Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Structural Characterisation of Transcription and Replication through Cisplatin Lesioned DNA

Aaron Alt

aus

Wiener Neustadt

2008

Erklärung

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

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbstständig, ohne unerlaubte Hilfe erarbeitet.

München, am 06.02.2008

(Aaron Alt)

Dissertation eingereicht am 7.2.2008

1. Gutachter Prof. Dr. Thomas Carell

2. Gutachter Prof. Dr. Karl-Peter Hopfner

Mündliche Prüfung am 13.3.2008

For my aunt Kathy Friedman,
who dreamt of becoming a Chemist,
but was deported to Auschwitz, where
her dreams were shattered.

Parts of this work were published or presented on conferences:

Publications:

Alt, Aaron; Lammens, Katja; Chiocchini, Claudia; Lammens, Alfred; Pieck, J. Carsten; Kuch, David; Hopfner, Karl-Peter; Carell, Thomas. Bypass of DNA Lesions Generated During Anticancer Treatment with Cisplatin by DNA Polymerase η. *Science*, **2007**, 318 (5852), 967-970.

Damsma, Gerke; Alt, Aaron; Brueckner, Florian; Carell, Thomas; Cramer Patrick. Mechanism of transcriptional stalling at cisplatin-damaged DNA. *Nature Structural & Molecular Biology*, **2007**, 14 (12), 1127-1133.

Lectures:

Alt Aaron. Processing of cisplatin- and oxaliplatin-DNA adducts at atomic resolution, CLUSTOXDNA, Annecy, France, March 2006

Alt Aaron. Processing of cisplatin- and oxaliplatin-DNA adducts at atomic resolution, CLUSTOXDNA, Milton Hill, Oxfordshire, United Kingdom, May 2006

Alt Aaron. Processing of cisplatin DNA adducts at atomic resolution, CLUSTOXDNA, Gandia, Spain, October 2006

Alt Aaron. Cancer drug resistance and translesion synthesis: The Yin and Yang of eukaryotic DNA polymerase η, CLUSTOXDNA, Plovdiv, Bulgaria, June 2007

Posters:

Aaron Alt, Carsten Pieck, Alfred Lammens, Karl-Peter Hopfner and Thomas Carell. Translesion Synthesis past Pt-DNA adducts by DNA polymerase η. 7th Winter Research Conference on Free Radicals in Biology, Les Houches, France, March 19-22, 2006.

Aaron Alt, Katja Lammens, J. Carsten Pieck, Alfred Lammens, Claudia Chiocchini, Karl-Peter Hopfner, Thomas Carell. Replication of a cisplatin DNA adduct by DNA polymerase η. The 72nd Annual Meeting of the Israel Chemical Society, Tel-Aviv, Israel, February 6-7, 2007.

Aaron Alt, Katja Lammens, J. Carsten Pieck, Alfred Lammens, Claudia Chiocchini, Karl-Peter Hopfner & Thomas Carell. Replication of a cisplatin DNA adduct by DNA polymerase η. The 2nd International Gene Center / SFB 646 Symposium, Munich, Germany, October 12-13, 2007.

Table of contents

1 Summary 4
2 Zusammenfassung 6
3 Introduction 8
3.1 DNA and DNA damages 9
3.1.1 Cisplatin lesions 11
3.1.2 CPD lesions 15
3.2 DNA Polymerases 17
3.2.1 Replicative polymerases 19
3.2.2 Low fidelity polymerases 20
3.3 Eukaryotic RNA polymerase II 30
3.4 Structure determination by X-ray crystallography 30
3.4.1 Crystallization 31
3.4.2 Theory of X-ray diffraction 34
3.4.3 The Phase problem 36
3.4.4 Solving the phase problem 37
3.5 Research objectives 39
4 Experimental part 40
4.1 Materials 40
4.1.1 Chemicals 40
4.1.2 Enzymes, Bacterial strains, Standards and Kits 40
4.1.3 Consumables 41
4.1.4 Chromatographic material 41
4.1.5 Laboratory instruments 42
4.1.6 Oligonucleotides 43
4.1.7 Buffers, Mediums, Solutions and Antibiotics 45
4.2 Methods 48
4.2.1 Platinum lesions 48
4.2.2 DNA synthesis, cleavage, purification and hybridization 49
4.2.3 Molecular Biology Methods 50
4.2.4 Microbiological methods 52
4.2.5 Protein biochemical methods 53

4.2.6 Primer extension assays 54
4.2.7 Crystallization 55
4.2.8 Data collection and processing 56
4.2.9 Structure solution and refinement 56
4.3 Bioinformatic Methods 57
4.3.1 Homology searches and alignments 57
4.3.2 Structure visualization and analyzing 57
4.3.3 Crystallographic software 57
5 Results 58
5.1 Preparation and purification of platin lesion containing DNA strands 58
5.2 DNA synthesis and purification 59
5.2.1 Preparation and purification of 2',3'-didesoxy primer strands 59
5.2.2 Preparation and purification of TT dimer lesion containing DNA strands 59
5.3 Primer extensions of 1, 3 GTG oxaliplatin adducts 60
5.4 Crystallization of DNA Pol η in ternary structure 62
5.4.1 Preliminary screenings 62
5.4.2 Crystallizations with lesion containing DNA and ddCTP 63
5.4.3 Crystallizations with lesion containing DNA and dNTPs 64
5.5 Structure solution and refinement 65
5.6 Crystal contacts 66
5.7 Crystal structure of the Pol η ternary complex with ddNTP 68
5.8 Crystal structure of the Pol η ternary complex with ddprimer 69
5.8.1 The catalytic center 70
5.8.2 The 3'dG pre-elongation step 71
5.8.3 The 3'dG elongation step 72
5.8.4 The 5'dG elongation step 73
5.8.5 The polymerase-associated domain 75
5.9 Biochemical studies 78
5.9.1 Nucleotide insertion studies 78
5.9.2 Structure based point mutation of Arg73 79
5.9.3 Functional hydrogen bonding studies with Zebularine 81
5.10 Cloning, expression and purification of DNA Pol η-513 82
5.11 RNA Pol II stalling at a cisplatin lesion 83
6 Discussion 87

6.1 Lesion bypass 87
6.1.1 Lesion bypass of a 1,3-d(GpTpG) oxaliplatin lesion 87
6.1.2 Lesion bypass of a 1,2-d(GpG) cisplatin lesion 87
6.2 Compare to the apoenzyme 91
6.3 Polymerase switch 93
6.4 Use of ddNTP leads to potential misinterpretations of the structures 94
6.5 DPO4 vs Pol η lesion bypass 96
6.6 Mechanism of transcriptional stalling at cisplatin lesioned DNA 98
6.6.1 RNA polymerase II stalling and AMP misincorporation 98
6.6.2 Impaired entry of lesions into the active site 99
6.6.3 Possible mechanisms for misincorporation 101
6.6.4 Nontemplated AMP incorporation and an 'A-rule' for RNAP II 102
6.6.5 Comparison to TT dimer damage recognition 103
6.7 Future perspectives and outlook 105
7 Appendix 106
7.1 Abbreviations 106
7.2 Crystallographic tables 108
8 References 110
9 Acknowledgements 121
10 Curriculum Vitae 123

1 Summary

Replication of the genome is strongly inhibited when high fidelity DNA polymerases encounter unrepaired DNA lesions, which can not be processed. The highly stringent active sites of these polymerases are unable to accommodate damaged bases and therefore DNA lesions block the replication fork progression. In order to overcome this problem, cells have evolved mechanisms for either repairing the damage, or synthesising past it with specially adapted polymerasases.

Eukaryotic DNA polymerase η (Pol η), belonging to the Y-family of DNA polymerases, is outstanding in its ability to replicate through a variety of highly distorting DNA lesions such as cyclobutane pyrimidine dimers (CPDs), which are the main UV-induced lesions. Also 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[1]. The bypass of such intrastrand crosslinks by high fidelity DNA polymerases is particularly difficult because two adjacent coding bases are simultaneously damaged. Thus, replication by Pol η allows organisms to survive exposure to sunlight or, in the case of cisplatin, gives rise to resistances against cisplatin treatment. Mutations in the human POLH gene, encoding Pol η, causes the variant form of xeroderma pigmentosum (XP-V), characterized by the failure to copy through CPDs. This leads to strongly increased UV sensitivity and skin cancer predisposition.

This thesis describes mechanistic investigations of the translesion synthesis (TLS) process by *S. cerevisiae* DNA Pol η at atomic resolution, which were undertaken in collaboration with the *Hopfner* group. To study this process, cisplatin lesioned DNA had to be prepared first. Once this technique was established, the catalytic fragment of Pol η was crystallized as ternary complex with incoming 2',3'-dideoxycytidine 5'-triphosphate (ddCTP) and an primer - template DNA containing a site specific Pt-GG adduct (Figure 1-1 A).

The first obtained structure shows the ddCTP positioned in a loosely bound conformation in the active site, hydrogen bonded to the templating base. Realizing the importance of the 3’-hydroxy group for positioning the NTP and the DNA correctly inside the polymerase, the complex was crystallized again with a 2’-deoxynucleoside 5’-triphosphate (dNTP). To prevent nucleotidyl transfer, primer strands which terminate at the 3’-end with a 2’,3’-dideoxy ribose were prepared by reverse DNA

synthesis and used for cocrystallization. The resulting crystals diffracted typically to 3.1-3.3Å resolution at a synchrotron light source (Figure 1-1 B and C).

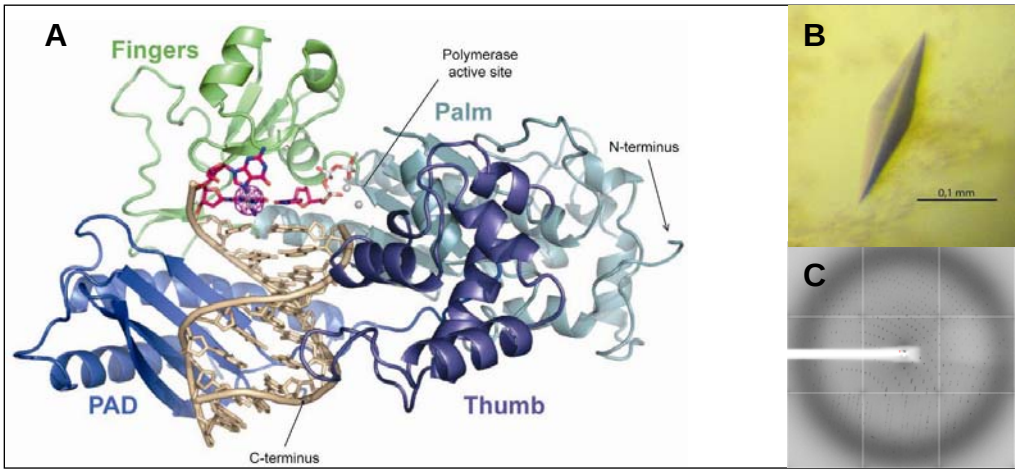


Figure 1-1. A. A cartoon depiction of the Pol η in ternary complex with cisplatin damaged DNA and an incoming desoxy nucleoside triphosphate. **B.** Crystal of Pol η. **C.** Diffraction pattern of Pol η.

A Pol η specific arginine (Arg73 in yeast Pol η) was identified for its importance to position the dNTP correctly in the active site and was shown to be necessary for lesion bypass. In contrast to the fixed preorientation of the dNTP in the active site, the damaged DNA is bound flexibly in a rather open DNA binding cleft. Nucleotidyl transfer requires a revolving of the DNA, energetically driven by hydrogen bonding of the templating base to the dNTP. For the 3'dG of the Pt-GG, this step is accomplished by bona fide Watson-Crick base pairs to dCTP and is biochemically efficient and accurate. In contrast, bypass of the 5'dG of the Pt-GG is less efficient and promiscuous for dCTP and dATP. Structurally, this can be attributed to misalignment of the templating 5'dG due to the rigid Pt crosslink.

In cooperation with the *Cramer* group the structural reasons for the blockage of RNA Polymerase II (RNAP II) by the cisplatin lesion were elucidated. Using structural as well as biochemical methods it could be shown that stalling results from a translocation barrier that prevents delivery of the lesion to the active site. AMP misincorporation occurs at the barrier and also at an abasic site, suggesting that it arises from nontemplated synthesis according to an 'A-rule' known for DNA polymerases. RNAP II can bypass a cisplatin lesion that is artificially placed beyond the translocation barrier, even in the presence of a G·A mismatch. Thus, the barrier prevents transcriptional mutagenesis.

2 Zusammenfassung

Die Verdopplung des Genoms wird stark gehemmt wenn replikative DNA Polymerasen auf unreparierte DNA Schäden treffen. Geschädigte Basen können nicht in die stringenten aktiven Zentren von „high fidelity“ Polymerasen gedreht werden und blockieren die Progression der Replikationsgabel. Um diesem Ereignis zuvorzukommen, haben Zellen Mechanismen zur Reperatur und zur fehlerfreien Überwindung von DNA Schäden entwickelt. Alle drei Reiche des Lebens verfügen über spezielle DNA Polymerasen der Y-Familie, die über Schäden hinweg replizieren können und so einen lebenswichtigen Schadenstoleranzmechanismus darstellen.
Die eukaryotische DNA Polymerase η (Pol η) ist in ihrer Fähigkeit über eine Vielzahl von stark DNA Helix verzerrenden Schäden hinweg zu lesen einzigartig. Sie kann zum Beispiel Cyclobutan Pyrimidindimere (CPDs), der häufigste UV verursachte Schaden, oder die durch das Chemotherapie-Wirkstoff Cisplatin verursachten 1,2-d(GpG) Addukte (Pt-GG) hinweg replizieren[1]. Die Replikation solcher Schäden ist besonders schwierig, weil zwei aufeinanderfolgende Basen gleichzeitig geschädigt sind.

Somit ermöglicht es die Aktivität von Pol η Organismen die Exposition von Sonnenlicht zu tolerieren. Mutationen im menschlichen POLH Gen, welches für Pol η kodiert, verursachen die variante Form von Xeroderma Pigmentosum (XP-V). Darüberhinaus trägt diese Eigenschaft von Pol η zur Resistenz gegenüber Cisplatin basierter Chemotherapie bei.

Im Rahmen dieser Arbeit, welche in Zusammenarbeit mit der Gruppe *Hopfner* durchgeführt wurde, wurde der Mechanismus der „Translesion Synthesis“ (TLS) durch *S. cerevisiae* DNA Pol η auf atomarer Ebene untersucht. Dazu wurde Cisplatin geschädigte DNA hergestellt. Diese wurden als Template im Komplex mit dem katalytischen Fragment von Pol η, einem ddCTP und Primer kristallisiert (Abbildung 1-1 A).
Bei der Analyse der erhaltenen Struktur zeigte sich, dass das fehlende 3’-OH für die korrekte Positionierung des NTPs und somit auch der DNA verantwortlich ist. Um den Komplex in einer katalytisch kompetenten Form zu erhalten, wurde die Struktur unter Verwendung von dNTPs und eines Primers mit terminaler 2’,3’-Didesoxyribose von

neuem mit Auflösungen von 3.1-3.3 Å bestimmt (Abbildung 1-1 B und C). Im Kristall liegt der Komplex in zwei Formen vor, wobei bei beiden das dNTP fest im aktiven Zentrum gebunden ist. Neben den für Polymerasen üblichen Komplexierung der Phosphate durch Metallionen, findet eine für Pol η spezifische Interaktion von einem Argenin mit den α- und β-Phosphaten statt. Die Bedeutung der Orientierung des NTPs durch diese Aminosäure für die TLS konnte in biochemischen Assays bestätigt werden. Im Gegensatz zu der rigiden Positionierung des NTPs interagiert die DNA flexibel mit dem Protein und liegt in zwei verschieden gebundenen Positionen im Kristall vor. Der Einbau des NTPs erfolgt durch rotierende Bewegung der DNA zum aktiven Zentrum, welche durch die Bildung von Wasserstoffbrücken zwischen NTP und Template getrieben wird. Der Einbau gegenüber dem 3'dG des Pt-GGs erfolgt durch die Bildung von Watson-Crick Wasserstoffbrücken und ist effizient und genau. Im Gegensatz dazu ist der Einbau gegenüber dem 5'dG des Pt-GGs weniger effizient und ungenau In *primer extension assays* werden dCTP und dATP gleich gut eingebaut. Strukturell lässt sich dies mit einer Fehlstellung des 5'dG vom Pt Crosslink erklären.

In Zusammenarbeit mit der Gruppe *Cramer* wurden die strukturellen Ursachen für die Blockade der RNA Polymerase II durch den Cisplatin Schaden untersucht. Dabei kamen sowohl kristallographische als auch biochemische Methoden zum Einsatz. Diese Untersuchungen zeigten, dass der Schaden nicht in die aktive Tasche eindringen kann, welches zur Blockade des Enzyms führt. Zusätzlich führt die Blockade zum Fehleinbau eines AMPs, welcher auch gegenüber einer *Abasic site* geschieht. Dies scheint der *A-Regel* zu folgen, welche bisher nur für DNA Polymerasen bekannt war. Da die Polymerase über einen Schaden, welcher artifiziel nach der translokations Barriere platziert wurde, selbst in anwesenheit einer G-A Fehlpaarung hinweg lesen kann, verhindert die Blockade die transkriptionelle Mutagenese.

3 Introduction

The cellular DNA is permanently damaged by environmental stress such as UV-light and genotoxic agents causing lesion sites. If not repaired, these damages can lead to mutations or replication arrest and consequently to cell death or cancer. To maintain the integrity of the genome, cells have developed multiple pathways, such as nucleotide excision repair (NER) and base excision repair (BER) to overcome different types of DNA lesions. Due to constant damage formation there are also lesions present during replication of the genome in the S phase, which have not been repaired. These pose an obstacle to high fidelity polymerases, which are unable to bypass such lesions in an error free manner or are stalled[2]. Low-fidelity polymerases, belonging to the Y-family, replace stalled replication machinery and synthesize DNA past the damaged site. Once the lesion is overcome, the replicative polymerases then continue with high fidelity and processivity[3].

Among Y-family polymerases, Pol η is specialized to bypass strongly helix disturbing DNA lesions such as ultraviolet-induced *cis-syn* thymine dimers (TT dimers) or 1,2-d(GpG) cisplatin intrastrand crosslinks. The translesion synthesis performed by Pol η proceeds with remarkable efficiency and accuracy[4-9]. In humans, the replication of Pol η through TT dimers establishes besides repair of these lesions a second line of defence against sunlight induced skin carcinogenesis, whereas mutations in Pol η cause the cancer-prone syndrome xeroderma pigmentosum (XP-V)[7, 9]. The beneficial effect of translesion synthesis by Pol η is contrasted during chemotherapy with cisplatin, where Pol η allows tumor cells to replicate through intrastrand crosslink. Although this lesion is normally repaired by the NER pathway, it is often overseen by the cells' repair mechanisms[10] and thus completely inhibits replicative DNA and RNA polymerases[11, 12]. Recently it has been shown by *in vivo* studies that Pol η plays a significant role in modulating cellular sensitivity to various anticancer agents and thus contributes to acquired drug resistance of cells to cisplatin in cancer chemotherapy[13, 14].

3.1 DNA and DNA damages

DNA

Deoxyribonucleic acid (DNA) contains the genetic instructions used in the development and functioning of all known living organisms. It is a linear polymer made up of four different monomers with a fixed backbone, built of repeating deoxyribose-phosphate units. Joined to each deoxyribose is one of four possible bases: adenine (A), cytosine (C), guanine (G) and thymine (T) (Figure 3-1.). The sequence of bases along a DNA strand constitutes the genetic information, which instructs indirectly for assembling proteins, which themselves orchestrate the synthesis of a host of other biomolecules such as RNA, sugars, lipids and many others that form cells and ultimately organisms[15, 16].

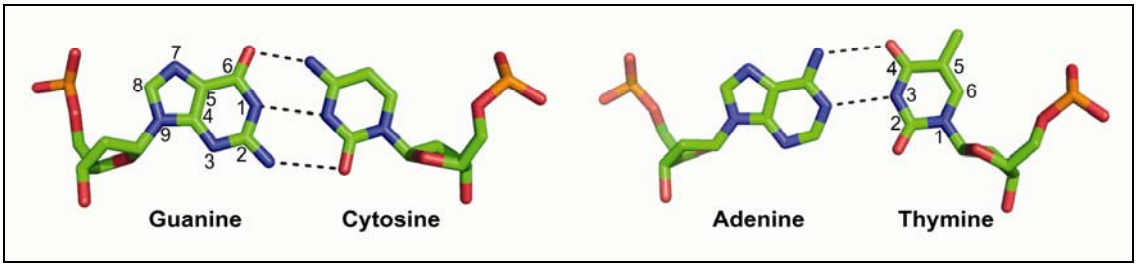


Figure 3-1. The four nucleotides in DNA. The hydrogen bonds established between the bases are shown as black dashes.

DNA exists in several possible conformations including: A-DNA, B-DNA, C-DNA, D-DNA[17], E-DNA[18], H-DNA[19], L-DNA[17], P-DNA[20], and Z-DNA[21, 22]. However, only A-DNA, B-DNA, and Z-DNA (Figure 3-2) have been observed in naturally occurring biological systems. The conformation, which DNA adopts, is depending on the sequence of the DNA, the amount and direction of supercoiling, chemical modifications of the bases and also solution conditions, such as salt concentrations. B-DNA is viewed as the native structure of DNA, as it prevails in cells, featuring a right handed double helix. On the outside of the double helix the spaces between the intertwined strands form two helical grooves of different widths described as the major groove and the minor groove.

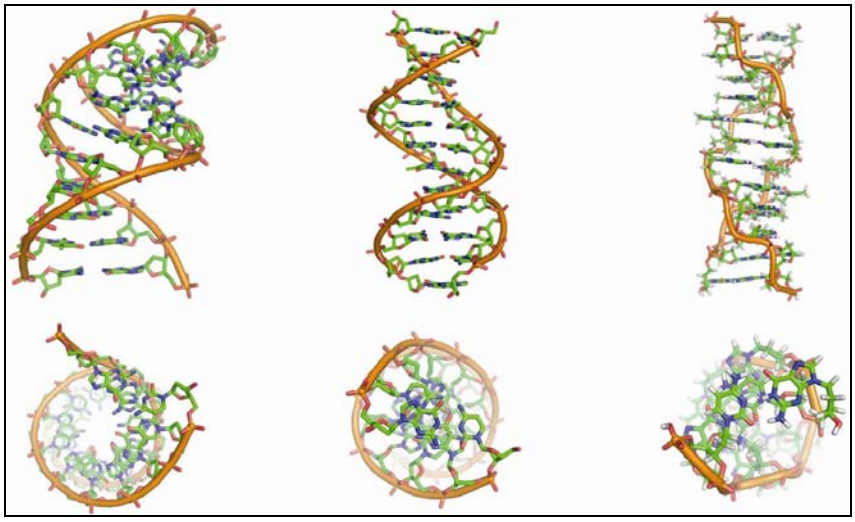


Figure 3-2. The three most important DNA secondary structure families in side view (top) and along the helix axis (bottom). Left: A-DNA, middle: B-DNA and right: Z-DNA

A-DNA features a shallow and wide minor groove and a narrower and deeper major groove and a central hollow cavity. It occurs in dehydrated samples of DNA, such as those used in crystallographic experiments, while in the cell it may be produced in hybrid pairings of DNA and RNA strands, as well as in enzyme-DNA complexes [23, 24]. Z-DNA features a left handed helix with pair wise clustered base pairs and is mainly found in CpG alternating sequences or at very high salt concentrations. The major and minor grooves, unlike in A- and B-DNA, show little difference in width.

DNA damages

In all living organisms cellular DNA continually incurs a variety of damage types. In the cell, DNA is subject to spontaneous hydrolysis of its phosphodiester, *N*-glycosylic and amino bonds. Moreover DNA is exposed to constant attack by various oxygen species and free radicals that damage the sugar and the bases. Exposure to sunlight and to chemical pollutants adds further to this load. In total, each cell undergoes about 10^6 DNA damages per day[25].

The type of DNA damage produced depends on the type of mutagen. For example, UV light mostly damages DNA by producing thymine dimers, which are cross-links between adjacent pyrimidine bases in a DNA strand. Another example is oxidants, such as free radicals or hydrogen peroxide, which produce multiple forms of damage, including base modifications, particularly of guanosine, as well as double-strand breaks after attackof the sugar. Of these oxidative lesions the most dangerous are double-

strand breaks, as these lesions are difficult to repair and can produce point mutations, insertions and deletions from the DNA sequence, as well as chromosomal translocations.

Many mutagens intercalate into the space between two adjacent base pairs. For an intercalator to fit between base pairs, the bases must separate, which distorts the DNA superstructure locally because of the unwinding of the double helix at the site of intercalation. These structural changes inhibit both transcription and DNA replication, causing toxicity and mutations. As a result, DNA intercalators are often carcinogens, with acridines and ethidium bromide being well-known examples. Nevertheless, due to their properties of inhibiting DNA transcription and replication, some of these chemicals are also used in chemotherapy to inhibit rapidly-growing cancer cells. These include cisplatin and nitrogen mustards, both belonging to the family of alkylating agents. These anticancer agents form crosslinks upon reaction with DNA.

Cells have devised ingenious mechanisms, such as nucleotide excision repair and base excision repair processes for repairing and tolerating damages of genomic DNA. Failure of these mechanisms can lead to serious disease consequenses, as well illustrated in the human hereditary disease xeroderma pigmentosum (XP), hereditary non-polyposis colon cancer (HNPCC) and some forms of breast cancer. XP is characterized by about a 10.000-fold increased risk of skin cancer associated with sunlight exposure, due to impaired nucleotide excision repair processes[26, 27]. Individuals with HNPCC manifest an increased hereditary predisposition to colon cancer.

3.1.1 Cisplatin lesions

cis-Diamminedichloroplatinum(II) (cisplatin) is a potent anticancer drug that is widely used to treat testicular, ovarian, head, neck, and small cell lung tumors[10, 28-30]. Although cisplatin was approved by the U.S. Food and Drug Administration (FDA) in 1978 and is one of the most successful anticancer drugs, side effects, natural and acquired resistance of patients toward the drug and its limited scope have motivated searches for structurally and functionally analogous alternatives[31-34]. These efforts include functionalization of ligands[33], bimetallic platinum systems[35] and different metal centers, most notably Rh and Ru[36-39]. Unfortunately, finding analogous compounds that outperform cisplatin has been proven to be difficult.

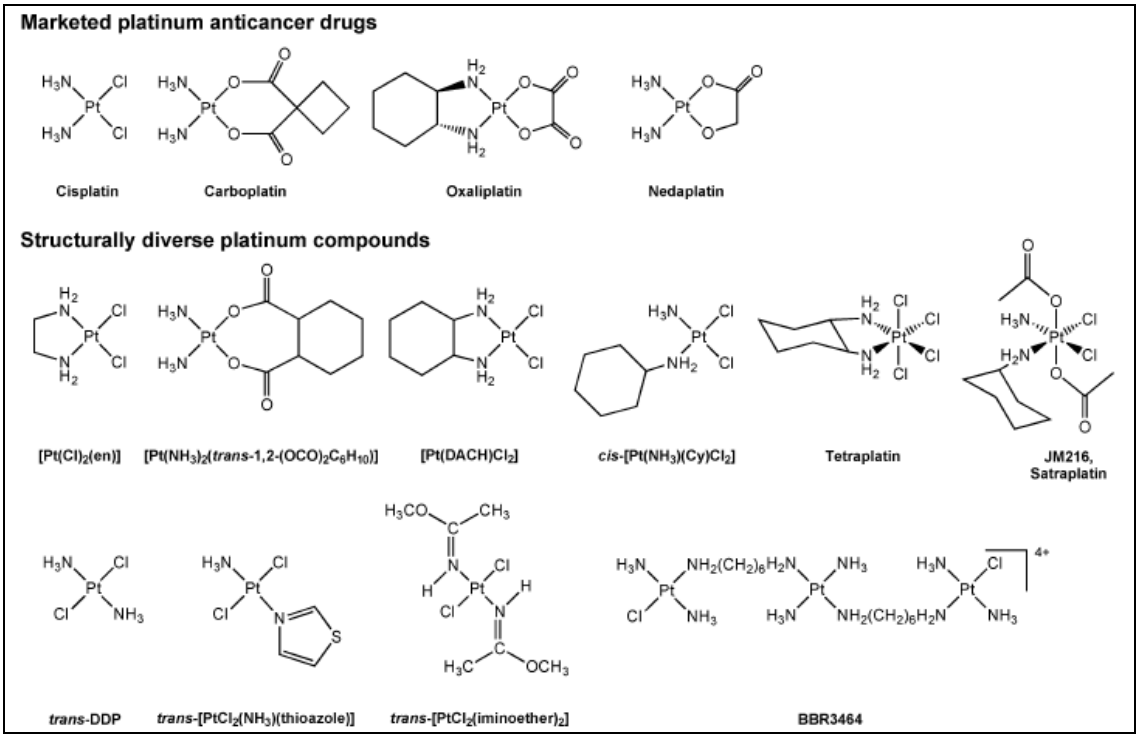


Figure 3-3. Chemical structures of platinum compounds including marketed platinum anticancer drugs[40].

Over 3000 platinum compounds have been synthesized and tested for their biological activity since the discovery of cisplatin[41, 42]. Of these, however, fewer than 30 compounds have entered clinical trials. At present, only four platinum drugs are registered as marketed drugs: cisplatin, carboplatin, oxaliplatin, and nedaplatin. Cisplatin, carboplatin and nedaplatin form the same DNA adduct *in vivo*[43-45]. Until recently, a new potentially promising drug belonging to the platinum family of compounds was satraplatin[46]. Unlike the platinum drugs currently on the market, which all require intravenous administration, it can be administered orally; however the drug failed the phase 3 of clinical trials. For an overview of platin based drugs see Figure 3-3.

Despite its widespread therapeutic use, the mode of action of cisplatin is still not completely understood, although the general reaction patterns are largely agreed on[10, 47-49]. Upon administration, cisplatin encounters a relatively high concentration of chloride ions (~100 mM) that suppresses hydrolysis of the two labile chlorido ligands that function as leaving groups and maintain the compound in a neutral state. Once inside a cell, however, the sharply decreased intracellular chloride concentration of around 4-12 mM causes cisplatin to undergo aquation,[21] whereby the chlorido ligands

are replaced by aqua ligands to form the activated complexes $[Pt(NH_3)_2Cl(H_2O)]^+$ and $[Pt(NH_3)_2(H_2O)_2]^{+2}$, which can react more readily with cellular targets (Figure 3-4). These complexes bind to various cellular components like DNA, RNA, proteins, and membrane phospholipids[50-54]. The major target leading to cell death, however, is genomic DNA[10, 55, 56]. Specifically, the N7 atom of purine bases is the main binding site, with guanine being preferred over adenine[57].

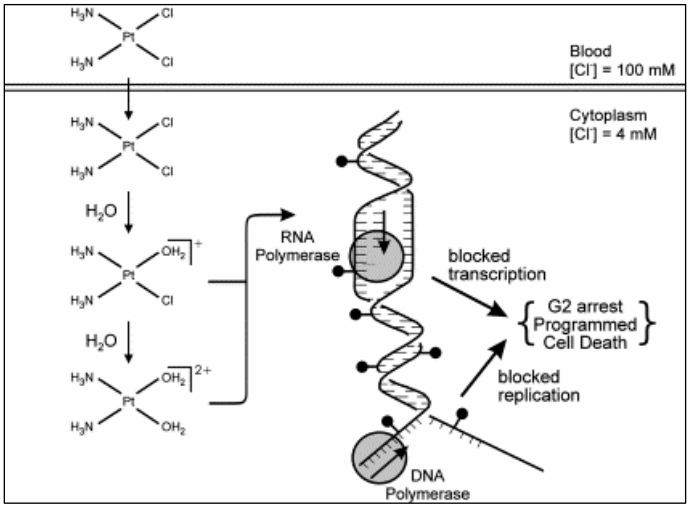


Figure 3-4. Schematic representation of cisplatin aquation chemistry in the cell and consequent cellular responses to cisplatin lesioned DNA. (Adapted from Kartalou[58]).

The possible binding modes include monofunctional binding to a single purine base, intra- and interstrand bifunctional binding, and DNA-protein cross-linking[48]. Due to the *cis* orientation of the leaving groups, intrastrand cross-links between two adjacent nucleobases are the most abundant products. The major lesions formed are 60-65% intrastrand GG, 25-30% intrastrand AG, 5-10% intrastrand GNG, and 1-3% interstrand adducts, with the underlined bases indicating the site of cisplatin binding[59]. Because of their abundance, the intrastrand GG adducts are believed to be the major determinants of the cytotoxic response to platinum anticancer agents[59-62].

In the case of the Pt-GG lesion the crosslink formation between two neighboured bases alters the DNA superstructure significantly. The lesion causes the adjacent guanine bases to roll towards one another, leading to an overall helix bend. In comparison to B-DNA, the 1,2-intrastrand GG cross-link unwinds the DNA duplex by about 10-20° in the vicinity of the site of platination and bends it by 60–80° towards the major groove.

This causes a widening and flattening of the minor groove opposite the platinum adduct, affording geometric parameters resembling those found in A-DNA[10, 63-65]. Structures of duplex DNA containing a 1,2- and 1,3-intrastrand cross-link are illustrated in Figure 3-5 A and B respectively.

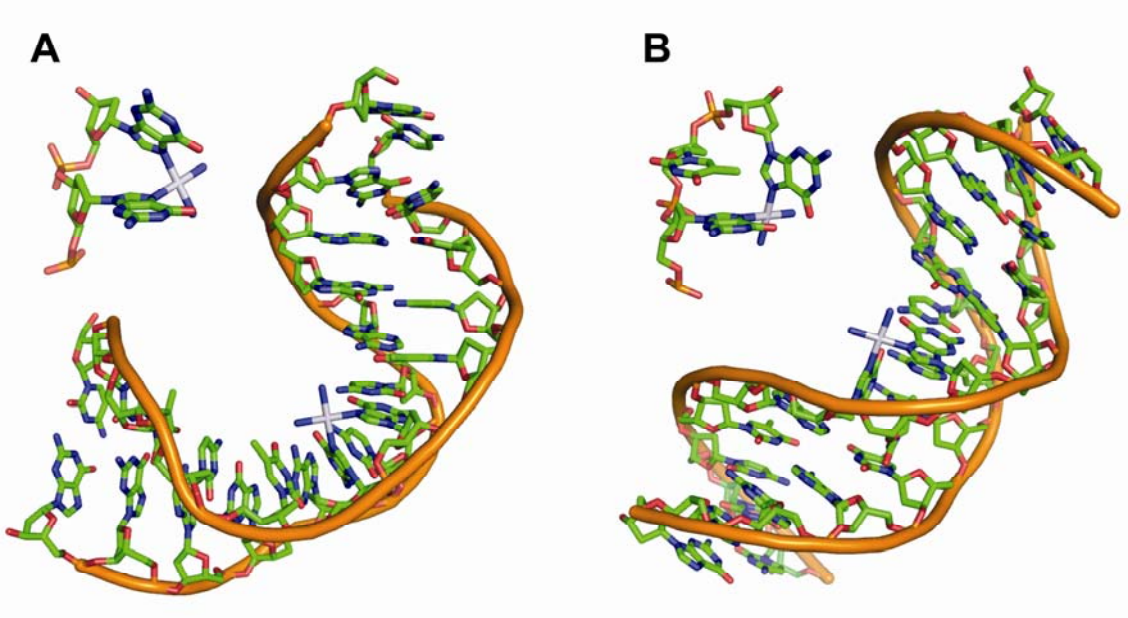


Figure 3-5. Platinum-DNA adduct structures and duplex DNA containing (**A**) cisplatin 1,2-d(GpG)[65]and (**B**) 1,3-d(GpTpG) intrastrand[66] cross-links, generated by PyMol.

3.1.1.1 Effects of Platinum Adducts on DNA Polymerases

Mammalian cells have the ability to synthesize DNA while ignoring various chemical lesions. The process, called translesion synthesis (TLS), demands specialized DNA polymerases, which are less stringent than the major replicative DNA polymerases and can accommodate damaged bases[67]. Various DNA polymerases have been tested *in vitro* for their ability to replicate past Pt-DNA adducts. The replicative DNA polymerases (Polymerases) α, ε and δ are completely blocked by cisplatin adducts[68, 69], even in the presence of the accessory proliferating cell nuclear antigen (PCNA) protein and replication protein A (RPA)[69]. Polymerases performing translesion synthesis past bulky adducts *in vivo* include members of the X, B, and Y families of DNA polymerases[2, 70, 71]. Among the Y-family polymerase, Pol ι, Pol κ, and Pol λ are incapable of inserting even a single deoxynucleotide triphosphate opposite cisplatin-DNA adducts[72-74]. Pol η is by far the most efficient enzyme at translesion synthesis past cisplatin- and oxaliplatin-DNA adducts *in vitro*[75-77].

