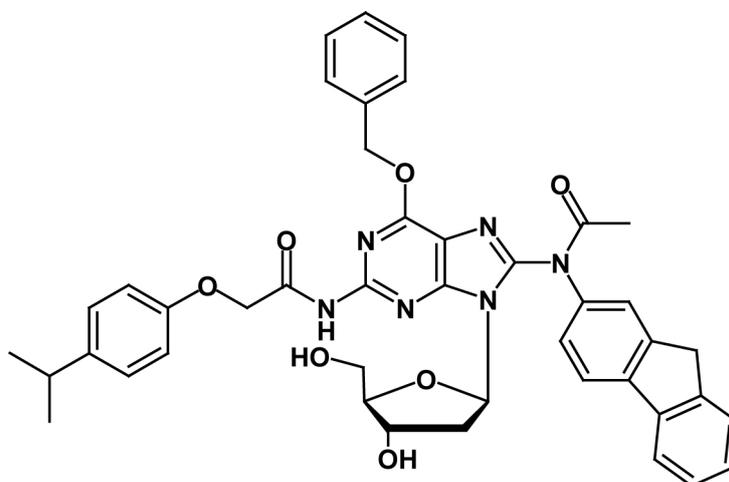
¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 156.42, 149.98, 141.21, 136.18, 130.99, 129.23, 128.87, 127.45, 124.95, 124.14, 115.21, 114.64, 113.32, 71.93, 68.53, 67.45, 55.39, 33.03, 26.07, 24.53, 18.30, 17.84, -4.64, -5.11.

IR ν (cm⁻¹) 3669 (*w*), 3410 (*m*), 3037 (*m*), 2955 (*s*), 2928 (*s*), 2887 (*m*), 2856 (*s*), 1695 (*s*), 1610 (*s*), 1586 (*s*), 1508 (*s*), 1461 (*s*), 1426 (*s*), 1343 (*s*), 1285 (*m*), 1224 (*br s*), 1178 (*m*), 1076 (*m*) 1030 (*w*), 967 (*w*), 907 (*w*), 833 (*s*), 776 (*m*), 696 (*w*), 681 (*w*), 637 (*w*).

HR-MS (ESI⁻) *m/z* for [C₅₈H₆₉N₆O₇Si₂]⁻ ([M-H]⁻): calc.: 1017.4766, found: 1017.4746.

10.7.8 Deprotection of the hydroxyl groups of the aromatic amine dG adducts

10.7.8.1 8-(*N*-acetyl-2-amino-fluorene)-*O*⁶-benzyl-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine 48



8-(*N*-acetyl-2-amino-fluorene)-*O*⁶-benzyl-3',5'-*O*-bis(*tert*-butyldimethylsilyl)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine (1.09 g, 1.11 mmol) **43** was dissolved in THF (26 ml) and cooled under argon atmosphere to 0 °C. Acetic acid (364 μL, 6.36 mmol) and a 1 M solution of tetrabutylammonium fluoride in THF (3.20 ml, 3.07 mmol) were added and the reaction mixture was allowed to reach room temperature and was stirred overnight. The solvent was subsequently removed *in vacuo* and the residue was purified by chromatography on silica gel (CH₂Cl₂:MeOH 40:1) to yield 470 mg (0.62 mmol, 57 %, lit. 79 %) of the product as a white powder.

R_f = 0.36 (CH₂Cl₂:MeOH 10:1)

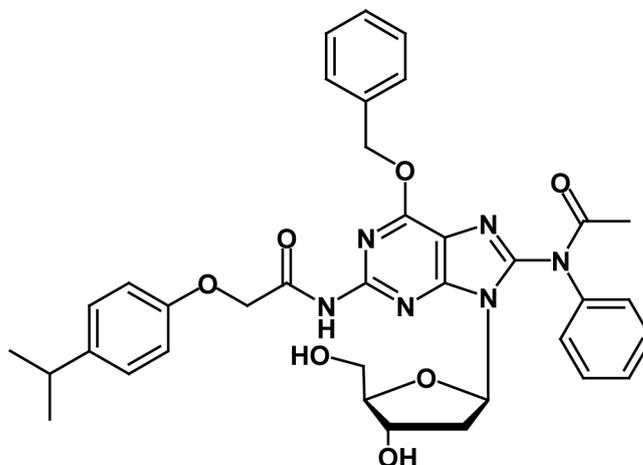
¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 7.91 (m, 1H, AF-1H), 7.88 (d, 1H, ³*J* = 7.5, AF-1H), 7.49 (m, 2H, AF-2H), 7.38 – 7.27 (m, 8 H, AF-3H + Bz-5H), 7.12 (d, ³*J* = 8.7, 2H, Pac-H2), 6.83 (d, ³*J* = 8.7, 2H, Pac-H2), 6.22 (m, 1H, C₁'-H), 5.56 (br s, 2H, Bz-CH₂), 5.20 (br d, 1H, C₃'-OH), 4.92 (s, 2H, Pac-CH₂), 4.57 (m, 2H, C₅'-OH + C₃'-H), 3.93 (s, 2H, AF-C9-H₂), 3.81 (m, 1H, C₄'-H), 3.65 (m, 1H, C₅'-H_a), 3.53 (m, 1H, C₅'-H_b), 3.18 (br m, 1H, C₂'-H_a), 2.79 (sept, ³*J* = 6.9, 1 H, *i*Pr-H), 2.07 (s, 3H, *N*⁸-Ac), 2.04 (m, 1H, C₂'-H_b), 1.14 – 1.12 (m, 6H, 2*xi*Pr-CH₃).

¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 167.77, 156.41, 152.40, 143.89, 141.34, 136.25, 129.24, 128.92, 128.80, 127.54, 127.31, 125.61, 116.28, 114.78, 88.66, 68.63, 67.78, 62.36, 36.91, 33.01, 24.53.

IR ν (cm⁻¹) 3745 (*w*), 3391 (*br m*), 2957 (*m*), 1695 (*s*), 1608 (*s*), 1507 (*s*), 1455 (*s*), 1426 (*s*), 1342 (*s*), 1287 (*m*), 1220 (*s*), 1179 (*m*), 1051 (*br w*), 940 (*w*), 828 (*m*), 734 (*s*), 697 (*w*).

HR-MS (ESI⁻) m/z for [C₄₃H₄₁N₆O₇]⁻ ([M-H]⁻): calc.: 753.3037, found 753.3074.

10.7.8.2 8-(*N*-acetyl-amino-benzene)-*O*⁶-benzyl-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine **46**



The deprotection of the 3' and 5' hydroxy functions of acetylaminobenzene-dG **41** was carried out according to the procedure described for the deprotection of acetylaminofluorene-dG (section 10.7.8.1).

formulation

1.19 g (1.33 mmol) **41**
 4.00 mL (3.83 mmol) 1 M solution of tetrabutylammonium fluoride in THF
 0.46 mL (8.04 mmol) acetic acid
 32 mL THF

yield: 560 mg, 0.84 mmol, 63 %

$R_f = 0.39$ (CH₂Cl₂:MeOH 10:1)

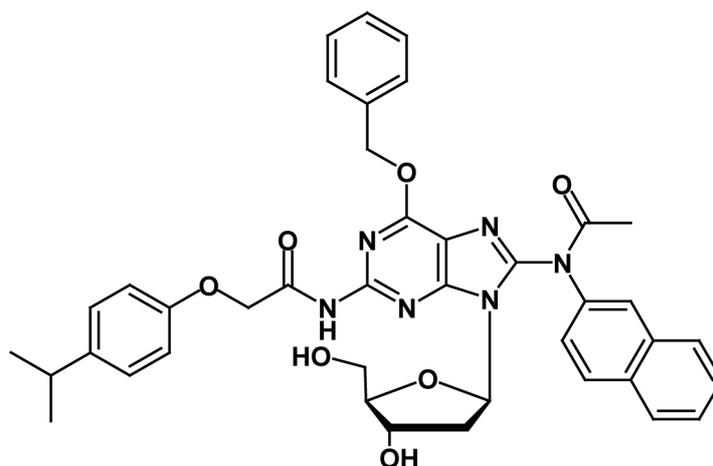
¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 10.64 (s, 1H, *N*²-H), 7.54 – 7.29 (m, 10H, Phe-5H + Bz-5H), 7.12 (d, ³*J* = 8.5, 2H, Pac-H₂), 6.83 (d, ³*J* = 8.7, 2H, Pac-H₂), 6.18 (m, 1H, C₁-H), 5.57 (br s, 2H, Bz-CH₂), 5.22 (br d, 1H, C₃'-OH), 4.94 (s, 2H, Pac-CH₂), 4.62 (br t, 1H, ³*J* = 5.75 Hz, C₅'-OH), 4.55 (m, 2H, C₃'-H), 3.78 (m, 1H, C₄'-H), 3.63 (m, 1H, C₅'-H_a), 3.52 (m, 1H, C₅'-H_b), 3.14 (m, 1H, C₂'-H_a), 2.80 (sept, ³*J* = 6.9, 1 H, *i*Pr-H), 2.03 (s, 3H, *N*⁸-Ac), 2.02 (m, 1H, C₂'-H_b), 1.14 (d, ³*J* = 6.9, 6H, 2*xi*Pr-CH₃).

¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 167.79, 156.39, 152.34, 141.32, 136.23, 129.95, 129.25, 128.94, 128.82, 127.56, 116.21, 114.74, 88.64, 84.41, 71.07, 68.63, 67.71, 62.31, 49.03, 37.11, 33.02, 31.14, 24.55.

IR ν (cm⁻¹) 3854 (w), 3744 (w), 3368 (br m), 2958 (m), 1696 (s), 1609 (s), 1507 (s), 1456 (m), 1342 (s), 1224 (s), 1051 (m) 695 (s).

HR-MS (ESI⁺) *m/z* for [C₃₆H₃₉N₆O₇]⁺ ([M+H]⁺): calc.: 667.2880, found 667.2859

10.7.8.3 8-(*N*-acetyl-2-amino-naphthalene)-*O*⁶-benzyl-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine **47**



The deprotection of the 3' and 5' hydroxy functions of acetylaminonaphthalene-dG **42** was carried out according to the procedure described for the deprotection of acetylaminofluorene-dG (section 10.7.8.1).

formulation

0.74 g (0.79 mmol) **42**
 2.38 mL (2.28 mmol) 1 M solution of tetrabutylammonium fluoride in THF
 0.27 mL (4.71 mmol) acetic acid
 20 mL THF

yield: 540 mg, 0.75 mmol, 95 %

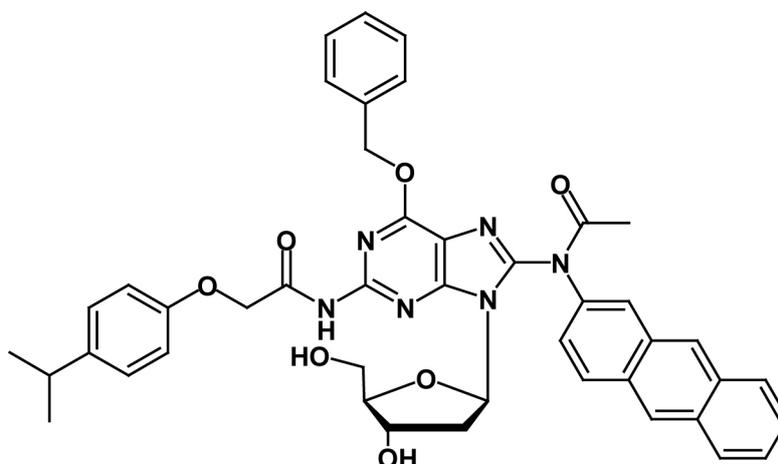
$R_f = 0.56$ ($\text{CH}_2\text{Cl}_2:\text{MeOH}$ 10:1)

¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 10.61 (s, 1H, *N*²-H), 7.52 (m, 7H, Nap-7H), 7.40-7.34 (m, 5H, Bn-5H), 7.09-7.11 (m, 2H, PAC- H_{3,5}), 6.80-6.82 (m, 2H, PAC- H_{2,6}), 6.25-6.23 (m, 1H, C₁-H), 5.58 (br s, 2H, Bz-CH₂), 5.24-5.20 (br d, 1H, C_{3'}-OH), 4.89-4.84 (m, 4H, Pac-CH₂, C_{5'}-OH + C_{3'}-H), 3.77-3.67 (m, 3H, C_{4'}-H, C_{5'}-H_a + H_b), 3.16 (m, 1H, C_{2'}-H_a), 2.73-2.84 (sept, ³*J* = 6.9, 1H, *i*Pr-H), 2.08 (s, 3H, *N*⁸-Ac), 1.96 (s, 1H, C_{2'}-H_b), 1.12-1.21 (m, 6H, 2*xi*Pr-CH₃).

IR ν (cm⁻¹) 3378 (*br m*), 3059 (*w*), 2957 (*m*), 1696 (*s*), 1609 (*s*), 1507 (*s*), 1457 (*m*), 1427 (*m*), 1340 (*s*), 1282 (*m*), 1224 (*s*), 1102 (*w*), 1052 (*w*), 941 (*w*), 828 (*w*), 793 (*w*), 747 (*m*), 696 (*w*), 652 (*w*).

HR-MS (ESI⁻) *m/z* for [C₄₀H₃₉N₆O₇]⁻ ([M-H]⁻): calc.: 715.2880, found: 715.2883.

10.7.8.4 8-(*N*-acetyl-2-amino-anthracene)-*O*⁶-benzyl-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine **49**



The deprotection of the 3' and 5' hydroxy functions of acetylaminoanthracene-dG **44** was carried out according to the procedure described for the deprotection of acetylaminoanthracene-dG (section 10.7.8.1).

formulation

596 mg (0.61 mmol) **44**
 1.80 mL (1.72 mmol) 1 M solution of tetrabutylammonium fluoride in THF
 0.20 mL (3.49 mmol) acetic acid
 15 mL THF

yield: 346 mg, 0.45 mmol, 74 %

$R_f = 0.55$ ($\text{CH}_2\text{Cl}_2:\text{MeOH}$ 10:1)

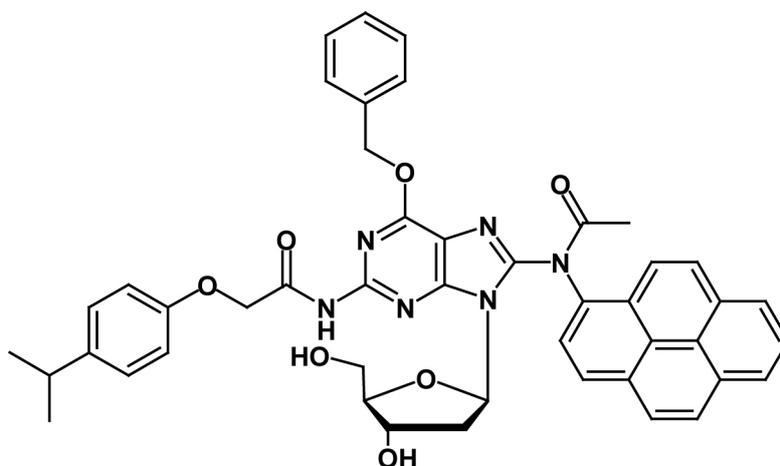
¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 10.60 (s, 1H, *N*²-H), 8.58 (s, 1H, Ant-H_{9/10}), 8.56 (s, 1H, Ant-H_{10/9}), 8.22 – 7.95 (m, 4H, Ant-4H), 7.62 – 7.28 (m, 8 H, Ant-3H + Bz-5H), 7.14 – 7.08 (m, 2 H, Pac-H_{3,5}), 6.86 – 6.80 (m, 2 H, Pac-H_{2,6}), 6.26 (br m, 1H, C_{1'}-H), 5.58 (br s, 2H, Bz-CH₂), 5.19 (br s, 1H, C_{3'}-OH), 4.93 (s, 2H, Pac-CH₂), 4.69 – 4.46 (br m, 2H, C_{5'}-OH + C_{3'}-H), 3.80 (br m, 1H, C_{4'}-H), 3.64 (br m, 1H, C_{5'}-H_a), 3.51 (br m, 1H, C_{5'}-H_b), 3.21 (br m, 1H, C_{2'}-H_a), 2.79 (sept, $J = 6.9$, 1H, *i*Pr-H), 2.13 (br s, 4H, *N*₈-Ac + C_{2'}-H_b), 1.13 (d, $J = 6.9$, 6H, 2*xi*Pr-CH₃).

¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 167.78, 160.26, 156.41, 152.48, 151.93, 146.22, 141.34, 137.13, 136.24, 131.99, 131.12, 130.15, 129.24, 128.92, 128.80, 128.54, 128.39, 127.54, 127.03, 126.60, 126.52, 126.44, 116.35, 114.78, 88.67, 84.56, 71.09, 68.67, 67.79, 62.32, 37.22, 33.01, 24.52, 23.21.

IR ν (cm⁻¹) 3746 (*w*), 3379 (*br s*), 3058 (*m*), 2957 (*m*), 1981 (*w*), 1696 (*s*), 1609 (*s*), 1507 (*s*), 1457 (*m*), 1427 (*m*), 1343 (*s*), 1287 (*m*), 1220 (*s*), 1100 (*w*), 1051 (*m*), 949 (*w*), 890 (*w*), 828 (*m*), 793 (*w*), 737 (*m*), 695 (*m*), 625 (*w*).