3.1.1.2 Effects of Platinum Adducts on RNA Polymerase II

Already in 1974 it was realized that platinum based drugs can inhibit transcription by bacterial RNA polymerases[78]. These *in vitro* studies reported the ability of platinum ethylenediamine dichloride and cisplatin DNA adducts to inhibit transcription elongation by RNA polymerases[78, 79]. Similar to the inhibition of DNA synthesis, RNA polymerases are strongly blocked by bifunctional adducts. They are believed to encounter platinum lesions at a relatively early stage in the DNA damage-response process, due to their high abundance in the cell. RNA polymerase II (RNAP II), which transcribes most eukaryotic genes, is present with about 300 000 copies in a single cell, with nearly a fourth constantly involved in transcription[80, 81]. Both, platinum 1,2-(GpG) and 1,3-(GpTpG) intrastrand cross-links strongly block the elongation complex. A cisplatin lesion in the DNA template strand blocks transcription elongation by the single-subunit RNA polymerase from phage T7[82] and by RNAP II [11, 83, 84] and leads to stable polymerase stalling[85]. The stalled RNAP II elongation complex can be bound by the elongation factor TFIIS, which stimulates polymerase back-tracking and 3' RNA cleavage[12, 86]. Nucleotide excision repair can then occur in vitro at the DNA damage site with the polymerases remaining on the DNA and recruiting repair factors[87, 88]. After the repair, RNAP II can re-elongate, although a small fraction of polymerases can apparently read through a cisplatin lesion also without repair[11]. Arrested polymerases at the site of the platinum lesion not only function as damage recognition factors, triggering transcription-coupled repair (TCR)[89], but also mediate programmed cell death[90].

3.1.2 CPD lesions

To ensure genetic integrity when living cells are exposed to bright sunlight, nature developed several strategies to repair and deal with different types of UV damage. The three major lesions generated are cyclobutane pyrimidine dimers, pyrimidine-pyrimidone (6-4) photoproducts (6-4 PPs), and their Dewar valence isomers[91].

Of these, the *cis-syn* cyclobutane pyrimidine dimers (CPD) are the major photoproducts induced by UV-B light ($290\ nm < \lambda < 320\ nm$) present in solar light[92]. The photochemical [2+2] cycloaddition reaction involves the C5-C6 double bonds of two adjacent pyrimidine bases in DNA to form a cyclobutane ring system (Figure 3-6)[93]. Due to the conformation of the bases in double stranded DNA (*anti* at the glycosidic bond) only the *cis-syn* cyclobutane stereoisomer is obtained. Although both, thymidine

and cytidine, can react with each other in all four possible options, the order of reaction is 5'-T=T-3' > 5'-T=C-3' > 5'-C=T-3' > 5'-C=C-3' [94]. Thus, *in vivo* the *cis-syn* thymine (T=T) dimer is the predominant CPD lesion formed.

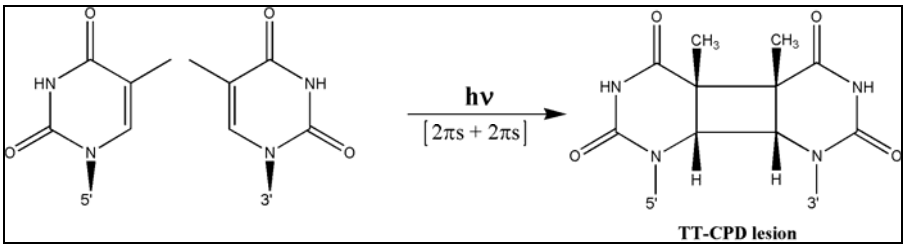


Figure 3-6. Formation of the TT-CPD-lesion by a photochemical [2+2] cycloaddition reaction.

Due to the covalent bonds between two neighboured bases, CPDs are steric demanding lesions, causing disturbance of the DNA superstructure. In comparison to B-DNA, CPD containing DNA is unwound by ~10° and the overall helical axis bends ~30° towards the major groove[95-98]. This results in a widening of both minor and major grooves both 3' and 5' of the CPD (Figure 3-7).

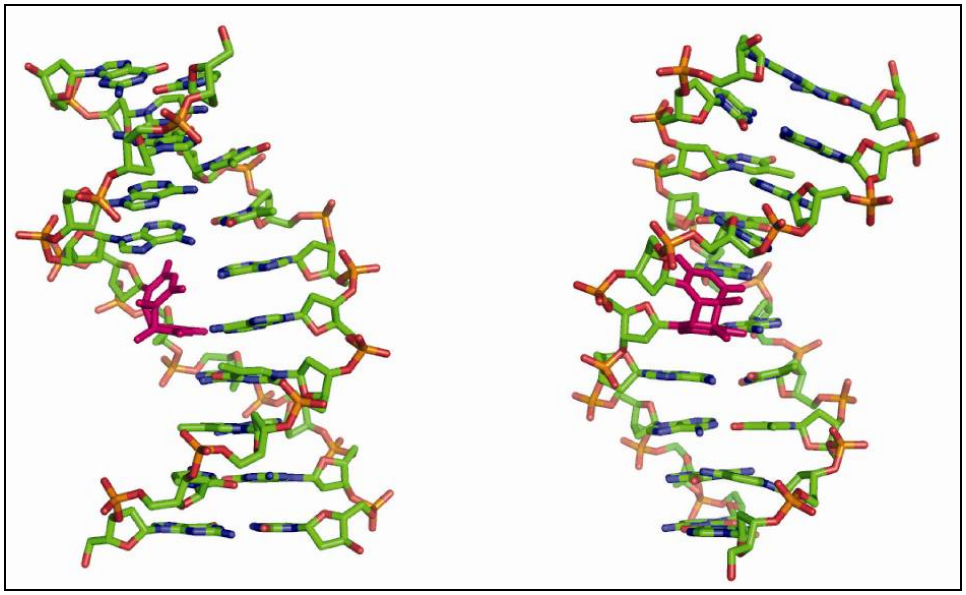


Figure 3-7. *The crystal structure of a DNA decamer containing a TT-CPD lesion. The structure is depicted from two different perspectives with the lesion shown in magenta. PDB entry 1N4E.*

3.1.2.1 Effects of CPDs on DNA Polymerases

The CPDs mostly exert their detrimental cytotoxic effects by blocking DNA replication and transcription and are important players in the formation of skin cancers[99]. Although the lesion introduces only a modest deformation of the DNA helix and the ability of the two Ts in the dimer to form Watson–Crick base pairs with the As is not significantly altered[95-98], a CPD still presents a strong block to DNA synthesis by most DNA polymerases, because the covalent linkage of the two Ts in the CPD prevents the kinking of the DNA backbone in such a way that the 5'dT of the dimer cannot be pushed out of the active site. This then hinders the classical replicative and repair Polymerases from replicating through this lesion, because they can accommodate only a single templating nucleotide in their active site[100-102].

The Y-family DNA Polymerases, on the other hand, can synthesize past DNA lesions, but they synthesize DNA with much lower fidelity and processivity than the replicative Polymerases. Pol η, for example, has the unique ability to replicate through CPDs proficiently and accurately[7, 8].

3.2 DNA Polymerases

Possibly the earliest enzymatic activity to appear in evolution was that of the polynucleotide polymerases. Faithful replication of DNA molecules by DNA polymerases into daughter nucleic acids is essential for genome integrity and stable transmission of genetic information in all living organisms and thus a prerequisite for known life. On the basis of phylogenetic relationships, DNA polymerases have been grouped into six families: A, B, C, D, X, and Y (Table 3-1).

Family	Examples	Error rate	Function
A	Pol I, T7, Taq, Pol γ	10^{-5} to 10^{-6}	Replication
B	Pol II, Pol α, δ, ε, ζ	10^{-5} to 10^{-6}	Replication
C	Pol III α subunit	10^{-5} to 10^{-6}	Replication
D	Pol D	10^{-5} to 10^{-6}	Replication
X	Pol β, λ, μ, σ	10^{-4} to 10^{-5}	Repair
Y	DinB, UmuCD, DPO4, Dbd, Pol η, ι, κ, Rev1	10^{-2} to 10^{-4}	TLS

Table 3-1. Overview of the 6 known DNA polymerase families, their fidelity and function.

The A-family is typified by *Escherichia coli* (*E. coli*) DNA polymerase I (Pol I), the B-family by *E. coli* Pol II, the C-family by the *E. coli* Pol III α-catalytic subunit, the D-family by archeal polymerases[103] and the X-family by eukaryotic Pol β. The prototypical Y-family polymerases include Pol IV and Pol V in *E. coli*, which are also known as DinB and UmuC respectively, and the eukaryotic Rev1 and RAD30. The latter has been renamed Pol η according to the alphabetic convention[104].

Structures of a number of high-fidelity replicative or repair polymerases have shown that they all share a similar architecture, which resembles a cupped right hand with palm, fingers, and thumb domains (Figure 3-8).

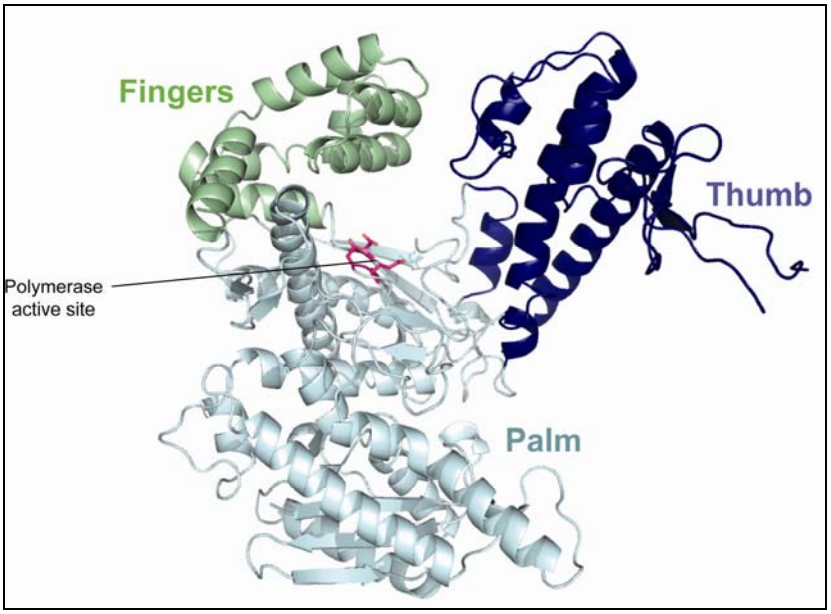


Figure 3-8. Structural features of T7 DNA polymerase[100]. The palm (*light blue*), fingers (*green*), and thumb (*dark blue*) domains are shown as cartoons. The catalytic center is shown in pink sticks.

In all these polymerases, the palm domain harbours the three conserved acidic residues that coordinate the binding of two divalent metal ions. This domain also contributes to the binding of the incoming dNTP.

3.2.1 Replicative polymerases

Replicative DNA polymerases are able to replicate the genome at high speed, with high processivity and with a very low error rate of about 10^{-8} to 10^{-10} per bases replicated[105], by using the replication clamp PCNA. To achieve this fidelity, a DNA polymerase must have an exceptional ability to discriminate against incorrect base pairs, which may exhibit only slight structural and energetic differences from the correct base pair. At least three distinct processes contribute to the high fidelity of DNA replication:

- *base selection:* the DNA polymerization reaction itself has a very low error frequency (~10^{-3} to 10^{-5}).
- *editing:* the proofreading reaction by the 3′→5′ exonucleases associated with replicative polymerases removes any base that might, on rare occasions, be mis-inserted. This reduces the error frequency by one or two orders of magnitude.
- *postreplicative mismatch repair:* this system further increases DNA replication fidelity by approximately another three orders of magnitude.

In high-fidelity replicative polymerases, the finger domain makes intimate contacts with the incoming dNTP, and the thumb domain contributes to duplex DNA binding. The active site of these enzymes fits very tightly with the templating base, the incoming dNTP and a few base pairs in the duplex DNA adjacent to the site of nucleotide incorporation. This puts a strong emphasis on the geometric shape complementarity of the incoming nucleotide with the templating base[100-102, 106-109]. As a result, geometric selection makes a predominant contribution to the nucleotide insertion specificities of these polymerases[110-112]. Most importantly, in all these polymerases, only a single unpaired template base is held in the active site, while the single-stranded template strand, including the next 5 unpaired template base, is flipped out of the active site at a 90° angle (Figure 3-9)[100-102]. All these features impose a high degree of geometric selectivity on the polymerases, which then accounts for their high fidelity. The downside to this is that they are highly sensitive to geometric distortions in DNA, and thus are unable to replicate through distorting DNA lesions.

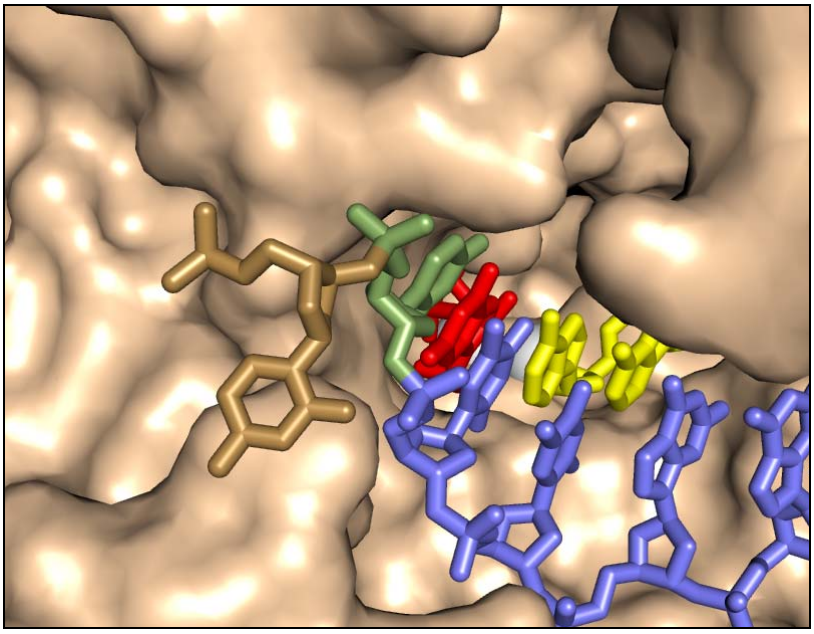


Figure 3-9. Protein surface representation of the T7 DNA polymerase with the DNA and dNTP depicted as sticks, showing how the nascent base pair is enveloped by the protein leaving little room for modifications on the templating base[100]. The primer is yellow and the dNTP to be inserted is red. The template strand is blue, with the templating nucleotide in olive and the nucleotide to its 5′-side in brown. The magnesium ions are depicted as grey spheres.

3.2.2 Low fidelity polymerases

3.2.2.1 Introduction

Although most types of DNA damage are removed by the cellular repair machinery, these processes are often slow and incomplete. Damages that are not repaired prior to the S phase are obstacles for the replication machinery. Most lesions cannot be accommodated into the active sites of replicative DNA polymerases; thereby the progression of the replication fork is blocked. Prolonged stalling of replication forks can lead to a collapse of the replication machinery. This might result in double-strand breaks and chromosomal rearrangements, or even to a permanent cell-cycle arrest and cell death. To avoid this potential crisis, cells have evolved bypass mechanisms that handle stalled replication forks[113].

An important mechanism for overcoming these blocks, particularly in mammalian cells, entails the use of specialized lesion bypass DNA polymerases to carry out translesion synthesis (TLS) past the damaged sites. Most of these polymerases belong

to the Y-family[114]. They all share conserved sequence motifs within the *N*-terminal ~350 residues, but their overall length and the *C*-terminal part vary considerably. The conserved active site, located at the *N*-terminal part of the Y-family polymerases, is characterized by a more open active site than in replicative polymerases. The variable *C*-terminus is involved in localization, recruitment and protein–protein interactions (Figure 3-10).

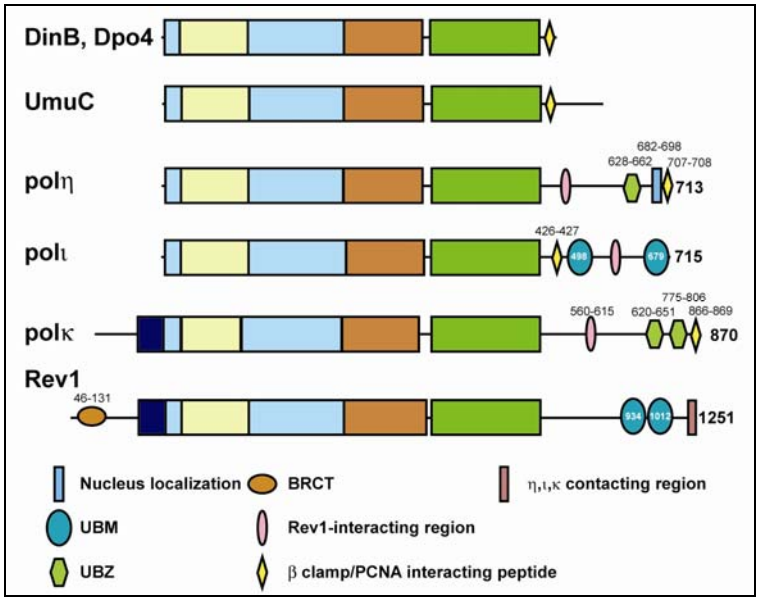


Figure 3-10. Structural domains of the Y-family polymerases. The polymerase domain is labeled in light blue (palm), yellow (finger), orange (thumb), green (PAD) and dark blue (*N*-terminal addition in Pol κ and Rev1). The regulatory units are color- and shape-coded as indicated at the bottom of the figure. UBM stands for ubiquitinbinding motif, UBZ for ubiquitin-binding zinc finger, and BRCT for Brca1 *C*-terminal domain. (Adapted from Yang [115])

Comparable to the replicative polymerases, the Y-family polymerases also contain the domains palm, fingers and thumb; however the last two are smaller and stubby. The Y-family polymerases bear an additional "little finger" domain, also termed polymerase associated domain (PAD) (Figure 3-11). This domain enlarges the potential protein-DNA interface and enhances DNA binding, which is still rather loose when compared to replicative polymerases. Y-family polymerases lack a $3' \rightarrow 5'$ exonuclease activity, which is an integral part of all replicative polymerases.

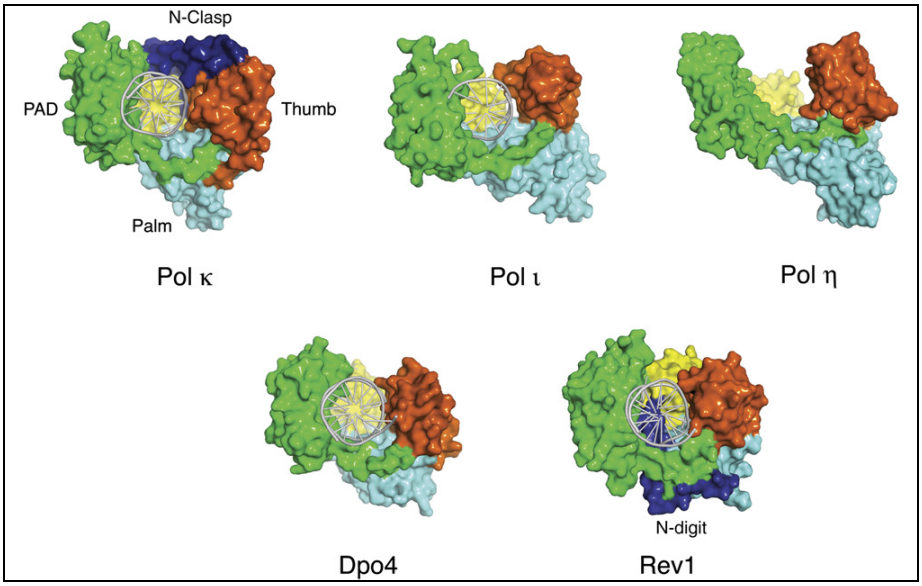


Figure 3-11. Surface diagrams comparing the structures of several Y-family polymerases. The colours of the domains coincide with those used in Figure 3-10.

Because low fidelity polymerases adopt a much more open structure than replicative polymerases, they are less stringent and can accommodate altered bases in their active sites, incorporating either a correct or an incorrect nucleotide opposite. A downside of this "sloppiness" is that translesion synthesis (TLS) is intrinsically error prone.

There are two Y-family polymerases in *Escherichia coli* (polymerases IV and V), two in *Saccharomyces cerevisiae* (Pol η and Rev1) and four in mammalian cells (Polymerases η, ι, κ and Rev1). In addition, the B-family DNA polymerase ζ, which is formed by the association of Rev3 and Rev7 proteins, also plays an important role in TLS in eukaryotes[116]. These five polymerases used in TLS in human cells have different substrate specificities, enabling them to deal with many different types of damaged bases[117].

3.2.2.2 Polymerase switch

In order for a TLS polymerase to take over synthesis when a replicative polymerase is blocked, the accurate replicative polymerase first needs to be replaced. PCNA plays thereby a central role in recruiting the TLS polymerases and effecting the polymerase switch from replicative to TLS polymerases[67]. Following replication fork blockage, Rad18 (an E3 ligase) binds to exposed single-stranded DNA at the fork and mediates together with Rad6 (an E2 ubiquitin conjugating enzyme) mono-ubiquitylation of PCNA. The affinity of PCNA for the TLS polymerases, which all have ubiquitin-binding motifs, is increased, thereby facilitating their engagement at the stalled fork. Thus replicative DNA polymerases are replaced and TLS is effected (Figure 3-12).

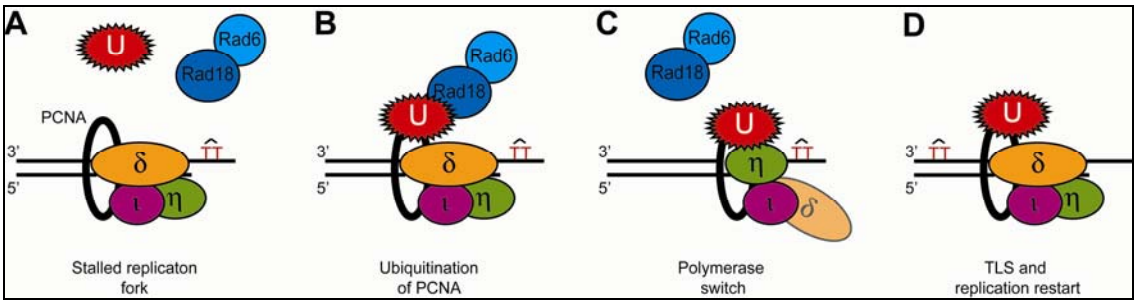


Figure 3-12. A proposed mechanism for translesion DNA synthesis. **A:** A replication fork, consisting of PCNA and Pol δ, is stalled when encountering a DNA lesion, for instance, a TT dimer. **B:** Blockage of the fork activates Rad6-Rad18, which mono-ubiquitinate PCNA. **C:** Modified PCNA then recruits the lesion bypass Y-family polymerase Pol η to carry out TLS. **D:** After the damage has been by-passed, replication by Pol δ restarts. (Adapted from Lehmann[67])

3.2.2.3 DNA polymerase η

Introduction

Pol η was first identified in yeast, where it was shown to replicate efficiently through a *cis-syn* thymine dimer by incorporating two adenines opposite the two thymines[8]. This observation, which implied that Pol η could promote error-free replication through UV-induced cyclobutane pyrimidine dimers (CPDs), was in accordance with the observation that deletion of the *RAD30* gene in yeast conferred an enhancement of UV mutagenesis[118]. This previously uncharacterized *S. cerevisiae* DNA repair gene, related to the *E. coli* dinB, umuC and *S. cerevisiae* REV1 genes, encodes Pol η.

Mutations in human Pol η result in a cancer-prone syndrome, the variant form of xeroderma pigmentosum (XP-V)[7-9]. Unlike cells from patients belonging to the classical xeroderma pigmentosum (XP) complementation groups, which harbour a defect in NER, XP-V cells are proficient in NER but defective in the replication of UV-damaged DNA[119]. Cells from XP-V individuals are deficient in replication through a *cis-syn* TT dimer, unlike cells from healthy individuals[120, 121]. XP-V cells are hypermutable with UV light, and the frequency of adenine incorporation opposite the TT photoproducts is reduced in XP-V cells compared to that in normal cells[122, 123].

Fidelity and Mismatch Extension Ability of Pol η

Pol η and other Y-family polymerases are able to replicate through DNA lesions since they are not inhibited by the geometric distortions imposed by the presence of lesions in DNA. The idea that TLS polymerases can better tolerate the geometric distortion of a mismatched base pair in their active sites was substantiated from steady-state kinetic analyses with yeast and human Pol η, indicating that they misincorporate nucleotides with a frequency of ~10^{-2} to 10^{-3} [6, 124]. In a subsequent study, human Pol η was shown to be highly mutagenic in an *in vitro* DNA synthesis reaction, making one base substitution error for every 18 to 380 nucleotides synthesized depending on the mismatch[125]. This very low fidelity indicates a relaxed requirement for correct base pairing geometry and indicates that the function of Pol η may be tightly controlled to prevent potentially mutagenic DNA synthesis.

The fact that Pol η incorporates wrong nucleotides at a fairly high rate seems incongruous with its role in promoting error-free bypass of CPDs and in suppressing UV mutagenesis. However, both yeast and human Pol η extend mismatched primer termini with a frequency of ~10^{-2} to 10^{-3} relative to extension from matched primer termini [126, 127]. Thus, immediately after a misincorporation, Pol η would dissociate from the DNA rather than extend from the mispair. The resulting primer-terminal mispair then could be subject to proofreading by a 3'→5' exonuclease. Replication through a TT dimer by Pol η then would be more accurate than predicted from the fidelity of nucleotide incorporation alone[126]. Consequently Pol η would have another opportunity to incorporate the correct nucleotide. Such an idea is supported from experiments done with the SV40 origin-dependent replication system in human extracts, in which it was shown that human Pol η induced replication errors are

enhanced if the proofreading exonucleases are inhibited by the addition of high amounts of dGMP to the replication reaction [128].

Lesion Bypass by Pol η

Pol η is unique among eukaryotic DNA polymerases in its proficient ability to replicate through a *cis–syn* thymine– thymine (TT) dimer. This proficiency of Pol η is derived from its unique structural feature - the ability to accommodate both template nucleotides of a CPD in its wide and open active site (Figure 3-13)[129].

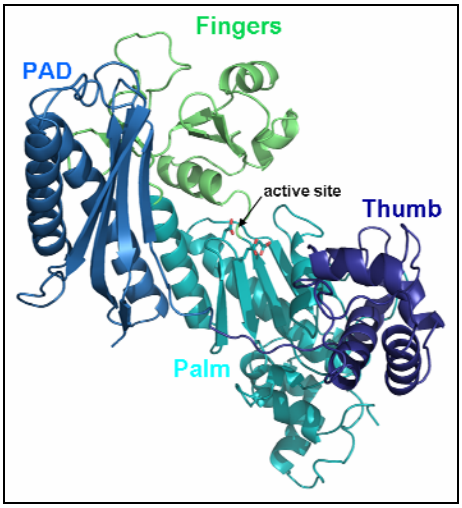


Figure 3-13. Structure of the Pol η apoenzyme[129]. The palm (*light blue*), fingers (*green*), and thumb (*dark blue*) domains are shown as cartoons. The polymerase associated domain (PAD), uniquely found in the Y-family of polymerases is also shown in blue cartoon. The catalytic center is shown as sticks.

Although the proficient ability to replicate through CPDs is the most prominent and biologically consequential feature of Pol η, it can promote replication through many other DNA lesions, such as Pt-GGs for example, as well.

CPDs

Pol η replicates through this lesion with the same efficiency and accuracy as it replicates through undamaged dTs. Steady-state kinetic studies have shown that both yeast and human Pol η insert two As opposite the 3'dT and the 5'dT of the TT dimer with the same efficiency and accuracy as they insert two As opposite two normal Ts in the undamaged sequence[5, 6]. Moreover, the polymerase exhibits the same processivity on the dimer containing DNA as on undamaged DNA. For both undamaged and

damaged DNAs, ~50% of the yeast Pol η molecules remain bound after the incorporation of two dNTPs opposite the 3'dT and 5'dTs of the dimer or opposite the two dTs in the analogous undamaged sequence, and ~30% of the enzyme molecules incorporate at least three dNTPs on both DNAs [5].

From presteady-state kinetic analyses of the different steps of the nucleotide incorporation reaction opposite the two dTs of the dimer, it has been determined that yeast Pol η binds the DNA when opposite the 3'dT of the dimer with about the same affinity as it binds the analogous site in undamaged DNA[4]. The nearly identical efficiency and accuracy of nucleotide incorporation opposite the damaged and undamaged dTs strengthens the inference derived from structural studies that Pol η can well accommodate both Ts of the dimer in its active site[5, 6, 129].

The idea that both Ts of a TT dimer stay in the active site of Pol η rather than the dimer being pushed out of the active site, as occurs in T7 polymerase[130-132], has also been examined in studies with N3-methyl derivatives of the 3'dT and the 5'dT of a TT dimer. These studies added further support to the assumption that both the Ts of the dimer remain in the active site and direct the insertion of an A[133].

In addition to the formation of CPDs at two adjacent thymines, UV also induces the formation of CPDs and (6–4) photoproducts at 5`-TC-3` and 5`–CC–3` sequences in the genome. The contribution of these lesions to UV mutagenesis is supported by the fact that the 3`-cytosine in both sequences is highly mutagenic and in yeast as well as in humans, C to T transitions are the predominant form of UV-induced mutations[134-136].

The ability of Pol η to accommodate the two template nucleotides of the TT dimer and to incorporate an A directly opposite the two Ts via Watson-Crick hydrogen bonding provides an elegant mechanism by which Pol η can also replicate efficiently and accurately opposite CPDs at 5'-TC-3' and 5'-CC-3' dipyrimidine sites. Pol η also functions in the error-free bypass of CPDs formed at these sites. This is supported by the observation wherein inactivation of Pol η in both yeast and humans leads to a large increase in the frequency of C to T transitions, which result from the misincorporation of an A opposite the 3'C of CPDs at 5'-TC-3' and 5'-CC-3' sites in the absence of Pol η[137, 138].

CPDs are responsible for a number of UV-induced mutations in mammalian cells. They are formed much more frequently than (6-4) photoproducts[134-136, 139-141]. Moreover, CPDs are removed much less efficiently than the (6-4) photoproducts by NER[142]. Hence, the proficient ability of Pol η to replicate efficiently and accurately opposite

CPDs formed at various dipyrimidine sites provides a large measure of protection from sunlight-induced skin cancers in humans.

Cisplatin

Translesion synthesis through cisplatin-DNA adducts has been an interesting aspect of DNA synthesis in cisplatin-treated cells because of its correlation to drug sensitivity[143]. Cisplatin-resistant cells exhibit more TLS than drug-sensitive cells[144-146]. Specialized DNA polymerases are overexpressed in many cancer cells and have a role in the cellular tolerance to cisplatin DNA damage[13, 147]. The process also plays a critical role in conveying the mutagenic properties of cisplatin because of the nature of TLS, which carries out both error-prone and error-free DNA synthesis[117]. The mutagenicity of cisplatin is closely related to the evolution of resistance of cell lines against the drug. In particular, the reduced ability to replicate cisplatin damaged DNA decreases the rate at which the cells become resistant to cisplatin. For example, suppression of human DNA polymerase involved in TLS, such as polymerase Rev 1[148] or Pol ζ[149, 150], increases the sensitivity of cells to cisplatin and reduces the rate of appearance of cisplatin resistance.

(6-4) TT photoproducts

Whereas a *cis-syn* TT dimer has only a modest effect on DNA structure and retains the ability to form Watson-Crick base pairs with the correct nucleotides[95-98], a (6-4) TT photoproduct induces a large structural distortion in DNA, with the 3'dT of the (6-4) lesion oriented perpendicular to the 5'dT[151, 152]. The O2 carbonyl of the 3'dT in the (6-4) TT lesion can hydrogen bond with the imino and amino protons of a guanine (G), whereas the 5'dT of the lesion retains the ability to form normal Watson-Crick hydrogen bonding with an A[153]. Although both, yeast and human Pol η are unable to replicate past a (6-4) TT lesion, they both incorporate a G opposite the 3'dT of the lesion about eightfold better than an A. However, the efficiency of G insertion opposite the 3'dT of this lesion is still about 50 to a 100-fold reduced compared to the insertion of an A opposite the corresponding undamaged T residue[154]. The subsequent extension step opposite the 5'dT is performed by Pol ζ[154, 155].

Because (6-4) photoproducts are formed more frequently at 5'-TC-3' and CC sequences than at TT sites [134-136, 139-141], the ability of Pol η to insert a G opposite the 3'dT of a (6-4) TT lesion strongly suggests that Pol η also contributes to the error-free bypass of

(6-4) lesions formed at 5'-TC-3' and CC sites. Similar to the 3'dT of (6-4) TT, the O2 carbonyl of the 3'dC in the (6-4) TC and CC photoproducts can also form hydrogen bonds with the opposing G. The incorporation of a G opposite the 3'dC of TC and CC photoproducts, followed by extension by Pol ζ, would result in the error-free bypass of (6-4) lesions formed at these sites[156].

8-Oxoguanine

Whereas the absorption of UV-B and UV-C radiation from sunlight by DNA results in the formation of CPDs, (6-4) photoproducts, and other photolesions, the action of UV-A radiation from sunlight on cellular chromophores generates reactive oxygen species[94, 157]. Hence, exposure of mammalian cells to sunlight produces, in addition to CPDs and other photolesions, a significant level of 8-oxoguanine (8-oxoG) lesions. Pol η plays a prominent role in efficient and accurate replication through the 8-oxoG lesion[158]. The efficiency by which yeast Pol η incorporates a C opposite the lesion and then extends from the inserted nucleotide is remarkably similar to that at an undamaged G. Evidence for the *in vivo* role of Pol η in the error-free bypass of the 8-oxoG lesion comes from the observation of the rate of spontaneous G:C to T:A transversions, a characteristic feature of mutations resulting from the mutagenic bypass of 8-oxoG, is synergistically enhanced upon the simultaneous inactivation of Pol η and the Ogg1 DNA glycosylase, which removes the 8-oxoG paired with a C. Human Pol η also predominantly incorporates a C opposite 8-oxoG. However, in this case a low but significant level of A is also inserted[158].

By promoting error-free replication through the 8-oxoG lesion, Pol η would contribute to minimizing the incidence of skin cancers, as well as of internal cancers, that would have otherwise resulted from the mutagenic bypass of this lesion by replicative DNA polymerases[159].

Other lesions

Pol η can replicate through an ^{6}O-methyl guanine (m6G) lesion, but opposite this lesion, it incorporates the C and T nucleotides nearly equally well. In contrast to the efficient bypass of CPDs and 8-oxoG, replication through the m6G lesion is inhibited ~20-fold at the nucleotide incorporation step[160].

An abasic site presents a severe block to replication by Pol η[161]. Even for a G or an A, the two nucleotides incorporated most often by Pol η opposite an abasic site, the

efficiency of nucleotide incorporation opposite the lesion is reduced by almost 1000-fold and the extension step is similarly affected[161].
Pol η is also inhibited by N^2-deoxyguanosine adducts of benzo[a]pyrene 7,8-diol 9,10-epoxide (BPDE)[162], butadiene epoxide[163] and by the acrolein-derived adduct γ-hydroxy-1,N^2-propano-deoxyguanosine (γ-HOPdG)[164]. Replication by Pol η is also inhibited across from 1,N^6-etheno-deoxyadenosine[165].

The ability of Pol η to efficiently replicate through TT dimers, cisplatin lesions, 8-oxoG lesions and also to replicate through other helix distorting DNA lesions has indicated that Pol η is rather insensitive to geometric alterations conferred on DNA by these lesions[161, 164, 165]. However, the efficiency and accuracy of nucleotide incorporation by Pol η is impaired by DNA lesions that severely impinge upon the Watson-Crick hydrogen bonding of base pairs[166, 167]. In this regard, Pol η differs strikingly from classical high-fidelity DNA polymerases.

3.3 Eukaryotic RNA polymerase II

The genetic information encoded in DNA needs to be processed in order to yield functional proteins. This process consists of transcribing the genetic information to messenger RNA (mRNA) by RNA polymerases. The mRNA is then translated into protein by the ribosome.

Bacteria and archaea contain one such RNA polymerase, whereas in eukaryotes, RNA is synthesized by three different types of polymerases, namely RNA polymerase I, II and III[168-170]. These three polymerases differ in promoter specificity, localisation in the cell and susceptibility to inhibitors. RNAP I transcribes 18S and 28S ribosomal RNA (rRNA) and is located in the nucleoli. RNAP III synthesizes small nuclear RNAs (snRNAs), transfer RNAs and the 5.7S rRNA and is located in the nucleoplasm[171].

RNAP II, also located in the nucleoplasm, transcribes all mRNAs in the course of gene transcription and some snRNAs[172, 173]. RNAP II consists of 12 subunits with a total mass of 513 kDa. Sequence analyses showed that two big subunits are highly conserved and are the only polypeptides present in all types of cellular RNA polymerases[174]. Various biochemical[175-177] and structural studies[178, 179] have shown that these highly conserved subunits contain the catalytic center responsible for DNA-directed RNA transcription.

3.4 Structure determination by X-ray crystallography

Knowing the three-dimensional structure of a given protein is a big step towards understanding the function of the same. Currently only few techniques are available for determining three-dimensional macromolecular structures. Standard light microscopy can not be utilized for the visualizing molecules at the atomic level, due to the long wavelengths of about 4-8x10^{-7} m of visible light. At the scale of angstroms, the wavelengths needed to analyze atomic distances lie within the spectral range of X-rays.

Protein crystallography requires large three-dimensional crystals of the sample suitable for diffraction of X-rays, which often represents the limiting step. Once obtained, the crystals can be measured at in house diffractometers or at synchrotrons.

Another method that permits analysis at high resolution level of a proteins ternary structure is nuclear magnetic resonance (NMR). As NMR allows structural elucidation of soluble proteins, it provides information of dynamic processes that cannot be

obtained from the rather rigid surroundings of a crystal. However, NMR is limited to proteins of small molecular weight of up to 35 kDa.

At certain speeds, neutrons and electrons can serve as radiation as well. These are employed in the electron microscopy technique, which allows elucidation of macromolecular surface structures, by focusing the scattered electrons with magnetic fields. However the method is limited to a maximal resolution of 6-8 Å due to the poor signal-to-noise ratio and the fast damaging of the sample. This makes the method more appropriate for very large assemblies, which are difficult to crystallize.

In summary, although crystallography, when compared to NMR, gives a more static description of the macromolecular structures, there are no limits in the size of the molecule to be analyzed. This makes X-ray crystallography the method of choice for studying macromolecules and their complexes at an atomic level.

3.4.1 Crystallization

Protein crystallography requires well ordered crystals of the sample suitable for diffraction of X-rays. Since proteins are usually large spherical or ellipsoidal objects with irregular surfaces, large crystals, wherein the protein is highly ordered in a three dimensional array, are quite difficult to obtain. The interactions of the macromolecule complex within a crystal are dependent on a plethora of parameters, such as the temperature, pH, nature of the solvent, precipitant and salts added, as well as the presence of ligands. Moreover, many proteins are composed of multiple functional domains with internal or terminal flexible regions. As now commonly believed and practiced, one usually attempts to remove all the flexible and non-functional parts and retains only the smallest functional domains to facilitate growth of crystals. In principle, this approach increases the probability of getting crystals because any flexible parts might inhibit the orderly packing of macromolecules in a crystalline array.

All these parameters may inhibit or support crystal growth. In addition, the conditions for crystal formation might be different from these optimal for crystal growth. Many crystallization conditions have to be tested during a crystallization experiment in order to obtain a condition which promotes growth of crystals suitable for X-ray diffraction analysis.

Crystals form when proteins are precipitated very slowly from supersaturated solutions. This thermodynamic driven process includes nucleation, crystal growth and growth termination. The most frequently used procedure for obtaining crystals is the vapour diffusion method[180].

In this thesis, the *hanging drop* vapour diffusion method was used. Therefore a buffered protein solution is mixed with precipitant solution from the reservoir in a 1:1 ratio on top of a cover slide. The reservoir contains a more concentrated precipitant solution and once the reaction chamber is sealed by turning the cover slide over and placing it on top of the chamber, slowly equilibrium between the reservoir and the hanging drop is reached. This optimally causes a saturation of the protein in the drop and if all conditions are right, protein crystals will occur in the drop.

3.4.1.1 Crystallizing a DNA polymerase in ternary structure

Crystallizing protein-DNA complexes is similar to crystallization of any macro-molecule. It depends on precipitant, ionic environment, pH and additives. DNA length is often the key factor that determines whether a protein-DNA complex crystallizes[181]. Since crystals are often grown at room temperature, one needs at least 7 bps to keep a DNA duplex stable at 20 °C. DNA sequences starting from the minimal length with increments of 1-2 bps can be screened for cocrystallzation.

Another variable is the sequence of the extra DNA that flanks the central essential sequence, particularly bases at the ends of DNA. It is often observed that DNA ends in crystals are packed against other DNA ends or protein molecules. Therefore, sticky ends, which contain single or double unpaired bases, are often employed so that the two ends of a single DNA are complementary. For example, one end has a 5’ protruding T and the other end has a 3’ overhanging A. Such DNA can polymerize in a head-to-tail fashion to form a repetitive linear array that potentially facilitates crystal growth[182].

To capture a DNA polymerase and substrate DNA binary complex, one only needs to mix the protein with a DNA that contains both a double-stranded region and a 5’ overhanging single stranded region. The junction between the double and single strand portions defines a specific site for polymerase to act[102]. For capturing a DNA polymerase, DNA substrate and incoming nucleotide in ternary complex, multiple approaches to stall the chemical reaction have been applied, which include:

- Using a dideoxynucleotide triphosphate as an incoming nucleotide. A number of polymerases discriminate against dideoxynucleotide triphosphates, and these nucleotide analogs reduce the chemical reaction rate so that an enzyme-substrate complex can be captured[183, 184].
- Using a 3' dideoxy primer strand, which does not contain an hydroxyl nucleophile to form a covalent bond with an incoming nucleotide[100].
- Replacing Mg^{2+}, which is often essential for a chemical reaction, by Ca^{2+}, which enables the DNA and nucleotide association but is ineffective in facilitating catalysis[185].
- Stabilizing protein-DNA complexes by covalent cross-linking. The basis for the design is the hypothesis that a protein α-helix tracks the DNA minor groove adjacent to the active site. By systematically replacing residues along that α-helix with Cys and modifying the DNA minor groove with a thiol group, a specific pair of modified protein and DNA can produce disulfide cross-linked complexes with high efficiency, while retaining the native conformation[186].

3.4.1.2 Crystal Seeding

If the concentration of a crystallizing protein is plotted against the concentration of the crystallizing agent, the resulting diagram divides the space into several areas depending on the physical state of the protein (Figure 3-14).

- At very high concentrations of both protein and crystallizing agent, the protein precipitates as an amorphous material.
- At lower concentrations, crystal nuclei may form, which can subsequently grow to diffracting crystals.
- At still lower concentrations, nuclei will not form, so generally no crystals appear. However, if a nucleus or crystal is placed in such a solution, it will continue growing. This area, where crystals grow but no nucleation takes place, is referred to as the metastable zone.
- At even lower concentrations the protein is completely soluble and nucleuses or crystals placed in such a solution dissolve.

It is often found that crystals grown in the metastable zone are better ordered and thus diffract better than crystals grown at higher concentrations. Also, sometimes it is difficult to obtain big enough crystals for diffraction in the nucleation zone, due to many but small crystals formed at these conditions. To grow crystals in the metastable zone, small crystals acting as nuclei, can be transferred into crystallization drops with lower protein and/or precipitant concentration. Crystal seeding can be performed as *microseeding*, by introducing only fragments of crushed crystals, or as *macroseeding*, by transferring an entire crystal to a new drop.

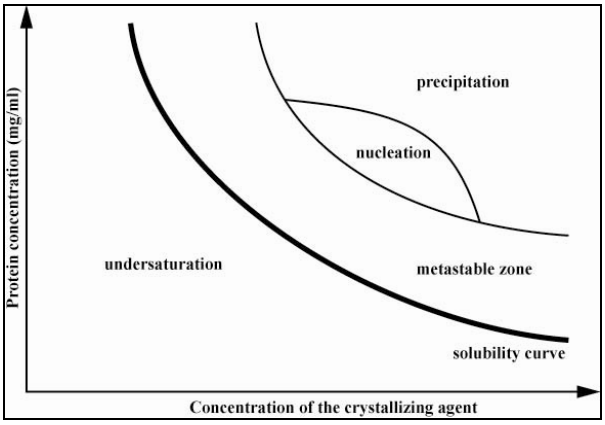


Figure 3-14. Crystallisation phase diagram. Schematic representation of a two-dimensional phase diagram, illustrating the change of protein molecules concentration against precipitating agent concentration. The concentration space is divided by the solubility curve into two areas corresponding to undersaturated and supersaturated state of a protein solution. The supersaturated area comprises of the metastable, nucleation and precipitation zones.

3.4.2 Theory of X-ray diffraction

X-rays are electromagnetic waves with a wavelength of atomic distances (10^{-10} m = 1 Å). For X-ray diffraction experiments of protein crystals, usually X-rays between 0.8-1.6 Å are used. X-rays interact with electrons in the electron sheath of atoms in the protein crystal and cause dipole oscillation of the electrons at the X-ray frequency. The oscillating electrons emit X-rays with the same wavelength in every direction (coherent or elastic scattering).

As a crystal is composed of molecules located in a unit cell, which is periodically repeated in three dimensions, the emitted waves scattered from different atoms interfere, which normally leads to their effacement. Only if waves have a retardation of

nλ (n = integer), constructive interference is observed. The retardation is thereby dependent on the distance between the scattering atoms. Thus, scattering can be described as reflection of imaginary lattice planes. These pervade the crystal lattice and its lattice points. Intersections with the unit cell axis of the crystal lattice are termed Miller Indices (hkl), which describe the orientation and the spacing between a set of parallel lattice planes. The conditions for constructive interference are given by Bragg's Law.

$$n \cdot \lambda = 2 \cdot d \cdot \sin\theta$$

Only if the distance d between parallel lattice planes and the angle θ between parallel lattice planes and the incident beam follow Bragg's law, a reflexion is observed, and results in one reflexion for each set of parallel lattice planes. The conditions of Bragg's law are graphically represented in a Ewald construction[187]. For this, the crystal lattice (real space) is transformed into a reciprocal lattice (Figure 3-15).

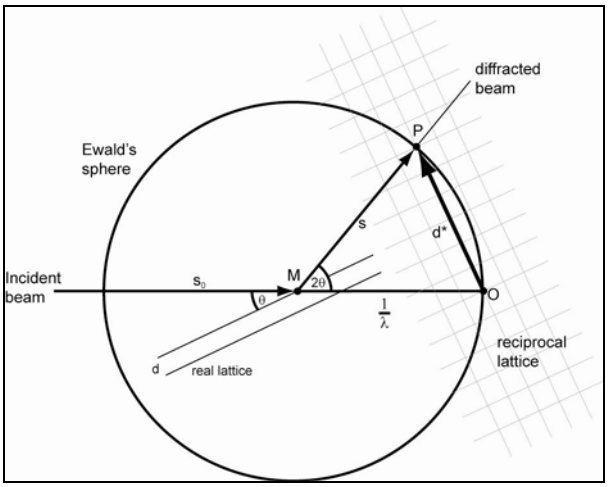


Figure 3-15. Two dimensional Ewald's sphere construction. The sphere has a radius of 1/λ. The origin of the reciprocal lattice is at **O**. $\mathbf{s_0}$ indicates the direction of the incident beam, which is scattered on a set of parallel lattice planes with distance **d** in the real crystal lattice. Scattering is only observed when a reciprocal lattice point **P** intersects with Ewald's sphere. In this case Bragg's law is fulfilled and the beam scatters into the direction **s.**

Each set of lattice planes can be described by a vector d_{hkl} , having the length $|d_{hkl}|$ between neighboring lattice planes and is perpendicular to hkl. According to Braggs law, the smaller d_{hkl} is, the bigger is the diffraction angle θ_{hkl}, and the resolution. In the reciprocal lattice the distance between lattice planes $d^*_{hkl} = 1/ d_{hkl}$, and vector d^*_{hkl} is also perpendicular to hkl. The end point of vector d^*_{hkl} is a lattice point of the reciprocal lattice, which is formed by the reciprocal unit vectors a^* ($a^* = 1/a$), b^* ($b^* = 1/b$) and c^* ($c^* = 1/c$), whereby a^* is perpendicular to hkl (100), b^* perpendicular to (010) and c^* perpendicular to (001). Each lattice point in the reciprocal lattice can be described by $d^*_{hkl} = ha^* + kb^* + lc^*$, and corresponds to one possible reflection hkl. According to Bragg's law, a reflection is observed, when a reciprocal lattice point P intersects with Ewald's sphere (radius $1/\lambda$), and the distance between the intersection point P and the origin O of the reciprocal (and crystal) lattice (OP) is d^*_{hkl}, and the angle between crystal M and intersection point P (KP) and crystal M and origin O (MO) is $2\theta_{hkl}$. With $2\sin\theta_{hkl} = MO/OP$ ($2\theta_{hkl} = \lambda/d$), Bragg's law is fulfilled. The diffraction beam results in a spot (h,k,l) on the detector. The intensity of the spot can be measured, which represents the overall scattering from a particular set of parallel lattice planes.

As both reciprocal and crystal lattice share the same origin O, the reciprocal is rotated with the crystal lattice. If the crystal lattice is rotated 180º around origin O, the reciprocal centrosymmetric lattice points will cross the Ewald's sphere. The lattice points d^*_{hkl} and $-d^*_{hkl}$ are referred to as Friedel mates.

3.4.3 The Phase problem

Light detectors, such as photographic plates or charge-coupled devices (CCDs), measure only the intensity of the light hitting upon them. This measurement is incomplete because a light wave has not only amplitude, but also a phase, which is systematically lost in a measurement. In diffraction or microscopy experiments, the phase part of the wave often contains valuable information on the studied specimen.

In x-ray crystallography, the diffraction data give, when properly assembled, the amplitude of the 3D Fourier transform of the molecule's electron density in the unit cell. If the phases are known, the electron density can be simply obtained by Fourier synthesis.

$$\rho(xyz) = \frac{1}{V}\sum_h\sum_k\sum_l \left|F_{(h,k,l)}\right| e^{i\varphi_{hkl}} e^{-2\pi i(hx+ky+lz)}$$

3.4.4 Solving the phase problem

Patterson function

The Patterson function *P(u,v,w)* is used to solve the phase problem in X-ray crystallography. It is the Fourier transform of $F^2(h,k,l)$ and thus can always be calculated from the recorded diffraction intensities. The function is defined as:

$$P(u,v,w) = \frac{1}{V}\sum_h \sum_k \sum_l |F_{hkl}|^2 \cos(2\pi[hu + kv + lw])$$

In the Patterson function, u, v and w represent the axis of the Patterson cell unit, whose dimensions are identical to those of the crystal unit cell x, y and z. Distances between atoms in the real structure show up as vectors from the origin to maxima in the Patterson cell unit, forming a Patterson map. The maxima are proportional to the heights of the peaks in the electron density map.

There are four methods for solving the phase problem:

1. Multiple isomorphous replacement (MIR)
2. Multiwavelength anomalous dispersion (MAD)
3. Molecular replacement (MR)
4. Direct methods

The availability of a three-dimensional structure of the Pol η apoenzyme allowed solving the crystal structures by the molecular replacement method.

Molecular replacement

The phases are obtained by molecular replacement using a previously solved model with sufficient homology to the structure to be determined with help of the Patterson function. With the coordinates of the homologous model positioned correctly in the unit cell, which is done by rotation and translation of the model, it is possible to calculate the amplitudes F_{calc} and the phases Φ_{calc}.

The vectors in the Patterson map can be divided into two categories: intramolecular vectors and intermolecular vectors. Intramolecular vectors, which depend only on the orientation of the molecule and not on its position in the cell, can be used in the rotation function. In this function, the Patterson map of the unknown structure is compared to Patterson maps from the model in different orientations. The highest correlation is

obtained once the two vectors are in similar orientations. In the subsequent applied translation function the model is translated to the correct coordinates in the asymmetric unit. For this, the intermolecular vectors can be exploited. These depend on both, the orientation of the molecule and on its position in the cell, therefore once the orientation of the model is known, the translation function can be applied for correct positioning of the model within the asymmetric unit.

With phases derived from the model and with the experimental observed amplitudes F_{obs} an electron density map is calculated. This model will then have to be carefully analyzed and, through several refinement and model rebuilding steps, modified to account for the differences between the model and the target molecule in order to maximize the agreement with the experimental data.

3.5 Research objectives

The framework of this doctor thesis was to determine how specialized translesion synthesis polymerases deal with DNA lesions, which block high fidelity polymerases. Cisplatin, a drug used in anticancer therapy, which forms DNA crosslinks mainly at sites of two adjacent purines, was chosen to be the model compound. Therefore the first research objective was to establish the preparation and purification of the most abundant cisplatin DNA intrastrand crosslinks. The 1,2-d(GpG) and 1,2-d(ApG) cisplatin adducts result in a structurally similar lesion, thus emphasis was put on the 1,2-d(GpG) cisplatin lesion due to its more common occurrence. Another lesion formed during cancer therapy is the 1,3-d(GpNpG) lesion, wherein the platinum binds to the two guanines usually separated by a thymine.

Once the lesions could be prepared selectively and site specific, the next step was to investigate the lesion bypass mechanism by co-crystallization of the lesion containing DNA with a low fidelity polymerase in order to obtain an understanding on the molecular basis of the translesion synthesis process past DNA helix disturbing lesions.

Cisplatin lesions are efficiently bypassed by translesion DNA polymerase η (Pol η) which belongs to the Y-family of polymerases. Pol η inserts preferentially two cytosines opposite the 1,2-d(GpG) cisplatin lesion. The polymerase also inserts faithfully two As opposite a TT dimer photolesion, thereby playing an important role in lesion tolerance. Although the crystal structure of a shortened version of the yeast Pol η apoenzyme, lacking regulatory domains, was determined in 2001 so far no structure of eukaryotic Pol η in ternary complex with DNA and NTP has been reported in the literature. Dpo4, an archeal DinB homolog, capable of replicating past cisplatin lesions has been crystallized in ternary complexes with lesion containing DNA and was used extensively as a model for eukaryotic DNA Pol η. However there are discrepancies upon the validity of this model.

Thus, the final aim of this work was to crystallize eukaryotic DNA polymerase η in ternary complex with a cisplatin lesion containing DNA and incoming nucleotide triphosphate and to solve the structure of this complex. This would provide important insights into lesion tolerance mechanisms displayed by this recently discovered family of DNA polymerases.

4 Experimental part

4.1 Materials

4.1.1 Chemicals

Unless otherwise mentioned, the chemicals used had at least analytic purity and were provided by Sigma-Aldrich (Deisenhofen), Fluka (Neu-Ulm), BioRad (Munich), Merk (Darmstadt), Qiagen (Hilden) and Roth (Karlsruhe). Phosphoramidites and supports for DNA synthesis were obtained from Glen Research (Sterling, USA). Crystallization screens and reagents were ordered from former NeXtal Biotechnologies (now QIAGEN, Hilden) and from Hampton Research (Aliso Viejo, USA).

4.1.2 Enzymes, Bacterial strains, Standards and Kits

4.1.2.1 Enzymes

DNA polymerase η for crystallization and primer extension experiments with cisplatin and oxaliplatin lesion containing DNA was kindly provided by Carsten Pieck and Claudia Chioccini[188, 189].

Pfu Turbo DNA polymerase was purchased from Stratagene, Amsterdam, Netherlands.

4.1.2.2 Bacterial strains

***E. coli* Strain**	**Genotyp**	**Company**
TOP10	F⁻ *mcr*A Δ((*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔ(M15 Δ(*lac*X74 *rec*A1 *ara*Δ(139 Δ((*ara-leu*)7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G	Invitrogen, Karlsruhe
Rosetta-gami™ (DE3)pLysS	Δ(*ara-leu*)*7697* Δ*lacX74* Δ*phoA Pvu*II *phoR araD139 ahpC galE galK rpsL* (DE3) F′[*lac*+ *lacI*^{q} *pro*] *gor522*::Tn*10 trxB* pLysSRARE (Cam^R, Kan^R, Str^R, Tet^R)	Novagene, Schwalbach

Table 4-1. *E. coli* Strains.

4.1.2.3 DNA- und Proteinstandards

Standard	**Company**
2-Log DNA Ladder (0,1–10,0 kb)	NEB, Frankfurt am Main
DNA Ladder 1 kb	NEB, Frankfurt am Main
SeeBlue Plus2 Pre-Stained Protein Standard	Invitrogen, Karlsruhe

Table 4-2. Standards.

4.1.2.4 Kits

Name	Use	Company
MinElute PCR Purification Kit	Purification of PCR products	Qiagen, Hilden
MinElute Gel Extraction Kit	Isolation of DNA from agarose gels	Qiagen, Hilden
QIAquick Miniprep Kit	Purification of plasmid DNA	Qiagen, Hilden
QIAquick PCR Purification Kit	Purification of PCR products	Qiagen, Hilden
Complete	Protease inhibitor	Roche, Mannheim
Gateway-System	Cloning	Invitrogen, Karlsruhe

Table 4-3. Isolation- und purification kits.

4.1.3 Consumables

Equipment	Company
Acrodisc 13 mm syringe filter 0.2 µm	VWR, Darmstadt
VDX ™ 24 well crystallization plates	Hampton Research, Aliso Viejo, USA
Crystal*EX* ™ 96 well sitting drop plates	Corning, New York, USA
NeXtal DWBlock	QIAGEN, Hilden
Cryschem™ 24 well crystallization plates	Hampton Research, Aliso Viejo, USA
Amicon Ultra centrifugal filter devices	Millipore, Schwalbach

Table 4-4. Consumables for crystallization and protein/DNA preparation.

4.1.4 Chromatographic material

Equipment	Company
Sepac-C18 cartridges	Waters, Eschborn
5µ Silica-C18 RP columns	Macherey-Nagel, Düren
7µ Silica-C18 RP columns	Macherey-Nagel, Düren
HiTrap™ Heparin HP Column	GE Healthcare, Munich
Strep-Tactin Superflow cartridge	IBA, Göttingen

Table 4-5. Columns for desalting oligonucleotides, HPLC and FPLC separations.

4.1.5 Laboratory instruments

Instrument	Company
Agarose gel electrophoresis chamber	Biorad, München
Äkta Oligopilot 10	Amersham Biosciences
Autoclave Vakulab S3000	Systec, Gießen
BioPhotometer 6131	Eppendorf, Hamburg
Breeze HPLC system	Waters, Eschborn
Bruker Autoflex II (MALDI-TOF)	Bruker Daltonik GmbH, Bremen
Cary 100 UV-Vis spectrometer	Varian, Darmstadt
Centrifuge 5415R	Eppendorf, Hamburg
Centrifuge 5810R	Eppendorf, Hamburg
Cooling trap CT 02-50	Martin Christ, Osterode am Harz
Electroporator, Micropulser	Biorad, München
Expedite 8900 Nucleic Acid Synthesis System	PerSeptive Biosystems, Framingham, MA, USA
Gel imager IDA	Raytest, Straubenhardt
Gel imager LAS3000	Raytest, Straubenhardt
Incubator 1S	Noctua, Wiesloch
Incubator 44R	New Brunswick,
Incubator MIR-553	Sanyo, Osaka, Japan
LaChrom HPLC system	Merck-Hitachi, Darmstadt
Single Channel Pipettes	Eppendorf, Hamburg
Mastercycler Personal	Eppendorf, München
Matrix Hydra II	Thermo Scientific, Cheshire, UK
Mini Protean 3 Cell	Biorad, München
PCR Realplex	Eppendorf, Hamburg
pH meter MP220	Mettler Toledo,Gießen
Power Supply	Biorad, München
Sonicator	Bandelin, Berlin
Sorvall centrifuge, Evolution RC	Kendro, Dreieich
Stereomicroscope Leica MZ16	Leica, Bensheim
Thermomixer Comfort	Eppendorf, Hamburg
Vacuum concentrator RVC 2-25	Martin Christ, Osterode am Harz
Vacuum concentrator SpeedVac	Savant
EmulsiFlex-C5 homogenizer	Manheim
Vortex mixer	VWR, Darmstadt
Waterbath	Labora, Mannheim
Waters Millipore System	Millipore, Schwalbach

Table 4-6. Instruments used and their manufacturing/distributing companies.

4.1.6 Oligonucleotides

4.1.6.1 Oligonucleotides for crystallization experiments.

Abbreviation	DNA sequence
eta A	5′ TCT GGC TCA TCC AC
eta A1	G AGT AGG TG 5′
eta A2	G AGT AGG TGA 5′
eta B	5′ TCT GGC TCA TCC ACT C
eta B1	G AGT AGG TGA G 5′
eta B2	G AGT AGG TGA GA 5′
eta C	5′ TCT GGC TCA TCC ACT CAC
eta C1	G AGT AGG TGA GTG 5′
eta C2	G AGT AGG TGA GTG A 5′
eta D	5′ TCT GGC TCA TCC ACT CAC CT
eta D1	G AGT AGG TGA GTG GA 5′
eta D2	G AGT AGG TGA GTG GAA 5′
eta E	5′ TCT CTG GCT CAT CCA C
eta E1	GA GTA GGT G 5′
eta E2	GA GTA GGT GA 5′
eta F	5′ TCT CTG GCT CAT CCA CTC
eta F1	GA GTA GGT GAG 5′
eta F2	GA GTA GGT GAG A 5′
eta G	5′ TCT CTG GCT CAT CCA CTC AC
eta G1	GA GTA GGT GAG TG 5′
eta G2	GA GTA GGT GAG TGA 5′
eta H	5′ TCT CTG GCT CAT CCA CTC ACC T
eta H1	GA GTA GGT GAG TGG A 5′
eta H2	GA GTA GGT GAG TGG AA 5′

Table 4-7. Initial screening for crystals in ternary complex with bound DNA.

Abbreviation	DNA sequence
eta O	5′ TCT TCT GGC TCA TAC CAC
eta I	5′ TC TCT GGC TCA TAC CAC
eta L	5′ C TCT GGC TCA TAC CAC
eta I1	GA GTA TGG TG 5′
eta P	5′ TCT TCT GGC TCA CCA C
eta J	5′ TC TCT GGC TCA CCA C
eta M	5′ C TCT GGC TCA CCA C
eta Q	5′ TCT GGC TCA CCA C
eta J1	G AGT GGT G 5′
eta N	5′ TCT TCT GGC TCA TCC AC
eta K	5′ C TCT GGC TCA TCC AC
eta S	5′ CT GGC TCA TCC AC
eta E1	G AGT AGG TG 5′
eta R	5′ CTC TGG CTC CCA C
eta R1	GAG GGT G 5′
eta T	5′ TCT CTG GCT CAT CCC A
eta T1	GA GTA GGG T 5′

Table 4-8. Refinements of the initial hit (etaE, Table 4-7).

4.1.6.2 Oligonucleotides for primer extensions.

Abbreviation	DNA sequence
DPO4_GG	5′ TCT GGA AAT CCT TCC CCC
GG_comp1	T TTA GGA AGG GGG-fluo 5′
GG_comp2	CT TTA GGA AGG GGG-fluo 5′
DPO4_GTG	5′ TCT GTG AAT CCT TCC CCC
GTG_comp1	TTA GGA AGG GGG-fluo 5′
GTG_comp2	C TTA GGA AGG GGG-fluo 5′
GTG_comp3	AC TTA GGA AGG GGG-fluo 5′
DPO4_CGG	5′ TCC GGA AAT CCT TCC CCC
eta E1fluo	GAG TAG GTG-fluo 5′
eta E2fluo	C GAG TAG GTG-fluo 5′
eta E3fluo	CC GAG TAG GTG-fluo 5′
eta E4fluo	A GAG ACC GAG TAG GTG-fluo 5′

Table 4-9. Oligonucleotides used for primer extension studies with Pol η. The site of the lesion is indicated in red letters. Primers are fluorescein labelled.

4.1.6.3 Oligonucleotides for crystallizations

Abbreviation	DNA sequence
eta Ep1	5′ TCT CTG GCT CAT CCA C GA GTA GGT G 5′
eta Ep2	5′ TCT TCT GGC TCA CCA C CG AGT GGT G 5′
eta Ett1	5′ TCT CTT TCT CAT CCA C GA GTA GGT G 5′
eta Ett2	5′ TCT UCT TTC TCA CCA C AG AGT GGT G 5′
eta Ett3	5′ CTC UTC TTT CTC ACA C AA GAG TGT G 5′
Pol2_GG	5'-ACC TCA ACT ACT TGG CCC TCC TCA TT-'3
Pol2_CGG	5'-ACC TCA ACT ACT CGG CCC TCC TCA TT-'3
Pol2_IG	5'-ACC TCA ACT ACT TGI CCC TCC TCA TT-'3

Table 4-10. Lesion containg oligonucleotides used for crystallizing Pol η and RNA Pol II in ternary complexes. The site of the lesion is indicated in red letters.

4.1.6.4 Oligonucleotides for cloning.

Abbreviation	DNA sequence
Rad30-Gate-N	5′-GGGGacaagtttgtacaaaaaagcaggctTCAT CGAAGGTCGTATGTCAAAATTTACTTGGAAGGAGT-3′
Pol η- 513_R	5′-ggggaccactttgtacaagaaagctgggtc TCATTTTTGTAAATCTATAATATCGAAATTAG-3′

Table 4-11. Primers used for cloning with the Gateway method. The *att*B site is indicated in small letters.

The 2'-deoxyInosin containing oligonucleotide Pol2_IG (table 4-10) was purchased RP-HPLC purified and lyophilized from biomers.net (Ulm). All other oligonucleotides were purchased RP-HPLC purified and lyophilized from metabion (Martinsried). For primer extension experiments the oligomers were 5'-fluorescein labelled.

4.1.7 Buffers, Mediums, Solutions and Antibiotics

4.1.7.1 Buffers

Heparin buffers A and B:

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

Strep-tag elution 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

Crystallization buffer:

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

Primer extension buffer:

40 mM Tris, 100 mM KCl, 5 mM $MgCl_2$, 10mM DTT, 2.5% (v/v) glycerol; pH 7.5

RP-HPLC buffers A and B:

Buffer A = 0.1 M $NHEt_3OAc_{(aq)}$

Buffer B = 0.1 M $NHEt_3OAc$ in 80% $MeCN_{(aq)}$

PBS-buffer:

140 mM NaCl, 8 mM Na_2HPO_4, 1.5 mM NaH_2PO_4; pH 7.2

10 x TBST- buffer:

100 mM Tris, 1.5 M NaCl, 0.5% (v/v) Tween 20; pH 8.0

Agarosegel- sample 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- sample buffer:

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

SDS- sample 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

10 x TBE buffer:

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

4.1.7.2 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 $MgCl_2$, 5.0 mM $MgSO_4$

Medium- Agar :

Medium with 1.5% (w/v) Agar

4.1.7.3 Solutions

MALDI matrix:

3-Hydroxypicolinic acid (50 mg), 15-Crown-5 (10 µl), Ammonium hydrogencitrate (10 mg) in ddH_2O (1ml)

Staining solution:

0.25% (v/v) Coomassie Brilliant Blue R 250, 10% (v/v) acetic acid, 40% (v/v) ethanol in ddH_2O

Destaining solution:

30% (v/v) ethanol, 10% (v/v) acetic acid in ddH_2O

4.1.7.4 Antibiotics and Inducers

Anhydrotetracycline 2 mg/ml in ethanol
Carbenicillin 100 mg/ml in ddH_2O
Chloramphenicol 34 mg/ml in ethanol

4.2 Methods

4.2.1 Platinum lesions

Cisplatin or oxaliplatin lesioned DNA can be obtained by directly allowing oligomers to react in solution with the platinum compound. In order to obtain a site specific lesion the oligomer has to be designed, so that flanking the site to be platinated no purine bases are present at all if possible. Since the GA lesion can not form[190], it is possible to have 3' to the lesion an A in the designed oligomer. Besides the Gs to be platinated in the reaction, it is favorable not to have any other Gs in the sequence at all, since these can result in monoadducts, which impede the purification process.

4.2.1.1 1,2-d(GpG) and 1,2-d(IpG) cisplatin lesion containing oligonucleotides

For obtaining the active aquated cisplatin species, *cis*-Diamminedichloroplatinum(II) (1.5mg) in water (1ml) was incubated with 19 µl of $AgNO_3$ (0.5 M) for 10h at 25 °C. The sample was centrifuged in a *5415R Centrifuge* for 10 minutes at 10000 rpm in order to sediment the AgCl. The activated cisplatin solution (1.5 equivalents) was incubated with 100 µM DNA, containing two site specific and adjacent guanines in the oligomer sequence for obtaining a site specific 1,2-d(GpG) lesion, or an inosine adjacent to a guanine for obtaining a site specific 1,2-d(IpG) lesion, in sodium perchlorate buffer (0.1 M, pH 6.0) and mixed overnight at 37 °C in the dark. The major reaction product was purified by HPLC and analysed by matrix-assisted laser desorption ionization–time of flight (MALDI–TOF) and contained a single cisplatin adduct located at the GG or the IG sequence.

4.2.1.2 1,3-d(GpTpG) oxaliplatin lesion containing oligonucleotides

Oxaliplatin reacts poorly with DNA *in vitro*, but its biotransformation product, Dichloro(1,2-diaminocyclohexane)platinum(II) (Pt(dach)Cl_2) reacts more readily with DNA *in vitro* and forms adducts with the same carrier ligands as those formed by oxaliplatin *in vivo*. Pt(dach)Cl_2 (1.9mg) was activated with 19 µl $AgNO_3$ (0.5 M) in water (1 ml) for 10h at 25 °C. The platination process is analogous to the one for obtaining the 1,2-d(GpG) cisplatin lesion, with the difference that the oligomer contains one site-specific GTG sequence. The major reaction product was purified by HPLC and

analysed by MALDI–TOF and contained a single oxaliplatin adduct located at the GTG sequence.

4.2.2 DNA synthesis, cleavage, purification and hybridization

DNA oligonucleotides were synthesized on a PerSeptive Biosystems Expedite 8900 Synthesizer and an Äkta Oligopilot 10 using UltraMILD bases and reagents and following standard phosphoramidite protocols (10 equivalents of phosphoramidite (0.1 M), 4 min recycling and 1 µmol scale synthesis; see Figure 4-1 for the reverse DNA synthesis reaction scheme). The phosphoramidite amounts for the 5-Bromo-2'-deoxyUridine (Br-dU), CPD and 2',3'-dideoxy nucleosides were equal to those of the natural bases but the coupling times were elongated to 15 minutes. While trityl values showed good incorporation of the Br-dU and TT dimer nucleosides, the coupling of the 2',3'-dideoxy nucleosides could only be verified by MALDI-TOF since these terminating nucleosides do not contain a trityl protecting group.

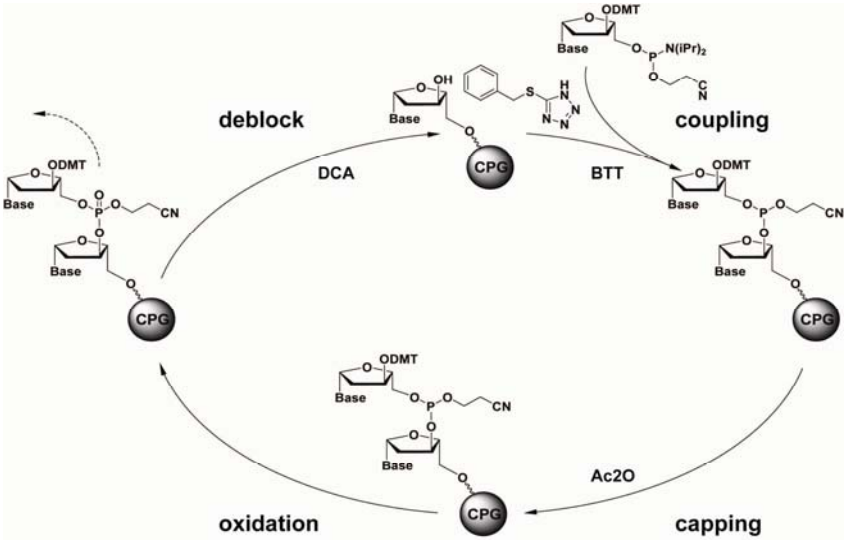


Figure 4-1. The solid-phase reverse DNA synthesis cycle.

The controlled pore size glass (CPG) solid support was subjected to a concentrated ammonia-ethanol (3:1) mixture for 12 h at r.t. for cleavage of the strands terminating with 2',3'-dideoxy nucleosides. The strands containing the Br-dU and CPD nucleosides were released from the CPG and deprotected by subjection to conc. NH_3 (aq) for 24 h at r.t. The solvents were removed in a SpeedVac concentrator and the resulting pellet

was redissolved in double destilled water. Subsequent filtration through a 0.2 µm filter removed the CPG. Analytics and purification were performed on Merck LaChrome HPLC systems using 5µ Silica-C18 RP columns for analytics and 7µ Silica-C18 RP columns for purification. For this, a gradient of 0-30% buffer B over 45 min, where buffer A = 0.1 M $NHEt_3OAc_{(aq)}$ and buffer B = 0.1 M $NHEt_3OAc$ in 80% $MeCN_{(aq)}$ was used. The purified fractions were concentrated, desalted on Waters Sepac-C18 cartridges and concentrated again. The concentration of the oligomers was estimated by UV spectroscopy following standard procedures on a Cary 100 UV-Vis spectrometer.
DNA was typically hybridized by depositing equimolar amounts of primer and template strand, both in double distilled water, into a vial. The water was evaporated in a vacuum concentrator SpeedVac and thereafter the DNA was resuspended in the crystallization- or primer extension buffer resulting in a 1 mM solution of dsDNA. Next the DNA was hybridized in a PCR Realplex cycler by heating to 85 °C and slow cooling to 4 °C. The dsDNA was then stored at 4 °C until use.

4.2.3 Molecular Biology Methods

4.2.3.1 Oligonucleotide design and Polymerase Chain Reaction

PCR primers Rad30-Gate-N and Pol η- 513_R (see Table 4-11) with *att*B containing sites for compatibility with the Gateway technology were used for amplification. As template served pDEST007-polη. The PCR reaction conditions are given below.