HR-MS (ESI⁻) m/z for [C₄₄H₄₁O₇N₆]⁻ ([M-H]⁻): calc.: 765.3037, found 765.3041.

10.7.8.5 8-(*N*-acetyl-1-amino-pyrene)-*O*⁶-benzyl-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine **50**



The deprotection of the 3' and 5' hydroxy functions of acetylaminopyrene-dG **45** was carried out according to the procedure described for the deprotection of acetylaminofluorene-dG (section 10.7.8.1).

formulation

950 mg (0.93 mmol) **45**
 2.80 mL (2.68 mmol) 1 M solution of tetrabutylammonium fluoride in THF
 0.32 mL (5.58 mmol) acetic acid
 22 mL THF

yield: 449 mg, 0.57 mmol, 62 %

$R_f = 0.52$ (CH_2Cl_2 :MeOH 10:1)

¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 8.42 – 8.12 (m, 5H, Pyr-5H), 7.46 – 7.22 (m, 8H, Bz-5H, Pyr-3H), 7.11 (d, $^3J = 8.6$, 2H, Pac-H2), 6.82 (d, $^3J = 8.6$, 2H, Pac-H2), 6.49 (m, 1H, C₁-H), 5.49 (br s, 2H, Bz-CH₂), 5.32 (br d, 1H, C₃-OH), 4.92 (s, 2H, Pac-CH₂), 4.70 (m, 2H, C₅-OH + C₃-H), 3.97 (m, 1H, C₄-H), 3.75 (m, 1H, C₅-H_a), 3.63 (m, 1H, C₅-H_b), 3.15 (br m, 1H, C₂-H_a), 2.78 (sept, $^3J = 6.8$, 1 H, *i*Pr-H), 2.28 (br s, 3H, *N*⁸-Ac), 2.01 (m, 1H, C₂-H_b), 1.12 (d, $^3J = 6.9$, 6H, 2*xi*Pr-CH₃).

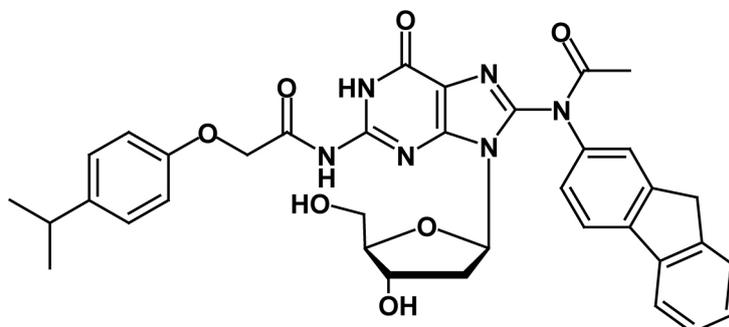
¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 167.77, 156.39, 152.49, 146.77, 141.30, 136.16, 130.98, 130.65, 129.20, 128.85, 127.55, 127.46, 126.54, 125.05, 123.88, 116.17, 114.73, 88.82, 71.18, 68.46, 67.71, 62.46, 49.04, 33.02, 24.54.

IR ν (cm⁻¹) 3854 (*w*), 3752 (*w*), 3392 (*br m*), 2958 (*m*), 1981 (*w*), 1696 (*s*), 1662 (*s*), 1533 (*m*), 1208 (*s*), 1428 (*w*), 1436 (*m*), 1384 (*s*), 1366 (*s*), 1343 (*s*), 1290 (*m*), 1242 (*s*), 1217 (*s*), 1109 (*w*), 1066 (*m*), 1026 (*m*), 952 (*m*), 856 (*s*), 828 (*m*), 792 (*m*), 751 (*m*), 697 (*w*), 623 (*w*).

HR-MS (ESI⁻) m/z for [C₄₆H₄₁N₆O₇]⁻ ([M-H]⁻): calc.: 789.3037, found 789.3036.

10.7.9 Deprotection of the O^6 -position of the aromatic amine dG adducts

10.7.9.1 8-(*N*-acetyl-2-amino-fluorene)- N^2 -isopropylphenoxyacetyl-2'-deoxyguanosine 53



8-(*N*-acetyl-2-amino-fluorene) - O^6 -benzyl - N^2 - isopropylphenoxyacetyl - 2'- deoxyguanosine (101 mg, 0.13 mmol) **48** was dissolved in EtOH (5 mL). A 10% powder of Pd/C (10 mg) and cyclohexene (504 μ L, 5 mmol) were added and the reaction mixture was refluxed for 5 h, cooled and purified by chromatography on silica gel (CH_2Cl_2 :MeOH 40:1) to provide 67 mg (0.10 mmol, 77 %, lit. 88%) of the product as a white powder.

R_f = 0.35 (CH_2Cl_2 :MeOH 10:1)

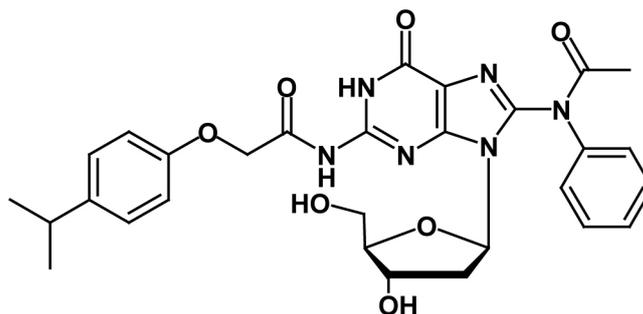
$^1\text{H-NMR}$ (400 MHz, DMSO-d_6): δ (ppm) = 11.99 (s, 1H, N1-H) 11.48 (s, 1H, N2-H), 7.95 (d, $^3J = 8.1$, 1H, AF-1H), 7.91 (d, 1H, $^3J = 7.6$, AF-1H), 7.56 (d, $^3J = 7.2$, 1H, AF-1H), 7.40 – 7.32 (m, 4 H, AF-4H), 7.12 (d, $^3J = 8.5$, 2H, Pac-H2), 6.87 (d, $^3J = 8.8$, 2H, Pac-H2), 6.42 (m, 1H, C_1 -H), 5.18 (br d, 1H, C_3' -OH), 4.90 (s, 2H, Pac- CH_2), 4.81 (m, 1H, C_5' -OH), 4.35 (m, 1H, C_3' -H), 3.92 (s, 2H, AF- C_9 - H_2), 3.81 (m, 1H, C_4' -H), 3.61 (m, 1H, C_5' - H_a), 3.49 (m, 1H, C_5' - H_b), 2.93 (br m, 1H, C_2' - H_a), 2.80 (m, 1 H, *i*Pr-H), 2.03 – 2.01 (m, 4H, N^8 -Ac + C_2' - H_b), 1.14 – 1.11 (m, 6H, 2*xi*Pr- CH_3).

$^{13}\text{C-NMR}$ (101 MHz, DMSO-d_6): δ (ppm) = 156.61, 143.82, 141.17, 127.63, 127.29, 125.60, 120.70, 114.85, 88.61, 84.45, 71.59, 63.27, 62.56, 36.90, 33.01, 32.78, 30.03, 24.51, 23.20.

IR ν (cm^{-1}) 3854 (*w*), 3752 (*w*), 3366 (*br s*), 3057 (*m*), 2957 (*m*), 1684 (*s*), 1607 (*s*), 1560 (*m*), 1506 (*s*), 1455 (*m*), 1431 (*w*), 1367 (*m*), 1336 (*m*), 1293 (*m*), 1247 (*s*), 1178 (*m*), 1099 (*w*), 1082 (*w*), 1031 (*m*), 952 (*w*), 862 (*w*), 828 (*m*), 789 (*w*), 768 (*w*), 745 (*m*), 699 (*w*).

HR-MS (ESI $^-$) m/z for $[\text{C}_{36}\text{H}_{35}\text{N}_6\text{O}_7]^-$ ($[\text{M-H}]^-$): calc.: 663.2567, found: 663.2574.

10.7.9.2 8-(*N*-acetyl-amino-benzene)-*N*²-isopropylphenoxyacetyl-2'-deoxy-guanosine 51



The deprotection of *O*⁶-position of acetylamino-benzene-dG **46** was carried out according to the procedure described for the deprotection of acetylamino-fluorene-dG (section 10.7.9.1).

formulation

560 mg (0.84 mmol) **46**
 56 mg Pd/C 10 %
 3.25 mL (32 mmol) cyclohexene
 33 mL ethanol

yield: 310 mg, 0.53 mmol, 63 %

$R_f = 0.74$ (CH₂Cl₂:MeOH 10:1)

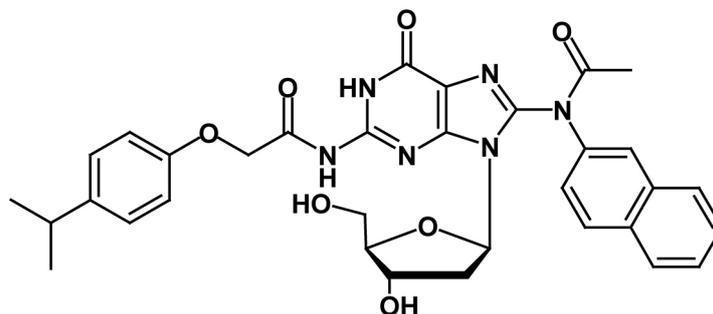
¹H-NMR (400 MHz, DMSO-d₆): δ (ppm) = 7.45 – 7.25 (m, 5H, Phe-5H), 7.14 (d, ³*J* = 8.5, 2H, Pac-H₂), 6.86 (d, ³*J* = 8.2, 2H, Pac-H₂), 6.14 (m, 1H, C_{1'}-H), 5.23 (br d, 1H, C_{3'}-OH), 4.83 (s, 2H, Pac-CH₂), 4.72 (br t, 1H, C_{5'}-OH), 4.43 (m, 2H, C_{3'}-H), 3.78 (m, 1H, C_{4'}-H), 3.61 (m, 1H, C_{5'}-H_a), 3.49 (m, 1H, C_{5'}-H_b), 3.14 (m, 1H, C_{2'}-H_a), 2.81 (m, 1H, *i*Pr-H), 2.02 – 2.00 (m, 4H, *N*⁸-Ac + C_{2'}-H_b), 1.14 (d, ³*J* = 6.7, 6H, 2*xi*Pr-CH₃).

¹³C-NMR (101 MHz, DMSO-d₆): δ (ppm) = 171.53, 156.11, 154.90, 141.72, 129.83, 127.64, 119.57, 114.81, 88.46, 71.22, 66.90, 62.36, 49.02, 33.03, 24.51

IR ν (cm⁻¹) 3370 (*br s*), 3241 (*br s*), 2959 (*m*), 2871 (*w*), 1981 (*w*), 1672 (*s*), 1599 (*s*), 1558 (*s*), 1509 (*m*), 1435 (*w*), 1370 (*w*), 1346 (*m*), 1309 (*m*), 1284 (*m*), 1260 (*m*), 1184 (*s*), 1053 (*m*), 963 (*w*), 833 (*w*), 807 (*w*), 785 (*w*), 760 (*m*), 696 (*m*), 624 (*m*).

HR-MS (ESI⁻) *m/z* for [C₂₉H₃₁N₆O₇]⁻ ([M-H]⁻): calc.: 575.2254, found 575.2273

10.7.9.3 8-(*N*-acetyl-2-amino-naphthalene)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine 52



The deprotection of *O*⁶-position of acetylaminonaphthalene-dG **47** was carried out according to the procedure described for the deprotection of acetylaminofluorene-dG (section 10.7.9.1).

formulation

540 mg (0.84 mmol) **47**
 54 mg Pd/C 10 %
 2.90 mL (28 mmol) cyclohexene
 30 mL ethanol

yield: 320 mg, 0.51 mmol, 68 %

R_f = 0.68 (CH₂Cl₂:MeOH 10:1)

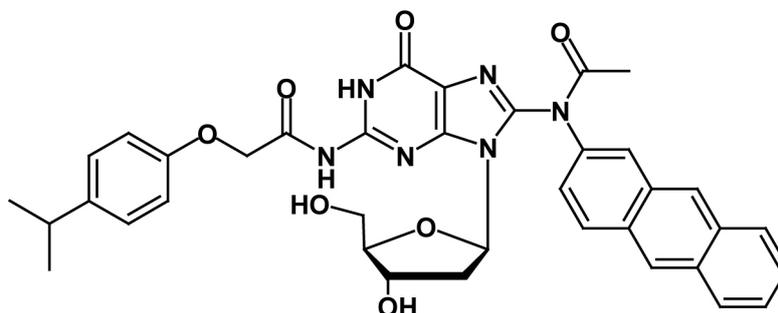
¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 11.92 (s, 1H, *N*¹-H), 11.49 (s, 1H, *N*²-H), 8.09 – 7.82 (m, 4H, Nap-4H), 7.67 – 7.44 (m, 3 H, Nap-3H), 7.17 – 7.08 (m, 2 H, Pac-H_{3,5}), 6.91 – 6.82 (m, 2 H, Pac-H_{2,6}), 6.21 (br m, 1H, C₁-H), 5.23 (br s, 1H, C₃-OH), 4.84 (s, 2H, Pac-CH₂), 4.72 (t, 1H, *J* = 5.7, C₅-OH), 4.43 (br m, 1H, C₃-H), 3.80 (br m, 1H, C₄-H), 3.68 – 3.58 (m, 1H, C₅-H_a), 3.53 – 3.44 (m, 1H, C₅-H_b), 3.01 (br m, 1H, C₂-H_a), 2.81 (sept, *J* = 6.9, 1H, *i*Pr-H), 2.08 (br s, 4H, *N*₈-Ac + C₂-H_b), 1.14 (d, *J* = 6.9, 6H, 2*xi*Pr-CH₃).

¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 171.97, 156.40, 141.32, 136.23, 133.38, 129.27, 128.94, 127.56, 116.29, 114.74, 109.99, 88.90, 70.93, 68.63, 67.66, 62.31, 49.03, 33.02, 24.55, 21.51.

IR ν (cm⁻¹) 3752 (*w*), 3354 (*br s*), 3232 (*br s*), 2957 (*m*), 2870 (*w*), 1988 (*w*), 1682 (*s*), 1599 (*s*), 1558 (*m*), 1507 (*s*), 1434 (*w*), 1367 (*m*), 1337 (*m*), 1281 (*br m*), 1178 (*m*), 1082 (*w*), 1051 (*m*), 959 (*w*), 861 (*w*), 828 (*w*), 784 (*m*), 749 (*m*), 651 (*w*).

HR-MS (ESI⁻) *m/z* for [C₃₃H₃₃N₆O₇]⁻ ([M-H]⁻): calc.: 625.2411, found: 625.2436.

10.7.9.4 8-(*N*-acetyl-2-amino-anthracene)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine 54



The deprotection of *O*⁶-position of acetylaminoanthracene-dG **49** was carried out according to the procedure described for the deprotection of acetylaminofluorene-dG (section 10.7.9.1).

formulation

192 mg (0.25 mmol) **49**
 60 mg Pd/C 10 %
 1.00 mL (9.65 mmol) cyclohexene
 10 mL ethanol

yield: 60 mg, 0.09 mmol, 36 %

$R_f = 0.42$ (CH₂Cl₂:MeOH 10:1)

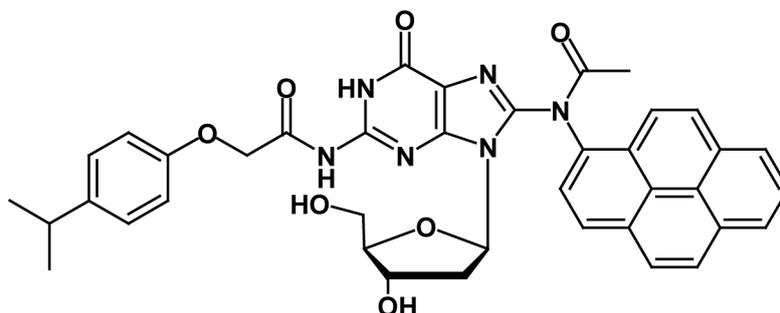
¹H-NMR (400 MHz, DMSO-d₆): δ (ppm) = 11.92 (s, 1H, N₁-H), 11.47 (s, 1H, N²-H), 8.59 (s, 1H, Ant-H_{9/10}), 8.57 (s, 1H, Ant-H_{10/9}), 8.18 – 7.94 (m, 4H, Ant-4H), 7.62 – 7.47 (m, 3 H, Ant-3H), 7.17 – 7.11 (m, 2 H, Pac-H_{3,5}), 6.89 – 6.83 (m, 2 H, Pac-H_{2,6}), 6.22 (br m, 1H, C₁-H), 5.20 (br s, 1H, C_{3'}-OH), 4.83 (s, 2H, Pac-CH₂), 4.69 (br t, 1H, C_{5'}-OH), 4.43 (br m, 1H, C_{3'}-H), 3.81 (br m, 1H, C₄-H), 3.62 (br m, 1H, C₅-H_a), 3.49 (br m, 1H, C₅-H_b), 3.01 (br m, 1H, C₂-H_a), 2.80 (sept, $J = 6.9$, 1H, *i*Pr-H), 2.11 (br s, 4H, N₈-Ac + C₂-H_b), 1.14 (d, $J = 6.9$, 6H, 2*xi*Pr-CH₃).

¹³C-NMR (101 MHz, DMSO-d₆): δ (ppm) = 156.13, 154.93, 141.76, 137.32, 132.00, 131.14, 130.06, 128.55, 128.38, 127.64, 126.95, 126.88, 126.82, 126.59, 126.50, 126.38, 124.95, 119.70, 114.86, 88.46, 84.37, 71.20, 66.96, 62.35, 37.10 33.02, 24.51, 23.23.