Reaction conditions:

2.5 µl 10x buffer
1 µl template DNA
1 µl Primer Rad30-Gate-N (10 pmol/µl)
1 µl Primer Pol η- 513_R (10 pmol/µL)
0,3 µl Pfu Turbo (2,5 U/mL)
0.5 µl dNTP mix (10 mM)
18.7 µl ddH_2O

PCR program:

1. 95°C 3 min
2. 95°C 15 sec
3. 58°C 30 sec
4. 72°C 5 min
5. GOTO 2 Repeat 5
6. 95°C 15 sec
7. 56°C 30 sec
8. 72°C 5 min
9. GOTO 6 Repeat 20
10. 72°C 10 min

The PCR product was purified by agarose gel electrophoresis. For this 0.4 g agarose in 50 ml TBE buffer was dissolved by boiling in a microwave oven. After the solution cooled down slightly 8 µl ethidium bromide were added. The gel was run at 95 V.
The purified DNA was extracted from the gel using the QIAquick PCR Purification Kit according to the manufacturers' protocol. The final DNA concentration was 30 ng/µl.

4.2.3.2 Cloning

Gateway™ Technology

Cloning of PCR products into the vector of choice was undertaken by using the Gateway technology, which is based on the site-specific recombination system used by phage λ to integrate into and excise from the *E. coli* chromosome. The attachment sites necessary for the homologous recombination event are provided by the PCR products and vectors, respectively.

The BP reaction was carried out according to the Invitrogen instruction manual for the Gateway™ Technology and contained:
2 µl PCR-produkt pDEST007-polη513(60ng)
1 µl pDONR 201 (150 ng/µl)
5µl TE-buffer pH 8
2 µl BP clonase

The LR reaction was carried out according to the Invitrogen instruction manual for the Gateway™ Technology and contained:
1 µl pDONOR-η 513 (120 ng/µl)
1 µl pDEST007 (150 ng/µl)
6 µl TE-buffer
2 µl LR clonase

4.2.4 Microbiological methods

4.2.4.1 Bacterial cultures

Culture of bacterial plates

Bacterial cultures were added with a sterile spreader onto antibiotic-containing plates. The plates were incubated overnight at 37°C. Thereafter plates were stored inverted at 4°C.

Bacterial overnight cultures

Over night cultures were grown in LB medium containing the relevant antibiotic in either round bottom Falcon tubes or in small Erlenmeyer flasks, allowing for air exchange. A cell culture was either picked using a sterile tip from a bacterial plate, or taken from a glycerol stock for inoculation. The cell cultures were then incubated over night at 37°C, shaking at 250 rpm.

4.2.4.2 Transformation of electro-competent cells

The electroporation cuvette was placed on ice, while the electrocompetent cells were allowed to thaw slowly on ice. The plasmid preparation to be transformed was also placed on ice. Meanwhile 1 ml of SOC medium was warmed in a round bottom Falcon tube to 37°C. 1 µl of plasmid DNA was mixed gently with 50 µl cells and incubated on ice for 1 min. The cells were transfered to the pre-cooled cuvette, and electroporated (2.5 kV, 3ms). Immediately afterwards the cells were transferred to the SOC medium and incubated for 1 hour shaking at 37°C. The cells were harvested after centrifugation at 4000 x g for 2 min and spread on a bacterial plate containing the relevant antibiotics for selection.

4.2.4.3 Transformation of heat competent cells

Transformation of plasmid DNA into heat competent *E.coli* TOP10 cells was performed by mixing 50 µl of competent cells with 1 µl of purified plasmid DNA for 5 minutes on ice. Cells were incubated at 42°C for 30 seconds and then immediately chilled on ice for 2 minutes. 250 µl of fresh LB medium was added, followed by incubation at 37°C for 1 hour in a thermo shaker. Cells were plated on LB-agar plates containing the respective antibiotics and incubated at 37°C overnight. Plasmid DNA

was isolated from a 3 ml overnight culture using the QIAquick Miniprep Kit. DNA-sequencing of the clones after the BP reaction was performed by GATC-Biotech, Konstanz.

4.2.5 Protein biochemical methods

4.2.5.1 Protein expression

An overnight bacterial cell culture was grown in 60 ml LB medium containing 100 µg/ml carbenicillin and 34 µg/ml chloramphenicol. Ten ml of this culture were added to 2 l Erlenmeyer flasks containing 1 l LB medium with 100 µg/ml carbenicillin and 34 µg/ml chloramphenicol. Six flasks were prepared in this manner, incubated at 37°C and shaken at 250 rpm until the optical density of the cells reached 0.6 at a wavelength of 600 nm against a blank of LB medium. At this point the temperature was decreased to 16°C and gene expression was induced by the addition of 0.2 µg/ml anhydrotetracyclin. Optimal expression of Pol η was achieved after 4-5 hours. The cells were then harvested by centrifugation at 8000 rpm for 12 min at 4°C and stored as pellets at -20°C until purification.

Before and after induction 100 µl samples were taken from the culture. The cells were pelleted and redissolved in 100 µl SDS sampling buffer after discarding the supernatant. The sample was then heated at 85°C for 10 minutes and afterwards stored at -20°C.

4.2.5.2 Cell rupture

The pellet of cells having expressed Pol η was redissolved in cold *Strep-tag elution buffer A* together with an added tablette of *Protease Inhibitor Complete* (without EDTA). For rupture the cells were passed through a high pressure cell homogenizer at 10000 psi 3-4 times. The suspension was then transferred into Oak Ridge centrifuge tubes and centrifuged at 18000 rpm for 30 min at 4°C in a Sorvall SS-34 rotor for removal of insoluble proteins and cell debris. After harvesting the supernatant this step was repeated and finally the supernatant was filtered through sterile 0.2 µm filters before being subjected to further protein purification on an Äkta FPLC system.

4.2.5.3 Protein purification

The protein solution was loaded onto a *Strep*-Tactin Superflow column. Proteins not bound were eluted by washing with 3.5 CV *Strep-tag elution buffer A* (1ml/min). The bound protein was released by eluting with 100% *Strep-tag elution buffer B* (1ml/min). The Pol η containing fractions were pooled, reduced to 500 µl in an *Amincon Ultra centrifuge filter device* at 4000 rpm and the buffer exchanged against *Heparin buffer A* by 5 ml addition of the same to the centrifuge filter and repetitive reduction to 500 µl. The protein was loaded onto a HiTrap™ Heparin HP ion exchange column which was washed with 5 CV *Heparin buffer A* for eluting DNA. Pol η was eluted with a linear gradient over 10 CV from 100% *Heparin buffer A* to 100% *Heparin buffer B*.

After the second step of purification the Pol η containing fractions were pooled and reduced to 500 µl in an *Amincon Ultra centrifuge filter device* at 4000 rpm. 5 ml of *crystallization buffer* were added and the protein was again reduced to 500 µl. This step was repeated twice and the polymerase was concentrated to at least 7 mg/ml. The protein preparation was aliquoted into 15 µl samples, shock frozen with liquid nitrogen and stored at -80°C until further use.

4.2.6 Primer extension assays

The 5' fluorescein labeled primers were annealed to the unlabeled templates containing a site specific lesion and to undamaged templates in the *primer extension buffer* at a molar ratio of 1:1.5 in order to ensure that all labeled primers are hybridized. Primer extension studies were carried out, unless otherwise specified, by incubating Pol η (0.25 µM) with 1 µM primer-template substrate and dNTP (200 µM) in the *primer extension buffer*. The reaction mixtures (30 µl) were incubated at 26 °C and the reaction was stopped by adding 50 mM EDTA pH 8.0. Quenched samples were heated to 95 °C for 3 min and analyzed by electrophoresis on denaturizing 22.5 % acrylamide gels, containing 7.5 M urea. Resolved extension products were detected by using a LAS-3000 imaging system.

4.2.7 Crystallization

4.2.7.1 DNA screening

Prior to the crystallization setup, the ternary complex was allowed to form by incubating dsDNA (0.24 mM) with ddCTP (1 mM), $MgCl_2$ (5 mM) and protein (6 mg/ml) at room temperature for 10 minutes. 13-22 base pair template DNA strands with complementary primer strands (Table 4-7) yielding sticky ends of 5-8 bases with or without a 1 base overlap at the 3' template strands were screened for cocrystallization with Pol η utilizing commercial NeXtal DWBlock Suite sparse matrix screens (Qiagen) in CrystalEX 96 well sitting drop plates. Setups were pipeted by a Hydra II semi-automatic protein crystallization robot (0.2 µl protein + 0.2 µl precipitant).

After obtaining an initial hit with *eta E* and the complementary primer *eta E1*, DNA strands containing 7-10 bp double strands with various lengths of overhangs were tested for yielding even better crystals (Table 4-8).

4.2.7.2 Crystallization of Pol η with lesion containing DNA and ddCTP

The first tetragonal crystals of the Pol η – dsDNA - ddCTP complex were obtained from solutions containing 20% PEG 3350 and 200 mM calcium chloride, at 18°C by vapour-diffusion using the sitting drop method. Prior to the crystallization setup, the ternary complex was allowed to form by incubating dsDNA (eta Ep1, Table 4-10) (2.5 equivalents) with ddNTP (1 mM), $MgCl_2$ (5 mM) and protein (6 mg/ml) at room temperature for 10 minutes.To improve crystal quality the conditions of the initial hit were refined by varying the protein concentration, concentrations of the components of the crystallization buffer, pH, temperature, drop size, mixing ratio between drop and precipitant and reservoir volume in 24 well formats either in sitting drop VDX™ plates or hanging drop Cryschem plates.

The biggest crystals appeared overnight in wells containing 2 µl droplets (1 µl of protein-DNA complex with 1 µl reservoir solution) and 100 µl of precipitant solution (160 mM calcium chloride and 16% (w/v) PEG 3350) and achieved maximum size after about 2-3 days using the hanging drop vapour-diffusion method at 18ºC.

Repetitive macro seeding was required to produce diffraction quality crystals. To do so, a crystal was fished out of the well, washed twice in precipitant solution for removing any potential present microcrystals and placed in a new 2 µl droplet. This procedure

was repeated at least 3 times, until the crystal reached its final size of approximate dimensions of 0.2 × 0.1 × 0.1 mm.

4.2.7.3 Crystallization of Pol η with lesion containing DNA and dNTP

Crystals of the Pol η – dsDNA - dNTP complexes were obtained by manually mixing 1 µl protein solution and 1 µl precipitant, setting up hanging drops and performing macro seeding as described in 4.2.7.2. Prior to the crystallization setup, the ternary complex was allowed to form by incubating dsDNA (2 equivalents) with dNTP (1 mM), $MgCl_2$ (5 mM) and protein (4.5 mg/ml) at room temperature for 10 minutes. The dsDNA used was a 2',3'-dideoxy primer strand (4.2.2) hybridized in an equimolar ratio to a site specific lesion containing template strand.
Cryocooling was achieved by soaking the crystals for 5-10 seconds in reservoir solution containing 15% D(-)-2,3-butanediol and flash-freezing in liquid nitrogen.

4.2.8 Data collection and processing

Redundant single wavelength anomalous dispersion (SAD) data at the platinum L_{III} absorption edge wavelength were collected at beamlines PX-1 at SLS (Swiss Light Source, Villingen, Switzerland), ID23-2 at ESRF (European Synchrotron Radiation Facility, Grenoble, France) and PX-14.1 at BESSY (Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung). Best wavelengths for the peak were determined by fluorescence scan on each crystal.

4.2.9 Structure solution and refinement

Data were processed with the XDS suite[191]. Initial phases were determined by molecular replacement method with the program Phaser[192] using the coordinates of the Pol η apoenzyme structure (1JIH)[129] as a search model. The topology and parameter data of the cisplatin lesion for refinement were generated from coordinates 1CKT[193] using XPLO2D[194]. All manual model building steps were performed with MAIN[195] and COOT[196]. After corrections for bulk solvent and anisotropic overall B-values, the model data was refined by iterative cycles of simulated annealing, positional refinement and individual B-factor refinement with the program CNS[197]. Initial non crystallographic symmetry (NCS) restraints were gradually removed in the final cycles

of the refinement, to allow some structural variations. Data collection and model statistics are summarized in Tables 7.1 and 7.2. PROCHECK[198] revealed no disallowed angles and good stereochemistry for the final model which comprises 509 residues.

4.3 Bioinformatic Methods

4.3.1 Homology searches and alignments

Protein sequences were retrieved from the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=protein). Multiple sequence alignments were performed with ClustalW (http://align.genome.jp/) and edited manually for structure based sequence alignments.

4.3.2 Structure visualization and analyzing

Least-squares superpositioning of polymerases were performed with LSQMAN[199], using the Cα atoms of the Palm residues for search models. Figures for visualization of the crystal structures were prepared with PyMol[200]. Difference distance matrix plots were produce using the DDMP program from the Center for Structural Biology at Yale University, New Haven, CT.

4.3.3 Crystallographic software

XDS[191]
Phaser[192]
XPLO2D[194]
MAIN[195]
COOT[196]
CNS[197]
PROCHECK[198]
LSQMAN[199]
PYMOL[200]

5 Results

5.1 Preparation and purification of platin lesion containing DNA strands

The DNA templates *DPO4_GG*, *DPO4_GTG* and *DPO4_CG* (Table 4-9) were prepared for primer extension experiments, whereas the DNA templates *eta Ep1* and *eta Ep2* (Table 4-10) were prepared for cocrystallizations with Pol η. The DNA templates *Pol2_GG*, *Pol2_CGG* and *Pol2_IG* (Table 4-10) were prepared for a collaboration with the group of Prof. Cramer. All strands were prepared according to the procedures described in 4.2.1. When platinating a commercial strand (Es1, Figure 5-1), besides the sought after product, other side products such as strands with 2 or more bound platinums are obtained (crude, Figure 5-1).

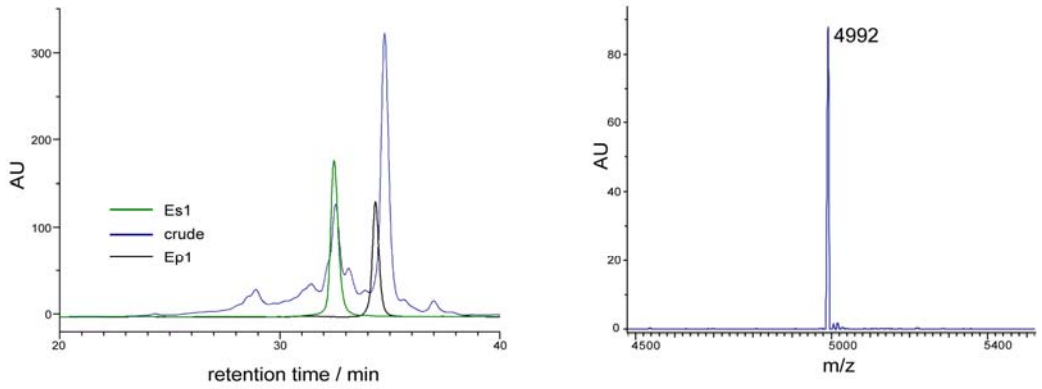


Figure 5-1. HPLC spectra of the unplatinated strand (Es1), unpurified reaction mixture (crude) and the platinated strand (Ep1) are overlayed on the left. On the right the MALDI-TOF spectra of Ep1 is shown (calculated 4997 amu).

After purification of the main product by HPLC the strands were approximately 99% pure and did not contain contaminations of unplatinated strands, which would falsly contribute to primer extensions (Ep1, Figure 5-1). The high purity of the strands is also of importance for crystallization setups, where impurities could inhibit crystal growth.

5.2 DNA synthesis and purification

5.2.1 Preparation and purification of 2',3'-didesoxy primer strands

The complementary primer strands of *eta Ep1/2* and *eta Ett 1/2/3* (Table 4-10) were synthesized with 2',3'-dideoxybases at the 3'end using 5'-CE-phosphoramidites. The strands were subsequently purified by HPLC as described in 4.2.2 (Figure 5-2).

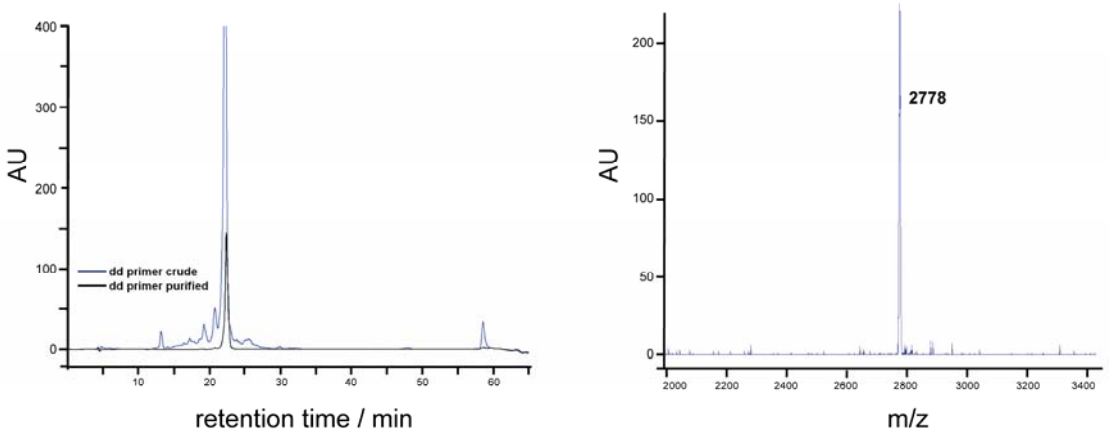


Figure 5-2. HPLC spectra of the unpurified synthesis product (crude) and the purified strand (purified) are overlayed on the left. On the right the MALDI-TOF spectra of the purified strand is shown (calculated 2779 amu).

By utilizing these primers for crystallization, the nucleotidyl transfer is inhibited due to the lack of the primers 3'-OH group. Thus, the Pol η – DNA – dNTP complex can be crystallized with the polymerase trapped at a defined position relative to the lesion.

5.2.2 Preparation and purification of TT dimer lesion containing DNA strands

The DNA templates *eta Ett1*, *eta Ett2* and *eta Ett3* (Table 4-10) were synthesized for cocrystallizations with Pol η. The chemical analog of the TT dimer, which was used for DNA synthesis, contains a formacetal linkage instead of the intradimer phosphate, which is highly useful in preparing the CPD lesion analog in quantities sufficient for crystallization studies. The strands were prepared according to the procedures described in 4.2.1. and after HPLC purification a purity of approximately 99% was obtained.

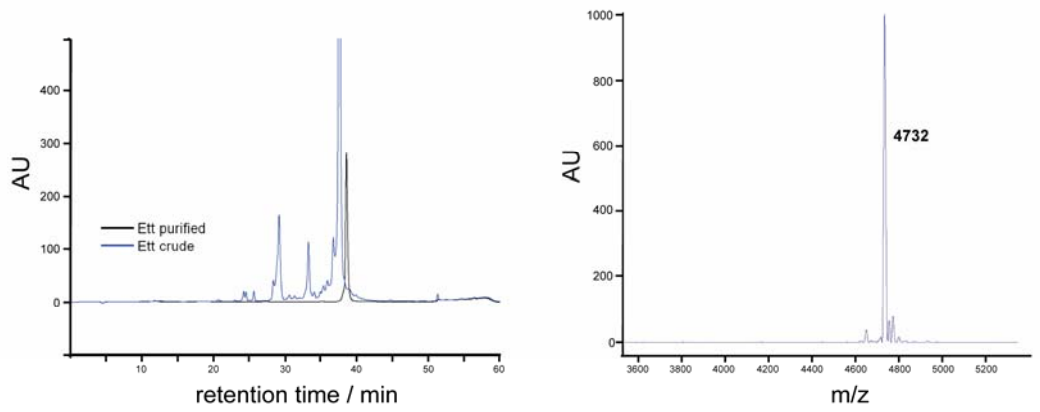


Figure 5-3. HPLC spectra of the unpurified synthesis product (crude) and the purified strand (purified) are overlayed on the left. On the right the MALDI-TOF spectra of the purified strand is shown (calculated 4734 amu).

5.3 Primer extensions of 1, 3 GTG oxaliplatin adducts

Template strand *DPO4_GTG* (Table 4-9) was prepared as described in 4.2.1.2 and purified by HPLC as described in 4.2.2. The template strand was annealed to the 5' fluorescein primer strands *GTG_comp 1/2/3* in the reaction buffer in a 1:1.5 ratio in order to ensure that all primer strands anneal. In order to evaluate the capability of Pol η to copy through a DNA strand containing a 1,3 GTG oxaliplatin adduct, primer extension experiments were performed as described in 4.2.6 with the following alterations: Increasing concentrations of Pol η were incubated for 20 min with dNTPs (100 μM) at 25 C°.

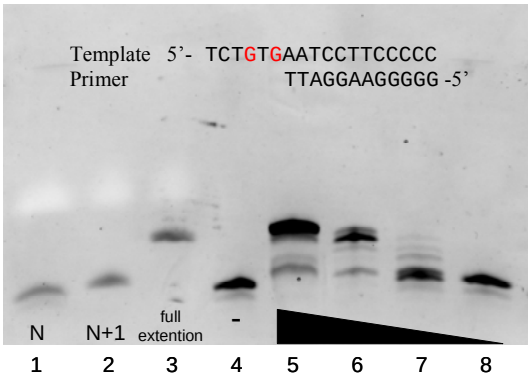
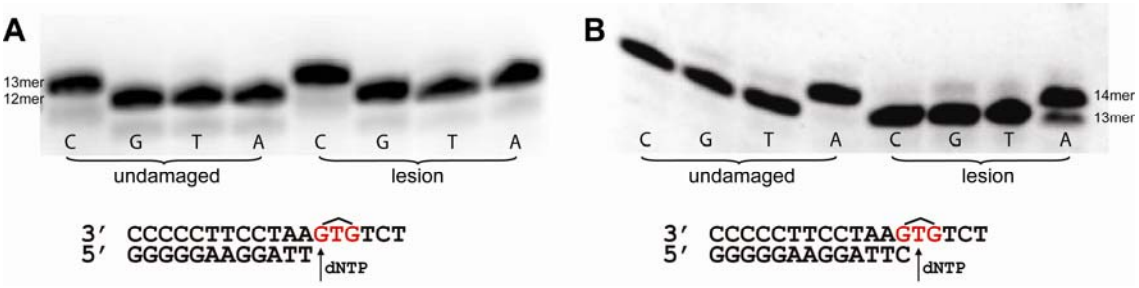


Figure 5-4. Concentration dependent primer extension studies with Pol η and DNA containing a site specific oxaliplatin adduct. In lane 1 the unextended primer (N), lane 2 the first extension (N+1), lane 3 the full extension product and in lanes 5 – 8 the reaction products obtained with decreasing concentrations of Pol η ($112x10^3$/$112x10^2$/1120/112 nM).

Once it was shown that Pol η can perform a full extension of *GTG_comp 1* (Figure 5-4), the specificity of incorporation opposite the various damaged bases was evaluated (Figure 5-5). *GTG_comp 1* was used for studying the incorporation opposite the lesions 3'dG, *GTG_comp 2* for the dT flanked by the two dGs and *GTG_comp 3* for the 5'dG. The primer extension studies were performed as described in 4.2.6 with the following alterations: Pol η (120 nM) was incubated for 15 min with 1 µM primer-template substrate and dNTP (100 µM) in the *primer extension buffer* at 35 C°.



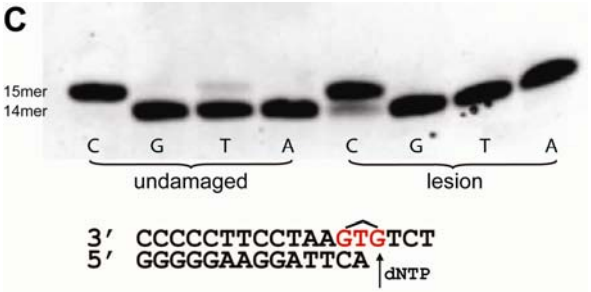


Figure 5-5. Primer extension studies with *DPO4_GTG* and primers terminating at the different positions opposite the lesion. The specificity of dNTP incorporation opposite the 3' **(A)**, T **(B)** and 5' **(C)** bases of the lesion was probed.

Primer extension studies with *DPO4_GTG* and *GTG_comp 1* revealed that Pol η inserts opposite the 3'dG mainly a dCTP in the absence of the lesion (Figure 5-5 A). Opposite a lesioned base as well mainly a dCTP was inserted.

Opposite the dT between the dGs, Pol η mainly inserts a dATP and to a much lesser extent dTTP and dGTP in the absence of the lesion. Mainly a dATP but also dGTP and dTTP were inserted in the case of a lesion present (Figure 5-5 B).

For the 5'dG mainly a dCTP and to a lesser extend a dTTP insertion in the absence of the lesion was observed. Opposite the lesioned base mainly a dCTP was inserted (Figure 5-5 C).

5.4 Crystallization of DNA Pol η in ternary structure

5.4.1 Preliminary screenings

The Pol η apoenzyme structure was solved in 2001[129], but no structure of this polymerase in ternary complex with the substrates was obtained subsequently.

We rationalized that Pol η, being a TLS polymerase, might be easier to be trapped in ternary complex by using a lesion containing DNA for cocrystallization experiments. Therefore we screened for crystal growing conditions as described in 4.2.7.2 using the cisplatin lesion containing DNA strand *DPO4_GG* (Table 4-9). The first microcrystals were obtained in conditions containing 0.2 M magnesium acetate and 20 % Peg 3350 w/v and were further optimized for obtaining diffracting crystals (Figure 5-6).

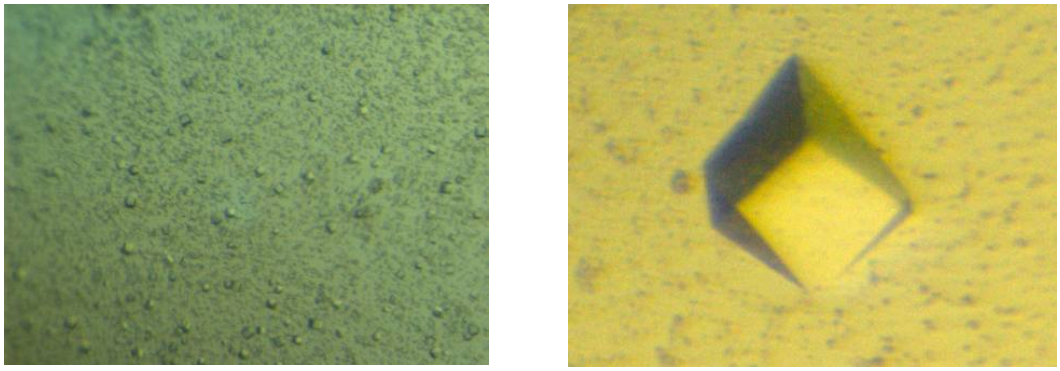


Figure 5-6. Microcrystals of Pol η with cisplatin lesion containing template – primer DNA strand *DPO4_GG*, which grew in the initial hit (left) and a crystal grown after refining the initial conditions (right).

The obtained crystals defracted to a limiting resolution of 2.4 Å, but after solving and analyzing the structure we realized that we did not obatin a structure of the ternary complex, since the polymerase did not contain DNA (Figure 5-7).

Next we decided to screen different DNAs, assuming that another factor for obtaining crystals of the ternary complex might be the length and/or sequence of the DNA to be crystallized. In order to test many different oligomers for cocrystallization rapidly, we decided to use undamaged oligomers for the initial screens (Table 4-7).

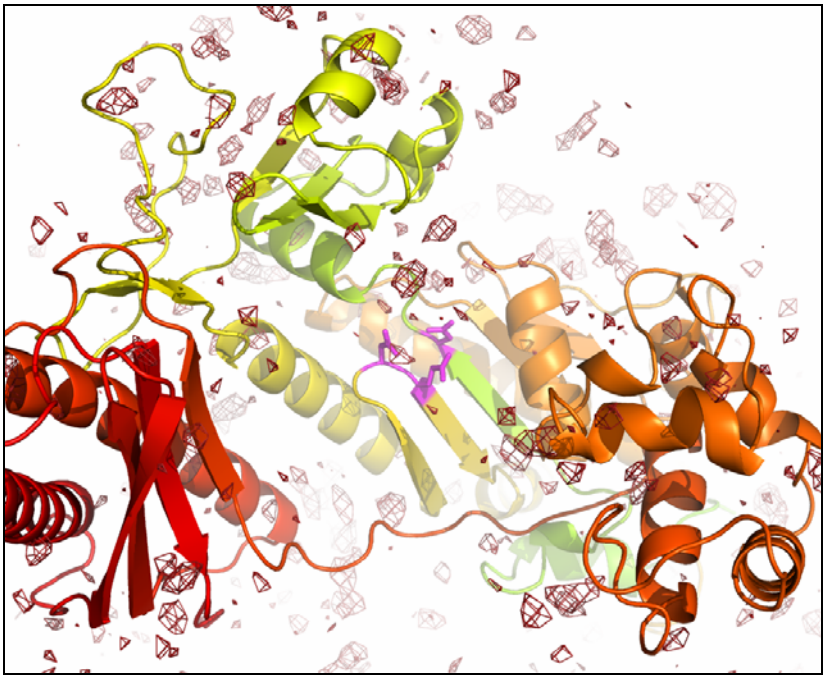


Figure 5-7. The F_o-F_c difference electron density shown in red after the initial molecular replacement with the Pol η apoenzyme[129], which was used as the model for replacement, depicted as a cartoon. The catalytic center is shown in magenta sticks for ease of orientation.

After screening different oligomers for cocrystallization, a 16 bp template (*eta E*) annealed to a 9 bp primer (*eta E1*) resulting in a 9 bp double strand with a sticky end of 7 bases without an overlap at the template 3' terminus led to a visually new crystal form with 0.2 M $CaCl_2$ and 20 % w/v PEG 3350. These conditions were subsequently refined, however further optimization of the DNA length or sequence (Table 4-8) did not result in better crystals.

5.4.2 Crystallizations with lesion containing DNA and ddCTP

With a DNA sequence at hand which crystallized, a cisplatin lesioned oligonucleotide with the same sequence was prepared as described in 4.2.1.1 and crystallization setups were performed.

The crystals of the Pol η – dsDNA - ddCTP complex were obtained overnight at 18 °C using the hanging-drop vapour diffusion technique (Figure 5-8 A). Macroseeding had to be performed at least 3 times in order to obtain larger crystals, suitable for diffraction (Figure 5-8 B).

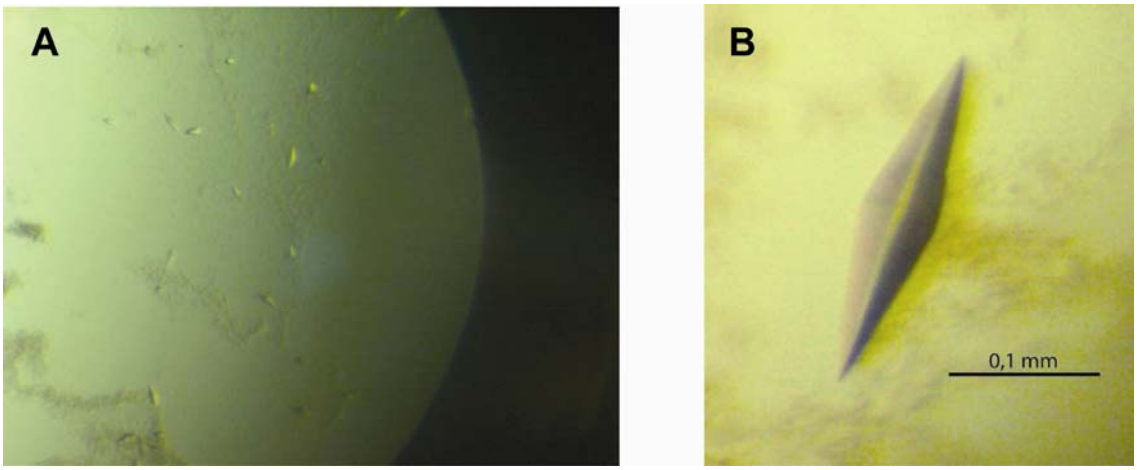


Figure 5-8. **A:** Photograph of a Pol η – dsDNA - ddCTP crystal illuminated with polarized light, which was obtained in reservoir solutions containing 200 mM $CaCl_2$ and 20 % (w/v) PEG 3350. **B:** The crystal in this photograph was macro seeded at 291 K in a crystallization solution consisting of 170 mM $CaCl_2$ and 16% (w/v) PEG 3350 three times to obtain a final size of 50 x 50 x 200 µm.

Cryocooling was achieved by soaking the crystals for 5-10 seconds in reservoir solution containing 15 % D(-)-2,3-butanediol and flash-freezing in liquid nitrogen. Crystals grown under these conditions diffracted to a limiting resolution of 3.3 Å and belong to space group $P4_12_12$ with two protein-DNA complexes per asymmetric unit.

5.4.3 Crystallizations with lesion containing DNA and dNTPs

The first obtained structure showed the ddCTP positioned in a loosely bound conformation in the active site, hydrogen bonded to the templating base. Realizing the importance of the nucleotides 3'-hydroxy group for correct positioning in the active site, which in turn positions the DNA correctly inside the polymerase we crystallized the complex once more, but with a dNTP in stead of the ddNTP. To prevent nucleotidyl transfer, primer strands terminating at the 3'-end with a 2',3' dideoxy ribose were prepared by reverse DNA synthesis as described in 4.2.2.

In order to obtain crystal structures of the *3'dG* and the *5'dG elongation* steps, the register of the lesion in the template strands was adjusted so that the crystallization occurs in each case with a 16 bp template containing a site specific cisplatin lesion annealed to a 9mer primer (*eta Ep1* and *eta Ep2* respectively, Figure 5-9). The *3'dG elongation* step was crystallized with a dCTP, whereas for the *5'dG elongation* step only crystals with dATP could be obtained.

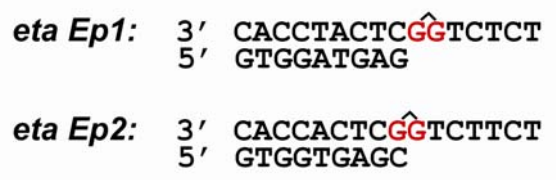


Figure 5-9. DNA constructs for cocrystallizing Pol η in the *3'dG elongation* (eta Ep1) and *5'dG elongation* (eta Ep2) steps.

The ternary complex was allowed to form by incubating 2.5 equivalents *eta Ep1* or *eta Ep2* (Table 4-10) with dNTP (1 mM), $MgCl_2$ (5 mM) and protein (0.1 mM) at room temperature for 10 minutes. Crystals were obtained overnight at 18 °C using the hanging-drop vapour diffusion technique by mixing 1 µl of protein-DNA complex with 1 µl reservoir solution containing 170 mM $CaCl_2$ and 16% (*w/v*) PEG 3350. The resulting crystals were macro seeded at least 3 times. However, repetitive seeding was limited by microcrystals, attached to the core crystal, which as well increased in size and prevented the core crystal to grow bigger. Cryocooling, for avoiding radiation damage during data collection, was achieved as described in 5.4.2. The obtained crystals diffracted typically to 3.1 - 3.3 Å resolution at a synchrotron light source and belong to space group $P4_12_12$. The crystals have two protein-DNA complexes per asymmetric unit, which have unit cell dimensions of $a = b = 104.1$ Å, $c = 293.0$ Å for the *3'dG elongation* (obtained with *eta Ep1*) and $a = b = 103.7$ Å, $c = 292.8$ Å for the *5'dG elongation* (obtained with *eta Ep2*), and $\alpha = \beta = \gamma = 90°$.

5.5 Structure solution and refinement

Data were collected from crystals of the yeast Pol η complex at 100 K. The crystals diffracted to a limiting resolution of 3.1 – 3.3 Å at the SLS and ESRF synchrotron facilities. Data were autoindexed, integrated and scaled with the XDS suite[191] and the phases were determined by molecular replacement method as described in 4.2.9. A picture of the diffraction pattern is shown below (Figure 5-10 A), and the data processing statistics are listed in chapter 7.2.

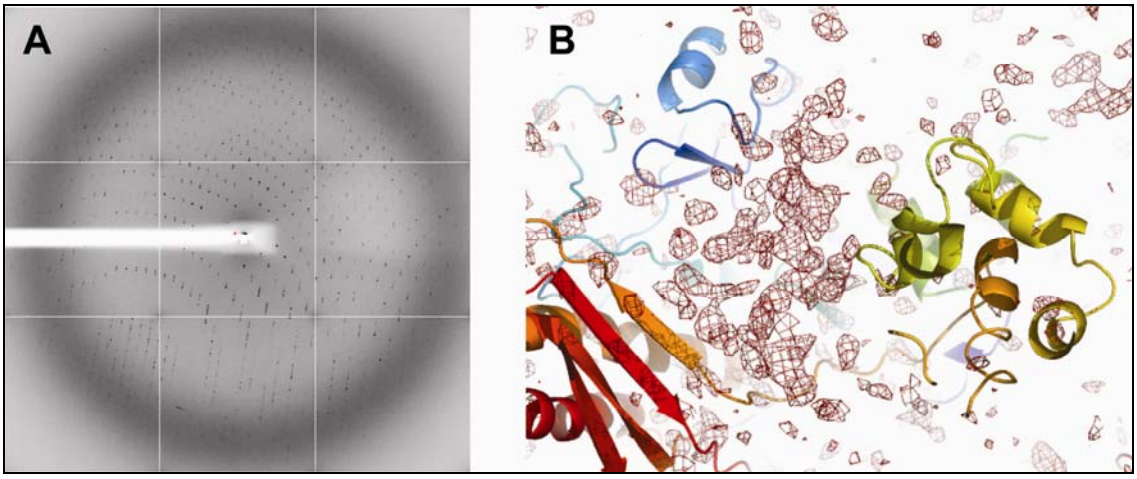


Figure 5-10. A: Diffraction pattern recorded at the Swiss Light Sourse (SLS) synchrotron beamline PX1. **B:** The F_o-F_c difference electron density shown in red after the initial molecular replacement with the Pol η apoenzyme[129], which was used as the model for replacement, depicted as a cartoon.

The initial model built into the density obtained by the replacement (Figure 5-10 B) was rebuilt manually with Coot[196] and refined with CNS[197] iteratively. The final model of the ddCTP structure (4.2.7.2) contains residues 1 – 509 of the polymerase. 28 residues at the *N* terminus, corresponding to the strep-tag and a linker, and 23 residues at the *C* terminus are missing in the electron density map. In the structures obtained with dNTPs (4.2.7.3), two more residues at the *N* terminus are visible in the electron density. The 5 bases of the overhang at the 5' end of the template are missing in the electron density of all structures. For more details of the refined models see chapter 7.2.

5.6 Crystal contacts

Interestingly the Pol η ternary complex crystallized in the same space group as the apoenzyme ($P4_12_12$) with very similar unit cell constants. The exact length of the DNA is crucial for obtaining crystals of the ternary complex as was realized from screeing experiments (4.2.7.1). The bound DNA is making tail to tail contacts with DNA belonging to another molecule (Figure 5-11).

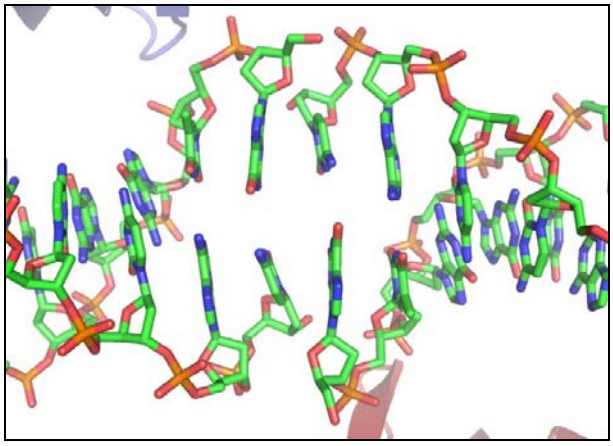


Figure 5-11. Tail to tail DNA-DNA contacts contributing to the crystal packing.

The resulting structure resembles a DNA encaged by 2 polymerases (Figure 5-12). The importance for the ideal length can now be derived from the obtained structures, wherein intermolecular DNA - DNA contacts contribute to the crystal packing.

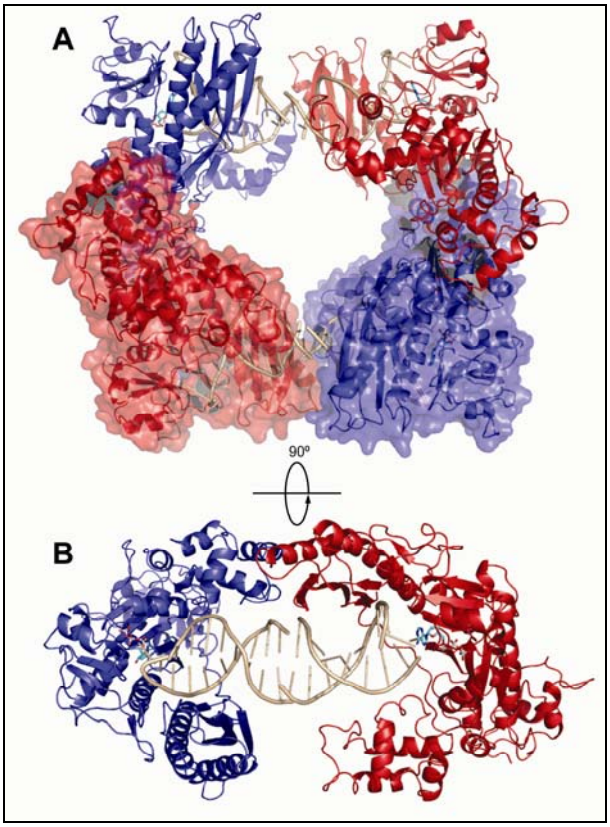


Figure 5-12. A: The two polymerases on the left side belong to the same asymmetric unit and are related through non crystallographic symmetry. **B:** The polymerases shown on top as a surface model are omitted for clarity. The DNA is shown in sand colored cartoon and the incoming dNTP is shown in cyan.

5.7 Crystal structure of the Pol η ternary complex with ddNTP

We realized after solving the structure, comprised of the polymerase, a cisplatin lesion containg template, a natural primer and a 2',3'-didesoxycytidinetriphosphate that the ddCTP is not orientationed correctly in the active site. Although the triphosphate is complexed to the two metal ions, the sugar protrudes into the DNA binding cleft and is not aligned with Phe35 as expected (Figure 5-13). However, the ddCTP still establishes hydrogen bonds to the templating cisplatin lesion containing 3'dG via Watson - Crick base pairing.

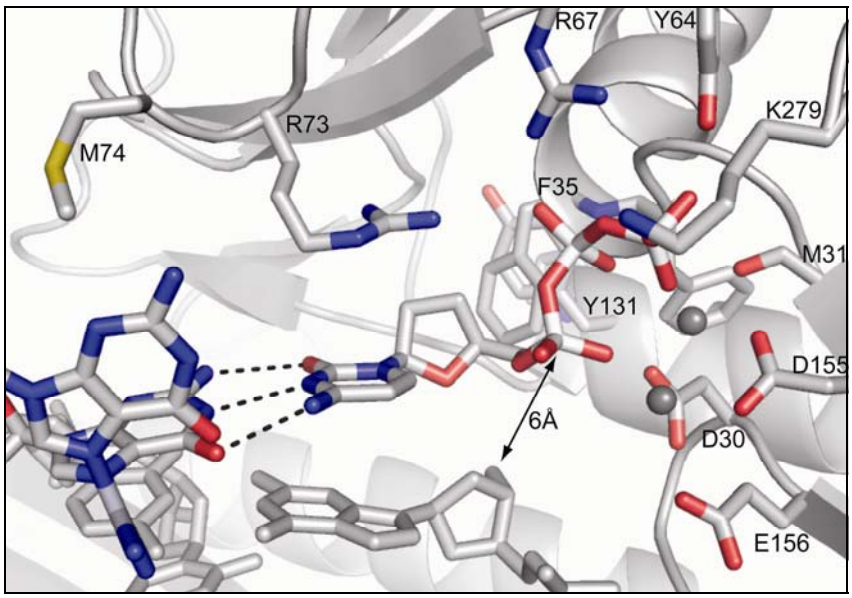


Figure 5-13. Close up view of the catalytic center. The enzyme is depicted as a cartoon with the catalytic residues shown as sticks. The Watson-Crick H-bonded ddCTP and the lesion containing bases are depicted as sticks. The two catalytic ions are shown as grey spheres.

The 3'-OH group of the dCTPs ribose unit would control the positioning of the incoming nucleotide in the wide active site to establish a correct set of H-bonds with the lesion containing base. In the obtained structure the complex is inactive. The primer 3'-OH group is 6 Å away from the dCTP α phosphate and thus unable to perform the nucleophilic attack.

5.8 Crystal structure of the Pol η ternary complex with ddprimer

By crystallizing Pol η with *eta Ep1* (with a dCTP) and *eta Ep2* (with a dATP) we obtained crystal structures of the *3'dG* and the *5'dG* elongation steps opposite a Pt-GG lesion. The overall structure of Pol η resembles a right hand, similar to T7 DNA polymerase with palm, fingers, and thumb but with an additional structural element termed the "polymerase-associated domain," which is a four-β-sheet-structure with two α-helices.

The palm, thumb and PAD form a DNA-binding groove that leads to the polymerase active site located within the palm domain. The orientation of the primer template places the primer 3' end against the dNTP at the confluence of the fingers and palm (Figure 5-14).

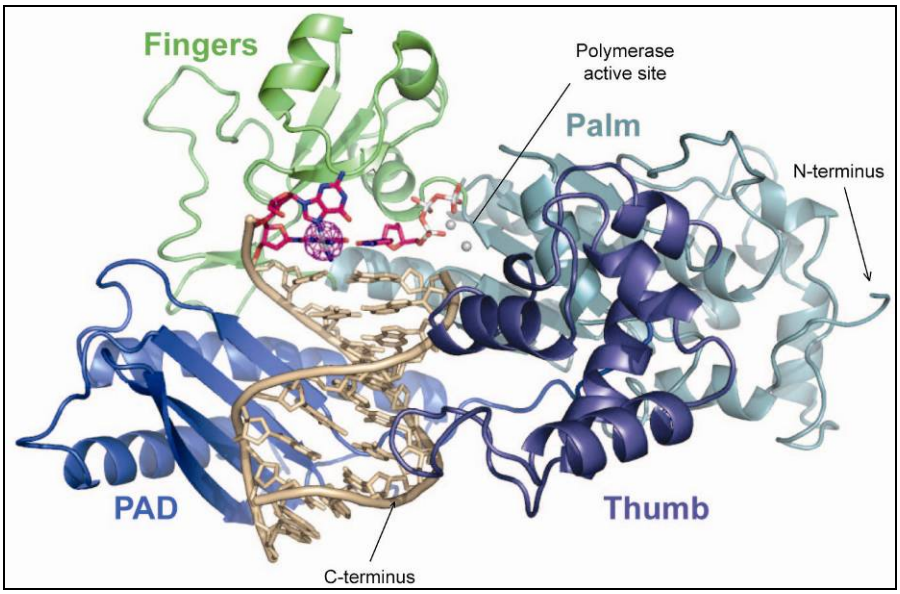


Figure 5-14. Overview of the Pol η structure in ternary complex with lesion containing DNA. The *"3'dG elongation"* complex is shown as a cartoon with the various domains of the protein color coded. In magenta, the cisplatin lesion is shown with the platinum anomalous electron density contoured at 10 σ. The primer and template strands of the DNA (brown) and the Watson-Crick H-bonded dCTP (magenta) are depicted as sticks. The two catalytic magnesium ions are shown as grey spheres.

By crystallizing Pol η together with the DNA template-primer construct *eta Ep1* (Table 4-10) in which the primer ends just before the Pt-GG lesion we obtained 2 structures, derived from the two different molecules of the asymmetric unit.

The first structure shows the Pt-GG lesion in a situation before it enters the active site and wherein the Watson-Crick base pair with the incoming dCTP has not yet been

formed. This structure can be viewed as a "*pre-elongation*" form and will be described in 5.8.2. The second structure contains the Pt-GG lesion partially situated in the active site, with the 3'dG of the lesion forming a Watson-Crick base pair with the incoming dCTP. This complex represents a good model for the relatively efficient bypass of the 3'dG of the lesion and will be described in more detail in 5.8.3.

In order to get information about the replication through the 5'dG of the Pt-GG lesion, we crystallized Pol η together with the template-primer construct *eta Ep2* (Table 4-10) in which the primer was designed to end just in front of the 5'dG base of the lesion. Again crystals were obtained, which provided two independent ternary complexes per asymmetric unit with a limiting resolution of 3.3 Å. In this case, the obtained structures are quite alike and will be described in 5.8.4.

The Pt-GG crosslink is in all complexes well defined in the electron density and the bases forming the Pt-GG lesion are positioned in an appr. 90° angle with respect to each other.

5.8.1 The catalytic center

The catalytic center contains three strictly conserved and functionally important acidic residues consisting of Asp30, Asp155 and Glu156, which hold two metals in the active site. In the active polymerase these would be two Mg^{+2} cations, but in our structures the metals are presumably Ca^{+2} cations originating from the crystallization conditions. These metals in turn ligate the phosphates of the bound dCTP (Figure 5-15). The fingers are making numerous interactions with the bound nucleotide and in addition wall off this end of the DNA-binding groove. The DNA primer 3' end is located in the joint between palm and fingers. The dCTPs 3' hydroxyl group is interacting with the peptide backbone of Phe35 whilst the dCTPs' ribose unit is interacting with the phenyl sidechain of Phe35. This residue is prohibiting ribonucleoside triphosphates from entering the active site because of unfavorable steric interaction with the 2'-OH and is termed the "steric gate"[201]. The catalytic residue Arg73, which is highly conserved in Pol η, interacts through electrostatic attraction of the side chain nitrogens with unesterified oxygens of the α and β phosphates and the esterified oxygen linking between them. In addition, the guanidinium group stacks on top of the base moiety of the incoming dNTP to establish a cation-π interaction [202].

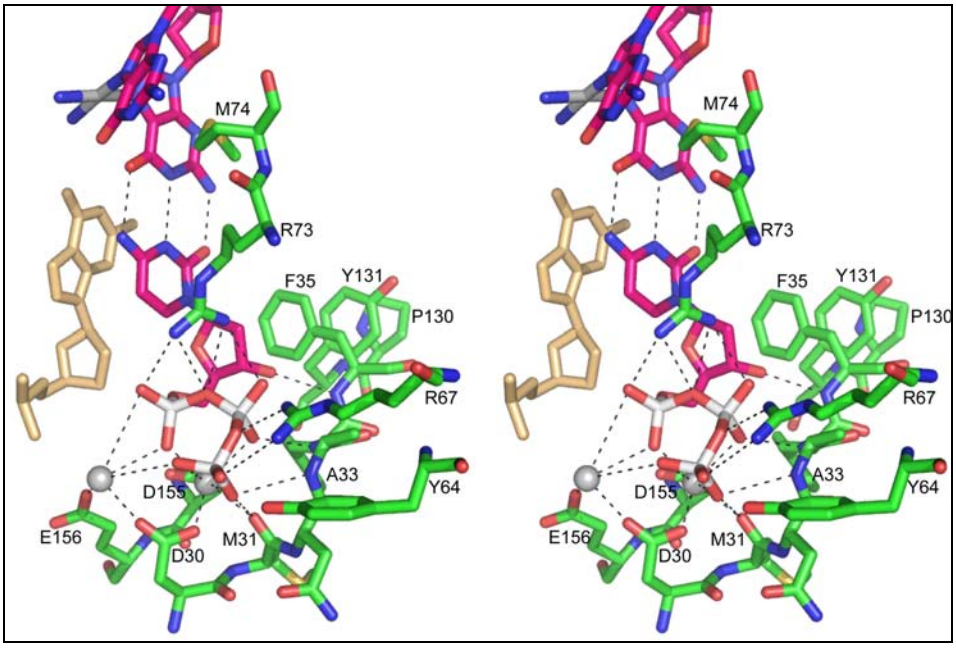


Figure 5-15. Detailed stereo view of the interaction of the dCTP with the surrounding at the 3'dG elongation status of the protein-DNA complex. Important is the H-bond between the 3'-OH and the protein backbone, which orients the dNTP in the binding pocket.

All these interactions contribute to the correct positioning of the incoming dNTP in the active site, positioning it in anticipation of the subsequent 3'-OH nucleophilic attack.

5.8.2 The 3'dG pre-elongation step

In the pre-elongation complex, the Pt-GG lesion is positioned outside the active site and the 3'dG of the lesion is held at a rather large distance from the dCTP of about 6 Å (3'dG-O6•••4NH_2-dCTP). Still, the dCTP is already positioned in the active site even in the absence of any templating base and complexed by two metal ions. These in turn are coordinated by the catalytically essential residues Asp30, Asp155 and Glu156 (Figure 5-16 A). In this state, the expected 3'-OH group (absent in our structure due to the dideoxy terminus) of the primer is located about 8.5 Å away from the α-phosphate of the dCTP and hence unable to perform the primer elongation. The primer - template is held in position on one side of the DNA binding cleft by interactions of the fingers and the PAD, over a strech of 4 bases, with the templating strand. These interactions are 2 bases away from the templating 3'dG into the 3' direction. On the other side of the DNA binding cleft, the thumb is making interactions primarily to the phosphodiester backbone of two bases, which are located 3 bases away from the incoming dCTP (Figure 5-16 B).

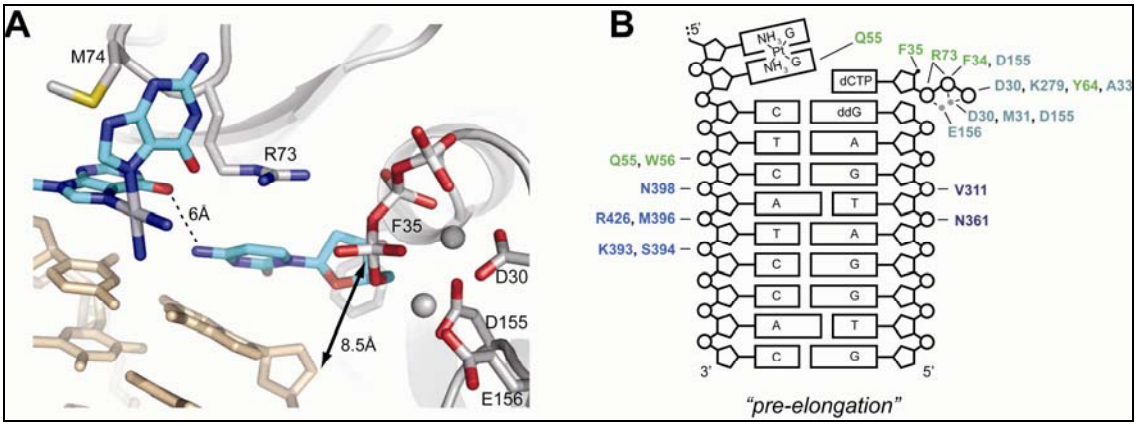


Figure 5-16. A: Active site of the *"pre-elongation"* complex. The catalytic residues together with the Ca^{+2} ions tightly bind the dCTP. Phe35 is placed upon the deoxyribose of the dNTP. The backbone carbonyl of Phe35 establishes a H-bond to the 3'-OH of the dNTP. Arg73 orients the nucleotide, which is still unpaired for H-bonding based recognition. **B:** Schematics of protein-DNA contacts. The primer-template DNA is represented as a ladder, with the bases represented as rectangles, and the phosphates as circles. Direct hydrogen-bond contacts are indicated by solid lines between the amino acid name (coloured as in Figure 5-14) and a circle for phosphate contacts. DNA contacts with the symmetry-related molecule are not shown.

5.8.3 The 3'dG elongation step

The 3'dG of the lesion establishes a perfect Watson-Crick base pair with the pre-orientated dCTP (Figure 5-17 A). The side chain of Met74 of yeast Pol η has slipped into the π-surface notch (S-π distances of 3.8 Å) created by the two dG bases of the Pt-GG lesion to create attractive S-π interactions[202]. Arg73 coordinates the dCTPs α- and β-phosphate and stacks with its base moiety. The distance between the putative 3'-OH nucleophile of the primer and the α-phosphate of the dCTP are reduced to only about 5.0 Å, placing the dCTP ready for nucleotidyl transfer. Correct positioning of the dCTP for the 3'dG translesion synthesis step is guided by a H-bond between the 3'-OH group of dCTP and the backbone of Pol η at Phe35, as well as by hydrophobic interactions of the dCTPs' ribose subunit with the phenyl ring of Phe35. The primer - template is held in position on one side of the DNA binding cleft by interactions of the fingers, palm and the PAD, over a strech of 3 bases, with the templating strand. These interactions are one base away from the templating 3'dG. On the other side of the DNA binding cleft, the thumb is making interactions primarily to the phosphodiester backbone of two

bases, which are located one base away from the incoming dCTP. The PAD is making a single interaction with the primer, one base from the 5' end (Figure 5-17 B).

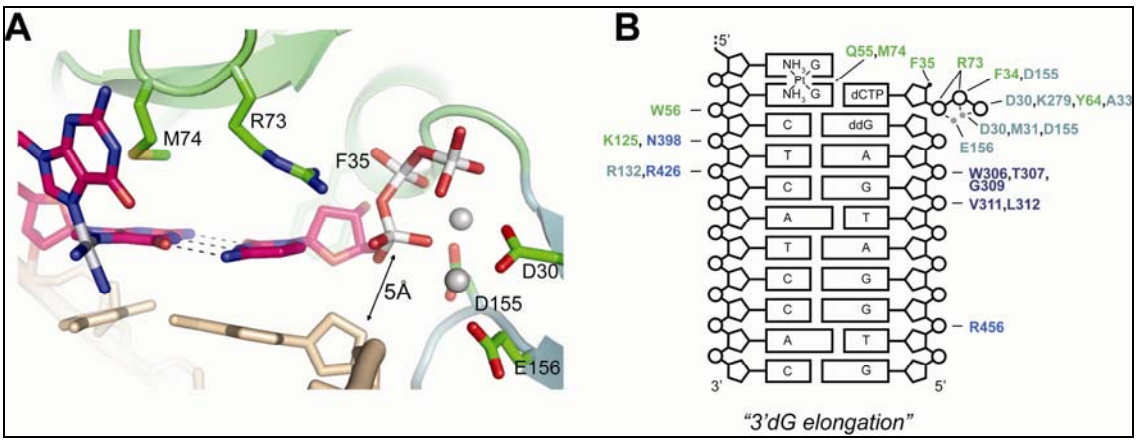


Figure 5-17. A: The catalytic residues together with the Ca^{2+} ions tightly bind the dCTP. R73 orients the nucleotide for H-bonding based recognition. The dCTP pairs perfectly with the 3'desoxyguanine of the lesion. Met74 stacks in the hydrophobic π-notch between the two cisplatin complexing guanines. **B:** Schematics of protein-DNA contacts. The primer-template DNA is represented as a ladder, with the bases represented as rectangles, and the phosphates as circles. Direct hydrogen-bond contacts are indicated by solid lines between the amino acid name (coloured as in Figure 5-14) and a circle for phosphate contacts. DNA contacts with the symmetry-related molecule are not shown.

5.8.4 The 5'dG elongation step

The DNA duplex and the position of the putative 3'-OH of the primer, are arranged in an intermediary conformation compared to the "pre-elongation" and "3'dG elongation" complexes described before (Figures 5-16 and 5-17). The 3'dG of the lesion remains stably bound via a Watson-Crick base pair to the primer dC base. The incoming dATP stacks on top of the newly formed base pair. The 5'dG of the lesion has moved into the active site but the crosslinking Pt-bond to the 3'-dG forces this base to stay perpendicular relative to the incoming dATP and to the 3'dG-dC base pair as shown in Figure 5-18. Only one H-bond is established between the exocyclic C(6)-NH_2 amino group of dATP and the C(6)=O carbonyl oxygen of the 5'dG. In this structure, the expected position of the primer 3'-OH is with 7.5 Å distance to the dATP larger as in the first elongation complex (5 Å) providing a rational for the slower second bypass step.

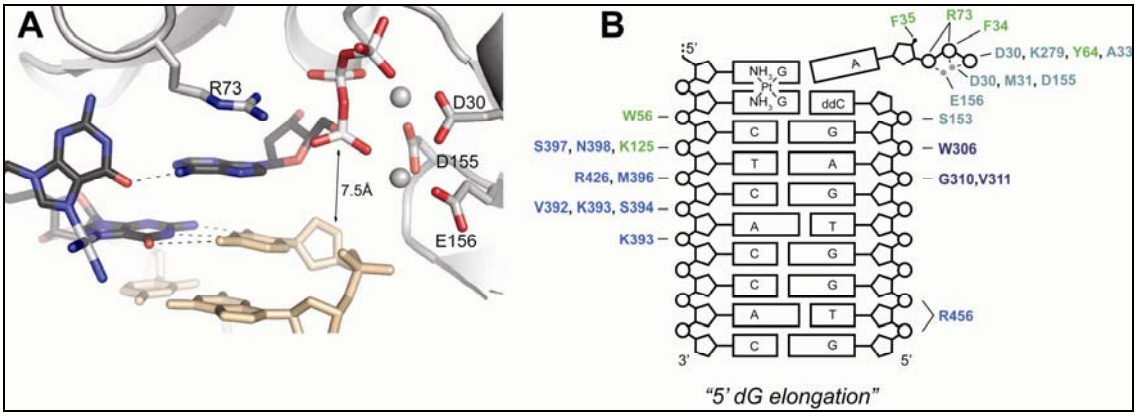


Figure 5-18. A: Active site of the *"5'dG elongation"* complex of Pol η. The catalytic residues together with the Ca^{2+} ions tightly bind the dATP. The nucleotide, ready for H-bonding based recognition, pairs with the lesions 5'desoxyguanine through only one hydrogen bond. **B:** Schematics of protein-DNA contacts. The primer-template DNA is represented as a ladder, with the bases represented as rectangles, and the phosphates as circles. Direct hydrogen-bond contacts are indicated by solid lines between the amino acid name (coloured as in Figure 5-14) and a circle for phosphate contacts. DNA contacts with the symmetry-related molecule are not shown.

The primer - template is held in position on one side of the DNA binding cleft by interactions of the fingers and the PAD, over a strech of 5 bases, with the templating strand. On the other side of the DNA binding cleft, the palm is interacting with the phosphate backbone 5' of the incoming dCTP. The thumb is making interactions to the following 2 bases and also the PAD contacts the primers' phosphate backbone at the last two bases (Figure 5-18 B).

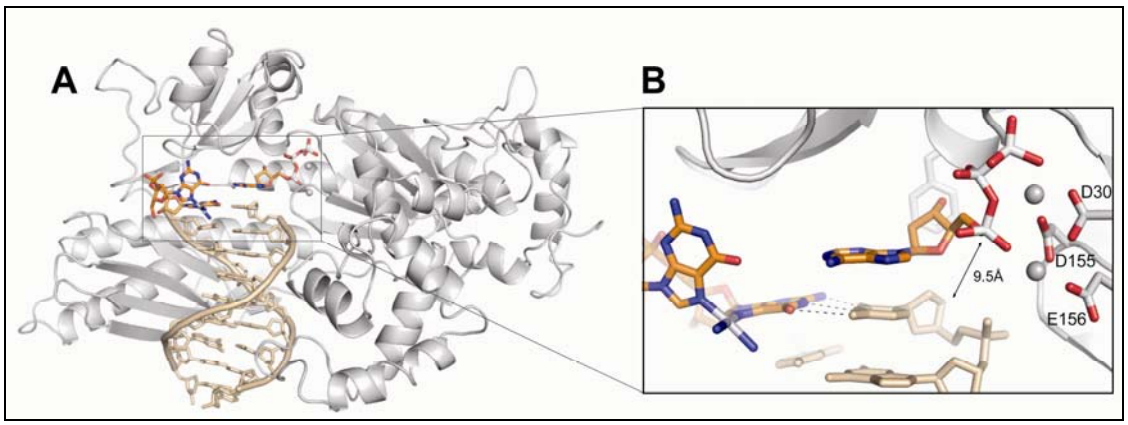


Figure 5-19. A: Overview of complex 2 of the 5'dG elongation step. **B:** The catalytic residues together with the Ca^{2+} ions bind and preorient the dATP in the active site. However, there is a large distance between the α-phosphate of the dNTP and the putative primer OH in the absence of the H-bond between the dNTP and the 5'desoxyguanine.

In complex 2 of this step, the 5'dG is with a 9.5 Å distance to the α-phosphate even larger and no hydrogen bond is formed to the dATP (Figure 5-19). Thus, the single H-bond formed between the two perpendicularly oriented heterocycles pulls the primer strand and the dATP together rationalizing the "slow" nucleophilic attack of the α-phosphate for primer extension. Such a H-bond is besides with dATP only possible with dCTP via the C(4)-NH_2 amino group, explaining why these two nucleotides are preferentially inserted opposite the 5'dG by the yeast Pol η during the second step of lesion bypass (5.9.1).

5.8.5 The polymerase-associated domain

The DNA binding surface area enclosed by the palm, fingers, and thumb domains in Pol η is with 675 $Å^2$ substantially less than in T7 Pol with 1630 $Å^2$. The size of the Pol η hand is augmented by an extra domain, the PAD consisting of residues 393–508. The PAD is joined to the thumb through a flexible linker that traverses the DNA binding groove from the thumb to the fingers side and has resemblance to the palm in containing a mixed β sheet and two long α helices. The two β sheets are roughly perpendicular to each other, and are the principal elements defining the floor and the wall of the DNA binding cleft (Figure 5-20).

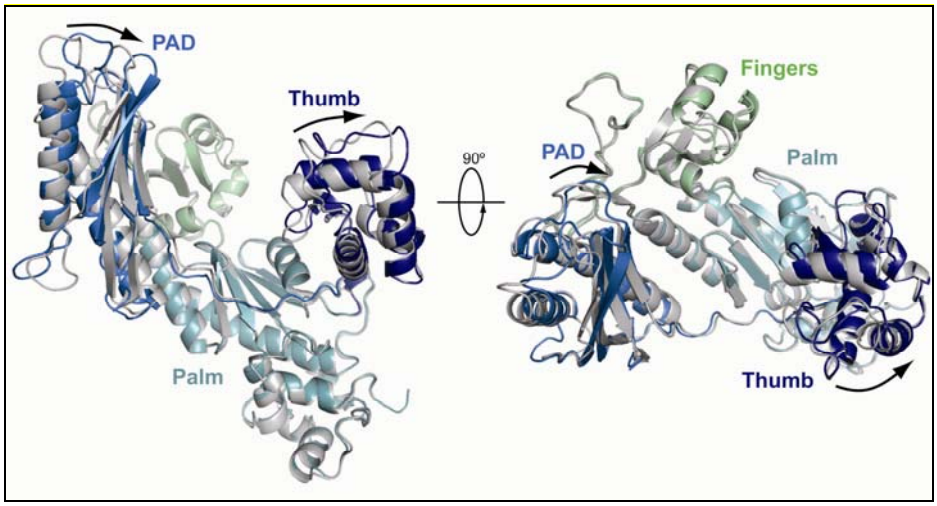


Figure 5-20. Overlay of the *3'dG elongation* (colour coded as in Figure 5-14) and the the *3'dG pre-elongation* (grey) complexes depicted as comics. The two views are related by a 90° rotation around a horizontal axis. The bound DNA is omitted for clarity.

Most importantly, the inclusion of the PAD increases the potential DNA binding surface of Pol η from 675 Å^2 to 1113 Å^2, which is comparable to that observed in other DNA polymerases.

The reorientation of the DNA from the pre-elongation into the 3'dG elongation state is accompanied by small motions of the thumb and PAD domains, relative to the palm and finger domains (Figure 5-20). The PAD is moving towards the DNA, performing a slight closing motion, whilst the thumb is moving as well and establishes more contacts to the DNA in proximity of the catalytic center (Figures 5-16 B and 5-17 B). These movements induce a shift of the DNA contacts relative to the PAD domain by one base pair (Figure 5-21).

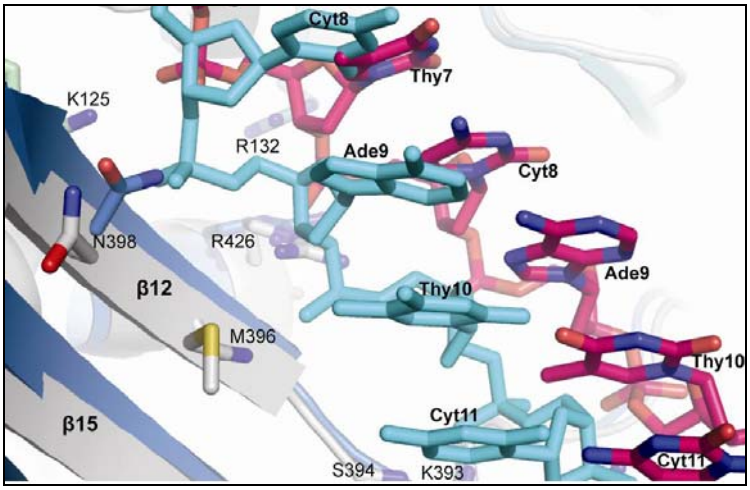


Figure 5-21. View of the DNA in proximity of the PAD domain. Overlay of the Pol η PAD domains of the 3'dG elongation state (colour coded as in Figure 5-15 with the DNA in magenta) and the the pre-insertion state (grey with the DNA in cyan). Peptide residues making functional contacts to the DNA are depicted as sticks with the carbon atoms shown in the colour of the corresponding peptide.

Comparing the *3'dG pre-elongation* with the *3'dG elongation* complex, using a different distance matrix plot[203], shows that relatively few changes occur within the individual domains, whilst marked shifts occur in the domains' positions relative to each other (Figure 5-22).

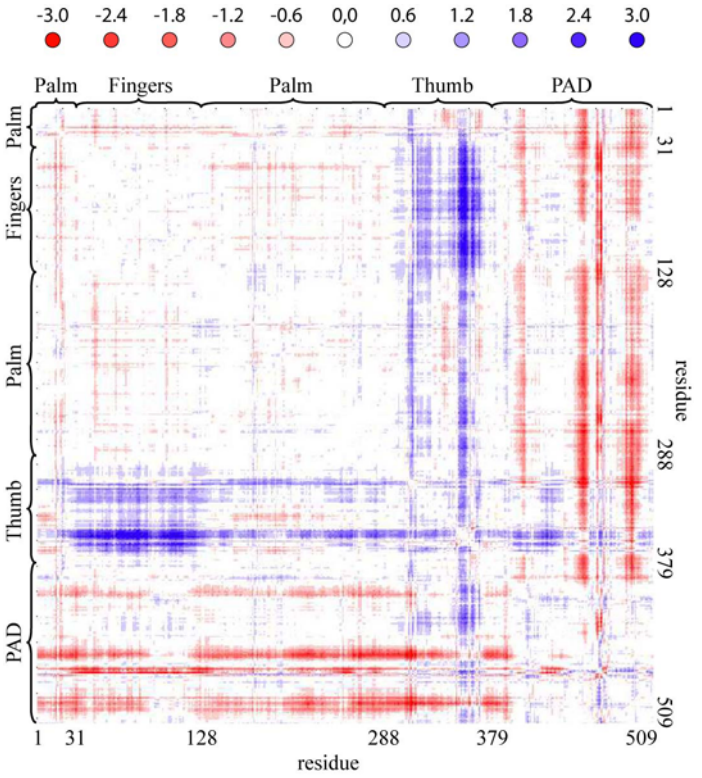


Figure 5-22. The difference distance matrix of the Pol η insertion complex was subtracted from the difference distance matrix of the Pol η pre-insertion complex. The delta values are plotted with colors representing CA differences of -3.0 to 3.0 Angstroms according to the scale at the top.

The areas that are essentially white (palm and fingers) represent regions of close structural similarity, without significant movement. The colors are interpreted to mean that the PAD domain of the *3'dG elongation* complex is closer (red) to the center of mass than the same domain of the *3'dG pre-elongation* structure. In contrast, the thumb domain of the *3'dG elongation* complex is farther away (blue) from the center of mass than the corresponding domain in the *3'dG pre-elongation* complex.

5.9 Biochemical studies

The primer extension studies were planned together with Claudia Chioccini, who carried them out. The reaction conditions are given in the legends.

5.9.1 Nucleotide insertion studies

The specificity of translesion synthesis opposite the damaged (lesion) and undamaged (native) bases was evaluated by using *GG_comp1* for studying the incorporation opposite the 3'dG (Figure 5-23 A) and *GG_comp2* for studying the incorporation opposite the 5'dG (Figure 5-23 B). Template strand *DPO4_GG* (Table 4-9) was prepared as described in 4.2.1.1 and purified by HPLC as described in 4.2.2. The template strand was annealed to the 5' fluorescein labeled primer strands *GG_comp1/2* in the reaction buffer in a 1:1.5 ratio in order to ensure that all primer strands anneal.

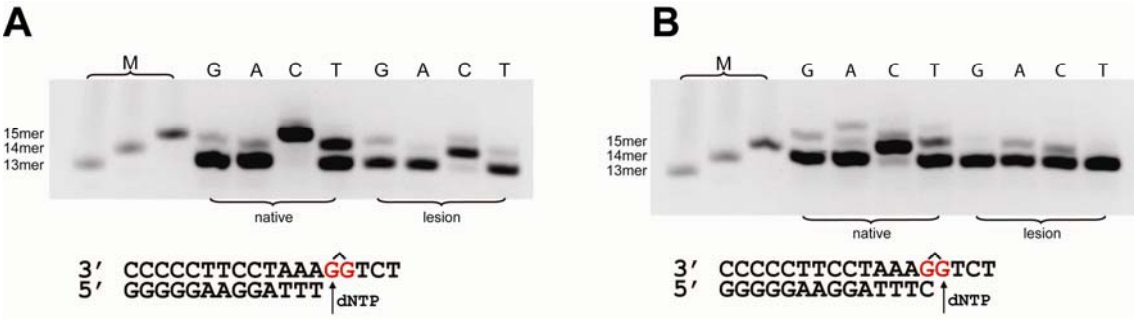


Figure 5-23. The reactions were carried out using 150nM Pol η, 1µM DNA, 500µM dNTPs and incubating at 35°C for 30 minutes. The markers (M) represent an insertion opposite the 3'dG (14mer) and the 5'dG (15mer). The dNTP used in each experiment is noted at the top of the lanes. **A:** Nucleotide insertion studies opposite the 3'guanine. **B:** Nucleotide insertion studies opposite the 5' guanine.