IR ν (cm⁻¹) 3354 (*br m*), 3231 (*br m*), 3050 (*w*), 2957 (*m*), 2926 (*m*), 2869 (*m*), 1682 (*s*), 1601 (*s*), 1557 (*m*), 1508 (*s*), 1458 (*w*), 1433 (*w*), 1367 (*m*), 1339 (*m*), 1282 (*br m*), 1254 (*br m*), 1179 (*w*), 1099 (*w*), 1082 (*w*), 1052 (*w*), 958 (*w*), 828 (*w*), 804 (*w*), 783 (*w*), 739 (*m*), 716 (*w*), 640 (*w*), 624 (*w*).

HR-MS (ESI⁻) m/z for [C₃₇H₃₅N₆O₇]⁻ ([M-H]⁻): calc.: 675.2567, found: 675.2566.

10.7.9.5 8-(*N*-acetyl-1-amino-pyrene)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine **55**



The deprotection of *O*⁶-position of acetylaminopyrene-dG **50** was carried out according to the procedure described for the deprotection of acetylaminofluorene-dG (section 10.7.9.1).

formulation

400 mg (0.50 mmol) **50**
 38 mg Pd/C 10 %
 2.00 mL (19 mmol) cyclohexene
 19 mL ethanol

yield: 287 mg, 0.41 mmol, 82 %

$R_f = 0.56$ (CH₂Cl₂:MeOH 10:1)

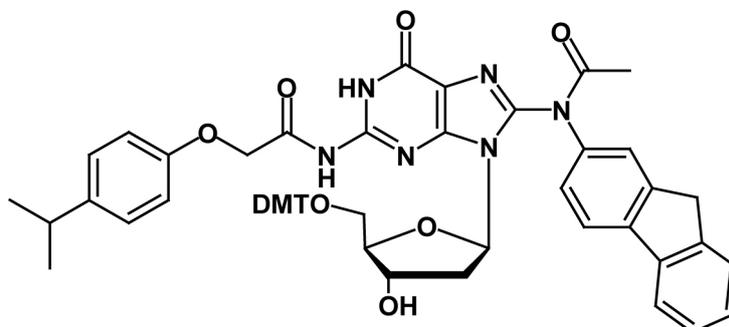
¹H-NMR (400 MHz, DMSO-d₆): δ (ppm) = 11.84 (s, 1H, *N*¹-H) 11.45 (s, 1H, *N*²-H), 8.50 – 8.11 (m, 8H, Pyr-8H), 7.14 (d, ³*J* = 8.6, 2H, Pac-H2), 6.86 (d, ³*J* = 8.6, 2H, Pac-H2), 6.45 (m, 1H, C_{1'}-H), 5.34 (br d, 1H, C_{3'}-OH), 4.83 (s, 2H, Pac-CH₂), 4.78 (m, 1H, C_{5'}-OH), 4.55 (m, 1H, C_{3'}-H), 3.98 (m, 1H, C_{4'}-H), 3.73 (m, 1H, C_{5'}-H_a), 3.61 (m, 1H, C_{5'}-H_b), 3.27 (br m, 1H, C_{2'}-H_a), 2.80 (sept, ³*J* = 6.7, 1 H, *i*Pr-H), 2.26 (m, 1H, C_{2'}-H_b), 1.97 (s, 3H, *N*⁸-Ac), 1.14 (d, ³*J* = 6.9, 6H, 2*xi*Pr-CH₃).

¹³C-NMR (101 MHz, DMSO-d₆): δ (ppm) = 171.43, 156.11, 154.63, 148.31, 141.73, 134.03, 130.99, 128.68, 127.65, 127.47, 126.36, 125.41, 123.90, 114.83, 88.66, 71.41, 66.88, 62.31, 56.01, 36.83, 33.02, 28.85, 24.52.

IR ν (cm⁻¹) 3391 (*br m*), 3234 (*br m*), 2958 (*m*), 1682 (*s*), 1602 (*s*), 1558 (*m*), 1507 (*s*), 1458 (*w*), 1435 (*w*), 1369 (*m*), 1344 (*w*), 1291 (*m*), 1257 (*m*), 1180 (*m*), 1052 (*m*), 965 (*w*), 850 (*m*), 829 (*w*), 784 (*w*), 710 (*w*), 682 (*w*), 632 (*w*).

HR-MS (ESI⁻) *m/z* for [C₃₉H₃₅N₆O₇]⁻ ([M-H]⁻): calc.: 699.2567 found: 699.2575

10.7.10 5'-DMT protection of the bulky dG adducts

10.7.10.18-(*N*-acetyl-2-amino-fluorene)-(5'-*O*-dimethoxytrityl)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine **58**

8-(*N*-acetyl-2-amino-fluorene)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine (133 mg, 0.20 mmol) **53** was dissolved in pyridine (1 mL). 4,4'-Dimethoxytrityl chloride (102 mg, 0.3 mmol, 1.5 eq.) was added and the reaction mixture was stirred for 2 h at room temperature, concentrated and purified by chromatography on silica gel (CH₂Cl₂:MeOH 40:1, 1 % pyridine) to provide 162 mg (0.17 mmol, 85 %, lit. 76 %) of a white foam.

R_f = 0.56 (CH₂Cl₂:MeOH 10:1)

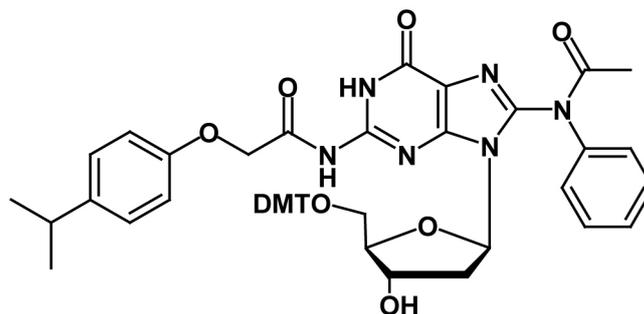
¹H-NMR (400 MHz, CDCl₃): δ (ppm) = 11.7 (br s, 2H, *N*¹-H + *N*²-H), 7.75 – 7.66 (m, 2H, AF-2H), 7.52 (d, ³*J* = 7.5 Hz, 1H, AF-1H), 7.40 – 7.23 (m, 6 H, AF-4H + DMT-2H), 7.21 – 7.09 (m, 9H, Pac-H2 + DMT-7H), 6.87 – 6.67 (m, 6H, Pac-H2 + DMT-4H), 6.19 (m, 1H, C₁-H), 5.25 (d, 1H, C₃-OH), 4.66 (s, 2H, Pac-CH₂), 4.48 (m, 1H, C₃-H), 4.05 (s, 2H, AF-C9-H₂), 3.82 (m, 1H, C₄-H), 3.79 (s, 3H, -OMe), 3.78 (s, 3H, -OMe), 3.42 (m, 1H, C₅-H_a), 3.26 (m, 1H, C₅-H_b), 3.11 (m, 1H, C₂-H_a), 2.87 (sept, ³*J* = 6.7 Hz, 1 H, *i*Pr-H), 2.15 (s, 3H, *N*⁸-Ac), 1.95 (m, 1H, C₂-H_b), 1.22 (m, 3H, *i*Pr-CH₃), 1.20 (m, 3H, *i*Pr-CH₃).

¹³C-NMR (101 MHz, CDCl₃): δ (ppm) = 158.74, 154.23, 149.82, 144.74, 143.73, 143.65, 139.46, 136.09, 135.60, 129.96, 129.96, 129.09, 128.00, 127.80, 127.73, 125.07, 123.71, 115.12, 113.14, 113.02, 86.25, 67.54, 55.23, 55.15, 36.96, 33.32, 29.68, 24.05, 22.60.

IR ν (cm⁻¹) 3855 (*w*), 3746 (*w*), 3198 (*br m*), 2958 (*m*), 2360 (*m*), 2338 (*m*), 1980 (*w*), 1682 (*s*), 1602 (*s*), 1557 (*m*), 1506 (*s*), 1455 (*w*), 1436 (*m*), 1366 (*w*), 1282 (*m*), 1246 (*s*), 1175 (*s*), 1031 (*m*), 959 (*w*), 827 (*m*), 767 (*w*), 736 (*m*), 702 (*m*).

HR-MS (ESI⁻) *m/z* for [C₅₇H₅₃N₆O₉]⁻ ([M-H]⁻): calc.: 965.3874, found 965.3834.

10.7.10.2 8-(*N*-acetyl-amino-benzene)-(5'-*O*-dimethoxytrityl)-*N*²-isopropylphenoxyacetyl-2'-deoxyguaosine **56**



The 5'-DMT protection of acetylaminobenzene-dG **51** was carried out according to the procedure described for the protection of acetylaminofluorene-dG (section 10.7.10.1).

formulation

320 mg (0.51 mmol) **51**
 260 mg (0.77 mmol, 1.5 eq.) DMT-Cl
 2.90 mL pyridine

yield: 472 mg, 0.50 mmol, 98 %

R_f = 0.81 (CH₂Cl₂:MeOH 10:1)

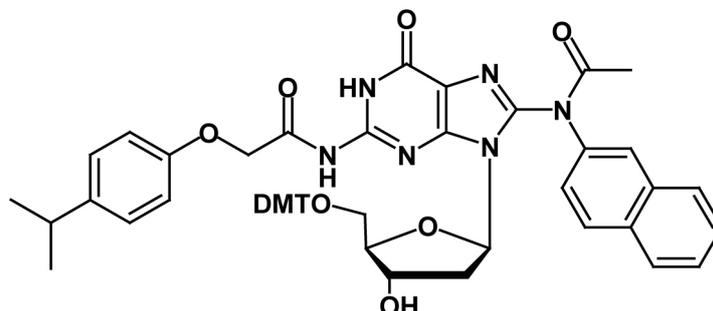
¹H-NMR (400 MHz, DMSO-d₆): δ (ppm) = 11.74 (s, 2H, *N*¹-H), 11.32 (s, 2H, *N*²-H), 7.37 – 7.11 (m, 16H, Phe-5H + DMT-9H + Pac-H₂), 6.85 (d, ³*J* = 8.6, 2H, Pac-H₂), 6.70 (m, 4H, DMT-4H), 6.24 (m, 1H, C₁-H), 5.17 (br d, 1H, C₃-OH), 4.79 (s, 2H, Pac-CH₂), 4.54 (m, 2H, C_{3'}-H), 3.99 (m, 1H, C₄-H), 3.64 (s, 6H, 2xOMe), 3.62 (m, 1H, C_{5'}-H_a), 3.36 (m, 1H, C_{5'}-H_b), 3.10 (m, 1H, C₂-H_a), 2.83 (sept, ³*J* = 6.9, 1 H, *i*Pr-H), 2.19 (br m, 1H, C₂-H_b), 2.04 (s, 3H, *N*⁸-Ac), 1.14 (d, ³*J* = 6.9, 6H, 2x*i*Pr-CH₃).

¹³C-NMR (101 MHz, DMSO-d₆): δ (ppm) = 150.07, 136.55, 130.12, 130.10, 129.73, 128.15, 127.79, 127.64, 127.63, 127.62, 126.79, 124.31, 124.30, 114.83, 114.82, 114.81, 114.81, 113.53, 113.16, 113.02, 87.08, 84.11, 71.15, 66.86, 58.54, 55.40, 32.91, 24.38, 22.92, 16.12.

IR ν (cm⁻¹) 3855 (*w*), 3746 (*w*), 3198 (*br m*), 3057 (*m*), 2958 (*m*), 2932 (*m*), 2869 (*m*), 2836 (*m*), 1683 (*s*), 1605 (*s*), 1556 (*m*), 1506 (*s*), 1456 (*w*), 1438 (*m*), 1367 (*m*), 1334 (*w*), 1282 (*m*), 1247 (*s*), 1175 (*m*), 1075 (*m*), 1021 (*m*), 960 (*w*), 827 (*m*), 784 (*w*), 754 (*w*), 727 (*m*), 698 (*m*), 614 (*w*).

HR-MS (ESI⁻) *m/z* for [C₅₀H₄₉O₉N₆]⁻ ([M-H]⁻): calc.: 877.3561, found: 877.3593

10.7.10.3 8-(*N*-acetyl-2-amino-naphthalene)-(5'-*O*-dimethoxytrityl)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine 57



The 5'-DMT protection of acetylaminonaphthalene-dG **52** was carried out according to the procedure described for the protection of acetylaminofluorene-dG (section 10.7.10.1).

formulation

322 mg (0.51 mmol) **52**
 260 mg (0.77 mmol, 1.5 eq.) DMT-Cl
 2.90 mL pyridine

yield: 460 mg, 0.49 mmol, 96 %

R_f = 0.76 (CH₂Cl₂:MeOH 10:1)

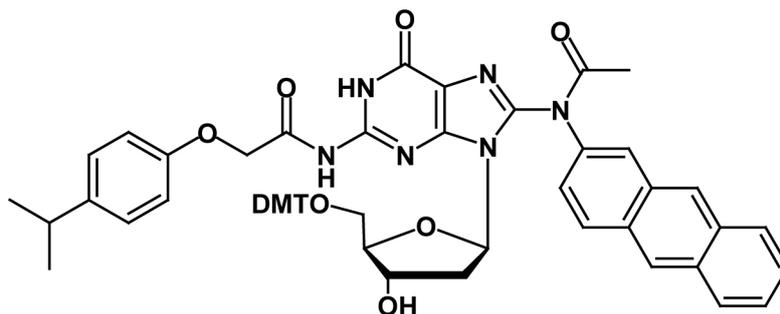
¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 11.74 (s, 1H, *N*¹-H), 11.33 (s, 1H, *N*²-H), 7.98 – 7.78 (br m, 4H, Nap-4H), 7.64 – 7.39 (br m, 3 H, Nap-3H), 7.30 – 7.01 (m, 11 H, DMT-9H, + Pac-2H), 6.91 – 6.77 (m, 2 H, Pac-2H), 6.73 – 6.50 (br m, 4H, DMT-4H), 6.31 (br m, 1H, C_{1'}-H), 5.17 (br s, 1H, C_{3'}-OH), 4.83 (d, *J* = 16, 1H, Pac-CH₂), 4.70 (d, *J* = 16, 1H, Pac-CH₂), 4.55 (br m, 1H, C_{3'}-H), 4.03 (br m, 1H, C_{4'}-H), 3.64 (s, 3H, OCH₃), 3.61 (s, 3H, OCH₃), 3.37 (br m, 1H, C_{5'}-H_a), 3.08 (br m, 1H, C_{5'}-H_b), 2.92 (very br m, 1H, C_{2'}-H_a), 2.82 (sept, *J* = 6.9, 1H, *i*Pr-H), 2.10 (br s, 4H, *N*₈-Ac + C_{2'}-H_b), 1.14 (d, *J* = 6.9, 6H, 2*xi*Pr-CH₃).

¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 150.05, 136.56, 130.10, 130.07, 129.34, 128.15, 128.01, 127.78, 127.63, 127.63, 127.63, 127.15, 126.78, 124.32, 124.31, 114.84, 113.21, 71.25, 66.85, 66.82, 65.39, 32.91, 24.39, 23.03.

IR ν (cm⁻¹) 3746 (w), 3204 (br m), 3058 (w), 2999 (w), 2958 (m), 2931 (m), 2869 (w), 2835 (w), 1981 (w), 1684 (s), 1604 (s), 1556 (m), 1506 (s), 1438 (m), 1367 (m), 1333 (w), 1282 (m), 1246 (s), 1175 (m), 1114 (w), 1075 (m), 1031 (m), 961 (w), 860 (w), 826 (m), 789 (w), 750 (m), 702 (m), 669 (w), 651 (w).

HR-MS (ESI⁻) *m/z* for [C₅₄H₅₁O₉N₆]⁻ ([M-H]⁻): calc.: 927.3718, found: 927.3768.

10.7.10.4 8-(*N*-acetyl-2-amino-anthracene)-(5'-*O*-dimethoxytrityl)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine 59



The 5'-DMT protection of acetylaminanthracene-dG **54** was carried out according to the procedure described for the protection of acetylaminofluorene-dG (section 10.7.10.1).

formulation

60 mg (0.09 mmol) **54**
 45 mg (0.13 mmol, 1.6 eq.) DMT-Cl
 0.50 mL pyridine

yield: 56 mg, 0.06 mmol, 66 %

R_f = 0.55 (CH₂Cl₂:MeOH 10:1)

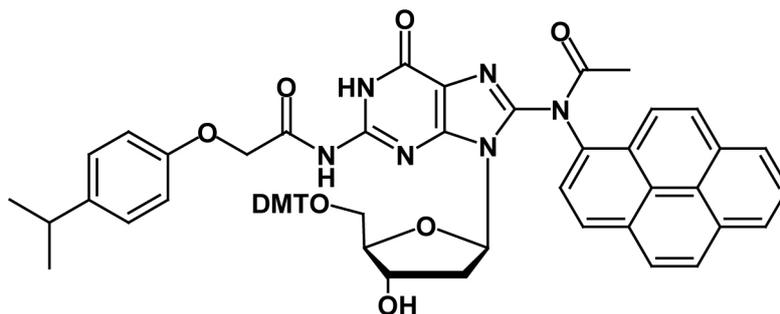
¹H-NMR (400 MHz, DMSO-d₆): δ (ppm) = 11.74 (s, 1H, *N*¹-H), 11.31 (s, 1H, *N*²-H), 8.46 (m, 2H, Ant-2H), 8.16 – 8.01 (m, 4H, Ant-4H), 7.66 – 7.42 (m, 3 H, Ant-3H), 7.28 – 7.22 (m, 2H, DMT-2H), 7.17 – 6.99 (m, 9 H, DMT-7H, + Pac-2H), 6.89 – 6.80 (m, 2 H, Pac-2H), 6.69 – 6.52 (br m, 4H, DMT-4H), 6.34 (br m, 1H, C₁-H), 5.15 (br s, 1H, C_{3'}-OH), 4.81 (d, *J* = 15.9, 1H, Pac-CH₂), 4.70 (d, *J* = 16.0, 1H, Pac-CH₂), 4.54 (br m, 1H, C₃-H), 4.02 (br m, 1H, C₄-H), 3.61 (br s, 6H, OCH₃), 3.37 (br m, 1H, C₅-H_a), 3.09 (br m, 1H, C₅-H_b), 3.01 (very br m, 1H, C₂-H_a), 2.82 (sept, *J* = 6.9, 1H, *i*Pr-H), 2.14 (br s, 4H, *N*₈-Ac + C₂-H_b), 1.14 (d, *J* = 6.9, 6H, 2*xi*Pr-CH₃).