Primer extension studies revealed that Pol η inserts opposite the lesioned 3'dG mainly a dCTP and in the absence of the lesion besides the preferential incorporation of dCTP also a dTTP. In both cases the bypass is highly efficient and mostly error free (Figure 5-23 A).

Opposite the lesioned 5'dG the bypass is much slower and promiscuous for dCTP and dATP, whereas for the unlesioned template mainly dCTP and to a lesser extend a dTTP insertion is observed (Figure 5-23 B).

5.9.2 Structure based point mutation of Arg73

To gain further insight into the biochemistry of the bypass reaction we analyzed a site directed mutant of Pol η. We noticed that the Arg73 residue seems to activate and stabilize the dNTP for the lesion bypass steps, from the side opposite of the two metal ions. Interestingly, Arg73 is highly conserved among Pol η homologs, but not among other Y-family polymerases (Figure 5-24). This raises the question how much this residue determines the efficiency of the lesion bypass process.

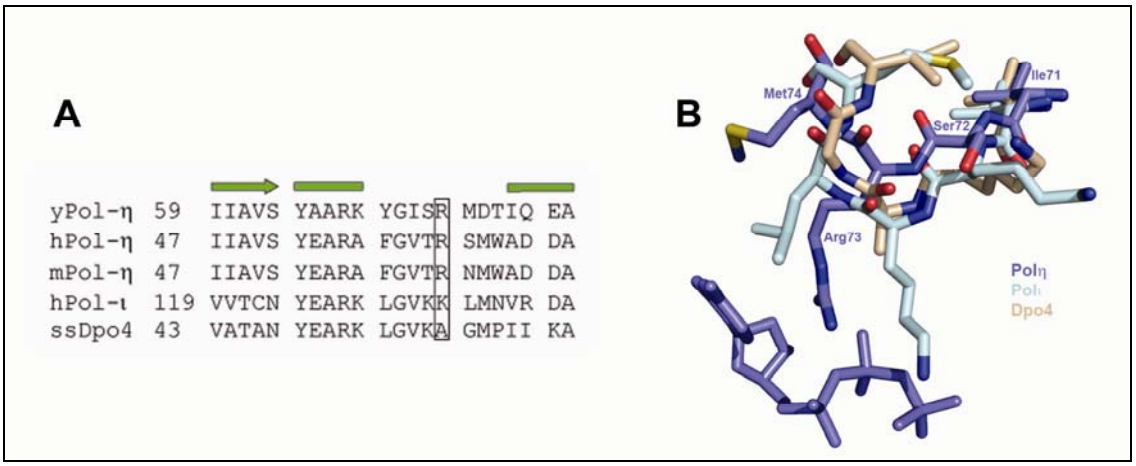


Figure 5-24. A. Sequence alignment of *S. cerevisae*, human and mouse Pol η, DPO4 and human Pol ι showing that R73 is conserved in Pol η homologs, but not in other Y-family polymerases. Secondary structure elements are depicted as rectangle for α helix and arrow for β strand. **B:** Structural alignment of the yeast Pol η, human Pol ι and DPO4 shows that Arg73 of Pol eta is in the Pol iota structure close to a Leu residue. By sequence alignment it occupies the same position as a Lysine residue. DPO4 possesses at that position an Ala.

A mutant of Pol η, wherein Arg73 was mutated to leucine (η*) was kindly prepared and provided by Carsten J. Pieck. This mutant also consisted of 532 amino acids, analogous to the unmutated polymerase.

The efficiency of translesion synthesis opposite undamaged and the damaged bases was evaluated by using *eta_Ep1* with and without a site specific cisplatin lesion for primer extension studies.

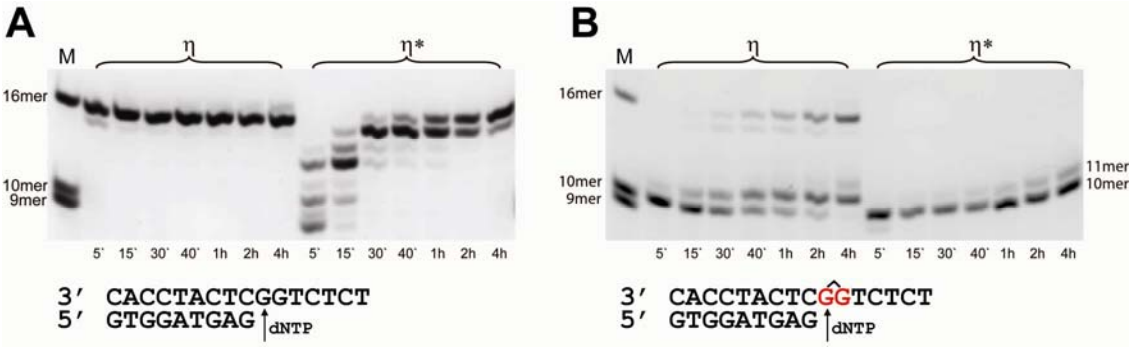


Figure 5-25. Reactions were carried out with 1µM DNA substrate, 250nM Pol η, and 200µM dNTPs at 26C° for the indicated amount of time. The markers (M) represent an insertion opposite the 3'dG (10mer), the 5'dG (11mer) and a full extension (16mer). **A:** Time dependent primer extension studies with Pol η (η) and Pol ηR73L (η*) using *eta Ep1* but without the site specific cisplatin lesion. **B:** Time dependent primer extension studies with Pol η (η) and Pol ηR73L (η*) using *eta Ep1* containing the site specific cisplatin lesion.

We found that Pol η* still can replicate undamaged templates, although the polymerization efficiency is substantially reduced when compared to the unmutated polymerase (Figure 5-25 A). Opposite a Pt-GG lesion, Pol η* is able to perform the first bypass step opposite the 3'dG slowly, in comparison with the wild-type enzyme, whereas the second step of lesion bypass, opposite the 5'dG is strongly compromised (Figure 5-25 B). The same biochemical properties are observed with CPD lesions (Figure 5-26), showing that the function of R73 is generally required for the replication through intrastrand cross-links.

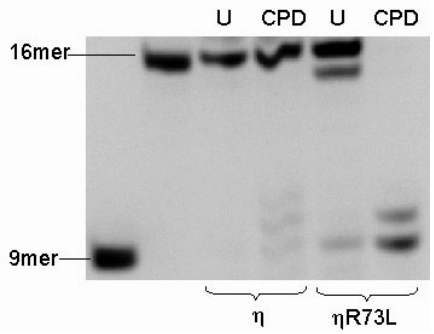


Figure 5-26. Primer extension experiment with Pol η and Pol ηR73L (η*) across a cyclobutane pyrimidine dimer (CPD) and undamaged DNA (U). The first two lanes contain markers for the unelongated primer (9mer) and the full extension product (16mer). Reactions were carried out using 250nM Pol η/Pol ηR73L, 1µM template DNA (*eta Ett1*), 200µM dNTPs and incubating at 26°C for 3h.

5.9.3 Functional hydrogen bonding studies with Zebularine

Primer extension studies were performed with Pt-GG containing templates and the 5'-triphosphate of 6-deaminocytidine (zebularine, dZTP), which was kindly prepared and provided by David Kuch. This base analog of cytidine, lacking the exocyclic C(4)-NH_2 amino group, can not form the H-bond to the C(6)=O carbonyl oxygen of the 5'dG, but can still form the two other H-bonds to dG.

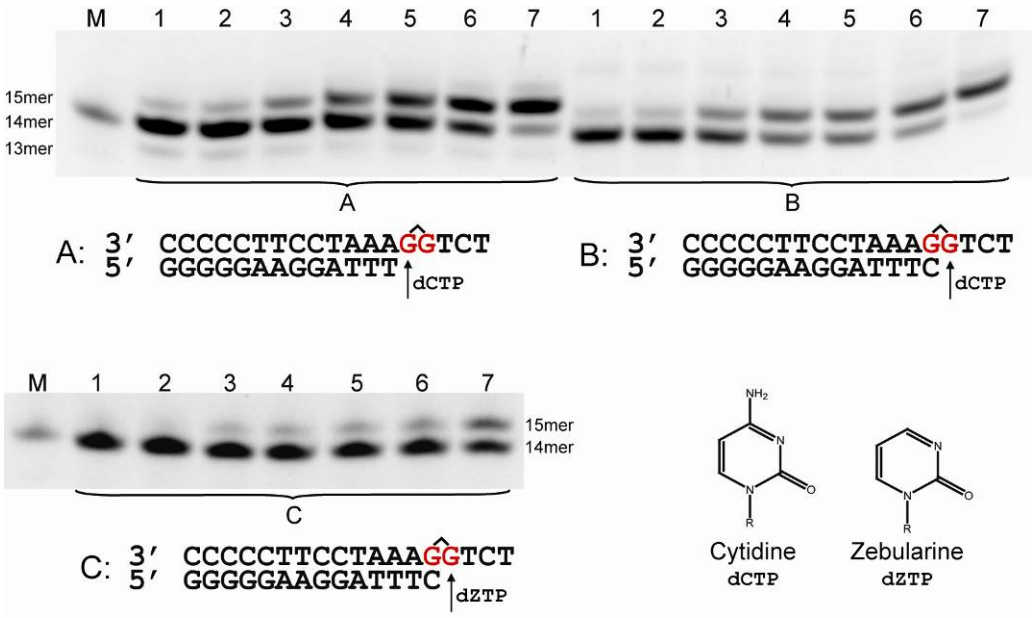


Figure 5-27. Efficiency of the translesion synthesis past the 3'dG and the 5'dG of the Pt-GG lesion. Reactions were carried out by incubating Pol η at increasing concentrations (lanes 1-7: 20/40/80/120/160/200/300 nM) for 20 minutes at 30°C with 500µM dCTP (**A** and **B**) or 500µM dZTP (**C**) and 1µM template DNA. A 13mer (**A**) or 14mer (**B** and **C**) hybridized to a 18mer DNA template (*DPO4_GG*) containing a site specific cisplatin lesion was used. The fluorescein labeled 14mer was used as a marker (M).

The translesion synthesis exhibited by Pol η past the 3'dG of the Pt-GG lesion is very efficient (Figure 5-27 A). Bypass past the 5'dG of the Pt-GG lesion with dCTP is slower than the preceding step (Figure 5-27 B) and is even more dramatically reduced when dZTP is being incorporated (Figure 5-27 C).

5.10 Cloning, expression and purification of DNA Pol η-513

The *S. cerevisiae Rad30* gene encodes the 632 residues of full length Pol η. A shortened, active version of 532 amino acids, truncated at the C-terminus was cloned and purified originally in our laboratory for primer extension experiments[188, 189].
For crystallization optimization we were in need of a slightly shortened version of the already shortened version at hand. These 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. They extend from the polymerase-associated domain and *in vivo* interact with the host replication complex, which targets Pol η to the replication fork through an interaction between PCNA[204] and residues 621–628 near the Pol η C-terminus [205].

A PCR product encoding Pol η 513 fused to a strep-tag was amplified by using the expression vector pDEST007-polη as a template and the primers Rad30-Gate-N and Pol η-513_R (Table 4-11). The expression clone was inserted by BP reaction into the entry clone, which was then subcloned in pExp007 by LR reaction. Both reactions are based on the Gateway technology. The resulting vector was transferred in *E. coli* Rosetta-gami and expression was performed identical to the protocol established by Dr. Claudia Chioccini (Figure 5-28).

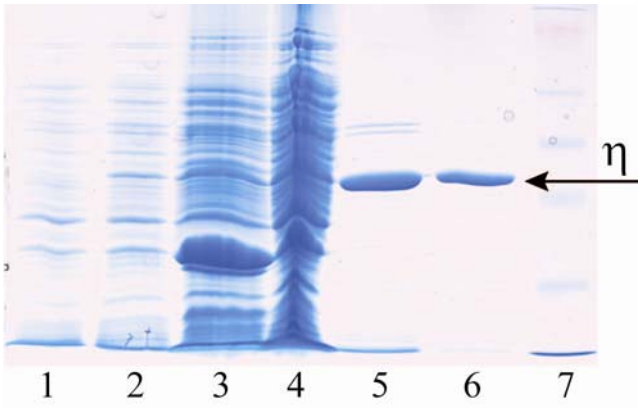


Figure 5-28. Expression and purification of Pol η-513. From left to right: cells before induction, cells after 4 hours induction, cell residues after lysis, cell lysate after lysis, Strep-tag purification, heparin purification and marker.

Pol η 513 was purified on a *Strep*-Tactin Superflow column identical to the protocol established by Dr. Claudia Chioccini. Thereafter the buffer was exchanged to *Heparin buffer A* and the protein was passed over a *HiTrap™ Heparin HP* ion exchange column for removing remaining DNA (Figures 5-28 and 5-29).

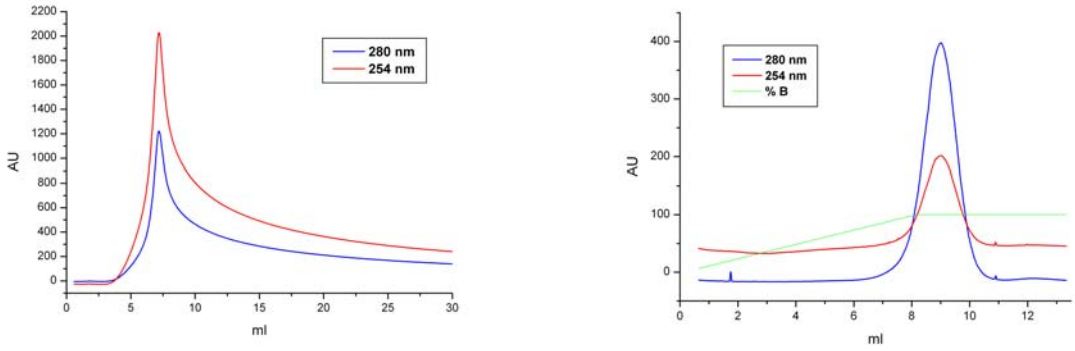


Figure 5-29. Elution profiles of Pol η-513 (absorption at 280 nm in blue; absorption at 254 nm in red) from a *Strep*-Tactin Superflow cartridge (left) and from a HiTrap™ Heparin HP Column (right).

After pooling and concentrating the fractions of the eluted protein, the concentration was determined on a *BioPhotometer 6131*.

5.11 RNA Pol II stalling at a cisplatin lesion

The detailed molecular mechanisms of cisplatin DNA adduct processing by nucleic acid polymerases are not understood. In order to derive the molecular mechanism of *S. cerevisiae* RNAP II stalling at cisplatin lesions, the DNA templates *Pol2_GG*, *Pol2_CGG* and *Pol2_IG* (Table 4-10) were prepared as described in 4.2.1.1 for a collaboration with the group of Prof. Cramer. These oligomers contained a site specific lesion at a GG or an IG sequence and were used in a combination of X-ray crystallography and RNA-extension assays by the Cramer group to elucidate recognition of cisplatin-induced DNA damage by transcribing RNAP II. Elongation complexes for crystallization were reconstituted from the 12-subunit *S. cerevisiae* RNAP II and the nucleic acid scaffolds as described[206, 207]. For this, the strands were planned with a cisplatin lesion incorporated at registers +2/+3 of the template strand, directly downstream of the NTP-binding site at register +1 (Figure 5-30).

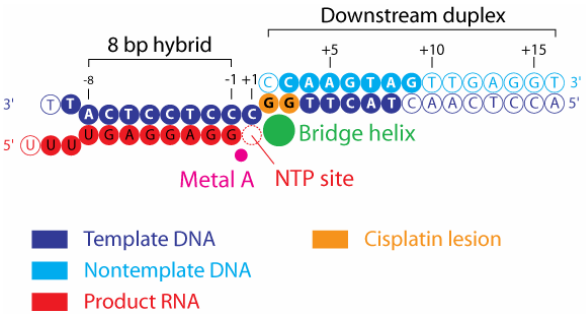


Figure 5-30. Nucleic acid scaffold. Filled circles denote nucleotides with interpretable electron density that were included in the structure (Figure 5-31); open circles denote nucleotides with uninterpretable or absent electron density.

The obtained crystal structure of the resulting cisplatin-damaged elongation complex was determined at 3.8 Å resolution. A very strong peak in the anomalous difference Fourier map revealed the location of the platinum atom, which was used as a marker for lesions positioning and determination of the register (Figure 5-31).

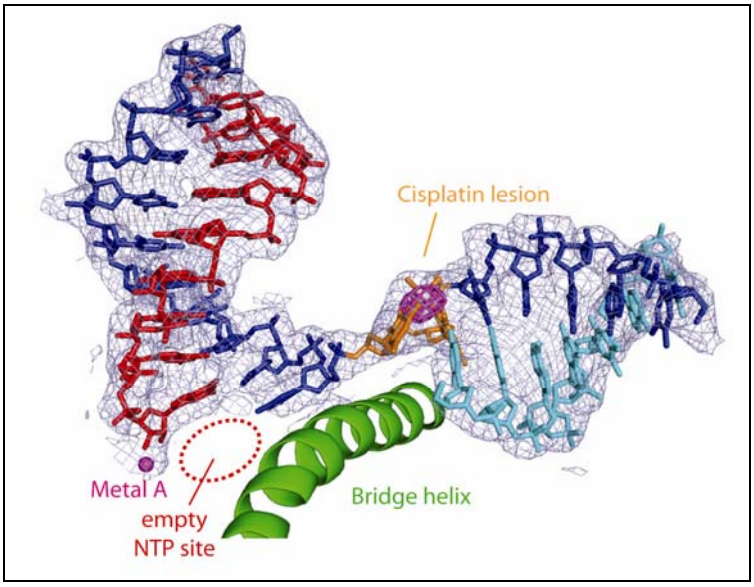


Figure 5-31. Structure of nucleic acids in RNAP II elongation complex A. Final $2F_o$-F_c electron density map for the nucleic acids is shown (blue, contoured at 1.0σ). Anomalous difference Fourier map reveals the location of the platinum atom (magenta, contoured at 15σ).

In the damaged elongation complex structure, the cisplatin lesion is bound at positions +2/+3, above the polymerase bridge helix (Figure 5-32).

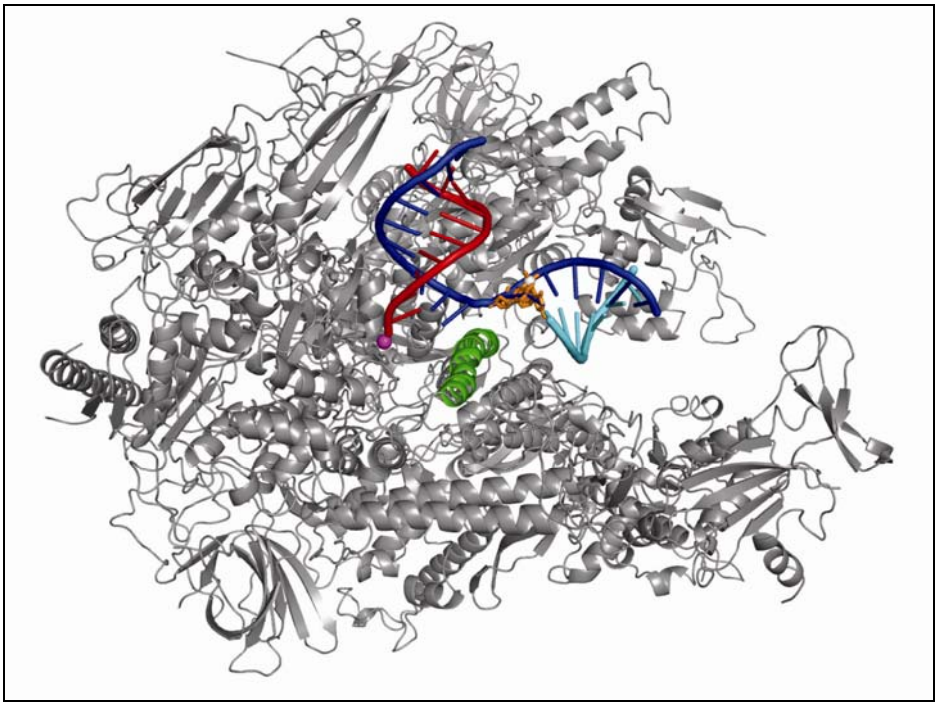


Figure 5-32. Overview of the cisplatin-damaged RNAP II elongation complex. RNAP II is shown as a silver ribbon, with the bridge helix highlighted in green. The nucleic acids are in color, and the cisplatin lesion is shown as a stick model in orange. A large portion of RNAP II was omitted for clarity.

The DNA-RNA hybrid occupies the upstream positions -1 to -8 (Figure 5-30). The hybrid structure is essentially identical to previous structures of the complete RNAP II elongation complex[207] and the elongation complex containing a CPD lesion at the polymerase active site[206]. However, the downstream DNA duplex adopts a slightly altered position (Figure 5-33).

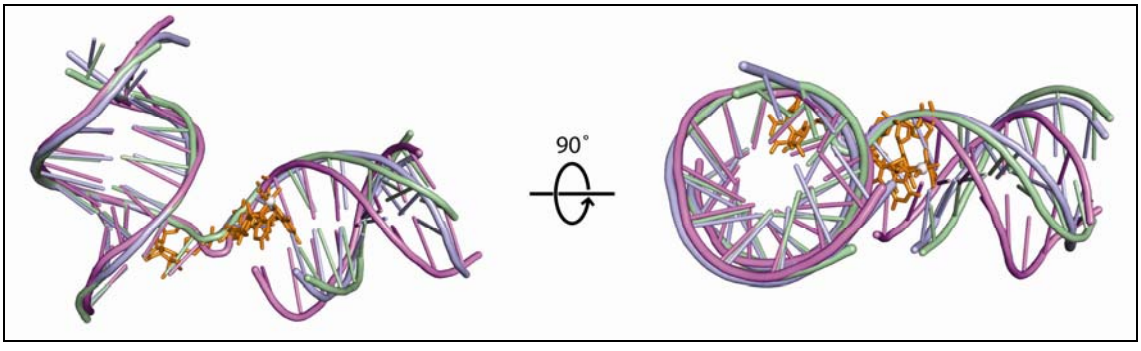


Figure 5-33. Cisplatin-induced changes in the downstream DNA. Comparison of the course of nucleic acids in the structures of the cisplatin-containing (pale green) with the CPD-containing complexes (violet and light blue). The proteins were superimposed on the basis of the active site region and then omitted. Nucleic acids are depicted as ribbons, lesions as orange sticks.

The change in the downstream DNA position results from the presence of the cisplatin lesion rather than from the scaffold design or the nucleic acid sequences, which were highly similar to those used previously and differed only in the substitution of Gs by Cs[206]. This was done in order to obtain a selective platination of the template strand and avoid unwanted side products. Structures of short DNA oligomers containing a cisplatin lesion reveal that the lesion at the center of the duplex leads to DNA bending (Figure 3-5)[65, 208]. In our structure, the lesion is located at the end of the duplex and induces a slight repositioning of the downstream DNA, without substantial changes in its internal structure.

6 Discussion

6.1 Lesion bypass

6.1.1 Lesion bypass of a 1,3-d(GpTpG) oxaliplatin lesion

Pol η was shown to bypass a 1,3 GTG oxaliplatin *in vitro* by inserting to a high degree the correct nucleotides (Figures 5-4 and 5-5). The lesion bypass seems to be slower than the bypass of an undamaged strand based on the gels which show still unelongated primers for the lesion containing strands. However kinetics need to be measured for the bypass of this lesion with Pol η to make a quantitive statement.

Solution structures of a double stranded DNA with a site specific 1,3 GTG cisplatin lesion show the T positioned in the minor groove and stacking with the 5'G. The T also does not hydrogen bond with its complimentary A, which stacks with its flanking Cs as in B-DNA (Figure 3-5)[66]. This could lead to the suggestion that Pol η inserts an A opposite the T in a non instructive fashion, a mechanism called A-rule[209, 210]. However, Pol η does not insert an A opposite an abasic site but a G, and this only in a very inefficient manner[161]. This led to the suggestion that Pol η does not follow the A-rule[210]. Pol η also requires instructions, based on hydrogen bonding, for inserting efficiently a complementary base[166]. Therefore it seems that the T, flanked by the two Gs of the lesion, is bypassed in an instructive fashion, rather than by following the A-rule.

One could argue that the whole lesion is flipped out of the active site and the next base, which is also a T, is responsible for instructing the insertion of an A. This could be easily investigated by conducting an experiment wherein a template strand containing the 1,3 GCG lesion is used for the primer extension experiments.

In summary, it seems that the active site of Pol η is big enough to accommodate the 1,3-GTG oxaliplatin lesion and that this lesion is processed in an efficient and instructive fashion.

6.1.2 Lesion bypass of a 1,2-d(GpG) cisplatin lesion

The structures of the ternary complex of yeast Pol η with a Pt-GG in the template DNA and an incoming dNTP have provided clear evidence that its active site is open enough to accommodate both residues of a Pt-GG (Figures 6-1 and 6-2). This is in contrast to

most DNA polymerases, which include not only the high-fidelity replicative polymerases but also the other members of the Y-family, in which only the templating residue can fit in the active site, whereas the next 5'-template base along with the rest of the unpaired template is pushed out of the active site at a 90° angle [100-102, 106-109, 211-213]. Because the two purines of a Pt-GG are covalently linked by the cisplatin and therefore cannot be separated, the active site of most polymerases is not equipped to handle this lesion. Interestingly, at the *3'dG elongation* step, the base 5' to the templating base is in a nearly 90° angle to the base located in the active site due to the conformation induced by the cisplatin crosslink. This highly resembles the conformation adapted by the unpaired template strand in high-fidelity replicative polymerases (Figure 3-9).

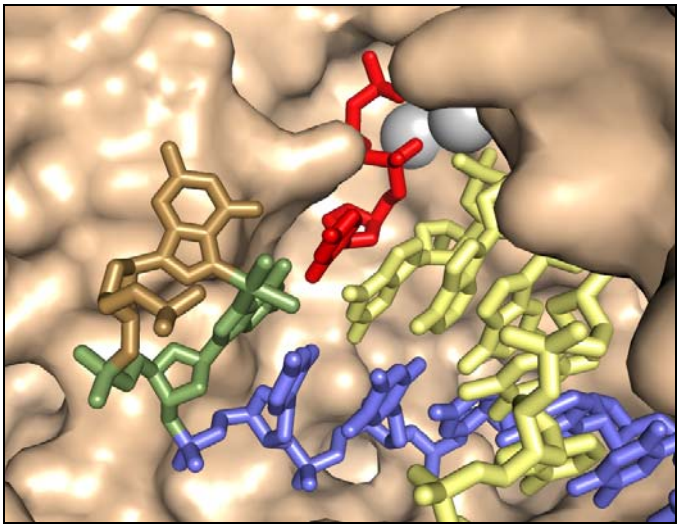


Figure 6-1. Protein surface representation of Pol η with the DNA and dNTP depicted as sticks, showing the *3'dG elongation* step. The primer is yellow and the dNTP to be inserted is red. The template strand is blue, with the templating nucleotide in olive and the nucleotide to its 5′-side in brown. The magnesium ions are depicted as grey spheres.

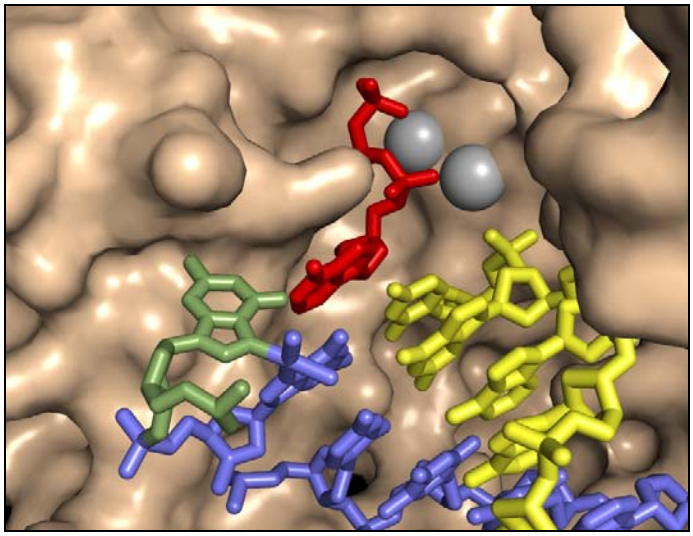


Figure 6-2. Protein surface representation of Pol η with the DNA and dNTP depicted as sticks, showing the *5'dG elongation* step. The primer is yellow and the dNTP to be inserted is red. The template strand is blue, with the templating nucleotide in olive. The magnesium ions are depicted as grey spheres.

The four obtained Pol η complexes plus the biochemical data allow us to suggest a plausible scenario of how Pol η manages to bypass Pt-GG lesions.

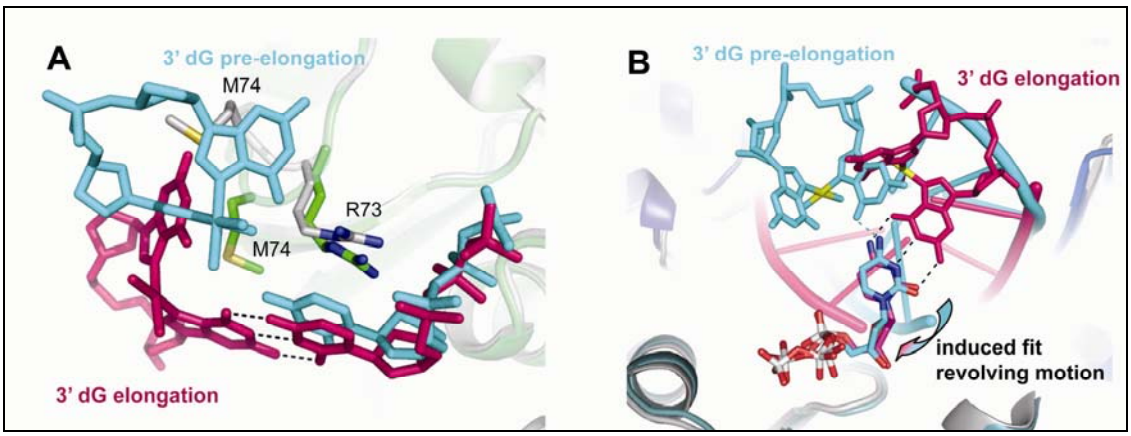


Figure 6-3. The 3'dG elongation process of Pol η. **A:** Detailed view of the lesion in the *"pre-elongation"* complex (cyan) superpositioned with the lesion in the *"3'dG elongation"* state (magenta). **B:** Correct Watson-Crick base pairing revolves the DNA to position the 3'-OH of the primer for nucleophilic attack on the α-phosphate of the dNTP. For clarity the finger domain has been omitted and the DNA molecules (cyan and magenta) are viewed in a simplified form.

The enzyme seems to bind the template-primer complex first in a non-productive manner described by the *3'dG pre-elongation* state (Figure 6-3 A). The dNTPs are

bound and presumably rapidly exchanged in the active site in the absence of a templating base. If the templating base is able to establish a sufficient number of H-bonds with the dNTP, it shifts the primer-template equilibrium toward the *3'dG elongation* state, in which nucleotidyl transfer occurs (Figure 6-3 A). This revolving motion (Figure 6-3 B) results in reorientation of the DNA from the *3'dG pre-elongation* into the *3'dG elongation* state and is accompanied by small motions of the thumb and PAD domains, relative to the palm and finger domains (Figure 5-20). The DNA contacts to the PAD are shifted by one base pair (Figure 5-21), which likely helps to maintain the contact between Pol η and the DNA template upon switching between both states.

In the second step of lesion bypass, described by the *5'dG elongation* complex, the primer template moves forward, with one H-bond apparently being the minimum requirement to pull the template primer complex toward the incoming dNTP in order to allow the 5'dG nucleotidyl transfer, albeit at much reduced efficiency and accuracy. This scenario is supported by various biochemical observations[166, 214]. A rate determining conformational adjustment was determined by pre-steady state kinetic experiments. Although these data were obtained with undamaged DNA templates, such a process could be the revolving motion of the DNA-template complex, which is essential in order to bring the dNTP in close proximity to the 3'-OH of the primer. This conformational rearrangement was indeed found to precede the chemical step. [214] Such a step would be mechanistically quite distinct from the dNTP directed induced fit observed in high fidelity polymerases [100, 101] and could be a specific adaptation for replication of templates containing intrastrand crosslink lesions.

Regardless, the structural mechanism of primer extension argues that Pol η selects correct dNTPs not by shape [215, 216] but by hydrogen bond complementarities, which is fully consistent with previous biochemical observations [166]. Pol η seems to share this property with the herpes simplex virus and human DNA primases, which both were shown to select the correct dNTP based exclusively on H-bonding complementarity[217, 218].

To test this key hypothesis, we performed functional hydrogen bonding studies with Pt-GG containing templates and the 5'-triphosphate of 6-deaminocytidine (dZTP). This base analog can not form the H-bond to the O(6) of dG.

As predicted from our structural model, bypass of the 5'dG with dZTP is dramatically reduced compared to bypass with dCTP or dATP (Figure 5-27), which verifies the essentiality of hydrogen bond formation for substrate recognition and incorporation by Pol η.

6.2 Compare to the apoenzyme

A comparison of the *3'dG elongation* complex and the Pol η apoenzyme[129] reveals small but significant changes upon DNA binding. In particular, the whole thumb domain moves towards the DNA (Figure 6-4). The loop, formed by the residues 306-312, contacts the DNA backbone and the loop 353-363 moves towards the minor groove. The β-sheet β_{14} of the PAD domain moves by about 4 Å towards the DNA backbone.

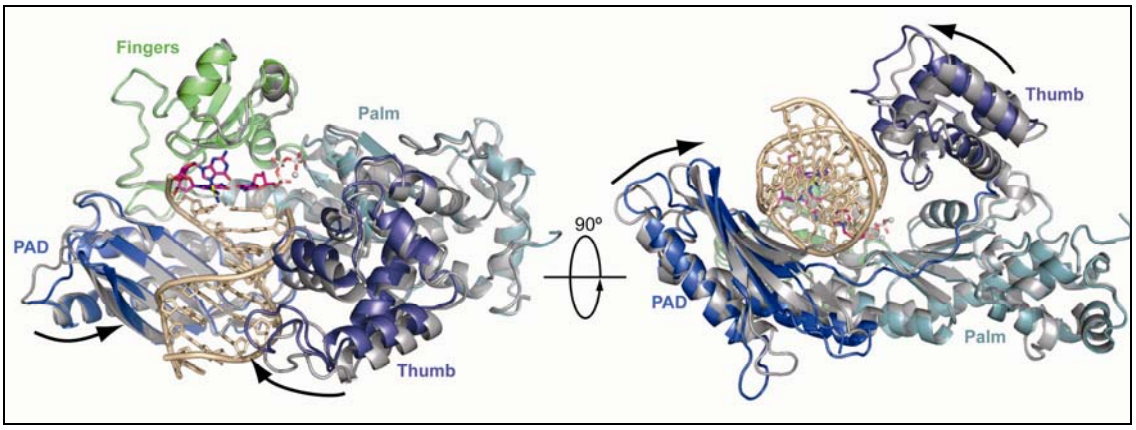


Figure 6-4. Overlay of the apoenzyme with the first elongation complex The ***"3'dG elongation"*** complex is shown as a cartoon with the various domains of the protein color coded and superpositioned with the apoenzyme (1JIH)[129] in grey. In magenta, the cisplatin lesion and dCTP are shown. The primer and template strands of the DNA in brown and the Watson-Crick H-bonded dCTP in magenta are depicted as sticks. The two catalytic magnesium ions are shown as grey spheres. The views are related by a 90° rotation around a horizontal axis.

Comparison of the apoenzyme with the *3'dG elongation* complex, using a different distance matrix plot[203], shows that relatively few changes occur within the individual domains, whilst marked shifts occur in the domains' positions relative to each other (Figure 6-5). One exception, clearly highlighted in the plot, is loop 73-74 consisting of

the structurally identified Arg73, which orients the dCTP in the active site and Met74, which slips in between the two cisplatin bound guanines.

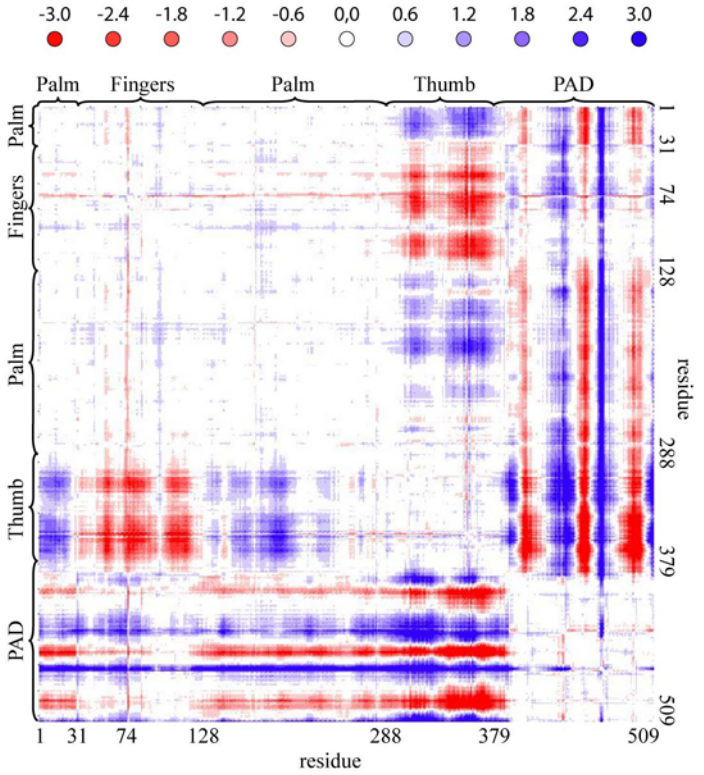


Figure 6-5. The difference distance matrix of the Pol η *3'dG elongation* complex was subtracted from the difference distance matrix of the Pol η apoenzyme. The delta values are plotted with colors representing CA differences of -3.0 to 3.0 Angstroms according to the scale at the top.

Those areas that are essentially white (palm and fingers) represent regions of close structural similarity. The colors are interpreted to mean that the PAD domain of the *3'dG elongation* complex is closer (red) to the center of mass than the same domain of the apoenzyme structure. In contrast, the thumb domain of the insertion complex is farther away (blue) from the center of mass than the corresponding domain in the apoenzyme structure (Figure 6-5).

6.3 Polymerase switch

Studies with yeast Pol η and its natural substrate, the TT dimer, reveiled that the polymerase inserts a correct nucleotide opposite the 3'dT of the dimer, which results in its stable association with the TLS intermediate. The Pol η - DNA complex remains stable after the incorporation of nucleotides opposite the 5'dT of the dimer and the next two nucleotides beyond the lesion[219]. Thereafter, the Pol η - DNA complex is no longer stable, and Pol η dissociates from the DNA. In addition, DNA synthesis by Pol η up to two nucleotides beyond the TT dimer was shown to be both, necessary and sufficient to allow Pol α to resume DNA replication[220].

This implies that Pol η senses the location of the dimer as synthesis proceeds. By analyzing our structures with a Pt-GG one notices that termination following incorporation of 2 nucleotides beyond the cisplatin lesion places the two damaged bases at the third and fourth base pairs of the duplex primer template (Figure 6-6). This positions the lesion in immediate proximity of the PAD domain, which is present only in Y-family polymerases and differs among family members.

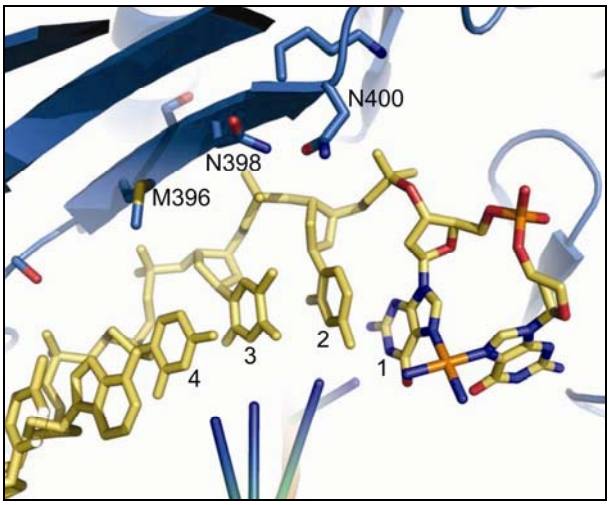


Figure 6-6. Interaction of the template strand with the PAD. The structure presented is the *3'dG insertion* complex, wherein the 3'dG of the Pt-GG is copied (position1). The insertion of another 3 bases would place the lesion in immediate proximity of the PAD (position 4). The primer strand is shown as a simplified cartoon for clarity.

The reason for the lowered affinity of Pol η to the DNA at this point of replication could be explained by unfavourable interactions of the lesions phosphodiester

backbone with the PAD domain. Due to the two guanines rigid connection through the platinum, the backbone of the lesion itself is also less flexible than in undamaged DNA. This might result in clashing of the backbone into the PAD as replication after the lesion continues, rather than sliding along the PAD. This interaction might then result in disassociation of Pol η from DNA, which triggers the switch to another polymerase. Whether or not PCNA must be specifically deubiquitinated in vivo to facilitate such dissociation and further polymerase switches remains an open question.

6.4 Use of ddNTP leads to potential misinterpretations of the structures

To investigate the mechanism that allows Pol η to replicate through helix disturbing DNA lesions, we crystallized a 16-base-pair DNA template strand containing a single Pt-GG intrastrand crosslink, hybridized to a 9mer primer strand and a dideoxycytidinetriphosphate (ddCTP) in complex with the large fragment of DNA polymerase η from yeast. This resulted in the first crystal structure of this eukaryotic Y-family polymerase in ternary complex.

However we realized that by using a natural primer and a ddCTP, which was used to inhibit nucleotidyl transfer, we obtained a structure wherein the ddCTP is not placed correctly in the active pocket for base pairing. Although the triphosphate stays complexed to the two metal ions and Arg73 residue coordinates the ddCTPs' phosphates from the side opposite of the two metals, the sugar triphosphate unit adopts a stretched out conformation, with the base protruding into the DNA binding cleft (Figure 6-7).

Since Pol η contains a wide DNA binding cleft and the DNA can slide along the PAD with some degree of freedom, the template adjusted itself to the location of the ddCTP and established bona fide Watson-Crick hydrogen bonding. This also resulted in the lesion being located outside the active pocket.

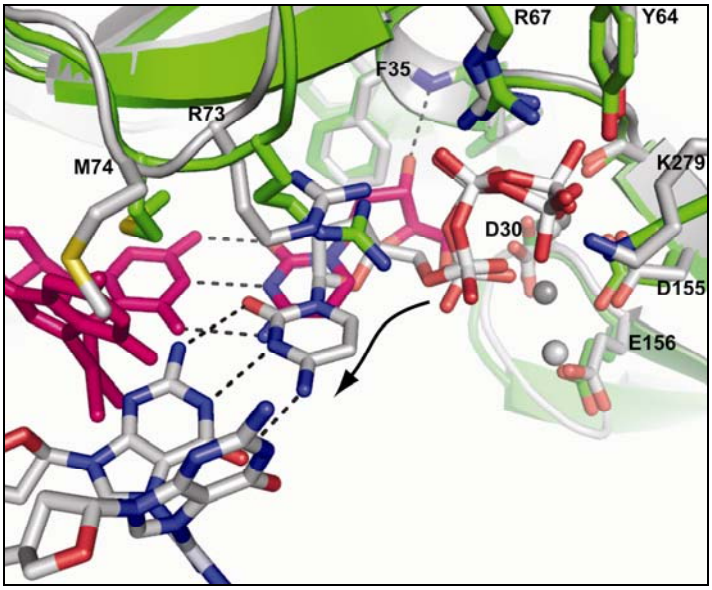


Figure 6-7. Overlays of Pol η,crystallized with a ddCTP (grey) with the Polη 3'dG elongation state (green and magenta), which was crystallized with a dCTP. The catalytic carboxylates, Ca^{+2} ions, lesions and the incoming nucleotide are shown in the stick model. The overlays were generated by superimposing the palm domains of the two proteins.

We postulated that by using a 2'-deoxycytidinetriphosphate (dCTP) for the crystallization setup, we would obtain a structure, wherein the dCTP would be confined to the active site by interactions of the 3'hydroxyl group with the peptide backbone of Phe35. The 3'-OH group would control the positioning of the incoming nucleotide in the wide active site to establish a correct set of H-bonds with the lesion containing base. Based on these structural results, obtained by cocrystallizations with ddCTP and the compare to the structures obtained by crystallizing with dCTP (Figure 6-7), we assume that parts of the reported discrepancies between biochemical[221] and structural[183] data of the translesion synthesis process of Y-family polymerases may be explainable with the use of dNTPs in the crystallization experiments. These will be discussed in further detail in chapter 6.5.

6.5 DPO4 vs Pol η lesion bypass

Dpo4 from archael *S. solfataricus* is phylogenetically related to the DinB group of Y-family polymerases, members of which, like *E. coli* Pol IV and human Pol κ, are highly inefficient at replicating through CPDs[222]. Although Dpo4 can replicate through a cis-syn TT dimer, it does so with a much more reduced efficiency than does Pol η. However, it has been suggested that in its lesion bypass properties, including its ability to bypass CPDs, it is more akin to Pol η than to the other Y-family polymerases[223].

Dpo4 was used as the model polymerase for clarifying the TLS past TT dimers and was crystallized in ternary complex with ddATP and a TT dimer containing template–primer DNA strand. The x-ray crystal structure of Dpo4 complexed with a cis-syn TT dimer has shown that, whereas the 3'dT of the CPD (TT-1) forms a Watson–Crick base pair with the incoming dideoxy ATP (Figure 6-8 A), the 5'dT (TT-2) forms a Hoogsteen base pair with the dideoxy ATP in syn conformation (Figure 6-8 B)[183].

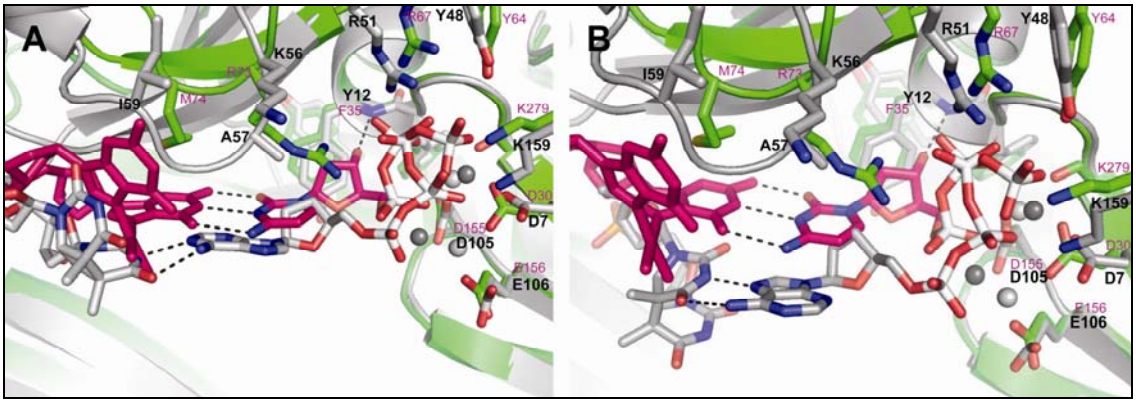


Figure 6-8. Overlays of DPO4 (grey with labels in black) with the Polη 3'dG elongation state (green and magenta with labels in magenta). **A:** DPO4 in ternary complex with a ddATP opposite the 3'dT of a TT dimer (TT-1). **B:** DPO4 in ternary complex with a ddATP opposite the 5'dT of a TT dimer coordinated via Hoogsteen base pairing (TT-2). The overlays of the TT-1 and TT-2 structures with Pol η were generated by superimposing the palm domains of the two proteins.

Based upon these structures, a similar mechanism for nucleotide incorporation opposite a TT dimer has been proposed for Pol η. However biochemical studies with dNTP analogues showed that both Dpo4 and Pol η indeed do proceed by different mechanisms, but neither of them engage in Hoogsteen base pairing for incorporating an

A opposite the 5'dT of the CPD[221]. The mechanism of CPD bypass opposite the 3'dT adopted by Dpo4 resembles that used by T7, in which nucleotide incorporation involves an abasic site-like intermediate[130], whereas Pol η uses the normal Watson–Crick base pairing[221]. Opposite the 5'dT both polymerases incorporate an A by using Watson-Crick base pairing.

The origin of the misinterpretation of the DPO4 crystal structures in ternary complex with a TT dimer might lie in the use of a ddATP, which lacks the 3' hydroxyl group, vital for correctly positioning the triphosphate in the polymerases' active pocket. When superimposing the structures of DPO4 with our structure obtained for the 3' elongation step one can clearly see the similarities to our structure which was obtained by cocrystallization with a ddCTP (Figure 6-7). The three catalytic carboxylates Asp7, Asp105 and Glu106 of DPO4 are superimposable with Asp30, Asp155 and Glu156 in Pol η for both complexes TT-1 and TT-2 as are Tyr12 and Phe35 which comprise the steric gate (Figure 6-8). In both DPO4 structures the ddATP protrudes out of the active site establishing hydrogen bonds to the templating CPD, in the case of TT-2 via Hoogsteen base pairing. As was observed for the TT-2 structure the templating thymine did not rotate 36° or rise by 3.4 Å relative to the preceding base pair as in normal DNA. The incoming ddATP assumed a syn conformation and formed a Hoogsteen base pair with the 5' thymine of the CPD and thereby maintained the triphosphate contact with Dpo4 for nucleotidyl transfer.

According to the biochemical observations in the case of TT-1 the templating CPD can not enter the active site. This is actually observed in the crystal structure, but the base pairing is still maintained since the ddATP is not confined to the active site due to the lack of the 3' hydroxyl group, which would establish a hydrogen bond to the backbone of Tyr12. Because of the same reason the TT-2 structure reveals a Hoogsteen base pairing. The ddATP is in a stretched out conformation and the template is in position for replication. Would the ddATP pair through Watson-Crick hydrogen bonds in this structure, the DNA would need to be pushed away from the active site. This way the DNA is correctly positioned, however the ddATP needs to adapt a syn conformation and thus Hoogsteen base pairing was observed for this structure.

Polymerases belonging to the Y-family are characterized by more open and relaxed active sites and thus the correct positioning of the dNTP in the active site is of higher

importance than in high fidelity polymerases. These rely on an induced fit mechanism and therefore the absence of the 3' hydroxyl group is of lesser importance.

Parts of the reported discrepancies between biochemical[221] and structural[183] data of the translesion synthesis process of Y-family polymerases may be explainable with the use of dNTPs in the crystallization experiments, which should be employed in crystallizations of polymerases belonging to the Y-family.

6.6 Mechanism of transcriptional stalling at cisplatin lesioned DNA

6.6.1 RNA polymerase II stalling and AMP misincorporation

The crystal structure of complex A which was crystallized with *scaffold A* (Figure 6-9) defines a state of the elongation complex in which the lesion lies downstream of the active site (Figures 5-31 and 5-32). Complex A was therefore used in RNA-extension assays to investigate the mechanism of RNAP II stalling.

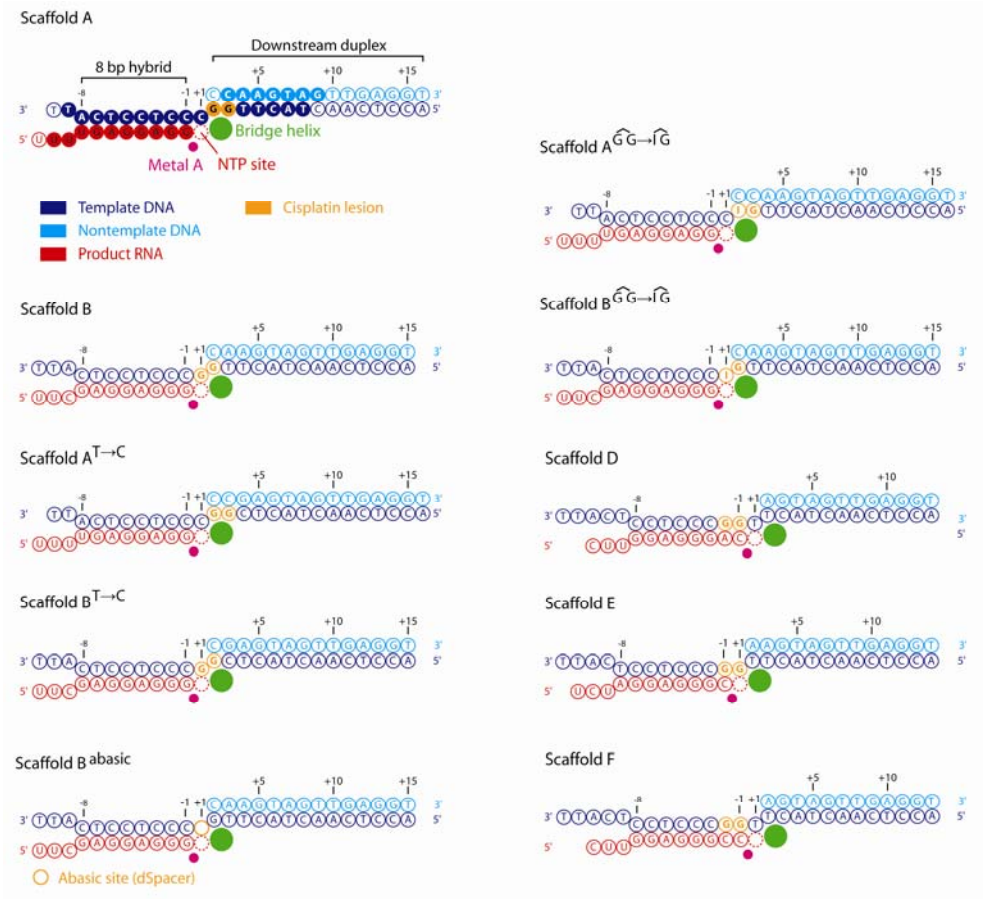


Figure 6-9. Nucleic acid scaffolds for crystallizations and primer extension studies.

Most of the RNA was extended by two nucleotides (Figure 6-10). Thus, the complex apparently stalled after nucleotide incorporation opposite the first guanosine (the 3'-guanosine) of the cisplatin lesion. Incubation of complex A with subsets of NTPs suggested that the terminal incorporation is a specific misincorporation of AMP (Scaffold A, Figure 6-10).

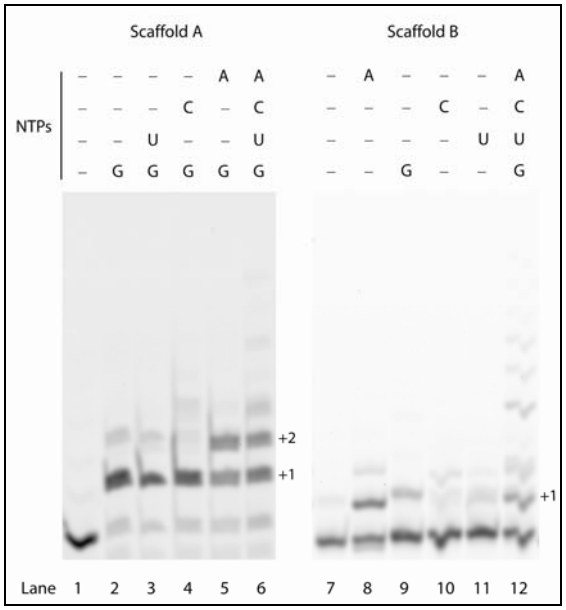


Figure 6-10. RNA extension with scaffolds A and B. Lanes 1 and 7 show the fluorescently labeled reactant RNA. In the other lanes, the scaffolds were incubated with RNAP II and indicated NTPs (1 mM) for 20 min.

To further investigate this, a scaffold that resulted in an elongation complex which was advanced by one position was prepared (*scaffold B*, Figure 6-9). Also incubation of this complex B with individual NTPs revealed AMP misincorporation (Scaffold B, Figure 6-10).

6.6.2 Impaired entry of lesions into the active site

If the AMP misincorporation is templated by the lesion, the lesion must adopt a position in the active site, at least transiently. However, structural considerations suggest that lesion entry into the active site would be impaired, because translocation of the cisplatin dinucleotide lesion from positions +2/+3 to positions +1/+2 is expected to be disfavored, as is the case for a dinucleotide photolesion[206]. Template bases in positions +1/+2 are twisted against each other by about 90° in structures of the

undamaged elongation complex[178, 207], but such twisting is impossible for nucleotides that are covalently linked in dinucleotide lesions, giving rise to a translocation barrier[206]. To determine whether translocation of the cisplatin lesion is indeed impaired complex B was crystallized. The lesion was designed to be placed at positions +1/+2 (*scaffold B*, Figure 6-9). The anomalous difference Fourier map revealed a platinum peak at the same location where one was seen in complex A. This indicates that the polymerase had apparently stepped backward by one position, placing the cisplatin lesion again at positions +2/+3 (Figure 6-11 A). A lower height of the platinum peak indicated partial occupancy of the nucleic acids. Indeed, the electron density for the nucleic acids was weak and fragmented, and did not allow for model-building. Thus, the cisplatin lesion was not stably accommodated at positions +1/+2, showing that translocation from positions +2/+3 to positions +1/+2 is apparently disfavored.

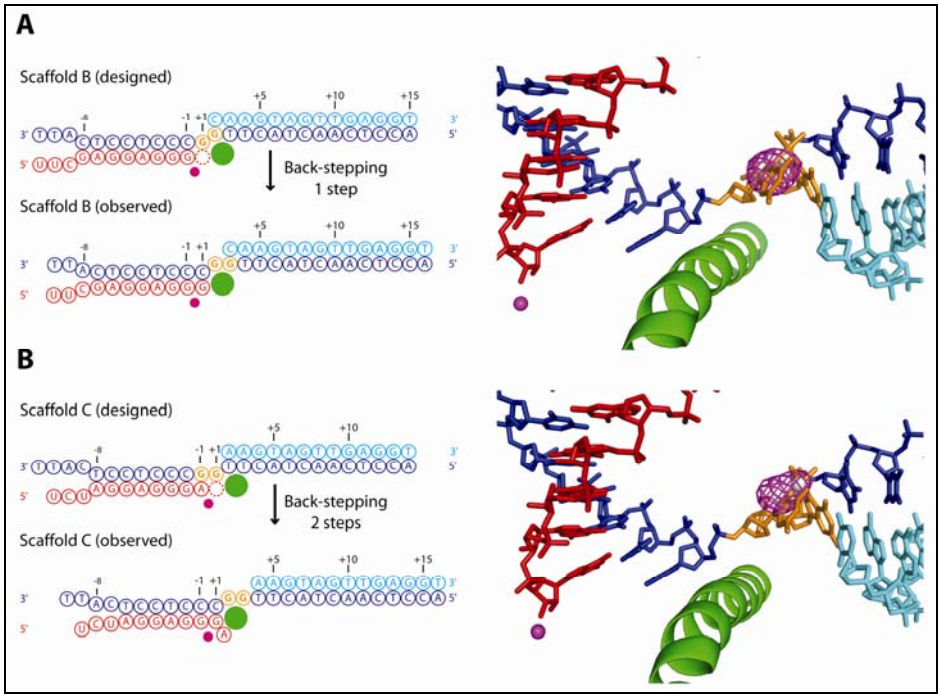


Figure 6-11. Anomalous difference Fourier maps of RNAP II elongation complexes B **(A)** and C **(B)**, contoured at 6σ. Model of complex A is shown, viewed from the side.

These results suggest that the lesion does not stably bind the active site at all. To test this, a scaffold that contained the lesion at predicted positions -1/+1 in the active site, and a G·A mismatch pair at position -1 *(scaffold C*, Figure 6-9) was prepared. The resulting complex C was crystallized, however the anomalous difference Fourier map

revealed a platinum peak at the same location as in complexes A and B. This indicated that the polymerase had stepped backward by two positions so that the cisplatin lesion again occupied positions +2/+3 (Figure 6-11 B). The height of the platinum peak was lower than for complexes A and B, and the electron density for the nucleic acids was again fragmented, preventing model-building. The attempt to solve a structure with the lesion placed at positions -2/-1 (*scaffold D*, Figure 6-9) resulted in a crystal structure with no electron density for nucleic acids. This again was consistent with a low affinity of this scaffold for the polymerase. Thus, the cisplatin lesion was not stably accommodated in the active site.

6.6.3 Possible mechanisms for misincorporation

As the template strand in our scaffold contained a thymidine immediately downstream of the cisplatin lesion, AMP misincorporation may have arisen from template misalignment, a recently characterized mechanism for misincorporation of nucleotides[224]. During misalignment, the cisplatin lesion would transiently adopt a flipped-out, extrahelical conformation, and the thymidine flanking the lesion on the 5' side would transiently occupy the position of the template base, at register +1 in the active site, to direct AMP incorporation. To test whether this was the case, a scaffold that was identical to scaffold A except that the flanking thymidine was replaced by cytidine (*scaffold* $A^{T\rightarrow C}$, Figure 6-9) was prepared. This altered scaffold gave rise to the same RNA-extension products, showing that template misalignment did not occur. This was also confirmed by altering scaffold B (*scaffold* $B^{T\rightarrow C}$, Figure 6-9) and repeating the RNA-extension assay[225].

Another possible explanation for AMP misincorporation is that the 3'-guanosine of the lesion could act as a templating base by adopting a position that allows it to form two hydrogen bonds with the Watson-Crick positions of an incoming ATP substrate (Figure 6-12 B). To test whether the third Watson-Crick position of the 3'-guanosine in the lesion, the extracyclic 2-amino group, is involved in templating AMP misincorporation, we replaced the 3'-guanosine in the lesion with inosine, which lacks the 2-amino group (Figure 6-12 B). RNA-extension analysis showed that AMP was still specifically misincorporated when scaffolds A and B were modified by replacement of the GG cisplatin lesion with an IG cisplatin lesion (Figure 6-12 A). Thus, AMP misincorporation does not involve the 2-amino group. These results were still

consistent with lesion-templated misincorporation involving a G•A base pair (Figure 6-12 B), but this could not be tested directly.

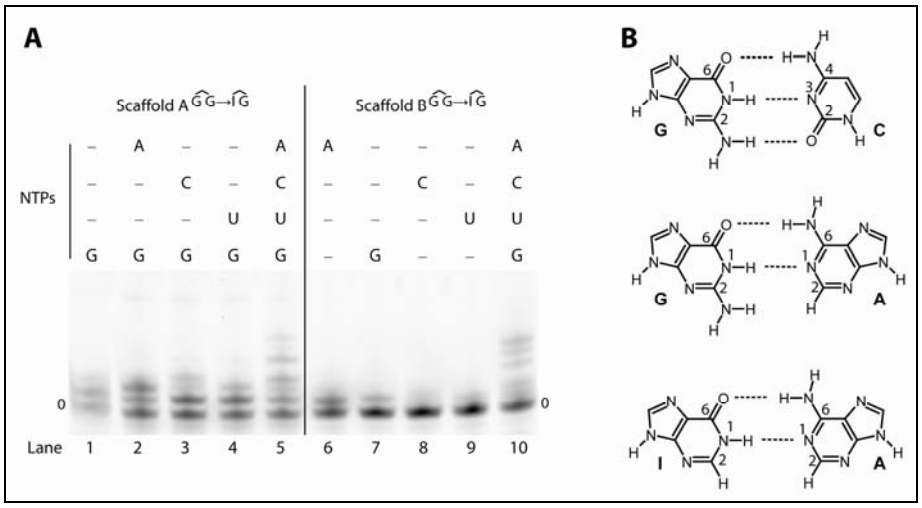


Figure 6-12. A: The 2-amino group at the 3' position of the lesion is not involved in directing misincorporation. Altered scaffolds were used in RNA-extension assays in which the guanosine at the 3' position of the lesion was replaced by inosine. **B:** Possible base pair formation. Hydrogen bonds are indicated with dashed lines.

6.6.4 Nontemplated AMP incorporation and an 'A-rule' for RNAP II

The crystallographic data strongly suggest impairment of translocation of the lesion to a position where it can direct AMP misincorporation. Could AMP incorporation be a result of nontemplated synthesis? Such preferential, nontemplated AMP incorporation occurs when DNA polymerases encounter a bulky DNA lesion and is known as A-rule[209, 210]. To test this hypothesis, a scaffold identical to *scaffold B*, but containing an abasic site at the templating position +1 and containing no cisplatin lesion (*scaffold* B^{abasic}, Figure 6-9) was prepared. Indeed, AMP, and also to a lesser extent GMP, could be incorporated into RNA opposite the abasic position (Figure 6-13), showing that RNAP II incorporates purine nucleotides, preferentially AMP, when a templating base is unavailable. Moreover, the efficiency of nontemplated AMP incorporation was comparable to that of the terminal AMP misincorporation during stalling at a cisplatin lesion, as judged from the slow rate of the reaction and its dependence on NTP concentration[225].

These observations suggest that RNAP II obeys an 'A-rule' for nontemplated nucleotide addition, which could explain the terminal AMP incorporation when the bulky cisplatin

lesion is encountered. This mechanism is consistent with the model in which the cisplatin lesion does not enter the active site during elongation, not even transiently to direct terminal AMP incorporation.

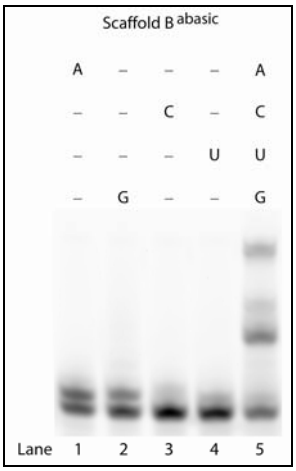


Figure 6-13. Nontemplated purine incorporation at an abasic site. The scaffold used corresponds to scaffold B but contains an abasic site at the templating position +1 and no cisplatin lesion.

6.6.5 Comparison to TT dimer damage recognition

Comparison of the results with the previously established mechanism of RNAP II stalling at a TT dimer reveals that RNAP II stalls at the two dinucleotide lesions for different reasons (Figure 6-14). First, the cisplatin lesion cannot overcome the translocation barrier, but the TT dimer can. Second, inefficient AMP incorporation occurs at both lesions, but it enables lesion binding at the active site positions -1/+1 only for the CPD, probably because a stabilizing T-A base pair is formed at position -1, in contrast to a G-A mismatch that would be formed at the cisplatin lesion. One should keep in mind that although primer extension studies reveiled a correct incorporation of an A opposite the 3' T of the TT dimer, in the structure itself the lesion was located outside the active center[206]. In the case of the cisplatin lesion, which is also located outside the active center, RNAP II incorporates an A opposite an abasic site, conforming to the A-rule[209, 210]. Even a slight bias for an A in base pairing at a series of consecutive steps could result in a cumulative large preference for As[226], so the incorporation of an A opposite the TT dimer might also follow the A-rule, resulting in a stabalizing T-A base pair. In this case a correct incorporation is achieved and the lesion can be translocated. One could easily test this hypothesis by conducting primer

extension studies with a CC dimer and check whether a G or an A is incorporated opposite the lesion. An incorporation of an A could then be contributed to the A-rule.

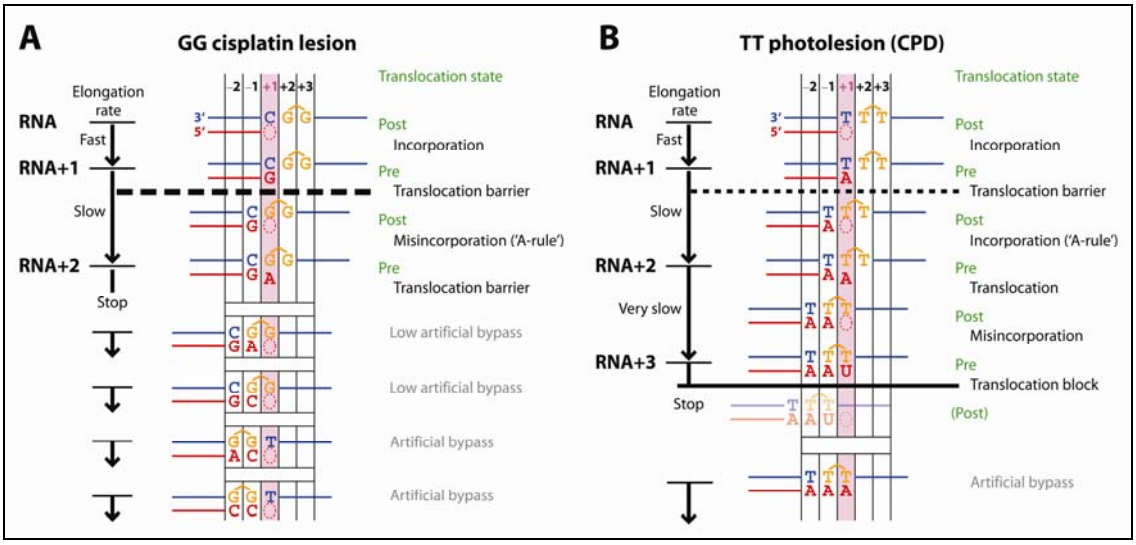


Figure 6-14. A: RNAP II stalling at the cisplatin lesion, shown as a schematic representation of RNA extension in complex A. The initial RNA (top) corresponds to the unextended RNA of scaffold A. Dashed line represents translocation barrier. The artificial conditions leading to lesion bypass are depicted at the bottom. **B:** RNAP II stalling at a CPD lesion[206].

Third, UMP misincorporation occurs opposite the 5' nucleotide of the CPD, whereas correct incorporation occurs opposite the 5' nucleotide of the cisplatin lesion, but only if the lesion is artificially placed beyond the translocation barrier. Fourth, AMP misincorporation at the cisplatin lesion is not required for stalling, but CPD-directed UMP misincorporation is required for stalling. Finally, the misincorporated nucleotide opposite the cisplatin lesion can be bypassed, whereas the misincorporated nucleotide opposite the CPD cannot.

6.7 Future perspectives and outlook

Platinum anticancer agents are widely used in cancer chemotherapy. These platinum complexes appear to kill dividing cells by forming platinum-DNA adducts which interfere with DNA replication, transcription and cell division. These agents form bulky DNA adducts which are thought to exert their cytotoxic effect by blocking DNA replication. Translesion synthesis, one of the pathways of postreplication repair, is thought to account for some resistance to DNA damage and much of the mutagenicity of bulky DNA adducts in dividing cells. The most efficient polymerase identified to date for the bypass of platinum-DNA adducts is Pol η.
Our data may guide chemistry programs directed to block the translesion synthesis of Pol η across Pt-GG lesions for example by using dinuclear Pt-complexes or complexes, which are sterically more demanding. Investigation of such compounds may allow deciphering the biological importance of translesion synthesis for the development of cancer resistances and may even allow overcoming these problems.

Pol η is able to bypass a variety of interstrand crosslinks. Primery results with a site specific 1, 3 oxaliplatin GTG lesion suggest that Pol η can bypass this lesion quite efficiently. Detailed primer extension experiments with Michaelis-Menten kinetics should be undertaken for each of the 3 bases involved in this lesion.

In conclusion, crystallizing Pol η in ternary complex with a TT dimer containing template – primer, the native substrate of Pol η, might shed further light on the translesion synthesis exhibited by this unique polymerase. First diffracting crystals have actually already beeen obtained for the 3'dT elongation step, although further work is needed for obtaining diffracting crystals for the step opposite the 5'dT. Furthermore other lesions such as 8-oxoG or 6-OMeG can be crystallized in ternary complex with Pol η, following the protocol which was established during this work.