¹³C-NMR (101 MHz, DMSO-d₆): δ (ppm) = 158.13, 156.13, 154.70, 150.04, 145.24, 141.92, 136.54, 136.03, 131.93, 131.15, 128.22, 127.81, 127.65, 124.32, 119.73, 114.90, 113.17, 113.06, 87.29, 85.59, 66.98, 55.31, 33.02, 24.50.

IR ν (cm⁻¹) 3200 (*br m*), 3054 (*m*), 3000 (*m*), 2958 (*m*), 2363 (*w*), 1681 (*s*), 1606 (*s*), 1504 (*s*), 1437 (*s*), 1367 (*w*), 1282 (*br m*), 1247 (*s*), 1175 (*m*), 1050 (*w*), 1031 (*m*), 901 (*br w*), 827 (*m*), 746 (*m*), 702 (*s*), 641 (*w*), 613 (*w*).

HR-MS (ESI⁻) *m/z* for [C₅₈H₅₃N₆O₉]⁻ ([*M*-H]⁻): calc.: 977.3874, found: 977.3848.

10.7.10.5 8-(*N*-acetyl-1-amino-pyrene)-(5'-*O*-dimethoxytrityl)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine 60



The 5'-DMT protection of acetylamino-pyrene-dG **55** was carried out according to the procedure described for the protection of acetylamino-fluorene-dG (section 10.7.10.1).

formulation

250 mg (0.36 mmol) **55**
 185 mg (0.55 mmol, 1.5 eq.) DMT-Cl
 2.00 mL pyridine

yield: 351 mg, 0.35 mmol, 99 %

R_f = 0.68 (CH₂Cl₂:MeOH 10:1)

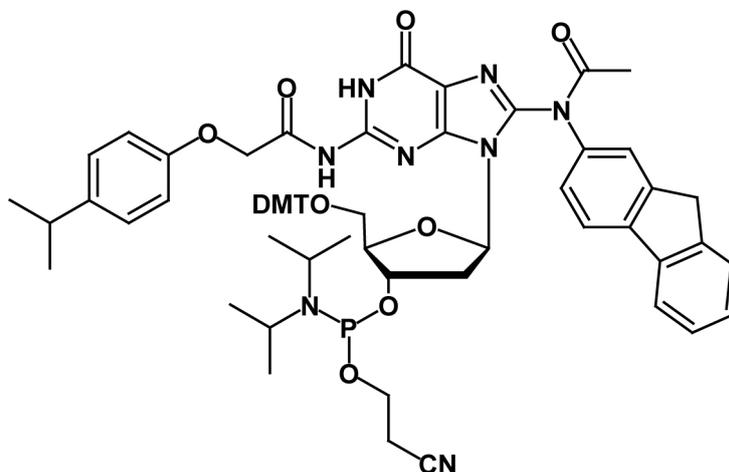
¹H-NMR (400 MHz, DMSO-*d*₆): δ (ppm) = 11.66 (s, 1H, *N*¹-H) 11.29 (s, 1H, *N*²-H), 8.49 – 8.05 (m, 8H, Pyr-8H), 7.37 – 7.01 (m, 11H, DMT-9H + Pac-H₂), 6.84 (d, ³*J* = 8.8, 2H, Pac-H₂), 6.65 – 6.57 (m, 4H, DMT-4H), 6.51 (m, 1H, C_{1'}-H), 5.27 (br d, 1H, C_{3'}-OH), 4.79 (s, 2H, Pac-CH₂), 4.69 (m, 1H, C_{3'}-H), 4.18 (m, 1H, C_{4'}-H), 3.70 (s, 6H, 2xOMe), 3.57 (m, 1H, C_{5'}-H_a), 3.49 (m, 1H, C_{5'}-H_b), 3.17 (br m, 1H, C_{2'}-H_a), 2.81 (sept, ³*J* = 6.9, 1 H, *i*Pr-H), 2.05 (br m, 1H, C_{2'}-H_b), 1.91 (s, 3H, *N*⁸-Ac), 1.14 (d, ³*J* = 6.9, 6H, 2x*i*Pr-CH₃).

¹³C-NMR (101 MHz, DMSO-*d*₆): δ (ppm) = 170.90, 158.26, 158.16, 156.12, 150.04, 145.31, 141.88, 140.65, 136.57, 136.00, 130.99, 130.19, 129.34, 128.21, 128.07, 127.85, 127.65, 126.80, 125.91, 124.34, 123.91, 114.86, 113.18, 87.33, 85.60, 84.63, 66.90, 55.43, 33.02, 24.51.

IR ν (cm⁻¹) 3397 (*br w*), 3203 (*br m*), 3052 (*m*), 2957 (*m*), 2931 (*m*), 2869 (*w*), 2835 (*w*), 1916 (*w*), 1682 (*s*), 1603 (*s*), 1556 (*m*), 1509 (*s*), 1460 (*w*), 1437 (*m*), 1367 (*m*), 1336 (*w*), 1284 (*m*), 1246 (*s*), 1174 (*s*), 1075 (*m*), 1031 (*m*), 963 (*w*), 901 (*w*), 846 (*m*), 826 (*m*), 783 (*w*), 753 (*w*), 726 (*w*), 702 (*m*), 681 (*w*), 633 (*w*).

HR-MS (ESI⁻) *m/z* for [C₆₀H₅₃N₆O₉]⁻ ([M-H]⁻): calc.: 1001.3874, found: 1001.3846.

10.7.11 Synthesis of the phosphoramidites

10.7.11.1 8-(*N*-acetyl-2-amino-fluorene)-3'-*O*-(2'-cyanoethoxydiisopropylamino-phosphino)-(5'-*O*-dimethoxytrityl)-*N*²-isopropylphenoxyacetyl-2'-deoxy-guanosine 63

8-(*N*-acetyl-2-amino-fluorene)-(5'-*O*-dimethoxytrityl)-*N*²-isopropylphenoxyacetyl-2'-deoxy-guanosine (150 mg, 0.16 mmol) **58** was dissolved in CH₂Cl₂ (1.7 mL). Diisopropylammonium tetrazolide (106 mg, 0.61 mmol, 4 eq., DIAT) was added and the suspension was degassed flushed with argon in a flame dried and degassed apparatus. The entire reaction was carried out under argon atmosphere. Bis-(diisopropylamino)-(2-cyanoethoxy)-phosphine (92 μL, 0.32 mmol, 2 eq.) was added and the reaction mixture was stirred for 2 h at room temperature. The solvents were subsequently removed *in vacuo* and the crude product was rapidly purified by chromatography on silica gel (CH₂Cl₂: MeOH 40:1, 1 % pyridine) to provide 163 mg (0.14 mmol, 90 %, lit. 76 %) of a white foam.

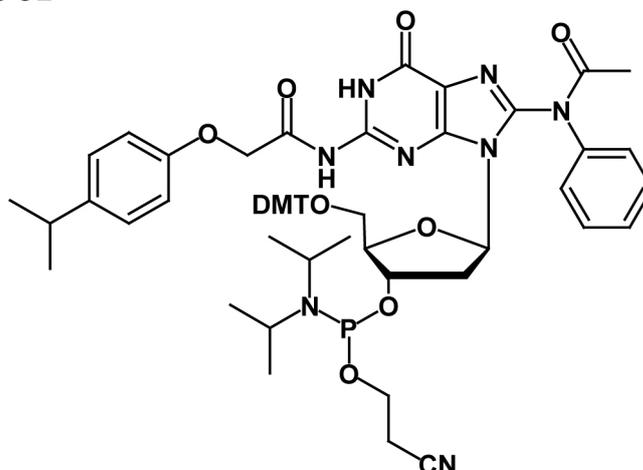
R_f = 0.80 (CH₂Cl₂/MeOH 10:1)

¹H-NMR (200 MHz, DMSO-*d*₆): δ (ppm) = 11.76 (s, 1H, *N*¹-H), 10.93 (s, 1H, *N*²-H), 7.91 – 6.53 (m, 24H, AF-7H + DMT-13H + Pac-H₄), 6.28 (m, 1H, C_{1'}-H), 4.78 (s, 2H, Pac-CH₂), 4.64 (m, 1H, C_{3'}-H), 4.13 (m, 1H, C_{4'}-H), 3.83 (m, 2H, AF-C9-2H), 3.63 – 3.62 (m, 6H, 2xOMe), 3.48 – 3.38 (m, 5H, C_{5'}-H_a + PO-CH₂), 3.22 – 3.05 (m, 2H, C_{5'}-H_b + C_{2'}-H_a), 2.85 (m, 1H, *i*Pr-H), 2.71 – 2.50 (m, 2H, N-2*i*Pr-2H), 2.34 (m, 1H, C_{2'}-H_b), 2.08 (s, 3H, *N*⁸-Ac), 1.19 – 1.02 (m, 18H, 2Pac-*i*Pr-CH₃ + 4N-*i*Pr-CH₃).

Impurities exist due to hydrolysis in DMSO; peaks which are unambiguously identified are given.

³¹P-NMR {¹H} (81 MHz, DMSO-*d*₆): δ (ppm) = 148.83, 148.33.

10.7.11.2 8-(*N*-acetyl-amino-benzene)-3'-*O*-(2'-cyanoethoxydiisopropylamino-phosphino)-(5'-*O*-dimethoxytrityl)-*N*²-isopropylphenoxyacetyl-2'-deoxy-guanosine 61



The phosphitylation of acetylaminobenzene-dG **56** was carried out according to the procedure described for the protection of acetylaminofluorene-dG (section 10.7.11.1).

formulation

100 mg (0.11 mmol) **56**
 80 mg (0.30 mmol) DIAT
 67 μ L (0.22 mmol, 2 eq) Bis-(diisopropylamino)-(2-cyanoethoxy)-phosphine
 1 mL CH₂Cl₂

yield: 98 mg, 0.10 mmol, 90 %

R_f = 0.55 (CH₂Cl₂:MeOH 10:1)

The ¹³C-NMR is very complicated; signal splitting is caused by phosphorus coupling and by the existence of two diastereomeric forms; significant signals are listed. Impurities exist due to hydrolysis in DMSO; peaks which are unambiguously identified are given.

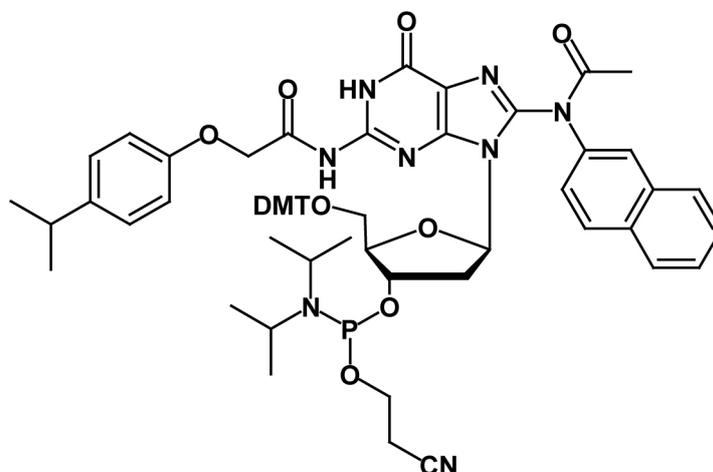
¹H-NMR (400 MHz, DMSO-d₆): δ (ppm) = 7.37 – 7.11 (m, 16H, Phe-5H + DMT-9H + Pac-H₂), 6.79 – 6.62 (br m, 6 H, Pac-2H + DMT-4H), 6.20 – 6.10 (br m, 1H, C₁-H), 4.75 – 4.66 (br m, 1H, C₃-H), 4.23 – 4.18 (m, 2H, Pac-CH₂), 3.98 – 3.81 (m, 3H, PO-CH₂ + C₄-H), 3.81–3.49 (m, 9H, OCH₃+ N-*i*Pr₂-2H + C₄-H), 3.46 – 3.35 (br m, 2H, C₅H₂), 3.24 (br m, 1H, C₂-H_a), 2.90 – 2.71 (m, 3H, *i*Pr-H + CH₂-CN), 2.38 (br m, 1H, C₂-H_b), 2.02 (br s, 3H, N₈-Ac); 1.20 – 1.07 (m, 18H, 2-Pac-*i*Pr-CH₃ + 4N-*i*Pr-CH₃).

¹³C-NMR (101 MHz, DMSO-d₆): δ (ppm) = 173.53, 162.62, 150.95, 150.55, 149.61, 146.33, 145.30, 119.44, 119.03, 115.19, 113.20, 79.73, 79.32, 78.91, 58.75, 58.60, 58.55, 55.36, 50.76, 50.59, 46.33, 44.95, 44.88, 42.84, 42.71, 24.92, 24.85, 24.78, 23.07, 23.05, 23.03, 20.33, 20.26, 19.82, 19.75.

³¹P-NMR {¹H} (81 MHz, CDCl₃): δ (ppm) = 150.57 (s)

HR-MS (ESI⁺) m/z for [C₅₉H₅₈N₈O₁₀P]⁺ ([M+H]⁺): calc.: 1079.4796 found 1079.4798.

10.7.11.3 8-(*N*-acetyl-2-amino-naphthalene)-3'-O-(2'-cyanoethoxydiisopropyl-amino-phosphino)-(5'-*O*-dimethoxytrityl)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine 62



The phosphitylation of acetylaminonaphthalene-dG **57** was carried out according to the procedure described for the protection of acetylaminofluorene-dG (section 10.7.11.1).

formulation

100 mg (0.11 mmol) **57**
 80 mg (0.30 mmol) DIAT
 67 μ L (0.22 mmol, 2 eq) Bis-(diisopropylamino)-(2-cyanoethoxy)-phosphine
 1 mL CH₂Cl₂

yield: 96 mg, 0.096 mmol, 87 %

R_f = 0.67 (CH₂Cl₂:MeOH 10:1)

¹H-NMR (400 MHz, CDCl₃): δ (ppm) = 11.32 (very br s, 1H, *N*¹-H), 8.70 (very br s, 1H, *N*²-H), 7.90 – 7.10 (complex m, 18H, Nap-7H + DMT-9H + Pac-2H), 6.84 – 6.77 (m, 2 H, Pac-2H), 6.69 – 6.57 (br m, 4H, DMT-4H), 6.19 – 6.05 (br m, 1H, C₁-H), 4.86 – 4.71 (br m, 1H, C_{3'}-H), 4.53 – 4.38 (m, 2H, Pac-CH₂), 4.26 – 4.07 (m, 3H, PO-CH₂ + C_{4'}-H), 3.68 (br s, 6H, OCH₃), 3.62–3.44 (m, 3H, *N*-*i*Pr₂-H₂ + C_{4'}-H), 3.37 - 3.19 (br m, 2H, C₅H₂), 2.92 (very br m, 1H, C₂-H_a), 2.86 (sept, J = 6.9, 1H, *i*Pr-H), 2.78 – 2.96 (m, 2H, CH₂-CN), 2.39 (br m, 1H, C₂-H_b), 2.19 (br s, 3H, *N*₈-Ac); 1.21 (d, J = 6.9, 6H, Pac-*i*Pr-CH₃), 1.19 – 1.05 (m, 12H, 4*N*-*i*Pr-CH₃).

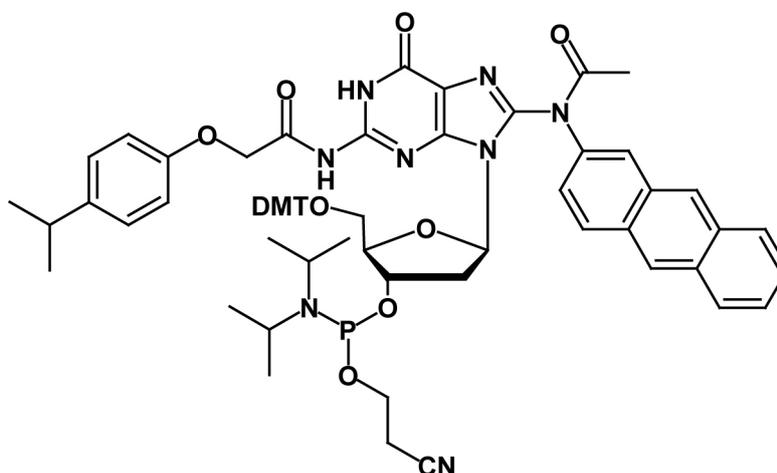
The ¹³C-NMR is very complicated; signal splitting is caused by phosphorus coupling and by the existence of two diastereomeric forms; significant signals are listed. Impurities exist due to hydrolysis in chloroform; peaks which are unambiguously identified are given.