7 Appendix

7.1 Abbreviations

Å	Angstrom (1 Å = 10^{-10} m)
A	adenosine
ACN	acetonitrile
BER	Base excision repair
bp	base pairs
C	cytosine
CCD	charge-coupled device
cp	cisplatin
CPD	cyclobutane pyrimidine dimer
CPG	controlled pore size glass
CV	column volume
dATP	2’-deoxyadenosine 5’-triphosphate
dCTP	2’-deoxycytidine 5’-triphosphate
ddATP	2',3'-dideoxyadenosine 5'-triphosphate
ddCTP	2',3'-dideoxycytidine 5'-triphosphate
ddH_2O	double distilled water
ddNTP	2',3'-dideoxynucleoside 5'-triphosphate
dGTP	2’-deoxyguanosine 5’-triphosphate
DCM	dichloromethane
DNA	deoxyribonucleic acid
dNTP	2’-deoxynucleoside 5’-triphosphate
ds	double strand
dTTP	2’-deoxythymidine 5’-triphosphate
G	guanine
g	gram
HPLC	High Performance Liquid Chromatography
I	inosine
kV	kilovolt
m	meter

MALDI-TOF	Matrix Assisted Laser Desorption Ionization- Time of Flight
ml	millilitres
mM	milimolar
mRNA	messenger RNA
ms	miliseconds
NER	Nucleotide excision repair
PCNA	Proliferating Cell Nuclear Antigen
PCR	polymerase chain reaction
Peg 3350	polyethyleneglycol 3350
pol	polymerase
Pol η	DNA polymerase η
RNA	ribonucleic acid
RNAP	RNA polymerase
RPA	replication protein A
rRNA	ribosomal RNA
SDS	sodium dodecyl sulfate
snRNA	small nuclear RNA
ss	single strand
T	thymine
TLS	translesion synthesis
TT dimer	*cis-syn* thymine dimer
UV	ultraviolet
µl	microliter
µM	micromolar
w/v	weight per volume
V	Volt
W	Watt

7.2 Crystallographic tables

Table 7.1. Crystallographic table of Pol η with CP containing DNA and ddCTP

	3'dG elongation with ddCTP
Crystallographic data collection and analysis	
Space group	$P4_12_12$
Cell dimensions	
a, b, c (Å)	a=b=103.5
	c= 292.1
α, β, γ (°)	90, 90, 90
X-ray source	
Wavelength	1.0702
Resolution (Å)	3.3
[a] R_{sym}	13.0
$I / \sigma I$	15.03
Completeness (%)	99.8
Redundancy	7.02
Refinement	
Resolution (Å)	20-3.3
No. reflections	24,769
[b] R_{work} / [c] R_{free}	22.0 / 29.8
No. atoms	
Protein	9066
Ligand/ion	DNA-template-primer
	2 ddCTP
	4 calcium
Water	92

[a] R_{sym} is the unweighted R value on I between symmetry mates

[b] $R_{work} = \Sigma_{hkl} \| F_{obs}(hkl) | - | F_{calc}(hkl) \| / \Sigma_{hkl} | F_{obs}(hkl) |$

[c] R_{free} = the cross validation R factor for 8% of reflections against which the model was not refined

Table 7.2: Crystallographic table of Pol η with CP containing template-$_{dd}$primer DNA

	First crystal *3'dG elongation*	Second crystal *5'dG elongation*
Crystallographic data collection and analysis		
Space group	$P4_12_12$	$P4_12_12$
Cell dimensions		
a, b, c (Å)	a=b=104.1 c= 293.0	a=b=103.7 c= 292.8
α, β, γ (°)	90, 90, 90	90, 90, 90
X-ray source	PX (SLS)	PX (SLS)
Wavelength	0.8726	1.07190
Resolution (Å)	3.1	3.3
R_{sym}[a]	15.4 (44.3)	14.7 (44.4)
$I / \sigma I$	16.98 (5.71)	11.78 (4.48)
Completeness (%)	99.6 (97.9)	98.4 (98.1)
Redundancy	14.2 (13.8)	5.5 (5.5)
Refinement		
Resolution (Å)	20-3.1	25-3.3
No. reflections	30,031	24,793
R_{work}[b] / R_{free}[c]	21.9 / 27.2	24.2 / 28.25
No. atoms		
Protein	8118 DNA- template-dd primer	8118 DNA-template-dd primer
Ligand/ion	2 dCTP 4 calcium	2 dATP 4 calcium
Water	121	-
B-factors		
Protein	33.1	32.3
DNA	46.65	60.03
dNTP	31.16	23.24
Ca	33.6	27.7
water	30.89	-
R.m.s deviations		
Bond lengths (Å)	0.00415	0.00887
Bond angles (°)	1.20265	1.33389

[a] R_{sym} is the unweighted R value on I between symmetry mates

[b] $R_{work} = \Sigma_{hkl} \| F_{obs}(hkl) | - | F_{calc}(hkl) \| / \Sigma_{hkl} | F_{obs}(hkl) |$

[c] R_{free} = the cross validation R factor for 8% of reflections against which the model was not refined

Coordinates and structure factors have been deposited for the 3' and 5'dG complexes in the Protein Data Bank under accession codes 2R8J and 2R8K.

8 References

[1] *Cisplatin: Chemistry and Biochemistry of a Leading Anticancer Drug*, 1 ed., Wiley-VCH, Zurich, **1999**.

[2] A. R. Lehmann, *Mutat. Res.* **2002**, *509*, 23.

[3] E. C. Friedberg, A. R. Lehmann, R. P. Fuchs, *Mol. Cell* **2005**, *18*, 499.

[4] M. T. Washington, L. Prakash, S. Prakash, *Proc. Natl. Acad. Sci. U S A* **2003**, *100*, 12093.

[5] M. T. Washington, R. E. Johnson, S. Prakash, L. Prakash, *Proc. Natl. Acad. Sci. U S A* **2000**, *97*, 3094.

[6] R. E. Johnson, M. T. Washington, S. Prakash, L. Prakash, *J. Biol. Chem* **2000**, *275*, 7447.

[7] C. Masutani, R. Kusumoto, A. Yamada, N. Dohmae, M. Yokoi, M. Yuasa, M. Araki, S. Iwai, K. Takio, F. Hanaoka, *Nature* **1999**, *399*, 700.

[8] R. E. Johnson, S. Prakash, L. Prakash, *Science* **1999**, *283*, 1001.

[9] R. E. Johnson, C. M. Kondratick, S. Prakash, L. Prakash, *Science* **1999**, *285*, 263.

[10] E. R. Jamieson, S. J. Lippard, *Chem. Rev.* **1999**, *99*, 2467.

[11] S. Tornaletti, S. M. Patrick, J. J. Turchi, P. C. Hanawalt, *J. Biol. Chem.* **2003**, *278*, 35791.

[12] A. Tremeau-Bravard, T. Riedl, J. M. Egly, M. E. Dahmus, *J. Biol. Chem.* **2004**, *279*, 7751.

[13] M. R. Albertella, C. M. Green, A. R. Lehmann, M. J. O'Connor, *Cancer Res.* **2005**, *65*, 9799.

[14] Y. W. Chen, J. E. Cleaver, F. Hanaoka, C. F. Chang, K. M. Chou, *Mol. Cancer Res.* **2006**, *4*, 257.

[15] J. D. Watson, F. H. Crick, *Nature* **1953**, *171*, 737.

[16] J. M. Berg, J. L. Tymoczko, L. Stryer, *Biochemistry*, 5 ed., W. H. Freeman, New york, **2002**.

[17] G. Hayashi, M. Hagihara, K. Nakatani, *Nucleic Acids Symp Ser (Oxf)* **2005**, 261.

[18] J. M. Vargason, B. F. Eichman, P. S. Ho, *Nat. Struct. Biol.* **2000**, *7*, 758.

[19] G. Wang, K. M. Vasquez, *Mutat. Res.* **2006**, *598*, 103.

[20] J. F. Allemand, D. Bensimon, R. Lavery, V. Croquette, *Proc. Natl. Acad. Sci. U S A* **1998**, *95*, 14152.

[21] A. H. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel, A. Rich, *Nature* **1979**, *282*, 680.

[22] A. Ghosh, M. Bansal, *Acta Crystallogr. D* **2003**, *59*, 620.

[23] X. J. Lu, Z. Shakked, W. K. Olson, *J. Mol. Biol.* **2000**, *300*, 819.

[24] M. C. Wahl, M. Sundaralingam, *Biopolymers* **1997**, *44*, 45.

[25] E. C. Friedberg, G. C. Walker, W. Siede, *DNA repair and mutagenesis*, ASM Press, Washington D.C., **1995**.

[26] A. M. Cordonnier, R. P. Fuchs, *Mutat. Res.* **1999**, *435*, 111.

[27] K. H. Kraemer, M. M. Lee, A. D. Andrews, W. C. Lambert, *Arch Dermatol* **1994**, *130*, 1018.

[28] R. S. Go, A. A. Adjei, *J. Clin. Oncol.* **1999**, *17*, 409.

[29] B. Rosenberg, L. VanCamp, J. E. Trosko, V. H. Mansour, *Nature* **1969**, *222*, 385.

[30] D. Wang, S. J. Lippard, *Nat. Rev. Drug Discov.* **2005**, *4*, 307.

[31] S. Akiyama, Z. S. Chen, T. Sumizawa, T. Furukawa, *Anticancer Drug Des.* **1999**, *14*, 143.

[32] M. A. Fuertes, C. Alonso, J. M. Perez, *Chem. Rev.* **2003**, *103*, 645.

[33] I. Judson, L. R. Kelland, *Drugs* **2000**, *59 Suppl 4*, 29.

[34] E. Wong, C. M. Giandomenico, *Chem. Rev.* **1999**, *99*, 2451.

[35] N. J. Wheate, J. G. Collins, *Coordination Chemistry Reviews* **2003**, *241*, 133.

[36] H. T. Chifotides, K. R. Dunbar, *Acc. Chem. Res.* **2005**, *38*, 146.

[37] M. J. Clarke, F. Zhu, D. R. Frasca, *Chem. Rev.* **1999**, *99*, 2511.

[38] P. Koepf-Maier, H. Koepf, *Chemical Reviews* **1987**, 1137.

[39] Y. K. Yan, M. Melchart, A. Habtemariam, P. J. Sadler, *Chem. Commun. (Camb)* **2005**, 4764.

[40] Y. Jung, S. J. Lippard, *Chem. Rev.* **2007**, *107*, 1387.

[41] B. Rosenberg, L. Vancamp, T. Krigas, *Nature* **1965**, *205*, 698.

[42] R. B. Weiss, M. C. Christian, *Drugs* **1993**, *46*, 360.

[43] D. Lebwohl, R. Canetta, *Eur. J. Cancer* **1998**, *34*, 1522.

[44] J. Reedijk, *J. Chem. Commun.* **1996**, 801.

[45] B. Spingler, D. A. Whittington, S. J. Lippard, *Inorg. Chem.* **2001**, *40*, 5596.

[46] L. Kelland, *Expert Opin. Investig. Drugs* **2007**, *16*, 1009.

[47] J. Reedijk, *Proc. Natl. Acad. Sci. U S A* **2003**, *100*, 3611.

[48] S. E. Sherman, S. J. Lippard, *Chemical Reviews* **1987**, 1153.

[49] Y. Zhang, Z. Guo, X. Z. You, *J. Am. Chem. Soc.* **2001**, *123*, 9378.

[50] M. Akaboshi, K. Kawai, H. Maki, K. Akuta, Y. Ujeno, T. Miyahara, *Jpn. J. Cancer Res.* **1992**, *83*, 522.

[51] D. V. Deubel, *J. Am. Chem. Soc.* **2002**, *124*, 5834.

[52] D. V. Deubel, *J. Am. Chem. Soc.* **2004**, *126*, 5999.

[53] J. M. Pascoe, J. J. Roberts, *Biochem. Pharmacol.* **1974**, *23*, 1345.

[54] G. Speelmans, R. W. Staffhorst, K. Versluis, J. Reedijk, B. de Kruijff, *Biochemistry* **1997**, *36*, 10545.

[55] H. C. Harder, B. Rosenberg, *Int. J. Cancer* **1970**, *6*, 207.

[56] J. A. Howle, G. R. Gale, *Biochem. Pharmacol.* **1970**, *19*, 2757.

[57] M. H. Baik, R. A. Friesner, S. J. Lippard, *J. Am. Chem. Soc.* **2003**, *125*, 14082.

[58] M. Kartalou, J. M. Essigmann, *Mutat. Res.* **2001**, *478*, 23.

[59] A. Eastman, *Biochemistry* **1983**, *22*, 3927.

[60] A. Eastman, *Cancer Cells* **1990**, *2*, 275.

[61] A. M. Fichtinger-Schepman, J. L. van der Veer, J. H. den Hartog, P. H. Lohman, J. Reedijk, *Biochemistry* **1985**, *24*, 707.

[62] S. E. Sherman, D. Gibson, A. H. J. Wang, S. J. Lippard, *J. Am. Chem. Soc.* **1988**, *110*, 7368.

[63] G. L. Cohen, W. R. Bauer, J. K. Barton, S. J. Lippard, *Science* **1979**, *203*, 1014.

[64] N. Poklar, D. S. Pilch, S. J. Lippard, E. A. Redding, S. U. Dunham, K. J. Breslauer, *Proc. Natl. Acad. Sci. U S A* **1996**, *93*, 7606.

[65] A. Gelasco, S. J. Lippard, *Biochemistry* **1998**, *37*, 9230.

[66] C. J. van Garderen, L. P. van Houte, *Eur. J. Biochem.* **1994**, *225*, 1169.

[67] A. R. Lehmann, *FEBS Lett.* **2005**, *579*, 873.

[68] G. Villani, U. Hubscher, J. L. Butour, *Nucleic Acids Res.* **1988**, *16*, 4407.

[69] J. S. Hoffmann, M. J. Pillaire, G. Maga, V. Podust, U. Hubscher, G. Villani, *Proc. Natl. Acad. Sci. U S A* **1995**, *92*, 5356.

[70] P. M. Burgers, E. V. Koonin, E. Bruford, L. Blanco, K. C. Burtis, M. F. Christman, W. C. Copeland, E. C. Friedberg, F. Hanaoka, D. C. Hinkle, C. W.

Lawrence, M. Nakanishi, H. Ohmori, L. Prakash, S. Prakash, C. A. Reynaud, A. Sugino, T. Todo, Z. Wang, J. C. Weill, R. Woodgate, *J. Biol. Chem.* **2001**, *276*, 43487.

[71] K. Bebenek, T. A. Kunkel, *Cell Mol. Life Sci.* **2002**, *59*, 54.

[72] J. P. McDonald, A. Tissier, E. G. Frank, S. Iwai, F. Hanaoka, R. Woodgate, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2001**, *356*, 53.

[73] E. Ohashi, T. Ogi, R. Kusumoto, S. Iwai, C. Masutani, F. Hanaoka, H. Ohmori, *Genes Dev.* **2000**, *14*, 1589.

[74] G. Maga, G. Villani, K. Ramadan, I. Shevelev, N. Tanguy Le Gac, L. Blanco, G. Blanca, S. Spadari, U. Hubscher, *J. Biol. Chem.* **2002**, *277*, 48434.

[75] A. Vaisman, S. E. Lim, S. M. Patrick, W. C. Copeland, D. C. Hinkle, J. J. Turchi, S. G. Chaney, *Biochemistry* **1999**, *38*, 11026.

[76] A. Vaisman, S. G. Chaney, *J. Biol. Chem.* **2000**, *275*, 13017.

[77] A. Vaisman, C. Masutani, F. Hanaoka, S. G. Chaney, *Biochemistry* **2000**, *39*, 4575.

[78] T. Giraldi, D. M. Taylor, *Biochem Pharmacol* **1974**, *23*, 1659.

[79] R. C. Srivastava, J. Froehlich, G. L. Eichhorn, *Biochemi* **1978**, *60*, 879.

[80] H. Kimura, Y. Tao, R. G. Roeder, P. R. Cook, *Mol Cell Biol* **1999**, *19*, 5383.

[81] T. I. Lee, R. A. Young, *Annu Rev Genet* **2000**, *34*, 77.

[82] Y. Jung, S. J. Lippard, *J Biol Chem* **2003**, *278*, 52084.

[83] Y. Corda, C. Job, M. F. Anin, M. Leng, D. Job, *Biochemistry* **1993**, *32*, 8582.

[84] Y. Corda, C. Job, M. F. Anin, M. Leng, D. Job, *Biochemistry* **1991**, *30*, 222.

[85] Y. Jung, S. J. Lippard, *J Biol Chem* **2006**, *281*, 1361.

[86] S. Tornaletti, D. Reines, P. C. Hanawalt, *J Biol Chem* **1999**, *274*, 24124.

[87] J. P. Laine, J. M. Egly, *Embo J* **2006**, *25*, 387.

[88] C. P. Selby, R. Drapkin, D. Reinberg, A. Sancar, *Nucleic Acids Res* **1997**, *25*, 787.

[89] J. Q. Svejstrup, *Nat Rev Mol Cell Biol* **2002**, *3*, 21.

[90] J. Q. Svejstrup, *J Cell Sci* **2003**, *116*, 447.

[91] J. Cadet, C. Anselmino, T. Douki, L. Voituriez, *J Photochem Photobiol B* **1992**, *15*, 277.

[92] J. Cadet, P. Vigny, *The Photochemistry of Nucleic Acids, Vol. 1*, Wiley, New York, **1990**.

[93] M. G. Friedel, M. K. Cichon, T. Carell, in *CRC Handbook of Organic Photochemistry and Photobiology (2nd Edition)* (Ed.: W. M. Horspool), CRC Press, Boca Raton, **2004**, pp. 141/1.

[94] J. L. Ravanat, T. Douki, J. Cadet, *J. Photochem. Photobiol. B* **2001**, *63*, 88.

[95] H. Park, K. Zhang, Y. Ren, S. Nadji, N. Sinha, J. S. Taylor, C. Kang, *Proc. Natl. Acad. Sci. U S A* **2002**, *99*, 15965.

[96] I. Husain, J. Griffith, A. Sancar, *Proc. Natl. Acad. Sci. U S A* **1988**, *85*, 2558.

[97] J. Kemmink, R. Boelens, T. Koning, G. A. van der Marel, J. H. van Boom, R. Kaptein, *Nucleic Acids Res.* **1987**, *15*, 4645.

[98] G. Ciarrocchi, A. M. Pedrini, *J. Mol. Biol.* **1982**, *155*, 177.

[99] K. M. Lima-Bessa, C. F. Menck, *Curr. Biol.* **2005**, *15*, R58.

[100] S. Doublie, S. Tabor, A. M. Long, C. C. Richardson, T. Ellenberger, *Nature* **1998**, *391*, 251.

[101] J. R. Kiefer, C. Mao, J. C. Braman, L. S. Beese, *Nature* **1998**, *391*, 304.

[102] Y. Li, S. Korolev, G. Waksman, *Embo J.* **1998**, *17*, 7514.

[103] I. K. Cann, Y. Ishino, *Genetics* **1999**, *152*, 1249.

[104] W. Yang, *Curr. Opin. Struct. Biol.* **2003**, *13*, 23.

[105] S. Brenowitz, S. Kwack, M. F. Goodman, M. O'Donnell, H. Echols, *J. Biol. Chem.* **1991**, *266*, 7888.

[106] M. C. Franklin, J. Wang, T. A. Steitz, *Cell* **2001**, *105*, 657.

[107] H. Pelletier, M. R. Sawaya, A. Kumar, S. H. Wilson, J. Kraut, *Science* **1994**, *264*, 1891.

[108] M. R. Sawaya, R. Prasad, S. H. Wilson, J. Kraut, H. Pelletier, *Biochemistry* **1997**, *36*, 11205.

[109] J. Trincao, R. E. Johnson, W. T. Wolfle, C. R. Escalante, S. Prakash, L. Prakash, A. K. Aggarwal, *Nat. Struct. Mol. Biol.* **2004**, *11*, 457.

[110] H. Echols, M. F. Goodman, *Annu. Rev. Biochem.* **1991**, *60*, 477.

[111] M. F. Goodman, *Proc. Natl. Acad. Sci. U S A* **1997**, *94*, 10493.

[112] E. T. Kool, *Annu. Rev. Biochem.* **2002**, *71*, 191.

[113] L. Barbour, W. Xiao, *Mutat. Res.* **2003**, *532*, 137.

[114] H. Ohmori, E. C. Friedberg, R. P. Fuchs, M. F. Goodman, F. Hanaoka, D. Hinkle, T. A. Kunkel, C. W. Lawrence, Z. Livneh, T. Nohmi, L. Prakash, S. Prakash, T. Todo, G. C. Walker, Z. Wang, R. Woodgate, *Mol. Cell* **2001**, *8*, 7.

[115] W. Yang, R. Woodgate, *Proc. Natl. Acad. Sci. U S A* **2007**, *104*, 15591.

[116] J. R. Nelson, C. W. Lawrence, D. C. Hinkle, *Science* **1996**, *272*, 1646.

[117] S. Prakash, R. E. Johnson, L. Prakash, *Annual Review of Biochemistry* **2005**, *74*, 317.

[118] J. P. McDonald, A. S. Levine, R. Woodgate, *Genetics* **1997**, *147*, 1557.

[119] A. R. Lehman, S. Kirk-Bell, C. F. Arlett, M. C. Paterson, P. H. Lohman, E. A. de Weerd-Kastelein, D. Bootsma, *Proc. Natl. Acad. Sci. U S A* **1975**, *72*, 219.

[120] M. Cordeiro-Stone, L. S. Zaritskaya, L. K. Price, W. K. Kaufmann, *J. Biol. Chem.* **1997**, *272*, 13945.

[121] N. Nikolaishvili-Feinberg, M. Cordeiro-Stone, *Biochemistry* **2001**, *40*, 15215.

[122] Y. C. Wang, V. M. Maher, D. L. Mitchell, J. J. McCormick, *Mol. Cell Biol.* **1993**, *13*, 4276.

[123] H. L. Waters, S. Seetharam, M. M. Seidman, K. H. Kraemer, *J. Invest. Dermatol.* **1993**, *101*, 744.

[124] M. T. Washington, R. E. Johnson, S. Prakash, L. Prakash, *J. Biol. Chem.* **1999**, *274*, 36835.

[125] T. Matsuda, K. Bebenek, C. Masutani, F. Hanaoka, T. A. Kunkel, *Nature* **2000**, *404*, 1011.

[126] M. T. Washington, R. E. Johnson, L. Prakash, S. Prakash, *Proc. Natl. Acad. Sci. U S A* **2001**, *98*, 8355.

[127] M. T. Washington, R. E. Johnson, S. Prakash, L. Prakash, *J. Biol. Chem.* **2001**, *276*, 2263.

[128] K. Bebenek, T. Matsuda, C. Masutani, F. Hanaoka, T. A. Kunkel, *J. Biol. Chem.* **2001**, *276*, 2317.

[129] J. Trincao, R. E. Johnson, C. R. Escalante, S. Prakash, L. Prakash, A. K. Aggarwal, *Mol. Cell* **2001**, *8*, 417.

[130] Y. Li, S. Dutta, S. Doublie, H. M. Bdour, J. S. Taylor, T. Ellenberger, *Nat. Struct. Mol. Biol.* **2004**, *11*, 784.

[131] C. A. Smith, J. Baeten, J. S. Taylor, *J. Biol. Chem.* **1998**, *273*, 21933.

[132] L. Sun, M. Wang, E. T. Kool, J. S. Taylor, *Biochemistry* **2000**, *39*, 14603.

[133] L. Sun, K. Zhang, L. Zhou, P. Hohler, E. T. Kool, F. Yuan, Z. Wang, J. S. Taylor, *Biochemistry* **2003**, *42*, 9431.

[134] J. D. Armstrong, B. A. Kunz, *Proc. Natl. Acad. Sci. U S A* **1990**, *87*, 9005.

[135] D. E. Brash, *Trends Genet.* **1997**, *13*, 410.

[136] K. A. Canella, M. M. Seidman, *Mutat. Res.* **2000**, *450*, 61.

[137] A. Stary, P. Kannouche, A. R. Lehmann, A. Sarasin, *J. Biol. Chem.* **2003**, *278*, 18767.

[138] S. L. Yu, R. E. Johnson, S. Prakash, L. Prakash, *Mol. Cell Biol.* **2001**, *21*, 185.

[139] F. Bourre, G. Renault, A. Sarasin, *Nucleic Acids Res.* **1987**, *15*, 8861.

[140] D. E. Brash, W. A. Haseltine, *Nature* **1982**, *298*, 189.

[141] D. E. Brash, S. Seetharam, K. H. Kraemer, M. M. Seidman, A. Bredberg, *Proc. Natl. Acad. Sci. U S A* **1987**, *84*, 3782.

[142] D. L. Mitchell, C. A. Haipek, J. M. Clarkson, *Mutat. Res.* **1985**, *143*, 109.

[143] S. G. Chaney, S. L. Campbell, B. Temple, E. Bassett, Y. Wu, M. Faldu, *J. Inorg. Biochem.* **2004**, *98*, 1551.

[144] G. R. Gibbons, W. K. Kaufmann, S. G. Chaney, *Carcinogenesis* **1991**, *12*, 2253.

[145] E. L. Mamenta, E. E. Poma, W. K. Kaufmann, D. A. Delmastro, H. L. Grady, S. G. Chaney, *Cancer Res.* **1994**, *54*, 3500.

[146] A. Vaisman, M. Varchenko, A. Umar, T. A. Kunkel, J. I. Risinger, J. C. Barrett, T. C. Hamilton, S. G. Chaney, *Cancer Res.* **1998**, *58*, 3579.

[147] M. R. Albertella, A. Lau, M. J. O'Connor, *DNA Repair* **2005**, *4*, 583.

[148] T. Okuda, X. Lin, J. Trang, S. B. Howell, *Mol. Pharmacol.* **2005**, *67*, 1852.

[149] F. Wu, X. Lin, T. Okuda, S. B. Howell, *Cancer Res.* **2004**, *64*, 8029.

[150] X. Lin, J. Trang, T. Okuda, S. B. Howell, *Clin. Cancer Res.* **2006**, *12*, 563.

[151] J. K. Kim, B. S. Choi, *Eur. J. Biochem.* **1995**, *228*, 849.

[152] J. K. Kim, D. Patel, B. S. Choi, *Photochem. Photobiol.* **1995**, *62*, 44.

[153] J. H. Lee, G. S. Hwang, B. S. Choi, *Proc. Natl. Acad. Sci. U S A* **1999**, *96*, 6632.

[154] R. E. Johnson, L. Haracska, S. Prakash, L. Prakash, *Mol. Cell Biol.* **2001**, *21*, 3558.

[155] A. Bresson, R. P. Fuchs, *Embo J.* **2002**, *21*, 3881.

[156] S. Prakash, L. Prakash, *Genes Dev.* **2002**, *16*, 1872.

[157] T. Douki, D. Perdiz, P. Grof, Z. Kuluncsics, E. Moustacchi, J. Cadet, E. Sage, *Photochem. Photobiol.* **1999**, *70*, 184.

[158] L. Haracska, S. L. Yu, R. E. Johnson, L. Prakash, S. Prakash, *Nat. Genet.* **2000**, *25*, 458.

[159] G. W. Hsu, M. Ober, T. Carell, L. S. Beese, *Nature* **2004**, *431*, 217.

[160] L. Haracska, S. Prakash, L. Prakash, *Mol. Cell Biol.* **2000**, *20*, 8001.

[161] L. Haracska, M. T. Washington, S. Prakash, L. Prakash, *J. Biol. Chem.* **2001**, *276*, 6861.

[162] D. Chiapperino, H. Kroth, I. H. Kramarczuk, J. M. Sayer, C. Masutani, F. Hanaoka, D. M. Jerina, A. M. Cheh, *J. Biol. Chem.* **2002**, *277*, 11765.

[163] I. G. Minko, M. T. Washington, L. Prakash, S. Prakash, R. S. Lloyd, *J. Biol. Chem.* **2001**, *276*, 2517.

[164] I. G. Minko, M. T. Washington, M. Kanuri, L. Prakash, S. Prakash, R. S. Lloyd, *J. Biol. Chem.* **2003**, *278*, 784.

[165] R. L. Levine, H. Miller, A. Grollman, E. Ohashi, H. Ohmori, C. Masutani, F. Hanaoka, M. Moriya, *J. Biol. Chem.* **2001**, *276*, 18717.

[166] M. T. Washington, S. A. Helquist, E. T. Kool, L. Prakash, S. Prakash, *Mol. Cell Biol.* **2003**, *23*, 5107.

[167] M. T. Washington, W. T. Wolfle, T. E. Spratt, L. Prakash, S. Prakash, *Proc. Natl. Acad. Sci. U S A* **2003**, *100*, 5113.

[168] R. G. Roeder, W. J. Rutter, *Nature* **1969**, *224*, 234.

[169] R. G. Roeder, W. J. Rutter, *Biochemistry* **1970**, *9*, 2543.

[170] R. G. Roeder, W. J. Rutter, *Proc Natl Acad Sci U S A* **1970**, *65*, 675.

[171] A. B. Lassar, P. L. Martin, R. G. Roeder, *Science* **1983**, *222*, 740.

[172] P. R. Cook, *Science* **1999**, *284*, 1790.

[173] F. J. Iborra, A. Pombo, D. A. Jackson, P. R. Cook, *J Cell Sci* **1996**, *109 (Pt 6)*, 1427.

[174] L. A. Allison, M. Moyle, M. Shales, C. J. Ingles, *Cell* **1985**, *42*, 599.

[175] J. M. Cho, R. K. Carlin, J. E. Evans, A. P. Kimball, *Biochem Pharmacol* **1982**, *31*, 2583.

[176] S. B. Carroll, B. D. Stollar, *J Mol Biol* **1983**, *170*, 777.

[177] M. Riva, A. R. Schaffner, A. Sentenac, G. R. Hartmann, A. A. Mustaev, E. F. Zaychikov, M. A. Grachev, *J Biol Chem* **1987**, *262*, 14377.

[178] A. L. Gnatt, P. Cramer, J. Fu, D. A. Bushnell, R. D. Kornberg, *Science* **2001**, *292*, 1876.

[179] P. Cramer, D. A. Bushnell, R. D. Kornberg, *Science* **2001**, *292*, 1863.

[180] A. McPherson, in *Preparation and Analysis of Protein Crystals*, John Wiley & Sons, New York, **1982**, pp. 94.

[181] S. R. Jordan, T. V. Whitcombe, J. M. Berg, C. O. Pabo, *Science* **1985**, *230*, 1383.

[182] P. A. Rice, S. Yang, K. Mizuuchi, H. A. Nash, *Cell* **1996**, *87*, 1295.

[183] H. Ling, F. Boudsocq, B. S. Plosky, R. Woodgate, W. Yang, *Nature* **2003**, *424*, 1083.

[184] Y. Li, G. Waksman, *Protein Sci.* **2001**, *10*, 1225.

[185] H. Ling, F. Boudsocq, R. Woodgate, W. Yang, *Mol. Cell* **2004**, *13*, 751.

[186] H. Huang, S. C. Harrison, G. L. Verdine, *Chem. Biol.* **2000**, *7*, 355.

[187] P. Ewald, *Z. Kristallogr.* **1921**, *56*, 129.

[188] M. Ober, H. Muller, C. Pieck, J. Gierlich, T. Carell, *J. Am. Chem. Soc.* **2005**, *127*, 18143.

[189] J. C. Pieck, LMU München (Munich), **2007**.

[190] Y. Mantri, S. J. Lippard, M. H. Baik, *J Am Chem Soc* **2007**, *129*, 5023.

[191] W. Kabsch, *J. Appl. Cryst.* **1993**, *26*, 795.

[192] A. J. McCoy, R. W. Grosse-Kunstleve, L. C. Storoni, R. J. Read, *Acta Crystallogr. D* **2005**, *61*, 458.

[193] U. Ohndorf, M. Rould, Q. He, C. Pabo, S. Lippard, *Nature* **1999**, *399*, 708.

[194] G. J. Kleywegt, T. A. Jones, *Acta Crystallogr. D* **1998**, *54*, 1119.

[195] D. Turk, thesis thesis, Technical University (Munich), **1992**.

[196] P. Emsley, K. Cowtan, *Acta Crystallogr. D* **2004**, *60*, 2126.

[197] A. T. Brunger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, G. L. Warren, *Acta Crystallogr. D* **1998**, *54*, 905.

[198] R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Cryst.* **1993**, *26*, 283.

[199] G. J. Kleywegt, *Acta Crystallogr. D* **1996**, *52*, 842.

[200] W. L. DeLano, DeLano Scientific LLC **2002**.

[201] A. M. DeLucia, N. D. Grindley, C. M. Joyce, *Nucleic Acids Res.* **2003**, *31*, 4129.

[202] E. A. Meyer, R. K. Castellano, F. Diederich, *Angew. Chem. Int. Ed.* **2003**, *42*, 1210.

[203] F. M. Richards, C. E. Kundrot, *Proteins* **1988**, *3*, 71.

[204] T. S. Krishna, X. P. Kong, S. Gary, P. M. Burgers, J. Kuriyan, *Cell* **1994**, *79*, 1233.

[205] L. Haracska, C. M. Kondratick, I. Unk, S. Prakash, L. Prakash, *Mol. Cell* **2001**, *8*, 407.

[206] F. Brueckner, U. Hennecke, T. Carell, P. Cramer, *Science* **2007**, *315*, 859.

[207] H. Kettenberger, K. J. Armache, P. Cramer, *Mol Cell* **2004**, *16*, 955.

[208] P. M. Takahara, A. C. Rosenzweig, C. A. Frederick, S. J. Lippard, *Nature* **1995**, *377*, 649.

[209] B. S. Strauss, *Bioessays* **1991**, *13*, 79.

[210] J. S. Taylor, *Mutat Res* **2002**, *510*, 55.

[211] H. Ling, F. Boudsocq, R. Woodgate, W. Yang, *Cell* **2001**, *107*, 91.

[212] D. T. Nair, R. E. Johnson, S. Prakash, L. Prakash, A. K. Aggarwal, *Nature* **2004**, *430*, 377.

[213] S. N. Uljon, R. E. Johnson, T. A. Edwards, S. Prakash, L. Prakash, A. K. Aggarwal, *Structure* **2004**, *12*, 1395.

[214] M. T. Washington, L. Prakash, S. Prakash, *Cell* **2001**, *107*, 917.

[215] J. C. Morales, E. T. Kool, *Nat. Struct. Biol.* **1998**, *5*, 950.

[216] S. Moran, R. X. Ren, E. T. Kool, *Proc. Natl. Acad. Sci. U S A* **1997**, *94*, 10506.

[217] K. A. Ramirez-Aguilar, C. L. Moore, R. D. Kuchta, *Biochemistry* **2005**, *44*, 15585.

[218] C. L. Moore, A. Zivkovic, J. W. Engels, R. D. Kuchta, *Biochemistry* **2004**, *43*, 12367.

[219] S. D. McCulloch, R. J. Kokoska, O. Chilkova, C. M. Welch, E. Johansson, P. M. Burgers, T. A. Kunkel, *Nucleic Acids Res.* **2004**, *32*, 4665.

[220] R. Kusumoto, C. Masutani, S. Shimmyo, S. Iwai, F. Hanaoka, *Genes Cells* **2004**, *9*, 1139.

[221] R. E. Johnson, L. Prakash, S. Prakash, *Proc. Natl. Acad. Sci. U S A* **2005**, *102*, 12359.

[222] W. T. Wolfle, M. T. Washington, L. Prakash, S. Prakash, *Genes Dev.* **2003**, *17*, 2191.

[223] F. Boudsocq, S. Iwai, F. Hanaoka, R. Woodgate, *Nucleic Acids Res.* **2001**, *29*, 4607.

[224] E. Kashkina, M. Anikin, F. Brueckner, R. T. Pomerantz, W. T. McAllister, P. Cramer, D. Temiakov, *Mol Cell* **2006**, *24*, 257.

[225] G. E. Damsma, A. Alt, F. Brueckner, T. Carell, P. Cramer, *Nat Struct Mol Biol* **2007**, *14*, 1127.

[226] J. S. Taylor, C. L. O'Day, *Biochemistry* **1990**, *29*, 1624.

9 Acknowledgements

I would like to thank all the people that supported me on my way and made this work possible.

I am grateful to Prof. Dr. Thomas Carell for allowing me to work on this highly interesting and challenging project, which at certain points did not look to be promising. I am also very grateful to Prof. Dr. Karl-Peter Hopfner for allowing me free access to his laboratory facilities and for fruitful discussions. I would also like to thank Prof. Dr. Patrick Cramer from the Gene Center, LMU and Prof. Stephen J. Benkovic from the Department of Chemistry at the Pennsylvania State University for fruitful collaborations. Sabiene Voß and Claudia Gräf I would like to thank for technical support and Mrs. Slava Gärtner for getting me through german bureaucracy whenever needed.

Special thanks go out to Markus Mueller, Stefanie Schorr, Katja Lammens, Alfred Lammens, Maria Angeles Izquierdo Arcusa, Gerke Damsma, Florian Brueckner and Sabiene Schneider for helpful discussions and corrections of this work presented.

Katja Lammens and Alfred Lammens for working together, teaching me crystallography and a fruitful collaboration not only at the lab, but also in the kitchen (is there a difference?!?). Gerke Damsma and Florian Brueckner from the Cramer group and Lisa Sproul from the Benkovic lab for fruitful collaborations on cisplatin damaged DNA.

Antonio Manetto for teaching me lots about Sicilian food, Italy and geography. And then there is that thingy that virtually all things originate from Italy. And if they are already associated with Italy, than they are probably originally from Sicily. Thank you Sebastian Bareyt for suffering these blabs together…

Heiko Mueller, Eva Jahn and Guido Clever for a great athmosphere at the lab and for suffering my musical taste without protesting too much. Carsten J. Pieck and Claudia Chioccini for a nice collaboration at the lab, and for teaching me some biochemical methods. Florian Klepper for organizing teams for soccer competitions and Johannes Gierlich for help on all high tech equippment. Phillipp Gramlich who is always helpful and a Litzka kind of guy. I want to thank Simon Warncke and Sascha “Paule” Breeger for really nice conversations, which usually were not related to work. More thanks go

out to all my labmates from the Carell lab for fun times after work, which usually included more than one beer.
The “Happy Hopfners” Christian C. Schiller, Diana Pipping, Gregor Witte, Katharina Buettner, Sophia Hartung, Wolfgang Reindl and all others affiliates for common fun also outside the lab and all over northern Europe, which did on the contrary not always include beer.
All the guys from CLUSTOXDNA network and last not least Mr. B from Mr. B’s who is running probably the best Jazz bar in town. Go and check it out!

My apologies to all others who I have not mentioned by name – thank you for your support!

10 Curriculum Vitae

NAME AND ADDRESS

Name: Aaron Alt

Date and Place of Birth: 17/12/1973, Wiener Neustadt, Austria

Nationality: Israel, Austria

ACADEMIC DEGREES

2005 – 2008: **Ph.D.** in Chemistry, Ludwig Maximilian University, Munich, Germany. Supervisor: Prof. Dr. Thomas Carell. Research Subject: Recognition, repair and translesion synthesis of bulky DNA lesions at molecular level.

2003 – 2005: **M.Sc**. in Chemistry "Cum Laude", Technion, Israel. Supervisor: Prof. Ehud Keinan. Research Subject: “Ozone scavengers as novel drug candidates for the treatment of asthma and other inflammatory diseases”.

2000 – 2003: **B.A.** in Chemistry "Cum Laude", Technion, Israel

HONORS

2003: Dean’s award for excellence during undergraduate studies.

LANGUAGES

Hebrew (fluent), German (fluent), English (fluent), Spanish (intermediate), Japanese (beginner)