$^{13}\text{C-NMR}$ (101 MHz, CDCl_3): δ (ppm) = 157.39, 153.78, 153.76, 143.78, 143.75, 142.52, 134.85, 134.82, 134.75, 132.47, 129.05, 128.99, 128.96, 128.11, 127.19, 127.15, 127.08, 126.82, 126.71, 126.67, 126.60, 125.70, 125.62, 119.56, 116.53, 116.44, 114.12, 114.05, 112.15, 112.01, 111.91, 111.87, 84.99, 66.35, 66.29, 63.05, 57.53, 57.33, 57.14, 57.08, 54.18, 54.12, 44.34, 44.27, 42.34, 42.21, 42.03, 41.90, 32.32, 23.69, 23.67, 23.60, 23.52, 23.44, 23.39, 23.09, 19.14, 19.07.

$^{31}\text{P-NMR}$ $\{^1\text{H}\}$ (81 MHz, CDCl_3): δ (ppm) = 150.10 (s)

HR-MS (ESI $^-$) m/z for $[\text{C}_{63}\text{H}_{68}\text{N}_8\text{O}_{10}\text{P}]^-$ ($[\text{M-H}]^-$): calc.: 1127.4796, found 1127.4740.

10.7.11.4 8-(*N*-acetyl-2-amino-anthracene)-3'-*O*-(2'-cyanoethoxydiisopropyl-amino-phosphino)-(5'-*O*-dimethoxytrityl)-*N*²-isopropylphenoxyacetyl-2'-deoxyguanosine 64



The phosphitylation of acetylaminanthracene-dG **59** was carried out according to the procedure described for the protection of acetylaminofluorene-dG (section 10.7.11.1).

formulation

56 mg (0.056 mmol) **59**

40 mg (0.15 mmol) DIAT

35 μL (0.12 mmol, 2 eq) Bis-(diisopropylamino)-(2-cyanoethoxy)-phosphine

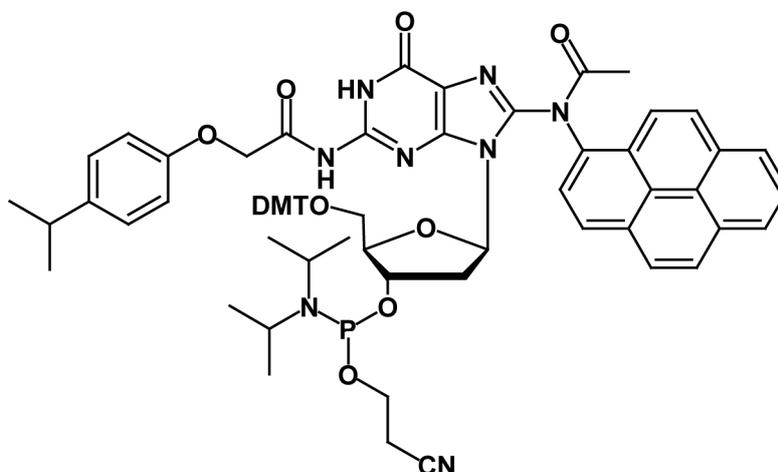
0.6 mL CH_2Cl_2

yield: 50 mg, 0.042 mmol, 75 %

R_f = 0.83 (CH_2Cl_2 :MeOH 10:1)

$^{31}\text{P-NMR}$ $\{^1\text{H}\}$ (81 MHz, CDCl_3): δ (ppm) = 150.54, 150.19.

10.7.11.5 8-(*N*-acetyl-1-amino-pyrene)-3'-*O*-(2'-cyanoethoxydiisopropylamino-phosphino)-(5'-*O*-dimethoxytrityl)-*N*²-isopropylphenoxyacetyl-2'-deoxy-guanosine 65



The phosphitylation of acetylaminopyrene-dG **60** was carried out according to the procedure described for the protection of acetylaminofluorene-dG (section 10.7.11.1).

formulation

160 mg (0.16 mmol) **60**
 110 mg (0.64 mmol) DIAT
 95 μ L (0.32 mmol, 2 eq) Bis-(diisopropylamino)-(2-cyanoethoxy)-phosphine
 1.5 mL CH₂Cl₂

yield: 153 mg, 0.12 mmol, 75 %

R_f = 0.56 (CH₂Cl₂:MeOH 10:1)

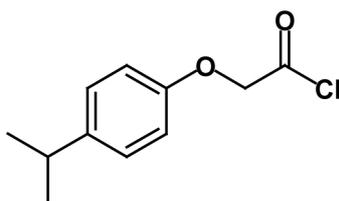
The ¹³C-NMR is very complicated; signal splitting is caused by phosphorus coupling and by the existence of two diastereomeric forms; significant signals are listed. Impurities exist due to hydrolysis in chloroform; peaks which are unambiguously identified are given.

¹H-NMR (600 MHz, CDCl₃): δ (ppm) = 8.50 – 8.10 (m, 8H, Pyr-8H), 7.00 – 6.51 (m, 17H, DMT-13H + 2 Pac-H₂), 6.48 – 6.39 (br m, 1H, C₁-H), 4.86 – 4.71 (br m, 1H, C₃'-H), 5.04 – 4.93 (m, 2H, Pac-CH₂), 4.54 – 4.36 (m, 3H, PO-CH₂ + C₄'-H), 3.74 – 3.25 (m, 11H, OCH₃ + 2*N*-*i*Pr-H + C₄'-H + C₅'-H₂), 2.92 (very br m, 1H, C₂'-H_a), 2.88 (sept, *J* = 6.74, 1H, *i*Pr-H), 2.79 – 2.73 (m, 2H, CH₂-CN), 2.05 (br m, 1H, C₂'-H_b), 1.96 (br s, 3H, N₈-Ac); 1.34 – 1.15 (m, 18H, Pac-*i*Pr-CH₃ + 4*N*-*i*Pr-CH₃).

¹³C-NMR (151 MHz, CDCl₃): δ (ppm) = 158.13, 154.74, 149.83, 143.30, 135.84, 130.70, 130.09, 129.30, 128.22, 127.55, 126.58, 126.15, 125.43, 123.80, 117.49, 116.81, 115.07, 113.06, 112.70, 85.95, 67.29, 59.83, 58.19, 58.07, 55.28, 50.58, 45.24, 43.24, 42.88, 33.30, 24.09, 22.98, 20.34, 19.99.

³¹P-NMR {¹H} (81 MHz, CDCl₃): δ (ppm) = 150.01, 149.98.

10.7.12 (4-Isopropyl)-phenoxy acetyl chloride



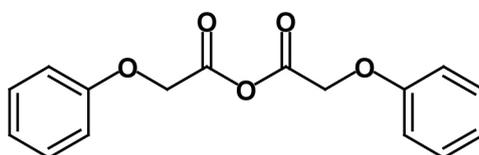
(4-Isopropyl)-phenoxy acetic acid (5 g, 25.7 mmol) and thionyl chloride (3.60 mL, 50 mmol, 2 eq) were heated to 70 °C for 30 min. After removal of excess thionyl chloride *in vacuo*, the product was purified by vacuum distillation (150 °C bath temperature, BP = 104 °C at 5 mbar) to provide 4.86 g (22.8 mmol, 89 %) of a colorless liquid.^[170]

¹H-NMR (300 MHz, CDCl₃): δ (ppm) = 7.21 (m, 2H, H_{3,5}), 6.68 (m, 2H, H_{2,6}), 4.95 (s, 1H, CH₂), 2.90 (sept, ³J = 6.9 Hz, 1H, CH), 1.52 (d, ³J = 6.9 Hz, 3 H, CH₃), 1.42 (d, ³J = 6.9 Hz, 3H, CH₃).

¹³C-NMR (75 MHz, CDCl₃): δ (ppm) = 170.51, 155.01, 143.29, 127.62, 114.76, 73.00.

IR ν (cm⁻¹) 3589 (w), 3128 (w), 3037 (w), 2961 (s), 2928 (w), 2872 (w), 1877 (br w), 1806 (s), 1609 (w), 1588 (w), 1509 (s), 1461 (w), 1417 (w), 1289 (w), 1264 (w), 1217 (s), 1182 (m), 1089 (w), 964 (w), 934 (s), 827 (m), 761 (m), 682 (w), 637 (w).

10.7.13 Phenoxyacetic anhydride



To a solution of phenoxyacetic acid (31 g, 209 mmol) in diethyl ether (150 mL) and dioxane (150 mL) was added *N,N'*-dicyclohexylcarbodiimide (22.3 g, 108 mmol, 0.52 eq.). The reaction mixture was stirred at room temperature for 3 h, subsequently cooled to 0 °C and stirred for an additional hour. The resultant dicyclohexylurea removed by filtration, the solution was concentrated *in vacuo* and kept overnight at -20 °C for crystallization. The product was purified by recrystallization from diethyl ether to provide 23.4 g (81 mmol, 87 %) of white crystals.^[171]

¹H-NMR (200 MHz, CDCl₃) δ (ppm) = 7.35 (dd (J = 7.3, 7.8), 4 H, *m*-aryl), 7.13 (t (J = 7.8), 2 H, *p*-aryl), 7.00 (d (J = 7.3), 4 H, *o*-aryl), 4.82 (s, 4 H, Ph-O-CH₂).

IR ν (cm⁻¹) 3853 (w), 3745 (w), 3061 (m), 2983 (w), 2929 (m), 2574 (m), 2188 (w), 2169 (w), 1827 (s), 1757 (m), 1731 (m), 1700 (s), 1596 (m), 1585 (m), 1494 (s), 1429 (m), 1418 (m), 1374 (m), 1307 (w), 1288 (w), 1269 (m), 1224 (s), 1089 (s), 1044 (s), 746 (s), 687 (s), 612 (w).

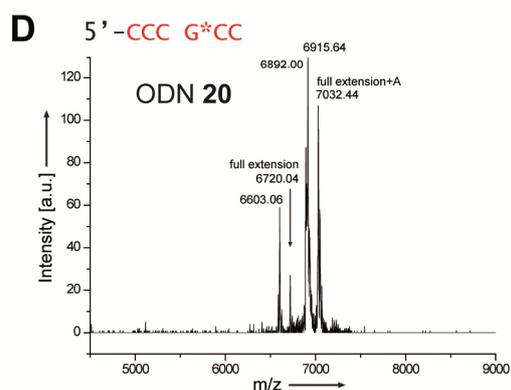
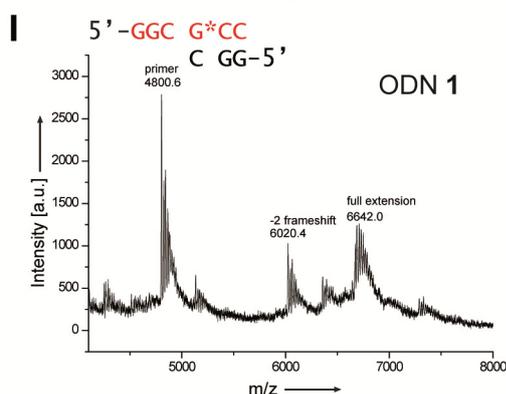
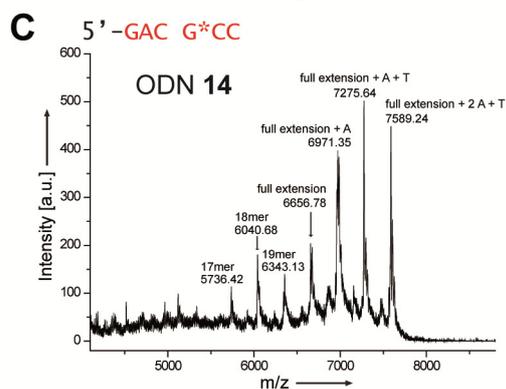
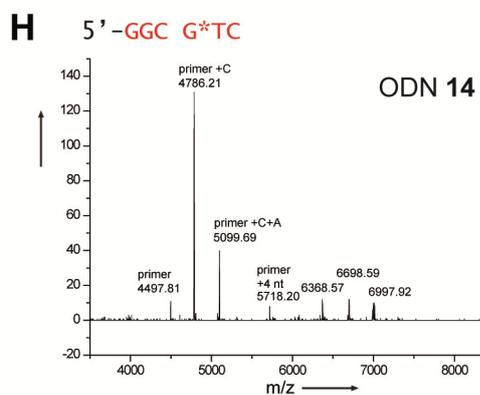
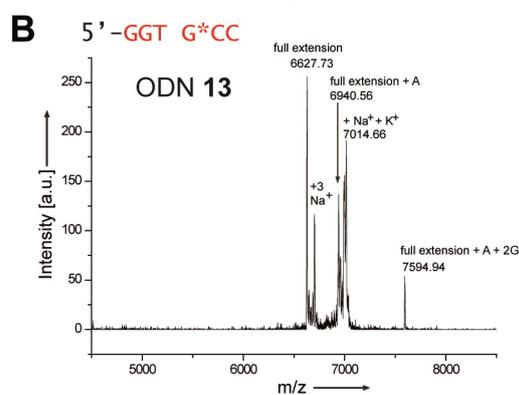
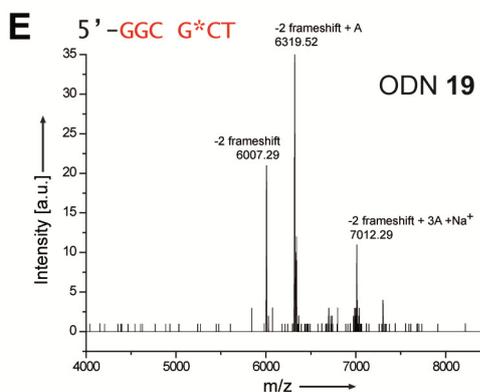
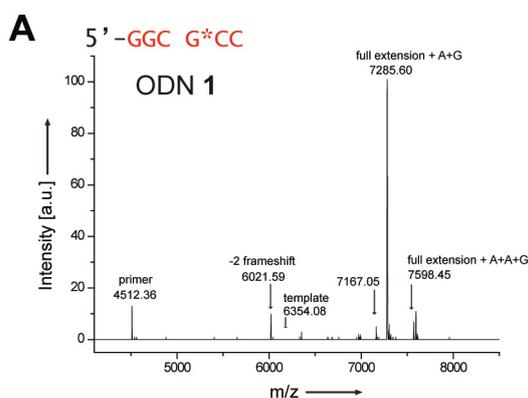
11 Appendix

11.1 Crystal structure data

	Pol η -AAA-dG	Pol η -AAF-dG
Data collection		
Space group	P 4 ₁ 2 ₁ 2	P4 ₁ 2 ₁ 2
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	103.5, 103.5, 292.7	103.44, 103.44, 292.35
α , β , γ (°)	90, 90, 90	90, 90, 90
Wavelength (Å)	0.919	0.9129
Resolution (Å)	48.0 - 2.7 (2.85-2.7)	49-2.7 (2.85-2.7)
<i>R</i> _{merge}	0.067 (0.377)	0.107 (0.523)
Mean <i>I</i> / σI	23.0 (6.0)	26.84 (5.99)
No. of observations	215,900 (30,490)	584,408 (92,986)
No. of unique reflections	44,732 (6,421)	44,372 (6,887)
Completeness (%)	99.9 (99.8)	98.2 (96.5)
Redundancy	4.8 (4.7)	13.1 (13.6)
Refinement		
Resolution (Å)	46.6 - 2.7	48.76 - 2.7
No. reflections	42,326	44,599
<i>R</i> _{work} / <i>R</i> _{free}	21.5 / 25.6	22.0 / 27.2
No. atoms		9,070
Protein	8,147	8,152
DNA	826	824
Waters	103	89
Ions	6 calcium	6 calcium
<i>B</i> -factors		
Protein	45.0	29.6
DNA	65.3	45.1
Waters	36.0	20.5
R.m.s deviations		
Bond lengths (Å)	0.0126	0.0120
Bond angles (°)	1.439	1.442

Table 10. Data collection, processing and structure refinement statistics. Numbers in parentheses correspond to the high resolution shell.

11.2 MALDI-Tof mass spectrometry analysis of the primer extension products with AAF-dG containing oligonucleotides



Calculated masses (m/z)

A: primer 4513.6; -2 frameshift 6023.6; template 6356.1; full extension +AG 7284.4; full extension + 2 A+G 7597.6
B: full extension 6626.0; full extension + A; 6939.2 full extension + A+2G 7597.6
C: full extension 6657.0; full extension + AT 7274.4; full extension + 2A+T 7587.6
D: full extension 6722.0; full extension + A 7035.2
E: -2 frameshift 6007.6; -2 frameshift + A 6939.2; -2 frameshift + A+2G+Na⁺ 7015.9
H: primer 4497.6; primer + C 4786.8, primer + CA 5100.0
I: primer 4802.8; -2 frameshift 6023.6; full extension + K⁺ 6681.1

11.3 Composition of the *LeMaster* Medium ^[164]

chemical	1 L	2 L	4 L
Alanine	0.5 g	1.0 g	2.0 g
Arginine HCl	0.58 g	1.16 g	2.32 g
Aspartic acid	0.41 g	0.80 g	1.6 g
Cystine	0.03 g	0.06 g	0.12 g
Glutamic acid	0.67 g	1.34 g	2.68 g
Glutamine	0.33 g	0.66 g	1.32 g
Glycine	0.54 g	1.08 g	2.16 g
Histidine	0.06 g	0.12 g	0.24 g
Isoleucine	0.23 g	0.46 g	0.92 g
Leucine	0.23 g	0.46 g	0.92 g
Lysine HCl	0.42 g	0.84 g	1.68 g
Phenylalanine	0.13 g	0.26 g	0.52 g
Proline	0.10 g	0.20 g	0.40 g
Serine	2.08 g	4.16 g	8.32 g
Threonine	0.23 g	0.46 g	0.92 g
Tyrosine	0.17 g	0.34 g	0.68 g
Valine	0.23 g	0.46 g	0.92 g
Adenine	0.50 g	1.0 g	2.0 g
Guanosine	0.67 g	1.34 g	2.68 g
Thymine	0.17 g	0.34 g	0.68 g
Uracil	0.50 g	1.0 g	2.0 g
Sodium acetate	1.50 g	3.0 g	6.0 g
Succinic acid	1.5 g	3.0 g	6.0 g
Ammonium chloride	0.75 g	1.50 g	3.0 g

Sodium hydroxide	1.08 g	2.16 g	4.32 g
anhydrous K₂HPO₄	8.0 g	16.0 g	32.0 g

add ddH₂O, autoclave and adjust pH to pH 7.5

non- autoclavable portion for 100 ml ddH₂O (sterile filtered)
(900 ml autoclaved medium + 100 ml non-autoclavable portion)

Glucose	10 g
MgSO₄ x 7 H₂O	0.25 g
FeSO₄	4.2 mg
conc. H₂SO₄	8 µL
Thiamine HCl	5 mg

DL-selenomethionine	50 mg/L
----------------------------	---------

11.4 List of abbreviations

Å	Angström
AAA	2-Acetylaminoanthracene
AAB	2-Acetylaminobenzene
AAF	2-Acetylaminofluorene
AAN	2-Acetylaminonaphthalene
AAP	1-Acetylaminopyrene
AAA-dG	Acetylaminoanthracene-C8-dG adduct
AAB-dG	Acetylaminobenzene-C8-dG adduct
AAF-dG	Acetylaminofluorene-C8-dG adduct
AF-dG	Aminofluorene-C8-dG adduct
AAN-dG	Acetylaminonaphthene-C8-dG adduct
AAP-dG	Acetylaminopyrene-C8-dG adduct
ATT	6-Aza-2-thiothymine
BNZ	Benzyl
BER	Base excision repair
BINAP	rac-2,2'-Bis(diphenylphosphino)-1,1'-binaphthyl
bp	Base pair
BSA	Bovine serum albumine
conc.	Concentrated
dA	2'-Deoxyadenosine
Da	Dalton
dba	Dibenzylideneacetone
dC	2'-Deoxycytidine
DCM	Dichloromethane
dG	2'-Deoxyguanosine
DMF	Dimethylformamide
DMT	Dimethoxytrityl
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
ddNTP	2', 3' Dideoxynucleoside triphosphate
ds	Double stranded
dT	2'-Deoxythymidine
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylendiamine tetraacetate
EI	Electron ionization
ESI	Elektrospray ionisation
EtBr	Ethidium bromide
EtOAc	Ethyl acetate
eq.	Equivalents
h	Hour
HPLC	High pressure liquid chromatography

HR-MS	High resolution mass spectrometry
IR	Infra red
IPTG	Isopropylthiogalactoside
kb	Kilobase
LB	Lysogeny broth
TFA	Trifluoroacetic acid
TLC	Thin layer chromatography
TLS	Translesion Synthesis
MeCN	Acetonitrile
min	Minute
M.p.	Melting point
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
n	Amount of substance
NaOtBu	Sodium <i>tert</i> -butoxide
NER	nucleotide excision repair
NMR	Nuclear magnetic resonance
OD	Optical density
PAGE	Polyacrylamid gel electrophoresis
PCR	Polymerase Chain Reaction
Pd ₂ (dba) ₃	Tris-(dibenzylideneacetone)-dipalladium (0)
ppm	Parts per million
RNA	Ribonucleic acid
rpm	Rotation per minute
rt	Room temperature
s	Second
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SDS	Sodium dodecylsulfate
ss	Single stranded
TBDMS	<i>tert</i> -Butyldimethylsilyl
TEMED	N,N,N',N'-Tetramethylenediamine
THF	Tetrahydrofurane
Tris	Tris(-hydroxymethyl)-aminomethane
U	Units
UV	Ultraviolet
vol.	Volume
w/v	Weight per volume

12 References

- [1] Lindahl, T.; Nyberg, B., Rate of depurination of native deoxyribonucleic acid. *Biochemistry* **1972**, 11, (19), 3610-8.
- [2] Lindahl, T.; Nyberg, B., Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry* **1974**, 13, (16), 3405-10.
- [3] Rahn, R. O.; Hosszu, J. L., Photochemical studies of thymine in ice. *Photochem. Photobiol.* **1969**, 10, (2), 131-7.
- [4] Rahn, R. O.; Setlow, J. K.; Hosszu, J. L., Ultraviolet inactivation and photoproducts of transforming DNA irradiated at low temperatures. *Biophys. J.* **1969**, 9, (4), 510-7.
- [5] Cadet, J.; Bellon, S.; Berger, M.; Bourdat, A. G.; Douki, T.; Duarte, V.; Frelon, S.; Gasparutto, D.; Muller, E.; Ravanat, J. L.; Sauvaigo, S., Recent aspects of oxidative DNA damage: guanine lesions, measurement and substrate specificity of DNA repair glycosylases. *Biol. Chem.* **2002**, 383, (6), 933-43.
- [6] Poirier, M. C., Chemical-induced DNA damage and human cancer risk. *Nat. Rev. Cancer* **2004**, 4, (8), 630-7.
- [7] Friedberg, E. C.; Walker, G. C.; Siede, W., DNA repair and mutagenesis. *ASM Press, Washington D.C* **1995**.
- [8] Case, R. A.; Hosker, M. E.; Mc, D. D.; Pearson, J. T., Tumours of the urinary bladder in workmen engaged in the manufacture and use of certain dyestuff intermediates in the British chemical industry. I. The role of aniline, benzidine, alpha-naphthylamine, and beta-naphthylamine. *Br. J. Ind. Med.* **1954**, 11, (2), 75-104.
- [9] Radomski, J. L., The primary aromatic amines: their biological properties and structure-activity relationships. *Annu. Rev. Pharmacol. Toxicol.* **1979**, 19, 129-57.
- [10] Pfeifer, G. P.; Denissenko, M. F.; Olivier, M.; Tretyakova, N.; Hecht, S. S.; Hainaut, P., Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene* **2002**, 21, (48), 7435-51.
- [11] Xue, W.; Warshawsky, D., Metabolic activation of polycyclic and heterocyclic aromatic hydrocarbons and DNA damage: a review. *Toxicol. Appl. Pharmacol.* **2005**, 206, (1), 73-93.
- [12] Schut, H. A.; Snyderwine, E. G., DNA adducts of heterocyclic amine food mutagens: implications for mutagenesis and carcinogenesis. *Carcinogenesis* **1999**, 20, (3), 353-68.
- [13] Saffhill, R.; Margison, G. P.; O'Connor, P. J., Mechanisms of carcinogenesis induced by alkylating agents. *Biochim. Biophys. Acta* **1985**, 823, (2), 111-45.
- [14] Hartter, D. R., The use and importance of nitroaromatic chemicals in the chemical industry. In: D.E. Rickert, Toxicity of Nitroaromatic Compounds, Hemisphere, New York **1985**, 1-13.
- [15] Rosenkranz, H. S., Mutagenic and carcinogenic nitroarenes in diesel emissions: risk identification. *Mutat. Res.* **1984**, 140, (1), 1-6.
- [16] Ames, B. N.; Gurney, E. G.; Miller, J. A.; Bartsch, H., Carcinogens as frameshift mutagens: metabolites and derivatives of 2-acetylaminofluorene and other aromatic amine carcinogens. *Proc. Natl. Acad. Sci. USA* **1972**, 69, (11), 3128-32.

- [17] Miller, J. A.; Miller, E. C., Some historical aspects of *N*-aryl carcinogens and their metabolic activation. *Environ. Health Perspect.* **1983**, 49, 3-12.
- [18] Beland, F. A.; Kadlubar, F. F., Formation and persistence of arylamine DNA adducts *in vivo*. *Environ. Health Perspect.* **1985**, 62, 19-30.
- [19] Hoffmann, G. R.; Fuchs, R. P., Mechanisms of frameshift mutations: insight from aromatic amines. *Chem. Res. Toxicol.* **1997**, 10, (4), 347-59.
- [20] Yasui, M.; Dong, H.; Bonala, R. R.; Suzuki, N.; Ohmori, H.; Hanaoka, F.; Johnson, F.; Grollman, A. P.; Shibutani, S., Mutagenic properties of 3-(deoxyguanosin-*N*2-yl)-2-acetylaminofluorene, a persistent acetylaminofluorene-derived DNA adduct in mammalian cells. *Biochemistry* **2004**, 43, (47), 15005-13.
- [21] Gill, J. P.; Romano, L. J., Mechanism for *N*-acetyl-2-aminofluorene-induced frameshift mutagenesis by *Escherichia coli* DNA polymerase I (Klenow fragment). *Biochemistry* **2005**, 44, (46), 15387-95.
- [22] Watt, D. L.; Utzat, C. D.; Hilario, P.; Basu, A. K., Mutagenicity of the 1-nitropyrene-DNA adduct *N*-(deoxyguanosin-8-yl)-1-aminopyrene in mammalian cells. *Chem. Res. Toxicol.* **2007**, 20, (11), 1658-64.
- [23] Heflich, R. H.; Neft, R. E., Genetic toxicity of 2-acetylaminofluorene, 2-aminofluorene and some of their metabolites and model metabolites. *Mutat. Res.* **1994**, 318, (2), 73-114.
- [24] Frederick, C. B.; Mays, J. B.; Ziegler, D. M.; Guengerich, F. P.; Kadlubar, F. F., Cytochrome P-450- and flavin-containing monooxygenase-catalyzed formation of the carcinogen *N*-hydroxy-2-aminofluorene and its covalent binding to nuclear DNA. *Cancer Res.* **1982**, 42, (7), 2671-7.
- [25] Hammons, G. J.; Guengerich, F. P.; Weis, C. C.; Beland, F. A.; Kadlubar, F. F., Metabolic oxidation of carcinogenic arylamines by rat, dog, and human hepatic microsomes and by purified flavin-containing and cytochrome P-450 monooxygenases. *Cancer Res.* **1985**, 45, (8), 3578-85.
- [26] Flammang, T. J.; Yamazoe, Y.; Benson, R. W.; Roberts, D. W.; Potter, D. W.; Chu, D. Z.; Lang, N. P.; Kadlubar, F. F., Arachidonic acid-dependent peroxidative activation of carcinogenic arylamines by extrahepatic human tissue microsomes. *Cancer Res.* **1989**, 49, (8), 1977-82.
- [27] Cramer, J. W.; Miller, J. A.; Miller, E. C., *N*-Hydroxylation: A new metabolic reaction observed in the rat with the carcinogen 2-acetylaminofluorene. *J. Biol. Chem.* **1960**, 235, 885-8.
- [28] Miller, E. C.; Miller, J. A.; Hartmann, H. A., *N*-Hydroxy-2-acetylaminofluorene: a metabolite of 2-acetylaminofluorene with increased carcinogenic activity in the rat. *Cancer Res.* **1961**, 21, 815-24.
- [29] Guengerich, F. P.; Shimada, T., Oxidation of toxic and carcinogenic chemicals by human cytochrome P-450 enzymes. *Chem. Res. Toxicol.* **1991**, 4, (4), 391-407.
- [30] Shimada, T.; Oda, Y.; Gillam, E. M.; Guengerich, F. P.; Inoue, K., Metabolic activation of polycyclic aromatic hydrocarbons and other procarcinogens by cytochromes P450 1A1 and P450 1B1 allelic variants and other human cytochromes P450 in *Salmonella typhimurium* NM2009. *Drug Metab. Dispos.* **2001**, 29, (9), 1176-82.

- [31] Kadlubar, F. F.; Miller, J. A.; Miller, E. C., Hepatic microsomal *N*-glucuronidation and nucleic acid binding of *N*-hydroxy arylamines in relation to urinary bladder carcinogenesis. *Cancer Res.* **1977**, *37*, (3), 805-14.
- [32] DeBaun, J. R.; Smith, J. Y.; Miller, E. C.; Miller, J. A., Reactivity *in vivo* of the carcinogen *N*-hydroxy-2-acetylaminofluorene: increase by sulfate ion. *Science* **1970**, *167*, (915), 184-6.
- [33] Lai, C. C.; Miller, J. A.; Miller, E. C.; Liem, A., *N*-sulfooxy-2-aminofluorene is the major ultimate electrophilic and carcinogenic metabolite of *N*-hydroxy-2-acetylaminofluorene in the livers of infant male C57BL/6J x C3H/HeJ F1 (B6C3F1) mice. *Carcinogenesis* **1985**, *6*, (7), 1037-45.
- [34] Kadlubar, F. F.; Unruh, L. E.; Flammang, T. J.; Sparks, D.; Mitchum, R. K.; Mulder, G. J., Alteration of urinary levels of the carcinogen, *N*-hydroxy-2-naphthylamine, and its *N*-glucuronide in the rat by control of urinary pH, inhibition of metabolic sulfation, and changes in biliary excretion. *Chem. Biol. Interact.* **1981**, *33*, (2-3), 129-47.
- [35] Flammang, T. J.; Kadlubar, F. F., Acetyl coenzyme A-dependent metabolic activation of *N*-hydroxy-3,2'-dimethyl-4-aminobiphenyl and several carcinogenic *N*-hydroxy arylamines in relation to tissue and species differences, other acyl donors, and arylhydroxamic acid-dependent acyltransferases. *Carcinogenesis* **1986**, *7*, (6), 919-26.
- [36] Shinohara, A.; Saito, K.; Yamazoe, Y.; Kamataki, T.; Kato, R., Acetyl coenzyme A dependent activation of *N*-hydroxy derivatives of carcinogenic arylamines: mechanism of activation, species difference, tissue distribution, and acetyl donor specificity. *Cancer Res.* **1986**, *46*, (9), 4362-7.
- [37] Bartsch, H.; Dworkin, M.; Miller, J. A.; Miller, E. C., Electrophilic *N*-acetoxyaminoarenes derived from carcinogenic *N*-hydroxy-*N*-acetyl aminoarenes by enzymatic deacetylation and transacetylation in liver. *Biochim. Biophys. Acta* **1972**, *286*, (2), 272-98.
- [38] King, C. M., Mechanism of reaction, tissue distribution, and inhibition of arylhydroxamic acid acyltransferase. *Cancer Res.* **1974**, *34*, (6), 1503-15.
- [39] Beland, F. A.; Allaben, W. T.; Evans, F. E., Acyltransferase-mediated binding of *N*-hydroxyarylamides to nucleic acids. *Cancer Res.* **1980**, *40*, (3), 834-40.
- [40] Boteju, L. W.; Hanna, P. E., Bioactivation of *N*-hydroxy-2-acetylaminofluorenes by *N,O*-acyltransferase: substituent effects on covalent binding to DNA. *Carcinogenesis* **1993**, *14*, (8), 1651-7.
- [41] Bartsch, H.; Miller, J. A.; Miller, E. C., *N*-acetoxy-*N*-acetyl aminoarenes and nitrosoarenes. One-electron non-enzymatic and enzymatic oxidation products of various carcinogenic aromatic acethydroxamic acids. *Biochim. Biophys. Acta* **1972**, *273*, (1), 40-51.
- [42] Howard, P. C.; Casciano, D. A.; Beland, F. A.; Shaddock, J. G., Jr., The binding of *N*-hydroxy-2-acetylaminofluorene to DNA and repair of the adducts in primary rat hepatocyte cultures. *Carcinogenesis* **1981**, *2*, (2), 97-102.
- [43] Jones, J. W.; Robins, R. K., Purine Nucleosides. Methylation Studies of Certain Naturally Occurring Purine Nucleosides. *J. Amer. Chem. Society* **1963**, *85*, 193-201.
- [44] Humphreys, W. G.; Kadlubar, F. F.; Guengerich, F. P., Mechanism of C8 alkylation of guanine residues by activated arylamines: evidence for initial adduct formation at the N7 position. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, (17), 8278-82.

- [45] Novak, M.; Kennedy, S. A., Selective Trapping of *N*-acetyl-*N*-(4-biphenyl)nitrenium and *N*-Acetyl-*N*-(2-fluorenyl)nitrenium Ions by 2'-Desoxyguanosine in Aqueous Solution. *J. Amer. Chem. Society* **1995**, 117, 574-575.
- [46] Bedard, L. L.; Massey, T. E., Aflatoxin B1-induced DNA damage and its repair. *Cancer Lett.* **2006**, 241, (2), 174-83.
- [47] Jerina, D. M.; Chadha, A.; Cheh, A. M.; Schurdak, M. E.; Wood, A. W.; Sayer, J. M., Covalent bonding of bay-region diol epoxides to nucleic acids. *Adv. Exp. Med. Biol.* **1991**, 283, 533-53.
- [48] Sage, E.; Leng, M., Conformation of poly(dG-dC) . poly(dG-dC) modified by the carcinogens *N*-acetoxy-*N*-acetyl-2-aminofluorene and *N*-hydroxy-*N*-2-aminofluorene. *Proc. Natl. Acad. Sci.* **1980**, 77, (8), 4597-601.
- [49] Santella, R. M.; Grunberger, D.; Broyde, S.; Hingerty, B. E., Z-DNA conformation of *N*-2-acetylaminofluorene modified poly(dG-dC).poly(dG-dC) determined by reactivity with anti cytidine antibodies and minimized potential energy calculations. *Nucleic Acids Res.* **1981**, 9, (20), 5459-67.
- [50] Marques, M. M.; Beland, F. A., Synthesis, characterization, and solution properties of ras sequences modified by arylamine carcinogens at the first base of codon 61. *Chem. Res. Toxicol.* **1990**, 3, (6), 559-65.
- [51] Hsu, G. W.; Kiefer, J. R.; Burnouf, D.; Becherel, O. J.; Fuchs, R. P.; Beese, L. S., Observing translesion synthesis of an aromatic amine DNA adduct by a high-fidelity DNA polymerase. *J. Biol. Chem.* **2004**, 279, (48), 50280–50285.
- [52] Fuchs, R.; Daune, M., Physical studies on deoxyribonucleic acid after covalent binding of a carcinogen. *Biochemistry* **1972**, 11, (14), 2659-66.
- [53] Fuchs, R. P.; Schwartz, N.; Daune, M. P., Hot spots of frameshift mutations induced by the ultimate carcinogen *N*-acetoxy-*N*-2-acetylaminofluorene. *Nature* **1981**, 294, (5842), 657-9.
- [54] O'Handley, S. F.; Sanford, D. G.; Xu, R.; Lester, C. C.; Hingerty, B. E.; Broyde, S.; Krugh, T. R., Structural characterization of an *N*-acetyl-2-aminofluorene (AAF) modified DNA oligomer by NMR, energy minimization, and molecular dynamics. *Biochemistry* **1993**, 32, (10), 2481-97.
- [55] Florian, J., Borden, J., Structure and torsional flexibility of the linkage between guanine and fluorene residues in the deoxyguanosine -aminofluorene and deoxyguanosine-acetylaminofluorene carcinogenic adducts. *Theor. Chem. Acc.* **2005**, 113, 28-34.
- [56] Livneh, Z., DNA damage control by novel DNA polymerases: translesion replication and mutagenesis. *J. Biol. Chem.* **2001**, 276, (28), 25639-42.
- [57] Masutani, C.; Kusumoto, R.; Iwai, S.; Hanaoka, F., Mechanisms of accurate translesion synthesis by human DNA polymerase eta. *Embo J.* **2000**, 19, (12), 3100-9.
- [58] Matsuda, T.; Bebenek, K.; Masutani, C.; Hanaoka, F.; Kunkel, T. A., Low fidelity DNA synthesis by human DNA polymerase-eta. *Nature* **2000**, 404, (6781), 1011-3.
- [59] Rothwell, P. J.; Waksman, G., Structure and mechanism of DNA polymerases. *Adv. Protein Chem.* **2005**, 71, 401-40.
- [60] Vooradi, V.; Romano, L. J., Effect of *N*-2-acetylaminofluorene and 2-aminofluorene adducts on DNA binding and synthesis by yeast DNA polymerase eta. *Biochemistry* **2009**, 48, (19), 4209-16.

- [61] Koffel-Schwartz, N.; Fuchs, R. P., Sequence determinants for -2 frameshift mutagenesis at NarI-derived hot spots. *J. Mol. Biol.* **1995**, *252*, (5), 507-13.
- [62] Fuchs, R. P.; Koffel-Schwartz, N.; Pelet, S.; Janel-Bintz, R.; Napolitano, R.; Becherel, O. J.; Broschard, T. H.; Burnouf, D. Y.; Wagner, J., DNA polymerases II and V mediate respectively mutagenic (-2 frameshift) and error-free bypass of a single *N*-2-acetylaminofluorene adduct. *Biochem. Soc. Trans.* **2001**, *29*, (Pt 2), 191-5.
- [63] Shibutani, S.; Suzuki, N.; Grollman, A. P., Mechanism of frameshift (deletion) generated by acetylaminofluorene-derived DNA adducts *in vitro*. *Biochemistry* **2004**, *43*, (50), 15929-35.
- [64] Fuchs, R. P.; Fujii, S., Translesion synthesis in Escherichia coli: lessons from the NarI mutation hot spot. *DNA Repair* **2007**, *6*, (7), 1032-41.
- [65] Garcia, A.; Lambert, I. B.; Fuchs, R. P., DNA adduct-induced stabilization of slipped frameshift intermediates within repetitive sequences: implications for mutagenesis. *Proc. Natl. Acad. Sci.* **1993**, *90*, (13), 5989-93.
- [66] Milhe, C.; Fuchs, R. P.; Lefevre, J. F., NMR data show that the carcinogen *N*-2-acetylaminofluorene stabilises an intermediate of -2 frameshift mutagenesis in a region of high mutation frequency. *Eur. J. Biochem.* **1996**, *235*, (1-2), 120-7.
- [67] Roy, D.; Hingerty, B. E.; Shapiro, R.; Broyde, S., A slipped replication intermediate model is stabilized by the syn orientation of *N*-2-aminofluorene- and *N*-2-(acetyl)aminofluorene-modified guanine at a mutational hotspot. *Chem. Res. Toxicol.* **1998**, *11*, (11), 1301-11.
- [68] Broschard, T. H.; Koffel-Schwartz, N.; Fuchs, R. P., Sequence-dependent modulation of frameshift mutagenesis at NarI-derived mutation hot spots. *J. Mol. Biol.* **1999**, *288*, (1), 191-9.
- [69] Bebenek, K.; Kunkel, T. A., Frameshift errors initiated by nucleotide misincorporation. *Proc. Natl. Acad. Sci.* **1990**, *87*, (13), 4946-50.
- [70] Shibutani, S.; Grollman, A. P., On the mechanism of frameshift (deletion) mutagenesis *in vitro*. *J. Biol. Chem.* **1993**, *268*, (16), 11703-10.
- [71] Yang, W.; Woodgate, R., What a difference a decade makes: insights into translesion DNA synthesis. *Proc. Natl. Acad. Sci.* **2007**, *104*, (40), 15591-8.
- [72] Burgers, P. M.; Koonin, E. V.; Bruford, E.; Blanco, L.; Burtis, K. C.; Christman, M. F.; Copeland, W. C.; Friedberg, E. C.; Hanaoka, F.; Hinkle, D. C.; Lawrence, C. W.; Nakanishi, M.; Ohmori, H.; Prakash, L.; Prakash, S.; Reynaud, C. A.; Sugino, A.; Todo, T.; Wang, Z.; Weill, J. C.; Woodgate, R., Eukaryotic DNA polymerases: proposal for a revised nomenclature. *J. Biol. Chem.* **2001**, *276*, (47), 43487-90.
- [73] Ohmori, H.; Friedberg, E. C.; Fuchs, R. P.; Goodman, M. F.; Hanaoka, F.; Hinkle, D.; Kunkel, T. A.; Lawrence, C. W.; Livneh, Z.; Nohmi, T.; Prakash, L.; Prakash, S.; Todo, T.; Walker, G. C.; Wang, Z.; Woodgate, R., The Y-family of DNA polymerases. *Mol. Cell* **2001**, *8*, (1), 7-8.
- [74] Prakash, S.; Johnson, R. E.; Prakash, L., Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. *Annu. Rev. Biochem.* **2005**, *74*, 317-53.
- [75] Bebenek, K.; Kunkel, T. A., Functions of DNA polymerases. *Adv. Protein Chem.* **2004**, *69*, 137-65.

- [76] Yang, W., Damage repair DNA polymerases Y. *Curr. Opin. Struct. Biol.* **2003**, 13, (1), 23-30.
- [77] Yang, W.; Woodgate, R., What a difference a decade makes: insights into translesion DNA synthesis. *Proc. Natl. Acad. Sci.* **2007**, 104, (40), 15591–15598.
- [78] Shaheen, M.; Shanmugam, I.; Hromas, R., The Role of PCNA Posttranslational Modifications in Translesion Synthesis. *J. of Nucleic Acids* **2010**, 2010, Article ID 761217, 8 pages, doi:10.4061/2010/761217.
- [79] Kannouche, P. L.; Wing, J.; Lehmann, A. R., Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage. *Mol. Cell* **2004**, 14, (4), 491-500.
- [80] Bi, X.; Barkley, L. R.; Slater, D. M.; Tateishi, S.; Yamaizumi, M.; Ohmori, H.; Vaziri, C., Rad18 regulates DNA polymerase kappa and is required for recovery from S-phase checkpoint-mediated arrest. *Mol. Cell Biol.* **2006**, 26, (9), 3527-40.
- [81] Plosky, B. S.; Vidal, A. E.; Fernandez de Henestrosa, A. R.; McLenigan, M. P.; McDonald, J. P.; Mead, S.; Woodgate, R., Controlling the subcellular localization of DNA polymerases iota and eta via interactions with ubiquitin. *Embo J.* **2006**, 25, (12), 2847-55.
- [82] Hoege, C.; Pfander, B.; Moldovan, G. L.; Pyrowolakis, G.; Jentsch, S., RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* **2002**, 419, (6903), 135-41.
- [83] Stelter, P.; Ulrich, H. D., Control of spontaneous and damage-induced mutagenesis by SUMO and ubiquitin conjugation. *Nature* **2003**, 425, (6954), 188-91.
- [84] Lehmann, A. R.; Niimi, A.; Ogi, T.; Brown, S.; Sabbioneda, S.; Wing, J. F.; Kannouche, P. L.; Green, C. M., Translesion synthesis: Y-family polymerases and the polymerase switch. *DNA Repair* **2007**, 6, (7), 891-9.
- [85] Livneh, Z.; Ziv, O.; Shachar, S., Multiple two-polymerase mechanisms in mammalian translesion DNA synthesis. *Cell Cycle* 9, (4), 729-35.
- [86] Johnson, R. E.; Washington, M. T.; Haracska, L.; Prakash, S.; Prakash, L., Eukaryotic polymerases iota and zeta act sequentially to bypass DNA lesions. *Nature* **2000**, 406, (6799), 1015-9.
- [87] Ziv, O.; Geacintov, N.; Nakajima, S.; Yasui, A.; Livneh, Z., DNA polymerase zeta cooperates with polymerases kappa and iota in translesion DNA synthesis across pyrimidine photodimers in cells from XPV patients. *Proc. Natl. Acad. Sci. USA* **2009**, 106, (28), 11552-7.
- [88] Silverstein, T. D.; Johnson, R. E.; Jain, R.; Prakash, L.; Prakash, S.; Aggarwal, A. K., Structural basis for the suppression of skin cancers by DNA polymerase eta. *Nature* **2010**, 465, (7301), 1039-43.
- [89] Biertümpfel, C.; Zhao, Y.; Kondo, Y.; Ramon-Maiques, S.; Gregory, M.; Lee, J. Y.; Masutani, C.; Lehmann, A. R.; Hanaoka, F.; Yang, W., Structure and mechanism of human DNA polymerase eta. *Nature* **2010**, 465, (7301), 1044-8.
- [90] Alt, A.; Lammens, K.; Chicchini, C.; Lammens, A.; Pieck, J. C.; Kuch, D.; Hopfner, K. P.; Carell, T., Bypass of DNA lesions generated during anticancer treatment with cisplatin by DNA polymerase eta. *Science* **2007**, 318, (5852), 967–970.
- [91] Bresson, A.; Fuchs, R. P., Lesion bypass in yeast cells: Pol eta participates in a multi-DNA polymerase process. *Embo J.* **2002**, 21, (14), 3881-7.

- [92] Cordonnier, A. M.; Lehmann, A. R.; Fuchs, R. P., Impaired translesion synthesis in xeroderma pigmentosum variant extracts. *Mol Cell Biol* **1999**, 19, (3), 2206-11.
- [93] Uljon, S. N.; Johnson, R. E.; Edwards, T. A.; Prakash, S.; Prakash, L.; Aggarwal, A. K., Crystal structure of the catalytic core of human DNA polymerase kappa. *Structure* **2004**, 12, (8), 1395-404.
- [94] Wolfle, W. T.; Washington, M. T.; Prakash, L.; Prakash, S., Human DNA polymerase kappa uses template-primer misalignment as a novel means for extending mispaired termini and for generating single-base deletions. *Genes Dev.* **2003**, 17, (17), 2191-9.
- [95] Ohashi, E.; Ogi, T.; Kusumoto, R.; Iwai, S.; Masutani, C.; Hanaoka, F.; Ohmori, H., Error-prone bypass of certain DNA lesions by the human DNA polymerase kappa. *Genes Dev.* **2000**, 14, (13), 1589-94.
- [96] Zhang, Y.; Yuan, F.; Wu, X.; Wang, M.; Rechkoblit, O.; Taylor, J. S.; Geacintov, N. E.; Wang, Z., Error-free and error-prone lesion bypass by human DNA polymerase kappa in vitro. *Nucleic Acids Res.* **2000**, 28, (21), 4138-46.
- [97] Shibutani, S.; Suzuki, N.; Grollman, A. P., Mutagenic specificity of (acetyl amino)fluorene-derived DNA adducts in mammalian cells. *Biochemistry* **1998**, 37, (35), 12034-41.
- [98] Tissier, A.; McDonald, J. P.; Frank, E. G.; Woodgate, R., poliota, a remarkably error-prone human DNA polymerase. *Genes Dev.* **2000**, 14, (13), 1642-50.
- [99] Zhang, Y.; Yuan, F.; Wu, X.; Taylor, J. S.; Wang, Z., Response of human DNA polymerase iota to DNA lesions. *Nucleic Acids Res.* **2001**, 29, (4), 928-35.
- [100] Washington, M. T.; Johnson, R. E.; Prakash, L.; Prakash, S., Human DNA polymerase iota utilizes different nucleotide incorporation mechanisms dependent upon the template base. *Mol. Cell. Biol.* **2004**, 24, (2), 936-43.
- [101] Nair, D. T.; Johnson, R. E.; Prakash, S.; Prakash, L.; Aggarwal, A. K., Replication by human DNA polymerase-iota occurs by Hoogsteen base-pairing. *Nature* **2004**, 430, (6997), 377-80.
- [102] Nair, D. T.; Johnson, R. E.; Prakash, L.; Prakash, S.; Aggarwal, A. K., Hoogsteen base pair formation promotes synthesis opposite the 1,N6-ethenodeoxyadenosine lesion by human DNA polymerase iota. *Nat. Struct. Mol. Biol.* **2006**, 13, (7), 619-25.
- [103] Nelson, J. R.; Lawrence, C. W.; Hinkle, D. C., Thymine-thymine dimer bypass by yeast DNA polymerase zeta. *Science* **1996**, 272, (5268), 1646-9.
- [104] Guo, D.; Wu, X.; Rajpal, D. K.; Taylor, J. S.; Wang, Z., Translesion synthesis by yeast DNA polymerase zeta from templates containing lesions of ultraviolet radiation and acetylaminofluorene. *Nucleic Acids Res.* **2001**, 29, (13), 2875-83.
- [105] Guo, D.; Xie, Z.; Shen, H.; Zhao, B.; Wang, Z., Translesion synthesis of acetylaminofluorene-dG adducts by DNA polymerase zeta is stimulated by yeast Rev1 protein. *Nucleic Acids Res.* **2004**, 32, (3), 1122-30.
- [106] Baynton, K.; Bresson-Roy, A.; Fuchs, R. P., Analysis of damage tolerance pathways in *Saccharomyces cerevisiae*: a requirement for Rev3 DNA polymerase in translesion synthesis. *Mol. Cell. Biol.* **1998**, 18, (2), 960-6.
- [107] Baynton, K.; Bresson-Roy, A.; Fuchs, R. P., Distinct roles for Rev1p and Rev7p during translesion synthesis in *Saccharomyces cerevisiae*. *Mol. Microbiol.* **1999**, 34, (1), 124-33.

- [108] Lawrence, C. W., Cellular functions of DNA polymerase zeta and Rev1 protein. *Adv Protein Chem.* **2004**, 69, 167-203.
- [109] Waters, L. S.; Minesinger, B. K.; Wilttrout, M. E.; D'Souza, S.; Woodruff, R. V.; Walker, G. C., Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiol. Mol. Biol. Rev.* **2009**, 73, (1), 134-54.
- [110] Nelson, J. R.; Lawrence, C. W.; Hinkle, D. C., Deoxycytidyl transferase activity of yeast REV1 protein. *Nature* **1996**, 382, (6593), 729-31.
- [111] Washington, M. T.; Minko, I. G.; Johnson, R. E.; Haracska, L.; Harris, T. M.; Lloyd, R. S.; Prakash, S.; Prakash, L., Efficient and error-free replication past a minor-groove N²-guanine adduct by the sequential action of yeast Rev1 and DNA polymerase zeta. *Mol. Cell Biol.* **2004**, 24, (16), 6900-6.
- [112] Lin, W.; Xin, H.; Zhang, Y.; Wu, X.; Yuan, F.; Wang, Z., The human REV1 gene codes for a DNA template-dependent dCMP transferase. *Nucleic Acids Res.* **1999**, 27, (22), 4468-75.
- [113] Levin, W.; Wood, A. W.; Chang, R. L.; Ittah, Y.; Croisy-Delcey, M.; Yagi, H.; Jerina, D. M.; Conney, A. H., Exceptionally high tumor-initiating activity of benzo(c)phenanthrene bay-region diol-epoxides on mouse skin. *Cancer Res* **1980**, 40, (11), 3910-4.
- [114] Bauer, J.; Xing, G.; Yagi, H.; Sayer, J. M.; Jerina, D. M.; Ling, H., A structural gap in Dpo4 supports mutagenic bypass of a major benzo[a]pyrene dG adduct in DNA through template misalignment. *Proc. Natl. Acad. Sci. USA* **2007**, 104, (38), 14905-10.
- [115] Ling, H.; Sayer, J. M.; Plosky, B. S.; Yagi, H.; Boudsocq, F.; Woodgate, R.; Jerina, D. M.; Yang, W., Crystal structure of a benzo[a]pyrene diol epoxide adduct in a ternary complex with a DNA polymerase. *Proc. Natl. Acad. Sci. USA* **2004**, 101, (8), 2265-9.
- [116] Batra, V. K.; Shock, D. D.; Prasad, R.; Beard, W. A.; Hou, E. W.; Pedersen, L. C.; Sayer, J. M.; Yagi, H.; Kumar, S.; Jerina, D. M.; Wilson, S. H., Structure of DNA polymerase beta with a benzo[c]phenanthrene diol epoxide-adducted template exhibits mutagenic features. *Proc. Natl. Acad. Sci. USA* **2006**, 103, (46), 17231-6.
- [117] Schneider, S.; Schorr, S.; Carell, T., Crystal structure analysis of DNA lesion repair and tolerance mechanisms. *Curr. Opin. Struct. Biol.* **2009**, 19, (1), 87-95.
- [118] Hsu, G. W.; Huang, X.; Luneva, N. P.; Geacintov, N. E.; Beese, L. S., Structure of a high fidelity DNA polymerase bound to a benzo[a]pyrene adduct that blocks replication. *J Biol Chem* **2005**, 280, (5), 3764-70.
- [119] Zang, H.; Chowdhury, G.; Angel, K. C.; Harris, T. M.; Guengerich, F. P., Translesion synthesis across polycyclic aromatic hydrocarbon diol epoxide adducts of deoxyadenosine by *Sulfolobus solfataricus* DNA polymerase Dpo4. *Chem. Res. Toxicol.* **2006**, 19, (6), 859-67.
- [120] Dutta, S.; Li, Y.; Johnson, D.; Dzantiev, L.; Richardson, C. C.; Romano, L. J.; Ellenberger, T., Crystal structures of 2-acetylaminofluorene and 2-aminofluorene in complex with T7 DNA polymerase reveal mechanisms of mutagenesis. *Proc. Natl. Acad. Sci. USA* **2004**, 101, (46), 16186-91.
- [121] Hsu, G. W.; Kiefer, J. R.; Burnouf, D.; Becherel, O. J.; Fuchs, R. P.; Beese, L. S., Observing translesion synthesis of an aromatic amine DNA adduct by a high-fidelity DNA polymerase. *J. Biol. Chem.* **2004**, 279, (48), 50280-5.

- [122] Rechkoblit, O.; Kolbanovskiy, A.; Malinina, L.; Geacintov, N. E.; Broyde, S.; Patel, D. J., Mechanism of error-free and semitargeted mutagenic bypass of an aromatic amine lesion by Y-family polymerase Dpo4. *Nat. Struct. Mol. Biol.* **2010**, 17, (3), 379-88.
- [123] Wang, L.; Broyde, S., A new anti conformation for *N*-(deoxyguanosin-8-yl)-2-acetylaminofluorene (AAF-dG) allows Watson-Crick pairing in the *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4). *Nucleic Acids Res.* **2006**, 34, (3), 785-95.
- [124] Fuchs, R. P.; Seeberg, E., pBR322 plasmid DNA modified with 2-acetylaminofluorene derivatives: transforming activity and in vitro strand cleavage by the *Escherichia coli* uvrABC endonuclease. *Embo J.* **1984**, 3, (4), 757-60.
- [125] Maher, V. M.; Birch, N.; Otto, J. R.; MacCormick, J. J., Cytotoxicity of carcinogenic aromatic amides in normal and xeroderma pigmentosum fibroblasts with different DNA repair capabilities. *J. Natl. Cancer. Inst.* **1975**, 54, (6), 1287-94.
- [126] Tang, M.; Lieberman, M. W.; King, C. M., uvr Genes function differently in repair of acetylaminofluorene and aminofluorene DNA adducts. *Nature* **1982**, 299, (5884), 646-8.
- [127] de Laat, W. L.; Jaspers, N. G.; Hoeijmakers, J. H., Molecular mechanism of nucleotide excision repair. *Genes. Dev.* **1999**, 13, (7), 768-85.
- [128] Maillard, O.; Camenisch, U.; Blagoev, K. B.; Naegeli, H., Versatile protection from mutagenic DNA lesions conferred by bipartite recognition in nucleotide excision repair. *Mutat. Res.* **2008**, 658, (3), 271-86.
- [129] Banerjee, A.; Santos, W. L.; Verdine, G. L., Structure of a DNA glycosylase searching for lesions. *Science* **2006**, 311, (5764), 1153-7.
- [130] Stivers, J. T., Extrahelical damaged base recognition by DNA glycosylase enzymes. *Chemistry* **2008**, 14, (3), 786-93.
- [131] Sugawara, K., XPC: its product and biological roles. *Adv. Exp. Med. Biol.* **2008**, 637, 47-56.
- [132] Min, J. H.; Pavletich, N. P., Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* **2007**, 449, (7162), 570-5.
- [133] Missura, M.; Buterin, T.; Hindges, R.; Hubscher, U.; Kasparkova, J.; Brabec, V.; Naegeli, H., Double-check probing of DNA bending and unwinding by XPA-RPA: an architectural function in DNA repair. *Embo J.* **2001**, 20, (13), 3554-64.
- [134] Kessler, K. J.; Kaufmann, W. K.; Reardon, J. T.; Elston, T. C.; Sancar, A., A mathematical model for human nucleotide excision repair: damage recognition by random order assembly and kinetic proofreading. *J. Theor. Biol.* **2007**, 249, (2), 361-75.
- [135] Petit, C.; Sancar, A., Nucleotide excision repair: from *E. coli* to man. *Biochimie* **1999**, 81, (1-2), 15-25.
- [136] Wood, R. D., DNA damage recognition during nucleotide excision repair in mammalian cells. *Biochimie* **1999**, 81, (1-2), 39-44.
- [137] Thoma, B. S.; Vasquez, K. M., Critical DNA damage recognition functions of XPC-hHR23B and XPA-RPA in nucleotide excision repair. *Mol. Carcinog.* **2003**, 38, (1), 1-13.
- [138] Hanawalt, P. C., Transcription-coupled repair and human disease. *Science* **1994**, 266, (5193), 1957-8.

- [139] Strasser, R., Ph.D. thesis in the group of Prof. T. Carell, LMU Munich, Germany, **2010**.
- [140] Camenisch, U.; Nageli, H., XPA gene, its product and biological roles. *Adv. Exp. Med. Biol.* **2008**, 637, 28-38.
- [141] Ikegami, T.; Kuraoka, I.; Saijo, M.; Kodo, N.; Kyogoku, Y.; Morikawa, K.; Tanaka, K.; Shirakawa, M., Solution structure of the DNA- and RPA-binding domain of the human repair factor XPA. *Nat. Struct. Biol.* **1998**, 5, (8), 701-6.
- [142] Buchko, G. W.; Ni, S.; Thrall, B. D.; Kennedy, M. A., Structural features of the minimal DNA binding domain (M98-F219) of human nucleotide excision repair protein XPA. *Nucleic Acids Res.* **1998**, 26, (11), 2779-88.
- [143] Buchko, G. W.; Daughdrill, G. W.; de Lorimier, R.; Rao, B. K.; Isern, N. G.; Lingbeck, J. M.; Taylor, J. S.; Wold, M. S.; Gochin, M.; Spicer, L. D.; Lowry, D. F.; Kennedy, M. A., Interactions of human nucleotide excision repair protein XPA with DNA and RPA70 Delta C327: chemical shift mapping and ¹⁵N NMR relaxation studies. *Biochemistry* **1999**, 38, (46), 15116-28.
- [144] Johnson, D. L.; Reid, T. M.; Lee, M. S.; King, C. M.; Romano, L. J., Preparation and characterization of a viral DNA molecule containing a site-specific 2-aminofluorene adduct: a new probe for mutagenesis by carcinogens. *Biochemistry* **1986**, 25, (2), 449-56.
- [145] Koehl, P.; Burnouf, D.; Fuchs, R. P., Construction of plasmids containing a unique acetylaminofluorene adduct located within a mutation hot spot. A new probe for frameshift mutagenesis. *J. Mol. Biol.* **1989**, 207, (2), 355-64.
- [146] Wang, Z.; Rizzo, C. J., Synthesis of the C8-deoxyguanosine adduct of the food mutagen IQ. *Org. Lett.* **2001**, 3, (4), 565-8.
- [147] Gillet, L. C.; Schärer, O. D., Preparation of C8-amine and acetylamine adducts of 2'-deoxyguanosine suitably protected for DNA synthesis. *Org. Lett.* **2002**, 4, (24), 4205-8.
- [148] Gillet, L. C.; Alzeer, J.; Schärer, O. D., Site-specific incorporation of *N*-(deoxyguanosin-8-yl)-2-acetylaminofluorene (dG-AAF) into oligonucleotides using modified 'ultra-mild' DNA synthesis. *Nucleic Acids Res.* **2005**, 33, (6), 1961-9.
- [149] Matteucci, M. D.; Caruthers, M. H., Synthesis of deoxyoligonucleotides on a polymer support. 1981. *Biotechnology* **1992**, 24, 92-8.
- [150] Caruthers, M. H.; Beaucage, S. L.; Becker, C.; Efcavitch, J. W.; Fisher, E. F.; Galluppi, G.; Goldman, R.; deHaseth, P.; Matteucci, M.; McBride, L.; et al., Deoxyoligonucleotide synthesis via the phosphoramidite method. *Gene Amplif. Anal.* **1983**, 3, 1-26.
- [151] Shibutani, S.; Gentles, R.; Johnson, F.; Grollman, A. P., Isolation and characterization of oligodeoxynucleotides containing dG-N2-AAF and oxidation products of dG-C8-AAF. *Carcinogenesis* **1991**, 12, (5), 813-8.
- [152] Johnson, F.; Huang, C. Y.; Yu, P. L., Synthetic and oxidative studies on 8-(arylamino)-2'-deoxyguanosine and -guanosine derivatives. *Environ. Health Perspect* **1994**, 102 Suppl 6, 143-9.
- [153] Philipps, G. R., Purification and characterization of phosphodiesterase from *Crotalus* venom. *Physiol. Chem.* **1975**, 356, (7), 1085-96.

- [154] Kadlubar, F. F.; Unruh, L. E.; Beland, F. A.; Straub, K. M.; Evans, F. E., *In vitro* reaction of the carcinogen *N*-hydroxy-2-naphthylamine with DNA at the C8 and N^2 atoms of guanine and at the N^6 atom of adenine. *Carcinogenesis* **1980**, 1, 139-150.
- [155] Kriek, E.; Westra, J. G., Structural identification of the pyrimidine derivatives formed from *N*-(deoxyguanosin-8-yl)-2-aminofluorene in aqueous solution at alkaline pH. *Carcinogenesis* **1980**, 1, (6), 459-68.
- [156] Trincão, J.; Johnson, R. E.; Escalante, C. R.; Prakash, S.; Prakash, L.; Aggarwal, A. K., Structure of the catalytic core of *S. cerevisiae* DNA polymerase ϵ : implications for translesion DNA synthesis. *Mol. Cell* **2001**, 8, (2), 417-426.
- [157] Reissner, T.; Schneider, S.; Schorr, S.; Carell, T., Crystal Structure of a Cisplatin-(1,3-GTG) Cross-Link within DNA Polymerase ϵ . *Angew. Chem. Int. Ed.* **2010**, 49, 3077-3080.
- [158] Norman, D.; Abuaf, P.; Hingerty, B. E.; Live, D.; Grunberger, D.; Broyde, S.; Patel, D. J., NMR and computational characterization of the *N*-(deoxyguanosin-8-yl)aminofluorene adduct [(AF)G] opposite adenosine in DNA: (AF)G[syn].A[anti] pair formation and its pH dependence. *Biochemistry* **1989**, 28, (18), 7462-76.
- [159] Eckel, L. M.; Krugh, T. R., 2-Aminofluorene modified DNA duplex exists in two interchangeable conformations. *Nat. Struct. Biol.* **1994**, 1, (2), 89-94.
- [160] Choi, J. Y.; Guengerich, F. P., Adduct size limits efficient and error-free bypass across bulky N2-guanine DNA lesions by human DNA polymerase ϵ . *J. Mol. Biol.* **2005**, 352, (1), 72-90.
- [161] Meyer, E. A.; Castellano, R. K.; Diederich, F., Interactions with aromatic rings in chemical and biological recognition. *Angew. Chem. Int. Ed.* **2003**, 42, (11), 1210-50.
- [162] Lone, S.; Townson, S. A.; Uljon, S. N.; Johnson, R. E.; Brahma, A.; Nair, D. T.; Prakash, S.; Prakash, L.; Aggarwal, A. K., Human DNA polymerase κ encircles DNA: implications for mismatch extension and lesion bypass. *Mol. Cell* **2007**, 25, (4), 601-14.
- [163] Vasquez-Del Carpio, R.; Silverstein, T. D.; Lone, S.; Swan, M. K.; Choudhury, J. R.; Johnson, R. E.; Prakash, S.; Prakash, L.; Aggarwal, A. K., Structure of human DNA polymerase κ inserting dATP opposite an 8-OxoG DNA lesion. *PLoS One* **2009**, 4, (6), e5766.
- [164] LeMaster, D. M.; Richards, F. M., ^1H - ^{15}N heteronuclear NMR studies of *Escherichia coli* thioredoxin in samples isotopically labeled by residue type. *Biochemistry* **1985**, 24, (25), 7263-8.
- [165] Kabsch, W., Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **1993**, 26, (6), 795-800.
- [166] McCoy, A. J.; Grosse-Kunstleve, R. W.; Storoni, L. C.; Read, R. J., Likelihood-enhanced fast translation functions. *Acta Crystallogr.* **2005**, 61, (Pt 4), 458-464.
- [167] Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and Development of Coot. *Acta Crystallogr.* **2010**, 66, 486-501.
- [168] Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C., PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr.* **2002**, 58, (Pt 11), 1948-1954.

- [169] Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* **1997**, 53, (Pt 3), 240-55.
- [170] Sarges, R.; Hank, R. F.; Blake, J. F.; Bordner, J.; Bussolotti, D. L.; Hargrove, D. M.; Treadway, J. L.; Gibbs, E. M., Glucose transport-enhancing and hypoglycemic activity of 2-methyl-2-phenoxy-3-phenylpropanoic acids. *J. Med. Chem.* **1996**, 39, (24), 4783-803.
- [171] Otto, S.; Furlan, R. L. E.; Sanders, J. K. M., Dynamic combinatorial libraries of macrocyclic disulfides in water. *J. Amer. Chem. Soc.* **2000**, 122, 12063-12064.

CURRICULUM VITAE

STEPHANIE SCHORR

PERSONAL DATA

Date of birth: 16th April, 1982
Place of birth: Frankfurt/Main, Germany
Nationality: German

STUDIES

- 12/2006 – 12/2010 Doctoral thesis in the group of *Prof. Dr. T. Carell, Ludwig-Maximilians-Universität München* on replication of bulky DNA adducts
- 09/2006 Master of Science in Biochemistry, *Ruhr-Universität Bochum*
- 03/2006 – 09/2006 Master thesis with the topic “Stabilisation of DNA nanostructures from trisoligonucleotides by covalent disulfide-crosslinking” in the group of *Prof. Dr. G. von Kiedrowski, Ruhr-Universität Bochum*
- 2005 project work in the group of *Prof. Dr. M. Havenith-Newen, Ruhr-Universität Bochum* on THz spectroscopy of water and aqueous carbohydrate solutions (3 month)
- 2004/2005 Erasmus-exchange program (6 month) at the *Sussex University of Brighton, UK*; project work with *Dr. M. Paget* (Characterization of the two Rex homologues ERex 1 and ERex 2 in *Enterococcus faecalis*).
- 10/2004 Beginning of the Master studies at the *Ruhr-Universität Bochum* (Biochemistry)
- 09/2004 Bachelor of Science in Biochemistry
Bachelor thesis: „Generation of RNA fragments with homogeneous ends via cis-cleaving ribozymes“ (*Prof. Dr. S. Müller, Ruhr-Universität Bochum*)
- 10/2001 – 09/2004 Bachelor studies (Biochemistry) at the *Ruhr-Universität Bochum*
- 07/2001 Abitur, *Gymnasium am Ostring, Bochum*

GRANTS

Kekulé Ph.D. fellowship of the *Fonds der Chemischen Industrie*

Scholarship: „*Studienanfängsstipendium*“ of the *Fonds der Chemischen Industrie*

LANGUAGE SKILLS

German (mother tongue), English (fluent), French (basic)